DISS. ETH NO. 23927

Structural Studies of Crosslinked Nucleosome Core Particles

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

Philip Barth

Dipl. Phys., Universität Konstanz

born on 29.06.1981

citizen of Germany

accepted on the recommendation of

Prof. Rudolf Glockshuber, examiner

Prof. Timothy J. Richmond, co-examiner

Prof. Donald Hilvert, co-examiner

2016

ACKNOWLEDGEMENTS

SUMMARY

The genome of every yeast (*S. cerevisiae*) cell contains about 70,000 individual nucleosomes, each with essentially a different DNA sequence. Although the X-ray crystal structure of the nucleosome core particle (NCP) containing human α-satellite DNA was solved almost 20 years ago, the number of high resolution X-ray crystal structures of NCPs with different DNA sequences is still very limited. The only other two DNA sequences on NCPs that were successfully crystallized and had their structure solved at high resolution are the artificially selected, strong positioning 601 sequence and the MMTV-A promoter sequence. In this work I was able to establish a method for the covalent cross-linking of the nucleosome DNA to the histone octamer core by introducing a non-native thiol group into the DNA via the convertible nucleoside approach. A robust protocol was developed to form a disulfide crosslink between a cysteine introduced into the N-terminal tail of H3 and the thiol of a convertible dG* base introduced into the DNA by oligonucleotide ligation. I was able to show that the crosslinked nucleosomes can be purified on a weak anion exchange column yielding highly pure crosslinked nucleosomes. The X-ray crystal structure of an NCP with α-satellite DNA crosslinked to the histone octamer containing the H3 R40C point mutation was subsequently determined at 2.8 Å.

The established protocol for nucleosome DNA-to-histone crosslinking was combined with another site-specific crosslink that covalently links the two copies of H2A via an introduced N38C point mutation. The structure of this double crosslinked NCP containing both the H3 R40C crosslink to the nucleosome DNA and the H2A N38C crosslink was also determined at 2.8 Å. In previous work, the H2A/H2B inter-dimer crosslink was shown to stabilize the nucleosome against dimer ejection during reconstitution of nucleosome with DNA longer than 147 base pairs (T. Frouws, PhD thesis [1]).

The crosslinking protocol was applied to the DNA sequence of a Met16 promoter nucleosome which was successfully crosslinked and purified according to the established method. The introduced cross-link could be shown to stabilize the Met16 NCP and crystals of Met16 promoter NCP were grown. The crystals diffracted to 3.2 Å on the in-house X-ray generator.

The DNA-to-histone H3 crosslink was shown to stabilize the nucleosome DNA against dissociation upon dilution. This added stability was taken advantage of to improve the purification protocol of the NCP:ΔDot1 complex which, in previous work, led to DNA dissociation and impure samples of the NCP:ΔDot1 crosslinked complex caused by high dilution upon preparative gel electrophoresis purification (V. Vogirala, PhD thesis [2]). A crystallization condition was found that yielded crystal hits, although always accompanied with heavy precipitation which was believed to be caused by DNA impurities (V. Vogirala, PhD thesis [2]). With the added DNA-to-histone crosslink, highly pure

ΔDot1:NCP complex samples devoid of free DNA were obtained. This material did not lead to improved crystal growth of the NCP:ΔDot1 complex.

ZUSAMMENFASSUNG

Das Genom jeder einzelne Hefezelle (*S. cerevisiae*) enthält rund 70,000 individuelle Nukleosomen und damit die gleiche Anzahl verschiedener nukleosomaler DNA-Sequenzen. Obwohl die Kristallstruktur des Nukleosoms (NCP) mit der α-Satelliten DNA vor fast 20 Jahren gelöst wurde, ist die Anzahl der Kristallstrukturen mit von α-Satelliten DNA verschiedenen DNA-Sequenzen bis heute sehr gering. Die einzigen anderen komplett unterschiedlichen DNA-Sequenzen, welche erfolgreich als Nukleosom kristallisiert werden konnten, sind die durch künstliche Selektion gefundene 601-Sequenz und die MMTV-A Promoter-Sequenz. In dieser Arbeit gelang es mir, eine Methode zur kovalenten Verknüpfung der nukleosomalen DNA mit dem Histon H3 mittels Disulfidbrücke zu etablieren, indem ich eine unnatürliche Thiolgruppe mittels konvertierbarem Nukleotid in die DNA einbrachte. Ich erarbeitete ein robustes Protokoll zur kovalenten Verknüpfung von DNA und Histon-Oktamer über eine Cysteinmutante im N-terminalen Bereich von H3 (R40C), welche mit der Thioethylgruppe einer konvertierbaren dG* Base verbunden werden konnte. Diese wird mittels Oligonukleotid an die DNA ligiert. Ich konnte zeigen, dass die kovalent verknüpften Nukleosomen mittels schwacher Anionenaustausch-Chromatographie gereinigt werden können und zu einem sehr hohen Reinheitsgrad der Probe führen. Die Kristallstruktur konnte mittels Röntgenstreuung bis zu einer Auflösung von 2.8 Å bestimmt werden.

Das erarbeitete Protokoll zur Herstellung von kovalent verbundenen Nukleosomen wurde erfolgreich mit einer zweiten spezifischen Disulfidbrücke kombiniert, welche zuvor im Richmond Labor entwickelt wurde (T. Frouws, PhD thesis [1]). Diese Disulfidbrücke verbindet die beiden H2A Monomere über eine Cysteinmutante (N38C). In vorheriger Arbeit konnte gezeigt werden, dass die kovalente H2A-H2A Verbindung rekonstituierte Nukleosomen gegen H2A/H2B-Auswurf schützt, speziell wenn diese mit DNA von mehr als 147 bp Länge rekonstituiert werden. Die Kristallstruktur des Nukleosoms mit beiden kovalenten Verbindungen (DNA zu H3 sowie H2A N38C) konnte ebenfalls mit einer Auflösung von 2.8 Å bestimmt werden.

Das Protokoll zur kovalenten Verknüpfung von DNA und Histon-Oktamer wurde testhalber auf die DNA-Sequenz des Met16 Promoters angewendet. Diese konnte erfolgreich verknüpft werden und die verknüpften Nukleosomen konnten erfolgreich gereinigt werden. Es konnte gezeigt werden, dass die kovalente Verknüpfung die Met16 Nukleosomen stabilisiert. Kristalle des Met16 Promoter-Nukleosoms streuten Röntgenstrahlen in der *in-house* Röntgenquelle bis zu einer Auflösung von 3.2 Å.

Desweiteren konnte nachgewiesen werden, dass die DNA zu H3-Verbindung die Nukleosomen gegen DNA-Dissoziation schützt, wenn diese stark verdünnt werden. Diese zusätzliche Stabilität wurde ausgenutzt um das existierendes Reinigungs-Protokoll für den ΔDot1:NCP Komplex zu verbessern, welches in der bisherigen Version zu einer starken DNA-Dissoziation durch Verdünnung bei der preparativen Gelelektrophorese führte und in der Folge zu einer Verunreinigung durch freie DNA in der ΔDot1:NCP Probe [2]. Eine Kristallisationsbedingung, welche zuvor gefunden wurde, aber immer mit starker Präzipitation einherging (V. Vogirala, PhD thesis [2]), wurde erneut mit reiner, durch die kovalente Verbindung stabilisierte Probe durchgeführt. Das Kristallwachstum in den zuvor gefundenen Bedingungen konnte nicht verbessert werden. Auch mit sehr reiner Probe blieb das Problem der Präzipitation bestehen.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

aa	amino acid
ATP	adenosine 5'- triphosphate
bp	base pair
BSA	bovine serum albumin
βME	2-mercaptoethanol
CV	column volume
Da	dalton (g/mol)
DEAE	diethylaminoethyl
dG	deoxyribo guanidine
dG*	convertible deoxyribo guanidine
DNA	deoxyribonucleic acid
ΔDot1	Dot1 (amino acids 158-582)
Dot1	disruptor of telomeric silencing
ds	double stranded
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
ESI	electrospray ionization
flDot1	full length Dot1
FPLC	fast protein liquid chromatography
g	gram
g	standard gravity (9.81 kg·m/s^2)
GLB	gel loading buffer
h	hour
HPLC	high pressure liquid chromatography
kDa	kilo Dalton (1 kg/mol)
LRS	loss of rDNA silencing
MALDI-TOF	matrix assisted laser desorption ionization – time of flight
MCS	multiple cloning site
mg	milligramme
MGy	megagray
μl	microliter (10^{-6} l)
μm	micrometer (10^{-6} m)
μM	micromolar (10^{-6} mol/l)
min	minute
mm	millimeter (10^{-3} m)
mM	millimolar (10^{-3} mol/l)
mmCIF	macromolecular Crystallographic Identification File
MPD	2-methyl-2,4-pentanediol
MS	mass spectroscopy
MW	molecular weight
MWCO	molecular weight cutoff
NCP	nucleosome core particle
NCP-x	nucleosome core particle with H3:DNA crosslink
NCP-xx	nucleosome core particle with H3:DNA crosslink and H2A N38C crosslink
nm	nanometer (10^{-9} m)
nM	nanomolar (10^{-9} mol/l)
OD	optical density

ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
pdb	protein data bank
PGLB	protein gel loading buffer
RMSD	root mean square deviation
RNA	ribonucleic acid
RNAse	ribonuclease
RP	reverse phase
RPM	revolutions per minute
s	second
SAH	S-adenosyl homocysteine
SDS	sodium dodecyl sulfate
SELEX	systematic evolution of ligands by exponential enrichment
SLS	swiss light source
ss	single stranded
TBE	tris borate EDTA
TEMED	tetramethylethylenediamine
TFA	trifluoroacetic acid
TSS	transcription start site
UV	ultraviolet
V	volts
wt	wild type

GLOSSARY

Nucleosome DNA nomenclature

The two strands making up the α-satellite DNA on the nucleosome core particle of the first high resolution NCP structure were labeled I and J. The 147 bases on each DNA strand are numbered from -73 to 73 with the dyad base pair having the number 0. For the palindromic 147 bp α-satellite sequence, e.g. positions 70.J (base number 70 away from the dyad axis on the J-strand) and 70.I are indistinguishable.

SHL

The superhelical locations (SHL) on the nucleosome DNA with integer numbers are the blocks of base pairs with their major groove opening towards the histone octamer (magenta in the schematic below). They are numbered in increments of full helical turns of the nucleosome DNA SHL 1-7. The block around the dyad base pair is defined as SHL 0. Correspondingly, the other side of the nucleosome DNA (base pairs 1 − 73 on the J-strand and base pairs -1 − -73 on the I-strand) has SHL positions named SHL -1 − SHL -7. The blocks of base pairs with their minor groove opening facing the histone octamer are numbered in half integer numbers (SHL +/-0.5 − SHL +/-6.5) and are colored grey in the schematic below. The dyad base pair is colored green.
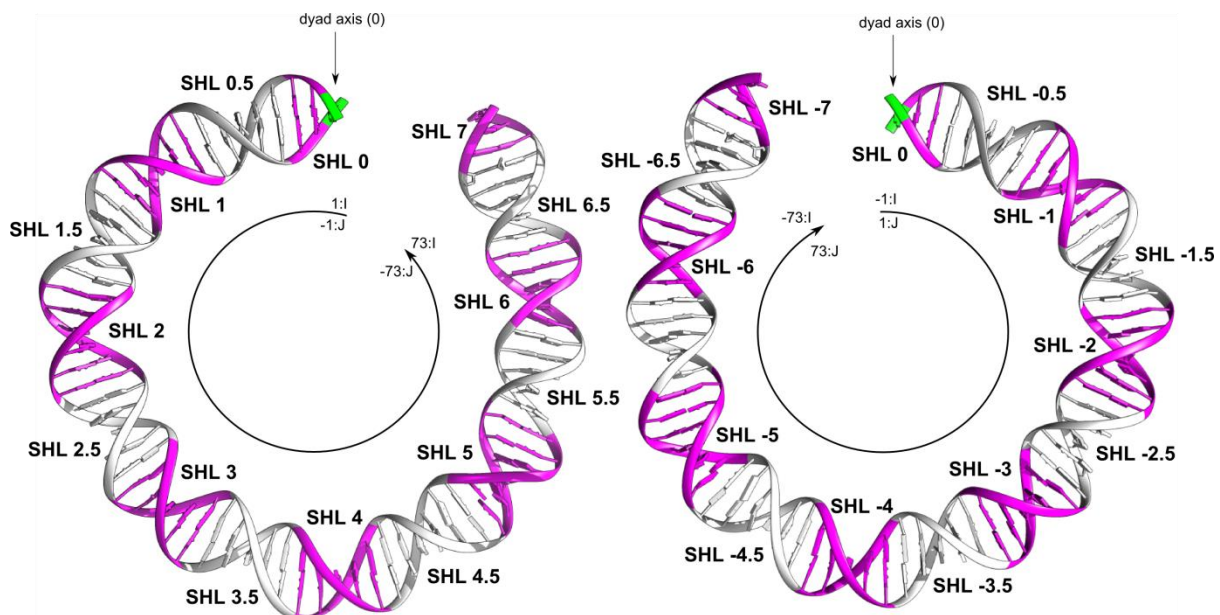
Local reference frame

The local reference frame is calculated by taking the laboratory frame and substracting the contribution from the superhelical pitch. Twist values Ω in the local reference frame Ω_{LRF} are calculated as

$$\Omega_{LRF} = \Omega_{laboratory} - \frac{2\pi sin\alpha}{N}$$

where α is the pitch angle and N is the number of base pair steps in one superhelical turn.

Crosslinked nucleosomes

A colon (:) denotes the formation of a disulfide crosslink between two molecules, e.g. H3 R40C:DNA dG*70.I/J indicates the crosslinking of the cysteine thiol group of the H3 R40C cysteine point mutation to the thiol group of convertible dG (dG*) at positions 70 on both the I-strand and the J-strand.

NCP-x indicates a single crosslinked NCP, containing only a crosslink between the two copies of H3 with one of the DNA strands, whereas NCP-xx indicates the double crosslinked NCP with both, the H3 to DNA crosslinks as well as the H2A N38C crosslink.

1. Introduction

1.1. Chromatin

The DNA of a eukaryotic cell, carrying all the information of the genome, is tightly packed into the cell's nucleus. The genome must serve the contradictory roles of fitting into each cell's nucleus which requires multiple levels of compaction and of making large parts of it accessible to enzymes during various parts of the cell cycle. These requirements are met in a eukaryotic cell by the formation of chromatin. This nucleoprotein complex is organized in a hierarchical fashion with the nucleosome being the most basic building block. The nucleosome, as the fundamental repeating unit of chromatin, was first described by Olins & Olins in 1974 with the 'beads-on-a-string' model for chromatin emerging based on electron micrographs taken from lysed cell nuclei [3]. Even earlier, chromatin digested with nucleases yielded DNA fragments of approximately 200 bp [4, 5]. Extensive micrococcal nuclease treatment digests the linker DNA between nucleosome cores and ultimately leads to DNA fragments no shorter than 146 bp [6]. The term nucleosome for the repeating unit of chromatin was introduced a year later by Oudet *et al.* [7]. Richmond *et al.* solved the X-ray crystal structure of the nucleosome isolated from chromatin with mixed sequence DNA in 1984 to 7 Å [8]. The first X-ray crystal structure of the nucleosome core particle (NCP) at high resolution with defined sequence DNA was published by Luger *et al.* in 1997 and showed much of the molecular detail at 2.8 Å resolution [9]. A structure with improved resolution (1.94 Å) was published in 2002 which allowed for the description of the many water-mediated interactions within the nucleosome and a detailed analysis of the structure of the nucleosome DNA [10, 11]. The arrangement of nucleosomes upon further compaction remains elusive and probably depends on the length of linker DNA between nucleosome cores. Several models of higher order structures forming what is known as the 30 nm fiber were proposed [12, 13]. A two-start helical conformation seems likely, corroborated by the X-ray crystal structure of a tetranucleosome [14, 15]. Fig. 1.1 shows the hierarchical levels of the chromatin structure leading from the chromosomal DNA being wrapped around the histone octamer, forming the nucleosome, up to the mitotic chromosome. The X-shaped chromosomes which are separated into two dividing daughter cells during mitosis are only one form of chromatin, representing the densest packing, whereas during most of the cell cycle chromatin exists as euchromatin.

1.1.1. Multiple forms of chromatin

Euchromatin represents the open form of chromatin and in this state most of the DNA is accessible to the transcriptional machinery of the cell and many of the genes in euchromatin are actively transcribed. The percentage of actively transcribed genes differs with the phase of the cell cycle, but also

largely differs between different species. Euchromatin is characterized by a high degree of acetylation (hyperacetylation) and generally low levels of methylation (hypomethylation) both on the histone tails and the DNA (reviewed in [16]). Several subsets of euchromatin can be distinguished based on the state of modification and form (reviewed in [17]).
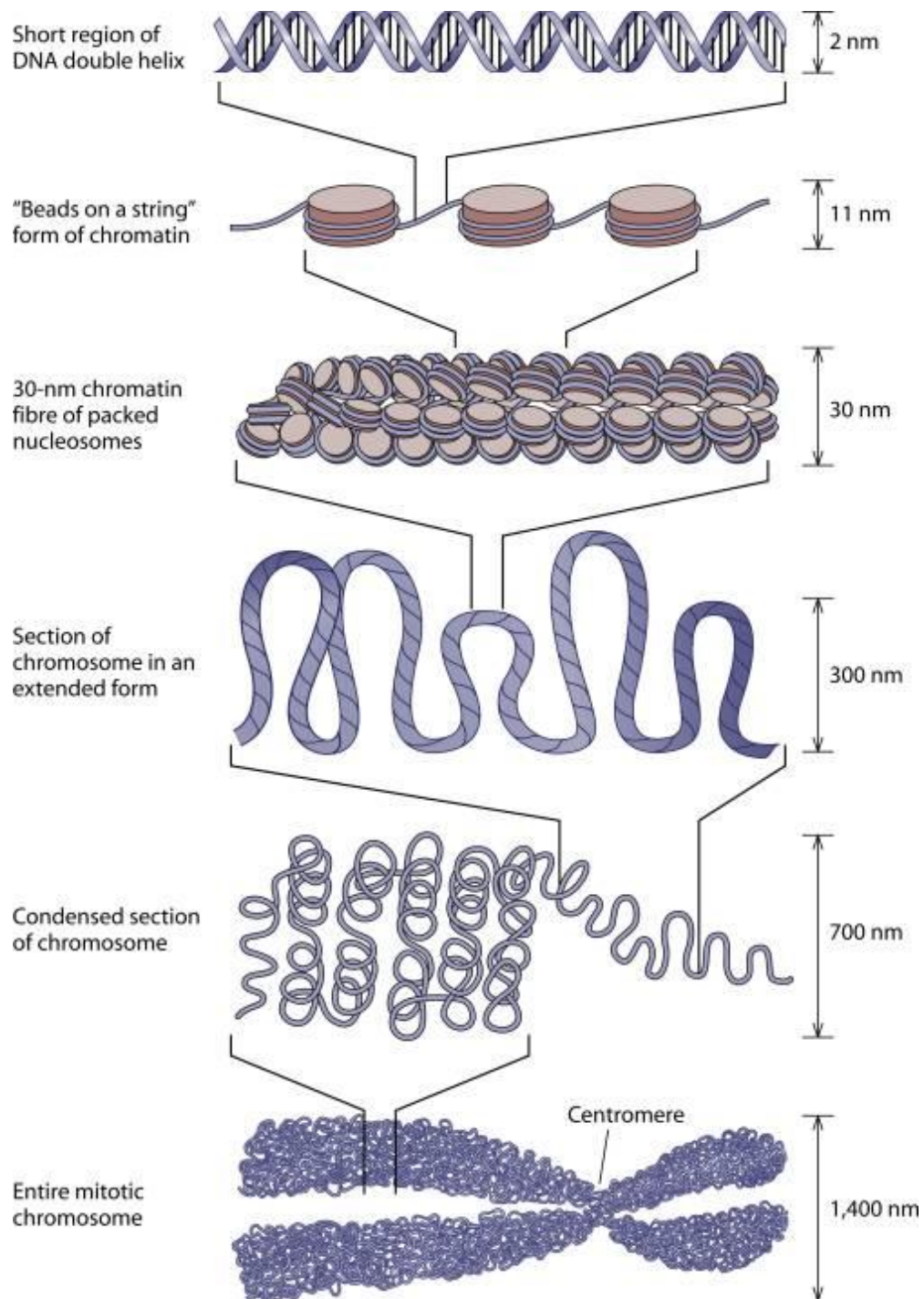
Figure 1.1: The hierarchical compaction of chromatin. It starts at the level of free DNA (top) with a diameter of 2 nm and leads up to the full mitotic chromosome (bottom) with a diameter of more than 1 mm. The diameter of the substructure is shown on the right. The 30 nm fiber is shown as a single start helical solenoid. Image taken from **[18]** with permission.

Heterochromatin represents the compacted, mostly inaccessible state of chromatin. Several characteristic locations on chromosomal DNA form constitutive heterochromatin. The centromers where the two chromatids of a chromosome are connected by a cohesion ring complex remain in a condensed form throughout the cell cycle [19]. At the ends of chromosomes telomeric DNA repeats also form constitutive heterochromatin [20]. A facultative form of heterochromatin can be realized in cell development e.g. with the inactivation and compaction of one X-chromosome in female cells. The genes of the inactivated chromosome show low levels of trancription and generally high levels of methylation, both of the DNA and histones H3 and H4 along with low levels of histone acetylation [16]. Heterochromatin can be characterized by its much more compact form and the accompanying inaccessability of the DNA. It is therefore generally transcriptionally inactive throughout the cell cycle.

1.2. The nucleosome core particle (NCP)

Since the first X-ray crystal structure of the NCP was solved at high resolution in 1997, several high resolution X-ray crystal structures of NCPs were published. Variations in NCP content fall mainly in two classes. Firstly, there are now a few structures of NCPs containing DNA sequences other than α-satellite DNA. Secondly, there are X-ray structures of NCPs assembled from histone octamers origininating from different species or containing histone variants or modifications. A third class of X-ray crystal structures elucidates the interactions of NCPs with protein factors or small molecule cofactors. An overview of the X-ray structures published to date is given in Table 1.1.

Variant	Origin species histone octamer	DNA sequence (bp)	Reference
NCPs with different histone variants			
Major histones	*X. laevis*	α-satellite (146)	[9]
Major histones	*G. gallus*	α-satellite (146)	[21]
Major histones	*S. cerevisiae*	α-satellite (146)	[22]
Major histones	*H. sapiens*	α-satellite (146)	[23]
Major histones	*D. melanogaster*	α-satellite (147)	[24]
H2A.Z	*X.laevis/M. musculus*	α-satellite (146)	[25]
Macro H2A	*M. musculus/H. sapiens*	α-satellite (146)	[26]
H3T	*H. sapiens*	α-satellite (146)	[27]
H3Y	*H. sapiens*	α-satellite (146)	[28]
CENP-A	*H. sapiens*	α-satellite (147)	[29]
NCPs with different DNA sequences			
Sequence variant	*X. laevis*	α-satellite (146)	[10]
DNA length	*X. laevis*	α-satellite (147)	[10]
DNA length	*X. laevis*	α-satellite (145)	[30]
A16	*X. laevis*	α-satellite (147)	[31]
TTTAA	*X. laevis*	α-satellite (147)	[32]
601	*X. laevis*	601 (145)	[33]
601 L	*X. laevis*	601L (145)	[34]
MMTV-A	*X. laevis*	MMTV-A (147)	[35]
NCPs in complex with protein factors			
LANA peptide	*X. laevis*	α-satellite (147)	[36]
RCC1	*X. laevis*	601 (145)	[37]
Sir3-BAH	*X. laevis*	601 (147)	[38]
CENP-C	*D. melanogaster*	α-satellite (147)	[39]
Set8	*X. laevis*	601 (149)	[40]
SAGA-DUB	*X. laevis*	601 (145)	[41]
Chromatosome	*D. melanogaster*	601 (167)	[42]

Table 1.1: Solved crystal structures containing NCP with histone variants, different DNA sequences or crystal structures of NCPs with bound protein factors.

The overall architecture of the NCP is conserved over the range of X-ray crystal structures solved so far. It consists of a histone octamer core, containing two copies of each of the core histones (H2A, H2B, H3 and H4) with 145 – 147 base pairs of DNA wrapped around the disk-shaped octamer core. The nucleosome DNA forms a left-handed superhelix making 1.67 turns.

1.2.1. The histone octamer core

The structure of the histone octamer core was first solved to high resolution in 1991 and for the first time showed the molecular architecture comprising four heterodimers (two of each H2A with H2B and H3 with H4) which are arranged in a disk-shaped quaternary structure with a 65 Å diameter and a height of between 10 – 60 Å [43]. Each individual histone is characterized by a helix-loop-helix motif with three helices, of which the first and last (α1 & α3) are short and the central helix (α2) is elongated. This arrangement was named the histone-fold. The histones in a heterodimer form a handshake motif where the central α2-helices cross and the terminal parts of the α2-helix contact either the α1 or the α3-helix (Fig. 1.2). The octamer core is arranged such that it forms a left-handed helix with three four-helix bundles stabilizing the histone octamer fold. The first four-helix bundle between H2B and H4 is followed by the central four-helix bundle between H3 and H3'. The third four-helix bundle is formed between H2B' and H4' (Fig. 1.2 D,E). Along the super-helical trace on the surface of the histone octamer a series of DNA binding sites are formed by the termini of the two α1-helices of each heterodimer (α1α1) and its two flanking sites formed by loops L1 and L2 (L1L2) which are exposed to the DNA binding surface (Fig. 1.1 A). A series of 14 arginine amino acids spaced out along the DNA binding surface insert their guanidinium moieties into the narrowed minor grooves of the nucleosome DNA where the electrostatic potential is strongest [44] (Fig. 1.1 B).
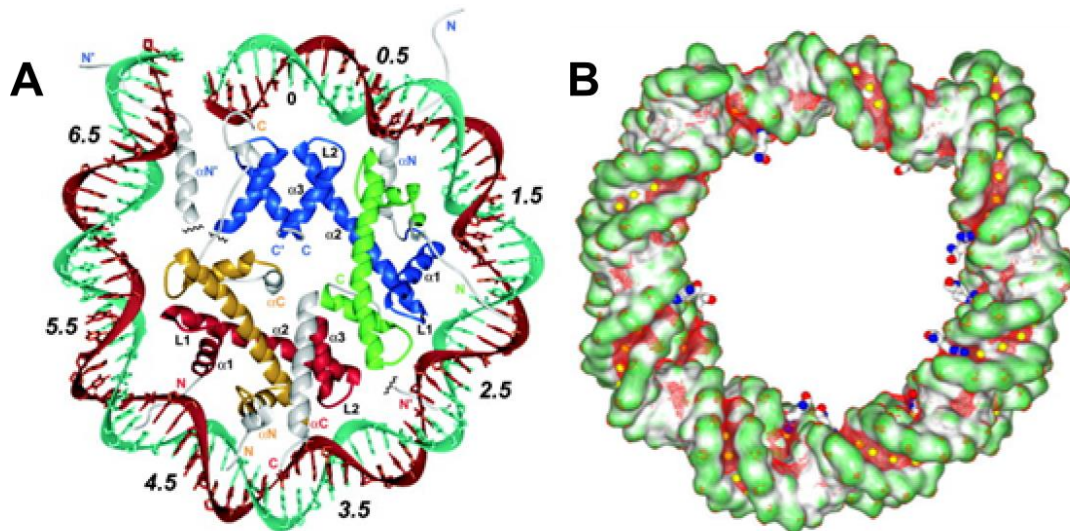
Figure 1.1: Histone binding to the nucleosome DNA. **A)** Histone binding sites on the nucleosome DNA formed by the α1α1 motifs and their flanking L1L2 motifs shown for one half of the NCP. The last turn of DNA, before it exits the nucleosome, is essentially straight and only bound by the H3 αN-helix and the H3 N-terminal tail tunneling through the minor groove channel between SHL -1/7. DNA is colored green and brown, H2A (yellow), H2B (red), H3 (blue) and H4 (green) as before with the H3 αN-helix and the H2B αC-helix colored in white **[10]**. **B)** Surface representation of the nucleosome DNA with the coloring according to the shape of the DNA surface. Grey indicates concave surfaces and green indicates convex surfaces. Arginine amino acids which insert into the minor grooves are shown in stick representation. Yellow dots are reference points where the electrostatic potential was calculated. Red mesh represents the electrostatic isosurface at -6 k$_B$T/e (from **[44]**).
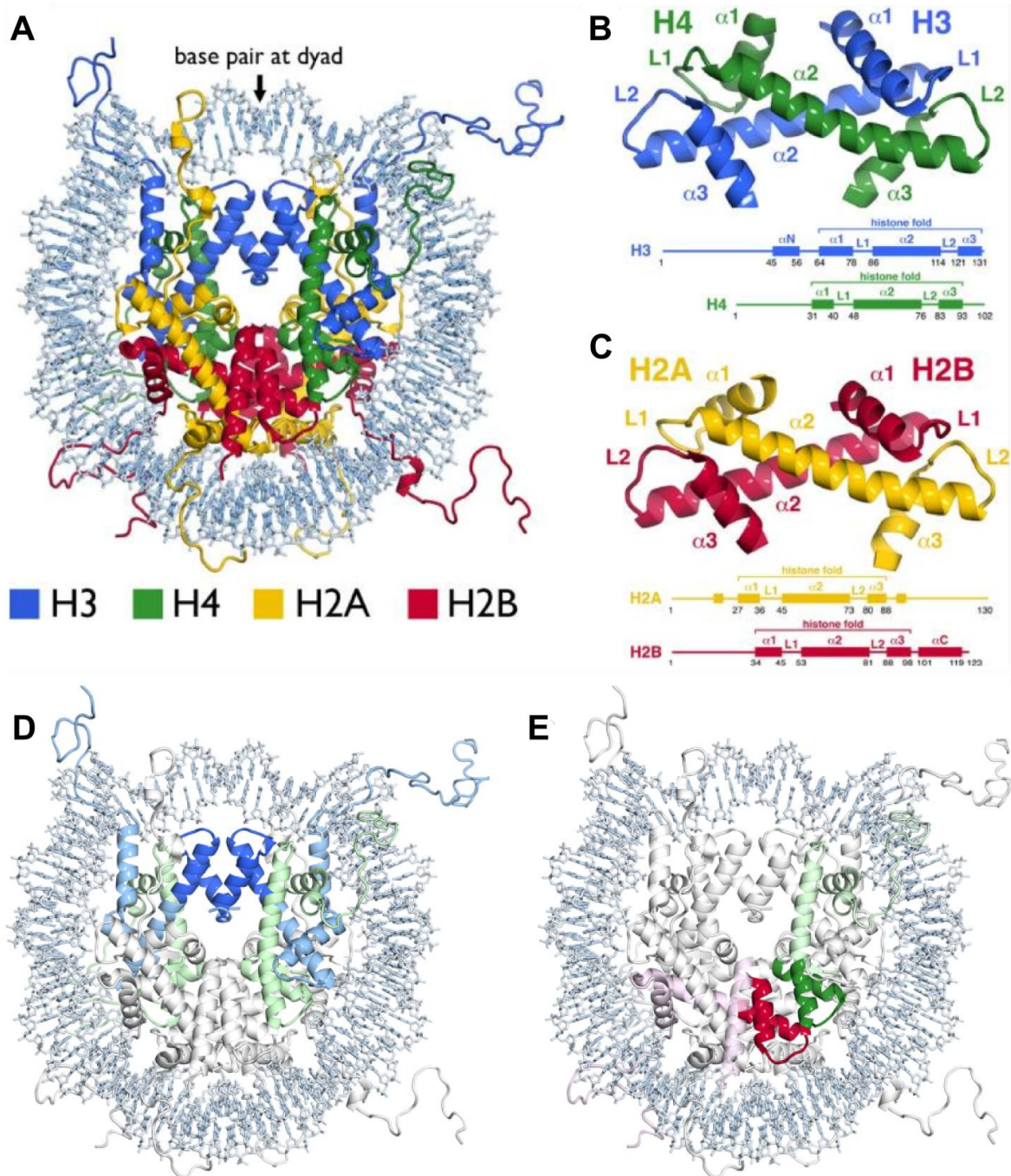
Figure 1.2: Structure of the histone octamer core. **A)** The overall structure shows the histone octamer core (colored) in the context of the NCP (DNA in gray). The top view shows the round shape which serves as a binding platform for DNA. The histone octamer core consists of four heterodimers which are made of H2A & H2B (**C**) or H3 & H4 (**B**) each forming a hand-shake motif. **D)** The two H3/H4 heterodimers form a tetramer via a four-helix bundle between the two copies of H3 (dark blue). **E)** Two more four-helix bundles are formed between H2B and H4. The mostly unstructured histone tails extrude from the core particle (adapted from **[45]** with permission).

All core histones feature an extended unstructured N-terminal tail but only H2A has also a long C-terminal tail. α-Helical regions that extend beyond the histone-fold motifs are found in H3, which contains an additional αN-helix, and in H2B, which contains an αC-helix. The surfaces of the histone octamer not bound by DNA form various motifs with specific distributions in shape and charge. An

acidic patch formed by amino acids E56, E61, E64, D90, E91 and E92 of H2A and E102 and E110 of H2B creates a negatively charged binding pocket (Fig. 1.3) on the mostly positively charged octamer surface.
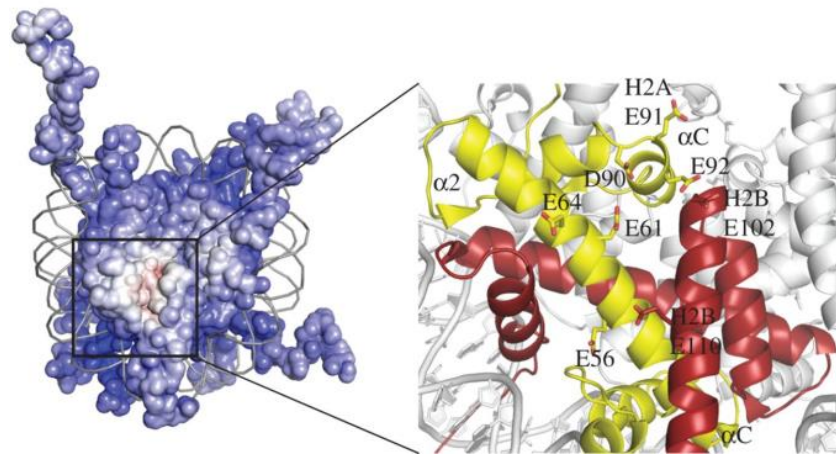
Figure 1.3: The acidic patch on the surface of the histone octamer disk. The histone octamer is shown in surface representation and the coloring is according to charge with red being negative and blue being positively charged. The DNA backbone is shown as grey lines. The right side shows a blow-up of the acidic patch region with the histones shown in ribbon representation. H2A is shown in yellow and H2B in red. The amino acids forming the acidic patch are additionally shown as sticks and are labeled (adapted from **[46]** with permission).

Several co-crystal structures of protein factors bound to NCPs show specific recognition of the histone octamer acidic patch with several arginine or lysine amino acids being inserted into the acidic patch [36, 47, 38]. This leads to the emerging picture of the acidic patch as a common recognition motif for nucleosome binding factors (reviewed in [46, 45]).
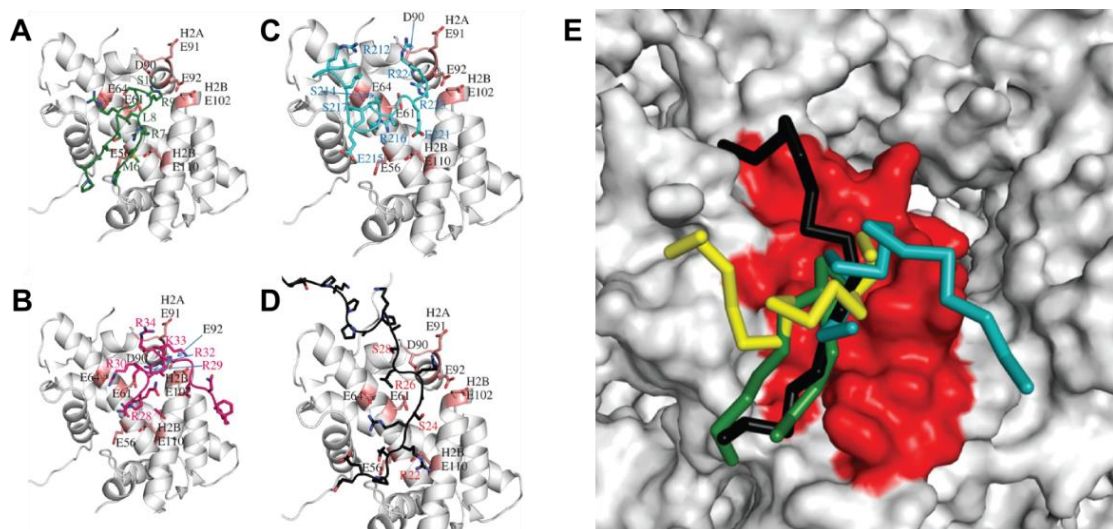
Figure 1.4: Histone octamer acidic patch recognition found in multiple NCP co-crystal structures. **A-D)** The binding of several NCP binding factors to the acidic patch of the NCP (shown in grey ribbons, acidic amino acids shaded in red). The acidic patch binding domains are shown in colored sticks (**A**: LANA peptide, green; **B**: Sir3 BAH domain, magenta; **C**: RCC1, cyan; **D**: HMGN2, black). The amino acids forming the factor:acidic patch interactions are labeled. **E)** Overlay image of the interacting regions on top of the acidic patch (red) on the octamer surface (grey). The color coding for the binding factor domains is as in **A-D**.

In the crystal structures of all NCPs, one interaction between the densely packed NCPs is formed between the H4 tail of one NCP with the acidic patch of an adjacent NCP, suggestive of an *in vivo* interaction between nucleosomes forming the 30 nm fiber, although this hypothesis is not substantiated by the crystal structure of the tetranucleosome in which the interaction between stacked nucleosomes is made between the αC-helices of H2B (Fig. 1.7) [14].
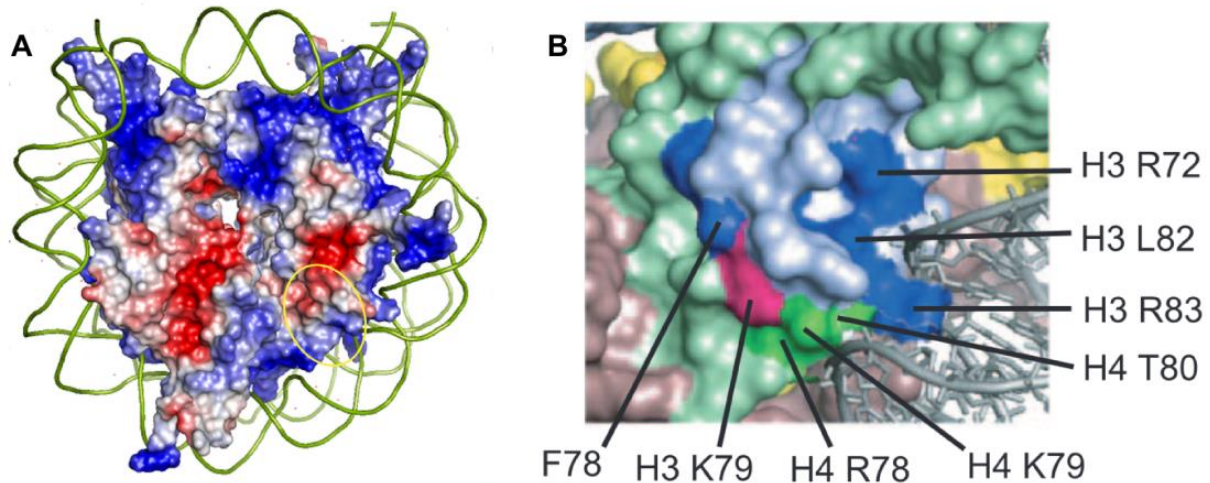
Figure 1.5: The LRS surface on the face of the histone octamer. **A)** Surface representation of the histone octamer lateral surface colored according to surface charge. Blue indicates positive charge, red indicates negative charge and white represents neutral surfaces. The DNA backbone is shown as green lines. The LRS surface, harboring the H3 K79 amino acid, is encircled in yellow **[48]**. **B)** Zoom-in on the LRS surface. Several amino acids comprising the LRS surface are labeled (image taken from **[49]** with permission).

Another binding surface that was identified on the face of the histone octamer is the LRS surface (Loss of rDNA silencing), which is located near the DNA binding site at SHL +/-2.5 (Fig. 1.5). It consists of several basic amino acids and is required for transcriptional silencing [49]. It also harbors amino acid H3 K79, which is the methylation target of Dot1, a key regulator of the disruption of silencing (sections 1.8 and 1.9).

1.2.2. The crystal structure of the nucleosome core particle

In 1997, the X-ray crystal structure of the NCP was solved to high resolution for the first time by Luger *et al*. [9]. It shows the histone octamer core with 146 bp of α-satellite DNA wrapped around it, basically unchanged compared to the free histone octamer structure [43]. Additional stretches of the histone tails running through adjacent DNA gyres of nucleosome DNA are resolved. The NCP with 146 bp of palindromic α-satellite DNA was designed to minimize two-fold disorder by making it pseudo two-fold symmetric. However, the X-ray crystal structure revealed a base pair located on the particles pseudo two-fold axis which makes the NCP asymmetric with respect to its two halves. One half contains 72 bp of DNA and the other half contains 73 bp. The numbering of base pairs is count-

ed from the base pair on the dyad axis, which is defined as 0. The positions on the DNA superhelix where the major groove faces the histone octamer were named super helical locations SHL +/-1 to SHL +/-7 depending on the number of full turns of the DNA helix away from the dyad. The superhelical positions where the minor groove faces inwards carry corresponding half-integer numbers. In the high resolution NCP structure, the DNA side with 72 base pairs shows a region around SHL 1.5 where the DNA is stretched by overtwisting to compensate for the missing base pair. This DNA stretching behavior is further described in section 1.3. The nucleosome DNA forming the left-handed superhelix around the histone octamer core has an average radius of 41.8 Å, although the bending is not uniform along the superhelix. The regions around SHL +/-1.5 and SHL +/-4.5 show maximum curvature, whereas the terminal 10 – 12 base pairs are essentially straight [9]. Along the binding surface of the DNA on the histone octamer, multiple contacts between amino acids and the DNA phosphate backbone are made. All of the interactions are sequence non-specific underscoring the ability of the histone octamer to bind a vast variety of different DNA sequences. The sequence of α-satellite DNA shows a regular spacing of G·C rich and A·T rich stetches. Blocks of 5 – 6 base pairs where the minor groove faces the histone octamer tend to be A·T rich, whereas the neighboring blocks of 5 – 6 base pairs, where the major groove faces inward, are found to be rich in G·C base pairs. This pattern, which is observed throughout eukaryotic genomes, can be explained by the anisotropic bendability of the respective base pairs towards the minor groove. A·T base pairs and base pair steps are more easily bent into the minor groove, thereby narrowing it, especially TA steps, but also CA steps [50]. Narrow minor grooves with their increased negative electrostatic potential increase the binding energy of the inserted arginine amino acids at those sites [44].

In 2002, the structure of the NCP was solved to 1.94 Å resolution. This structure features a wild type *X.laevis* histone octamer and 147 base pairs of palindromic α-satellite DNA [10]. Again, a base pair is located on the dyad position, but this structure posseses molecular two-fold symmetry with respect to the dyad and does not show DNA stretching. This 147 bp state of the NCP can be considered to be saturated with DNA and is depicted in Fig. 1.6.
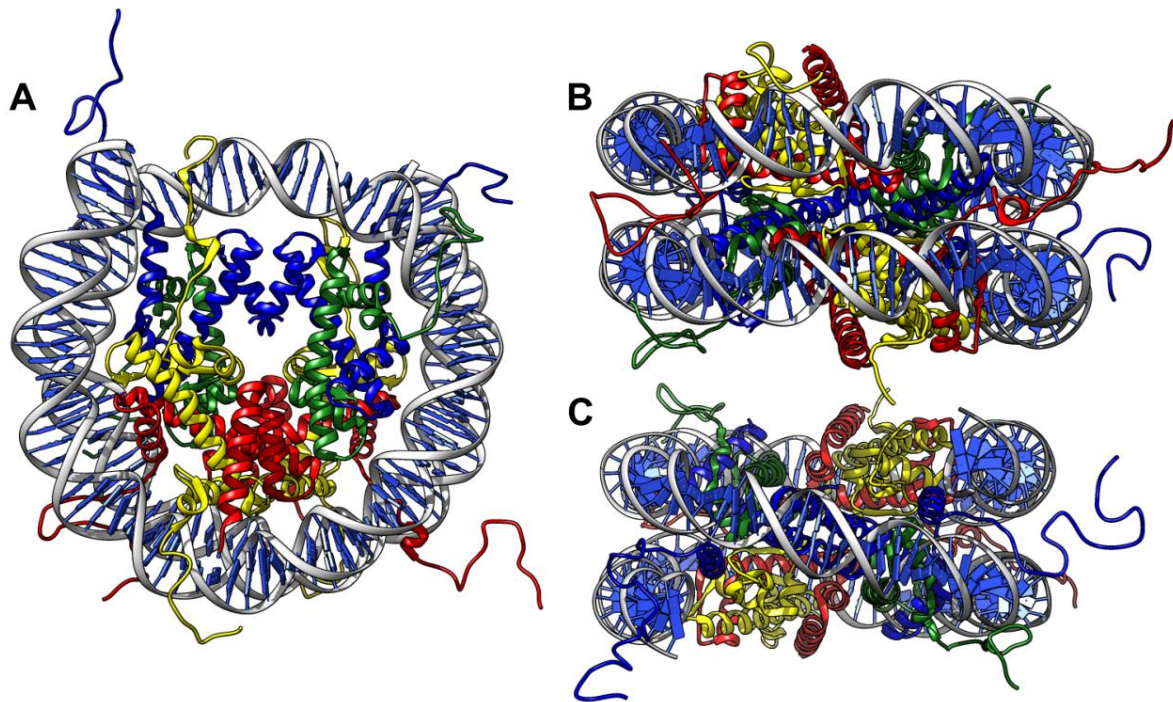
Figure 1.6: X-ray crystal structure of the NCP solved to 1.94 Å **[10]** (PDB ID: 1KX5) shown in three views. **A)** NCP from a top view. **B)** Side view facing the side opposite of the dyad. **C)** Side view facing the dyad axis. Histones are shown in yellow (H2A), red (H2B), dark blue (H3) and green (H4). DNA is shown in grey with the nucleobases colored blue.

1.2.3. Histone tails

All four core histones have considerable amino-terminal tail regions. H2A has also an extended C-terminal tail. The tail domains amount to approximately 25% of the total mass of the histones, but posttranslational modifications of amino acids occur almost exclusively on the histone tails due to the general inaccessability of the octamer core to modifying enzymes [51]. The histone tails, which protrude from the nucleosome surface, can be easily accessed by these enzymes. All of the H3 and H2A N-terminal tails protrude through narrow channels formed by two aligned minor grooves from adjacent gyres of the DNA superhelix. Most of the other tails leave the nucleosome disk following DNA minor grooves. The histone tails which extrude from the NCP are probably unstructured *in vivo* although some propensity to adopt α-helical shape has been suggested (e.g. [52]). In the X-ray crystal structure of the NCP, the histone tails are mostly not resolved unless they are involved in crystal contacts or protrude through aligned minor groove channels between DNA gyres.

The acetylation of amino acids in the histone tails was first discovered to contain a pattern encoding the state of chromatin [53]. Related patterns for other histone tail modifications, such as methylation, phosphorylation or ubiquitination have been found. The distribution of modified amino acids on the histone tails is widely believed to form a code (the histone-code) which is important for the regulation of chromatin (see section 1.7) [54].

1.2.4. H1 and the higher order structure of chromatin

The nucleosome represents the basic building block of chromatin. The X-ray structures of the NCPs with various DNA sequences or histone octamer variants have been determined. The inter-nucleosome linker DNA and the linker histone H1, however, remained unresolved until recently when the X-ray crystal structure of the chromatosome was published detailing the interaction of the linker histone with the NCP and the linker DNA, albeit at relatively low resolution and with averaged electron density due to two-fold disorder [42].

A low-resolution X-ray crystal structure of a tetranucleosome (9 Å) was published, showing the arrangement of four individual nucleosomes with a 167 bp repeat length [14]. The structure represents a truncated version of a two-start helix for the 30 nm fiber and is not compatible with a solenoid model of a one-start helix. The stacked nucleosomes within the tetranucleosome interact via their H2B αC-helices. An interaction between the H4 tail and the acidic patch of an adjacent nucleosome, as seen in the crystal packing of NCP crystals, is not observed.
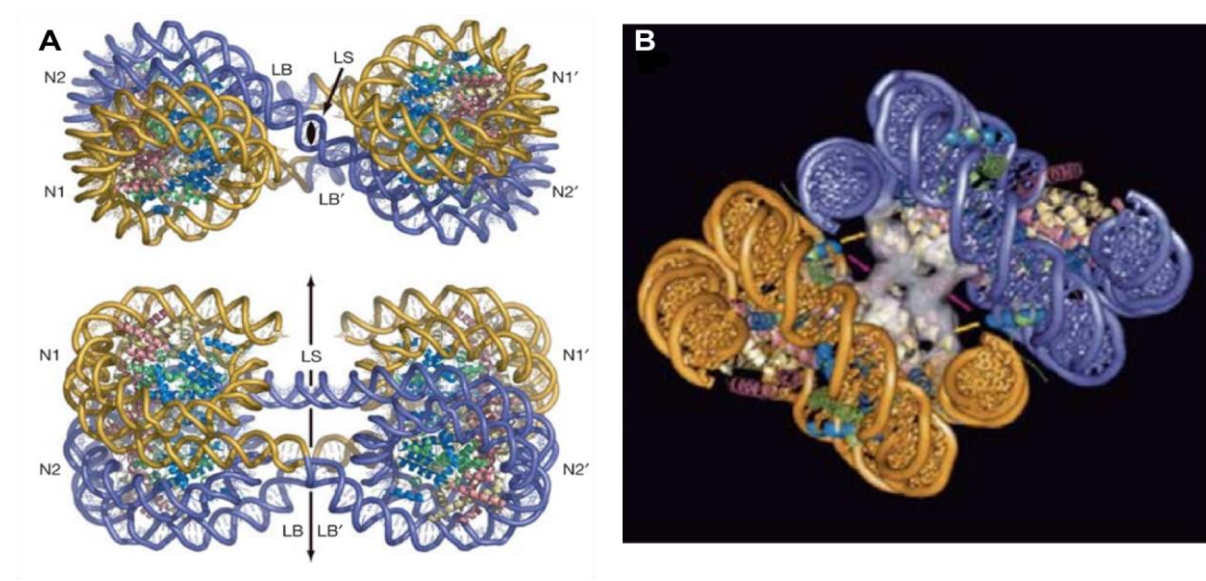
Figure 1.7: Structural model based on the 9 Å X-ray crystal structure of a tetranucleosome with 167 bp repeat length. **A)** Two views of the model rotated by 90°. The general packing of individual nucleosomes in the tetranucleosome and the trajectories of the linker DNA are illustrated. The linker DNA between N2 and N2' is straight (LS), whereas the linkers between N1 and N1' and between N2 and N1' are bent (LB). The two-fold axis is indicated by a black arrow. **B)** Model build from the electron density showing the inter-nucleosome interactions in the tetranucleosome which occur mainly between the H2B αC-helices (red) (Images taken from **[14]** with permission).

1.3. Structure of DNA on the nucleosome core particle

The standard number of base pairs in a full turn of free B-DNA (helical pitch) is 10.5 bp/turn. In contrast, DNA in bulk chicken chromatin has an average helical periodicity of 10.17 +/- 0.05 bp/turn, a value that is reproduced in hydroxyl footprinting experiments [55, 56]. The structure of the NCP containing recombinant *X.laevis* histones and 146 bp of α-satellite DNA clearly show the DNA to be stretched in the shorter half of the NCP around SHL +2, whereas the NCP structure containing 147 bp of α-satellite DNA shows no stretching [9]. The 146 bp NCP structures with stretched DNA show helical periodicities of 10.23 and 10.15 bp/turn, similar to the value found in bulk chromatin, indicating that the DNA on nucleosomes *in vivo* is in a stretched state [11]. An NCP structure containing wild type *X.laevis* histone octamer and 145 bp of α-satellite DNA shows two stretching sites at SHL +/-2 and a helical twist value of 10.13 bp/turn in the local reference frame (LRF, see glossary for details) [30]. The X-ray crystal structure of the NCP containing the 145 bp strong positioning 601 sequence shows two DNA stretching sites located at SHL +/-5 and no stretching at SHL +/-2 and a helical twist value of 10.09 bp/turn [33].

DNA identity	# base pairs	Helical twist[LRF]	Reference/PDB identifier
Free B-DNA	N/A	10.50	[56]
α-satellite	147	10.30	[10]/1KX5
α-satellite	146	10.15	[10]/1KX3
α-satellite	146 (b)	10.23	[10]/1KX4
α-satellite	145	10.13	[30]/2NZD
601	145	10.09	[57]/3LZ0

Table 1.2: Overview of the double-stranded helical twist values for the DNA in several NCP X-ray crystal structures containing DNA of various length. The value for free DNA is given as a reference in the top row. The values for the helical periodicities of the nucleosome DNA are given in the local reference frame (LRF) of the nucleosome superhelix.

Table 1.2 summarizes the results of the previously solved set of NCP structures containing various DNA lengths with respect to DNA stretching. Since the nucleosomes in an NCP crystal form an important crystal contact at the DNA termini creating a pseudo-continuous DNA molecule, the question arises whether the stretching seen in the nucleosome DNA stems from crystal packing requirements. The X-ray crystal structure of an NCP with a protein factor (RCC1) bound contains 147 bp 601 DNA in a crystal packing arrangement which does not form end-to-end stacking of the DNA termini [47]. It still shows stretching at SHL +/-5 which is indicative of the stretched state not being an artifact of the crystal packing but further corroborates the model of the naturally stretched state of DNA in a nucleosome context.

1.4. Nucleosome positioning

All NCP crystal structures, including the recently determined X-ray crystal structure of the MMTV-A NCP, contain strong positioning sequences. These are either the α-satellite centromeric positioning sequence or the artificial 601 positioning sequence which was derived from SELEX experiments. The haploid human genome with its approximately 3.2×10^9 base pairs and an average nucleosome repeat length (NRL) of close to 200 base pairs contains around 16,000,000 nucleosomes [58]. Another well-studied genome is the *Saccharomyces cerevisiae* (yeast) genome. The haploid yeast genome comprises approximately 12.1×10^6 base pairs with an average nucleosome repeat length of 167 base pairs resulting in over 70,000 nucleosomes [58]. In 2012, Brogaard *et al*. published a base-pair-resolution map of the nucleosome dyad positions of the yeast genome which was based on a site-directed hydroxyl radical footprinting method developed earlier [59, 60]. The nucleosome position map shows a linker length distribution strongly favoring multiples of 10n+5 base pairs. This map comprises the dyad positions of every nucleosome within the yeast genome, although it does not contain information on the stretching states of the DNA of each nucleosome. The number of possible conformations the DNA can take on (i.e. stretched or not stretched) for a single mapped dyad position increases to four when one stretched base pair is permitted per NCP half. With multiple stretching sites on a single NCP half, the number of possible conformations increases further. To allow for the crystallization into well-diffracting NCP crystals, the same packing as in the high resolution α-satellite NCP structure (PDB ID: 1KX5) is probably necessary, including the DNA end-to-end stacking of adjacent NCPs [9]. Another feature, which was shown in the nucleosome dyad mapping experiment, was the clustering of most dyad positions spaced by integral multiples of the DNA helical repeat of approximately 10 base pairs, representing different translational positions of the nucleosome DNA. Multiple, almost equally stable rotational positions lead to an inhomogenous translational distribution of DNA on the histone octamer. In the case of 147 base pairs of α-satellite DNA, the DNA has multiple translational positions but can be shifted into a single translational orientation by heating the NCPs at 37 – 50 °C [61].

There is still no consensus as to which factors contribute to which extent to the positioning of nucleosomes *in vivo*. Several factors have been proposed, including sequence elements such as poly(dA:dT) tracts, as major determinants of nucleosome organization [62]. The *in vivo* distribution of yeast nucleosomes can be recreated by reconstituting purified yeast genomic DNA on chicken octamer [63]. This is an indication of the dominant role the DNA sequence has in the formation of nucleosomes. An extensive body of work exists looking at nucleosome positioning from a computational perspective. The shape of DNA was suggested to dominate the sequence affinity in nucleo-

some formation [64] and models were developed to recreate the observed strong positioning of the 601, 603 and 605 clones selected for high positioning power based on computer algorithms [65, 66]. More ambitious computer models tried to recreate the entire *in vivo* distribution of micrococcal nuclease digestion maps of nucleosome positions to create a universal nucleosome code [67]. Other researchers argue for a more important role of the transcriptional machinery and nucleosome re-modeling factors on nucleosome positioning (summarized in [68]).

1.5. Nucleosome stability

NCPs are destabilized upon an increase in salt concentration and/or temperature as well as a decrease in particle concentration. Due to the many electrostatic interactions between the histone octamer and the nucleosome DNA, higher salt concentrations will shield those interactions and hence weaken the interaction leading to DNA dissociation. For purified nucleosomes from chicken erythrocytes, the salt induced dissociation of DNA reaches its half point around 500 mM monovalent salt concentration with the NCP concentration in the μM range [69]. Strong positioning DNA sequences possess a higher salt dissociation half-point. The DNA sequence with the highest affinity for the histone octamer (the 601L palindromic sequence) has a salt dissociation half-point of 1.585 M NaCl [34]. Apart from the DNA sequence, the identity of the histones in the histone octamer has an effect on the stability against salt-induced dissociation. It was shown that the NCP with a histone octamer containing H2A.Z (a variant of the major H2A histone) increases the salt-induced dissociation midpoint from 440 mM NaCl (for major histones) to 506 mM NaCl (for the H2A.Z octamer), implying that the histone – histone interactions within the histone octamer play a role in nucleosome dissociation [70]. The nucleosome concentration is another important parameter governing DNA dissociation from the NCP. Below concentrations of 1 μM NCP, the fraction of free DNA after a short incubation time increases dramatically [69]. Even for the strongly positioning 601 DNA, it was found that samples diluted to 1 – 10 nM only retained 20% – 30% intact nucleosomes (at 200 mM NaCl and 23 °C) [71].

