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

Nucleosome-Chd1 structure and implications for chromatin remodelling

Lucas Farnung
aus
Dortmund, Deutschland
2017

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

Nucleosome-Chd1 structure and implications for chromatin remodelling

Lucas Farnung
aus
Dortmund, Deutschland
2017

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Cramer betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

Göttingen, den 29.09.2017

Lucas Farnung

Dissertation eingereicht am: 09.10.2017
Erstgutachter: Prof. Dr. Patrick Cramer
Zweitgutachter: PD Dr. Dietmar Martin
Tag der mündlichen Prüfung: 06.11.2017

Acknowledgements

I want to thank Patrick Cramer for providing an outstanding research environment that allows the tackling of difficult problems in molecular biology. Patrick allows young scientists to strive and grow while providing excellent advice and guidance, making the stay in his laboratory an invaluable experience.

I want to explicitly thank a number of former and current members of the Cramer lab: Franz Fischer for insect cell expression and help with purifying full-length Chd1 for the first time, Carina Burzinski for purification of Pol II, Ute Neef for production of baculovirus and insect cell expression, Hauke Hillen and Kayo Nozawa for collection of X-ray crystallographic data and advice on data processing of X-ray crystallographic data, Youwei Xu for cloning and establishing Paf1C purification in the lab, and Dimitry Tegunov for providing software packages used in the validation of cryo-EM data. I want to thank Christoph Wigge for providing support, help and advice for cryo-EM sample preparation and data collection. His help allowed quick and high-quality data acquisition that ultimately resulted in the NCP-Chd1 structure.

Solving challenging problems requires both courage and brilliance, I was incredibly fortunate to meet Seychelle M. Vos who unifies both qualities and was able to learn from her vast biochemical knowledge. Without the constant scientific exchange with Seychelle, it would have been impossible to realize the work presented here. The act of advancing science is not just a mere accumulation of additional knowledge. Sometimes it transmits something in addition. These are the moments that are truly important. Thank you.

I am thankful to the members of my thesis committee: Dr. Dietmar Martin, Prof. Dr. Karl-Peter Hopfner, Dr. Philipp Korber, Prof. Dr. Andreas Ladurner, and Prof. Dr. Veit Hornung.

I am incredibly grateful to my family. Without the support of my parents and my brother, completion of this endeavor would have not been possible.

Summary

Nuclear eukaryotic genomic DNA is organized in a structure called chromatin, which is composed of a fundamental unit, the nucleosome. The nucleosome is a protein-DNA complex consisting of a histone octamer core with 145-147 basepairs of DNA wrapped around it.

Here, I report the 4.8 Å resolution cryo-electron microscopy reconstruction of the single-subunit full-length *S. cerevisiae* chromatin remodeller Chd1 in complex with a *X. laevis* nucleosome core particle in the presence of the transition state mimicking adduct ADP·BeF_3. The nucleosome was formed with Widom 601 nucleosomal DNA and 63 bp of extranucleosomal DNA. This is the first high-resolution structure of a full-length chromatin remodeller bound to the nucleosome. The structure reveals an altered nucleosome where two helical turns of DNA are peeled away from the histone octamer. The cryo-EM structure elucidates the mechanism of Chd1 regulation by the regulatory double chromodomain of the chromatin remodeller to control efficient chromatin remodelling. The structure allows for a model describing the molecular mechanism of the Chd1 ATPase motor, where ATP binding leads to closure of the ATPase motor and translocation of DNA in one base pair steps. ATP hydrolysis and dissociation of ADP resets the ATPase motor to allow for a new round of the enzymatic cycle. Chd1 is a 3'-5' DNA translocase that shifts DNA towards the octamer dyad, resulting in centering of the nucleosome. My structure provides the foundation for understanding the mechanism of chromatin remodelling. The gained insights extend beyond Chd1 to other CHD family members, providing a framework to interpret other chromatin remodellers.

In addition, I report the crystal structure of a *S. cerevisiae* RNA polymerase II elongation complex with the modified base 3d-Napht-A in the +1 site. The crystal structure was solved at a resolution of 3.2 Å allowing unambiguous placement of 3d-Napht-A. The structure revealed that the modified base impairs closure of the trigger loop and prevents translocation, blocking further elongation. The structure provides the molecular basis for understanding the effects of minor groove DNA alkylations on transcription and provides a platform for the targeted design of drugs to impair transcription for therapeutic benefit.

In conclusion, my studies provide an initial framework for the structural understanding of chromatin remodelling and extend our knowledge of RNA polymerase II stalling in certain mutational contexts.

Publications

Part of this work has been published or is in the process of being published:

Lucas Farnung, Seychelle M. Vos, Christoph Wigge, Patrick Cramer. (2017) Nucleosome-Chd1 structure and implications for chromatin remodelling. *Nature.* doi: 10.1038/nature24046 (in press)

Author contributions: **LF designed and carried out experiments and performed cryo-EM data acquisition and analysis.** SMV developed the protein expression strategy, performed baculovirus production, and insect cell expression. CW assisted with cryo-EM grid preparation and data collection. PC designed and supervised research. **LF and PC interpreted the data and wrote the manuscript, with input from all authors.**

Stefano Malvezzi*, **Lucas Farnung***, Claudia Aloisi, Todor Angelov, Patrick Cramer, Shana J. Sturla. (2017) Mechanism of RNA polymerase II stalling by DNA alkylation *PNAS.* (accepted in principle)
*Both authors contributed equally.

Author contributions: S.M. synthesized the modified DNA strands, and performed and analyzed the experiments using isolated enzyme. **L.F. designed and performed all structural experiments including crystallization trials, data processing, model building and interpretation.** C.A. designed and performed the cell studies. T.A. synthesized and purified modified phosphoramidites. P.C. and S.J.S. designed and supervised research. **S.M., L.F., P.C., and S.J.S. wrote the manuscript.**

Additional publications:

Veronika Fitz, Jaeoh Shin, Christoph Ehrlich, **Lucas Farnung**, Patrick Cramer, Vasily Zaburdaev, Stephan W. Grill. (2016) Nucleosomal arrangement affects single-molecule transcription dynamics. *PNAS* doi: 10.1073/pnas.1602764113

Author contributions: V.F., P.C., V.Z., and S.W.G. designed research; **V.F., J.S., and L.F. performed research**; **V.F., C.E., and L.F. contributed new reagents/analytic tools**; V.F. and J.S. analyzed data; and V.F., J.S., P.C., V.Z., and S.W.G. wrote the paper.

Fangjie Zhu, **Lucas Fanung**, Yimeng Yin, Eevi Kaasinen, Svetlana Dodonova, Patrick Cramer, and Jussi Taipale. (2017) The interaction landscape between transcription factors and the nucleosome. (submitted)

Author contributions: J.T., F.Z. and P.C. conceived the experiments. F.Z. performed the NCAP-SELEX, MNase-seq and data analysis. **L.F produced the histone octamers.** Y.Y. helped with the curation and production of TF proteins. E.K. analysed the MNase-seq data. S.D. performed EMSA validation for SOX. F.J. and J.T. interpreted the data and wrote the manuscript. All authors discussed the findings and contributed to the manuscript.

Table of Contents

1 Introduction

Deoxyribonucleic acid (DNA) stores the genetic information required for all cellular processes. Genomes are inherently large and must be compacted and organized to fit within their host cell. Bacteria fit their ∼3 mm of genomic DNA within 0.5-1 μm using proteineacous factors within the cytoplasm. Eukaryotes can have genomes that reach 150 billion base pairs (equivalent to 91 meters of non-compacted DNA) and are segregated in an organelle known as the nucleus. To allow for proper organization, eukaryotes organize their genomes in a spatio-temporal manner employing a protein-DNA structure called chromatin. Although compaction of the DNA is necessary for storage, accessibility is also required to allow for the proper function of cellular processes such as transcription, DNA replication, and repair.

1.1 Chromatin

Twenty years ago, the Richmond lab solved the crystal structure of a histone octamer from *X. laevis* bound to a 146 base pair palindromic DNA from α-satellite DNA (Luger et al., 1997). This first structure of the nucleosome core particle presented a breakthrough in the understanding of chromatin at an atomic resolution. A plethora of nucleosome structures from different organisms have been solved over the last two decades highlighting a conserved architecture of the nucleosome that extends from Archaea to apes (Luger et al., 1997; White et al., 2001; Tsunaka et al., 2005; Vasudevan et al., 2010). In contrast, crystallizing nucleosomes with additional factors bound proved extremely challenging and resulted only in a handful of structures (Barbera et al., 2006; Makde et al., 2010; Tan and Davey, 2011; Armache et al., 2011; Kato et al., 2013; McGinty et al., 2014). With the recent advent of high-resolution single particle cryo-electron microscopy (cryo-EM), it is now possible to capture structures of nucleosomes bound to individual factors or within larger assemblies.

1.1.1 Organization of the genome in eukaryotes

Eukaryotes organize their nuclear genome in a densely packed but readily accessible structure called chromatin. Chromatin is organized in a hierarchical structure. The fundamental

unit is the nucleosome (Kornberg, 1974). The nucleosomes form a "beads-on-a-string"-like structure and multiple nucleosomes pack into higher order structures in the presence of linker histones (e.g. H1) (Simpson, 1978). Both *in vitro* and *in vivo* studies converge on the importance of the tetranucleosome as a packaging unit (Schalch et al., 2005). However, it remains elusive if the 30 nm chromatin fiber, a hallmark of higher order packaging *in vitro*, also exists *in vivo* (Ou et al., 2017). Nucleosomal arrays are capable of compaction and can form even more compact structures that are mostly inaccessible to the transcription and DNA replication machinery (heterochromatin) (van Steensel, 2011). The highest and most compacted state of chromatin is found in the metaphase chromosome.

In addition to its packaging function, chromatin also plays a role in preventing DNA damage and provides regulated access to the genome for important cellular processes such as DNA replication and DNA-dependent transcription of genes.

1.1.2 Architecture of the nucleosome

The nucleosome is a DNA-protein complex, consisting of ∼147 bp of DNA that are wrapped around a protein core (Figure 1) (Kornberg, 1974; Luger et al., 1997). The protein core consists of the core histones H2A, H2B, H3, and H4. The core histones share a common architecture. The histones contain a centrally located histone-fold motif with N- and C-terminal extensions. The histone fold itself consists of three α helices that are connected by two flexible loops. The shorter first and third helix pack against the longer central helix. H2A and H2B form dimers, whereas two H3 and two H4 histones form a tetramer (Figure 1) (Arents et al., 1991). Two H2A-H2B dimers and one H3-H4 tetramer together form the histone octamer around which 1.65 turns of DNA are wrapped in a left-handed superhelical arrangement.

Histone-DNA contacts are mediated by direct hydrogen bonds, ionic interactions, and nonpolar contacts including hydrogen bonds from water (McGinty and Tan, 2015). The DNA encloses the histone octamer in two parallel DNA gyres with one side of the nucleosome showing a diagonally running DNA that connects the two gyres of nucleosomal DNA. Overall, this architectural arrangement results in a pseudo-twofold axis which is referred to as the dyad axis (Flaus et al., 1996). DNA locations are labeled by their superhelical locations (SHL) with the location at the dyad axis being designated SHL 0. SHLs range from -7 to +7. The flexible regions of the histones at their N- and C-terminal ends (called tails) can be covalently post-translationally modified with modifications including phosphorylations, acetylation, methylations, and ubiquitinations (Taverna et al., 2007). These modifications are established or removed by proteins or protein complexes called histone

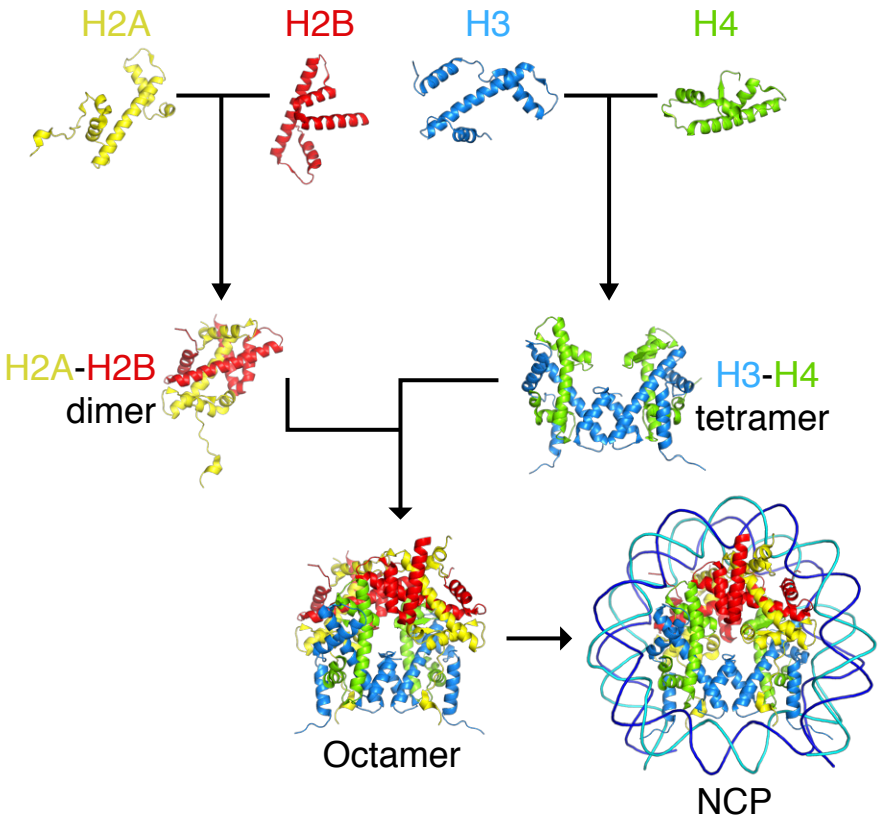

Figure 1: Architecture of the nucleosome core particle. The nucleosome core particle consists of the histones H2A, H2B, H3, and H4. 147 bp of DNA are wrapped around the histone octamer in a left-handed helix.

writers and erasers. Modifications can be read by histone readers (Allis and Jenuwein, 2016).

A hotspot for binding of chromatin factors to the nucleosome is the H2A/H2B acidic patch. The acidic patch, a signature feature of the nucleosome, is a surface formed by the H2A/H2B dimer. The surface is formed by ten acidic residues and is a prominent surface for the binding of chromatin factors. In fact, all available crystal structures of factors bound to the nucleosome show an interaction between the factor and the acidic patch (Barbera et al., 2006; Makde et al., 2010; Armache et al., 2011; Kato et al., 2013; McGinty et al., 2014; Fang et al., 2016).

1.1.3 ATP-dependent chromatin remodellers

The compact packaging of genomic DNA around the nucleosome requires molecular machines to facilitate and regulate access to the DNA by shifting nucleosomes for important cellular processes such as DNA replication and transcription. These molecular machines are called chromatin remodellers. The shifting of nucleosomes happens in an ATP-dependent manner, driven by an ATPase motor that shows DNA translocase activity. Chromatin remodellers are essential for the proper maintenance of chromosome function and their deregulation results in a variety of diseases including cancer and neurodevelopmental disorders (Weiss et al., 2016; Sugathan et al., 2014).

The ATPases involved in chromatin remodelling are part of the superfamily 2 RNA/DNA helicase family and are classified in four sub-groups (Clapier et al., 2017; Flaus et al., 2006): the chromodomain helicase-DNA binding (CHD) (Tran et al., 2000), the switch/sucrose non-fermentable (SWI/SNF) (Neigeborn and Carlson, 1984), the inositol requiring mutant 80 (INO80) (Ebbert et al., 1999; Shen et al., 2000), and the imitation switch (ISWI) subfamily (Elfring et al., 1994) (Figure 2). All chromatin remodellers share a common ATP hydrolysis-driven DNA translocase motor with two RecA-like lobes which bind the nucleosome at SHL $\pm$ 2 (Schwanbeck et al., 2004). Differences in their phenotype are explained by additional auxiliary domains and factors that play a regulatory role for the activity of the chromatin remodeller.

Chromatin remodeller subfamilies

The SWI/SNF subfamily contains a centrally located ATPase motor flanked N-terminally by a HSA (helicase/SANT-associated) and a post-HSA domain and C-terminally by an AT-hook and a bromodomain (Figure 2a). The HSA domain has been shown to bind actin or actin-related proteins (Schubert et al., 2013). One of the most prominent members of this subfamily is the RSC complex (Cairns et al., 1996). The SWI/SNF subfamily facilitates access to the chromatin by sliding and ejecting nucleosomes from the DNA (Lorch et al., 1999) and is responsible for activating or repressing gene expression (Mitra et al., 2006).

Whereas the two RecA-like lobes of the SWI/SNF subfamily only contain a short linker between them, the INO80 subfamily of chromatin remodellers contain a variable linker length between the two lobes of the ATPase motor (Figure 2b) (Morrison and Shen, 2009). The N-terminus is organized similarly to the SWI/SNF subfamily with a HSA and a post-HSA domain. Chromatin remodellers belonging to the INO80 subfamily play critical roles in nucleosomal spacing and accessibility. They are able to edit nucleosomes by evicting H2A-H2B dimers in exchange for other histone variants (Mizuguchi et al.,

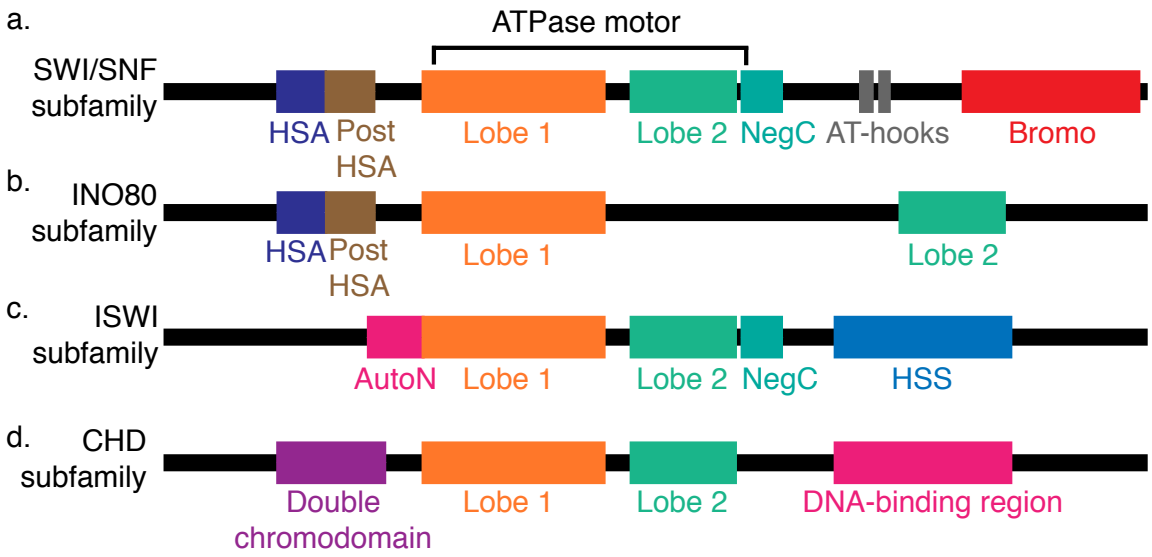

Figure 2: Domain architecture of the chromatin remodeller subfamilies SWI/SNF, INO80, ISWI, and CHD. Domains present in the subfamilies are indicated and colored. This figure has been adapted from Clapier et al. (2017).

2004). Through this editing activity, the INO80 subfamily plays a role in DNA repair by replacing the histone variant H2A.X (van Attikum et al., 2007). Remodellers of the INO80 subfamily have also been implicated in telomere maintenance (Yu et al., 2007) and chromosome segregation (Krogan et al., 2004).

The ISWI subfamily ATPase motor is flanked by a regulatory AutoN domain at the N-terminus and a regulatory NegC domain at the C-terminus (Figure 2c) (Dang and Bartholomew, 2007). A HSS (HAND-SANT-SLIDE) domain located at the C-terminus is capable of binding DNA as well as H3 tails (Boyer et al., 2004). ISWI complexes assemble and space nucleosomes on DNA. The subfamily plays a critical role in gene expression by regulating access to chromatin (Hochheimer et al., 2002).

The CHD subfamily and its member Chd1 in *S. cerevisiae* will be discussed in a specific subsection (Figure 2d).

Mechanism of DNA translocation

Chromatin remodelling is defined by the energy landscape of the nucleosome. Biophysical assays revealed that the DNA is positioned on the nucleosome with a relative stability of approximately 11-14 kcal mole^{-1} (Gottesfeld and Luger, 2001). This stability is conferred by a total of 14 histone-DNA contacts, yielding an energy contribution of $\sim$1 kcal mole^{-1}

per contact of DNA with the octamer. Estimation of forces required to break one of these interactions has been estimated to be on the order of ∼1 pN. ATP provides approximately 7 kcal mole^{-1} of Gibbs free energy, thus chromatin remodellers are not able to provide the required energy to translocate DNA and remodel a full nucleosome with a single ATP hydrolysis event. Multiple rounds of ATP hydrolysis are required to successfully remodel a nucleosome.

Since structures of full-length chromatin remodellers bound to nucleosomes are not yet available, current models describing the mechanism for DNA translocation are primarily derived from structural and biochemical characterization of SF2 family helicases (Dürr et al., 2006; Sengoku et al., 2006; Gu and Rice, 2010). The conserved ATPase motor is formed by the two RecA-like lobes which together form a DNA binding surface and constitute an ATP binding site. The two lobes bind the DNA in an arrangement where both bind the same strand (tracking strand) while the lobes are slightly offset. ATP binding and hydrolysis would then result in a movement of the lobes and cause DNA translocation (Saha et al., 2002). In this model, the ATPase motor moves unidirectionally relative to the DNA. If the remodeller is held at a fixed position, the DNA is translocated in one direction while the ATPase motor "walks" in the opposite direction.

Biochemical evidence suggests that the motor movement requires one ATP hydrolysis event per 1-2 bp of translocated DNA (Zhang et al., 2006; Harada et al., 2016). On a nucleosome, the ATPase motor seems to be stably bound at SHL 2 (Kagalwala et al., 2004; Schwanbeck et al., 2004). DNA translocation would then result in creation of torsional stress of DNA on the entry and exit side of nucleosomal DNA. Based on these assumptions, a model called the wave-ratchet-wave model has been proposed which is signified by one-dimensional diffusion of DNA along the nucleosome to shift nucleosomes (Clapier et al., 2017). The model proposes that the introduction of torsional stress by the ATPase motor breaks histone-DNA contacts. The DNA segment closer to the dyad then propagates in a wave-like fashion toward the DNA exit side, while the torsional restraints on the other side of the ATPase motor results in the "pulling in" of entry side DNA (Saha et al., 2005). This ultimately resolves the torsional stress introduced by the chromatin remodeller. It remains, however, unclear what size such a wave has (Saha et al., 2006). Many repeated enzymatic cycles of the ATPase motor eventually result in shifting or eviction of the nucleosome.

Chd1

Chd1 is a member of the CHD subfamily of chromatin remodellers. This subfamily is named after its unique tandem chromodomains that are arranged at the amino terminus

of the remodeller. This double chromodomain is followed by the conserved ATPase motor consisting of two RecA-like lobes. It is followed by a DNA-binding region with two domains (SANT and SLIDE). Recently, an additional domain was discovered at the carboxy terminus of *S. cerevisiae* Chd1 called CHCT (Mohanty et al., 2016). Whereas the CHD subfamily has eight members in *H. sapiens* that have different additional auxiliary domains, the baker's yeast *S. cerevisiae* only has one member in the CHD subfamily, Chd1. CHD chromatin remodellers have been implicated in a range of cellular processes including maintenance of pluripotency in stem cells (Gaspar-Maia et al., 2009) and Chd1-dependent emergence of hematopoietic progenitor cells from endothelial cells (Koh et al., 2015). Chd1 plays also a role during transcription (Zentner et al., 2013). System-wide genomic studies have shown that Chd1 is located both at the promotor and within the coding regions of genes, probably accompanying RNA polymerase II to maintain and optimize nucleosomal localization during transcription (Zentner et al., 2013; Smolle et al., 2012).

Chd1 in *S. cerevisiae* is a single-subunit chromatin remodeller with the classical CHD subfamily domain organization. The CHCT domain has a positively charged surface and is able to interact with DNA and nucleosomes (Mohanty et al., 2016). Whereas the double chromodomain of human CHD1 binds H3 tails trimethylated at residue K4, yeast Chd1 cannot bind the modification due to a mutation in the aromatic cage that is required for binding the modified tail (tryptophan in human substituted to glutamate in yeast) (Sims et al., 2005). Chd1 has been shown to be recruited to open reading frames by the transcription elongation factor complex Paf1C (polymerase associated factor 1 complex) through its subunit Rtf1 (Simic et al., 2003).

Chd1 shifts nucleosomes away from DNA ends and biochemical data suggests that this characteristic is mediated by the DNA-binding region (McKnight et al., 2011; Patel et al., 2013). Chd1 shows ATPase activity on linear DNA, however, it is greatly stimulated in the presence of nucleosomes, suggesting that Chd1 can distinguish linear DNA from nucleosomal DNA. Recently, the Bowman lab has shown via site-specific cross-linking experiments that Chd1 binds both DNA gyres and can "communicate over an ∼90 bp loop of nucleosomal DNA" (Nodelman et al., 2017). Structural data for *S. cerevisiae* Chd1 exists. A crystal structure of the double chromodomain with the ATPase motor revealed an autoinhibited state where a region in the double chromodomain (the acidic wedge) sequesters a basic region in ATPase lobe 2 (Hauk et al., 2010) (Figure 3a). This forces ATPase lobe 2 to stay in a conformation that is not catalytically competent. A model proposes that the presence of nucleosomal DNA could alleviate this process of inhibition. Additionally, a crystal structure of the DNA-binding region of Chd1 with the SANT and SLIDE domains was solved in 2011 (Sharma et al., 2011) (Figure 3b). The structure revealed that

the SANT and SLIDE domains preferentially bind linear DNA which is consistent with a preference of the DNA-binding region for extranucleosomal DNA.

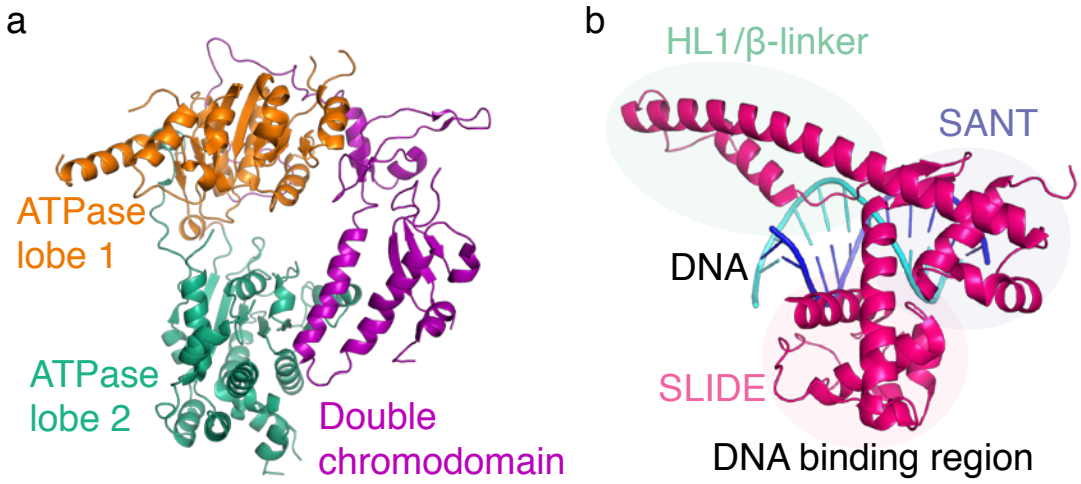

Figure 3: Crystal structures of the domains of Chd1. **a.** Crystal structure of the ATPase motor with the double chromodomain of Chd1 (PDB code 3MWY, Hauk et al. (2010)). ATPase lobe1, lobe2 and the double chromodomain are indicated in orange, green, and purple. The acidic wedge in the double chromodomain interacts with a basic region of ATPase lobe 2. **b.** Crystal structure of the DNA-binding region of Chd1 in complex with double-stranded DNA (PDB code 3TED, Sharma et al. (2011)). The SANT, SLIDE, and the HL1/β-linker are indicated with blue, magenta, and green ovals. The DNA-binding region is colored in pink.

1.2 Transcription

Transcription is the first step of gene expression, whereby RNA chain assembly is catalysed by DNA-dependent RNA polymerases. RNA polymerases are conserved in prokaryotes, archaea, and eukaryotes (Werner and Grohmann, 2011). Whereas bacteria and archaea only possess one RNA polymerase (Pol), eukaryotes have RNA polymerase I, II and III (Roeder and Rutter, 1969). Plants even possess two additional RNA polymerases (IV and V). RNA polymerase I and III primarily transcribe rRNA (Pol I and Pol III), tRNAs (Pol III) and other small RNAs (Pol III). Pol II synthesizes nascent mRNA, snRNA and microRNAs. RNA polymerase II is a $\sim$500 kDa protein complex, consisting of 12 subunits. Pol II activity is tightly regulated by a great number of transcription factors to tune

transcription of genes depending on the environmental context. I will discuss factors and activities of the Pol II transcription cycle (Figure 4).

