Molecular characterization of Mst77F and implication in Drosophila spermatogenesis

Dissertation
for the award of the degree
“Doctor rerum naturalium” (Dr. rer. nat.)
Division of Mathematics and Natural Sciences
of the Georg-August-Universität Göttingen

submitted by
Nils Kost

from Lübben, Germany
Göttingen 2012

Committee members:

Dr. Wolfgang Fischle (1st reviewer), Research group Chromatin Biochemistry
Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Peter Rehling (2nd reviewer), Department of Biochemistry II
Georg-August-University Göttingen

Prof. Dr. Steven Johnson, Department of Molecular Oncology
Georg-August-University Göttingen

Date of the oral examination: August 03, 2012

I affirm that the presented thesis "***Molecular characterization of Mst77F and implication in Drosophila spermatogenesis***" has been written independently and with no other sources and aids than quoted.

June 30, 2012, Göttingen

Acknowledgments

First and foremost, I am grateful to Dr. Wolfgang Fischle, for his trust, constant support and fruitful discussions on several projects.

I would like to thank my PhD Thesis committee members Prof. Peter Rehling and Prof. Steven Johnson for their support and guidance throughout this project.

Many thanks go to all members of the Chromatin Biochemistry group for providing a great working atmosphere. In particular, I would like to thank Claudia Fahlbusch, Lydia Abdelhalim and Winfried Lendeckel for their help in the lab.

I want to thank Prof. Renate Renkawitz-Pohl, Dr. Christina Rathke and Sophie Kaiser for their collaboration on the Mst77F project.

Special thanks goes to Dr. Alf Herzig for providing me with reagents and his help on *Drosophila* methodology

I thank the GGNB administration for the constant support, organization of lectures, method courses as well as retreats.

I am grateful to my parents, Karin and Michael Scholta, for their constant hold up.

Lastly, I want to thank my own family, Daniela and Maxi, for their love, support and cheering me up when it was necessary.

Table of contents

List of figures

List of tables

Abbreviations

°C	degree centigrade
A	absorbance
aa	amino acid
AFM	atomic force microscopy
amp	ampère
APS	Ammoniumperoxodisulfat
bp	basepaire
BSA	bovine serum albumine
C	concentration
cDNA	complementary DNA
CTD	C-terminal domain
Dm	*Drosophila melanogaster*
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotides
DTT	Dithiothreitol
ΔG	Gibb's free energy
ΔH	change in enthalpy (heat energy)
$T\Delta S$	change in entropy
ε	Extinction coefficient
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
et al.	et alteres; et alii
fig.	figure
FP	fluorescence polarisation
FPLC	fast protein liquid chromatography
fwd	forward
g	gram
g	gravitation
h	hour
H1	histone 1; linker histone
H2A	histone 2A
H2B	histone 2B
H3	histone 3

H4	histone 4
HA	hemagglutinin
HMG	high mobility group
HP1	heterochromatin protein 1
ITC	isothermal titration calorimetry
kb	kilobase
kDa	kilodalton
LB	Laura-Bertani-media
m	milli
M	molar
min	minutes
ml	milliliter
mm	milimeter
mM	millimolar
Mst	male specific transcribed
MW	molecular weight
ng	nanogram
nm	nanometer
OD	optical density
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pH	potentium hydrogenii
PMSF	phenylmethanesulphonylfluoride
pp.	paginae
rev	reverse
RNA	ribonucleic acid
rpm	rounds per minute
RT	room temperature
s	seconds
SMART	simple modular architecture research tool
tabl.	table
TBE	TRIS-Borat-EDTA-buffer
TEMED	tetramethylethylendiamin
U	Units
v/v	volume/volume
WT	wild type

w/v	weight/volume
x	times
YT-medium	yeast - tripton media
µ	micro
µg	microgram
µM	micromolar

Amino acids and nucleic acids are shortened according to the international one letter - codes.

1. Introduction

Fertilization of female oocytes requires extraordinary specialized male gametes, the spermatides. In the course of spermatogenesis spermatids undergo a series of morphological as well as molecular rearrangements that are unique and found in no other cell type. In this process the DNA becomes structurally reorganized in a completely distinct manner that avoids the ordinary somatic histone configuration. The inherited program that underlies this development ensures the faithful transmission of the genetic material from one generation to another. Though, the molecular mechanisms that drive differentiation are unclear in many aspects.

1.1 Spermatogenesis

Spermatogenesis, or the development from a stem cell towards a heavily specified, differentiated male gamete takes place in the testis. This organ can be considered a one-dimensional spacio-temporal array of all spermatid developmental stages and has the morphological characteristics of a coiled up, thickened tube. Spermatogenesis follows a gradual differentiation program with dramatic changes in morphology, gene expression and cell cycle dynamics. In *Drosophila* this whole process is manifested by the transformation of a 15 µm diameter round spermatid into a 1.8 mm long needle shaped, motile sperm accompanied by a 200-fold condensation of the genome. About 50% of the genes expressed during this development are testis specific. On the cellular level spermatogenesis is apparently well described in *Drosophila* as well as in mice. However, the molecular mechanisms that drive these processes and in particular the condensation of the genome, are very little understood.

1.1.1 Premeiotic stages and meiosis

The germinal proliferation center is situated at the apical tip of the testis. About eight germ-line stem cells, each associated with two cyst progenitor cells, are situated in close vicinity to a set of twenty somatic cells called “the hub”. Spermatogenesis commences with the simultaneous division of the germ-line- and the cyst cells

resulting in a spermatogonium encapsulated by two cyst cells. The cyst cells stop dividing whereas the spermatogonium undergoes four mitotic divisions followed by the premeiotic S-phase. At this point the cells interconnected by cytoplasmic bridges and still associated with the initial two cyst cells form a cyst of sixteen primary spermatocytes (McKearin, 1997). Cyst maturation is characterized by cell growth, extensive gene expression and storage of translationally repressed transcripts for

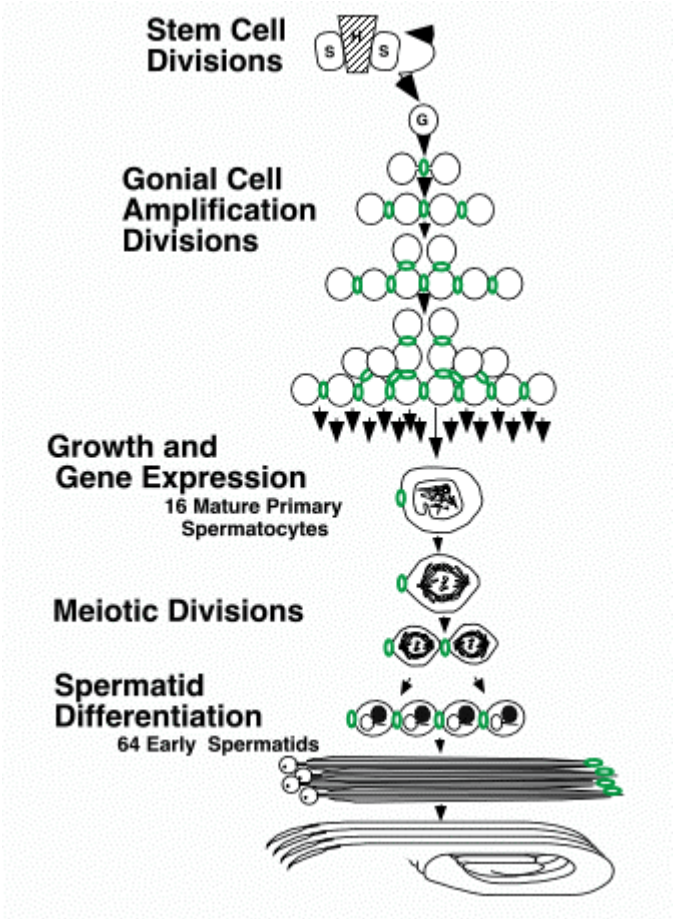


Fig. 1.1 Schematic illustration of Drosophila Spermatogenesis

The gonadial stem cells are situated at the apical tip of the testis in close vicinity to somatic hub cells. The gonial stem cells undergo four mitotic divisions and form a cyst of 16 primary spermatocytes that are interconnected by cytoplamic bridges. Prior to meiosis, as the cyst matures, the cells grow and undergo heavy transcription. After two meiotic divisions a cyst of 64 early spermatocytes enters postmeiotic differentiation. The chromatin is restructured and condenses 200-fold. Concomitantly morphological changes towards needle shape in mature sperm cells are initiated. Finally the previously interconnected spermatids individualize.
Scheme taken and adapted from http://www.sickkids.ca/research/brilllab/sper_pop2.asp.

for later stages of spermatogenesis. Importantly, at the end of this stage bulk transcription is shut down (Gould-Somero, 1974; Olivieri and Olivieri, 1965). The mature cyst enters the first meiotic divisions resulting in a secondary spermatocyte cyst of 32 diploid cells still interconnected and encapsulated by the two cyst cells. After a short interphase the second meiotic division cycle ensues. Eventually, the end product is a cyst of 64 haploid interconnected spermatids surrounded by the two cyst cells (Fig. 1.1).

The premeiotic stages are essentially cellular amplification steps. Nevertheless, the cells are also configured for the postmeiotic differentiation that involves dramatic morphological as well as molecular rearrangements.

1.1.2 Structural organization of the DNA after meiosis

1.1.2.1 The nucleosome core particle - the basic unit of chromatin

As spermatids exit meiosis they carry a single set of chromosomes. Nevertheless, the linear array of genomic DNA exceeds the size of a spermatid nucleus by a multitude. The condensation of the genetic material is achieved through packaging into nucleic acid protein complexes, referred to as chromatin. This macromolecular entity is the basis for all DNA related metabolic processes and constantly subject of regulatory processes.

The simplest unit of chromatin is the nucleosome core particel (Kornberg, 1974). Each core particel consists of 147 bp of DNA that is wrapped around an octameric proteinacious core built up from small basic proteins, the (core) histones, in approximately 1.7 helical turns. Each octamer comprises two copies each of the highly conserved histones H2A, H2B, H3 and H4.

Apart from the globular core that serves as the interaction surface for the DNA, the N- and C-terminal unstructured histone “tails” protrude out of the nucleosome. This is a fundamental feature since mainly the tails are target of posttranslational modification and interaction partners thereby serving as functionalized units in chromatin related processes (Grant, 2001; Hacques et al., 1990).

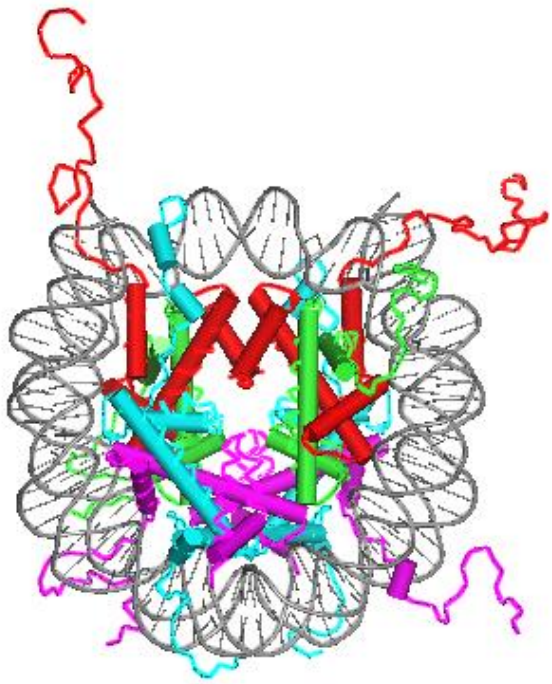


Fig. 1.2 The structure of the nucleosome core particel:

Two copies each of the histones H2A (cyan), H2B (pink), H3 (red) and H4 (green) form the histone octamer. 147 bp DNA are wrapped around the histone octmers. Together they form the nucleosome core particel. Figure generated with the Pymol program from PDB entry 1KX5 (Davey et al., 2002).

1.1.2.2 Histone posttranslational modifications in early differentiating spermatids

Histone posttranslational modifications (PTMs) are set by specific enzymes or enzyme complexes mainly within the unstructured N-terminal tails that protrude out of the globular nucleosome core. Though, an increasing number of modified residues in the globular domain are identified (Kouzarides, 2007). Over the last decade a huge number of different, site-specific modifications has been identified and correlated to specific processes. Up to date the covalent addition of small methyl-, acetyl- and phosphate groups was found to be the most abundant set of modifications but also many other PTMs were detected (Kouzarides, 2007). Histone PTMs either directly alter the structure of chromatin by themselves through changing the physical parameters of the nucleosome complex that effects its interaction with other nucleosomes (“cis” regulation), or the modification functions as a recruiting site for proteins that in turn exert their specific effects in the related process. Functionally, the existence of combinatorial sets of modifications led to the proposition of a histone

code (Jenuwein and Allis, 2001). In this context a differential read out of one modification is generated through intra- and intermolecular crosstalk between modifications.

The early stages of postmeiotic spermatid differentiation are characterized by extensively modified chromatin. Methylation, phosphorylation, acetylation and ubiquitinylation were detected. Whereas the importance of methylated histone H3 (K4, K9 and K27) and H4 (R3) or phosphorylated H4 (S1) is unclear, global H4 hyperacetylation (K5 K8, K12, K16) and H2A ubiquitinylation are thought to exert "cis" effects by opening up the chromatin structure thereby preparing the DNA for subsequent condensation steps (Braun, 2001; Rathke et al., 2007). Apart from the proposed physical alteration of the chromatin fiber by acetylation and ubiquitinylation nothing is known about the contribution of additional modifications and their putative binding partners in the DNA condensation process.

1.1.3 Postmeiotic spermatid maturation

The terminal stages of sperm development encompass the most intriguing differentiations steps of eukaryotic cells. These steps are best studied in mice and *Drosophila* whereas *Drosophila* as a spermatogenesis model system only emerged recently. However, the basic concepts were described being conserved even though the degree of DNA condensation was found to be different. Generally, the final differentiation steps that involve DNA condensation and morphological changes towards needle shaped cells are proposed to fulfill the following three functions: (a) The strong condensation of the DNA accompanied by the morphological rearrangements are suggested to create a hydrodynamic favorable shape that assists sperm motility. (b) The tight condensation of the genome assures protection from mutagenic damage. (c) The removal of histones resets the genome in respect of a functional histone modification status. Upon fertilization this allows *de novo* deposition of maternal histones.

In *Drosophila,* postmeiotic spermatid maturation can be divided into different morphological stages according to the nuclear shape of the cells: round spermatids directly after meiosis II, young elongating spermatids, early canoe stage spermatids, late canoe stage spermatids, spermatids during individualization and mature sperm

cells (Fig. 1.2). The gradual morphological change is accompanied by several molecular rearrangements that affect amongst others the DNA.

1.1.4 Sperm nuclear proteins involved in chromatin condensation

The switch from the nucleosomal towards a highly condensed DNA organization is essential to obtain small hydrodynamic sperm heads that are capable to fertilize the oocyte. The condensation of DNA is thought to be accomplished by two classes of small very basic, nuclear proteins: transition proteins and protamines. Whereas the transition proteins constitute short-term histone replacement factors that are not present in differentiated spermatids, the protamines are the major structural protein in mature sperm cells (Braun, 2001). However, this model is not valid for all species and a coherent pattern is missing. In mammals and flies, transition proteins as well as protamines have been described (Braun, 2001; Sunil Jayaramaiah Raja, 2005). Fishes and birds lack transition proteins and protamines directly condense the DNA. Annelids and echinoderms keep the nucleosomal configuration and abandon spermatid specific genome organizing proteins (Wouters-Tyrou et al., 1998). This inconsistent picture is further confused by the fact that protamines are dispensable for fertility and vitality in some organisms whereas in others they are essential (Rathke et al., 2010). Furthermore, in some species additional proteins are proposed to contribute to condensation processes. In mice and humans the linker histone H1-like protein (Hils1) is suggested to participate in chromatin remodeling during mice and human spermatogenesis.

The functional mechanism of how DNA is condensed by transition proteins and protamines is unknown. The high content of Cystein residues in protamines was suggested to cause intermolecular oxidation that in turn leads to condensation. *In vitro,* DNA protamine complexes have been reconstituted and analyzed by Atomic force Microscopy. The imaged structures resembled stacked doughnuts (Allen et al., 1997; Braun, 2001). However, if these structures are formed in vivo is not known.

1.1.5 The details of spermatid differentiation

1.1.5.1 The condensation of the DNA

As illustrated in Fig. 1.3 the major step in chromatin reorganization takes place in the canoe stage. First, histone proteins are short term replaced by the transition proteins. These proteins in turn are substituted by protamines. In this configuration the *Drosophila* genome displays a 200-fold stronger condensation state (Fuller, 1993). The removal of histone proteins is preceded by global hyperacetylation of histone H4 and ubiquitinylation of H2A and is also conserved across species. For histone acetylation it could be shown that it actually is necessary but not sufficient for chromatin rearrangements (Awe and Renkawitz-Pohl, 2010). Both modifications are thought to open up chromatin structure thereby facilitating histone eviction. The exact mechanism of the eviction process is unclear. However, concomitant with histone removal numerous DNA breaks are detectable and transition proteins start associating with the DNA (Rathke et al., 2007). The current model derived from these data also involves chromatin remodeling complexes that gain access to the DNA upon ubiquitinylation and acetylation. The DNA breaks are suggested to enhance the accessibility so that the remodelers can evict the histone proteins that in turn are replaced by transition proteins as intermediate genome organizers. Apart from the implications of acetylation and ubiquitinylation also other histone posttranslational modifications (phosphorylation, methylation) as well as high levels of SUMO and CTCF proteins are detectable. Their contribution to chromatin restructuring is unknown.

Postmeiotic spermatid differentiation is an only little understood process. The functional integration of cellular events like histone modifications, histone eviction, DNA breaks, transition protein- and protamine deposition is to date hardly possible. The signaling pathways involved in these processes as well as the structural organization of DNA intermediate- as well as highly compacted states (*in vivo*) are unknown.

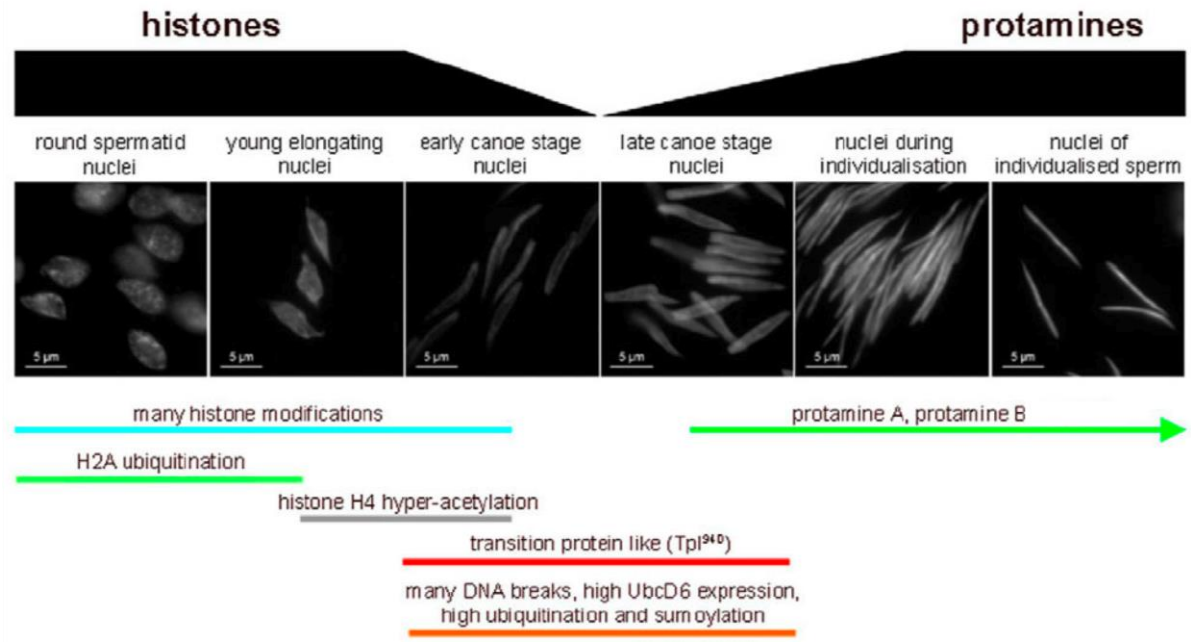


Fig. 1.3 The postmeiotic differentiation of *Drosophila* spermatids

Spermatids leaving meiosis carry round nuclei. Histones are the major structural proteins of chromatin. The histone tails display a differential modification pattern. As cells differentiate d*e novo* H2A ubiquitinylation occurs followed by global H4 hyperacetylation and nuclear shaping commences. With the onset of the canoe stage histones become degraded accompanied by occurrence of DNA breakes, transition protein deposition, high levels of sumo and ubiquitin. At the end of the canoe stage protamines start to replace the transition proteins and nuclear shaping is almost finished. Finally individualized spermatids represent the differentiated state ready for oocyte fertilization. Figure adapted from Rathke et al., JCS 2007.

1.1.5.2 Nuclear shaping (Fuller, 1993)

The nuclear shaping is a microtubule based function. Upon flagellar elongation of the spermatid the nucleus becomes concave on one site and convex on the opposite site. Initially, perinuclear microtubules accumulate at both sites of the nucleus and as the development proceeds microtubules shift to the convex surface and form a laterally aligned bundle. Continuing nuclear transformation gradually condenses the DNA at the inner site of the nuclear envelope right next to the perinuclear microtubules. As DNA condensation proceeds it adapts a net like structure in the nucleoplasm which is stepwise remove from the concave site in form of vesicles. Condensation and nuclear shaping simultaneously continue until the needle shape of the nucleus is obtained.

Importantly, it could be shown that the accumulation of perinuclear microtubules is required for a proper nuclear shaping. However, a possible functional interconnection

of nuclear shaping and DNA condensation is elusive.

1.1.6 Mst77F - a suggested chromosomal architectural protein

The Mst77F gene was discovered by Russel and Kaiser twenty years ago (Russell and Kaiser, 1993). The Mst77F protein is a small, very basic protein that is exclusively expressed in spermatids. Sequence alignments classify the protein as a distant relative of H1/H5 linker histones, major structural proteins of chromatin. Accordingly functions in chromatin compaction were proposed. Expression studies showed that Mst77F mRNA is expressed already at the primary spermatocyte stage, translationally inhibited and stored until translation in young elongating spermatids (Fig. 1.2) (Sunil Jayaramaiah Raja, 2005). At this time histone proteins are still the major genome organizers and Mst77F displays a clear histone colocalization suggesting an association with chromatin (Rathke et al., 2010). The observations that Mst77F stays associated with the DNA in differentiated, mature sperm cells and that *Drosophila* mutants in a protamine$^{-/-}$ background condense chromatin normally and are fertile shed light on Mst77F as a protein of redundant protamine function involved in DNA condensation (Rathke et al., 2010). However, in contrast to protamines the nuclear Mst77F distribution in mature sperm cells is not homogeneous. Additionally, protamine$^{-/-}$ flies are more susceptible to genome damage suggesting at least in part diverging functions of Mst77F and protamines. Additionally, along with the suggested contribution to DNA condensation a second Mst77F function was proposed. This function was implicated by the discovery of a mutant Mst77F allele termed ms(3)nc3 that was identified in a non-complementation screen of β_2-tubulin mutants. This Mst77F mutant is characterized by a single amino acid substitution of serine 149 toward threonine. As already described nuclear shaping is a tubulin dependent process and in the Mst77F ms(3)nc3 background the cells fail to undergo nuclear elongation and display a roundish phenotype (Rathke et al., 2010). Nevertheless, the mutant displays a wild-type like spaciotemporal expression pattern. Proper DNA condensation is observed and Mst77F ms(3)nc3 is also associated with the tightly compacted DNA in differentiated spermatids. A direct link between Mst77F and tubulin was suggested by the observation that in young elongating nuclei Mst77F specifically localizes to the convex site of the nucleus in a parallel alignment with the

perinuclear microtubule bundle. This colocalization along with nuclear shaping defects in the background of the Mst77F ms(3)nc3 mutant led to the hypothesis that Mst77F also is a factor involved in morphological shaping of differentiating spermatids (Rathke et al., 2010). A stabilizing function on β_2-tubulin, the major tubulin isoform in testis, was suggested. This hypothesis was consolidated by reduced β_2-tubulin levels in Mst77F ms(3)nc3 background (Rathke et al., 2010). Alternatively, a coordinating function in positioning β_2-tubulin and chromatin in close vicinity prior to nuclear elongation was proposed. However, no evidence for a simultaneous interaction of Mst77F with chromatin and tubulin could be presented.

In *Drosophila* as well as in mice our current knowledge of the molecular mechanisms that drive spermatogenesis is very limited. Especially the details of histone eviction and the genome condensation by transition proteins and finally protamines are not described at all. The proposed dual functionality of Mst77F in DNA condensation and morphogenesis during spermatid maturation impede the proposition of Mst77F being a classical structural component in this system. However, Mst77F associates with DNA shortly before condensation processes commence. Along with protamines Mst77F is a component of differentiated sperm cells. A strong argument for Mst77F being a DNA condensing protein in *Drosophila* spermatogenesis is the fact that protamine$^{-/-}$ flies show normal DNA condensation and Mst77F to date is the only known protein associated with this DNA condensation state. The distant Mst77F relationship towards H1/H5 linker histone family proteins that are implicated in somatic and spermatid specific condensation processes, makes a related Mst77F function in spermatogenesis conceivable.