The pathway of NCP disassembly can take one of several routes, the most likely one being the dissociation of H2A/H2B dimer(s) and subsequent unwrapping of the DNA. The last step consists of the dissociation of the $(H3/H4)_2$ tetramer from the DNA. The reverse path is taken upon NCP assembly. Fig. 1.8 shows the 'butterfly model' of NCP assembly [72]. From the initial condition of separate DNA, H2A/H2B heterodimer and $(H3/H4)_2$ tetramer (state VI), several paths to assembled NCPs can be taken. A transition from state VI to state III is unlikely, since the histone octamer is not soluble at low (physiological) salt concentrations. The most likely path is the initial binding of the $(H3/H4)_2$ tetramer to the DNA and subsequent binding of the two H2A/H2B heterodimers either to the DNA

(state IV) or to the $(H3/H4)_2$ tetramer (state II), followed by final assembly into the folded nucleosome (state I).
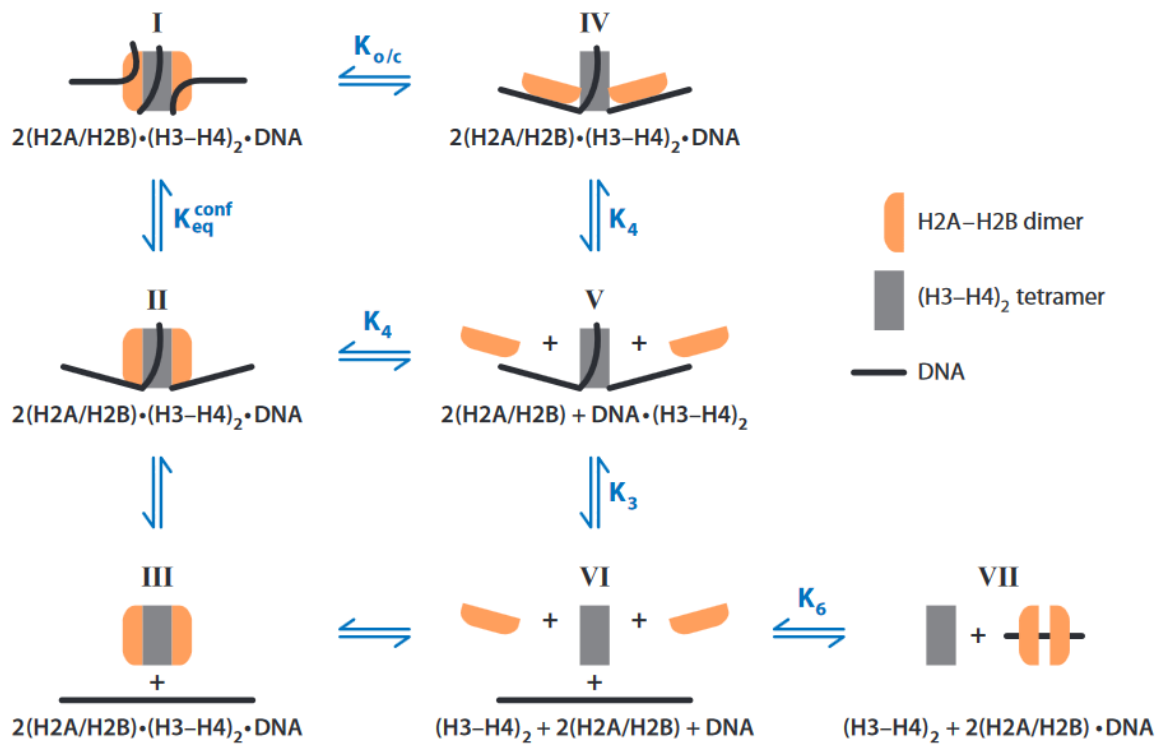
Figure 1.8: 'Butterfly model' of NCP (dis)assembly. All possible pathways start at position VI and the most likely pathway is from VI − V − (IV or II) − I **[72]**.

Upon disassembly of the nucleosome, the reverse pathway is likely to take place with the initial dissociation of H2A/H2B heterodimer, followed by the unwrapping of DNA and subsequent dissociation of $(H3/H4)_2$ tetramer from the DNA.

1.6. Nucleosome interactions with protein factors

Ever since the X-ray crystal structure of the NCP was solved in 1997, a series of discoveries has furthered our knowledge about the structural basis of chromatin. Fig. 1.9 shows the timeline of a selection of important events in chromatin structural biology.
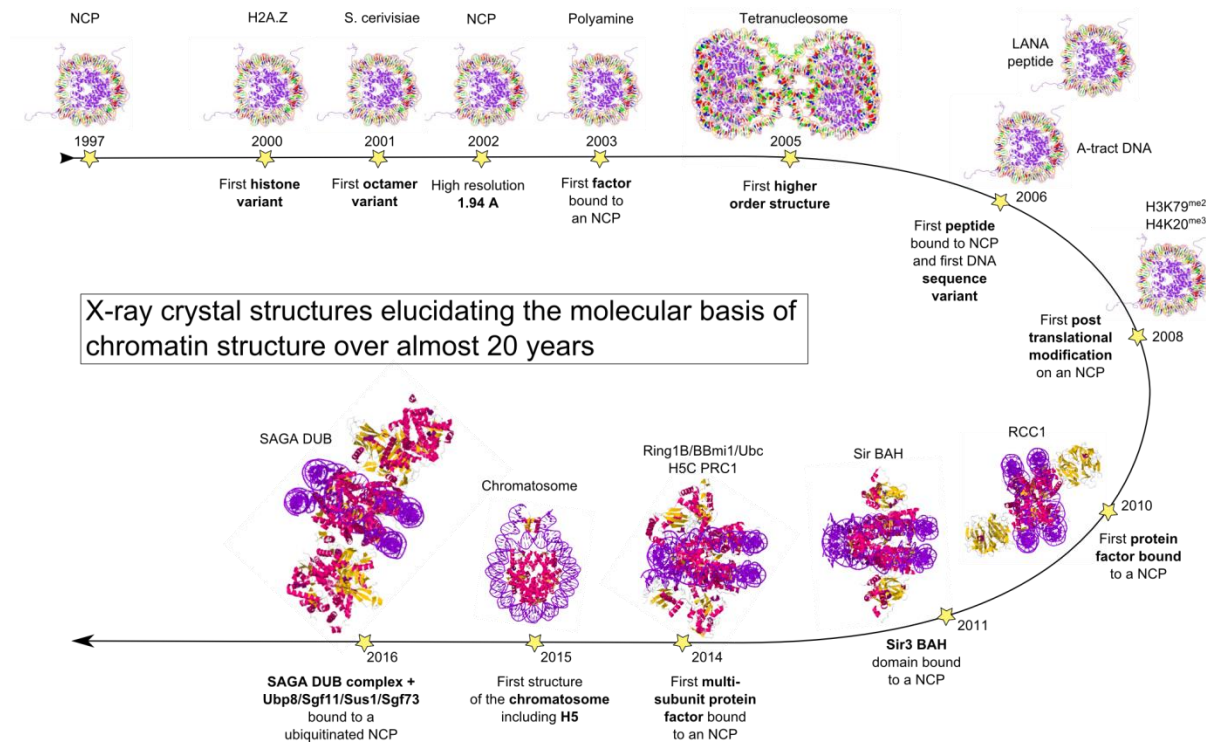
Figure 1.9: Timeline of a selection of discoveries in chromatin structural biology over a range of almost 20 years. The selection is based on structures deposited in the RSCB protein data bank as of August 2016. Structures solved by electron microscopy are not shown.

The NCP structure solved in 1997 contained human α-satellite-based DNA and a histone octamer core assembled from recombinant *X. laevis* histones. The first X-ray crystal structure of an NCP with a histone variant was solved in 2000, containing the H2A variant H2A.Z, and the first nucleosome structure containing a histone octamer core of a different species (*S. cerevisiae*) followed a year later in 2001 [22, 25]. In 2003, the first NCP X-ray crystal structure with a ligand (a polyamine in this case) bound to the NCP was published [73]. An important structure for the elucidation of higher order chromatin organization was solved in 2005 with the tetranucleosome [14]. Although only solved to 9 Å, the structure gave important insights into a possible form of the elusive 30 nm fiber. Further important structures were published in 2006, one of them showing the nucleosome core particle with an altered DNA sequence containing a poly(dA:dT)-tract and a metal response element (MRE) and the other one showing the binding of a peptide (LANA) to the acidic patch on the octamer surface, a motif that turned out to be an important binding site for many chromatin regulating enzymes [46, 31]. In 2008, the Luger lab published structures of posttranslationally modified NCPs containing

16

the histone marks H3 K79^{me2} and H4 K20^{me3} respectively [74]. The first structure of a full length chromatin regulating enzyme (RCC1 (Regulator of Chromatin Condensation 1)) bound to a nucleosome was solved in 2010 [47]. Since then, several more nucleosome co-crystal structures were solved. Of note are the structure of the Sir3 BAH domain bound to the nucleosome in 2011 [38], the structure of the polycomb repressive complex 1 (PRC1) ubiquitination module bound to the nucleosome in 2014 [75], the SAGA DUB complex bound to a ubiquitylated nucleosome (2016) [41] and also the first X-ray crystal structure of the chromatosome containing the linker histone H5 [42] in 2015. With the steady improvement of electron microscopic methods and equipment, structural information is also beginning to be generated by the emerging technique of cryo electron microscopy, evidenced by the cryo-EM structure of a retroviral integrase bound to a nucleosome or the very recently published cryo-EM structure of the nucleosome, resolved at 3.9 Å [76, 77].

1.7. Chromatin regulation

The positioning of nucleosomes and the corresponding accessibility of transcription factor binding sites on nucleosome DNA are important determinants of transcription levels of genes. The mechanism(s) of DNA positioning are mainly driven by intrinsic features of the DNA, such as its form and sequence. Another important factor in the regulation of transcriptional levels and many other regulatory mechanisms in chromatin is the presence of posttranslational modifications (PTMs) on histones. The discovery of correlations between certain posttranslational modifications and the levels of genetic activity led to the suggestion of a ´histone-code´ [54]. This hypothesis postulates a code consisting of PTMs which are written and read by histone writers and histone readers, respectively. Posttranslational modifications can consist of the addition of small moieties (e.g. the methylation of lysines or arginines or the phosphorylation of threonines, serines or tyrosines), but may also include the covalent attachment of small proteins such as ubiquitin to lysine amino acids. Most of the PTMs occur on the unstructured histone tail regions extruding from the nucleosome (Fig. 1.10). Prominent PTMs correlating with transcriptional repression are H3 K9^{me2,3} and H3 K27^{me2,3} whereas monomethylation of these lysine amino acids correlates with transcriptional activation. On the other hand, PTMs such as H3 K4^{me2,3} and H3 K36^{me3} strongly correlate with actively transcribed genes. Evidence is accumulating that posttranslational modifications might be involved in gene regulation in a synergistic manner. There are known cases of histone modification crosstalk. For example, the higher methylation states of H3 K79 depend on the ubiquitination of H2B K123 (K117 in yeast). Many more cases of crosstalk are known (Fig. 1.10 B) (reviewed in [78]).
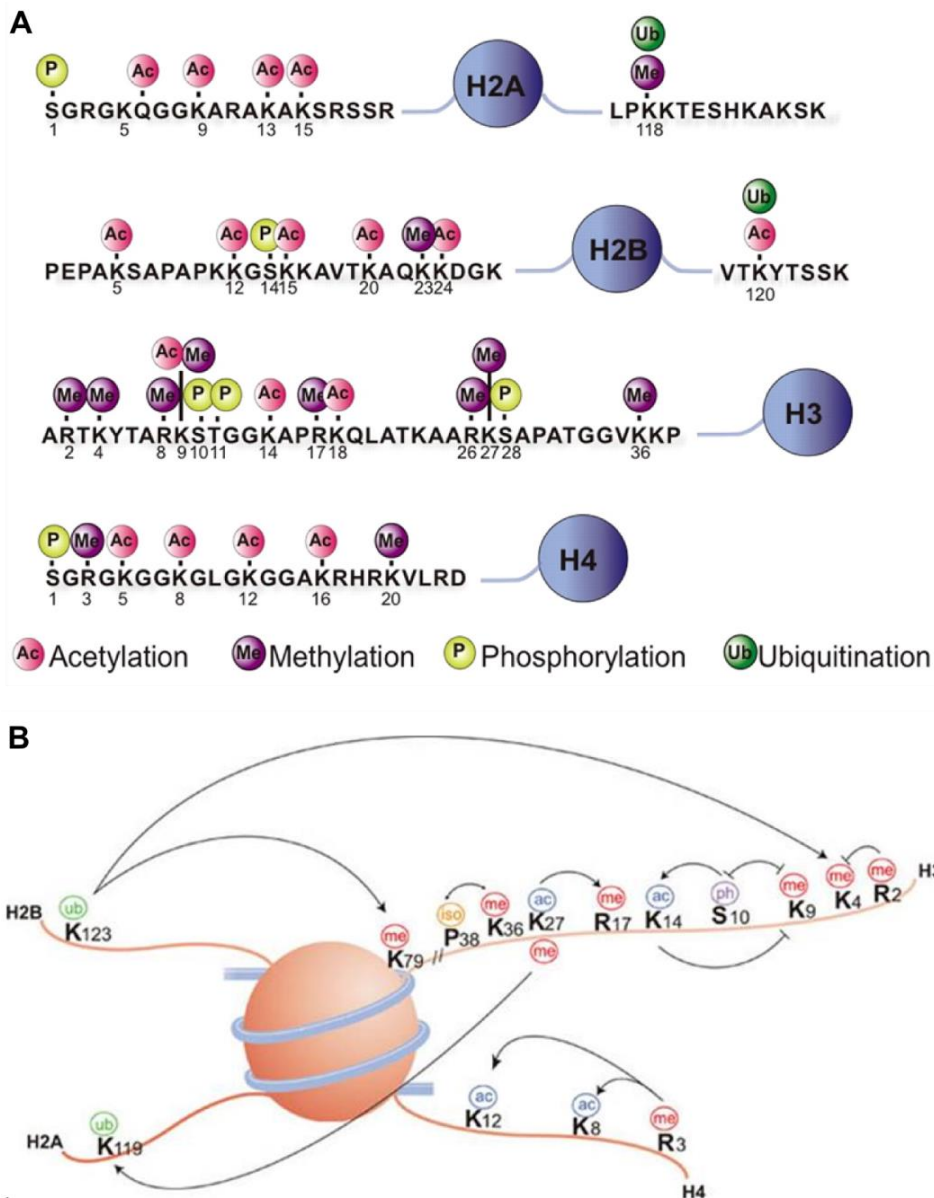
Figure 1.10: Posttranslational modifications on the histone tails. **A)** The core regions of the histones are drawn as blue circles. All N-terminal tails are targets of numerous posttranslational modification. The H3 N-terminal tails are important targets for lysine methylations. These can occur as mono- di- or trimethylations. H2A and H2B are also posttranslationally modified on their C-terminal tails (from **[79]**). **B)** Crosstalk between histone posttranslational modifications. Known cases of crosstalk are indicated by black arrows. Histone tail and amino acid identities are indicated. me = methylation, ac = acetylation, ph = phosphorylation, ub = ubiquitination, iso = isomerization **[78]**.

1.8. H3 K79 Methylation by Dot1

The PTM H3 K79^{me1,2,3} is associated with actively transcribed genes. This PTM is unique as it is a modification of an amino acid which lies in the core of the histone-fold and not in its tail region, yet still on the surface of the nucleosome. The enzyme exclusively responsible for the establishment of the various methylation states of H3 K79 in yeast was identified as Dot1, named due to its disrupting

effect on telomeric silencing (DOT) upon overexpression [80, 81]. Many different roles of the Dot1 methyltransferase activity have since been discovered, ranging from cell development and reprogramming to differentiation and proliferation (reviewed in [82, 83]). Dot1 plays important roles during several stages of the cell cycle. Dot1 expression peaks during G_1, but is also implicated in S phase progression as well as mitosis and meiosis (reviewed in [83]). The yeast (*S. cerevisiae*) Dot1 enzyme is biochemically well characterized. It is responsible for the formation of all three methylation states of H3 K79[(1,2,3)], although it was found that the di- and trimethylation of H3 K79 are dependent on the ubiquitination of H2B K120 (K123 in mammals) [84, 85]. Dot1 was found to employ a distributive methylation mechanism which is conserved among studied homologues of Dot1 [86]. Furthermore, a charge-based interaction between a basic stretch on the H4-tail ($R_{17}H_{18}R_{19}K_{20}$) and an acidic patch on Dot1 close to its C-terminus (aa 557-561) was identified to be important for the methylation of H3 K79 but it was not found to be necessary for Dot1 binding to NCP [87]. Moreover, Dot1 will only methylate K79 on histone H3 in the context of a nucleosome and not on free H3 in solution. This is probably due to the interaction with the H4 tail, necessary for methylation activity [81].

In 2003, the X-ray crystal structure of human Dot1L was solved to 2.5 Å [88]. A year later, the C-terminal 425 amino acids of yeast Dot1 (ΔDot1) were solved at 2.2 Å resolution [89]. The substrate binding site of Dot1 does not resemble those of other lysine methyltransferases, but is more similar to that found in class I arginine methyltransferase (PRMT1) [82].
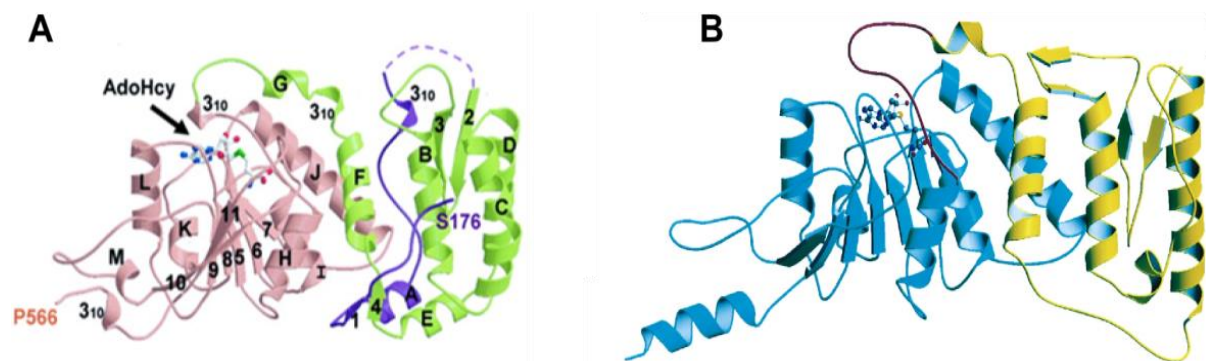
Figure 1.11: X-ray crystal structures of yeast Dot1 and human Dot1L. **A)** X-ray crystal structure of N-terminally truncated Dot1 (Δ157). The two main domains are colored light green (α-helical domain) and pink (catalytic domain with Rossmann fold and substrate (SAM) binding site). The substrate SAH (AdoHcy) is shown in balls and sticks. Only amino acids 176 – 566 are resolved in the structure (from **[89]** with permission). **B)** X-ray crystal structure of human Dot1L (aa 1 – 332) showing a very similar fold as the yeast homologue. The catalytic domain is colored in cyan and the α-helical N-terminal domain is colored yellow. The methyl group donor SAM is shown in sticks-and-balls representation (from **[89]** with permission).

The global architecture of Dot1Δ157 features two main domains. The N-terminal domain consists mostly of α-helices (αA - αG). The C-terminal domain harbors the catalytic center with its S-adenosyl methionine (SAM) binding pocket proximal to seven β-strands (β5 - β11) surrounded by two α-helices (αJ and αL) forming a Rossmann fold. The active center of Dot1 features a narrow channel

(approximately 4 × 5 Å) surrounded by hydrophobic amino acids (Fig. 1.12 C). The product of the methylation substrate (S-adenosyl homocysteine, SAH) is bound in this channel with the cysteine moiety facing towards the bottom of the channel. If bound similarly, this would place the methyl group of active S-adenosyl methionine (SAM) at the bottom of the binding channel, where it could donate its methyl group to H3 K79 upon insertion of the lysine sidechain into the binding channel (Fig. 1.12). The exit of the substrate binding channel is surrounded by a flexible loop which could potentially regulate the interaction of the active site of Dot1 with its methylation target.
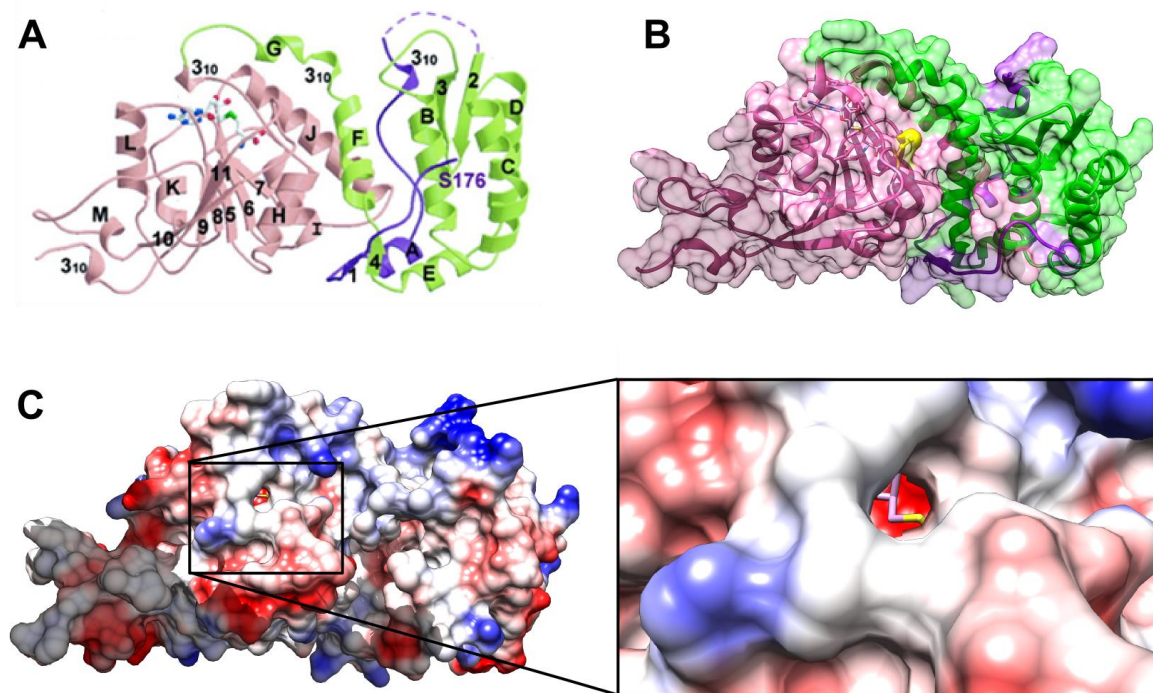
Figure 1.12: Crystal structure of ΔDot1 with putative H3 K79 binding channel. **A)** Model of ΔDot1 with color coding as in Fig. 1.11 A. **B)** Surface representation with surfaces colored as model in A). The SAM sulfur atom is colored yellow. **C)** Coulomb surface of ΔDot1 looking into the putative binding channel for H3 K79 (blow-up box). Blue denotes positive surface charge, red denotes negative surface charge and white represents neutral surfaces.

The missing N-terminal amino acids (1-157) are predicted to be mostly α-helical and putatively form a DNA binding domain [89]. ΔDot1 carries a net positive charge at neutral pH with a theoretical pI of 8.37 (ExPASy [90]).
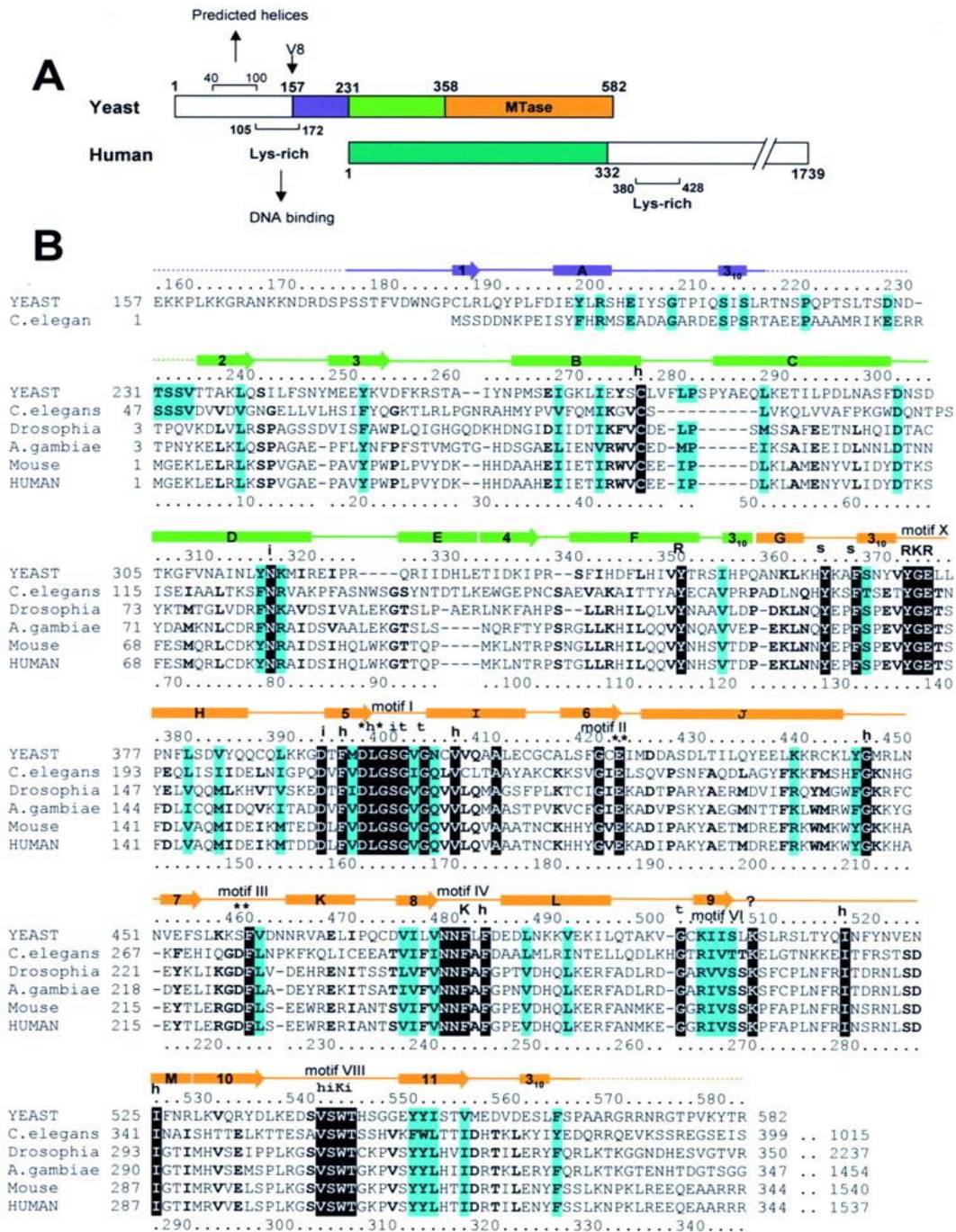
Figure 1.13: Schematic representation of the yeast and human homologues of Dot1 with sequence alignment. **A)** The N-terminal 157 amino acids (white) are not part of the crystallized protein. The proteolytic cleavage site at position 157 is labeled (V8). Following the first 74 N-terminal amino acids (blue), the mainly α-helical domain (light green) has a homologous counterpart in human Dot1L (dark green). The C-terminus harbors the catalytic center with its methyltransferase activity (orange). **B)** Sequence alignment of yeast Dot1 with its homologues from C. *elegans, D. melanogaster, A. gambiae, M. musculus* and *H. sapiens*. The numbering as well as the secondary structure annotation (orange) corresponds to the *S. cerivisiae* homologue Dot1 (from **[89]** with permission).

Taking into account the structural information and the well characterized interactions found between Dot1 and various histones, a model for the bound state of Dot1 to the NCP was suggested (Fig. 1.14) [89]. It places the Dot1 molecule in a position so that the methyl group of SAM can be

transferred to H3 K79 and the C-terminal region of Dot1 close to the H4 tail, where the charge-based interaction was identified. The N-terminal domain of Dot1, which is speculated to be a DNA binding domain, is placed close to the nucleosome DNA at the entry/exit sites from the NCP.
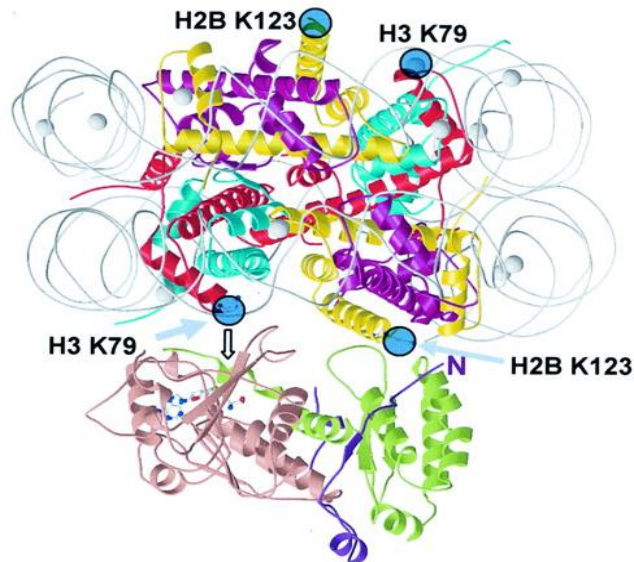
Figure 1.14: Proposed model of Dot1 binding to the NCP based on structural information and biochemically derived data. Dot1 is shown in pink, light green and blue. The N-terminus is positioned close to the ubiquitination site on H2B (yellow) K120/123 (labeled and highlighted by blue circles) and the active site with the SAH (represented as sticks and balls) is placed close to H3 (red) K79 (labeled and highlighted by blue circles). The H4 (cyan) N-terminal tail is in close proximity to the C-terminus of Dot1, consistent with the observed interaction of an acidic patch (aa 557-561) on Dot1 and a basic patch on the H4 tail (17-20) (from **[89]** with permission).

1.9. Regulation of silencing by Dot1

Dot1 with its H3 K79 methyltransferase activity is known to disrupt silencing in telomeres upon overexpression. The effects of silencing disruption can be brought about by two different mechanisms of action. Firstly, Dot1 is responsible for the mono- di- and trimethylation of H3 K79 on the previously identified LRS (loss of rDNA silencing) surface on the histone octamer which is also the binding target for the BAH (bromo adjacent homology) domain of Sir3, the silencing information regulator in yeast [38, 48]. The methylation of K79 within this surface changes the shape of the LRS surface, potentially leading to the effects of silencing [74]. Secondly, the presence of the Dot1 molecule itself can potentially lead to silencing by competing for access to the histone octamer LRS surface or histone tail regions which are important for binding of silencing proteins, e.g. the binding of the basic stretch on the H4 N-terminal tail (R17 - K20), which was shown to bind to Sir3 as well as to Dot1 [38, 87]. Sir3 binding is effectively suppressed by methylation of H3 K79, independent of the degree of methylation (mono- di- or trimethylation). Sir3 binding to the nucleosome is also abolished

by acetylation of H4 K16, a PTM which prohibits chromatin folding into the 30 nm fiber [91]. These data suggest a mechanism for the disruption of silencing by Dot1 which involves suppression of Sir3 binding to nucleosomes by acetylation of H4 K16.
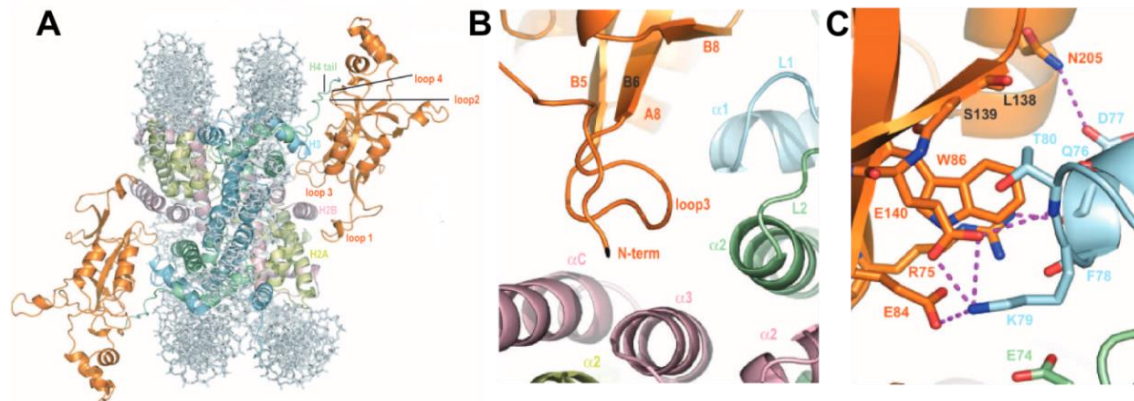
Figure 1.15: X-ray crystal structure of the Sir3 BAH domain bound to an NCP. **A)** Two copies of Sir3 BAH (orange) are bound on either side of the NCP. **B)** Interaction of the BAH domain with the LRS surface on the histone octamer which harbors H3 K79, the methylation target of Dot1. **C)** Zoomed-in image of (B) with the side chains shown. The interactions of the BAH domain amino acids to K79 on H3 are shown as purple broken lines (from **[38]** with permission).

In this configuration, Dot1 can methylate H3 K79 which prevents binding of the Sir3 BAH domain. Since there is no known demethylase for the H3 K79 methylation mark, H3 K79^{me} acts like a permanent barrier for the spreading of heterochromatin formation. Dot1 also competes directly with Sir3 for binding to the basic patch on the H4 N-terminal tail. This way, Dot1 is able to establish a boundary to spreading heterochromatin domains which are silenced by the Sir2/3/4 complex (see [92]).
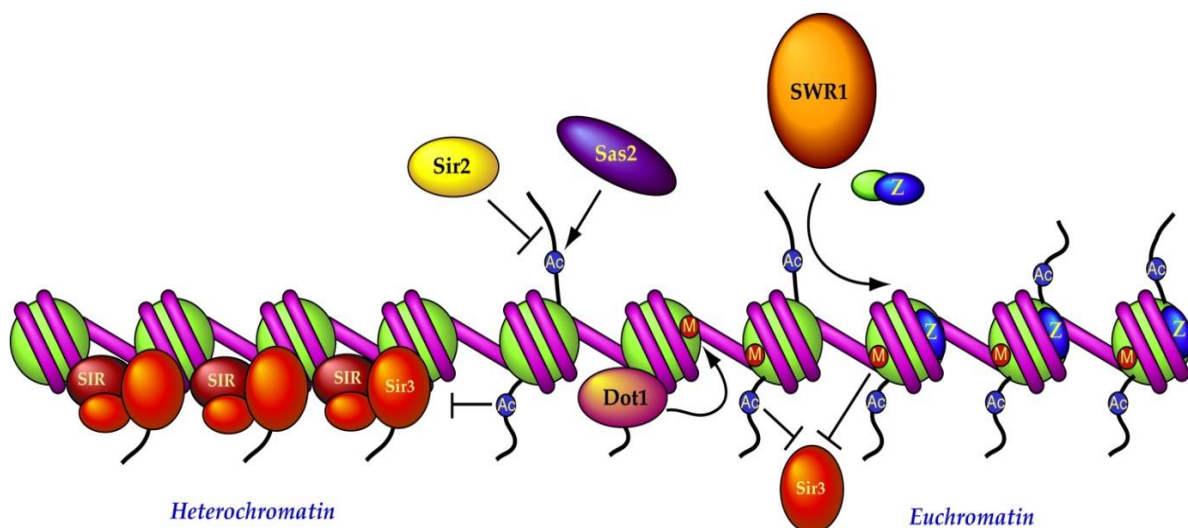
Figure 1.16: Model for the maintenance of the boundary between hetero- and euchromatin. The NCPs (histone octamer: green, DNA: magenta) are silenced by binding of the SIR complex. Sir2 (yellow) enables binding of the SIR complex by deacetylating H4 K16, a PTM (Ac) which is written by Sas2 (purple). Dot1 on the other hand methylates H3 K79 (M), which suppresses SIR binding. SWR1, a remodeling enzyme exchanges H2A.Z/H2B dimers with regular H2A/H2B dimers. **[92]**.

1.10. Thesis outline

During the work on this PhD thesis I was able to establish a working protocol for the crosslinking of nucleosome DNA to the histone octamer and show potential applications of the method.

Chapter 2 lists the materials, enzymes and bacterial strains used during this work.

Chapter 3 summarizes the methods used during the course of this thesis.

Chapter 4 is divided into four parts. Part 1 describes the establishment of the crosslinking protocol and the optimization of the crosslink location on the histone octamer and on the DNA. The disulfide crosslink was engineered by introducing a cysteine point mutation into the N-terminal tail of histone H3 in position 40 (R40C) where it passes through a minor groove channel formed by the nucleosome DNA between SHLs +1/-7 as well as between SHLs -1/+7. It is crosslinked to a convertible nucleotide which is introduced into the nucleosome DNA in direct proximity to the H3 N-terminal tail at position 70.I/J.

Part 2 of chapter 4 describes the production of an NCP which contains the DNA to H3 crosslink and an additional H2A N38C crosslink between the two copies of the H2A/H2B dimer and details the purification of the NCP.

In part 3 of chapter 4 the X-ray crystal structures of both crosslinked versions of the nucleosome are discussed. Both structures were determined to 2.8 Å.

Part 4 of chapter 4 discusses the application of the nucleosome crosslinking approach to a DNA sequence which has not been crystallized on a nucleosome before. Met16 promoter DNA was prepared and crosslinked to the histone octamer and set up for crystallization. Initial X-ray diffraction to 3.2 Å indicates successful growth of well-diffracting crystals.

Finally, chapter 5 describes how the crosslinked nucleosome was used to overcome the problem of DNA dissociation from the nucleosome upon dilution during preparative gel electrophoresis of a crosslinked complex between the NCP and a truncated version of Dot1 (ΔDot1). In previous experiments, the complex of nucleosome and ΔDot1 could be purified by preparative gel electrophoresis, but was not stable under the conditions of purification. This instability led to dissociation of DNA from the complex and most likely occurred after elution of the complex from the preparative gel in the highly dilute state. The protocol for crosslinking the nucleosome DNA to the histone octamer was combined with a previously established protocol for site-specific crosslinking of ΔDot1 to the

histone H3 via a T80C cysteine point mutation (V. Vogirala, PhD thesis). By incorporating two cysteine point mutations in H3 (R40C, T80C) into the histone octamer, a highly stable, double crosslinked complex of ΔDot1 to NCP could be reconstituted, effectively preventing DNA dissociation. Highly pure samples of the two binding states of ΔDot1, containing either one ΔDot1 molecule or two ΔDot1 molecules bound to the NCP, were used to repeat crystallization trials of the ΔDot1:NCP complex using the crystallization conditions previously found to yield crystal hits (V. Vogirala, PhD thesis [2]).

2. Materials

2.1. Chemicals

Chemical name	Supplier	Cat. No.
1,4-Dithio-DL-Threit(ol) (DTT)	AppliChem	A1101
100 bp DNA standard	New England Biolabs	N3231L
2-log DNA ladder	New England Biolabs	N3200L
2-Propanol	Merck	109634
2-Mercaptoethanol (βME)	Fluka	63690
3-Methyl butanol (isoamylalcohol)	Fluka	59090
50 bp DNA standard	New England Biolabs	N3236L
Acetic acid	Merck	101830
Acetonitrile	Merck	100030
Acrylamide 4K solution (30%) - Mix 37.5 : 1	AppliChem	A1672
Adenosine 5'-Triphosphoric Acid Disodium Salt (ATP)	AppliChem	A1348
Ammonium hydroxide 32%	AppliChem	A5169
Ampicillin Na-salt	AppliChem	A0939
Ammonium persulfate (APS)	Sigma-Aldrich	A3678
Benzamidine hydrochloride hydrate	Sigma-Aldrich	B6506
Bis-(2-hydroxyethyl)-imino-tris-(hydroxymethyl)-methan	AppliChem	A1025
Boric acid	Merck	100165
Bromphenol blue	TCI America	34725
Calcium chloride dehydrate	Sigma-Aldrich	C3881
Chloroform	Sigma-Aldrich	132950
Complete protease inhibitor	Roche	13760700
EGTA (Ethylene glycol-bis(2-aminoethylether)-tetraacetic acid)	Sigma-Aldrich	E4378
Ethidium bromide	Sigma-Aldrich	E8751
Ethanol	Merck	100983
Ethylenediaminotetraacetic Acid Disodium Salt Dihydrate	AppliChem	A2937
Gel red	Biotium	41003
Glucose	Riedel-de Haën	16301
Glycerol	AppliChem	A1123
Guanidinium hydrochloride	Sigma-Aldrich	G4505
Hydrochloric acid (fuming)	Merck	100317
Imidazole	Sigma-Aldrich	50750
Magnesium dichloride ($MgCl_2$)	Fluka	63072
Manganese dichloride tetrahydrate ($MnCl_2 \cdot 4H_2O$)	Sigma-Aldrich	M3634
Paraffin oil	Hampton Research	HR3-411
PEG6000	Sigma-Aldrich	81260
PEG8000	Sigma-Aldrich	89510
Phenol, buffered at pH 8.0	Sigma-Aldrich	P4557
PMSF (Phenylmethanesulfonyl fluoride)	AppliChem	A0999
Potassium acetate	Fluka	60034

Potassium chloride	Merck	104936
S-adenosyl homocysteine	Sigma-Aldrich	A9384
Silicon oil	Hampton Research	HR3-415
Sodium acetate	Sigma-Aldrich	236500
Sodium cacodylate	Sigma-Aldrich	20840
Sodium chloride	Merck	106404
Sodium dodecyl sulfate (SDS)	Sigma-Aldrich	71728
Sodium hydroxide	Merck	106469
TCEP (Tris(2-carboxyethyl)phosphine)	Sigma-Aldrich	C4706
TEMED (Tetramethylethylenediamine)	Fluka	87689
Tris (Tris(hydroxymethyl)aminomethane) base	Sigma-Aldrich	T1503

2.2. Enzymes

Enzyme name	Supplier	Cat. No.
AvaII	New England Biolabs	R0153M
BsaI-HF	New England Biolabs	R3535L
Calf intenstine phosphatase (CIP)	New England Biolabs	M0290L
EcoRV	New England Biolabs	R3195M
HinfI	New England Biolabs	R0155M
RNAse A	Sigma Aldrich	R6513
T4 Ligase	Richmond lab	-
T4 Polymerase	New England Biolabs	M0203L
Phusion Polymerase	New England Biolabs	M0530L

2.3. Size standards

Size standards A (2-log ladder, New England Biolabs), B (100 bp ladder, New England Biolabs) and C (50 bp ladder, New England Biolabs) are for gels separating DNA and size standard D (Dual color protein marker, Bio-Rad) is for SDS PAGE gels, separating protein.
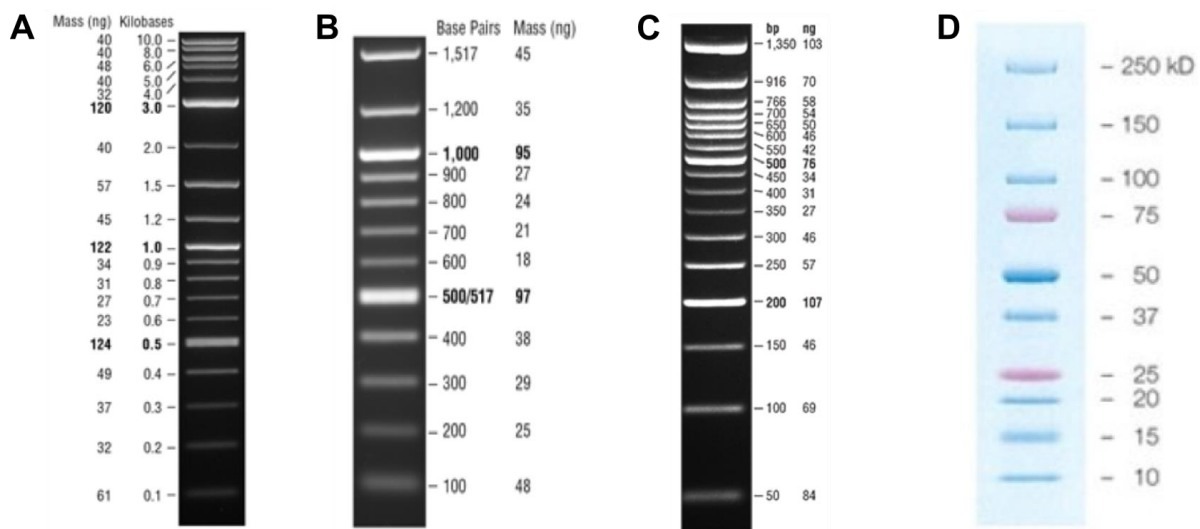
2.4. Laboratory consumables

Item	Usage	Specification
Amico Ultra-4	Sample concentration	10,000 Da MWCO, $V_{end} \cong 200\ \mu l$
Amico Ultra-15	Sample concentration	10,000 Da MWCO, $V_{end} \cong 500\ \mu l$
Vivaspin500 (Sartorius)	Sample concentration	5,000 Da MWCO, $V_{end} \cong 25\ \mu l$
Miracloth (Millipore)	Coarse filtration	Pore size: 22 μm – 25 μm

2.5. Buffers and solutions

Buffers (general)	Composition
10x TBE	0.89 M Tris base, 0.89 M boric acid, 15 mM EDTA
5x SDS gel running buffer	0.25 M Tris base, 1.92 M glycine, 0.5% SDS (w/v)
6x gel loading dye	2.5 mg/ml bromphenol blue, 30% glycerol (v/v), 60 mM EDTA
SDS gel loading buffer	1.1 mM bromphenole blue, 8% SDS (w/v), 50 mM EDTA pH 8.0, 0.2 M Tris/HCl pH 6.8, 35% glycerole
Buffers (enzyme reactions)	**Composition**
10x NEB2 buffer	500 mM NaCl, 100 mM Tris/HCl pH 7.9, 100 mM $MgCl_2$, 10 mM DTT
10x NEB3 buffer	1 M NaCl, 500 mM Tris/HCl pH 7.9, 100 mM $MgCl_2$, 10 mM DTT
10x NEB4 buffer	500 mM KOAc, 100 mM Tris/HOAc pH 7.9, 100 mM MgOAc, 10 mM DTT
10x Cutsmart buffer	500 mM KOAc, 200 mM Tris/HOAc pH 7.9, 100 mM MgOAc, 1 mg/ml BSA
10x EcoRV buffer	1 M NaCl, 500 mM Tris/HCl pH 7.5, 100 mM $MgCl_2$
10x T4 Ligase buffer –DTT	500 mM Tris/HCl pH 7.5, 100 mM $MgCl_2$, 20 mM ATP
Buffers (Chromatography)	
ΔDot1 buffer A	50 mM Tris/HCl pH 7.5, 500 mM NaCl, 15 mM imidazole, 10% glycerol, 1 mM benzamidine, 1 mM PMSF, 1 tablet Complete protease inhibitor
ΔDot1 buffer B	50 mM Tris/HCl pH 7.5, 500 mM NaCl, 10% glycerol, 250 mM imidazole
ΔDot1 buffer C	50 mM Tris/HCl pH 7.5, 200 mM NaCl, 10% glycerol, 5 mM DTT, 1 mM EDTA
ΔDot1 buffer D	50 mM Tris/HCl pH 7.5, 1.5 M NaCl, 10% glycerol, 5 mM DTT, 1 mM EDTA
ΔDot1 buffer E	20 mM Tris/HCl pH 7.5, 500 mM NaCl, 5% glycerol, 5 mM DTT, 1 mM EDTA, 1 mM S-Adenosyl homocysteine (SAH)
Histone C-18 buffer A	5% acetonitrile, 0.3% TFA
Histone C-18 buffer B	95% acetonitrile, 0.3% TFA
Octamer S-200 buffer	10 mM Tris/HCl pH 7.5, 2 M NaCl, 1 mM EDTA, 10 mM DTT
S-200 buffer (Histones)	6 M guanidine/HCl, 20 mM NaOAc pH 5.2, 200 mM NaCl, 1 mM EDTA
TES50	20 mM Tris/HCl pH 7.5, 0.1 M EDTA, 50 mM NaCl
TES250	20 mM Tris/HCl pH 7.5, 0.1 M EDTA, 250 mM NaCl
TES300	20 mM Tris/HCl pH 7.5, 0.1 M EDTA, 300 mM NaCl
TES600	20 mM Tris/HCl pH 7.5, 0.1 M EDTA, 600 mM NaCl

TES1500	20 mM Tris/HCl pH 7.5, 0.1 M EDTA, 1.5 M NaCl
TEK250	20 mM Tris/HCl pH 7.5, 0.1 M EDTA, 250 mM KCl
TEK600	20 mM Tris/HCl pH 7.5, 0.1 M EDTA, 600 mM KCl
Buffers (DNA production)	**Composition**
Cell resuspension buffer	25 mM Tris/HCl pH 8.0, 10 mM EDTA, 55 mM glucose
Cell lysis buffer	0.2 M NaOH, 1% SDS
Neutralization buffer	2 M HAc, 0.6 M KOAc, 0.4 M $CaCl_2$, 0.4 M $MnCl_2$
TEG buffer	50 mM Tris/HCl pH 8.0, 50 mM EDTA, 50 mM EGTA
Buffers (Histones)	**Composition**
Wash buffer	50 mM Tris/HCl pH 7.5, 2 M NaCl, 1 mM EDTA, 1 mM benzamidine
Triton wash buffer	Same as wash buffer, except with 1% Triton X-100 added
Unfolding buffer	6 M guanidine/HCl, 20 mM NaOAc pH 5.2
Buffers (Histones)	**Composition**
Unfolding buffer	6 M guanidine/HCl, 20 mM Tris/HCl pH 7.5, 10 mM DTT
Refolding buffer/S-200	2 M NaCl, 20 mM Tris/HCl pH 7.5, 1 mM EDTA, 10 mM DTT
Buffers (NCP)	**Composition**
TEK10	10 mM Tris/HCl pH 7.5, 0.1 mM EDTA, 10 mM KCl
TEK1.4	10 mM Tris/HCl pH 7.5, 0.1 mM EDTA, 1.4 M KCl
TEK2.0	10 mM Tris/HCl pH 7.5, 0.1 mM EDTA, 2.0 M KCl
TCS buffer –DTT	20 mM Tris/HCl pH 7.5, 1 mM EDTA
Buffers (Cells)	**Composition**
CCMB80	10 mM KOAc pH 7.0, 80 mM $CaCl_2$, 20 mM $MnCl_2$, 10 mM $MgCl_2$, 10% glycerol, pH adjusted to 6.4 with HCl
Buffers (Crystals)	**Composition**
HL-2	10 mM KCacodylate pH 6.0, 37 mM $MnCl_2$, 40 mM KCl, 2% MPD (V/V)
HL-24/2	10 mM KCacodylate pH 6.0, 37 mM $MnCl_2$, 40 mM KCl, 24% MPD (v/v), 2% Trehalose (w/v)
HL-40	10 mM KCacodylate pH 6.0, 37 mM $MnCl_2$, 40 mM KCl, 40% MPD (V/V)

2.6. Crystallization screens

2.6.1. Standard NCP screen (Mn^{2+} vs. K^+)

	1	2	3	4	5	6
A	40 mM KCaco pH 6.0 100 mM KCl 130 mM $MnCl_2$	40 mM KCaco pH 6.0 100 mM KCl 140 mM $MnCl_2$	40 mM KCaco pH 6.0 100 mM KCl 150 mM $MnCl_2$	40 mM KCaco pH 6.0 100 mM KCl 160 mM $MnCl_2$	40 mM KCaco pH 6.0 100 mM KCl 170 mM $MnCl_2$	40 mM KCaco pH 6.0 100 mM KCl 180 mM $MnCl_2$
B	40 mM KCaco pH 6.0 120 mM KCl 130 mM $MnCl_2$	40 mM KCaco pH 6.0 120 mM KCl 140 mM $MnCl_2$	40 mM KCaco pH 6.0 120 mM KCl 150 mM $MnCl_2$	40 mM KCaco pH 6.0 120 mM KCl 160 mM $MnCl_2$	40 mM KCaco pH 6.0 120 mM KCl 170 mM $MnCl_2$	40 mM KCaco pH 6.0 120 mM KCl 180 mM $MnCl_2$
C	40 mM KCaco pH 6.0 140 mM KCl 130 mM $MnCl_2$	40 mM KCaco pH 6.0 140 mM KCl 140 mM $MnCl_2$	40 mM KCaco pH 6.0 140 mM KCl 150 mM $MnCl_2$	40 mM KCaco pH 6.0 140 mM KCl 160 mM $MnCl_2$	40 mM KCaco pH 6.0 140 mM KCl 170 mM $MnCl_2$	40 mM KCaco pH 6.0 140 mM KCl 180 mM $MnCl_2$
D	40 mM KCaco pH 6.0 160 mM KCl 130 mM $MnCl_2$	40 mM KCaco pH 6.0 160 mM KCl 140 mM $MnCl_2$	40 mM KCaco pH 6.0 160 mM KCl 150 mM $MnCl_2$	40 mM KCaco pH 6.0 160 mM KCl 160 mM $MnCl_2$	40 mM KCaco pH 6.0 160 mM KCl 170 mM $MnCl_2$	40 mM KCaco pH 6.0 160 mM KCl 180 mM $MnCl_2$

2.6.2. Elevated salt NCP screen (Mn^{2+} vs. K$^+$)

	1	2	3	4	5	6
A	40 mM KCaco pH 6.0 140 mM KCl 160 mM MnCl$_2$	40 mM KCaco pH 6.0 140 mM KCl 170 mM MnCl$_2$	40 mM KCaco pH 6.0 140 mM KCl 180 mM MnCl$_2$	40 mM KCaco pH 6.0 140 mM KCl 190 mM MnCl$_2$	40 mM KCaco pH 6.0 140 mM KCl 200 mM MnCl$_2$	40 mM KCaco pH 6.0 140 mM KCl 210 mM MnCl$_2$
B	40 mM KCaco pH 6.0 160 mM KCl 160 mM MnCl$_2$	40 mM KCaco pH 6.0 160 mM KCl 170 mM MnCl$_2$	40 mM KCaco pH 6.0 160 mM KCl 180 mM MnCl$_2$	40 mM KCaco pH 6.0 160 mM KCl 190 mM MnCl$_2$	40 mM KCaco pH 6.0 160 mM KCl 200 mM MnCl$_2$	40 mM KCaco pH 6.0 160 mM KCl 210 mM MnCl$_2$
C	40 mM KCaco pH 6.0 180 mM KCl 160 mM MnCl$_2$	40 mM KCaco pH 6.0 180 mM KCl 170 mM MnCl$_2$	40 mM KCaco pH 6.0 180 mM KCl 180 mM MnCl$_2$	40 mM KCaco pH 6.0 180 mM KCl 190 mM MnCl$_2$	40 mM KCaco pH 6.0 180 mM KCl 200 mM MnCl$_2$	40 mM KCaco pH 6.0 180 mM KCl 210 mM MnCl$_2$
D	40 mM KCaco pH 6.0 200 mM KCl 160 mM MnCl$_2$	40 mM KCaco pH 6.0 200 mM KCl 170 mM MnCl$_2$	40 mM KCaco pH 6.0 200 mM KCl 180 mM MnCl$_2$	40 mM KCaco pH 6.0 200 mM KCl 190 mM MnCl$_2$	40 mM KCaco pH 6.0 200 mM KCl 200 mM MnCl$_2$	40 mM KCaco pH 6.0 200 mM KCl 210 mM MnCl$_2$

2.6.3. 601 NCP screen (Mn^{2+} vs. K$^+$)

	1	2	3	4	5	6
A	40 mM KCaco pH 6.0 100 mM KCl 80 mM MnCl$_2$	40 mM KCaco pH 6.0 100 mM KCl 90 mM MnCl$_2$	40 mM KCaco pH 6.0 100 mM KCl 100 mM MnCl$_2$	40 mM KCaco pH 6.0 100 mM KCl 110 mM MnCl$_2$	40 mM KCaco pH 6.0 100 mM KCl 120 mM MnCl$_2$	40 mM KCaco pH 6.0 100 mM KCl 130 mM MnCl$_2$
B	40 mM KCaco pH 6.0 120 mM KCl 80 mM MnCl$_2$	40 mM KCaco pH 6.0 120 mM KCl 90 mM MnCl$_2$	40 mM KCaco pH 6.0 120 mM KCl 100 mM MnCl$_2$	40 mM KCaco pH 6.0 120 mM KCl 110 mM MnCl$_2$	40 mM KCaco pH 6.0 120 mM KCl 120 mM MnCl$_2$	40 mM KCaco pH 6.0 120 mM KCl 130 mM MnCl$_2$
C	40 mM KCaco pH 6.0 140 mM KCl 80 mM MnCl$_2$	40 mM KCaco pH 6.0 140 mM KCl 90 mM MnCl$_2$	40 mM KCaco pH 6.0 140 mM KCl 100 mM MnCl$_2$	40 mM KCaco pH 6.0 140 mM KCl 110 mM MnCl$_2$	40 mM KCaco pH 6.0 140 mM KCl 120 mM MnCl$_2$	40 mM KCaco pH 6.0 140 mM KCl 130 mM MnCl$_2$
D	40 mM KCaco pH 6.0 160 mM KCl 80 mM MnCl$_2$	40 mM KCaco pH 6.0 160 mM KCl 90 mM MnCl$_2$	40 mM KCaco pH 6.0 160 mM KCl 100 mM MnCl$_2$	40 mM KCaco pH 6.0 160 mM KCl 110 mM MnCl$_2$	40 mM KCaco pH 6.0 160 mM KCl 120 mM MnCl$_2$	40 mM KCaco pH 6.0 160 mM KCl 130 mM MnCl$_2$

2.6.4. ΔDot1:NCP screen 1 (2-propanol, sodium citrate tribasic, buffer)

	1	2	3	4	5	6
A	100 mM Hepes pH 7.5 150 mM Na citrate tb 14% 2-propanol	100 mM Hepes pH 7.5 175 mM Na citrate tb 16% 2-propanol	100 mM Hepes pH 7.5 200 mM Na citrate tb 18% 2-propanol	100 mM Hepes pH 7.5 225 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 7.5 250 mM Na citrate tb 22% 2-propanol	100 mM Hepes pH 7.5 275 mM Na citrate tb 24% 2-propanol
B	100 mM Hepes pH 7.5 150 mM Na citrate tb 14% 2-propanol	100 mM Hepes pH 7.5 175 mM Na citrate tb 16% 2-propanol	100 mM Hepes pH 7.5 200 mM Na citrate tb 18% 2-propanol	100 mM Hepes pH 7.5 225 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 7.5 250 mM Na citrate tb 22% 2-propanol	100 mM Hepes pH 7.5 275 mM Na citrate tb 24% 2-propanol
C	100 mM Hepes pH 7.5 150 mM Na citrate tb 14% 2-propanol	100 mM Hepes pH 7.5 175 mM Na citrate tb 16% 2-propanol	100 mM Hepes pH 7.5 200 mM Na citrate tb 18% 2-propanol	100 mM Hepes pH 7.5 225 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 7.5 250 mM Na citrate tb 22% 2-propanol	100 mM Hepes pH 7.5 275 mM Na citrate tb 24% 2-propanol
D	100 mM Hepes pH 7.5 150 mM Na citrate tb 14% 2-propanol	100 mM Hepes pH 7.5 175 mM Na citrate tb 16% 2-propanol	100 mM Hepes pH 7.5 200 mM Na citrate tb 18% 2-propanol	100 mM Hepes pH 7.5 225 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 7.5 250 mM Na citrate tb 22% 2-propanol	100 mM Hepes pH 7.5 275 mM Na citrate tb 24% 2-propanol

2.6.5. ΔDot1:NCP screen 2 (20% 2-propanol, sodium citrate tribasic (tb), buffer)

	1	2	3	4	5	6
A	100 mM Hepes pH 8.0 150 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 8.0 175 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 8.0 200 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 8.0 225 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 8.0 250 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 8.0 275 mM Na citrate tb 20% 2-propanol
B	100 mM Hepes pH 7.5 150 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 7.5 175 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 7.5 200 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 7.5 225 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 7.5 250 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 7.5 275 mM Na citrate tb 20% 2-propanol
C	100 mM Tris pH 7.0 150 mM Na citrate tb 20% 2-propanol	100 mM Tris pH 7.0 175 mM Na citrate tb 20% 2-propanol	100 mM Tris pH 7.0 200 mM Na citrate tb 20% 2-propanol	100 mM Tris pH 7.0 225 mM Na citrate tb 20% 2-propanol	100 mM Tris pH 7.0 250 mM Na citrate tb 20% 2-propanol	100 mM Tris pH 7.0 275 mM Na citrate tb 20% 2-propanol
D	100 mM KCaco pH 6 150 mM Na citrate tb 20% 2-propanol	100 mM KCaco pH 6 175 mM Na citrate tb 20% 2-propanol	100 mM KCaco pH 6 200 mM Na citrate tb 20% 2-propanol	100 mM KCaco pH 6 225 mM Na citrate tb 20% 2-propanol	100 mM KCaco pH 6 250 mM Na citrate tb 20% 2-propanol	100 mM KCaco pH 6 275 mM Na citrate tb 20% 2-propanol

2.6.6. Screen 3 (100 mM Hepes pH 7.5, 20% 2-propanol vs. sodium citrate tb)

	1	2	3	4	5
A	100 mM Hepes pH 7.5 150 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 7.5 175 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 7.5 200 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 7.5 225 mM Na citrate tb 20% 2-propanol	100 mM Hepes pH 7.5 250 mM Na citrate tb 20% 2-propanol

2.7. Equipment

2.7.1. General equipment

Apparatus	Supplier	Model
Gel documentation	Carestream	Gel Logic 212 PRO
Heatblock	Techne	Dri-Block DB-3
Lyophilizer	Christ	Alpha 1-2 LD Plus
PCR cycler	Biorad	C1000 Thermal cycler
Rocking platform	VWR	Rocker 100
Rotation platform	Denley	Spiramax 5
Spectrophotometer	Thermo Scientific	Nanodrop 2000
UV transilluminator	LKB	2011 Macrovue
Vacuum pump	KNF Neuberger	Laboport
Vortexer	Merck	MELB 1719
Waterbath	Techne	TE-10A
Water purification	Millipore	Millipak 40

2.7.2. Purification

Apparatus	Supplier	Model
HPLC	Gilson	155/321/GX-271
HPLC	Gilson	112/119/305/306/805/811C/FC204
Peristaltic pump	Gilson	Minipuls-3
Plotter	Kipp & Zonen	BD41

2.7.3. Chromatography columns

Column	Type	Supplier	Specifications
5PW DEAE	Weak anion exchange	Tosoh Biosciences LLC	60 ml
S-200	Size exclusion	GE Healthcare	XK 16/70 (120 ml)
S-200 sephacryl	Size exclusion	GE Healthcare	XK 50/1000 (1600 ml)
HiTrap Q HP	Strong anion exchange	GE Healthcare	5 ml
HiTrap SP HP	Cation exchange	GE Healthcare	5 ml
HisTrap excel	Affinity tag (6xHIS)	GE Healthcare	5 ml
4PW C18	Reverse Phase	Tosoh Biosciences LLC	60 ml

2.7.4. Cell culture

Apparatus	Supplier	Model
Incubator shaker	New Brunswick Scientific	Innova 4330
Incubator	Heraeus	B 5050 E
Spectrophotometer	Pharmacia	Novaspec II

2.7.5. Cell lines

Cells	Genotype	Usage
BL21(DE3)	F$^-$ *ompT hsdS*$_B$(r$_B$-m$_B$-) *gal dcm* (DE3)	DNA production
BL21(DE3)pLysS	F$^-$ *ompT hsdS*$_B$(r$_B$-m$_B$-) *gal dcm* (DE3)pLysS(Cam^R)	Histone expression
BL21-CodonPlus - (DE3) RIPL strain	B F$^-$ *ompT hsdS*$_B$(r$_B$-m$_B$-) *dcm* Tetr *gal* λ(DE3) *endA* Hte [*argU proL* Cam^r][*argU ileY leuW* Strep/Spec^r]	ΔDot1 expression

2.7.6. Centrifugation

Apparatus	Supplier	Model
Centrifuge	Heraeus	Biofuge pico
Eppendorf centrifuge	Eppendorf	Model 5424
Tabletop centrifuge	Hettich	Rotina 38R
Tabletop centrifuge	Hettich	Rotanta 46R
Centrifuge	Sorvall	RC 6+
Centrifuge	Sorvall	RC 26 PLUS

2.7.7. Rotors

Rotor	Supplier	Model
For 30 ml tubes	Sorvall	SS-34
For 500 ml bottles	Sorvall	SLA-3000

3. Methods

3.1. Analytical methods

3.1.1. Agarose gel electrophoresis

Agarose gels were cast by adding 1 g of agarose to 100 ml of 1x TBE buffer for a 1% agarose gel. 0.8 g or 2.0 g of agarose per 100 ml 1x TBE buffer yielded 0.8% or 2% agarose gels, respectively. The agarose was allowed to soak for 5 min and was then heated in a microwave oven until the agarose had fully dissolved. The agarose was then allowed to cool down to approximately 50 °C. 10,000x Gelred (Biotium) was added to a final concentration of 1x. The liquid agarose was poured into a 100 mm x 70 mm x 10 mm cast. A suitable sample well comb (10-well or 15-well) was placed in the liquid agarose. After setting, the gel was immersed in 1x TBE buffer. DNA samples were mixed with 0.2x sample volumes of 6x GLD. As size reference 5 µl of a 1:10 dilution of 2-log ladder or 100 bp ladder (New England Biolabs) were loaded. Electrophoresis was carried out at 120 V (80 V for 0.8% agarose gels) until the marker dye reached the end of the gel. Imaging of agarose gels was done on a gel imager (Carestream). Cutting out of gel bands for gel extraction was carried out on a UV illuminator.

