Mechanisms of transcriptional stalling and mutagenesis at DNA lesions

Gerke E. Damsma
aus Hellendoorn, Niederlande
2009

Erklärung

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

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig und ohne unerlaubte Hilfe erarbeitet.

München, am 26. November 2009

Gerke Damsma

Dissertation eingereicht am 26. November 2009
1. Gutachter: Prof. Dr. Patrick Cramer
2. Gutachter: Prof. Dr. Dietmar Martin
Mündliche Prüfung am 27. Januar 2010

Acknowledgements

First and most of all, I want to thank Patrick for giving me the opportunity to work in his lab and for his continuous personal support. His excitement about science and his encouraging and respectful attitude towards his co-workers create an extremely pleasant and motivating atmosphere in the lab. This was crucial for keeping my motivation high and turning my efforts into success.

I am thankful to all present and former members of the Cramer lab for their highly collaborative attitude and all their help, for inspiring scientific discussions and for the great times together. I particularly want to thank Florian Brückner for his work on Pol II nucleic acid complexes in this lab, which formed the basis of my projects.

Very special thanks to my other Pol II co-workers for great team work and sharing many ideas. I thank Alan for all his help on crystallographic matters and for proofreading this thesis. I thank Elisabeth for great discussions and efficient Pol II purifications. I thank Jasmin for introducing me to the world of bead-assays and for her high motivation.

Special thanks to Stefan Benkert for his technical support in producing huge amounts of yeast, Claudia Buchen and Kristin Leike for helping with many of the everyday problems in the lab.

Thanks to Dietmar Martin, Heinrich Leonhardt, Karl-Peter Hopfner, Klaus Förstemann and Roland Beckmann for being my PhD examiners.

I am grateful for having a great and supporting little family that made it possible for me to do my work. I want to thank Tjaard for being my son; it was great to write my thesis together with you. I want to thank Hans for all his love and support. I am grateful to my parents, for their continuous support during all my life and for seriously trying to understand what I am doing.

Summary

RNA polymerase II (Pol II) is the eukaryotic enzyme responsible for transcribing protein-coding genes into messenger RNA (mRNA). This thesis describes the study on the molecular mechanisms of Pol II interacting with DNA damages. The two damages investigated are 1,2-d(GpG) DNA intrastrand cross-links (cisplatin lesions), induced by the anticancer drug cisplatin and 8-oxoguanine (8oxoG), the most encountered DNA lesion resulting from oxidative stress.

We performed a structure-function analysis of Pol II stalling at a cisplatin lesion in the DNA template. Pol II 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. Pol 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.

In addition, we combined structural and functional data to derive the molecular mechanism of Pol II transcription over 8oxoG. When Pol II encounters 8oxoG in the DNA template strand, it correctly incorporates cytosine in most instances, but it also misincorporates adenine. The misincorporated adenine forms a Hoogsteen base pair with 8oxoG at the active center. This misincorporation requires rotation of the 8oxoG base from the standard *anti-* to an uncommon *syn*-conformation, which likely occurs during 8oxoG loading into the active site at a lower rate. X-ray analysis shows that the misincorporated adenine forms a Hoogsteen base pair with 8oxoG in the polymerase active center. Mass spectrometric analysis of RNA extension products shows that the misincorporated adenine escapes the intrinsic proofreading function of Pol II, and remains in the RNA product after polymerase bypass, resulting in transcriptional mutagenesis. Mutagenesis is suppressed by the transcript cleavage-stimulatory factor TFIIS, which is essential for cell survival during oxidative stress.

Previously, the mechanism of Pol II stalling at a cyclobutane pyrimidine dimer photolesion was investigated by our group. In this thesis we show that the stalling mechanism at a cisplatin lesion differs from that of Pol II stalling at a photolesion, which allows delivery of the lesion to the active site but blocks transcription by lesion-templated misincorporation. In case of 8oxoG, no stalling occurs at all, which leads to transcriptional mutagenesis. Together, these results lead to the conclusion that it is impossible to predict the mechanisms of transcriptional stalling or mutagenesis at other types of lesions.

Publications

Parts of this work have been published or are in the process of publication:

Hirtreiter, A., Damsma, G.E., Cheung, A., Klose, D., Grohmann, D., Vojnic, E., Martin, A.C.R., Cramer, P., Werner, F. (2010) Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled coil motif. *Submitted*

Damsma, G.E., Cramer, P. (2009) Molecular basis of transcriptional mutagenesis at 8-oxoguanine. *J. Biol. Chem.* **284**(46): 31658-31663.

Sydow, J. F., Brueckner, F., Cheung, A. C., Damsma, G. E., Dengl, S., Lehmann, E., Vassylyev, D., Cramer, P. (2009). Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol Cell* **34**(6):710-21.

Brueckner, F., Armache, K. J., Cheung, A., Damsma, G. E., Kettenberger, H., Lehmann, E., Sydow, J. F., Cramer, P. (2009). Structure-function studies of the RNA polymerase II elongation complex. *Acta Crystallogr D Biol Crystallogr.* **65**, 112-120.

Cramer, P., Armache, K.-J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C.-D., Lehmann, E., Leike, K., Sydow, J. and Vannini, A. (2008). Structure of Eukaryotic RNA Polymerases. *Annu. Rev. Biophys.* **37***,* 337-352.

Damsma, G.E., Alt, A., Brueckner, F., Carell, T., Cramer, P. (2007). Mechanism of transcriptional stalling at cisplatin-damaged DNA. *Nat Struct Mol Biol* **14**, 1127-33.

Table of Contents

1 Introduction

1.1 Eukaryotic mRNA transcription

The process of DNA transcription into messenger RNA (mRNA) is catalyzed by DNA dependent RNA polymerases, and specifically in eukaryotes by RNA polymerase II (Pol II). The mRNA transcription cycle consists of three stages: initiation, elongation and termination. Initiation involves binding Pol II to the promoter, local melting, and forming the first few phosphodiester bonds. The initiation phase is subject to the most regulation. To allow initiation, appropriate modification of chromatin at the promoter region is essential (Li et al., 2007). Pol II has then to be recruited to the promoter. In eukaryotes, the core promoter is the basis for the assembly of the transcription preinitiation complex (PIC). Additionally, regulatory factors, namely activators and repressors, bind to enhancer and silencer elements on the DNA respectively, to allow transmission of regulatory signals via the coactivators. The PIC comprises the general transcription factors (GTFs) TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, and Pol II (Thomas and Chiang, 2006). These factors function together to initiate transcription at the transcription start site (Figure 1.1). PIC formation begins with the binding of transcription factor TFIID to the TATA box via the TATA-binding protein (TBP) subunit, to the initiator (Inr) and/or to the downstream promoter element (DPE). The entry of other general transcription factors follows by one of two possible pathways; either a sequential assembly pathway or a preassembled RNA polymerase II holoenzyme pathway. The promoter-bound complex is sufficient for a basal level of transcription. However, general cofactors are required to transmit regulatory signals between gene-specific activators and the general transcription machinery in the case of regulated, activator-dependent transcription (Thomas and Chiang, 2006). There are three classes of general cofactors: the TBP-associated factors (TAFs), the Mediator, and the upstream stimulatory activity (USA)-derived positive cofactors and negative cofactor 1 (Figure 1.1). Promoter activity in a gene-specific or cell-type-specific manner is usually adjusted by the independent or combined function of the general cofactors. After PIC formation, TFIIH phosphorylates serines 2 and 5 in the CTD of Pol II and this process is stimulated by TFIIE. During the shift from initiation to elongation, phosphorylation on serine 5 of the CTD is lost (Weaver, 2008).
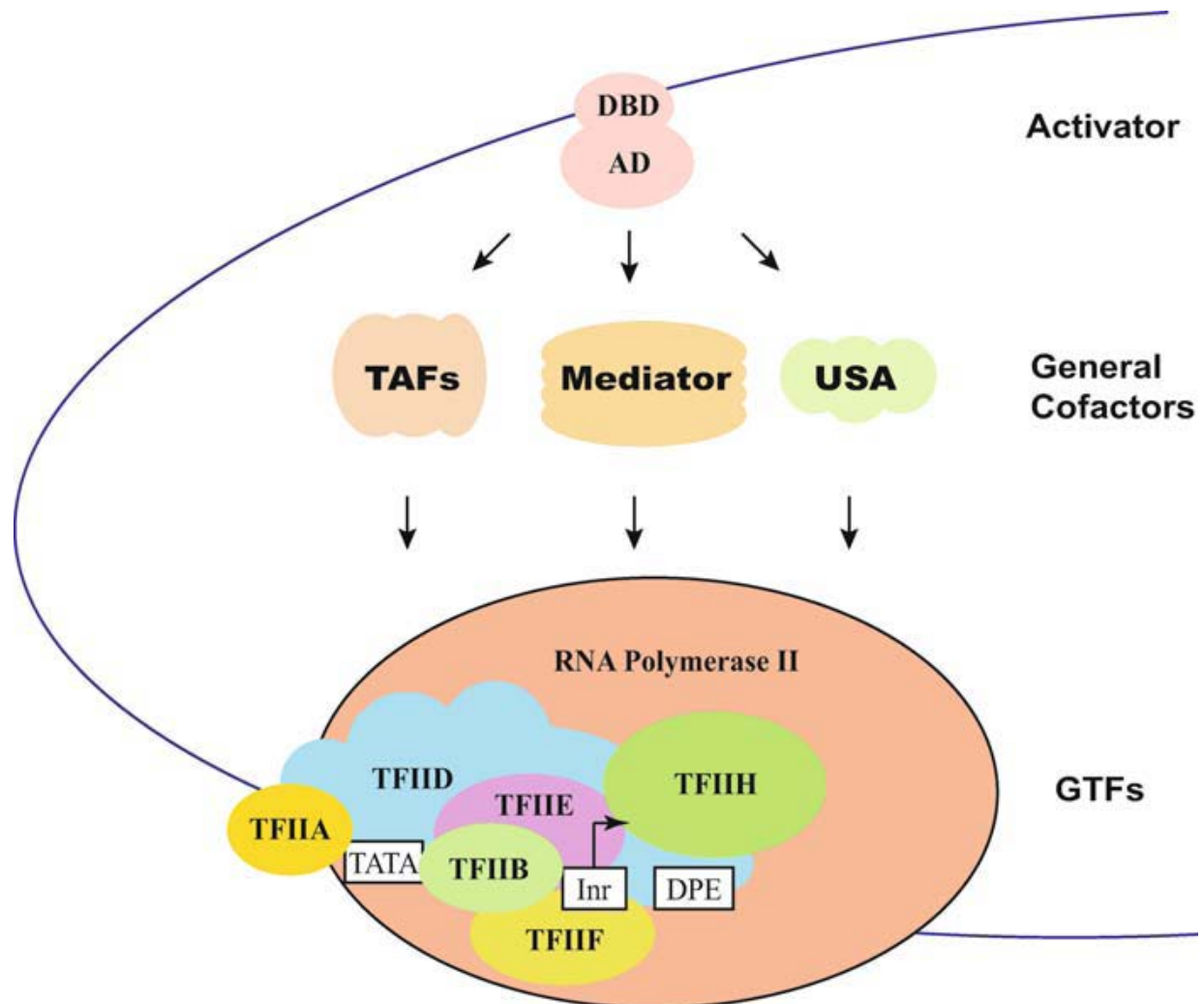
Figure 1.1 The eukaryotic transcription machinery.

General cofactors serve as molecular bridges in activator-dependent transcription. General cofactors (TAFs, Mediator, and USA) are required for transducing signals between gene-specific activators and components of the general transcription machinery. An activator normally contains a DNA-binding domain (DBD) contacting specific DNA sequences and an activation domain (AD) interacting with general cofactors or with components of the general transcription machinery. It should be noted that TAFs normally function as an integral part of TFIID, not as a free entity in mammalian cells as drawn here. Adopted from (Thomas and Chiang, 2006).

During elongation, Pol II polymerises ribonucleotides in the 5'→3' direction to synthesize the remaining RNA. Transcription can also be controlled at the elongation level. TFIIS stimulates elongation by limiting arrest at discrete sites that produce a backtracked RNA that is extruded past the active site. It does this by inserting a hairpin loop into the active site of Pol II and stimulating an RNAse-activity that cleaves off the extruded 3' end of the nascent RNA, which is causing transcription arrest. TFIIF also stimulates elongation, by limiting transient pausing (Weaver, 2008).

Finally in termination, Pol II and the RNA product dissociate from the DNA template. An intact polyadenylation site and active factors that cleave at the polyadenylation site are required for transcription termination. Cleavage at the poly(A) site provides an entry site for the 5'→3' exonuclease Rat1, which degrades the RNA until it catches the polymerase and terminates transcription (Weaver, 2008).

1.2   Eukaryotic DNA-dependent RNA polymerases

Gene transcription in eukaryotic cells is carried out by the three different DNA dependent RNA polymerases Pol I, Pol II, and Pol III. Pol I produces ribosomal RNA, Pol II synthesizes messenger RNAs and small nuclear RNAs, and Pol III produces transfer RNAs and other small RNAs. A fourth RNA polymerase, Pol IV, which was recently discovered in plants, is not included here, as its composition and structure are currently unknown. The RNA polymerases are multisubunit enzymes. Pol I, II, and III comprise 14, 12, and 17 subunits, and have a total molecular weight of 589, 514, and 693 kDa, respectively (Table 1.1). Ten subunits form a structurally conserved core, and additional subunits are located on the periphery.

Table 1.1 RNA polymerase subunits.

RNA polymerase	Pol I	Pol II	Pol III
Ten-subunit core	A190	Rpb1	C160
	A135	Rpb2	C128
	AC40	Rpb3	AC40
	AC19	Rpb11	AC19
	A12.2	Rpb9	C11
	Rpb5 (ABC27)	Rpb5	Rpb5
	Rpb6 (ABC23)	Rpb6	Rpb6
	Rpb8 (ABC14.5)	Rpb8	Rpb8
	Rpb10 (ABC10α)	Rpb10	Rpb10
	Rpb12 (ABC10β)	Rpb12	Rpb12
Rpb4/7 subcomplex	A14	Rpb4	C17
	A43	Rpb7	C25
TFIIF-like subcomplex[a]	A49	(Tfg1/Rap74)	C37
	A34.5	(Tfg2/Rap30)	C53
Pol III-specific subcomplex	–	–	C82
	–	–	C34
	–	–	C31
Number of subunits	14	12	17

[a]The two subunits in Pol I and Pol III are predicted to form heterodimers that resemble part of the Pol II initiation/elongation factor TFIIF, which is composed of subunits Tfg1, Tfg2, and Tfg3 in *Saccharomyces cerevisiae*, and of subunits Rap74 and Rap30 in human.

1.3   Structure of RNA polymerase II

Pol II consists of a 10-subunit core enzyme and a peripheral heterodimer of subunits Rpb4 and Rpb7 (Rpb4/7 subcomplex, Table 1.1). The core enzyme comprises subunits Rpb1, Rpb2, Rpb3, and Rpb11, which contain regions of sequence and structural similarity in Pol I, Pol III, bacterial RNA polymerases (Vassylyev et al., 2002; Zhang et al., 1999), and the archaeal RNA polymerase (Kusser et al., 2008). The Pol II core also comprises subunits Rpb5, Rpb6, Rpb8, Rpb10, and Rpb12, which are shared between Pol I, II, and III (common subunits, Table 1.1). Counterparts of these common subunits, except Rpb8, exist in the archaeal polymerase. But only a counterpart of Rpb6 exists in the bacterial enzyme (Minakhin et al., 2001). Finally, homologues of the core subunit Rpb9 exist in Pol I and Pol III, but not in the archaeal or bacterial enzyme.

Initial electron microscopic studies of Pol II revealed the overall shape of the enzyme (Darst et al., 1991). The core Pol II could subsequently be crystallized, leading to an electron density map at 6 Å resolution (Fu et al., 1999). Crystal improvement by controlled shrinkage and phasing at 3 Å resolution resulted in a backbone model of the Pol II core (Cramer et al., 2000). This revealed that Rpb1 and Rpb2 form opposite sides of a positively charged active center cleft, whereas the smaller subunits are arrayed around the periphery. Refined atomic structures of the core Pol II were obtained in two different conformations and revealed domain-like regions within the subunits, as well as surface elements predicted to have functional roles (Cramer et al., 2001) (Figure 1.2 and Figure 1.3). The active site and the bridge helix, which spans the cleft, line a pore in the floor of the cleft. The Rpb1 side of the cleft forms a mobile clamp, which was trapped in two different open states in the free core structures (Cramer et al., 2001) but was closed in the structure of a core complex that included DNA and RNA (Gnatt et al., 2001). The mobile clamp is connected to the body of the polymerase by five switch regions that show conformational variability. The Rpb2 side of the cleft consists of the lobe and protrusion domains. Rpb2 also forms a protein wall that blocks the end of the cleft.

The Pol II core structures lacked subunits Rpb4 and Rpb7, which can dissociate from the yeast enzyme (Edwards et al., 1991). A structure of the archaeal homologue of the Rpb4/7 heterodimer showed that Rpb7 contains an N-terminal domain, later called the tip domain, and a C-terminal domain that includes an oligosaccharide-binding (OB) fold (Todone et al., 2001). The approximate location of Rpb4/7 on the core polymerase was first determined by electron microscopy (EM) of two-dimensional crystals (Jensen et al., 1998). Later, EM analysis of single particles revealed a closed clamp and showed that the Rpb4/7 subcomplex protrudes from outside the core enzyme below the clamp (Craighead et al., 2002). A different open-closed transition that involved the polymerase jaws was observed by EM of two-dimensional crystals (Asturias et al., 1997). Crystallographic backbone models of the complete Pol II then revealed the exact position and orientation of Rpb4/7 and showed that it formed a wedge between the clamp and the linker to the unique tail-like C-terminal repeat domain (CTD) of Rpb1 (Armache et al., 2003; Bushnell and Kornberg, 2003). The crystal structure of free Rpb4/7 (Figure 1.2d), together with an improved resolution of the complete Pol II crystals, finally enabled refinement of a complete atomic model of Pol II (Armache et al., 2005) (Figure 1.2b).
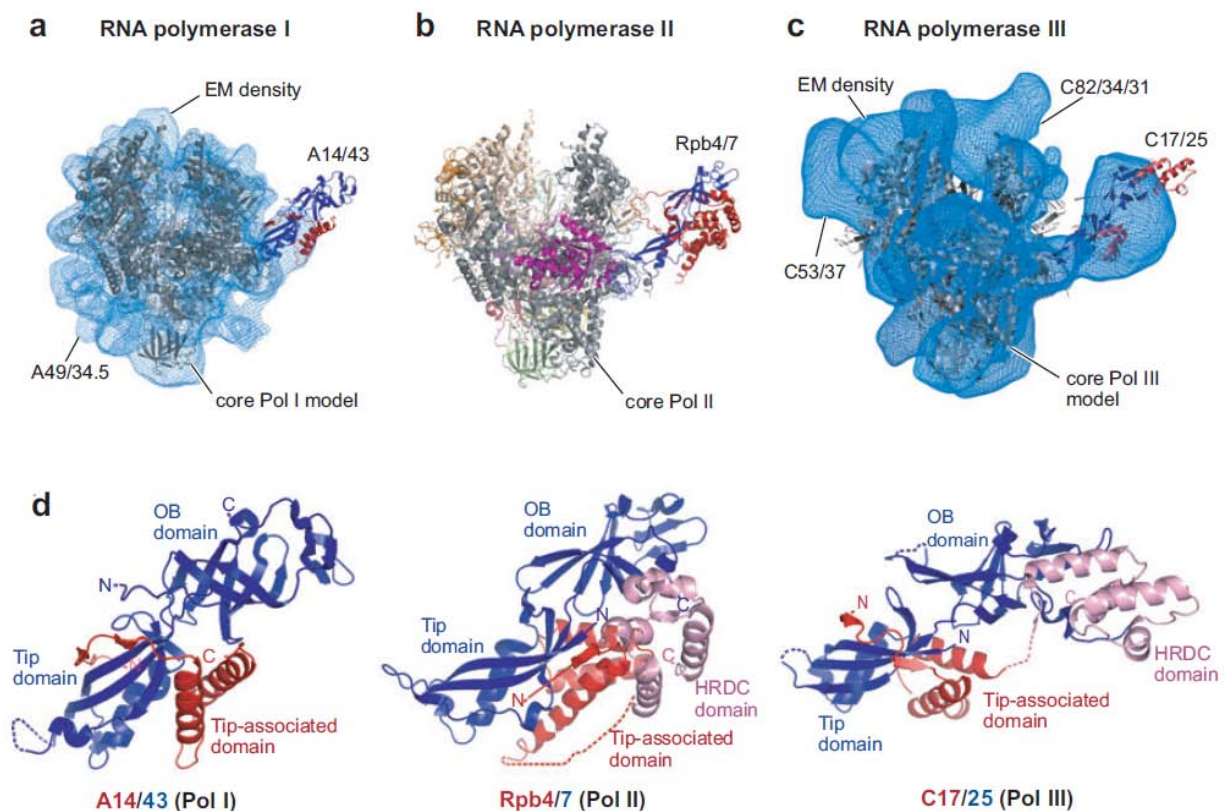
Figure 1.2 RNA polymerase structures.

(a) Pol I hybrid structure (47).

(b) Ribbon model of the refined complete Pol II crystal structure (Armache et al., 2005).

(c) Pol III EM structure (Fernandez-Tornero et al., 2007) with the Pol II-based homology model and the crystal structure of C17/25 (Jasiak et al., 2006) fitted as published (Fernandez-Tornero et al., 2007).

(d) Rpb4/7 subcomplex structures: A14/43 (left), Rpb4/7 (center) (Armache et al., 2005), C17/25 (right). Adopted from (Cramer et al., 2008).

The CTD of Rpb1 is flexible and consists of heptapeptide repeats of the consensus sequence YSPTSPS. The CTD integrates nuclear events by binding proteins involved in mRNA biogenesis (for reviews see References (Buratowski, 2003; Hirose and Manley, 2000; Meinhart et al., 2005)). CTD-binding proteins recognize a specific CTD phosphorylation pattern, which changes during the transcription cycle. Structural and functional studies of CTD-binding and CTD-modifying proteins and their complexes with CTD peptides elucidated CTD structure and revealed some of the mechanisms underlying CTD function.
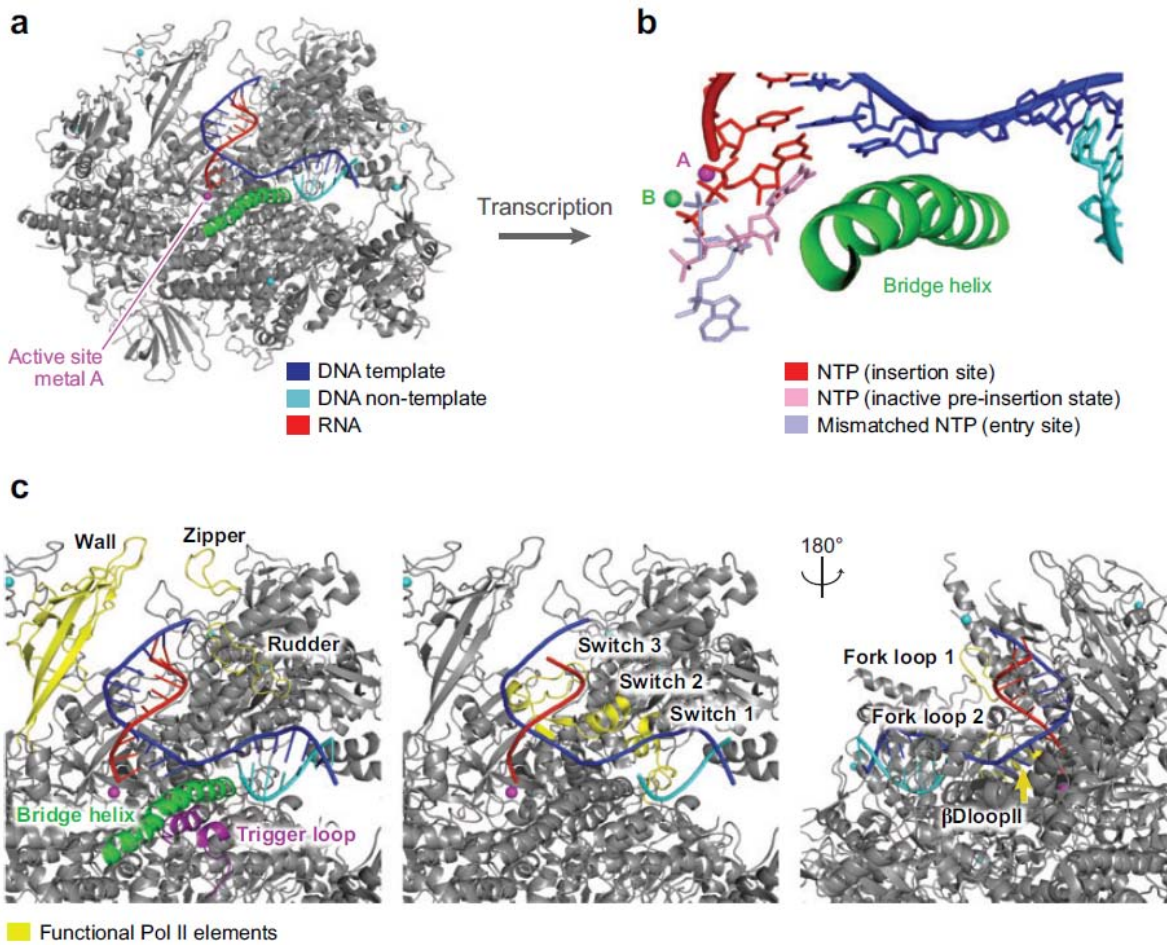
Figure 1.3 Structure of the Pol II elongation complex (EC).

(a) Overview of the EC structure (Kettenberger et al., 2004; Wang et al., 2006; Westover et al., 2004).

(b) Superposition of NTP-binding sites [red, insertion site; violet, entry site (Kettenberger et al., 2004; Wang et al., 2006; Westover et al., 2004); pink, inactive preinsertion-like state (Kettenberger et al., 2004; Wang et al., 2006; Westover et al., 2004).

(c) Functional Pol II surface elements in the EC.

1.4   The Pol II elongation complex and nucleotide incorporation

During the elongation phase of transcription, the polymerase moves along a DNA template and synthesizes a complementary chain of ribonucleotides. RNA extension begins with binding of a nucleoside triphosphate (NTP) substrate to the transcription elongation complex (EC) that is formed by the polymerase, DNA, and RNA. Catalytic addition of the nucleotide to the growing RNA 3' end then releases a pyrophosphate ion. Finally, translocation of DNA and

RNA frees the substrate site for binding of the next NTP. The EC is characterized by an unwound DNA region, the transcription bubble. The bubble contains a short hybrid duplex formed between the DNA template strand and the RNA product emerging from the active site. The mechanism of RNA elongation was elucidated by structural studies of Pol II-nucleic acid complexes (Figure 1.3). EM first revealed the point of DNA entry to the Pol II cleft (Poglitsch et al., 1999). The first crystal structure of a Pol II–nucleic acid complex which was elucidated was that of the core Pol II transcribing a tailed template DNA (Gnatt et al., 2001), this template allows for promoter-independent transcription initiation. This structure revealed downstream DNA entering the cleft and an 8 to 9 base pair DNA-RNA hybrid in the active center. Comparison with the high-resolution core Pol II structure (Cramer et al., 2001) revealed protein surface elements predicted to play functional roles. Later, polymerase EC structures utilized synthetic DNA-RNA scaffolds (Kettenberger et al., 2004; Westover et al., 2004) and revealed the exact location of the downstream DNA and several nucleotides upstream of the hybrid. Mechanisms were suggested for how Pol II unwinds downstream DNA and how it separates the RNA product from the DNA template at the end of the hybrid. Although Pol II generally uses DNA as a template, there is also evidence that Pol II can use RNA templates. Recent structures showed that an RNA template-product duplex can bind to the site normally occupied by the DNA-RNA hybrid and provided the structural basis for the phenomenon of RNA dependent RNA synthesis by Pol II (Lehmann et al., 2007).

Additional structures of Pol II ECs included the NTP substrate (Kettenberger et al., 2004; Wang et al., 2006; Westover et al., 2004). These studies suggested how Pol II selects the correct NTP and how it incorporates a nucleotide into RNA. The NTP was crystallographically trapped in the insertion site (Wang et al., 2006; Westover et al., 2004), which is apparently occupied during catalysis, but a separate study also revealed the NTP in an overlapping, slightly different location, suggesting an inactive NTP-bound preinsertion state of the enzyme (Kettenberger et al., 2004). Both NTPs form Watson-Crick interactions with a base in the DNA template strand. Binding of the NTP to the insertion site involves folding of the trigger loop (Wang et al., 2006), a mobile part of the active center first observed in free bacterial RNA polymerase (Vassylyev et al., 2002), and in the Pol II-TFIIS complex (Kettenberger et al., 2003). Folding of the trigger loop closes the active site and may be involved in selection of the correct NTP. The NTP complex structures revealed contacts of the nucleotide with the polymerase, which explain discrimination of ribonucleotides against deoxyribonucleotides, and provided insights into the selection of the correct nucleotide complementary to the templating DNA base. Catalytic nucleotide incorporation apparently follows a two-metal ion mechanism suggested for all polymerases (Steitz, 1998). The Pol II

active site contains a persistently bound metal ion (metal A) and a second, mobile metal ion (metal B) (Cramer et al., 2001). Metal A is held by three invariant aspartate side chains and binds the RNA 3' end (Cramer et al., 2001), whereas metal B binds the NTP triphosphate moiety (Westover et al., 2004). Recent studies of functional complexes of the bacterial RNA polymerase revealed the close conservation of the EC structure (Vassylyev et al., 2007a) and provided additional insights into nucleotide incorporation (Vassylyev et al., 2007b). As for Pol II, NTP binding to the insertion site can induce folding of the trigger loop. In the presence of the antibiotic streptolydigin, however, the NTP binds in the inactive, preinsertion state, in which the triphosphate and metal B are too far from metal A to permit catalysis. This finding supported a two step mechanism of nucleotide incorporation (Kettenberger et al., 2004; Vassylyev et al., 2007b). The NTP would first bind in the inactive preinsertion state to an open active center conformation. Complete folding of the trigger loop then leads to closure of the active center, delivery of the NTP to the insertion site, and catalysis. An alternative model for nucleotide addition involves binding of the NTP to a putative entry site in the pore, in which the nucleotide base is oriented away from the DNA template, and rotation of the NTP around metal ion B directly into the insertion site (Westover et al., 2004).