1.2 Linker histone H1

1.2.1 H1 localisation and structural function

Apart from the canonical core histones and their variants a fifth different kind of histone, linker histone H1 is associated with the nucleosome. H1 exhibits much less conservation across species and in contrast to the core histones possesses no histone-fold domain. It exhibits a short N- and a long C-terminal basic domain discontinued by a globular domain. The globular domain binds to the nucleosome

where the DNA enters/leaves the histone octamer. Additional contact is made by the C-terminal domain that interacts with the linker DNA in a range of approximately 20 bp aside the nucleosome core particle. Because of its “exterior” association with the nucleosome it shows much higher mobility in comparison to the core histones (Catez et al., 2004; Phair et al., 2004).

In mammals eleven H1 subtypes have been identified including specialized isoforms in testis and ovaries (Happel and Doenecke, 2009). Besides the knowledge about their existence very little is known about functional differences on the molecular level. In yeast, *C.elegans* and some other model organsisms they have been proven to be nonessential, but defects in development and life span were observed (Jedrusik and Schulze, 2001; Patterton et al., 1998). However, recent work in higher eukaryots such as mice and *Drosophila* indicate their necessity for viability (Fan et al., 2005; Lu et al., 2009b).

The exact mechanism how H1 proteins exert their functions is not resolved. However, much data suggest its involvement in regulation of chromatin and chromosome structure. In *Tetrahymena thermophila* deletion of the somatic H1 gene resulted in increased nuclear volume, suggesting chromatin condensation defects (Shen et al., 1995). Likewise, similar effects were observed in embryonic stem cells derived from H1 knockout mice (Fan et al., 2005). Moreover, H1 was also reported to be required for metaphase chromosome structure in *X.leavis.* Depletion led to aberrant morphology and segregation defects *(Maresca et al., 2005).*

However, our today’s understanding of H1 related chromatin condensation processes is limited. Recent work set out to investigate the molecular details which enable H1 proteins to achieve their proposed functions that are mainly attributed to the C-terminal domain.

1.2.2 H1 structure and function - the globular domain

The structure of the globular domains of linker histones H1 and isoform H5 have been resolved two decades ago by x-ray crystallography and NMR studies (Cerf et al., 1993) (Cerf et al., 1993). These domains consist of a three-helix fold for which the term “winged-helix” has been coined. This fold adapts the classical helix-turn-helix motif. In comparison to the N- and C-terminal domains of H1 the “winged-helix” is

well conserved across species and necessary to generate chromatosome stops in nuclease digestion experiments. Additionally, it could be shown that the isolated globular domain is sufficient for nucleosome binding *(Allan et al., 1980)*. The exact interaction position of the globular domain on the nucleosome is controversial. However, generally accepted is its contact with two strands of DNA at the nucleosome where the DNA enters/exits. This position is implicated to stabilize the wrapping of the DNA around the histone octamer (Brown et al., 2006; Crane-Robinson, 1997; Fan and Roberts, 2006; Syed et al., 2010).

1.2.3 H1 structure and function - the N-terminal domain

Depending on the isoform the N-terminal domain is 20 to 35 amino acids in length. Sequence analysis revealed a bipartite amino acid distribution with a high content of basic amino acids proximal the “winged-helix” globular domain (Bohm and Mitchell, 1985). This part of the N-terminal domain was suggested to contribute to binding stability of H1 (Allan et al., 1986; Vila et al., 2001). In aqueous solutions the N-terminal domain shows an unstructured coiled coil conformation that under stabilizing conditions adapts a α-helical conformation. In light of the emerging field of intrinsically unstructured proteins a folding upon interaction with DNA was proposed (Vila et al., 2001; Vila et al., 2002).

1.2.4 H1 structure and function - the C-terminal domain

The C-terminal domain (CTD) is approximately 100 amino acids in length and displays sequence variation between isoforms and across species (Ponte et al., 2003). A striking feature is the high content of basic residues that turn the C-terminal domain into a highly charged moiety with 30 to 50 evenly distributed net positive charges (Subirana, 1990). The sequence content of the CTD resembles that of intrinsically unstructured proteins (Hansen et al., 2006; Lu et al., 2009a).

1.2.4.1 Intrinsically unstructured proteins (IUPs)

“Structure determines function” is the classical definition of functional proteins. In the

recent years more and more proteins were discovered that deviate from this classical definition. Intrinsic unstructured proteins - that is the hindrance of spontaneous folding into well-organized structures in the absence of stabilizing interactions - emerge in the field of protein science. Even though these characteristics were put forward recently, they were originally discovered more than 25 years ago (Mitchell and Tjian, 1989; O'Hare and Williams, 1992). In contrast to the prediction of 3D structures of globular proteins that to date remains a key challenge, the identification of sequences that adapt unstructured characteristics is straightforward. The fingerprint of unstructured domains is their low sequence complexity. This often involves repetitive or periodic elements of a limited set of amino acids and a

Tabl. 1.1 Functional implications of intrinsically unstructured domains

Listed are the different function of intrinsically unstructured domains and the number of proteins found for each function. A common function of intrinsically unstructured domains is their involvement in protein protein interactions. However, multiple examples were also found for their role in DNA binding events. Posttranslational modifications in intrinsically unstructured domains might reflect regulatory mechanisms that modulate the function. Table modified from Duner et al., Biochemistry 2002.

function	no. of examples
protein–protein binding	54
protein–DNA binding	19
protein–rRNA binding	5
protein–tRNA binding	1
protein–mRNA binding	1
protein–genomic RNA binding	3
protein–lipid interaction	6
polymerization	4
substrate/ligand binding	6
cofactor/heme binding	1
metal binding	9
autoregulatory	7
regulation of proteolysis in vivo	7
acetylation	4
fatty acylation (myristolation and palmitoylation)	4
glycosylation	3
methylation	1
phosphorylation	16
ADP-ribosylation	1
flexible linkers/spacers	7
entropic spring	2
entropic bristle	1
entropic clock	1
structural mortar	>10
self-transport through channel	3
DNA unwinding	1
DNA bending	1
protein detergent	3
disordered region is not essential for protein function	6
unknown	16

compositional bias towards a low content of bulky hydrophobic amino acids and a high proportion of polar/charged amino acids (Obradovic et al., 2003; Romero et al., 2001).

The molecular function of IUPs is versatile (Tabl. 1.1). The region of disorder can manifest in flexible linkers that connect globular domains within one polypeptide chain. This allows for rotational and conformational flexibility in search for binding partners and additionally allows proteins that bind the linker to induced interdomain conformational changes. Another important aspect of intrinsically unstructured domains is the coupled binding and folding mechanism upon recognition of interaction partners. The folding occurs locally but also can concern entire domains (Gunasekaran et al., 2003). Moreover, intrinsically unstructured domains are regions that are targeted by regulatory mechanisms. Binding of a small molecule or a posttranslational modification alters the inducible structures and this correlates with functional diversity (Sandhu and Dash, 2007).

The growing field of intrinsic unstructured proteins illustrates the surprisingly manifold aspect of protein structure and function that is not appreciated by protein crystallography approaches. Even though only a fractional amount of intrinsically unstructured proteins has been discovered, their vital role across cellular functions already becomes clear.

1.2.4.2 Functional implication of the intrinsically unstructured histone H1 CTD

Recently, spectroscopic experiments could show that in the presence of secondary structure stabilizing reagents the H1 CTD acquires a substantial α-helical conformation (Verdaguer et al., 1993). Moreover, the addition of DNA under physiological conditions also induces structures in the CTD (Roque et al., 2005). Further experiments identified two regions within the long H1 CTD that are primarily responsible for alteration of the linker DNA structure and chromatin condensation (Lu and Hansen, 2004). It could be shown that these subdomain adapt α-helical structures upon molecular recognition of DNA (Lu et al., 2009a). Furthermore, posttranslational modifications in the H1 C-terminal domain are proposed to “switch” between distinct H1 functions (Happel and Doenecke, 2009). Roque and coworkers

could show in 2008 that cyclin-dependent kinase-mediated phosphorylation of specific sites in the C-terminal domain cause a decrease in the proportion of α-helix and an increase in the β-sheet content (Roque et al., 2008). This effect depended on the amount of introduced phosphate groups and interestingly the fully phosphorylated CTD had a higher chromatin aggregation potential.

1.4 Objectives of the thesis

DNA condensation and structural reorganization is a conserved mechanism in the process of spermatogenesis. The exact molecular mechanisms and determinants underlying this process, such as histone displacement, degradation, incorporation of transition proteins and protamines together with the role of linker histone like proteins are poorly understood.

In *Drosophila* the Mst77F protein is put forward as an additional DNA condensing protein. The cellular spaciotemporal expression pattern has been described whereas the molecular function is unknown. In this work I set out to characterize Mst77F on a molecular level by biochemical and biophysical methods. I wanted to answer the question if Mst77F is a protein involved in DNA condensation during Drosphila spermatogenesis and by what mechanism it conducts its function. To obtain this goal the following question were addressed:

1. Mst77F colocalizes with DNA in vivo but does it directly interact with DNA?
2. Which part of the protein binds DNA?
3. What is the interaction mode? Is it similar to linker histone H1 proteins, the proposed homologs?
4. Does Mst77F introduce structural changes in DNA?
5. How does Mst77F alter DNA structure?
6. Mst77F colocalizes with DNA when histones are still the structural organizers. In this respect, is Mst77F sufficient to evict histones from the DNA?
7. How inert is the DNA Mst77F complex? Does Mst77F serve as a general transcriptional inhibitor?

2 Material and Methods

2.1 Laboratory equipment

The laboratory equipment used during the course of this study is summarized in table 2.1.

Tabl. 2.1: Laboratory equipment used for the experiments

Name of Equipment	Manufacturer
Aekta Systems (Prime / Purifier and Exlorer)	GE Healthcare
balances	Mettler - Toledo
CD spectrometer ChiraScan	Applied Photophysics
centrifuges • 5424 / 5415R / 5810R • analytical ultracentrifuge XL-A • Sorvall Evolution RC • Sorvall Discovery 90SE • table top mini	 Eppendorf Beckmann ThermoScientific ThermoScientific Roth
freezer • -20ºC • -80ºC	 Liebherr ThermoScientific
french press EmulsiFlex - C5	Avestin
Hereaus Kelvitron®t Incubator	ThermoScientific
iTC 200	GE Healthcare
laserscanner FLA-5100	FUJIFILM
Mini – Protean Cells	Biorad
Multitron Shaker	HT Infors
Nanodrop ND – 100	ThermoScientific
Nanoscope V Multi Mode AFM	Bruker

peristaltic pump	Ismatec
pH – meter	Mettler - Toledo
plate reader Chameleon V	Hidex
PowerPac universal power supply	Biorad
Perfection V750 Pro Scanner	Epson
PhosphoImager Typhoon 8600	Molecular Dynamics
pipettors	Eppendorf
rotator Roto Shake Gene	Scientific Industries
speed vac Savant SPD131DDA	ThermoScientific
Stuart Gyrorocker SSL3	Sigma
Sub - Cell – GT Agarose Gel Electrophoresis	Biorad
thermocycler Epgradient S	Eppendorf
thermomixer Comfort	Eppendorf
UV Transilluminator	Biorad
vortex Genie 2	Scientific industries
waterbath	Julabo

2.2 Chemicals / Reagents

All chemicals/reagents used during the course of this study are summarized in table 2.2.

Tabl. 2.2: Chemicals / reagents used for the experiments

Chemical / Reagent	Manufacturer / Vendor
acetic acid	Merck
acetyl – CoA	Sigme - Aldrich
acrylamide 30%	Roth
acrylamide / bisacrylamide (37,5 : 1)	Merck
agar	Roth
agarose	Serva
alkaline phosphatase	ThermoScientific
albumin, bovine	Sigma - Aldrich
Amberlite MB3 resin	Merck
ammonium chloride	Merck
ammonium peroxisulfate	AppliChem
ampicillin	AppliChem
bacto - tryptone	Roth
bisacrylamide 2%	Roth
bis[sulfosuccinimidyl]suberate	ThermoScientific
boric acid	Merck
bromophenol blue	Sigma - Aldrich
cesium chloride	Merck
chloramphenicol	Sigma - Aldrich

chloroform : phenol : isoamylalcohol	Sigma - Aldrich
coomassie brilliant blue G-250	Biorad
dimethylsulfoxid	Sigm - Aldrich
dithiothreitol	AppliChem
DNA Ladder 1kb	LifeTechnology
ethanol	Merck
ethidium bromide	Roth
ethylenediaminetetraacetate	Roth
glucose	Merck
glutaraldehyde	Electron Microscopy Sciences
glycerol	Merck
guanidin hydrochloride	Sigma - Aldrich
HisPur™ Cobalt resin	ThermoScietific
hydrochloric acid	Merck
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	VWR
imidazol	Roth
isopropyl-β-D-thiogalactopyranosid	Sigma - aldrich
magnesium chloride	Merck
magnesium sulfate	Merck
2-mercaptoethanol	Sigma - Aldrich
Micrococcal nuclease (MNase)	Sigma-aldrich
milk powder	Regilait
nonidet P-40	Roche
NucleoSpin® extract plasmid purifcation kit	Macherey - Nagel
oligonucleotides	Sigma

peptone	Roth
Pfu DNA Polymerase (native)	ThermoScientific
phenylmethylsulfonylfluorid	Serva
polyethylenglycol 6000	Merck
potassium chloride	Merck
protein ladder Seablue®plus 2	LifeTechnology
QiaEX II gel extraction kit	Qiagen
radionucleotides	Perkin - Elmar
reaction tubes 1.5 ml low binding properties	Nerbe
restriction endonucleases	New England Biolabs / ThermoScientific
RNase T1	Roche
sodium acetate	Merck
sodium chloride	Merck
sodium dihydrogen phosphate	Merck
sodium dodecyl sulfate	Roth
sodium fluoride	Merck
sodium hydroxide	Merck
sodium monohydrogen phosphate	Merck
streptavidin coated paramagnetic particels	Promega
streptavidin coated 96 well plates	ThermoScientific
T4 DNA Ligase	ThermoScientific
tetramethylethylendiamine	Sigma - Aldrich
triethanolamine	VWR
tris(hydroxymethyl)aminoethane	Roth

triton X-100	Merck
urea	Merck
well plates 96 format (transparent)	Greiner
well plates 384 format (black)	Corning
yeast extract	Mobio

2.3 Bioinformatic tools

Protein structural domain prediction was carried out using the **S**imple **M**odular **A**rchitecture **R**easearch **T**ool (**SMART**) algorithm (http://smart.embl-heidelberg.de/)(Schultz et al., 1998).

The relative protein disorder was calculated by the **I**ntrinsic **U**nstructured **P**rotein **re**lative **d**isorder (**IUPred**) algorithm (http://iupred.enzim.hu/) on the basis of the primary structure by estimation of the pairwise energy content between amino acids. The algorithm predicts the structural properties of a given protein by identifying the number of favorable interactions all amino acids can form (Dosztanyi et al., 2005a, b).

2.4 Preparation of SDS-PAGE Gels and Electrophoresis

SDS-PAGE was performed using standard protocols (Gallagher, 2006). The resolving gel contained 0.4 M TrisHCl pH = 8.8, 0.1% (w/v) SDS, 15% acrylamide-bisacrylamide (37.5:1), 0.1% APS, 0.04% TEMED and the stacking gel was composed of 0.68 M TrisHCl pH = 6.8, 0.1% (w/v) SDS, 5% acrylamide-bisacrylamide (37.5:1), 0.1% APS, 0.1% TEMED. Casting and running of the gels was performed with the Mini-Protean System. The sample loading buffer contained 62.5 mM TrisHCl pH = 6.8, 8.5% (v/v) glycerol, 2 % (w/v) SDS, 4 mM DTT and 0.05% bromophenol blue. The protein samples were boiled in loading buffer for 5 min at 95 ºC. The gel running buffer contained 25 mM Tris, 200 mM glycine and 0.1 % (w/v) SDS. Electrophoresis was performed at a constant current of 25 mA until the bromophenol blue migrated out of the gel. SeeBlue Plus 2 marker was used as a size

standard.

2.5 Protein staining within SDS Gels

For protein visualization the SDS-PAGE gels were soaked with coomassie G-250 protein staining solution (Meyer, 1965) at RT for 15 min and constant agitation (0.05 % Coomassie R-250, 10% (v/v) acetic acid, 50% (v/v) methanol). The staining solution was replaced by destaining solution (10 % (v/v) acetic acid, 7.5 % (v/v) methanol) until the desired contrast was reached.

2.6 Determination of Nucleic Acid and Protein Concentrations

Nucleic acid concentrations were determined with the NanoDrop Spectrophotometer (ThermoScientific) using the DNA/RNA program implemented in the NanoDrop software. In order to determine protein concentrations the OD of the samples was measured with the UV-Vis program of the NanoDrop software. Extinction coefficients and theoretical molecular weights of proteins were calculated on the basis of the primary proteins sequences using the ProtParam online tool (Gasteiger E., 2005). Protein concentrations were calculated using the Lambert - Beer equation:

$$C = \frac{A}{\varepsilon * d}$$

with c being the concentration, A the measured absorbence, ε the molar extinction coefficient and d the pathlenght of the cuvette. Table 2.3 summarizes the molecular weights and the extinction coefficients of the proteins analyzed.

Tabl. 2.3: Molecular weights and extinction coefficients of the proteins analyzed

Protein	Extinction coefficient (ε)	MW (Da)
Mst77F WT	17420	26421
Mst77F S149T	17420	26435
Mst 77F shuffled C	17420	26421

Mst77F ∆20C	17420	24222
Mst77F ∆40C	17420	21973
Mst77F ∆60C	17420	19748
Mst77F ∆110C	17420	14338
Mst77F ∆100N	2980	16589
hH1.4	2980	22529
xPr-Set7	27850	42246
xH2A	4050	13960
xH2B	6070	13774
xH3	4040	15273
xH4	5040	11236

2.7 Molecular Cloning

Tabl. 2.4: Bacterial cells that were used throughout this study

Strain	Genotyp	Application	Vendor
DH5α	*E. coli F- φ80lacZΔM15 Δ(lacZYAargF)U169 deoR recA1 endA1 hsdR17 (rk-, mk+) phoA supE44 λ- thi-1 gyrA96 relA1*	molecular cloning DNA production	LifeTechnologies
BL21-Codon Plus (DE3)-RIL	*E. coli B F– ompT hsdS(rB– mB–) dcm+ Tetr gal endA Hte [argU ileY leuW Camr*	protein expression	Agilent

Tabl. 2.5: Formulations of the bacterial growth media used

Type	Formulation	Application
LB	1% w/v bacto-tryptone, 0.5% w/v yeast extract, 1% w/v NaCl	cloning, DNA production
2xYT	1.6% w/v bacto-tryptone, 1% w/v yeast extract, 0.5% w/v NaCl	protein expression
SOC	2% w/v bacto-tryptone, 0.5% w/v yeast extract, 10 mM NaCl 2.5 mM KCl 10 mM $MgCl_2$ 10 mM $MgSO_4$	cloning

2.7.1 Bacterial transformation

Transformation of plamids (Griffith, 1928) was carried out with chemically competent *E.coli* cells throughout this work. The bacterial cells (table 2.4), were prepared according to Sambrook and Russell (Sambrook, 2001). The cells were thawed on ice and 10 ng DNA or 3 µL of a ligation reaction of were added. The DNA was allowed to attach to the bacterial membrane for 30 min on ice and transformation was induced by heat-shock for 30 s at 42ºC in a water bath. After 2 min incubation on ice, 250 µL of SOC medium were added. The bacteria were incubated in a thermomixer at 37ºC for 1 h and shaking at 1000 rpm to induce resistance. The cells were plated on agar-plates containing the respective antibiotic. Incubation of the plates was carried out over night at 37ºC.

2.7.2 PCR based DNA amplification

Standard polymerase chain reactions (Mullis et al., 1986) were carried out in a 50 µL reaction volume with 0.2 µM of forward and reverse primer, 0.2 µM dNTP's, 1 U Pfu native DNA Polymerase, 1x Pfu buffer and 10 – 100 ng template DNA. PCR reactions were performed in an Eppendorf MasterCycler Epgradient S using the protocol in table 2.6.

Tabl. 2.6: Typical PCR program used for Pfu – Polymerase (native)

Step	Temperature (°C)	Time (s)	cycles
1	95	180	1
2	95	30	35
3	50-65	30	
4	72	120/kb	
5	72	120/kb	1
6	10	∝	-

2.7.3 Purification of PCR products

The PCR products were purified using the NucleoSpin® PCR and gel extraction kit according to the manufacturers protocol.

2.7.4 Restriction digest of DNA

PCR derived DNAs and plasmid DNAs were digested using New England Biolabs Restriction Endonucleases (Hartl, 2001) according to the manufacturers recommendations. In brief, 1 ug of Plasmid DNA was digested with 5 Weiss Units each of the appropriate set of Restriction Endonucleases for 1 h at their respective temperature. PCR products were digested according to the plasmid digestion parameters but over night. The restriction reaction was terminated by thermal inactivation in all possible cases.

2.7.5 Dephosphorylation of Plasmid DNA

Restriction endonuclease treated plasmid DNA was dephosphorylated with Fast Alkaline Phosphatase in the respective restriction buffer for 15 min at 37°C followed by thermal inactivation at 75°C for 5 min.

2.7.6 Purification of DNAs

For molecular cloning reactions DNAs subjected to restriction endonucleases were purified by agarose gel electrophoresis and subsequent extration from the gel with the QiaEX II gel extraction kit according to the manufacturers recommendations.

2.7.7 Agarose electrophoresis & ethidium bromid visualization of DNA

Typically, 0.5 % to 2 % (w/v) agarose was dissolved by boiling in TBE (90 mM TrisHCl pH = 8.0, 90 mM Sodiumborate, 2 mM EDTA) and gels were prepared using a Sub-Cell GT electrophoresis system. Prior to loading onto the gel the DNA samples were mixed with 10x loading dye (TBE with 30 % (v/v) glycerol and 0.05% (w/v) bromphenolblue). Electrophoresis was carried out at a constant voltage of 80 V in 1x TBE buffer. As size standard served the 1kb Plus DNA ladder. The DNA within the gel was stained with ethidiumbromide solution (0.5 µg / ml ethidium bromide in 1X TBE - buffer) for 30 min followed by visualization on a gel - doc system in the Department of Cellular Biochemistry at the Max – Planck Institute for Biophysical Chemistry, Goettingen, Germany.
For native agarose gel electrophoresis 0.5% to 1.5% agarose gels (0.5% - 1.5% w/v agarose, 18 mM Tris, 18 mM boric acid, pH 8.0) were run in 0.2X TB buffer (18 mM Tris, 18 mM boric acid, pH 8.0) at a constant voltage of 80 V for 2 hrs at 4°C. Staining and visualization was carried out as discribed above.

2.7.8 Ligation

Ligation of PCR products with the destination vector was carried out using 30 ng of plasmid vector and a three fold molar excess of the insert DNA in a 20 µL reaction volume. 1 Unit of T4 DNA Ligase catalyzed the reaction for 1 h at 22°C. The ligation reaction was transformation into *E.coli* DH5α according to paragraph 2.7.1.

2.8 Mini - Preparation of Plasmid DNA

The preparation of plasmid DNA from E.coli was carried out according to the principle of alkaline lysis (Birnboim and Doly, 1979). The work was conducted with the NucleoSpin extract kit.

2.9 Site - Directed - Mutagenesis of Plasmid DNA

PCR based site directed plasmid mutagenesis was performed as described in the QuickChange Site Directed Mutagenesis Kit from Agilent. Primers were designed containing the desired mutation, flanked by 22 ± 2 nucleotides in each direction, preferentially ending with a C or G. The PCR reaction was followed by a DpnI restriction digest (1 h, 37°C) to eliminate the methylated parental plasmid. After transformation into *E.coli* DH5α and plasmid preparation the clones were tested by sequencing to validate the mutation.