3.1.2. SDS polyacrylamide electrophoresis

A mini-PROTEAN assembly (Biorad) with 0.75 mm spacers was used to pour the resolving gel first. To prepare two 15% (18%) SDS PAGE gels, 1.8 ml (1.0 ml) of MilliQ water was mixed with 4 ml (4.8 ml) 30% acrylamide (37.5:1). 2 ml 3 M Tris/HCl pH 8.8 and 80 µl 10% SDS were added. To start the polymerization, 32 µl of 25% APS and 8 µl of TEMED were added. The acrylamide solution was topped with 2-propanol to straighten the meniscus and prevent radical scavenging by air oxygen. Upon complete polymerization, the 2-propanol was removed and the stacking gel was poured on top of the resolving gel. The stacking gel was prepared by mixing 2.5 ml MilliQ water, 0.75 ml 30% acrylamide (37.5:1), 0.96 ml 0.5 M Bis-Tris buffer pH 6.8 and 40 µl 10% SDS. The polymerization reaction was started by adding 16 µl 25% APS and 4 µl TEMED. A suitable sample well comb was inserted until the stacking gel was set. Samples for SDS PAGE were mixed in a 1:1 ratio with PGLB and heated for 5 min at 100 °C. As a size reference, 5 µl of dual color protein standard (Bio-Rad) was loaded. Electrophoresis was carried out at 250 V in 1x SDS PAGE running buffer until the dye front reached the end of the gel. Staining of SDS gels was done in coomassie stain and microwaving for 60 s and consecutive shaking on a rocking platform for 15 min. The gel was destained in a mixture of acetic acid and ethanol, also heated in the microwave and then shaken on the rocking platform until sufficient destaining was achieved. Imaging of SDS PAGE gels was done on a gel imager (Carestream) by white light illumination.

3.1.3. Native polyacrylamide gel electrophoresis (PAGE)

DNA samples, especially of small sizes, were also visualized on polyacrylamide gels. Native PAGE gels were cast in a Mini-PROTEAN assembly (BioRad) with 0.75 mm spacers. To pour two native PAGE gels, 6 ml MilliQ water were mixed with 1 ml 10x TBE buffer and 3 ml 30% acrylamide (37.5:1). The polymerization reaction was started by addition of 40 µl 25% APS and 10 µl TEMED. Appropriate sample well combs were inserted into the gel until it was set. Samples for native PAGE were mixed with 0.2x sample volumes of 6x GLD. As size reference 3 µl of a 1:10 dilution of 2-log ladder (New England Biolabs) was loaded. For oligonucleotides, 3 µl of a 1:10 dilution of 100 bp ladder or 50 bp ladder (New England Biolabs) was loaded. Electrophoresis was run for 25 – 30 min at 200 V in 1x TBE buffer for 10% gels and for 40 min at 200 V in 1x TBE for 15% gels. Staining of native PAGE gels was done for 15 min in 100 ml of a 1x Gelred (Biotium) solution. Imaging was done on a gel imager (Carestream) by UV illumination.

3.1.4. Electrophoretic mobility shift assay (EMSA)

To visualize the shifting of nucleosomes by the binding of proteins or by crosslinking, nucleosome samples were run on a 200 mm x 200 mm x 1 mm 5% native PAGE with continuous buffer recircula-tion in 0.25x TBE buffer at 4 °C. EMSA gels were prepared by mixing 39 ml MilliQ water with 1.25 ml 10x TBE buffer and 8.33 ml acrylamide (37.5:1). Polymerization was started by adding 160 µl 25% APS and 50 µl TEMED. After setting, the EMSA gel was prerun in 0.25x TBE for at least 3 h with con-stant buffer recirculation at 4 °C and 250 V. The samples were mixed with 0.2x sample volumes of 25% sucrose and loaded on the EMSA gel. As size reference, 10 µl of a 1:10 dilution of 2-log ladder (New England Biolabs) was loaded. Electrophoresis was run for at least 3 h at 200 V in 0.25x TBE buffer with continuous buffer recirculation at 4 °C. Staining of EMSA gels was done in ethidium bro-mide for at least 15 min at 22 °C. Imaging was done on a gel imager (Carestream) with UV back illu-mination.

3.1.5. Preparative gel electrophoresis

For preparative gel electrophoresis a Model 491 prep gel (Bio-Rad) was used. A 5% preparative poly-acrylamide gel was prepared in a glass column with an inner diameter of 28 mm and a bed height of approximately 130 mm. The preparation of the gel solution was essentially as described in section 3.1.4., although for the preparative gel the acrylamide solution was thoroughly degassed before casting and the poured gel was topped with 2-butanol instead of 2-propanol. The gel column was water cooled during casting to prevent warping of the setting gel. Once it was set, the 2-butanol was removed and the gel surface was washed and then topped with 0.25x TBE buffer. The gel column

was assembled into the running tank as described in the user manual [93]. 0.25x TBE was used as running buffer and the preparative gel was pre-run for 12 h at 220 V and 4 °C with continuous buffer recirculation. Sucrose was added to the samples to a final concentration of 5%. The sample was centrifuged at maximum speed for 5 min at 4 °C before loading. The electrophoresis ran at 200 – 250 V for 15 h while the eluate was collected at 1.0 – 1.5 ml/min with 5 min of collection per fraction. The UV absorption of the eluting fractions was spectroscopically measured at 260 nm and the UV trace was recorded on a chart recorder (Kipp & Zonen).

3.2. DNA methods

3.2.1. Preparation of chemically competent cells

20 µl of chemically competent TOP10 cells (Invitrogen) were spread on a TYE plate without antibiotics. The plate was incubated upside down for 12 h at 22 °C. 100 ml of 2x TY medium were inoculated with a single colony from the TYE plate using sterile technique. The culture was grown at 20 °C and 195 RPM in the shaking incubator to an OD_{600} of 0.2. The cells were then spun down at 2,200 g and 4 °C for 10 min. The pellet was then gently resuspended in 50 ml CCMB80 buffer and incubated on ice for 20 min. The cells were centrifuged again at 2,200 g and 4 °C for 10 min. The pellet was resuspended again in 10 ml ice cold CCMB80 buffer and aliquoted in small reaction tubes which were stored at –80 °C until use.

3.2.2. Polymerase chain reaction (PCR)

The typical reaction volume of a PCR reaction was 50 µl in a thin walled PCR tube to allow for fast equilibration times in the sample upon temperature change. During operation the lid was heated to 95 °C to prevent evaporation and recondensation of the sample. The PCR samples contained 1x concentration of PCR buffer as well as 200 µM of dNTPs as raw material for DNA synthesis. The vector template was added to a final concentration of 0.1 ng/µl and PCR primers were added to 0.5 µM final concentration. Finally, one unit of Phusion DNA polymerase (New England biolabs) was added. The initial denaturation step was typically 30 s at 95 °C, followed by a variable number of cycles (between 25 – 35), each consisting of a denaturation phase (20 s) at 95 °C, a primer annealing phase (20 s) at temperatures depending on the sequence of the primer and a DNA polymerization phase (15 s/kb) at 72 °C. After the last PCR cycle was finished, a final DNA polymerization phase (5 min) at 72 °C ensued after which the sample was cooled to 4 °C. 1 µl of DpnI was added to the PCR reaction to digest the methylated template plasmid. The sample was incubated for 60 s at 37 °C. If needed, the PCR reaction was purified from excess primer and free dNTPs using a PCR cleanup kit (Sigma-

Aldrich), or run on an agarose gel after which the bands could be excised and the DNA extracted from the gel slabs.

3.2.3. Cloning methods

3.2.3.1. SLIC cloning

To be able to ligate oligonucleotides containing a convertible nucleotide to the core segment of the target DNA sequence, a restriction site was introduced into the template plasmid containing the target sequence. In the case of α-satellite DNA, an AvaII restriction site was introduced 54 base pairs away from the dyad axis. Oligonucleotides containing the mutated sequence along with 20 - 30 base pairs homology to the sequence of the target plasmid were purchased and used for SLIC cloning. In SLIC cloning, the T4 polymerase 3'-exonuclease activity in the absence of free nucleotides was used to digest nucleotides from the 3' end of a linearized plasmid and an insert for the same amount of time, leaving complementary ends of single stranded DNA for reannealing and transformation. Commonly, the T4 polymerase digest of the plasmid and the insert are done separately and then added together for annealing and transformation after the digest (for single-tube cloning, see section 3.2.3.2). In a typical reaction volume of 20 µl, 20 - 40 ng/µl plasmid were digested with 1x BSA added in 1x NEB2 buffer. Addition of 25 - 50 units of T4 DNA polymerase started the digest. To stop the exonuclease digest, dCTP was added to a final concentration of 1 mM. In a separate reaction tube, approximately 10 ng/µl insert was simultaneously digested in 20 µl 1x NEB2 with 1x BSA added. The reaction was started by addition of 25 - 50 units of T4 DNA polymerase. Of crucial importance was the correct timing of both digests to ensure matching overhangs on the plasmid and template DNA samples for annealing. Plasmid and insert were then mixed and 10x T4 ligase buffer was added to 1x final concentration. The mixture was then incubated at 42 °C for 30 min before being placed on ice. 250 µl of rich medium was added and the cells were incubated for 30 min at 37 °C and 200 RPM in the shaking incubator. 25 µl of the incubated cells were then spread on a TYE plate containing appropriate antibiotic and were incubated upside down for 12 h at 37 °C.

3.2.3.2. QUICK-SLIC cloning

The protocol for SLIC cloning was streamlined to be a one-pot reaction with T4 polymerase digestion of plasmid and insert occuring simultaneously in the same reaction tube in 1x NEB2 buffer. A typical reaction volume was 20 µl and contained 5 ng/µl insert (which can be a synthesized String® of DNA or a PCR product) and 25 ng/µl linearized target plasmid. 25 - 50 units of T4 polymerase were added to start the digest. After quenching the T4 polymerase exonuclease activity by addition of dCTP to a final concentration of 1 mM, the reaction was incubated for 10 min on ice before it was used for bacterial transformation directly.

36
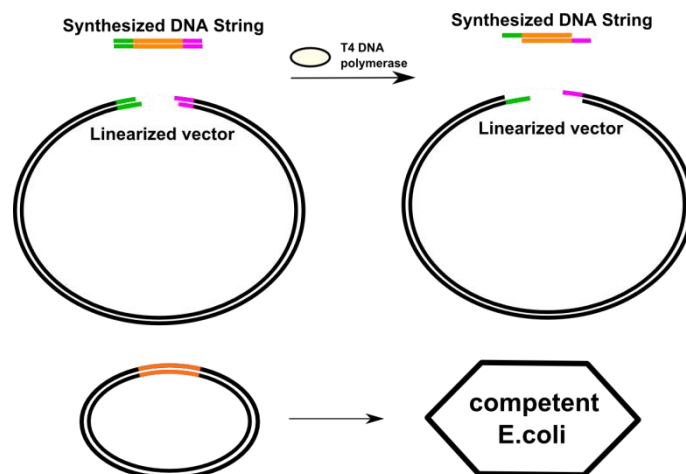
Figure 3.1: Schematic representation of SLIC cloning. The target vector which receives the DNA sequence of interest was linearized by restriction digestion. The insert (which can be a PCR product or a synthesized STRING® of DNA) carries the desired DNA sequence (orange), flanked by sequences of 20 – 30 base pairs homologous to the target vector (green and magenta). T4 DNA polymerase (champaign) with its 3'-exonuclease activity digests nucleotides at a speed of approximately 1 bp/min. After the digest, the digested ends annealed and the annealed product was used for transformation into competent cells.

3.2.4. Bacterial transformation

Plasmids acquired by either cloning method were introduced into *E.coli* strains by chemical transformation via heat shock. Either commercially available (Invitrogen) chemically competent TOP10 *E.coli* cells (for higher transformation efficiencies) or self-made chemically competent *E.coli* cells (if lower transformation efficiencies were sufficient) were used. Typically, 50 μl of chemically competent cells were thawed on ice. 0.1 – 1.0 μg of plasmid DNA was added to the cell suspension and incubated on ice for 30 min. Then the competent cells were exposed to 42 °C in a water bath for exactly 30 s with no mixing. After the heat shock the cells were placed on ice for 2 min. 500 μl of 2x TY medium was added to the cells which were allowed to recover their phenotype for 60 min at 37 °C shaking at 190 RPM. After phenotype recovery, the cells were spread on a TYE plate containing the appropriate antibiotic and incubated upside down at 37 °C for 12 h.

3.2.5. Ethanol precipitation of nucleic acids

0.1x sample volumes of 3 M sodium acetate pH 5.2 were added to the DNA sample. After mixing, 2.5x – 3x sample volumes of 100% ethanol were added. The sample was vortexed and incubated for 60 min on ice. For larger recoveries, incubation was done for 12 h at –20 °C. The precipitated DNA was pelleted by centrifugation at 4,000 g for 30 min and 4 °C. DNA pellets were washed with 70% ethanol to remove co-precipitated salt and then resuspended in 10 mM Tris/HCl pH 8.0.

3.2.6. 2-Propanol precipitation of nucleic acids

0.1x samples volumes of 3 M sodium acetate pH 5.2 were added to the DNA sample. After mixing, 0.5x - 1x sample volumes of 100% 2-propanol were added. The sample was vortexed and incubated for 60 min at 22 °C. For larger recoveries, incubation was done for 12 h at 22 °C. The precipitated DNA was pelleted by centrifugation at 4,000 g for 30 min and 20 °C. DNA pellets were resuspended in 10 mM Tris/HCl pH 8.0.

3.2.7. Miniprep

A culture from successfully transformed *E.coli* cells was grown by picking a single colony with a sterile inoculation loop and growing it in 5 ml 2x TY medium containing the appropriate antibiotic for 12 - 16 h at 37 °C and 190 RPM. 2 ml of cell culture were spun down in a 2 ml tube at maximum speed for 5 min. The miniprep was carried out as described in the supplier's manual (Sigma-Aldrich Cat.no. PLN350): The supernatant was discarded and the cells were resuspended in 200 µl cold resuspension buffer containing RNAse A. Cell lysis was carried out by adding 200 µl of lysis buffer containing SDS and NaOH. Within 5 min the cell lysate was neutralized by adding 350 µl of neutralization buffer. Cell debris was pelleted by centrifugation at maximum speed for 10 min. The supernatant was carefully transferred to a prepared DNA binding column. The column was attached to a vacuum manifold and washed with 500 µl wash solution 1 and then with 750 µl wash solution 2 containing ethanol. Remaining ethanol was removed by centrifugation in a fresh collection tube for 1 min at full speed. Elution of the plasmid DNA was done by addition of 100 µl of elution buffer (10 mM Tris/HCl pH 8.0) and centrifugation at maximum speed for 1 min.

3.2.8. DNA restriction digest

Analytical restriction digests were carried out in a volume of 50 µl containing 1x concentration of the buffer suitable for the digest (section 2.5). 1 µl of sample containing between 50 – 200 ng/µl DNA were digested with 10 units of restriction enzyme for 60 min at the temperature appropriate for the restriction enzyme. For DNA digests at a preparative scale 10,000 units of restriction enzyme were added per 5 mg of DNA. Preparative DNA digests were typically incubated for 12 - 16 h at the appropriate temperature. More enzyme was added until the completion of the digest was verified by gel electrophoresis.

3.2.9. **Tandem repeat cloning**

To maximize yields, the number of copies of the desired DNA sequence within the carrier plasmid was doubled to yield up to 32 copies. The pUC57-based plasmid containing one copy of the target DNA sequence was digested in two separate reactions with PstI/BamHI and PstI/BglII, respectively. PstI/BamHI cuts at both ends of the insert that was to be multiplied. PstI/BglII cuts the plasmid so that it contained the starting amount of copies of the target DNA sequence and had compatible sticky ends to the insert cut by PstI/BamHI. After the restriction digest the samples were run on a 1% agarose gel. The bands containing the linearized vector and the insert were cut out of the gel on a UV table. The DNA was recovered from the gel slices with a gel extraction kit (Sigma-Aldrich). Vector and insert were then mixed at a 1:2 ratio in 1x T4 ligase buffer. 1,500 units of T4 ligase were added and the reaction was incubated for 60 min at 22 °C. The ligation reaction was then used for the next round of bacterial transformation. Depending on the size of the insert and the carrier plasmid, up to 32 copies of the target DNA sequence could be successfully cloned into the carrier plasmid. The final plasmid carrying multiple copies of the target DNA sequence was sent for sequencing using T7 forward and T7 reverse sequencing primers. Long term storage of plasmids was at -20 °C.

3.2.10. **Large scale plasmid preparation**

For large scale preparation of DNA, plasmid containing up to 32 copies of the target DNA sequence was freshly transformed into chemically competent TOP10 *E.coli* cells (section 3.2.4) which were then streaked out on a TYE plate containing the appropriate antibiotic and incubated for 12 h at 37 °C. Starter cultures of several clones were grown for 12 h at 37 °C and 195 RPM in 100 ml 2x TY medium containing the appropriate antibiotic. 2 ml culture were centrifuged and the plasmid DNA extracted with a miniprep kit (section 3.2.7). The plasmid DNA was test digested with PstI/BamHI and the digested DNA was run on a 10% DNA PAGE to check the size of the insert. The starter culture carrying the plasmid with the most copies (ideally 32 copies) of target DNA was then used to inoculate 12 x 1 l 2x TY medium containing the appropriate antibiotic. The main culture was then grown for 12 – 16 h at 37 °C and 190 RPM. The cells were harvested by centrifugation at 6,000 g – 8,000 g for 6 – 7 min in a SORVALL SLA3000 rotor. The cell pellets were resuspended in a total of 500 ml cell resuspension buffer. The cell suspension was distributed to four 500 ml centrifugation bottles. Cell lysis was carried out by addition of 125 ml cell lysis solution. After mixing by inversion the solution was incubated at 22 °C for 10 min. Neutralization of the lysis solution was achieved by adding 125 ml ice cold neutralization buffer. Each bottle was shaken until the solution was homogenous and then incubated in the cooled (4 °C) SLA3000 rotor for 30 min after which the cell debris was pelleted by centrifugation at 11,000 g for 30 min at 4 °C. The supernatant was filtered through four layers of

Miracloth. 500 µl of 10 mg/ml RNAse A were added and the solution was incubated for 60 min at 37 °C in a water bath with repeated stirring. The DNA was precipitated from solution by addition of 40% PEG8000 to a final concentration of 10%. After incubation at 4 °C for at least 60 min the DNA was pelleted by centrifugation for 30 min at 4 °C in an SLA3000 rotor at 11,000 g. The DNA pellet was collected and resuspended in 50 ml TEG buffer. 5 ml of 3 M NaOAc pH 5.2 were added and then the DNA was precipitated by adding 0.52x sample volumes of 2-propanol. After incubation at 22 °C for at least 60 min, the DNA was spun down in 50 ml polypropylene tubes at 4,000 g and the pellets were resuspended in 9 ml MilliQ water.

3.2.11.   EcoRV release of insert DNA

10x EcoRV buffer was added to a final concentration of 1x. 100x BSA was added to a final concentration of 1x. Per 10 mg of plasmid DNA 3 µl of 1,000 U/µl EcoRV were added to cut out the insert pieces of DNA containing the target DNA sequence. The sample was incubated for 12 h at 37 °C. The completion of the digest was checked by running 1 µl of sample on a 1% agarose gel or a 10% native PAGE. If the digest was complete, the solution was brought to 500 mM NaCl by adding the appropriate amount of 5 M NaCl. 40% PEG6000/ 500 mM NaCl was then added to yield 10% PEG6000. The solution was incubated for at least 60 min at 4 °C. The precipitated vector DNA was pelleted by centrifugation at 4 °C for 30 min in an SS-34 rotor at 17,000 g. The supernatant was collected and the DNA was precipitated by addition of 0.52x sample volumes of 2-propanol. Centrifugation at 4 °C in 50 ml propylene tubes at 4,000 g pelleted the insert DNA which was then resuspended in 9 ml MilliQ water or in 3 ml TES300 for HPLC purification. The insert DNA is purified from remaining vector on a 60 ml 5PW DEAE ion exchange column (Tosoh Biosciences) using the salt gradient shown is Table 3.1. The peak fractions containing the insert DNA are pooled and ethanol precipitated. The pelleted insert DNA is resuspended in 9 ml MilliQ.

Time [min]	% Buffer B
0	0
9	20
49	65
51	100
65	100
66	0

Table 3.1: Salt gradient for the separation of insert DNA from remaining vector DNA on a 60 ml 5PW DEAE column (Tosoh Biosciences) with buffer A = TES300 and buffer B = TES600. Pump speed was 4 ml/min during all steps.

3.2.12.   Production of core DNA for ligation

The DNA was brought to 1x Cutsmart buffer by addition of 10x CutSmart. For the production of α-satellite based DNA, 100 µl of both HinfI and AvaII were added per 5 mg of insert DNA. The re-

striction digest was done for 12 h at 37 °C. Completion was tested on a 10% native PAGE gel. More enzyme was added until the restriction digest was complete. For BsaI constructs, 100 µl of BsaI were added instead of AvaII and HinfI. The digest was also carried out for 12 h at 37 °C. Upon completion of the restriction digest, the DNA was ethanol precipitated and resuspended in 3 ml TES300 for injection into a 60 ml 5PW DEAE ion exchange column. The digested DNA was separated from the short spacer DNA between restriction sites by running the same salt gradient as for the insert separation shown in Table 3.1. Peak fractions containing the core DNA were pooled and ethanol precipitated and subsequently resuspended in MilliQ water.

3.2.13.   Oligonucleotide design

The oligonucleotides used in the preparation of crosslinkable DNA were designed with regard to two factors. Firstly, the oligonucleotide sequence had to be compatible with the sticky end of the digested core part of the DNA sequence (AvaII or BsaI). Secondly, a convertible nucleotide was inserted at a position suitable for a DNA to histone H3 N-terminal cysteine crosslink. The bulk of the sequence was designed to match the native DNA sequence as closely as possible.

Oligonucleotide construct	Sequence 5'- 3'		T_m [°C]
α-satellite 147bp dG*=70	pGAC CAG GTG GAT AXT GAT	with X = convG	39.2
α-satellite 147bp dG*=69	pGAC CAG GTG GAT ATX GAT	with X = convG	39.2
601 145bp dG*=69	pCGT GTC AGA TAT ATA CAT XGA T	with X = convG	38.9
Met16 promoter+ 147bp dG*=70	pTCT ATG AGC TTT TAC TTT GTX GAG	with X = convG	54.0
Met16 promoter- 147bp dG*=70	pGCA ACT CAT AAA ATG AAG ACX TAT	with X = convG	50.0

Table 3.2: Sequences of oligonucleotides used for the production of crosslinkable DNA. X is the convertible dG* nucleotide. The melting temperature is given (calculated as nearest-neighbor method [94]).

3.2.14.   Oligonucleotide synthesis

The designed oligonucleotide sequences were synthesized at a 1 µmol scale (Microsynth). Depending on the DNA strand, some oligonucleotides were synthesized with a 5' phosphate necessary for ligation to the core DNA. The 5' end of the oligonucleotide forming the terminus of the DNA sequence was left unphosphorylated to prevent self-ligation during the oligonucleotide ligation process. The purified oligonucleotides were delivered still bound to the solid support of the synthesis column.

3.2.15.   Oligonucleotide conversion

The raw material for the production of crosslinkable oligonucleotides was 2-F-dI (Glen Research). The conversion to convertible dG* (dG containing a cystamine moiety at the N2 position) was carried

out according to a previously reported conversion strategy [95, 96]. While still attached to the solid support of the synthesis column, the unconverted, protected oligonucleotide containing the 2-F-dI was exposed to a solution of 0.5 M cystamine in acetonitrile at a pH of 9 at 22 °C. The cystamine solution was repeatedly pushed through the column by two attached 1 ml syringes. After 48 h, the cystamine solution was replaced with fresh solution. After another 48 h of incubation, the amine solution was eluted by flushing the column with DMSO and then with acetonitrile. After drying the column with argon gas, the oligonucleotides were deprotected by applying 1 ml of a 1 M DBU solution in acetonitrile twice for 60 min at 22 °C. The solid support was washed again, this time twice with methanol and thrice with acetonitrile. After drying with argon gas, the oligonucleotides on the solid support were eluted into a 15 ml polypropylene screw cap tube by flushing approximately 500 µl of 32% ammonia through the synthesis column. The deprotection reaction was carried out for 5 h at 60 °C. The oligonucleotides were then cooled down and the ammonia was evaporated in an evacuated centrifuge (Speedvac). The dried oligonucleotides were sent back to Microsynth for final purification and quality control (MALDI-TOF).
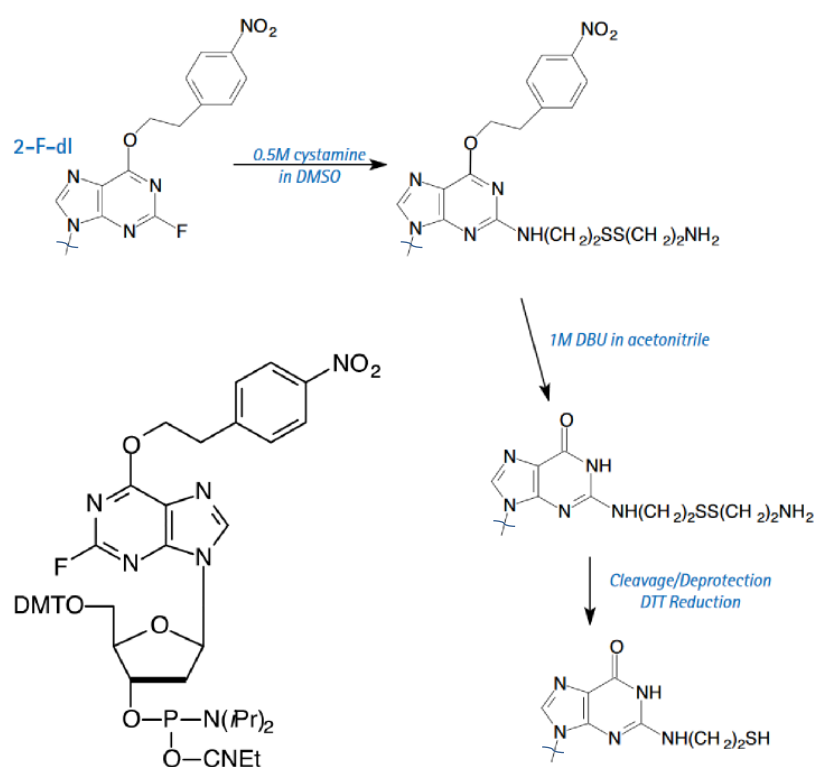
Figure 3.2: Conversion pathway from 2-F-dI to converted dG (dG*). The 2-F-dI-CE-phosphoramidite, which was used as the raw material for DNA synthesis (lower left corner), was purchased from Glen Research. The O6 moiety is protected by a nitrobenzene group and the C2 is bound by a fluorine leaving group. The phsophoramidite is protected with standard cyanoethyl (CNEt) and N,N-diisopropyl groups on the 3' end and a DMT group at the 5'-oxygen. The exchange of the 2-F group with cystamine in DMSO is followed by deprotection with 1 M DBU and subsequent cleavage and general deprotection in 32% ammonia [97]. The pathway only shows the base moieties.

3.2.16. Oligonucleotide annealing

Oligonucleotides with complementary sequences were added in equimolar amounts. 5 M NaCl stock solution was added to yield a final NaCl concentration of 100 mM. The oligonucleotide mixture was heated to 95 °C for 5 min and left to cool down to 22 °C in the heat block. When the samples reached 22 °C they were transferred to the fridge to further cool down to 4 °C. Once the oligonucleotides were annealed and cooled down to 4 ° they were kept at 4 °C for purification and storage.

3.2.17. Annealed oligonucleotide purification

Once the complementary oligonucleotides had been annealed, they were purified on a 5 ml HiTrap Q column (GE Healthcare Life Sciences). The elution gradient ran from 50 mM NaCl to 1.5 M NaCl over 25 column volumes at 2 ml/min. Peak fractions were analyzed on a 15% native PAGE gel. All the buffers including the loading dye had to be precooled to 4 °C to prevent melting of annealed oligonucleotides. Gel electrophoresis was also carried out at 4 °C.

3.2.18. Test ligations

After purification and determination of concentration, the oligonucleotides containing the convertible nucleotide were test ligated to the core DNA. Different ratios of oligonucleotide to core DNA ends were tested. Test ligations were usually done at a concentration of 2 µM DNA ends. Oligonucleotide:DNA ratios between 1:1 and 2:1 were tested. 1 µl of 1,000 U/µl T4 DNA ligase was added into a total reaction volume of 50 µl. The ligation reactions were incubated for 60 min at temperatures between 4 °C and 22 °C. The ligation reaction was analyzed by running the samples on a 15% native PAGE or a 2% agarose gel. PAGE gels were stained after the electrophoresis for 10 min in a 1x Gelred (Biotium) solution. Agarose gels were prestained by adding Gelred (Biotium) to 1x concentration into the agarose solution after heating.

3.2.19. Large scale ligations

The ligation of oligonucleotides to the core DNA was done with an excess of oligonucleotide. The minimum excess ratio was determined in a set of test ligations. A trade-off between optimal ligation conditions and wasting expensive convertible oligonucleotide had to be made. Once the optimal excess oligonucleotide ratio was determined, the core DNA was brought to a final concentration of 5 – 10 µM in 1x T4 ligase reaction buffer without DTT present in a reaction volume of typically 10 ml. Oligonucleotide was then added at the predetermined excess ratio. 50 µl of 1,000 U/µl T4 DNA ligase was added per 1 mg of core DNA. The completion of the ligation reaction was highly dependent on the incubation temperature, which in turn was dependent on the sequence of the sticky DNA ends that were ligated. Typically, an incubation temperature of 16 °C was ideal, although some oli-

gonucleotide ligation reactions would only complete at 8 °C. Incubation of the ligation reaction was mostly for 12 h. Test ligations were completed after 60 min at the desired temperature. The fully ligated product was then ethanol precipitated and resuspended in 3 ml TES300 and injected to the 60 ml 5PW DEAE ion exchange column (Tosoh Biosciences) for purification. Typical yields ranged between 1 – 3 mg of final product.

Time [min]	% Buffer B
0	0
9	20
49	65
51	100
65	100
66	0

Table 3.3: Salt gradient for the separation of ligated 147 bp DNA from unligated material and free oligonucleotide on a 60 ml 5PW DEAE column (Tosoh) with buffer A = TES300 and buffer B = TES600. Pump speed was 4 ml/min during all steps.

The peak fractions were analyzed on a 10% native PAGE gel and the fractions containing the desired DNA product were pooled and ethanol precipitated. The precipitated DNA was centrifuged and the pellet was resuspended in 10 mM Tris pH 8.0. The DNA concentration was determined spectroscopically. The complete flowchart for the production of crosslinkable DNA is shown in Fig. 3.3.
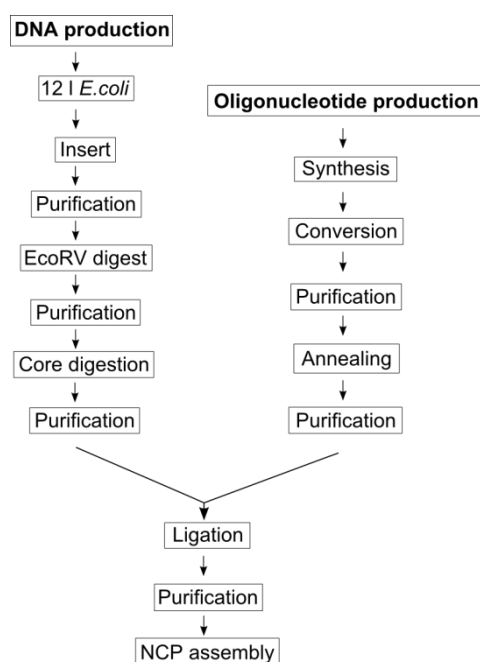
Figure 3.3: Flowchart for the production of crosslinkable DNA. The DNA was produced in *E.coli* culture. The purified core DNA and oligonucleotide were ligated to yield crosslinkable DNA which was assembled into nucleosomes for subsequent crosslinking.

3.3. Protein methods

3.3.1. Histone expression

3.3.1.1. Bacterial transformation

2 µl of product from the mutagenesis PCR reaction (section 3.2.2) were taken for the transformation of 100 µl BL21 (DE3) pLysS *E.coli* cells. The competent cells were thawed on ice and incubated for 30 min on ice after addition of plasmid DNA. The transformation was done by heat shock for 40 s at 42 °C. The cells were cooled down for 2 min on ice before 500 µl of 2x TY medium were added. The cell phenotype was allowed to recover by shaking the cells for 60 min at 37 °C. The cells were then spread on a TYE plate containing 100 µg/ml ampicillin and 33 µg/ml chloramphenicol and then incubated upside down for 12 h at 37 °C.

3.3.1.2. Test expression

A single colony from the plate (as described in section 3.3.1.1) was then used to inoculate a 100 ml 2x TY starter culture containing 100 µg/ml ampicillin and 33 µg/ml chloramphenicol for 16 - 18 h at 20 °C and 190 RPM until the OD_{600} reached 0.2 - 0.4. Induction of protein overexpression was started by addition of 0.4 M IPTG to a final concentration of 400 µM. Samples which were induced at an OD_{600} of 0.2, 0.3 and 0.4 respectively were grown for 2 h, 3 h and 4 h each. 500 µl of each induction condition were spun down for 5 min at 20 °C and 20,000 g and then resuspended in 5 µl SDS GLB and placed in boiling water for 5 min and then run on an 18% SDS PAGE gel at 250 V for 55 min.

3.3.1.3. Large scale expression

A 1 l 2x TY starter culture containing 100 µg/ml ampicillin and 33 µg/ml chloramphenicol was inoculated from the same colony as the starter cultures from the test expression (section 3.3.1.2) and incubated at 37 °C and 190 RPM until the OD600 reached 0.3 – 0.5. It was then distributed to twelve flasks each containing 1 l 2x TY with 100 µg/ml ampicillin and 33 µg/ml chloramphenicol and grown at 37 °C and 190 RPM until the OD_{600} reached 0.3 - 0.4. The histone overexpression was started by adding 0.4 M IPTG to a final concentration of 400 µM. The cells were harvested in an SLA3000 rotor for 7 min at 4 °C and 8,000 g after growing for 3 h at 37 °C and 190 RPM shaking. The cell pellets were resuspended in histone wash buffer and stored at –20 °C.

3.3.2. Histone inclusion body preparation

The frozen cells were thawed in a water bath at 37 °C. The cell suspension was sonicated for ten cycles of 60 s with 60 s pauses in between at a power output of 50 W (level 8). The homogenized cell lysate was then centrifuged at 4 °C for 30 min in an SS-34 rotor at 17,000 g. The pelleted inclusion

bodies were resuspended in 30 – 50 ml of triton wash buffer and dounce homogenized. The solution was centrifuged in an SS-34 rotor at 4 °C for 15 min at 17,000 g. The process was repeated until the supernatant after centrifugation was clear. A final step of resuspension in detergent-free buffer was carried out to remove the detergent. The pellet after centrifugation was now resuspended in 30 ml unfolding buffer and stirred for 60 min at 22 °C. The solution was again centrifuged in an SS-34 rotor at 4 °C for 15 min at 17,000 g and the supernatant containing the unfolded histones was collected.

3.3.3. Sulfitolysis of histones

1.03 g sodium sulfite and 0.51 g sodium tetrathionate were added to the supernatant of the histone inclusion body preparation (section 3.3.2) amounting to final concentrations of sulfite and tetrathionate of 300 mM and 60 mM, respectively. The S-alkylation reaction mixture was incubated for 12 - 16 h at 22 °C with constant stirring in the dark. During the reaction, the sulfur atoms of the histone cysteines form a stable disulfide to a sulfite group (SO_3^-). This protective disulfide bond was only reduced again upon refolding of histone octamer by addition of DTT.

3.3.4. Histone purification

The sulfitolysis reaction mixture was again centrifuged for 15 min at 4 °C in a SS-34 rotor at 17,000 g. The supernatant was filtered through a 0.45 µm syringe filter and then injected into an equilibrated S-200 histone column. The chromatography was run at 22 °C at a pump speed of 3 ml/min with 4 min/fraction. The A_{280} signal was recorded on a chart recorder (Kipp & Zonen). Remaining chromosomal DNA usually eluted in a first peak and was followed by the histone protein. Every third peak fraction was run on an 18% SDS PAGE. The histone peak was pooled and dialyzed against 5 L of MilliQ water with two buffer changes after 4 h each. One dialysis step was usually overnight (for 12 h). The dialyzed sample was then lyophilized and resuspended in 3 ml of RP HPLC buffer A. It was then loaded on a pre-equilibrated reverse phase chromatography column (4PW-C18, Tosoh). A 10 µl aliquot of every second peak fraction was ethanol precipitated and resuspended in 5 µl SDS PAGE GLB and then analyzed on an 18% SDS PAGE. The pure histone fractions were pooled and dialyzed against 5 L of MilliQ water at 4 °C thrice for 4 h with two buffer changes. The final sample was collected and the concentration was determined spectroscopically. Aliquots of 4.5 mg were prepared and lyophilized. A sample of the lyophilized sample was sent for MALDI analysis. The sulfite group on each cysteine adds 80 g/mol to the molecular mass. Storage of the lyophilized histones was at –80 °C.

3.3.5. Histone octamer assembly

One aliquot of each core histone (H2A, H2B, H3 and H4) was resuspended in 2 ml histone unfolding buffer and incubated for 30 min at 22 °C. The concentrations were then spectroscopically determined against unfolding buffer as reference, using the molar extinction coefficients shown in Table 3.4.

Histone	ε
wt xH2A	4,050
wt xH2B	6,070
wt xH3	4,040
wt xH4	5,400

Table 3.4: Extinction coefficients (ε) for the wt histones used for protein quantitation by UV absorption at λ = 276 nm.

The extinction coefficient of H3 cysteine point mutations was modified to account for the change in ε and the total extinction coefficient was calculated as the sum of the individual coefficients. A pipetting scheme was calculated to ensure the correct stoichiometry of all four core histones. Since separating refolded octamer from excess H2A-H2B dimer is easier than separating it from excess (H3-H4)$_2$ tetramer, an excess of H2A-H2B dimer is used in the octamer assembly reaction. Effectively, the molar ratios of the core histones upon assembly were

$$H2A : H2B : H3 : H4 \ = \ 1 : 1 : 0.95 : 0.95.$$

The histone mix was then filled into an equilibrated dialysis bag with an MWCO of 6,000 – 8,000 Da and placed in 2 L of histone octamer refolding buffer. The dialysis to refolding buffer was done for 4 h thrice with buffer changes in between. Upon successful assembly of histone octamer, the sample was harvested and spun down for 15 min at 4 °C and 4,000 g. The supernatant was transferred to a concentrator tube with an MWCO of 10,000 Da and concentrated to approximately 1 ml. The concentrated histone octamer was centrifuged again for 15 min at 4 °C and 18,000 g and then injected to an equilibrated 120 ml S-200 column (GE Healthcare). The chromatography run was done with the parameters given in Table 3.5.

Run parameter	Value
Pump speed	1 ml/min
Fraction collection	2 min/fraction
Detector wavelength	276 nm
Detector sensitivity	0.1 mV (full-scale)

Table 3.5: Run parameters used for the separation of refolded histone octamer on a 120 ml S-200 column (GE Healthcare).

Typically, a small peak of aggregated histone eluted first and was followed by the main histone oc-tamer peak. Another peak of free H2A-H2B dimer usually followed the main peak. 5 µl of each peak fraction was mixed with 5 µl SDS PAGE GLB and incubated for 5 min in boiling water before being run on an 18% SDS PAGE. Pure histone octamer fractions were pooled and concentrated using a concen-trator with an MWCO of 10,000 Da. The histone octamer concentration was spectroscopically meas-ured against flow-through from the concentrator tube and then stored at 4 °C.

3.3.6. ΔDot1 methods

3.3.6.1. Bacterial transformation

1 µl of pET28a vector containing the ΔDot1S542C gene was taken for the transformation of 100 µl BL21 (DE3) Codon Plus RIPL *E.coli* cells. The competent cells were thawed on ice and incubated for 30 min on ice after addition of plasmid DNA. The transformation was done by heat shock for 25 s at 42 °C. The cells were cooled down for 2 min on ice before 500 µl of 2x TY medium were added. The cell phenotype was allowed to recover by shaking the cells for 60 min at 37 °C. The cells were then spread on a TYE plate containing 50 µg/ml kanamycin and 33 µg/ml chloramphenicol and then incu-bated upside down for 12 h at 37 °C.

3.3.6.2. ΔDot1 S542C expression

A single colony containing a clone carrying the ΔDot1S542C gene was transferred into a 1 l starter culture of 2x TY medium containing 50 µg/ml kanamycin and 33 µg/ml chloramphenicol and was incubated for 16 – 18 h at 23 °C and 195 RPM in the shaking incubator. Upon reaching an OD_{600} of 0.2 the starter culture was distributed to 10 x 1l 2xTY, containing 50 µg/ml kanamycin and 33 µg/ml chloramphenicol and grown to an OD_{600} of 0.6. At this point the protein overexpression was induced by adding IPTG to a final concentration of 1 mM. Protein overexpression was done for 12 h at 22 °C and 195 RPM in the shaking incubator. Cells were harvested by centrifugation in an SLA-3000 rotor at 8,000 g and 4 °C for 6 min. The cell pellets were resuspended in ΔDot1 wash buffer and stored at −80 °C.

3.3.6.3. ΔDot1 S542C purification

Frozen cell pellets were thawed in a water bath and spun down for 5 min at 4,000 g and 4 °C. The supernatant was discarded and the pellets were resuspended in ten pellet volumes of ΔDot1 lysis buffer. The resuspended cells were lyzed by sonication for 10x 60 s in the ultrasound homogenizer (Janke & Kunkel AG) at approximately 50 W (level 8) while being immersed in an ice bath. Immedi-ately after sonication benzamidine and PMSF were added to final concentrations of 1 mM each, to

suppress protease activity. The cell lysate was centrifuged for 45 min at 18,000 g and 4 °C in an SS-34 rotor. The cleared lysate was then loaded onto a 5 ml Ni-NTA cartridge column at a pump speed of < 1 ml/min. After one full loading cycle, the flow-through was continuously cycled through the column again at 3 ml/min. Unspecifically bound protein was washed off by running seven column volumes of wash buffer containing 25 mM imidazole through the column. The loaded Ni-NTA column was then attached to an equilibrated AEKTA system on which the elution of ΔDot1 was done by running an imidazole gradient from 25 mM imidazole to 250 mM imidazole over twelve column volumes at 1 ml/min pump speed. Fractions were collected over 90 s each and the peak fractions were immediately brought to 1 mM S-adenosyl homocysteine (SAH) to prevent ΔDot1 precipitation. Fractions containing the overexpressed ΔDot1 were pooled and concentrated in an Amicon Ultra-15 concentrator (10,000 Da MWCO, Millipore) and then buffer exchanged to ΔDot1 buffer C using the same concentrator. The sample was then loaded onto a HiTrap Q HP cartridge column (GE Healthcare) which was connected to a HiTrap S HP cartridge column (GE Healthcare) using a peristaltic pump set to 1 ml/min pump speed. The HiTrap Q HP column acted as a trap for anionic impurities in the sample. During loading, the protein sample was placed on ice. The HiTrap Q HP column was removed and the fully loaded HiTrap S HP column was then attached to an equilibrated AEKTA system where it was washed with three column volumes of ΔDot1 buffer C. Elution of ΔDot1 was achieved by running a gradient from 200 mM NaCl (ΔDot1 buffer C) to 1.5 M NaCl (ΔDot1 buffer D) over ten column volumes at 1 ml/min pump speed. Peak fractions containing ΔDot1 were pooled and concentrated using an Amicon Ultra-15 concentrator (10,000 MWCO, Millipore). The sample was buffer exchanged to ΔDot1 buffer E using the same concentrator and then injected into a 120 ml S-200 column equilibrated to ΔDot1 buffer E. Elution from the S-200 size exclusion column was done at 1 ml/min and 2 ml/fraction over 120 min. Peak fractions were pooled and concentrated and SAH was added to a final concentration of 1 mM. The sample was stored at 4 °C under argon gas for short term storage or used immediately.

3.4. Nucleosome Core Particle (NCP) methods

3.4.1. Nucleosome assembly

The NCPs were assembled from folded histone octamer and free DNA by dialyzing the sample from a high salt starting buffer (TEK1.4 for α-satellite-based DNA constructs and TEK2.0 for 601-based DNA constructs) into a final buffer containing 10 mM potassium chloride (TEK10). The histone octamer is only soluble at low salt concentrations. The interactions of the histone octamer with the corresponding DNA are effectively shielded by high salt concentrations so that during the reduction of salt con-

centration the interactions between the histone octamer and the DNA can slowly establish. The protocol used adheres to those previously published [61].

Initially, a set of ratios of histone octamer:DNA was titrated to determine the optimal ratio for getting the least amount of free DNA (lower ratios) and losing material due to precipitation (higher ratios). Typically, ratios of 0.8 to 1.2 in increments of 0.1 were set up for assembly. DNA was added to the assembly mix to a final concentration of 3 - 5 µM and the amount of histone octamer was calculated accordingly. The starting concentration of potassium chloride in the assembly mix depends on the DNA construct and was 2 M for palindromic α-satellite constructs and 2.5 M for 601-based constructs. DTT was added to a final concentration of 5 mM. The assembly buffer consisted of 20 mM Tris/HCl pH 7.5 and it contained 1 mM EDTA. The dialysis schemes are shown in Table 3.6. Since 601-based NCPs have a considerably higher salt stability with the half point of dissociation approximately 30% higher than for the α-satellite based NCPs, the starting salt concentration is chosen higher accordingly [34].

α-satellite based DNA	
Time [min]	[KCl]
30	2.0 M
70	1.4 M
90	1.2 M
600	1.0 M
90	0.8 M
90	0.6 M
Indefinitely	0.01 M

601-based DNA	
Time [min]	[KCl]
180	2.5 M
30	2.0 M
70	1.4 M
90	1.2 M
600	1.0 M
90	0.8 M
90	0.6 M
Indefinitely	0.01 M

Table 3.6: Time courses of the stepwise salt dialysis for assembly of NCPs from histone octamer and DNA (left table for α-satellite-based NCPs and right table for 601-based NCPs).

To ensure completion of dialysis, the dialysis buffer:sample ratio was at least 100:1. Assembly took place at 4 °C. After completion of dialysis, the samples were centrifuged shortly at maximum speed to pellet any precipitate that occurred during assembly. NCP samples containing α-satellite based DNA sequences had to be heat shifted for 120 - 180 min at 37 °C to establish a single translational position [61]. 601-based DNA constructs were uniformly positioned directly after salt dialysis and did not require heat shifting.

3.4.2. Nucleosome core particle crosslinking

Crosslinking of the nucleosome DNA to the histone octamer was achieved by reacting the convertible nucleotide to the cysteine point mutation on the H3 N-terminal tail. To stabilize the NCP even

further, a second crosslink between the two copies of H2A N38C can be introduced. This double crosslinked NCP was also used in the production of crosslinked ΔDot1 S542C to NCP complex.

3.4.2.1. DNA:H3 crosslinking

The assembly of crosslinked NCPs is basically identical to the NCP assembly described in section 3.4.1., except the histone octamer used for NCP reconstitution had two copies of H3 R40C and the TEK10 buffer had to be DTT-free and degassed thoroughly. Sodium tetrathionate was added as oxidizing agent to the assembled NCPs to a final concentration of 4 mM. The crosslinking reaction took place at 50 °C for 3 h during the heat shifting of the NCPs. Crosslinked NCPs could subsequently be purified as described in section 3.4.3.

3.4.2.2. Double crosslinked nucleosome core particle

The stability of the NCP can be increased further by introducing a second crosslink between the two copies of H2A. The octamer used for assembly additionally contained the H2A N38C point mutation. The H2A N38C crosslink has been shown to increase the stability of the octamer during reconstitutions of chromatin higher order structures [1]. The assembly and purification of histone octamer containing H2A N38C was essentially as described in section 3.3.5. The crosslinking reaction and purification for the doubly crosslinked nucleosome was done as described in sections 3.4.2 and 3.4.3 respectively.

3.4.2.3. Crosslinking ΔDot1 S542C to the nucleosome core particle

To stabilize the complex of ΔDot1 to the NCP, a nucleosome construct with multiple crosslinking sites was prepared. For crosslinking of the nucleosome DNA to the histone octamer, an R40C mutation on histone H3 was used. To covalently bind ΔDot1 S542C to the histone octamer, a previously described crosslink via H3 T80C was used (V. Vogirala, PhD thesis [2]). Histone octamer containing both crosslinking sites (H3 R40C and H3 T80C) was prepared as described in section 3.3.5. NCPs were assembled as described in section 3.4.1. Great care had to be taken to prevent air oxidation of the free sulfhydryls on the NCP. ΔDot1S542C samples were filled into dialysis bags (10,000 Da MWCO) and dialyzed into degassed TEK10 (pH 8.8) containing 1 µM SAH for 12 h with two buffer changes after a minimum of 4 h. Thorough degassing was essential to prevent air oxidation of the sulfhydryl groups on the ΔDot1 S542C point mutation. The dialysis of the ΔDot1 samples was done in a 50 ml polypropylene tube with argon gas layered on top. The 50 ml tube was placed in an evacuated desiccator at 4 °C.