1.5   Overcoming obstacles during elongation

During active transcription, Pol II must overcome intrinsic DNA arrest sites, which are generally rich in A-T base pairs and pose a natural obstacle to transcription. At such sites, Pol II moves backward along DNA and RNA, resulting in extrusion of the RNA 3' end through the polymerase pore beneath the active site and transcriptional arrest. The RNA cleavage stimulatory factor TFIIS can rescue an arrested polymerase by creating a new RNA 3' end at the active site from which transcription can resume. The mechanism of TFIIS function was elucidated with the structures of Pol II and a Pol II EC in complex with TFIIS (Kettenberger et al., 2003, 2004). TFIIS inserts a hairpin into the polymerase pore and complements the active site with acidic residues, changes the enzyme conformation, and repositions the RNA transcript (Kettenberger et al., 2003, 2004). These studies supported the idea that the Pol II active site is tunable, as it can catalyze different reactions, including RNA synthesis and RNA cleavage (Kettenberger et al., 2003; Sosunov et al., 2003).

Other obstacles to transcription are bulky lesions in the DNA template strand. Structural studies of Pol II ECs that contain a lesion in the template strand unraveled the mechanism of polymerase stalling at a bulky lesion, a UV light-induced thymine-thymine cyclobutane

pyrimidine dimer (CPD) (Brueckner et al., 2007). Cells efficiently eliminate CPDs by transcription-coupled DNA repair (TCR), which begins with Pol II stalling at the lesion. TCR continues with assembly of the nucleotide excision repair machinery, removal of a lesion-containing DNA fragment, and repair of the resulting DNA gap. The structural studies revealed that the structure of the EC is generally not influenced by the presence of the lesion, except for some changes in the downstream DNA, arguing against allosteric models for the assembly of the repair machinery during TCR. The study also demonstrated a translocation barrier that impairs delivery of the CPD to the active site. This study also showed that Pol II can, under certain artificial conditions, bypass a distorting CPD lesion.

1.6   Transcriptional mutagenesis

When DNA damage is located in transcribed regions of the genome, RNA polymerase can stall, leading to TCR, or RNA polymerase can continue transcribing its product past a lesion (Figure 1.4), in this case mistakes may be introduced. If mistakes are introduced, for example by a miscoding or non-informational damage-site, transcriptional mutagenesis can occur. Furthermore, analysis of the miscoding properties of many lesions indicate that they will miscode similarly during both replication and transcription (Viswanathan et al., 1999). Transcriptional mutagenesis that is induced by DNA damage can in principle generate a pool of mutant mRNAs that could result in the production of mutant proteins that would alter the cellular phenotype (Figure 1.5). These mutant proteins might give the cell a growth advantage, or the ability to escape growth suppression, and the ensuing DNA replication past the lesion could convert it into a heritable mutation, giving rise to a mutant cell population now permanently expressing this advantageous protein (Saxowsky and Doetsch, 2006). This process has been called retromutagenesis since a transcriptional event could lead to a permanent DNA sequence change (Viswanathan and Doetsch, 1998). In nonproliferating cells, the contribution of transcriptional mutagenesis to the mutant protein pool, is likely to be much more apparent, because the capacities of certain DNA repair pathways are diminished in nondividing cells (Saxowsky and Doetsch, 2006).
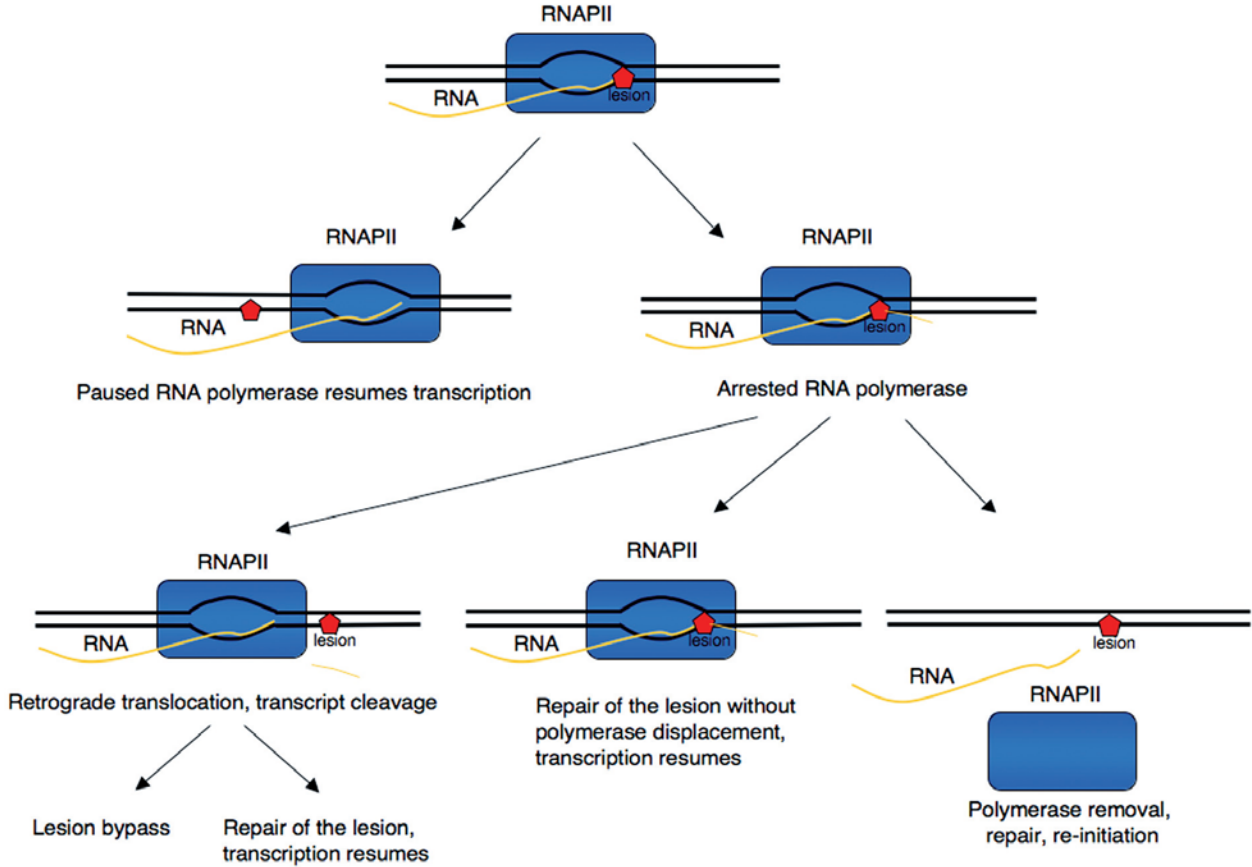
Figure 1.4 Transcriptional encounters with DNA damage: What are the options?

The presence of a lesion in the transcribed strand may cause different transcription outcomes depending on whether the lesion transiently pauses or permanently arrests the elongating polymerase. Adopted from (Tornaletti, 2009).
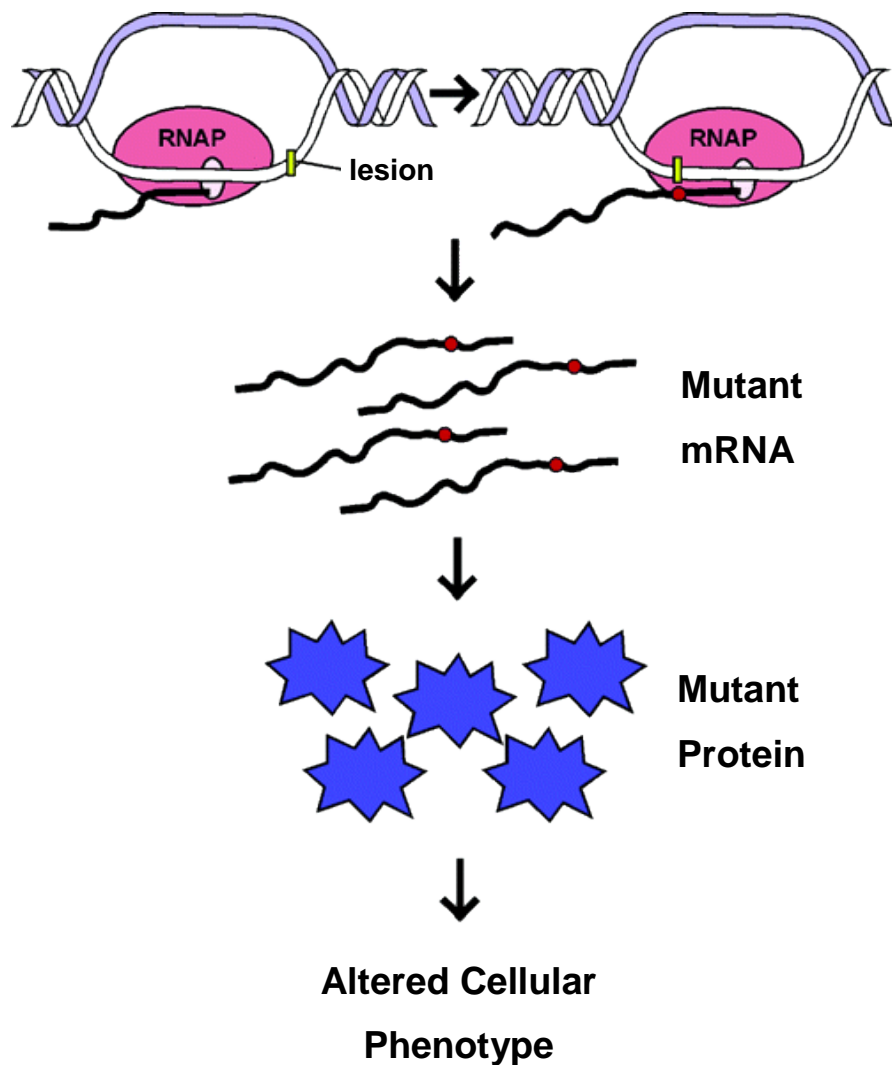
Figure 1.5 Transcriptional mutagenesis.

Transcription past a DNA lesion (yellow box) with altered base pairing properties may lead to the production of a population of mutant transcripts. These transcripts can, in turn, be translated into mutant proteins that could alter the phenotype of the cell. Adopted from (Saxowsky and Doetsch, 2006).

1.7 Scope of this work

Prior to this work, the mechanisms of recognition and bypass of CPD DNA lesions was analyzed (Brueckner et al., 2007). However, at that point it was unknown whether the mechanisms of transcriptional stalling or mutagenesis at other types of lesions could be predicted from this study. In addition, it was unknown whether transcriptional mutagenesis could occur at other lesions. To further elucidate these issues, the scope of this work was to introduce different lesions, specifically cisplatin and 8-oxoguanine, into the DNA template strand at several positions around the polymerase active site and to study the resulting Pol II ECs. This was achieved structurally, by solving the structure of Pol II in complex with these lesions, and functionally, by performing RNA elongation and cleavage assays. Together the mechanism of DNA damage recognition and the mechanism of transcriptional mutagenesis can be obtained.

2   Mechanism of transcriptional stalling at cisplatin-damaged DNA

2.1   Introduction

Cisplatin (cis-diamminedichloroplatinum(II)) is a widely used anticancer drug that forms DNA adducts that interfere with replication and transcription (Wang and Lippard, 2005). The most frequent cisplatin DNA adducts are 1,2-d(GpG) intrastrand cross-links, or cisplatin lesions, in which platinum coordinates the N7 atoms of adjacent guanosines in a DNA strand (Kartalou and Essigmann, 2001) (Figure 2.1). A cisplatin lesion in the DNA template strand blocks transcription elongation by the single-subunit RNA polymerase from phage T7 (Jung and Lippard, 2003) and by Pol II (Corda et al., 1991, 1993; Tornaletti et al., 2003), and leads to stable polymerase stalling (Jung and Lippard, 2006). The stalled Pol II elongation complex can be bound by the elongation factor TFIIS, which stimulates polymerase back-tracking and 3' RNA cleavage (Tornaletti et al., 2003). A small fraction of polymerases can apparently read through a cisplatin lesion (Tornaletti et al., 2003). The detailed molecular mechanisms of cisplatin DNA adduct processing by nucleic acid polymerases are not understood. Here we have used a combination of X-ray crystallography and RNA-extension assays to derive the molecular mechanism of *Saccharomyces cerevisiae* Pol II stalling at a cisplatin lesion. Comparison of the results with our previous analysis of Pol II stalling at a DNA photolesion, a TT cyclobutane pyrimidine dimer (Brueckner et al., 2007) (CPD, Figure 2.1), reveals that the two types of dinucleotide lesions trigger transcriptional stalling by different mechanisms.

GG cisplatin lesion   TT photolesion
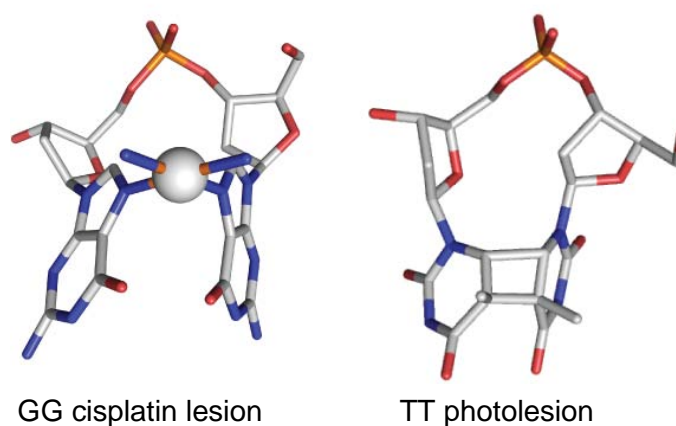
Figure 2.1 Structure of the 1,2-d(GpG) cisplatin lesion and the thymine-thymine CPD photolesion.
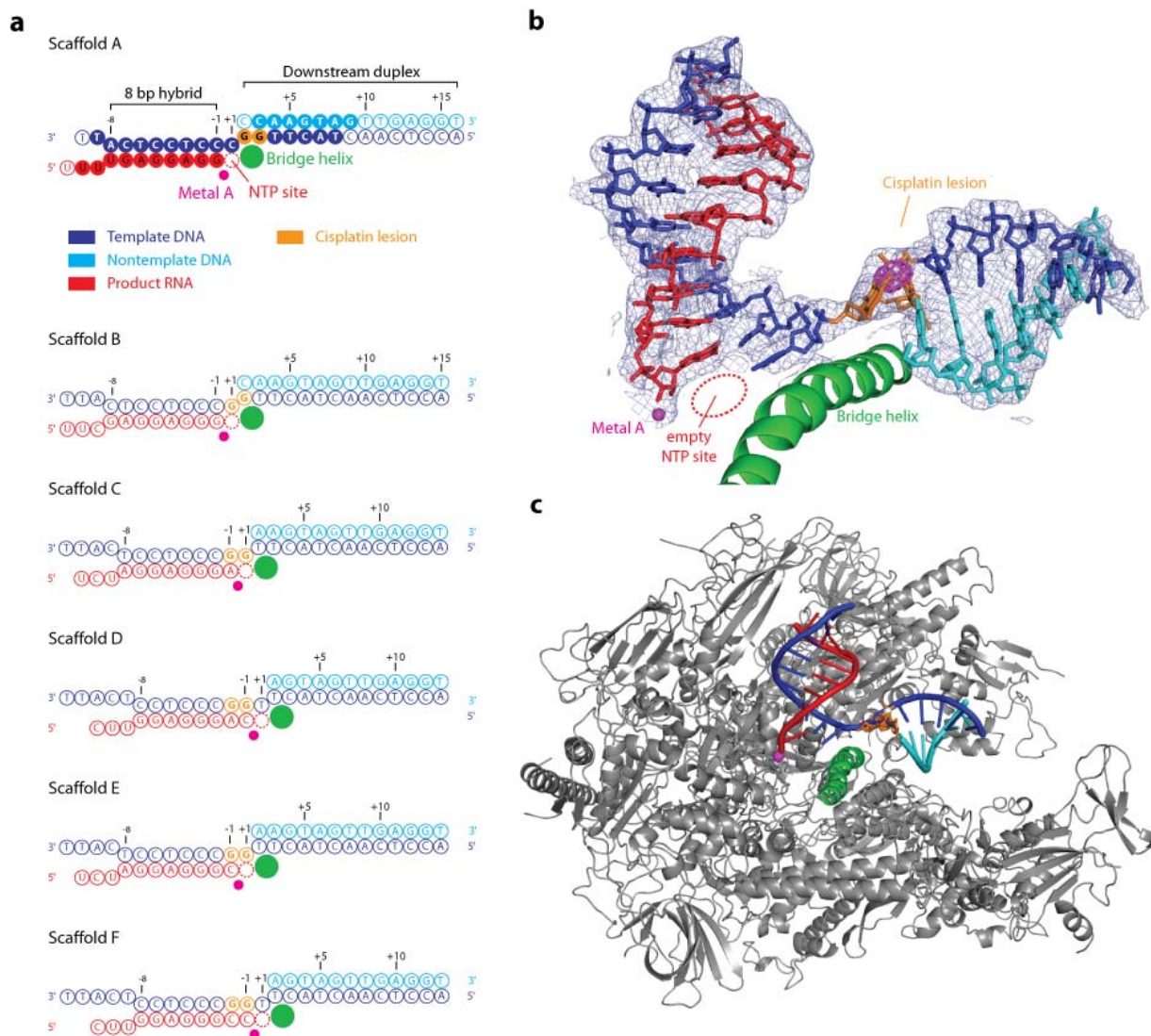
Figure 2.2 Structure of a cisplatin-damaged Pol II elongation complex.

(a) Nucleic acid scaffolds. The color code is indicated and used throughout. Filled circles denote nucleotides with interpretable electron density that were included in the structure in C. Open circles denote nucleotides with non-interpretable or lacking electron density.

(b) Structure of nucleic acids in Pol II elongation complex A. The final $2F_o$-F_c electron density map is shown for the nucleic acids (blue, contoured at 1.0σ). The anomalous difference Fourier map reveals the location of the platinum atom (magenta, contoured at 15σ).

(c) Overview of the cisplatin-damaged Pol II EC. Pol II is shown as a ribbon model in silver, with the bridge helix in green. The nucleic acids are in color, and the cisplatin lesion is shown as a stick model in orange.

2.2   Results

2.2.1   Structure of a cisplatin-damaged Pol II elongation complex

To elucidate recognition of cisplatin-induced DNA damage by transcribing Pol II, we carried out a structure-function analysis of elongation complexes containing a cisplatin lesion in the template DNA strand. Elongation complexes were reconstituted from the 12-subunit *S. cerevisiae* Pol II and nucleic acid scaffolds as described (Brueckner et al., 2007; Kettenberger et al., 2004). A cisplatin lesion was first incorporated at registers +2/+3 of the template strand, directly downstream of the NTP-binding site at register +1 (scaffold A, Figure 2.2a). The crystal structure of the resulting cisplatin-damaged elongation complex (complex A) was determined at 3.8 *Å* resolution (Figure 2.2b,c, Table 2.1 and Methods). A very strong peak in the anomalous difference Fourier map revealed the location of the platinum atom (Figure 2.2b).
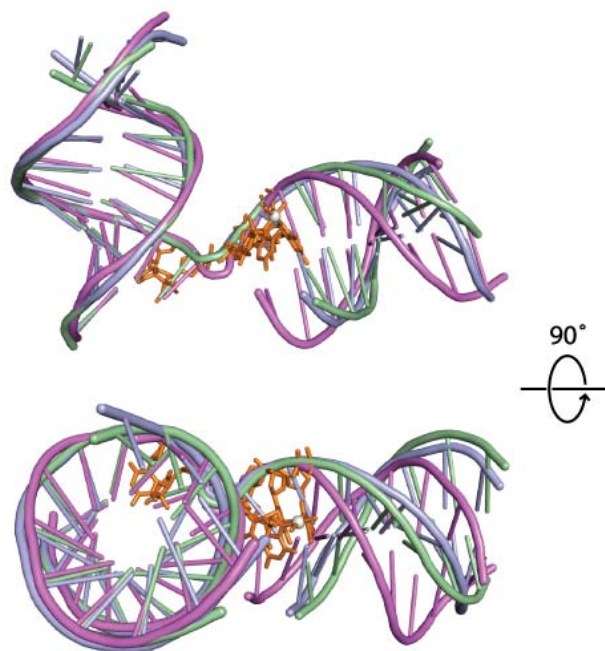
Figure 2.3 Cisplatin-induced changes in the downstream DNA.
Comparison of the course of nucleic acids in complex A (pale green) with that in the structures of the CPD-containing complexes B and C (Brueckner et al.) (violet and light blue, respectively). The proteins were superimposed based on the active site region and then omitted.

In the damaged elongation complex structure, the cisplatin lesion is bound at positions +2/+3, above the polymerase bridge helix (Figure 2.2b). The DNA-RNA hybrid occupies the upstream positions −1 to −8. The hybrid structure is essentially identical to our previous structures of the complete Pol II elongation complex (Kettenberger et al., 2004) and the elongation complex containing a CPD lesion at the polymerase active site (Brueckner et al., 2007). However, the downstream DNA duplex adopts a slightly altered position (Figure 2.3). 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 (Brueckner et al., 2007). Structures of free DNA containing a cisplatin lesion (Gelasco and Lippard, 1998; Takahara et al., 1995) reveal that the lesion at the center of the duplex leads to DNA bending. 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.

Table 2.1 Crystallographic data and refinement statistics for elongation complex A (scaffold A, cisplatin lesion at positions +2/+3).

	Pol II EC with cisplatin lesion
Data collection	
Space group	$C222_1$
Cell dimensions *a*, *b*, *c* (Å)	222.1, 393.1, 283.7
Resolution (Å)	50.0-3.80 (4.02-3.80)[1]
R_{sym}	8.8 (36.2)
$I / \sigma I$	11.2 (3.4)
Completeness (%)	99.1 (97.0)
Redundancy	3.4 (3.2)
Refinement	
Resolution (Å)	50.0-3.80
No. reflections	235,052
R_{work} / R_{free}[2]	21.5 / 24.0
No. atoms	
Protein	31,102
Nucleic acids	698
B-factors	
Protein	115.5
Nucleic acids	144.0
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.5
Pt peak in anomalous difference Fourier (σ)	23.5

Diffraction data were collected from a single crystal. [1]Values in parentheses are for the highest-resolution shell. [2]For free R-factor calculation, we excluded from refinement the same set of reflections that had been excluded from previous Pol II structure determinations (Armache et al., 2005; Brueckner et al., 2007; Kettenberger et al., 2004).

2.2.2   RNA polymerase II stalling and AMP misincorporation

The crystal structure of complex A defines a state of the elongation complex in which the lesion lies downstream of the active site. We therefore used complex A in RNA-extension assays to investigate the mechanism of Pol II stalling. The RNA was labeled with a fluorescent dye at its 5' end, complex A was incubated with NTP substrates, and the RNA-extension products were separated on polyacrylamide gels and visualized with a fluorimager (Figure 2.4). After incubation with a physiological concentration of 1 mM NTPs for 20 min, most of the RNA was extended by two nucleotides (Figure 2.4a). 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 (Figure 2.4a). To further investigate this, we prepared a scaffold that resulted in an elongation complex that was advanced by one position (scaffold B, Figure 2.2a). As expected, the results of incubation of complex B with individual NTPs again revealed AMP misincorporation (Figure 2.4a).

A time course of RNA extension with complex A showed that the first incorporation event was fast, whereas the second incorporation - that is, the apparent misincorporation - was much slower and incomplete (Figure 2.4b, upper). Apparently, the initial incorporation is fast because GMP is correctly incorporated opposite a template cytidine and because the substrate site is free in complex A (Figure 2.2b), and translocation of the polymerase is not required. The slow rate of the second incorporation event was confirmed in a time course of RNA extension with complex B (Figure 2.4b, lower). This misincorporation event also required high ATP concentrations and was not observed in the presence of only 10 mM ATP (Figure 2.4b).
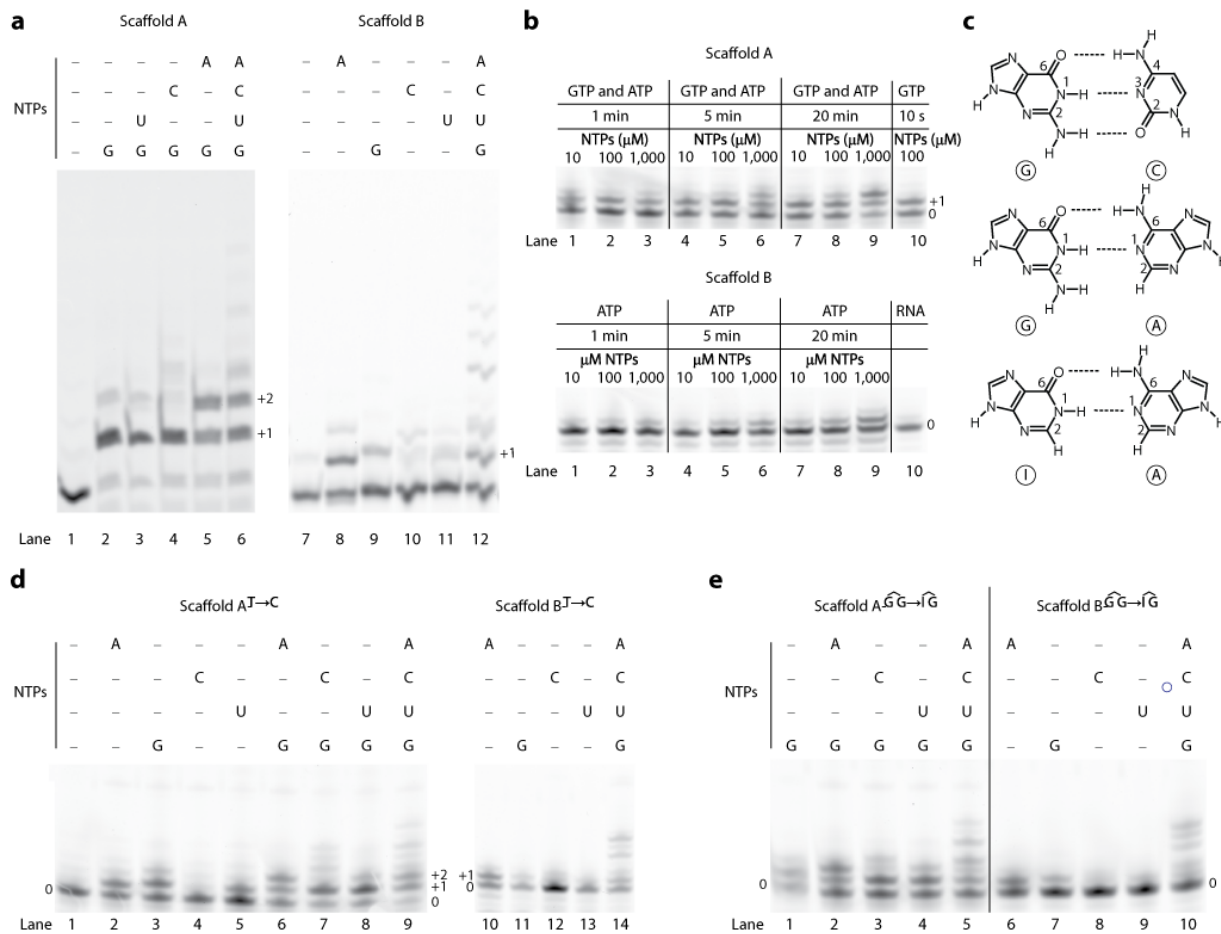
Figure 2.4 Pol II stalling and lesion-directed misincorporation

(a) 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 Pol II and different types of NTPs (1 mM) for 20 min as indicated (Methods).

(b) Time courses of the RNA extension reactions shown in A and dependence on the NTP concentration.

(c) Possible base pair formation. Hydrogen bonds are indicated with dashed lines.

(d) Adenine misincorporation does not result from template misalignment. Altered scaffolds were used, in which the thymine at the 5' flanking position of the lesion was replaced by cytosine (Figure 2.5). The misincorporation seen in lane 2 is also observed with an undamaged scaffold (Figure 2.6) and is therefore not due to the lesion.