Tabl. 2.7: Primers used for the mutagenesis of Mst77F serin149 into threonine

Primer Name	Sequence – mutated site indicated by small highlighted letter
Mst77F S149T fwd	5’ - CCCTCGTAAAGAGAACAAATGT**a**CGAAACCTCGTGTC CGTAAAAGTTG - 3’
Mst77F S149T rev	5’ - CAACTTTTACGGACACGAGGTTTCG**t**ACATTTGTTCTC TTTACGAGGG - 3’

Tabl. 2.8: PCR program used for the mutagenesis of pET3a Mst77F S149T

Step	Temperature (°C)	Time (s)	cycles
1	95	210	1
2	95	30	

3	65	30	18
4	72	720	
5	72	720	1
6	10	∝	-

2.10 Expression and Purification of Recombinant Proteins

2.10.1 Bacterial Expression

A single colony of *E.coli* BL21 (DE3) RIL cells (table 2.4) transformed with the plasmid carrying the protein encoding sequence was used to inocculate a 50 mL pre-expression culture of LB medium containing the respective antibiotic. The pre-culture was grown over night at 37°C and 130 rpm. The next morning the pre-culture was diluted 1:100 in prewarmed 2xYT medium (table 2.5) and grown to an OD of 0.6 at 37°C and 130 rpm. Protein expression was induced by the addition of IPTG to a final concentration of 0.3 mM. The protein production was allowed to go on for 1.5 h followed by centrifugation of the *E.coli* cells in a Sorvall Evolution centrifuge using a SLC-1500 rotor (10 min, 6000 x g). Cells were subsequently either directly lysed for protein purification or shock frozen with liquid nitrogen and stored at -80°C until the proteins were to be purified.

2.10.2 HIS-tag protein purification

Bacterial pellets from *E.coli* BL21 (DE3) RIL cultures (see previous paragraph) were thoroughly resuspended in HIS - lysis/wash buffer (20 mM Hepes pH = 7.4, 1 M NaCl, 40 mM Imidazole, 1 mM β-Mercaptoethanol, 1 mM PMSF) and lysed by three passages through an EmulsiFlex - C5 cell disrupter (Avestin) at 4°C using a pressure setting of 100 - 150 bar. Insoluble material was removed by centrifugation for 25 min at 20000 x g. The supernatant was either loaded onto a 1 mL Ni-NTA column using the Äkta FPLC system or purified on a 1 mL Cobalt resin by gravity flow. In either case the column material was washed with at least 100mL of HIS - lysis/wash buffer. The bound proteins were eluted with HIS – elution buffer (20 mM Hepes (NaOH) pH = 7.4, 1 M NaCl, 250 mM Imidazole, 1 mM β-Mercaptoethanol, 1 mM PMSF) in 15 mL. Directly after elution all proteins

were extensivly dialysed against storage buffer (50 mM Imidazol, 300 mM NaCl, 5 mM β-Mercaptoethanol, 10% Glycerol (v/v), pH = 6.4 adjusted with 3 M NaAc pH = 5.2 (final NaAc 40 mM)), concentrated to 20 – 100 µM, aliquoted, snap frozen in liquid nitrogen and finally stored at -80°C until usage.

2.10.3 Histone Inclusion Body Purification

The purification of histone proteins under denaturing conditions was largely performed as documented (Luger et al., 1999), with some modifications that were introduced in the Tsukiyama Lab by Marnie E. Gelbart. Plasmids encoding the *X. laevis* histone proteins H2A (GeneBank: CAD89676), H2B (GeneBank: CAD89678), H3 (GeneBank: CAD89679), and H4 (GeneBank: CAD89677) were obtained from Karolin Luger, Colorado State University, Dept. of Biochemistry and Molecular Biology. 2 L of *E.coli* BL21 (DE3) RIL Bacteria were grown as described in paragraph 2.10.1 with a protein expression period of 4 h and lysed in Wash Buffer (50 mM TrisCl pH = 7.5, 100 mM NaCl, 1 mM EDTA, 2 mM DTT, 1 mM DTT) by three passages through an EmulsiFlex - C5 cell disrupter at 4°C using a pressure setting of 100 – 150 bar. Inclusion bodies were pelleted by centrifugation for 20 min at a speed of 20000 x g. The pellet was washed two times with TW - Buffer (Wash Buffer with 1 % (v/v) Triton X-100) and subsequently two times with Wash Buffer. To solubilize the histones from the inclusion body pellets, 350 µL DMSO was added to the pellet. The sample was stirred thoroughly with a spatula and incubated for 30 min at RT. Then, 13.3 mL unfolding buffer (7 M guanidine - HCl, 20 mM TrisHCl pH = 7.5, 10 mM DTT) were added, the sample was mixed and the suspension was rotated for 1 h at RT. Insoluble material was removed by centrifugation for 15 min at 25000 x g and 4°C. The supernatant was dialysed 3 times against 2 L each of Urea Dialysis Buffer (7 M Urea, 1 mM EDTA, 10 mM TrisHCl pH = 7.5, 100 mM NaCl, 2 mM DTT, 0.2 mM PMSF) with one dialysis step over night.

2.10.4 Chromatographic Purification of Histones

The denatured histone proteins were purified by ion exchange chromatography on a XK26/20 Q sepharose and a XK26/20 SP sepharose column tandem setup. The columns were equilibrated in 90 % Urea Buffer A (7 M Urea, 10 mM TrisHCl pH =

7.5, 1 mM EDTA, 2 mM DTT, 0.2 mMPMSF) and 10 % Urea Buffer B (Urea Buffer A with 1 M NaCl). The dialyzed protein was cleared from precipitate by centrifugation at 20000 x g for 10 min and 4°C. The sample was loaded at a constant flow rate of 2 mL/min and both columns were washed with 10 column volumes buffer. The Q sepharose column with the bound DNA and contaminating proteins was removed. The elution was carried out with a isocratic gradient from 10 – 100 % Urea Buffer B over 10 column volumes. 5 mL fractions were collected and analyzed by SDS-PAGE, pooled and dialyzed 3 times against 2 L each ddH2O for 48 h in total. For storage, the histones were aliquoted, lyophilized and stored at -80°C until usage.

2.11 In vitro chromatin reconstitution & analysis

2.11.1 Preparation of DNA templates for chromatin reconstitution

The pG5ML plasmid used in *in vitro* transcription assays was obtained from (An, 2004) and prepared in quantities by using a Plasmid Giga Kit according to the manufacturers recommendations. The super coiled form of the plasmids was isolated by CsCl gradient centrifugation (Sambrook, 2001).

For the reconstitution of mono- and oligonucleosomal complexes the non-natural nucleosome positioning sequence “601” was used in order to generate high affinity binding sites for the histone octamers (Lowary and Widom, 1998). The 12 x 200 x 601 template was generated from a pUC18 plasmid containing a 12 x 200 x 601 insert. This plasmid was kindly provided by Daniela Rhodes, MRC Laboratory of Molecular Biology, Cambridge, UK. The DNA was prepared in quantities using the Plasmid Giga Kit (Qiagen) according to the manufacturers recommendations. 10 mg of the pUC18 12 x 200 x 601 plasmid were applied for the enzymatic digestion with the restriction endonucleases DdeI, HaeII, BfuCI, and EcoRI (New England Biolabs) over night. The obtained 2437 bp 12 x 200 x 601 repeat sequence and small fragments which correspond to the pUC18 backbone were separated by sequential PEG precipitation as described previously (Lis and Schleif, 1975) with a final PEG concentration of 8%.

The 147 bp 601 DNA sequence without linker DNA was used for mononucleosome reconstitutions. It was generated by PCR applying a monomeric 187 bp 601

sequence as DNA template and was kindly provided by Dr. Alexandra Stuetzer, Max - Planck Institute for Biophysical Chemistry, Goettingen, Germany.

2.11.2 Histone octamer reconstitution

The reconstitution of histone octamers was performed according to (Luger et al., 1999). In brief, the lyophilized histones were dissolved in unfolding buffer (see paragraph 2.10.3) and mixed in equimolar amounts according to their UV absorbance spectra (paragraph 2.6 and table 2.3). After dialysis 3 times against 2 L each of RB high buffer (10 mM TrisHCl pH = 7.5, 1 mM EDTA, 2 M NaCl, 1 mM DTT), the sample was concentrated to a final volume of 1 mL and subjected to size exclusion chromatography.

2.11.3 Size exclusion chromatographie of histone octamers

The histone octamers reconstituted by salt dialysis (previous paragraph 2.11.2) were separated from H2A/H2B dimers on a HR16/60 Superdex 200 column (GE-Healthcare) in RB high buffer and a constant flow of 1 mL/min. Peak fractions were collected in 1 mL volumes and analysed by SDS – PAGE. The fractions containing all four histone proteins in stoichometric ratios were pooled and concentrated. The histone octamers were either used immediately, stored for short term at 4°C, or were diluted 1:1 with 100 % glycerol and stored for long term use at -20°C.

2.11.4 Reconstitution of mono and oligonucleosomes

In order to assemble regularly spaced nucleosomes on DNA, a method of continuous dialysis was utilized (Luger et al., 1999). The concentration of the histone octamer was determined photometrically assuming that an OD276 = 0.45 corresponds to 1 mg/mL of histone octamer. For a test assembly typically 50 µg of DNA were used. The reconstituted octamers (paragraph 2.11.5) were mixed with the respective DNA in a 0.6 : 1 to 1.3 : 1 with 0.1 increments of octamer/DNA molar ratio for mononucleosomal DNA and 0.8 : 1 to 1.2 : 1 molar ratio for 12 x 200 x 601 oligonucleosomal arrays in RB high buffer. The dialysis vessels were placed in 400 mL RB high buffer and the buffer was slowly exchanged against 2 L RB low buffer

(RB high but 10 mM NaCl) over a period of 36 h using a peristaltic pump. The samples were subsequently dialyzed against TEA 20 storage buffer (10 mM Triethanolamine-HCl pH = 7.5, 20 mM NaCl, 0.1 mM EDTA), analyzed by native agarose gel electrophoresis and kept at 4°C. Typically, a ratio of 0.9 to 1 resulted in saturated but not aggregated material. The ratio that was determined in the pre-assembly was then used to produce larger quantities.

2.11.5 In vitro chromatin transcription assay

The chromatin *in vitro* transcription assay is based on a protocol that was developed in the lab of Dr. James Manley, Columbia University, New York, USA (Hirose and Manley, 1998). The method was adapted by Dr. Adrian Schomburg within the Lab of Chromatin Biochemistry, Max-Planck Institute for Biophysical Chemistry, Goettingen, Germany in order to apply *in vitro* assembled pG5ML plasmid chromatin as a transcriptional template.

To evaluate the effects of Mst77F on the transcription process the pG5ML plasmid (An, 2004) was either used in it's "naked" form (plasmid only) or was assembled into chromatin. 1 µL/reaction of the "naked" pG5ML or pG5ML chromatin (0.1 µg/µL) was diluted in 9 µL/reaction transcription mix (2 mM MgCl2, 2 mM spermidine, 20 mM creatinphosphate, 3 % (w/v) polyvinylalcohol, 500 µM ATP, 500 µM UTP, 500 µM GTP, 20 µM CTP, 80 nM TSA, 10 mM sodiumbutyrate, 20 ng Gal4-VP16). In the case of the "naked" plasmid 1µL/reaction of Mst77F in BC200 buffer (20 mM TrisHCl pH = 7.9, 20% glycerol, 0.2 mM EDTA, 200 mM KCl, 1 mM DTT, 0.5 mM PMSF) at varying concentrations was added and the whole reaction was incubated for 20 min at 30°C. When pG5ML chromatin was the respective template two sequential incubation periods became necessary. The first incubation was carried out at 30°C for 20 min after the addition of 1 µL/reaction p300 HAT (20 ng/µL). This allowed p300 mediated histone acetylation that turned the transcriptional repressed pG5ML chromatin into a bona fide transcriptional template. Subsequently varying concentrations of Mst77F in BC200 buffer were added to the reaction. After another 20 min incubation at 30°C the transcription reaction was initiated by addition of 1 µL/reaction 10 µCi α^{32}P-CTP (3000 Ci/mM) and 10 µL/reaction HeLa nuclear extract, which was obtained from the Department of Cellular Biochemisty, Max Planck

Institute for Biophysical Chemistry, Goettingen, Germany (Dignam et al., 1983). After incubation at 30°C for 1 h, 2 µL/reaction RNAseT1 (10 U/µL) were added and the incubation was continued for another 15 min at 30°C. 1/10 volume of a loading control (pG5MLΔ50, generated by inserting a G into the G-less cassette at position -50) that was transcribed in a separate reaction were added to each sample. The transcripts were recovered using the RNeasy Mini Kit according to the manufacturers recommendations. The purified RNA was dried in a speed - vac and resuspended in RNA loading dye (95 % formamide, 50 mM TrisHCl pH = 7.5, 0.05% (w/v) bromphenol blue and xylenecyanol). After boiling for 5 min, the samples were loaded onto a 5 % urea PAGE gel (50 % w/v urea, 5 % acrylamid-bisacrylamide 30:1 in TBE) and electrophoresis at 25 W for 30 min. The gel was subsequently dried at 80°C for 2 h and the radioactive signal was visualized by autoradiography on a storage phosphor screen for 1h. Quantifications of signal intensities were performed using ImageJ version 1.42q.

2.11.6 Micrococcus nuclease digestion assay

Digestion of mononucleosomes was performed according to [154] with variations. 5 µg mononucleosomes were incubated with 0.2 units Micrococcal nuclease (MNase) in a 1 ml reaction volume in MNase buffer (20 mM TrisHCl pH=7.2, 5 mM $MgCl_2$, 3 mM $CaCl_2$) at 20°C for 1 min, 2 min, 4 min and 8 min. After the corresponding time a 200 µL aliquot of the reactionmixture was collected and mixed with 400 µl of binding buffer (taken from the Macherey&Nagel PCRpurification kit) to instantaneously stop the reaction. DNA was purified withthe Macherey&Nagel PCR purification kit according to manufacturer's protocol, and samples were analyzed by 5% TBE-PAGE.

2.12 Biochemical and biophysical binding assays

2.12.1 Generation of DNA templates

In order to investigate aspects of the protein DNA interaction several different DNA templates were used. The short DNAs used for most of the experiments (6mer duplex, 12mer duplex and 20mer duplex, poly AT duplex, poly GC duplex) were

generated by thermal annealing from oligonucleotides. The oligo nucleotides were mixed in equimolar ratios, heated to 95°C in a thermomixer and incubated at this temperature for 10 min. Then, the thermomixer was switched off and the samples were allowed to cool to RT slowly. Formation of the duplex DNA was checked by 15% TBE - PAGE.

Tabl. 2.9: Oligonucleotides used for thermal annealing

DNA	Sequence 5' – 3'
6mer duplex fluorescine fwd	[Flc]CGGTAG
6mer duplex rev	CTACCG
12mer duplex fluorescine fwd	[Flc]GTACCACGGTAG
12mer duplex rev	CTACCGTGGTAC
20mer duplex fluorescine fwd	[Flc]GTACCACGGTAGTGAATGCT
20mer duplex rev	AGCATTCACTACCGTGGTAC
poly AT duplex fluorescine fwd	[Flc]AAAAAAAAAAAA
poly AT duplex rev	TTTTTTTTTTTT
poly GC duplex fluorescine fwd	[Flc]GGGGGGGGGGGG
poly GC duplex rev	CCCCCCCCCCCC

The long 12 x 200 x 601 template used for AFM imaging was prepared by PEG precipitation (Lis and Schleif, 1975).

For the *in vitro* fluorescence DNA cross linking assay the biotinylated / fluorescence 234 bp DNA was generated by large scale PCR using 1x Taq buffer (ThermoFisher), 0.2 mM each "234 x 601" biotinylated / fluorescence forward and unmodified reverse primer (table 2.9) (Sigma, 0.2 mM each dNTP, 2 mM $MgCl_2$, 1 U Taq DNA Polymerase (ThermoFisher) and 10 ng/reaction DNA template containing the "601" nucleosome positioning sequence flanked by linker sequences allowing for efficient PCR amplification. The program given in table 2.10 was applied for the PCR reaction. The PCR products were purified by phenol : chlorophorm : isoamylalcohol extraction and subsequently ethanol precipitated.

Tabl. 2.10: Primer sequences used for the PCR based synthesis of 234 bp DNA used in DNA cross linking assays

Primer Name	Sequence 5' – 3'
234 x 601 biotin fwd	[Biotin]GGTTATGTGATGGACCCTATACG
234 x 601 fluorescine fwd	[Flc]GGTTATGTGATGGACCCTATACG
234 x 601 rev	ATGATTACGAATTCGAGCTCGGTAC

Tabl 2.11: PCR program used for the PCR based synthesis of 234 bp DNA with Pfu – Polymerase (native)

Step	Temperature (°C)	Time (s)	cycles
1	95	180	1
2	95	30	35
3	55	30	
4	72	60	
5	72	120	1
6	10	∝	-

2.12.2 12mer duplex DNA Pulldown experiments

Typically, 2.5 µg of biotinylated 12mer duplex DNA were pre-incubated with 50 µL of streptavidin coated paramagnetic particles in binding buffer (10 mM Triethanolamine-HCl = pH 7.5, 150 mM NaCl, 0.1 % v/v Nonidet P – 40, 1 mM DTT) O/N at 4°C. Unbound DNA was removed by two washes with 1 mL binding buffer each. Prior to the incubation with protein the beads were blocked with 0.5% (w/v) milk dissolved in binding buffer for 1 h at RT. Then, the immobilized 12mer duplex DNA was incubated with 7.5 µg recombinant Mst77F protein for 2 h at RT in blocking buffer. The beads were washed three times with washing buffer (10 mM triethanolamine-HCl = pH 7.5, 300 mM NaCl, 0.2 % v/v Nonidet P – 40, 1 mM DTT) and bound proteins were eluted

by boiling the beads in protein loading buffer. Samples were analyzed by SDS - PAGE and visualized by Coomassie staining.

2.12.3 Chromatin co-precipitation experiments

10 ng/µL of nucleosomal arrays (OD260 = 0.2) were incubated in a titration series with recombinant Mst77F protein for 1 hr at 4°C and a total volume of 100 µL in binding buffer (10 mM Triethanolamine pH = 7.4. 150 mM NaCl, 1mM DTT). 5 mM $MgCl_2$ was added to the nucleosomal arrays with Mst77F to induce nucleosome clustering (Schwarz et al., 1996). Oligonucleosomes were pelleted by centrifugation for 30 min at 16,000 x *g* and 4°C. Pellets were washed once with 1 mL of binding buffer containing 5 mM $MgCl_2$ and centrifuged for 15 min at 16,000 x g and 4°C. Pellets were solubilized in protein loading buffer. Samples were run on a SDS - PAGE and proteins were visualized by Coomassie staining.

2.12.4 Fluorescence polarisation (FP)

Fluorescence polarisation binding assays were performed according to (Jacobs, 2004) with 5 nM fluorescein-labeled 12mer duplex DNA (table 2.8) and 1 nM to 10 µM of purified Mst77F protein in 10 µL of 10 mM triethanolamine-HCl pH = 7.5 and 150 mM NaCl at 4°C in black microtiter plates. Fluorescence polarization was measured using a Chameleon multi-label counter equipped with a 485 nm fluorescence polarization excitation filter and 535 nm fluorescence polarization emission filters in parallel and perpendicular orientation. Anisotropy (A) values were calculated using the equation:

$$A = \frac{2P}{3 - P}$$

P is the degree of fluorescence polarization calculated by the equation:

$$P = \frac{I_{parallel} - I_{perpendicular}}{I_{parallel} + I_{perpendicular}}$$

*I*parallel and *I*perpendicular are the intensities of the emitted fluorescence light in parallel and perpendicular orientation. For the analysis of binding curves, non-linear

least-square fitting of the data was carried out using the equation:

$$A = A_f + \left(A_b + A_f\right)\frac{c_{prot}}{K_d + c_{prot}}$$

where A_f and A_b are the measured anisotropy values of free and bound peptides, $c_{protein}$ is the known protein concentration and K_d is the dissociation constant. All data sets were plotted and fitted with the Kaleidagraph software.

2.12.5 Isothermal Titration Calorimetry (ITC)

The 12mer duplex DNA used in this assay was produced by thermal annealing according to paragraph 2.12.1 and table 2.8. The sample was concentrated to 50 µM using Amicon Ultra centrifugation filter units (Millipore) and dialyzed extensively against 10 mM TriethanolamineHCl = pH 7.5, 150 mM NaCl, 1mM TCEP. Recombinant Mst77F was concentrated to 100 µM and dialyzed in parallel against the exact same buffer as the DNA sample. All ITC measurements were performed on an iTC200 calorimeter at 20°C in the Department of Physical Chemistry at the Max-Planck Institute for Biophysical Chemistry, Goettingen, Germany. The thermodynamic parameters of the protein : DNA interaction were recorded through the emitted heats of 20 sequential protein injections into the 12mer duplex DNA. At each injection step 2 µL of 34.4 µM Mst77F protein solution were released into the Calorimeter cell containing 13.8 µM 12mer duplex DNA. Each injection was spaced by a 120 s interval of constant stirring at 1000 rpm to allow for baseline stabilisation. The heats of dilution, obtained by titration of Mst77F into buffer, were subtracted from raw data before analysis. Raw data were integrated with respect to time, normalized per mol of injected Mst77F and the apparent heat change (Δq) was plotted against the molecular ratio of Mst77F vs. 12mer duplex DNA by the implemented Origin analysis software. For determination of apparent enthalpy changes(ΔH_{app}), the molar association constant (K_A), and the stoichiometry (n) of the interaction, nonlinear least-square fitting of the Δq values was performed by the Origin software using a one set of identical binding sites model. The heat change for each injection is described by the equation:

$$\Delta q_i = \frac{nc_{b,i}\Delta HV}{2}\left[1+\frac{c_{a,i}}{nc_{b,i}}+\frac{1}{nKc_{b,i}}-\sqrt{\left(1+\frac{c_{a,i}}{nc_{b,i}}+\frac{1}{nKc_{b,i}}\right)^2-\frac{4c_{a,i}}{nc_{b,i}}}\right]$$

q_i is the molar heat change, V the cell volume, $c_{a,i}$ the total concentration of the injected component in the cell after the i^{th} injection step, and $c_{b,i}$ the total concentration of the macromolecule in the cell after the i^{th} injection (Jelesarov and Bosshard, 1999).

2.12.6 Electrophoretic Mobility Shift Assay (EMSA)

The EMSA experiment was carried out on the basis of the original protocoll (Garner and Revzin, 1981) and adopted in the following way. The 12mer duplex fluorescine labeled DNA template (random sequence) used in this assay was assembled by thermal annealing according paragraph 2.12.1 and table 2.8. Mononucleosomes and 12 x 200 x 601 nucleosomal arrays were assembled according to the paragraph 2.11.4. 50 ng / reaction 12 mer duplex DNA, 250 ng / reaction mononucleosomes and 12 x 200 x 601 nucleosomal arrays, respectivly, were used as binding partners in a titration series of various Mst77F and control proteins. The protein titration was ranging from 20 nM to 5.12 µM in 2-fold dilution steps. All reactions were carried out in EMSA buffer (10 mM Hepes (NaOH) pH = 7.4, 50 mM KCl, 0.25 mg / mL BSA, 5 mM DTT and 5% (v/v) glycerol) (Hayashihara et al., 2010) with the proteins to be added last. EMSA reactions were incubated at room temperature for 20 min and subsequently were loaded either onto 15% 0.5X TBE – PAGE gels (12 mer duplex DNA), 5% 0.5X TBE – PAGE gels (Mononucleosomes) or 1% 0.5X TBE – Agarose gels (12 x 200 x 601 oligonucleosomal arrays). All gels were run at a constant 100 V at 4°C in the cold room with variation in the run time (90 min for the PAGE gels and 240 min for the Agarose gels). The EMSA reactions carried out with the 12mer fluorescine duplex DNA were visualized with the FLA-5100 imaging system in the Department of Physical Biochemistry at the Max – Planck Institute for Biophysical Chemistry, Goettingen, Germany. Mononucleosome and Oligonucleosome EMSAs were stained and visualized according to paragraph 2.7.7. Quantifications of signal intensities were performed using ImageJ version 1.42q (Rasband, 1997 - 2011) and plotted with the Kaleidagraph software.