NCPs containing H3 R40C and T80C were mixed with different ratios of ΔDot1 S542C at a fixed NCP concentration of 1 μM in a volume of 30 μl. The potassium chloride concentration was brought to 50 mM and the buffer was 100 mM Tris/HCl pH 8.8. SAH was added to 1 μM final concentration acting as cofactor for ΔDot1. Then, sodium tetrathionate was added to a final concentration of 4 mM to start the crosslinking reaction. Test crosslinking reactions were incubated for 60 min at 22 °C. 10 μl of sample per ratio were analyzed on a 5% polyacrylamide EMSA gel (section 3.1.4). Once the optimal ratio of ΔDot1 S542C:NCP was determined, the sample volume was scaled up to produce sample amounts sufficient for crystallization. After addition of sodium tetrathionate, the scaled-up reaction was incubated for 12 h at 4 °C.

3.4.3. Nucleosome core particle purification

The reconstituted NCP samples were purified on a 5PW DEAE ion-exchange column (Tosoh Biosciences). The NCP samples were filtered through a 0.22 μm syringe filter and then loaded onto an equilibrated ion exchange column (5PW DEAE, Tosoh Biosciences). The elution from the column was done by running the salt gradient shown in Table 3.7.

Time [min]	% Buffer B
0	0
8	0
43	100
57	100
58	0

Table 3.7: Salt gradient for purification of crosslinked NCPs on a 60 ml 5PW-DEAE column (Tosoh Biosciences) with buffer A = TEK250 and buffer B = TEK600. Pump speed was 4 ml/min during all steps.

Peak fractions were immediately pooled and concentrated in an Amicon concentrator (10,000 Da MWCO, Millipore) and then buffer exchanged to TEK10 by adding 10 ml TEK10 to approx. 500 μl of concentrated sample. The buffer exchange was repeated twice and then the sample was further concentrated in a vivaspin500 concentrator (5 kDa MWCO, Sartorius) to a final concentration of 6 – 8 mg/ml NCP.

3.4.4. Purification of crosslinked ΔDot1 S542C:NCP complex

The crosslinked sample was centrifuged for 15 min at 4,000 g at 4 °C to pellet any precipitate. The cleared crosslinked sample in the supernatant was then concentrated in an Amicon Ultra-15 concentrator (10,000 MWCO, Millipore) to a sample volume of approximately 500 μl. 25% sucrose was added to a final concentration of 5%. Purification of crosslinked ΔDot1:NCP complex was carried out on a Model 491 preparative gel, as described in section 3.1.5.

3.5. Crystallization methods

3.5.1. Crystallization drop setup

Crystallization experiments were generally conducted as sitting drop vapor diffusion experiments in 24 well polystyrene plates (Cryschem, Hampton Research). Crystallization setups were topped with argon gas and sealed with clear tape (Manco Inc.) before being placed in a nitrogen atmosphere at 22 °C. Exclusion of air oxygen was important to prevent oxidation of possibly unreacted sulfhydryls in the samples. Regulation of equilibration dynamics between drop and reservoir solution was achieved by applying a mixture of silicon oil and paraffin oil on top of the drop and/or the reservoir solution. Mostly a mixture of equal amounts of paraffin oil and silicon oil (Al's oil) was used. NCP samples were set up in crystallization conditions which were used before to crystallize NCPs [9]. The standard crystallization screen for NCPs ranges in $MnCl_2$ concentration from 130 mM to 180 mM in increments of 10 mM and the KCl concentration ranges from 100 mM to 160 mM in increments of 20 mM while all screen conditions contains 40 mM of sodium cacodylate (pH 6.0). Sample drops were set up by mixing 1 µl of sample containing 6 – 8 mg/ml NCP with 1 µl of screen solution. 200 µl of screen solution were then diluted fourfold with purified water. The ionic strength in the sample drop decreases slowly over the course of several days establishing a reversed salting-in transition for the NCP crystals to grow in. Crystal growth was typically complete within two weeks.

3.5.2. Post crystallization treatment

NCP crystals were dehydrated after having fully matured. The sealed wells containing the crystals were opened and the crystals were transferred to a drop of harvest liquid (HL-2) in a glass depression dish using a glass capillary attached to a syringe filled with oil. The crystals were incubated in the harvest liquid for 10 – 30 min before increasing amounts of soaking solution containing 40% MPD (HL-40) were added. The soaking process followed the scheme shown in table Table 3.8.

Time [min]	40% MPD added	[MPD]$_{final}$
0	0 µl	2%
3	5 µl	5.5%
6	7 µl	9.5%
9	11 µl	14.1%
12	20 µl	19.6%
15	21 µl	23.4%

Table 3.8: Time course of the post-crystallization treatment. The initial volume of 50 µl HL-2 was supplemented with the indicated (column 2) amount of HL-40 to yield the final MPD concentration indicated in the third column.

After 30 min of incubation in the final soaking solution now containing 23.4% MPD, the crystals were transferred into a fresh glass depression dish with 50 µl of soaking solution containing 24% MPD and

2% trehalose. Overnight the crystals were allowed to heal any micro-cracks which might occur during the dehydration procedure. On the following day, the crystals were collected with a microloop (MiTeGen) of appropriate size and then plunged into liquid propane at -120 °C. Crystals were stored at liquid nitrogen temperature.

3.6. Diffraction data collection and computing methods

3.6.1. Data collection

X-ray diffraction data was collected at the SLS synchrotron at the Paul-Scherrer institute in Vilingen, Switzerland. Datasets were collected under cryogenic conditions either on beamline X06SA (PX I) equipped with a Pilates 6M-F detector or on beamline X06DA (PXIII) equipped with a Pilates 2M-F detector. Several crystals were screened to check the diffraction quality and full non-overlapping datasets were collected over multiple well-diffracting crystals. A dataset was typically acquired over several crystals. Several positions on each crystal were exposed to the unattenuated X-ray beam to a total dose of approximately 10 MGy in rotational increments of 0.1° - 0.3°. Initial X-ray diffraction analysis of Met16 NCPs was carried out in-house under cryogenic temperatures using a MicroMax-007HF X-ray generator generating X-rays of 1.5418 Å wavelength (Rigaku). The X-ray diffraction was recorded on a MAR345 image plate detector (MAR Research) mounted on a MAR DTB Base.

3.6.2. Indexing

Reflections from a dataset were indexed using the xdsHelper suite (T. J. Richmond) which uses xds [98, 99]. Individual framesets (positions on a crystal) were assigned individual resolution cutoffs based on the diffraction quality. Merging of datasets from different crystals was also carried out in the xdsHelper suite using scala (CCP4).

3.6.3. Phasing

The crosslinked NCPs were presumed to be mostly identical to the previously solved structure of the nucleosome core at 1.94 Å (PDB ID: 1KX5, [11]). The phases could therefore be obtained by molecular replacement. The CCP4 Phaser module was used for phasing by molecular replacement with the 1KX5 structure as a search model [100].

3.6.4. Refinement and model building

Refinement of the crosslinked nucleosome structures was done in Phenix [101]. A first round of rigid body refinement and bulk solvent correction was followed by five macrocycles of real space refinement, simulated annealing (cartesian), atomic displacement parameter (ADP) refinement and occupancy refinement of the sulfur atoms comprising the disulfide bonds in the structure. Manual rebuilding of parts of the model was done in Coot [102]. The convertible dG* nucleotide was introduced into the DNA model by specifying a *pdb* file and an *mmCIF* file for the convertible base which were derived from the structure of a crosslinked oligonucleotide [103]. The disulfide crosslink between the modified base and the protein sulfhydryl was specified in the *Phenix.refine* file. After the first stage of refinement, ions and ordered water molecules were placed in regions of positive density in the F_o-F_c map. The unresolved histone tails were deleted from the model and a final round of five refinement macrocycles in Phenix was performed.

3.6.5. Validation

After completed refinement, the resulting structural model was validated using Phenix [101]. The model quality is assessed by real-space correlation and geometry inspection using MolProbity tools [104]. The structures of the crosslinked NCP and the double-crosslinked NCP were also compared to the existing high resolution NCP structure (PDB ID: 1KX5) using Chimera [105]. Due to the palindromic nature of the DNA and the resulting pseudo two-fold symmetry, the two pseudo-symmetric halves of the NCP were compared to one another.

4. Crosslinked nucleosomes

4.1. DNA crosslink to the N-terminal tail of histone H3

4.1.1. Introduction

Most of the DNA in chromatin is usually tightly wrapped around histone octamers forming a chain of nucleosomes. The nucleosome DNA is not covalently bound to the histone octamer, but many interactions, electrostatic as well as hydrophobic, between the histones and the nucleosome DNA lead to the formation of a stable complex. Most of the contacts between the histones and the DNA are sequence non-specific and are formed between the histones and the DNA sugar-phosphate backbone [9]. The DNA sequence does, however, play an important role in nucleosome formation since the anisotropic bendability is determined by the DNA sequence [106]. Well positioning sequences show a strong preference for base pair steps that are easily bent into the minor groove at positions where the minor groove faces the histone octamer. These sequence preferences can be found at intervals of 10.2 bp and minimize the energy penalty resulting from the strong bending around the histone octamer core [9]. The binding of the DNA molecule is enhanced by the insertion of the positively charged guanidinium moieties of arginine amino acids at all of the 14 positions where the DNA minor groove faces the histone octamer core.

The formation of nucleosomes that are stable and homogenous enough to form well-diffracting crystals has limited the number of solved NCP X-ray crystal structures to those containing strong positioning sequences, such as the 601 sequence or α-satellite DNA or versions thereof [9, 57]. The recently determined MMTV-A promoter NCP also contains a sequence that can be considered a well positioning sequence [35].

The first major goal of this work was to establish a method to covalently crosslink nucleosome DNA to the histone octamer, thereby overcoming the problem of DNA dissociation under various conditions such as low complex concentration or high salt conditions. Another major advantage of a working protocol for DNA crosslinking to the histone octamer is the possibility of enlarging the sample size of DNA sequences that can be crystallized on an NCP and yield well-diffracting crystals. Sample inhomogeneities due to dispersed translational positioning of the DNA will not lead to high resolution diffraction. A crosslink restrains the translation of the DNA around the histone octamer and should yield stable and homogenous NCP samples. The termini of the DNA were a suitable location for a DNA-to-histone crosslink so that DNA breathing at the ends is constrained. The location focused on were the regions where the H3 N-terminal tails pass through the minor groove channels

formed at SHLs 7/-1 and SHLs -7/1. At those locations, amino acids 38 – 41 of H3 come into close proximity to base pairs 68 – 71 nt away from the dyad base pair (Fig. 4.1).

The method chosen for establishing the nucleosome-to-histone crosslink was disulfide bond formation between the thiol group of a cysteine point mutation engineered into the H3 N-terminal tail and a convertible nucleotide introduced into the nucleosome DNA via ligation of a modified oligonucleotide at the ends of the DNA. The convertible nucleotide approach is an established methodology for crosslinking of DNA to DNA binding proteins [107]. A variety of convertible nucleotides is now available for the crosslinking of different DNA bases, although the non-native thiol groups introduced differ in their ability to be crosslinked to proteins that bind to the minor groove or major groove. For the crosslinking of the H3 N-terminal tail to the DNA, a convertible dG* was introduced into the nucleosome DNA sequence that can form a crosslink in the minor groove via a thioethyl tether at the N2 position. Previous work on crosslinking of minor groove binding proteins to DNA containing convertible dG* bases shows the feasibility of the approach and several X-ray crystal structures have been solved to date by the implementation of the convertible nucleotide approach (Table 4.1).
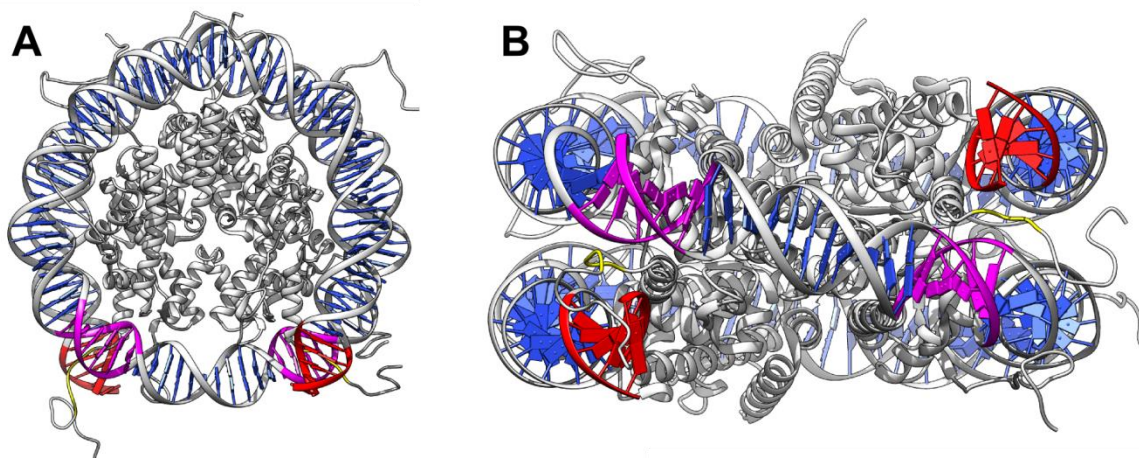
Figure 4.1: Selected position for the DNA to Histone H3 crosslink**. A)** The part of the H3 N-terminal tail (yellow) that passes through the minor groove channels formed by SHL +/-1 (magenta) and the DNA termini (SHL +/-7, red) seen from a top view. **B)** Side view looking at the DNA dyad axis (PDB ID: 1KX5).

Structure	Convertible nucleotide	PDB ID	Ref.
Covalently Trapped Catalytic Complex of HIV-1 Reverse Transcriptase	Convertible dG	1RTD	[108]
8-oxoguanine DNA glycosylase I (hOGG1) covalently bound to target DNA	Convertible dC	1YQK-L 1YQR	[109]
Dimeric p53-DNA Complex	Convertible dC	3EXJ	[110]
DNA decamer with an engineered crosslink in the minor groove	Convertible dG	1CW9	[111]

Table 4.1: List of structures that were solved by implementation of crosslinking via a convertible nucleotide. The second column indicates the kind of convertible nucleotide used. The third and fourth columns give the PDB identifier and the literature reference respectively.

4.1.2. Optimization of crosslink location

Different approaches to crosslinking the nucleosome DNA to the histone octamer have been tried by previous lab members in the Richmond group. Phosphorothionates and internal DNA backbone dithiols proved unsuitable in crosslinking attempts with amino acid point mutations H39C, R40C, Y41C, R42C and P43C on the H3 tail. Convertible dG* was introduced at positions +/-65, +/-67, +/-69 and +/-73 from the dyad axis and various crosslinking conditions were tested. The best results were previously achieved by crosslinking convertible dG* at position 69 from the dyad to R40C (Approximately 50% crosslinking efficiency, unpublished data). In order to increase the crosslinking efficiency, optimization of the location of the crosslink was attempted by scrutinizing the path of the H3 tail in the high resolution structure of the NCP containing 147 bp α-satellite DNA (PDB ID: 1KX5). The upper limit for the N2 (dG*) – S (dG*) distance was measured to be 5.38 Å based on the molecular topology of the dG* base (Fig. 4.2). Adding 2.05 Å for the mean distance of a standard disulfide bond results in approximately 7.5 Å as the upper bound for the cysteine sulfhydryl to N2 (dG*) distance.
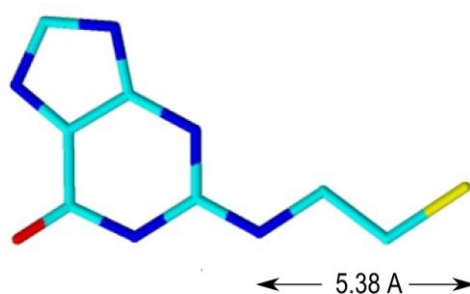
$\longleftarrow$ 5.38 Å $\longrightarrow$

Figure 4.2: The distance between the N2 position of convertible dG* and the sulfur atom of its thioethyl linker. In the maximally extended conformation it is 5.38 Å. The drawing and calculation was done in ChemSketch (ACD/Labs, Version 2.0). Only the base moiety is shown.

Examination of an existing structure of an oligonucleotide complex containing a crosslink between two adjacent convertible dG* bases yielded N2 – S distances of 3.8 Å – 5.0 Å (PDB ID: 1CW9 [103]). Distances between the N2 atoms of guanine bases model built into positions 69 and 70 and the

sulfhydryls of cysteine point mutations H39C, R40C were measured in the high resolution NCP struc-
ture (Fig. 4.3) [11]. They range from 6.3 Å for the H39C to dG*(J70) crosslink to 8.4 Å for the H39C to
dG*69.J crosslink (Table 4.2).

Residue 1 - Atom	Residue 2 - Atom	Distance [Å]
dG*69.J - N2	H3 H39C - S	6.3 - 8.4
dG*69.J - N2	H3 R40C - S	5.7 - 6.9
dG*70.J - N2	H3 H39C - S	4.9 - 7.3
dG*70.J - N2	H3 R40C - S	4.7 - 6.5

Table 4.2: Distances measured between the N2 of the introduced dG* base at positions 69 and 70 on the J-strand to cyste-
ine point mutations on the H3 N-terminal tail at positions H39C and R40C in the high resolution NCP structure (PDB ID:
1KX5 **[10]**). The distances between N2 and the sulfur atom of the cysteine point mutations are given as a range for all pos-
sible side chain rotamers of the cysteine. R40C was modeled as a flipped peptide since the arginine side chain points in the
opposite direction in the model structure (PDB ID: 1KX5).

In the configuration that the H3 tail adopts in the high resolution 147 bp α-satellite NCP structure,
the side chain of amino acid R40 points away from base 70 in the J-strand and measures 12.4 Å from
the R40C sulfhydryl to dG*70.J. R40C was still screened for crosslinking since the R40C sulfhydryl
would be in the appropriate distance range for a disulfide crosslink if the peptide backbone adopted
a flipped conformation at position 40. This seemed to be a reasonable assumption given previous
crosslinking data for R40C to dG*69.I/J (data not shown). The flipped peptide at position 40 was
modeled and the S(CYS) − N2(dG*) distances were determined to be between 4.7 Å and 6.5 Å to
dG*70 on the J-strand and between 5.7 Å and 6.9 Å to dG*69 on the J-strand. All allowed cysteine
rotamer conformations were analyzed. Based on the distances observed in other published struc-
tures, the combination of H3 R40C and dG*70.J seemed like the most promising position for a his-
tone:DNA crosslink.
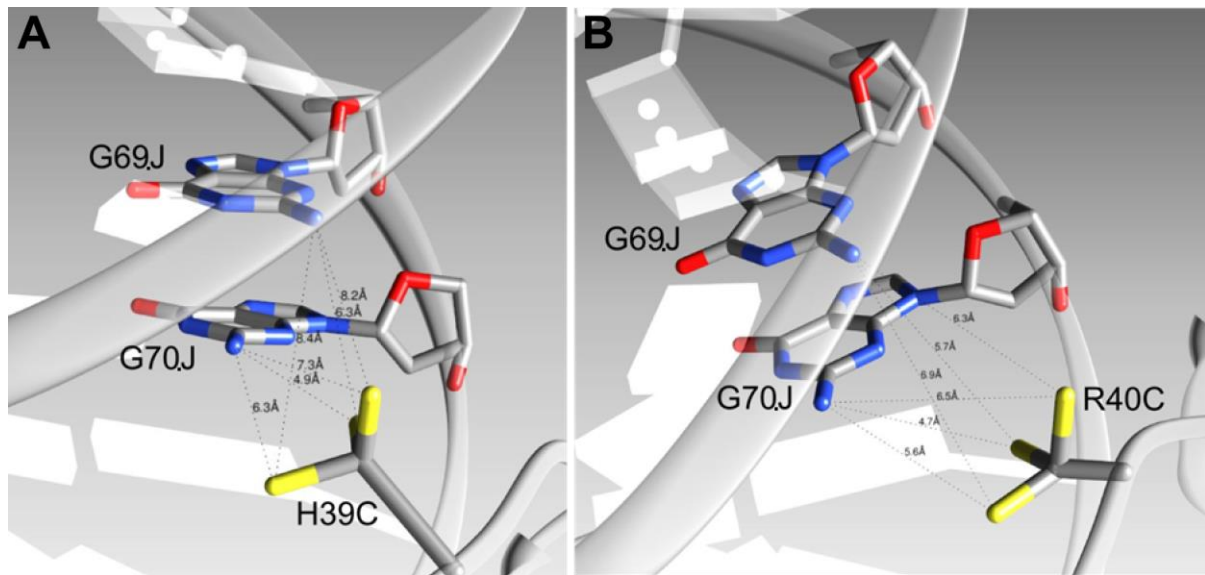
Figure 4.3: Distances measured between the N2 of the introduced dG* base at positions 69 and 70 on the J-strand to cysteine point mutations on the H3 N-terminal tail at positions H39C and R40C in the high resolution NCP structure. **A)** The distances between N2 and the sulfur atom of the cysteine in H39C are given as a range for all possible side chain rotamers of the cysteine. **B)** R40C was modeled as a flipped peptide since the arginine side chain points in the opposite direction in the model structure (PDB ID: 1KX5).

4.1.3. Cloning and overexpression of histones with cysteine point mutations

The H3 H39C/C110A, H3 R40C/C110A and H3 R40C/T80C/C110A histones were cloned by site-directed mutagenesis of the wild type histone plasmids (pET3a based) as described in section 3.2.2 (for primer sequences refer to appendix 8.5). For the H3 R40C/T80C/C110A point mutation, the single mutation plasmids were taken as templates. Plasmids carrying the histone point mutation gene were transformed into competent cells (*E.coli* BL21 (DE3) TOP10) (section 3.2.4). The sequence of the histone containing the cysteine point mutation was verified by plasmid sequencing using T7 sequencing primers. The expression vector *E.coli* BL21 (DE3) pLysS was used for overexpression of the histone. The inclusion bodies formed by the overexpressed histones were extracted and purified as described in section 3.3.2. After sulfitolysis (section 3.3.3), the unfolded histones were purified over a Sephacryl S-200 column (XK 50/1000, Pharmacia). Sulfitolysis of the cysteine amino acids was essential for protection against air oxidation during long-term storage and thereby improved crosslinking efficiency.
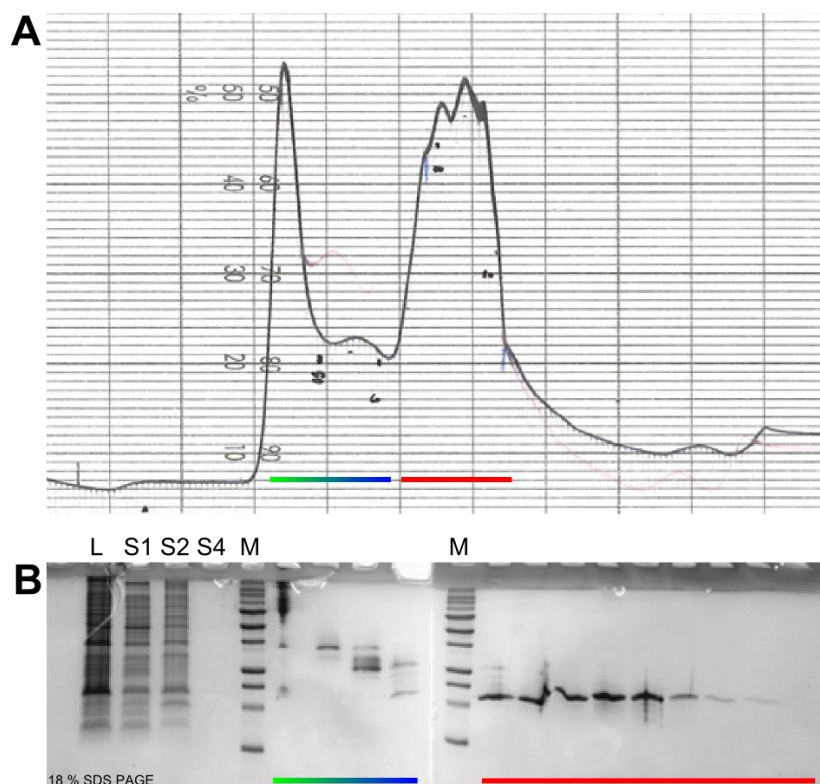
Figure 4.4: Sephacryl S-200 purification of histone protein. **A)** UV trace (λ = 280 nm) of the histones from an inclusion body preparation separated on a sephacryl S-200 column (top). The early fractions contain DNA impurities as well as non-histone proteins (green-blue gradient line). Histone protein eluted late as relatively pure sample (red line). **B)** Peak samples were analyzed on an 18% SDS PAGE (bottom) along with the cleared lysate (L) and samples from the supernatants (S1-S4) from the inclusion body wash steps. Dual color protein standard was loaded as reference (M).

The histone-containing fractions were pooled, dialyzed against water and lyophilized as described in section 3.3.4. They were further purified on a reversed-phase chromatography column (4PW-C18, Tosoh Biosciences)(Fig. 4.5). Pure histone fractions were pooled and dialyzed against water and then lyophilized in aliquots. The aliquots were stored at -80 °C until use.
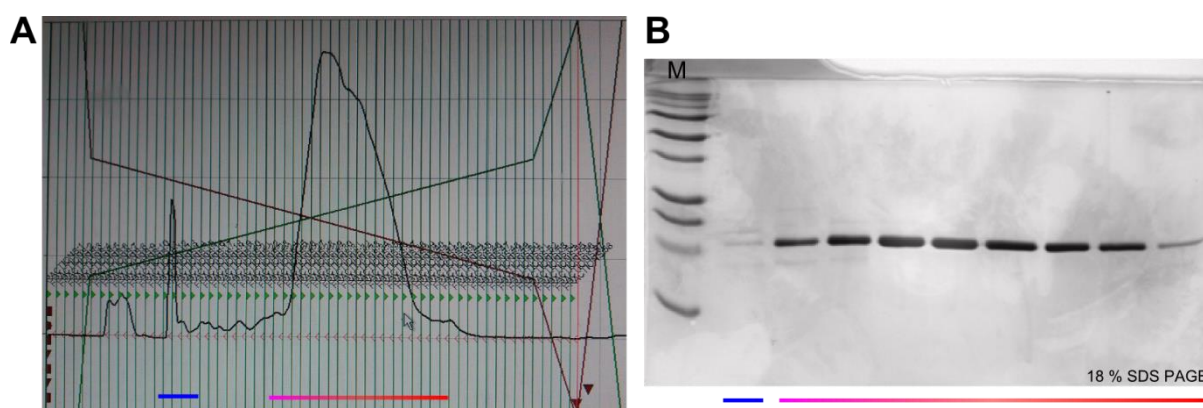
Figure 4.5: Reversed-phase purification of histone protein. **A)** UV trace (λ = 280 nm) of the histone purification (left) on a 4PW-C18 RP column (Tosoh Biosciences). The elution gradient runs from 35% buffer A (5% acetonitrile, 0.3% TFA) to 65% buffer B (95% acetonitrile, 0.3% TFA). Protein and DNA impurities eluted first (blue line) followed by the main histone peak (pink-red gradient line). **B)** Peak samples were analyzed on an 18% SDS PAGE.

4.1.4. Production of a crosslinkable histone octamer

Crosslinkable histone octamer was assembled with the combinations of histone proteins listed in Table 4.3. Those with cysteine point mutations were sulfonated (section 3.3.3) and assembled under reducing conditions to create free sulfhydryls. Purification of the folded histone octamers on a 120 ml S-200 column was also carried out under reducing conditions (5 mM DTT). Histone octamers with cysteine point mutations exhibit the same elution behavior as the wild type octamer, as for example shown for the H3 R40C/C110A histone octamer (Fig. 4.6).

Point mutants	Crosslink	H2A	H2B	H3	H4
H3 R40C	DNA:H3 R40C	wt	wt	R40C/C110A	wt
H3 R40C H2A N38C	DNA:H3 R40C+ H2A N38C	N38C	wt	R40C/C110A	wt
H3 T80C	ΔDot1:H3 T80C	wt	wt	T80C/C110A	wt
H3 T80C/R40C	ΔDot1:H3 T80C+ DNA:H3 R40C	wt	wt	R40C/T80C/ C110A	wt
H3 T80C/R40C H2A N38C	ΔDot1:H3 T80C+ DNA:H3 R40C+ H2A N38C	N38C	wt	R40C/T80C/ C110A	wt

Table 4.3: Combinations of histones used to assemble the various histone octamer versions for different crosslinked versions of the NCP.
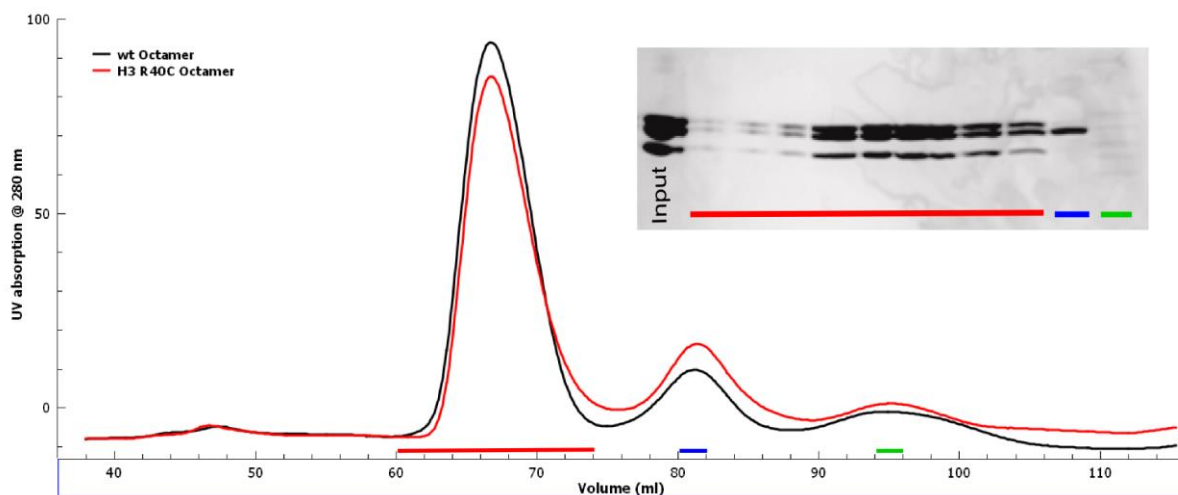
Figure 4.6: Elution from an S-200 column of histone octamer carrying cysteine point mutations (red) compared to wild type octamer (black). The retention times were unchanged, whereas the relative intensities varied. After a very small aggregate peak at 45 ml elution volume, the main histone octamer peak eluted (red lanes) between 60 – 74 ml, followed by free H2A/H2B heterodimer (blue lane, ca. 80 ml elution volume) and free histones (green lane, > 90 ml elution volume). The fractions were visualized on an 18% SDS PAGE (insert). The input material was loaded as reference.

The content of the elution peaks was examined by SDS PAGE with subsequent Coomassie staining (section 3.1.2). The peak fractions of the histone octamer containing H3 H39C/C110A are shown as an example in Fig. 4.6. The peak fractions containing the histone octamer were pooled and concen-

trated (Amicon Ultra-4, 10,000 Da MWCO, Millipore) to a volume of approximately 1 ml, typically yielding concentrations between 50 - 100 µM.

4.1.5. Production of crosslinkable oligonucleotides

The oligonucleotides containing 2-F-dI at the desired locations were purchased from Microsynth and were converted as described in section 3.2.15. After conversion and PAGE purification, the oligonucleotides underwent MALDI-TOF quality control.
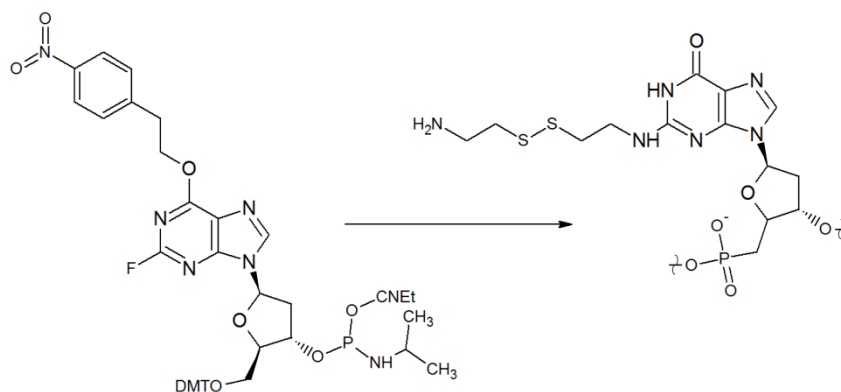
Figure 4.7: Starting material for the production of crosslinkable nucleotides. The 2-F-dI CE phosphoramidite (left) has a fluorine atom as a leaving group at the N2-position. The phsophoramidite is protected by capping groups for oligonucleotide synthesis under standard conditions (nitrophenyl-ethyl capping group at O6, DMTO at 3' and CNEt and N(i-Pr)$_2$ at 5'). The 2-F was substituted by a cystamine moiety and then the nucleotide was deprotected under standard deprotection conditions (right) leading to convertible dG (dG*).
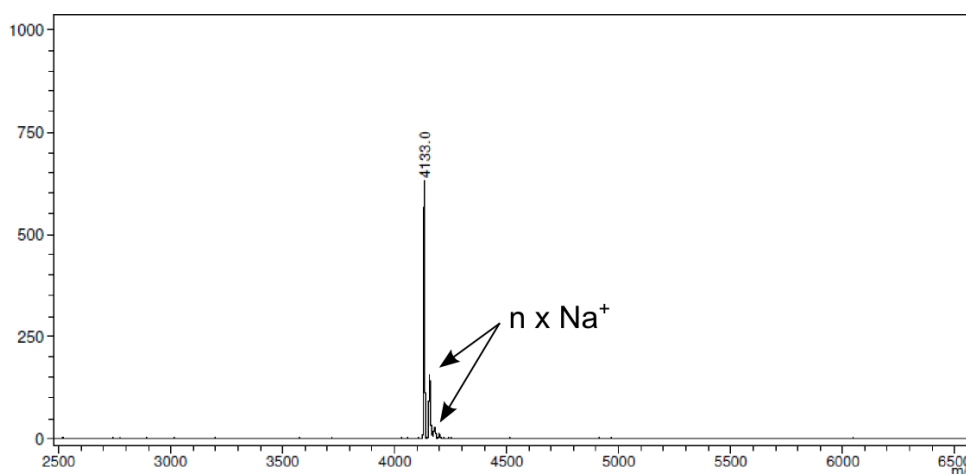
Figure 4.8: Representative MALDI-TOF spectrum of the converted and PAGE purified oligonucleotide containing dG*. Several side peaks from sodium ions attatched to the ionized oligonucleotide are visible, each spaced approx. 23 Da apart. The theoretical mass for the converted oligonucleotide in this example is 4,130 Da. The calibration accuracy of the MALDI-TOF data is +/- 0.5% ($\triangleq$ 20 Da).

4.1.6. Production of crosslinkable DNA

The crosslinking strategy chosen in this work is based on the convertible nucleotide approach [107]. A guanine base with a C2-thioalkyl tether at the N2 position (dG*) was introduced at the desired location via oligonucleotide ligation. For the crosslinking of 147 bp α-satellite DNA to cysteine point mutations on the H3 N-terminal tail, positions 69 or 70 on both the I-strand and the J-strand were changed to dG* via the ligation of a convertible dG* oligonucleotide. The base on the opposite strand oligonucleotide was changed to C to allow for Watson-Crick base pairing. After annealing and purification of the oligonucleotides (sections 3.2.16, 3.2.17), the DNA containing dG* and a sticky end compatible to the AvaII or BsaI restriction site was ligated to the core piece of DNA as prepared by restriction digest with AvaII or BsaI (section 3.2.10). The ligation of palindromic material could not be done sequentially, since the sticky ends from the AvaII digest on the core DNA self-ligated in the absence of oligonucleotide (Fig. 4.9, 4.10). Since both ligations (core DNA palindrome ligation and oligo to core DNA ligation) occur simultaneously and the core DNA overhangs from the AvaII digest tended to self-ligate, the oligonucleotide had to be added in excess to favor the ligation of an oligo-nucleotide over core DNA self-ligation (Fig. 4.9). The minimum excess of oligo to core DNA needed to avoid contamination by higher ligation states was determined while minimizing the expenditure of oligonucleotide. A ratio of oligonucleotide to core DNA of 2:1 during the ligation reaction was appropriate to yield a sample which was mostly devoid of higher ligation and incomplete ligation products.
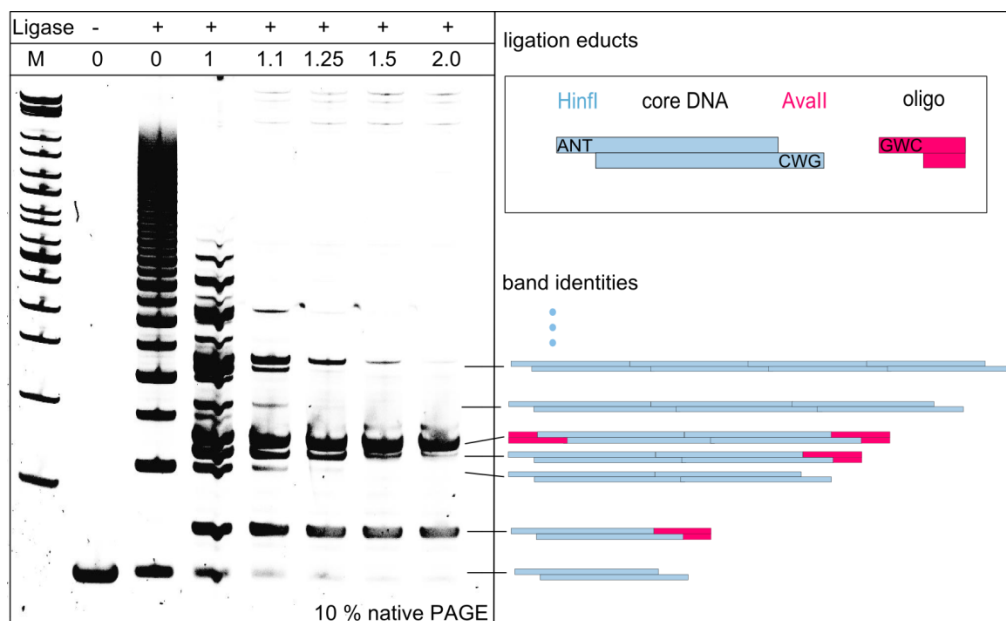
Figure 4.9: Ligation reaction to produce 147 bp α-satellite DNA containing a convertible dG* at position 70 of the J-strand. Since the core DNA self-ligates, which can be seen by the ladder-formation in lane 3, the oligonucleotide has to be added at the same time. To ensure minimal self-ligation of core DNA AvaII sites, the oligonucleotide was added in excess. Different ratios of oligonucleotide to core DNA were tested. Samples were run on a 10% native PAGE gel.

A

```
ATCCATATCCACCTG
TAGGTATAGGTGGACCAG-Ⓟ

Ⓟ-GTCC.....54bp.....G
      G...core DNA...CTWA-Ⓟ
```

B

T4 Ligase

(Core DNA 54 bp : Oligo)x2
Core DNA 54 bp : Oligo
Core DNA 54 bp
Oligo 15 bp

10 % native PAGE
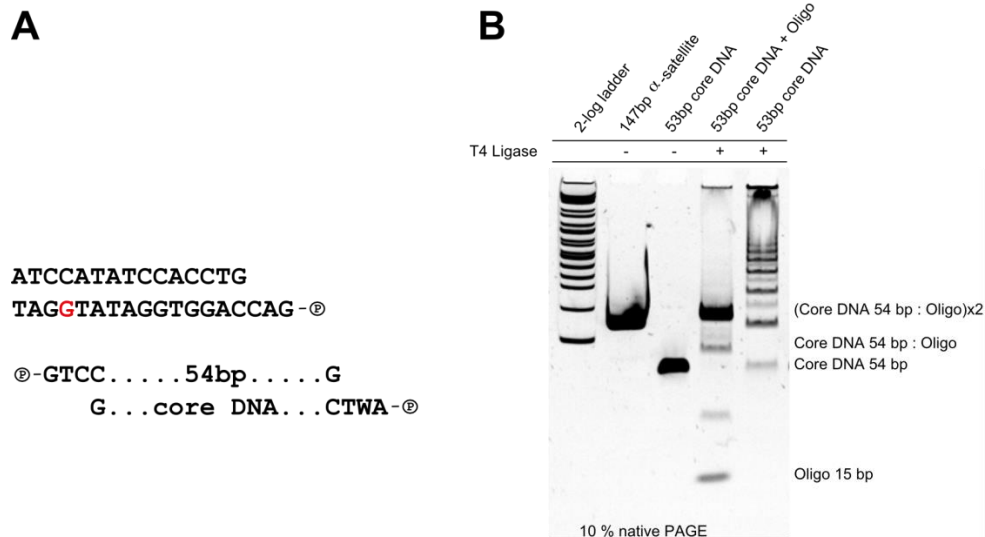
Figure 4.10: Oligonucleotide ligation to obtain crosslinkable 147 bp nucleosome DNA. **A)** Sequence of annealed oligonucleotide containing the convertible dG* (red). Encircled P designates phosphorylated 5'-ends. Below, the sticky ends of the core DNA are shown. W denotes either A or T. **B)** Ligation reactions to get crosslinkable palindromic α-satellite DNA. Lane 5 shows self-ligation of the core DNA. Simultaneous ligation of core DNA and oligonucleotide shows successful ligation (lane 4).

After successful ligation of oligonucleotide to core DNA, the ligation reaction was purified on an ion exchange column (60 ml 5PW-DEAE, Tosoh Biosciences) as described in section 3.2.19.
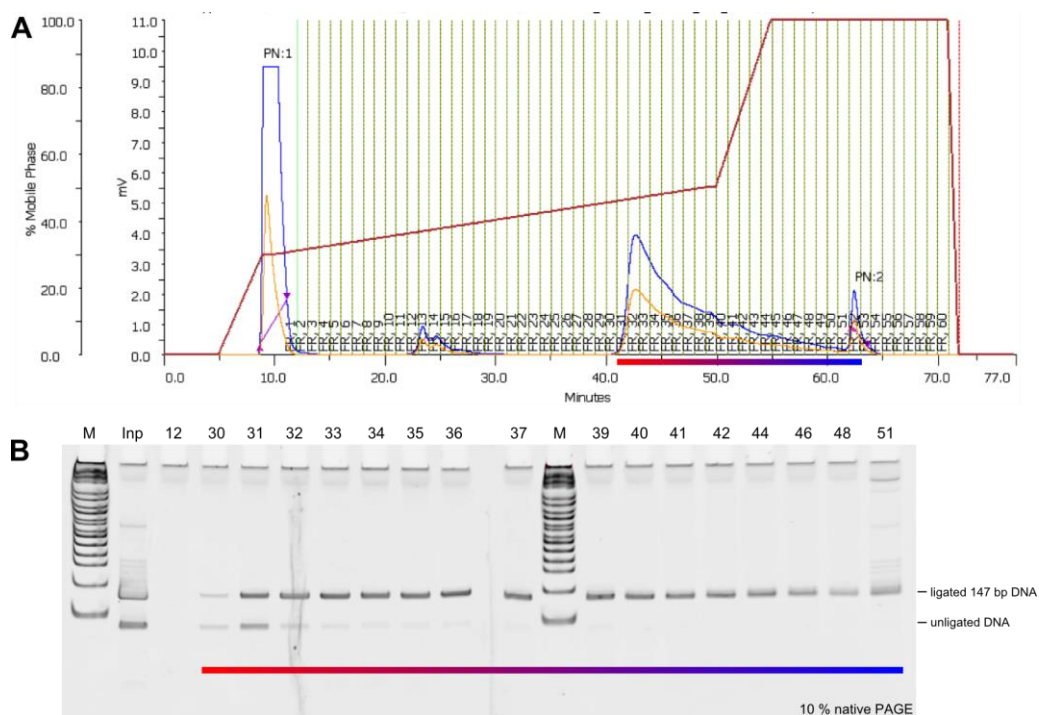
Figure 4.11: Ion exchange purification of ligated nucleosome DNA. **A)** UV traces (blue: λ = 260 nm, orange: λ = 280 nm) of the eluted fractions during purification of 147 bp α-satellite DNA on a 60 ml 5PW DEAE ion exchange column. The first peak, before the start of the fraction collection, is ATP which is present at high concentrations in the ligation buffer and co-precipitates during the preparation. The first small peak (fractions 11 – 16) is from remaining oligonucleotide. The unligated halves of the α-satellite palindrome are hard to separate from the ligated product but a reasonably pure sample can be obtained by discarding the first few fractions of the fully ligated product. **B)** Peak fractions on a 10% native PAGE showing the separation of the different species from the ligation reaction.

65

4.1.7. Nucleosome assembly and crosslinking

After purification on a 5PW DEAE (Tosoh Biosciences) HPLC column (Fig. 4.11), the crosslinkable DNA was assembled into NCPs with histone octamer containing one of the mutated H3 histones. Nucleosome reconstitution was carried out as described in section 3.4.1. The assembled NCPs were analyzed on both 5% native PAGE gel and 18% SDS PAGE. All four permutations of crosslinks between convertible dG* at positions 69 and 70 on the I and J-strand to H3 point mutations H39C and R40C were tested (Fig. 4.12).
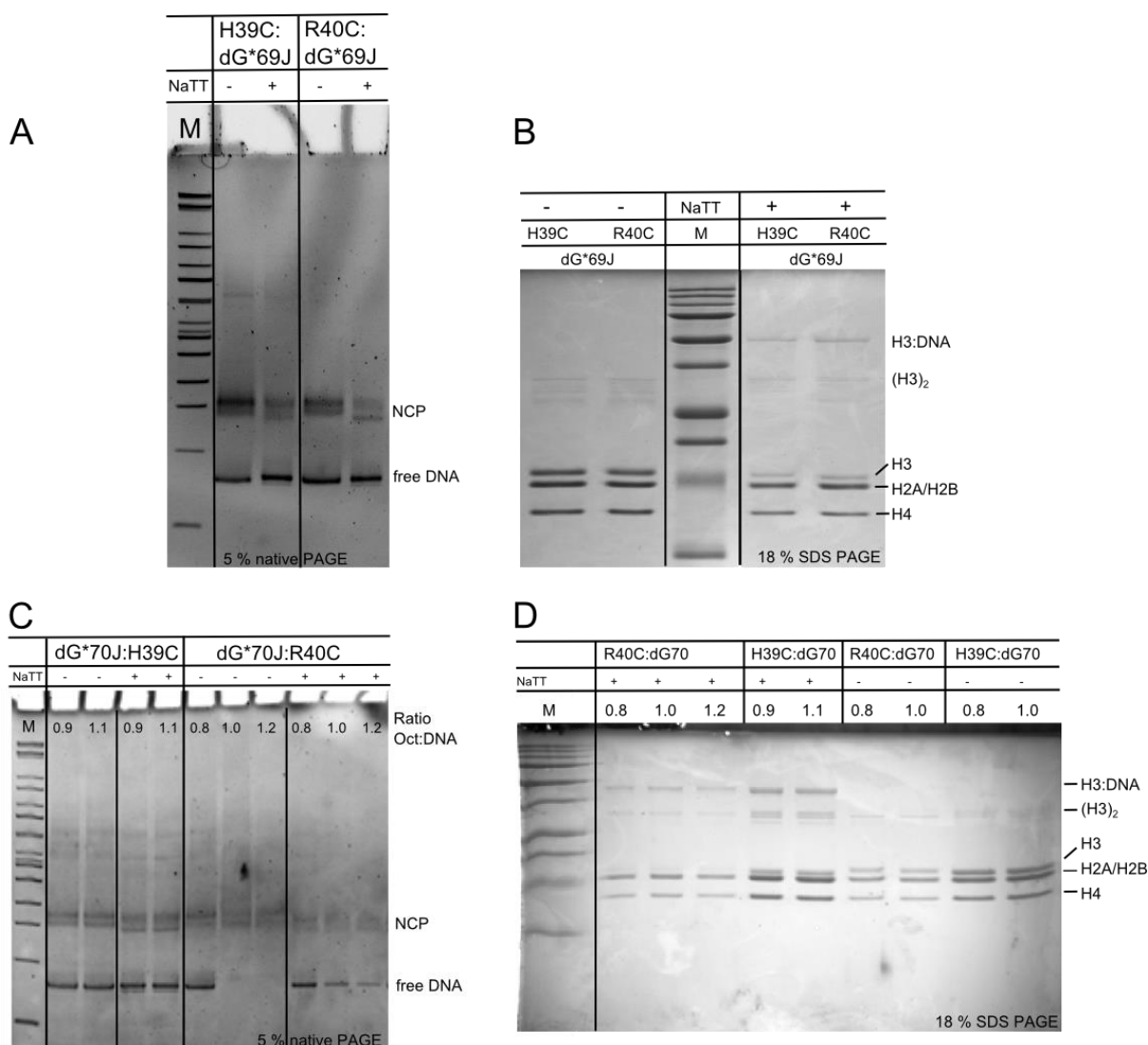
Figure 4.12: Biochemical characterization of different crosslinking locations. **A)** 5% native PAGE gel of reconstituted NCPs in the reduced state (-NaTT) compared to the crosslinked state (+NaTT) for NCPs with a convertible dG* at position 69.I and 69.J and H3 point mutations H39C or R40C. **B)** Same samples run on an 18% SDS PAGE in the reduced state (-NaTT) and the oxidized state (+NaTT). Band identities as indicated on the right side of each gel. **C)** 5% native PAGE gel of reconstituted NCPs in the reduced state (-NaTT) compared to the crosslinked state (+NaTT) at different octamer:DNA ratios for the convertible dG* in positions 70.I and 70.J. Lane 2 and 3 are reduced NCPs with the H3 H39C point mutation, lane 4 and 5 are the same sample but crosslinked with 4 mM sodium tetrathionate (+NaTT). Lanes 6 – 8 are NCPs at different octamer:DNA ratios with the H3 R40C point mutation to dG*70.I/J and lanes 9 – 11 are the same sample as in lanes 6 – 8 but crosslinked with 4 mM sodium tetrathionate (+NaTT). Band identities as indicated on the right side of the gel. **D)** The same samples as in (C) run on an 18% SDS PAGE. Some ratios from the reduced samples were omitted.

To obtain an estimate of the band intensities as an indicator of crosslinking efficiency, the images of the SDS PAGE gels were inverted and the background was subtracted using a rolling paraboloid (diameter 50 pixels, ImageJ [112]). The band intensities were integrated into horizontal bin values and plotted against the direction of the gel run. The resulting distributions were fitted as a set of Gaussians (SciDavis). The resulting intensities are shown in Table 4.4.

Band idendity	H39C dG*70 reduced (D, lane 9)	H39C dG*70 oxidized (D, lane 6)	R40C dG*70 reduced (D, lane 8)	R40C dG*70 oxidized (D, lane 4)	H39C dG*69 reduced (B, lane 1)	H39C dG*69 oxidized (B, lane 4)	R40C dG*69 reduced (B, lane 2)	R40C dG*69 oxidized (B, lane 5)
H2A/H2B	349	436	214	209	1431	1075	1439	1139
H3	288	214	108	32	1093	206	1070	266
H4	239	312	109	111	1024	694	1030	726
H3:H3	48	195	22	27	51	83	57	73
H3:DNA	0	235	0	105	0	191	0	255

Table 4.4: Intensities of gel bands on 18% SDS PAGE (Fig. 4.12, lanes indicated in table). The values are integrated intensities and are in arbitrary units.

The percentage of crosslinked H3 is estimated by evaluating the band intensities of H3 normalized against the band intensity of H4 (as a reference value):

$$Crosslinking\ efficiency\ estimate = 1 - \left\{ \frac{|H3ox|}{|H4ox|} \middle/ \frac{|H3red|}{|H4red|} \right\} \qquad (1)$$

The resulting crosslinking efficiency estimates are shown in Table 4.5.

Crosslink	Crosslinking efficiency estimate
H3 H39C:dG*69	72%
H3 H39C:dG*70	57%
H3 R40C:dG*69	35%
H3 R40C:dG*70	71%

Table 4.5: List of calculated crosslinking efficiency estimates for permutations of crosslinks from dG*69 and dG*70 to H3 H39C and H3 R40C based on SDS PAGE gel analysis. Calculation of values are according to equation (1).

Initially, the crosslinking pairs H3 H39C:dG*69 and H3 R40C:dG*70 were judged equally effective, but further experiments showed more efficient crosslinking for the H3 R40C:dG*70 pair. Crosslinking efficiencies of 90% for the H3 to DNA crosslink were achieved (Fig. 4.13). The H3 R40C:dG*70.I/J crosslink pair was subsequently used to produce up to 1 mg of crosslinked NCP for crystallization.
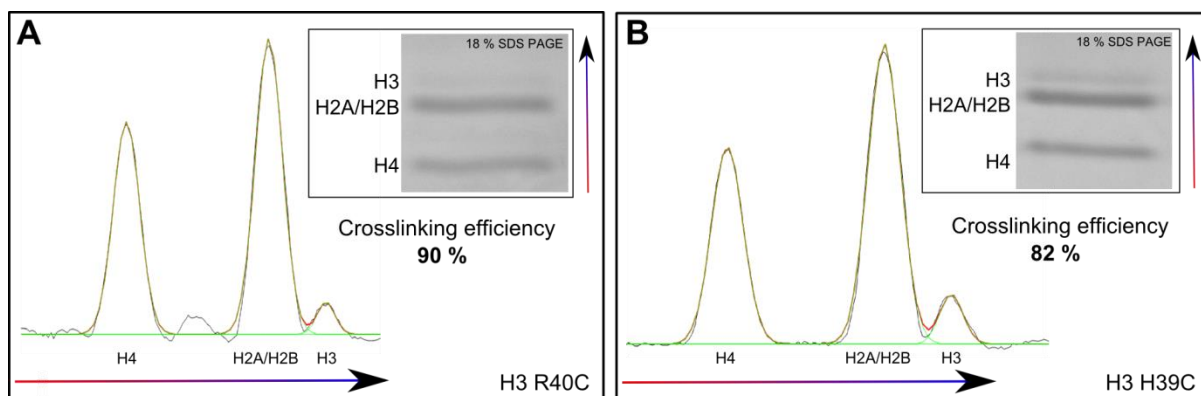
Figure 4.13: Determination of crosslinking efficiencies according to equation (1). **A)** The crosslinking efficiency for H3 R40C:dG*70.I/J determined by SDS PAGE (insert) gel analysis. The binned gray values (gray curve) are fitted (red curve) with multiple gaussians (green curves). **B)** Same analysis as in A) for the H3 H39C:dG*70.I/J crosslink.

4.1.8. Purification of crosslinked NCPs

The crosslinked NCPs were purified over a 60 ml 5PW DEAE weak ion exchange column (Tosoh Biosciences) as described in section 3.4.3. The increasing monovalent salt concentration during the elution from the DEAE column destabilizes nucleosomes that are not crosslinked and leads to dissociation [113]. This made the purification over a DEAE weak anion exchange chromatography column superior to gel filtration purification, since crosslinked NCPs were selectively purified. The crosslinked NCPs eluted as a single peak well separated from the remaining material, which consisted of NCPs with only one DNA terminus crosslinked or NCPs with no crosslink at all. Free DNA elutes last. A representative elution profile from the ion exchange column (60 ml 5PW DEAE, Tosoh Bioscience) is shown in Fig. 4.14.
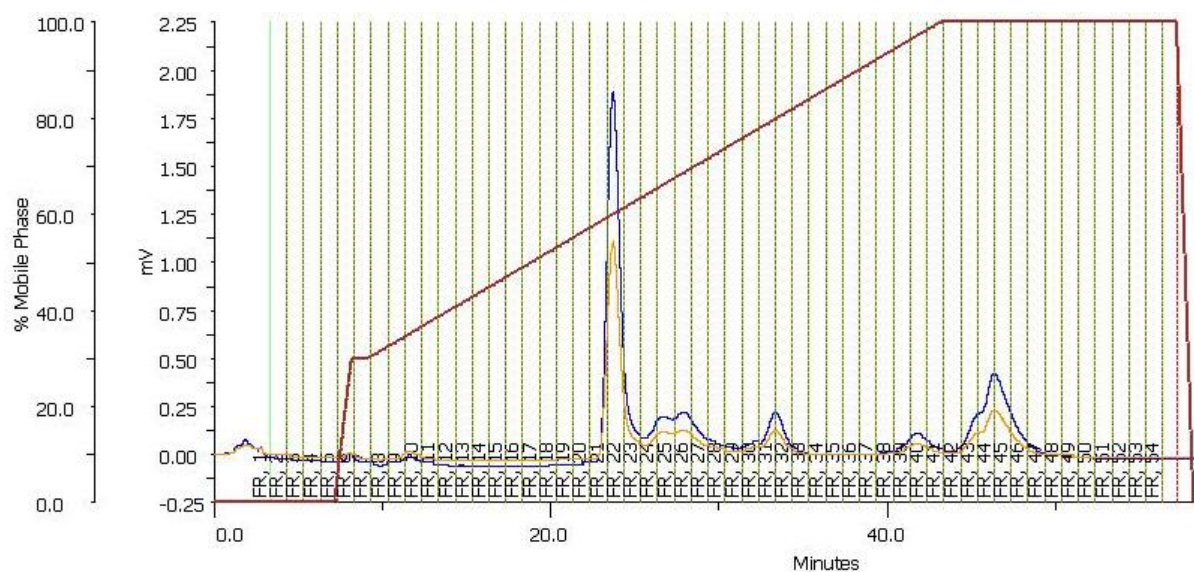
Figure 4.14: Elution profile of a crosslinked NCP sample from the DEAE column. The red trace shows the gradient from Buffer A (250 mM KCl) to Buffer B (600 mM KCl). The blue and yellow traces are the UV absorption readings at 260 nm and 280 nm respectively. Collected fractions are separated by vertical dotted lines and represent 4 ml fractions.