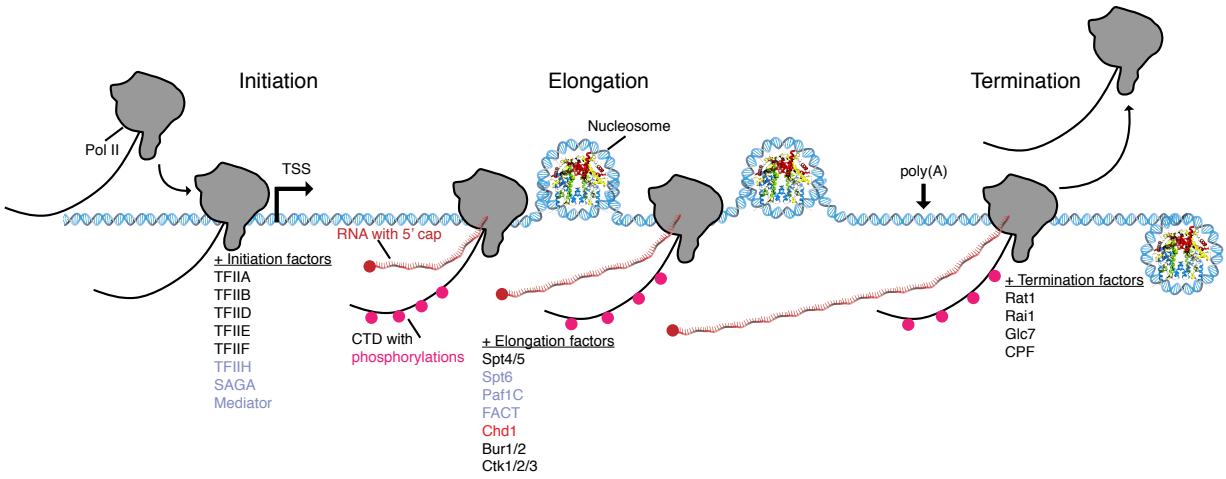

Figure 4: Schematic of the transcription cycle. Initiation, elongation and termination are shown schematically. Yeast factors that have been shown to be involved in the transcription cycle are shown. Factors that interact with chromatin are shown in light blue. Chd1 is highlighted in red.

1.2.1 Transcription initiation

Transcription of protein-coding genes starts with the first step of the transcription cycle, called transcription initiation (Hantsche and Cramer, 2017). During transcription initiation in the Pol II system, general transcription factors and RNA polymerase II assemble together on a promoter (Sainsbury et al., 2015). This process has been studied extensively using functional and structural approaches (Sainsbury et al., 2013; He et al., 2016; Plaschka et al., 2016). RNA polymerase II assembles with TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH on promotor DNA. Most protein-coding genes, additionally, are controlled by the co-activator Mediator (Plaschka et al., 2015). Pol II and the initiation factors melt the DNA in a nucleotide dependent manner. The melted DNA allows for Pol II to initiate nascent RNA synthesis from within a ”transcription bubble”.

1.2.2 Transcription elongation

After transcription initiation and promoter escape, RNA polymerase II enters the elongation phase of the transcription cycle (Jonkers and Lis, 2015). During this phase of

transcription, RNA polymerase II has escaped the promoter, maintains the transcription bubble and catalyses the addition of nucleotides to the growing 3' end of RNA by a Brownian ratchet mechanism (Martinez-Rucobo and Cramer, 2013). The molecular mechanism of elongation and its regulation have been studied in a wide range of crystallographic studies in the yeast system as well as the bacterial system, revealing a conserved catalytic center and mechanism of transcription among bacteria and eukaroytes (Figure 5) (Cramer et al., 2001; Vassylyev et al., 2007a).

Elongation undergoes allosteric regulation by a range of elongation factors such as Spt4/5 in yeast (DSIF in human), Spt6, Elf1, TFIIS, and Paf1C (Clark-Adams and Winston, 1987). Recently, the study of transcription elongation has expanded to cryo-EM to better understand the regulation of this process in the presence of additional regulatory factors (Xu et al., 2017; Bernecky et al., 2017; Ehara et al., 2017).

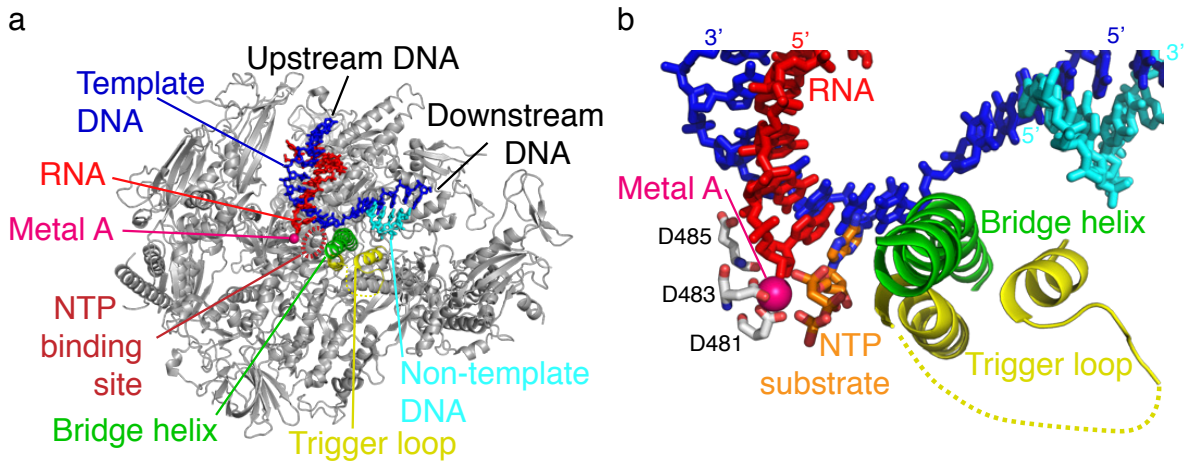

Figure 5: Architecture of the Pol II EC and Pol II active site. **a.** Overview of the *S. cerevisiae* RNA polymerase II elongation complex (PDB code 3HOV) (Sydow et al., 2009). Template strand, non-template strand, RNA, bridge helix, trigger loop, and metal A are colored in dark blue, cyan, red, green, yellow and pink, respectively. Color scheme is used throughout the figure. Missing amino acids for the trigger loop are indicated as yellow dotted line. NTP binding site is indicated as a red, dotted oval. **b.** Active site of the Pol II EC. Conserved aspartate residues that coordinate metal A are shown as sticks in white. A non-hydrolyzable NTP (AMPCPP) was used to obtain a structure with the NTP substrate bound (PDB code 4A3F) (Cheung et al., 2011).

Nucleotide addition cycle

During elongation, Pol II catalyses the addition of nucleotides to a growing RNA chain in a DNA-directed manner (Cheung and Cramer, 2012). Nucleotide addition can be thought of as a catalytic cycle (nucleotide addition cycle) (Hantsche and Cramer, 2016). In the cycle, a nucleotide enters the active site of Pol II at the nucleotide addition site. Pol II selects the correct NTP using hydrogen bonding between the template DNA and the NTP, catalyzes the addition of the selected NTP to the 3' end of the RNA, translocates along the template DNA to free the catalytic site, and allows for the next NTP to bind at the nucleotide addition site (Martinez-Rucobo and Cramer, 2013). The cycle is then repeated.

The NTP substrate binds at the nucleotide addition site between the 3' end of the RNA and the bridge helix. NTP selection is then performed in a two-step mechanism (Westover et al., 2004; Kettenberger et al., 2004; Sydow and Cramer, 2009; Cheung et al., 2011). Presence of the NTP in the nucleotide addition site causes a structural rearrangement of the mobile trigger loop. The rearrangement closes the active site and the NTP moves to a position (insertion site) where all contacts required for catalysis are satisfied. Catalysis is then realized by a two-metal ion mechanism (Vassylyev et al., 2007b). One of the two magnesium ions involved in catalysis is coordinated by three conserved aspartate residues and the RNA 3' end, while metal B appears to be more flexible. Upon catalysis, metal B is bound by the NTP substrate. Nucleotide addition then follows a S_N2 nucleophilic substitution (Yee et al., 1979; Armstrong et al., 1979). The 3'-OH of the RNA acts as the nucleophile that attacks the α-phosphate of the NTP. The nucleotide forms a covalent bond with the RNA and is now part of the growing RNA chain. The catalysis releases one pyrophosphate. It has been proposed that the pyrophosphate release causes unfolding of the trigger loop to allow for a new round of the nucleotide addition cycle after translocation (Brueckner and Cramer, 2008; Martinez-Rucobo and Cramer, 2013).

Mechanism of translocation

After one nucleotide addition cycle, the insertion site of the active site of RNA polymerase II is occupied (pre-translocated state) and translocation of the DNA and RNA is required to allow for the next round of substrate addition. Translocation is performed in one base steps and is based on a Brownian ratchet mechanism. The mechanism of translocation was elucidated through a number of structural studies (Brueckner and Cramer, 2008; Bushnell et al., 2002). The initial Pol II core structure bound to nucleic acid captured the pre-translocated state (Gnatt et al., 2001) and subsequent structural studies have captured the enzyme in the post-translocated state. Additionally, a structure of Pol II bound by the

mushroom toxin α-amanitin captured an intermediate state between the pre- and post-translocated state (Brueckner and Cramer, 2008). This structure led to the proposal of a two-step translocation mechanism where the release of the pyrophosphate causes unfolding of the trigger loop. This unfolding leads to a cooperative movement of the newly formed hybrid base pair out of the active site and bending of the bridge helix which was previously locked (ratchet-like), ultimately resulting in the translocation to the post-translocated state. Overall, translocation frees the nucleotide addition site and the next NTP can bind.

1.2.3 Transcription termination

Successful transcription elongation is followed by transcription termination to conclude the transcription cycle of RNA polymerase II (Porrua and Libri, 2015; Proudfoot, 2016). The 3' end of genes is marked by the poly(A) site. In contrast to Pol I and Pol III, RNA polymerase II does not terminate in a sequence-specific manner but rather termination sites are located in an approximately several thousand base pairs window past the poly(A) signal (Nojima et al., 2013). When Pol II transcribes over the poly(A) site, the RNA is cleaved and a poly(A) tail is added to the 3' end of the RNA. It remains unclear how transcription termination works mechanistically and structures are largely lacking. It has been postulated that termination may require Pol II clamp opening and conformational changes in the elongation factor Spt4/5 during termination (Richard and Manley, 2009; Schreieck et al., 2014). The pre-termination complex is postulated to be inherently labile, making its study difficult.

Based on biochemical and functional genomics data, two models for transcription termination are proposed. The torpedo model proposes that a 5'-3' exonuclease (Rat1 in yeast, XRN2 in human) is recruited by a CTD-interacting protein (Rtt103 in yeast) and degrades the nascent RNA after cleavage by the CPF-CF complex (Kim et al., 2004; West et al., 2004). The exonuclease then dissociates Pol II from the DNA-RNA hybrid by catching up with polymerase and driving Pol II off of its substrate in a torpedo-like fashion (Connelly and Manley, 1988). The allosteric model proposes that a change in the elongation factors that associate with Pol II or conformational changes in the polymerase itself destabilize the previously stable elongation complex, resulting in dissociation of RNA polymerase II from the DNA-RNA hybrid (Richard and Manley, 2009). These two models are not mutually exclusive and there is evidence that both models could play a role in a unified mechanism of transcription termination (Luo et al., 2006).

1.3 Aims and scope of this work

Chromatin remodelling is required for DNA replication, DNA repair and transcription. The single-subunit chromatin remodeller Chd1 centers nucleosomes on DNA and can induce a regular nucleosome pattern. It plays a key role in cell pluripotency (Gaspar-Maia et al., 2009) and is essential for RNA polymerase II passage through chromatin (Skene et al., 2014).

I determined the structure of full-length *S. cerevisiae* Chd1 bound to a nucleosome core particle at a nominal resolution of 4.8 Å. Chd1 binds the NCP on one side of the nucleosome and peels two full helical turns of nucleosomal DNA away from the histone octamer. The ATPase motor binds the NCP at SHL +2. Based on previously published structures, I was able to determine a mechanism of regulation for Chd1 where the regulatory double chromodomain contacts DNA at SHL +1 and releases ATPase lobe 2 to allow for lobe 2 closure. Comparison of our structure with the structure of a NCP-Snf2 complex (Liu et al., 2017) revealed a model for the mechanism of ATPase translocation where binding of ATP results in translocation of ATPase lobe 2 from a pre- to a post-translocated state that causes DNA translocation. ATP hydrolysis results in resetting of the ATPase motor and the enzymatic cycle can begin again. The observed conformation allowed assignment of the translocase activity of Chd1. Chd1 is a 3'-5' translocase, which translocates DNA towards the octamer dyad.

It has been shown that DNA alkylations present in minor groove alkylation adducts can stall transcription elongation. In order to understand how these alkylation adducts alter transcription, I solved the 3.2 Å resolution crystal structure of a *S. cerevisiae* RNA polymerase II elongation complex with an alkylated base present opposite the nucleotide addition site. Based on biochemical work and structural interpretation, it was possible to characterize a novel mechanism of Pol II stalling in the presence of this class of modified bases. Pol II likely stalls due to a clash between the DNA adduct and the mobile trigger loop which is required for proper RNA chain synthesis and translocation during the nucleotide addition cycle.

2 Materials & Methods

This chapter lists strains, buffers, oligonucleotides and other materials used in the presented work. It additionally provides protocols for all presented work that is part of this dissertation.

Parts of this section have been published or are in the process of being published:

Lucas Farnung, Seychelle M. Vos, Christoph Wigge, Patrick Cramer. (2017) Nucleosome-Chd1 structure and implications for chromatin remodelling. *Nature.* doi: 10.1038/nature24046 (in press)

Stefano Malvezzi*, **Lucas Farnung***, Claudia Aloisi, Todor Angelov, Patrick Cramer, Shana J. Sturla. (2017) Mechanism of RNA polymerase II stalling by DNA alkylation *PNAS.* (accepted in principle)
*Both authors contributed equally.

2.1 Materials

2.1.1 Strains

Bacterial strains

Table 1: *E. coli* strains used in this study

Strain	Genotype	Source
BL21-Codon Plus(DE3)-RIL	B F^- *ompT* *hsdS*(r_b^- m_b^-) *dcm*$^+$ Tet^r *gal* λ(DE3) *endA* Hte [*argU ileY LeuW* Cam^r]	Stratagene
XL1-Blue	*rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 *sup*E44 *rel*A1 *lac* [F' *pro*AB *lac*I^q ZΔM15 Tn*10* (Tet^r)]	Stratagene

DH10EMBacY	F⁻*mcrA*Δ(*mrr*-hsdRMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ⁻*rps*L *nup*G/EMBacY/pMON7124	Geneva Biotech

Insect cell strains

Table 2: Insect cell strains used in this study

Strain	Species	Source
Sf9	*Spodoptera frugiperda*	Life Technologies
Sf21 (IPLB-Sf-21-AE)	*Spodoptera frugiperda*	Expression Systems, LLC
Hi5 (*T. ni*)	*Trichoplusia ni*	Expression Systems, LLC

Yeast strains

Table 3: *Saccharomyces cerevisiae* strain used in this study

Strain	Genotype	Source
BJ5464	BJ5464 Rpb3 His-Bio tag introduced at the N-terminus of Rpb3	Kireeva et al., 2000

2.1.2 Plasmids

Table 4: Plasmids used in this study

Vector	Insert	Type	Source
438-A	Untagged	438 series	UC Berkeley
438-B	N-terminal His 6× tag and a TEV protease cleavage site	438 series	UC Berkeley
438-C	N-terminal His 6× tag, MBP tag, N10 linker and a TEV protease cleavage site	438 series	UC Berkeley
Spt16	438-C vector with *S. cerevisiae* Spt16 with N-terminal His 6× tag, MBP tag, N10 linker and a TEV protease cleavage site	438 series	This study
Pob3	438-A vector with *S. cerevisiae* Pob3	438 series	This study

FACT	*S. cerevisiae* Spt16 with N-terminal His 6× tag, MBP tag, N10 linker and a TEV protease cleavage site, and *S. cerevisiae* Pob3	438 series	This study
Chd1	*S. cerevisiae* Chd1 with N-terminal His 6× tag, MBP tag, N10 linker and a TEV protease cleavage site	438 series	This study
H2A	pET3A vector with *X. laevis* H2A	pET series	Halic Lab
H2B	pET3A vector with *X. laevis* H2B	pET series	Halic Lab
H3	pET3A vector with *X. laevis* H3	pET series	Halic Lab
H4	pET3A vector with *X. laevis* H4	pET series	Halic Lab

2.1.3 Oligonucleotides

Table 5: Oligonucleotides used for generation of nucleosomal DNA, crystallization, and transcription assays

Type	**Sequence (5'-3')**	**Application**
Template DNA	ACC TCA ACT ACT TG(3-deaza-3-NAPHT-dA) CCC (5-bromo-U)CC TCA TT	Crystallization
Non-template DNA	CAA GTA GTT GAG GT	Crystallization
RNA	UUC GAG GAG GG	Crystallization
Forward Primer	CGC TGT TTT CGA ATT TAC CCT TTA TGC GCC GGT ATT GAA CCA CGC TTA TGC CCA GCA TCG TTA ATC GAT GTA TAT ATC TGA CAC GTG CCT	Nucleosome reconstitution
Reverse Primer	ATC AGA ATC CCG GTG CCG AG	Nucleosome reconstitution

2.1.4 Buffers and solutions

Table 6: Buffers used for purification of *S. cerevisiae* Chd1

Name	Description
Lysis buffer	300 mM NaCl, 20 mM Na·HEPES pH 7.4, 10 % (v/v) glycerol, 1 mM DTT, 30 mM imidazole pH 8.0, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
High Salt Buffer	1000 mM NaCl, 20 mM Na·HEPES pH 7.4, 10 % (v/v) glycerol, 1 mM DTT, 30 mM imidazole pH 8.0, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
Nickel elution buffer	300 mM NaCl, 20 mM Na·HEPES pH 7.4, 10 % (v/v) glycerol, 1 mM DTT, 500 mM imidazole pH 8.0, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
Dialysis buffer	300 mM NaCl, 20 mM Na·HEPES pH 7.4, 10 % (v/v) glycerol, 1 mM DTT, 30 mM imidazole pH 8.0
Gel Filtration buffer	300 mM NaCl, 20 mM Na·HEPES pH 7.4, 10 % (v/v) glycerol, 1 mM DTT

Table 7: Buffers used for purification of *S. cerevisiae* FACT

Name	Description
Lysis buffer	300 mM NaCl, 20 mM Na·HEPES pH 7.4, 10 % (v/v) glycerol, 1 mM DTT, 30 mM imidazole pH 8.0, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
High Salt Buffer	1000 mM NaCl, 20 mM Na·HEPES pH 7.4, 10 % (v/v) glycerol, 1 mM DTT, 30 mM imidazole pH 8.0, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
Nickel elution buffer	300 mM NaCl, 20 mM Na·HEPES pH 7.4, 10 % (v/v) glycerol, 1 mM DTT, 500 mM imidazole pH 8.0, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
Dialysis buffer	150 mM NaCl, 20 mM Na·HEPES pH 7.4, 10 % (v/v) glycerol, 1 mM DTT, 30 mM imidazole pH 8.0
Gel Filtration buffer	300 mM NaCl, 20 mM Na·HEPES pH 7.4, 10 % (v/v) glycerol, 1 mM DTT

Table 8: Buffers used for purification of *S. cerevisiae* Paf1C

Name	Description
Lysis buffer	50 mM Tris-HAc (pH 8.7 at 4 °C), 600 mM KAc, 2 mM $MgCl_2$, 10 mM imidazole, 2 mM DTT, pH 8.0, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
Denatured protein wash buffer	2mg/mL denatured protein, 50 mM Tris-HAc (pH 8.7 at 4 °C), 600 mM KAc, 2 mM $MgCl_2$, 10 mM imidazole, 2 mM DTT, pH 8.0, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
Ni elution buffer	50 mM Tris-HAc (pH 8.7 at 4 °C), 150 mM KAc, 2 mM $MgCl_2$, 150 mM imidazole, 2 mM DTT, pH 8.0, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
HiTrap SP Buffer 0	50 mM Tris-HAc (pH 8.7 at 4 °C), 2 mM $MgCl_2$, 2 mM DTT
HiTrap SP Buffer 2000	50 mM Tris-HAc (pH 8.7 at 4 °C), 2000 mM KAc, 2 mM $MgCl_2$, 2 mM DTT
Gel filtration buffer	10 mM Tris-HCl (pH 8.5 at 4 °C), 200 mM NaCl, 2 mM $MgCl_2$, 10 % (v/v) glycerol, 2 mM DTT

Table 9: Buffers used for purification of *X. laevis* histones

Name	Description
Wash buffer	100 mM NaCl, 50 mM Tris-HCl pH 7.5, 2 mM DTT, 1 mM Na-EDTA pH 8.0, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
TW buffer	100 mM NaCl, 50 mM Tris-HCl pH 7.5, 2 mM DTT, 1 mM Na-EDTA pH 8.0, 1 % (v/v) Triton X-100 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
Unfolding buffer	7 M guanidine hydrochloride, 20 mM Tris-HCl pH 7.5, 10 mM DTT
SAU-200	7 M deionized urea (freshly deionized), 20 mM sodium acetate pH 5.2, 200 mM NaCl, 1 mM Na-EDTA pH 8.0, 2 mM DTT

Name	Description
SAU-1000	7 M deionized urea (freshly deionized), 20 mM sodium acetate pH 5.2, 1000 mM NaCl, 1 mM Na-EDTA pH 8.0, 2 mM DTT
Dialysis Buffer	15 mM Tris-HCl pH 7.5, 2 mM DTT

Table 10: Buffers used for purification of *S. cerevisiae* Rpb4/7

Name	Description
Freezing buffer	50 mM Tris pH 7.5 at 4 °C, 150 mM NaCl, 10 % (v/v) glycerol, 0.852 μg/ml leupeptin, 4.11 μg/ml pepstatin A, 0.51 mg/ml PMSF, 0.99 mg/ml benzamidine
Buffer0	50 mM Tris pH 7.5 at 4 °C, 150 mM NaCl, 10 mM β-mercaptoethanol, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
Buffer10	50 mM Tris pH 7.5 at 4 °C, 150 mM NaCl, 10 mM β-mercaptoethanol, 10 mM imidazole, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
Buffer20	50 mM Tris pH 7.5 at 4 °C, 150 mM NaCl, 10 mM β-mercaptoethanol, 20 mM imidazole, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
Buffer50	50 mM Tris pH 7.5 at 4 °C, 150 mM NaCl, 10 mM β-mercaptoethanol, 50 mM imidazole, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
Buffer200	50 mM Tris pH 7.5 at 4 °C, 150 mM NaCl, 10 mM β-mercaptoethanol, 200 mM imidazole, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
Salt buffer	50 mM Tris pH 7.5 at 4 °C, 2 M NaCl, 10 mM β-mercaptoethanol, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
SourceQ0	20 mM Tris pH 7.5 at 4 °C, 1 mM EDTA, 10 mM DTT
SourceQ2000	20 mM Tris pH 7.5 at 4 °C, 2 M NaCl 1 mM EDTA, 10 mM DTT
Gel Filtration Buffer	5 mM HEPES pH 7.25 at 20 °C, 40 mM $(NH_4)SO_4$, 2 M NaCl, 10 μM $ZnCl_2$, 10 mM DTT

Table 11: Buffers used for purification of *S. cerevisiae* RNA polymerase II (BJ5464 strain)

Name	Description
Freezing buffer	150 mM Tris pH 7.9 at 4 °C, 3 mM EDTA, 30 μM $ZnCl_2$, 30 % (v/v) glycerol, 3 % (v/v) DMSO, 30 mM DTT, 0.852 μg/ml leupeptin, 4.11 μg/ml pepstatin A, 0.51 mg/ml PMSF, 0.99 mg/ml benzamidine
HSB150	50 mM Tris pH 7.9 at 4 °C, 150 mM KCl, 1 mM EDTA, 30 μM $ZnCl_2$, 10 % (v/v) glycerol, 2.5 mM DTT, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
HSB1000/7	50 mM Tris pH 7.9 at 4 °C, 1000 mM KCl, 7 mM imidazole, 1 mM EDTA, 30 μM $ZnCl_2$, 10 % (v/v) glycerol, 2.5 mM DTT, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine
Ni buffers	20 mM Tris pH 7.9 at 4 °C, 150 mM KCl, 7/50/100 mM imidazole
MonoQ 150	20 mM Tris-acetate pH 7.9 at 4 °C, 150 mM KAc, 10 % (v/v) glycerol, 0.5 mM EDTA, 10 μM $ZnCl_2$, 10 mM DTT
MonoQ 2000	20 mM Tris-acetate pH 7.9 at 4 °C, 2000 mM KAc, 10 % (v/v) glycerol, 0.5 mM EDTA, 10 μM $ZnCl_2$, 10 mM DTT
Gel Filtration Buffer	10 mM HEPES pH 7.0 at 20 °C, 200 mM KCl, 5 % (v/v) glycerol, 2 mM DTT

2.1.5 Media and additional components

Table 12: Growth media for *E. coli*

Media	Components
LB	1 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 0.5 % (w/v) NaCl
2× YT	1.6 % (w/v) Bacto tryptone, 1 % (w/v) Bacto yeast extract, 0.5 % (w/v) NaCl

Table 13: Media for cell expression

Media	Company
Sf-900 SFM III	Life Technologies

ESF921	Expression Systems, LLC.

Table 14: Antibiotics and additives used for *E. coli*

Antibiotic/Additive	**Stock concentration**	**Final concentration**
Ampicillin	100 mg/mL	100 μg/mL
Chloramphenicol	34 mg/mL	34 μg/mL
Kanamycin	50 mg/mL	50 μg/mL
Gentamycin	10 mg/mL	50 μg/mL
X-Gal	150 mg/mL	150 μg/mL
IPTG	1 M	1 mM
X-TremeGENE9 (Sigma-Aldrich)		

2.2 Methods

2.2.1 Cloning and related techniques

Polymerase Chain Reaction (PCR)

Polymerase chain reactions were carried out with Phusion High Fidelity DNA Polymerase in 50 μ L reactions, unless stated otherwise. Primer annealing temperatures were calculated with SnapGene and designed to have an annealing temperature of 60 °C. The following reaction scheme was used:

Table 15: Reaction for PCR

Component	**Volume**
5× HF Buffer	10 μL
10× 10 mM dNTP	10 μL
10 μM Primer 1	2.5 μL
10 μM Primer 2	2.5 μL
Template DNA	150 ng
DMSO	3 μL
H_2O	Up to 50 μL

The 2-step PCR program was run with the following parameters:

Table 16: Two-step PCR program

T in °C	Time	Loop
98	1 min	
98	10 sec	
72	8 min	Back to step 2, Loop 33×
72	10 min	
4	Pause	

If applicable, PCR templates were digested with DpnI. PCR products were mixed with loading dye and separated by agarose electrophoresis. Relevant products were subsequently purified using the QIAquick Gel Extraction Kit (Qiagen).

Electrophoretic separation of DNA using agarose gels

Agarose gel electrophoresis was performed to separate DNA products. The gel contained 1 % (v/v) agarose, 1 × TAE buffer and 0.5 μg/mL SybrSafe (Invitrogen). Based on the sample volume, appropriate combs were chosen for gel electrophoresis. DNA samples were loaded on the gel with a size marker (1 kb DNA ladder, NEB). The gel was subsequently run in 1 × TAE buffer at 120 V until satisfactory separation of the DNA was achieved. Gel visualization was achieved using a gel imaging system (Intas Göttingen, Germany).

Restriction digests

Restriction digests with the appropriate restriction enzymes are performed in 20 μL or 50 μL reactions. Restriction enzymes from the manufacturer NEB (New England Biolabs) were used in this study. Restriction enzyme buffers were chosen based on recommendations by NEB. Restriction digests were performed for 3 hours at the appropriate temperature. Digested vectors were applied to an agarose gel and purified using the QIAquick Gel Extraction Kit (Qiagen).

Ligation independent cloning

Ligation independent cloning (LIC) is a cloning technique that does not require endonuclease activity for overhang generation or ligase activity. LIC is an attractive technique to clone genes because inherent restriction sites present in the gene sequence do not affect cloning.

To perform LIC, 438-series vectors were digested with the restriction enzyme SspI in the following reaction:

Table 17: Reaction for LIC restriction digest of a plasmid

Component	Volume
Vector (2 μ g)	variable
CutSmart Buffer	5 μL
SspI-HF	1 μL
100 mM DTT	1 μL
H_2O	Up to 50 μL

Inserts and linearized plasmids are treated with T4 DNA polymerase in a LIC reaction using the exonuclease activity of T4 DNA polymerase to generate sequence specific overhangs.

Table 18: LIC reaction for plasmid

Component	Volume
gel purified linearized vector (50-150 ng)	10 μL
25 mM dGTP stock	2 μL
T4 DNA pol 10× buffer	2 μL
100 mM DTT	1 μL
T4 DNA polymerase	0.4 μL
H_2O	0.4 μL

The reaction is incubated at 22 °C for 30 minutes and 20 minutes at 75 °C. Inserts (e.g. generated by PCR) are LIC-treated in the following reaction:

Table 19: Reaction for LIC reaction of insert

Component	Volume
gel purified linearized insert (50-150 ng)	10 μL
25 mM dCTP stock	2 μL
T4 DNA pol 10× buffer	2 μL
100 mM DTT	1 μL
T4 DNA polymerase	0.4 μL
H_2O	0.4 μL

The reaction is incubated at 22 °C for 30 minutes and 20 minutes at 75 °C.

After LIC treatment, 2 μL of LIC-treated plasmid and 2 μL of LIC-treated insert are combined and incubated for 10 min at RT to anneal the generated overhangs. 2.5 μL of the annealed mixture are then transformed into chemically competent cells.

To combine multiple genes into one vector, one plasmid containing one gene is linearized with SwaI generating the "vector" DNA, and the gene to be added, "insert" is generated with PmeI digest. The restriction digested fragments are then gel-purified, and undergo an additional round of LIC cloning.