2.12.7 Protein cross - linking assay

Protein multimerization was investigated in the presence of DNA. The protein specific amine reactive chemical cross – linker BS^3 (Bis[sulfosuccinimidyl]suberate) was used to stabilize the protein – protein interaction and to exclude any unspecific reaction with the DNA present in the reaction mix (Sinz, 2003). Prior to the cross – linking all proteins were dialysed into 10 mM triethanolamineHCl pH = 7.4, 150 mM NaCl and 1 mM DTT containing buffer. A final protein concentration of 20 µM in a 10 µL reaction volume was used for each cross – linking reaction in the above specified buffer. 12 x 200 x 601 DNA was titrated to the proteins. At the lowest ratio a 500-fold molar excess of protein over DNA was used. The DNA amount was increased 2-fold in each titration ending up with 30 fold molar excess of protein over DNA. After addition of the DNA the reaction was incubated at room temperature for 30 min. Then, the chemical cross linker BS^3 was added to the reaction in a 20-fold molar excess (400 µM) over the protein and the reaction was incubated at room temperature for another 90 min. The cross - linking was terminated by the addition of 100 mM final TrisHCl pH = 7.4 buffer and an incubation period of 15 min at room temperature. The SDS – PAGE and protein staining was carried out according to paragraphs 2.4 and 2.5

2.13 DNA / Chromatin compaction assays

2.13.1 Atomic Force Microscopy

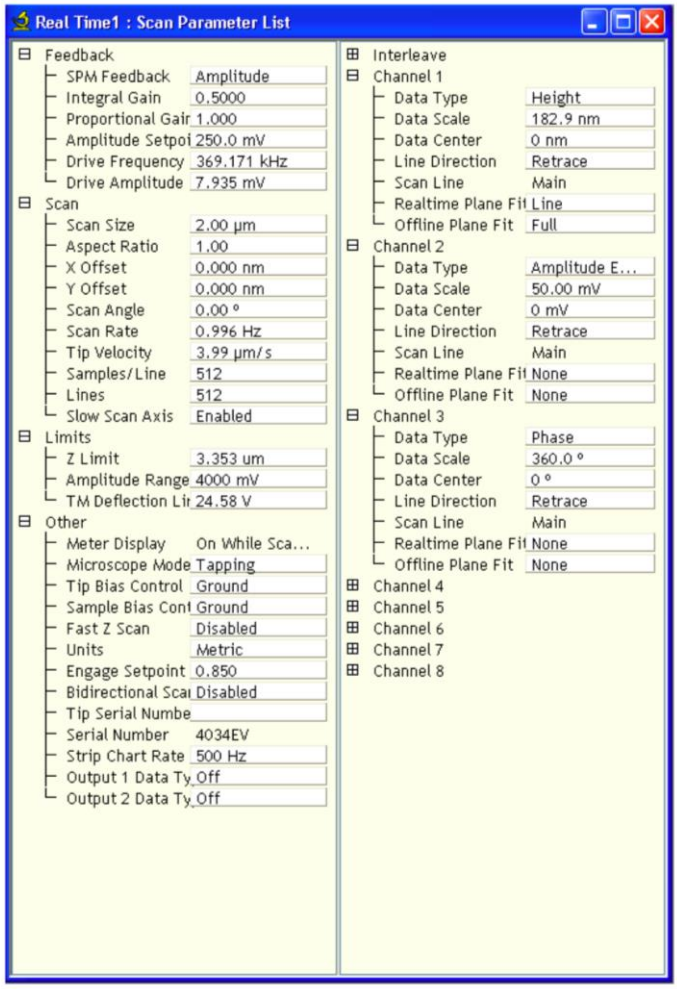
Sample preparation and AFM imaging was basicly performed according to published protocols (Leuba and Bustamante, 1999) with modifications towards the imaging of DNA. All buffers used for sample preparation were filtered through a Anotop 10 - 0.02 µM filter Unit. The 12 x 200 x 601 DNA and the respective proteins were dialysed against 10 mM triethanolamineHCl pH = 7.4, 150 mM NaCl and 1 mM DTT over night. The next day 3.71×10^{-15} mol (0.3 ng/µL) of the DNA were incubated with varying amounts of Mst77F and control proteins (4-fold, 20-fold and 100-fold molar excess of protein over DNA) in a total volume of 20 µL. The reactions were incubated at RT for 30 min and subsequently deposited onto the surface (~ 0.5 cm^2) of a freshly cleaved mica. The sample was incubated for 10 min at RT. The surface was drained

with the help of a KIMTECH paper towel and washed with 2 mL of ddH_2O. Prior to imaging the mica was dried for 20 min in a stream of filtered compressed air.

The 12 x 200 x 601 nucleosomal arrays and recombinant Mst77F protein were dialyzed against 10 mM triethanolamineHCl pH = 7.4, 50 mM NaCl and 1 mM DTT. 3.44 x 10^{-13} mol (10 ng / µL) chromatin were preincubated for 1 h at 4°C with 1.38 x 10^{-12} mol (4-fold molar excess) of Mst77F protein in a total volume of 100 µL. Fresh glutaraldehyde was added to a final concentration of 0.05% (v/v) and the sample was incubated over night at 4°C. The next day the glutaraldehyde was removed by extensive dialysis against 10 mM triethanolamineHCl pH = 7.4 and 50 mM NaCl. 10 µL of the reaction was deposited onto the surface (~ 0.5 cm^2) of a freshly cleaved mica. The sample was incubated for 10 min at RT. The surface was drained with the help of a paper towel and washed carefully with 2 mL of ddH_2O. Prior to imaging the mica was dried for 20 min in a stream of filtered compressed air.

The Mst77F WT and Mst77F Δ100N proteins in conjunction with the 12mer duplex DNA were complexed at concentrations and conditions that were applied for the differential centrifugation assay (paragraph 2.13.2). 150 ng of 12mer duplex DNA were incubated with either 40 nM or 640 nM of protein in a final volume of 50 µL. Prior to deposition the samples were diluted 1 : 10 in the reaction buffer and 20 µL were pipetted onto the surface of a freshly cleaved mica. As before, the surface was drained with the help of a KIMTECH paper towel and washed carefully with 2 mL of ddH_2O. Prior to imaging the mica was dried for 20 min in a stream of filtered compressed air.

Tabl. 2.11: AFM settings applied for the recorded images



All Images were recorded in air on a Nanoscope V Multi Mode atomic force microscope in tapping mode using an “E“ - scanner and silicon etched probe tips with a spring constant of 40 N/m and a resonance frequency of 325 kHz. All images were captured using the parameter settings shown in table 2.11 and flattened by the Nanoscope V software.

2.13.2 Centrifugation Fractionation Assay

The clustering of the Mst77F protein triggered by its association with DNA was analysed by centrifugation fractionation. The orignal assay (Schwarz et al., 1996) was adopted as follows: 150 ng of 12mer duplex DNA was bound to either the wild - type Mst77F, Mst77F Δ100N, hH1.4 or xPR-Set7 in a total volume of 50 µL in 10 mM

triethanolamineHCl pH = 7.4 buffer supplemented with 150 mM NaCl and 1 mM DTT. The proteins were applied at concentrations of 40 nM to 1.28 µM in concentration doubling increments. Immediatly after addition of the proteins and thorough mixing 20 µL of each sample were removed and saved as input. All reactions were subsequently kept on ice for 30 min to allow for equilibration. Then the protein DNA complexes were pelleted by centrifugation for 30 min at 16000 x g and 4°C. Again, 20 µL of the supernatant were removed and saved as output. To correlate the amounts of DNA within the respective input and corresponding output the DNA of all samples was stained with 0.5 µg / mL ethidium bromide (Bonasera et al., 2007). The fluorescence was measured on a Chameleon multi-label counter equipped with a 360 nm excitation filter and a 612 nm emission filter. For each sample the ratio of input DNA vs. output DNA was determined and plotted vs. the respective protein concentration with the Kaleidagraph software.

2.13.3 In vitro DNA cross linking assay

The assays were performed according to (Vogel et al., 2011). All incubations were carried out at room temperature. In brief, black 96 well plates coated with Streptavidin were pre-washed three times with 200 µL each of 1X PBS. 500 ng of biotinylated 234 bp duplex DNA were synthesized according to paragraph 2.12.1 (table 2.9 and 2.10) and immobilized on the surface of the 96 well plate in immobilization buffer (10 mM TrisHCl pH = 7.4, 2 M NaCl, 1 mM EDTA) for 1 h under constant agitation. The wells were briefly washed 3 times with 200 µL each of 1X PBS. 500 ng of 5' fluorescine 234 bp duplex DNA (synthesized according to paragraph 2.12.1 (table 2.9 and 2.10)) were mixed in a titration series ranging from 160 nM to 5.12 µM with recombinant Mst77F proteins or control proteins in binding buffer (10 mM triethanolamineHCl pH = 7.4, 150 mM NaCl, 1 mM DTT), and were immediatly transfered to the wells with the immobilized biotinylated DNA. The reactions were incubated for 1 h under constant agitation. After measuring the input fluorescence for all proteins the wells were washed 10 times for 5 min each with 200 µL 1X PBS under constant agitation and the remaining fluorescence was measured. The fluorescence was recorded using a Chameleon multi - label counter (Hidex) equipped with a 485 nm excitation filter and a 535 nm emission filter. Relative

fluorescence was calculated from the mean absolute values (output fluorescence / input fluorescence) of three independant experiments. Control measurements were carried out using either fluorescence DNA alone without any immobilized biotin - DNA or fluorescence DNA + the respective protein without biotin - DNA or fluorescence DNA in the background of immobilized biotin - DNA.

2.14 Other techniques

2.14.1 Circular Dichroism (CD)

The protein samples were dialyzed into 10 mM triethanolamineHCl pH = 7.4, 150 mM sodiumfluoride buffer and concentrated / diluted to a final concentration of 30 µM. The Circular Dichroism spectra were recorded on a ChiraScan CD spectrometer at 20°C in a 2 mm Quartz Suprasil CD cell in the Department of Neurobiology, Max - Planck Institute for Biophysical Chemistry, Göttingen, Germany.

2.14.2 Analytical Ultra Centrifugation

For protein analysis by the sedimentation velocity method (Schuck, 2000), 0.8 OD260 of recombinant Mst77 was prepared in 400 µL triethanolamineHCl pH = 7.4, 150 mM NaCl and 1 mM β-mercaptoethanol. Double-sector charcoal filled epon cells were filled with 412 µL of buffer in the reference sector and 392 µL protein solution in the sample sector. After loading the centrifuge with the rotor, the bucket of the instrument was evacuated to 1 micron vacuum pressure. Prior to starting the run the sample was equilibrated to 20°C for a 1 h. The sample was centrifuged at a speed of 35.000 rpm and sedimentation of the molecules was monitored over 200 scans. During the run, scans were continuously acquired. The data analysis was performed with the SEDFIT software (Schuck, 2000) using a partial specific volume of Mst77 calculated from its' amino acid composition with the software SEDNTERP.
Generally, after positioning the meniscus and the bottom, a simplex fit for the meniscus position was performed at a resolution of 50-100. The frictional ratio was fitted with the simplex algorithm. The initial fits were further refined by the simulated annealing algorithm until the root mean square deviation converged at a minimum. Figures of the analytical ultracentrifugation analyses were prepared by exporting the

raw data at a resolution of 200 to the Origin software.

3. Results

The Mst77F protein is exclusively expressed in the testis of *Drosophila* males. During postmeiotic spermatid maturation it arises in the transition phase from a nucleo-histone towards a protamine based organization of the genome. *In vivo* observations suggest a role in the condensation of the DNA in order to reach a high level of compaction. Beyond the descriptive work that characterized the gene and its timely and spacial expression no information about the proteins structure or its exact function *in vivo* are available. In this study I set out to characterize Mst77F on the molecular level and its functional implication in the compaction of DNA by *in vitro* techniques.

3.1 Mst77F is a protein of bivalent structural organization

Since the discovery of the *Drosophila* Mst77F gene in 1993 (Russell and Kaiser, 1993) little knowledge about the molecular structure or the functional role of the Mst77F protein during *Drosophila* spermatogenesis could be obtained. To develop hypotheses that help to characterize the protein in experimental assays I applied, as a starting point several online available protein domain prediction algorithms to the Mst77F protein sequence. The only tool amongst several that identified putatively structured domains of Mst77F was the **S**imple **M**odular **A**rchitecture **R**esearch **T**ool (SMART) (Schultz et al., 1998)(Fig. 3.1 panal A) that assigned a N-terminal coiled-coil motif to the region of amino acids 26 to 49 based on the *coils2* algorithm (Lupas et al., 1991). Coiled-coils represent a class of α-helical protein- protein interaction interfaces in a multitude of proteins with different functions (Mason and Arndt, 2004). The relationship between the amino acid sequence and the finally folded structure adopting a coiled-coil conformation is well studied (Mason and Arndt, 2004; Nishikawa and Scheraga, 1976) and a high confidence prediction is therefore possible.

The second motif which was annotated within the Mst77F sequence identified by SMART shows two regions of low compositional complexity between amino acids 171 to 191 and 200 to 214 (Wootton JC, 1996) (Fig. 3.1 panal A). Such regions

display only little diversity in their amino acid composition and show loosely clustered, irregularly spaced or periodic patterns of these residues (Coletta et al., 2010). Structural information is not available, because a) there is no sequence conservation due to rapid evolution across species (Ekman et al., 2006) and b) those regions do not readily crystallize, arguing towards their structural undefined nature (Huntley and Golding, 2002). Nevertheless, low compositional complexity regions are found in a number of distinct proteins that either adopt folded conformations or display intrinsically unstructured properties (Tatham and Shewry, 2000; Tompa, 2002). For some proteins the pivotal biological function of low complexity regions could be demonstrated (Phatnani and Greenleaf, 2006; Wanker et al., 1995).

To confirm the protein motif prediction and to obtain additional structural properties I conducted a secondary structure analysis using the secondary structure prediction online tool (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sspred.html) (Combet et al., 2000) in conjunction with the determination of the relative protein disorder by the IUPred tool (see paragraph 2.3) (Dosztanyi et al., 2005a, b). Secondary structure and relative disorder prediction were coherent in their results (Fig. 3.1 panel B, data not shown for sec. structure prediction). The N-terminal region (amino acids 10 to 90) of Mst77F adopts an ordered confirmation due to several predicted α-helices and a few extended strands. This is in good agreement with the putative coiled-coil motif identified in the SMART blast. Additionally, the possible α-helical organization of Mst77F is extended up to amino acid 90. More then half of the protein (amino acids 91 to 215) cannot faithfully be correlated with any regularly folded secondary structure but random coils. This is also reflected in the relative disorder prediction and by the SMART annotation predicting an overall highly bivalent protein organization: A putatively structured N-terminus organized by α-helices and an unstructured C-terminus in flexible random coil conformation.

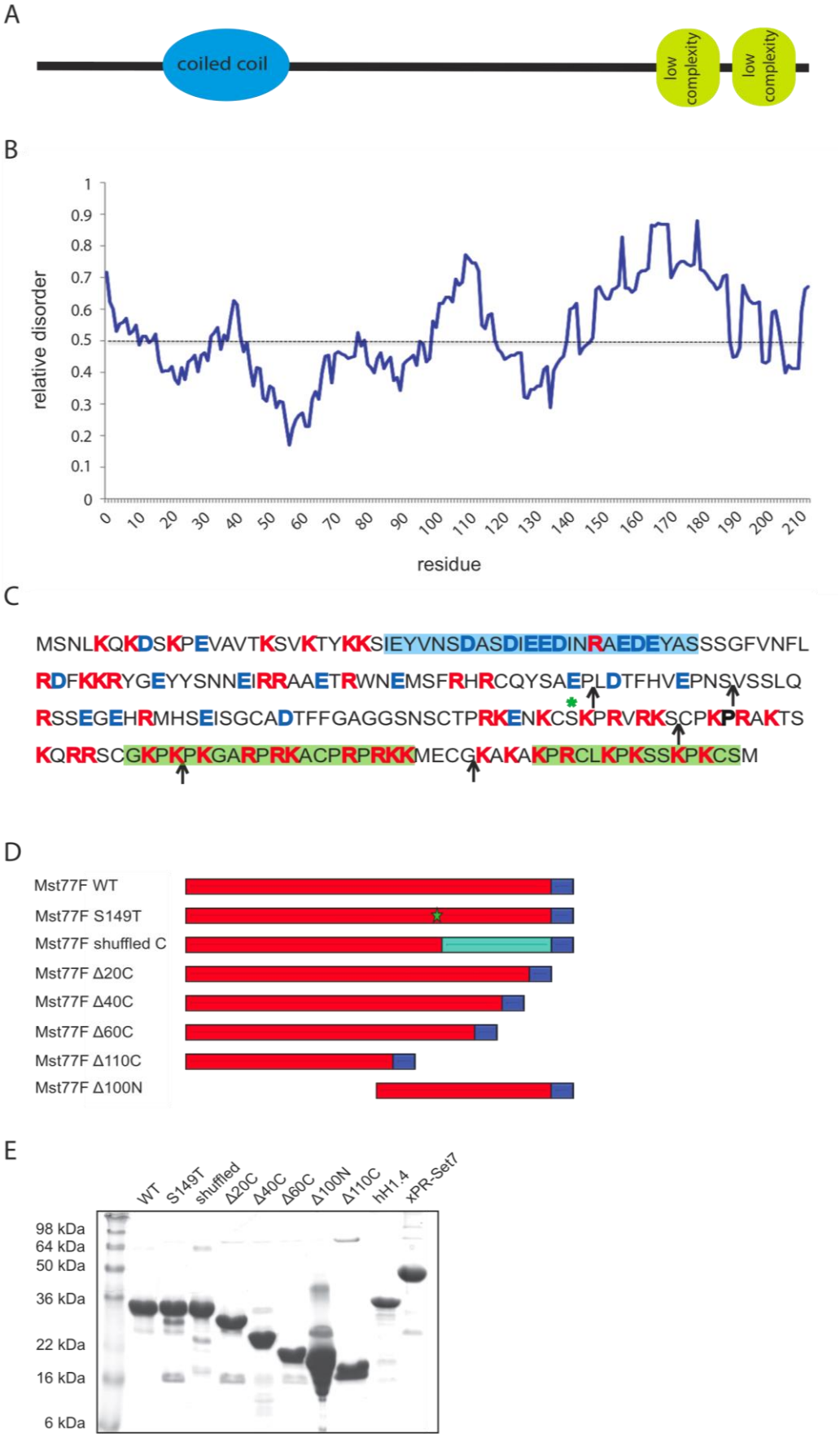

A
coiled coil
low complexity
low complexity
B
relative disorder
residue
C
MSNLKQKDSKPEVAVTKSVKTYKKSIEYVNSDASDIEEDINRAEDEYASSSGFVNFL
RDFKKRYGEYYSNNEIRRAAETRWNEMSFRHRCQYSAEPLDTFHVEPNSVSSLQ
RSSEGEHRMHSEISGCADTFFGAGGSNSCTPRKENKCSKPRVRKSCPKPRAKTS
KQRRSCGKPKPKGARPRKACPRPRKKMECGKAKAKPRCLKPKSSKPKCSM
D
Mst77F WT
Mst77F S149T
Mst77F shuffled C
Mst77F Δ20C
Mst77F Δ40C
Mst77F Δ60C
Mst77F Δ110C
Mst77F Δ100N
E
WT
S149T
shuffled
Δ20C
Δ40C
Δ60C
Δ100N
Δ110C
hH1.4
xPR-Set7
98 kDa
64 kDa
50 kDa
36 kDa
22 kDa
16 kDa
6 kDa

Fig. 3.1: Structural parameters of Mst77F

(**A**) Mst77F domain architecture according to SMART prediction: N-terminal coiled-coil motif (blue) as possible protein-protein interaction module and C-terminal low compositional complexity regions (green) with undefined structure and function for Mst77F. (**B**) Protein relative disorder prediction for Mst77F shows a bivalent organization of the protein. Roughly the first half of the protein except the very N-terminal 10 amino acids are assumed to have folded characteristics whereas the C-terminus is predicted to be largely unstructured. (**C**) Mst77F primary structure annotations: coiled-coil (blue background), low complexity region (green background), positively charged amino acids (red), negatively charged amino acids (blue), arrows represent the cloned Mst77F truncation mutants from the C-terminus and N-terminus respectively, Mst77F S149T point mutant (green asterisk). (**D**) Schematic representation of the Mst77F construct expressed in *E.coli*. 6xHis affinity purification tag (blue box), Mst77F with compositionally intact but shuffled amino acid order in the highly charged C-terminus (cyan box). (**E**) Affinity purified Mst77F proteins, hH1.4 and xPR-Set7 resolved on 15% Tris-Glycine SDS-PAGE

By taking a closer look at the possibly unstructured C-terminus it was striking that this part of the protein is very rich in basic amino acids (Fig. 3.1 panel C). 60% of all positive charges are situated within the last 70 amino acids of Mst77F. This local positive charge accumulation in the C-terminal domain in conjunction with its predicted random coil conformation resembles the structural properties of histone H1 family proteins and some members of HMG-box family of chromosomal, architectural proteins (Dow et al., 2000; Dragan et al., 2003). In these proteins the positively charged N- or C-terminal tails execute a major contribution to DNA binding via electrostatic interactions with the negatively charged phosphodiester backbone of the DNA (Churchill, 1999; Lu et al., 2009a). For histone H1 it could be demonstrated that the CTD undergoes structural rearrangements upon interaction with DNA.
Despite the physical and structural relationship of the Mst77F C-terminus with histone H1 proteins and some HMG proteins no conserved H1/H5 globular domain or a HMG-box motif could be detected in Mst77F.

3.2 The Mst77F-DNA interaction

Previous studies reported Mst77F to be an important factor in postmeiotic spermatid maturation in *Drosophila.* Functions in chromatin condensation and nuclear shaping were proposed (Rathke et al., 2010; Sunil Jayaramaiah Raja, 2005). These findings are solely based on cytogenetic experiments and no molecular prove for a direct interaction with DNA has been found. However, the Mst77F sequence analysis

revealed a physiochemical similarity of the C-terminal tail towards a class of HMG proteins. These proteins functionally act as structural components of chromatin through a direct interaction with DNA in an unspecific, charge-mediated manner (Churchill, 1999; Dragan et al., 2003). Also, histone H1 type proteins, major structural components of cellular chromatin exhibit a highly charged C-terminus that is only little structured in solution (Allan et al., 1980; Dootz et al., 2011; Happel and Doenecke, 2009; Lu et al., 2009a). Additionally, large patches of positive charges have been suggested to be indicative for protein nucleic acid interfaces (Honig et al., 1989; Jones et al., 2003; Shazman et al., 2007). On the basis of these data I hypothesized that Mst77F binds DNA via its extremely basic C-terminal domain. Due to its high positive charge density and putatively unstructured conformation the interaction might be unspecific and mainly of ionic nature.

3.2.1 The Mst77F C-terminal domain is necessary and sufficient for DNA binding

To test these hypotheses I expressed along with the wild-type several deletion-mutant Mst77F proteins (Fig. 3.1 panel D, E). These proteins were subsequently tested in two conceptually different assays. On the one hand I performed pulldown experiments with the recombinant Mst77F proteins and immobilized dodecamer duplex DNA (paragraph 2.12.2). This assay revealed qualitative and kinetic information about the respective interactions. Alternatively, I conducted quantitative measurements of the interaction under equilibrium conditions by fluorescence polarization (FP) using the same dodecameric duplex DNA but a 5' fluorescine instead of a biotin-tag (paragraph 2.12.4). In both assays the wild-type protein displayed strong binding to the respective DNA template (Fig. 3.2 A and B left) and the interaction was quantified with an equilibrium dissociation constant of 90 $\pm$ 20 nM (Tabl. 3.1). Mst77F Δ20C, Mst77F Δ40C and Mst77F Δ60C are characterized by their stepwise deletions of the C-terminal basic domain thereby reducing the local, high-density charge in 25% intervals. These proteins should reveal the functional impact of the tail in the DNA binding reaction. In pulldown assays Mst77F Δ20C and Mst77F Δ40C showed a similar recovery rate compared to the wild-type protein even though the overall charge sum is reduced by 25% and 50%, respectively. Mst77F Δ60C

(75% charge reduction) abolished binding completely giving a clear indication towards a dominating role of the C-terminal domain in the interaction with DNA (Fig 3.2 panel A). This result was supported by the data of the analyzed control proteins. Mst77F Δ110C lacks the

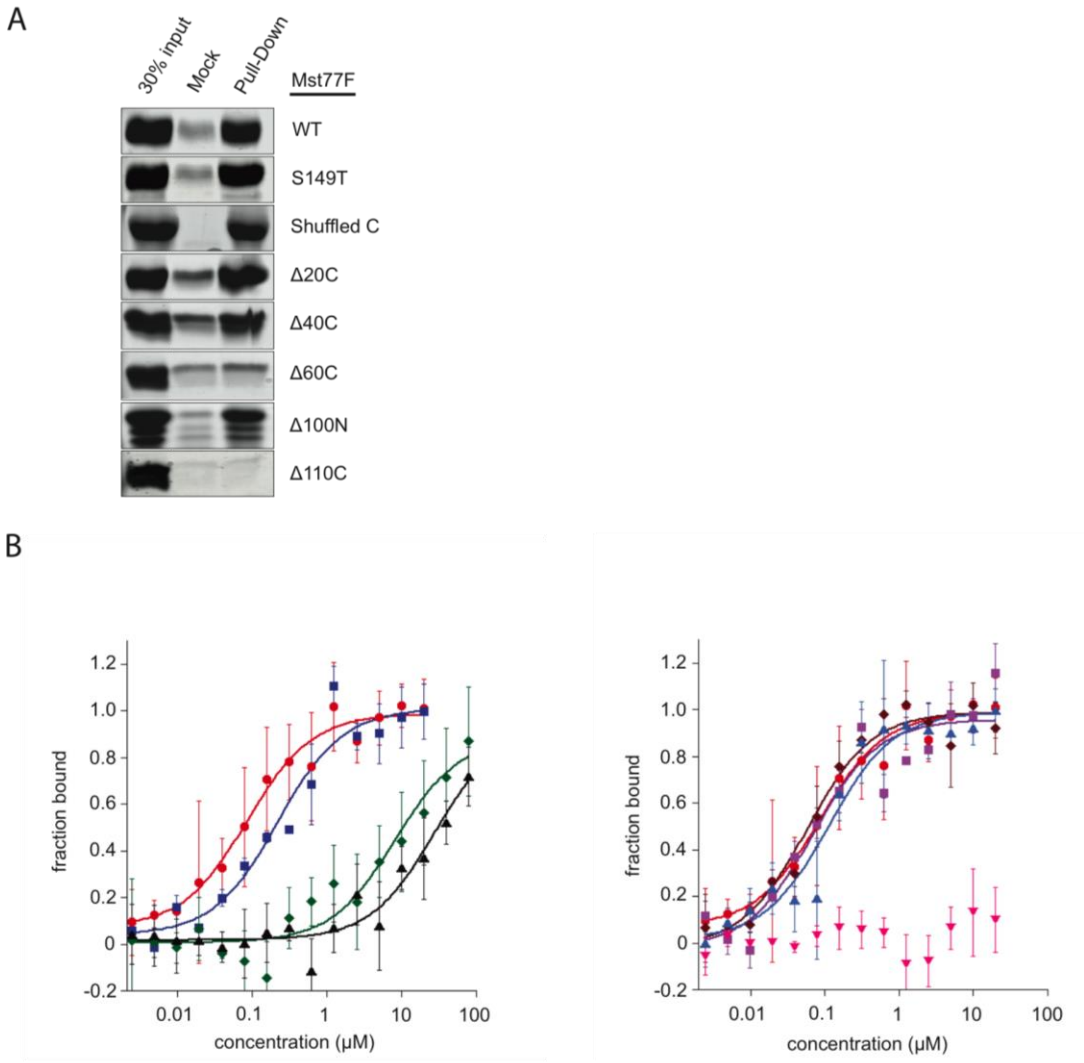


Fig. 3.2 Mst77F binds DNA with its C-terminal positively charged tail

(**A**) Pulldown experiment with DNA dodecamers immobilized on the surface of paramagnetic beads. Mock represents beads without DNA incubated with the respective protein. (**B**) Fluorescence Polarization equilibrium binding experiments with 12mer duplex DNA. The average bound fraction of 3 independant experiments is plotted. Curve fitting was performed as described in 2.12.4 using the KaleidaGraph software *left plot*: binding curves of Mst77F wild-type (red), Mst77F Δ20C (blue), Mst77F Δ40C (green), Mst77F Δ60C (black); *right plot*: Mst77F S149T (purple), Mst77F shuffled C (brown), Mst77F Δ100N (light blue) Mst77F wild-type (red), Mst77F Δ110C (pink).