The peaks containing crosslinked NCPs were pooled and concentrated and the buffer was exchanged to a TE buffer containing 10 mM KCl, as described in section 3.4.3. An aliquot of the purified sample was analyzed on a 5% native PAGE and on an 18% SDS PAGE (Fig. 4.15).
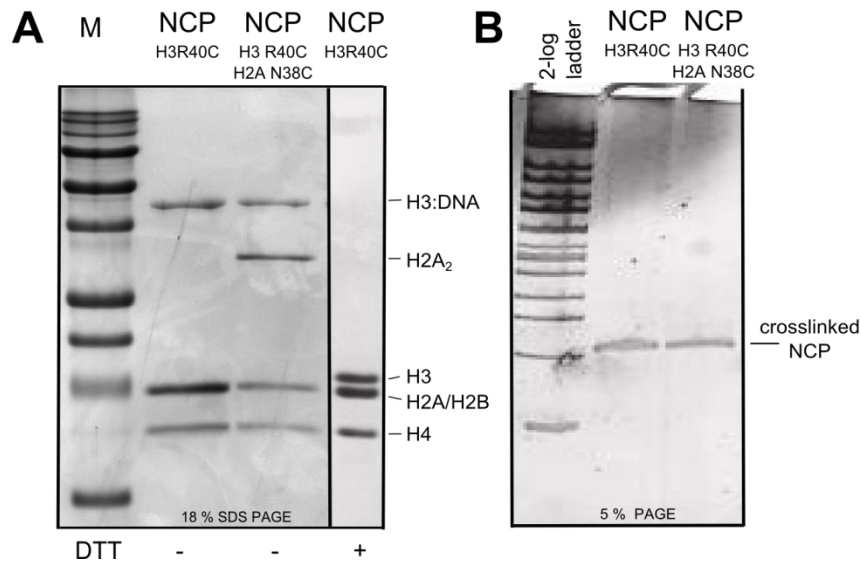
Figure 4.15: Biochemical characterization of purified crosslinked NCP samples. **A)** 18% SDS PAGE gel of purified crosslinked NCPs. **B)** 5% native PAGE EMSA gel of purified crosslinked NCPs.

The purified NCPs showed a high degree of homogeneity with no uncrosslinked H3 detectable. The purification of double crosslinked NCPs also yielded highly homogenous sample. Both single crosslink NCP and double crosslinked NCP (see section 4.2) run as homogenous well-defined bands on a 5% native PAGE gel (Fig. 4.15).

4.1.9. Stability of NCP against DNA dissociation upon dilution

The stability of the nucleosomes against DNA dissociation upon dilution was tested with a dilution series assay. A sample of 1 µM concentration of H3 R40C:DNA crosslinked sample treated with DTT to release the disulfides was prepared to be used as a reference for uncrosslinked nucleosome sample. 1 µM stock samples of both the single crosslink sample (H3 R40C:DNA) and the double crosslinked sample (H3 R40C:DNA, H2A N38C) were prepared. 25 µl of each sample were taken and diluted with TEK10 in ratios of 1:10 (100 nM), 1:50 (20 nM), 1:250 (4 nM) and 1:1250 (0.8 nM). The samples were incubated in the diluted state for 12 h at 4 °C before being concentrated back to 25 µl. 10 µl of each reconcentrated sample was mixed with 2 µl 25% sucrose and analyzed on a 5% native PAGE EMSA gel (Fig. 4.16). The uncrosslinked reference sample showed free DNA in each dilution condition, whereas both the single crosslink and the double crosslinked NCP samples remained intact over the entire range of dilutons. Therefore, the crosslink between H3 and DNA effectively stabilized the nucleosome against DNA dissociation upon dilution. The highest dilution of the sample with

the crosslink reversed showed the least amount of free DNA (Fig. 4.15, uncrosslinked NCP, 1:1250), which was not expected. The total amount of sample, as judged by band intensity, is also the lowest, which indicates that significant amounts of sample precipitated upon dissociation of DNA from the histone octamer.
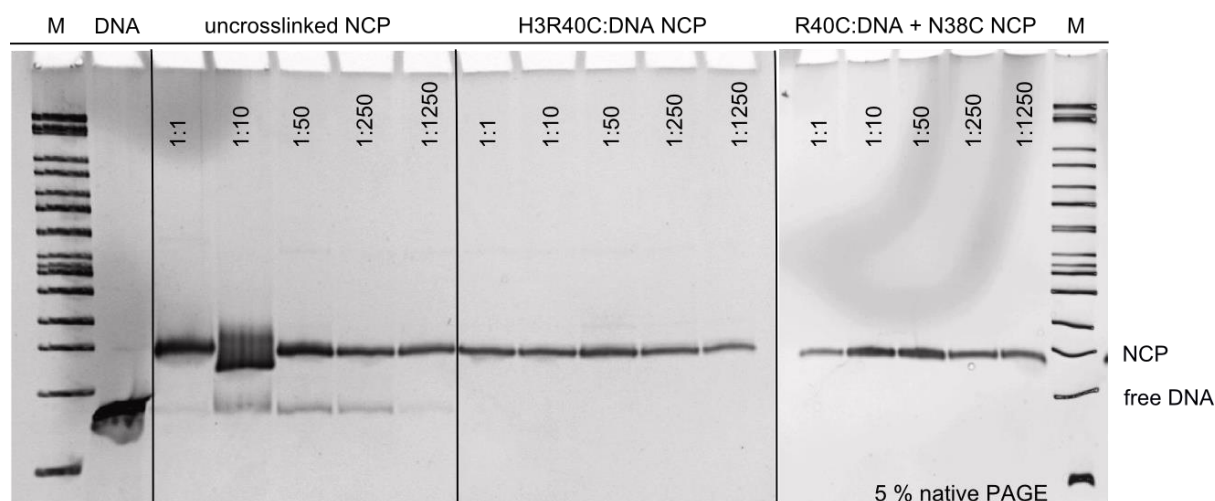
Figure 4.16: Dilution series assay of crosslinked nucleosomes. The uncrosslinked NCP sample consisted of H3 R40C:DNA crosslinked sample with the disulfides released by addition of DTT. Single crosslink and double crosslink NCP samples are labeled. The dilution factors are indicated in each lane. Starting concentration of each sample was 1 µM. Band identities are as indicated on the right side of the gel.

4.1.10.   Crystallization of crosslinked NCP

The crosslinked NCP samples were set up for crystallization under the same conditions as the crystals of uncrosslinked NCPs that gave well-diffracting crystals previously [9]. The NCP sample had a stock concentration of $4 - 6$ mg/ml ($A_{260}/A_{280} = 1.77$) in TE buffer and was set up for crystallization in sitting drop vapor diffusion experiments in a 24 well plate (Cryschem, Hampton) at 22 °C. Screen solutions (Standard NCP screen, K^+ vs. Mn^{2+}) were filtered through a 0.22 µm syringe filter before use. The drops were set up as 1 µl sample + 1 µl screen solution and equilibrated against the reservoir which consisted of 200 µl screen solution + 600 µl MilliQ water. The single wells were flushed with argon gas and sealed immediately and stored in a vibrationally isolated box with a nitrogen gas atmosphere. Crystals typically grew within a few days and fully matured within $2 - 3$ weeks to a size of up to 100 x 100 x 400 µm.
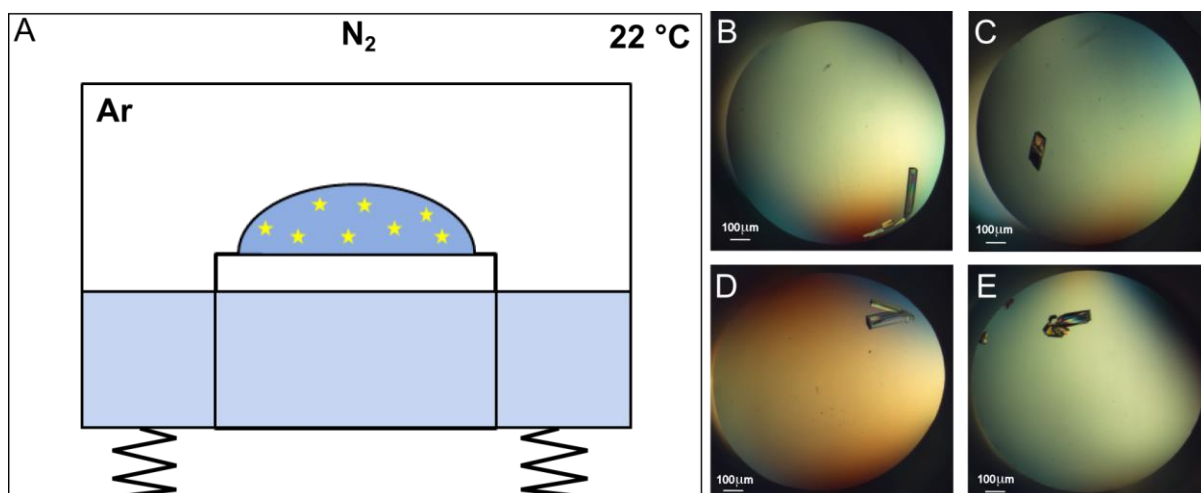
Figure 4.17: Sitting drop vapor diffusion setup leading to crystals of crosslinked NCPs. **A)** Vibrationally suspended sitting-drop crystallization setup in an N_2 environment. The crystallization wells were filled with argon gas before being sealed. The sitting drops were set up in a salting-in environment so that the sample (yellow stars) concentration decreases as the drop size grows when water from the well solution (light blue) enters the drop solution (darker blue). **B-E)** Crystals grown from 4 – 6 mg/ml crosslinked NCP in various salt conditions of the standard NCP screen. The crystal in C) stands upright in the well showing the hollow hexagonal rod morphology.

4.1.11.   Post-crystallization treatment

Crystals of NCPs containing the DNA:H3 crosslink were harvested and subjected to established dehydration conditions for improved diffraction as described in section 3.5. The crystals containing the DNA:H3 crosslink were subjected to a final short (4 min) incubation in the final dehydration condition with 0.5 M final concentration of sodium nitrate added (10 mM KCacodylate pH 6.0, 37 mM $MnCl_2$, 40 mM KCl, 24% MPD, 2% trehalose, 500 mM sodium nitrate). Nitrate ions (NO_3^-) act as a potent scavenger of free radicals and were shown to increase the resistance of disulfide crosslinks to radiation damage in lysozyme crystals [114].

4.1.12.   Data collection and processing

A complete dataset with 3.3-fold redundancy was collected at beamline X06DA (PXI) of the Swiss Light Source (SLS) at the Paul Scherrer Institute over seven different crystals under cryogenic conditions. The diffraction images were collected on a Pilatus 6M-F detector with a crystal to detector distance of 500 mm. Per position on the crystal, six frames were collected, each spanning 0.3° of rotation for 5 s of unattenuated beam. The total dose per crystal position is estimated to be 10 MGy at a beam flux of 2.7 x 10^{12} photons/s and a beam diameter of approximately 50 μm x 90 μm. The dose rate in this case amounts to 366.6 kGys^{-1}.
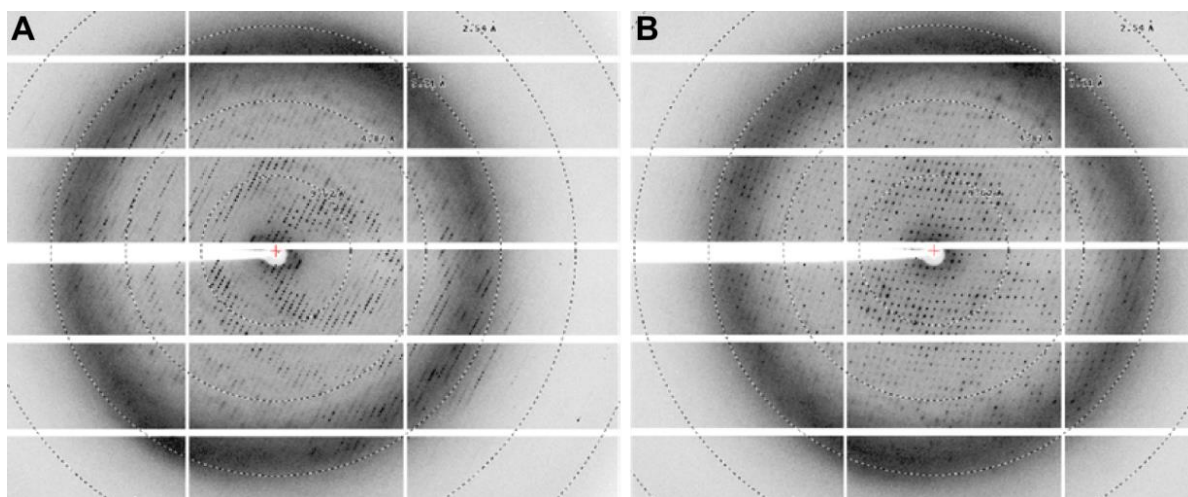
Figure 4.18: Diffraction images of NCP crystals containing a DNA to histone crosslink. **A)** The X-ray diffraction was recorded on a 6M-F Pilatus detector. Resolution spots are visible to 2.6 Å. **B)** Another view rotated by 90°.

The dataset was indexed using XDS [99, 98] which was controlled by the XDSHelper suite (T. J. Richmond). The cell dimensions for the different crystals were within 0.4% of the average value (Table 4.6). Structural isomorphism between single crystals could therefore be assumed and the framesets of the single crystals were merged using XDSHelper.

Crystal	Avg. a [Å]	Avg. b [Å]	Avg. c [Å]
1	106.17	183.22	109.50
2	106.08	182.47	109.47
3	106.06	182.81	109.52
4	106.55	183.19	109.48
5	106.32	183.28	109.44
6	105.92	182.99	109.49
7	106.43	183.48	109.50
Average	**106.25**	**183.11**	**109.49**

Table 4.6: Average unit cell dimensions of the single crystals making up the full dataset. The overall average is displayed in the bottom row.

The merged dataset was then used for phasing by molecular replacement. The high resolution NCP structure (PDB ID: 1KX5) served as a search model for the CCP4 Phaser module [100]. Refinement of the structure was done entirely in Phenix [101]. Model building and visualization was done in Coot [102]. Introduction of the convertible dG* base was done by using a *pdb*-file from a previously solved structure [111] containing a convertible dG* crosslink and creating a *pdb-file* and an *mmCIF*-file containing restraints for the convertible dG* using the eLBOW module in Phenix [115]. The first round of refinement was done with the convertible dG* base modeled as a regular dG. Refinement macrocycles in Phenix consisted of the following refinement steps:

- Bulk solvent correction
- Rigid body refinement (only first macrocycle)
- Simulated annealing (only first macrocycle)
- Real-space refinement
- TLS refinement
- B-factor refinement (individual)

During refinement cycles, the DNA base pairs were constrained to adhere to standard Watson-Crick base pairing. The minimization algorithms all employed maximum likelihood approaches.

Data collection	Single crosslinked NCP (NCP-x)
Framesets	56
Resolution [Å]	29.54 – 2.80
Resolution of last shell [Å]	2.95 – 2.80
Multiplicity	3.3
No. of unique hkl	47,952
Completeness (last shell)	90.3 (68.0)
Rmerge (last shell) [%]	8.4 (23.1)
I/σ (last shell)	12.7 (2.5)
Spacegroup	P $2_12_12_1$
a [Å]	106.25
b [Å]	183.13
c [Å]	109.49
Refinement	
Resolution [Å]	29.5 – 2.8
R-factor (last shell) [%]	19.97 (38.05)
R-free (last shell) [%]	25.11 (42.83)
No. of Reflections work/free	42,969 (4,812)
No. of atoms in model	
Total (Avg. B-factor [Å^2]	12,218 (91.4)
RMSD from ideality	
Bond length [Å]	0.011
Bond angles [°]	1.120

Table 4.7: Statistics of data collection and refinement of the crosslinked NCP.

4.1.13.   Model validation

The refined model was analyzed using the MolProbity toolbox [104]. The model geometry was inspected and the protein backbone geometry was checked for outliers. Iterative refinement cycles were carried out until no further improvement of model quality could be achieved.

Basic statistic	Value
Ramachandran outliers (%)	0.7
Ramachandran favored (%)	96.3
Rotamer outliers (%)	2.5
Clashscore	13.04

Table 4.8: Basic statistics of the refined structure determined by MolProbity. The clashscore is the number of serious clashes per 1,000 atoms [104]. A clashscore below 20 is considered well-resolved.

4.2. Double crosslinked nucleosome core particle

4.2.1. Introduction

The DNA-H3 crosslink stabilizes the NCP and makes it resistant to DNA dissociation from the histone octamer. An additional stabilizing crosslink was introduced into the NCP by crosslinking the two copies of H2A to each other via an N38C point mutation. As previously reported, the H2A crosslink stabilizes the histone octamer and keeps the H2A-H2B dimer from dissociating from the histone octamer [1]. The occurence of hexasomes by dissociation of an H2A-H2B dimer was effectively prevented by introduction of the N38C crosslink [1]. This crosslink was introduced into the NCP in addition to the DNA-H3 crosslink with the ultimate goal of creating highly robust NCPs.
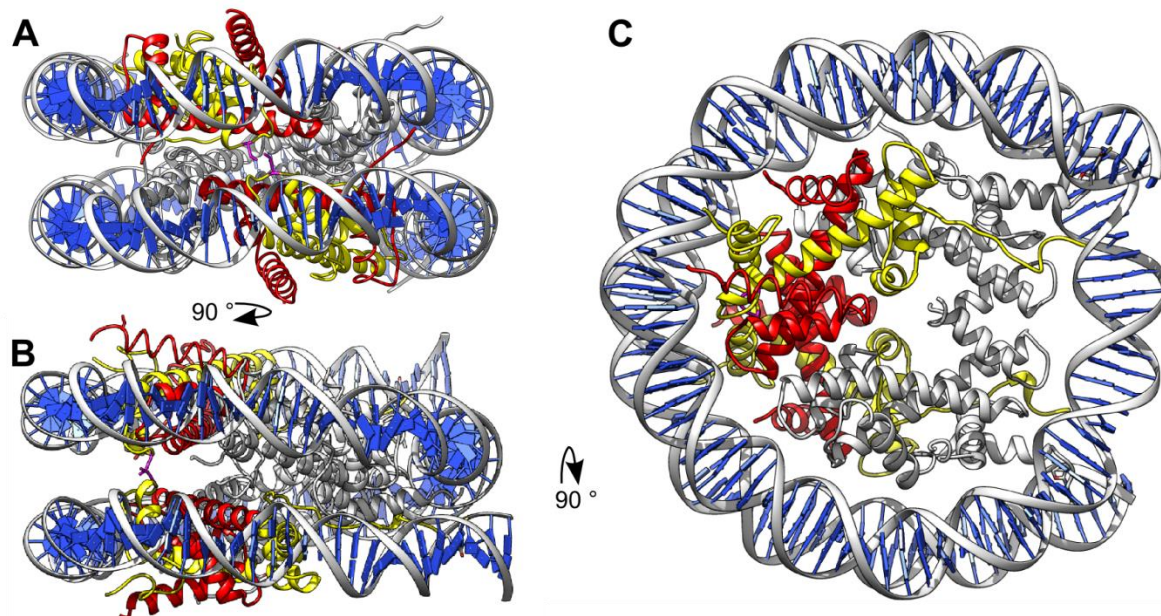
Figure 4.19: Model of the NCP carrying the H2A N38C point mutation. **A)** Side view of the NCP facing the H2A-H2B inter-dimer interface. **B)** Second side view of the NCP rotated by 90°. **C)** Top view of the NCP-xx model. The H2A (yellow)/H2B (red) heterodimers are highlighted with the mutated amino acid (N38C, shown in magenta) positioned at the only contact site between the NCPs two heterodimers. H3 and H4 are shown in grey ribbons. The convertible nucleoside is shown in sticks while the rest of the nucleosome DNA is shown in slab/ribbon representation.

4.2.2. Nucleosome core particle preparation and crosslinking

The DNA used for generating the double crosslinked NCP (NCP-xx) was identical to the DNA used for the DNA-H3 crosslinked NCPs (section 4.1.4). The histone octamer was assembled from H2A N38C, H3 R40C/C110A and wild type histones of H2B and H4 (as described in section 4.1.4). The octamer was assembled under a reducing environment to deprotect the sulfonated cysteine amino acids and keep the thiols reduced. The reducing agent was only removed by dialysis after the histone octamer was assembled into NCPs. The assembly of NCPs followed the standard protocol as described in section 3.4.1. The assembled NCPs were first heat-shifted for 120 min at 50 °C while still in the dialysis

bag. The crosslinking was started by dialysis into degassed TEK10 buffer supplemented with 4 mM tetrathionate. The crosslinking reaction took place for 12 h at 22 °C. After pelleting any precipitate, the double crosslinked NCP sample was concentrated to approximately 3 ml and brought to 250 mM KCl.

4.2.3. Double crossliked nucleosome core particle purification

The double crosslinked NCP was purified on a weak anion exchange column (5PW DEAE, Tosoh Biosciences) using the same protocol and parameters as in the purification of the DNA:H3 crosslinked NCP (section 4.1.8).

4.2.4. Crystallization of double crosslinked NCP

The purified double crosslinked NCPs were set up for crystallization as vapor diffusion experiments in a sitting drop setup using a 24 well plate (CrysChem, Hampton Research) as described in section 4.1.10. Initial crystallization trials were undertaken using the crystallization conditions from previous non-crosslinked NCPs. The screening covers a range of concentrations of monovalent salt (KCl) and divalent ions ($MnCl_2$). The 'standard' NCP screen covers a range of 100 mM – 160 mM KCl in 20 mM increments and a range of 130 mM – 180 mM $MnCl_2$ in 10 mM increments. The sample drops were diluted twofold by mixing them with an equal volume of screen solution. The reservoir solution is diluted fourfold with purified water, effectively creating a reverse salting-in environment for the crystallization. Crystal growth took place rather quickly with the first crystals growing within two days and fully mature crystals occurring in almost all conditions within a week. The crystal morphology was remarkably different from that of previously grown NCP crystal, which diffracted to high resolution. The crystals of crosslinked NCPs appeared flatter in one dimension and showed prominent saw-tooth like morphology (Fig. 4.20). A single crystallization condition (highest monovalent and divalent salt concentration) showed a single crystal of similar morphology to known well-diffracting NCP crystals (hexagonal hollow rods).
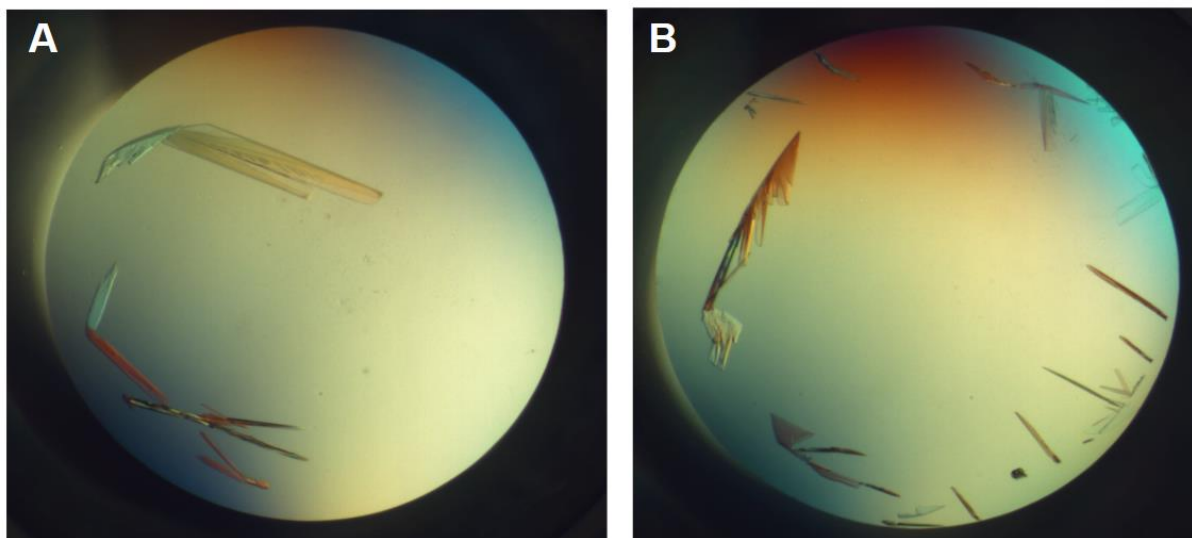
Figure 4.20: Crystals of double crosslinked NCP grown in the standard NCP screen. **A)** The morphology is markedly different to the standard NCP morphology, i.e. hexagonal hollow rods. The crystals grew very flat, maximally extending 10 µm in the smallest dimension. **B)** Many very thin, flat crystals grew.

Crystals from most wells were harvested and subjected to the usual dehydration procedure (as described in section 3.5) for standard NCP crystals, where the soaking in 24% MPD and 2% trehalose reduced water content in the unit cell and improved the diffraction quality of the crystals. Crystals of the double crosslinked NCP were taken to the Swiss Light Source (SLS) and data was collected at beamline X06DA (PXIII) at a wavelength of 1 Å under cryogenic conditions. The diffraction quality of most crystals was poor with reflections visible only up to a resolution of 20 Å (Fig. 4.21).
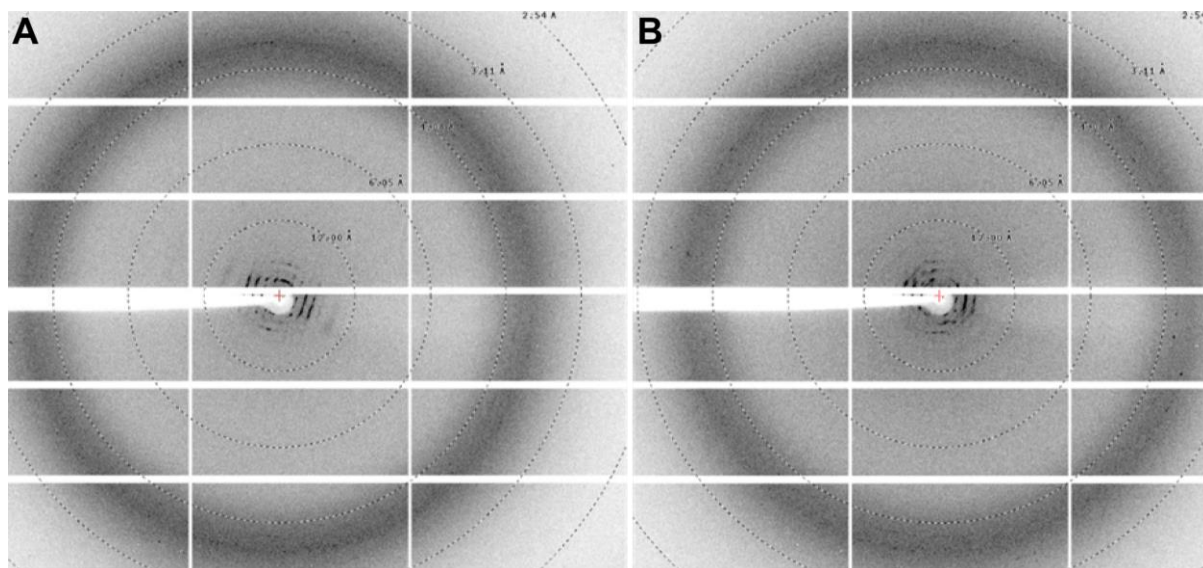
Figure 4.21: Initial diffraction of double crosslinked NCP crystals grown under standard NCP conditions. **A)** The reflections were recorded on a 2M-F Pilatus detector. The resolution of the crystals was below 20 Å. **B)** Second diffraction image rotated by 90°.

The crystal with the 'hollow rod morphology' in the highest salt condition diffracted to the highest resolution (approximately 7 Å), suggesting that the double crosslinked NCPs would grow crystals that diffract to high resolution at higher salt concentrations, probably due to the slower crystal growth compared to the growth in lower salt conditions. Hence another crystallization plate was set up with screening conditions ranging from 140 mM – 200 mM KCl in 20 mM increments and from 160 mM – 210 mM $MnCl_2$ in 10 mM increments (section 2.6.2). Crystallization drops were set up as for the previous plate. Crystal growth commenced more slowly in the higher salt conditions with the first crystals appearing three days after setting up of the plate. In addition to a few large crystals per well, a large number of small crystals appeared. Within a week most crystals were fully matured. The crystal morphology was of the hollow hexagonal rod type with crystal sizes ranging from 100 μm x 100 μm to 150 μm x 500 μm. The newly grown crystals of the double crosslinked NCPs were harvested and subjected to standard dehydration conditions as described before (section 3.5).
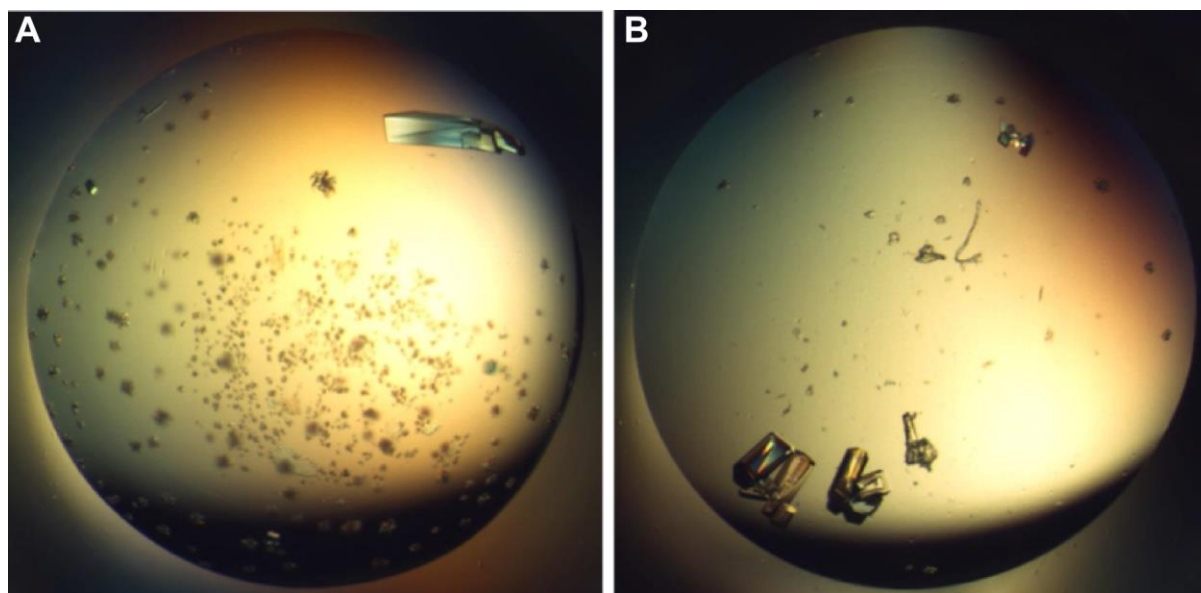
Figure 4.22: Crystals of double crosslinked NCPs grown in the extended NCP screen. **A)** Apart from a few large crystals, the elevated [K^+] and [Mn^{2+}] NCP screen yielded many small crystals. **B)** In some conditions, the crystals grew as hexagonally shaped hollow rods.

4.2.5. Data collection and processing

A full data set was collected at the Swiss Light Source (SLS) at beamline X06DA (PXIII) under cryogenic conditions and an X-ray wavelength of 1 Å. The diffraction images were recorded on a Pilatus 2M-F detector positioned 260 mm away from the crystal. The dataset was recorded using two crystals with an average exposure of 540 s per crystal position at a beam flux of $2x10^{11}$ photons/s. The rotational increment per exposure was 0.1°. The total dose per crystal position was approximately 10 MGy, which constitutes a compromise between high resolution and preservation of the disulfide bonds. Statistics of the dataset are shown in Table 4.10. The framesets of the dataset were indexed

78

using XDS [99] via the XDSHelper software suite (T. J. Richmond). Framesets from the two individual crystals were scaled and merged also using XDSHelper. The unit cell dimensions of the individual crystals are shown in Table 4.9 and show good agreement (< 0.2% difference/axis), justifying the merging of the framesets.

Crystal	Avg. a [Å]	Avg. b [Å]	Avg. c [Å]
1	106.43	183.73	109.46
2	106.69	183.94	109.52

Table 4.9: Unit cell dimensions of both crystals contributing to the dataset after indexing with XDS. All three unit cell dimensions are within 0.2% of each other for both crystals.

The final dataset was phased by molecular replacement using the high resolution structure of the NCP as the search model (PDB ID: 1KX5) and the program Phaser from the CCP4 suite [100]. The high degree of sequence identity to the search model (> 99%) gave an unambiguous solution in the molecular replacement. Refinement of the solution was done entirely using Phenix [101]. The starting model was modified from the high resolution NCP structure (PDB ID: 1KX5). The regions of the tails with zero occupancy were first deleted as were the ordered solvent molecules. The N-terminal tails of histone H3 were deleted up to R40. The amino acids N38 on the copies of H2A were intitally modelled as alanines. The DNA sequence was modified to include four changed base pairs at the introduced AvaII restriction sites and the base pair containing the convertible dG* (base pair 70). The convertible nucleotides were initially modelled as conventional dG bases. The first round of refinement consisted of five macrocycles of refinement, each carrying out the following refinement steps:

- Bulk solvent correction
- Rigid body refinement (only first macrocycle)
- Simulated annealing (only first macrocycle)
- Real-space refinement
- TLS refinement
- B-factor refinement

The refinement was performed with DNA base pair restraints in place to enforce standard Watson-Crick base pairing distances. The search algorithms all employed maximum likelihood approaches.

Data collection	Double crosslinked NCP (NCP-xx)
Framesets	10
Resolution [Å]	29.49 – 2.80
Resolution of last shell [Å]	2.95 – 2.80
Multiplicity	3.6
No. of unique hkl	52,834
Completeness (last shell)	98.6 (93.7)
Rmerge (last shell) [%]	6.2 (32.8)
I/σ (last shell)	13.9 (2.8)
Spacegroup	P $2_12_12_1$
a [Å]	106.56
b [Å]	183.84
c [Å]	109.49
Refinement	
Resolution [Å]	29.5 – 2.8
R-factor (last shell) [%]	21.83 (29.20)
R-free (last shell) [%]	25.56 (33.70)
No. of Reflections work/free	52,753 (4,852)
No. of atoms in model	
Total (Avg. B-factor [Å^2]	12,092 (109.8)
RMSD from ideality	
Bond length [Å]	0.0059
Bond angles [°]	0.896

Table 4.10: Statistics of data collection and refinement of the double crosslinked NCP.

4.2.6. Model validation

The refined structure was analyzed using MolProbity [104]. General model geometry (bond lengths, bond angles and dihedrals) was inspected and protein backbone geometry was checked for outliers.

Basic statistic	Value
Ramachandran outliers (%)	0.9
Ramachandran favoured (%)	95.8
Rotamer outliers (%)	4.9
Clashscore	15.22

Table 4.11: Basic statistics of the refined structure determined by MolProbity.
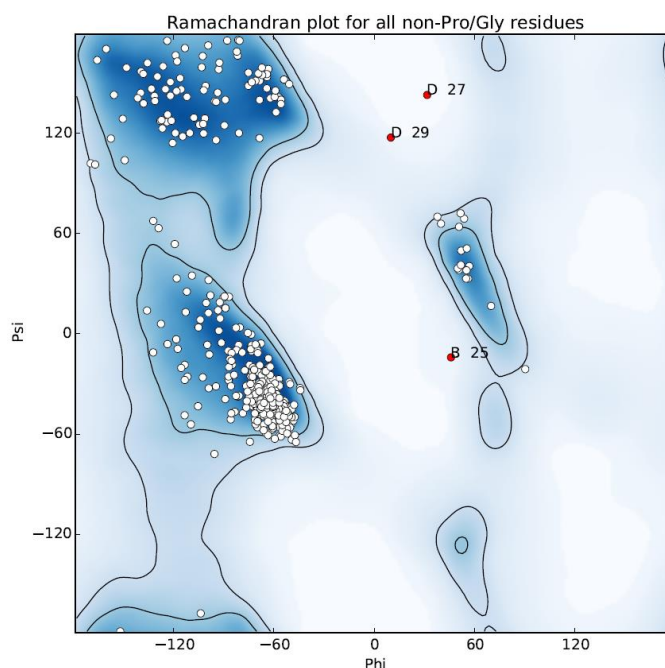
Figure 4.23: Ramachandran plot for all non-proline and non-glycine amino acids in the double crosslinked NCP structure. Darker blue color indicates more favorably oriented peptide bonds. Indicated in red are the few outliers labeled with their chain ID and amino acid number. White dots represent amino acids with Φ and Ψ values that are within the favored regions (encircled dark blue) or allowed regions (encircled light blue). The plot was created with Coot [102].

4.3. X-ray crystal structures of crosslinked NCPs

The structures of the NCPs with either the DNA:H3 crosslink (NCP-x) or with the DNA:H3 crosslink and the H2A N38C crosslink (NCP-xx) were determined to 2.8 Å. The a-axis dimensions of both crosslinked NCP structures were within 0.6% of the a-axis in the high resolution NCP structure (PDB ID: 1KX5, [9]). The b-axis was slightly larger in both crosslinked NCP crystals (+ 1.1% for NCP-x and + 1.5% for NCP-xx). The c-axis was identical in all three structures (Table 4.12).

Structure	a	b	c
α-satellite 147 bp (1KX5)	105.95	181.17	109.49
α-satellite 147 bp H3 R40C : dG*70.I/J	106.25	183.13	109.49
α-satellite 147 bp H3 R40C : dG*70.I/J + H2A N38C	106.56	183.84	109.49

Table 4.12: Comparison of the unit cell dimensions of the two crosslinked NCP structures with those of the high resolution NCP structure (PDB ID: 1KX5) [9]. Both crosslinked NCP structures differ by no more than 0.6% in the a-axis. The b-axis of crosslinked NCP crystals is approximately 1% larger than for 1KX5. The c-axis is identical to the 1KX5 c-axis in both crosslinked NCP structures.

The structures of the crosslinked NCPs overall showed little difference compared to the high resolution NCP structure (PDB ID: 1KX5) except for a few regions discussed below. The quality of the electron density ($2F_o$-F_c) for both crosslinked structures is representatively shown in Fig. 4.24 for the α1-helix of H2A and for the DNA around SHL 1.5 (NCP-x) and SHL 3.5 (NCP-xx).
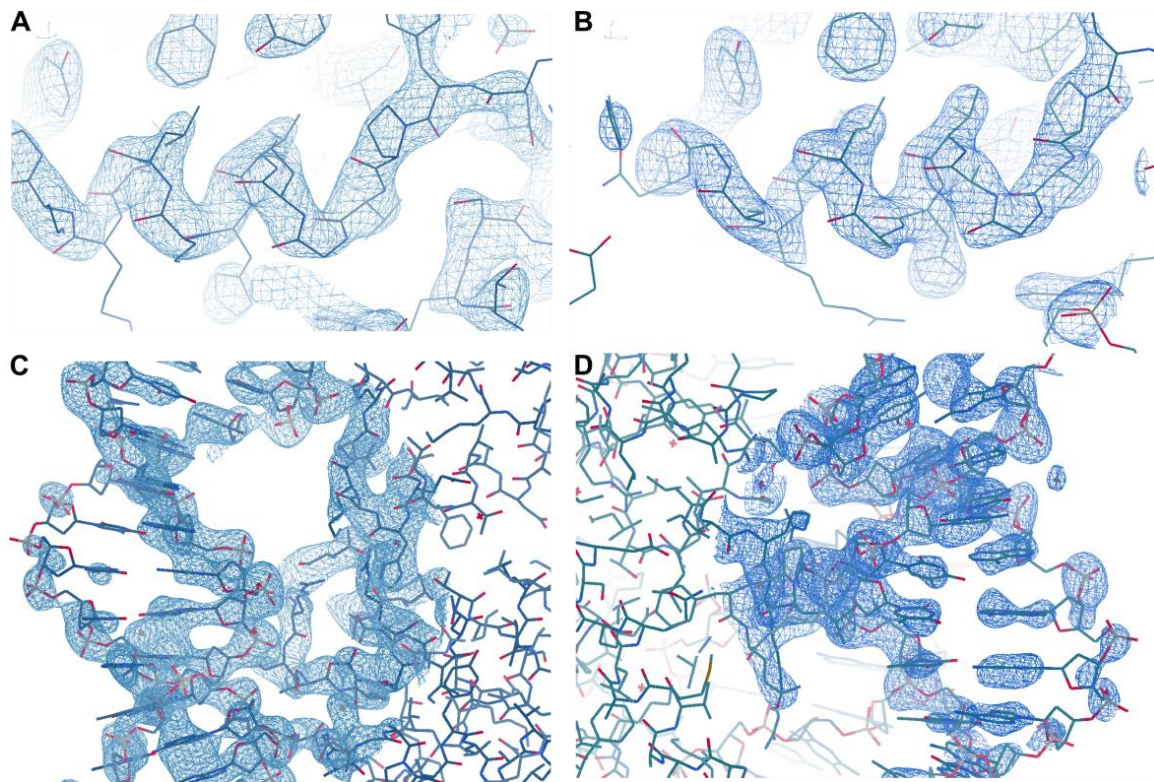
Figure 4.24: Examples of the electron density at several regions of the crosslinked NCP structures. **A)** The α1-helix of H2A in the NCP-x structure. **B)** The same location as in A) in the NCP-xx structure. **C)** Nucleosome DNA at SHL 3.5 (NCP-xx). The phosphates facing outward from the NCP have lower overall density and correspondingly higher B-factors (Fig. 4.25). **D)** SHL 1.5 region in the NCP-xx structure.

The overall B-factor of the crosslinked NCP structures are 91.4 Å^2 for NCP-x and 109.8 Å^2 for NCP-xx. The B-factor distribution is bimodal with the two main contributions from the protein and DNA parts of the NCP. Protein B-factors peak around 50 and 60 for NCP-x and NCP-xx respectively, while DNA B-factors peak around 125 - 150 for NCP-x and NCP-xx respectively. Fig. 4.25 displays the distribution of B-factors in a worm representation. It is apparent that the minor groove outwards positions of the DNA have a higher B-factor than the regions facing the histone octamer. This was to be expected since the interactions with the histones stabilize the DNA and reduce its disorder. This result agrees with the distribution of B-factors seen in other NCP crystal structures. Also, the N and C-terminal amino acids of all histones in both structures are relatively high, which reflects the fact that the histone tails extending beyond the NCP boundaries are too disordered to be resolved in the crystal structure. Regions of the histone octamer tend to adopt a higher B-factor where they are interacting with the nucleosome DNA.
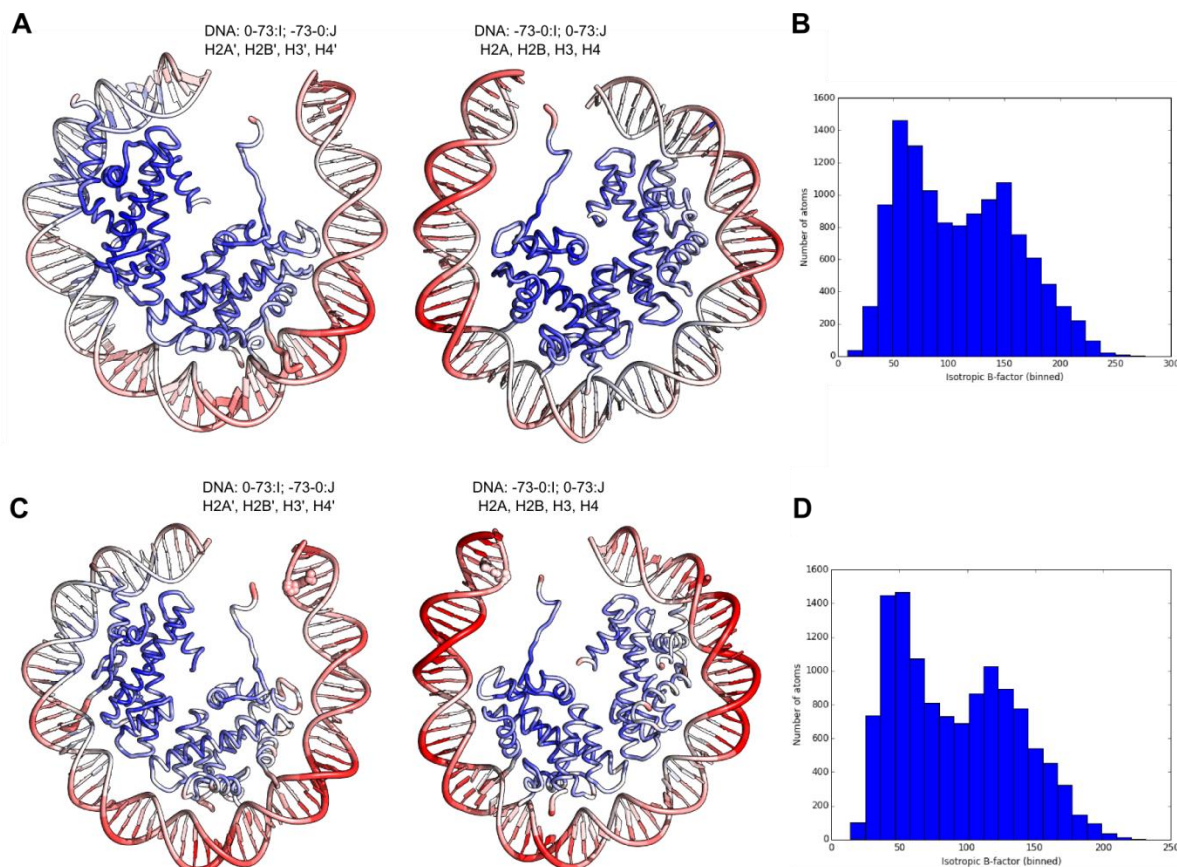
Figure 4.25: B-factors of the crosslinked NCPs. For better visibility, the NCPs in **A)** and **C)** are split in halves with the indicated parts of the NCP shown. **A)** The B-factors for the single crosslink NCP in a color-coded worm representation. The thickness of the worm represents the B-factor value of the amino acid. Dark blue represents a B-factor of 0, white represents B = 100 and dark red worms have a B-factor of 200. The same color coding applies to **C)**, where the B-factors of the double crosslinked NCP are shown. **B)** and **D)** Bimodal B-factor distributions with the histone octamer having average B-factors around 50 – 60, whereas the average DNA B-factors are 125 – 150 in the NCP-x structure and the NCP-xx structure respectively. The binned B-factor distributions were calculated in Phenix **[101]**.

4.3.1. Structural isomorphism

To analyze any structural differences in the NCPs due to the introduction of the crosslinks, the structures of the single (NCP-x) and double crosslinked (NCP-xx) NCP were compared to the high resolution NCP structure (PDB ID: 1KX5). The RMSD (root mean square deviation) between the two crosslinked and the reference NCP (PDB ID: 1KX5) was plotted after alignment based on the octamer part of the structures (MatchMaker, UCSF Chimera [116], Fig. 4.27) using standard Needleman-Wunsch algorithms [117]. The RMSD for each chain was calculated as

$$RMSD = \sqrt{\frac{1}{n}\sum_i d_i^2}$$

with n being the amount of amino acids i which is summed over and d being the distance between the C_α atoms (for proteins) or C4' atoms (for DNA) which are compared. The RMSD values for the single chains are shown in Table 4.13.
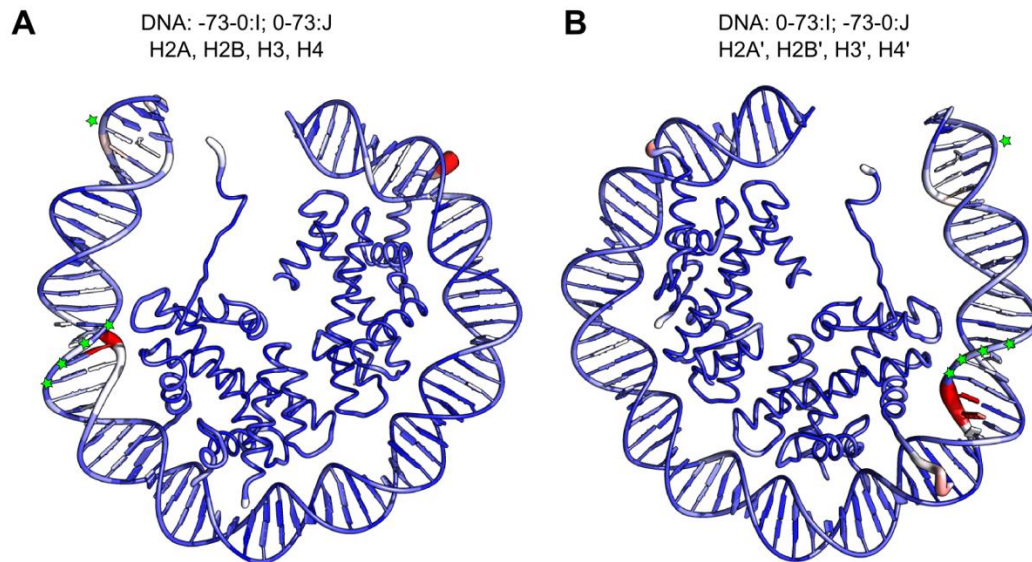
Figure 4.26: RMSD of the two halves of the structure of NCP-x compared to 1KX5. **A)** Nucleotides $0 - 73$.J and nucleotides -$73 - 0$.I as well as chains A – D. **B)** Nucleotides -$73 - 0$.J and nucleotides $0 - 73$.I as wells as chains E – H. The thickness of the worm representation corresponds to the RMSD to the aligned structure of the high resolution α-satellite NCP (PDB ID: 1KX5). The color code also represents the RMSD of the backbones (C_α for protein and C4' for DNA) ranging from 0 (dark blue) through white (2.0) to red (4.0). The nucleotides with altered sequence to the original α-satellite sequence are labeled with a green asterisk.

Figure 4.27: RMSD of the two halves of the structure of the double crosslinked NCP-xx. **A)** Nucleotides $0 - 73$.J and nucleotides -$73 - 0$.I as well as chains A – D. **B)** Nucleotides -$73 - 0$.J and nucleotides $0 - 73$.I as well as chains E – H. The thickness of the worm representation corresponds to the RMSD to the aligned structure of the high resolution α-satellite NCP (PDB ID: 1KX5). The color code also represents the RMSD of the backbones ($C\alpha$ for protein and C4' for DNA) ranging from 0 (dark blue) through white (2.0) to red (4.0). The nucleotides with altered sequence to the original α-satellite sequence are labeled with a green asterisk.

Chain	Identity	Residue range NCP-x	Residue range NCP-xx	RMSD NCP-x	RMSD NCP-xx
A	H3	38 - 135	39 – 134	0.590	0.578
B	H4	25 - 102	24 – 101	0.173	0.511
C	H2A	15 - 120	15 – 118	0.253	0.354
D	H2B	26 - 122	29 – 125	0.230	0.426
E	H3'	39 - 135	39 – 133	0.249	0.525
F	H4'	19 - 102	15 – 102	0.241	0.711
G	H2A'	14 - 119	14 – 118	0.234	0.376
H	H2B'	24 - 122	32 – 124	0.363	0.397
I	DNA	-73 – 73	-73 – 73	0.442	1.000
J	DNA	-73 – 73	-73 – 73	0.473	0.951

Table 4.13: RMSD values for the individual chains of the crosslinked NCPs. The indicated range of amino acids/nucleotides of the crosslinked NCPs was compared to the high resolution α-satellite structure (PDB ID: 1KX5). The RMSD is calculated based on the Cα atoms in proteins and C4' in DNA according to equation (1).

The DNA strands (chains I and J) tend to have larger RMSDs than the histone chains which can be explained by the larger amount of mobility of the DNA molecules as seen in the comparably higher B-factors. There are two positions in the DNA sequence where alterations were introduced. Firstly, the convertible nucleotide is covalently bound to H3, potentially leading to distortions in the DNA path towards the DNA termini. The RMSD for the DNA termini, however, are not elevated which indicates no structural changes due to the DNA:H3 crosslink. Secondly, the sequence at SHL +/-5.5 was altered to accommodate the restriction site necessary for the introduction of the oligonucleotide. These are also the regions of largest RMSD. The change in base identity cannot directly be responsible since only the C4' atoms of each base are used for the RMSD calculation. The change in sequence can, however, alter the local DNA parameters, which might explain the high RMSD values in that region. Importantly, the change of sequence created two GG steps, which form binding sites for divalent metal ions (section 4.3.4).

4.3.2. Crystal packing and end-to-end stacking

The crosslinked nucleosomes crystallized in the same space group (P2$_1$2$_1$2$_1$) as the 147 bp α-satellite NCP (PDB ID: 1KX5). The NCP molecules with pseudo two-fold symmetry can pack in the crystal in two different orientations leading to two-fold disorder. For palindromic 147 bp DNA sequences with a base pair positioned on the molecular dyad axis, the only molecular difference between the two orientations is the identity of the central base pair. In the case of the α-satellite sequence, the central base pair is an A·T base pair, and the fact that the two orientations are averaged leads to a su-

perposition of an A·T with a T·A base pair resulting in a tubular electron density for the central base pair plane (Fig. 4.28).
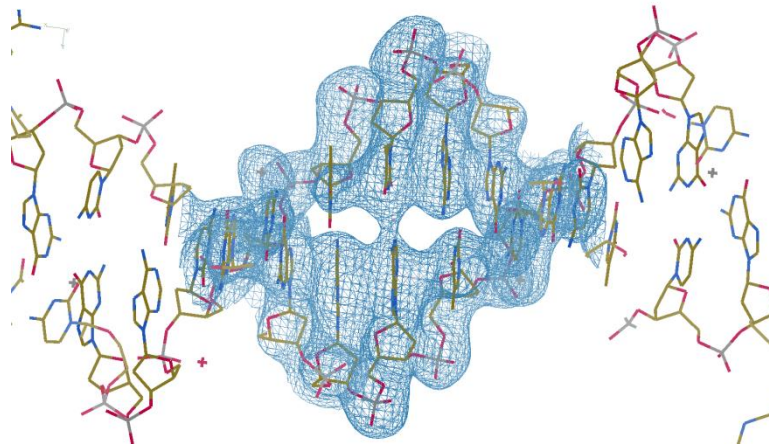
Figure 4.28: Electron density of the σ-weighted $2F_o$-F_c map around the central base pair (A·T) contoured at 1.4 σ. The density is the averaged electron density of the two possible orientations of the dyad base pair and therefore overlaps at the base pair center. The electron density is overlaid over the modeled structure (shown in stick representation). Only one orientation of the dyad base pair is modeled.

The two-fold disorder poses a large obstacle for obtaining high resolution structural information for non-palindromic DNA sequences (such as the 601 sequence), for which the electron density is a superposition of two orientations and shows poor quality at positions of the DNA where the two halves of the nucleosome deviate in sequence or in form. On the other hand, the previously published crystal structure of the NCP containing the MMTV-A promoter sequence shows substantial crystal packing asymmetry [118], yet the reason for the asymmetric packing remains elusive.