Circular extension polymerase cloning

Circular extension polymerase cloning (CPEC) is a cloning technique to generate multi-part DNA assemblies (Quan and Tian, 2011). It is based on generating overhangs that prime each other during a PCR reaction.

To design a CPEC reaction, overlaps need to be generated with an annealing temperature of 70 °C using the NEB Tm Calculator.

The PCR reaction is then setup in the following way:

Table 20: CPEC reaction

Component	**Volume**
Q5 5× buffer	10 μ L
dNTPs	4 μ L
gel purified, linear plasmid	150 ng vector
gel purified PCR product/insert	200 ng vector
Q5 DNA polymerase	0.5 μ L
H_2O	Fill up to 50 μL

The following PCR program is used:

Table 21: PCR program for CPEC reaction

Temperature in °C	**Time**	**Loop**
98	30 sec	
98	10 sec	
70-55	0.1 s for each temperature change for 3 min	
55	30 sec	Back to step 1, Loop 26×

72	5 min
4	Pause

After the PCR 40 μL of the CPEC reaction are added to 100 μL of chemically competent XL1 blue cells. The competent cells are transformed as described in the next section.

Transformation of chemically competent cells

DNA of interest was added to 100 μL of chemically competent cells. The cells were incubated for 25 minutes on ice. Subsequently, the cells were heat-shocked at 42 °C for 45 seconds and incubated on ice for 2 minutes. 900 μL of LB medium without any selection marker were added to the cells. The cell suspension was incubated for 1-4 hours at 37 °C and 300 rpm.

Transformation of electrocompetent cells

500 ng of purified plasmid were added to 100 μL of electrocompetent cells. The cells were incubated for 10 minutes on ice. After incubation, the cells were transferred to a Gene Pulser/MicroPulser Cuvettes (0.1 cm gap) (Biorad) and pulsed using a Biorad MicroPulser with the following settings: 2.5 μF, 1.8 kV. After pulsing, 1 mL of LB medium was added to the pulsed cell resuspension and transferred to a 15 mL falcon. The cells were outgrown for six hours before being plated on the appropriate selection media.

Isolation of plasmid DNA

Transformed cells were plated on agarose plates with appropriate, selective antibiotics. Single colonies were picked from the plates and 5 mL or 50 mL LB medium with the appropriate antibiotics were inoculated. Cultures were grown overnight at 37 °C and 160 rpm. Plasmid DNA was isolated using the QIAprep Spin Miniprep Kit.

Verification of plasmid DNA using sequencing

Isolated plasmid DNA was sequenced for presence of the insert without any mutations using Seqlab (Göttingen).

Cloning of *S. cerevisiae* Chd1

A vector encoding full-length *S. cerevisiae* Chd1 was obtained through the MRC PPU Reagents and Services facility (MRC PPU, College of Life Sciences, University of Dundee, Scotland). The vector was used as a PCR template for cloning Chd1 into a modified pFastBac vector via ligation independent cloning (LIC) [a gift of Scott Gradia, UC Berkeley, vector 438-C (Addgene: 55220)]. The construct contains an N-terminal 6x His tag followed by a maltose binding protein (MBP) tag and a tobacco etch virus protease cleavage site.

Cloning of *S. cerevisiae* FACT

gBlocks encoding *Trichoplusia ni* codon-optimized Spt16 and Pob3 were designed using Integrated DNA Technologies (IDT) Codon Optimization Tool and synthesized by IDT. Two gBlocks encoding the N- and C-terminal part of Spt16 were cloned into vector 438-C using CPEC (Quan and Tian, 2011). The gBlock encoding Pob3 was cloned into vector 438-A (Addgene: 55218) using LIC. Combination of Spt16 and Pob3 into a single vector was achieved by using successive rounds of LIC. Each subunit is preceded by a PolH promoter and followed by a SV40 termination site. Spt16 has an N-terminal 6x His tag, followed by a maltose binding protein (MBP) tag, and a tobacco etch virus protease cleavage site.

2.2.2 Bacmid preparation

Purified plasmids (500 ng) were electroporated into DH10EMBacY (Geneva Biotech, Geneva, Switzerland) cells to generate bacmids containing full-length Chd1 or FACT constructs. Bacmids were prepared from positive clones using blue/white selection, and alkaline lysis followed isopropanol precipitation. V_0 and V_1 virus productions were performed as described (Vos et al., 2016).

2.2.3 Insect cell expression

600 mL of Hi5 cells grown in ESF-921 media (Expression Systems, Davis, CA, United States) were infected with 300 μL of V_1 virus for protein expression. The cells were grown for 48-72 hrs at 27 °C. Cells were harvested by centrifugation (238 xg, 4 °C, 30 min) and resuspended in lysis buffer (300 mM NaCl, 20 mM Na·HEPES pH 7.4, 10% glycerol (v/v), 1 mM DTT, 30 mM imidazole pH 8.0, 0.284 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine). The cell resuspension was snap frozen and stored at -80 °C.

2.2.4 Protein purifications and related techniques

All protein purifications were performed at 4 °C, unless stated otherwise. Purified proteins were aliquoted, snap-frozen in liquid nitrogen and stored at -80 °C unless stated otherwise.

Protein concentration determination

Absorption coefficients of proteins and protein complexes were determined by their primary amino acid sequence and the ExPasy ProtParam tool at 280 nm. Protein concentration were determined by measuring the absorption at 280 nm using a NanoDrop-200 spectrophotometer (Thermo Scientific). The calculated absorption coefficient was applied to give the final protein concentration.

SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyze protein samples. Before SDS-PAGE, protein samples were mixed with 5× LDS sample buffer (Invitrogen). Appropriate amounts of sample were loaded onto a 4-12 % NuPAGE Bis-Tris gradient gel with 10 or 15 wells (Invitrogen). A protein standard was used to indicate molecular weights (PageRuler Prestained Protein Ladder, Thermo Scientific). Gels were run in 1× MES SDS Running Buffer (Invitrogen) at 200 V for 30-40 minutes. Gels were subsequently stained with InstantBlue (expedeon) and imaged using a scanner (Epson Perfection V700 Photo).

X. laevis histone purification

X. laevis histones were expressed and purified as described previously (Luger et al., 1999; Dyer et al., 2004). Briefly, inclusion bodies were resuspended by using a manual Dounce tissue grinder (Sigma-Aldrich). Histones were aliquoted, flash-frozen, lyophilized, and stored at -80 °C prior to use. Lyophilized histones were resuspended in unfolding buffer (7 M guanidine hydrochloride, 20 mM Tris-HCl pH 7.5, 10 mM DTT) to a concentration of 1.5 mg/mL. H2A, H2B, H3, and H4 were then combined at a molar ratio of 1.2:1.2:1:1. The sample was incubated on ice for 30 minutes before it was dialyzed against 3 x 600 mL refolding buffer (2 M NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8, 5 mM β-mercaptoethanol) for a total of 18 hours at 4 °C. Dialyzed sample was recovered and applied to a GE S200 16/600 pg size exclusion column, pre-equilibrated in refolding buffer. Peak fractions containing histone octamer were pooled and concentrated to 30 μM.

S. cerevisiae Chd1 purification

Protein purifications were performed at 4 °C. Frozen cell pellets were thawed and lysed by sonication. Lysates were cleared by centrifugation (18,000 xg, 4 °C, 30 min) and ultracentrifugation (235,000 xg, 4 °C, 60 min). The supernatant containing Chd1 was filtered using 0.8 μm syringe filters (Millipore) and applied to a GE HisTrap HP 5 mL (GE Healthcare, Little Chalfont, United Kingdom), pre-equilibrated in lysis buffer. After sample application, the column was washed with 10 CV lysis buffer, 5 CV high salt buffer (1 M NaCl, 20 mM Na·HEPES pH 7.4, 10% glycerol (v/v), 1 mM DTT, 30 mM imidazole pH 8.0, 0.284 μg/mL leupeptin, 1.37 μg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL benzamidine), and 5 CV lysis buffer. The protein was eluted with a gradient of 0-100% elution buffer (300 mM NaCl, 20 mM Na·HEPES pH 7.4, 10% glycerol (v/v), 1 mM DTT, 500 mM imidazole pH 8.0, 0.284 μg/mL leupeptin, 1.37 μg/ml pepstatin A, 0.17 mg/ml PMSF, 0.33 mg/ml benzamidine). Peak fractions were pooled and dialyzed for 16 hours against 600 mL dialysis buffer (300 mM NaCl, 20 mM Na·HEPES pH 7.4, 10% glycerol (v/v), 1 mM DTT, 30 mM imidazole) in the presence of 2 mg His6-TEV protease. The dialyzed sample was applied to a GE HisTrap HP 5 mL. The flow-through containing Chd1 was concentrated using an Amicon Millipore 15 ml 50,000 MWCO centrifugal concentrator and applied to a GE S200 16/600 pg size exclusion column, pre-equilibrated in gel filtration buffer (300 mM NaCl, 20 mM Na·HEPES pH 7.4, 10 % (v/v) glycerol, 1 mM DTT). Peak fractions were concentrated to 100 μM, aliquoted, snap frozen, and stored at -80 °C. Typical yields of S. cerevisiae Chd1 from 1.2 L of insect cell culture are 7-10 mg.

S. cerevisiae FACT purification

FACT was purified as described for Chd1, with minor modifications. After dialysis, the sample was applied to tandem GE HisTrap HP 5 mL, GE HiTrap Q 5 mL columns. After washing with 5 CV of dialysis buffer, the HisTrap was removed. FACT was eluted from the HiTrap Q 5 mL by applying a gradient of 0-100% high salt buffer (1 M NaCl, 20 mM Na·HEPES pH 7.4, 10 % (v/v) glycerol, 1 mM DTT, 30 mM imidazole pH 8.0). Peak fractions were pooled and applied to a GE S200 16/600 pg size exclusion column. Pure fractions containing full-length FACT were concentrated as described above to a concentration of 60 μM, aliquoted, flash frozen, and stored at -80 °C. Typical preparations yield 10-15 mg of full-length S. cerevisiae FACT (Spt16 + Pob3) from 1.2 L of insect cell culture.

S. cerevisiae Paf1C purification

S. cerevisiae Paf1C (ΔCtr9-913) was expressed as described previously (Xu et al., 2017). Combined cell pellets were resuspended in lysis buffer. The cell resuspension was lysed by sonication and centrifuged (27,000 ×g, 30 min, 4 °C). The supernatant was filtered using a 0.8 μ syringe filter and applied to a GE HisTrap HP 5 mL (GE Healthcare), pre-equilibrated in lysis buffer. The column was washed with lysis buffer for 5 CV, denatured protein wash buffer for 5 CV, and lysis buffer for 5 CV. The protein of interest was then eluted using Ni Elution buffer. The elution was collected in a 50 mL falcon. The conductivity of the sample was then adjusted to the same conductivity as a HiTrap SP Buffer 0 mixture with 3.5 % (v/v) of HiTrap SP Buffer 2000 using HiTrap SP Buffer 0. The diluted sample was then applied to a HiTrap SP HP 5 mL (GE Healthcare) and washed for 40 CV with a HiTrap SP Buffer 0 mixture with 3.5 % (v/v) of HiTrap SP Buffer 2000. Elution was achieved by applying a gradient from 3.5 % (v/v) HiTrap SP Buffer 2000 to 40 % (v/v) HiTrap SP Buffer 2000. Fractions were collected and the presence of Paf1C complex was monitored by SDS-PAGE. Fractions containing Paf1C were pooled and concentrated (Amicon Millipore 15 ml 50,000 MWCO centrifugal concentrator). The concentrated sample was applied to a Superose 6 10/300 (GE Healthcare), pre-equilibrated in gel filtration buffer. Peak fractions were analyzed using SDS-PAGE. Fractions with stoichiometric amounts of all subunits of Paf1C were subsequently concentrated. The concentrated sample containing the whole Paf1C was aliquoted, snap-frozen and stored at -80 °C.

S. cerevisiae Rpb4/7 purification

S. cerevisiae Rpb4/7 was expressed in BL21-Codon Plus(DE3)-RIL cells. Cells were harvested and resuspended in 50 mL of Rpb4/7 freezing buffer per 2 L of culture. The cells were lysed by sonication. The lysate was centrifuged (30 min, 27,000 ×g, and 4 °C). The centrifugation was repeated once and the supernatant was pooled. The pooled sample was applied to 6 mL Ni-NTA resin, pre-equilibrated in Buffer 0. The Ni-NTA resin was washed with salt buffer (3 CV), Buffer10 (3 CV), Buffer20 (3 CV), and Buffer 50 (3 CV). The sample was subsequently eluted using Buffer200 over 6 column volumes. The conductivity of the elution sample was adjusted to the conductivity of SourceQ0. The sample was filtered using a 0.22 μm sterile filter unit. The filtered protein sample was applied to a MonoQ 10/100 column (GE Healthcare), pre-equilibrated in a mixture of 95 % SourceQ0 and 5 % SourceQ2000. The column was washed for 2 CV with a mixture of 95 % SourceQ0 and 5 % SourceQ2000. Elution was performed via a linear gradient over 10 column volumes from 95 % SourceQ0 and 5 % SourceQ2000 to 100 % SourceQ2000.

The peak elution was monitored using SDS-PAGE. The fractions containin Rpb4/7 were concentrated using an Amicon Millipore 15 ml 10,000 MWCO centrifugal concentrator and applied to a pre-equilibrated GE Superdex 75 10/300. Fractions were collected and analyzed by SDS-PAGE. Peak fractions with Rpb4/7 were concentrated with an Amicon Millipore 15 ml 10,000 MWCO centrifugal concentrator. Protein concentration was estimated as described. The protein was subsequently aliquoted, flash-frozen and stored at -80 °C.

S. cerevisiae RNA polymerase II purification

S. cerevisiae cells of the BJ5464 strain were lysed using bead beaters (Hamilton Beach Brands, Inc.). 200 mL of cell suspension were filled into a bead beater metal chamber. An equal amount of glass beads with a diameter of 0.5 mm were added to the cell suspension. The cell lysis was carried out at 4-8 °C for a total of 90 minutes. The bead beater were cycled in the following pattern: 30 s ON, 90 s OFF for 90 minutes. The metal chambers of the bead beater were constantly cooled using an ice-salt mixture. The resulting cell lysate was subsequently centrifuged for 30 min at 13,689 ×g and 4 °C. Centrifugation was repeated once. The supernatant was ultracentrifuged at 76,221 ×g and 4 °C for 90 min in a Ti 45 rotor (Beckman-Coulter). The supernatant was pooled and an ammonium sulphate precipitation was performed by addition of 50 % (w/v) ammonium sulphate under constant stirring. After an incubation time of 12 hours, the protein sample was centrifuged at 34,200 ×g and 277 K for 45 minutes. The centrifugation was repeated once, and the supernatant was discarded. The resulting ammonium sulphate pellet was resuspended in 140 mL HSB 0/7 per 100 g of ammonium sulphate pellet. The conductivity was measured and adjusted to the conductivity of the HSB1000/7 buffer. The conductivity adjusted sample was applied to Ni-NTA resin (Qiagen), pre-equilibrated in HSB1000/7 and 3 CV of Ni Buffer7. Rpb3 His-tagged Pol II was eluted with Ni buffer 200 and elution fractions were collected. The conductivity of the elution sample was adjusted to the conductivity of Mono Q 150. The sample was subsequently applied to a pre-equilibrated MonoQ 10/100 column (GE Healthcare). Pol II was eluted using a gradient over 12 column volumes from a salt concentration of 150 mM KOAC to 1500 mM KOAc at a flow rate of ∼2 mL/min. Fractions containing Pol II were collected in 1 mL fractions in 96-well blocks. Presence of Pol II was monitored by SDS-PAGE. Recombinantly expressed and purified Rpb4/7 was added at a two-fold molar excess and incubated for 40 min to form a homogeneous 12-subunit RNA polymerase II complex. The sample was applied to a Superose 6 10/300 size exclusion column (GE Healthcare), which was pre-equilibrated in gel filtration buffer.

Fractions were collected in 96-well blocks with a 0.5 mL fraction size. Presence of Pol II was again monitored by SDS-PAGE. The peak elution containing Pol II was pooled and concentrated to a final concentration of 4 mg/mL. The sample was then aliquoted, snap-frozen in liquid nitrogen and stored as described.

Mass spectrometry

Proteins were subjected to SDS-PAGE. Individual protein bands were selected for mass spectrometric determination and processed by the mass spectrometric core facility at the MPI for Biophysical Chemistry (Göttingen, Germany). Results were analyzed using the software Scaffold 4.

2.2.5 Preparation of DNA for nucleosome reconstitutions

DNA fragments for nucleosome reconstitution were generated by PCR, essentially as described previously (Maskell et al., 2015). A vector containing the Widom 601 sequence was used as a template for PCR. In-house expressed and purified Phusion polymerase was used for the PCR reaction with two primers (Forward: CGC TGT TTT CGA ATT TAC CCT TTA TGC GCC GGT ATT GAA CCA CGC TTA TGC CCA GCA TCG TTA ATC GAT GTA TAT ATC TGA CAC GTG CCT, Reverse: ATC AGA ATC CCG GTG CCG AG). The PCR program had the following steps: 1. 98° for 1 min, 2. 98 °C for 10 sec, 3. 72 °C for 45 sec, cycle between step 2 and 3 for 35 times, 4. 72 °C for 10 min, 5. Pause at 5 °C. PCR products were pooled from three 48-well PCR plates (100 μL per well). The products were ethanol precipitated and resuspended in 1 mL TE buffer (10 mM Tris pH 8.0, 1 mM EDTA pH 8.0). The resuspended DNA was applied to a ResourceQ 6 mL (GE Healthcare) and eluted with a gradient from 0-100 % TE high salt buffer (10 mM Tris pH 8.0, 1 M NaCl, 1 mM EDTA pH 8.0). Peak fractions were analyzed on a 1 % (v/v) TAE agarose gel and fractions containing the desired DNA product were pooled. The sample was ethanol precipitated, resuspended in 200 μL TE buffer, and stored at -20° prior to use.

2.2.6 Preparation of reconstituted nucleosomes

Nucleosome reconstitution was performed as described (Dyer et al., 2004), with minor modifications. Histone octamer and DNA were mixed at a 1:1 molar ratio in 2 M KCl, and transferred to Slide-A-Lyzer MINI Dialysis Units 20,000 MWCO (Thermo Scientific, Waltham, MA, United States). The sample was gradient dialyzed against low salt buffer (30 mM KCl, 20 mM Na·HEPES pH 7.5, 1 mM EDTA pH 8, 1 mM DTT) over 18 hours.

The sample was dialyzed for another four hours against low salt buffer, recovered, and stored at 4 °C. Quantification of the reconstituted nucleosome was achieved by measuring absorbance at 280 nm. Molar extinction coefficients were determined for protein and nucleic acid components and were summed to yield a molar extinction coefficient for the reconstituted NCP.

2.2.7 Preparation of nucleosomal complexes

To prepare a nucleosome-Chd1-FACT-Paf1C complex, FACT, Chd1, and Paf1C were mixed at a molar ratio of 1:1.2:1.4 and incubated for 10 minutes. Zero monovalent salt buffer (2 mM $MgCl_2$, 20 mM Na·HEPES pH 7.5, 5 % (v/v) glycerol , 1 mM DTT) was added within 6 minutes to achieve a final monovalent salt concentration of 30 mM. Reconsituted NCP was added at a 0.5 molar ratio of the FACT concentration. The sample was incubated for 10 minutes, centrifuged (21,000 ×g, 4 °C, 10 min), and applied to a Superose 6 Increase 3.2/300 column equilibrated in gel filtration buffer (30 mM NaCl, 2 mM $MgCl_2$, 20 mM Na·HEPES pH 7.5, 5 % (v/v) glycerol, 1 mM DTT). Peak fractions were pooled, ADPÂůBeF3 was added to a concentration of 1 mM ADP and 3 mM BeF3-, and incubated for 10 minutes. The sample was cross-linked with 0.1 % (v/v) glutaraldehyde, incubated for 10 minutes on ice. The cross-linking reaction was quenched for 10 min using a concentration of 90 mM Tris-HCl (pH 7.9), 9 mM lysine and 9 mM aspartate. The sample was transferred to a Slide-A-Lyzer MINI Dialysis Unit 20,000 MWCO (Thermo Scientific), and dialyzed for 6 hours against 600 mL dialysis buffer (30 mM NaCl, 2 mM $MgCl_2$, 20 mM Na·HEPES pH 7.4, 1 mM DTT).

2.2.8 Cryo-EM and image processing

Cryo-EM

The nucleosome-Chd1-FACT-Paf1C complex sample was applied to R2/2 gold grids (Quantifoil). The grids were glow-discharged for 45 seconds before sample application of 2 μL on each side of the grid. The sample was subsequently blotted for 8.5 seconds and vitrified by plunging into liquid ethane with a Vitrobot Mark IV (FEI Company, Hillsboro, OR, United States) operated at 4 °C and 100 % humidity. Cryo-EM data was acquired on a FEI Titan Krios transmission electron microscope (TEM) operated at 300 keV, equipped with a K2 summit direct detector (Gatan, Pleasanton, CA, United States). Automated data acquisition was carried out using FEI EPU software at a nominal magnification of 105,000x. Image stacks of 40 frames were collected in counting mode over 10s. The dose

rate was 3 e^- per Ångström^2 per second for a total dose of 30 e^- Å^{-2}. A total of 3806 image stacks were collected.

Image processing

Frames were stacked and subsequently processed with MotionCor2 (Zheng et al., 2017). CTF correction was performed with Gctf (Zhang, 2016). Image processing was performed with RELION 2.0.4 (Scheres, 2012; Kimanius et al., 2016), unless noted otherwise. Post-processing of refined models was performed with automatic B-factor determination in RELION. Particles were picked using projections of an initial reconstruction (∼400,000 particles, FEI Falcon 2, not shown), yielding 990,020 particle images. Particles were extracted with a box size of 224^2 pixel, normalized, and screened using iterative rounds of reference-free 2D classification, yielding a total of 773,326 particles. Particle images were sub-divided into three batches and processed individually. Using a 40 Å low-pass filtered model from an initial reconstruction (not shown), I performed iterative rounds of hierarchical 3D classification with image alignment as outlined in (Figure 8. The three particle image batches were subsequently merged, re-extracted with a box size of 240^2 pixel and subjected to another round of 3D classification with image alignment. The best two classes were combined and subjected to a 3D refinement with a mask that encompasses the entire NCP-Chd1 complex. The NCP-Chd1 reconstruction was obtained from 67,032 particles with a resolution of 4.8 Å(gold-standard Fourier shell correlation 0.143 criterion). The map was sharpened with a B-factor of -204 Å^2. Local resolution estimates were determined using a sliding window of 40^3 voxels as previously described (Plaschka et al., 2016). Resolutions for individual Chd1 domains were determined by masking the respective regions and perfoming B-factor sharpening (gold-standard Fourier shell correlation 0.143 criterion) using RELION.

Model building

Crystal structures of the *X. laevis* nucleosome with Widom 601 sequence (Vasudevan et al., 2010) (PDB code 3LZ0), the S. cerevisiae Chd1 DNA-binding domains13 (PDB code 3TED), and *S. cerevisiae* Chd1 core (Hauk et al., 2010) (double chromodomain and ATPase motor, PDB code 3MWY) were placed into the electron density using UCSF Chimera (Pettersen et al., 2004). The individual Chd1 domains (SANT, SLIDE, double chromodomain, ATPase lobe 1, ATPase lobe 2) were fitted as rigid bodies. Residues 842-922 were removed from the double chromodomain-ATPase motor structure (PDB code 3MWY) due to weak density. We did not observe assignable density for the CHCT do-

main of Chd1. We did not assign weak density near H3 (residues 46-56), and H2A (residues 56-71). Extranucleosomal DNA, nucleosomal DNA from SHL -7 to SHL -5, and the H4 N-terminal tail residues 16-20 were built using COOT (Emsley et al., 2010). Three rounds of flexible fitting were performed with vmd (Humphrey et al., 1996) and MDFF (Trabuco et al., 2008), resulting in good fits of the electron density. Secondary structure restraints were applied and the model was real-space refined against the post-processed EM map using PHENIX48. ADP·BeF_3^- was built by superpositioning ATP-γ-S from the inactive Chd1 structure (PDB code 3MWY) onto our model, and replacing the ATP analogue with ADP·BeF_3^- (PDB code 3ICE) (Thomsen and Berger, 2009). BeF_3^- was modeled in a tetrahedral conformation for simplicity but is likely planar when it mimics part of the pentavalent transition state of ATP hydrolysis. While ADP is remarkably well-resolved at the given resolution, BeF_3^- has weaker density and was modelled based on previous structural data from other ATPases. R804 and R807 were fitted manually. The complete structure was geometry-optimized with PHENIX. Figures were generated using PyMol (Schrodinger„ LLC, 2015) and UCSF Chimera (Pettersen et al., 2004). Electron density was shown for the local resolution filtered map, if not stated otherwise.

2.2.9 Crystallography and data analysis of crystallographic data

Crystallography of Pol II EC

12-subunit Pol II (4.1 mg/mL) was incubated with a 1.5-fold molar excess of nucleic acid scaffold containing the 3d-Napht-A analog and 5-Br-U. The sample was incubated for 20 min on ice before crystallization by hanging drop vapor diffusion using 4-7 % PEG 6000, 200 mM ammonium acetate, 300 mM sodium acetate, 50 mM HEPES pH 7.0 and 5 mM TCEP as reservoir solution. Crystals were grown for 4-8 days, cryo-protected in mother solution supplemented with 22 % glycerol and nucleic acid scaffold. Crystals were incubated overnight at 8 °C before they were harvested and flash-cooled in liquid nitrogen.

Data collection of crystallographic data

Diffraction data were collected at beamline X06SA of the SLS (Swiss Light Source, Villigen, Switzerland) or beamline P13 at the Deutsches Elektronensynchrotron (DESY, Hamburg, Germany). The native data set was collected at a wavelength of 0.976 Å. The bromine peak data set was collected at a wavelength of 0.919 Å.

Data processing and model building

Diffraction images were processed with XDS. The structure was solved with molecular replacement using a 12-subunit Pol II structure (PDB 3HOX) without nucleic acids. Refinement was performed using Phenix.Refine (Adams et al., 2010). Refinement statistics are summarized in Supplementary Table 1. The final model was analyzed using MolProbity 45. 92 % of residues were in Ramachandran plot favored regions, 7.1 % in allowed regions. 0.9 % Ramachandran outliers were observed. Figures were generated using PyMOL (Schrodinger„ LLC, 2015).

3 Results & Discussion

3.1 Nucleosome-Chd1 structure and implications for chromatin remodelling

Results presented in this section are currently in the process of being published *(in press)*:

Lucas Farnung, Seychelle M. Vos, Christoph Wigge, Patrick Cramer. (2017) Nucleosome-Chd1 structure and implications for chromatin remodelling. *Nature.* doi: 10.1038/nature24046

The results presented in this section are based on the cited publication. Authors contributions are stated in the Publications chapter.

3.1.1 Abstract

Chromatin remodelling factors change nucleosome positioning and facilitate DNA transcription, replication, and repair (Narlikar et al., 2013). The conserved remodelling factor Chd (Hauk and Bowman, 2011) can shift nucleosomes and induce a regular nucleosome spacing (Lieleg et al., 2015; Hughes and Rando, 2015; Lusser et al., 2005). Chd1 is required for RNA polymerase II passage through nucleosomes (Skene et al., 2014) and for cellular pluripotency (Gaspar-Maia et al., 2009).

Chd1 contains the DNA-binding domains SANT and SLIDE, a bilobal motor domain that hydrolyses adenosine triphosphate (ATP), and a regulatory double chromodomain. Here we report the cryo-electron microscopy (cryo-EM) structure of Chd1 from the yeast S. cerevisiae bound to a nucleosome at a resolution of 4.8 Å. Chd1 detaches two turns of DNA from the histone octamer and binds between the two DNA gyres in a state poised for catalysis. The SANT and SLIDE domains contact detached DNA around superhelical location (SHL) -7 of the first DNA gyre. The ATPase motor binds the second DNA gyre at SHL +2 and is anchored to the N-terminal tail of histone H4 as in a recent nucleosome-Snf2 ATPase structure (Liu et al., 2017).

Comparison with published results reveals that the double chromodomain swings towards nucleosomal DNA at SHL +1, resulting in ATPase closure (Nodelman et al., 2017).

The ATPase can then promote translocation of DNA towards the nucleosome dyad, thereby loosening the first DNA gyre and remodelling the nucleosome. Translocation may involve ratcheting of the two lobes of the ATPase, which is trapped in a pre- or post-translocated state in the absence (Liu et al., 2017) or presence, respectively, of transition state-mimicking compounds.

3.1.2 NCP-Chd1-FACT-Paf1C complex

To investigate how RNA polymerase II transcribes through chromatin, we prepared factors that facilitate chromatin transcription in the yeast *S. cerevisiae* (Methods). These included the chromatin-remodelling enzyme Chd1 (chromodomain-helicase-DNA binding protein 1) (Figure 6), the histone chaperone FACT (facilitates chromatin transcription) and the transcription elongation factor Paf1C (polymerase-associated factor 1 complex). We formed a complex of these factors in the presence of the transition state-mimicking adduct ADP·BeF_3 and a nucleosome with DNA comprising the Widom 601 sequence (Lowary and Widom, 1998) and 63 base pairs (bp) of extranucleosomal DNA (Methods).