C-terminal domain and displayed no binding to DNA. Vice versa the Mst77F Δ100N mutant that misses the putative α-helical N-terminus restored binding to WT level. Additionally, the shuffled C-terminal domain Mst77F mutant with a randomized order of the C-terminal domain amino acids showed equal recovery in respect to the WT protein ruling out a possible structural contribution coming from the C-terminus.

The Mst77F S149T point mutant has been reported to show severe phenotypes in cytogenetic experiments in *Drosophila.* Nuclear shaping defects of spermatids and male sterility is observed but chromatin condensation appears to be normal (Rathke et al., 2010). In conformity with the *in vivo* observed chromatin compaction in the Mst77F S149T background, the S149T mutant interacted with DNA to a similar extent as the wild-type protein.

Quantitatively, even though the charge decline occurred in steps of 25% in the Mst77F Δ20C, Mst77F Δ40C and Mst77F Δ60C mutants, the decrease in measured affinity was not proportional to the percentage of charge reduction (Fig 3.2 panel B left and Tabl. 3.1). The changes in the equilibrium dissociation constants observed between Mst77F WT and Mst77F Δ20C mutant (2.5 fold decrease; WT 90 $\pm$ 20 nM vs. Δ20C 230 $\pm$ 70 nM) or Mst77F Δ40C and Mst77F Δ60C (3.5 fold decrease; Δ40C 8.42 $\pm$ 3.13 µM vs. Δ60C 28.6 $\pm$ 12.5 µM) are rather mild when compared to the transition from Mst77F Δ20C to Mst77F Δ40C (40 fold decrease; Δ20C 230 $\pm$ 70 nM vs. Δ40C 8.42 $\pm$ 3.13 µM). However, this binding effect with its abrupt decrease in affinity in the transition from Δ20C to Δ40C was not reflected by the protein recovery in the pulldown assay pointing towards kinetic effects under the applied experimental conditions. The differences observed might manifest by altered association kinetics that became apparent in one assay but not the other. Strikingly, bioinformatic analysis assigned a low compositional complexity motif to this region. The recorded effect might therefore be brought about by a functional sub-domain within the unstructured region. Such functional sub-domains within intrinsically disordered domains have been identified (Lu and Hansen, 2004).

Coherent to the pulldown results the equilibrium dissociation constants for Mst77F Δ100N, Mst77F shuffled C-terminal domain and Mst77F S149T were not significantly different from the wild-type protein and no binding for the Mst77F Δ110C protein could be observed (Fig. 3.2 panel B right and Tabl. 3.1).

Taken together, the results from the pulldown experiment and the FP measurements strongly indicate a high affinity interaction of Mst77F with DNA mediated by the C-terminal domain of the protein.

3.2.2 Mst77F binding to DNA is based on sequence unspecific ionic interactions

Binding experiments with Mst77F and short DNAs identified the C-terminal domain to be sufficient for the interaction (paragraph 3.2.1). The high positive charge content along with a predicted structurally undefined conformation allows for an ionic interaction type that is sequence unspecific. Binding that virtually depends on opposing charges should not be influenced by structural properties of the DNA or sequence. Therefore, I tested the binding of Mst77F to different subtypes of B-DNA.
Poly(dA)*poly(dT) duplexes have been described to form a narrow minor groove with a wider major groove whereas poly(dG)*poly(dC) DNA structurally displays a wide minor groove and a narrow major groove (Nelson et al., 1987; Yoon et al., 1988). The interactions of Mst77F with poly(dA)*poly(dT) as well as poly(dG)*poly(dC) dodecamers were compared to regular B form DNA (random sequence) of the same length (same negative charge sum) in fluorescence polarization binding assays (Fig. 3.3 panel A). If major and minor groove are the predominant high affinity binding parameters in the protein DNA complex changing their dimensions should result in altered binding affinities. This was not observed. The measured equilibrium dissociation constants are not significantly different (Tabl. 3.1).
Along with natural structural variations of B-DNA caused by repetitive AT or GC sequences the canonical Watson & Crick B-form DNA was structurally modified by pre-incubation with ethidium bromid. Moreover, also single strand DNA that has no helical structure was measured in a Mst77F binding reaction (Fig. 3.3 panel B). The observed deviations of the measured equilibrium dissociation constants were marginal from the Mst77F interaction with regular B-type DNA and not significant (Tabl. 3.1).
If the conformational status and the sequence of the DNA are irrelevant for high affinity interactions, the origin must be different. In fact, a salt titration series that stepwise compensates the opposing charges of DNA and protein resulted in loss of binding (Fig. 3.3 panel C). The binding strength of an ionic interaction type is based

on the size of the available interaction interface in terms of accessible charges. To this point all measurements were carried out with dodecameric DNAs. In an ionic interaction shorter and longer DNAs should result in lower and higher equilibrium dissociation constants, respectively. This indeed could be observed for affinity measurements with hexamers and eicosamers (Fig. 3.3 panel D). The similar K_d values for various structurally different DNAs used in the FP measurements are strongly indicative for a Mst77F binding mechanism that is independent of high

Tabl. 3.1 Equilibrium dissociation constants measured by fluorescence polarization.

Values represent means of 3 independent experiments obtained by non-linear least square fitting using the Kaleidagraph software.

DNA	K_d (nM)
6mer duplex	1260 ± 170
12mer duplex	90 ± 20
20mer duplex	20 ± 10
12mer single strand	240 ± 60
NaCl (mM)	**K_d (nM)**
20	110 ± 50
150	90 ± 20
300	500 ± 260
450	17950 ± 7500
12mer duplex	**K_d (nM)**
poly-GC	170 ± 60
poly-AT	70 ± 30
random	90 ± 20
Protein	**K_d (nM)**
Mst77F WT	90 ± 20
Mst77F S149T	70 ± 30
Mst77F shuffled C	50 ± 10
Mst77F d20C	230 ± 70
Mst77F d40C	8420 ± 3130
Mst77F d60C	28590 ± 12500
Mst77F d100N	120 ± 40
Mst77F d110C	could not be determined

affinity major and minor groove binding. In fact, experiments with increased salt concentrations abolished the interaction and shorter/longer DNAs with a lower/higher charge sum resulted in weaker/stronger equilibrium dissociation constants.
All together these experiments strongly point towards an unspecific ionic interaction mechanism of Mst77F with DNA.

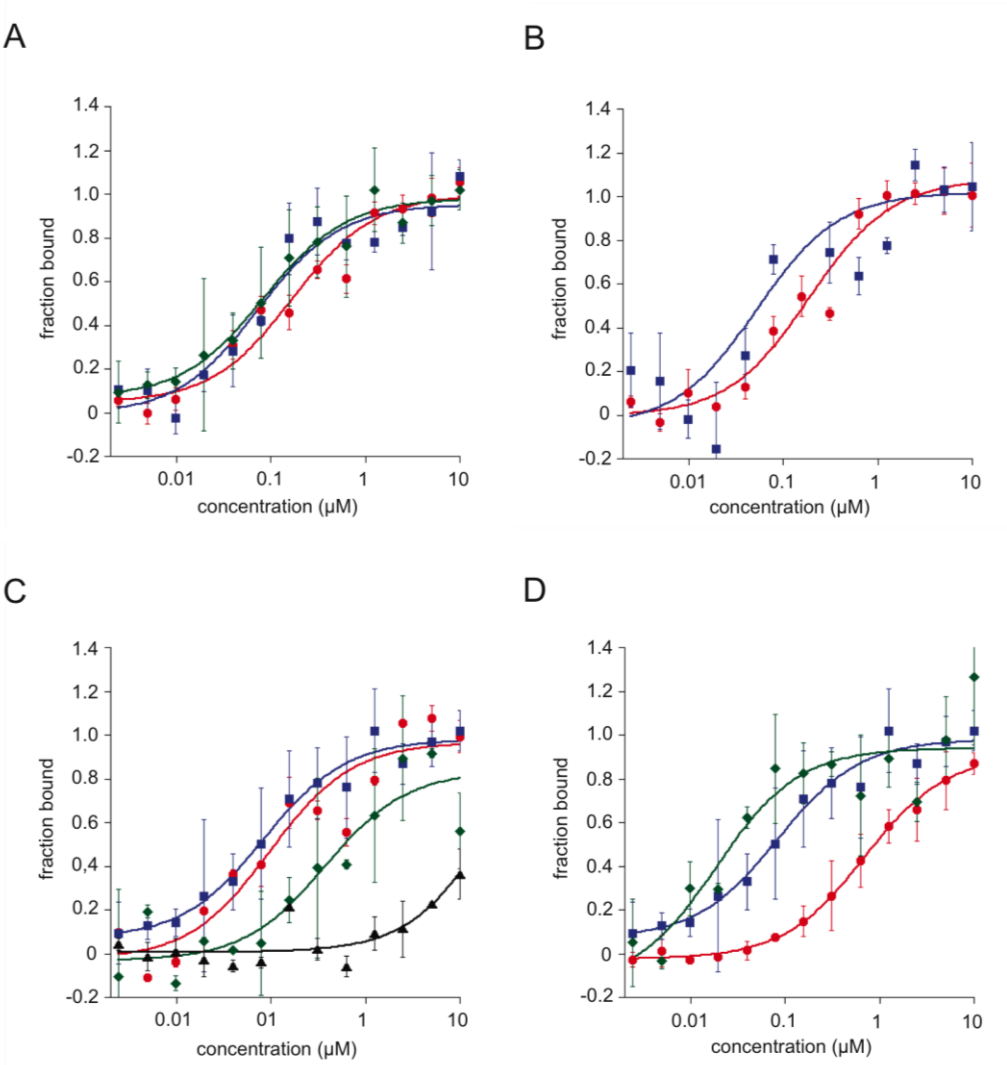
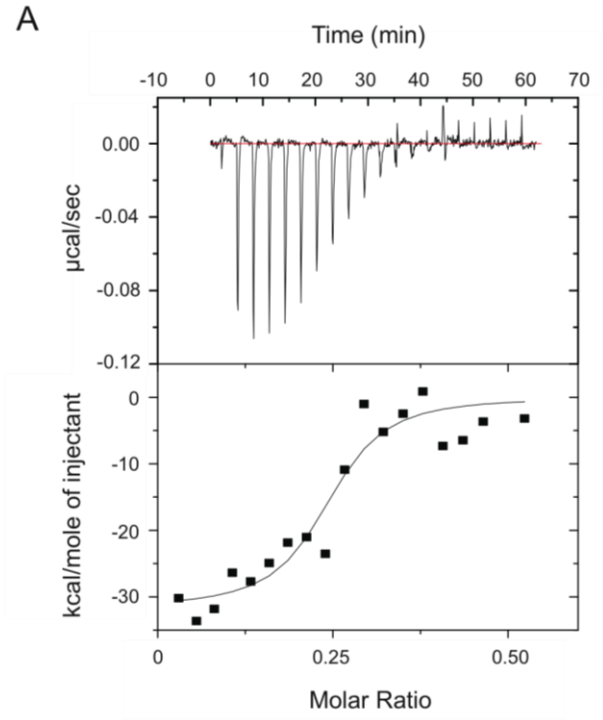


Fig. 3.3 Mst77F binding to DNA is based on an ionic interaction mechanism

Fluorescence polarization binding experiments. Plotted are average bound fractions of 3 independent experiments. Curve fitting was performed as described in 2.12.4 using the KaleidaGraph software (**A**) Mst77F wild-type binding to short B-form DNAs: 12mer poly GC (red), 12mer poly AT (blue), 12mer random (green) (**B**) Mst77F interacts with the phosphate backbone of nucleic acids: single stranded DNA (red) and structurally distorted DNA mediated by ethidium bromide intercalation (blue) (**C**) Increasing NaCl concentrations abolish Mst77Fs interaction with DNA: 20 mM NaCl (red), 150 mM NaCl (blue), 300 mM NaCl (green), 450 mM NaCl (black) (**D**) Longer DNAs are bound by MSt77F with higher affinity: 6mer (red), 12mer (blue), 20mer (green)

3.3 Thermodynamics of the Mst77F DNA interaction

According to the previous experiments the Mst77F DNA interaction is strongly based on electrostatic binding. Nevertheless, recent studies on intrinsically unstructured proteins reported secondary structures induced upon recognition of bona fide interaction partners (Lu et al., 2009a; Shewmaker et al., 2008; Tompa, 2002). Within this recognition process ionic interaction play a fundamental role. To investigate the possible structural induction upon protein DNA interaction I conducted isothermal titration calorimetry (ITC). ITC results give the complete set of thermodynamic parameters that are the signature of an interaction. Changes in Mst77F structure introduced upon binding to DNA dodecamers should be identified by a characteristic entropic contribution. Additionally, beyond the thermodynamic fingerprint the stoichiometric ration of the interaction is identified. It therefore is possible to characterize the size of the interaction interface.



B

Kd (µM)	0,10 +/- 0,16
N	0,24 +/- 0,01
ΔG	- 9.35
ΔH (kcal/mol)	-31,65 +/- 2,38
TΔS (kcal/mol)	-22,3

Fig. 3.4 Isothermal Titration Calorimetry with Mst77F and DNA dodecamers

(**A - top**) Raw data of heat release upon stepwise injection of recombinant Mst77F wild-type**.** 2 µL of protein solution was injected into 250 µL of DNA solution. Intervalls between injections were 120 s. (**bottom**) Plot and fit of the integrated heats of Mst77F injections vs. molar ratio of Mst77F per DNA molecule calculated by the MicroCal software ; (**B**) data table of thermodynamic parameters of the Mst77F wild-type - 12mer DNA interaction derived from fitting the data with a *one set of identical binding sites* model. Measurements were carried out at 20°C in 10 mM Triethanolamine, 150 mM NaCl, 1 mM TCEP

From the measured thermodynamic parameters of Mst77F binding DNA dodecamers a stoichiometric ratio of one protein binding four short DNAs could be derived (Fig. 3.4 panel B). This suggests a brought interface across the charges of the C-terminal domain, even though information about the exact interaction positions is not available. Noticeable are the high negative enthalpy (ΔH) and entropy ($T\Delta S$) values indicating the type of interaction. Whereas the strong negative ΔH value reflects the ionic interaction mode (salt bridges, hydrogen bonding), the high negative $T\Delta S$ is difficult to explain. A strong negative contribution is usually a measure of high flexibility or hydrophobic interactions whereas positive contributions in entropy reflect protein folding. To this point there is no indication of structural induction upon binding to DNA. Due to the proposed random coil conformation it is likely a combination of high flexibility due to electrostatic repulsion between amino acids of the same charge and additional hydrophobic interactions between amino acids and DNA bases. In agreement with the FP measurements the derived equilibrium dissociation constant determined by ITC is 0.1 µM. This result confirms the previously obtained K_d values by FP and underpins the validity of the thermodynamic parameters.

3.4 The mechanism of the Mst77F-DNA interaction

Despite the characterization of the Mst77F binding domain and the determination of the interaction parameters with DNA by FP and ITC the question of functional relevance of this unspecific interaction emerged. Usually unspecific binding mechanisms are involved in one-dimensional scanning along the DNA. This lateral diffusion in search for a bona fide binding site that is eventually recognized with high affinity is a general feature of sequence-specific DNA binding proteins (e.g. transcription factors). However, this usually also involves amino acids of additional protein motifs that insert into major or minor groove and convey the cognate sequence specificity (Winter et al., 1981). These extensively studied mechanisms were also confirmed for structural DNA binding proteins that show no sequence preferences but exhibit their function by motifs that recognize major/minor groove and thereby introduce structural changes (Klass et al., 2003). The previous binding experiments with Mst77F did not point towards an interaction mode that involves

groove insertion of particular amino acids. However, Mst77F has been implicated in genomic shaping during spermatid maturation (Rathke et al., 2010). Previous studies suggested Mst77F being a distant member of the histone H1 family of structural proteins and sequence alignments show a similarity in the C-terminal domains (CTD) of both proteins (Russell and Kaiser, 1993). Importantly, the H1 CTD is assumed to be intrinsically unstructured and its necessity for H1 function in chromatin compaction could be demonstrated (Lu et al., 2009a). The initial interaction between protein and DNA/chromatin is therein mediated by electrostatic mechanisms. On this basis I hypothesized that the Mst77F DNA interaction is similar to histone H1 interactions with DNA. Moreover, I wished to demarcate effects that are based on simple adhesion of positively charged patches on their surface of proteins that are not necessarily involved in a direct interaction with DNA.

3.4.1 Mst77F aggregates DNA

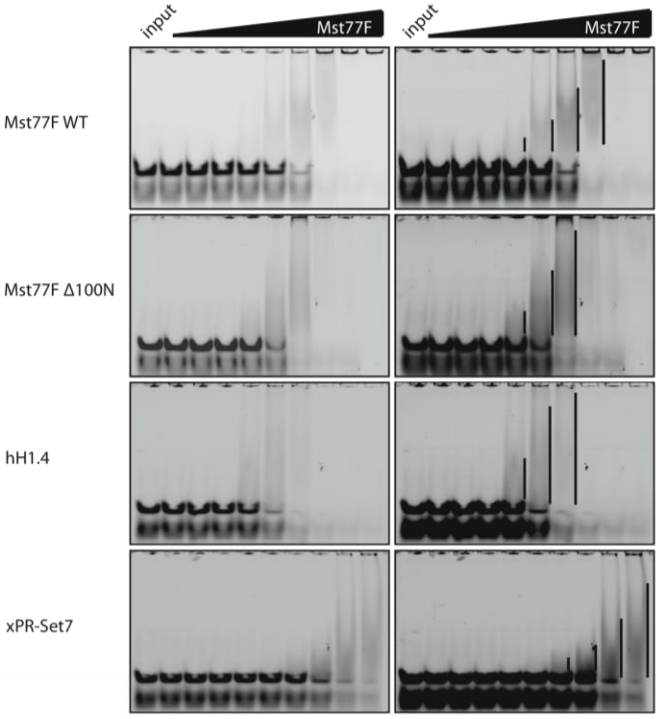
To further refine the binding of Mst77F to DNA I studied the gel migration behavior of dodecameric fluorescine tagged duplex DNA in the context of previously tested Mst77F proteins in electrophoretic mobility shift assays (EMSA) (Fig. 3.5 panel C and partially B). Ordinary EMSA experiments usually result in a discrete shifted band that appears upon binding of a protein. In a 1:1 interaction model a single band is expected and multiple interaction result in a complex pattern of numerous bands. However, the shifting of the free DNA upon binding of a protein is concentration dependant and typically follows the law of mass action. Figure 3.5 panel C and panel B (partially) show that Mst77F did not shift the free DNA towards a discrete high molecular weight band. This was surprising since the ITC experiment uncovered a stoichiometric ratio of four short DNA dodecamers per protein. Instead it generated smeary shifting patterns starting at K_d concentrations in a narrow window. At saturating conditions the DNA was removed from the gel reflecting the initial formation of distinct stoichiometric complexes and finally their aggregation. The observed dose effect within the tested Mst77F concentrations depended on the length of the C-terminal tail (Fig. 3.5 panel C). As expected the Mst77F Δ110C mutant that misses the charged tail did not “shift” the DNA. Mst77F shuffled C-terminal domain and Mst77F S149T behave very similar to the wild-type protein (Fig.

A

Ⅰ MSETAPAAPAAPAPAEKTPVKKKARKSAGAAKRKASGPPVSELITKAVAASKERSGVSLAALKKALAAAGYD
VEKNNSRIKLGLKSLVSKGTLVQTKGTGASGSFKLNKKAASGEAKPKAKKAGAAKAKKPAGAAKKPKKATG
AATPKKSAKKTPKKAKKPAAAAGAKKAKSPKKAKAAKPKKAPKSPAKAKAVKPKAAKPKTAKPKAAKPKK
AAAKKK

Ⅱ MGRGKKMSKPGDGRSGDVSDTGRNGGTNENHPKTNGEVVHCGQAKIYSYMSPTKSPSARPPLQEENSV
THHESKCLGKPSTETRKKAEVEKKKILSTELSVKPSEQRETECNSIGEFLEPKLELNDVQRNLALPPEDKLQ
SQKMVKNKPLRKKTQRQKSPNRKLTDYYPVRRSSRKNKTEIESEEKKRIDELIQTGKEEGIKMHMITGKGR
GVIATRDFQRGEFVVEYHGDLIEITDAKRREASYAQDSATGCYMYYFQYLNTSYCIDATRETGRLGRLINHSK
SGNCHTKLHNINNVPHLILVASRDINVGEELLYDYGDRRKSSIDAHPWLKN

B



C

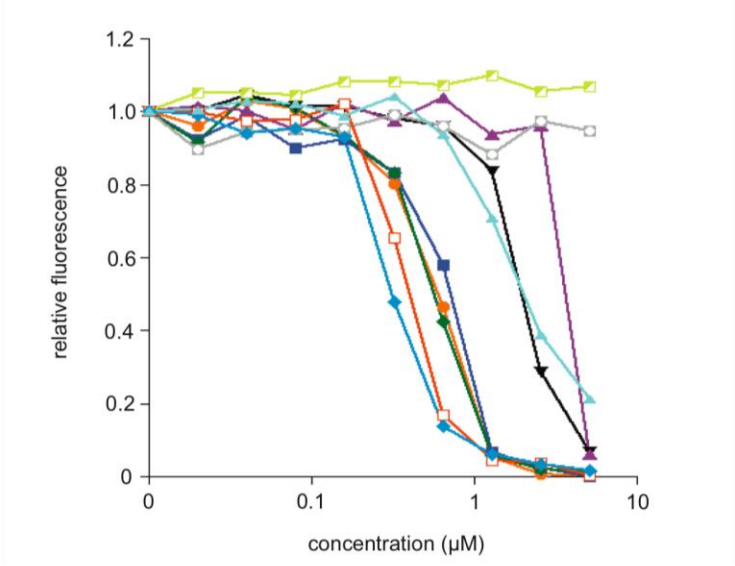


Fig. 3.5 Unspecific charge mediated protein – DNA interactions in EMSA experiments

(**A**) Protein sequences of hH1.4 (I) and xPR-Set7 (II); basic amino acids are annotated in red, acidic amino acids are displayed in blue (**B**) EMSA experiment: Mst77F wild-type, Mst77F Δ100N, hH1.4 and xPR-Set7 were complexed with DNA dodecamers in a concentration titration range from 0.02 to 5.12 µM in 2 fold increments. Samples were run on 15% TBE-PAGE gels for 90 min @ 80 volts. Input represents DNA dodecamers only. Parallel arranged gel within the same row show different exposures times. Unspecific DNA protein complexes are marked with bars (**C**) DNA dodecamer band intensities of EMSA experiments were integrated using the ImageJ software and plotted with Kaleidagraph: Mst77F wild-type (orange), Mst77F S149T (blue), Mst77F shuffled C (green), Mst77F Δ20C (black). Mst77F Δ40C (purple), Mst77F Δ60C (yellow), Mst77F Δ100N (brown), Mst77F Δ110C (light blue), hH1.4 (), xPR-Set7 (turquoise).