One crystal contact which is formed in the NCP X-ray crystal structure is portrayed in Fig. 4.29 and shows the pseudo-continuous stacking of the DNA termini of adjacent nucleosomes. The electron density at the DNA termini is shown in Fig. 4.30. Since the crosslinks between the histone octamer and the DNA are close to the DNA ends, it was important to show that the introduction of the crosslink did not disturb the important crystal contact between DNA ends. The RMSD between the crosslinked NCP structures and the high resolution NCP structure (PDB ID: 1KX5) is not elevated at the DNA termini (Fig. 4.26, 4.27), implying no distortions due to the introduced crosslinks.
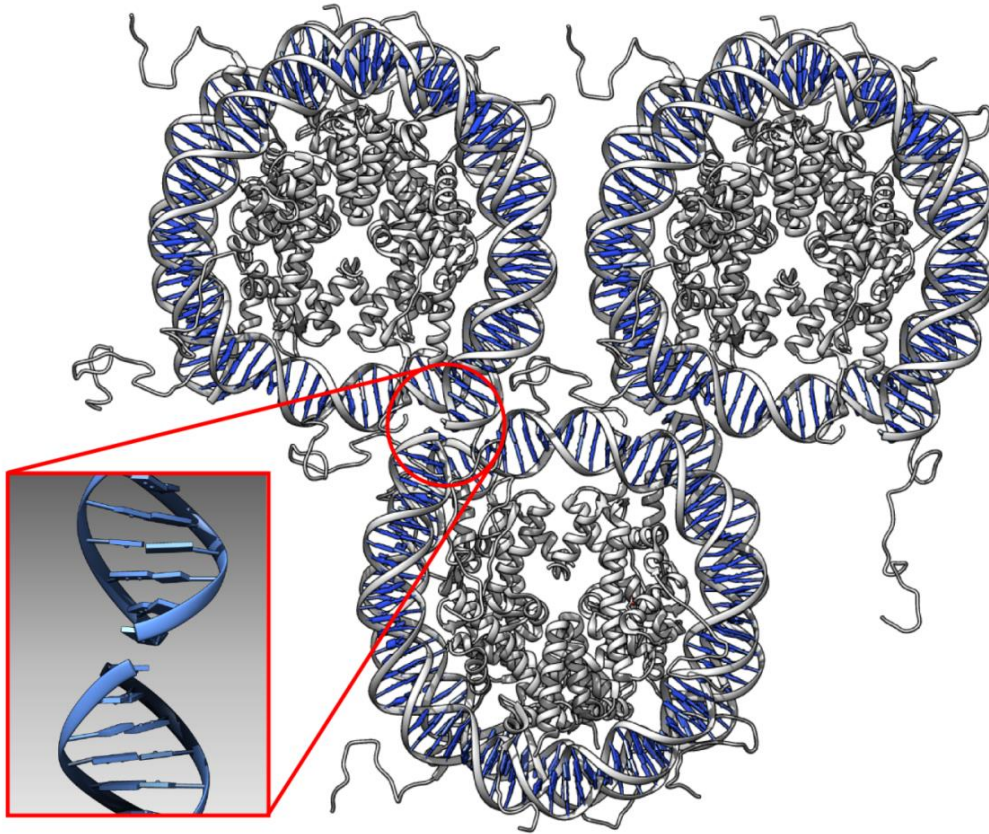
Figure 4.29: Pseudo-continuous end-to-end stacking of DNA termini in the NCP X-ray crystal structure. One contact site is highlighted (red circle) and again enlarged looking perpendicular to the DNA double helical axis as well as to the NCP super helical axis (blow-up). The 5´-ends of the DNA are dephosphorylated.
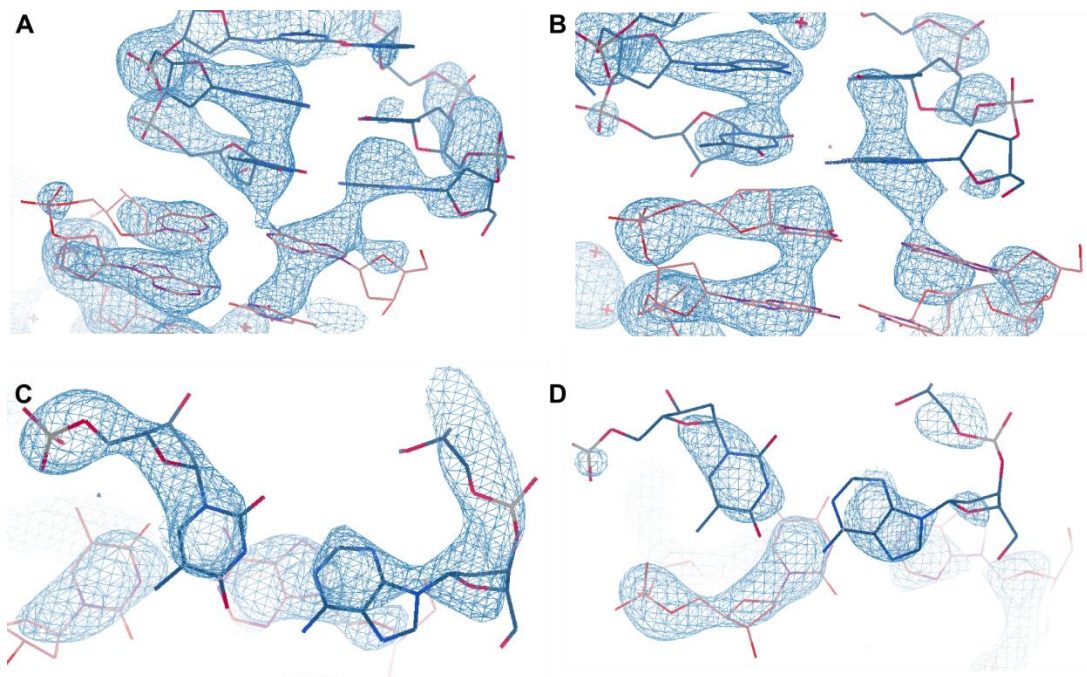
Figure 4.30: Electron density at the DNA termini which form a pseudo-continuous stacking interface. **A,C)** and **B,D)** The two interfaces are depicted perpendicular to the DNA axis (upper panels) and along the DNA axis (lower panels). The DNA molecules from the adjacent nucleosomes are colored orange and blue, respectively. The $2F_o-F_c$ map at the termini is contoured at 1.5 σ.

Other crystal contacts within the NCP crystal involve the sugar-phosphate backbone near SHL 6.5 straddling the backbone of an adjacent NCP molecule at SHL -2.5. This interaction is mediated by a Mn^{2+} ion (blue sphere in Fig. 4.31). A third type of crystal contact is established between SHL -2 and SHL -6 (Fig. 4.31, lower panel). The regions involved in crystal contacts tend to have a lower B-factor, due to reduced flexibility (compare Fig. 4.25).
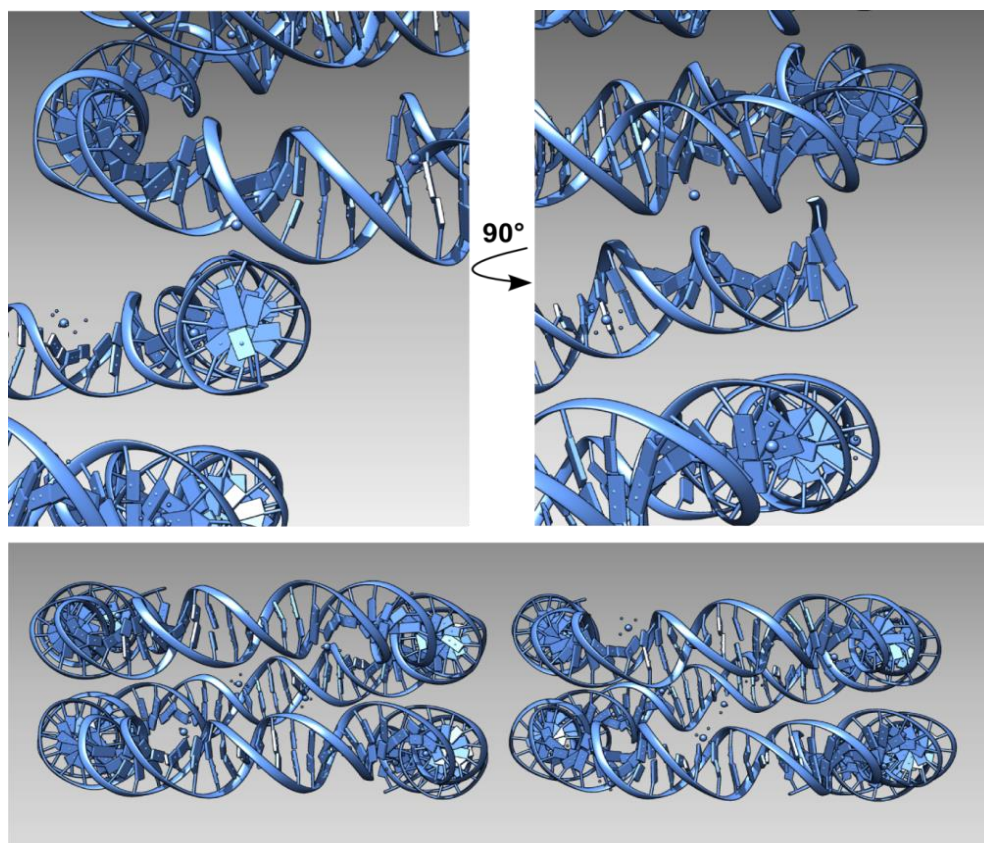
Figure 4.31: Crystal contacts between adjacent NCP molecules in a nucleosome crystal. The top two images display the contact between SHL 6.5 on one NCP molecule and SHL -2.5 on an adjacent NCP from two perspectives rotated by 90°. The sugar-phosphate backbones are inserted into the neighboring NCPs grooves with a Mn^{2+} ion bridging the contact. The lower panel shows the third type of crystal contact where two adjacent NCPs form a contact between their SHL -2 and the SHL -6 position of the neighboring NCP.

4.3.3. Crosslinks

The total X-ray exposure of the crystals had to be carefully chosen in order to be able to visualize the disulfide crosslinks in the crystal structure. The method used to obtain the high resolution structure of the NCP (PDB ID: 1KX5) consisted of maximal exposure of single crystal positions until a drop-off in resolution was observed [10]. Although this approach maximizes the resolution obtained, it would most likely destroy the disulfide crosslink, as it is the most vulnerable chemical group with respect to radiation damage due to the large absorption cross-section of sulfur atoms to X-rays in the 1 Å-wavelength range [119]. Therefore, a total dose per crystal position of 10 MGy was chosen as a con-servative estimate by taking the beam cross section and the photon flux into consideration.

4.3.3.1. The H3:DNA disulfide crosslink

The disulfide crosslink which was formed between the convertible dG* and the H3 N-terminal tail was not found to alter the overall structure of the nucleosome, although the formation of the cross-link forces the region N-terminal to the crosslink into a different conformation. The 1KX5 structure of the NCP shows the sidechain of R40 pointing away from the DNA terminus towards SHL 1 (opposite of the crosslinking location). The electron density at the DNA termini is comparatively low, but the continuous electron density for the DNA:H3 crosslink can be observed in the structure (Fig. 4.32 A,B).
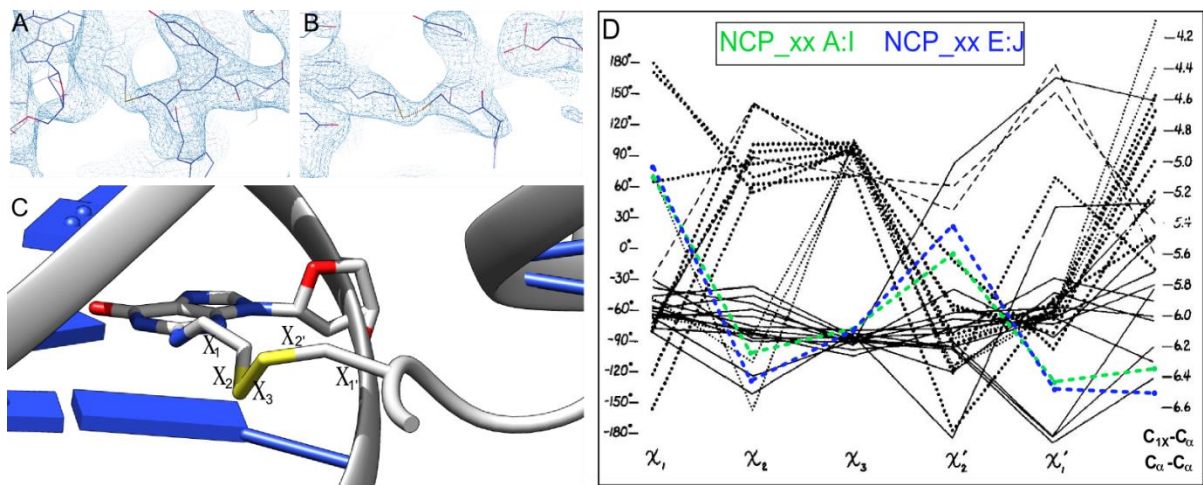
Figure 4.32: Analysis of the DNA:H3 crosslink in the double crosslinked NCP structure. **A)** Electron density along the DNA:H3 crosslink contoured at 0.7 σ of the $2F_o$-F_c map. **B)** Crosslink seen from different angles rotated by 90°. The model was build into the density and is shown in stick representation. **C)** A larger representation of the DNA:H3 crosslink region. The disulfide dihedral angles (χ_1, χ_2, χ_3, χ_2' and χ_1') are labeled. **D)** The dihedral angle distribution for the two DNA:H3 crosslinks in the NCP-xx structure overlaid on a dataset of solved disulfide crosslink structures (from **[120]**).

The crosslinks at both DNA ends vary slightly, but generally follow the same path. The crosslink as it was modelled with two χ-angles positive and three negative, cannot be considered a left-handed spiral which would represent the most often found disulfide geometry [120]. Their χ_3 angles are close to –90° indicative of a stable disulfide bridge.

4.3.3.2. The N38C disulfide crosslink

The disulfide crosslink between the two copies of H2A, via an N38C point mutation, crosslink the two H2A/H2B heterodimers together. Amino acid 38 lies in the L1 region which, together with its sym-metrical counterpart on the second copy of H2A, forms the lone contact between the two H2A/H2B heterodimers. The H2A N38C crosslink stabilizes the assembled histone octamer and it was shown previously that it successfully prevents H2A/H2B dimer dissociation (T. Frouws, PhD thesis [1]). The geometry of the crosslink is shown in Fig. 4.33 and is in good agreement with the geometry observed in the crosslinked NCP structure containing the N38C disulfide but no DNA:H3 crosslink [1]. The di-

sulfide bridge follows a left-handed spiral with all χ-angles negative and the χ_3 angle close to -90°, indicating the absence of strain and therefore high stability of the disulfide.
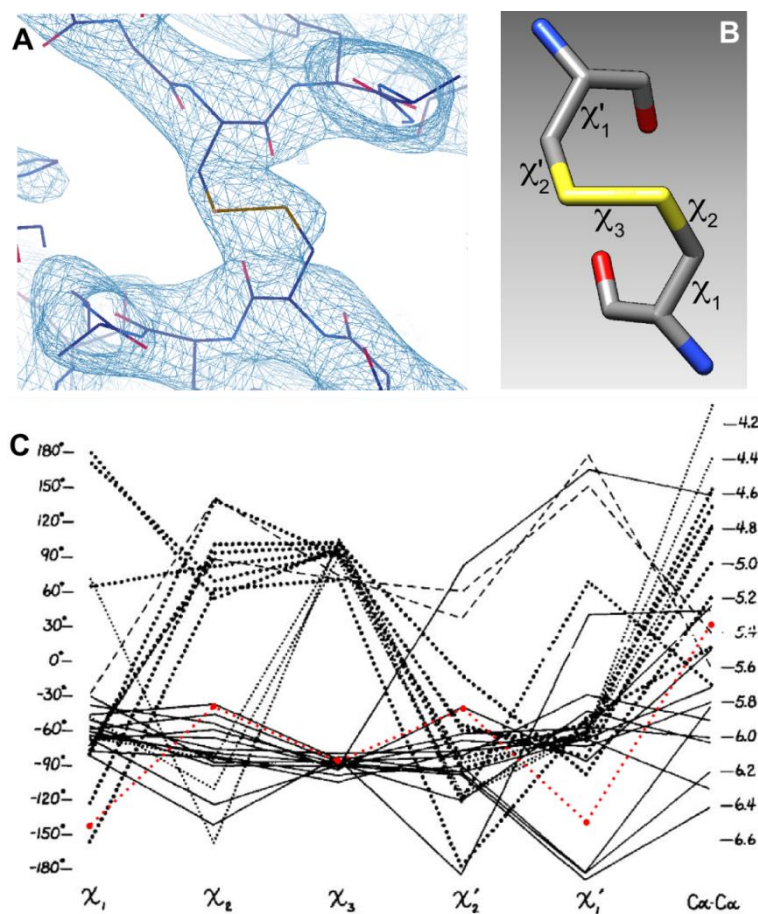
Figure 4.33: Geometry of the N38C crosslink in the double crosslinked nucleosome structure. **A)** Electron density ($2F_o$-F_c map contoured at 1.5 σ) in the crosslink region. **B)** Atomic model of the disulfide crosslink. Sulfur atoms are shown in yellow, oxygen in red and nitrogen in blue. The nomenclature for the dihedral angles (χ-angles) is shown next to the respective dihedral. **C)** Values for the dihedral angles plotted (red dots) overlaid on a dataset of 72 sets of χ-angles found in other structures featuring disulfide bonds (from **[120]**).

The occupancies of the two sulfur atoms were refined to 0.73 and 0.93, although they should be identical, since the structure is two-fold averaged and there should be no preference for one sulfur atom being ionized by strong X-ray radiation over the other. The average occupancy of 0.83 indicates some amount of radiation damage occurring at the applied dose.

4.3.4. SHL 5

The positions on the DNA which differ the most between the two crosslinked NCP stuctures compared to the 1KX5 structure are SHL +/-5 (Fig. 4.26, 4.27). These are adjacent to the parts of the sequence which were changed to introduce the AvaII restriction site. The two crosslinked structures show clear deviations in the backbone trajectories compared to 1KX5 (Fig. 4.34).
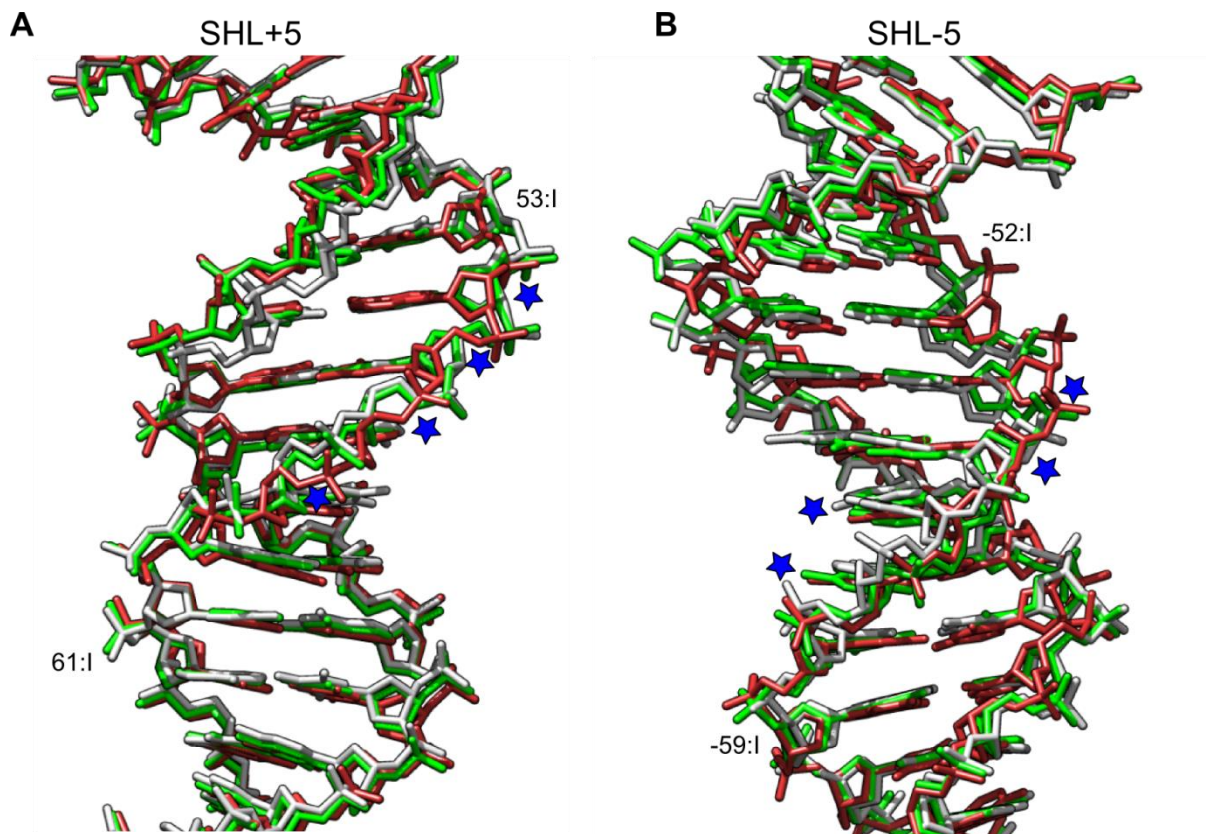
Figure 4.34: Comparison of the two crosslinked structures (NCP-x, green; NCP-xx, brown) to the high resolution α-satellite structure (1KX5) shown in white. **A)** Aligned SHL +5 positions of all three structures. **B)** SHL -5 position of all three structures. The positions of two base pairs are indicated for reference. The alignment was done by the Matchmaker algorithm based on the octamer part of the structures **[116]**. Blue stars mark the positions of changed DNA sequence.

The change in sequence to accommodate the AvaII restriction site led to the occurrence of two GG base pair steps, resulting in the formation of partial negative charge distributions between the O6 and N7 moieties of the two neighboring dG bases which favor the binding of a Mn^{2+} ion in the minor groove of a GG base pair step (Fig. 4.35). The Mn^{2+} ion probably leads to the observed distortions of the DNA backbone in the SHL +/-5 positions.

A

α-satellite   ACCTGCAGATACTACC

α-satellite convG   ACCTG**GTCC**TACTACC

Positions above sequence: 63 60 55 50

B

kcal mol⁻¹ scale: +110, +90, +70, +50, +30

C 59 58

D 55 56
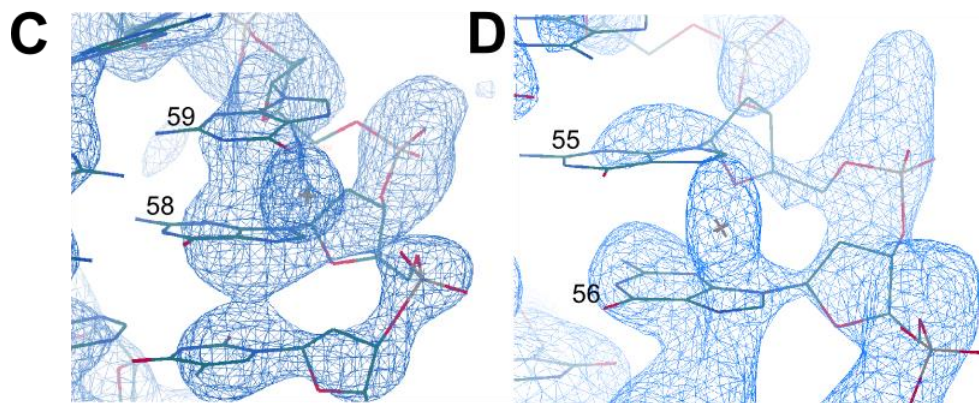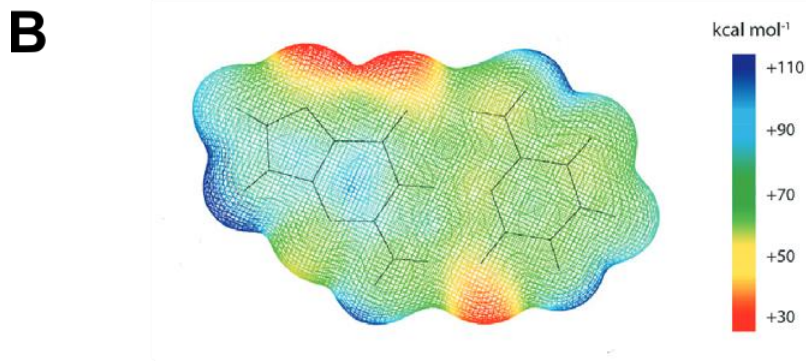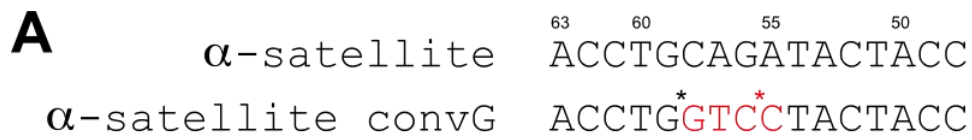
Figure 4.35: Newly created GG base pair steps with additional Mn²⁺ binding sites. **A)** Change in sequence of the crosslinkable DNA (α-satellite convG) to the original 147 bp α-satellite DNA used to crystallize the nucleosome. The positions of the bases are indicated above the sequence. The changed base pairs are colored red and the asterisks indicate the introduced GG steps, which bind Mn²⁺ ions. **B)** Partial charge distribution in the plane of a G·C base pair. The color code is given in the image (taken from **[121]**). **C)** Electron density of NCP-xx at the Mn²⁺ binding site (bases 58,59.I/J). The Mn²⁺ ion is shown as a white cross in the center of the image. **D)** Electron density of the the Mn²⁺ ion bound at the second introduced GG step. The base pair numbers are as indicated in the images.

4.4. Crosslinking of the Met16 promoter nucleosome

To test the assembly and crosslinking of nucleosome core particles loaded with DNA sequences that have not been crystallized before, the sequence of the NCP at the promoter position of the Met16 gene in yeast was chosen as a target. The Met16 gene maps to a region on chromosome 16 spanning base pairs 877632 – 876847 (Fig. 4.36) and it encodes for a phosphoadenylyl-sulfate reductase [122]. The gene is regulated by a TATA box, 100 base pairs upstream of the TSS, located within the -1 nucleosome (Fig. 4.36). A transcription factor binding site (AP-1) is located just outside of the NCP sequence, assuming 147 base pairs on the NCP.

A 110 bp long sequence symmetrically arranged around the dyad position was commercially synthesized with BsaI restriction sites at positions 55 - 60 bp away from the dyad (Fig. 4.37). BsaI was chosen specifically to avoid changing the native DNA sequence, which was achieved by the fact that BsaI, as a type IIS restriction enzyme, cuts outside of its target sequence.
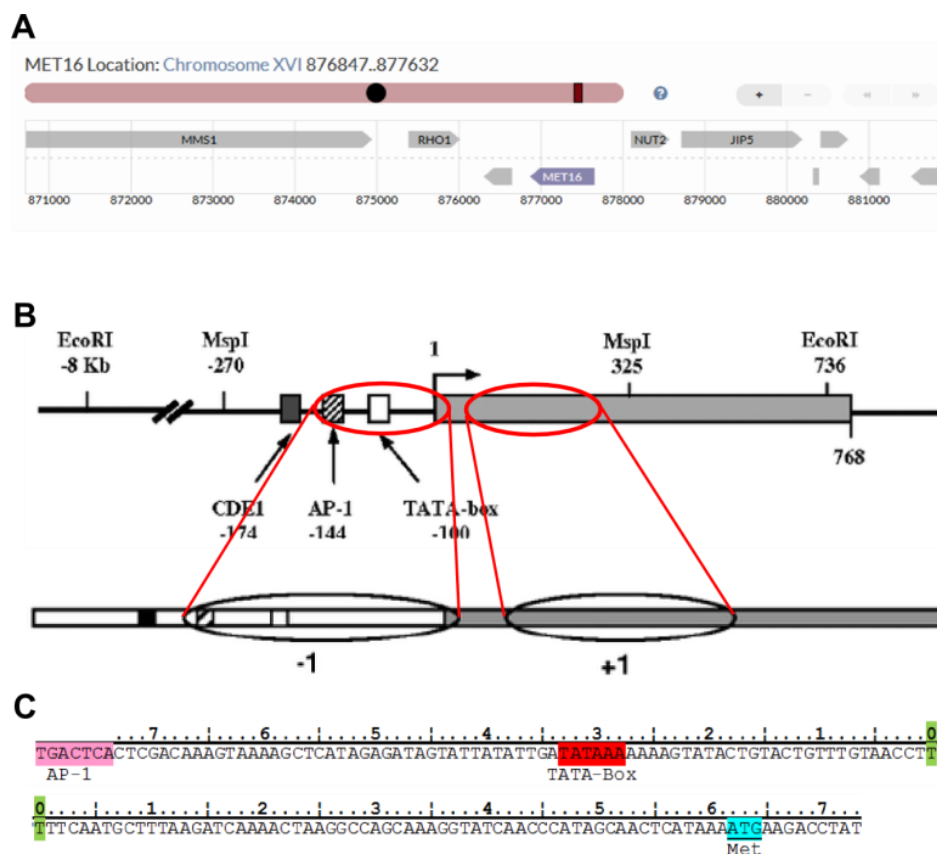
Figure 4.36: Location and sequence of the Met16 promoter. **A)** Location of the Met16 gene in the yeast genome. Image taken from www.yeastgenome.org. **B)** Organization of the Met16 promoter region. A TATA box (white box) is located 100 bp upstream of the TSS (arrow). An AP-1 binding site (striped) is located at position -144 from the TSS. The positions of the -1 and +1 nucleosome as mapped by micrococcal nuclease digest are indicated by ovals (adapted from **[123]** with permission). **C)** DNA sequence of the -1 promoter nucleosome containing the TATA box (red) and the AP-1 binding site just outside the NCP (pink) **[124]**. The dyad position is highlighted green and the TSS is shown in cyan.

In the yeast nucleosome map by Broogard *et al.*, the dyad postions found at the Met16 promoter fall mainly on two highly populated translational positions (highlighted in red in Table 4.14) [60]. The two positions are surrounded by several less populated nucleosome dyad positions. The exact dyad positions are given in Table 4.14.

Chromosome	Dyad position	NCP score	NCP score/noise ratio
chrXVI	877683	0.85	0.512
chrXVI	877684	0.95	0.579
chrXVI	877694	12.68	7.827
chrXVI	877706	0.76	0.469
chrXVI	877714	4.11	2.704
chrXVI	877720	0.62	0.428

Table 4.14: List of all dyad positions mapped in the *S. cerevisiae* genome within the Met16 gene promoter region **[60]**. The NCP score is a measure of the relative amount of nucleosomes centered at this position.

The sequence with its dyad positioned at base pair 877694, showing the highest NCP score, was chosen for DNA production, although it was not clear if the large NCP score was due to the sequence being a strong positioning sequence or if the promoter nucleosome is positioned *in vivo* by the action of remodeling factors or the pre-initiation complex (PIC).

The central 110 base pairs of the sequence were ordered as a DNA String® (Life Technologies) with 30 base pairs of sequence homology to the acceptor plasmid on each end for SLIC cloning into the pUC57 vector for DNA tandem repeat cloning (section 3.2.9). BsaI restriction sites for the seemless ligation of oligonucleotides containing convertible dG* were inserted into the Met16 promoter sequence about 20 base pairs away from both ends. PstI and BamHI sites were placed outside of the BsaI sites within a stretch of 25 base pairs homologous to the target plasmid pUC57 for SLIC cloning. The sequence is shown in appendix 8.6.
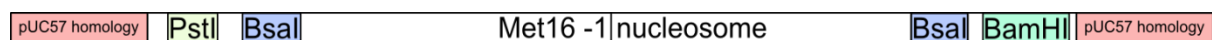
Figure 4.37: Design of the ordered DNA String® containing the Met16 -1 nucleosome sequence. The sequence is truncated about 55 bp from the dyad position by BsaI restriction sites. PstI and BamHI restriction sites were inserted for tandem repeat cloning and 25 base pairs of homologous sequence to the BglII/BamHI cut pUC57 were added at each end.

Tandem repeat cloning of inserts up to 16 copies was done as described in section 3.2.9. The final plasmid containing 16 copies was sequence verified using M13 forward and reverse sequencing primers. Large scale production of insert DNA was done essentially as described in section 3.2.10, except the EcoRV digest is replaced with a BsaI digest in 1x Cutsmart buffer. BsaI cuts the carrier plasmid (pUC57) once inside the BLA gene, which leaves the vector in two pieces of 1,300 – 1,400 bp in size.

The oligonucleotides containing the convertible nucleotide were ordered from Microsynth (Balgach, Switzerland) and treated post-synthesis to exchange the 2-F leaving group with the cystamine moiety required for crosslinking as described in section 3.2.15. The PAGE purified oligonucleotides were subjected to MALDI analysis and showed a single peak at the expected molecular weight. The annealing and purification of double stranded oligonucleotide was done as described in section 3.2.17. The purified annealed oligonucleotides for both core DNA ends were ligated to the core DNA in two-fold excess. The ligation reaction was purified on an anion exchange column and the fractions containing pure 147 bp Met16 promoter DNA were pooled and ethanol precipitated yielding approximately 250 µl of pure Met16 DNA with a convertible dG* at position 70.I/J. Reconstitution of Met16 nucleosomes was done at a 1.05:1 excess of histone octamer (containing H3 R40C C110A) by stepwise dialysis from TEK1.4 to TEK10 (as described in section 3.4.1). The reconstituted Met16 nucleosomes were crosslinked during a 120 min heat-shift at 37 °C by addition of sodium tetrathionate to a final concentration of 4 mM. The crosslinked Met16 NCPs were purified over an anion exchange column (60 ml 5PW DEAE, Tosoh Biosciences) with the same gradient as the crosslinked α-satellite NCPs (Table 3.7). The peak fraction containing crosslinked Met16 NCP was concentrated first in an Amicon Ultra-4 concentrator (10,000 MWCO, Millipore) and then further concentrated in a Vivaspin500 (5 kDa MWCO, Sartorius).

To test the behavior of crosslinked Met16 NCP compared to an uncrosslinked sample (by addition of 10 mM DTT), samples of the crosslinked NCP and of Met16 NCP with the disulfide released by DTT were analyzed on a 5% native PAGE gel and an 18% SDS PAGE (Fig. 4.38). The shifting behavior of crosslinked Met16 NCPs compared to Met16 NCPs with the crosslink released by DTT shows the stabilizing effect of the crosslink, as judged by the sharpness of the band. The more diffuse shifting behavior upon release of the crosslink might indicate redistribution to different rotational positions of the Met16 DNA on the histone octamer. No free DNA occurs, indicating that the Met16 DNA is still bound to the histone octamer.

The purified sample was concentrated to 15 µl yielding an NCP concentration of 3 mg/ml. The sample was set up for crystallization in a subset of conditions from the standard NCP crystallization screen (section 2.6.1). Several conditions yielded crystals ranging in size from 50 × 50 µm up to 200 × 200 µm. Fig. 4.38 E-G shows different sizes and morphologies of the Met16 NCP crystals. All crystals feature pronounced birefringence. Initial diffraction images obtained at the in-house Rigaku X-ray generator with a MAR detector yielded diffraction up to 3.2 Å. A complete dataset has yet to be collected at the Swiss Light Source synchrotron.
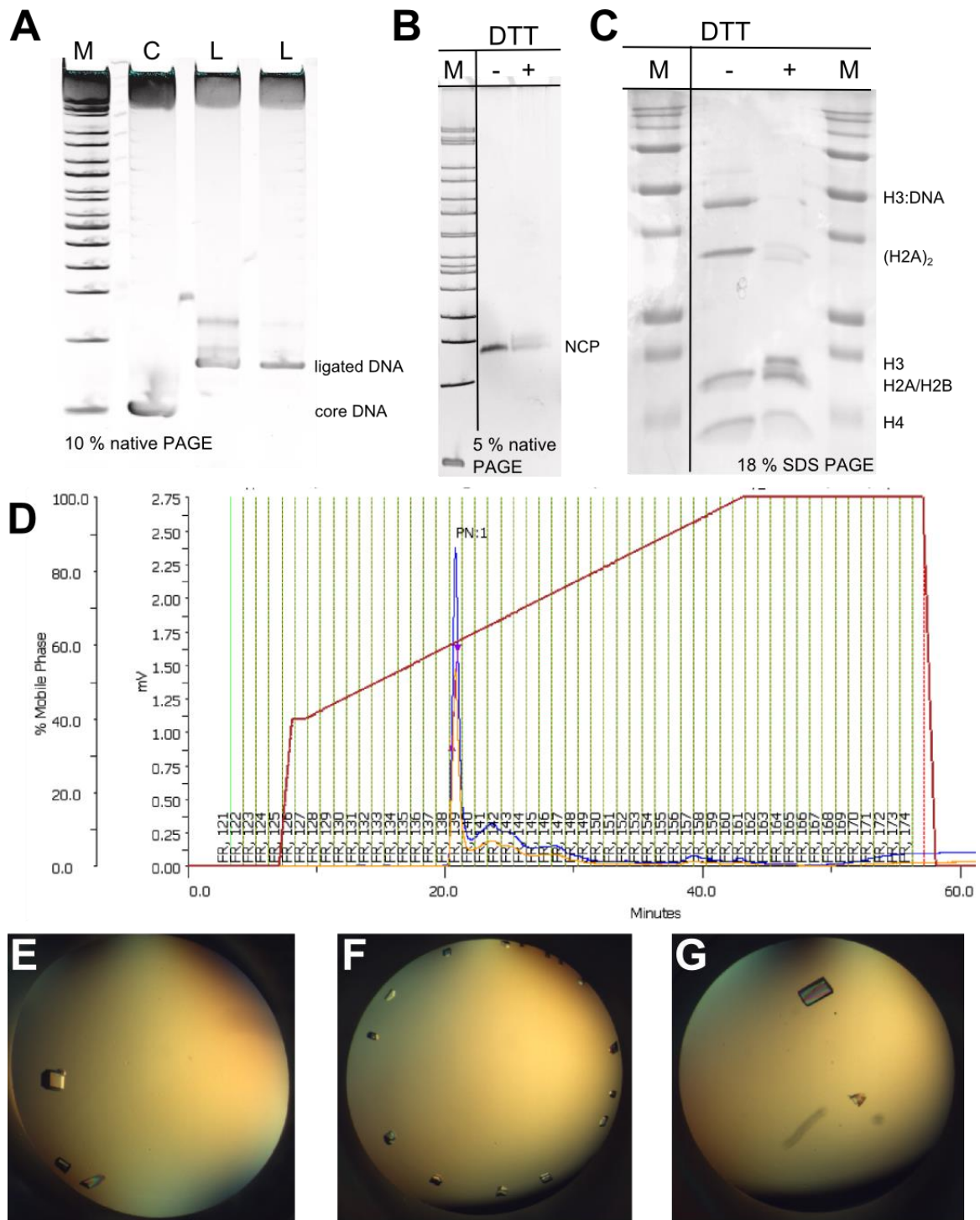
Figure 4.38: Production and crystallization of crosslinked Met16 NCPs. **A)** 10% native PAGE of the ligation reaction. M is 2-log ladder (New England Biolabs), C is the unligated core DNA (110 bp) and L is the ligation reaction loaded in two different amounts. Band identities as indicated on the side of the gel. **B)** 5% native PAGE gel of the crosslinked Met16 NCP with and without DTT. **C)** 18% SDS PAGE of the purified crosslinked Met16 NCP with and without DTT added. **D)** Elution profile of the crosslinked Met16 NCPs from the DEAE column. The crosslinked NCPs eluted in a single fraction (138). The elution gradient (in % buffer B) is shown in red. Fractions are separated by broken green lines. The blue line is the sample absorption at λ = 260 nm and the orange line is the sample absorption at λ = 280 nm. **E)** Crystals grown in screen solution containing 120 mM KCl, 180 mM $MnCl_2$, 40 mM Cacodylate pH 6.0. **F)** Crystals grown in screen solution containing 100 mM KCl, 170 mM $MnCl_2$, 40 mM Cacodylate pH 6.0. **G)** Crystals grown in screen solution containing 160 mM KCl, 160 mM $MnCl_2$, 40 mM Cacodylate pH 6.0.

5. Crosslinking of the ΔDot1:NCP complex

The X-ray crystal structure of the complex formed by Dot1 bound to the NCP has been elusive and several attempts to grow well-diffracting crystals have failed. Progress has been made in this lab by a previous lab member in establishing a site-specific disulfide crosslink between ΔDot1 and histone H3 (V. Vogirala, PhD thesis [2]). The crosslink was made between a cysteine point mutation (H3 T80C) introduced adjacent to the methylation target site of Dot1 (H3 K79) and a cysteine point mutation (S542C) on ΔDot1 located in motif VIII, a loop between helices $\alpha 10$ and $\alpha 11$, forming part of the active site. The crosslinked complex between Dot1 and NCP could be purified on a preparative gel with good separation of the one-bound state (NCP:ΔDot1$_1$) from the two-bound state (NCP:ΔDot1$_2$), although the dilution of sample during preparative gel purification repeatedly led to dissociation of DNA from the complex as evidenced by the presence of free DNA in the pooled peak samples from the preparative gel. The purified crosslinked sample was screened extensively using the NCCR crystallization facility and crystallization plates were set up in this lab. Initial crystal hits could be obtained for both the one-bound sample as well as the two-bound sample, although the hits for the one-bound state were not reproducible. Further screening around the condition that led to a hit for the two-bound state reliably gave small crystals (< 20 μm) but always occurred with heavy precipitation in the well. It was concluded that precipitation was probably caused by the presence of large amounts of free DNA in the sample due to dilution-driven DNA dissociation (V. Vogirala, PhD thesis [2]).

The introduction of the crosslink between the nucleosome DNA and the histone octamer, as described in this thesis, provides a potentially straightforward solution for the problem of DNA dissociation since it prevents the dissociation of DNA from the nucleosome caused by dilution. It was therefore combined with the existing crosslinking protocol for the ΔDot1 to H3 crosslink (V. Vogirala, PhD thesis [2]) as well as the H2A N38C crosslink (T. Frouws, PhD thesis [1]). The system is schematically shown in Fig. 5.1.
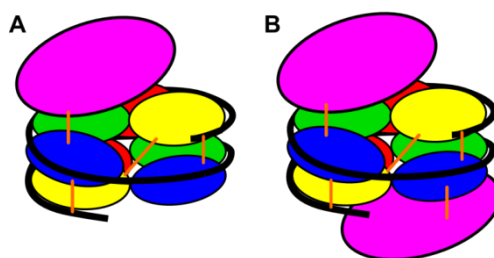
Figure 5.1: Schematic representation of the two species of one-bound Dot1 to NCP **(A)** and two-bound Dot1 to NCP **(B)**. The color coding is magenta for ΔDot1, black for DNA, H2A is yellow, H2B is red, H4 is green and H3 is blue. The crosslinks are represented by orange lines.

5.1. Preparation of ΔDot1 S542C

Transformation of the pET28a-based expression plasmid carrying the ΔDot1 S542C gene was carried out as described in section 3.3.2. The successfully transformed cells were grown, harvested and lysed. The soluble ΔDot1 protein was purified from the cleared lysate by a series of purification steps (as described in section 3.3.6.3). The first step was a metal affinity purification of ΔDot1 S542C via its N-terminal His-tag (6x His).
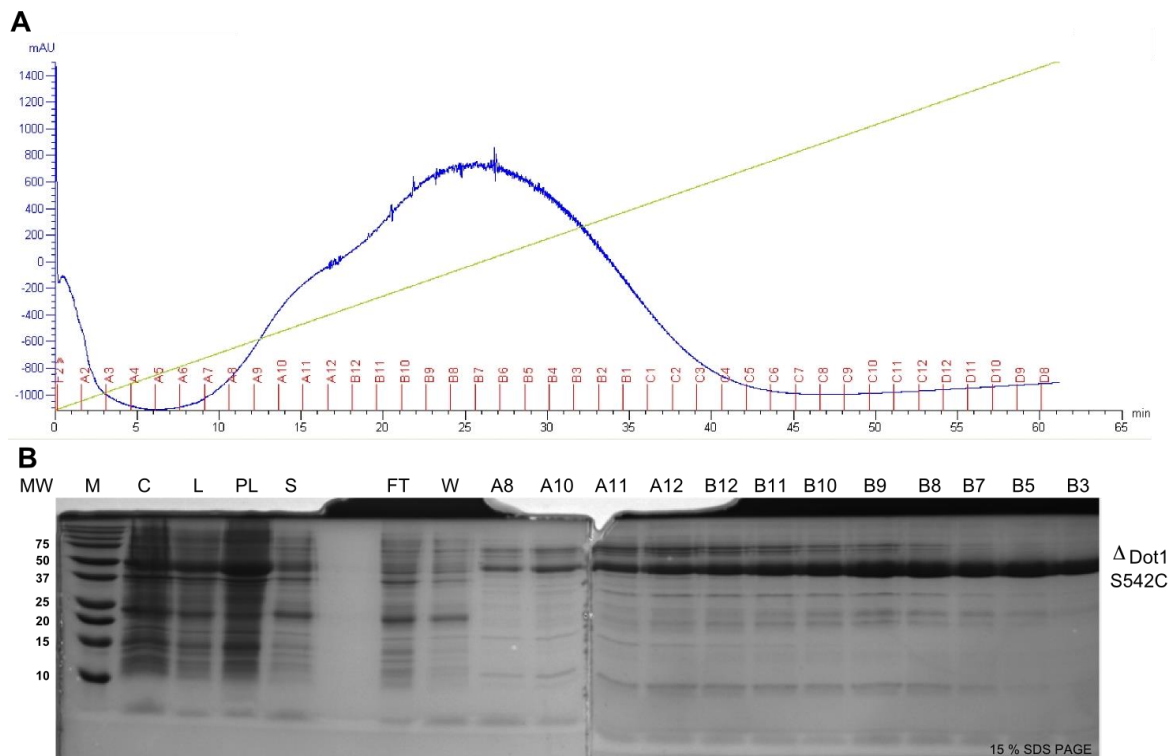
Figure 5.2: Metal affinitiy chromatography purification of overexpressed ΔDot1. **A)** First purification step of ΔDot1 S542C over a metal affinity chromatography column (5 ml Ni-NTA, GE Healthcare). The UV trace (λ = 280 nm) is shown in blue and the elution gradient running from 25 mM imidazole to 250 mM imidazole over 12 column volumes is shown in yellow. Fractions are indicated as red separators. **B)** SDS PAGE analysis of the purification steps (MW = molecular weight, M = protein marker, C = cell pellet, L = lysate, PL = pellet lysate, S = supernatant lysate, FT = flow-through from column loading, W = column wash and fractions as indicated). The position of ΔDot1 S542C running on the gel is indicated on the right side and the masses of the reference proteins are indicated on the left.

After pooling and concentration of the fractions containing ΔDot1 S542C, the sample was buffer exchanged to the starting buffer of the next purification step. A strong anion exchange column (5 ml HiTrap Q HP) was added in front of the cation exchange column to serve as a trap for anionic impurities. At pH 7.5 ΔDot1 S542C has a net positive charge and eluted from the Q column in the flow-through to then bind to the cation exchange column. The anion exchange column was detached from the cation exchanger prior to sample elution.
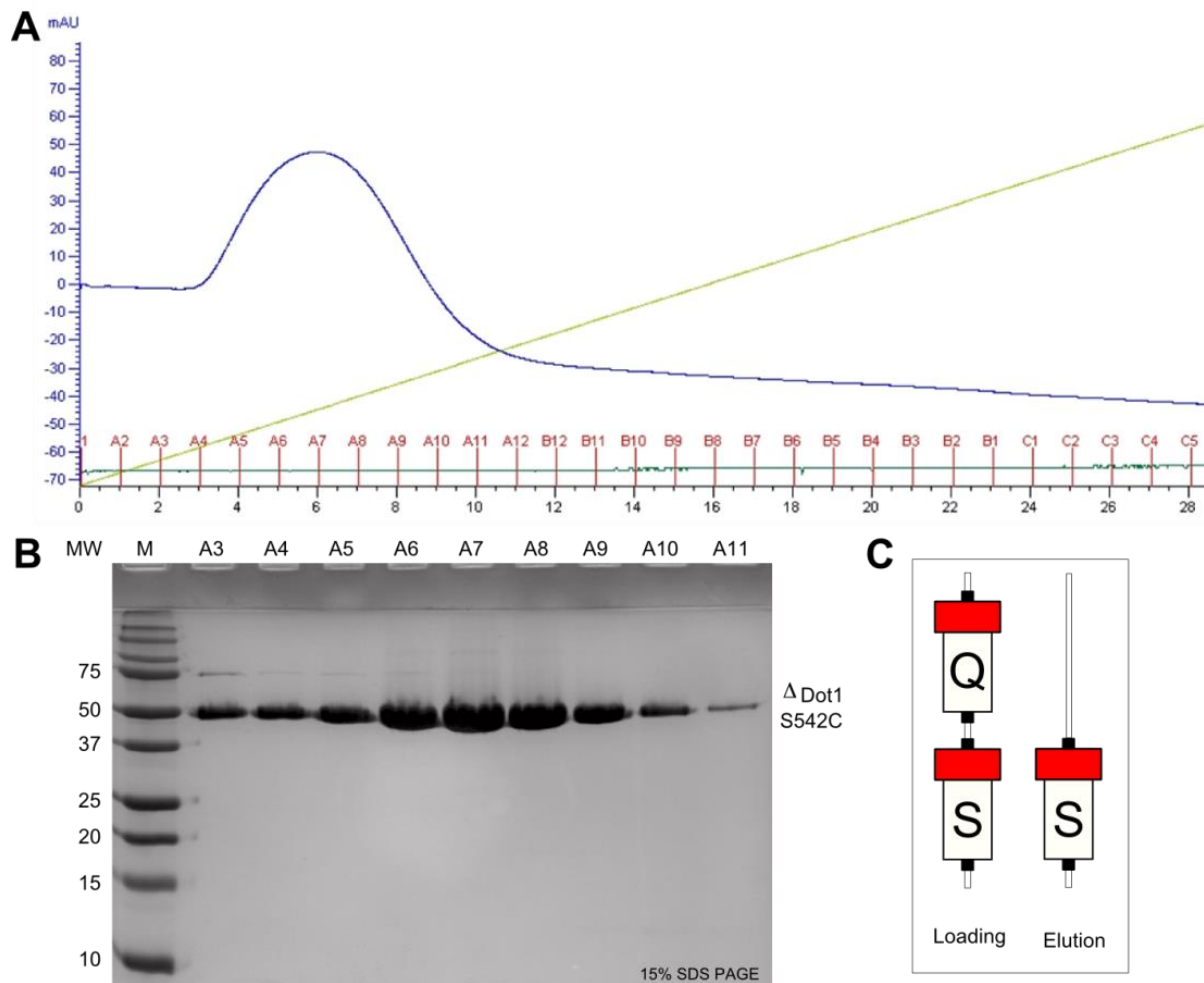
Figure 5.3: Ion exchange chromatography purification of ΔDot1. **A)** Elution profile (blue trace) of ΔDot1 S542C from a cation exchange chromatography column (5 ml HiTrap S HP, GE Healthcare) over a gradient from 200 mM NaCl to 1.5 M NaCl. **B)** 15% SDS PAGE analysis of the peak fractions. **C)** Arrangement of the anion exchange (Q) and cation exchange (S) columns during sample loading and elution. Flow in both cases was from top to bottom.

The pooled and concentrated fractions were supplemented with 1 µM S-adenosyl homocysteine (SAH) to prevent precipitation and then loaded on a 120 ml S-200 size exclusion chromatography column (GE Healthcare). The pure ΔDot1 fractions were pooled and concentrated and 1 µM SAH was added to prevent precipitation. The sample was stored at 4 °C for immediate use or at -20 °C for mid-term storage.
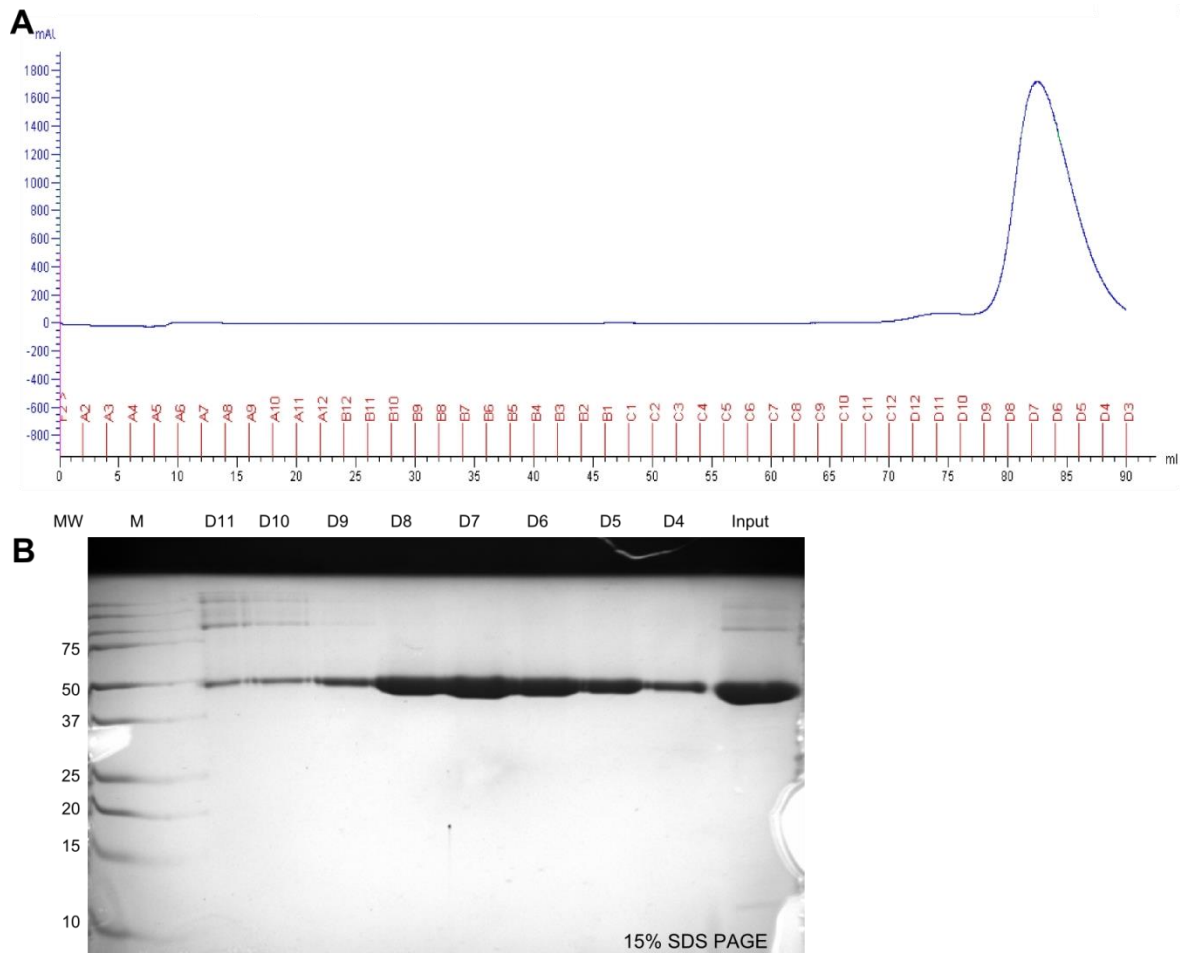
Figure 5.4: Size exclusion chromatography purification of ΔDot1. **A)** Final purification step of ΔDot1 S542C over a 120 ml S-200 size exclusion chromatography column (GE Healthcare). UV trace (λ = 280 nm) is shown in blue and fractions are indicated in red. **B)** Peak fractions as analyzed on a 15% SDS PAGE.

5.2. Crosslinking strategy

Compared to the protocol for the single crosslink between DNA and H3, there was the added complexity of ΔDot1-to-H3 crosslink formation. The protocol for crosslinking nucleosome DNA to H3 (even also including the H2A N38C crosslink) involves establishing the crosslink(s) during the heat-shift procedure which led to population of a single translational position of the DNA molecules. The crosslinking buffer had a basic pH (8.8) to facilitate disulfide exchange, and the crosslinking reaction of DNA:H3 ran efficiently at 37 °C and even at 55 °C. For the inclusion of the additional H3-to-ΔDot1 crosslink, several reaction parameters were tried to find conditions which yield optimal results. Different crosslinking conditions were tested for NCPs containing α-satellite DNA with a convertible dG* at position 70.I/J and histone octamer containing the H3 R40C/T80C/C110A point mutations. The NCP sample was either subjected to a 120 min long heat shift at 50 °C with subsequent addition of ΔDot1 and crosslinking at 4 °C and 22 °C, or crosslinked with ΔDot1 added during the heat-shift at 37 °C and 50 °C. For each crosslinking condition a range of ΔDot1 to NCP ratios (0:1, 1:1, 2:1 and 3:1)

was tested. After the crosslinking reaction, the samples were analyzed on a 5% native EMSA gel (Fig. 5.5).
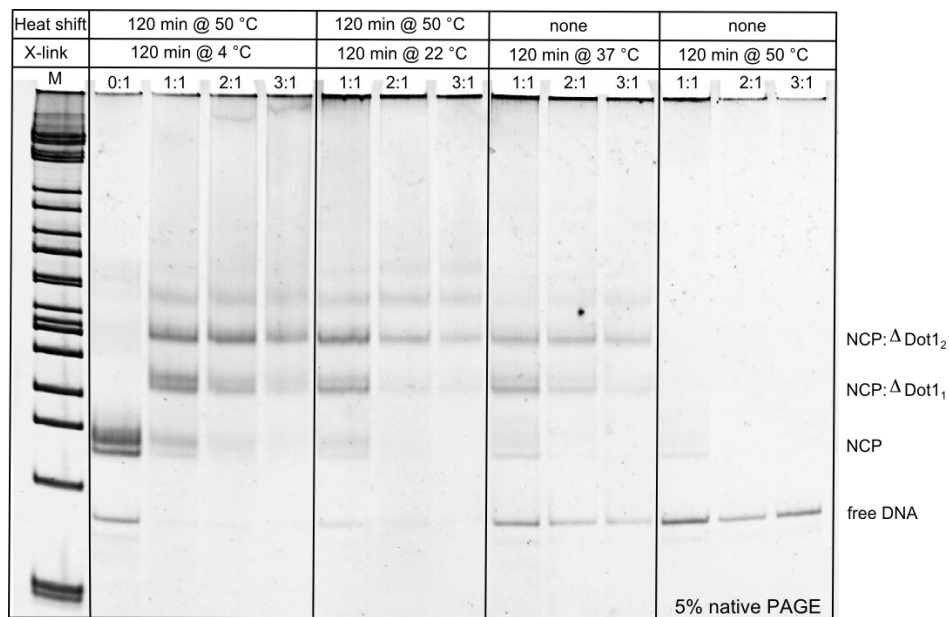
Figure 5.5: Different crosslinking conditions to test the ΔDot1 to NCP crosslinking. 2-log ladder (M) is loaded for size reference. The first block of samples was heat shifted for 120 min at 50 °C and then crosslinked for 120 min at 4 °C. The second block was heat shifted for 120 min at 50 °C and then crosslinked for 120 min at 22 °C. The last two blocks were crosslinked (at 37 °C and 50 °C respectively) with no prior heat shifting. ΔDot1 to NCP ratios of 1:1, 2:1 and 3:1 were tested for each condition.

With no prior heat-shift of the nucleosomes, the crosslinking protocol for ΔDot1 added at 37 °C led to dissociation of nucleosome DNA and subsequent precipitation of the sample. This problem was exacerbated at 55 °C. Crosslinking at 55 °C led to complete dissociation and only free DNA was visible on the gel (4th block in Fig. 5.5). Apparently, nucleosomes are destabilized by the high temperature conditions in the presence of ΔDot1 and sodium tetrathionate. The temperature of the crosslinking reaction after the initial heat-shift did not seem to influence the distribution of different sample species although the overall yield was higher when the crosslinking reaction took place at 4 °C (1st block in Fig. 5.5).