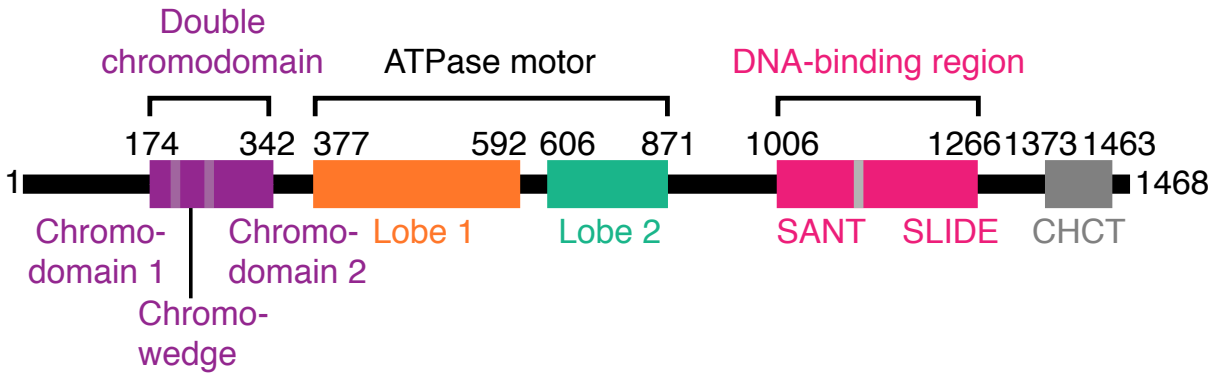

Figure 6: Chd1 domain architecture. Residues at domain boundaries are indicated.

3.1.3 NCP-Chd1 structure at 4.8 Å

Cryo-EM analysis revealed nucleosome-Chd1 particles that had lost FACT and Paf1C (Figure 8, A1). The resulting reconstruction of the nucleosome-Chd1 complex at an overall resolution of 4.8 Å revealed protein secondary structure (Figure A2, Table A1). Crystal structures of the nucleosome (Vasudevan et al., 2010; Lowary and Widom, 1998) and Chd1 domains (Hauk et al., 2010; Sharma et al., 2011) were unambiguously placed into the density. Only a minor, unassigned density remained that was located near histones H3

(residues 46-56) and H2A (residues 56-71) and may arise from a C-terminal domain in Chd1 (Mohanty et al., 2016). A detailed structure was obtained after flexible fitting and real-space refinement.

The structure reveals an altered nucleosome with one engaged Chd1 molecule (Figure 7). Two turns of nucleosomal DNA at SHL -5 to -7 are detached from the histone octamer. This alters the trajectory of extranucleosomal DNA by $\sim 60°$and breaks DNA interactions with histones H2A, H2B, and H3 (Figure 9).

The ability of Chd1 to detach DNA depends on the presence of an ATP analogue or ADP· BeF_3, indicating that our structure trapped Chd1 in a state poised for activity. The histone octamer is unaltered compared to the free nucleosome, whereas it adopts an altered conformation in a nucleosome-ACF remodelling complex with ADP· BeF_3 (Sinha et al., 2017) (Figure A3f).

Chd1 binds between extranucleosomal DNA and the second DNA gyre at SHL +2 (Figure 11), 10), consistent with lower-resolution information (Sundaramoorthy et al., 2017). Chd1 domains assemble between the two DNA gyres and form multiple DNA interactions. The SANT and SLIDE domains contribute to Chd1 affinity for the nucleosome (McKnight et al., 2011) and contact the first turn of extranucleosomal DNA in a way that was observed for free DNA (Sharma et al., 2011). The ATPase engages with DNA at SHL +2, consistent with the structure of the related Snf2 ATPase bound to the nucleosome (Liu et al., 2017) and with biochemical data (Nodelman et al., 2017; McKnight et al., 2011). The double chromodomain contacts DNA at SHL +1 (Figure 12) and binds between the SANT domain and ATPase lobe 1. The structure is incompatible with binding of linker histone H1 (Bednar et al., 2017), explaining why H1 can repress Chd1-dependent remodelling (Lusser et al., 2005).

The ATPase motor adopts a closed conformation with the ADP·BeF_3 adduct bound between lobes 1 and 2 (Figure 15). Compared to the free Chd1 structure (Hauk et al., 2010), lobe 2 rotates by $\sim 40°$ towards lobe 1. This rotation closes the active site and positions the catalytic arginine 'fingers' in lobe 2 (R804 and R807) close to the ATP-binding site (Figure 13) (Gu and Rice, 2010; Sengoku et al., 2006). One of these arginine fingers is mutated in human CHD1 in prostate cancers (Huang et al., 2012). These observations indicate that the structure trapped Chd1 in a functional state poised for catalysis.

The ATPase motor interacts extensively with DNA (Figure 10). Based on biochemical and structural observations (Gu and Rice, 2010; Singleton et al., 2007), we define the 'tracking strand' as the DNA strand running in the 5' to 3' direction from SHL +2 towards the histone octamer dyad. Lobe 1 contacts the tracking strand backbone with three protein regions containing ATPase motifs Ia and Ic, and with a loop (residues 457-461) located

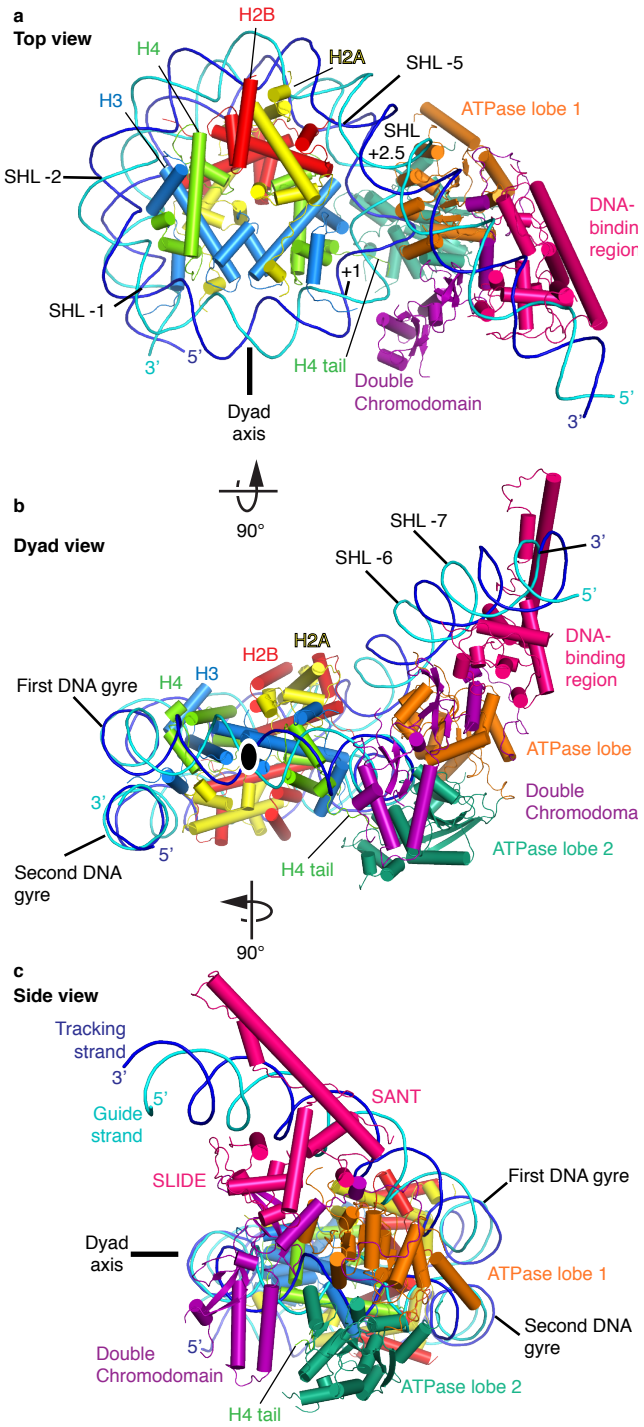

Figure 7: Three views of the NCP-Chd1 structure. a-c. Chd1 domains are colored as in Fig X. H2A, H2B, H3, H4, tracking strand, and guide strand are in yellow, red, light blue, green, dark blue, and cyan, respectively. The histone octamer dyad axis is indicated as black line or black oval circle. SHL, superhelical location.

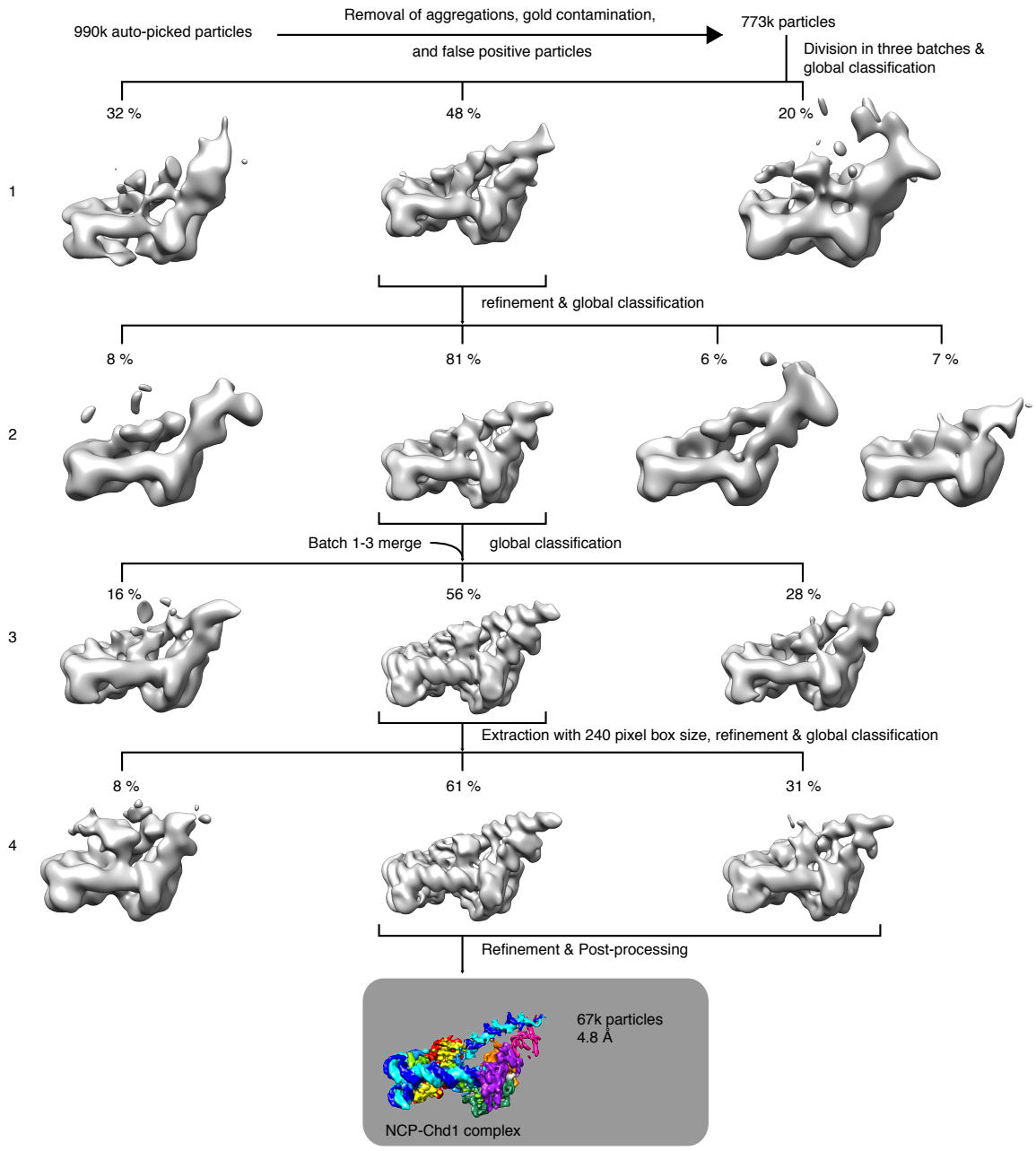

Figure 8: Sorting and classification tree used to reconstruct the nucleosome-Chd1 particle at 4.8 Å resolution. Steps 1 and 2 of batch 1 global classification are shown representatively for all three batches.

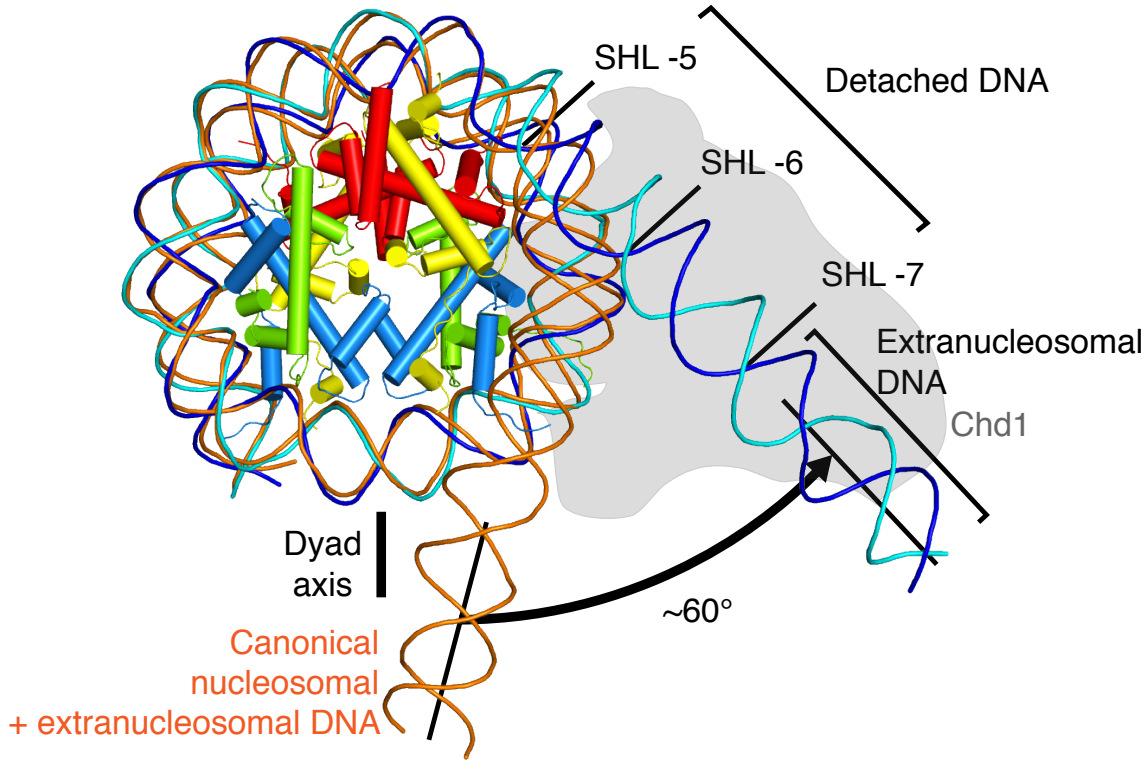

Figure 9: Detachment of nucleosomal DNA from the histone octamer at SHL -7 to -5. Extranucleosomal DNA rotates by 60° with respect to its location in the absence of Chd1 (orange, modelled by extending nucleosomal DNA with B-DNA). The position of Chd1 is indicated in grey color.

between motifs Ia and Ib. The lobe 1 regions formed by motifs IIa and III contact the complementary 'guide' DNA strand. Lobe 2 interacts with the tracking strand via loops formed by motifs IV, IVa, and V. Residue Trp793 in motif Va inserts into the minor groove and contacts the guide strand backbone (Figure 14). These ATPase-DNA interactions resemble the 'primary' interactions in a nucleosome-Snf2 complex (Liu et al., 2017) and interactions observed in a distantly related ATPase-DNA complex (Dürr et al., 2005). The interactions support the model that Chd1 translocates along the DNA minor groove from SHL +2 away from the octamer dyad, thereby moving DNA towards the dyad (Narlikar et al., 2013; Nodelman et al., 2017; Saha et al., 2005).

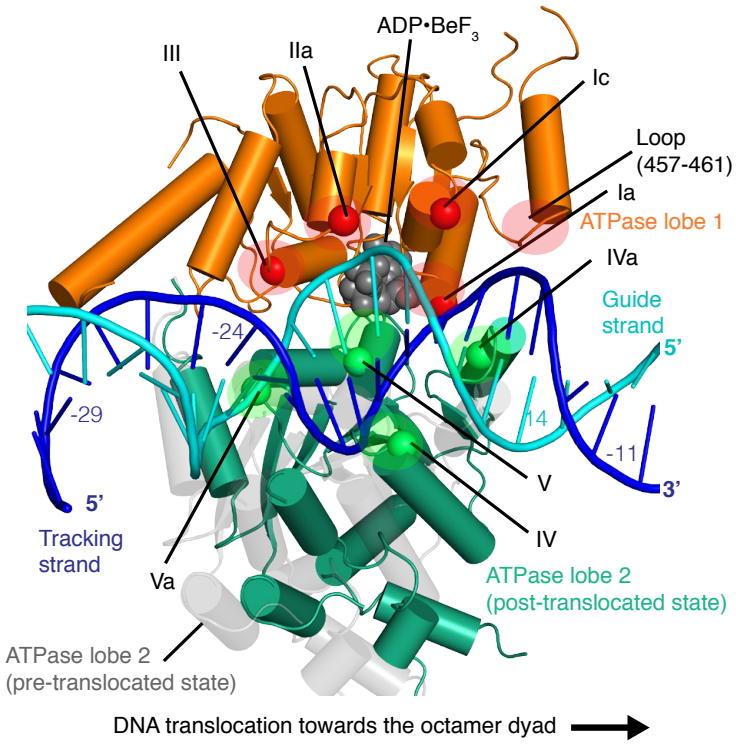

Figure 10: Primary ATPase-DNA interactions. Location of ATPase motifs on lobe 1 and lobe 2 are highlighted in red and green, respectively. The view is from the center of the histone octamer onto nucleosomal DNA. DNA register is indicated by numbering next to DNA bases. Color code is as in Figure 7. ADP·BeF_3 is shown as grey spheres. The model of lobe 2 in the pre-translocated position (grey) was derived from superposition of the nucleosome-Snf2 structure (PDB code 5X0Y) (Liu et al., 2017).

3.1.4 Regulation of the Chd1 ATPase motor activity by the double chromodomain

The structure also reveals the basis for ATPase activation by nucleosome binding. In the absence of the nucleosome, ATPase lobe 2 is sequestered in an open conformation by the 'chromo-wedge', an acidic region in the double chromodomain (Hauk et al., 2010). In the presence of the nucleosome, the double chromodomain swings by 15°and binds nucleosomal DNA. The chromo-wedge contacts the DNA backbone at SHL +1 (Figure 12) using a region that contains cancer mutations in the human homologue CHD4. Thus, binding of Chd1 to nucleosomal DNA induces swinging of the double chromodomain that releases lobe 2 and allows for ATPase closure and activation. Chd1 recognizes bent nucleosomal DNA because

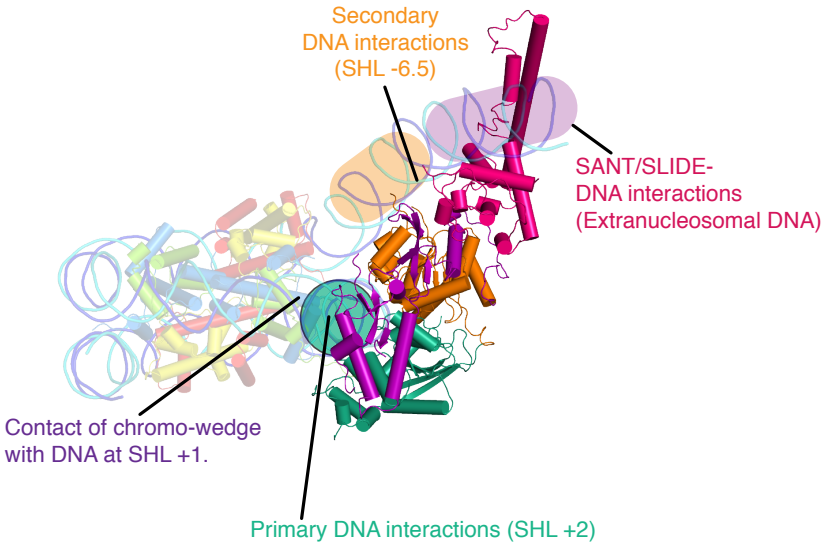

Figure 11: Overview of Chd1-DNA interactions. Primary and secondary interactions of Chd1 with the DNA are indicated. Interaction of the double chromodomain with DNA is shown.

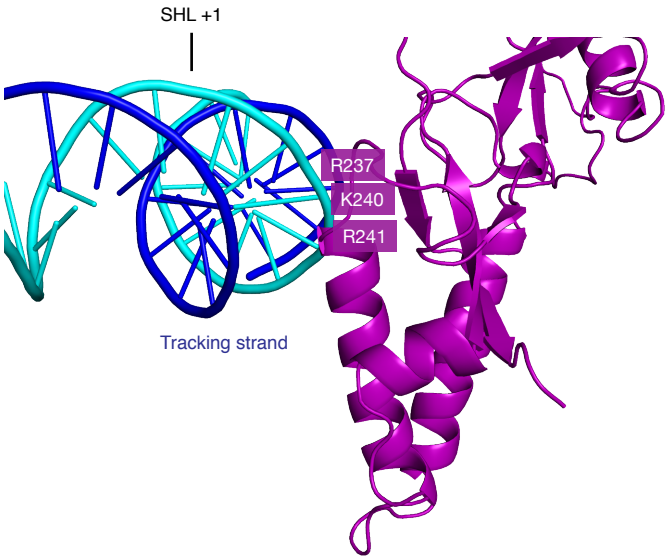

Figure 12: Contact of chromo-wedge with DNA at SHL +1. Basic residues important for DNA interaction are highlighted.

free DNA only weakly activates the ATPase (Hauk et al., 2010), and straight B-DNA would clash with the double chromodomain (Figure 16).

Interactions of the double chromodomain with other Chd1 domains may compensate

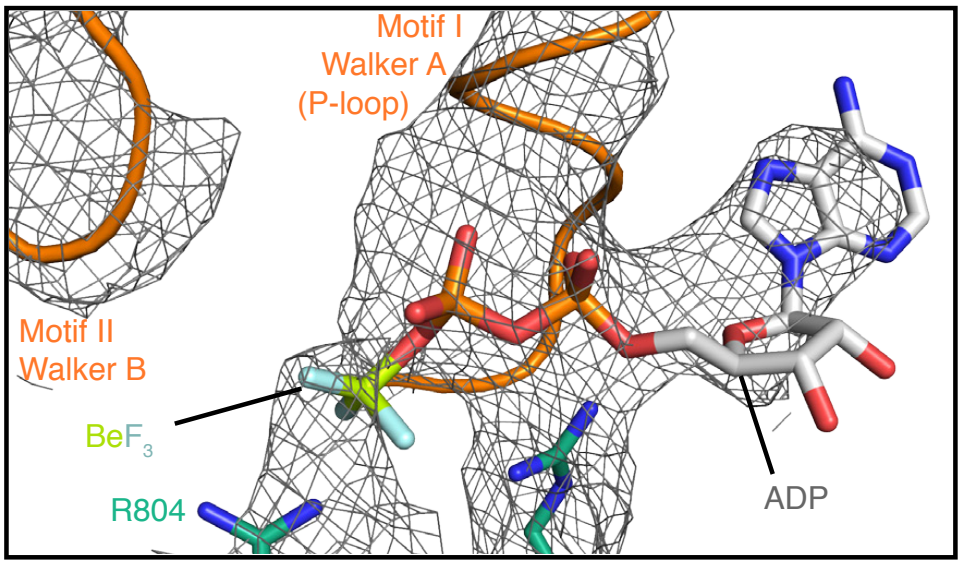

Figure 13: ADP·BeF_3 binds in the active site of the Chd1 ATPase motor. Electron density is shown for ADP·BeF_3, motif I (Walker A, P-loop, residues 403-410), motif II (Walker B, residues 510-515), and the arginine fingers (R804 + R807). Motifs I and II are shown in ribbon representation. ADP·BeF_3 and the arginine finger residues are shown as sticks. Density for ADP is strong, whereas density for BeF3- is weaker and thus we cannot formally rule out that BeF3- is not bound or shows only partial occupancy.

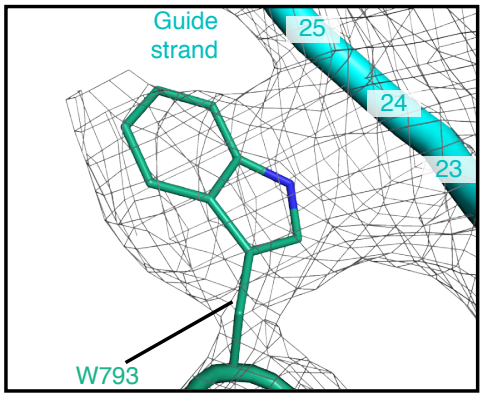

Figure 14: Contact of W793 with the phosphate backbone of the guide strand at SHL +2. Electron density is shown as a grey mesh. Side chain of W793 is shown as a stick representation.

for the loss of histone-DNA contacts upon detaching nucleosomal DNA. The double chromodomain binds the SLIDE domain as predicted (Nodelman et al., 2017) (Figure 18).

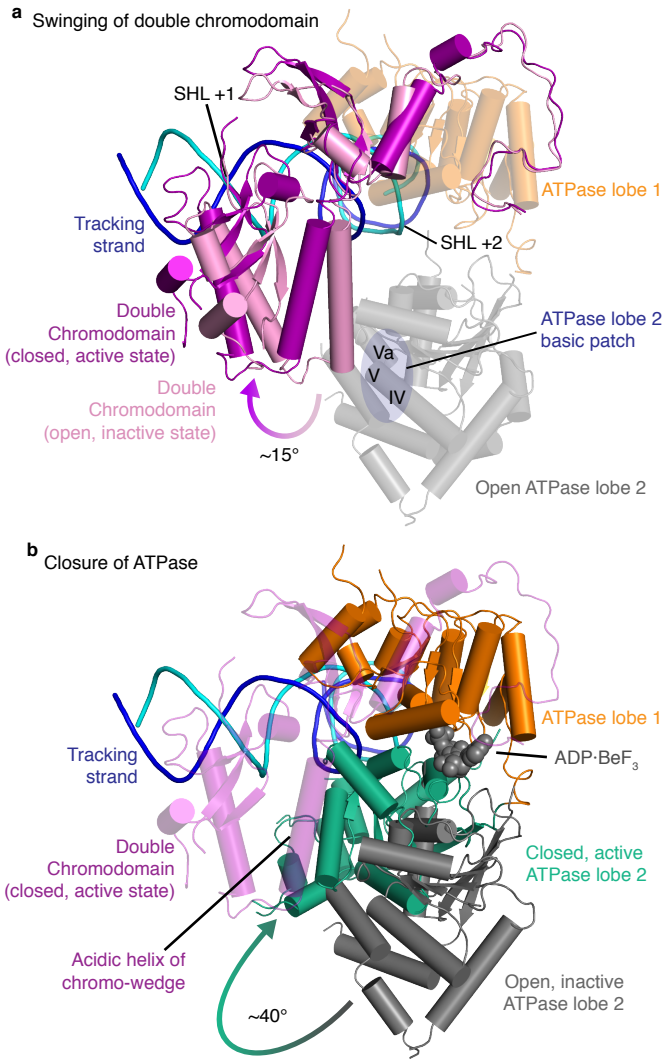

Figure 15: Chd1 structural changes and ATPase activation. **a.** Swinging of double chromodomain (open state, light pink; closed state, purple) onto DNA liberates ATPase lobe 2 (grey). The structure of free Chd1 in its inactive state (Hauk et al., 2010) (PDB code 3MWY) was placed by superimposing ATPase lobe 1 (orange). In the inactive state, the chromo-wedge binds to a basic patch on lobe 2. View as in Figure 7. **b.** ATPase closure and activation. Lobe 2 (sea green) rotates by ∼40° to allow for binding of ADP·BeF_3 (grey spheres). BeF3- was modeled in a tetrahedral conformation for simplicity but is likely planar when it mimics part of the pentavalent transition state of ATP hydrolysis.

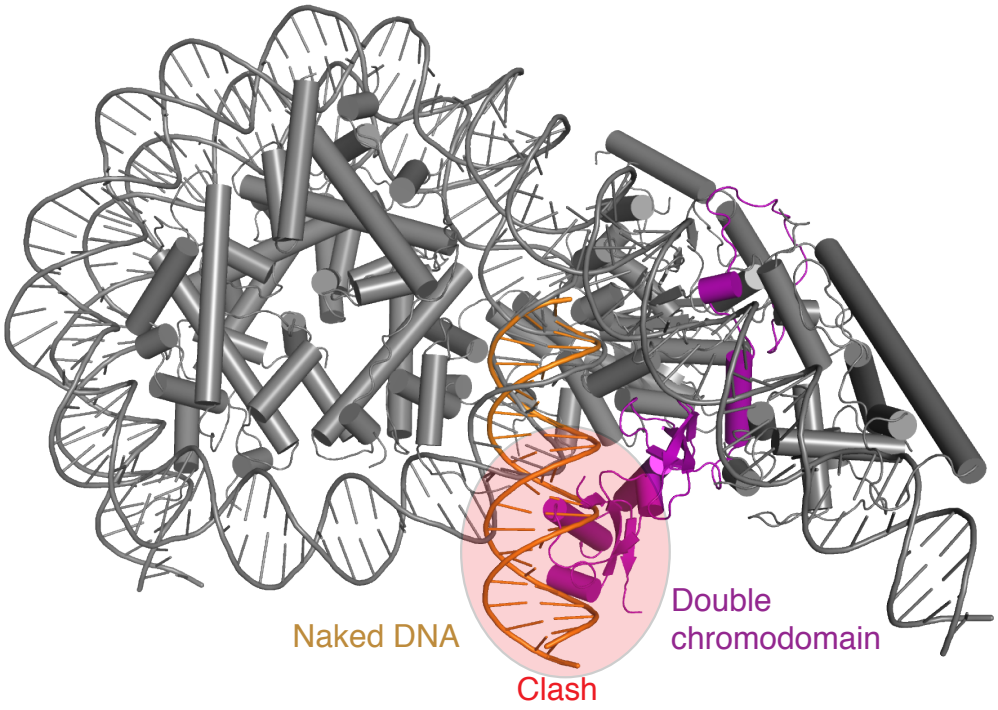

Figure 16: Modeling linear B-DNA (orange) onto the ATPase motor in the nucleosome-Chd1 structure leads to a clash with the double chromodomain (purple). B-DNA was superimposed onto nucleosomal DNA at SHL +2.