3.5 panel C). Very interesting was the gel migration behavior of DNA Mst77F Δ100N complexes. It was only slightly different from the wild-type protein towards a stronger dose response (Fig. 3.5 panel B). A functional contribution of the suggested α-helical N-terminus in the DNA binding event became not obvious.

The unexpected behavior of Mst77F DNA interaction in the EMSA experiments raised the question weather the observed effects are explainable by the ionic type of interaction. On this account I compared Mst77F to positively charged proteins that either have been described to interact with DNA unspecifically by charge or proteins that have not been reported to bind DNA at all.

Histone H1, a major structural component of chromatin (Happel and Doenecke, 2009; van Holde and Zlatanova, 1996) is highly basic (pI = 11.4) and has been shown to strongly bind DNA unspecifically (Dootz et al., 2011). Distinct from Mst77F it displays an even higher charge density that is not limited to the C-terminal tail but almost spans the whole sequence (Fig 3.5 panel A I). On the other hand xPR-Set7 is a known histone modifier with an enzymatic activity that introduces mono-methylation on histone H4 lysine 20 in nucleosomal assemblies (Nishioka et al., 2002). xPR-Set7 resembles in its basicity Mst77F with a pI of 9.4 (Mst77F pI = 9.7) but shows a uniform positive charge distribution across the sequence. Importantly, it has not been demonstrated to physically interact with DNA (Fig 3.5 panel A II). Apart from their common net positive charge those proteins show little structural similarity (Allan et al., 1980; Ramakrishnan et al., 1993; Xiao et al., 2005). Surprisingly, the migration properties of DNA complexed with hH1.4 as well as xPR-Set7 showed the same smear like distribution within a narrow concentration window. However, the dose response of hH1.4 became visible at lower concentrations finally resulting in loss of the DNA from the gel. Within the tested concentrations the loss of DNA was not seen for xPR-Set7 (Fig 3.5 panel B and C).

Taken together, the EMSA assays did not reveal a specific Mst77F on DNA upon binding. Analogous effects were observed for the hH1.4 and xPR-Set7 DNA interactions pointing towards a phenomenon that generally holds true for ion guided interaction, no matter if unspecific or adhesive, and reflects the differential stoichiometric complexes that are formed. However, the observed formation of high molecular weight complexes/aggregates might demarcate the charge mediated unspecific binding effect from simple charge adhesion effects. These complexes were further investigated in the upcoming paragraphs.

3.4.2 Mst77F effects on DNA are mediated by the N-terminus of the protein

Mst77F and hH1.4 are thought to contribute to the structural organization of DNA and chromatin, respectively (Rathke et al., 2010{Thoma, 1977 #62)}(Shen et al., 1995). Their role in the formation of condensed DNA/chromatin states was suggested or has been shown and differences within these complexes due to functional deviation of the proteins are possible. On the other hand xPR-Set7 as a histone modifying enzyme is not expected to have direct structural effects on DNA or chromatin. On this basis I hypothesized that unspecific charge mediated binding exerts effects that can be discriminated from simple charge based adhesion. Moreover, the functions of different unspecific DNA binding proteins and their effects on DNA are also distinct. To test these hypotheses I performed centrifugation fractionation assays to uncover molecular weight/density differences in the formed protein DNA complexes that became not apparent in EMSA experiments. Additionally, I conducted Atomic Force Microscopy (AFM) studies that convey visual information on the respective size of the complexes but also provide useful information on molecular shape. To guarantee consistency throughout the different experiment these assays were carried out with the DNA dodecamers also used for previous experiments.

The principle of the centrifugation assay is outlined in Fig. 3.6 panel A and has been adapted from (Carruthers et al., 1998; Nikitina et al., 2007). The sedimentation behavior of DNA dodecamers in dependence of increasing concentrations of Mst77F wild-type, Mst77F Δ100N, hH1.4 and xPR-Set7 was investigated (Fig 3.6 panel B). At protein concentrations from 0.04 µM to 0.32 µM no significant difference in the sedimentation of the DNA became visible across proteins. The DNA was floating in

the supernatant. At 0.64 µM protein concentration an Mst77F effect that can be attributed to the N-terminal region occurred. At this concentration, congruent with the EMSA assays, a massive sedimentation was observed and only approximately 20% of the input DNA was retained in the supernatant after centrifugation (red curve). Mst77F Δ100N that mostly consists of the charged tail moiety and lacks the N-terminal part of the protein did not show this behavior (blue curve). However, similar characteristics in gel migration were detected for this protein in the EMSA assays. hH1.4 showed a 2-fold stronger dose response in gel shifts compared to Mst77F wild-type and exhibited only mild precipitation of DNA protein complexes at 0.64 µM. 80% of the input DNA were still in the supernatant after centrifugation (green curve). Upon further increase of the protein concentrations to 1.28 µM, there was no DNA detectable in the Mst77F wild-type experiment and only a mild decrease of DNA in the supernatants of Mst77F Δ100N and hH1.4. Throughout the whole titration series the measurable DNA in dependence of increasing PR-Set7 concentrations was constantly high (black curve) reflecting the adhesion between counter-ions.

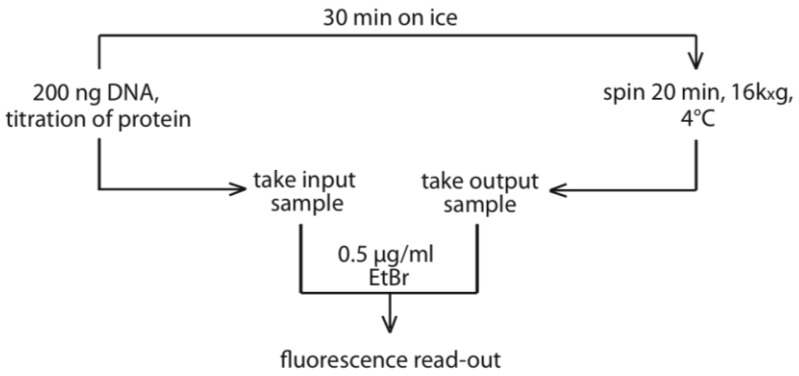
I wished to consolidate the data obtained by the centrifugation fractionation assay with an independent method. Additionally, I was interested to acquire information about the molecular shape of the complexes that might help to develop a hypothesis how Mst77F interacts with DNA. Structural effects of hH1.4 on DNA were reported before. The interaction with DNA is thought to be mediated mostly by the globular H1/H5 domain. Mst77F possesses no H1/H5 fold but exhibits little sequence but major physiochemical similarity in its CTD. On the basis of similar effects observed in EMSA and centrifugation fractionation assays for hH1.4 and Mst77F Δ100N I decided to focus on the structural impact of an isolated CTD and compare it to the wild-type protein.

The sedimentation of the different DNA protein complexes within a gravity field reflected their differential molecular weight/density and therefore varying sizes that should be apparent in optical techniques. I compared Mst77F wild-type with Mst77F Δ100N at two concentrations that resulted in major readout differences in the centrifugation assay. At low protein concentrations (40 nM) the measured DNA fluorescence in the supernatant was similar for both proteins reflecting complexes that cannot readily be precipitated in a gravity field. In AFM differences became visible. The wild-type protein formed rod like structures with the DNA that vary in size

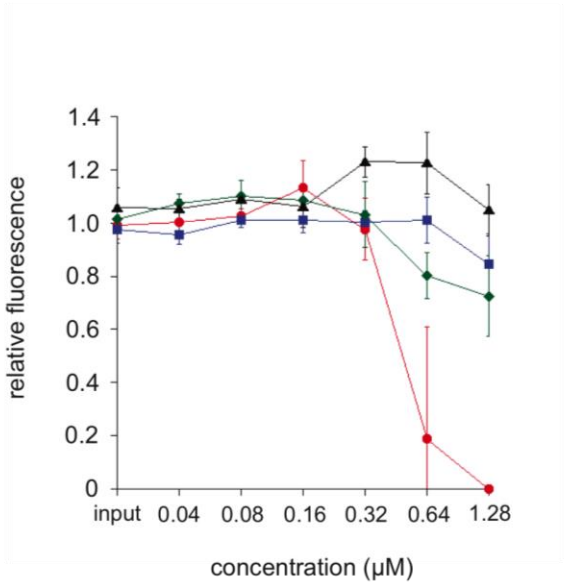
from a few nanometers up to several hundred nanometers. Mst77F Δ100N in contrast showed small globular complexes with very short DNA (Fig. 3.6 panel C). Importantly, in DNA binding experiments no differences between both proteins could be detected neither in pulldown nor FP assays. The structural information on the formed complexes clearly showed differences pointing towards detached functions of N- and C-terminus. However, the density of both structures was not sufficient to precipitate under the applied conditions. Increased protein concentrations (640 nM) that led in the centrifugation assay to rapid sedimentation of DNA Mst77F wild-type complexes were in AFM visible as massive clusters of more than five hundred nanometers in XY dimensions. This was not visible with the Δ100N mutant. Analogous to the centrifugation assay that shows no DNA precipitation evoked by the protein, no structural differences compared to lower concentrations were detectable (Fig. 3.6 panel C right).

In summary, Mst77F exerts specific effects on DNA that are depending on its N-terminus *in vitro.* These effects can be characterized by hydrodynamic assays and imaging techniques and become visible in form of DNA protein aggregates. These aggregates are not seen with similar charged proteins or proteins of high positive charge that already have been reported to structurally alter DNA.

A


30 min on ice
200 ng DNA,
titration of protein
spin 20 min, 16kxg,
4°C
take input
sample
take output
sample
0.5 µg/ml
EtBr
fluorescence read-out

B


relative fluorescence
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
input
0.04
0.08
0.16
0.32
0.64
1.28
concentration (µM)

C

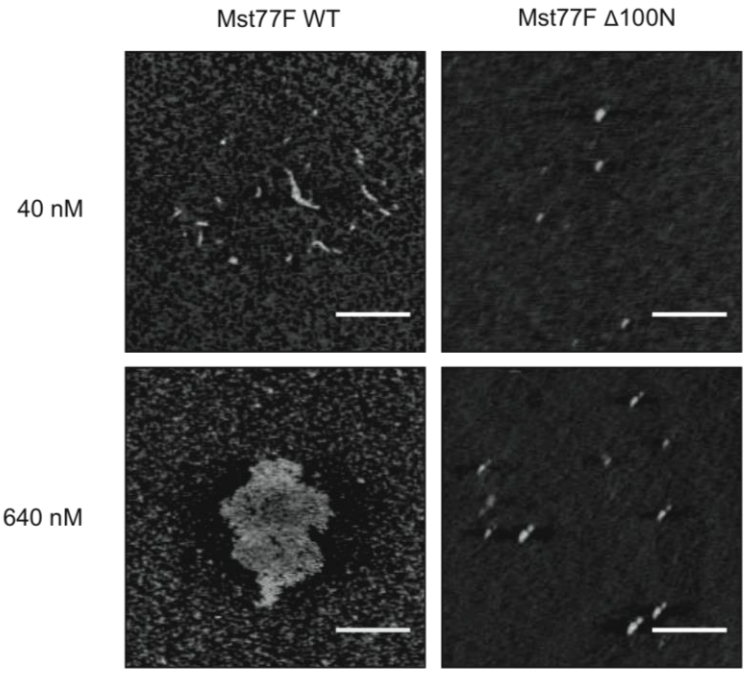

Mst77F WT
Mst77F Δ100N
40 nM
640 nM

Fig 3.6 Mst77F induces aggregation of DNA

(**A**) Schematic representation of the centrifugation fractionation experiment (**B**) Mst77F forms with DNA high molecular weight complexes that can be pelleted by centrifugation. DNA dodecamers were incubated with Mst77F wild – type, Mst77F Δ100N, hH1.4 and xPR-Set7 from 0.04 to 1.28 μM protein concentration in 2 fold increments. The ratio of floating DNA before and after centrifugation was plotted with the Kaleidagraph software. Data represents the mean of three independent experiments: Mst77F wild-type (red), Mst77F Δ100N (blue), hH1.4 (green) and xPR-Set7 (black) (**C**) Mst77F – DNA complexes pelleted by centrifugation can be visualized by Atomic Force Microscopy. Mst77F wild-type and Mst77F Δ100N DNA dodecamer complexes were imaged at 40 nM and 640 nM, respectively. Images were recorded in tapping mode. Scale bars represent 500 nm

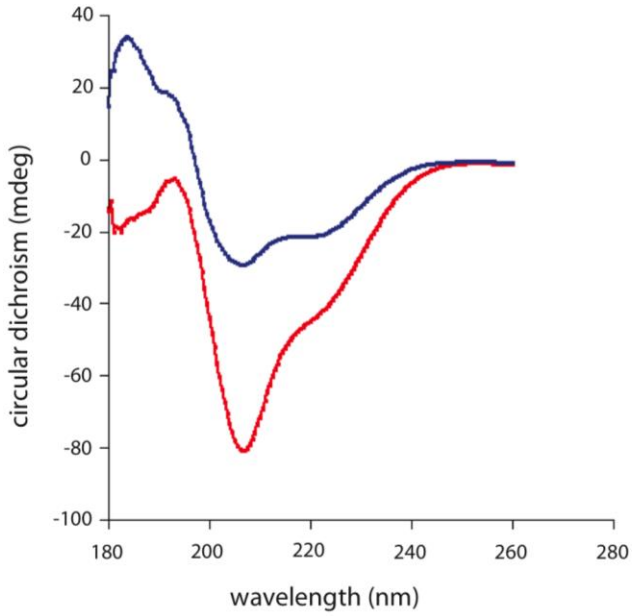
3.5 The Mst77F N-terminus functions as multimerization interface upon DNA recognition

The differences between Mst77F wild-type and Δ100N mutant seen in fractionated centrifugation and AFM experiments leading to a massive DNA aggregation strongly point towards an additional functional contribution of the N-terminus. This effect seems not to be related to a binding event since DNA binding assays did not uncover a N-terminal contribution. However, the bio- informatics sequence analysis predicted a α-helical coiled-coil motif, a well-studied protein-protein interaction interface, establishing the possibility of Mst77F self-interaction (Mason and Arndt, 2004). These facts led me to hypothesize that Mst77F multimerizes through its N-terminal domain. This multimerizaton is necessary to effectuate DNA aggregation.

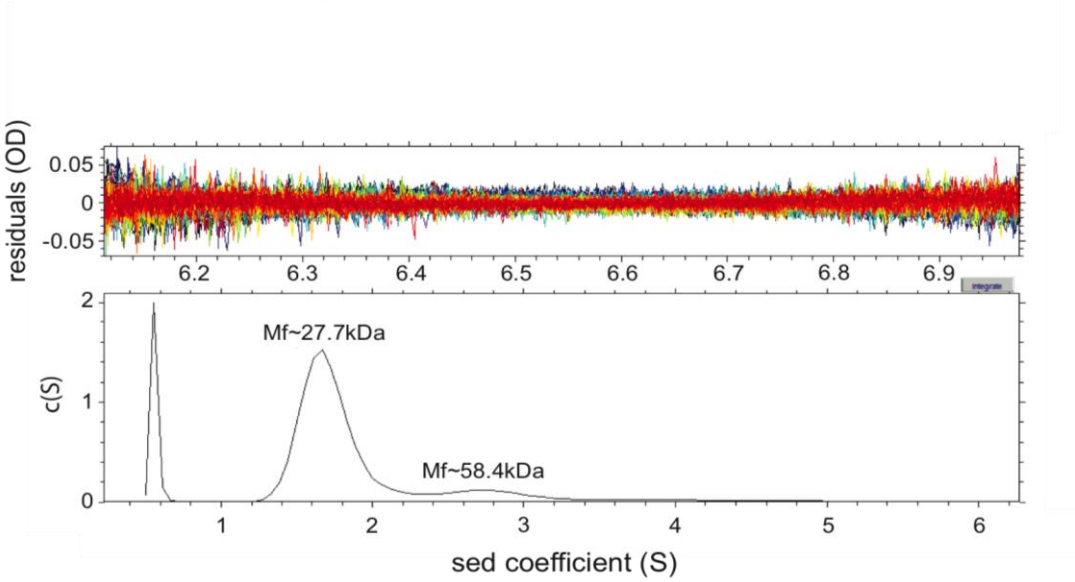
To test this, I first performed circular dichroism (CD) spectroscopy. I compared the Δ110C mutant that according to the prediction should have a high α-helical content to the full-length protein that is suggested to contain an overall elevated random coil structure (Fig. 3.7 panel A). Indeed, the CD spectrum of the Δ110C mutant corroborated the predicted helical structure that dominates the N-terminus as seen for the minima around 220 nm and 205 nm (blue curve). Nevertheless, β-sheets also contributed to the overall N-terminal fold reflected by the plateau around 215 nm. However, as expected within the overall structure of the full-length protein random coils constituted an abundant fraction (red curve) (Johnson, 1990). Nonetheless, the verified helices do not automatically imply a functional multimerization of the protein. A sedimentation velocity analytical ultracentrifugation experiment that records the hydrodynamic sedimentation behavior in a gravity field, revealed a small fraction of the protein that represents a dimer (6%) (Fig. 3.7 panel B). The low ratio of Mst77F

dimer to monomer in solution questions the functional relevance and probably fails to explain the observed aggregation effects on the very short DNA dodecamers. However, the analytical ultracentrifugation experiment was carried out with Mst77F in the absence of DNA. A conceivable explanation is that the binding of the protein to DNA induces a conformational change or a lattice that promotes its multimerization. This was shown for histone H1 that has a similar CTD (Fang et al., 2012; Lu et al., 2009a). To test this hypothesis I conducted protein-protein cross linking experiments in the presence of increasing DNA concentrations (Fig. 3.7 panel C). I compared Mst77F wild-type to the Δ100N mutant, h.H1.4 and xPR-Set7. The wild-type protein displayed a clear DNA dependant multimerization pattern. The titration of DNA led to the formation of increased dimers, tetramers, octamers and also even multitudes of these stoichiometric complexes (Fig. 3.7 panel C very left). As expected the Δ100N mutant showed no unambiguous multimerization pattern independent of the DNA concentration (Fig. 3.7 panel C middle left) pointing towards a specific contribution of the N-terminus in Mst77F function. However, increasing DNA concentrations led to unspecific cross linking of Mst77F Δ100N reflected by loss of protein band intensity and suggesting close proximity between the single molecules upon DNA binding. The observed difference between the wild-type protein and the Δ100N mutant sets the differential DNA interaction mode apart that results in distinct structures as recorded in centrifugation fractionation and AFM experiments. Importantly, previous H1 cross linking experiments showed a oligomerization of the H1/H5 globular domain in the presence of DNA (Clark and Thomas, 1986; Jean O.Thomas, 1991). This became not apparent in my experiments (panel C middle right). In contrast to these studies I worked with the full-length protein including the CTD that is solely rich in highly reactive lysine residues. Considering that I used a 10- fold molar excess of cross linker over protein I reason that the CTD adsorbed all the crosslinker. At the same time that suggests that the CTD of hH1.4 is not in close proximity for cross linking. Moreover even though oligomerization upon DNA binding similar to Mst77F wild-type has been reported, the structural impact on very short DNA was distinct as could be seen by centrifugation fractionation. These results suggest differential interaction modes of hH1.4 and Mst77F with themselves and probably also with DNA. xPR-Set7 has not been reported to form dimers or higher oligomeric states and consistently also showed no specific multimerization in this assay (panel C very right).

A


40
20
0
-20
-40
-60
-80
-100
circular dichroism (mdeg)
180
200
220
240
260
280
wavelength (nm)

B


residuals (OD)
0.05
0
-0.05
6.2
6.3
6.4
6.5
6.6
6.7
6.8
6.9
c(S)
2
1
0
Mf~27.7kDa
Mf~58.4kDa
1
2
3
4
5
6
sed coefficient (S)

C

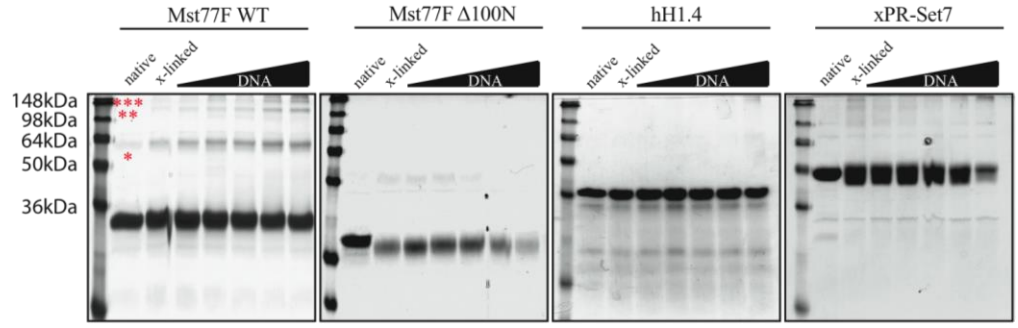

Mst77F WT
Mst77F Δ100N
hH1.4
xPR-Set7
native
x-linked
DNA
148kDa
98kDa
64kDa
50kDa
36kDa

Fig 3.7 The N-terminus of Mst77F multimerizes

(**A**) Circular Dichroism spectroscopy identifies a α-helical conformation of the N-terminal 100 amino acids of Mst77F. Proteins far UV CD spectra were recorded in 10 mM triethanolamine, 150 mM NaF at concentrations of 30 µM. Mst77F Δ110C (blue) and Mst77F wild-type (red) (**B**) Sedimentation velocity analytical ultracentrifugation experiment: plotted is the sedimentation coeficient of Mst77F as a function of the protein concentration. Residuals represent the deviation of the fitted data from the raw date along the measured pathlenght. Molecular weights (MW) were calculated from the SEDFIT program by fitting the sedimentation raw data. Derived protein masses are within the accepted 10% error margin of the method. (**C**) Protein-protein cross linking experiment identified the Mst77F N-terminus as multimerization interface. Cross linking was carried out with the protein specific amine cross linker BS^3 in the presence of increasing amounts of DNA for Mst77F wild-type, Mst77F Δ100N, hH1.4 and xPR-Set7. Cross linker was used in a 10 fold molar excess over protein and cross linking was going on for 1h @ RT. Proteins were analysed on a 15% tris-glycine SDS=GAGE gel. Asterisks indicate dimers (*), tetramers (**) and octamers (***).

In summary, in the presence of DNA Mst77F showed enhanced multimerization via its N-terminus that adapts a α-helical conformation. This multimerization constitutes functional specificity of Mst77F that is different from hH1.4 which also has been reported to multimerizes in the presence of DNA via its globular N-terminal domain. Moreover, these results demarcated the function of the CTD as a binding module from the N-terminal effector module of the protein thereby emphazising the importance of a folded domain in protein function.

3.6 Mst77F induces DNA clustering through its multimerization

The previous experiments clearly demonstrated the necessity of N- and C-terminal domain in Mst77F function as multimerization and binding module, respectively. Functional differences between the proteins could be highlighted. The fact that the Mst77F Δ100N mutant also formed smaller complexes with DNA raises the possibility that additional contribution, that exceeds the initial binding function, comes from the CTD of Mst77F in DNA aggregation. The intrinsically unstructured Mst77F CTD resembles the CTD of histone H1 and for the latter a structural induction upon DNA binding could be demonstrated (Roque et al., 2005). Importantly, this is necessary for the function of H1 in chromatin structure. To propose a model of the Mst77F DNA interaction and its structural impact, the aggregation effects and the contribution of the single domains have to be understood. To further distinguish the binding function from the higher order aggregation effect I conducted protein induced DNA cross

linking experiments (Fig. 3.8) (Vogel et al., 2011). The experimental set up is outlined in Fig. 3.8 panel A and described in the Material and Methods paragraph 2.13.2 in detail. 234 bp DNA was immobilized in a 96 well plate through biotin-streptavidin coupling. The same DNA but with a fluorescine instead of biotin-tag was added in equal quantities. Protein induced DNA cross linking was examined in protein titration experiments and analyzed by plotting the output vs. input signal ratio after washing.

In good agreement to previous assays the functional difference between the wild type protein and the Δ100N mutant became apparent. However, a difference in the dose response in comparison to the other assays (EMSA, centrifugation fractionation, AFM) was also obvious. Basal fluorescence was observed for the initial three titration points (0.16 µM, 0.32 µM, 0.64 µM - less then 10% retained input signal), reflecting low level immobilization of the tracer DNA. A steep increase in the fluorescence signal was seen at 1.28 µM for both proteins, with approximately 30% retained fluorescence. Within a 2-fold concentration margin from 1.28 µM to 2.56 µM the wild-type protein recovered 90% of the input fluorescence signal. 100% immobilization of the tracer DNA in the well was accomplished with 5.12 µM Mst77F wild-type protein. In contrast to the Mst77F wild-type, the recovered fluorescence signal did not increase with elevated Δ100N concentrations and stayed steady at approximately 30% of the input signal over the last three concentrations tested. I reasoned that Mst77F Δ100N mutant indeed formed stable complexes with independent DNAs but this capability is limited. The additive effect of the N-terminal domain in the wild type protein provokes highly efficient aggregation.