(e) The 2-amino group at the 3' position of the lesion is not involved in directing misincorporation. Altered scaffolds were used in RNA extension, in which the guanine at the 3' position of the lesion was replaced by inosine.
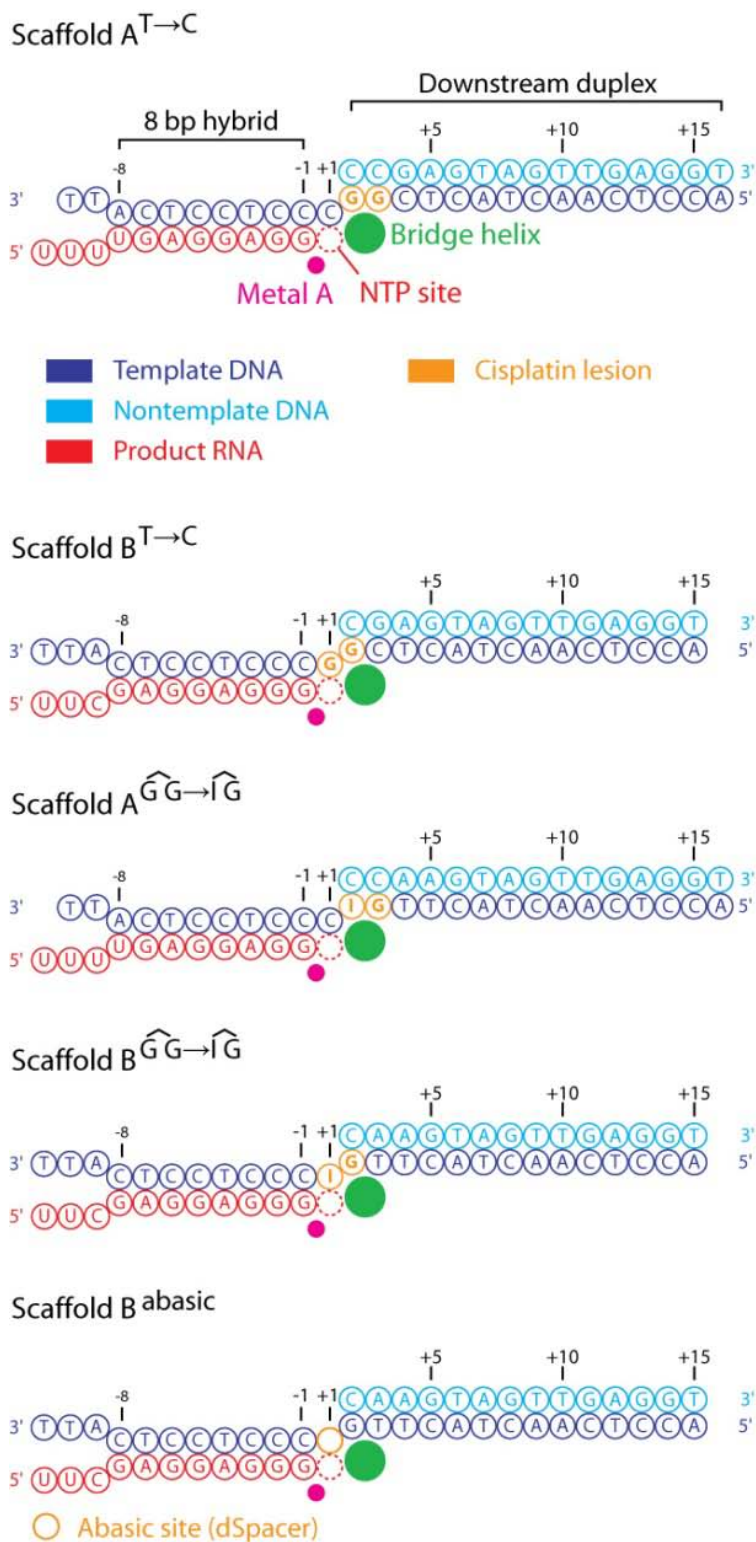
Figure 2.5 Additional scaffolds used in RNA extension assays.
The color code is indicated and used throughout.
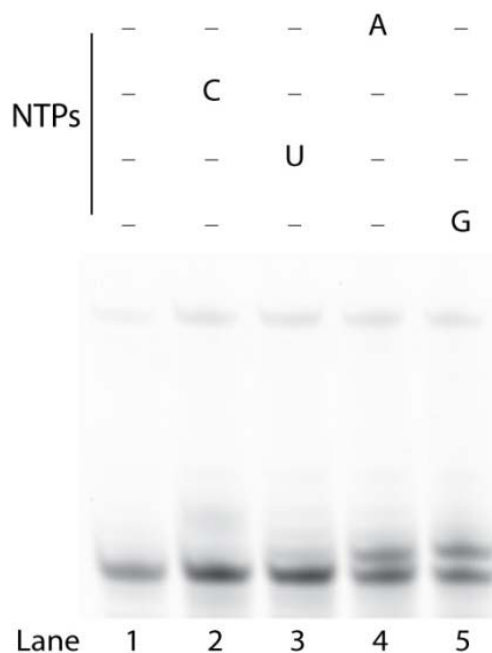
Figure 2.6 Nucleotide incorporation using an undamaged scaffold.
RNA extension assays with a scaffold that is identical to scaffold A but lacks the lesion. This experiment showed that A misincorporation occurs with this scaffold in the absence of a lesion at a position where templated incorporation leads to G incorporation. Incubation time was 5 min, NTP concentration was 1 mM.

2.2.3   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 (Kashkina et al., 2006). 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, we prepared a scaffold that was identical to scaffold A except that the flanking thymidine was replaced by cytidine (scaffold A$^{T->C}$, Figure 2.5). This altered scaffold gave rise to the same RNA-extension products (Figure 2.4d), however, showing that template misalignment did not occur. We confirmed this by altering scaffold B and repeating the RNA-extension assay (Figure 2.4d). 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 2.4c). 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 2.4c). 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 2.4e). Thus, AMP misincorporation does not involve the 2-amino group. These results were consistent with lesion-templated misincorporation involving a G-A base pair (Figure 2.4c), but we could not test this directly.

2.2.4 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 (Brueckner et al., 2007). Template bases in positions +1/+2 are twisted against each other by about 90° in structures of the undamaged elongation complex (Gnatt et al., 2001; Kettenberger et al., 2004), but such twisting is impossible for nucleotides that are covalently linked in dinucleotide lesions, giving rise to a translocation barrier (Brueckner et al., 2007). To determine whether translocation of the cisplatin lesion is indeed impaired, we crystallized complex B, which was designed to contain the dimer at positions +1/+2 (scaffold B, Figure 2.2a). The anomalous difference Fourier map revealed a platinum peak at the same location where one was seen in complex A, indicating that the polymerase had apparently stepped backward by one position, so that the cisplatin lesion again occupied positions +2/+3 (Figure 2.7a). A lower height of the platinum peak (Table 2.2) 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, and translocation from positions +2/+3 to positions +1/+2 was apparently disfavored. These results suggest that the lesion does not stably bind the active site at all. To test this, we prepared 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 2.2a), and crystallized the resulting complex C. The anomalous difference Fourier map revealed a platinum peak at the same location as in complexes A and B, indicating that the polymerase had stepped backward by two positions so that the cisplatin lesion again occupied positions +2/+3 (Figure 2.7b). The height of the platinum peak was lower than for complexes

A and B (Table 2.2), and the electron density for the nucleic acids was again fragmented, preventing model-building. We also tried to solve a structure with the lesion placed at positions –2/–1 (scaffold D, Figure 2.2a). However, crystal structure analysis (Table 2.2) revealed no electron density for nucleic acids, consistent with a low affinity of this scaffold for the polymerase. Thus, the cisplatin lesion was not stably accommodated in the active site.

Table 2.2: Additional crystallographic data statistics.

Scaffold/complex[1]	B	C	D
Designed position of cisplatin lesion	+1/+2	-1/+1	-2/-1
Data collection			
Space group	$C222_1$	C2	C2
Unit cell axes (Å)	222.6, 394.1, 283.6	393.3, 222.3, 283.1	393.6, 222.4, 285.1
Unit cell ß angle (°)	90	90.6	90.6
Wavelength (Å)	1.07158	1.07155	1.07158
Resolution range (Å)	50.0-4.0 (4.14-4.0)[2]	50.0-3.8 (3.94-3.8)	50.0-3.8 (3.94-3.8)
Unique reflections	100,722 (9,484)	219,371 (22,754)	238,532 (24,245)
Completeness (%)	98.7 (93.9)	98.1 (99.1)	99.8 (98.7)
R_{sym} (%)	10.9 (38.8)	6.0 (27.2)	12.4 (35.5)
$I/\sigma(I)$	21.4 (5.1)	12.0 (3.5)	9.0 (4.5)
Pt peak in anomalous Fourier (σ)	18.2	12.5	-

[1]Diffraction data were collected at the Swiss Light Source beamline PX1 and were processed with program XDS (Kabsch, 1993) or DENZO (Otwinowski, 1996) (complex B).

[2]Numbers in parenthesis correspond to the highest resolution shells.
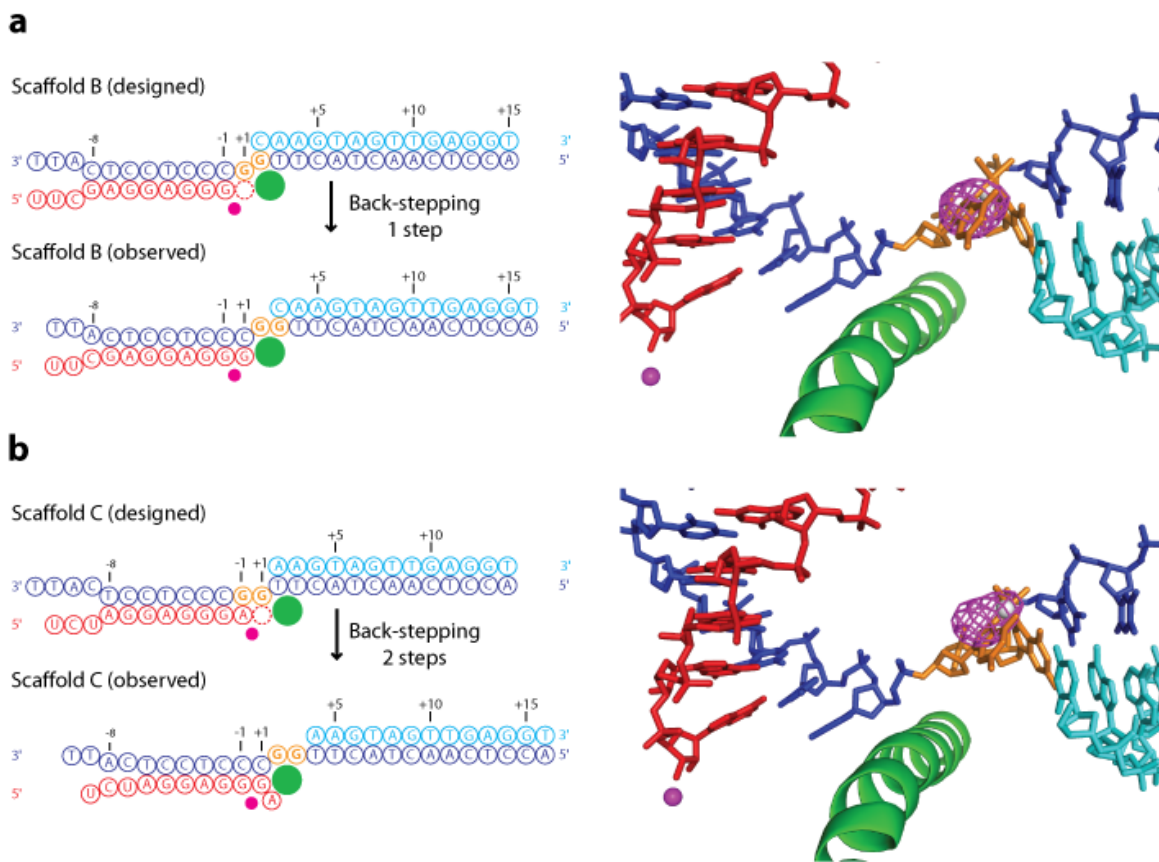
Figure 2.7 The cisplatin lesion is not stably accommodated in the active site.

(a) Anomalous difference Fourier maps of the Pol II elongation complex B. The contour levels of the anomalous difference Fourier maps is 6σ. The model of complex A is shown. The view is from the side.

(b) Anomalous difference Fourier map of the Pol II elongation complex C. The contour levels of the anomalous difference Fourier maps is 6σ. The model of complex A is shown. The view is from the side.

2.2.5 Nontemplated AMP incorporation and an 'A-rule' for Pol II

The crystallographic data strongly suggest impairment of translocation of the lesion to a position where it can direct AMP misincorporation. We therefore asked ourselves whether AMP incorporation resulted from nontemplated synthesis. Such preferential, nontemplated AMP incorporation occurs when DNA polymerases encounter a bulky DNA lesion and is known as A-rule (Strauss, 1991; Taylor, 2002). We prepared 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 2.5). Indeed, AMP, and also to a lesser extent GMP, could be incorporated into RNA opposite the abasic position (Figure 2.8a), showing that Pol II

incorporates purine nucleotides, preferentially AMP, when a templating base is unavailable. 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 (Figure 2.4b, Figure 2.8a). In the presence of the cleavage factor TFIIS, template GMP incorporation with scaffold A occurred normally, but nontemplated AMP incorporation was suppressed (Figure 2.8b), consistent with a role of TFIIS in transcriptional proofreading (Thomas et al., 1998; Wind and Reines, 2000). However, TFIIS was unable to promote passage over the translocation barrier. Together, these observations suggest that Pol 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.
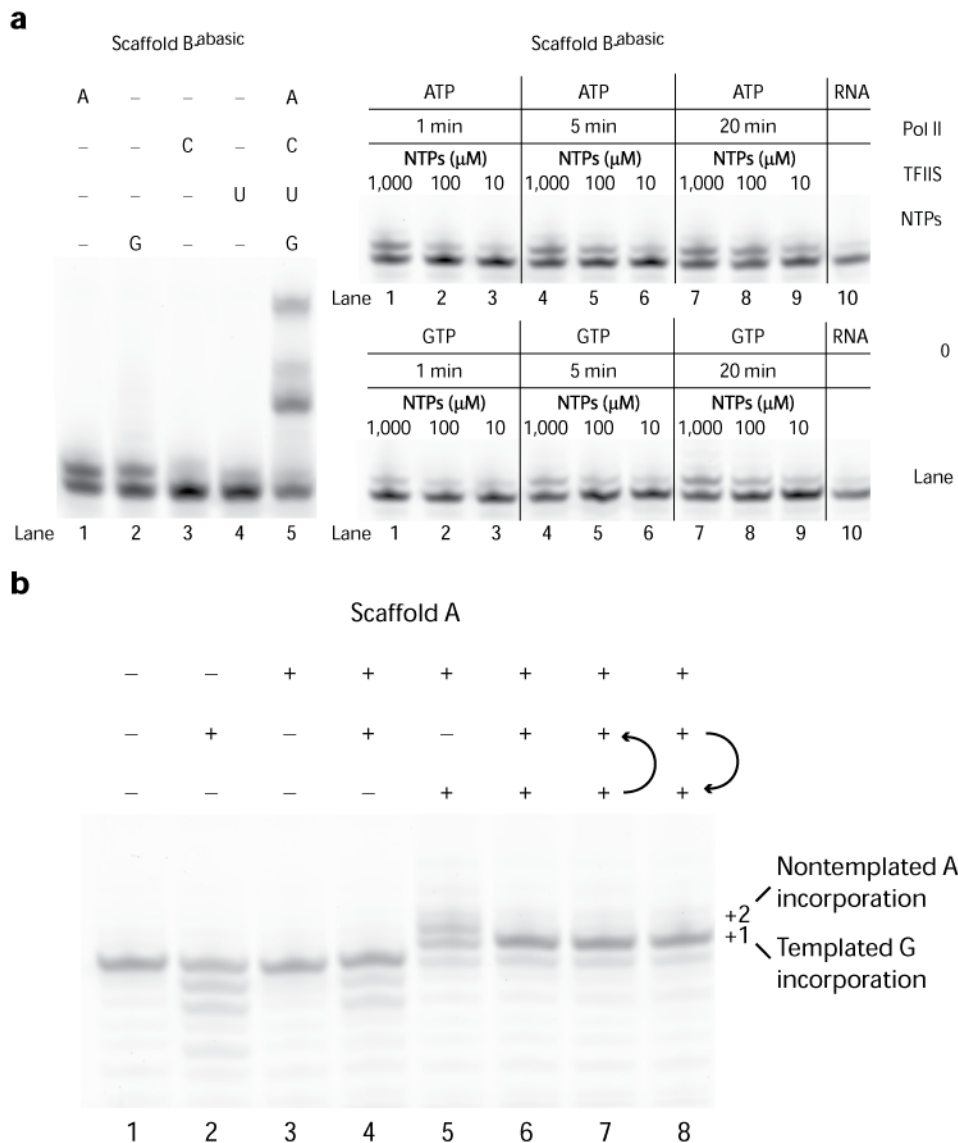
Figure 2.8 RNA extension assays

(a) 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 (Scaffold B^{abasic}, Figure 2.5).

(b) Effect of RNA cleavage factor TFIIS on RNA elongation with scaffold A. NTPs and TFIIS were either added simultaneously (lane 6), or sequentially as indicated by arrows with individual incubation times of 20 min (lanes 7-8).

2.2.6   Artificial bypass of a cisplatin lesion

The above data suggest that impaired translocation and delivery of the lesion to the active site causes polymerase stalling, but do not exclude the possibility that stalling is due to the terminal AMP misincorporation and impaired elongation of the resulting G-A mismatch. To test whether polymerase stalling depends on the misincorporation event, we carried out RNA-extension studies with scaffold C, which contained the misincorporated adenosine opposite the 3' guanosine of the cisplatin lesion, and with a scaffold that contained the cognate cytidine at this position (scaffold E, Figure 2.2a). Incubation of complexes C and E with individual NTPs showed that CMP was correctly incorporated opposite the 5' guanosine of the lesion, although AMP misincorporation also occurred, albeit less efficient (Figure 2.10a). The majority of elongation complexes C and E did not allow for RNA extension past the lesion (Figure 2.10a), indicating that they are stably stalled and that stalling is independent of the G-A mismatch. Thus, misincorporation is not required for stalling. Prolonged incubation of scaffolds C and E with Pol II and NTPs, however, led to a small but reproducible amount of run-off products (Figure 2.10a). These products may be a result of lesion bypass. Alternatively, they may arise from deplatination of the DNA, maybe by transfer of the platinum to an amino acid side chain of Pol II. To investigate this, we analyzed the nucleic acid products resulting from prolonged incubation of complex A with Pol II and NTPs by MALDI mass spectrometry. Only intact, fully platinated template strand was observed (Figure 2.9), and this represents strong evidence against deplatination. Thus, the cisplatin lesion apparently can occupy the active site transiently if a nucleotide is present in the RNA opposite the 3' guanosine of the lesion, enabling RNA extension past the lesion for a fraction of complexes. Modeling showed that entry of the bulky cisplatin lesion into the active site would require conformational changes but that the lesion, once entered, could be accommodated in the active site with only slight structural changes (Figure 2.10b). To provide further support for possible lesion bypass, we used scaffolds in which the RNA contained nucleotides opposite both guanosines of the cisplatin lesion (scaffolds D and F, Figure 2.2a). Incubation of the resulting elongation complexes with NTPs gave rise to efficient lesion bypass (Figure 2.10a). Notably, bypass was possible not only with a G-C match at the 3' position of the lesion, but also with a G-A mismatch at this position (Figure 2.10a). These data show that lesion bypass is facilitated by RNA nucleotides that are already present in the scaffold opposite the lesion, most probably because this facilitates binding of the lesion beyond the translocation barrier, at positions −2/−1, just upstream of the active site. Because Pol II can pass the lesion and a mismatch, stalling is not due to impaired extension of a lesion-containing hybrid or a mismatch. This supports the

model in which stalling results from impaired translocation and entry of the lesion into the active site.
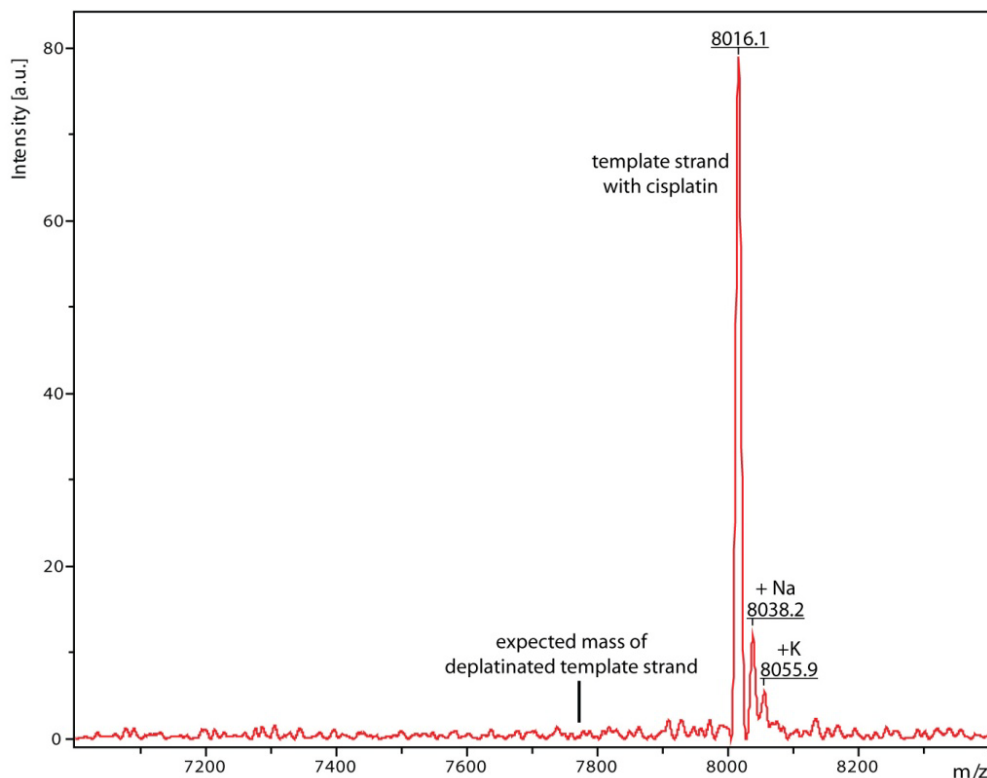
Figure 2.9 MALDI analysis of DNA platination.
Complex A was incubated with NTPs overnight and the nucleic acids were subsequently subjected to MALDI mass spectrometry. The measured mass/charge ratio of the DNA template strand containing the cisplatin lesion (8016.1) compares well with the theoretical molecular weight of 8006 Da. There is no peak near the expected mass for a deplatinated strand (7777 Da).
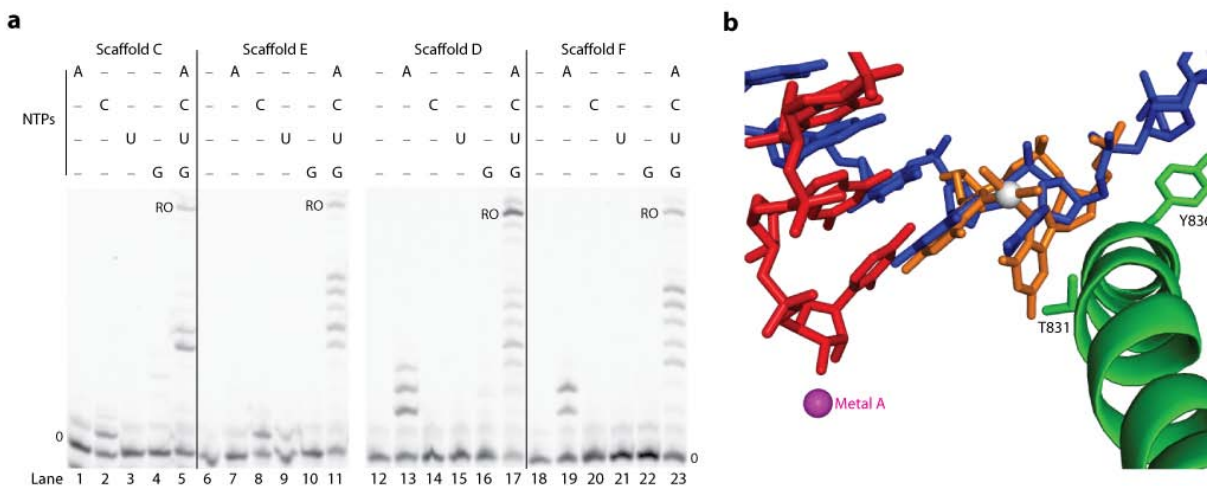
Figure 2.10 Lesion bypass

(a) RNA extension with scaffolds C, E, D, and F. Lanes 6, 12 and 18 show the fluorescently labeled reactant RNA. In the other lanes, the scaffolds were incubated with Pol II and different types of NTPs (1 mM) for 20 min as indicated (Methods). Run-off products are indicated (RO). Additional RNA extension assays confirmed these results (Figure 2.11).

(b) Modeling of the cisplatin lesion at positions +1/-1. A cisplatin lesion was manually placed such that it superimposes with template nucleotides at the active center positions +1 and −1 of complex A. Due to some close contacts, in particular with threonine T831 in the bridge helix, minor conformational changes are required to accommodate the lesion.
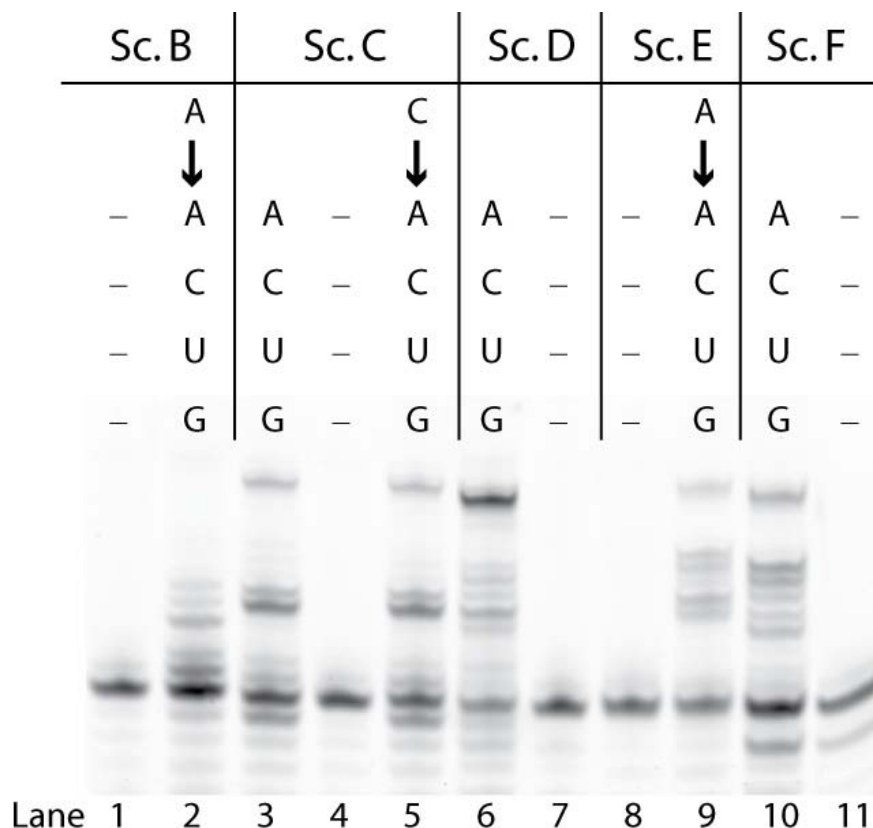
Figure 2.11 Additional RNA extension assays.

Additional RNA extension assays support the conclusions about cisplatin-induced Pol II stalling. The pattern of RNA products obtained after extension with scaffold C is different from the pattern obtained after pre-incubation of scaffold B with ATP, followed by all four NTPs. This difference can be explained by the translocation barrier. Scaffold B causes Pol II to backstep by one position (Figure 2.7a). This results in a location of the lesion before the translocation barrier. Preincubation with ATP results in nontemplated and inefficient A addition according to an A-rule, but there is no further extension since the translocation barrier cannot be overcome. In contrast, in scaffold C, there is an A residue already present at the 3' end of the RNA. This results in a small population of complexes that accommodate the lesion at the active site ("lesion placed beyond the translocation barrier") and which can be extended. As controls, very similar patterns of RNA products were observed for extension of scaffold D, and pre-incubation of scaffold C with CTP, followed by extension with all four NTPs, demonstrating the reliability of our assays. The run-off products deviate by a single nucleotide, as expected from the experimental design. Also, a very similar pattern of RNA products was obtained for extension of scaffold F, and pre-incubation of scaffold E with CTP, followed by extension with all four NTPs. Again, the run-off products deviate by a single nucleotide as expected.

2.3   Discussion

In the structural and functional studies described here, we used cisplatin lesion–containing Pol II elongation complexes to investigate the detailed mechanism of transcriptional stalling at the most frequent cisplatin-induced DNA adduct. We show that Pol II stalls in front of a cisplatin lesion because it does not pass a translocation barrier that impairs delivery of the bulky lesion into the active site. When elongation complexes are prepared in which the lesion is positioned beyond the translocation barrier, bypass of the lesion is possible. To our surprise, stalled Pol II was able to incorporate AMP in an apparently nontemplated manner according to an A-rule that has been described previously for DNA polymerases (Strauss, 1991; Taylor, 2002). The A-rule may explain inefficient AMP incorporation not only at the cisplatin lesion, but also at the same position during transcription of another dinucleotide lesion, the TT CPD photolesion (Brueckner et al., 2007). This model supersedes the non-intuitive assumption that when CPD enters the active site, its 3' thymidine acts as a template to direct AMP incorporation. AMP incorporation according to an A-rule enables formation of a matched T-A base pair at the 3' position of the CPD, and this may stabilize the lesion in the active site (Brueckner et al., 2007). In contrast, at the 3' position of the cisplatin lesion, A-rule incorporation would result in a G-A mismatch if the lesion was able to cross the translocation barrier. Pol II bypassed the cisplatin lesion and the mismatch when they were artificially placed beyond the translocation barrier, showing that the barrier prevents efficient transcriptional mutagenesis and synthesis of substantial amounts of erroneous RNAs. Comparison of the results with the previously established mechanism of Pol II stalling at a CPD (Brueckner et al., 2007) reveals that Pol II stalls at the two dinucleotide lesions for different reasons (Figure 2.12). First, the cisplatin lesion cannot overcome the translocation barrier, but the CPD can, and the CPD binds in the active site. 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. 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.