It also binds and buttresses lobe 1, which not only contacts SHL +2 but also detached DNA around SHL -6 on the second DNA gyre (Figure 17). In particular, motif Ib and residue 506 (between motifs Ic and II) bind the DNA backbone. These additional contacts between the ATPase and the second DNA gyre resemble the 'secondary' contacts in the nucleosome-Snf2 complex (Liu et al., 2017).

3.1.5 Model for ATPase motor activity

Comparison of our structure with the nucleosome-Snf2 complex (Liu et al., 2017) suggests a model for how ATP binding and hydrolysis result in DNA translocation. In the absence of ATP (Liu et al., 2017), the ATPase is partially closed, whereas in the presence of ADP·BeF_3 it is entirely closed (Figure 13). Superposition of lobe 1 in these two structures results in different positions of lobe 2, which are offset along DNA by approximately one base pair in the direction of translocation (Figure 10). Provided that ATPases move in steps of one base pair (Singleton et al., 2007; Saikrishnan et al., 2009; Hopfner and Michaelis, 2007),

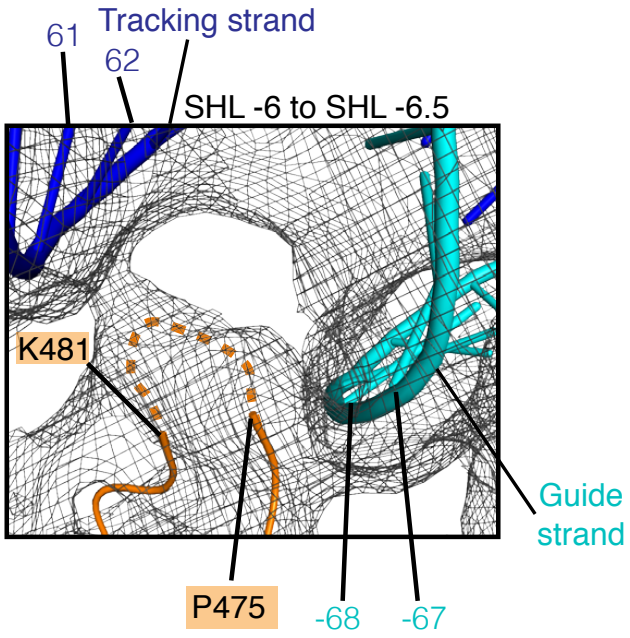

Figure 17: Secondary DNA contacts of ATPase lobe 1. Contact of motif Ib with first DNA gyre around SHL -6.

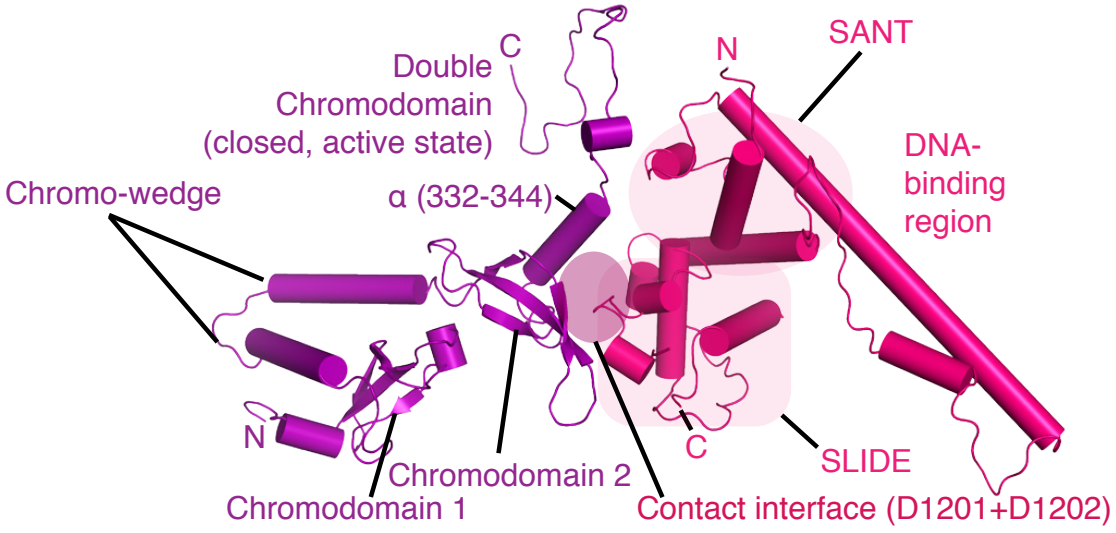

Figure 18: Interface between the double chromodomain and the SANT/SLIDE domains of the DNA binding region. Chd1 domains are colored as in Figure 7.

these observations suggest that the conformational 'ratcheting' between the ATPase lobes underlies DNA translocation (Wigley and Bowman, 2017).

According to this translocation model, the ATPase first binds DNA in a partially closed conformation (pre-translocation state). ATP binding then leads to complete closure of the ATPase and lobe 2 movement, which triggers DNA translocation by one base pair (post-translocation state). ATP hydrolysis then dissociates ADP and resets the ATPase to the pre-translocated state at the new DNA position. We speculate that directional translocation within this enzymatic cycle results from non-equivalent lobe 2 movements during translocation and ATPase resetting.

The structure also reveals the basis for ATPase activation by nucleosome binding (Figure 15). In the absence of the nucleosome, ATPase lobe 2 is sequestered in an open conformation by the 'chromo-wedge', an acidic region in the double chromodomain (Figure 15a) (Hauk et al., 2010). In the presence of the nucleosome, the double chromodomain swings by 15° and binds nucleosomal DNA (Figure 15b) (Nodelman et al., 2017). The chromo-wedge contacts the DNA backbone at SHL +1 (Figure 12) using a region that contains cancer mutations in the human homologue CHD4 (Le Gallo et al., 2012). Thus, binding of Chd1 to nucleosomal DNA induces swinging of the double chromodomain that releases lobe 2 and allows for ATPase closure and activation. Chd1 recognizes bent nucleosomal DNA because free DNA only weakly activates the ATPase (Hauk et al., 2010), and straight B-DNA would clash with the double chromodomain (Figure 16).

3.1.6 Molecular interactions of Chd1 and histones

Our structure also reveals Chd1 interactions with histones. ATPase lobe 2 contacts highly conserved residues in helix α 1 of histone H3. Lobe 2 also uses an acidic pocket to bind to the basic N-terminal tail of histone H4 (Figure 19). This predicts that H4 acetylation or methylation at residues K16 and K20, respectively, alter Chd1 binding. A similar lobe 2-H4 tail interaction is observed in the nucleosome-Snf2 complex (Liu et al., 2017), and the H4-binding pocket is conserved in ISWI (Clapier et al., 2001), suggesting that H4 tail binding is a general feature of remodelling enzymes.

3.1.7 Model for chromatin remodelling by Chd1

Our structural observations and published biochemical data (Nodelman et al., 2017; Sundaramoorthy et al., 2017; McKnight et al., 2011; Saha et al., 2005; Nodelman et al., 2016) converge on a model for nucleosome remodelling by Chd1. Chd1 positions its ATPase

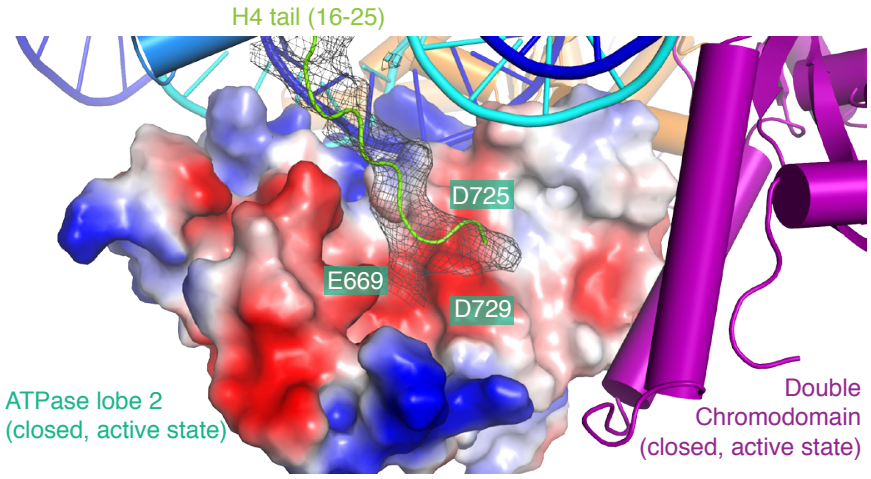

Figure 19: Chd1 binds the N-terminal tail of histone H4 (green) with ATPase lobe 2 (surface representation coloured according to electrostatic surface potential; red, negative, white, neutral, blue, positive). The view is the inverse of that in Figure 7, i.e. after a 180° rotation.

motor at SHL +2 and uses a ratcheting cycle to move on the tracking strand in the 3'-5' direction away from the octamer dyad. As Chd1 holds onto histones, this results in DNA translocation towards the octamer dyad. Progression of the ATPase by one nucleotide per catalytic event (Singleton et al., 2007; Saikrishnan et al., 2009; Hopfner and Michaelis, 2007) leads to a helical rotation of DNA. This may generate a short DNA region that is slightly peeled away from the octamer surface. Propagation of this dissociated region would reposition the octamer, consistent with proposed models (Narlikar et al., 2013; Clapier and Cairns, 2009).

This model for nucleosome remodelling, however, does not explain how Chd1 centers nucleosomes on a DNA fragment and how it induces a regular nucleosome spacing. One possibility is that two Chd1 molecules act from opposite sides of the nucleosome to center it by shifting it away from both DNA ends (Nodelman et al., 2017). Alternatively, a single Chd1 molecule may center the nucleosome if the ATPase motor (McKnight et al., 2011) could swing between two positions on the nucleosome. It is also possible that instead or in addition the DNA-binding region can be repositioned as observed for SNF2h (Leonard and Narlikar, 2015).

In conclusion, our structure of the nucleosome-Chd1 complex provides a framework for understanding nucleosome remodelling and its coupling to other nuclear events. The conservation of Chd1 domains across species and homologues indicates that our results

are relevant for understanding all proteins of the CHD family. The high conservation of the ATPase motor (Figure 20) further suggests that our results can inform mechanistic analysis of other chromatin-remodelling factors, including those of the ISWI family (Yan et al., 2016), which resemble Chd1 in domain architecture.

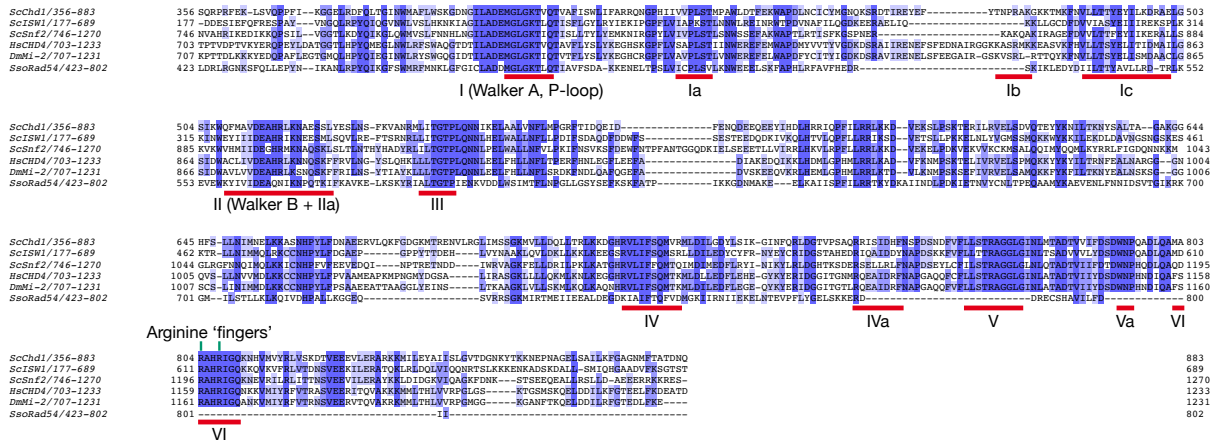

Figure 20: Sequence alignment of ATPase regions in ScChd1 (356-883), ScIsw1 (177-689), ScSnf2 (746-1270), HsChd4 (703-1233), DmMi-2 (707-1231), and SsoRad54 (423-802). Arginine 'fingers' of ScChd1 (R804+R807) are indicated and ATPase motifs are underlined. Sequence coloured according to identity. Darker shades of blue indicate higher conservation, whereas lighter shades of blue indicate less conservation. Alignment was generated with MAFFT (Katoh and Standley, 2013) and visualized using JalView (Waterhouse et al., 2009).

3.2 Mechanism of RNA polymerase II stalling by DNA alkylation

Results presented in this section are in the process of being published:

Stefano Malvezzi*, **Lucas Farnung***, Claudia Aloisi, Todor Angelov, Patrick Cramer, Shana J. Sturla. (2017) Mechanism of RNA polymerase II stalling by DNA alkylation *PNAS.* (accepted in principle)

The results presented in this section are based on the cited publication. Authors contributions are stated in the Publications section. Abstract and introduction are given for context. Experiments presented in the results section were performed by Lucas Farnung, unless otherwise stated. Results that are part of the stated publication but were not obtained by Lucas Farnung are found in the Appendix for context.

3.2.1 Abstract

Several anticancer agents form DNA adducts in the minor groove, interfering with DNA replication and transcription, and inducing apoptosis. Therapeutic resistance can occur, however, when cells are proficient in the removal of drug-induced damage. Acylfulvenes are a class of experimental anticancer agents with a unique repair profile suggesting their capacity to stall RNA polymerase (Pol) II and trigger transcription-coupled nucleotide excision repair. Here we show how different forms of DNA alkylation impair transcription by RNA PolÂăII in cells and with the isolated enzyme, and unravel a new mode of RNA Pol II stalling that is due to alkylation of DNA in the minor groove. We incorporated a model for acylfulvene adducts, the stable 3-deaza-3-methoxynaphtylethyl-adenosine analog (3d-Napht-A), and smaller 3-deaza-adenosine analogs, into DNA oligonucleotides to assess RNA Pol II transcription elongation in vitro. RNA Pol II was strongly blocked by a 3d-Napht-A analog but bypassed smaller analogs. Crystal structure analysis revealed that a DNA base containing 3d-Napht-A can occupy the +1 templating position and impair closing of the trigger loop in the Pol II active center and polymerase translocation into the next template position. These results show how RNA Pol II copes with minor groove DNA alkylation and establish a mechanism for drug resistance.

3.2.2 Introduction

The catalysis of transcription by RNA polymerase is fundamental to the viability of growing cells. In conventional cancer chemotherapy, DNA damage products that interfere with genomic processes initiate cell death. However, therapeutic resistance of cancer cells to DNA alkylating agents may result when cells are proficient in the repair of drug-induced damage (Salehan and Morse, 2013; Cirauqui et al., 2016; Yu et al., 2017).

Acylfulvenes (AFs) are a class of experimental anticancer drugs that appear to be selectively repaired by the transcription-coupled sub-pathway of nucleotide excision repair (TC-NER) (Jaspers et al., 2002; Koeppel et al., 2004; Otto et al., 2017). Thus, AF-induced DNA adducts are preferentially removed from actively transcribed regions of the genome but are largely ignored by global-genome repair. These findings suggest that the AF-DNA adduct does not perturb the duplex structure, yet impedes the progress of RNA polymerase. Understanding the chemical basis of how DNA alkylation stalls RNA polymerase to initiate repair can provide key insight for the design of therapeutics less prone to failure due to resistance.

AFs are semi-synthetic derivatives of the fungal sesquiterpene illudin S and alkylate DNA in the minor groove (Gong et al., 2007), a common property for several anticancer agents such as distamycins, lexitropsins, duocarmycins, and ecteinascidin 743, as well as peptide-based minor groove binders (Hargrove et al., 2015). Hydroxymethylacylfulvene (HMAF, irofulven) was tested previously in clinical trials for several cancers and is anticipated to re-enter clinical trials with biomarker-driven patient stratification strategies. Acylfulvenes are pro-drugs that are reductively activated by prostaglandin reductase 1 (PTGR1) to alkylate primarily position 3 of adenine (Figure B2a) (Gong et al., 2007). Inhibition of DNA synthesis and capacity to induce cell cycle arrest of minor-groove alkylating agents is attributed primarily to alkylation of DNA, and levels of 3-AF-A in AF-treated cancer cells are proportional to cytotoxicity (Gong et al., 2007; Neels et al., 2007; Pietsch et al., 2013; van Midwoud and Sturla, 2013; Woynarowski et al., 1997).

Cells have evolved various DNA repair functions, and large helix-distorting adducts are often removed by the nucleotide excision repair (NER) pathway. NER is divided into two sub-pathways: transcription coupled- (TC-) and global genome- (GG-) NER (Hanawalt, 2002). TC-NER-deficient human fibroblast cells are more sensitive to illudin S, AF and HMAF compared to GG-NER-deficient human fibroblast cells, suggesting that TC-NER selectively repairs AF adducts (Figure B2a) (Jaspers et al., 2002; Koeppel et al., 2004; Otto et al., 2017). Moreover, siRNA-mediated down-regulation of TC-NER in a cancer cell line greatly increased their sensitivity to AF, whereas down-regulation of GG-NER did

not (Figure B2a) (Otto et al., 2017). Since AF adducts are selectively repaired by TCR and ignored by GG-NER, a large portion of adducts that are not repaired in surviving cells are expected to be a basis of the increased mutation frequency observed in cell lines (Glatt et al., 2014). Indeed, error-prone bypass of minor groove modifications has been characterized for certain translesion synthesis DNA polymerases (Malvezzi et al., 2017). Finally, exposing a cancer cell line to UCN-01, a compound that inhibits Chk1, enabling checkpoint evasion and reduction of NER function, increased sensitivity to AF. Moreover, AF adducts persisted longer in co-treated cells (van Midwoud and Sturla, 2013; Jiang and Yang, 1999). The repair and persistence profiles for AF-DNA adducts thus suggests that the basis of selective removal is associated with the DNA damage recognition step.

The main difference between the TC- and GG-sub-pathways of NER is the DNA damage recognition step: GG-NER is activated by XPC-RAD23B factors that sense the altered helix conformation induced by DNA damage, whereas TC-NER is activated by the stalling of RNA polymerase II (Pol II) (Hanawalt, 2002; Ljungman and Lane, 2004). Therefore, a working hypothesis is that 3-AF-A stalls Pol II but does not induce significant helix distortion (Otto et al., 2017). Pol II efficiently transcribes over small DNA lesions like O6-methyl-guanine (O6-Me-G), 8-oxo-guanine (8-oxo-G), N2-1-carboxyethyl-guanine and N3-carboxymethyl-thymine, among others (Burns et al., 2010; Tornaletti et al., 2004; You et al., 2012, 2015), however, larger modifications can induce it to stall. Upon Pol II stalling, downstream factors are recruited to the damage site, and the repair process then progresses in the same manner as for GG-NER (Hanawalt, 2002; Ljungman and Lane, 2004).

The mechanism of Pol II stalling at several DNA lesions has been elucidated. Cyclobutane pyrimidine dimers (CPDs), cisplatin 1,2-d(GPG) intrastrand crosslinks, monofunctional pyriplatin-guanine adducts and 8,5'-cyclo-2'-deoxyadenosine strongly inhibit Pol II progression and crystal structures of yeast Pol II bound to DNA containing these lesions were solved to elucidate the basis of polymerase stalling or bypass (Brueckner et al., 2007; Damsma et al., 2007; Walmacq et al., 2015; Wang et al., 2010; Shin et al., 2017, 2016). Two flexible elements of the Pol II active site, the bridge helix and the trigger loop, are required for nucleotide addition and nucleic acid translocation to the next template position (Xu et al., 2014; Fouqueau et al., 2013; Brueckner et al., 2009; Da et al., 2016). Wang and co-workers reported the blockage of Pol II by minor groove binding pyrrole-imidazole polyamides and identified their interaction with the residues Arg1386 and His1387 in the conserved switch 1 region of Rpb1 by molecular modelling (Xu et al., 2016). Despite these insights, there is no direct observation of Pol II encountering a DNA template with minor groove adducts, and it is unknown how Pol II deals with minor-groove-binding and -modifying agents.

In this study, the question of how minor groove DNA alkylation impedes RNA synthesis was addressed. Evaluating transcription activity in a human cancer cell confirmed the inhibited RNA synthesis. Due to the chemical instability of the native AF adduct (Gong et al., 2007; Woynarowski et al., 1997), we characterized the behavior of Pol II in transcription over its stable analog 3-deaza-3-methoxynaphtylethyl-adenosine (3d-Napht-A) (Figure B2b) (Malvezzi et al., 2017). We tested the capacity of 3d-Napht-A to stall purified RNA Pol II with DNA constructs mimicking the transcription bubble and ternary elongation complexes, and tested the limits of Pol II's size tolerance of minor groove alkyl adducts by carrying out the same investigation with other 3-deaza-alkyl-adenosine analogs of systematically smaller size, i.e. 3-deaza-3-phenethyl-adenosine (3d-Phen-A) and 3-deaza-3-methyl-adenosine (3d-Me-A) (Figure B2c). The mechanism of stalling was furthermore investigated by crystallographic analysis of the stalled Pol II in the ternary elongation complex (EC) with bound DNA template and RNA transcript. These data reveal a previously unobserved mode of Pol II stalling and provide insights into the chemical topology of minor groove modifications required to stall transcription in cancer cells.

3.2.3 Pol II stalling at minor-groove alkylation adducts

We assessed the capacity of the major 3-AF DNA adduct to impede the progress of RNA Pol II using purified yeast Pol II and DNA that contains 3d-Napht-A, a chemically stable analog of the adduct, in constructs of a transcription bubble with 3d-Napht-A four or nine bases downstream of the growing RNA 3'-end (Figure 21b) (Gaykalova et al., 2012). When RNA was extended in the presence of four NTPs, Pol II stalled after nucleotide incorporation opposite the 3d-Napht-A, whereas it transcribed efficiently over unmodified A or the smaller methylated analog 3d-Me-A (Figure 21 b). RNA extension of a transcription bubble with 3d-Napht-A nine bases downstream of the RNA 3'-end behaved similarly (Figure 21 c,d).

3.2.4 Structural study of RNA polymerase II stalling by DNA alkylation

To investigate the mechanism of Pol II stalling opposite an alkylated template base, we determined the structure of a Pol II elongation complex in the presence of 3d-Napth-A. We reconstituted a 12-subunit *S. cerevisiae* Poll II elongation complex (EC) with a modified nucleic acid scaffold (Figure 22).

The resulting complex was crystallized, and the structure was determined using molecular replacement. Electron density was observed for 8 bp of downstream DNA, for the

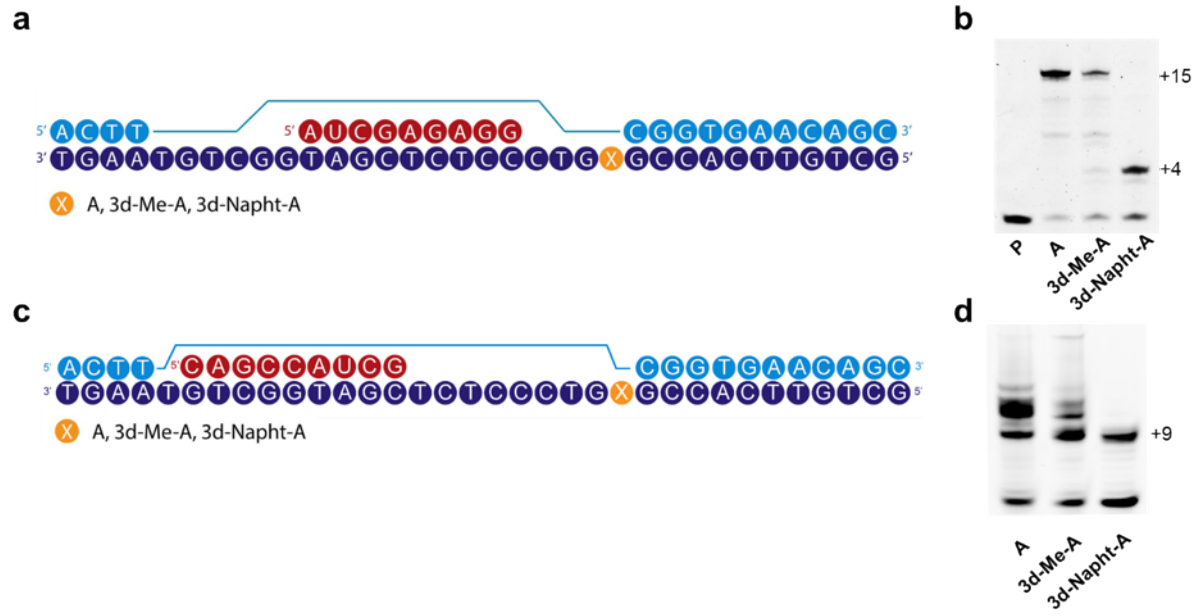

Figure 21: a, c. DNA constructs used for testing the impact of 3d-alkyl-A analogs on the propensity of Pol II to form run-off transcripts. DNA template (dark blue) and RNA primer (red) were annealed, then allowed to react with Pol II (10 min at 25ÂřC). The non-templating DNA strand (light blue) was then annealed to create the transcription bubble. Alkylated adenine analogs were placed four nucleotides (a, b) or nine nucleotides (c, d) downstream of the transcription start on the DNA template. **b, d.** Denaturing gel electrophoresis representing Pol II transcription over A, 3d-Me-A and 3d-Napht-A in presence of all four NTPs (1 mM). Reactions were quenched after 25 min. The site of the modified base is indicated with an x. Experiments in a, b were performed by S. Malvezzi. Experiments performed in c, d were performed by Lucas Farnung.

DNA template strand up to upstream position -10, and for the entire RNA except for the 5'-terminal base at the upstream end of the DNA-RNA hybrid (Figure 23).

The register of nucleic acids was unambiguously defined by bromine labelling of the DNA template strand at position -4 and anomalous diffraction (Figure 24). The DNA modification was clearly observed in the electron density (Figure 24). The structure was refined to a free R factor of 22 % at a resolution of 3.2 Å, resulting in an atomic model with very good stereochemistry (Table B2).

The structure revealed a previously unobserved mode of Pol II interaction with a modified DNA nucleotide. The EC adopts the post-translocation state with the modified adenine base being accommodated in the templating position opposite of the +1 site that binds the

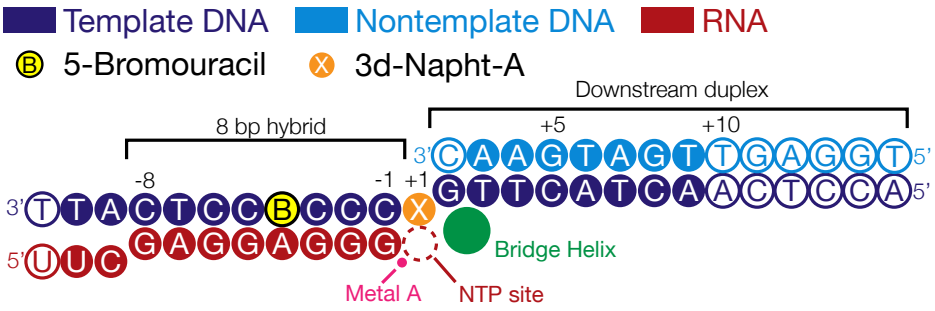

Figure 22: The nucleic acid scaffold used in the crystal structure is depicted schematically with respect to the +1 site containing the 3d-Napht-A base modification. The colour scheme is used throughout the figure.

nucleoside triphosphate substrate (Figure 23). The 3-deaza-3-methoxynaphtylethyl group contacts the central bridge helix in the Pol II active center from underneath. In particular, the second aromatic ring of 3d-Napht-A forms van der Waals contacts with the side chain methyl group of bridge helix residue Thr831. The trigger loop adopts an open conformation, similar to that observed in a previous EC structure that adopts the post-translocation state (PDB 1Y1W) (Kettenberger et al., 2004).

The structure immediately suggests the mechanism of impaired Pol II progression and stalling at an alkylated template base and explains the structure-activity profiles observed herein. Superposition of a Pol II EC structure containing a closed trigger loop (PDB 2E2H) (Wang et al., 2006) showed that the closed state cannot be accommodated in the presence of the alkylated nucleotide due to clashes between trigger loop residues Thr1080 and Leu1081 and the 3-deaza-3-methoxynaphtylethyl group (Figure 25). Inefficient nucleotide incorporation can apparently still take place, because catalysis can occur in a trigger loop-independent, low-fidelity fashion (Toulokhonov et al., 2007), explaining the observed misincorporations. After a single (mis)incorporation event, translocation is strongly impaired, however, as this would result in major clashes of the 3d-Napht-A moiety with Pol II residues that very intimately interact with the minor groove of the hybrid base pair in the post-translocation position -1. In addition, movement of the bridge helix, also required for translocation (Brueckner and Cramer, 2008), is likely impaired by the observed alkyl-bridge helix contact.