To further dissect the functional contribution of the CTD in DNA aggregation, I also tested the Mst77F shuffled C-terminal domain, Δ20C and Δ40C mutants (Fig 3.8 panel B). These mutants showed equivalent recovery in pulldown assays towards the wild-type protein. Importantly, Mst77F shuffled C-terminal domain has a similar equilibrium dissociation constant as the wild-type protein but displayed only 2/3 of the DNA aggregation potential. Additionally, consecutive deletions of the CTD resulted in 50% and 90% reduction of DNA-DNA crosslinking mediated by Mst77F Δ20C and Δ40C, respectively. These effects cannot be explained by the equilibrium dissociation constants of the respective protein, since under the applied conditions 96% of Δ20C and 40% of Δ40C were bound to the DNA. However, an induced structure within the

CTD of Mst77F upon DNA recognition was not discovered but can also not be excluded by the experiments carried out so far. The functional impairment of Mst77F

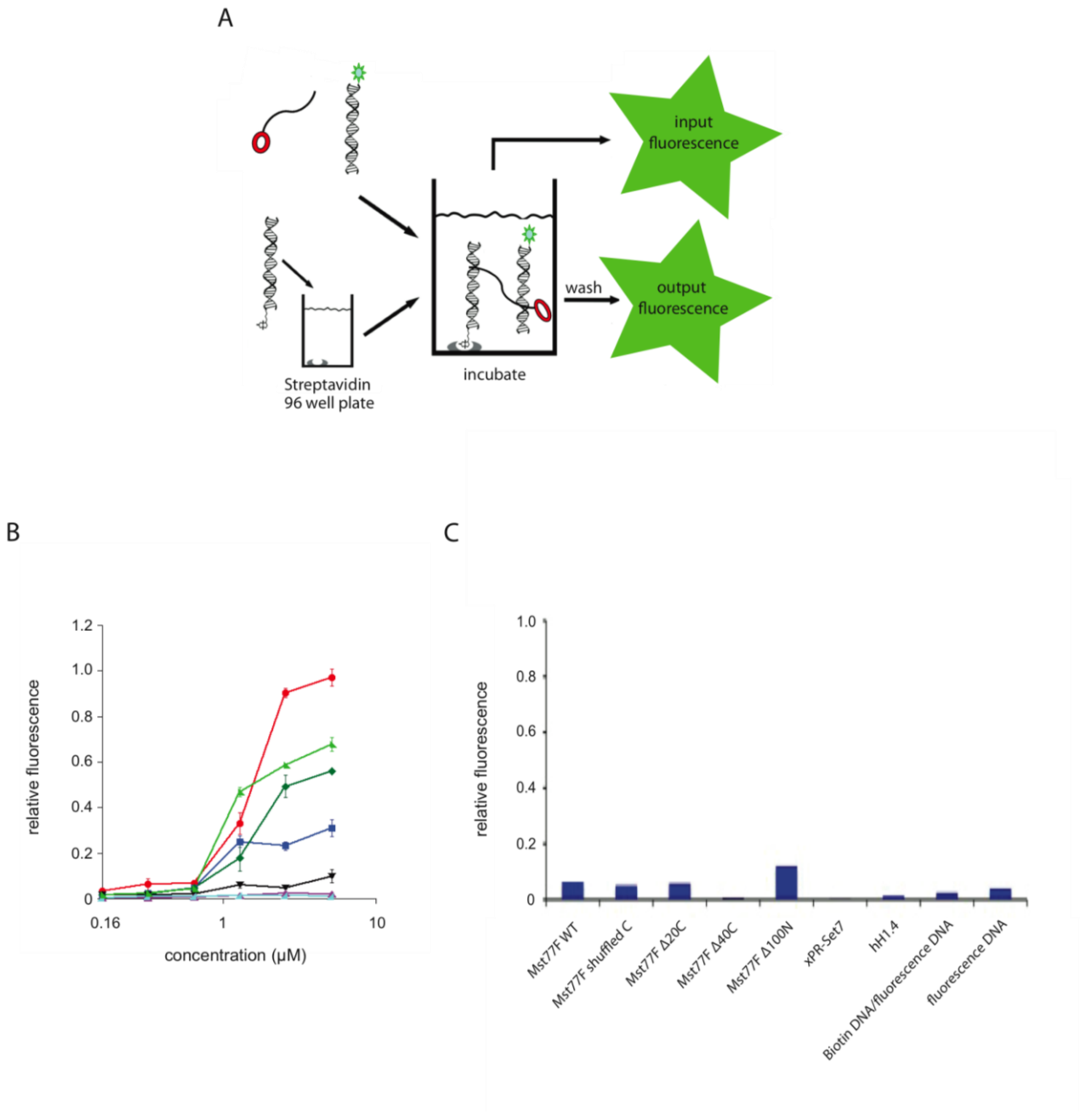


Fig.3.8 Mst77Fs N-terminus triggers quantitatively DNA aggregation

(**A**) Schematic representation of the experiment: 234bp DNA is immobilized in a 96 well plate. The same amount of fluorescine DNA is added to the well and addition of protein induces inter DNA cross linking. Retained DNA due to protein cross linking is quantified by plotting the input vs,. output ration of the fluorescence signal before and after washing. (**B**) Experiment as outlined in A. Mst77F wild-type (red), Mst77F shuffled C (light green), Mst77F Δ20C (green), Mst77F Δ40C (black), MSt77F Δ100N (blue), hH1.4 (cyan) and xPR-Set7 (purple) have been tested in dependence of increasing protein concentrations from 0.16 µM to 5.12 µM in 2 fold increments. Plots show means of three independent experiments (**C**) Unspecific sticking of the fluorescence DNA to the well or in dependence of the indicated proteins.

shuffled C-terminal domain, Δ20C and Δ40C pointed towards an additional component within the CTD that is needed for highly efficient DNA aggregation. Even though hH1.4 and Mst77F Δ100N displayed similar effects on DNA in other experiments, hH1.4 had no steady cross linking potential in this assay. The impact of H1 proteins on DNA structure has been documented extensively and the predominant model is that H1 forms sandwich structures with DNA bridging DNA molecules through a layer of histone H1 (Jean O.Thomas, 1991). However, these studies were solely based on imaging techniques and did not include kinetic aspects that are an important aspect of this assay. It is conceivable that hH1.4 forms the mentioned complexes but that they dissociated in the consecutive washing steps. Also, xPR-Set7 had no cross linking potential and it is likely that previous interactions with DNA represented simple electrostatic adhesion.

In summary, Mst77F efficiently interconnected individual DNA molecules *in vitro*. This effect relied on the integral function of the C-terminal tail binding to DNA and the N-terminal protein multimerization domain as the Δ100N mutant showed strongly reduced cross linking. However, the impaired cross linking potential of Mst77F shuffled C-terminal domain strongly pointed towards an additional functional contribution of the CTD that exceeds DNA binding. Importantly, hH1.4 displayed no cross linking at all, emphasizing different functionality in DNA/chromatin organization.

3.7 Mst77F tightly compacts long DNA *in vitro*

Within the nucleus of spermatids the carrier of the genetic information is several orders of magnitude longer than the very short DNAs that have been used for the previous experiments. During the postmeiotic maturation process of *Drosophila* spermatids the DNA adapts a 200-fold higher compaction status compared to somatic cells. This state requires extraordinary tight arrangement of the DNA. According to the data I generated in the course of this study which clearly show Mst77Fs potential to aggregate and interconnect even short DNAs, I proposed a massive structural impact of the Mst77F wild-type protein on longer DNA. This effect

should be discriminative from the Δ100N multimerization mutant that displays less potential to link up individual DNA molecules, the Mst77F Δ60C binding mutant, hH1.4 and xPR-Set7.

To investigate the structural effects that Mst77F and the other proteins exert on longer, linear DNA I conducted again AFM with DNA that is 2434 bp in lenght. The structural properties of Mst77F wild-type DNA complexes were compared to DNA complexes that are shaped by the Mst77F Δ100N multimerization mutant and the Mst77F Δ60C DNA binding mutant. Histone hH1.4 as a member of the H1 family proteins that have been reported to structurally impact DNA/chromatin, was used to discriminate the Mst77F effect on DNA from the effects of well-studied structural proteins (Dootz et al., 2011; Shen et al., 1995). Previous experiments within this study showed clear functional differences between Mst77F and hH1.4 and on this basis I expected also structurally different complexes. Last but not least, xPR-Set7 that was not implicated in direct DNA binding, though displayed certain DNA binding properties in EMSA assays, was used as a further control. The structural properties of the DNA were imaged in the presence of 4-fold, 20-fold and 100-fold molar excess of the indicated proteins over DNA (Fig. 3.9).

At little molar excess of protein over DNA (4:1) the observed effects were limited to single molecules. Compared to free DNA imaging that showed DNA molecules in their extended but flexible conformation, a 4-fold molar excess of Mst77F wild-type structurally induced bending and looping of a fraction of molecules. Mst77F formed hubs that interconnected parts within one molecule (top row 1:4). Importantly this effect was intra-molecular and limited to individual molecules. A similar effect was observed for the Δ100N mutant that also exerted intra-molecular connections between parts of the same DNA molecule. DNA complexed with Mst77F Δ60C was indistinguishable from DNA “only” structures reflecting its mutant DNA association. Histone H1 has previously been shown to form fibrillar, branched and network-like structures (“tramtracks”) with DNA mediated through sandwiching of DNA molecules by the protein (Lu et al., 2009a; Lucius et al., 2001). At a molar ration 1 : 4 the fibrillar and branched species were seen with low frequency. The predominant form was indistinguishable from DNA “only” images. As expected for a non DNA binding protein xPR-Set7 DNA complexes showed similar results to DNA only at the lowest molar DNA protein ratio.

Increased molar DNA-protein rations of 1:20 uncovered the specific Mst77F effect on DNA that can be attributed to the N-terminal domain of the protein (Fig. 3.9 ; 1:20 column). Individual DNA molecules clustered and formed a “nucleus” with protruding DNA “tentacles” that seemed to be free of protein. Within this “nucleus” substructures were not recognizable suggesting a tightly packed, dense DNA protein entity. With lower frequency the second observable species was individual DNA molecules that displayed, by the majority high intra-molecular associations. Interestingly, the DNA structures built up by the Mst77F Δ100N mutant were very much alike the structures formed by hH1.4 suggesting a similar interaction mode with the proteins packaging DNA molecules in a parallel aligned manner. Both proteins converted individual DNA molecules into fibrillar, branched complexes consisting of several DNAs. The recorded differences between Mst77F wild-type, Mst77F Δ100N and hH1.4 nicely reflected the specific function that extend the simple charge mediated binding to DNA. Again, as predicted no significant differences between DNA incubated with Mst77F Δ60C, xPR-Set7 and DNA “only” species became obvious.

Further increase of the molar DNA protein ratios to 1:100 led to an amplification of the formed structures (Fig. 3.9 ; 1:100 column). The “nucleus” structures elicited by the Mst77F wild-type protein expanded their dimensions in the Z plane. Individual DNA molecules could not be detected anymore. A major change from scattered fibrillar, branched towards tightly arranged, network like structures was seen for the Δ100N mutant possibly reflecting the impact of an excess of flexible, highly charged peptide moieties on their counter-ion polymer. Even though hH1.4 has a similar charge density to the Δ100N mutant and showed previously similar results, the formed complexes with long DNA were distinct. At the highest monitored molar ratio fibrillar, branched conformations were still evident but the major species constituted spacious, parallel (and maybe lateral) interconnected DNA molecules with a small, central high-density entity. Fundamentally, this structure was very different from the structures built up by Mst77F wild-type. As previously shown for the lower molar ratios there was also no effect of the Mst77F Δ60C protein on DNA structure at the 1:100 molar ratio. However, a 100-fold molar excess of xPR-Set7 led to branched interconnection of DNA molecules distinct from structures seen for lower molar ratios

of Mst77F Δ100N or hH1.4. I reasoned this is attributed to the attractive charges of the protein towards DNA.

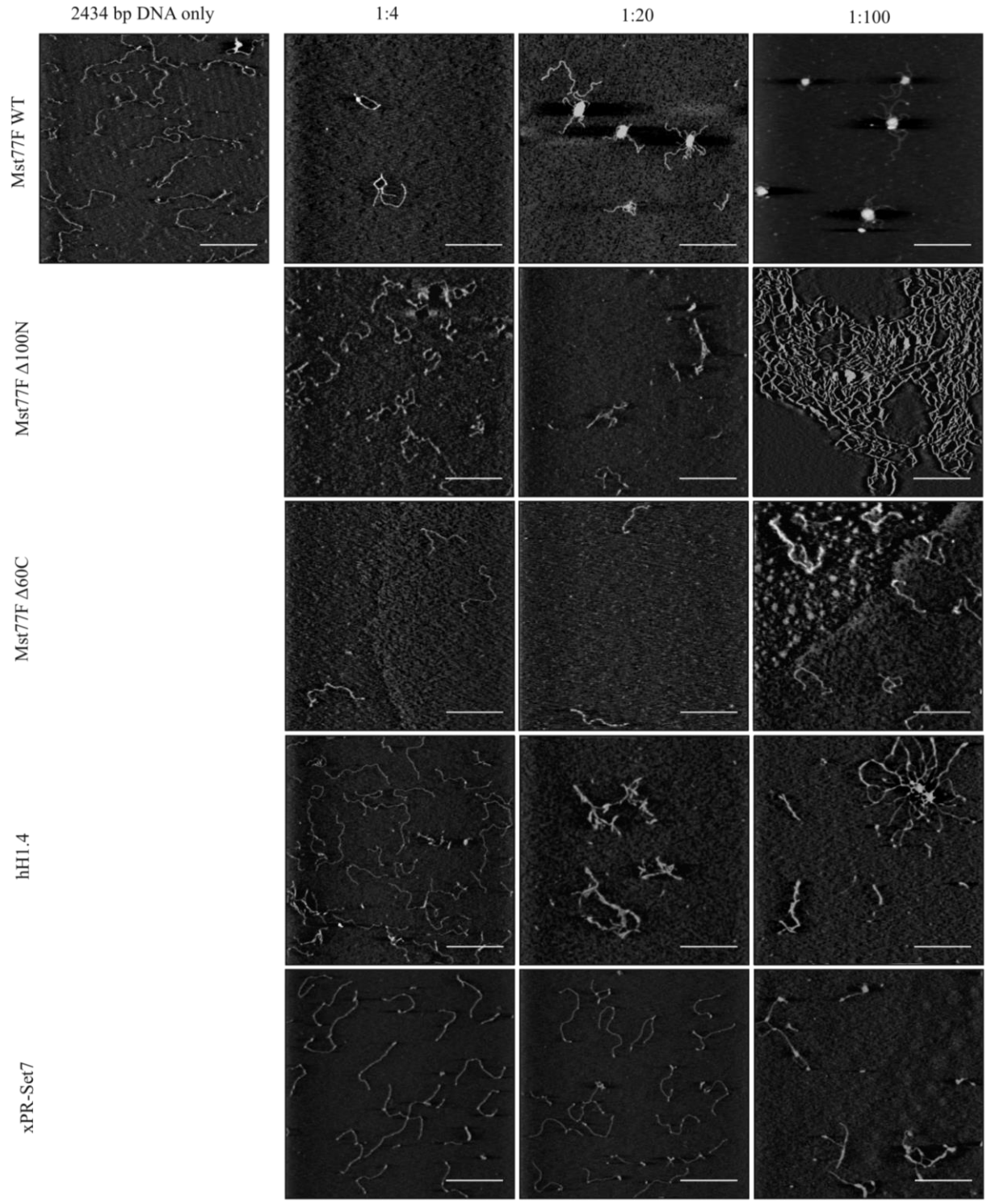


Fig. 3.9 AFM uncovers structural differences caused by Mst77F from effects triggered by other charged proteins

Mst77F wild-type, Mst77F Δ100N, Mst77F Δ60C, hH1.4 and xPR-Set7 were complexed with 12 x 200 x 601 DNA at 4-fold, 20-fold and 100-fold molar excess of the protein over DNA. Images were recorded in

tapping mode. The scale bar represents 500 nm. The height scale of all images is 3 nm but for Mst77F wild-type 1:100: due to illustration the height scale was increased to 20 nm.

Taken together, Mst77F tightly compacted DNA *in vitro*. In very good agreement with the centrifugation fractionation and DNA crosslinking assays this effect can be attributed to the N-terminal region of the protein that constitutes a multimerization module. Consequently, mutants that were impaired in DNA binding show no structural effect on DNA. The Mst77F mutant that lacked the N-terminus induced network like complexes with long DNAs. Importantly, histone H1 proteins that have been reported to elicit structural effects on DNA showed different structures.

3.8 Mst77F inhibits transcription *in vitro*

The postmeiotic spermatid maturation requires timely onset of sequential processes that lead to a fundamental reorganization of the genome. Mst77F has been suggested to participate in the compaction of DNA in the transition from nucleo-histone to a protamine organized genome (Rathke et al., 2010). My *in vitro* data clearly identified Mst77F as a potent compactor of DNA, corroborating its proposed *in vivo* function. However, immunocytology experiments on developing spermatids clearly showed a heterogenous nuclear distribution of Mst77F questioning its exclusive role as a structural protein involved in nuclear condensation (Rathke et al., 2010). Transcriptional processes in *Drosophila* postmeiotic spermatids are thought to be inhibited. Though, recent work identified elevated transciption from some loci (Barreau et al., 2008; Vibranovski et al., 2010; Vibranovski et al., 2009). This raised the possibility that Mst77F functions as a general repressor of transcriptional processes.

To test the Mst77F impact on transcriptional processes, I applied an *in vitro* transcription system that is based on the plasmid pG5ML (Fig. 3.10 panel A) (An, 2004). This plasmid contains five sequential Gal4 binding sites, which are used to recruit the Gal4-VP16 transcriptional activator (Sadowski et al., 1988). The Gal4 sequences are followed by a strong viral adML-promotor and a G-less transcriptional cassette. The transcripts of this cassette selectively can be recovered after RNAse T1 digestion that only cleaves after G - nucleotides. The experimental scheme is

outlined in Fig. 3.10 panel B and described in detail in Material and Methods paragraph 2.11.5. To the DNA, the Gal4-VP16 activator was added and allowed to bind as indicated. Mst77F was added and the reactions were again incubated as shown. Then, HeLa nuclear extract was added, which contained the RNA polymerase II complex, general transcription

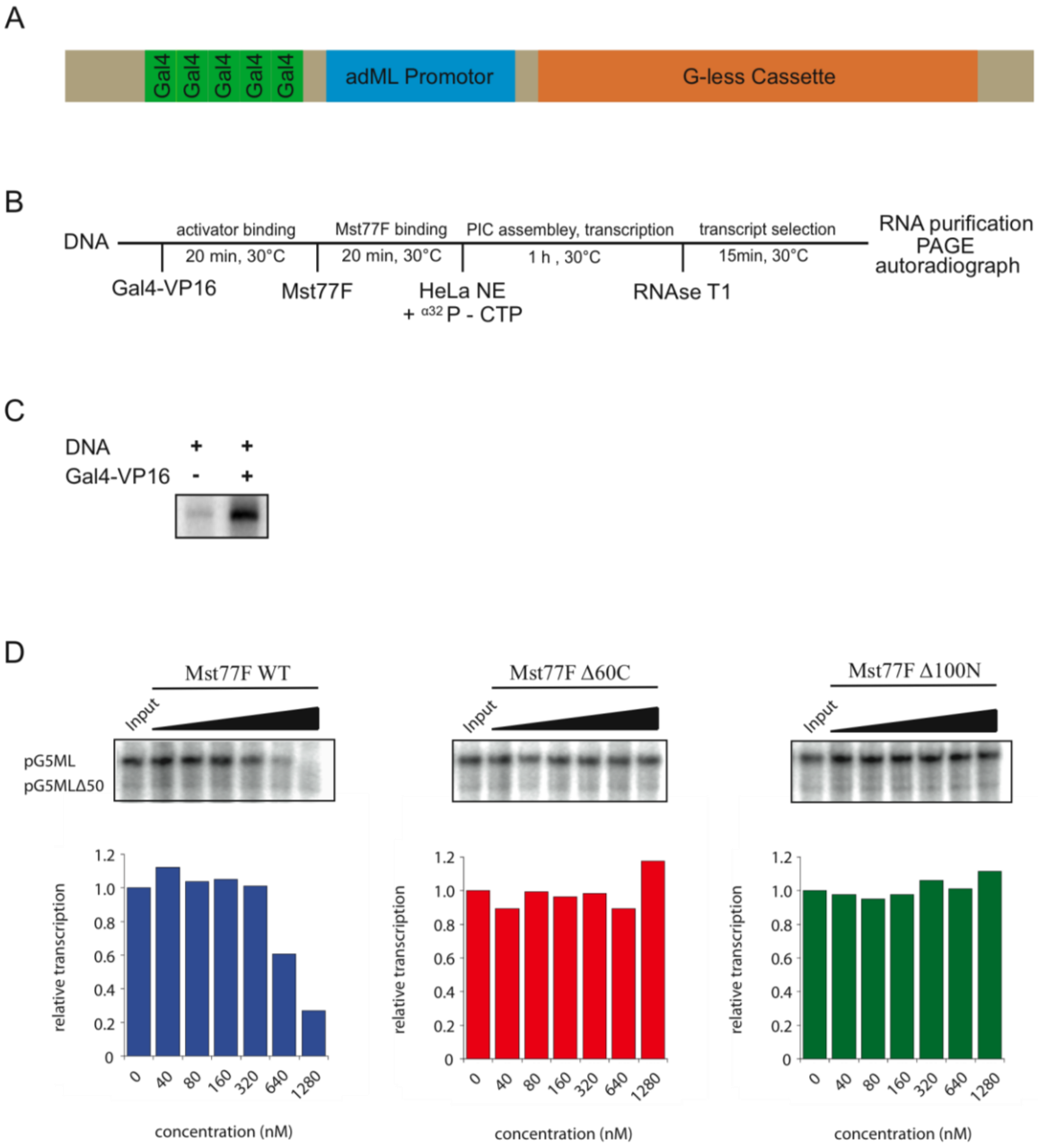


Fig. 3.10 Mst77F inhibits transcription *in vitro*

(**A**) Schematic representation of the pG5ML plasmid transcriptional cassette with the Gal4 binding sites, the strong adenovirus major late (adML) promotor and the G-less transcription unit. (**B**) The flow scheme

of the transcription experiment with DNA as transcriptional template: the system is entirely composed of recombinant components until addition of the HeLa nuclear extract. (**C**) Transcription reactions require an activator (Gal4-VP16): transcripts are only abundant in reactions that incl. Gal4-VP16. (**D**) Transcription experiments with MSt77F wild type (left), Mst77F ∆60C (middle) and Mst77F ∆100N (right): Transcription reaction was carried out according to the scheme in (B). Mst77F proteins were titrated from 40 nM up to 1.28 µM final concentrations in 2-fold increments. Purified RNAs were run on 5% urea TBE-PAGE gels. The gel was dried and exposed for 1h to a "storage-phosphor" visualization screen. The transcript band intensities were integrated with ImageJ and normalized towards the loading control pG5ML ∆50. The signal ratio was plotted with the Kaleidagraph software.

factors etc. that are necessary for transcription. Along with the nuclear extract a radioactively labeled nucleotide was included and the transcription was allowed to proceed for 1h. Unspecific transcripts were removed by RNAseT1 treatment. The G-less cassette transcript was purified and resolved on an urea-PAGE gel and visualized by autoradiography. In order to account for unequal recovery of transcripts, a control transcript pG5ML∆50 that was co-transcribed and not incubated with Mst77F served as purification control.

To test the functionality of the complex transcription system I compared two reaction conditions. One transcription reaction was conducted in the presence of the transcriptional activator Gal4-VP16. The other one missed Gal4-VP16. As expected, in the presence of a transcriptional activator RNA transcripts could be observed (Fig 3.10 panel C). Therefore, the impact of Mst77F on transcription processes could be tested.

Clear structural differences in DNA Mst77F complexes could be observed for Mst77F wild-type, Mst77F ∆60C and Mst77F ∆100N. These proteins were tested for their capability to suppress transcription (Fig 3.10 panel D). In this experiment, Mst77F wild-type caused a dose dependent transcription repression similar to the dose responses seen with other experiments with basically no transcripts detectable at 1.28 µM protein concentration. As expected from other experiments, Mst77F ∆60C had no impact on the transcription process. No reduction in transcripts was observed. Surprisingly, Mst77F ∆100N that formed network like structures with DNA was also not capable to reduce the transcription within the concentrations tested.

Take together, the structurally different effects of MSt77F wild-type, ∆60C and ∆100N on DNA functionally also translated into effects that effect DNA metabolism, more precisely transcription. Whereas the wild type protein was a potent repressor both deletion mutants, effecting either DNA binding or multimerization of the protein failed to inhibit transcriptional processes.