Thus, the detailed mechanisms of transcriptional stalling at two different dinucleotide lesions differ in many respects. As a consequence, it is impossible to predict the mechanisms of transcriptional stalling or mutagenesis at other types of lesions. Instead, the detailed stalling mechanisms remain to be investigated for other lesions, including the less frequent 1,3-d(GpNpG) intrastrand platinum cross-link (Jung and Lippard, 2006; Laine and Egly, 2006; Tornaletti et al., 2003; Tremeau-Bravard et al., 2004). Nevertheless, this study and our previous study of CPD-induced transcriptional stalling (Brueckner et al., 2007) suggest general aspects of the recognition of bulky dinucleotide lesions by Pol II, including a translocation barrier between positions +2/+3 and +1/+2, an A-rule for nontemplated nucleotide incorporation, and the possibility of lesion bypass under conditions that circumvent the natural stalling mechanism.
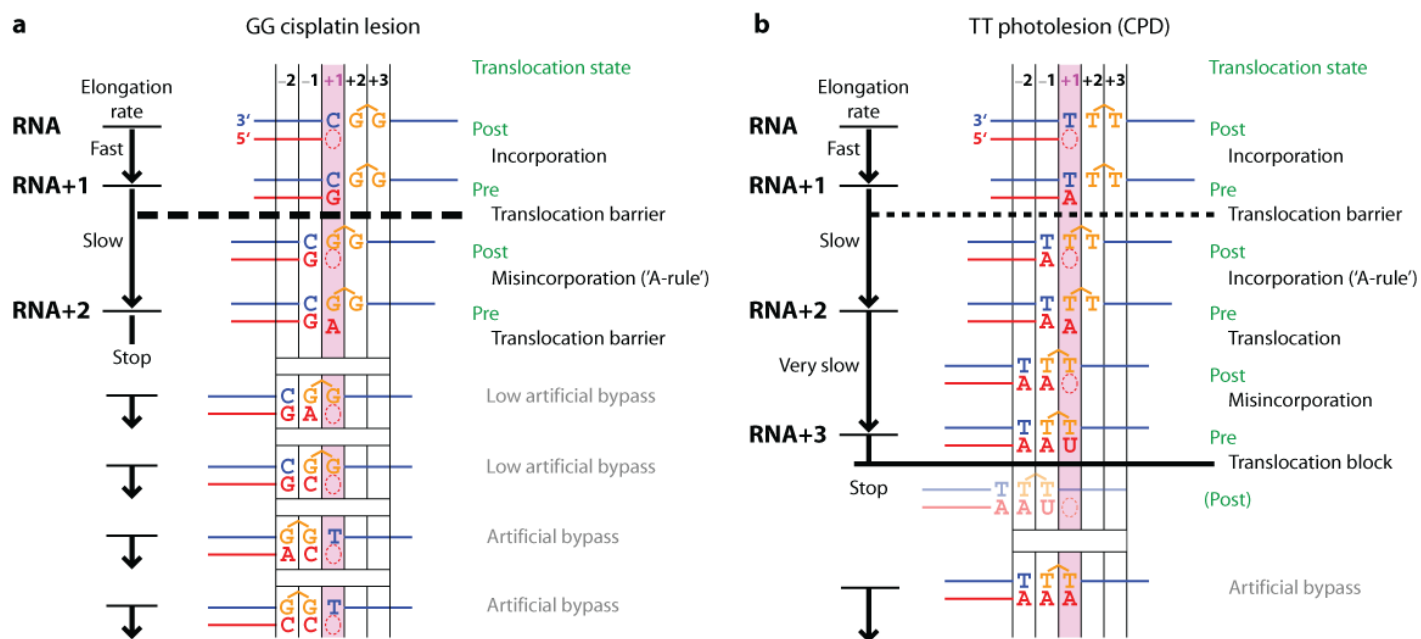
Figure 2.12 Different mechanisms of Pol II stalling at dinucleotide lesions
(a) Pol II stalling at the cisplatin lesion (this study) shown as a schematic representation of RNA extension in complex A. The initial RNA (top) corresponds to the non-extended RNA of scaffold A. The translocation barrier is indicated with a dashed horizontal line. The artificial situations leading to lesion bypass are depicted at the bottom.
(b) Pol II stalling at a CPD lesion, adapted from (Brueckner et al., 2007).

3   Molecular basis of transcriptional mutagenesis at 8-oxoguanine

3.1   Introduction

Reactive oxygen species are generated in cells by aerobic respiration, and by exposure to ionizing radiation, transition metals, chemical oxidants or free radicals. Reactive oxygen species can damage DNA, in particular by converting guanine to 8-oxoguanine (8oxoG). 8oxoG is the major mutagenic lesion in the genome (Lindahl, 1993) and can form a Watson-Crick base pair with cytosine, but also a Hoogsteen base pair with adenine, since oxidation at the C8-position converts the N7 hydrogen bond acceptor to a hydrogen bond donor (Figure 3.1a)(Hsu et al., 2004). Previous studies revealed that mammalian Pol II incorporates cytosine or adenine opposite 8oxoG, and can bypass the lesion (Kathe et al., 2004; Kuraoka et al., 2003; Tornaletti et al., 2004). Lesion bypass is enhanced by TFIIS, which stimulates Pol II backtracking and 3' RNA cleavage, although this enhancement apparently does not lead to a substantial removal of misincorporated adenine (Charlet-Berguerand et al., 2006; Kuraoka et al., 2007). The molecular mechanism of 8oxoG processing by RNA polymerases however remained to be elucidated.

Here we combined functional assays and X-ray crystallography to derive the detailed molecular mechanism of transcription by Pol II from the yeast *S. cerevisiae* over DNA that contains an 8oxoG lesion in the template strand. We show that Pol II can misincorporate adenine at 8oxoG, and that the misincorporated adenine escapes intrinsic polymerase proofreading and remains in the transcript upon elongation, resulting in transcriptional mutagenesis. The mutant transcript is observed directly by mass spectrometry. We then use X-ray analysis to show that the misincorporated adenine forms a Hoogsteen base pair with 8oxoG at the Pol II active center. In contrast, 8oxoG forms a standard Watson-Crick base pair with a correctly incorporated cytosine. Finally, we suggest a mechanism for mutagenic transcription and its suppression by TFIIS.

3.2   Results

3.2.1   Yeast Pol II slowly misincorporates adenine at 8oxoG.

To elucidate recognition and processing of 8oxoG-induced DNA damage by transcribing Pol II, we carried out a structure-function analysis of reconstituted Pol II elongation complexes (ECs) containing an 8oxoG lesion at defined positions in the DNA template strand. ECs were reconstituted from pure, complete 12-subunit *Saccharomyces cerevisiae* Pol II and minimal nucleic acid scaffolds as described (Damsma et al., 2007). An 8oxoG lesion was included at position +1 of the template strand (Figure 3.1b, scaffold A). In the resulting EC (complex A), the lesion lies in the polymerase active site, directly opposite the site that binds the nucleoside triphosphate (NTP) substrate, the nucleotide addition site (Figure 3.1b, scaffold A). To monitor RNA extension, we added a fluorescent label to the RNA 5' end. EC A was incubated with NTP substrates, and the RNA products were separated on a polyacrylamide gel and visualized with a fluorimager (Figure 3.1c). Incubation with a physiological concentration of 1 mM NTPs led to RNA extension and accumulation of the run-off product (Figure 3.1c, lanes 1-3). Very similar results were obtained in RNA extension assays with an undamaged template (Figure 3.2), showing that 8oxoG does not strongly interfere with Pol II elongation, consistent with published data (Kathe et al., 2004; Tornaletti et al., 2004). Lesion bypass was possible with scaffolds that contained 8oxoG opposite register +1, but also with scaffolds that contained the lesion within the upstream DNA, at register +3 (not shown).
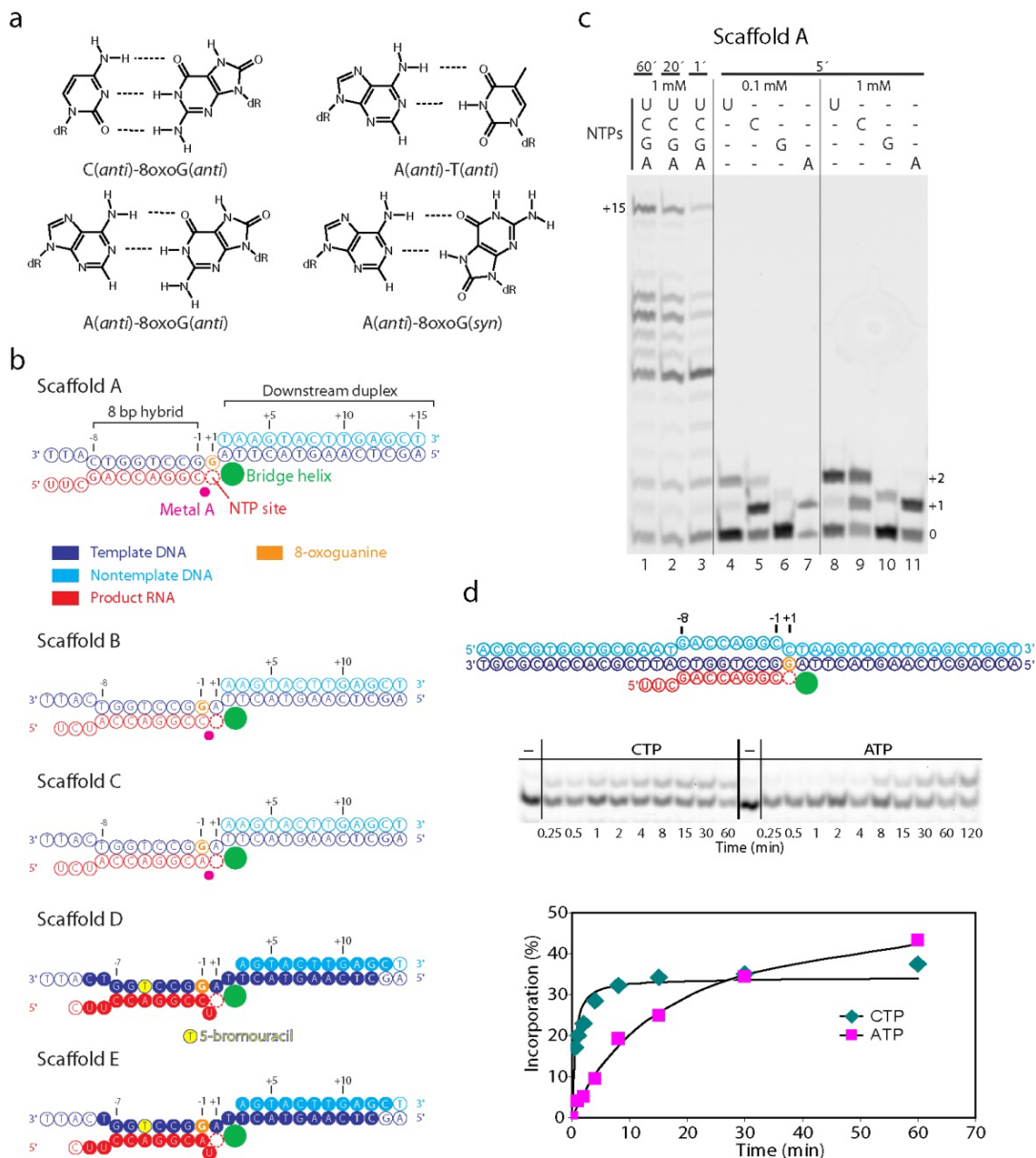
Figure 3.1. 8oxoG can direct slow adenine misincorporation.

(a) Base pairing properties of 8-oxoG. 8oxoG differs from guanine by an additional oxygen atom connected to C8, which results in an NH group at N7.

(b) Nucleic acid scaffolds. The indicated color code is used throughout. The RNA was 5' end-labeled with 6-carboxyfluoresceine (FAM) and an UGCAU linker. Filled circles denote nucleotides present in the structures (Figure 3.5).

(c) RNA extension with EC A.

(d) Time courses of incorporation with CTP or misincorporation with ATP. The indicated complementary scaffold was used. The DNA nontemplate was 5' end-labeled with Biotin with the use of a TTTTT linker. The RNA was 5' end-labeled with 6-carboxyfluoresceine (FAM) and an UGCAU linker.

Incubation with individual NTPs showed that both cytosine and adenine are incorporated opposite the lesion (Figure 3.1c, lanes 4-11). Uridine was also incorporated, but this was also observed for undamaged DNA (Figure 3.2), and is caused by a lesion-independent mechanism called template misalignment (Kashkina et al., 2006). Incubation with GTP also resulted in some misincorporation (Figure 3.1c, lanes 4-11) and for undamaged DNA some misincorporation of ATP was also observed (Figure 3.2), but these misincorporations are at a level expected from yeast Pol II in the absence of TFIIS (Sydow et al., 2009). To further investigate adenine misincorporation and cytosine incorporation opposite the lesion, we performed a time course analysis with an EC that contained a fully complementary complete scaffold, which results in a natural, complete EC coupled to magnetic beads (Dengl and Cramer, 2009). This revealed that adenine misincorporation was about 34-fold slower than cytosine incorporation (Figure 3.1d), suggesting a misincorporation frequency of a few percent, consistent with data for mammalian Pol II (Charlet-Berguerand et al., 2006). Thus misincorporation events are rare but occur at a significant frequency.
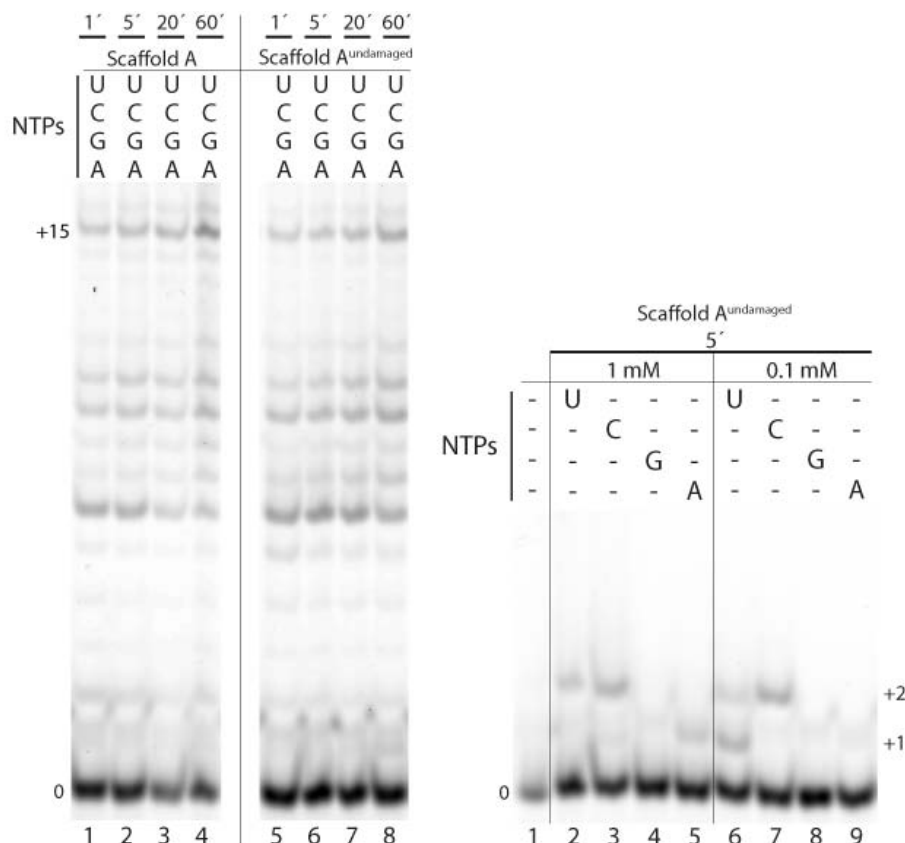
Figure 3.2 RNA extension assays.

The lanes show the fluorescently labeled RNA. The scaffolds were incubated with Pol II and indicated NTPs for the time that is indicated above. On the left, 1 mM NTP mix was used to elongate scaffold A with an 8-oxoguanine or with a normal guanine (Scaffold A^undamaged). On the right, single NTPs were added to undamaged scaffold A, different concentrations where used, as indicated.

3.2.2 Adenine misincorporation results in transcriptional mutagenesis.

The above results suggest that both adenine and cytosine can be incorporated opposite 8oxoG, but could not clarify whether adenine misincorporation also occurs in the presence of the correct CTP substrate. To test this, we performed RNA extension assays in the presence of both CTP and ATP. We established a protocol to analyze the RNA products, which contained the additionally incorporated nucleotide, by MALDI mass spectrometry. The method allowed us to distinguish between the RNA product that included the correctly incorporated cytosine and the product that included the misincorporated adenine, although the expected mass difference was only 24 Da (Figure 3.3a).

When the ratio of ATP to CTP was 1:1, mass spectrometry was unable to unambiguously detect adenine misincorporation, since the mass of the misincorporation product is indistinguishable from the mass of the cytosine incorporation product associated with a sodium ion that has a mass of 23 Da (Figure 3.3a). However, when the ratio of ATP over CTP was increased, product RNA that contained the misincorporated adenine was readily detected (Figure 3.3a). When a 1000-fold excess of ATP over CTP was used, essentially all incorporation events were misincorporations. Thus, in the presence of both CTP and ATP, both cytosine and adenine can be incorporated opposite 8oxoG, although the correct cytosine incorporation is more frequent, consistent with the faster rates observed above. These competition assays and the kinetic data show that 8oxoG triggers adenine misincorporation even in the presence of the correct substrate, and suggest that the frequency of misincorporation is in the range of a few percent.

3.2.3   8oxoG-triggered misincorporation results in transcriptional mutagenesis

The above data show that both cytosine and adenine are incorporated by Pol II opposite 8oxoG, but do not exclude that before further elongation the misincorporated adenine is removed by the intrinsic proofreading function of the polymerase. We therefore investigated whether a misincorporated adenine is maintained in the RNA upon elongation. Complex A was incubated with all four NTPs, and the resulting run-off product was analysed by MALDI mass spectrometry. This revealed the expected mass for the correct RNA run-off product, but again was unable to unambiguously detect adenine misincorporation, since the mass of the misincorporation product was indistinguishable from the mass of the cytosine incorporation product associated with a sodium ion (Figure 3.3b). To facilitate the analysis, we aimed to increase the level of adenine misincorporation opposite 8oxoG. Complex A was pre-incubated with ATP, followed by addition of the other three NTPs. Subsequent MALDI analysis showed that the run-off product contained the misincorporated adenine (Figure 3.3b). These data show that 8oxoG can cause adenine misincorporation in the presence of CTP, and the inserted adenine escapes polymerase-intrinsic proofreading and remains in the RNA, leading to transcriptional mutagenesis.
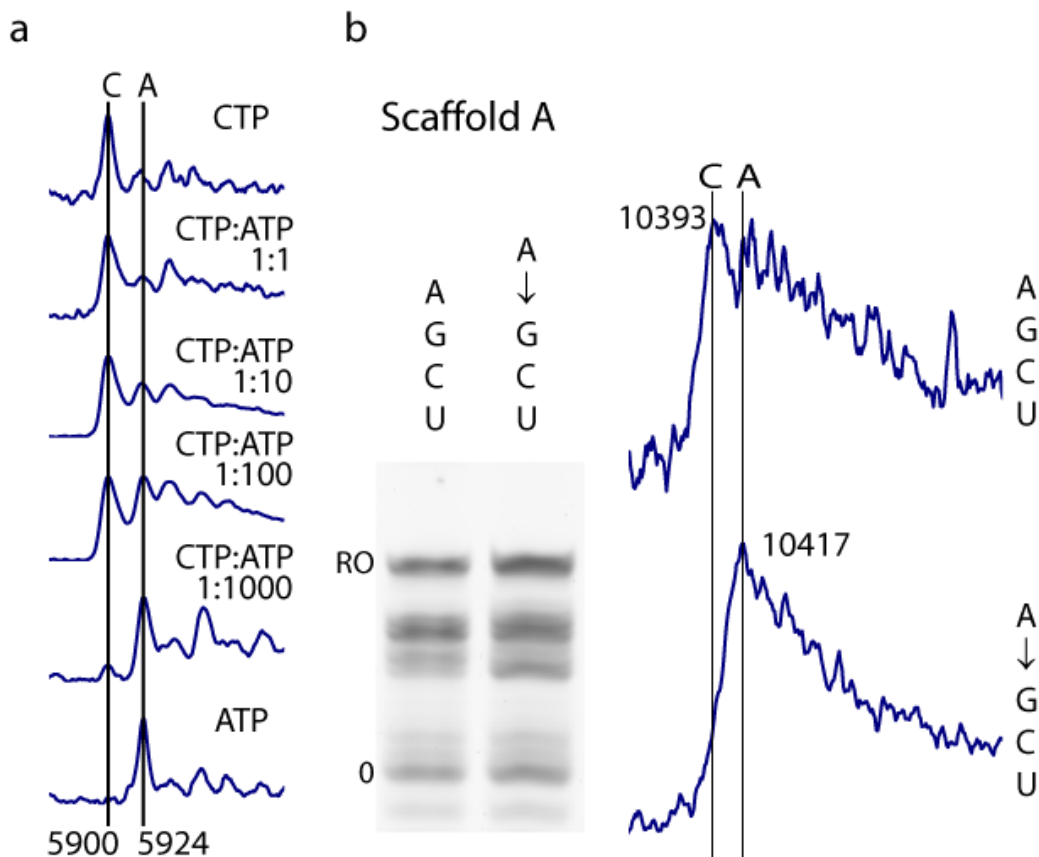
Figure 3.3 Adenine misincorporation can lead to transcriptional mutagenesis.

(a) Complex A was incubated for 5 minutes with mixtures of ATP (1 mM) and CTP at different ratios as indicated. The nucleic acids were desalted and subjected to MALDI mass spectrometry. The observed mass/charge curves are given. The expected mass of the RNA strand was 5900 Da if a cytosine was incorporated and 5924 Da if an adenine was incorporated. Decreasing the amount of competing CTP in the reaction mixture increased the extent of adenine misincorporation.

(b) Misincorporated adenine can remain in a run-off RNA product. On the left, RNA extension assays with complex A are shown. Complex A was incubated with all four NTPs (1 mM) for 60 min or preincubated with ATP for 5 minutes, and then incubated with all four NTPs for 60 min. The run-off product is indicated (RO). The nucleic acids were desalted and subjected to MALDI mass spectrometry. On the right, mass spectrometric detection of the run-off RNA is shown. The calculated mass of the run-off product was 10393 Da if a cytosine was incorporated opposite 8oxoG and 10417 Da if an adenine was misincorporated opposite 8oxoG. Preincubation with ATP increases the amount of mutated RNA run-off product and thus enables direct observation of the mutated RNA.

3.2.4   TFIIS removes a misincorporated adenine.

The above results show that an 8oxoG-adenine mismatch base pair, once formed in the polymerase active site, escapes the intrinsic proofreading function of Pol II. However, since the RNA cleavage-stimulatory factor TFIIS can enhance Pol II proofreading (Jeon and Agarwal, 1996; Thomas et al., 1998), we investigated whether TFIIS can stimulate removal of a misincorporated adenine opposite the 8oxoG lesion. For RNA extension assays in the presence of TFIIS we first used scaffolds B and C, which already contained cytosine or adenine, respectively, at the RNA 3' end opposite the 8oxoG lesion (Figure 3.1b). We added recombinant TFIIS to the obtained ECs B and C, and incubated the mixture for 5 minutes. This led to cleavage of a RNA 3' dinucleotide, which was not observed for undamaged DNA containing a G-C match (scaffold B[undamaged], Figure 3.4). Dinucleotide cleavage generally occurs in the presence of TFIIS (Izban and Luse, 1993), and was also observed for undamaged DNA containing a G-A mismatch (scaffold C[undamaged], Figure 3.4). Upon subsequent incubation with NTPs for 5 minutes, RNA was elongated (Figure 3.4). MALDI analysis of run-off RNA produced by EC C, incubated with both TFIIS and NTPs for 5 minutes, indicated replacement of the RNA 3' adenine by cytosine (data not shown). This showed that TFIIS had stimulated removal of the misincorporated adenine from the RNA prior to elongation. These data suggest that TFIIS suppresses transcriptional mutagenesis by stimulating proofreading, but do not demonstrate proofreading under rapid elongation conditions. During rapid elongation, proofreading apparently does not occur to a large extent, at least with mammalian Pol II (Charlet-Berguerand et al., 2006).
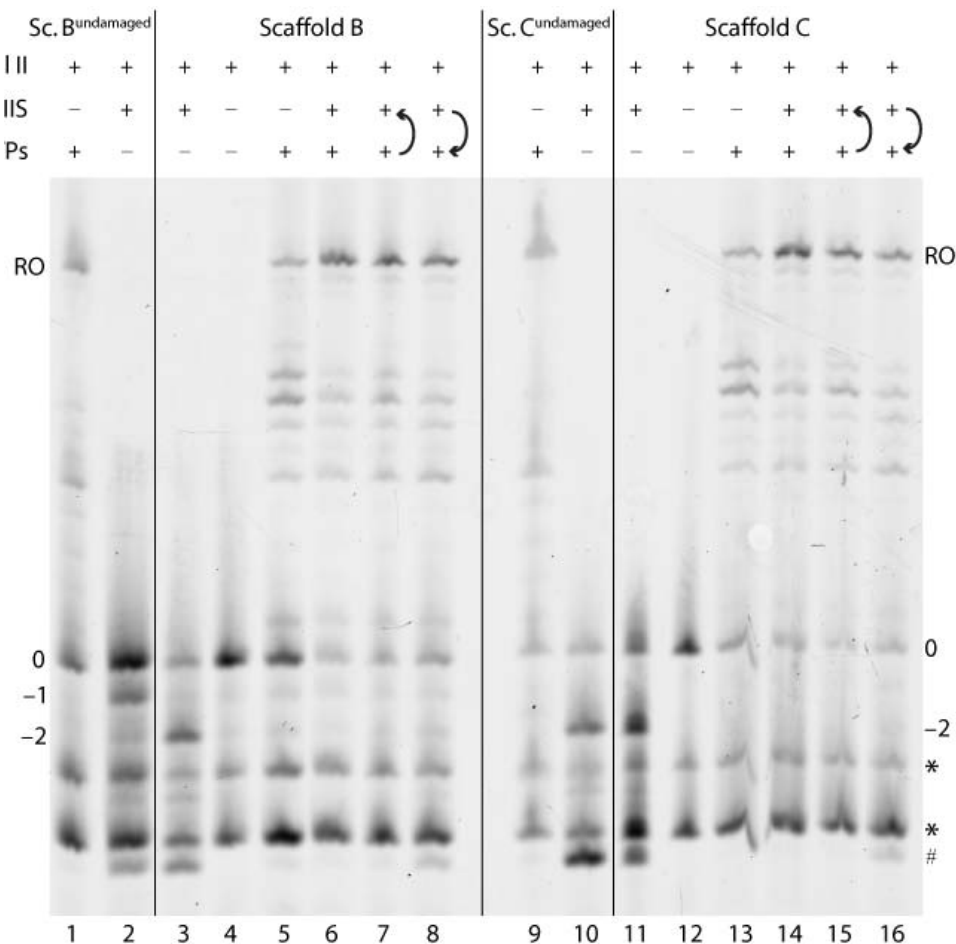
Figure 3.4 Effect of cleavage factor TFIIS on RNA elongation with ECs B and C.

Lanes 4 and 12 show the control experiments where fluorescently labelled RNA was incubated with Pol II and no RNA extension had taken place. In lanes 2 and 10, TFIIS was added for 5 min to undamaged complexes B^{undamaged} and C^{undamaged}, respectively, which contained guanine instead of 8oxoG. In lanes 3 and 11, TFIIS was added for 5 min to complexes B and C, respectively. As expected, cleavage products were observed in lanes 2, 3, 10, and 11 (a product that arises from TFIIS-induced backtracking and cleavage is indicated by #). In other lanes, the scaffolds were incubated with Pol II, NTPs (1 mM) and/or TFIIS for 5 min as indicated. Run-off products are indicated (RO). NTPs and TFIIS were added either simultaneously (lanes 6 and 14) or sequentially as indicated by arrows, with individual incubation times of 5 min (lanes 7, 8, 15, and 16). The stars on the right indicate RNA impurities that were already present in the starting RNA material. The "#" indicates a cleavage product after backtracking.

3.2.5   8oxoG and the misincorporated adenine form a Hoogsteen pair.