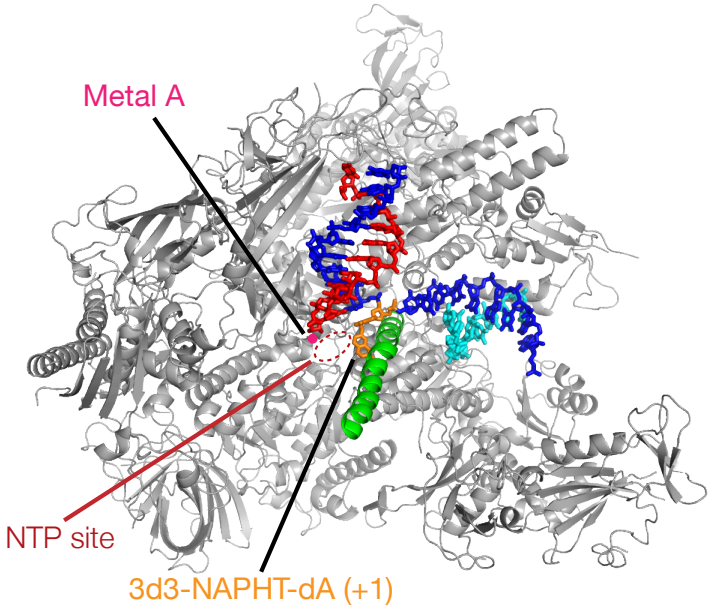

Figure 23: Overview of the Pol II elongation complex structure with the modified adenine present in the +1 site, the nucleotide addition site. Pol II is shown as a silver ribbon model in a side view. The bridge helix is coloured in green. Nucleic acids are presented as stick models. 3d-Napht-A is highlighted in orange.

3.2.5 Discussion

The stalling of isolated Pol II at 3d-Napht-A observed in this study supports the assertion that in cells treated with AFs, transcription is stalled and the stalled Pol II initiates TC-NER and reduces drug action. The chemical basis of this cellular process appears to be associated with the alkyl adduct preventing translocation of the nascent +1 base pair to the next template position due to clashes with Pol II residues binding the minor groove edge of the DNA-RNA base pair at position -1, and by a clash of the second ring on the methoxynaphtyl group with two residues of the mobile trigger loop (Figure 25). These interactions highlight a different Pol II stalling mechanism than predicted by Dervan and Wang in the case of polyamide minor-groove binding agents, wherein Pol II was stalled upstream of the polyamide-bound DNA by interactions between the Switch 1 region and the minor-groove binder impeding translocation (Xu et al., 2016). The structure analysis and structure-activity relationships elucidated in the current study suggests the smaller phenethyl group of 3d-Phen-A does not prevent the trigger loop from closing, allowing Pol II to bypass analogs with only one aromatic ring. Finally, Pol II preferentially incorporated UMP opposite the smaller analogs 3d-Me-A and 3d-Phen-A, and RNA could be extended.

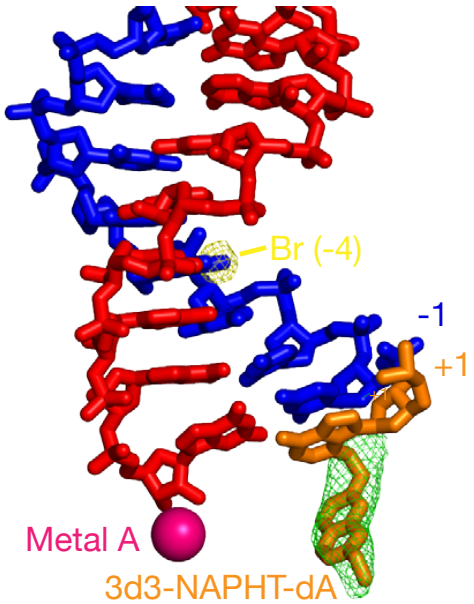

Figure 24: Simulated annealing omit map after removing the 3d-Napht moiety at the +1 site from the model (Fo-FC map, contoured at 3σ, positive density is coloured in green). A peak in the anomalous difference Fourier map (density is coloured in yellow, contoured at 6σ) reveals the exact position of the bromine atom at position -4, allowing unambiguous assignment of the post-translocated state.

The mechanism of Pol II stalling reported herein should inform the design of improved anticancer alkylating agents that inhibit DNA synthesis without impeding the progression of RNA polymerase, thus evading TC-NER. Based on our results, 3-adenosine adducts with only one ring are tolerated by Pol II but do block the replicative DNA polymerase hPol α (Malvezzi et al., 2017). Moreover, the chemical modifications did not destabilize DNA duplexes, suggesting that 3-AF-A adducts do not induce helix distortion and evade GG-NER (Malvezzi et al., 2017). Thus, 3-adenosine adducts containing one phenyl ring may have a desired balance of properties in order to inhibit DNA replication but avoid TC-NER. The detailed insight derived from this study reveals how clashes between DNA adducts and residues binding the DNA-RNA minor groove edge and the mobile trigger loop of RNA Pol II prevent translocation after insertion of a nucleotide opposite an adduct. Knowledge concerning structural characteristics that impede Pol II improves our understanding of how cells initiate the repair of damaged DNA and may also support the design of more effective cancer therapeutics.

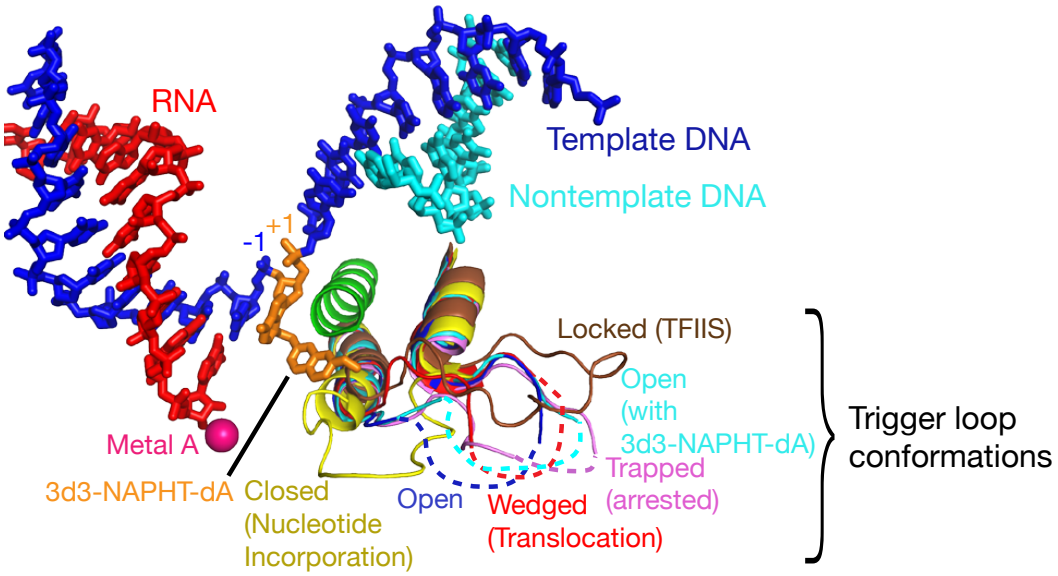

Figure 25: Trigger loop conformations. Comparison of the conformation of the trigger loop observed here (cyan) with alternative conformations observed previously for Pol II. The closed trigger loop (PDB 2E2H), open trigger loop in the post-translocation state (PDB 1Y1W), wedged trigger loop (PDB 2VUM), trapped trigger loop (PDB 3PO2), and locked trigger loop (PDB 3PO3) are depicted in yellow, blue, red, violet, and brown, respectively. Metal A is shown as a magenta sphere.

4 Conclusions and Outlook

The recently published cryo-EM reconstruction of the budding yeast Snf2 ATPase motor bound to a nucleosome provided the first glimpse into the mechanism of a chromatin remodeller on its natural substrate (Liu et al., 2017). The structure confirmed a great number of biochemical data on the proper positioning of the ATPase motor on the nucleosome (Kagalwala et al., 2004; Nodelman et al., 2017). The lack of an ATP analogue, however, prevented the authors from visualizing the engaged ATPase. Furthermore, the authors used a truncated form of Snf2. This impeded elucidation of the exact mechanism of chromatin remodelling and prevented the authors from understanding how regulatory domains influence remodelling activity.

Our NCP-Chd1 structure now presents the first structure of a full-length chromatin remodeller bound to its natural target in the presence of a non-hydrolyzable ATP derivative. The structure, in combination with the already available biochemical and structural data, begins to reveal the intricate intra- and intermolecular regulatory mechanisms that control chromatin remodellers. The use of full-length Chd1 reveals how auxiliary domains control remodelling activity. It further expands our understanding of the conserved ATPase motor shared by SF2 family members (Wigley and Bowman, 2017); Comparison of the apo Snf2 ATPase motor bound to the nucleosome with the ADP·BeF_3 bound structure of Chd1 reveals the pre- and post-translocated state of the ATPase motor and confirms a one-base-pair-per-ATP DNA translocation mechanism. Ratcheting explains the translocation movement which ultimately provides the basis for chromatin remodelling. The ATPase first binds DNA in a partially closed conformation. Binding of ATP then leads to a full closure of the ATPase motor, resulting in fully closed lobe 2. This is the structural rearrangement that provides the basis for DNA translocation by a single base pair step. Hydrolysis of ATP and dissociation of ADP and P_i resets the ATPase. In conclusion, cycles of ATP hydrolysis lead to a movement of the ATPase motor in one direction, while shifting DNA towards the dyad. This pulls DNA in from the opposite site and makes the DNA longer on the proximal side where Chd1 binds, ultimately resulting in chromatin remodeling. On this basis, a full picture for a mechanism of chromatin remodelling starts to emerge that has a broad foundation in biochemical data and structural observations (Figure 26) (Hauk et al., 2010; Nodelman et al., 2017; Sundaramoorthy et al., 2017).

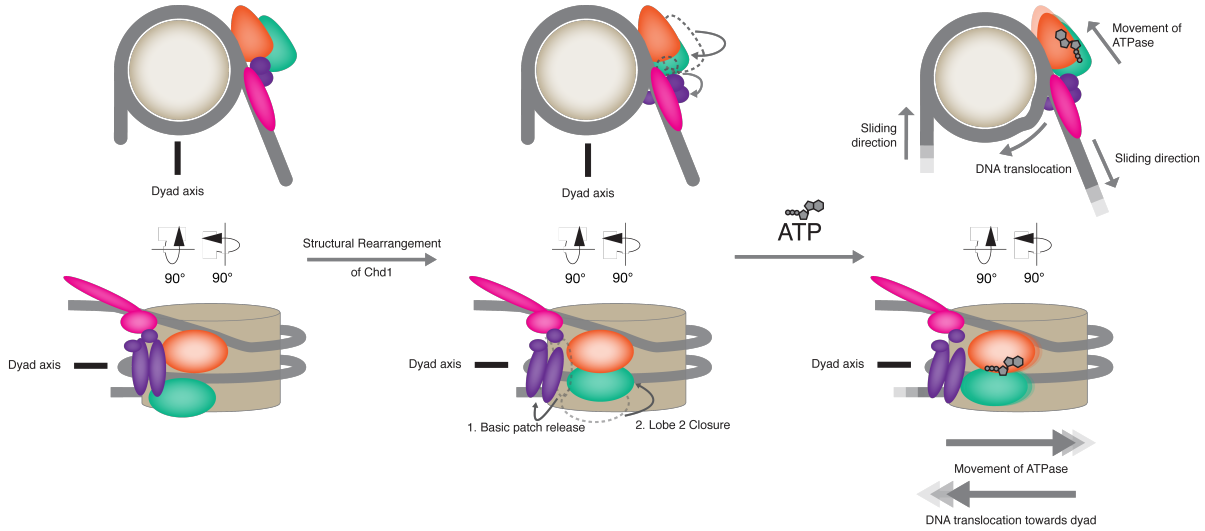

Figure 26: Schematic for a model of the mechanism of chromatin remodelling by Chd1.

4.1 Towards a model for chromatin remodelling

To further elucidate the mechanism of chromatin remodelling, it is necessary to establish the full enzymatic cycle of a single remodeller in the presence of its substrate. So far, two structures of chromatin remodellers bound to the nucleosome have been published/are in press. However, the two structures have trapped the chromatin remodeller in two different states. Snf2 was trapped in a partially closed state with no ATP present (Liu et al., 2017), while Chd1 presented in this study was trapped in a closed, post-translocated state in the presence of the transition state mimicking adduct ADP·BeF_3. To allow for precise understanding of the remodelling mechanism, additional structures of the same remodeller should be solved in the presence of other ATP analogues or ADP. Chd1 seems to be an appropriate target because it allows for study of the ATPase motor as wells as its auxiliary domains. Most chromatin remodellers are components of larger assemblies, where other polypeptides in the complex carry the regulatory functions found in the auxiliary domains of Chd1 (Clapier et al., 2017). To allow proper comparison between the Snf2 structure and our structure, it is important to capture Chd1 bound to a nucleosome in the absence of nucleotide. This would allow for comparison of the translocation state and confirm that the structural rearrangements observed in our structure are a result of the presence of nucleotide.

Chromatin remodellers translocate DNA, hence DNA distortion is expected to occur during the process of remodelling. To capture these distortions it is necessary to obtain

structures at even higher resolutions. Higher resolutions could be obtained by use of a phase plate (Chua et al., 2016), collection at higher magnifications or greater particle number. Perhaps these approaches also require time-sensitive sample preparations where ATP is added to the complex for a short period of time before freezing the sample to trap Chd1 while it is actively remodelling the nucleosome. Such a time-sensitive approach would also require extensive data collection followed by thorough data analysis with a strong emphasis on classifying the captured states to obtain homogenous states at high enough resolutions to shed light on the intricate mechanism of ATPase motor movement and DNA translocation.

Higher resolution reconstructions should also further elucidate the interactions of the chromatin remodellers with the histone octamer. The H4 tail appears to be an important hub for interactions of ATPase lobe 2 with the histone octamer. Additionally, interactions have been also observed with H3. The human CHD1 is able to interact with modified tails of H3 (Sims et al., 2005). There is strong biochemical evidence that the chromatin remodeller auxiliary domains associate with the acidic patch of H2A/H2B, however, this interaction has not been observed in any presently available chromatin remodeller structure (Leonard and Narlikar, 2015; Dann et al., 2017).

4.2 Increasing complexity: chromatin biology and transcription

With recent advances in recombinant protein expression and cryo-EM, it is possible to target larger complex assemblies for structural characterization at resolutions well below ten Ångströms (Vos et al., 2016; Scheres, 2012). Formation of well-defined complexes, which have characteristics favorable for cryo conditions, appears to be the most substantial challenge in visualizing such complexes. Preparation techniques such as GraFix or size exclusion chromatography have proven to be amenable for complex formation (Kastner et al., 2008; Martinez-Rucobo et al., 2015). Complexes can be further stabilized by addition of crosslinkers, specific post translational modifications, additional binding factors, and through use of specific grid chemistries for freezing (Russo and Passmore, 2014). Thus far, each structure that has been successfully obtained by cryo-EM has required different biochemical approaches to stabilize the components.

Studies of RNA polymerase II in its initiation and elongation context have proven to be feasible subject of single particle analysis (Plaschka et al., 2016; He et al., 2016; Bernecky et al., 2017). The cryo-EM reconstruction shown here illustrates that nucleosome-

based complexes are also favorable targets. Indeed, nucleosome complexes are significantly easier to study using cryo-EM than crystallography. However, domains and subcomplexes obtained by crystallographic methods are required for the interpretation of cryo-EM maps, and efforts to obtain such structures should not be reduced.

With the broad evidence suggesting a direct physical interaction between Chd1 and transcription elongation-related factors such as Paf1C and FACT that can also bind Chd1 and the nucleosome, it is now appropriate to focus research on obtaining a stable complex with Pol II, related elongation factors, Chd1, and the nucleosome. An initial approach where the nucleosome is bound to the elongation machinery in *trans* could provide the first insights and provide a blueprint for designing a nucleosomal scaffold that can bind Pol II in *cis*. Ultimately, a RNA polymerase II-nucleosome core particle complex should elucidate how eukaryotic RNA polymerases are able to transcribe through chromatin and unify structural studies of chromatin biology and transcription.

A Appendix: Nucleosome-Chd1 structure

A.1 Supplementary Table

Table A1: Cryo-EM data collection, refinement, and validation statistics

	Nucleosome-Chd1 structure (EMDB-3765) (PDB 5O9G)
Data collection and processing	
Microscope	FEI Titan Krios
Detector	Gatan K2 Summit
Magnification	105,000×
Voltage (kV)	300
Electron exposure ($e^-/Å^2$)	30
Defocus range (μm)	1.25 to 2.75
Pixel size (Å)	1.35
Symmetry imposed	C1
Initial particle images (no.)	990,020
Final particle images (no.)	67,032
Map resolution (Å)	4.8
FSC threshold (Å)	0.143
Refinement	
Initial models used (PDB code)	3LZ0, 3MWY, 3TED
Map sharpening *B* factor ($Å^2$)	-204
Model composition	
Non-hydrogen atoms	19,667
Protein residues	1934
Ligands	2
Validation	
MolProbity Score	1.93

Clashscore	7
Poor rotamers (%)	0.36
Ramachandran plot	
Favored (%)	90.81
Allowed (%)	9.06
Disallowed (%)	0.13

A.2 Supplementary Figures

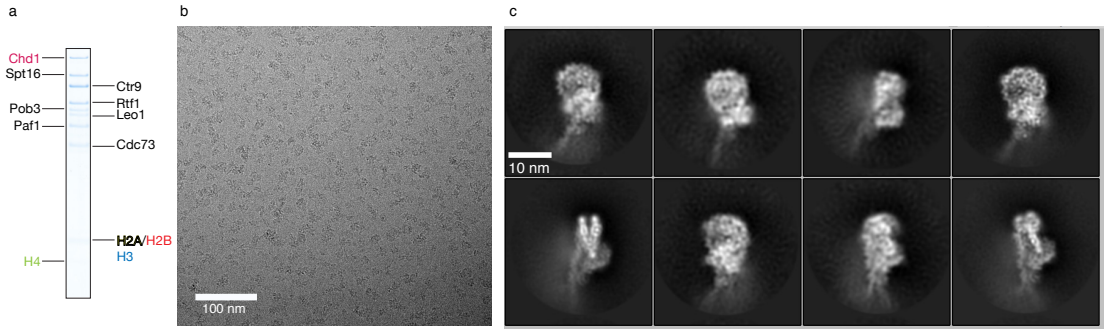

Figure A1: Complex formation and data quality. **a.** Formation of the nucleosome-Chd1-FACT-Paf1C complex. SDS-PAGE of peak fraction used for cryo-EM grid preparation containing Chd1, FACT subunits, Paf1C subunits and histones. Identity of the bands was confirmed by mass spectrometry. For gel source data, see Supplementary Figure 1. **b.** Representative cryo-EM micrograph of data collection. **c.** 2D class averages contain nucleosome-like shapes.

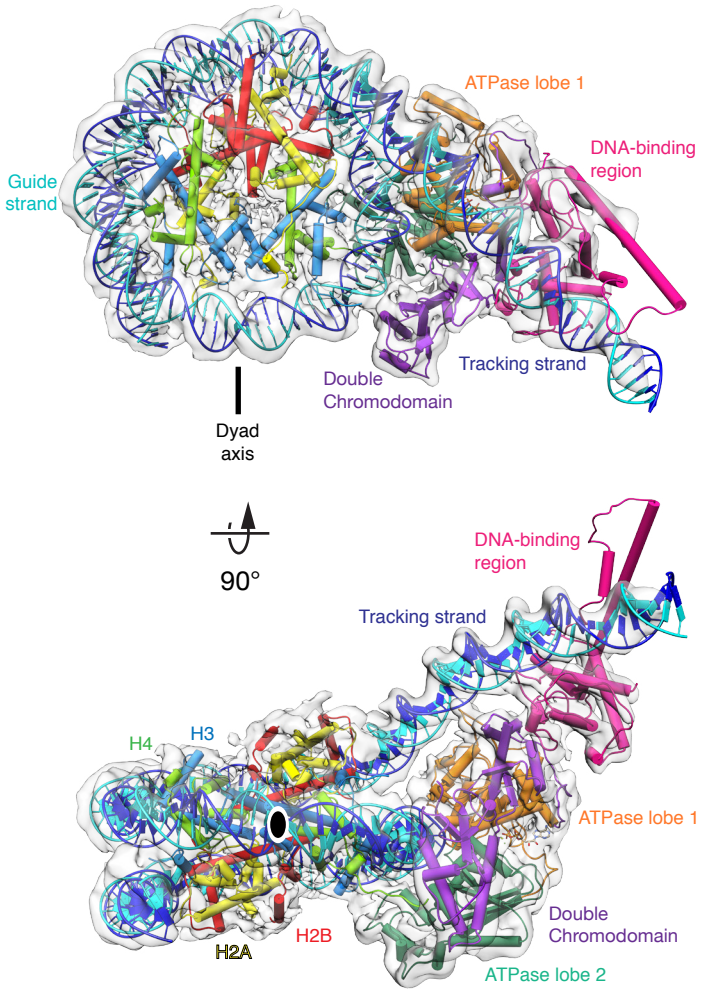

Figure A2: Overall fit of the nucleosome-Chd1 structure. Two views are depicted as in Fig. 7b, c.

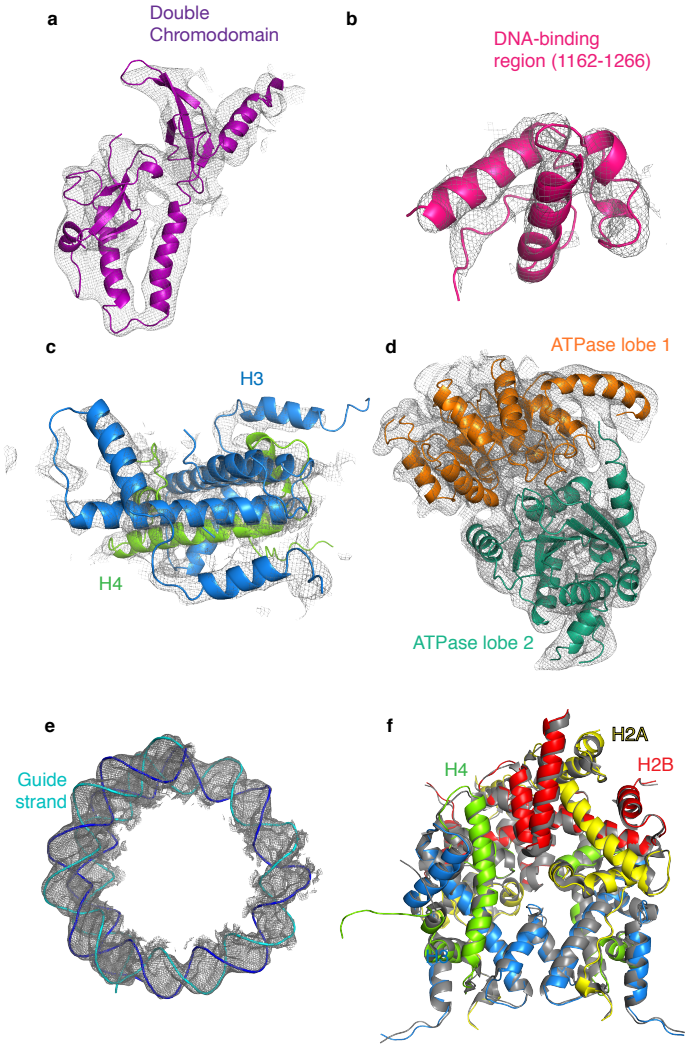

Figure A3: Quality of the nucleosome-Chd1 structure. a-e. Electron density (grey mesh) for various Chd1 domains reveals secondary structure and a good fit for DNA (SHL -4 to SHL +7). f. Superposition of the histone octamer core with canonical octamer core (PDB code 3LZ0). The canonical octamer core is rendered in grey.

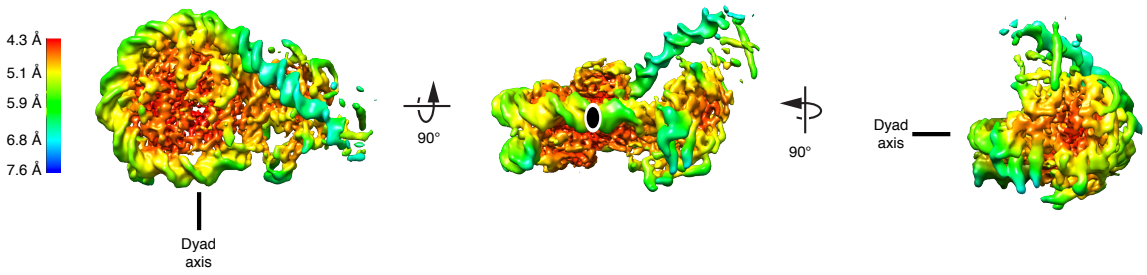

Figure A4: Nucleosome-Chd1 reconstruction colored according to local resolution.

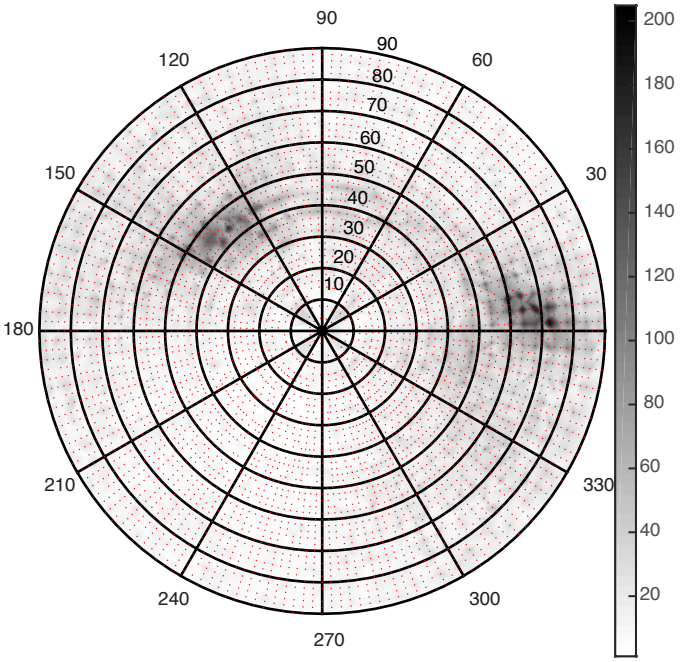

Figure A5: Angular distribution of particles. Red dots indicate the presence of at least one particle image assigned within ± 1°. Shading from white to black indicates the density of particle images at a given orientation.

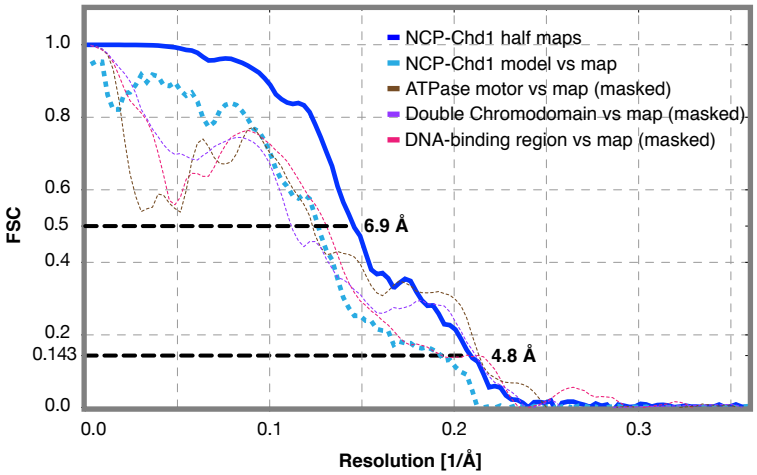

Figure A6: Estimation of the average resolution. The dark blue line indicates the Fourier shell correlation between the half maps of the reconstruction. The dotted light blue line indicates the Fourier shell correlation between the derived model and the reconstruction. Resolutions are given for the FSC 0.143 and the FSC 0.5 criterion. The dotted lines show the Fourier shell correlation between the derived Chd1 domains and the corresponding masked regions.

B Appendix: Mechanism of RNA polymerase II stalling by DNA alkylation

Stefano Malvezzi*, **Lucas Farnung***, Claudia Aloisi, Todor Angelov, Patrick Cramer, Shana J. Sturla. (2017) Mechanism of RNA polymerase II stalling by DNA alkylation *PNAS.* (accepted in principle)

The additional text and figures presented in this section are based on the cited pulication. The figures and text are included to provide additional background on the experiments presented in the Results & Discussion section.

B.1 Supplementary text

B.1.1 DNA alkylation impairs RNA synthesis in cells

It has been established that cells proficient in TC-NER are more resistant to cytotoxicity induced by AFs (Jaspers et al., 2002; Koeppel et al., 2004; Otto et al., 2017). To evaluate, therefore, how AF impacts RNA synthesis in cancer cells relative to a simple methylating agent, we compared RNA synthesis activity following treatment with HMAF vs. MMS. MMS can form various DNA methylation adducts, mainly the non-cytotoxic 7-Me-G, but also 3-Me-A, and to a lesser extent O6-Me-G (Beranek, 1990; Brink et al., 2007). A significant reduction of RNA synthesis in colon adenocarcinoma SW480 cells overexpressing PTGR1 (SW480-PTGR1), an enzyme required for metabolic bioactivation of AFs, was observed after treatment with 1 μM HMAF, but not 1 μM MMS (Figure B6), concentrations at which all cells remained viable (Figure B5). Because MMS is much less potent than HMAF, we also evaluated RNA synthesis at doses that reduce cell viability equally with both compounds (Figure B7b). Again, RNA synthesis was significantly reduced in cells treated with 0.5 μM HMAF, but not in cells treated with 500 μM MMS. These results provide evidence for Pol II stalling after minor groove DNA alkylation in cancer cells.