5.3. Purification of crosslinked ΔDot1:NCP complex

Purification of the separate stoichiometric species of ΔDot1 to NCP was shown to be effective on a preparative gel of adequate size and running conditions. The crosslink between ΔDot1 and H3 effectively counteracted the dissociation of ΔDot1 from the complex, although dilution of the complex during purification led to the dissociation of nucleosome DNA and subsequent precipitation of the entire complex under low salt conditions (Fig. 5.6 A), as evidenced by the appearance of free DNA in the pooled and concentrated peak fractions of the preparative gel.
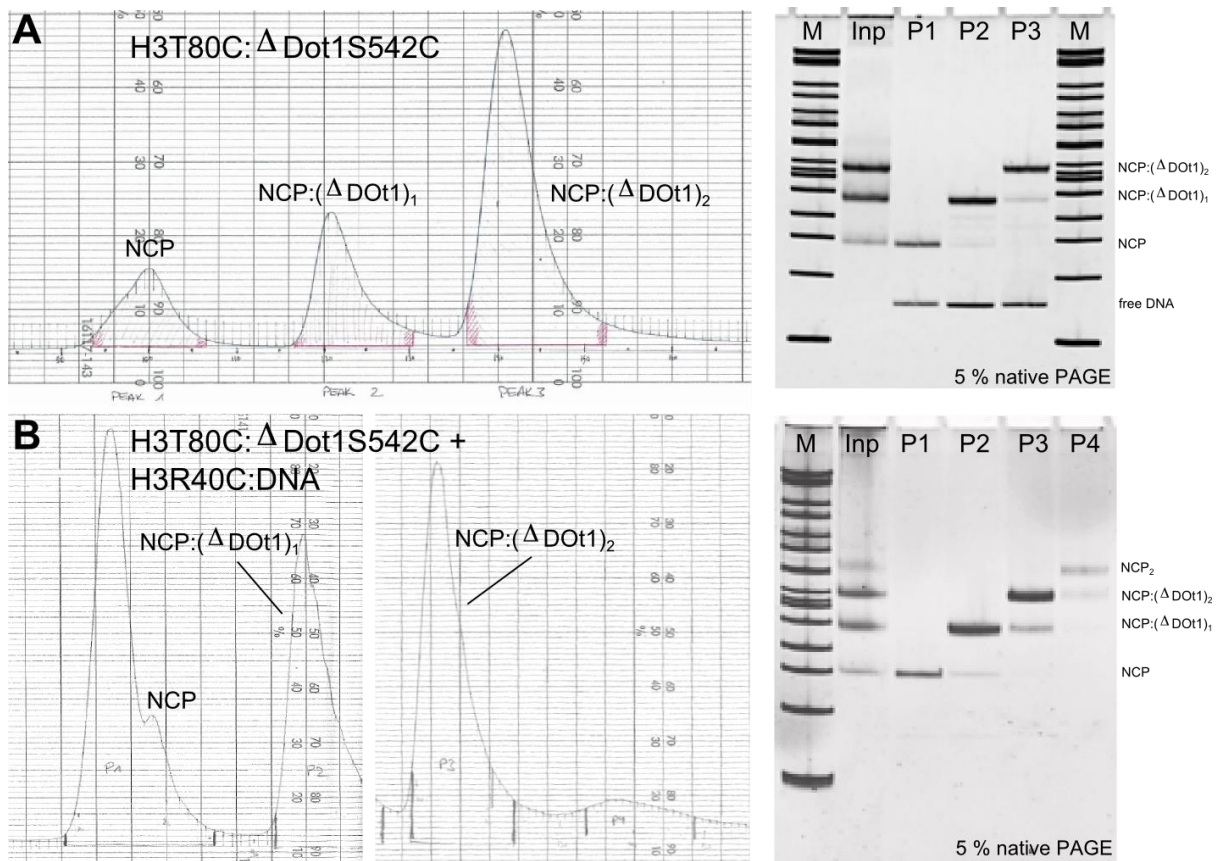
Figure 5.6: Preparative gel purification of crosslinked ΔDot1:NCP complex with and without the DNA:H3 crosslink. **A)** UV absorption (λ = 260 nm) for the purification of crosslinked ΔDot1S542C to NCP without the DNA:H3 R40C crosslink present (upper two panels). **B)** Purification of the complex with the introduction of the DNA:H3 R40C crosslink. The preparative gel consisted of a 5% native polyacrylamide matrix and was run in 0.25x TBE buffer. The fractions to be pooled from each eluted peak are indicated at the bottom of the UV trace.

With the introduction of the DNA to H3 crosslink, DNA dissociation could be effectively prevented. Samples of one-bound ΔDot1 to NCP as well as two-bound ΔDot1 to NCP of even higher purity were obtained after reinjection into a fresh preparative gel (Fig. 5.7). The different peaks from the preparative gel were analyzed on a 15% SDS PAGE and correspond to free NCP (P1), mostly NCP:ΔDot1$_1$ (P2), mostly NCP:ΔDot1$_2$ (P3), and mostly NCP dimers (P4) (Fig. 5.8).
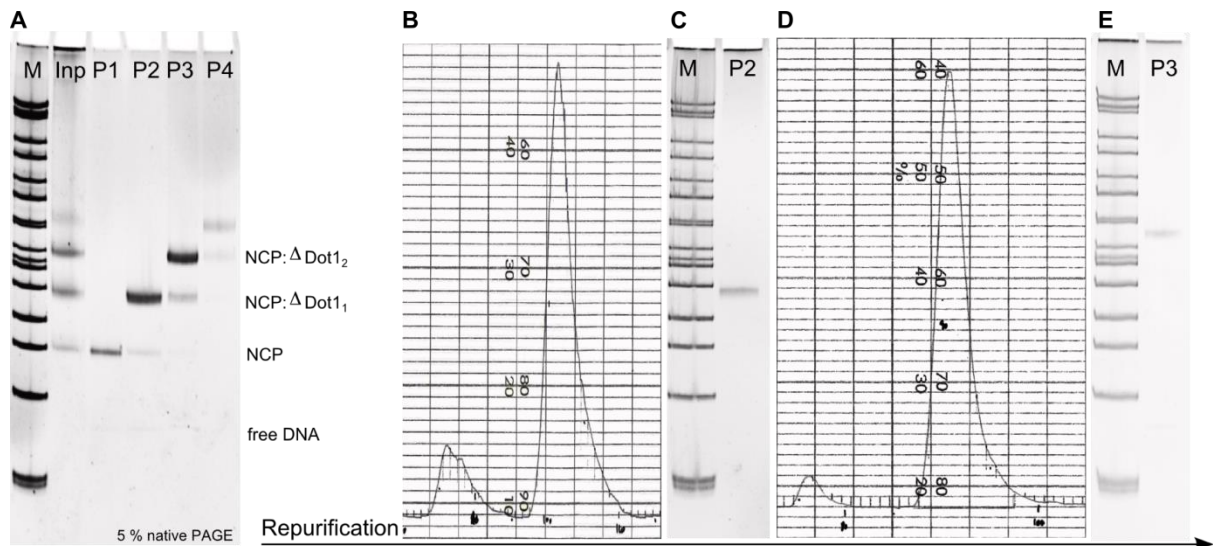
Figure 5.7: Pooled fractions of the preparative gel purification of the crosslinked complex between ΔDot1 and NCP. **A)** Pooled and concentrated peak fractions of a sample with the ΔDot1S542C:H3T80C crosslink and the additional DNA:H3 R40C crosslink. **B,C)** A second purification step was able to resolve the one-bound species and yielded very homogenous sample. The middle UV trace **(B)** (λ = 260 nm) and gel **(C)** show the repurification of the peak containing NCP:ΔDot1$_1$ (P2) and the UV trace **(D)** and gel **(E)** on the right side show the repurification of two-bound ΔDot1 to NCP (P3). All gels are 5% native PAGE gels.
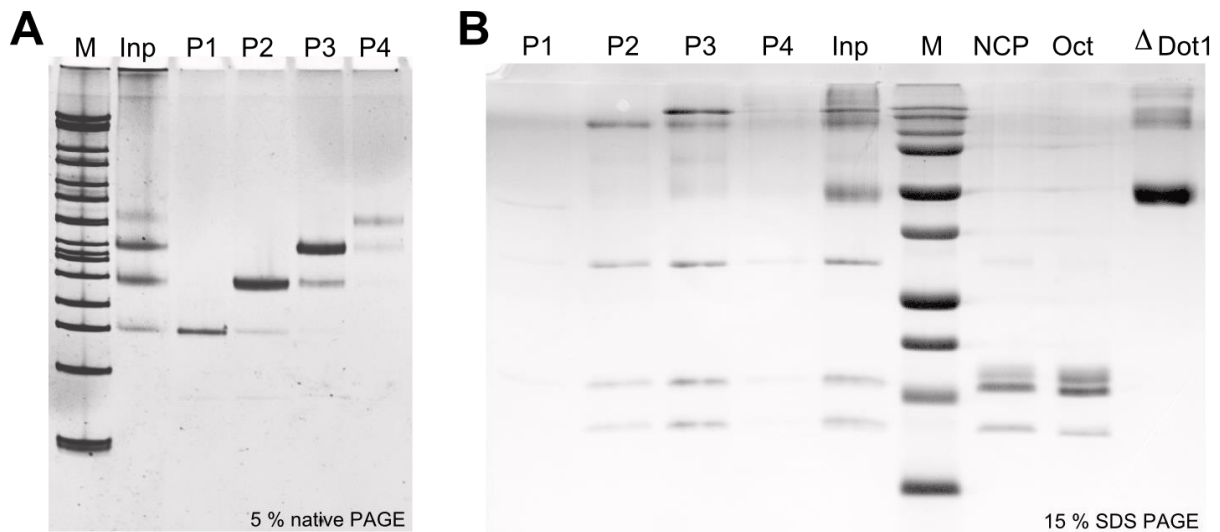
Figure 5.8: Peak fractions of crosslinked ΔDot1:NCP complex with H3:DNA complex eluting from the preparative gel. **A)** The peak fractions show no free DNA. **B)** SDS PAGE analysis of the peak fractions. The concentrations of P1 and P4 are rather low but seem to show no ΔDot1 present, which aligns with the assumption of P1 being NCP and P4 being a NCP dimer. Samples of one-bound and two-bound ΔDot1 to NCP can be resolved on the gel (P2 vs. P3). The absence of free ΔDot1 shows the high efficiency of crosslinking of ΔDot1 to H3.

The running conditions for the preparative gel purification of crosslinked ΔDot1 to NCP complex were important parameters for the separation of the different complex stoichiometries as well as for the stability of the complex upon sample dilution. Dilution during preparative gel purification could not be circumvented with the existing setup (Model 491 preparative gel, Bio-Rad), however.
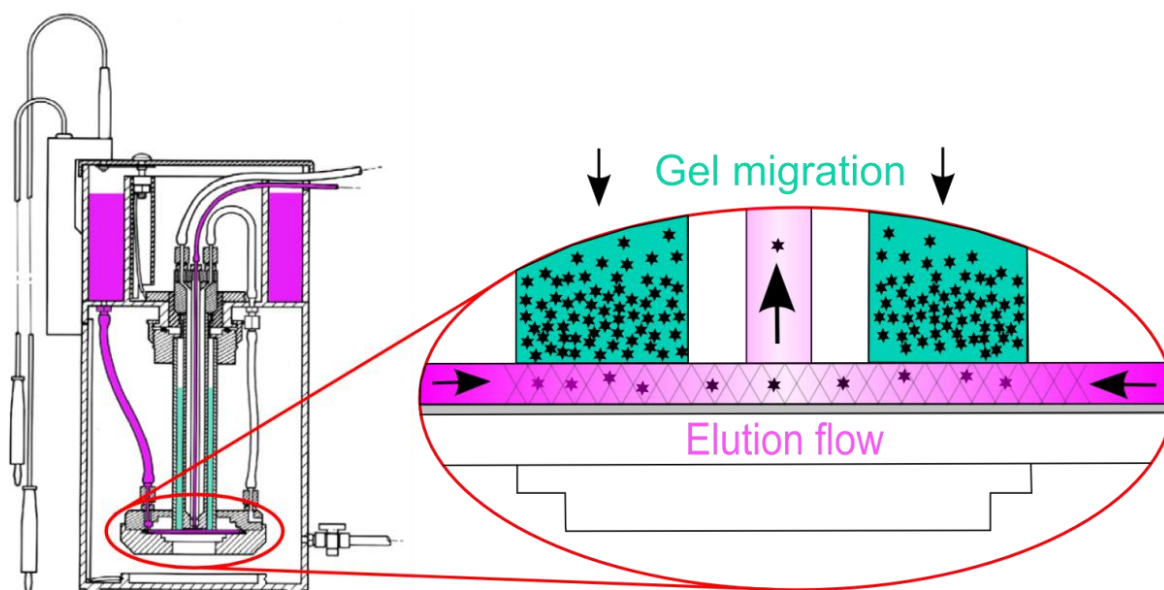
Figure 5.9: Schematic drawing of the preparative gel assembly (modified from **[93]**) with the elution pathway highlighted in violet. The gel matrix is drawn in turquoise. The elution chamber is shown in a blow-up box. Elution buffer enters the elution frit (violet, shaded) from the outer edge and gets pumped up the elution line (center). The sample (black stars) elutes from the gel matrix (turquoise) into the elution frit and is transported to the fraction collector along with the elution buffer.

Fig. 5.9 shows the reason for the dilution of the sample upon elution into the elution frit. The two major determinants of the degree of sample dilution were, firstly, the migration speed of the sample through the gel matrix driven by the applied voltage, and, secondly, the speed of the elution pump which determined the amount of buffer that the sample eluted into. Upon increase of the applied voltage the gel migration proceeded faster, limiting the time of elution for a single sample band from the gel, but also decreasing the resolution of separated bands and thereby jeopardizing sample purity. The separation of single gel bands could be increased by increasing the elution pump speed but thereby also increasing the degree of dilution of the sample. For the elution of complex containing both the H3 R40C-to-DNA crosslink as well as the H3 T80C-to-ΔDot1S542C crosslink the set of parameters shown in Table 5.1 was used for preparative gel purification. The two rounds of preparative gel electrophoresis yielded homogenous sample of both one-bound ΔDot1 to NCP complex and two-bound ΔDot1 to NCP complex. However, the yields of the samples were low due to the large number of purification steps that eventually led to the pure complex. The relatively high cost of oligonucleotides with convertible nucleotides is prohibitive for the production of large quantities of sample which would be needed for large scale rescreening of the pure crosslinked complex.

Parameter	Value
Gel parameters	
Gel matrix	5% native PAGE
Running buffer	0.25x TBE
Bed height	120 – 130 mm
Prerun	12 h @ 225 V
Run parameters	
Temperature	4 °C
Applied voltage	225 V
Elution pump speed	1.5 ml/min
Buffer recirculation	continous
Fraction collection parameters	
Fraction size	5 min $\triangleq$ 7.5 ml
Absorption measurement	λ = 260 nm
Recording sensitivity	20 mV / full scale
Chart recorder speed	0.5 mm/min

Table 5.1: Set of parameters for the preparative gel purification of crosslinked complex between ΔDot1 and NCP.

5.4. Crystallization trials of pure crosslinked ΔDot1 to NCP samples

The screening of crystallization conditions by a former lab member (V. Vogirala, PhD thesis, 2012) yielded a hit for the two-bound ΔDot1 to NCP complex under screening conditions containing 100 mM Hepes pH 7.5, 200 mM sodium citrate tribasic and 20% v/v 2-propanol (Hampton Research, HR2-110-27), although heavy precipitation was observed before the onset of crystal growth (V. Vogirala, PhD thesis [2]). Initially, another hit for the one-bound ΔDot1 to NCP complex under screening conditions containing 100 mM sodium citrate tribasic dihydrate pH 5.6 and 2.5 M 1,6-hexanediol (Hampton Research, HR2-112-19) was found, but it was not reproducible. He created a fine screen around the initial hit conditions and set up crystallization drops at varying screen-solution:sample ratios as sitting-drop vapor diffusion experiments. To slow down equilibration between the well solution and the drop they were covered with Al's oil, a 50:50 mixture of water-permeable silicon oil and water-impermeable paraffin oil. The crystals he obtained under the screened conditions were small (< 20 µm in the largest dimension) and were always accompanied by heavy precipitation. When the wells containing the crystals were opened, the crystals dissolved within 1 – 2 min, probably due to fast evaporation of the volatile 2-propanol in the sample. *In situ* diffraction at the Swiss Light Source synchrotron (beamline X06DA, PXIII) yielded a macromolecular diffraction pattern to 24 Å resolution. One crystal that he harvested under oil and soaked in a solution containing 100 mM Hepes, pH 7.5, 200 mM sodium citrate tribasic, 20% 2-propanol and 25% 1,6-hexanediol and then

flash-cooled into liquid propane at -120 °C diffracted anisotropically up to 8 Å (V. Vogirala, PhD thesis [2]).

The samples containing crosslinked ΔDot1 S542C to H3 T80C and the DNA dG*70:H3 R40C crosslink were used for crystallization after the complex was concentrated to 6 – 7 mg/ml. A screen was created to test all the conditions that previously yielded crystals with the sample not containing the DNA:H3 crosslink. The screen conditions are listed in sections 2.6.4 and 2.6.5. and contain the screen condition (Hampton Research, HR2-110-27) which yielded the initial hit as condition B4. Variations in the concentration of sodium citrate tribasic were screened against various concentrations of 2-propanol (ΔDot1:NCP screen 1, section 2.6.4). Crystallization drops were set up at different sample to screen solution ratios, which previously was shown to affect crystal size [2]. The drops were covered with 20 µl of Al's oil to slow equilibration between the drop and the reservoir. Drops which were not covered with Al's oil did not yield any crystals. All crystallization conditions were layered with argon gas and immediately sealed. The crystallization plates were placed in a nitrogen atmosphere. Only the conditions containing 20% 2-propanol yielded crystals of size larger than 10 µm.
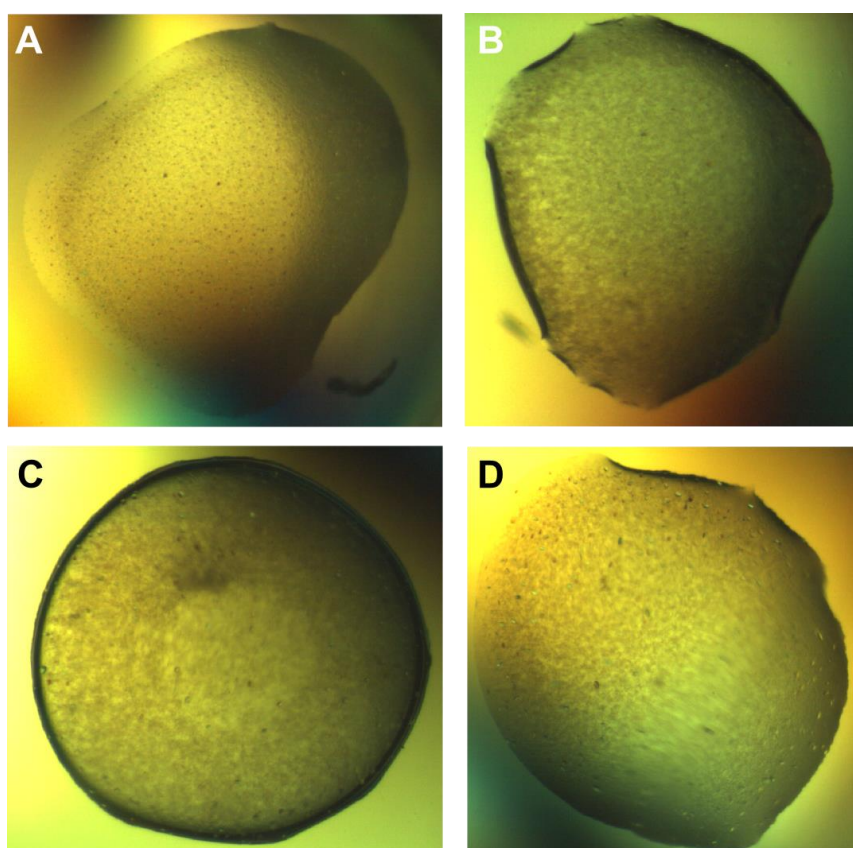
Figure 5.10: Crystals grown from sample containing ΔDot1S542C crosslinked to H3 T80C and the DNA crosslinked to H3 via R40C in different screen conditions: **A)** 100 mM Hepes pH 7.5, 225 mM sodium citrate tribasic, 20% 2-propanol, sample/screen ratio of 0.5. **B)** 100 mM Hepes pH 7.5, 175 mM sodium citrate tribasic, 20% 2-propanol, sample/screen ratio of 0.5, **C)** 100 mM Hepes pH 8.0, 225 mM sodium citrate tribasic, 20% 2-propanol, sample/screen ratio of 0.5, **D)** 100 mM Hepes pH 7.5, 225 mM sodium citrate tribasic, 20% 2-propanol, sample/screen ratio of 0.3.

All conditions that were set up for crystallization yielded heavy precipitate within a few hours of setup. The initial crystallization conditions could not be improved significantly with all conditions yielding crystals not larger than 10 – 20 μm. Attempts were made to further slow down equilibration of drop and reservoir by covering the drops with 20 μl of different paraffin:silicon oil ratios. Drops with a screen-solution:sample ratio of 0.3 were set up and covered with 20 μl of 90% paraffin oil/10% silicon oil and 95% paraffin oil/5% silicon oil. There was heavy precipitation under all conditions and the crystal size did not change significantly.

To test the influence of the complex concentration, the crosslinked ΔDot1:NCP complex was set up for crystallization at various sample concentrations against a screen of conditions containing 100 mM Hepes pH 7.5, 20% 2-propanol and 150 mM – 250 mM sodium citrate tribasic (Screen 3, section 2.6.6). Sample concentrations of 4.5 mg/ml, 6.5 mg/ml, 8.5 mg/ml and 10.5 mg/ml were used. Increasing complex concentration led to the growth of even smaller crystals with the same amount of precipitation as observed for lower complex concentrations. Crystallization drops set up at complex concentrations below 4 mg/ml did not lead to crystal growth but still showed heavy pre-cipitatation.

The precipitation seen under all conditions is due to the presence of 2-propanol. A drop of cross-linked ΔDot1:NCP was brought to 225 mM sodium citrate tribasic in 100 mM Hepes pH 7.5 and cov-ered in 100% paraffin oil, essentially preventing evaporation. No precipitation was observed even after two months. Conversely, a drop of crosslinked ΔDot1:NCP was brought to 20% 2-propanol in 100 mM Hepes pH 7.5 and covered in 100% paraffin oil. The sample immediately started to precipi-tate within seconds after setup.

The crystals which did grow in 150 mM – 225 mM sodium citrate tribasic, 100 mM Hepes pH 7.5 – 8.0 and 20% 2-propanol grew exclusively within the skin of precipitate formed at the interface be-tween the drop and Al's oil or in the precipitate formed at the bottom of the well. There were no crystals growing in regions of the drop devoid of precipitate, as was observed previously [2]. It was possible, however, to transfer some of the precipitate containing a few crystals of approximately 10 μm in size into a drop of soaking solution containing 100 mM Hepes pH 7.5, 225 mM sodium citrate tribasic and either 25% MPD or 20% 2-propanol, or both together. Crystals within the skin of precipi-tate were stable up to 5 min after the transfer into the conditions containing MPD. They dissolved after prolonged incubation.

The highly pure sample was also set up for crystallization using new conditions with the MIDAS screen (Molecular Dimensions [125]), a crystallization screen tailored for the crystallization of pro-

tein:DNA complexes using alternative precipitants. The screen yielded two initial hits in conditions containing poly(acrylic acid sodium salt) 2100, poly(acrylic acid sodium salt) 5100 or SOKALAN® CP 42 as precipitants.
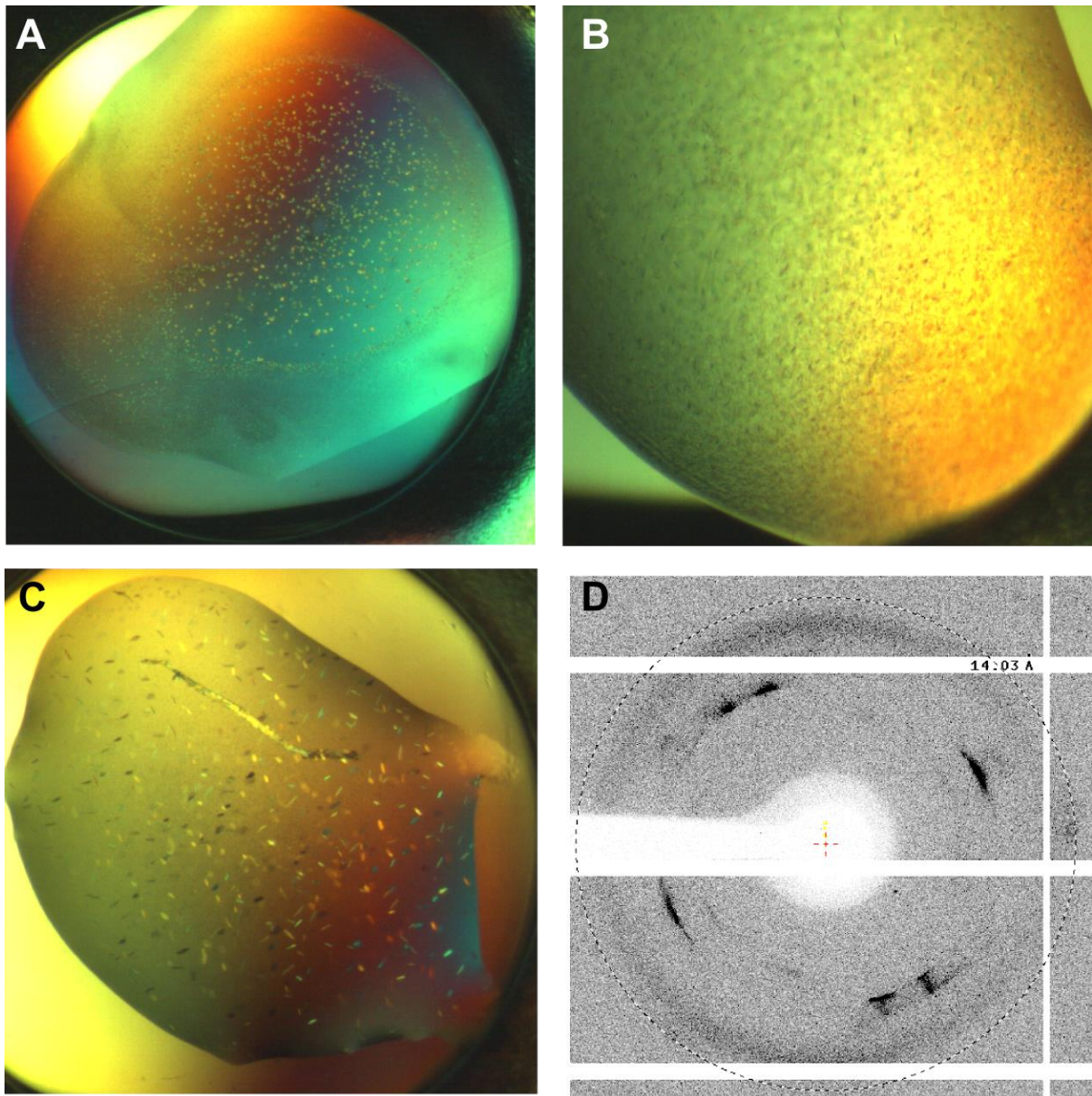
Figure 5.11: Initial hits of the ΔDot1:NCP complex in the Midas screen. **A)** Two-bound ΔDot1:NCP in condition 1-3 [100 mM Hepes pH 6.5, 45% w/v Poly(acrylic acid sodium salt) 2100]. **B)** Two-bound ΔDot1:NCP in condition 2-33 [100 mM Tris pH 7.0, 25% w/v SOKALAN® CP 42, 10% v/v tetrahydrofuran]. **C)** One-bound ΔDot1:NCP in the Midas screen condition 2-21 [100 mM Tris pH 8.0, 20% w/v Poly(acrylic acid sodium salt) 5100]. **D)** The crystals yielded no diffraction indicating macromolecular content. Probably the diffraction around 21 Å was due to polymeric acrylates.

The crystals were too small to be analyzed on SDS PAGE. None of the initial hit conditions yielded well-diffracting crystals. The only diffraction came from crystals grown under conditions containing poly(acrylic acid sodium salt) and was limited to below 20 Å (Fig. 5.11).

6. Discussion and outlook

In the present work a protocol for the site-specific crosslinking of NCPs was successfully developed. The convertible nucleoside approach was applied to establish crosslinks between the nucleosome DNA close to the termini and a cysteine introduced into the H3 N-terminal tail. The crosslinked nucleosome could be purified to homogeneity and crystallization trials yielded well-diffracting crystals. The crystal structure was determined at 2.8 Å and the disulfide crosslinks could be successfully visualized. The crosslink between the DNA termini and the H3 N-terminal tail did not significantly change the overall structure of the NCP, although some differences appear in the SHL +/-5 regions of the DNA where the sequence was changed to accommodate an internal DNA restriction site. Two newly formed GG base pair steps form additional binding sites for Mn^{2+} ions, which are probably mostly responsible for the observed local changes in the structure. Additionally, the convertible nucleotide (dG*) changes the DNA sequence wherever it replaces a non-dG base. By using the BsaI restriction enzyme, as was done to produce Met16 promoter DNA, changing the native sequence can be avoided altogether. The BsaI approach could potentially be utilized to create individual modifications at the DNA ends, since the sticky ends left after BsaI digestion are generally different from one another if the DNA sequence is non-palindromic. The oligonucleotides which are ligated to the core DNA add a level of flexibility in designing experiments by introducing fluorescent labels or functional groups, such as biotin, to the nucleosome DNA. The crosslink to the H3 tail probably changes the trajectory of the H3 tail N-terminal of the crosslinking site. The overall effect of this potential change in trajectory is unclear since most parts of the H3 N-terminal tail are not resolved even in the 1.94 Å structure.

Another variant of the crosslinked NCP which contains an additional crosslink between the two H2A chains at position 38 (H2A N38C) was produced. The H2A N38C crosslink, effectively linking the two H2A/H2B heterodimers together and thereby preventing the dissociation of dimer from the NCP, was previously shown to prevent the occurrence of sub-stoichiometric nucleosomes [1]. The double crosslinked nucleosome is a very stable complex against dissociation of H2A/H2B dimer or DNA, opening some intriguing possibilities for future research.

The experimental design to covalently bind nucleosome DNA to the histone octamer was shown to be successful and did not introduce disturbances in the structure of the NCP. This method could potentially be used to crystallize additional DNA sequences and solve their X-ray crystal structures in an NCP context. The Met16 promoter sequence was chosen to test the crosslinking protocol. It could be shown that the crosslink stabilizes the Met16 NCP and crystals were grown that diffracted to 3.2 Å on the in-house X-ray generator. A higher resolution dataset could be obtained at the Swiss Light

Sourse synchrotron. With the information from the single base pair-resolution nucleosome map in yeast it is theoretically possible to produce crosslinked nucleosomes for every individual nucleosome of the yeast genome by using the crosslinking protocol described in this thesis. A potential bottleneck might be the two-fold disorder introduced by averaging of two non-palindromic sequence orientations. This problem might be overcome by solving the X-ray crystal structures of palindromic half-nucleosomes, as was done for the non-palindromic 601 sequence (601L; [34]). Ideally, there would be a way to crystallize nucleosomes in an asymmetric packing, leading to unambiguous electron density for the DNA part of the NCP structure, as occurred for the X-ray crystal structure of the MMTV-A NCP. The reasons for the asymmetric packing are not completely understood, although progress has been made towards an explanation [1]. Another bottleneck is the lack of knowledge about the stretching states of nucleosomes. This information is not explicitly available from the nucleosome mapping data published so far. Potentially, a site-specific hydroxyl radical footprinting method is conceivable, which would map the DNA ends on the nucleosome, similar to the mapping of nucleosome dyad positions, yielding the exact length of the DNA sequence on the nucleosome. This additional positional information of the genomic DNA would also eliminate the uncertainty in the linker length between positioned nucleosomes due to DNA buffering by stretching on the nucleosomes, yielding more accurate information about the conformation of chromatin higher order structure. It is presently unclear if the crosslinking of the DNA ends to the histone octamer can suppress DNA stretching or if crosslinking is only possible if the 'natural' number of DNA base pairs is assembled on the histone octamer.

NCPs with their DNA crosslinked to the histone octamer are stable against DNA dissociation down to at least the sub-nanomolar range, which could be useful in single molecule studies of nucleosomes that depend on low sample concentrations. The stability against DNA dissociation upon elevation of the salt concentration remains to be tested systematically, although the fact that there is no detectable free DNA in the purified crosslinked NCPs after elution from an anion exchange column is a strong indication that the NCPs are stable against salt-induced DNA dissociation in up to $300 - 400$ mM monovalent salt. This makes the crosslinked NCP a useful sample for usage at low concentrations in physiological salt conditions.

Successfull protection against DNA dissociation upon dilution by crosslinking the nucleosome DNA to the histone octamer was used to produce homogenous sample of ΔDot1:NCP complex with no free DNA occurring. The improved sample homogeneity did not lead to improved crystal growth, however. Large scale rescreening of previously tested crystallization conditions will require the production of large quantities of the crosslinked ΔDot1:NCP complex. An alternative approach to solving the

structure of the ΔDot1:NCP complex is the quickly developing field of cryo electron microscopy. With the structure of the nucleosome core particle determined to 3.9 Å using a Titan Krios electron microscope (FEI) with a K2 Summit detector (Gatan) and a Volta phase plate (FEI) [76], it seems possible that the ΔDot1:NCP structure can be solved to high enough resolution to analyze the molecular details of the complex using cryo-EM.

7. Bibliography

[1]   T. Frouws, "Nucleosome Structure, Positioning and Remodeling," ETH Zürich, Zürich, 2014.

[2]   V. Vogirala, "Biochemical and Structural Studies of the yDot1/NCP Complex," ETH Zürich, Zürich, 2011.

[3]   A. Olins and D. Olins, "Spheroid Chromatin Units (v Bodies)," *Science,* vol. 183, no. 4122, pp. 330-332, 1974.

[4]   R. Williamson, "Properties of rapidly labelled deoxyribonucleic acid fragments isolated from the cytoplasm of primary cultures of embryonic mouse liver cells," *J Mol Biol,* vol. 51, no. 1, pp. 157 - 68, 1970.

[5]   D. Hewish and L. Burgoyne, "Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease," *Biochem Biophys Res Commun,* vol. 52, no. 2, pp. 504 - 10, 1973.

[6]   M. Noll and R. Kornberg, "Action of micrococcal nuclease on chromatin and the location of histone H1.," *J Mol Biol,* vol. 109, no. 3, pp. 393 - 404, 1977.

[7]   P. Oudet, M. Gross-Bellard and P. Chambon, "Electron microscopic and biochemical evidence that chromatin structure is a repeating unit," *Cell,* vol. 4, no. 4, pp. 281-300, 1975.

[8]   T. Richmond, J. Finch, B. Rushton, D. Rodes and A. Klug, "Structure of the nucleosome core particle at 7 A resolution," *Nature,* vol. 311, pp. 532 - 7, 1984.

[9]   K. Luger, A. Maeder, R. Richmond, D. Sargent and T. Richmond, "Crystal structure of the nucleosome," *Nature,* vol. 389, pp. 251 - 60, 1997.

[10]   C. Davey, D. Sargent, K. Luger, A. Maeder and T. Richmond, "Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 A resolution.," *J Mol Biol,* vol. 319, no. 5, pp. 1097-113, 2002.

[11]   T. Richmond and C. Davey, "The structure of DNA in the nucleosome," *Nature,* vol. 423, pp. 145 - 50, 2003.

[12]   J. Finch and A. Klug, "Solenoidal model for superstructure in chomatin," *Proc Natl Acad Sci U.S.A.,* vol. 73, no. 6, pp. 1897 - 901, 1976.

[13]   C. Woodcock, L. Frado and J. Rattner, "The higher-order structure of chromatin: evidence for a helical ribbon arrangement.," *J Cell Biol,* vol. 99, no. 1, pp. 42 - 52, 1984.

[14]   T. Schalch, S. Duda, D. Sargent and T. Richmond, "X-ray structure of a tetranucleosome and its implications for the chromatin fibre," *Nature,* vol. 436, pp. 138 - 41, 2005.

[15] B. Dorigo, T. Schalch, A. Kulangara, S. Duda, R. Schroeder and T. Richmond, "Nucleosome arrays reveal the two-start organization of the chromatin fiber," *Science,* vol. 306, no. 5701, pp. 1571 - 3, 2004.

[16] F. van Leeuwen and D. Gottschling, "Genome-wide histone modifications: gaining specificity by preventing promiscuity.," *Curr Opin Cell Biol,* vol. 14, no. 6, pp. 756 - 62, 2002.

[17] C. Allis and T. Jenuwein, "The molecular hallmarks of epigenetic control," *Nat Rev Genetics,* vol. 17, pp. 487 - 500, 2016.

[18] A. Jansen and K. Verstrepen, "Nucleosome Positioning in Saccharomyces cerevisiae," *Microbiol Mol Biol Rev.,* vol. 75, no. 2, pp. 301-320, 2011.

[19] J. Peters, A. Tedeschi and J. Schmitz, "The cohesin complex and its roles in chromosome biology," *Genes Dev,* vol. 22, pp. 3089 - 114, 2008.

[20] J. Craig, "Heterochromatin--many flavours, common themes.," *Bioessays,* vol. 27, no. 1, pp. 17 - 28, 2005.

[21] J. Harp, B. Hanson, D. Timm and G. Bunick, "Asymmetries in the nucleosome core particle at 2.5 A resolution.," *Acta Crystallogr D Biol Crystallogr.,* vol. 56, no. 12, pp. 1513-34, 2000.

[22] C. White, R. Suto and K. Luger, "Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions.," *EMBO J,* vol. 20, no. 18, pp. 5207-18, 2001.

[23] Y. Tsunaka, N. Kajimura, S. Tate and K. Morikawa, "Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle.," *Nucleic Acids Res,* vol. 33, no. 10, pp. 3424-34, 2005.

[24] C. Clapier, S. Chakravarthy, C. Petosa, C. Fernandez-Tornero, K. Luger and C. Müller, "Structure of the Drosophila nucleosome core particle highlights evolutionary constraints on the H2A-H2B histone dimer.," *Proteins,* vol. 71, no. 1, pp. 1-7, 2008.

[25] R. Suto, M. Clarkson, D. Tremethick and K. Luger, "Crystal structure of a nucleosome core particle containing the variant histone H2A.Z.," *Nat Struct Biol,* vol. 7, no. 12, pp. 1121-4, 2000.

[26] S. Chakravarthy, S. Gundimella, C. Caron, P. Perche, J. Pehrson, S. Knochbin and K. Luger, "Structural characterization of the histone variant macroH2A.," *Mol Cell Biol,* vol. 25, no. 17, pp. 7616-24, 2005.

[27] H. Tachiwana, W. Kagawa, A. Osakabe, K. Kawaguchi, T. Shiga, Y. Hayashi-Takanaka, H. Kimura and H. Kurumizaka, "Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T.," *Proc Natl Acad Sci U.S.A,* vol. 107, no. 23, pp. 10545-9, 2010.

[28] T. Kujirai, N. Horikoshi, K. Sato, K. Maehara, S. Machida, A. K. H. Osakabe, Y. Ohkawa and H. Kurumizaka, "Crystal structure of human nucleosome containing H3.Y," *Nucleic Acids Res,* no. E-pub, 2016.

[29] H. Tachiwana, W. Kagawa, T. Shiga, A. Osakabe, Y. Miya, K. Saito, Y. Hayashi-Takanaka, T. Oda, M. Sato, S. Park, H. Kimura and H. Kurumizaka, "Crystal structure of the human centromeric nucleosome containing CENP-A.," *Nature,* vol. 476, pp. 232-5, 2011.

[30] M. Ong, T. Richmond and C. Davey, "DNA Stretching and Extreme Kinking in the Nucleosome Core," *J Mol Biol,* vol. 368, no. 4, pp. 1067-74, 2007.

[31] Y. Bao, C. White and K. Luger, "Nucleosome core particles containing a poly(dA.dT) sequence element exhibit a locally distorted DNA structure," *J Mol Biol,* vol. 361, no. 4, pp. 617-24, 2006.

[32] B. Wu, K. Mohideen, D. Vasudevan and C. Davey, "Structural Insight into the Sequence Dependence of Nucleosome Positioning," *Structure,* vol. 18, no. 4, pp. 528-536, 2010.

[33] D. Vasudevan, E. Chua and C. Davey, "Crystal structures of nucleosome core particles containing the '601' strong positioning sequence," *J Mol Biol,* vol. 403, no. 1, pp. 1-10, 2010.

[34] E. Chua, D. Vasudevan, G. Davey and C. Davey, "The mechanics behind DNA sequence-dependent properties of the nucleosome," *Nucleic Acids Res,* vol. 40, no. 13, pp. 6338-52, 2012.

[35] T. D. Frouws, S. C. Duda and T. J. Richmond, "X-ray structure of the MMTV-A nucleosome core," *Proc Natl Acad Sci U S A,* 2016.

[36] A. Barbera, J. Chodaparambil, B. Kelley-Clarke, V. Joukov, J. Walter, K. Luger and K. Kaye, "The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA.," *Science,* vol. 311, pp. 856-61, 2006.

[37] R. Makde, J. England, H. Yennawar and S. Tan, "Structure of RCC1 chromatin factor bound to the nucleosome core particle.," *Nature,* vol. 467, pp. 562-6, 2010.

[38] K. Armache, J. Garlick, D. Canzio, G. Narlikar and R. Kingston, "Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution.," *Science,* vol. 334, no. 6058, pp. 977-82, 2011.

[39] H. Kato, J. Jinag, B. Zhou, M. Rozendaal, H. Feng, R. Ghirlando, T. Xiao, A. Straight and Y. Bai, "A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C," *Science,* vol. 340, no. 6136, pp. 1110-3, 2013.

[40] T. Girish, R. McGinty and S. Tan, "Multivalent Interactions by the Set8 Histone Methyltransferase With Its Nucleosome Substrate," *J Mol Biol,* vol. 428, no. 8, pp. 1531-43, 2016.

[41] M. Morgan, M. Haj-Yahya, A. Ringel, P. Bandi, A. Brik and C. Wolberger, "Structural basis for histone H2B deubiquitination by the SAGA DUB module," *Science,* vol. 351, no. 6274, pp. 725-8, 2016.

[42] B. Zhou, J. Jiang, H. Feng, R. Ghirlando, T. Xiao and Y. Bai, "Structural Mechanisms of Nucleosome Recognition by Linker Histones," *Mol Cell,* vol. 59, no. 4, pp. 628-38, 2015.

[43] G. Arents, R. Burlingame, B. Wang, W. Love and E. Moudrianakis, "The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix," *Proc Natl Acad Sci U.S.A.,* vol. 88, no. 22, pp. 10148-52, 1991.

[44] S. West, R. Rohs, R. Mann and B. Honig, "Electrostatic interactions between arginines and the minor groove in the nucleosome," *J Biomol Struct Dyn,* vol. 27, no. 6, pp. 861 - 6, 2010.

[45] R. McGinty and S. Tan, "Nucleosome Structure and Function," *Chem Rev.,* vol. 115, no. 6, pp. 2255 - 73, 2015.

[46] A. Kalashnikova, M. Porter-Goff, U. Muthurajan, K. Luger and J. Hansen, "The role of the nucleosome acidic patch in modulating higher order chromatin structure," *J R Soc Interface,* vol. 10, 2013.

[47] R. Makde, J. England, H. Yennawar and S. Tan, "Structure of RCC1 chromatin factor bound to the nucleosome core particle," *Nature,* vol. 467, pp. 562-6, 2010.

[48] A. Norris, M. Bianchet and J. Boeke, "Compensatory Interactions between Sir3p and the Nucleosomal LRS Surface Imply Their Direct Interaction," *PLOS Genet,* vol. 4, no. 12, 2008.

[49] C. Fry, A. Norris, M. Cosgrove, J. Boeke and C. Peterson, "The LRS and SIN Domains: Two Structurally Equivalent but Functionally Distinct Nucleosome Surfaces Required for Transcriptional Silencing," *Mol Cell Biol,* vol. 26, no. 23, pp. 9045 - 59, 2006.

[50] C. Calladine and H. Drew, *Understanding DNA*, Academic Press, 1997.

[51] A. Wolffe, "Chromatin disruption and modification," *Nucleic Acids Res,* vol. 27, no. 3, pp. 711 - 20, 1999.

[52] Y. Zheng and C. Qiang, "The histone H3 N-terminal tail: a computational analysis of the free energy landscape and kinetics," *Phys Chem Chem Phys,* vol. 17, pp. 13689 - 98, 2015.

[53] K. Luger and T. Richmond, "The histone tails of the nucleosome," *Curr Opin Genetics Dev,* vol. 8, pp. 140 - 6, 1998.

[54] B. Strahl and C. Allis, "The language of covalent histone modifications," *Nature,* vol. 403, pp. 41 - 5, 2000.

[55] J. Hayes, T. Tullius and A. Wolffe, "The structure of DNA in a nucleosome," *Proc Natl Acad Sci*

U.S.A., vol. 87, pp. 7405 - 9, 1990.

[56] A. Travers and F. Klug, "The bending of DNA in nucleosomes and its wider implications," *Phil Trans R Soc Lond B,* vol. 317, pp. 537 - 61, 1987.

[57] D. Vasudevan, E. Y. Chua and C. A. Davey, "Crystal Structures of Nucleosome Core Particles Containing the '601' Strong Positioning Sequence," *J Mol Biol,* vol. 401, no. 1, 2010.

[58] J. Compton, M. Bellard and P. Chambon, "Biochemical evidence of variability in the DNA repeat length in the chromatin of higher eukaryotes," *Proc Natl Acad Sci U.S.A.,* vol. 73, no. 12, pp. 4382-6, 1976.

[59] A. Flaus, K. Luger, S. Tan and T. Richmond, "Mapping nucleosome position at single base-pair resolution by using site-directed hydroxyl radicals," *Proc Natl Acad Sci U.S.A,* vol. 93, no. 4, pp. 1370-5, 1996.

[60] K. Brogaard, L. Xi, J. Wang and J. Widom, "A map of nucleosome positions in yeast at base-pair resolution," *Nature,* vol. 486, pp. 496 - 501, 2012.

[61] K. Luger, T. Rechsteiner and T. Richmond, "Preparation of nucleosome core particle from recombinant histones," *Methods Enzymol,* vol. 304, 1999.

[62] E. Segal and J. Widom, "Poly(dA:dT) tracts: major determinants of nucleosome organization," *Curr Opin Struct Biol,* vol. 19, pp. 65 - 71, 2009.

[63] N. Kaplan, I. Moore, Y. Fondufe-Mittendorf, A. Gossett, D. Tillo, Y. Field, E. LeProust, T. Hughes, J. Lieb, J. Widom and E. Segal, "The DNA-encoded nucleosome organization of a eukaryotic genome," *Nature,* vol. 458, pp. 362 - 6, 2009.

[64] G. Freeman, J. Lequieu, D. Hinckley, J. Whitmer and J. de Pablo, "DNA Shape Dominates Affinity in Nucleosome Formation," *Phys Rev Let,* vol. 113, pp. 168101-1-5, 2014.

[65] A. Morozov, K. Fortney, D. Gaykalova, V. Studitsky, J. Widom and E. Siggia, "Using DNA mechanics to predict the in vitro nucleosome positions and formation energies," *Nucleic Acids Res,* vol. 37, no. 14, pp. 4707 - 22, 2009.

[66] P. Lowary and J. Widom, "New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning," *J Mol Biol,* vol. 276, no. 1, pp. 19 - 42, 1998.

[67] E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thastrom, Y. Field, I. Moore, J. Wang and J. Widom, "A genomic code for nucleosome positioning," *Nature,* vol. 442, pp. 772 - 8, 2006.

[68] Y. Zhang, Z. Moqtaderi, B. Rattner, G. Euskirchen, M. Snyder, J. Kadonaga, X. Liu and K. Struhl, "Evidence against a genomic code for nucleosome positioning Reply to "Nucleosome sequence preferences influence in vivo nucleosome organization"," *Nat Struct Mol Biol,* vol. 17, pp. 920 - 3, 2010.

[69] J. Ausio, D. Seger and H. Eisenberg, "Nucleosome Core Particle Stability and Conformational Change," *J Mol Biol,* vol. 176, pp. 77-104, 1984.

[70] Y. Park, P. Dyer, D. Tremethick and K. Luger, "A New Fluorescence Resonance Energy Transfer Approach Demonstrates That the Histone Variant H2AZ Stabilizes the Histone Octamer within the Nucleosome," *J Biol Chem,* vol. 279, pp. 24274 - 82, 2004.

[71] A. Thaström, J. Gottesfeld, K. Luger and J. Widom, "Histone-DNA Binding Free Energy Cannot Be Measured in Dilution-Driven Dissociation Experiments," *Biochemistry,* vol. 43, pp. 736 - 41, 2004.

[72] A. Andrews and K. Luger, "Nucleosome Structure(s) and Stability: Variations on a Theme," *Annu Rev Biophys,* vol. 40, pp. 99 - 117, 2011.

[73] R. Suto, R. Edayathumangalam, C. White, C. Melander, J. Gottesfeld and P. L. K. Dervan, "Crystal Structured of Nucleosome Core Particles in Complex with Minor Groove DNA-binding Ligands," *J Mol Biol,* vol. 326, pp. 371 - 80, 2003.

[74] X. Lu, M. Simon, J. Chodaparambil, J. Hansen, K. Shokat and K. Luger, "The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure," *Nat Struct Mol Biol,* vol. 15, no. 10, pp. 1122 - 4, 2008.

[75] R. McGinty, R. Henrici and S. Tan, "Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome.," *Nature,* vol. 514, no. 7524, pp. 591 - 6, 2014.

[76] E. Chua, V. Vogirala, O. Inian, A. Wong, L. Nordenskiöld, J. Plitzko, R. Danev and S. Sandin, "3.9 A structure of the nucleosome core particle determined by phase-plate cryo-EM," *Nucleic Acids Res,* vol. 44, no. 7, pp. 8013 - 9, 2016.

[77] D. Maskell, L. Renault, E. Serrao, P. Lesbats, R. Matadeen, S. Hare, D. Lindemann, A. Engelman, A. Costa and P. Cherepanov, "Structural basis for retroviral integration into nucleosomes.," *Nature,* vol. 523, pp. 366 - 9, 2015.

[78] A. Bannister and T. Kouzarides, "Regulation of chromatin by histone modifications," *Cell Res,* vol. 21, pp. 381 - 95, 2011.

[79] S. Kato, K. Inoue and M. Youn, "Emergence of the osteo-epigenome in bone biology," *IBMS Bone KEy,* vol. 7, pp. 314 - 24, 2010.

[80] M. Singer, A. Kahana, A. Wolf, L. Meisinger, S. Peterson, C. Goggin, M. Mahowald and D. Gottschling, "Identification of high-copy disruptors of telomeric silencing in Saccharomyces cerevisiae.," *Genetics,* vol. 150, no. 2, pp. 613 - 32, 1998.

[81] F. van Leeuwen, P. Gafken and D. Gottschling, "Dot1p Modulates Silencing in Yeast by Methylation of the Nucleosome Core," *Cell,* vol. 109, no. 6, pp. 745 - 56, 2002.

[82] A. Nguyen and Y. Zhang, "The diverse functions of Dot1 and H3K79 methylation," *Genes Dev,* vol. 25, pp. 1345 - 58, 2011.

[83] W. Kim, M. Choi and J. Kim, "The histone methyltransferase Dot1/DOT1L as a critical regulator of the cell cycle," *Cell cycle,* vol. 13, no. 5, pp. 1 - 13, 2014.

[84] H. Ng, Q. Feng, H. Wang, H. Erdjument-Bromage, P. Tempst, Y. Zhang and K. Struhl, "Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association," *Genes Dev.,* vol. 16, no. 12, pp. 1518 - 27, 2002.

[85] S. Briggs, T. Xiao, Z. Sun, J. Caldwell, J. Shabanowitz, D. Hunt, C. Allis and B. Strahl, "Gene silencing: trans-histone regulatory pathway in chromatin," *Nature,* vol. 418, p. 498, 2002.

[86] I. Strulemeijer, D. De Vos, K. van Hartem, O. Joshi, O. Blomberg, T. van Welsem, M. Terweij, H. Vlaming, E. de Graaf, A. Altelaar, B. Bakker and F. van Leeuwen, "Dot1 histone methyltransferases share a distributive mechanism but have highly diverged catalytic properties," *Sci Rep,* vol. 5, no. 9824, pp. 1 - 11, 2015.

[87] I. Fingerman, H. Li and S. Briggs, "A charge-based interaction between histone H4 and Dot1 is required for H3K79 methylation and telomere silencing: identification of a new trans-histone pathway," *Genes Dev,* vol. 21, pp. 2018 - 29, 2007.

[88] J. Min, Q. Feng, Z. Li, Y. Zhang and R. Xu, "Structure of the Catalytic Domain of Human DOT1L, a Non-SET Domain Nucleosomal Histone Methyltransferase," *Cell,* vol. 112, pp. 711 - 23, 2003.

[89] K. Sawada, Z. Yang, J. Horton, R. Collins, X. Zhang and X. Cheng, "Structure of the Conserved Core of the yeast Dot1p, a Nucleosome Histone H3 Lysine 79 Methyltransferase," *J Biol Chem,* vol. 279, no. 41, pp. 43296 - 306, 2004.

[90] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. Wilkins, R. Appel and A. Bairoch, "Protein Identification and Analysis Tools on the ExPASy Server," in *John M. Walker (ed): The Proteomics Protocol Handbook*, Humana Press, 2005, pp. 571 - 607.

[91] M. Braunstein, R. Sobel, C. Allis, B. Turner and J. Broach, "Efficient transcriptional silencing in Saccharomyces cerevisiae requires a heterochromatin histone acetylation pattern," *Mol Cell Biol,* vol. 16, no. 8, pp. 4349 - 56, 1996.

[92] M. Atlaf, R. Utley, N. Lacoste, S. Tan, S. Briggs and J. Côte, "Interplay of Chromatin Modifiers on a Short Basic Patch of Histone H4 Tail Defines the Boundary of Telomeric Heterochromatin," *Mol Cell,* vol. 28, no. 6, pp. 1002 - 14, 2007.

[93] Bio-Rad Laboratories Inc., "Model 491 Prep Cell - Instruction Manual," www.bio-rad.com, accessed 08/2016.

[94] J. J. SantaLucia, "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics.," *Proc Natl Acad Sci U.S.A.,* vol. 95, no. 4, pp. 1460 - 5, 1998.

[95] D. Erlanson, L. Chen and G. Verdine, "DNA Methylation through a Locally Unpaired Intermediate," *J. Am. Chem. Soc.,* vol. 115, pp. 12583-12584, 1993.

[96] D. Erlanson, J. Glover and G. Verdine, "Disulfide Cross-linking as a Mechanistic Probe for the B <-> Z Transition in DNA," *J. Am. Chem. Soc.,* vol. 119, pp. 6927-6928, 1997.

[97] Glen Research, "Use od 2-F-dI to produce N2-modified dG derivatives," www.glenresearch.com, last accessed 08/2016.

[98] W. Kabsch, "Integration, scaling, space-group assignment and post-refinement," *Acta Cryst. D,* 2010.

[99] W. Kabsch, "XDS," *Acta Crystallogr D,* vol. 66, no. 2, pp. 125 - 32, 2010.

[100] A. McCoy, R. Grosse-Kunstleve, P. Adams, M. Winn, L. Storoni and R. Read, "Phaser crystallographic software," *J Appl Cryst,* vol. 40, pp. 658 - 74, 2007.

[101] P. Adams, P. Afonine, G. Bunkoczi, V. Chen, I. Davis, N. Echols, J. Headd, L. Hung, G. Kapralm, A. Grosse-Kunstleve, A. McCoy, N. Moriarty, R. Oeffner, R. Read, D. Richardson, J. Richardson, T. Terwilliger and P. Zwart, "PHENIX: a comprehensive Python-based system for macromolecular structure solution.," *Acta Crystallogr D,* vol. 66, no. 2, pp. 213 - 21, 2010.

[102] P. Emsley, B. Lohkamp, W. Scott and K. Cowtan, "Features and Development of Coot," *Acta Crystallogr D,* vol. 66, no. 4, pp. 486 - 501, 2010.

[103] D. M. van Alten, D. A. Erlanson, G. L. Verdine and L. Joshua-Tor, "A structural snapshot of base-pair opening in DNA," *Proc Natl Acad Sci U S A.,* vol. 96, no. 21, 1999.

[104] V. Chen, W. Arendall, J. Headd, D. Keedy, R. Immormino, G. Kapral, L. R. J. Murray and D. Richardson, "MolProbity: all-atom structure validation for macromolecular crystallography," *Acta Cryst. D,* 2010.

[105] E. Pettersen, T. Goddard, H. CC, G. Couch, D. Greenblatt, E. Meng and T. Ferrin, "UCSF Chimera--a visualization system for exploratory research and analysis," *J Comput Chem,* vol. 25, no. 13, pp. 1605 - 12, 2004.

[106] K. Struhl and E. Segal, "Determinants of nucleosome positioning," *Nat Struct Mol Biol,* vol. 20, no. 3, pp. 267 - 73, 2013.

[107] A. Ferentz and G. Verdine, "The Convertible Nucleoside Approach: Structural Engineering of Nucleic Acids by Disulfide Cross-linking," in *Nuclear Acids and Molecular Biology*, Berlin Heidelberg, Springer, 1994, pp. 14 - 40.

[108] H. Huang, R. Chopra, G. L. Verdine and S. C. Harrison, "Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance," *Science,* vol. 282, no. 5394, pp. 1669 - 75, 1998.

[109] A. Banerjee, W. Yang, M. Karplus and G. Verdine, "Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA.," *Nature,* vol. 434, no. 7033, pp. 612 - 8, 2005.

[110] K. Malecka, W. Ho and R. Marmorstein, "Crystal structure of a p53 core tetramer bound to DNA.," *Oncogene,* vol. 28, no. 3, pp. 325 - 33, 2009.

[111] D. van Alten, D. Erlanson, G. Verdine and L. Joshua-Tor, "A structural snapshot of base-pair opening in DNA," *Proc. Natl. Acad. Sci. U.S.A.,* vol. 96, no. 21, 1999.

[112] C. Schneider, W. Rasband and K. Eliceiri, "NIH Image to ImageJ: 25 years of image analysis," *Nature Methods,* 2012.

[113] T. Richmond, M. Searles and R. Simpson, "Crystals of a nucleosome core particle containing defined sequence DNA.," *J Mol Biol,* vol. 199, no. 1, pp. 161 - 70, 1988.