To investigate the molecular basis for 8oxoG transcription, we solved crystal structures of Pol II ECs containing 8oxoG (Methods). When we included 8oxoG at position -1 of the DNA template strand (Figure 3.1b, scaffold C), the obtained electron density for the nucleic acids was not interpretable, indicating a mixture of pre- and post-translocation states (Brueckner and Cramer, 2008). An interpretable electron density map was however obtained when we included an 8oxoG-C DNA-RNA base pair at the penultimate position in the hybrid (Figure 3.1b, scaffold D). The structure was solved at 3.7 Å resolution (Table 3.1, Figure 3.5). A bromine atom that had been included in the template strand as a marker gave rise to a single peak in an anomalous difference Fourier map and unambiguously defined the nucleic acid register. The EC was not post-translocated but adopted a state with a frayed RNA 3' end at register +1. This RNA fraying does not influence our analysis as it occurred also in a control structure with undamaged DNA (Figure 3.6). We recently published a detailed structural analysis of RNA fraying and showed that it occurs with different types of terminal RNA nucleotides and that frayed nucleotides can be stabilized by a neighboring mismatch pair (Sydow et al., 2009). The overall structure was generally similar to that of the undamaged complete Pol II EC (Kettenberger et al., 2004) (Figure 3.5a, b). The 8oxoG-C base pair adopted standard Watson-Crick geometry at register -1, consistent with normal cytosine incorporation opposite guanine (Figure 3.5c). An additional structure was solved at 3.9 Å resolution with the 8oxoG as above, but containing an adenine, instead of cytosine, in the RNA opposite the lesion (Figure 3.1b, Figure 3.5, scaffold E,Table 3.1). The unbiased difference electron density map for the nucleic acids showed a distinct density for the nascent base pair that did not adopt Watson-Crick geometry. When the 8oxoG residue was placed in the standard *anti* conformation it did not fit the density. However, when it was rotated by 180 degrees and placed in a *syn* conformation, it fitted the density very well and was in a good position to form a Hoogsteen base pair with the adenine in the opposite RNA strand (Figure 3.5c). Taken together, the structural analyses showed that 8oxoG can form a Watson-Crick base pair with a correctly incorporated cytosine, but can also form a Hoogsteen base pair with a misincorporated adenine (Figure 3.5c).
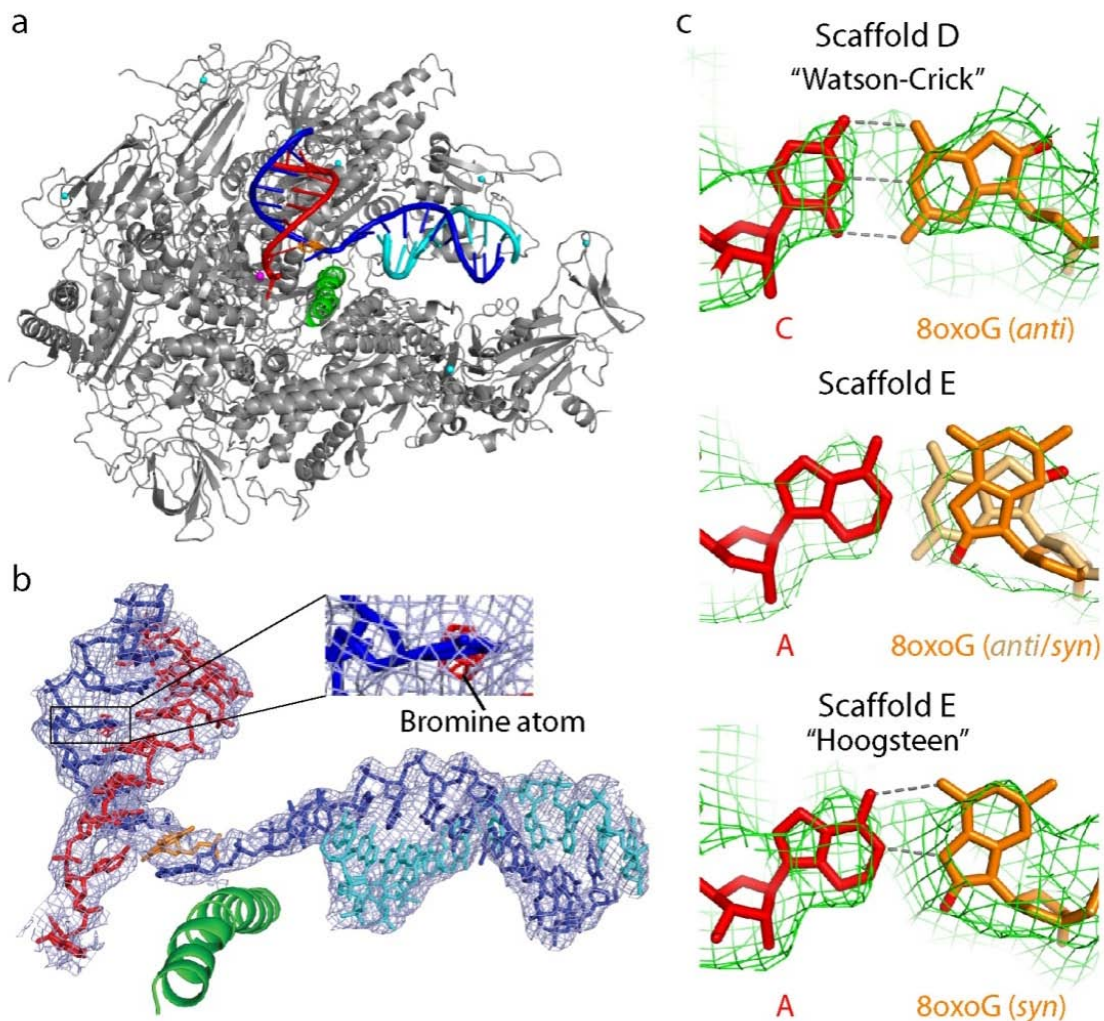
Figure 3.5 Structures of 8oxoG-containing Pol II ECs

(a) Overview of the structure of the 8oxoG-containing Pol II EC D. Pol II is shown as a ribbon model in silver, with the bridge helix highlighted in green. The nucleic acids are in colour, and 8oxoG is shown as a stick model in orange. A large portion of Pol II was omitted for clarity.

(b) Structure of nucleic acids in the Pol II EC E. The final $2F_o$-F_c electron density map is shown for the nucleic acids (blue, contoured at 1.0σ). The anomalous difference Fourier map reveals the location of the bromine atom as indicated (red, contoured at 3.5σ).

(c) 8oxoG-containing base pairs observed in EC structures. On the top, the Watson-Crick 8oxoG-C base pair observed in EC D is shown. On the bottom, the 8oxoG-A Hoogsteen base pair observed in EC E is depicted. The unbiased F_o-F_c difference electron density map is shown as a green mesh contoured at 2.7σ. Hydrogen bonds are indicated with dashed lines.
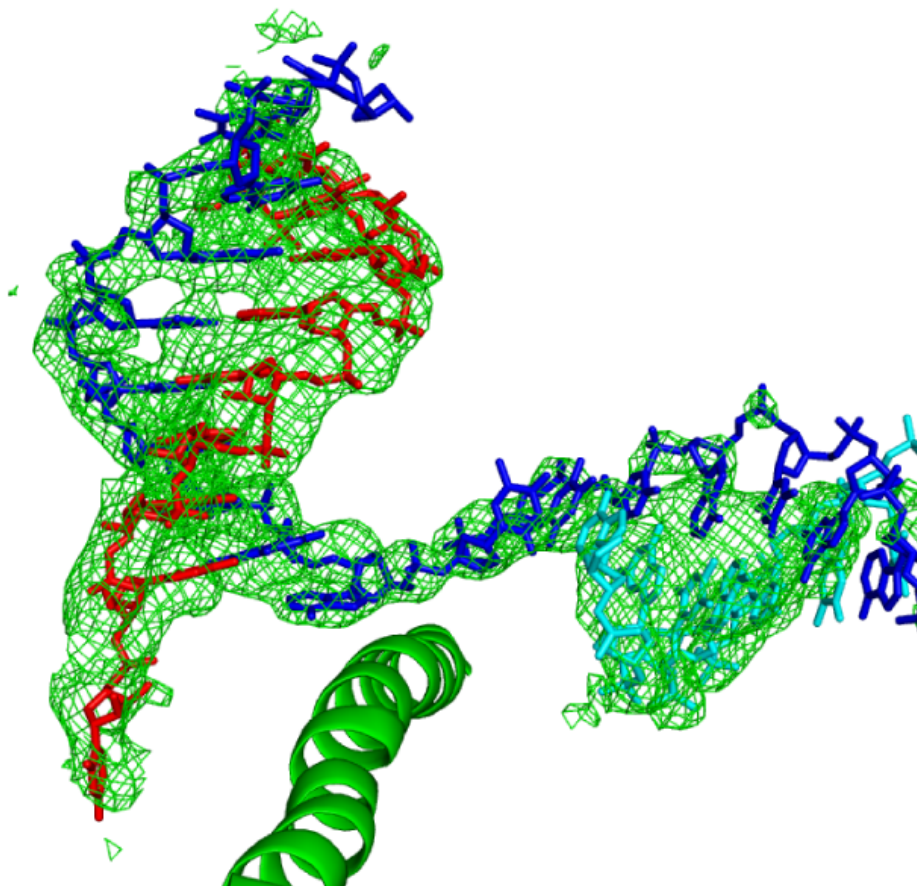
Figure 3.6 Control structure of an undamaged Pol II elongation complex.

The density map of undamaged Pol II elongation complex D (scaffold D^{undamaged}). The difference F$_O$-F$_C$ density map of the nucleic acids is shown (green, contoured at 3.0σ). The model of complex D is shown. The frayed RNA which was found in the structure of complex D, can also be seen for the undamaged structure.

Table 3.1 Crystallographic data and refinement statistics for complete Pol II elongation complexes (ECs)

Data collection	EC D	EC E
Space group	$C222_1$	$C222_1$
Unit cell axes (Å)	220.6, 392.0, 281.5	221.2, 394.1, 282.3
Wavelength (Å)	0.9188	0.9186
Resolution range (Å)	50.0-3.7 (3.8-3.7)[2]	50.0-3.9 (4.0-3.9)[2]
Unique reflections	250,702 (19,429)	216,612 (15,791)
Completeness (%)	99.6 (99.9)	99.7 (99.8)
R_{sym} (%)	8.2 (56.7)	7.8 (86.3)
$I/\sigma(I)$	11.9 (2.7)	13.4 (2.6)
Redundancy	3.8 (3.8)	3.9 (4.0)
Refinement		
Nonhydrogen atoms	32,348	32,298
RMSD bonds (Å)	0.008	0.008
RMSD angles (°)	1.50	1.48
Peak in anomalous difference Fourier (σ)	6.7	5.1
R_{cryst} (%)	22.5	22.8
R_{free} (%)	25.8	26.6

[1]Diffraction data were collected at the Swiss Light Source beamline PX1 (X06SA) and were processed with program XDS (Kabsch, 1993).

[2]Numbers in parenthesis correspond to the highest resolution shells.

3.3   Discussion

Here we presented structural and functional studies of 8oxoG-containing Pol II ECs to investigate the detailed mechanism of 8oxoG lesion processing during gene transcription. We have shown that yeast Pol II can misincorporate adenine at an 8oxoG lesion in the DNA template strand, that the misincorporated adenine forms a Hoogsteen base pair with the lesion at the polymerase active center, and that the misincorporated adenine escapes polymerase-intrinsic proofreading and remains in the RNA as a product of transcriptional mutagenesis.

How may misincorporation and proofreading occur? In incoming DNA, 8oxoG likely adopts a standard *anti* conformation within a Watson-Crick base pair, but its base must rotate

by 180º to the *syn* conformation in the 8oxoG-A DNA-RNA Hoogsteen base pair at the active center. Modelling shows that base rotation in the templating position at register +1 would result in a clash with the bridge helix and the DNA base at register -1. We therefore propose that base rotation occurs during 8oxoG translocation from the downstream position +2 to the templating position +1 in the recently described state of the EC that was suggested to be an intermediate of translocation (Brueckner and Cramer, 2008). DNA translocation from +2 to +1 apparently passes through an intermediate state, which contains the unpaired template base in a pre-templating position (Brueckner and Cramer, 2008). In this position the 8oxoG base would be unpaired and free to rotate from *anti* to *syn* before being loaded into the templating position. The loaded 8oxoG in *syn* conformation then triggers adenine misincorporation by Hoogsteen base pair formation with an incoming ATP substrate. The resulting 8oxoG-A base pair mimics a T-A base pair and is accommodated in the polymerase active center and in the DNA-RNA hybrid, leading to an escape from intrinsic proofreading and enabling synthesis of a mutated RNA and thus transcriptional mutagenesis. TFIIS induces backtracking of 8oxoG from the active center to downstream DNA. When repeated translocation preserves the *anti* conformation, correct cytosine incorporation occurs, and RNA is elongated error-free.

Transcriptional mutagenesis as demonstrated here *in vitro* may be a threat *in vivo*, published results indicate that the molecular mechanism of transcriptional mutagenesis described here is relevant *in vivo*. In particular, it was reported that a yeast strain lacking *dst1*, the gene encoding TFIIS, is sensitive to oxidative stress (Koyama et al., 2003; Koyama et al., 2007). Our study suggests that this sensitivity stems from the accumulation of mutant RNA transcripts, which result from the failure of the intrinsic Pol II proofreading activity to remove adenine that was misincorporated opposite 8oxoG. An additional contribution to the sensitivity to oxidative stress may come from conversion of GTP to 8oxoGTP and subsequent incorporation of 8oxoG into RNA opposite a templating adenine (Koyama et al., 2003). In *E. coli*, 8oxoG-triggered transcriptional mutagenesis also occurs *in vivo* (Bregeon et al., 2003).

Comparison of our results with previously published biochemical studies of 8oxoG processing reveals that the ratio of 8oxoG-dependent incorporation of adenine over cytosine differs for different RNA polymerases. We showed here that *S. cerevisiae* Pol II mainly incorporates cytosine opposite 8oxoG but can also misincorporate adenine to a low extent. Cytosine and adenine are however incorporated at similar levels by mammalian Pol II (Kuraoka et al., 2003; Yamaguchi et al., 2007) and by *E. coli* RNA polymerase (Viswanathan and Doetsch, 1998), whereas the RNA polymerase of bacteriophage T7 even prefers to incorporate adenine opposite 8oxoG (Chen and Bogenhagen, 1993).

The investigation of three different DNA template lesions with Pol II allows us to compare their effect on transcription and the mechanisms involved in each case. A UV-induced thymine-thymine cyclobutane pyrimidine dimer (CPD) lesion can enter the polymerase active site and trigger uridine misincorporation, which leads to transcriptional stalling (Brueckner et al., 2007). In contrast, a cisplatin-induced 1,2-guanine-guanine crosslink lesion stalls Pol II because it cannot be translocated from the downstream DNA into the active site, resulting in nontemplated adenine misincorporation (Damsma et al., 2007). Thus, misincorporation occurs at all three lesions studied thus far, but for different reasons and with different consequences. At a CPD, misincorporation is required for stalling (Brueckner et al., 2007). At the cisplatin crosslink, the lesion alone can stall Pol II, and misincorporation is a result of stalling (Damsma et al., 2007). At 8oxoG, misincorporation occurs in addition to correct incorporation, but does not stall the polymerase and rather leads to transcriptional mutagenesis when TFIIS is unavailable. Thus, as predicted (Brueckner et al., 2007; Damsma et al., 2007), the mechanisms of lesion processing differ for all lesions analyzed in mechanistic detail thus far.

Like RNA polymerases (Chen and Bogenhagen, 1993; Kuraoka et al., 2007; Viswanathan and Doetsch, 1998; Yamaguchi et al., 2007), DNA polymerases can incorporate cytosine or adenine opposite 8oxoG to various extents (Broyde et al., 2008; Hsu et al., 2004; Maga et al., 2007; Rechkoblit et al., 2006). High-fidelity DNA polymerases accommodate and bypass an 8oxoG-C base pair (Brieba et al., 2004; Freisinger et al., 2004; Hsu et al., 2004; Krahn et al., 2003). One structure of a high-fidelity DNA polymerase revealed an 8oxoG-A Hoogsteen base pair (Hsu et al., 2004). This suggests that the same mechanism of adenine misincorporation at 8oxoG occurs during replication and transcription, and that stable Hoogsteen pairing can occur within nucleic acid polymerases even though the active centers are structurally complementary to very different B- and A-form template-product duplexes in these structurally unrelated DNA and RNA polymerases, respectively.

4   Structure of the archaeal Spt4/5 core complex

4.1   Introduction

In all living organisms, transcription of protein-encoding genes is mediated by a common multiprotein machinery (Allison et al., 1985). There are numerous factors that regulate transcription by influencing the ability of RNA polymerase to access, bind, and transcribe specific genes in response to appropriate signals. Prokaryotic N-utilization substance G (NusG) is a well-conserved regulator of gene expression, which has a complex role in transcription elongation. NusG accelerates the elongation rate of *E. coli* RNA polymerase by suppression of specific polymerase pause sites (Burova et al., 1995). NusG was also found to facilitate Rho-dependent transcription termination by bridging Rho to RNA polymerase (Nehrke and Platt, 1994; Sullivan and Gottesman, 1992). Furthermore, NusG is involved in anti-termination of transcription, which is induced by the N protein of bacteriophage λ (Li et al., 1992). NusG proteins have an N-terminal NusG (NGN) domain and a C-terminal KOW (Kyprides, Ouzounis, Woese) domain, a nucleic-acid binding domain also found in some ribosomal proteins (Ponting, 2002; Steiner et al., 2002).
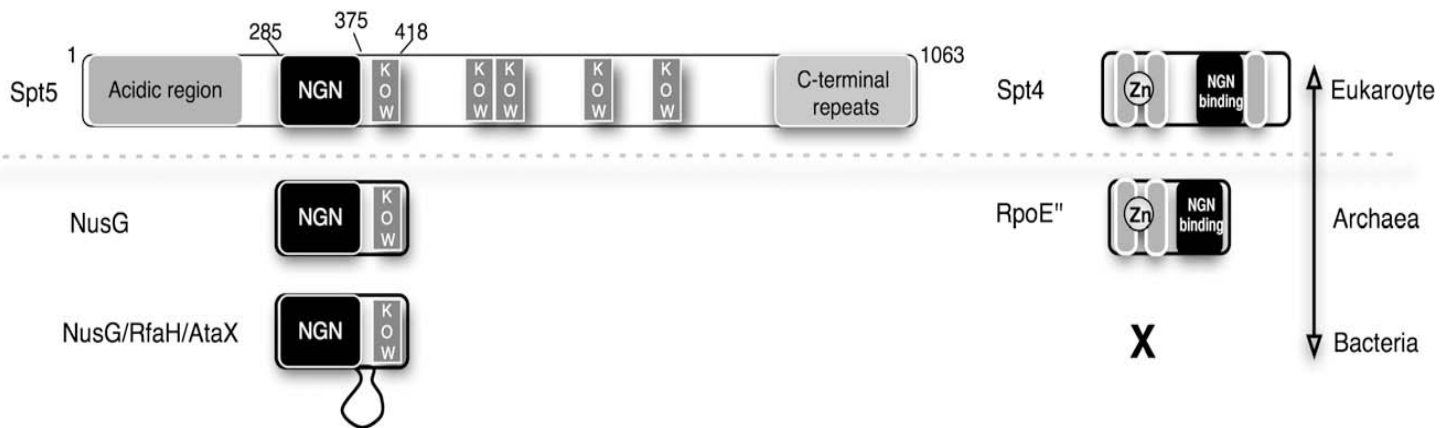
Figure 4.1 Overview of the eukaryotic Spt4, Spt5 and the prokaryotic NusG, RpoE".
Conserved motifs are shown as boxes. The C-terminal part of Spt4 is absent in RpoE". Residue
numbers refer to *S. cerevisiae* Spt5. Adopted from (Guo et al., 2008).

The eukaryotic ortholog of NusG is Spt5. Spt5 has an acidic N-terminus, a single NGN domain, five or six KOW domains and a C-terminal sequence with a set of simple repeats that are targets of regulatory kinases (Figure 4.1)(Ponting, 2002; Saunders et al., 2006). Spt4 and Spt5 together, form the stable heterodimer "DRB Sensitivity Inducing Factor" (DSIF) (Wada et al., 1998), which interacts with Pol II (Hartzog et al., 1998; Yamaguchi et al., 1999b) and is involved in inhibiting or activating transcription elongation depending on the existence of Negative Elongation Factor (NELF) (Cheng and Price, 2007; Yamaguchi et al., 1999a). It was also shown that DSIF is involved in rRNA transcription, since mutations in DSIF caused defects of Pol I function (Schneider et al., 2006). Another function of Spt5 is to suppress senescence and apoptosis, which is exerted through the Spt5 association with Spt4 and Pol II (Komori et al., 2009).

Recently, the structure of an archaeal Spt5NGN domain (*Methanocaldococcus jannaschii* Spt5NGN) was published (Zhou et al., 2009). Also the yeast structure of the Spt4/5 heterodimer was published as a fusion protein (Guo et al., 2008). To further investigate the Spt4/5 core complex from the archaea *M. jannaschii* we elucidated the structure of the first genuine Spt4/5 core complex at 1.9 Å resolution.

4.2   Results & Discussion

4.2.1   Solving the structure of the archaeal Spt4/5 core complex

In collaboration with Dr. Finn Werner (University College London), crystals were obtained from the core Spt4/5 complex from the archaea *Methanocaldococcus jannaschii* (Methods) and exposed to synchrotron radiation (X06SA, SLS). All the crystals obtained were very thin plates, resulting in poorly defined diffraction spots when exposed through the shortest axis of the plate, and often resulted in datasets that were less than 100% complete. Subsequently, the anomalous signal from the predicted protein-bound zinc atom was too low for experimental phasing. Therefore, several attempts were made to solve the structure of the archaeal Spt4/5 core complex using molecular replacement with the *S. cerevisiae* structure as a search model (Guo et al., 2008). Although these attempts initially failed, publication of the structure of the core Spt5 from *Methanocaldococcus jannaschii* (Zhou et al., 2009) provided a better search model and made molecular replacement possible (Methods). The structure was build and refined using the programs Coot (Emsley and Cowtan, 2004), Refmac (Vagin et al., 2004) and phenix.refine (Afonine et al., 2005). The final resolution was 1.9 Å (Table 4.1).

Table 4.1: Crystallographic data and refinement statistics for core Spt4/5

Data collection	
Space group	P21
Unit cell axes (Å)	64.4 64.0 83.0
	90.0 107.0 90.0
Wavelength (Å)	1.000
Resolution range (Å)	44.4 – 1.9 (2.0 – 1.9)
Unique reflections	47,050
Completeness (%)	92.1% (86.4%)
R_{sym} (%)	4.6 (59.6)
$I/\sigma(I)$	9.65 (1.40)
Redundancy	1.9
Refinement	
Nonhydrogen atoms	4745
RMSD bonds (Å)	0.008
RMSD angles (°)	1.039
R_{cryst} (%)	20.8
R_{free} (%)	25.5

[1]Diffraction data were collected at the Swiss Light Source beamline PX1 (X06SA) and were processed with program XDS (Kabsch, 1993).

[2]Numbers in parenthesis correspond to the highest resolution shells.

4.2.2   Details concerning the structure of the archaeal Spt4/5 core complex

The Spt5NGN domain consists of a four-stranded antiparallel beta sheet flanked by three alpha helices (Figure 4.2). Spt4 has four beta strands in its center, with an alpha helix at the edge of the sheet. Spt4 contains a zinc finger, formed by four conserved cysteine residues and is located at the opposite side of the beta sheet. Zinc binding was confirmed by the X-ray absorption edge scan of Spt4/Spt5NGN crystals (Figure 4.3). Mutations altering any of the four finger cysteines of *S. cerevisiae* Spt4 cause loss-of-function phenotypes and in several instances greatly decrease Spt4 levels (Basrai et al., 1996; Malone et al., 1993). Thus, it is likely that the zinc finger is an essential structural element of Spt4 (Guo et al., 2008).
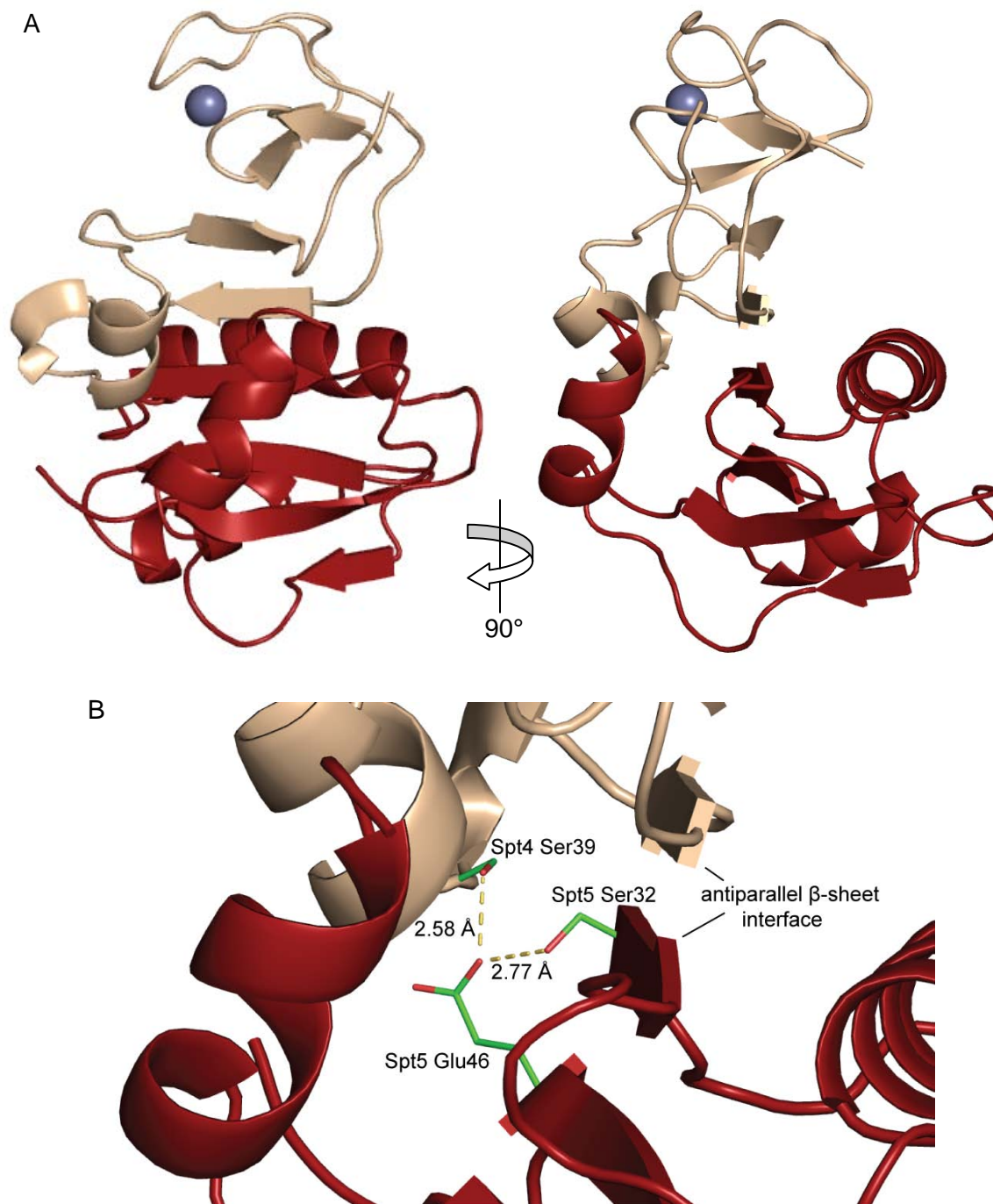
Figure 4.2 Structure of the Spt4/5 core complex from *M. jannaschii*.

(a) The structure is shown as a cartoon representation, with Spt4 in wheat, Spt5 in red, and the zinc atom is shown as a blue sphere.

(b) Close-up of the Spt4-Spt5NGN interface. The view is as in the right part of (A).
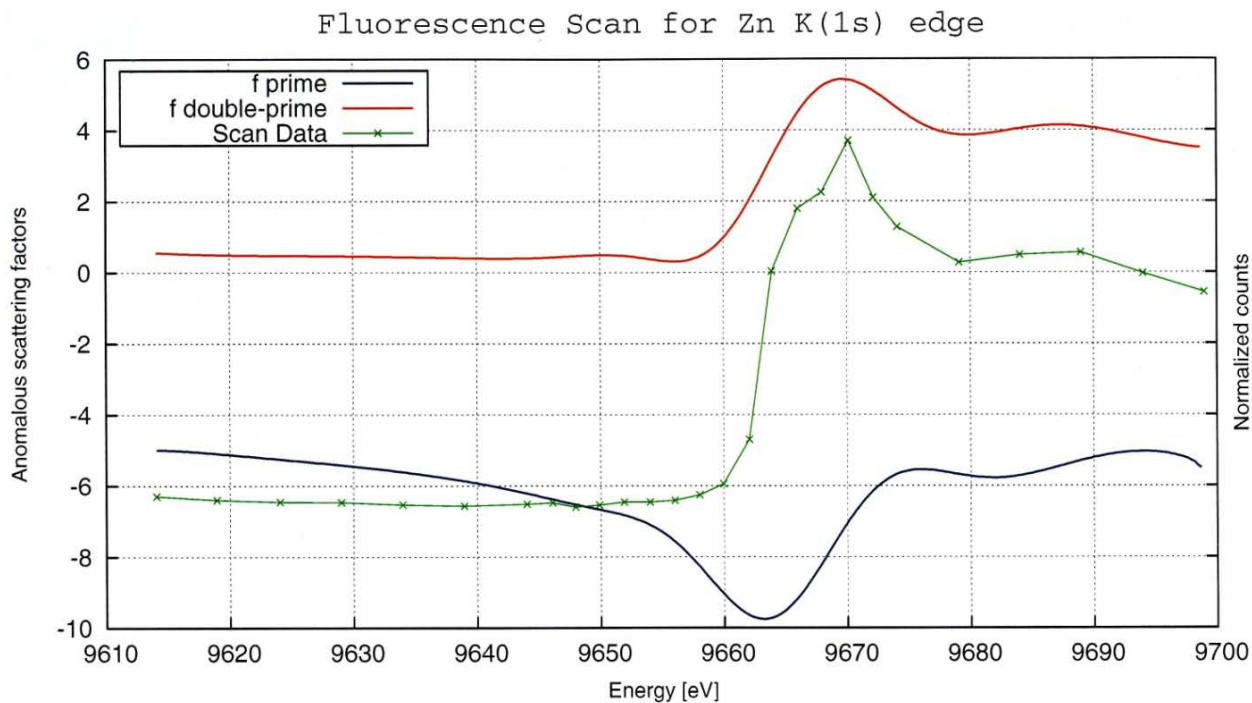
Figure 4.3 X-ray absorption edge scan of Spt4/5NGN crystals.
The peak around 9670 eV corresponds to the K absorption edge of Zn (theoretical value 9659 eV).

The Spt4-Spt5 heterodimer consists of an 8-stranded antiparallel β sheet formed by the four stranded β sheet of Spt4 and the four stranded β sheet of Spt5, aligned edge to edge (Figure 4.2). In addition to this β sheet interface, several hydrogen bonds contribute to the specificity of Spt4-Spt5NGN interactions. In particular, Ser39 from Spt4 forms a conserved hydrogen-bonding interaction with Glu46 in Spt5NGN (Figure 4.2b). This hydrogen bond fixes Glu46, which results in a strong acid-α-helix dipole interaction (Guo et al., 2008; Nicholson et al., 1988). Thus, Spt4 and Spt5NGN interact via hydrophobic interfaces that are held in register by polar interactions and alignment of their β sheets.