B.1.2 Pol II transcription can stall at minor-groove alkylation adducts

We next characterized RNA extension with ternary elongation complexes composed of an RNA annealed to a DNA template containing the analog at the nucleotide incorporation site +1 (Figure B3a). In the presence of all four NMPs, Pol II stalled at 3d-Napht-A after incorporating one NMP (Figure B3b) and no bypass was observed even after 60 min. When the same reactions were carried out with individual nucleotides, Pol II incorporated UMP and misincorporated CMP opposite 3d-Napht-A but not GMP and AMP (Figure B7c).

B.1.3 Pol II is tolerant to smaller modifications but errors arise

To derive the maximum adduct size tolerated by Pol II, RNA extension was evaluated for templates containing smaller alkyl groups. Pol II efficiently extended RNA annealed with templates containing A or 3d-Me-A, forming full-length products (Figure B7b). 3d-Phen-A was bypassed, but product accumulated after incorporation of the first nucleotide, suggesting the intermediate size adduct slows Pol II but does not impede it completely.

In the presence of individual nucleotides, UMP was very efficiently incorporated, leading to a +2 band resulting from insertion of UMP opposite A plus the next G. For modified substrates with 3d-Me-A or 3d-Phen-A, however, only the +1 band was observed for UMP, suggesting that the 3-alkyl groups reduced transcription fidelity. A similar degree of CMP misincorporation was observed in all cases, regardless of the DNA being modified or not (Figure B8). Moreover, Pol II efficiently misincorporated CMP opposite the next templating G forming a +2 band product, whereas in the case of 3d-Napht-A, only one nucleotide was inserted opposite the analog.

B.1.4 DNA alkylation alters Pol II incorporation and extension efficiency

Having established that Pol II incorporates UMP and CMP to varying extents depending on the identity of DNA modification, we measured apparent reaction rates for their incorporation by Pol II (Figure B4). For insertion of UTP opposite A or 3d-Me-A, incorporation rates were similar (9.0 ± 1.4 min-*1* and 8.0 ± 1.2 min-*1* respectively), whereas rates opposite 3d-Phen-A and 3d-Napht-A were significantly reduced (2.5 ± 0.2 min^{-1} and 1.2 ± 0.2 min^{-1}, respectively). The CMP misincorporation profile was quite different, however, with more than 10-fold greater incorporation opposite 3d-Me-A vs. A (5.8 ± 0.4

min^{-1} vs. 0.5 ± 0.1 min^{-1}, whereas 3d-Phen-A and 3d-Napht-A were similarly low as A (0.28 ± 0.03 min^{-1} and 0.39 ± 0.05 min^{-1}) (Figure B4a). The increased misincorporation of CMP opposite the 3-alkyl-adenosine analogs relative to A (Figure B4a) suggests a potential basis of transcriptional mutagenesis (TM) in cells (You et al., 2012; Burns et al., 2010). To understand the impact of 3-deaza-3-alkyl-adenosine analogs on Pol II activity during the extension steps following incorporation opposite an adduct, we evaluated RNA synthesis using DNA constructs with the RNA transcript containing either U or C opposite 3-deaza-3-alkyl-adenosine analogs and measured rates of CMP incorporation opposite the next templating G (Figure B4b). Extension after 3d-Me-A was similar to after A, but after 3d-Phen-A, extension from U or C was 170- or 250-fold lower, respectively, and after 3d-Napht-A, extension by Pol II was totally blocked, with no incorporation even after 60 min. This observation was consistent with the initially observed stalling after insertion of one nucleotide opposite the modified base. From these data, it could be concluded that Pol II is stalled by the larger analog 3d-Napht-A, whereas it bypasses smaller analogs with the efficiency decreasing in step with the size of the 3-alkyl group size. The efficient bypass of 3d-Me-A was consistent with the observation that exposing cells to MMS did not reduce transcription.

B.2 Supplementary table

Table B2: X-ray data collection and refinement statistics. Values in parentheses are for the highest-resolution shell.

	EC I with 3-deaza-3-methoxynaphtylethyl group	**EC II with 3-deaza-3-methoxynaphtylethyl group - Bromine peak**
Data collection		
Space group	$C222_1$	$C222_1$
Cell dimensions		
a, b, c Å	221.4, 394.9, 283.7	220.4, 394.2, 283.5
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	49.88 - 3.2 (3.31 -3.2	49.7 - 3.61 (3.73-3.61
$\mathbf{R_{merge}}$	0.137 (2.29)	0.403 (2.103)
$\mathbf{R_{meas}}$	0.142 (2.38)	0.41 (2.18)
$\boldsymbol{I/\sigma(I)}$	16.4 (1.4)	12.9 (1.5)
$\boldsymbol{CC_{1/2}}$	0.999 (0.53)	0.992 (0.711)
Completeness (%)	100 (100)	100 (100)
Redundancy	13.5 (13.3)	38.9 (14.2)
Refinement		
Resolution (Å)	49.88 - 3.2 (3.51 - 3.2)	
No. reflections	202,994 (20,129)	
$\mathbf{R_{work}/R_{free}}$	0.18 (0.33)	0.22 (0.36)
No. atoms	31,670	
Protein/DNA	31,626	
Ligand/Ion	44	
***B* factors (Å²)**		
Protein	139.6	
Ligand/Ion	140.8	
R.m.s. deviations		
Bond lengths (Å)	0.033	
Bond lengths (°)	1.03	

B.3 Supplementary figures

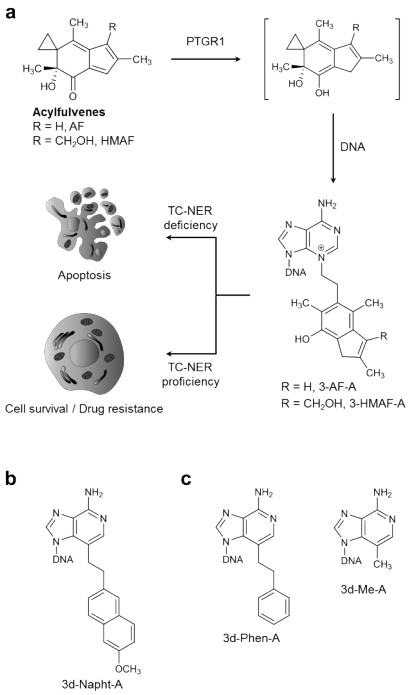

Figure B2: Mechanism of minor groove DNA adduct formation from acylfulvenes and structures of the minor groove adduct models used in this study. **a.** Acylfulvenes are bioactivated in the cytosol by the reductase PTGR1 to form a reactive intermediate that preferentially alkylates position 3 of adenine in genomic DNA. The cytotoxicity of acylfulvenes is dependent on the DNA repair proficiency of exposed cells, with cells deficient in TC-NER being more susceptible than cells proficient in TC-NER. **b.** Structure of the 3-deaza-3-methoxynaphtylethyl-adenosine analog (3d-Napht-A) incorporated in oligonucleotides and used as model of acylfulvene-derived DNA adducts in biochemical and crystallographic studies. **c.** Structure of 3-deaza-adenosine analogs of decreasing size, 3-deaza-3-phenethyl-adenosine (3d-Phen-A) and 3-deaza-3-methyl-adenosine (3-Me-A) used to enable structure-activity analysis of Pol II.

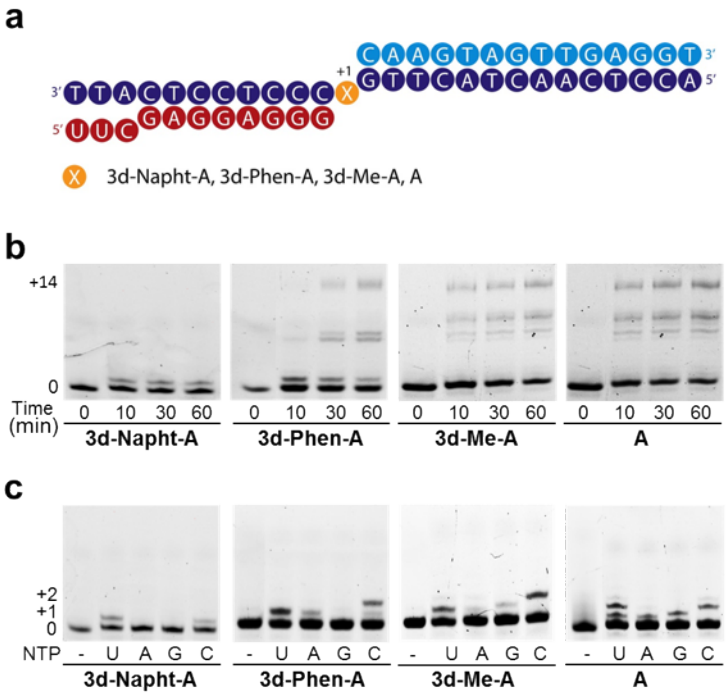

Figure B3: Pol II full-length synthesis and single nucleotide incorporation. **a.** Ternary elongation complexes containing A, 3d-Me-A, 3d-Phen-A or 3d-Napht-A at the underlined position, used for primer extension reactions. DNA template is dark blue, non-template DNA is light blue, RNA primer is red. **b.** Denaturing gel electrophoresis of Pol II transcription products in the presence of all four NTPs (1mM). Extension of the RNA primer to the end of the DNA template forms the indicated run-off product (+14). **c.** Single nucleotide incorporation opposite A, 3d-Me-A, 3d-Phen-A and 3d-Napht-A. Reactions were performed in presence of a single NTP (1mM) and were quenched after 20 min.

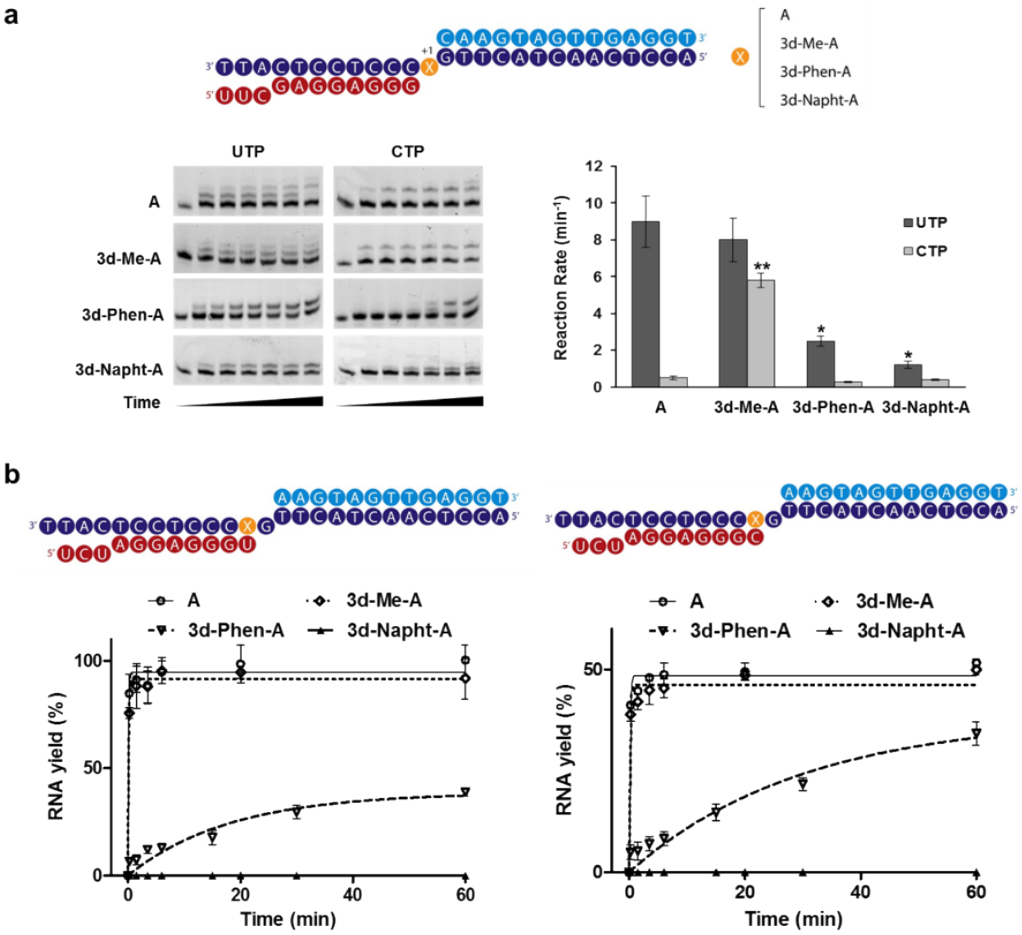

Figure B4: Single nucleotide incorporation and extension kinetics. **a.** Gel denaturing electrophoresis of primer extension kinetics by Pol II for UMP and CMP incorporation opposite unmodified adenine or 3-deaza-adenosine analogs (left). Reactions contained 1 mM NTP and were stopped at various time points between 0 and 60 min. Schematic representation of apparent rate constants for UMP and CMP incorporation kinetics (right). Error bars represent standard deviation (N=3) and statistical significance for NMP incorporation opposite the 3-deaza-3-alkyl-adenosine adducts compared to A was determined by unpaired Student t-test with Welch's correction (*p $<$ 0.05, ** p $<$ 0.01). **b.** Rates (min-1) of RNA primer extension from the correct base pair Aanalog:U (left) and the mismatched base pair Aanalog:C (right). The DNA template contained either A, 3d-Me-A, 3d-Phen-A or 3d-Napht-A at the indicated position. RNA yield is plotted as a function of time (100 % RNA yield corresponds to 50 % of primer extension). Error bars represent standard deviation (N=3).

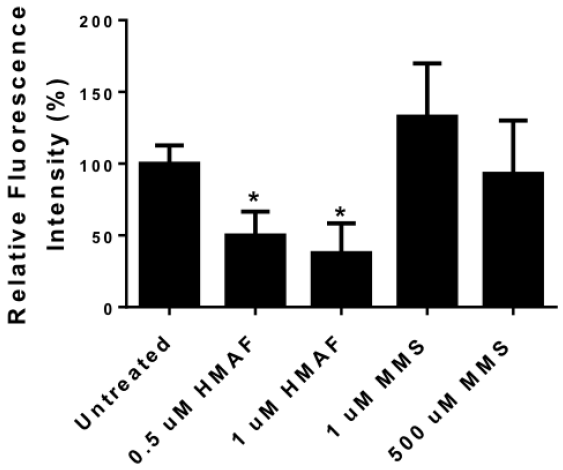

Figure B5: Levels of RNA synthesis in SW480-PTGR1 cells treated with HMAF and MMS. RNA synthesis was quantified on the basis of emitted fluorescence indicating incorporation of 5-ethynyl uridine in RNA during transcription, followed by coupling with the fluorescent azide dye Alexa488. Fluorescence signal was normalized to the signal from untreated cells. Error bars represent standard deviation (N$\geq$8). Statistical significance of treated cells compared to untreated cells was determined by one-way ANOVA with Dunnett post hot test (*p $<$ 0.0001).

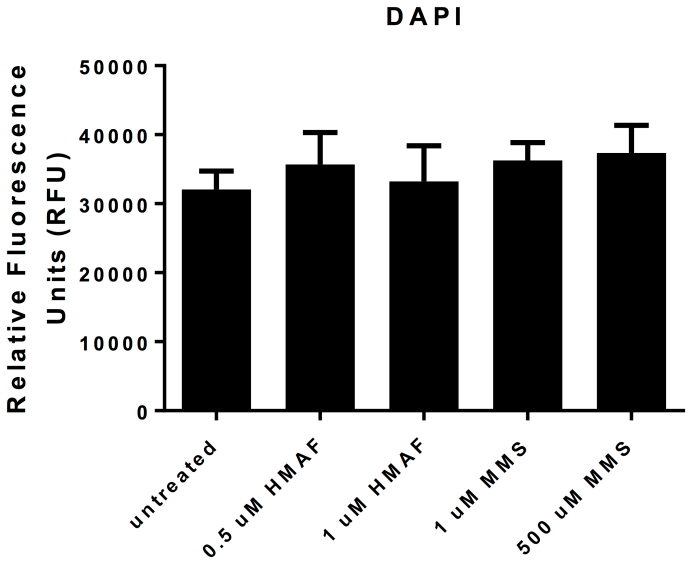

Figure B6: DAPI fluorescence signal in SW480-PTGR1 cells after 6 h of exposure to HMAF (0.5 and 1 μM) or MMS (500 and 1 μM). The average of fluorescence is represented with standard deviation (N=3).

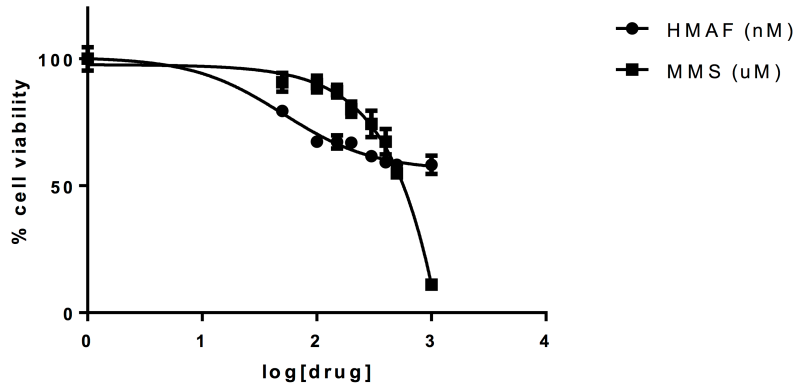

Figure B7: Cell viability of SW480-PTGR1 cells after 6 h of exposure to increasing concentration of HMAF (0-1000 nM) or MMS (0-1000 μM) and 24 of post-incubation. Cell viability was determined with the CellTiter Glow assay. Average of cell viability is represented with standard deviation (N=3).

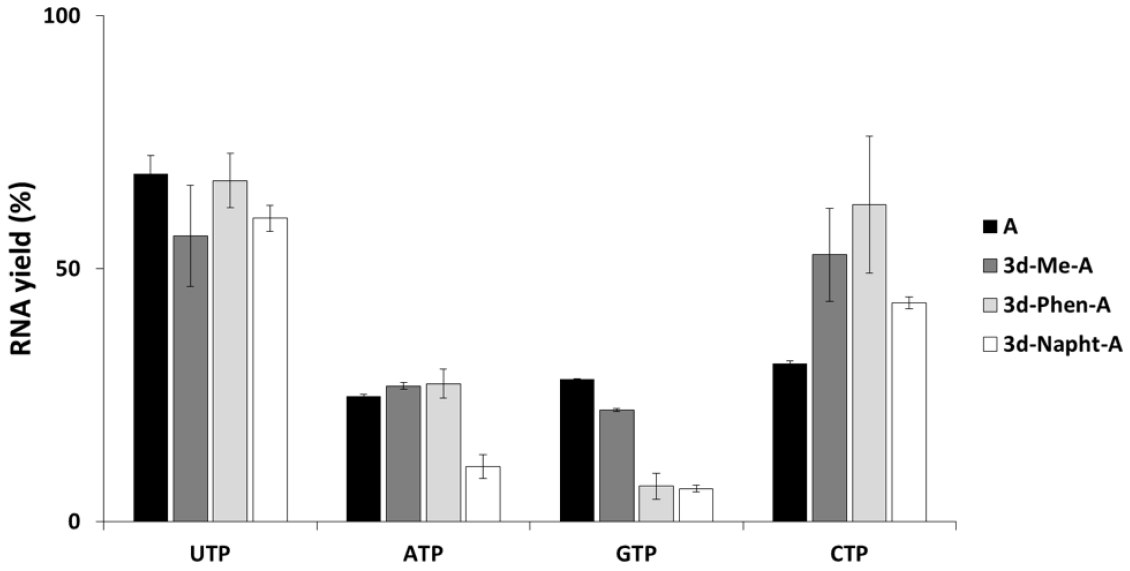

Figure B8: Quantification of single nucleotide incorporation opposite the 3-alkyl adenosine analogs. Reactions were performed in presence of a single NTP (1 mM) and were quenched after 20 min. The average of RNA yield is represented with standard deviation (N=2).

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, 66(Pt 2):213–221.

Allis, C. D. and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.*, 17(8):487–500.

Arents, G., Burlingame, R. W., Wang, B. C., Love, W. E., and Moudrianakis, E. N. (1991). The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. *PNAS*, 88(22):10148–10152.

Armache, K. J., Garlick, J. D., Canzio, D., Narlikar, G. J., and Kingston, R. E. (2011). Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution. *Science*, 334(6058):977–982.

Armstrong, V. W., Yee, D., and Eckstein, F. (1979). Mechanistic studies on deoxyribonucleic acid dependent ribonucleic acid polymerase from Escherichia coli using phosphorothioate analogues. 2. The elongation reaction. *Biochemistry*, 18(19):4120–4123.

Barbera, A. J., Chodaparambil, J. V., Kelley-Clarke, B., Joukov, V., Walter, J. C., Luger, K., and Kaye, K. M. (2006). The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. *Science*, 311(5762):856–861.

Bednar, J., Garcia-Saez, I., Boopathi, R., Cutter, A. R., Papai, G., Reymer, A., Syed, S. H., Lone, I. N., Tonchev, O., Crucifix, C., Menoni, H., Papin, C., Skoufias, D. A., Kurumizaka, H., Lavery, R., Hamiche, A., Hayes, J. J., Schultz, P., Angelov, D., Petosa, C., and Dimitrov, S. (2017). Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Mol. Cell*, 66(3):384–397.e8.

Beranek, D. T. (1990). Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. *Mutat. Res.*, 231(1):11–30.

Bernecky, C., Plitzko, J. M., and Cramer, P. (2017). Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nature Structural & Molecular Biology*, 47:5.6.

Boyer, L. A., Latek, R. R., and Peterson, C. L. (2004). The SANT domain: a unique histone-tail-binding module? *Nat Rev Mol Cell Biol*, 5(2):158–163.

Brink, A., Schulz, B., Stopper, H., and Lutz, W. K. (2007). Biological significance of DNA adducts investigated by simultaneous analysis of different endpoints of genotoxicity in L5178Y mouse lymphoma cells treated with methyl methanesulfonate. *Mutat. Res.*, 625(1-2):94–101.

Brueckner, F. and Cramer, P. (2008). Structural basis of transcription inhibition by α-amanitin and implications for RNA polymerase II translocation. *Nat Struct Mol Biol*, 15(8):811–818.

Brueckner, F., Hennecke, U., Carell, T., and Cramer, P. (2007). CPD damage recognition by transcribing RNA polymerase II. *Science*, 315(5813):859–862.

Brueckner, F., Ortiz, J., and Cramer, P. (2009). A movie of the RNA polymerase nucleotide addition cycle. *Curr. Opin. Struct. Biol.*, 19(3):294–299.

Burns, J. A., Dreij, K., Cartularo, L., and Scicchitano, D. A. (2010). O6-methylguanine induces altered proteins at the level of transcription in human cells. *Nucleic Acids Research*, 38(22):8178–8187.

Bushnell, D. A., Cramer, P., and Kornberg, R. D. (2002). Structural basis of transcription: alpha-amanitin-RNA polymerase II cocrystal at 2.8 A resolution. *PNAS*, 99(3):1218–1222.

Cairns, B. R., Lorch, Y., Li, Y., Zhang, M., Lacomis, L., Erdjument-Bromage, H., Tempst, P., Du, J., Laurent, B., and Kornberg, R. D. (1996). RSC, an Essential, Abundant Chromatin-Remodeling Complex. *Cell*, 87(7):1249–1260.

Cheung, A. C. M. and Cramer, P. (2012). A movie of RNA polymerase II transcription. *Cell*, 149(7):1431–1437.

Cheung, A. C. M., Sainsbury, S., and Cramer, P. (2011). Structural basis of initial RNA polymerase II transcription. *The EMBO Journal*, 30(23):4755–4763.

Chua, E. Y. D., Vogirala, V. K., Inian, O., Wong, A. S. W., Nordenskiöld, L., Plitzko, J. M., Danev, R., and Sandin, S. (2016). 3.9 Å structure of the nucleosome core particle determined by phase-plate cryo-EM. *Nucleic Acids Research*, 44(17):8013–8019.

Cirauqui, B., Margelí, M., Quiroga, V., Quer, A., Karachaliou, N., Chaib, I., Ramírez, J. L., Muñoz, A., Pollán, C., Planas, I., Drozdowsky, A., and Rosell, R. (2016). DNA repair pathways to regulate response to chemoradiotherapy in patients with locally advanced head and neck cancer. *Tumour Biol.*, 37(10):13435–13443.

Clapier, C. R. and Cairns, B. R. (2009). The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.*, 78(1):273–304.

Clapier, C. R., Iwasa, J., Cairns, B. R., and Peterson, C. L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Publishing Group*, 18(7):407–422.

Clapier, C. R., Längst, G., Corona, D. F., Becker, P. B., and Nightingale, K. P. (2001). Critical role for the histone H4 N terminus in nucleosome remodeling by ISWI. *Mol. Cell. Biol.*, 21(3):875–883.

Clark-Adams, C. D. and Winston, F. (1987). The SPT6 gene is essential for growth and is required for delta-mediated transcription in Saccharomyces cerevisiae. *Mol. Cell. Biol.*, 7(2):679–686.

Connelly, S. and Manley, J. L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.*, 2(4):440–452.

Cramer, P., Bushnell, D. A., and Kornberg, R. D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*, 292(5523):1863–1876.

Da, L.-T., Pardo-Avila, F., Xu, L., Silva, D.-A., Zhang, L., Gao, X., Wang, D., and Huang, X. (2016). Bridge helix bending promotes RNA polymerase II backtracking through a critical and conserved threonine residue. *Nat Commun*, 7:11244.

Damsma, G. E., Alt, A., Brueckner, F., Carell, T., and Cramer, P. (2007). Mechanism of transcriptional stalling at cisplatin-damaged DNA. *Nat Struct Mol Biol*, 14(12):1127–1133.

Dang, W. and Bartholomew, B. (2007). Domain architecture of the catalytic subunit in the ISW2-nucleosome complex. *Mol. Cell. Biol.*, 27(23):8306–8317.

Dann, G. P., Liszczak, G. P., Bagert, J. D., Müller, M. M., Nguyen, U. T. T., Wojcik, F., Brown, Z. Z., Bos, J., Panchenko, T., Pihl, R., Pollock, S. B., Diehl, K. L., Allis, C. D., and Muir, T. W. (2017). ISWI chromatin remodellers sense nucleosome modifications to determine substrate preference. *Nature*, 548(7669):607–611.

Dürr, H., Flaus, A., Owen-Hughes, T., and Hopfner, K.-P. (2006). Snf2 family ATPases and DExx box helicases: differences and unifying concepts from high-resolution crystal structures. *Nucleic Acids Research*, 34(15):4160–4167.

Dürr, H., Körner, C., Müller, M., Hickmann, V., and Hopfner, K.-P. (2005). X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA. *Cell*, 121(3):363–373.

Dyer, P. N., Edayathumangalam, R. S., White, C. L., Bao, Y., Chakravarthy, S., Muthurajan, U. M., and Luger, K. (2004). Reconstitution of nucleosome core particles from recombinant histones and DNA. *Meth. Enzymol.*, 375:23–44.

Ebbert, R., Birkmann, A., and Schüller, H. J. (1999). The product of the SNF2/SWI2 paralogue INO80 of Saccharomyces cerevisiae required for efficient expression of various yeast structural genes is part of a high-molecular-weight protein complex. *Mol. Microbiol.*, 32(4):741–751.

Ehara, H., Yokoyama, T., Shigematsu, H., Yokoyama, S., Shirouzu, M., and Sekine, S.-i. (2017). Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science*, 357(6354):921–924.

Elfring, L. K., Deuring, R., McCallum, C. M., Peterson, C. L., and Tamkun, J. W. (1994). Identification and characterization of Drosophila relatives of the yeast transcriptional activator SNF2/SWI2. *Mol. Cell. Biol.*, 14(4):2225–2234.

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, 66(Pt 4):486–501.

Fang, Q., Chen, P., Wang, M., Fang, J., Yang, N., Li, G., and Xu, R.-M. (2016). Human cytomegalovirus IE1 protein alters the higher-order chromatin structure by targeting the acidic patch of the nucleosome. *eLife*, 5:213.

Flaus, A., Luger, K., Tan, S., and Richmond, T. J. (1996). Mapping nucleosome position at single base-pair resolution by using site-directed hydroxyl radicals. *PNAS*, 93(4):1370–1375.

Flaus, A., Martin, D. M. A., Barton, G. J., and Owen-Hughes, T. (2006). Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Research*, 34(10):2887–2905.

Fouqueau, T., Zeller, M. E., Cheung, A. C., Cramer, P., and Thomm, M. (2013). The RNA polymerase trigger loop functions in all three phases of the transcription cycle. *Nucleic Acids Res.*, 41(14):7048–7059.

Gaspar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M. J., Heidersbach, A., Ramalho-Santos, J., McManus, M. T., Plath, K., Meshorer, E., and Ramalho-Santos, M. (2009). Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature*, 460(7257):863–868.

Gaykalova, D. A., Kulaeva, O. I., Pestov, N. A., Hsieh, F.-K., and Studitsky, V. M. (2012). Experimental analysis of the mechanism of chromatin remodeling by RNA polymerase II. *Meth. Enzymol.*, 512:293–314.