[114] E. De la Mora, I. Carmichael and E. F. Garman, "Effective scavenging at cryotemperatures: further increasing the dose tolerance of protein crystals," *J Synchrotron Radiat,* vol. 18, no. 3, pp. 346 - 57, 2011.

[115] N. Moriarty, R. Grosse-Kunstleve and P. Adams, "electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation," *Acta Crystallogr D Biol Crystallogr,* vol. 65, pp. 1074 - 80, 2009.

[116] E. Meng, E. Pettersen, G. Couch, C. Huang and T. Ferrin, "Tools for integrated sequence-structure analysis with UCSF Chimera," *BMC Bioinformatics,* vol. 7, no. 339, 2006.

[117] S. Needleman and C. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *J Mol Biol,* vol. 48, no. 3, pp. 443 - 53, 1970.

[118] T. Frouws, S. Duda and T. Richmond, "X-ray structure of the MMTV-A nucleosome core.," *Proc Natl Acad Sci U.S.A.,* vol. 113, no. 5, pp. 1214-9, 2016.

[119] M. Gerstel, C. Deane and E. Garman, "Identifying and quantifying radiation damage at the atomic level," *J Synchrotron Radiat,* vol. 22, no. 2, pp. 201 - 12, 2015.

[120] J. Richardson, "The anatomy and taxonomy of protein structure," *Adv Prot Chem,* vol. 34, pp. 167 - 339, 1981.

[121] H. Jaeger and H. Schaefer III, "Characterizing Radiation-Induced Oxidation of DNA by Way of the Monohydrated Guanine-Cytosine Radical Cation," *J Phys Chem,* vol. 113, pp. 8142 - 8, 2009.

[122] D. Thomas, S. Barbey and Y. Surdin-Kerjan, "Gene-enzyme relationship in the sulfate assimilation pathway of Saccharomyces cerevisiae. Study of the 3'-phosphoadenylylsulfate reductase structural gene," *J Biol Chem,* vol. 265, no. 26, pp. 15518 - 24, 1990.

[123] J. Ferreiro, N. Powell, N. Karabetsou, J. Mellor and R. Waters, "Roles for Gcn5p and Ada2p in transcription and nucleotide excision repair at the Saccharomyces cerevisiae MET16 gene," *Nucleic Acids Res,* vol. 34, no. 3, pp. 976 - 85, 2006.

[124] M. Ameyar, J. Wisniewska and J. Weitzman, "A role for AP-1 in apoptosis: the case for and against," *Biochimie,* vol. 85, no. 8, pp. 747 - 52, 2003.

[125] C. Grimm, A. Chari, K. Reuter and U. Fischer, "A crystallization screen based on alternative polymeric precipitants," *Acta Crystallogr D,* vol. D66, no. 6, pp. 685 - 97, 2010.

[126] E. Chua, D. Vasudevan, G. Davey, B. Wu and C. Davey, "The mechanics behind DNA sequence-dependent properties of the nucleosome.," *Nucleic Acids Research,* vol. 40, no. 13, pp. 6338-52, 2012.

[127] B. Schmidt, L. Ho and P. Hogg, "Allosteric Disulfide Bonds," *Biochemistry,* vol. 45, pp. 7429-33, 2006.

[128] T. Hagerman, Q. Fu, B. Molinié, J. Denvir, S. Lindsay and P. Georgel, "Chromatin Stability at Low Concentration Depends on Histone Octamer Saturation Levels," *Biophys J,* vol. 96, pp. 1944 - 51, 2009.

[129] N. Hazan, T. Tomov, R. Tsukanov, M. Liber, Y. Berger, R. Massoud, K. Toth, J. Langowski and E. Nir, "Nucleosome Core Particle Disassembly and Assembly Kinetics Studies Using Single-Molecule Fluorescence," *Biophys J,* vol. 109, pp. 1676-85, 2015.

[130] F. Frederiks, M. Tzouros, G. Oudgenoeg, T. van Welsem, M. Fornerod, J. Krijgsveld and F. van Leeuwen, "Nonprocessive methylation by Dot1 leads to functional redundancy of histone H3K79 methylation states," *Nat Struct Mol Biol,* vol. 15, no. 6, pp. 550 - 7, 2008.

[131] M. Altaf, R. Utley, N. Lacoste, S. Tan, S. Briggs and J. Côte, "Interplay of Chromatin Modifiers on a Short Basic Patch of Histone H4 Tail Defines the Boundary of Telomeric Heterochomatin," *Molecular Cell,* vol. 28, pp. 1002 - 14, 2007.

[132] G. Zentner and S. Henikoff, "Regulation of nucleosome dynamics by histone modifications," *Nat Struct Mol Biol,* vol. 20, no. 3, pp. 259 - 66, 2013.

8. Appendix

8.1. DNA sequences

α-satellite (α8s) 147 bp palindrome

```
1        ATCAATATCC ACCTGCAGAT ACTACCAAAA GTGTATTTGG AAACTGCTCC ATCAAAAGGC
61       ATGTTCAGCT GGAATCCAGC TGAACATGCC TTTTGATGGA GCAGTTTCCA AATACACTTT
121      TGGTAGTATC TGCAGGTGGA TATTGAT
```

α-satellite (α8s) 147 bp palindrome convG +70

```
1        ATCCATATCC ACCTGGTCCT ACTACCAAAA GTGTATTTGG AAACTGCTCC ATCAAAAGGC
61       ATGTTCAGCT GGAATCCAGC TGAACATGCC TTTTGATGGA GCAGTTTCCA AATACACTTT
121      TGGTAGTAGG ACCAGGTGGA TATGGAT
```

601 – 145 bp

```
1        ATCAGAATCC CGGTGCCGAG GCCGCTCAAT TGGTCGTAGA CAGCTCTAGC ACCGCTTAAA
61       CGCACGTACG CGCTGTCCCC CGCGTTTTAA CCGCCAAGGG GATTACTCCC TAGTCTCCAG
121      GCACGTGTCA GATATATACA TCGAT
```

601 – 145 bp convG +69

```
1        ATCCCAATCC CGGTCCCGAG GCCGCTCAAT TGGTCGTAGA CAGCTCTAGC ACCGCTTAAA
61       CGCACGTACG CGCTGTCCCC CGCGTTTTAA CCGCCAAGGG GATTACTCCC TAGTCTCCAG
121      GCACGTGTCG GACCGGGATT GGGAT
```

Met16 promoter NCP DNA 147 bp convG +70

```
1        ATACGTCTTC ATTTTATGAG TTGCTATGGG TTGATACCTT TGCTGGCCTT AGTTTTGATC
61       TTAAAGCATT GAAAAGGTTA CAAACAGTAC AGTATACTTT TTTTATATCA ATATAATACT
121      ATCTCTATGA GCTTTTACTT TGTGGAG
```

8.2. Plasmid sequences

pUC57 (2,710 bp) – DNA production

```
1        TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA
61       CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG
121      TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
181      ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC
241      ATTCGCCATT CAGGCTGCGC AACTGTTGGG AAGGGCGATC GGTGCGGGCC TCTTCGCTAT
301      TACGCCAGCT GGCGAAAGGG GGATGTGCTG CAAGGCGATT AAGTTGGGTA ACGCCAGGGT
361      TTTCCCAGTC ACGACGTTGT AAAACGACGG CCAGTGAATT CGAGCTCGGT ACCTCGCGAA
421      TGCATCTAGA TATCGGATCC CGGGCCCGTC GACTGCAGAG GCCTGCATGC AAGCTTGGCG
481      TAATCATGGT CATAGCTGTT TCCTGTGTGA AATTGTTATC CGCTCACAAT TCCACACAAC
541      ATACGAGCCG GAAGCATAAA GTGTAAAGCC TGGGGTGCCT AATGAGTGAG CTAACTCACA
601      TTAATTGCGT TGCGCTCACT GCCCGCTTTC CAGTCGGGAA ACCTGTCGTG CCAGCTGCAT
661      TAATGAATCG GCCAACGCGC GGGGAGAGGC GGTTTGCGTA TTGGGCGCTC TTCCGCTTCC
721      TCGCTCACTG ACTCGCTGCG CTCGGTCGTT CGGCTGCGGC GAGCGGTATC AGCTCACTCA
781      AAGGCGGTAA TACGGTTATC CACAGAATCA GGGGATAACG CAGGAAAGAA CATGTGAGCA
841      AAAGGCCAGC AAAAGGCCAG GAACCGTAAA AAGGCCGCGT TGCTGGCGTT TTTCCATAGG
901      CTCCGCCCCC CTGACGAGCA TCACAAAAAT CGACGCTCAA GTCAGAGGTG GCGAAACCCG
961      ACAGGACTAT AAAGATACCA GGCGTTTCCC CCTGGAAGCT CCCTCGTGCG CTCTCCTGTT
```

```
1021      CCGACCCTGC CGCTTACCGG ATACCTGTCC GCCTTTCTCC CTTCGGGAAG CGTGGCGCTT
1081      TCTCATAGCT CACGCTGTAG GTATCTCAGT TCGGTGTAGG TCGTTCGCTC CAAGCTGGGC
1141      TGTGTGCACG AACCCCCCGT TCAGCCCGAC CGCTGCGCCT TATCCGGTAA CTATCGTCTT
1201      GAGTCCAACC CGGTAAGACA CGACTTATCG CCACTGGCAG CAGCCACTGG TAACAGGATT
1261      AGCAGAGCGA GGTATGTAGG CGGTGCTACA GAGTTCTTGA AGTGGTGGCC TAACTACGGC
1321      TACACTAGAA GAACAGTATT TGGTATCTGC GCTCTGCTGA AGCCAGTTAC CTTCGGAAAA
1381      AGAGTTGGTA GCTCTTGATC CGGCAAACAA ACCACCGCTG GTAGCGGTGG TTTTTTTGTT
1441      TGCAAGCAGC AGATTACGCG CAGAAAAAAA GGATCTCAAG AAGATCCTTT GATCTTTTCT
1501      ACGGGGTCTG ACGCTCAGTG GAACGAAAAC TCACGTTAAG GGATTTTGGT CATGAGATTA
1561      TCAAAAAGGA TCTTCACCTA GATCCTTTTA AATTAAAAAT GAAGTTTTAA ATCAATCTAA
1621      AGTATATATG AGTAAACTTG GTCTGACAGT TACCAATGCT TAATCAGTGA GGCACCTATC
1681      TCAGCGATCT GTCTATTTCG TTCATCCATA GTTGCCTGAC TCCCCGTCGT GTAGATAACT
1741      ACGATACGGG AGGGCTTACC ATCTGGCCCC AGTGCTGCAA TGATACCGCG AGACCCACGC
1801      TCACCGGCTC CAGATTTATC AGCAATAAAC CAGCCAGCCG GAAGGGCCGA GCGCAGAAGT
1861      GGTCCTGCAA CTTTATCCGC CTCCATCCAG TCTATTAATT GTTGCCGGGA AGCTAGAGTA
1921      AGTAGTTCGC CAGTTAATAG TTTGCGCAAC GTTGTTGCCA TTGCTACAGG CATCGTGGTG
1981      TCACGCTCGT CGTTTGGTAT GGCTTCATTC AGCTCCGGTT CCCAACGATC AAGGCGAGTT
2041      ACATGATCCC CCATGTTGTG CAAAAAAGCG GTTAGCTCCT TCGGTCCTCC GATCGTTGTC
2101      AGAAGTAAGT TGGCCGCAGT GTTATCACTC ATGGTTATGG CAGCACTGCA TAATTCTCTT
2161      ACTGTCATGC CATCCGTAAG ATGCTTTTCT GTGACTGGTG AGTACTCAAC CAAGTCATTC
2221      TGAGAATAGT GTATGCGGCG ACCGAGTTGC TCTTGCCCGG CGTCAATACG GGATAATACC
2281      GCGCCACATA GCAGAACTTT AAAAGTGCTC ATCATTGGAA AACGTTCTTC GGGGCGAAAA
2341      CTCTCAAGGA TCTTACCGCT GTTGAGATCC AGTTCGATGT AACCCACTCG TGCACCCAAC
2401      TGATCTTCAG CATCTTTTAC TTTCACCAGC GTTTCTGGGT GAGCAAAAAC AGGAAGGCAA
2461      AATGCCGCAA AAAAGGGAAT AAGGGCGACA CGGAAATGTT GAATACTCAT ACTCTTCCTT
2521      TTTCAATATT ATTGAAGCAT TTATCAGGGT TATTGTCTCA TGAGCGGATA CATATTTGAA
2581      TGTATTTAGA AAAATAAACA AATAGGGGTT CCGCGCACAT TTCCCCGAAA AGTGCCACCT
2641      GACGTCTAAG AAACCATTAT TATCATGACA TTAACCTATA AAAATAGGCG TATCACGAGG
2701      CCCTTTCGTC
```

pET3a (4,640 bp) – Histone expression

```
1         TTCTCATGTT TGACAGCTTA TCATCGATAA GCTTTAATGC GGTAGTTTAT CACAGTTAAA
61        TTGCTAACGC AGTCAGGCAC CGTGTATGAA ATCTAACAAT GCGCTCATCG TCATCCTCGG
121       CACCGTCACC CTGGATGCTG TAGGCATAGG CTTGGTTATG CCGGTACTGC CGGGCCTCTT
181       GCGGGATATC GTCCATTCCG ACAGCATCGC CAGTCACTAT GGCGTGCTGC TAGCGCTATA
241       TGCGTTGATG CAATTTCTAT GCGCACCCGT TCTCGGAGCA CTGTCCGACC GCTTTGGCCG
301       CCGCCCAGTC CTGCTCGCTT CGCTACTTGG AGCCACTATC GACTACGCGA TCATGGCGAC
361       CACACCCGTC CTGTGGATAT CCGGATATAG TTCCTCCTTT CAGCAAAAAA CCCCTCAAGA
421       CCCGTTTAGA GGCCCCAAGG GGTTATGCTA GTTATTGCTC AGCGGTGGCA GCAGCCAACT
481       CAGCTTCCTT TCGGGCTTTG TTAGCAGCCG GATCCGCGAC CCATTTGCTG TCCACCAGTC
541       ATGCTAGCCA TATGTATATC TCCTTCTTAA AGTTAAACAA AATTATTTCT AGAGGGAAAC
601       CGTTGTGGTC TCCCTATAGT GAGTCGTATT AATTTCGCGG GATCGAGATC TCGATCCTCT
661       ACGCCGGACG CATCGTGGCC GGCATCACCG GCGCCACAGG TGCGGTTGCT GGCGCCTATA
721       TCGCCGACAT CACCGATGGG GAAGATCGGG CTCGCCACTT CGGGCTCATG AGCGCTTGTT
781       TCGGCGTGGG TATGGTGGCA GGCCCCGTGG CCGGGGGACT GTTGGGCGCC ATCTCCTTGC
841       ATGCACCATT CCTTGCGGCG GCGGTGCTCA ACGGCCTCAA CCTACTACTG GGCTGCTTCC
901       TAATGCAGGA GTCGCATAAG GGAGAGCGTC GACCGATGCC CTTGAGAGCC TTCAACCCAG
961       TCAGCTCCTT CCGGTGGGCG CGGGGCATGA CTATCGTCGC CGCACTTATG ACTGTCTTCT
1021      TTATCATGCA ACTCGTAGGA CAGGTGCCGG CAGCGCTCTG GGTCATTTTC GGCGAGGACC
1081      GCTTTCGCTG GAGCGCGACG ATGATCGGCC TGTCGCTTGC GGTATTCGGA ATCTTGCACG
1141      CCCTCGCTCA AGCCTTCGTC ACTGGTCCCG CCACCAAACG TTTCGGCGAG AAGCAGGCCA
1201      TTATCGCCGG CATGGCGGCC GACGCGCTGG GCTACGTCTT GCTGGCGTTC GCGACGCGAG
1261      GCTGGATGGC CTTCCCCATT ATGATTCTTC TCGCTTCCGG CGGCATCGGG ATGCCCGCGT
1321      TGCAGGCCAT GCTGTCCAGG CAGGTAGATG ACGACCATCA GGGACAGCTT CAAGGATCGC
1381      TCGCGGCTCT TACCAGCCTA ACTTCGATCA CTGGACCGCT GATCGTCACG GCGATTTATG
1441      CCGCCTCGGC GAGCACATGG AACGGGTTGG CATGGATTGT AGGCGCCGCC CTATACCTTG
1501      TCTGCCTCCC CGCGTTGCGT CGCGGTGCAT GGAGCCGGGC CACCTCGACC TGAATGGAAG
1561      CCGGCGGCAC CTCGCTAACG GATTCACCAC TCCAAGAATT GGAGCCAATC AATTCTTGCG
1621      GAGAACTGTG AATGCGCAAA CCAACCCTTG GCAGAACATA TCCATCGCGT CCGCCATCTC
```

```
1681      CAGCAGCCGC ACGCGGCGCA TCTCGGGCAG CGTTGGGTCC TGGCCACGGG TGCGCATGAT
1741      CGTGCTCCTG TCGTTGAGGA CCCGGCTAGG CTGGCGGGGT TGCCTTACTG GTTAGCAGAA
1801      TGAATCACCG ATACGCGAGC GAACGTGAAG CGACTGCTGC TGCAAAACGT CTGCGACCTG
1861      AGCAACAACA TGAATGGTCT TCGGTTTCCG TGTTTCGTAA AGTCTGGAAA CGCGGAAGTC
1921      AGCGCCCTGC ACCATTATGT TCCGGATCTG CATCGCAGGA TGCTGCTGGC TACCCTGTGG
1981      AACACCTACA TCTGTATTAA CGAAGCGCTG GCATTGACCC TGAGTGATTT TTCTCTGGTC
2041      CCGCCGCATC CATACCGCCA GTTGTTTACC CTCACAACGT TCCAGTAACC GGGCATGTTC
2101      ATCATCAGTA ACCCGTATCG TGAGCATCCT CTCTCGTTTC ATCGGTATCA TTACCCCCAT
2161      GAACAGAAAT CCCCCTTACA CGGAGGCATC AGTGACCAAA CAGGAAAAAA CCGCCCTTAA
2221      CATGGCCCGC TTTATCAGAA GCCAGACATT AACGCTTCTG GAGAAACTCA ACGAGCTGGA
2281      CGCGGATGAA CAGGCAGACA TCTGTGAATC GCTTCACGAC CACGCTGATG AGCTTTACCG
2341      CAGCTGCCTC GCGCGTTTCG GTGATGACGG TGAAAACCTC TGACACATGC AGCTCCCGGA
2401      GACGGTCACA GCTTGTCTGT AAGCGGATGC CGGGAGCAGA CAAGCCCGTC AGGGCGCGTC
2461      AGCGGGTGTT GGCGGGTGTC GGGGCGCAGC CATGACCCAG TCACGTAGCG ATAGCGGAGT
2521      GTATACTGGC TTAACTATGC GGCATCAGAG CAGATTGTAC TGAGAGTGCA CCATATATGC
2581      GGTGTGAAAT ACCGCACAGA TGCGTAAGGA GAAAATACCG CATCAGGCGC TCTTCCGCTT
2641      CCTCGCTCAC TGACTCGCTG CGCTCGGTCG TTCGGCTGCG GCGAGCGGTA TCAGCTCACT
2701      CAAAGGCGGT AATACGGTTA TCCACAGAAT CAGGGGATAA CGCAGGAAAG AACATGTGAG
2761      CAAAAGGCCA GCAAAAGGCC AGGAACCGTA AAAAGGCCGC GTTGCTGGCG TTTTTCCATA
2821      GGCTCCGCCC CCCTGACGAG CATCACAAAA ATCGACGCTC AAGTCAGAGG TGGCGAAACC
2881      CGACAGGACT ATAAAGATAC CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG
2941      TTCCGACCCT GCCGCTTACC GGATACCTGT CCGCCTTTCT CCCTTCGGGA AGCGTGGCGC
3001      TTTCTCATAG CTCACGCTGT AGGTATCTCA GTTCGGTGTA GGTCGTTCGC TCCAAGCTGG
3061      GCTGTGTGCA CGAACCCCCC GTTCAGCCCG ACCGCTGCGC CTTATCCGGT AACTATCGTC
3121      TTGAGTCCAA CCCGGTAAGA CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAACAGGA
3181      TTAGCAGAGC GAGGTATGTA GGCGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG
3241      GCTACACTAG AAGGACAGTA TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA
3301      AAAGAGTTGG TAGCTCTTGA TCCGGCAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTTG
3361      TTTGCAAGCA GCAGATTACG CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTTT
3421      CTACGGGGTC TGACGCTCAG TGGAACGAAA ACTCACGTTA AGGGATTTTG GTCATGAGAT
3481      TATCAAAAAG GATCTTCACC TAGATCCTTT TAAATTAAAA ATGAAGTTTT AAATCAATCT
3541      AAAGTATATA TGAGTAAACT TGGTCTGACA GTTACCAATG CTTAATCAGT GAGGCACCTA
3601      TCTCAGCGAT CTGTCTATTT CGTTCATCCA TAGTTGCCTG ACTCCCCGTC GTGTAGATAA
3661      CTACGATACG GGAGGGCTTA CCATCTGGCC CCAGTGCTGC AATGATACCG CGAGACCCAC
3721      GCTCACCGGC TCCAGATTTA TCAGCAATAA ACCAGCCAGC CGGAAGGGCC GAGCGCAGAA
3781      GTGGTCCTGC AACTTTATCC GCCTCCATCC AGTCTATTAA TTGTTGCCGG GAAGCTAGAG
3841      TAAGTAGTTC GCCAGTTAAT AGTTTGCGCA ACGTTGTTGC CATTGCTGCA GGCATCGTGG
3901      TGTCACGCTC GTCGTTTGGT ATGGCTTCAT TCAGCTCCGG TTCCCAACGA TCAAGGCGAG
3961      TTACATGATC CCCCATGTTG TGCAAAAAAG CGGTTAGCTC CTTCGGTCCT CCGATCGTTG
4021      TCAGAAGTAA GTTGGCCGCA GTGTTATCAC TCATGGTTAT GGCAGCACTG CATAATTCTC
4081      TTACTGTCAT GCCATCCGTA AGATGCTTTT CTGTGACTGG TGAGTACTCA ACCAAGTCAT
4141      TCTGAGAATA GTGTATGCGG CGACCGAGTT GCTCTTGCCC GGCGTCAACA CGGGATAATA
4201      CCGCGCCACA TAGCAGAACT TTAAAAGTGC TCATCATTGG AAAACGTTCT TCGGGGCGAA
4261      AACTCTCAAG GATCTTACCG CTGTTGAGAT CCAGTTCGAT GTAACCCACT CGTGCACCCA
4321      ACTGATCTTC AGCATCTTTT ACTTTCACCA GCGTTTCTGG GTGAGCAAAA ACAGGAAGGC
4381      AAAATGCCGC AAAAAAGGGA ATAAGGGCGA CACGGAAATG TTGAATACTC ATACTCTTCC
4441      TTTTTCAATA TTATTGAAGC ATTTATCAGG GTTATTGTCT CATGAGCGGA TACATATTTG
4501      AATGTATTTA GAAAAATAAA CAAATAGGGG TTCCGCGCAC ATTTCCCCGA AAAGTGCCAC

4561      CTGACGTCTA AGAAACCATT ATTATCATGA CATTAACCTA
```

pET28a (5,369 bp) – ΔDot1 S542C expression

```
1         ATCCGGATAT AGTTCCTCCT TTCAGCAAAA AACCCCTCAA GACCCGTTTA GAGGCCCCAA
61        GGGGTTATGC TAGTTATTGC TCAGCGGTGG CAGCAGCCAA CTCAGCTTCC TTTCGGGCTT
121       TGTTAGCAGC CGGATCTCAG TGGTGGTGGT GGTGGTGCTC GAGTGCGGCC GCAAGCTTGT
181       CGACGGAGCT CGAATTCGGA TCCGCGACCC ATTTGCTGTC CACCAGTCAT GCTAGCCATA
241       TGGCTGCCGC GCGGCACCAG GCCGCTGCTG TGATGATGAT GATGATGGCT GCTGCCCATG
301       GTATATCTCC TTCTTAAAGT TAAACAAAAT TATTTCTAGA GGGGAATTGT TATCCGCTCA
361       CAATTCCCCT ATAGTGAGTC GTATTAATTT CGCGGGATCG AGATCTCGAT CCTCTACGCC
```

```
 421        GGACGCATCG TGGCCGGCAT CACCGGCGCC ACAGGTGCGG TTGCTGGCGC CTATATCGCC
 481        GACATCACCG ATGGGGAAGA TCGGGCTCGC CACTTCGGGC TCATGAGCGC TTGTTTCGGC
 541        GTGGGTATGG TGGCAGGCCC CGTGGCCGGG GGACTGTTGG GCGCCATCTC CTTGCATGCA
 601        CCATTCCTTG CGGCGGCGGT GCTCAACGGC CTCAACCTAC TACTGGGCTG CTTCCTAATG
 661        CAGGAGTCGC ATAAGGGAGA GCGTCGAGAT CCCGGACACC ATCGAATGGC GCAAAACCTT
 721        TCGCGGTATG GCATGATAGC GCCCGGAAGA GAGTCAATTC AGGGTGGTGA ATGTGAAACC
 781        AGTAACGTTA TACGATGTCG CAGAGTATGC CGGTGTCTCT TATCAGACCG TTTCCCGCGT
 841        GGTGAACCAG GCCAGCCACG TTTCTGCGAA AACGCGGGAA AAAGTGGAAG CGGCGATGGC
 901        GGAGCTGAAT TACATTCCCA ACCGCGTGGC ACAACAACTG GCGGGCAAAC AGTCGTTGCT
 961        GATTGGCGTT GCCACCTCCA GTCTGGCCCT GCACGCGCCG TCGCAAATTG TCGCGGCGAT
1021        TAAATCTCGC GCCGATCAAC TGGGTGCCAG CGTGGTGGTG TCGATGGTAG AACGAAGCGG
1081        CGTCGAAGCC TGTAAAGCGG CGGTGCACAA TCTTCTCGCG CAACGCGTCA GTGGGCTGAT
1141        CATTAACTAT CCGCTGGATG ACCAGGATGC CATTGCTGTG GAAGCTGCCT GCACTAATGT
1201        TCCGGCGTTA TTTCTTGATG TCTCTGACCA GACACCCATC AACAGTATTA TTTTCTCCCA
1261        TGAAGACGGT ACGCGACTGG GCGTGGAGCA TCTGGTCGCA TTGGGTCACC AGCAAATCGC
1321        GCTGTTAGCG GGCCCATTAA GTTCTGTCTC GGCGCGTCTG CGTCTGGCTG GCTGGCATAA
1381        ATATCTCACT CGCAATCAAA TTCAGCCGAT AGCGGAACGG GAAGGCGACT GGAGTGCCAT
1441        GTCCGGTTTT CAACAAACCA TGCAAATGCT GAATGAGGGC ATCGTTCCCA CTGCGATGCT
1501        GGTTGCCAAC GATCAGATGG CGCTGGGCGC AATGCGCGCC ATTACCGAGT CCGGGCTGCG
1561        CGTTGGTGCG GATATCTCGG TAGTGGGATA CGACGATACC GAAGACAGCT CATGTTATAT
1621        CCCGCCGTTA ACCACCATCA AACAGGATTT TCGCCTGCTG GGGCAAACCA GCGTGGACCG
1681        CTTGCTGCAA CTCTCTCAGG GCCAGGCGGT GAAGGGCAAT CAGCTGTTGC CCGTCTCACT
1741        GGTGAAAAGA AAAACCACCC TGGCGCCCAA TACGCAAACC GCCTCTCCCC GCGCGTTGGC
1801        CGATTCATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG GAAAGCGGGC AGTGAGCGCA
1861        ACGCAATTAA TGTAAGTTAG CTCACTCATT AGGCACCGGG ATCTCGACCG ATGCCCTTGA
1921        GAGCCTTCAA CCCAGTCAGC TCCTTCCGGT GGGCGCGGGG CATGACTATC GTCGCCGCAC
1981        TTATGACTGT CTTCTTTATC ATGCAACTCG TAGGACAGGT GCCGGCAGCG CTCTGGGTCA
2041        TTTTCGGCGA GGACCGCTTT CGCTGGAGCG CGACGATGAT CGGCCTGTCG CTTGCGGTAT
2101        TCGGAATCTT GCACGCCCTC GCTCAAGCCT TCGTCACTGG TCCCGCCACC AAACGTTTCG
2161        GCGAGAAGCA GGCCATTATC GCCGGCATGG CGGCCCCACG GGTGCGCATG ATCGTGCTCC
2221        TGTCGTTGAG GACCCGGCTA GGCTGGCGGG GTTGCCTTAC TGGTTAGCAG AATGAATCAC
2281        CGATACGCGA GCGAACGTGA AGCGACTGCT GCTGCAAAAC GTCTGCGACC TGAGCAACAA
2341        CATGAATGGT CTTCGGTTTC CGTGTTTCGT AAAGTCTGGA AACGCGGAAG TCAGCGCCCT
2401        GCACCATTAT GTTCCGGATC TGCATCGCAG GATGCTGCTG GCTACCCTGT GGAACACCTA
2461        CATCTGTATT AACGAAGCGC TGGCATTGAC CCTGAGTGAT TTTTCTCTGG TCCCGCCGCA
2521        TCCATACCGC CAGTTGTTTA CCCTCACAAC GTTCCAGTAA CCGGGCATGT TCATCATCAG
2581        TAACCCGTAT CGTGAGCATC CTCTCTCGTT TCATCGGTAT CATTACCCCC ATGAACAGAA
2641        ATCCCCCTTA CACGGAGGCA TCAGTGACCA AACAGGAAAA AACCGCCCTT AACATGGCCC
2701        GCTTTATCAG AAGCCAGACA TTAACGCTTC TGGAGAAACT CAACGAGCTG GACGCGGATG
2761        AACAGGCAGA CATCTGTGAA TCGCTTCACG ACCACGCTGA TGAGCTTTAC CGCAGCTGCC
2821        TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA
2881        CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG
2941        TTGGCGGGTG TCGGGGCGCA GCCATGACCC AGTCACGTAG CGATAGCGGA GTGTATACTG
3001        GCTTAACTAT GCGGCATCAG AGCAGATTGT ACTGAGAGTG CACCATATAT GCGGTGTGAA
3061        ATACCGCACA GATGCGTAAG GAGAAAATAC CGCATCAGGC GCTCTTCCGC TTCCTCGCTC
3121        ACTGACTCGC TGCGCTCGGT CGTTCGGCTG CGGCGAGCGG TATCAGCTCA CTCAAAGGCG
3181        GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG AGCAAAAGGC
3241        CAGCAAAAGG CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC
3301        CCCCCTGACG AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA
3361        CTATAAAGAT ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TGCGCTCTCC TGTTCCGACC
3421        CTGCCGCTTA CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC GCTTTCTCAT
3481        AGCTCACGCT GTAGGTATCT CAGTTCGGTG TAGGTCGTTC GCTCCAAGCT GGGCTGTGTG
3541        CACGAACCCC CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACTATCG TCTTGAGTCC
3601        AACCCGGTAA GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA
3661        GCGAGGTATG TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT
3721        AGAAGGACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT
3781        GGTAGCTCTT GATCCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTTTTT TGTTTGCAAG
3841        CAGCAGATTA CGCGCAGAAA AAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG
3901        TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAA CAATAAAACT
3961        GTCTGCTTAC ATAAACAGTA ATACAAGGGG TGTTATGAGC CATATTCAAC GGGAAACGTC
4021        TTGCTCTAGG CCGCGATTAA ATTCCAACAT GGATGCTGAT TTATATGGGT ATAAATGGGC
```

125

```
4081      TCGCGATAAT GTCGGGCAAT CAGGTGCGAC AATCTATCGA TTGTATGGGA AGCCCGATGC
4141      GCCAGAGTTG TTTCTGAAAC ATGGCAAAGG TAGCGTTGCC AATGATGTTA CAGATGAGAT
4201      GGTCAGACTA AACTGGCTGA CGGAATTTAT GCCTCTTCCG ACCATCAAGC ATTTTATCCG
4261      TACTCCTGAT GATGCATGGT TACTCACCAC TGCGATCCCC GGGAAAACAG CATTCCAGGT
4321      ATTAGAAGAA TATCCTGATT CAGGTGAAAA TATTGTTGAT GCGCTGGCAG TGTTCCTGCG
4381      CCGGTTGCAT TCGATTCCTG TTTGTAATTG TCCTTTTAAC AGCGATCGCG TATTTCGTCT
4441      CGCTCAGGCG CAATCACGAA TGAATAACGG TTTGGTTGAT GCGAGTGATT TTGATGACGA
4501      GCGTAATGGC TGGCCTGTTG AACAAGTCTG GAAAGAAATG CATAAACTTT TGCCATTCTC
4561      ACCGGATTCA GTCGTCACTC ATGGTGATTT CTCACTTGAT AACCTTATTT TTGACGAGGG
4621      GAAATTAATA GGTTGTATTG ATGTTGGACG AGTCGGAATC GCAGACCGAT ACCAGGATCT
4681      TGCCATCCTA TGGAACTGCC TCGGTGAGTT TTCTCCTTCA TTACAGAAAC GGCTTTTTCA
4741      AAAATATGGT ATTGATAATC CTGATATGAA TAAATTGCAG TTTCATTTGA TGCTCGATGA
4801      GTTTTTCTAA GAATTAATTC ATGAGCGGAT ACATATTTGA ATGTATTTAG AAAAATAAAC
4861      AAATAGGGGT TCCGCGCACA TTTCCCCGAA AAGTGCCACC TGAAATTGTA AACGTTAATA
4921      TTTTGTTAAA ATTCGCGTTA AATTTTTGTT AAATCAGCTC ATTTTTTAAC CAATAGGCCG
4981      AAATCGGCAA AATCCCTTAT AAATCAAAAG AATAGACCGA GATAGGGTTG AGTGTTGTTC
5041      CAGTTTGGAA CAAGAGTCCA CTATTAAAGA ACGTGGACTC CAACGTCAAA GGGCGAAAAA
5101      CCGTCTATCA GGGCGATGGC CCACTACGTG AACCATCACC CTAATCAAGT TTTTTGGGGT
5161      CGAGGTGCCG TAAAGCACTA AATCGGAACC CTAAAGGGAG CCCCCGATTT AGAGCTTGAC
5221      GGGGAAAGCC GGCGAACGTG GCGAGAAAGG AAGGGAAGAA AGCGAAAGGA GCGGGCGCTA
5281      GGGCGCTGGC AAGTGTAGCG GTCACGCTGC GCGTAACCAC CACACCCGCC GCGCTTAATG

5341      CGCCGCTACA GGGCGCGTCC CATTCGCCA
```

8.3. Protein sequences

Xenopus laevis histone H2A.1 (130 amino acids, 14,047 Da)

```
              10        20        30        40        50        60
1       ATGTCAGGAAGAGGCAAACAAGGCGGTAAAACCCGCGCTAAGGCCAAGACTCGCTCATCT
1         M   S   G   R   G   K   Q   G   G   K   T   R   A   K   A   K   T   R   S   S

              70        80        90       100       110       120
61      CGGGCTGGGCTACAGTTCCCTGTTGGCCGTGTTCACCGGCTGTTAAGGAAAGGCAATTAT
21        R   A   G   L   Q   F   P   V   G   R   V   H   R   L   L   R   K   G   N   Y

             130       140       150       160       170       180
121     GCAGAGCGGGTGGGAGCTGGAGCTCCAGTCTATCTGGCTGCAGTGTTGGAGTATCTGACC
41        A   E   R   V   G   A   G   A   P   V   Y   L   A   A   V   L   E   Y   L   T

             190       200       210       220       230       240
181     GCTGAGATTTTGGAATTGGCCGGGAATGCGGCCCGTGATAACAAGAAGACTCGCATTATC
61        A   E   I   L   E   L   A   G   N   A   A   R   D   N   K   K   T   R   I   I

             250       260       270       280       290       300
241     CCCAGACACCTGCAGCTCGCTGTGCGCAACGATGAGGAACTGAACAAACTGCTCGGAAGA
81        P   R   H   L   Q   L   A   V   R   N   D   E   E   L   N   K   L   L   G   R

             310       320       330       340       350       360
301     GTCACTATCGCTCAGGGCGGGGTCCTGCCCAACATCCAGTCCGTGCTGCTGCCCAAGAAA
101       V   T   I   A   Q   G   G   V   L   P   N   I   Q   S   V   L   L   P   K   K

             370       380       390
361     ACCGAGAGTTCCAAGTCGGCCAAGAGCAAGTGA
121       T   E   S   S   K   S   A   K   S   K   *
```

Xenopus laevis histone H2A.1 N38C (130 amino acids, 14,036 Da)

```
                  10        20        30        40        50        60
1        ATGTCAGGAAGAGGCAAACAAGGCGGTAAAACCCGCGCTAAGGCCAAGACTCGCTCATCT
1          M   S   G   R   G   K   Q   G   G   K   T   R   A   K   A   K   T   R   S   S

                  70        80        90       100       110       120
61       CGGGCTGGGCTACAGTTCCCTGTTGGCCGTGTTCACCGGCTGTTAAGGAAAGGCTGCTAT
21         R   A   G   L   Q   F   P   V   G   R   V   H   R   L   L   R   K   G   C   Y

                 130       140       150       160       170       180
121      GCAGAGCGGGTGGGAGCTGGAGCTCCAGTCTATCTGGCTGCAGTGTTGGAGTATCTGACC
41         A   E   R   V   G   A   G   A   P   V   Y   L   A   A   V   L   E   Y   L   T

                 190       200       210       220       230       240
181      GCTGAGATTTTGGAATTGGCCGGGAATGCGGCCCGTGATAACAAGAAGACTCGCATTATC
61         A   E   I   L   E   L   A   G   N   A   A   R   D   N   K   K   T   R   I   I

                 250       260       270       280       290       300
241      CCCAGACACCTGCAGCTCGCTGTGCGCAACGATGAGGAACTGAACAAACTGCTCGGAAGA
81         P   R   H   L   Q   L   A   V   R   N   D   E   E   L   N   K   L   L   G   R

                 310       320       330       340       350       360
301      GTCACTATCGCTCAGGGCGGGGTCCTGCCCAACATCCAGTCCGTGCTGCTGCCCAAGAAA
101        V   T   I   A   Q   G   G   V   L   P   N   I   Q   S   V   L   L   P   K   K

                 370       380       390
361      ACCGAGAGTTCCAAGTCGGCCAAGAGCAAGTGA
121        T   E   S   S   K   S   A   K   S   K   *
```

Xenopus laevis histone H2B.1 (123 amino acids, 13,591 Da)

```
                  10        20        30        40        50        60
1        ATGGCCAAGTCCGCTCCAGCCCCCGAAGAAAGGCTCCAAGAAAGCGGTGACCAAGACTCAG
1          M   A   K   S   A   P   A   P   K   K   G   S   K   K   A   V   T   K   T   Q

                  70        80        90       100       110       120
61       AAGAAAGACGGGAAAAAGCGCAGGAAGACAAGGAAGGAGAGTTATGCCATTTACGTGTAC
21         K   K   D   G   K   K   R   R   K   T   R   K   E   S   Y   A   I   Y   V   Y

                 130       140       150       160       170       180
121      AAGGTGCTGAAGCAGGTGCACCCCGATACCGGCATCTCGTCCAAGGCCATGAGCATCATG
41         K   V   L   K   Q   V   H   P   D   T   G   I   S   S   K   A   M   S   I   M

                 190       200       210       220       230       240
181      AACTCCTTTGTCAACGATGTGTTTGAGCGCATCGCAGGGGAAGCCTCCCGCCTGGCTCAT
61         N   S   F   V   N   D   V   F   E   R   I   A   G   E   A   S   R   L   A   H

                 250       260       270       280       290       300
241      TACAACAAGCGCTCCACCATCACCTCCCGGGAGATCCAGACCGCGGTCCGACTGCTGCTG
81         Y   N   K   R   S   T   I   T   S   R   E   I   Q   T   A   V   R   L   L   L

                 310       320       330       340       350       360
301      CCTGGGGAGTTGGCCAAACACGCCGTGTCCGAGGGCACCAAGGCTGTCACCAAGTACACC
101        P   G   E   L   A   K   H   A   V   S   E   G   T   K   A   V   T   K   Y   T

                 370
361      AGCGCCAAGTAA
121        S   A   K   *
```

127

Xenopus laevis histone H3 C110A (wt) (136 amino acids, 15,334 Da)

```
                  10        20        30        40        50        60
1       ATGGCCCGTACCAAGCAGACCGCCCGTAAATCCACCGGAGGGAAGGCTCCCCGCAAGCAG
1        M   A   R   T   K   Q   T   A   R   K   S   T   G   G   K   A   P   R   K   Q

                  70        80        90       100       110       120
61      CTGGCCACCAAGGCAGCCAGGAAGTCCGCTCCTGCTACCGGCGGAGTCAAGAAACCTCAC
21       L   A   T   K   A   A   R   K   S   A   P   A   T   G   G   V   K   K   P   H

                 130       140       150       160       170       180
121     CGTTACCGGCCCGGCACAGTCGCTCTCCGCGAGATCCGCCGCTACCAGAAATCCACCGAG
41       R   Y   R   P   G   T   V   A   L   R   E   I   R   R   Y   Q   K   S   T   E

                 190       200       210       220       230       240
181     CTGCTCATCCGCAAACTGCCTTTCCAGCGCCTGGTCCGGGAGATCGCTCAGGACTTCAAG
61       L   L   I   R   K   L   P   F   Q   R   L   V   R   E   I   A   Q   D   F   K

                 250       260       270       280       290       300
241     ACCGACCTGCGCTTCCAGAGCTCGGCCGTTATGGCTCTGCAGGAGGCCAGCGAGGCTTAT
81       T   D   L   R   F   Q   S   S   A   V   M   A   L   Q   E   A   S   E   A   Y

                 310       320       330       340       350       360
301     CTGGTCGCTCTCTTTGAGGACACCAACCTGGCTGCCATCCACGCCAAGAGGGTCACCATC
101      L   V   A   L   F   E   D   T   N   L   A   A   I   H   A   K   R   V   T   I

                 370       380       390       400       410
361     ATGCCCAAGGACATCCAGCTGGCCCGCAGAATCCGAGGCGAGAGGGCTTAG
121      M   P   K   D   I   Q   L   A   R   R   I   R   G   E   R   A   *
```

Xenopus laevis histone H3 R40**C** C110A (136 amino acids, 15,281 Da)

```
                  10        20        30        40        50        60
1       ATGGCCCGTACCAAGCAGACCGCCCGTAAATCCACCGGAGGGAAGGCTCCCCGCAAGCAG
1        M   A   R   T   K   Q   T   A   R   K   S   T   G   G   K   A   P   R   K   Q

                  70        80        90       100       110       120
61      CTGGCCACCAAGGCAGCCAGGAAGTCCGCTCCTGCTACCGGCGGAGTCAAGAAACCTCAC
21       L   A   T   K   A   A   R   K   S   A   P   A   T   G   G   V   K   K   P   H

                 130       140       150       160       170       180
121     TGCTACCGGCCCGGCACAGTCGCTCTCCGCGAGATCCGCCGCTACCAGAAATCCACCGAG
41       C   Y   R   P   G   T   V   A   L   R   E   I   R   R   Y   Q   K   S   T   E

                 190       200       210       220       230       240
181     CTGCTCATCCGCAAACTGCCTTTCCAGCGCCTGGTCCGGGAGATCGCTCAGGACTTCAAG
61       L   L   I   R   K   L   P   F   Q   R   L   V   R   E   I   A   Q   D   F   K

                 250       260       270       280       290       300
241     ACCGACCTGCGCTTCCAGAGCTCGGCCGTTATGGCTCTGCAGGAGGCCAGCGAGGCTTAT
81       T   D   L   R   F   Q   S   S   A   V   M   A   L   Q   E   A   S   E   A   Y

                 310       320       330       340       350       360
301     CTGGTCGCTCTCTTTGAGGACACCAACCTGGCTGCCATCCACGCCAAGAGGGTCACCATC
101      L   V   A   L   F   E   D   T   N   L   A   A   I   H   A   K   R   V   T   I

                 370       380       390       400       410
361     ATGCCCAAGGACATCCAGCTGGCCCGCAGAATCCGAGGCGAGAGGGCTTAG
121      M   P   K   D   I   Q   L   A   R   R   I   R   G   E   R   A   *
```

128

Xenopus laevis histone H3 H39**C** C110A (136 amino acids, 15,300 Da)

```
                10        20        30        40        50        60
1       ATGGCCCGTACCAAGCAGACCGCCCGTAAATCCACCGGAGGGAAGGCTCCCCGCAAGCAG
1        M   A   R   T   K   Q   T   A   R   K   S   T   G   G   K   A   P   R   K   Q

                70        80        90        100       110       120
61      CTGGCCACCAAGGCAGCCAGGAAGTCCGCTCCTGCTACCGGCGGAGTCAAGAAACCTTGC
21       L   A   T   K   A   A   R   K   S   A   P   A   T   G   G   V   K   K   P   C

                130       140       150       160       170       180
121     CGTTACCGGCCCGGCACAGTCGCTCTCCGCGAGATCCGCCGCTACCAGAAATCCACCGAG
41       R   Y   R   P   G   T   V   A   L   R   E   I   R   R   Y   Q   K   S   T   E

                190       200       210       220       230       240
181     CTGCTCATCCGCAAACTGCCTTTCCAGCGCCTGGTCCGGGAGATCGCTCAGGACTTCAAG
61       L   L   I   R   K   L   P   F   Q   R   L   V   R   E   I   A   Q   D   F   K

                250       260       270       280       290       300
241     ACCGACCTGCGCTTCCAGAGCTCGGCCGTTATGGCTCTGCAGGAGGCCAGCGAGGCTTAT
81       T   D   L   R   F   Q   S   S   A   V   M   A   L   Q   E   A   S   E   A   Y

                310       320       330       340       350       360
301     CTGGTCGCTCTCTTTGAGGACACCAACCTGGCTGCCATCCACGCCAAGAGGGTCACCATC
101      L   V   A   L   F   E   D   T   N   L   A   A   I   H   A   K   R   V   T   I

                370       380       390       400       410
361     ATGCCCAAGGACATCCAGCTGGCCCGCAGAATCCGAGGCGAGAGGGCTTAG
121      M   P   K   D   I   Q   L   A   R   R   I   R   G   E   R   A   *
```

Xenopus laevis histone H3 R40C T80C C110A (136 amino acids, 15,283 Da)

```
                10        20        30        40        50        60
1       ATGGCCCGTACCAAGCAGACCGCCCGTAAATCCACCGGAGGGAAGGCTCCCCGCAAGCAG
1        M   A   R   T   K   Q   T   A   R   K   S   T   G   G   K   A   P   R   K   Q

                70        80        90        100       110       120
61      CTGGCCACCAAGGCAGCCAGGAAGTCCGCTCCTGCTACCGGCGGAGTCAAGAAACCTCAC
21       L   A   T   K   A   A   R   K   S   A   P   A   T   G   G   V   K   K   P   H

                130       140       150       160       170       180
121     TGCTACCGGCCCGGCACAGTCGCTCTCCGCGAGATCCGCCGCTACCAGAAATCCACCGAG
41       C   Y   R   P   G   T   V   A   L   R   E   I   R   R   Y   Q   K   S   T   E

                190       200       210       220       230       240
181     CTGCTCATCCGCAAACTGCCTTTCCAGCGCCTGGTCCGGGAGATCGCTCAGGACTTCAAG
61       L   L   I   R   K   L   P   F   Q   R   L   V   R   E   I   A   Q   D   F   K

                250       260       270       280       290       300
241     TGTGACCTGCGCTTCCAGAGCTCGGCCGTTATGGCTCTGCAGGAGGCCAGCGAGGCTTAT
81       C   D   L   R   F   Q   S   S   A   V   M   A   L   Q   E   A   S   E   A   Y

                310       320       330       340       350       360
301     CTGGTCGCTCTCTTTGAGGACACCAACCTGGCTGCCATCCACGCCAAGAGGGTCACCATC
101      L   V   A   L   F   E   D   T   N   L   A   A   I   H   A   K   R   V   T   I

                370       380       390       400       410
361     ATGCCCAAGGACATCCAGCTGGCCCGCAGAATCCGAGGCGAGAGGGCTTAG
121      M   P   K   D   I   Q   L   A   R   R   I   R   G   E   R   A   *
```

Xenopus laevis histone H4 (103 amino acids, 11,336 Da)

```
                    10        20        30        40        50        60
1    ATGTCCGGTCGTGGTAAAGGTGGTAAAGGTCTGGGTAAAGGTGGTGCTAAACGTCACCGT
1      M   S   G   R   G   K   G   G   K   G   L   G   K   G   G   A   K   R   H   R

                    70        80        90       100       110       120
61   AAAGTTCTGCGTGACAACATCCAGGGTATCACCAAGCCGGCTATCCGTCGTCTGGCTCGT
21     K   V   L   R   D   N   I   Q   G   I   T   K   P   A   I   R   R   L   A   R

                   130       140       150       160       170       180
121  CGTGGTGGTGTTAAACGTATCTCCGGTCTGATCTACGAAGAAACCCGCGGTGTTCTGAAA
41     R   G   G   V   K   R   I   S   G   L   I   Y   E   E   T   R   G   V   L   K

                   190       200       210       220       230       240
181  GTTTTCCTGGAAAACGTTATCCGTGACGCTGTTACCTACACCGAACACGCTAAACGTAAA
61     V   F   L   E   N   V   I   R   D   A   V   T   Y   T   E   H   A   K   R   K

                   250       260       270       280       290       300
241  ACCGTTACCGCTATGGACGTTGTTTACGCTCTGAAACGTCAGGGTCGTACCCTGTACGGT
81     T   V   T   A   M   D   V   V   Y   A   L   K   R   Q   G   R   T   L   Y   G

                   310
301  TTCGGTGGTTAA
101    F   G   G   *
```

Saccharomyces cerevisiae Dot1 S542C (Δ157) ≙ ΔDot1 S542**C** (426 amino acids, 48,895 Da)

```
                    10        20        30        40        50        60
1    ATGAAAAAGCCATTGAAAAAAGGTCGGGCAAACAAGAAGAATGATCGCGATAGTCCTTCA
1      M   K   K   P   L   K   K   G   R   A   N   K   K   N   D   R   D   S   P   S

                    70        80        90       100       110       120
61   TCAACATTTGTTGATTGGAATGGCCCGTGTCTACGGTTACAATATCCATTATTTGACATA
21     S   T   F   V   D   W   N   G   P   C   L   R   L   Q   Y   P   L   F   D   I

                   130       140       150       160       170       180
121  GAGTACTTAAGATCACATGAAATATATTCTGGAACTCCTATACAATCCATTAGTTTAAGA
41     E   Y   L   R   S   H   E   I   Y   S   G   T   P   I   Q   S   I   S   L   R

                   190       200       210       220       230       240
181  ACAAATTCTCCACAGCCAACGAGCTTGACATCAGATAACGACACTTCCTCAGTAACGACA
61     T   N   S   P   Q   P   T   S   L   T   S   D   N   D   T   S   S   V   T   T

                   250       260       270       280       290       300
241  GCAAAGTTGCAGAGTATTTTATTTTCAAATTATATGGAAGAGTACAAAGTCGACTTCAAA
81     A   K   L   Q   S   I   L   F   S   N   Y   M   E   E   Y   K   V   D   F   K

                   310       320       330       340       350       360
301  AGGTCAACAGCCATTTATAATCCAATGAGTGAAATTGGTAAATTAATTGAATACAGCTGC
101    R   S   T   A   I   Y   N   P   M   S   E   I   G   K   L   I   E   Y   S   C

                   370       380       390       400       410       420
361  CTGGTCTTTTTTACCTTCACCTTATGCTGAACAATTGAAGGAAACTATACTACCGGACCTA
121    L   V   F   L   P   S   P   Y   A   E   Q   L   K   E   T   I   L   P   D   L

                   430       440       450       460       470       480
421  AATGCATCATTTGATAACTCTGACACGAAAGGTTTCGTGAATGCTATAAATTTATACAAC
141    N   A   S   F   D   N   S   D   T   K   G   F   V   N   A   I   N   L   Y   N
```

130

```
               490       500       510       520       530       540
481    AAAATGATTCGTGAAATTCCTAGGCAAAGAATAATTGACCATTTAGAAACAATTGATAAA
161        K   M   I   R   E   I   P   R   Q   R   I   I   D   H   L   E   T   I   D   K

               550       560       570       580       590       600
541    ATTCCTCGTTCATTCATTCATGACTTCTTGCATATCGTCTATACCAGGAGTATCCATCCG
181        I   P   R   S   F   I   H   D   F   L   H   I   V   Y   T   R   S   I   H   P

               610       620       630       640       650       660
601    CAGGCGAATAAATTGAAACATTACAAAGCATTCAGCAATTATGTTTATGGAGAACTTTTG
201        Q   A   N   K   L   K   H   Y   K   A   F   S   N   Y   V   Y   G   E   L   L

               670       680       690       700       710       720
661    CCCAATTTCCTATCTGATGTATATCAACAATGCCAGTTGAAGAAGGGTGACACTTTCATG
221        P   N   F   L   S   D   V   Y   Q   Q   C   Q   L   K   K   G   D   T   F   M

               730       740       750       760       770       780
721    GATCTCGGTTCGGGAGTAGGTAATTGCGTAGTACAAGCTGCGTTGGAATGTGGATGTGCA
241        D   L   G   S   G   V   G   N   C   V   V   Q   A   A   L   E   C   G   C   A

               790       800       810       820       830       840
781    TTAAGCTTCGGATGTGAAATCATGGATGATGCTAGCGATTTAACTATACTGCAGTACGAG
261        L   S   F   G   C   E   I   M   D   D   A   S   D   L   T   I   L   Q   Y   E

               850       860       870       880       890       900
841    GAACTAAAGAAGAGGTGTAAGTTATATGGGATGCGTTTGAACAACGTGGAGTTTTCATTG
281        E   L   K   K   R   C   K   L   Y   G   M   R   L   N   N   V   E   F   S   L

               910       920       930       940       950       960
901    AAGAAAAGCTTTGTGGACAATAACAGGGTCGCTGAACTAATTCCTCAGTGCGATGTTATC
301        K   K   S   F   V   D   N   N   R   V   A   E   L   I   P   Q   C   D   V   I

               970       980       990       1000      1010      1020
961    CTCGTAAATAATTTTTTTATTTGATGAAGATTTGAATAAAAAAGTCGAAAAGATACTACAA
321        L   V   N   N   F   L   F   D   E   D   L   N   K   K   V   E   K   I   L   Q

               1030      1040      1050      1060      1070      1080
1021   ACGGCAAAAGTTGGATGTAAGATCATAAGTTTGAAAAGTTTAAGAAGCCTCACTTATCAG
341        T   A   K   V   G   C   K   I   I   S   L   K   S   L   R   S   L   T   Y   Q

               1090      1100      1110      1120      1130      1140
1081   ATCAACTTCTACAATGTTGAGAACATCTTCAATAGATTAAAGGTGCAAAGGTATGATCTT
361        I   N   F   Y   N   V   E   N   I   F   N   R   L   K   V   Q   R   Y   D   L

               1150      1160      1170      1180      1190      1200
1141   AAGGAGGATAGTGTTTGCTGGACGCATAGTGGCGGAGAGTATTATATATCAACAGTGATG
381        K   E   D   S   V   C   W   T   H   S   G   G   E   Y   Y   I   S   T   V   M

               1210      1220      1230      1240      1250      1260
1201   GAGGATGTGGACGAAAGTTTATTCAGCCCTGCTGCAAGAGGTAGGAGGAACAGAGGTACG
401        E   D   V   D   E   S   L   F   S   P   A   A   R   G   R   R   N   R   G   T

               1270      1280
1261   CCGGTGAAGTATACCAGATAA
421        P   V   K   Y   T   R   *
```

8.4. Sequencing primers

Sequencing of pUC57 based plasmids:

Forward primer   pUC57 Forward   5'-GTAAAACGACGGCCAGTG

Reverse primer   pUC57 Reverse   5'-GGAAACAGCTATGACCATG

Sequencing of pET3/pET28 based plasmids :

Forward primer   T7 Forward   5'-TAATACGACTCACTATAGGG

Reverse primer   T7 Terminal   5'- GCTAGTTATTGCTCAGCGG

8.5. Primers for site-directed mutagenesis

Site-directed mutagenesis of H3 T80C C110A to H3 R40C T80C C110A

Forward primer   5'-AGTCAAGAAACCTCACTGCTACCGGCCCGGCACAG

Reverse primer   5'-CCGCCGGTAGCAGGAGCGGACTTCCTGGCT

8.6. DNA String® sequences

Cloning of **Met16** promoter DNA

5'-GACTAGCCAT GGTCATAGAT CTGATGGTCT CGTTGCTATG GGTTGATACC TTTGCTGGCC

   TTAGTTTTGA TCTTAAAGCA TTGAAAAGGT TACAAACAGT ACAGTATACT TTTTTTATAT

   CAATATAATA CTATCTCTAT GAGACCATCG GATCCATCAA GCTTGGCGTA A

8.7. Oligonucleotide sequences

Oligonucleotide	convG position	Sequence (5' – 3') G = convertible dG*
a8s convG 70	70	GAC CAG GTG GAT AT**G** GAT
a8s oppo 70	-	ATC CAT ATC CAC CTG
a8s convG 69	69	GAC CAG GTG GAT A**G**T GAT
a8s oppo 69	-	ATC ACT ATC CAC CTG
601 convG 69	69	GAC CGG GAT TG**G** GAT
601 oppo 69	-	ATC CCA ATC CCG
Met16 +	70	TCT ATG AGC TTT TAC TTT GT**G** GAG
Met16 + oppo	-	CTC CAC AAA GTA AAA GCT CA
Met16 -	70	GCA ACT CAT AAA ATG AAG AC**G** TAT
Met16 - oppo	-	ATA CGT CTT CAT TTT ATG AG