4.2.3   Structural conservation of Spt4 and Spt5NGN domain

Comparison of the sequence and structure of the archaeal Spt5NGN domain with those of bacterial NusG and of eukaryotic Spt5NGN reveals conservations between bacteria, archaea and eukaryotes (Figure 4.4). Comparison of the structure of Spt5NGN of *M. jannaschii*, *E. coli* and *S. cerevisiae* also showed this conservation (Figure 4.5). The Spt5 protein is found in all three kingdoms of life and is therefore one of the most ancient

components of the transcriptional machinery (Ciampi, 2006; Guo et al., 2008; Harris et al., 2003; Ponting, 2002). The structural similarities of the NGN domain in archaea, bacteria and eukaryotes imply that they also share similar mechanisms of transcription regulation.

Comparison of the sequence of archaeal Spt4 and eukaryotic Spt4 reveals conservation between archaea and eukaryotes. In contrast to Spt5, there are no Spt4 sequence homologs in bacteria. Comparison of the structure of Spt4 of *M. jannaschii* and *S. cerevisiae* also showed conservation, for example the zinc finger is highly conserved (Figure 4.6). The *S. cerevisiae* Spt4 is also larger; the extra residues form helices and lie adjacent to the zinc finger, which is located on the opposite edge of the sheet that forms the Spt4/Spt5NGN binding interface, hence maintaining the interface in both structures. The lack of these helices in *M. jannaschii* Spt4 exposes the zinc finger although, whether the zinc finger has a nucleic-acid binding ability in the transcription elongation complex needs to be studied further. When comparing the sequence of Spt4 and Spt5NGN the binding interface is conserved between eukaryotes and archaea (shaded area in Figure 4.4). For bacteria however there is no corresponding binding partner, although the bacterial Spt5NGN does have an appendant domain, which folds into a hairpin at the position close to where Spt4 locates (Figure 4.5). This appendant domain was suggested to work as a structural element, such as a post or a stack, rather than as a site for residue-specific contacts in *E. coli* Spt5 (Richardson and Richardson, 2005; Zhou et al., 2009). Whether Spt4 and the appendant domain have a similar function remains to be investigated.

Spt4

```
                              10                20                30                40                50
Mj  -------------MRACLKCKYLTND-----EICPICH------SP-TSENWIGLLIVINPEKSEIAKKAGID--IKGKYALSVKE------------------
Af  -------------MAELACKNCRFINVDT---NICRNCG------STDLTKEWYGYVVIVDPEKSEIAKRLEIK--IPGKYALRV-------------------
Tv  ----------MKVQYRACKKCKRLTPE-----KTCPIHP------DEKTTTEWFGFLIITEPEFSAIAKRAGIS--EPGMYAIKVRQ----------------
Tg  --------------MVKERACRHCHYITTE---DRCPVCG-----SRDLSDEWFDLVIIIDPEKSRIAKKLGVK--VPGKYAIRVR----------------
Pf  --------------MSEKACRHCHYITSE----DRCPVCG-----SRDLSEEWFDLVIIVDVENSEIAKKIGAK--VPGKYAIRVR----------------

Sc  -----MSSERACMLCGIVQTTNEFNRDGCPNCQGIFEEAG--VSTMECTSPSFEGLVGMCKPTKSWVAKWLSVDHSIAGMYAIKVDGRLPAEVVELLPH-----YKPRDGS
Kl  -----MSTDRACMLCGLVQSTAEFNRNGCPNCQSIFEEAG--VSAVECTSPSFEGLVGMCKPSRSWVARWMSIDSYIPGMYAVKIDGRLPIEVTELLPH----YKPRDGS
Hs  MALETVPKDIRHLRACLLCSLVKTIDQFEYDGCDNCDAYLQMKGNREMVYDCTSSSFDGIIAMMSPEDSWVSKWQRVSNFKPGVYAVSVTGRLPQGIVRELKSRGVAYKSRD--
Dr  MSLETVPKDLRHLRACLLCSLVKTIDQFEYDGCDNCESYLQMKGNREMVYECTSSSFDGVIAMMSPEDSWVAKWQRIGNFKPGVYAVTVTGRLPPGVVRELKSRGVIYRSRD--
Dm  MAFDAIPKDLRGLRACLVCSLVKTFDQFETDGCENCEEFLRMKNNKDNVYDHTSNNFDGIIALTTPTDSWVAKWQRLSRFTRGIYAISVSGTLPQSTIREMKNRGIAYKTRDRS
```

Zinc Finger

Spt4-Spt5NGN interface

Spt5NGN

```
                                      10                      20                30                40                50                60                70                80
Mj 1    -MIFAVRTMVGQEKNIAGLMASRAEKEQ------------LDVYSILASES--------------LKGYVLVEAETKGDVEELIKGMPRVRGI------VPGTIAIEEIEPLLTP-  82
Af 2    SRFFAVRTTANQERVVANLMDMAARKHN------------LKVYSILAPKE--------------LKGYIIVEAETPDDLLKAIRGLPHVKGV------VKGEISFSEIEHFLTPK  85
Tv 134  PVIVALKTTVGSEVQVATDLNNKAERDRDERSQGGEKQPSEVFAIMAPYE-----------IKGYIFVETMHPDRISYLAKDIRGYKGM------VSGKIDLEEIAHYLTPK  228
Tg 4    GKIFTVRVTVGQQEETTAKLIYSKVKTYN------------LPVYAIIAPSK--------------VKGYIFVEAPSKSAVDEAIKGIRHARGT------LPGEVKFEEIEHFLEEK  87
Pf 3    GKIFAVRVTHGQEETTAKLIYSKVRTYN------------LPIYAILAPSR--------------VKGYIFVEAPNKGVVDEAIRGIRHARGV------LPGEVPFKEIEHFLEEK  86

Sc 285  ATIWGVRCRPGKEKELIRKLIKKKFNLDRAMG-------KKKLKILSIFQRDN--------------YTGRIYIEAPKQSVIEKFCNGVPDIYIS-----QKLLIPVQELPLLIKPN  375
Kl 243  AIIWSVRCRPGKEKELVRKLLNKKFNLDKSMG-------SKKLKILSIFQRDN--------------YTGRIYIEAPKQSVIEKFVNGVPDVYSN-----QKLLIPVQELPLLLKPT  333
Hs 177  PNLWTVKCKIGEERATAISLMR-KF-IAYQFT-------DTPLQIKSVVAPEH--------------VKGYIVEAYKQTHVKQAIEGVGNLRLGYW-----NQQMVPIKEMTDVLKVV  267
Dr 176  PNLWTVKCKIGEERATAISLMR-KF-VAYQCT-------DTPLQIKSVVAPEH--------------VKGYIVEAYKQTHVKAAIEGVGNLRMGFW-----NQQMVPIKEMTDVLKVV  266
Dm 214  PNLWMVKCRIGEEKATALLLMR-KF-LTYLNT-------DDPLQIKSIIAPEG--------------VKGYIYLEAYKQTHVKTAIDNVGNLRMGKW-----KQEMVPIKEMTDVLKVV  304

Ec 7    KRWYVVQAFSGFEGRVATSLREHIKLHNMEDLFG--EVMVPTEEVVEIRGGQRRKSERKFFPGYVLVQMVMNDASWHLVRSVPRVMGFIGGTSTDRPAPISDKEVDAIMNRL  115
Pp 7    KRWYVVQAFSGFEGRVAQSLREHIKMHSMEDYFD--EVILVPTEEVVEMRAGQRRKSERKFFPGYVLVQMVMNDQTWHLVRSIPRVMGFIGGTSDRPAPISDKEADAIINRL  115
Hp 16   MRWFVLQAFSGFEARVAMTLREYIKLHNMQDQFG--EVILVPTEEVVENVGGRRRKTERKFFPGYVLVQMEMNDDTWHLVKSVPRVMGFIGGTADKPAPISQSEEAERIINRV  124
```

Figure 4.4: Alignment of Spt4 and Spt5 homologs.

Residues that are part of the binding interface between Spt4 and Spt5NGN are noted with a shaded box. Highly conserved residues are depicted in bold. Mj, *Methanocaldococcus jannaschii;* **Af,** *Archaeoglobus fulgidus;* **Tv,** *Thermoplasma volcanium;* **Tg,** *Thermococcus gammatolerans;* **Pf,** *Pyrococcus Furiosus;* **Sc,** *Saccheromyces cerevisiae;* **Kl,** *Kluyveromyces lactis;* **Hs,** *homo sapiens;* **Dr,** *Danio rerio;* **Dm,** *Drosophila mojavensis;* **Ec,** *Escherichia coli;* **Pp,** *Photobacterium profundum;* **Hp,** *Haemophilus parasuis.*
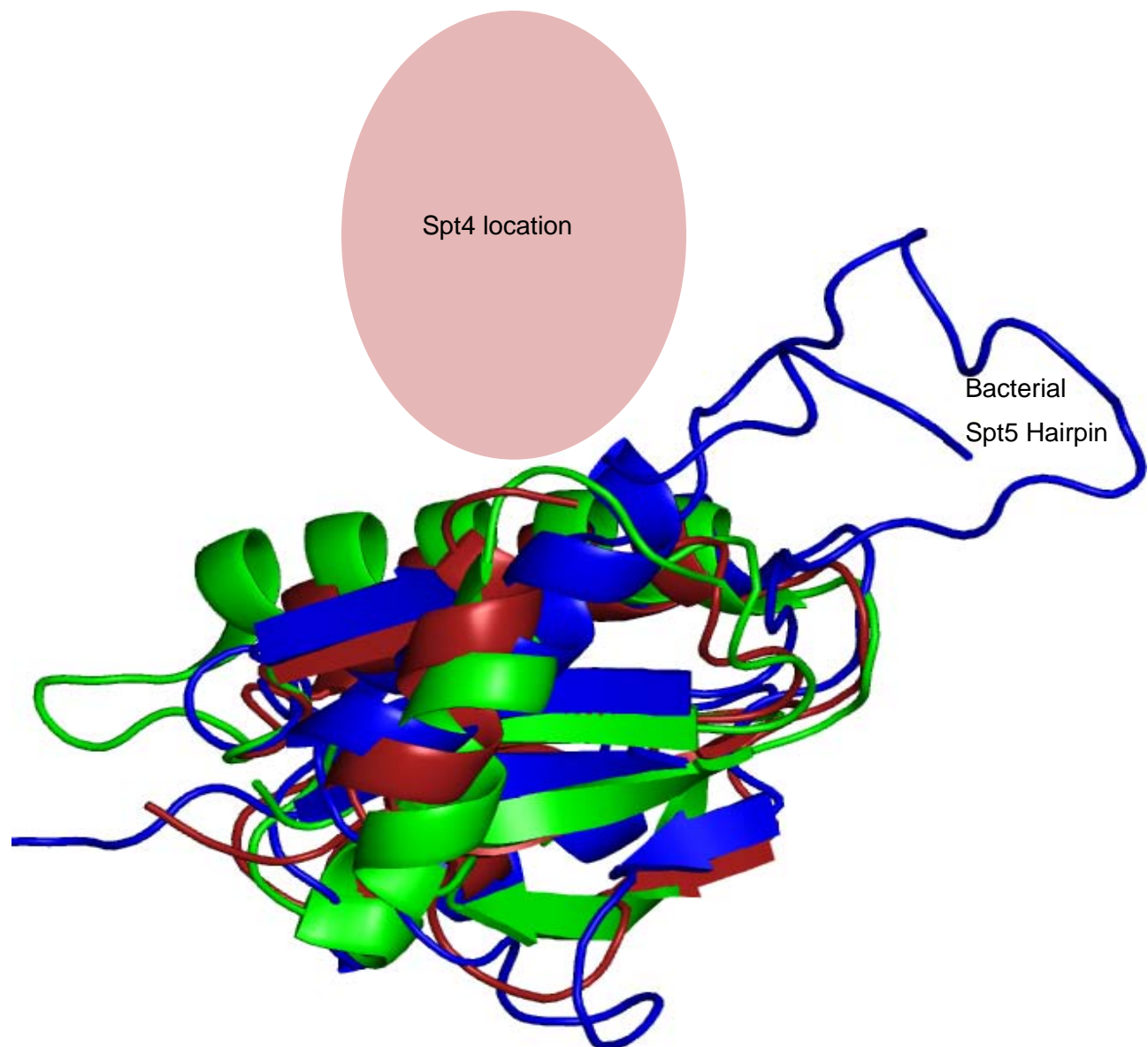
Figure 4.5 Structural superimposition of the Spt5NGN domain.
The *M. jannaschii* Spt5NGN domain is shown in firebrick. The *E.coli* NusG (pdb-code 2OUG) is
shown in blue, and the *S. cerevisiae* Spt5NGN domain (pdb-code 2EXU) is shown in green.
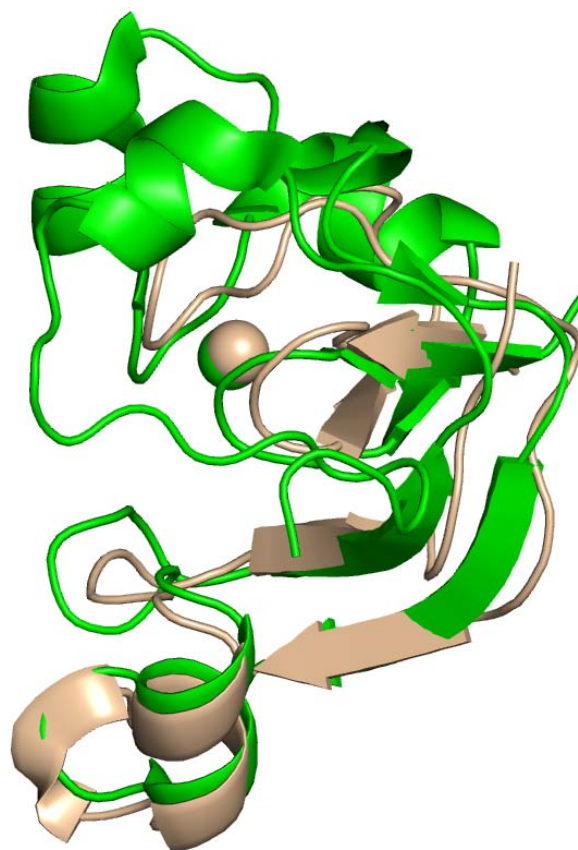
Figure 4.6 Structural superimposition of Spt4.
The *M. jannaschii* Spt4 is shown in wheat. The *S. cerevisiae* Spt4 (pdb-code 2EXU) is shown in green.

4.2.4   Surface analysis of the hydrophobic patch

The Spt5NGN domain was suggested to bind the largest subunit of RNA polymerase II (Hartzog et al., 1998). Based on analogy to *E. coli* RfaH (Belogurov et al., 2007) and on bacterial two-hybrid interaction assays (Nickels, 2009), the *E. coli* Spt5NGN domain is proposed to bind RNA polymerase via interaction with the loop between two antiparallel helices in the β' clamp domain (the β' clamp helices; β' CH, residues 260–309 in *E. coli* RNA polymerase) (Mooney et al., 2009). A hydrophobic pocket on the NGN domain is the proposed site of β' CH interaction (Belogurov et al., 2007).

Surface analysis of the *M. jannaschii* Spt5NGN structure also revealed this hydrophobic pocket in the center of the protein, and this hydrophobic pocket is structurally

conserved in the *E. coli* and *S. cerevisiae* Spt5NGN proteins (Figure 4.7) (Mooney et al., 2009).

Since the *E. coli* Spt5NGN domain is proposed to bind RNA polymerase via interaction with the loop between two antiparallel helices in the β' clamp domain (Mooney et al., 2009) it is proposed that *M. jannaschii* Spt5NGN will bind at the corresponding loop of the archaeal RNA polymerase. Since the structure of the *M. jannaschii* RNA polymerase is unknown, we used the structure from the archaea *Sulfolobus shibatae* to indicate the putative binding site of *M. jannaschii* Spt5 (Figure 4.8). This proposal however, remains to be tested experimentally.

M. jannaschii   *S. cerevisiae*   *E. coli*
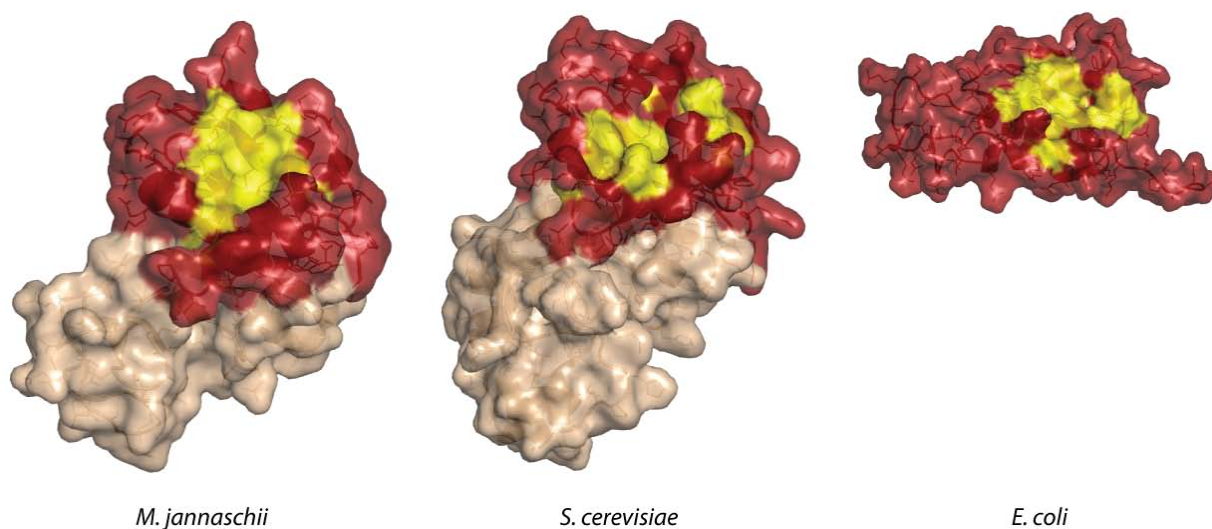
Figure 4.7 Surface representation of the Spt4/5 core structures from the indicated species.
The hydrophobic residues defining the pocket are colored yellow. Spt5NGN domains are in red and Spt4 is in wheat. The hydrophobic residues from *S. cerevisiae* and from *E. coli* are selected according to (Mooney et al., 2009).
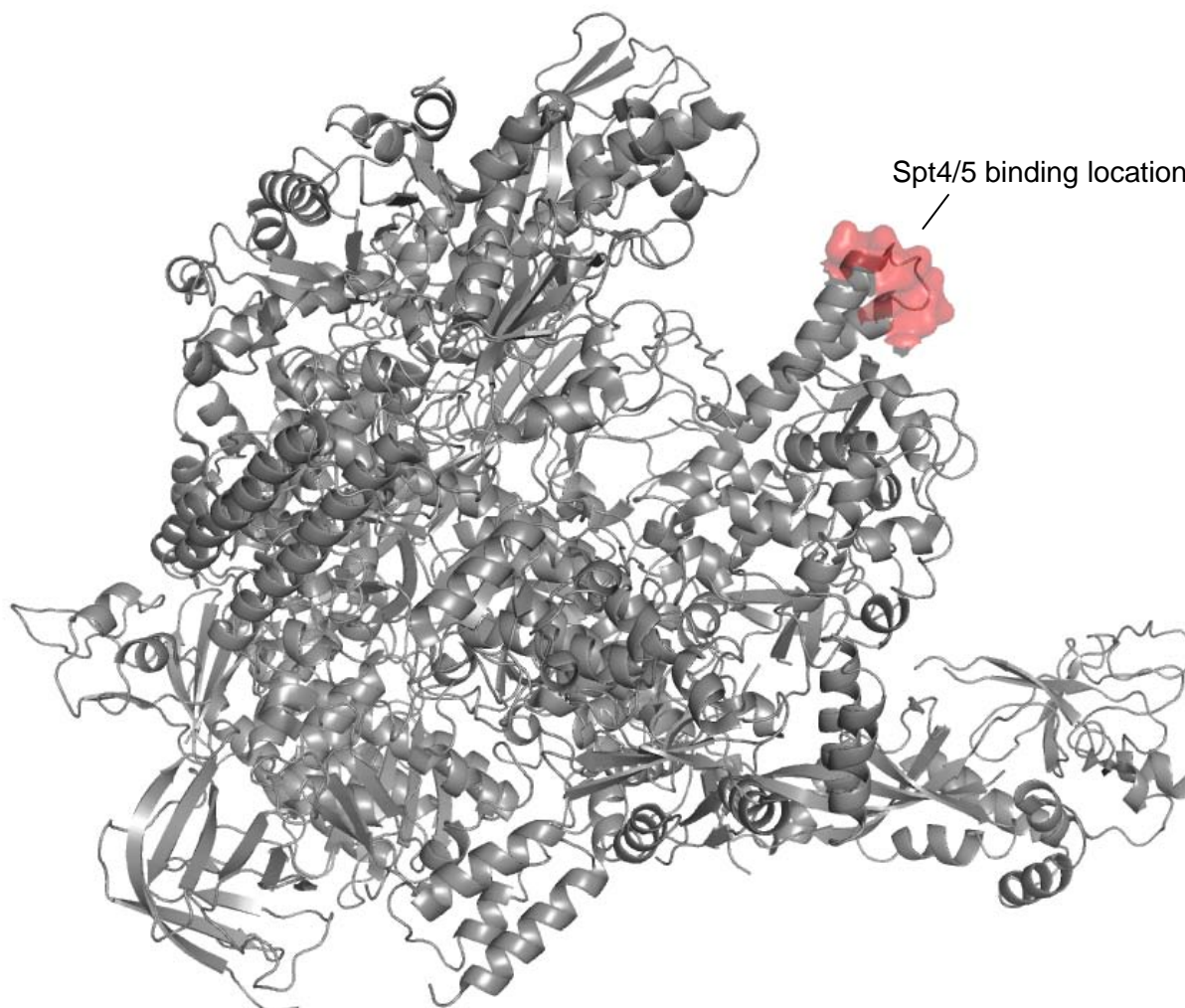
Figure 4.8 Structure of the archaeal RNA polymerase.

The structure of the *Sulfolobus shibatae* RNA polymerase (pdb-code 2WAQ) (Korkhin et al., 2009) is shown as a ribbon representation. The loop corresponding to the loop between two antiparallel helices in the β' clamp domain in *E. coli* RNA polymerase is highlighted in red.

4.3   Conclusion

Even though quite some structural information was available from Spt4/5 core complexes we present the first structure of a genuine Spt4/5 heterodimer. The Spt5NGN domain consists of a four-stranded antiparallel beta sheet flanked by three alpha helices. Spt4 has four beta strands in its center, with an alpha helix at the edge of the sheet. Spt4 contains a zinc finger, formed by four conserved cysteine residues and is located at the opposite side of the beta sheet. It is likely that the zinc finger is an essential structural element of Spt4. Spt4 and

Spt5NGN interact via hydrophobic interfaces that are held in register by polar interactions and alignment of their β sheets.

The Spt5NGN domain is conserved in archaea, bacteria and eukaryotes both structurally and functionally. This implies that Spt5 also shares similar mechanisms of transcription regulation. Spt4 reveals conservation between archaea and eukaryotes both structurally and functionally, there is no Spt4 homolog in bacteria. Surface analysis of the *M. jannaschii* Spt5NGN structure revealed a hydrophobic pocket in the center of the protein. This hydrophobic pocket is structurally conserved in the *E. coli* and *S. cerevisiae* Spt5NGN proteins and is suggested to bind RNA polymerase via interaction with the loop between two antiparallel helices in the β' clamp domain. This work will be implemented within the work from Dr. Finn Werner (University College London).

5 Material & Methods

5.1 Purification of RNA polymerase II

5.1.1 Fermentation of yeast

Two types of fermenters were available for producing up to 2.0 kg of yeast pellet per batch. The small fermenter (ISF200, Infors) has a nominal volume of 20 l and is ideally run with up to 15 l of media. The large fermenter (ABEC, Infors) has a nominal volume of 200 l and can be filled with up to 160 l of media. Table 5.1 shows the media composition and the culture parameters of both fermenters:

Table 5.1 Conditions for the fermentation of yeast

	small fermenter (20 l)	large fermenter (200 l)
YPD media	300 g peptone	3200 g peptone
	300 g glucose	3200 g glucose
	222 g yeast extract	2370 g yeast extract
	15 l desalted water	160 l desalted water
Antibiotics[1]	0.75 g ampicillin	8.0 g ampicillin
	0.15 g tetracycline HCl	1.6 g tetracycline HCl
typical inoculation volume	0.3 l with OD600 ≈ 2	4-5 l with OD600 ≈ 2
air flow	8 l/min	20 l/min
stirrer speed	800 rpm	200 rpm
typical growth time	12 – 15 hours	12 – 15 hours

[1]added after sterilization, prior to inoculation

Yeast was harvested at late logarithmic/early stationary phase, monitored by OD600 measurement. Cells were collected by a continuous flow centrifuge (Padberg Z4IG, 20,000 rpm). The cell pellet was resuspended in 330 ml of 3x freezing buffer (see chapter 5.1.2.1) per

kg cells and stirred at 4 °C for 30 min, before shock-freezing in liquid nitrogen. Cells were stored at -80°C until use.

5.1.2   Purification of 10-subunit core RNA polymerase II

Core RNA polymerase II was isolated from the *Saccharomyces cerevisiae* strain CB010ΔRpb4 (MATa *pep4*::HIS3/*prb1*::LEU2, *prc1*::HISG, *can1*, *ade2*, *trp1*) (Edwards et al,1990; Fu et al, 1999). This strain carries knockouts of several cellular proteases and of Rpb4. In the absence of Rpb4, Rpb7 dissociates from core Pol II during purification, giving rise to homogeneous 10-subunit core Pol II.

5.1.2.1   Buffers used for the purification

100 x protease inhibitor mix (p.i.)
1.42 mg Leupeptin
6.85 mg Pepstatin A
850 mg PMSF
1650 mg benzamidine
dry ethanol to 50 ml
stored at –20 °C; added immediately before use

3x freezing buffer
150 mM Tris-HCl, pH 7.9 @ 4 °C
3 mM EDTA
30 % glycerol
30 µM $ZnCl_2$
3 % DMSO
30 mM DTT
3 x protease inhibitor mix

1 x HSB150 buffer
50 mM Tris-HCl, pH 7.9 @ 4 °C
150 mM KCl
1 mM EDTA

10 % glycerol

10 µM $ZnCl_2$

10 mM DTT

1 x protease inhibitor mix

1 x HSB600 buffer

50 mM Tris-HCl, pH 7.9 @ 4 °C

600 mM KCl

1 mM EDTA

10 % glycerol

10 µM $ZnCl_2$

10 mM DTT

1 x protease inhibitor mix

TEZ buffer

50 mM Tris-HCl, pH 7.5 @ 20 °C

1 mM EDTA

10 µM $ZnCl_2$

1 mM DTT

1 x protease inhibitor mix

1 x Pol II buffer

5 mM Hepes pH 7.25 @ 20 °C

40 mM ammonium sulfate

10 µM $ZnCl_2$

10 mM DTT

Acetate buffer

100 mM sodium acetate pH 4.0

500 mM sodium chloride

PBS

4.3 mM Na_2HPO_4

1.4 mM KH_2PO_4

137 mM sodium chloride

2.7 mM potassium chloride

pH 7.4

Coupling buffer

100 mM sodium bicarbonate pH 8.3

500 mM sodium chloride

5.1.2.2   Purification of core RNA Polymerase II

Day1

For three bead-beaters (BioSpec), up to 600 ml of cell suspension were thawed in warm water. Each bead-beater was filled with 200 ml of borosilicate glass beads (0.45-0.50 mm diameter), 1 ml of protease inhibitor mix and 200 ml of the cell suspension. HSB150 was added to fill the bead-beater completely, taking care to avoid any remaining air bubbles. Lysis was achieved within 60-75 min of bead-beating (30 s on/90 s off) while the beater chambers were submersed in a salt/ice mixture. Glass beads were removed by filtration through a mesh funnel. The beads were washed with HSB150 until the flowthrough was clear. The lysate was cleared by two rounds of centrifugation (45 min at 9000 rpm in a GS3 rotor or 30 min at 12000 rpm in a SLA1500 rotor). Lipids were then removed by filtration of the supernatant through two layers of paper filter discs underneath a dressing cloth. The cleared lysate was applied onto a column packed with 250 ml of Heparin Sepharose 6 FF (GE Healthcare) (flow rate: 6-8 ml/min), pre-equilibrated with 750 ml of HSB150. Elution was accomplished with 500 ml of HSB600 (flow rate: 6-8 ml/min). Proteins in the eluate were precipitated by adding 291 g of fine-ground ammonium sulphate per litre of eluate (= 50 % saturation), followed by 60 min of stirring at 4 °C, over-night incubation at 4 °C and finally centrifugation (45 min. at 12000 rpm in a SLA1500 rotor). The heparin column was restored by washing with 1 l of 6 M urea, followed by water, and stored in 5 mM potassium acetate in 20 % (v/v) ethanol. Every five runs, the heparin column was regenerated by a brief wash with 500 ml of 0.1 M NaOH, followed by water and 5 mM potassium acetate in 20 % (v/v) ethanol.

Day 2

The next day the ammonium sulphate pellet of day 1 was dissolved in 50 ml of buffer TEZ. More TEZ was added to adjust the conductivity below the conductivity of TEZ containing additionally 400mM ammonium sulphate (TEZ400). This sample was centrifuged (15 min at 14000 rpm in an SLA1500 rotor) to remove undissolved constituents and then loaded by gravity flow onto the immunoaffinity column at 4 °C (see chapter 5.1.2.3, flow rate: 0.5-1.0 ml/min). The column was pre-equilibrated with 20 ml of TEZ containing 250 mM ammonium sulphate (TEZ250). The flowthrough was loaded onto a second column to increase the yield of Pol II. The columns were brought to room temperature, washed with 25 ml of TEZ500 at room temperature and Pol II was eluted in 1 ml fractions with TEZ500 containing additionally 50 % (v/v) glycerol (ca. 15 ml). Directly afterwards, 9 mM DTT was added to the elution fractions containing Pol II (monitored with the Bradford assay), and they were stored at 4 °C over night. The columns were washed with 5 ml of TEZ500 containing 70 % (v/v) ethylene glycol but no DTT, and reequilibrated with 25 ml of TEZ250 containing 0.02 % sodium azide. Generally, the recovery of Pol II decreased with each use of the column starting already from the first use. One reason is probably the sensitivity of the antibody towards DTT.