Glatt, H., Pietsch, K. E., Sturla, S. J., and Meinl, W. (2014). Sulfotransferase-independent genotoxicity of illudin S and its acylfulvene derivatives in bacterial and mammalian cells. *Arch. Toxicol.*, 88(1):161–169.

Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A., and Kornberg, R. D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science*, 292(5523):1876–1882.

Gong, J., Vaidyanathan, V. G., Yu, X., Kensler, T. W., Peterson, L. A., and Sturla, S. J. (2007). Depurinating acylfulvene-DNA adducts: characterizing cellular chemical reactions of a selective antitumor agent. *J. Am. Chem. Soc.*, 129(7):2101–2111.

Gottesfeld, J. M. and Luger, K. (2001). Energetics and affinity of the histone octamer for defined DNA sequences. *Biochemistry*, 40(37):10927–10933.

Gu, M. and Rice, C. M. (2010). Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. *Proc. Natl. Acad. Sci. U.S.A.*, 107(2):521–528.

Hanawalt, P. C. (2002). Subpathways of nucleotide excision repair and their regulation. *Oncogene*, 21(58):8949–8956.

Hantsche, M. and Cramer, P. (2016). The Structural Basis of Transcription: 10âĂĚYears After the Nobel Prize in Chemistry. *Angew. Chem. Int. Ed. Engl.*, 55(52):15972–15981.

Hantsche, M. and Cramer, P. (2017). Conserved RNA polymerase II initiation complex structure. *Curr. Opin. Struct. Biol.*, 47:17–22.

Harada, B. T., Hwang, W. L., Deindl, S., Chatterjee, N., Bartholomew, B., and Zhuang, X. (2016). Stepwise nucleosome translocation by RSC remodeling complexes. *eLife*, 5:e10051.

Hargrove, A. E., Martinez, T. F., Hare, A. A., Kurmis, A. A., Phillips, J. W., Sud, S., Pienta, K. J., and Dervan, P. B. (2015). Tumor Repression of VCaP Xenografts by a Pyrrole-Imidazole Polyamide. *PLoS ONE*, 10(11):e0143161.

Hauk, G. and Bowman, G. D. (2011). Structural insights into regulation and action of SWI2/SNF2 ATPases. *Curr. Opin. Struct. Biol.*, 21(6):719–727.

Hauk, G., McKnight, J. N., Nodelman, I. M., and Bowman, G. D. (2010). The Chromodomains of the Chd1 Chromatin Remodeler Regulate DNA Access to the ATPase Motor. *Mol. Cell*, 39(5):711–723.

He, Y., Yan, C., Fang, J., Inouye, C., Tjian, R., Ivanov, I., and Nogales, E. (2016). Near-atomic resolution visualization of human transcription promoter opening. *Nature*.

Hochheimer, A., Zhou, S., Zheng, S., Holmes, M. C., and Tjian, R. (2002). TRF2 associates with DREF and directs promoter-selective gene expression in Drosophila. *Nature*, 420(6914):439–445.

Hopfner, K.-P. and Michaelis, J. (2007). Mechanisms of nucleic acid translocases: lessons from structural biology and single-molecule biophysics. *Curr. Opin. Struct. Biol.*, 17(1):87–95.

Huang, S., Gulzar, Z. G., Salari, K., Lapointe, J., Brooks, J. D., and Pollack, J. R. (2012). Recurrent deletion of CHD1 in prostate cancer with relevance to cell invasiveness. *Oncogene*, 31(37):4164–4170.

Hughes, A. L. and Rando, O. J. (2015). Comparative Genomics Reveals Chd1 as a Determinant of Nucleosome Spacing in Vivo. *G3 (Bethesda)*, 5(9):1889–1897.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph*, 14(1):33–8– 27–8.

Jaspers, N. G. J., Raams, A., Kelner, M. J., Ng, J. M. Y., Yamashita, Y. M., Takeda, S., McMorris, T. C., and Hoeijmakers, J. H. J. (2002). Anti-tumour compounds illudin S and Irofulven induce DNA lesions ignored by global repair and exclusively processed by transcription- and replication-coupled repair pathways. *DNA Repair (Amst.)*, 1(12):1027–1038.

Jiang, H. and Yang, L. Y. (1999). Cell cycle checkpoint abrogator UCN-01 inhibits DNA repair: association with attenuation of the interaction of XPA and ERCC1 nucleotide excision repair proteins. *Cancer Res.*, 59(18):4529–4534.

Jonkers, I. and Lis, J. T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nature Publishing Group*, pages 1–11.

Kagalwala, M. N., Glaus, B. J., Dang, W., Zofall, M., and Bartholomew, B. (2004). Topography of the ISW2-nucleosome complex: insights into nucleosome spacing and chromatin remodeling. *The EMBO Journal*, 23(10):2092–2104.

Kastner, B., Fischer, N., Golas, M. M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., Uchtenhagen, H., Urlaub, H., Herzog, F., Peters, J.-M., Poerschke, D., Lührmann, R., and Stark, H. (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. 5(1):53–55.

Kato, H., Jiang, J., Zhou, B.-R., Rozendaal, M., Feng, H., Ghirlando, R., Xiao, T. S., Straight, A. F., and Bai, Y. (2013). A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. *Science*, 340(6136):1110–1113.

Katoh, K. and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, 30(4):772–780.

Kettenberger, H., Armache, K. J., and Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell*, 16(6):955–965.

Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedea, E., Greenblatt, J. F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*, 432(7016):517–522.

Kimanius, D., Forsberg, B. O., Scheres, S. H., and Lindahl, E. (2016). Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife*, 5:e18722.

Koeppel, F., Poindessous, V., Lazar, V., Raymond, E., Sarasin, A., and Larsen, A. K. (2004). Irofulven cytotoxicity depends on transcription-coupled nucleotide excision repair and is correlated with XPG expression in solid tumor cells. *Clin. Cancer Res.*, 10(16):5604–5613.

Koh, F. M., Lizama, C. O., Wong, P., Hawkins, J. S., Zovein, A. C., and Ramalho-Santos, M. (2015). Emergence of hematopoietic stem and progenitor cells involves a Chd1-dependent increase in total nascent transcription. *Proc. Natl. Acad. Sci. U.S.A.*, 112(14):E1734–43.

Kornberg, R. D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science*, 184(4139):868–871.

Krogan, N. J., Baetz, K., Keogh, M.-C., Datta, N., Sawa, C., Kwok, T. C. Y., Thompson, N. J., Davey, M. G., Pootoolal, J., Hughes, T. R., Emili, A., Buratowski, S., Hieter, P., and Greenblatt, J. F. (2004). Regulation of chromosome stability by the histone H2A variant Htz1, the Swr1 chromatin remodeling complex, and the histone acetyltransferase NuA4. *PNAS*, 101(37):13513–13518.

Le Gallo, M., O'Hara, A. J., Rudd, M. L., Urick, M. E., Hansen, N. F., O'Neil, N. J., Price, J. C., Zhang, S., England, B. M., Godwin, A. K., Sgroi, D. C., NIH Intramural Sequencing Center (NISC) Comparative Sequencing Program, Hieter, P., Mullikin, J. C., Merino, M. J., and Bell, D. W. (2012). Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes. *Nat. Genet.*, 44(12):1310–1315.

Leonard, J. D. and Narlikar, G. J. (2015). A nucleotide-driven switch regulates flanking DNA length sensing by a dimeric chromatin remodeler. *Mol. Cell*, 57(5):850–859.

Lieleg, C., Ketterer, P., Nuebler, J., Ludwigsen, J., Gerland, U., Dietz, H., Mueller-Planitz, F., and Korber, P. (2015). Nucleosome spacing generated by ISWI and CHD1 remodelers is constant regardless of nucleosome density. *Mol. Cell. Biol.*, 35(9):1588–1605.

Liu, X., Li, M., Xia, X., Li, X., and Chen, Z. (2017). Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. *Nature*, 544(7651):440–445.

Ljungman, M. and Lane, D. P. (2004). Transcription - guarding the genome by sensing DNA damage. *Nat. Rev. Cancer*, 4(9):727–737.

Lorch, Y., Zhang, M., and Kornberg, R. D. (1999). Histone octamer transfer by a chromatin-remodeling complex. *Cell*, 96(3):389–392.

Lowary, P. T. and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.*, 276(1):19–42.

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, 389(6648):251–260.

Luger, K., Rechsteiner, T. J., and Richmond, T. J. (1999). Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol. Biol.*, 119:1–16.

Luo, W., Johnson, A. W., and Bentley, D. L. (2006). The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev.*, 20(8):954–965.

Lusser, A., Urwin, D. L., and Kadonaga, J. T. (2005). Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. *Nat Struct Mol Biol*, 12(2):160–166.

Makde, R. D., England, J. R., Yennawar, H. P., and Tan, S. (2010). Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature*, 467(7315):562–566.

Malvezzi, S., Angelov, T., and Sturla, S. J. (2017). Minor Groove 3-Deaza-Adenosine Analogues: Synthesis and Bypass in Translesion DNA Synthesis. *Chemistry*, 23(5):1101–1109.

Martinez-Rucobo, F. W. and Cramer, P. (2013). Structural basis of transcription elongation. 1829(1):9–19.

Martinez-Rucobo, F. W., Kohler, R., van de Waterbeemd, M., Heck, A. J. R., Hemann, M., Herzog, F., Stark, H., and Cramer, P. (2015). Molecular Basis of Transcription-Coupled Pre-mRNA Capping. *Mol. Cell*, pages 1–12.

Maskell, D. P., Renault, L., Serrao, E., Lesbats, P., Matadeen, R., Hare, S., Lindemann, D., Engelman, A. N., Costa, A., and Cherepanov, P. (2015). Structural basis for retroviral integration into nucleosomes. *Nature*.

McGinty, R. K., Henrici, R. C., and Tan, S. (2014). Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature*, 514(7524):591–596.

McGinty, R. K. and Tan, S. (2015). Nucleosome Structure and Function. *Chem. Rev.*, 115(6):2255–2273.

McKnight, J. N., Jenkins, K. R., Nodelman, I. M., Escobar, T., and Bowman, G. D. (2011). Extranucleosomal DNA binding directs nucleosome sliding by Chd1. *Mol. Cell. Biol.*, 31(23):4746–4759.

Mitra, D., Parnell, E. J., Landon, J. W., Yu, Y., and Stillman, D. J. (2006). SWI/SNF binding to the HO promoter requires histone acetylation and stimulates TATA-binding protein recruitment. *Mol. Cell. Biol.*, 26(11):4095–4110.

Mizuguchi, G., Shen, X., Landry, J., Wu, W.-H., Sen, S., and Wu, C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science*, 303(5656):343–348.

Mohanty, B., Helder, S., Silva, A. P. G., Mackay, J. P., and Ryan, D. P. (2016). The Chromatin Remodelling Protein CHD1 Contains a Previously Unrecognised C-Terminal Helical Domain. *J. Mol. Biol.*, pages 1–17.

Morrison, A. J. and Shen, X. (2009). Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes. *Nature Publishing Group*, 10(6):373–384.

Narlikar, G. J., Sundaramoorthy, R., and Owen-Hughes, T. (2013). Mechanisms and Functions of ATP-Dependent Chromatin-Remodeling Enzymes. *Cell*, 154(3):490–503.

Neels, J. F., Gong, J., Yu, X., and Sturla, S. J. (2007). Quantitative correlation of drug bioactivation and deoxyadenosine alkylation by acylfulvene. *Chem. Res. Toxicol.*, 20(10):1513–1519.

Neigeborn, L. and Carlson, M. (1984). Genes affecting the regulation of SUC2 gene expression by glucose repression in Saccharomyces cerevisiae. *Genetics*, 108(4):845–858.

Nodelman, I. M., Bleichert, F., Patel, A., Ren, R., Horvath, K. C., Berger, J. M., and Bowman, G. D. (2017). Interdomain Communication of the Chd1 Chromatin Remodeler across the DNA Gyres of the Nucleosome. *Mol. Cell*, 65(3):447–459.e6.

Nodelman, I. M., Horvath, K. C., Levendosky, R. F., Winger, J., Ren, R., Patel, A., Li, M., Wang, M. D., Roberts, E., and Bowman, G. D. (2016). The Chd1 chromatin remodeler can sense both entry and exit sides of the nucleosome. *Nucleic Acids Research*, page gkw406.

Nojima, T., Dienstbier, M., Murphy, S., Proudfoot, N. J., and Dye, M. J. (2013). Definition of RNA polymerase II CoTC terminator elements in the human genome. *CellReports*, 3(4):1080–1092.

Otto, C., Spivak, G., Aloisi, C. M. N., Menigatti, M., Naegeli, H., Hanawalt, P. C., Tanasova, M., and Sturla, S. J. (2017). Modulation of Cytotoxicity by Transcription-Coupled Nucleotide Excision Repair Is Independent of the Requirement for Bioactivation of Acylfulvene. *Chem. Res. Toxicol.*, 30(3):769–776.

Ou, H. D., Phan, S., Deerinck, T. J., Thor, A., Ellisman, M. H., and O'Shea, C. C. (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science*, 357(6349):eaag0025.

Patel, A., Chakravarthy, S., Morrone, S., Nodelman, I. M., McKnight, J. N., and Bowman, G. D. (2013). Decoupling nucleosome recognition from DNA binding dramatically alters the properties of the Chd1 chromatin remodeler. *Nucleic Acids Research*, 41(3):1637–1648.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem*, 25(13):1605–1612.

Pietsch, K. E., van Midwoud, P. M., Villalta, P. W., and Sturla, S. J. (2013). Quantification of acylfulvene- and illudin S-DNA adducts in cells with variable bioactivation capacities. *Chem. Res. Toxicol.*, 26(1):146–155.

Plaschka, C., Hantsche, M., Dienemann, C., Burzinski, C., Plitzko, J., and Cramer, P. (2016). Transcription initiation complex structures elucidate DNA opening. *Nature*.

Plaschka, C., Lariviere, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E. V., Borchers, C. H., Baumeister, W., Herzog, F., Villa, E., and Cramer, P. (2015). Architecture of the RNA polymerase II– Mediator core initiation complex. *Nature*, pages 1–20.

Porrua, O. and Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. *Nature Publishing Group*, pages 1–13.

Proudfoot, N. J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science*, 352(6291):aad9926–aad9926.

Quan, J. and Tian, J. (2011). Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat Protoc*, 6(2):242–251.

Richard, P. and Manley, J. L. (2009). Transcription termination by nuclear RNA polymerases. *Genes Dev.*, 23(11):1247–1269.

Roeder, R. G. and Rutter, W. J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature*, 224(5216):234–237.

Russo, C. J. and Passmore, L. A. (2014). Ultrastable gold substrates for electron cryomicroscopy. *Science*.

Saha, A., Wittmeyer, J., and Cairns, B. R. (2002). Chromatin remodeling by RSC involves ATP-dependent DNA translocation. *Genes Dev.*, 16(16):2120–2134.

Saha, A., Wittmeyer, J., and Cairns, B. R. (2005). Chromatin remodeling through directional DNA translocation from an internal nucleosomal site. *Nat Struct Mol Biol*, 12(9):747–755.

Saha, A., Wittmeyer, J., and Cairns, B. R. (2006). Chromatin remodelling: the industrial revolution of DNA around histones. *Nat Rev Mol Cell Biol*, 7(6):437–447.

Saikrishnan, K., Powell, B., Cook, N. J., Webb, M. R., and Wigley, D. B. (2009). Mechanistic basis of 5'-3' translocation in SF1B helicases. *Cell*, 137(5):849–859.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nature Publishing Group*, pages 1–15.

Sainsbury, S., Niesser, J., and Cramer, P. (2013). Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature*, 493(7432):437–440.

Salehan, M. R. and Morse, H. R. (2013). DNA damage repair and tolerance: a role in chemotherapeutic drug resistance. *Br. J. Biomed. Sci.*, 70(1):31–40.

Schalch, T., Duda, S., Sargent, D. F., and Richmond, T. J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, 436(7047):138–141.

Scheres, S. H. W. (2012). A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.*, 415(2):406–418.

Schreieck, A., Easter, A. D., Etzold, S., Wiederhold, K., Lidschreiber, M., Cramer, P., and Passmore, L. A. (2014). RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nat Struct Mol Biol*, 21(2):175–179.

Schrodinger„ LLC (2015). The PyMOL Molecular Graphics System, Version 1.8.

Schubert, H. L., Wittmeyer, J., Kasten, M. M., Hinata, K., Rawling, D. C., Heroux, A., Cairns, B. R., and Hill, C. P. (2013). Structure of an actin-related subcomplex of the SWI/SNF chromatin remodeler. *Proc. Natl. Acad. Sci. U.S.A.*, 110(9):3345–3350.

Schwanbeck, R., Xiao, H., and Wu, C. (2004). Spatial contacts and nucleosome step movements induced by the NURF chromatin remodeling complex. *Journal of Biological Chemistry*, 279(38):39933–39941.

Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S., and Yokoyama, S. (2006). Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. *Cell*, 125(2):287–300.

Sharma, A., Jenkins, K. R., Heroux, A., and Bowman, G. D. (2011). Crystal structure of the chromodomain helicase DNA-binding protein 1 (Chd1) DNA-binding domain in complex with DNA. *J. Biol. Chem.*, 286(49):42099–42104.

Shen, X., Mizuguchi, G., Hamiche, A., and Wu, C. (2000). A chromatin remodelling complex involved in transcription and DNA processing. *Nature*, 406(6795):541–544.

Shin, J. H., Xu, L., and Wang, D. (2016). RNA polymerase II acts as a selective sensor for DNA lesions and endogenous DNA modifications. *Transcription*, 7(3):57–62.

Shin, J. H., Xu, L., and Wang, D. (2017). Mechanism of transcription-coupled DNA modification recognition. *Cell Biosci*, 7(1):9.

Simic, R., Lindstrom, D. L., Tran, H. G., Roinick, K. L., Costa, P. J., Johnson, A. D., Hartzog, G. A., and Arndt, K. M. (2003). Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. *The EMBO Journal*, 22(8):1846–1856.

Simpson, R. T. (1978). Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry*, 17(25):5524–5531.

Sims, R. J., Chen, C.-F., Santos-Rosa, H., Kouzarides, T., Patel, S. S., and Reinberg, D. (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *Journal of Biological Chemistry*, 280(51):41789–41792.

Singleton, M. R., Dillingham, M. S., and Wigley, D. B. (2007). Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.*, 76(1):23–50.

Sinha, K. K., Gross, J. D., and Narlikar, G. J. (2017). Distortion of histone octamer core promotes nucleosome mobilization by a chromatin remodeler. *Science*, 355(6322):eaaa3761.

Skene, P. J., Hernandez, A. E., Groudine, M., and Henikoff, S. (2014). The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. *eLife*, 3:e02042.

Smolle, M., Venkatesh, S., Gogol, M. M., Li, H., Zhang, Y., Florens, L., Washburn, M. P., and Workman, J. L. (2012). Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nat Struct Mol Biol*, 19(9):884–892.

Sugathan, A., Biagioli, M., Golzio, C., Erdin, S., Blumenthal, I., Manavalan, P., Ragavendran, A., Brand, H., Lucente, D., Miles, J., Sheridan, S. D., Stortchevoi, A., Kellis, M., Haggarty, S. J., Katsanis, N., Gusella, J. F., and Talkowski, M. E. (2014). CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proc. Natl. Acad. Sci. U.S.A.*, 111(42):E4468–77.

Sundaramoorthy, R., Hughes, A. L., Singh, V., Wiechens, N., Ryan, D. P., El-Mkami, H., Petoukhov, M., Svergun, D. I., Treutlein, B., Quack, S., Fischer, M., Michaelis, J., Böttcher, B., Norman, D. G., Owen-Hughes, T., and Formosa, T. (2017). Structural reorganization of the chromatin remodeling enzyme Chd1 upon engagement with nucleosomes. *eLife*, 6:e22510–28.

Sydow, J. F., Brueckner, F., Cheung, A. C. M., Damsma, G. E., Dengl, S., Lehmann, E., Vassylyev, D., and Cramer, P. (2009). Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol. Cell*, 34(6):710–721.

Sydow, J. F. and Cramer, P. (2009). RNA polymerase fidelity and transcriptional proofreading. *Curr. Opin. Struct. Biol.*, 19(6):732–739.

Tan, S. and Davey, C. A. (2011). Nucleosome structural studies. *Curr. Opin. Struct. Biol.*, 21(1):128–136.

Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., and Patel, D. J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature Structural & Molecular Biology*, 14(11):1025–1040.

Thomsen, N. D. and Berger, J. M. (2009). Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell*, 139(3):523–534.

Tornaletti, S., Maeda, L. S., Kolodner, R. D., and Hanawalt, P. C. (2004). Effect of 8-oxoguanine on transcription elongation by T7 RNA polymerase and mammalian RNA polymerase II. *DNA Repair (Amst.)*, 3(5):483–494.

Toulokhonov, I., Zhang, J., Palangat, M., and Landick, R. (2007). A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Mol. Cell*, 27(3):406–419.

Trabuco, L. G., Villa, E., Mitra, K., Frank, J., and Schulten, K. (2008). Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure*, 16(5):673–683.

Tran, H. G., Steger, D. J., Iyer, V. R., and Johnson, A. D. (2000). The chromo domain protein chd1p from budding yeast is an ATP-dependent chromatin-modifying factor. *The EMBO Journal*, 19(10):2323–2331.

Tsunaka, Y., Kajimura, N., Tate, S.-i., and Morikawa, K. (2005). Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle. *Nucleic Acids Research*, 33(10):3424–3434.

van Attikum, H., Fritsch, O., and Gasser, S. M. (2007). Distinct roles for SWR1 and INO80 chromatin remodeling complexes at chromosomal double-strand breaks. *The EMBO Journal*, 26(18):4113–4125.

van Midwoud, P. M. and Sturla, S. J. (2013). Improved efficacy of acylfulvene in colon cancer cells when combined with a nuclear excision repair inhibitor. *Chem. Res. Toxicol.*, 26(11):1674–1682.

van Steensel, B. (2011). Chromatin: constructing the big picture. *EMBO J.*, 30(10):1885–1895.

Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H., and Artsimovitch, I. (2007a). Structural basis for transcription elongation by bacterial RNA polymerase. *Nature*, 448(7150):157–162.

Vassylyev, D. G., Vassylyeva, M. N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007b). Structural basis for substrate loading in bacterial RNA polymerase. *Nature*, 448(7150):163–168.

Vasudevan, D., Chua, E. Y. D., and Davey, C. A. (2010). Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J. Mol. Biol.*, 403(1):1–10.

Vos, S. M., Pöllmann, D., Caizzi, L., Hofmann, K. B., Rombaut, P., Zimniak, T., Herzog, F., and Cramer, P. (2016). Architecture and RNA binding of the human negative elongation factor. *eLife*, 5:e14981.

Walmacq, C., Wang, L., Chong, J., Scibelli, K., Lubkowska, L., Gnatt, A., Brooks, P. J., Wang, D., and Kashlev, M. (2015). Mechanism of RNA polymerase II bypass of oxidative cyclopurine DNA lesions. *Proc. Natl. Acad. Sci. U.S.A.*, page 201415186.

Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D., and Kornberg, R. D. (2006). Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell*, 127(5):941–954.

Wang, D., Zhu, G., Huang, X., and Lippard, S. J. (2010). X-ray structure and mechanism of RNA polymerase II stalled at an antineoplastic monofunctional platinum-DNA adduct. *Proc. Natl. Acad. Sci. U.S.A.*, 107(21):9584–9589.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9):1189–1191.

Weiss, K., Terhal, P. A., Cohen, L., Bruccoleri, M., Irving, M., Martinez, A. F., Rosenfeld, J. A., Machol, K., Yang, Y., Liu, P., Walkiewicz, M., Beuten, J., Gomez-Ospina, N., Haude, K., Fong, C.-T., Enns, G. M., Bernstein, J. A., Fan, J., Gotway, G., Ghorbani, M., DDD Study, van Gassen, K., Monroe, G. R., van Haaften, G., Basel-Vanagaite, L., Yang, X.-J., Campeau, P. M., and Muenke, M. (2016). De Novo Mutations in CHD4, an ATP-Dependent Chromatin Remodeler Gene, Cause an Intellectual Disability Syndrome with Distinctive Dysmorphisms. *Am. J. Hum. Genet.*, 99(4):934–941.

Werner, F. and Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Micro*, 9(2):85–98.

West, S., Gromak, N., and Proudfoot, N. J. (2004). Human 5' –¿ 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, 432(7016):522–525.

Westover, K. D., Bushnell, D. A., and Kornberg, R. D. (2004). Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science*, 303(5660):1014–1016.

White, C. L., Suto, R. K., and Luger, K. (2001). Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *The EMBO Journal*, 20(18):5207–5218.

Wigley, D. B. and Bowman, G. D. (2017). A glimpse into chromatin remodeling. *Nature Structural & Molecular Biology*, 24(6):498–500.

Woynarowski, J. M., Napier, C., Koester, S. K., Chen, S. F., Troyer, D., Chapman, W., and MacDonald, J. R. (1997). Effects on DNA integrity and apoptosis induction by a novel antitumor sesquiterpene drug, 6-hydroxymethylacylfulvene (HMAF, MGI 114). *Biochem. Pharmacol.*, 54(11):1181–1193.

Xu, L., Butler, K. V., Chong, J., Wengel, J., Kool, E. T., and Wang, D. (2014). Dissecting the chemical interactions and substrate structural signatures governing RNA polymerase II trigger loop closure by synthetic nucleic acid analogues. *Nucleic Acids Research*, 42(9):5863–5870.

Xu, L., Wang, W., Gotte, D., Yang, F., Hare, A. A., Welch, T. R., Li, B. C., Shin, J. H., Chong, J., Strathern, J. N., Dervan, P. B., and Wang, D. (2016). RNA polymerase II senses obstruction in the DNA minor groove via a conserved sensor motif. *Proc. Natl. Acad. Sci. U.S.A.*, 113(44):12426–12431.

Xu, Y., Bernecky, C., Lee, C.-T., Maier, K. C., Schwalb, B., Tegunov, D., Plitzko, J. M., Urlaub, H., and Cramer, P. (2017). Architecture of the RNA polymerase II-Paf1C-TFIIS transcription elongation complex. *Nat Commun*, 8:15741.

Yan, L., Wang, L., Tian, Y., Xia, X., and Chen, Z. (2016). Structure and regulation of the chromatin remodeller ISWI. *Nature*, 540(7633):466–469.

Yee, D., Armstrong, V. W., and Eckstein, F. (1979). Mechanistic studies on deoxyribonucleic acid dependent ribonucleic acid polymerase from Escherichia coli using phosphorothioate analogues. 1. Initiation and pyrophosphate exchange reactions. *Biochemistry*, 18(19):4116–4120.

You, C., Dai, X., Yuan, B., Wang, J., Wang, J., Brooks, P. J., Niedernhofer, L. J., and Wang, Y. (2012). A quantitative assay for assessing the effects of DNA lesions on transcription. *Nat. Chem. Biol.*, 8(10):817–822.

You, C., Wang, J., Dai, X., and Wang, Y. (2015). Transcriptional inhibition and mutagenesis induced by N-nitroso compound-derived carboxymethylated thymidine adducts in DNA. *Nucleic Acids Research*, 43(2):1012–1018.

Yu, E. Y., Steinberg-Neifach, O., Dandjinou, A. T., Kang, F., Morrison, A. J., Shen, X., and Lue, N. F. (2007). Regulation of telomere structure and functions by subunits of the INO80 chromatin remodeling complex. *Mol. Cell. Biol.*, 27(16):5639–5649.

Yu, W.-K., Wang, Z., Fong, C.-C., Liu, D., Yip, T.-C., Au, S.-K., Zhu, G., and Yang, M. (2017). Chemoresistant lung cancer stem cells display high DNA repair capability to remove cisplatin-induced DNA damage. *Br. J. Pharmacol.*, 174(4):302–313.

Zentner, G. E., Tsukiyama, T., and Henikoff, S. (2013). ISWI and CHD chromatin remodelers bind promoters but act in gene bodies. *PLoS Genet*, 9(2):e1003317.

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *J. Struct. Biol.*, 193(1):1–12.

Zhang, Y., Smith, C. L., Saha, A., Grill, S. W., Mihardja, S., Smith, S. B., Cairns, B. R., Peterson, C. L., and Bustamante, C. (2006). DNA translocation and loop formation mechanism of chromatin remodeling by SWI/SNF and RSC. *Mol. Cell*, 24(4):559–568.

Zheng, S. Q., Palovcak, E., Armache, J.-P., Verba, K. A., Cheng, Y., and Agard, D. A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. 14(4):331–332.

List of Abbreviations

Å Ångstrom
S. cerevisiae *Saccharomyces cerevisiae*
T. ni *Trichoplusia ni*
X. laevis *Xenopus laevis*

ADP Adenosine diphosphate
ATP Adenosine triphosphate

bp base pair(s)

cryo-EM cryo-electron microscopy
CV Column volume(s)

DESY Deutsches Elektronen-Synchrotron
DNA Deoxyribonucleic acid
dNTP 2'-deoxyribonucleoside triphosphate
DTT Dithiothreitol

EC Elongation complex

HCl Hydrochloric acid
HEPES 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid

IPTG Isopropyl β-D-1- thiogalactopyranoside

LB Lysogeny broth
LIC Ligation independent cloning

MES 2-(N-morpholino)ethanesulfonic acid
mRNA messenger RNA
MWCO Molecular weight cut-off

NCP Nucleosome core particle
NER Nucleotide excision repair
NTP Nucleoside triphosphate

PCR Polymerase chain reaction
Pol Polymerase

SDS-PAGE Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SLS Swiss Light Source

T Temperature
TAE Tris-acetate-EDTA
Tris Tris-(hydroxymethyl)-aminomethane

v/v volume per volume

List of Figures

List of Tables