Day 3

The next day the peak fractions were combined and a buffer exchange procedure was performed. The buffer in the antibody column elution fractions was exchanged for 1x Pol II buffer using centrifugal ultrafiltration devices (MWCO 100,000 Da, Millipore Amicon Ultra-15). Completeness of the buffer exchange was monitored by measuring the conductivity of the flowthrough. Finally, Pol II was concentrated to 1-2 mg/ml. The Pol II sample was divided into aliquots of 100-500 µg Pol II. The aliquots were mixed with 1.13 times the volume of a saturated ammonium sulphate solution. The mixture was incubated for at least one hour at 4 °C and centrifuged for 45 min. at 4 °C in a table-top centrifuge at 13000 rpm. Most of the supernatant was decanted so that the pellet was still covered with supernatant, before it was shock-frozen in liquid nitrogen and stored at –80 °C. Pol II stored this way is stable for at least 3 months. From 600 g yeast pellet, a yield of 0.5-4 mg of highly purified Pol II was achieved. The yield was very dependent of the amount of times the immunoaffinity columns had been used before.

5.1.2.3   *Preparation of Pol II immunoaffinity resin*

The monoclonal antibody 8WG16 (NeoClone, Madison/USA), described in (Thompson & Burgess, 1996) is specific for the unphosphorylated CTD of Pol II and optimized to release Pol II upon treatment with glycerol or ethylene glycol at room temperature ("polyol responsive antibody"). The antibodies were purified from mouse ascites and immobilized on activated chromatography media according to the following procedure:

Lyophilized ascites were dissolved in PBS to its original volume and filtered through 0.2 μm membrane filters. The solution was passed > 3 times through a protein-A sepharose column (5 ml column volume, Sigma), pre-equilibrated in PBS. The column was washed with 50 ml PBS and antibodies were eluted with 20 ml of 0.75 M acetic acid. Fractions of 1 ml were collected into tubes containing 200 μl of 2 M Hepes (pH 7.9) to neutralize the acid. Peak fractions were pooled and the protein-A sepharose column was regenerated by washing for 5 minutes with 1 M acetic acid, followed by PBS with 0.02 % sodium azide.

The matrix for immunoaffinity columns was cyanogen bromide (CNBr)-activated sepharose 4B (Sigma), which reacts with free amines, e.g. accessible -NH_2 groups on proteins. Care was taken to avoid other sources of free amines (e.g. Tris) and to use a sealed bottle of activated sepharose. For each immunoaffinity column, 5 ml of gel was prepared by suspending 1.43 g of CNBr-sepharose in several ml of 1 mM HCl in a disposable gravity-flow column. The suspended CNBr-sepharose was first washed with 100 ml of 1 mM HCl, then with 20 ml of coupling buffer. Coupling was performed with 10 mg of purified antibodies per column for 2 hours at 20 °C or over night at 4 °C. When the coupling reaction was completed, no protein was detectable in the supernatant. The column was then washed with 25 ml of 1 M Tris, pH 8 and incubated for 2 hours at room temperature or over night at 4 °C. Finally, the column was washed with 20 ml of coupling buffer, followed by acetate buffer and coupling buffer. Columns were stored at 4 °C in TEZ60 with 0.02 % sodium azide. The columns could be used several times, if DTT exposure was reduced to a minimum, but in general a decrease in Pol II yield was observed already after the first use with further decrease after subsequent uses.

5.1.3 Purification of His-tagged RNA polymerase II

Ni buffer

20 mM Tris-HCl pH 7.9

150 mM KCl

10 µM $ZnCl_2$

10% v/v glycerol

10 mM DTT

1x p.i. (see chapter 5.1.2.1)

High salt buffer

20 mM Tris-HCl pH 7.9

1000 mM KCl

7 mM imidazole

10 µM $ZnCl_2$

10% v/v glycerol

10 mM DTT

1x p.i. (see chapter 5.1.2.1)

Ni7 buffer

20 mM Tris-HCl pH 7.9

150 mM KCl

7 mM imidazole

10 µM $ZnCl_2$

10 mM DTT

1x p.i. (see chapter 5.1.2.1)

Elution buffer

20 mM Tris-HCl pH 7.9

150 mM KCl

100 mM imidazole

10 µM $ZnCl_2$

10 mM DTT

MonoQ buffer

20 mM Tris-acetate pH 7.9

0.5 mM EDTA

10 µM $ZnCl_2$

10% v/v glycerol

10 mM DTT

For two bead-beaters (BioSpec), up to 400 ml of cell suspension were thawed in warm water. Each bead-beater was filled with 200 ml of borosilicate glass beads (0.45-0.50 mm diameter), 1 ml of protease inhibitor mix (see chapter 5.1.2.1) and 200 ml of the cell suspension. HSB150 (see chapter 5.1.2.1) was added to fill the bead-beater completely, taking care to avoid any remaining air bubbles. Lysis was achieved by bead beating for 80 min using intervals of 30 seconds followed by 90 second pauses while the beater chambers were submersed in a salt/ice mixture. Glass beads were removed by filtration through a mesh funnel. The beads were washed with HSB150 until the flowthrough was clear. The lysate was cleared by two rounds of centrifugation (2x45 min at 9000 rpm in a GS3 rotor). The lysate was ultracentrifuged for 90 min at 4 °C (76,221 xg (24,000 rpm in SW-28 (Beckmann-Coulter), swing-out rotor, or 125,171 xg (40,000 rpm in Ti- 45 (Beckmann-Coulter), fixed-angle rotor). The fat layer was aspired from the top of the tubes and the aqueous phase was collected without disturbing the DNA pellet. To the lysate ammonium sulphate was added to 50% saturation, proteins were allowed to precipitate overnight.

The next day, the precipitated lysate was centrifuged (45 min at 23,761 xg, 12,500 rpm in SLA-1500 rotor, Sorvall), and the ammonium sulphate pellet was dissolved in Ni buffer. To the solution 2 x 8 ml fresh Ni-NTA was added and stirred for at least 60 min. The Ni-NTA solution was loaded on two columns using gravity flow. After washing with high salt buffer and with Ni7 buffer, the protein was eluted with elution buffer. The eluted protein was 2x diluted with MonoQ buffer, filtered (0.22 µm), and subjected to anion exchange chromatography (MonoQ, GE healthcare) using a gradient from 150 mM to 1500 mM KOAc overnight.

The next day the last elution peak (at a conductivity of 50 mS/cm) was collected and concentrated to 1-2 mg/ml using centrifugal ultrafiltration devices (MWCO 100,000 Da, Millipore Amicon Ultra-15). The Pol II sample was divided into aliquots of 100-500 µg Pol II. The aliquots were mixed with 1.13 times the volume of a saturated ammonium sulfate solution. The mixture was incubated for at least 60 min at 4 °C and centrifuged for 45 min at 4 °C in a table-top centrifuge at 13000 rpm. Most of the supernatant was decanted so that the pellet was still covered with supernatant, before it was shock-frozen in liquid nitrogen and

stored at –80 °C. Pol II stored this way is stable for at least 3 months. From 300 g yeast pellet, a yield of 3-7 mg of highly purified Pol II was achieved.

5.2   Purification of Rpb4/7

Buffer 1
150 mM NaCl
5 % (v/v) glycerol
50 mM Tris pH 7.5
10 mM β-mercaptoethanol
1x p.i. (see chapter 5.1.2.1)

Buffer 2
50 mM Tris pH 7.5
5 mM DTT
1 mM EDTA

Recombinant yeast Rpb4/7 was expressed in *E. coli* BL21(DE3) RIL (Stratagene) using a bicistronic vector (Armache et al., 2005; Sakurai et al., 1999). Cells were grown in 2 x 2 L of culture in auto-induction medium (Studier, 2005). After approx. 4 h, when OD600 $\approx$ 0.6 was reached, the temperature was shifted from 30 °C to 20 °C. After 11 h, the cells were harvested by centrifugation (15 min at 5000 rpm in a SLC6000 rotor), resuspended in buffer 1 and lysed by means of a French Press. The lysate was cleared by centrifugation (30 min at 15000 rpm in a SS34 rotor) and applied onto a NiNTA column (Quiagen; 1 ml column volume). The column was washed subsequently with 3 ml of buffer 1, 3 ml buffer 1 containing additionally 10 mM imidazole and 3 ml of buffer 1 containing additionally 20 mM imidazole. Elution was performed subsequently with 3 ml of buffer 1 containing additionally 50 mM imidazole and 3 ml of buffer 1 containing additionally 200 mM imidazole. Peak fractions were pooled, diluted 1:3 with buffer 2 and applied on a ResourceQ column (GE Healthcare, 6 ml column volume), pre-equilibrated in buffer 2. Rpb4/7 was eluted with a linear gradient from 0-1000 mM NaCl in buffer 2. Peak fractions were concentrated and applied on a Superose12 10/300 GL gel filtration column (GE Healthcare), pre-equilibrated in Pol II buffer (see chapter 5.1.2.1). The purified Rpb4/7 heterodimer was concentrated to 10 mg/ml and aliquots were stored at –80°C.

5.3   Purification of TFIIS

Buffer A
50 mM Hepes pH 7.5
300 mM NaCl
5% glycerol
10 µM $ZnCl_2$
1x p.i. (see chapter 5.1.2.1)
10 mM ß-mercaptoethanol

For the expression of full-length TFIIS, transformed BL21(DE3) RIL (Stratagene) cells were used. Cells were grown in LB medium, supplemented with chloramphenicol (30 mg/l) and kanamycin (30 mg/l), to an OD600 of 0.7 and cooled down to 20 °C before expression was induced with 1mM IPTG, and continued over night. Cells from 1 l of culture were lysed by sonication in 50 ml of buffer A. The lysate was cleared by centrifugation (45 min, 13000 rpm, SS34 rotor) and was applied to a His Trap HP column (1 ml; GE Healthcare). The column was washed with buffer A containing 500 mM NaCl, and the protein was eluted over a total of 20 ml with a gradient of 0 mM to 500 mM imidazole in buffer A containing 500 mM NaCl. Peak fractions were diluted 5-fold in buffer A and loaded onto a Mono-S 5/5HR anion exchange column (GE Healthcare), pre-equilibrated with buffer A containing 100 mM NaCl. The TFIIS variant was eluted over a total of 15 column volumes with a gradient of 100-500 mM NaCl in buffer A. Peak fractions were pooled and subsequently applied onto a Superose 12 10/300GL gel filtration column (GE Healthcare), previously equilibrated in 1 x transcription buffer (chapter 5.6). Elution fractions (2 mg/ml protein) were shock-frozen in liquid nitrogen and stored at –80°C.

5.4   Purification of core Spt4/5.

The core Spt4/5 from *Methanocaldococcus jannaschii* was purified both using normal LB medium (in cooperation with the group of Finn Werner at University College London) and also using minimal medium with selenomethionine instead of normal methionine. Below the purification with selenomethionine is described. The LB medium purification was performed the same, except that LB medium was used throughout the entire purification.

Lysis buffer

20 mM Tris-HCl pH 7.9

300 mM potassium acetate

7 mM magnesium acetate

1 mM EDTA

0.1 mM $ZnCl_2$

10 mM DTT

10% glycerol

Crystallization buffer

150 mM NaCl

25 mM Tris/HCl pH 7.5

100 µM $ZnCl_2$

Day 1

On the first day methionine auxotroph competent cells were used to transform a pGEX plasmid containing the core Spt4/5 from *Methanococcus jannaschii*. The complete Spt4 and the first 82 residues of Spt5 were present on the plasmid. Spt5 was cloned with an N-terminal GST-tag, after cleavage a GSRRASVGSH linker remains on the N-terminal Spt5. The plasmid was provided by the Finn laboratory from University College London.

Day 2

The next day a colony was picked and used for a 50 ml overnight culture in LB medium (37°C, with ampicillin).

Day 3

The next morning 30 ml of the overnight culture was used to inoculate 2.5l of LB, supplemented with ampicillin. Cells were grown at 37°C until OD600 was around 0.6. Cells were collected (5000 rpm for 10 min using a SLC6000 rotor). Cells were resuspended in minimal media containing selenomethionine. Cells were grown at 20°C, when the OD increased by 0.2, 1µM IPTG was added. Cells were grown overnight (at least 16 hours).

Day 4

Cells were collected (5000 rpm for 15 min using a SLC6000 rotor). The pellet was dissolved in lysis buffer (~10ml/l) and 200 units benzonase was added. Cells were sonicated until they were lysed properly. The cell debrid was removed by centrifugation (20,000 rpm for 20 min using a sorvall SS34 rotor). The supernatant was filtered through a 0.22 µm filter and applied to a 5 ml GSTRAP FF column (GE healthcare). The column was washed with lysis buffer, and the protein was eluted using a 10 ml gradient from 0 mM to 10 mM reduced glutathione. Peak fractions were collected. The GS tag was cleaved by the addition of 2 µl Thrombin stock solution (1U/µl in PBS) per ml elution incubated for one hour at 37 °C or overnight at 4°C.

Day 5

The next day a sample was loaded on an SDS-PAGE gel to verify cleavage of the tag. Thrombin, GST and other impurities were removed by incubating at 70°C for 30 min, followed by centrifugation. The supernatant was filtered and applied to a Superose 12 10/300GL gel filtration column (GE Healthcare) (concentrated if necessary) using crystallization buffer. The protein fractions were collected and concentrated to 30 mg/ml. Aliquots (50-100µl) were shock-frozen in liquid nitrogen and stored at –80°C.

5.5   Crystallization of core Spt4/5

The purified core Spt4/5 from *Methanocaldococcus jannaschii* was concentrated to 30 mg/ml and crystallization screens were performed using the crystallization facility (MPI, Martinsried). Screens were performed at 4 °C and 20 °C as indicated in Table 5.2.

Table 5.2: Screens used for crystallization of the Spt4/5 core complex.

Screens Magic1 and Magic2 are screens from the E. Conti crystallization facility (MPI Martinsried). The other screens are commercially available screens.

	Hits at 20°C	Hits at 4°C
Magic 1	C1, B3, D9, G3, H1	C1, D1, D2
Magic 2	F3	B3, H6
Nextal Classic	H2	F11
Hampton Index	E10	C3, E10, F10
Wizard Screen	B4, B9, F12	Not performed
Cations	No hits	Not performed
Anions	No hits	Not performed
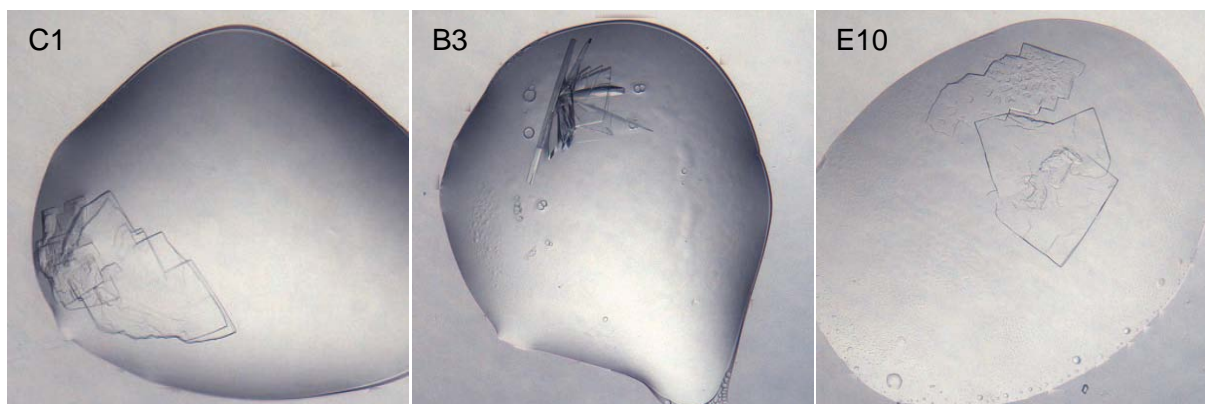
Figure 5.1 Selected hits for crystal optimization.

From left to right: Magic 1, C1 (50 mM MES pH 6.0 30% PEG 3350), Magic 2, B3 (50 mM MES pH 6.0 25% PEG400), Hampton Index, E10 (0.1 M Bis-Tris pH 6.5 45% v/v PPG400).

Table 5.3: Optimized crystallization conditions.

Crystallization condition	Solution1	Solution2	Crystals
PEG3350	50 mM MES pH 5.5-7.0	12-32% PEG3350	Thin, star like, intergrown crystals
PEG400	50 mM MES pH 5.5-7.0	20-30% PEG400	Thin, plate like intergrown crystals
PPG400	100 mM Bis-Tris pH 5.5-7.0	16-26% PPG400	Thin, large single crystals

The most promising looking hits are shown in Figure 5.1. These conditions were optimized by hand. Crystals were grown at 22 °C with the hanging drop vapour diffusion method by mixing 1 µl of sample solution with 1 µl of reservoir solution (optimized crystallization conditions shown in Table 5.3). Crystals were harvested in mother solution after 2-4 days, when they had reached their maximum size (approx. 0.3 x 0.3 x 0.01 mm). Crystals were transferred to mother solution containing additionally 25 % glycerol (only for the PEG3350 containing crystallization conditions). Crystals were flash-cooled by plunging into liquid nitrogen. Large single crystals were obtained for the PPG400 containing crystallization condition, as shown in Figure 5.2. All crystals obtained were very thin, resulting in difficult to handle crystals, and also difficult to process data sets, which were often not 100% complete.
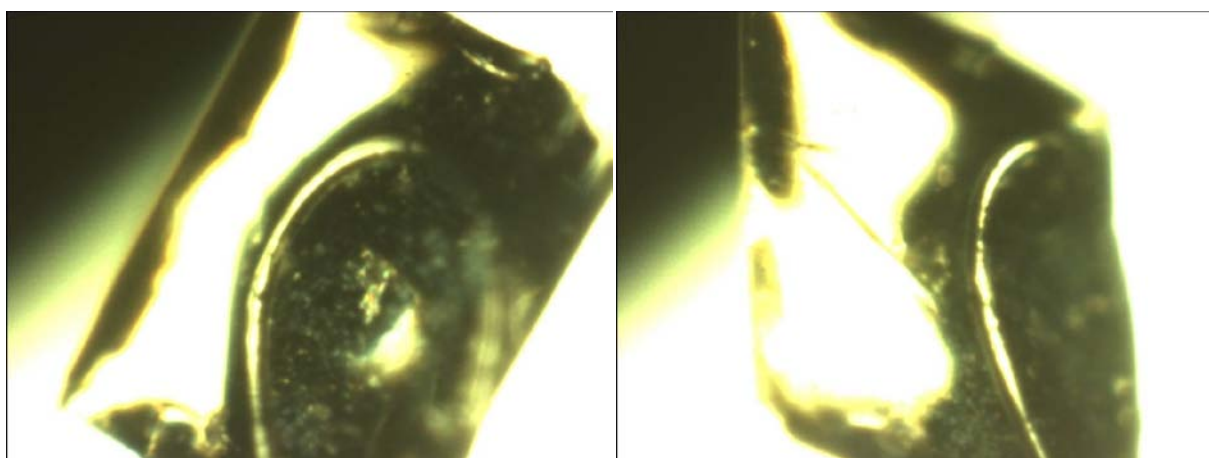
Figure 5.2 Crystals of the Spt4/5 core complex in a loop.
The crystal shown was large, but thin (around 500 x 500 x 10 µm). Diffraction was reasonable in one direction, but not in the thin part, this was a general problem.

Diffraction data were collected in 0.1°-0.25° increments at the protein crystallography beamline X06SA of the Swiss Light Source using a Pilatus 6M pixel detector (Broennimann et al., 2006). Raw data were processed with XDS (Kabsch, 1993). Several attempts were made to solve the structure using molecular replacement, when this was unsuccessful attempts were made to solve the structure using heavy metals. Unfortunately the anomalous signal from the internal zinc was too low; therefore heavy metal soaking was tried. The compounds mersalyl acid, sodium tetrachloroaurate and potassium bromide were used, these attempts failed because the heavy metal signals were too weak and/or incomplete. Then the structure of the core Spt5 from *Methanocaldococcus jannaschii* was published. Using this structure only, molecular replacement still failed. However, using the *M. jannaschii* Spt5 structure together with a Spt4 homologous structure (1RYQ, 46% sequence identity), molecular replacement was succesfull. The 1RYQ pdb file was used with trimmed side chains, only if the sequence was identical to the *M. jannaschii* Spt4 the side chain was kept, the program Phaser (McCoy et al., 2005) was used for molecular replacement. The structure was build and refined using the programs Coot (Emsley and Cowtan, 2004) and Refmac (Vagin et al., 2004).

5.6   Assembly of Pol II elongation complexes

TE buffer
10 mM Tris pH 7.4
1 mM EDTA

Assembly buffer
50 mM Hepes pH 7.5
40 mM $(NH_4)_2SO_4$
10 µM $ZnCl_2$
5 % (v/v) glyerol
10 mM DTT

Pol II buffer
5 mM Hepes pH 7.25 at 20 °C
40 mM $(NH_4)_2SO_4$
10 µM $ZnCl_2$
10 mM DTT

Transcription buffer

20 mM Hepes pH 7.6

60 mM $(NH_4)_2SO_4$

8 mM $MgSO_4$

10 µM $ZnCl_2$

10 % (v/v) glycerol

10 mM DTT

Nucleic acid scaffolds were annealed by mixing equimolar amounts of synthetic template DNA, nontemplate DNA, and RNA in TE buffer at a final concentration of 100 µM, heating the mixture to 95 °C for 5 min, and slow-cooling to 4 °C in a thermocycler. Stoichiometric Pol II ECs were assembled by incubating core Pol II for 10 min with 2 molar equivalents of nucleic acid scaffold, followed by 20 min incubation with 5 molar equivalents of recombinant Rpb4/7 in assembly buffer at 20 °C. The complexes were purified by size exclusion chromatography (Superose 6 10/300 GL) in Pol II buffer.

For the bead-based assays, the ECs containing complete complementary scaffolds were assembled essentially as described (Kireeva et al., 2003)). Briefly, the DNA nontemplate was 5' end-labeled with Biotin with the use of a TTTTT linker. The RNA was 5' end-labeled with 6-carboxyfluoresceine (FAM). For EC assembly, RNA polymerase II was incubated with a hybrid of the DNA template strand annealed to the RNA (2-fold excess) in transcription buffer for 15 min at 20 °C, subsequently with the biotinylated nontemplate DNA strand (4-fold excess) for 10 min at 25 °C, and then with recombinant Rpb4/7 (5-fold excess) for 10 min at 25 °C.

5.7   RNA extension and cleavage assays

Transcription buffer

20 mM Hepes pH 7.6

60 mM $(NH_4)_2SO_4$

8 mM $MgSO_4$

10 µM $ZnCl_2$

10 % (v/v) glycerol

10 mM DTT

Loading buffer

90 % (v/v) formamide

50 mM EDTA pH 7.4

Breaking buffer

50 mM Tris/HCl pH 8.0

150 mM NaCl

0.1% (w/v) triton X-100

5% (w/v) glycerol

0.5 mM DTT

Blocking buffer

50 mM Tris/HCl pH 8.0

150 mM NaCl

2 mM EDTA pH 8.0

0.1% (w/v) triton X-100

5% (w/v) glycerol

0.5% (w/v) BSA

200 µg/ml insulin

0.1 mg/ml heparin

0.5 mM DTT

5.7.1 Extension and cleavage assays using minimal scaffolds

Stoichiometric ECs of complete Pol II, containing the 10-subunit core and Rpb4/7, were assembled and purified as described in chapter 5.6. The RNAs used for extension assays were identical to those used for structural studies, except for five additional nucleotides (5'-UGCAU-3') and a fluorescent linker (6-Carboxyfluorescein) at the 5' end. For transcript extension, ECs were incubated with NTPs at 28 °C in transcription buffer. Reactions were stopped by incubating with an equal volume of gel loading buffer for 5 min at 95 °C. The RNA products were separated by denaturing gel electrophoresis (0.5 pmol RNA per lane, 0.4 mm 20 % polyacrylamide gels containing 7 M urea, 50-55 °C) and visualized with a Typhoon 9400 scanner (GE Healthcare).

5.7.2 MALDI-TOF analysis of minimal scaffold assays

For MALDI-TOF analysis, RNA extension was performed as described in chapter 5.7.1, except that the reactions were stopped by the addition of an equal volume of 100 mM EDTA. Samples were slowly heated to 95 °C, and incubated for 5 min at 95 °C. After centrifugation (15 min, 13,000 rpm), the supernatant was loaded twice on a C18 ZipTip (Millipore, pre-equilibrated with 3x 80% acetonitril, followed by 4x 100 mM TEAA) to desalt the nucleic acids, after washing 3x with 100 mM TEAA and 4x with H_2O, nucleic acids were eluted with 80% acetonitril. The nucleic acids were subjected to MALDI-TOF analysis using 3-hydroxypicolinic acid (3-HPA) and pyrazinecarboxylic acid (PCA) (4:1) as a matrix (Zhou et al., 2004).

5.7.3 Bead based extension assays

For bead based extension assays, beads (Dynabeads MyOneTM Streptavidin T1 from Invitrogen) were washed with breaking buffer and incubated overnight with blocking buffer at 4 °C. Assembly was performed as described in chapter 5.6. Beads were added to the assembled ECs and incubated for 30 min at 25 °C. Beads were subsequently washed with transcription buffer containing 0.1% Triton-X, transcription buffer containing 0.2 M $(NH_4)_2SO_4$, and with transcription buffer. Beads were resuspended in transcription buffer. For RNA extension assays, different amounts of NTPs were added, the mixture was incubated at 28 °C and reactions were stopped at different time points by addition of an equal volume of 100 mM EDTA, essentially as described (Brueckner et al., 2007). For cleavage assays, the bead-coupled ECs were incubated at 28 °C in transcription buffer for 5 min and stopped as described above. The beads were transferred into urea loading buffer, samples were heated to 95 °C and loaded on a 20% polyacrylamid gel containing 7 M Urea. The FAM 5' labeled RNA products were visualized with a Typhoon 9400 scanner (GE Healthcare). Gel bands were quantified using ImageQuant (GE healthcare). For MALDI-TOF analysis, the reaction was incubated, stopped and analyzed as described in chapter 5.7.2.

5.8   Crystallization set-up

NH$_4$-Mg-acetate crystallization solution (= Natrix #38)

50 mM Hepes pH 7.0

140-200 mM ammonium-acetate

150 mM magnesium-acetate

3.5-5.5 % (v/v) PEG 6000

5 mM TCEP

for cryocooling additionally 20-22 % glycerol

NH$_4$-Na-acetate crystallization solution (= Natrix #38woMg)

50 mM Hepes pH 7.0

140-200 mM ammonium-acetate

300 mM sodium-acetate

4.0-6.0 % (v/v) PEG 6000

5 mM TCEP

for cryocooling additionally 20-22 % glycerol

Purified Pol II ECs were concentrated to 3.5-4.5 mg/ml and an additional amount of the nucleic acid scaffold was added prior to crystallization to a final concentration of 2 µM. Crystals were grown at 22 °C with the hanging drop vapour diffusion method by mixing 2 µl of sample solution with 1 µl of reservoir solution (NH$_4$-Mg-acetate crystallization solution). Crystals were harvested in mother solution after 10-20 days, when they had reached their maximum size (approx. 0.4 x 0.3 x 0.2 mm). Crystals were transferred stepwise over 5 hours to mother solution containing additionally 0-22 % glycerol. Crystals were slowly cooled to 8 °C and flash-cooled by plunging into liquid nitrogen.

For 8-oxoguanine containing crystals, the NH$_4$-Na-acetate crystallization solution was used, which did not contain magnesium. Under both crystallization conditions (with or without magnesium) Pol II crystallizes well. There are however differences, whether the scaffold is pre- or posttranslocated (or if it is a mixture, see (Brueckner and Cramer, 2008)) might depend on the crystallization condition used.

5.9   Crystal structure analysis

Diffraction data were collected in 0.25° increments at the protein crystallography beamline X06SA of the Swiss Light Source using the new Pilatus 6M pixel detector (Broennimann et al., 2006). Raw data were processed with XDS (Kabsch, 1993). Structures were solved by molecular replacement with the program Phaser (McCoy et al., 2005), using the structure of the complete 12-subunit Pol II elongation complex without nucleic acids as a search model (PDB 1Y1W, (Kettenberger et al., 2004)). The molecular replacement solution was subjected to rigid body refinement with CNS version 1.2 (Brunger et al., 1998). Model building was carried out with programs Coot (Emsley and Cowtan, 2004), and Moloc (Gerber and Muller, 1995). The nucleic acids were built into the initial F_o-F_c electron density map. The register of the 8oxoG containing nucleic acids was unambiguously defined by bromine labeling. A thymine residue in the template strand was replaced for 5-bromouracil, diffraction data of the resulting complex were recorded at the wavelength of the bromine K absorption edge, and the resulting anomalous difference Fourier maps revealed single peaks demarking the positions of the bromine atom (Figure 3.5b). The register of the cisplatin containing nucleic acids was clearly defined by the platinum atom itself. The DNA lesions (cisplatin and 8oxoG) were manually built into the initial F_o-F_c maps and further adjusted manually and by restrained real space refinement using Moloc. Atomic positions and B-factors were refined with CNS version 1.2. Refinement was monitored with the free R factor, calculated from the same 2 % set of excluded reflections as in refinement of complete Pol II (Armache et al., 2005) and the complete Pol II EC (Brueckner et al., 2007; Kettenberger et al., 2004). Figures were prepared with Pymol (DeLano, 2002).

6   Conclusion

This thesis described the study on the molecular mechanisms of Pol II encountering DNA damages. The damages investigated were the anticancer drug cisplatin, which forms 1,2-d(GpG) DNA intrastrand cross-links (cisplatin lesions) and 8-oxoguanine (8oxoG), the major DNA lesion resulting from oxidative stress.

We performed a structure-function analysis of Pol II stalling at a cisplatin lesion in the DNA template. Pol II 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. Pol 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.

In addition, we combined structural and functional data to derive the molecular mechanism of Pol II transcription over 8oxoG. When Pol II encounters 8oxoG in the DNA template strand, it correctly incorporates cytosine in most instances, but it also misincorporates adenine. The misincorporated adenine forms a Hoogsteen base pair with 8oxoG at the active center. This misincorporation requires rotation of the 8oxoG base from the standard *anti-* to an uncommon *syn*-conformation, which likely occurs during 8oxoG loading into the active site at a lower rate. X-ray analysis showed that the misincorporated adenine forms a Hoogsteen base pair with 8oxoG in the polymerase active center. Mass spectrometric analysis of RNA extension products showed that the misincorporated adenine escapes the intrinsic proofreading function of Pol II, and remains in the RNA product after polymerase bypass, resulting in transcriptional mutagenesis. Mutagenesis is suppressed by the transcript cleavage-stimulatory factor TFIIS, which is essential for cell survival during oxidative stress.

We showed that the stalling mechanism at a cisplatin lesion differs from that of Pol II stalling at a photolesion, which involves delivery of the lesion to the active site and lesion-templated misincorporation that blocks transcription. In case of 8oxoG no stalling occurs at all, which leads to transcriptional mutagenesis. Together, these results lead to the conclusion that it is impossible to predict the mechanisms of transcriptional stalling or mutagenesis at other types of lesions.

7   Abbreviations

8oxoG	8-oxoguanine
AD	Activation domain
AMP	adenosine monophosphate
ATP	adenosine triphosphate
CPD	cyclobutane pyrimidine dimer
CTD	C-terminal domain of Rpb1 of Pol II
CTP	cytosine triphosphate
DBD	DNA binding domain
DNA	deoxyribonucleic acid
DPE	downstream promoter element
DSIF	DRB sensitivity-inducing factor
DTT	dithiothreitol
E. coli	*Escherichia coli*
EC	elongation complex
EDTA	ethylene diamine tetraacetic acid
EM	electron microscopy
GMP	guanosine monophosphate
GTF	general transcription factor
GTP	guanosine triphosphate
Hepes	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Inr	initiator element
IPTG	isopropyl β-D-1-thiogalactopyranoside
MALDI	matrix-assisted laser desorption ionization
MALDI-TOF	matrix-assisted laser desorption ionization with time-of-flight analysator mass spectrometry
Min	minutes
mRNA	messenger RNA
MWCO	molecular weight cutoff
NTP	nucleotide triphosphate
PDB	protein data bank
PEG	polyethylene glycol (number indicates average molecular weight in Da)
PIC	preinitiation complex

Pol	eukaryotic DNA-dependent RNA polymerase
PPG	polypropylene glycol
RMSD	root mean square deviation
RNA	ribonucleic acid
Rpb	subunit of Pol II (=RNA polymerase B)
S. cerevisiae	*Saccharomyces cerevisiae*
Spt	Suppressor of Ty
TAF	TBP-associated factor
TBP	TATA binding protein
TCEP	tris(2-carboxyethyl)phosphine
TCR	transcription coupled repair
TFII	transcription factor of Pol II transcription
Tris	trishydroxymethylaminomethane
UMP	uridine monophosphate

8   References

Afonine, P.V., Grosse-Kunstleve, R.W., Adams, P.D., 2005, A robust bulk-solvent correction and anisotropic scaling procedure. Acta crystallographica 61, 850-855.

Allison, L.A., Moyle, M., Shales, M., Ingles, C.J., 1985, Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell 42, 599-610.

Armache, K.J., Kettenberger, H., Cramer, P., 2003, Architecture of initiation-competent 12-subunit RNA polymerase II. Proceedings of the National Academy of Sciences of the United States of America 100, 6964-6968.

Armache, K.J., Mitterweger, S., Meinhart, A., Cramer, P., 2005, Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. The Journal of biological chemistry 280, 7131-7134.

Asturias, F.J., Meredith, G.D., Poglitsch, C.L., Kornberg, R.D., 1997, Two conformations of RNA polymerase II revealed by electron crystallography. J Mol Biol 272, 536-540.

Basrai, M.A., Kingsbury, J., Koshland, D., Spencer, F., Hieter, P., 1996, Faithful chromosome transmission requires Spt4p, a putative regulator of chromatin structure in Saccharomyces cerevisiae. Mol Cell Biol 16, 2838-2847.

Belogurov, G.A., Vassylyeva, M.N., Svetlov, V., Klyuyev, S., Grishin, N.V., Vassylyev, D.G., Artsimovitch, I., 2007, Structural basis for converting a general transcription factor into an operon-specific virulence regulator. Molecular cell 26, 117-129.

Bregeon, D., Doddridge, Z.A., You, H.J., Weiss, B., Doetsch, P.W., 2003, Transcriptional mutagenesis induced by uracil and 8-oxoguanine in Escherichia coli. Molecular cell 12, 959-970.

Brieba, L.G., Eichman, B.F., Kokoska, R.J., Doublie, S., Kunkel, T.A., Ellenberger, T., 2004, Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase. The EMBO journal 23, 3452-3461.

Broennimann, C., Eikenberry, E.F., Henrich, B., Horisberger, R., Huelsen, G., Pohl, E., Schmitt, B., Schulze-Briese, C., Suzuki, M., Tomizaki, T., Toyokawa, H., Wagner, A., 2006, The PILATUS 1M detector. Journal of synchrotron radiation 13, 120-130.

Broyde, S., Wang, L., Rechkoblit, O., Geacintov, N.E., Patel, D.J., 2008, Lesion processing: high-fidelity versus lesion-bypass DNA polymerases. Trends Biochem Sci 33, 209-219.

Brueckner, F., Cramer, P., 2008, Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. Nature structural & molecular biology 15, 811-818.

Brueckner, F., Hennecke, U., Carell, T., Cramer, P., 2007, CPD damage recognition by transcribing RNA polymerase II. Science (New York, N.Y 315, 859-862.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L., 1998, Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta crystallographica 54, 905-921.

Buratowski, S., 2003, The CTD code. Nat Struct Biol 10, 679-680.

Burova, E., Hung, S.C., Sagitov, V., Stitt, B.L., Gottesman, M.E., 1995, Escherichia coli NusG protein stimulates transcription elongation rates in vivo and in vitro. J Bacteriol 177, 1388-1392.

Bushnell, D.A., Kornberg, R.D., 2003, Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. Proceedings of the National Academy of Sciences of the United States of America 100, 6969-6973.

Charlet-Berguerand, N., Feuerhahn, S., Kong, S.E., Ziserman, H., Conaway, J.W., Conaway, R., Egly, J.M., 2006, RNA polymerase II bypass of oxidative DNA damage is regulated by transcription elongation factors. The EMBO journal 25, 5481-5491.

Chen, Y.H., Bogenhagen, D.F., 1993, Effects of DNA lesions on transcription elongation by T7 RNA polymerase. The Journal of biological chemistry 268, 5849-5855.

Cheng, B., Price, D.H., 2007, Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. The Journal of biological chemistry 282, 21901-21912.

Ciampi, M.S., 2006, Rho-dependent terminators and transcription termination. Microbiology 152, 2515-2528.

Corda, Y., Job, C., Anin, M.F., Leng, M., Job, D., 1991, Transcription by eucaryotic and procaryotic RNA polymerases of DNA modified at a d(GG) or a d(AG) site by the antitumor drug cis-diamminedichloroplatinum(II). Biochemistry 30, 222-230.

Corda, Y., Job, C., Anin, M.F., Leng, M., Job, D., 1993, Spectrum of DNA--platinum adduct recognition by prokaryotic and eukaryotic DNA-dependent RNA polymerases. Biochemistry 32, 8582-8588.

Craighead, J.L., Chang, W.H., Asturias, F.J., 2002, Structure of yeast RNA polymerase II in solution: implications for enzyme regulation and interaction with promoter DNA. Structure 10, 1117-1125.

Cramer, P., Armache, K.J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C.D., Lehmann, E., Leike, K., Sydow, J.F., Vannini, A., 2008, Structure of eukaryotic RNA polymerases. Annual review of biophysics 37, 337-352.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., Kornberg, R.D., 2000, Architecture of RNA polymerase II and implications for the transcription mechanism. Science (New York, N.Y 288, 640-649.

Cramer, P., Bushnell, D.A., Kornberg, R.D., 2001, Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science (New York, N.Y 292, 1863-1876.

Damsma, G.E., Alt, A., Brueckner, F., Carell, T., Cramer, P., 2007, Mechanism of transcriptional stalling at cisplatin-damaged DNA. Nature structural & molecular biology 14, 1127-1133.

Darst, S.A., Edwards, A.M., Kubalek, E.W., Kornberg, R.D., 1991, Three-dimensional structure of yeast RNA polymerase II at 16 A resolution. Cell 66, 121-128.

Dengl, S., Cramer, P., 2009, The torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II in vitro. The Journal of biological chemistry.

Edwards, A.M., Kane, C.M., Young, R.A., Kornberg, R.D., 1991, Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. The Journal of biological chemistry 266, 71-75.

Emsley, P., Cowtan, K., 2004, Coot: model-building tools for molecular graphics. Acta crystallographica 60, 2126-2132.

Fernandez-Tornero, C., Bottcher, B., Riva, M., Carles, C., Steuerwald, U., Ruigrok, R.W., Sentenac, A., Muller, C.W., Schoehn, G., 2007, Insights into transcription initiation and termination from the electron microscopy structure of yeast RNA polymerase III. Molecular cell 25, 813-823.

Freisinger, E., Grollman, A.P., Miller, H., Kisker, C., 2004, Lesion (in)tolerance reveals insights into DNA replication fidelity. The EMBO journal 23, 1494-1505.

Fu, J., Gnatt, A.L., Bushnell, D.A., Jensen, G.J., Thompson, N.E., Burgess, R.R., David, P.R., Kornberg, R.D., 1999, Yeast RNA polymerase II at 5 A resolution. Cell 98, 799-810.

Gelasco, A., Lippard, S.J., 1998, NMR solution structure of a DNA dodecamer duplex containing a cis-diammineplatinum(II) d(GpG) intrastrand cross-link, the major adduct of the anticancer drug cisplatin. Biochemistry 37, 9230-9239.

Gerber, P.R., Muller, K., 1995, MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. J Comput Aided Mol Des 9, 251-268.

Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., Kornberg, R.D., 2001, Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. Science (New York, N.Y 292, 1876-1882.

Guo, M., Xu, F., Yamada, J., Egelhofer, T., Gao, Y., Hartzog, G.A., Teng, M., Niu, L., 2008, Core structure of the yeast spt4-spt5 complex: a conserved module for regulation of transcription elongation. Structure 16, 1649-1658.

Harris, J.K., Kelley, S.T., Spiegelman, G.B., Pace, N.R., 2003, The genetic core of the universal ancestor. Genome Res 13, 407-412.

Hartzog, G.A., Wada, T., Handa, H., Winston, F., 1998, Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. Genes Dev 12, 357-369.

Hirose, Y., Manley, J.L., 2000, RNA polymerase II and the integration of nuclear events. Genes Dev 14, 1415-1429.

Hsu, G.W., Ober, M., Carell, T., Beese, L.S., 2004, Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. Nature 431, 217-221.

Izban, M.G., Luse, D.S., 1993, SII-facilitated transcript cleavage in RNA polymerase II complexes stalled early after initiation occurs in primarily dinucleotide increments. The Journal of biological chemistry 268, 12864-12873.

Jasiak, A.J., Armache, K.J., Martens, B., Jansen, R.P., Cramer, P., 2006, Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. Molecular cell 23, 71-81.

Jensen, G.J., Meredith, G., Bushnell, D.A., Kornberg, R.D., 1998, Structure of wild-type yeast RNA polymerase II and location of Rpb4 and Rpb7. The EMBO journal 17, 2353-2358.

Jeon, C., Agarwal, K., 1996, Fidelity of RNA polymerase II transcription controlled by elongation factor TFIIS. Proceedings of the National Academy of Sciences of the United States of America 93, 13677-13682.

Jung, Y., Lippard, S.J., 2003, Multiple states of stalled T7 RNA polymerase at DNA lesions generated by platinum anticancer agents. The Journal of biological chemistry 278, 52084-52092.

Jung, Y., Lippard, S.J., 2006, RNA polymerase II blockage by cisplatin-damaged DNA. Stability and polyubiquitylation of stalled polymerase. The Journal of biological chemistry 281, 1361-1370.

Kabsch, W., 1993, Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J Appl Cryst 26, 795-800.

Kartalou, M., Essigmann, J.M., 2001, Recognition of cisplatin adducts by cellular proteins. Mutat Res 478, 1-21.

Kashkina, E., Anikin, M., Brueckner, F., Pomerantz, R.T., McAllister, W.T., Cramer, P., Temiakov, D., 2006, Template misalignment in multisubunit RNA polymerases and transcription fidelity. Molecular cell 24, 257-266.

Kathe, S.D., Shen, G.P., Wallace, S.S., 2004, Single-stranded breaks in DNA but not oxidative DNA base damages block transcriptional elongation by RNA polymerase II in HeLa cell nuclear extracts. The Journal of biological chemistry 279, 18511-18520.

Kettenberger, H., Armache, K.J., Cramer, P., 2003, Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. Cell 114, 347-357.

Kettenberger, H., Armache, K.J., Cramer, P., 2004, Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. Molecular cell 16, 955-965.

Kireeva, M.L., Lubkowska, L., Komissarova, N., Kashlev, M., 2003, Assays and affinity purification of biotinylated and nonbiotinylated forms of double-tagged core RNA polymerase II from Saccharomyces cerevisiae. Methods Enzymol 370, 138-155.

Komori, T., Inukai, N., Yamada, T., Yamaguchi, Y., Handa, H., 2009, Role of human transcription elongation factor DSIF in the suppression of senescence and apoptosis. Genes Cells 14, 343-354.

Korkhin, Y., Unligil, U.M., Littlefield, O., Nelson, P.J., Stuart, D.I., Sigler, P.B., Bell, S.D., Abrescia, N.G., 2009, Evolution of Complex RNA Polymerases: The Complete Archaeal RNA Polymerase Structure. PLoS Biol 7, e102.

Koyama, H., Ito, T., Nakanishi, T., Kawamura, N., Sekimizu, K., 2003, Transcription elongation factor S-II maintains transcriptional fidelity and confers oxidative stress resistance. Genes Cells 8, 779-788.

Koyama, H., Ito, T., Nakanishi, T., Sekimizu, K., 2007, Stimulation of RNA polymerase II transcript cleavage activity contributes to maintain transcriptional fidelity in yeast. Genes Cells 12, 547-559.

Krahn, J.M., Beard, W.A., Miller, H., Grollman, A.P., Wilson, S.H., 2003, Structure of DNA polymerase beta with the mutagenic DNA lesion 8-oxodeoxyguanine reveals structural insights into its coding potential. Structure 11, 121-127.

Kuraoka, I., Endou, M., Yamaguchi, Y., Wada, T., Handa, H., Tanaka, K., 2003, Effects of endogenous DNA base lesions on transcription elongation by mammalian RNA polymerase II. Implications for transcription-coupled DNA repair and transcriptional mutagenesis. The Journal of biological chemistry 278, 7294-7299.

Kuraoka, I., Suzuki, K., Ito, S., Hayashida, M., Kwei, J.S., Ikegami, T., Handa, H., Nakabeppu, Y., Tanaka, K., 2007, RNA polymerase II bypasses 8-oxoguanine in the presence of transcription elongation factor TFIIS. DNA repair 6, 841-851.

Kusser, A.G., Bertero, M.G., Naji, S., Becker, T., Thomm, M., Beckmann, R., Cramer, P., 2008, Structure of an archaeal RNA polymerase. J Mol Biol 376, 303-307.

Laine, J.P., Egly, J.M., 2006, Initiation of DNA repair mediated by a stalled RNA polymerase IIO. The EMBO journal 25, 387-397.

Lehmann, E., Brueckner, F., Cramer, P., 2007, Molecular basis of RNA-dependent RNA polymerase II activity. Nature 450, 445-449.

Li, B., Carey, M., Workman, J.L., 2007, The role of chromatin during transcription. Cell 128, 707-719.

Li, J., Horwitz, R., McCracken, S., Greenblatt, J., 1992, NusG, a new Escherichia coli elongation factor involved in transcriptional antitermination by the N protein of phage lambda. The Journal of biological chemistry 267, 6012-6019.

Lindahl, T., 1993, Instability and decay of the primary structure of DNA. Nature 362, 709-715.

Maga, G., Villani, G., Crespan, E., Wimmer, U., Ferrari, E., Bertocci, B., Hubscher, U., 2007, 8-oxo-guanine bypass by human DNA polymerases in the presence of auxiliary proteins. Nature 447, 606-608.

Malone, E.A., Fassler, J.S., Winston, F., 1993, Molecular and genetic characterization of SPT4, a gene important for transcription initiation in Saccharomyces cerevisiae. Mol Gen Genet 237, 449-459.

McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., Read, R.J., 2005, Likelihood-enhanced fast translation functions. Acta crystallographica 61, 458-464.

Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S., Cramer, P., 2005, A structural perspective of CTD function. Genes Dev 19, 1401-1415.

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H., Severinov, K., 2001, Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. Proceedings of the National Academy of Sciences of the United States of America 98, 892-897.

Mooney, R.A., Schweimer, K., Rosch, P., Gottesman, M., Landick, R., 2009, Two structurally independent domains of E. coli NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. J Mol Biol 391, 341-358.

Nehrke, K.W., Platt, T., 1994, A quaternary transcription termination complex. Reciprocal stabilization by Rho factor and NusG protein. J Mol Biol 243, 830-839.

Nicholson, H., Becktel, W.J., Matthews, B.W., 1988, Enhanced protein thermostability from designed mutations that interact with alpha-helix dipoles. Nature 336, 651-656.

Nickels, B.E., 2009, Genetic assays to define and characterize protein-protein interactions involved in gene regulation. Methods 47, 53-62.

Otwinowski, Z.M., W., 1996, Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276, 307-326.

Poglitsch, C.L., Meredith, G.D., Gnatt, A.L., Jensen, G.J., Chang, W.H., Fu, J., Kornberg, R.D., 1999, Electron crystal structure of an RNA polymerase II transcription elongation complex. Cell 98, 791-798.

Ponting, C.P., 2002, Novel domains and orthologues of eukaryotic transcription elongation factors. Nucleic Acids Res 30, 3643-3652.

Rechkoblit, O., Malinina, L., Cheng, Y., Kuryavyi, V., Broyde, S., Geacintov, N.E., Patel, D.J., 2006, Stepwise translocation of Dpo4 polymerase during error-free bypass of an oxoG lesion. PLoS Biol 4, e11.

Richardson, L.V., Richardson, J.P., 2005, Identification of a structural element that is essential for two functions of transcription factor NusG. Biochim Biophys Acta 1729, 135-140.

Sakurai, H., Mitsuzawa, H., Kimura, M., Ishihama, A., 1999, The Rpb4 subunit of fission yeast Schizosaccharomyces pombe RNA polymerase II is essential for cell viability and

similar in structure to the corresponding subunits of higher eukaryotes. Mol Cell Biol 19, 7511-7518.

Saunders, A., Core, L.J., Lis, J.T., 2006, Breaking barriers to transcription elongation. Nat Rev Mol Cell Biol 7, 557-567.

Saxowsky, T.T., Doetsch, P.W., 2006, RNA polymerase encounters with DNA damage: transcription-coupled repair or transcriptional mutagenesis? Chemical reviews 106, 474-488.

Schneider, D.A., French, S.L., Osheim, Y.N., Bailey, A.O., Vu, L., Dodd, J., Yates, J.R., Beyer, A.L., Nomura, M., 2006, RNA polymerase II elongation factors Spt4p and Spt5p play roles in transcription elongation by RNA polymerase I and rRNA processing. Proceedings of the National Academy of Sciences of the United States of America 103, 12707-12712.

Sosunov, V., Sosunova, E., Mustaev, A., Bass, I., Nikiforov, V., Goldfarb, A., 2003, Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. The EMBO journal 22, 2234-2244.

Steiner, T., Kaiser, J.T., Marinkovic, S., Huber, R., Wahl, M.C., 2002, Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. The EMBO journal 21, 4641-4653.

Steitz, T.A., 1998, A mechanism for all polymerases. Nature 391, 231-232.

Strauss, B.S., 1991, The 'A rule' of mutagen specificity: a consequence of DNA polymerase bypass of non-instructional lesions? Bioessays 13, 79-84.

Studier, F.W., 2005, Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41, 207-234.

Sullivan, S.L., Gottesman, M.E., 1992, Requirement for E. coli NusG protein in factor-dependent transcription termination. Cell 68, 989-994.

Sydow, J.F., Brueckner, F., Cheung, A.C., Damsma, G.E., Dengl, S., Lehmann, E., Vassylyev, D., Cramer, P., 2009, Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. Molecular cell 34, 710-721.

Takahara, P.M., Rosenzweig, A.C., Frederick, C.A., Lippard, S.J., 1995, Crystal structure of double-stranded DNA containing the major adduct of the anticancer drug cisplatin. Nature 377, 649-652.

Taylor, J.S., 2002, New structural and mechanistic insight into the A-rule and the instructional and non-instructional behavior of DNA photoproducts and other lesions. Mutat Res 510, 55-70.

Thomas, M.C., Chiang, C.M., 2006, The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol 41, 105-178.

Thomas, M.J., Platas, A.A., Hawley, D.K., 1998, Transcriptional fidelity and proofreading by RNA polymerase II. Cell 93, 627-637.

Todone, F., Brick, P., Werner, F., Weinzierl, R.O., Onesti, S., 2001, Structure of an archaeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex. Molecular cell 8, 1137-1143.

Tornaletti, S., 2009, DNA repair in mammalian cells: Transcription-coupled DNA repair: directing your effort where it's most needed. Cell Mol Life Sci 66, 1010-1020.

Tornaletti, S., Maeda, L.S., Kolodner, R.D., Hanawalt, P.C., 2004, Effect of 8-oxoguanine on transcription elongation by T7 RNA polymerase and mammalian RNA polymerase II. DNA repair 3, 483-494.

Tornaletti, S., Patrick, S.M., Turchi, J.J., Hanawalt, P.C., 2003, Behavior of T7 RNA polymerase and mammalian RNA polymerase II at site-specific cisplatin adducts in the template DNA. The Journal of biological chemistry 278, 35791-35797.

Tremeau-Bravard, A., Riedl, T., Egly, J.M., Dahmus, M.E., 2004, Fate of RNA polymerase II stalled at a cisplatin lesion. The Journal of biological chemistry 279, 7751-7759.

Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F., Murshudov, G.N., 2004, REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta crystallographica 60, 2184-2195.

Vassylyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S., Yokoyama, S., 2002, Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. Nature 417, 712-719.

Vassylyev, D.G., Vassylyeva, M.N., Perederina, A., Tahirov, T.H., Artsimovitch, I., 2007a, Structural basis for transcription elongation by bacterial RNA polymerase. Nature 448, 157-162.

Vassylyev, D.G., Vassylyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I., Landick, R., 2007b, Structural basis for substrate loading in bacterial RNA polymerase. Nature 448, 163-168.

Viswanathan, A., Doetsch, P.W., 1998, Effects of nonbulky DNA base damages on Escherichia coli RNA polymerase-mediated elongation and promoter clearance. The Journal of biological chemistry 273, 21276-21281.

Viswanathan, A., Liu, J., Doetsch, P.W., 1999, E. coli RNA polymerase bypass of DNA base damage. Mutagenesis at the level of transcription. Ann N Y Acad Sci 870, 386-388.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., Buratowski, S., Handa, H., 1998, DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes Dev 12, 343-356.

Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D., Kornberg, R.D., 2006, Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. Cell 127, 941-954.

Wang, D., Lippard, S.J., 2005, Cellular processing of platinum anticancer drugs. Nat Rev Drug Discov 4, 307-320.

Weaver, R.F., 2008, Molecular biology, 4th Edition. McGraw-Hill, Boston, xxii, 890 p. 829 cm. pp.

Westover, K.D., Bushnell, D.A., Kornberg, R.D., 2004, Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. Cell 119, 481-489.

Wind, M., Reines, D., 2000, Transcription elongation factor SII. Bioessays 22, 327-336.

Yamaguchi, Y., Mura, T., Chanarat, S., Okamoto, S., Handa, H., 2007, Hepatitis delta antigen binds to the clamp of RNA polymerase II and affects transcriptional fidelity. Genes Cells 12, 863-875.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., Handa, H., 1999a, NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell 97, 41-51.

Yamaguchi, Y., Wada, T., Watanabe, D., Takagi, T., Hasegawa, J., Handa, H., 1999b, Structure and function of the human transcription elongation factor DSIF. The Journal of biological chemistry 274, 8085-8092.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., Darst, S.A., 1999, Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. Cell 98, 811-824.

Zhou, H., Liu, Q., Gao, Y., Teng, M., Niu, L., 2009, Crystal structure of NusG N-terminal (NGN) domain from Methanocaldococcus jannaschii and its interaction with rpoE". Proteins 76, 787-793.

Zhou, L., Deng, H., Deng, Q., Zhao, S., 2004, A mixed matrix of 3-hydroxypicolinic acid and pyrazinecarboxylic acid for matrix-assisted laser desorption/ionization time-of-flight mass spectrometry of oligodeoxynucleotides. Rapid Commun Mass Spectrom 18, 787-794.

9   Curriculum Vitae - Gerke Luinge-Damsma

Personal Information

Full name:	Gerke Elisabeth Luinge-Damsma
Surname:	Luinge-Damsma
First name:	Gerke
Gender:	Female
Nationality:	Dutch
Address:	Jollystr. 13
Place:	81545 Munich, Germany
Date of birth:	August, 13th 1982
Place of birth:	Hellendoorn, The Netherlands
E-mail:	damsma@lmb.uni-muenchen.de
Marital status:	Married
Foreign languages:	English, German

Education

2006-present   PhD student, Gene Center Munich, Ludwig-Maximilians-Universität München. Member in the group of prof. Patrick Cramer. PhD project: Mechanisms of transcriptional mutagenesis.

2008   Winner of the Römer prize from the Department of Chemistry and Biochemistry, Ludwig-Maximilians-University Munich in the category dissertation.

since 2006   Member of the international graduate program "NanoBioTechnology" as part of the Elite Network of Bavaria (ENB).

2004-2006   Topmaster Biomolecular Sciences, Groningen Biomolecular Sciences and Biotechnology Institute (GBB) and University of Groningen.
               First master project in the group of prof. Dick Janssen (University of Groningen): Cloning and expression of a novel lactose oxidase.

Second master project in the group of prof. Dieter Oesterhelt (MPI, Munich): Investigation of the Arginine Deiminase Pathway using proteomic tools.

2001-2004   Bachelor Chemistry, University of Groningen.
2001: Propaedeutic certificate
2003: Bachelor in Biochemistry;
Bachelor project in the group of prof. Bert Poolman (University of Groningen): Investigation of the OpuA transporter in *Lactococcus lactis*.

2000-2001   EF International Language School, San Diego (USA)
Intensive English School, including grammar, literature and public speaking.

1994-2000   College Reggesteyn, Nijverdal (The Netherlands), High School diploma, with the main subjects: Dutch, English, Mathematics A, Mathematics B, Economics, Physics, Chemistry and Biology

Publications

Damsma, G.E., Cramer, P. (2009) Molecular basis of transcriptional mutagenesis at 8-oxoguanine. *J. Biol. Chem.* **284**(46): 31658-31663.

Sydow, J. F., Brueckner, F., Cheung, A. C., Damsma, G. E., Dengl, S., Lehmann, E., Vassylyev, D., Cramer, P. (2009). Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol Cell* **34**(6):710-21.

Brueckner, F., Armache, K. J., Cheung, A., Damsma, G. E., Kettenberger, H., Lehmann, E., Sydow, J. F., Cramer, P. (2009). Structure-function studies of the RNA polymerase II elongation complex. *Acta Crystallogr D Biol Crystallogr.* **65**, 112-120.

Cramer, P., Armache, K.-J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C.-D., Lehmann, E., Leike, K., Sydow, J. and Vannini, A. (2008). Structure of Eukaryotic RNA Polymerases. *Annu. Rev. Biophys.* **37**, 337-352.

Heuts D.P., Winter R.T., <u>Damsma G.E.</u>, Janssen D.B., Fraaije M.W. (2008). The role of double covalent flavin binding in chito-oligosaccharide oxidase from *Fusarium graminearum*. *Biochem J.* **1**; 413(1):175-83

<u>Damsma, G.E.</u>, Alt, A., Brueckner, F., Carell, T., Cramer, P. (2007). Mechanism of transcriptional stalling at cisplatin-damaged DNA. *Nat Struct Mol Biol* **14**, 1127-33.

Conferences/Courses

2009 Workshop speedreading (Munich, Germany)

2009 Workshop self-management (Munich, Germany)

2008 EMBO conference. 'Gene transcription in yeast'. Including poster presentation: Mechanism of transcriptional stalling at cisplatin-damaged DNA. (St. Feliu de Guixols, Spain)

2008 CCP4 conference. 'Low Resolution Structure Determination and Validation'. (Leeds, UK)

2007 CeNS Winter School. 'Nanosystems: From Quantum Devices to Biological Engines' Including poster presentation: How does RNA polymerase II respond to DNA damage. (Mauterndorf, Austria)

2006 GBB Masterclass. 'Proteomics and Structural Genomics of Membrane Proteins'. Including oral presentation: Investigation of the Arginine Deiminase Pathway in *Halobacterium salinarum* using proteomic tools. (Groningen, Netherlands)

2006 Q-TOF user meeting, Waters (Munich, Germany)

2005 EGC Masterclass. 'Protein structure and structure-function relationships'. (Groningen, Netherlands)