3.9 Mst77F displays similar function in the context of recombinant chromatin

Previous experiments were solely carried out with “naked” DNA as Mst77F interaction partner because previous immunocytology experiments in *Drosophila* showed that Mst77F mainly exists when histone proteins already have been degraded (Rathke et al., 2010; Sunil Jayaramaiah Raja, 2005). However, there is a short overlap of both protein signals *in vivo* that might functionally be important for the transition from a nucleo-histone based DNA structure towards a highly condensed protamine organization. To test the hypothesis that Mst77F is an important factor for the progression of postmeiotic spermatid maturation, I conducted experiments that monitored its competence to bind to recombinant chromatin. I was interested to characterize its interaction mode with nucleosomes in comparison to histone H1, its suggested homolog. Mst77Fs potential to cause histone eviction from recombinant oligonucleosomal arrays was investigated along with AFM of the resulting structures. Finally, I looked at the previously reported repressive function on transcriptional processes in the context of recombinant chromatin.

The recombinant chromatin was assembled according to material and methods paragraph 2.11.4 from linear 2434 bp of DNA named 12x200x601 or 147 bp mononucleosome core particle templates. EMSA binding experiments with 12x200x601 oligonucleosomal arrays and the Mst77F wild-type protein showed the same binding behavior as the EMSA experiments with the DNA dodecamers (Fig. 3.11 panel A left). Clear band shifting was not observable and the dose response was very similar. At 0.64 µM protein most of the chromatin already disappeared from the gel with complete clearance at 1.28 µM. Interestingly within a similar experiment but with nucleosome core particles (147 bp DNA complexed with histone octamers) that have no linker DNA Mst77F showed the exact same effect. This suggests that the proposed unspecific electrostatic interaction mode holds also true for Mst77F interaction with DNA in the context of chromatin (Fig. 3.11 panel A right). Histone H1 binds the nucleosome at the dyad axis where the DNA enters/leaves the histone octamer and additionally binds approximately 20 bp of the linker DNA between nucleosomes with its C-terminal domain (Allan et al., 1980). Mst77F was suggested to be a distant H1 relative on the basis of sequence alignments that indeed showed a similarity in the CTD of both proteins. My *in vitro* experiments uncovered fundamental

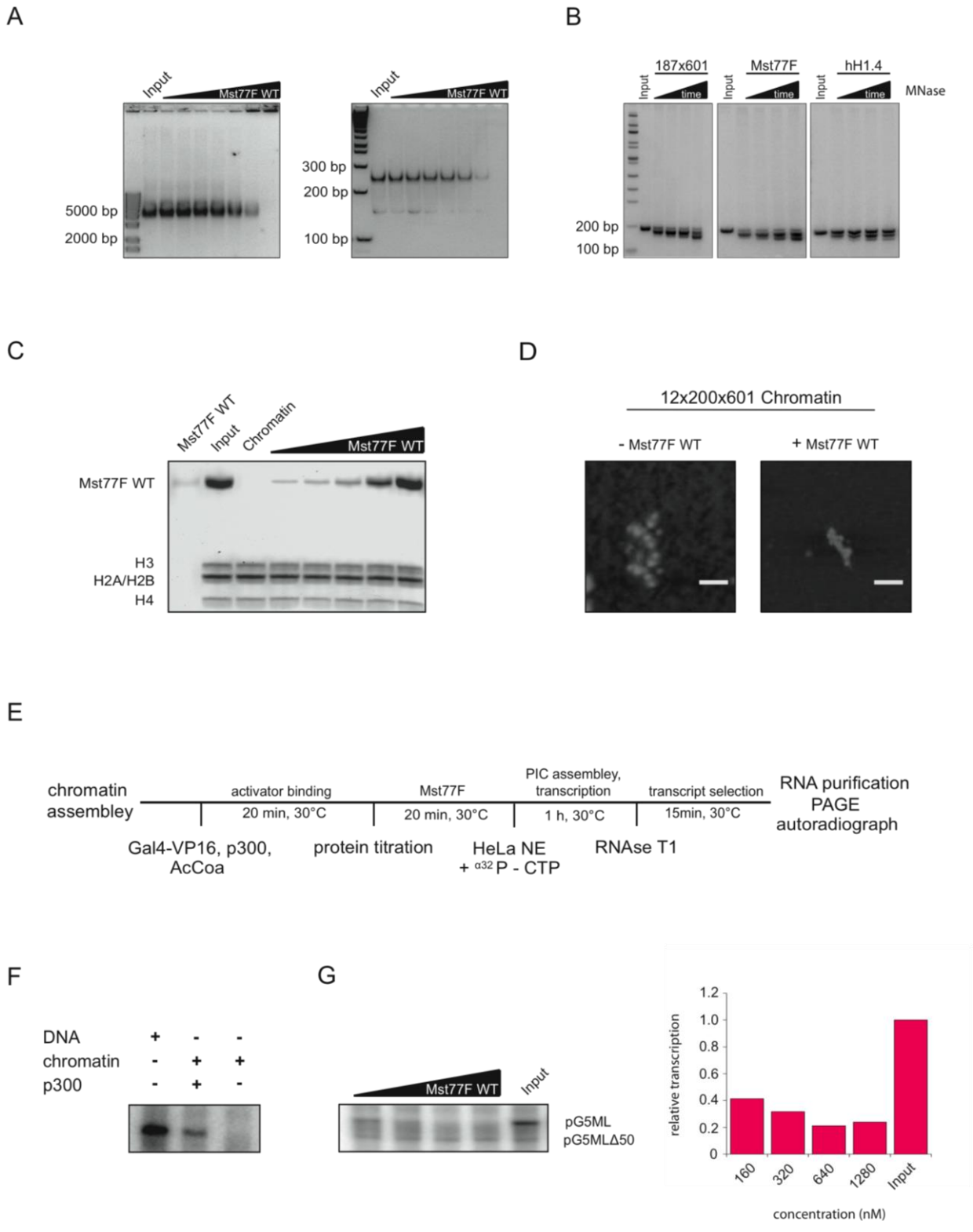


Fig. 3.11 Mst77F effects on chromatin are similar to the observed effects with DNA

(**A**) EMSA experiments with Mst77F wild-type and 12x200x601 oligonucleosomes (left) or nucleosome core particles (right): the protein was titrated from 0.02 µM to 1.28 µM in 2-fold increments. The samples were run at 80 volts for 90 min on 0.6% TB-Agarose in the case of 12x200x601 chromatin or on a 5% TB-PAGE gel for the nucleosome core particles. Both gels were stained with ethidium bromide for 15 min. (**B**) MNase digest of 187x601 mononucleosomes alone (left), in the presence of Mst77F (middle) or hH1.4

(right). Input means no MNase in the reaction. 200 mU of Mnase were incubated for 1 min, 2 min, 4 min or 8 min with the mononucleosomes or mononucleosome Mst77F and hH1.4 complexes, respectively, as indicated by the time bar. The DNA was phenol:chlorophorm:isoamylalcohol extracted and ethanol precipitated. For fragment size analysis DNA was run on a 5% TBE-PAGE gel for 1h @ 80 volts. (**C**) Mst77F chromatin coprecipitation: 12x200x601 oligonucleosomes were incubated with increasing concentrations of Mst77F wild-type protein from 0.08 to 1.28 µM in 2-fold increments. Clustering of chromatin complexes was triggered by the addition of 5 mM $MgCl_2$. Complexes were pelleted by centrifugation and analyzed on 15% Tris-Glycine SDS-PAGE. (**D**) AFM with 12 x 200 x 601 chromatin and Mst77F wild-type: The addition of the protein causes intramolecular compaction at a 4-fold molar excess of Mst77F over DNA. Images were recorded in tapping mode and scale bars represent 100 nm (**E**) Outline of the order of addition of components during the transcription experiment with chromatin being the transcriptional template. (**F**) Transcription from chromatin templates requires histone acetylation: Gal4-VP16 was present in all reactions. Autoradiograph of the resolved RNAs purified after the transcription reaction on a 5% Urea TBE-PAGE gel. (**G**) Mst77F also inhibits transcription from pG5ML chromatin templates. Autoradiograph of purified RNAs after transcription from pG5ML chromatin and increasing Mst77F concentrations. RNAs were resolved on a 5% Urea TBE-PAGE gel and the gel was dried. Transcript band intensities were integrated and normalized towards the pG5MLΔ50 signal with the help of the ImageJ software. Bar plot was generated with the Kaleidagraph software.

functional differences between both proteins. However, I did not work with testis specific H1 and that might account for the different observations. To investigate the localization of Mst77F on the nucleosome I performed MNase protection assays with 187x601 mononucleosomes, mononucleosomes complexed with hH1.4 and mononucleosomes complexed with Mst77F (paragraph 2.11.6) (Fig. 3.11 panel B). Without protection of the linker DNA by a protein MNase rapidly degraded the overlapping DNA to nucleosome core particle size of 147 bp within the time course of 8 min. Addition of hH1.4 prior to the MNase resulted, as expected, in a chromatosome stop with DNA length of approximately 167 bp caused by H1 protection. Preincubation of the mononucleosomes with Mst77F showed a Mononucleosome only like digestion pattern with only slide protection. This result strongly suggested that Mst77F does not bind specifically to the nucleosome in a H1 like manner.

Mst77Fs capability to evict histone proteins from the DNA was tested in chromatin co-precipitation assays (Material and Methods paragraph 2.12.3)(Schwarz et al., 1996). Recombinant 12x200x601 chromatin arrays were incubated with increasing amounts of Mst77F. $MgCl_2$ was added to induce internucleosomal clustering. Along with the clustered oligonucleosomes associated proteins were pelleted by centrifugation. The Mst77F and histone recovery as a result of increasing Mst77F concentrations were visualized by SDS-PAGE (Fig. 3.11 panel C). The addition of increasing amounts of Mst77F showed no reduction in the recovered amounts of histone proteins compared to the absolute amounts in the input lane and the

precipitation of 12x200x601 chromatin alone (chromatin lane). Therefore, under the applied conditions Mst77F exerted no histone eviction.

Since Mst77F binding to chromatin could be shown by EMSA and co-precipitation experiments I wondered weather the structural effects seen for its interaction with DNA also hold true for its interaction with recombinant chromatin. 12x200x601 chromatin alone and with a 4-fold molar excess of Mst77F was imaged by AFM (Fig. 3.11 panel D). At the applied conditions with 50 mM NaCl the chromatin fiber appeared loosly bend but single nucleosomes were recognizable (left). The addition of Mst77F triggered intranucleosomal compaction (right). Remarkably, the effect of a 4-fold molar excess of Mst77F appeared stronger compared to its effect on DNA at the same molar ratio.

Mst77Fs potential to inhibit transcriptional processes from DNA templates *in vitro* could be demonstrated (paragraph 3.8). The eviction of histones in the transition phase from histone to protamine organization during spermatid maturation may lead to a global leakage of transcription. To counteract this effect Mst77F might function as a structural repressor that holds up the repressed state until complete reorganization of the genome. Its inhibitory effect on chromatin transcription was tested again with the *in vitro* transcription assay. The experiment was basically carried out as described before but had to be adapted in order to allow the transcription from a chromatin template that is inherently silenced. Figure 3.11 panel E shows the flow of the experiment. Along with the transcriptional activator Gal4-VP16 the coactivator p300 (histone acetyl transferase) and its substrate acetyl-CoA were added in the first incubation step. In this set up Gal4-VP16 recruits the added p300 that in turn acetylates the histones. Additionally Gal4-VP16 also recruits chromatin remodeling complexes required for ongoing transcription from the nuclear extract. The inherently silenced state that can be activated by p300 mediated acetylation of histones is shown in figure 3.11 panel F. That showed that the system can actually be activated and that observed silencing effects are indeed mediated by Mst77F. In the subsequent transcription experiment Mst77F displayed a stronger inhibitory effect compared to the transcription from DNA templates. At 160 nM Mst77F only 40% of the transcripts observed for the control reaction without Mst77F are detected. At 640 nM no transcripts are detectable anymore (Figure 3.11 panel G).

Taken together, Mst77F also interacted with recombinant chromatin arrays. This interaction seemed to be similar to the binding of short DNAs. Additionally, mononucleosomes without linker DNA were bound equally well emphasizing the ionic character of the interaction. Unlike H1 Mst77F showed no association preference for the nucleosomal dyad axis at the entry/exit point of the DNA suggesting a functional distinction also in the chromatin context. Under the applied conditions Mst77F was not capable to evict histones from oligonucleosomal complexes suggesting a function after histone removal. The strong Mst77F mediated compaction of DNA was also seen for chromatin templates even though the Mst77F effect seemed to be stronger at equal molar ratios. Finally, Mst77F also inhibited transcription from chromatin templates but again showed stronger inhibition at lower concentrations. The latter observations possibly can be accounted to the special physiochemical properties of the chromatin fiber that enhance the Mst77F function.

4. Discussion

A fundamental conserved process of evolution is the sexual reproduction of organisms. In higher eukaryots this involves the fusion of the male and female genomes along with their respective gametes in a process described as fertilization. Male germ cells, the spermatids undergo a series of developmental processes in order to be capable of fertilizing the female oocyte. This involves mitotic as well as meiotic divisions and finally results in a superior compaction of the sperm DNA along with substantial morphological rearrangements, a process referred to as spermatid maturation.

Although the *Mst77F* gene has been discovered almost two decades ago and the spaciotemporal expression pattern of the protein has been investigated, its function in the process of spermatid maturation is not resolved. The very few existing studies on the protein suggest an involvement in DNA compaction and nuclear shaping but by what mechanism is unclear. Due to limited experimental tools *in vivo* no prove for these hypothesis could be demonstrated. To circumvent the existing *in vivo* limitations I characterized the molecular organization of Mst77F and its interaction with DNA by *in vitro* techniques. I identified Mst77F to be a protein of bivalent structural functionality. The intrinsically unstructured C-terminus binds DNA unspecifically with high affinity in a charge-mediated manner. This interaction induces homo-multimerization of Mst77 through its N-terminus and results in tight compaction of DNA. Importantly, these effects are distinct from histone H1, the proposed homolog. My results support the suggested function of Mst77F as a structural protein involved in the condensation of the *Drosophila* sperm genome during postmeiotic spermatid maturation.

4.1 The implication of intrinsically unstructured, charged domains in DNA binding

The spatiotemporal expression pattern of Mst77F in *Drosophila* postmeiotic spermatid maturation has been investigated (Rathke et al., 2010; Sunil Jayaramaiah Raja, 2005). The protein becomes expressed prior to tight condensation of the

genome. Due to its sequence similarities of the C-terminal domain (CTD) towards histone H1 family proteins, major structural proteins of cellular chromatin, it has been proposed to be part of the DNA compaction process. This requires a physical interaction with DNA in the first place. In my experiments I could show that Mst77F binds DNA *in vitro* via its highly charged C-terminal domain (Fig. 3.2). Protein structural analyses that were based on the Mst77F sequence and CD spectroscopy identified an intrinsically unstructured random coil conformation of the recombinant Mst77F C-terminus in solution (Fig. 3.1 and 3.7). Binding experiments with a Mst77F sequence shuffled CTD protein showed wild-type like binding affinity strongly pointing towards little structural contribution in the binding process. Additional binding experiments with DNA protected by histone octamers, distorted DNA and single stranded DNA resulted in wild-type like interactions underpinning the electrostatic binding determinant (Fig. 3.2, 3.3 and Tabl. 3.1).

On first sight this proposed interaction mode appears unusual. However, intrinsically unstructured domains often are involved in binding and macromolecular recognition processes (Dunker and Obradovic, 2001; Hansen et al., 2006). Within such domains an elevated level of charged residues is characteristic and suggests strong electrostatic contributions in binding events (Vucetic et al., 2003). In particular, it could be shown that proteins that alter chromatin folding and compaction frequently make use of long, charged, intrinsically disordered domains. Recent work demonstrated that the side specific repressive function of Polycomb Repressive Complex 1 that compacts chromatin and inhibits remodeling processes, relies on one of its subunits that interacts with DNA through an intrinsically unstructured, highly charged domain (Beh et al., 2012; Brown et al., 2011). Also, studies on global genome architectural proteins show that unspecific protein DNA interactions guided by electrostatic binding modes are common mechanisms within this group (Dow et al., 2000; Lu et al., 2009a; Luger et al., 1997).

In the nucleus of almost every eukaryotic cell the organization of DNA is predominantly achieved through core histone DNA complexes, the nucleosomes. These complexes are held together by a multitude of electrostatic interactions. Moreover, beyond the repetitive nucleosomal units the DNA is organized through additional proteins that physically interact with DNA and further affect the genome structure. Importantly, some of these auxiliary proteins share common

physiochemical properties that are also present in Mst77F: (a) they are in part unstructured and (b) additionally show highly basic domains of low compositional complexity. Histone H1, amongst others implicated in structural stabilization and compaction of chromatin, shows high affinity binding to the nucleosomal linker DNA largely mediated by its long unstructured basic C-terminus (Allan et al., 1980; Fang et al., 2012). Moreover, the HMG-D protein that is associated with condensed chromatin during developmental processes, binds unspecifically to DNA, in part via its highly charged, unstructured C-terminus (Dow et al., 2000; Ner et al., 2001).
The involvement of intrinsically disorder domains with high charge density seems to be a general mechanism in recognition and organization of DNA on the genome- and local level, respectively. Recently, unstructured domains were also implicated in protein-protein interactions and the observed physiochemical deviations across proteins may therefore reflect functionalized domains that convey dedicated effects. However, the rapid evolution of unstructured domains results in low sequence conservation and makes a functional prediction difficult (Coletta et al., 2010; Tompa, 2002).

4.2 Induction of structural elements in intrinsically unstructured domains is functionally relevant

A proposed feature of intrinsically unstructured domains is their transition from disordered to ordered conformations (Uversky, 2002). This process is coupled to recognition of their respective binding partners. For histone H1 proteins it could be shown that the C-terminal domain adapts α-helical structures upon binding to DNA. However, the conformational stabilization does not occur across the long C-terminal domain but is limited to two short, discontinued sub-domains. The formation of these local secondary structural elements is essential for protein function even though the initial protein DNA interaction is considered to be purely ionic (Clark et al., 1988; Lu et al., 2009a; Lu and Hansen, 2004; Vila et al., 2001).
The suggested function of Mst77F in sperm DNA compaction makes it conceivable that Mst77F compacts DNA by a similar mechanism as histone H1 influences chromatin structure. This essentially involves the structural refinement of H1 CTD sub-domains upon DNA binding through electrostatic interactions (Fang et al., 2012).

Whereas the electrostatic contribution in the binding reaction could be demonstrated for Mst77F, an induction of secondary structural elements became not apparent in my experiments. The thermodynamic parameters of the Mst77F DNA interaction determined by microcalorimetry experiments could underpin the ionic interaction mode. But, ITC did not hint the formation of structural elements in the CTD as would be reflected by a positive entropy of the binding reaction. However, the high negative entropy value suggests a very flexible conformation of Mst77F, something that is also seen for histone H1 (Fang et al., 2012). A possible reason why I could not detect structural formation might be due to the small DNA dodecamers used in this experiment. The size of the continuous interaction interface might be critical for secondary structure formation.

However, the non-gradual loss in binding affinity observed for the transition from Mst77F ∆20C towards ∆40C does not correlate with the proportional decrease in the charge sum and might indeed reflect functional rearrangements upon DNA binding in the region between ∆20C and ∆40C. Interestingly, the Mst77F bioinformatic sequence analysis revealed two regions of low compositional complexity that can be correlated with this part of the protein. In comparing Mst77F wild-type proteins with amino acid shuffled CTD Mst77F no significant change in binding affinity was observed. However, experiments that addressed intermolecular cross linking of DNA by these proteins show less efficient interconnection of individual DNA molecules by the shuffled CTD Mst77. These observations point towards separation of molecular recognition and function. Importantly, in the histone H1 CTD secondary structure formation seems not to depend on the amino acid sequence. Rather the overall composition and the relative intra-molecular position towards other structural or functional reference points was found to be relevant (Lu et al., 2009a and references therein). By shuffling the last 70 amino acids of the Ms77F C-terminus the low compositional complexity regions received a distinct composition of amino acids. For the histone H1 CTD it could be shown that upon replacement of nonpolar amino acids in one of its structurally inducible regions, H1 effects on chromatin folding were partially disrupted. Similarly, the Mst77F shuffled CTD mutant was less efficient in interconnecting/aggregating individual DNA molecules. This finding might indeed point towards an induced secondary structure within the last 40 amino acids of the Mst77F CTD upon DNA binding.

Unfortunately, only little is known about the protein chemistry that underlies the connections between amino acid composition, intrinsic disorder, molecular recognition and folding. To what extend the CTD contributes to Mst77F function, effects that exceed molecular recognition, could not be clearified in this study. On the assumption that the Mst77F CTD is a intrinsically unstructured domain I propose a induced structure upon DNA binding. This structure might be necessary for high efficient interconnection of DNA and therefore DNA condensation during *Drosophila* spermatogenesis. Therefore, further experiments that address structural details within the Mst77F CTD and how they possibly impact Mst77Fs function are necessary to underpin these interpretations.

4.3 Structural aspects of Mst77F DNA complexes

AFM imaging of Mst77F and H1 DNA complexes identified major differences in their ability to alter DNA structures. hH1.4 induces so called “tramtrack” structures with long DNA (Fang et al., 2012; Jean O.Thomas, 1991), structures of parallel aligned DNA molecules sandwiched by H1 proteins. Mst77F in contrast induced massive nucleation of individual DNA molecules forming highly condensed structures. Importantly, this effect relied on the N-terminal domain of Mst77F and the DNA molecules protruding out of these “nuclei” are void of Mst77F. Bioinformatic analysis of the Mst77F sequence in conjunction with CD spectroscopy of Mst77F Δ110C identified predominant α-helical structures in this part of the protein. In particular, a coiled coil motif could be annotated. Chemical cross linking of the wild-type protein in presence of DNA revealed a stoichiometric multimerization pattern that was not seen without DNA or with the Δ100N mutant. In solution only little Mst77F dimer and no multimers could be detected suggesting either a weak coiled coil multimerization stabilized through immobilization of molecules on the DNA via the CTD or an induced cooperative multimerization mode upon DNA binding. The latter model seems more likely since AFM with long DNA and Mst77F clearly showed non-distributive binding of Mst77F along the DNA. Some regions displayed massive DNA condensation whereas others were free of protein and adapted extended conformations. Therefore,

the CTD interaction with DNA likely causes high affinity homomultimerization of Mst77F that is several orders of magnitude stronger than in solution.

Supportive data came from the centrifugation fractionation experiments with Mst77F and very short DNA. Mst77F and DNA dodecamers formed fast sedimenting high molecular weight complexes clearly visible in AFM. The formation of lattice like structures through simple electrostatic interactions can be ruled out since the Mst77F Δ100N mutant and hH1.4 did not show aggregation. However, for histone H1 a cooperative binding mode to DNA could be demonstrated and the interaction interface was mapped to the H1/H5 globular domain (Clark and Thomas, 1986; Jean O.Thomas, 1991). Importantly, although H1 cooperativity has been reported aggregation of the short DNA dodecamers was not observed. If both proteins, Mst77F and H1 induce altered DNA structures through cooperative binding, and both need their globular domains what are the molecular determinants for the observed major differences? **1.** Cooperative DNA binding between H1 molecules is mediated by the globular domain and the contribution of the CTD is unclear (Clark and Thomas, 1986; Jean O.Thomas, 1991). In contrast, binding of Mst77F to DNA takes place through its CTD and induces cooperativity that likely involves the N-terminus (Fig. 3.6 and 3.8). The structural requirements for cooperativity are unknown and might involve structural adaptation of the CTD upon DNA recognition. From my experiments it became not clear whether the Mst77F CTD undergoes structural stabilization even though the protein mediated DNA aggregation assay indicates functional relevant contribution of the Mst77F CTD. This might indeed be due to induced structures in this region of the protein **2.** Additionally, the association with DNA seems to result in differential kinetic stability of the different protein DNA complexes. Surprisingly, even though H1 proteins strongly associate with DNA and also form aggregates, the complex is not stable. In the protein induced DNA cross linking assay extensive washing removed all fluorescence DNA in the wells were hH1.4 was tested for its cross-linking potential. This is actually in good agreement with previous studies that identified an average H1 residence time of 100 s *in vivo* (Stasevich et al., 2010). Mst77F Δ100N DNA complexes displayed kinetic stability even though the overall recovery was largely reduced compared to the wild-type DNA complexes. The suggested kinetic stability of the protein DNA complex could

also be shown by SPR experiments with the wild-type protein and DNA dodecamers that showed virtually no dissociation (data not shown).

According to the binding, thermodynamic and structural data I propose the following model how Mst77F interacts with DNA:

In solution the Mst77F CTD is flexible and extended due to electrostatic repulsion between positively charged amino acids. In this conformation protein protein interactions via the N-terminus are rare and the protein samples its environment for interaction partners. Upon recognition of DNA by the CTD, Mst77F undergoes a conformational change that increases its affinity for self-interactions dramatically. The CTD on it own can aggregate multiple individual DNA molecules in a one-dimensional manner. However, spreading of aggregation towards highly condensed structures is mediated by cooperative protein multimerization (Fig 4.1).

Two scenarios how cooperativity is induced are conceivable. The first scenario implies a self-inhibitory mechanism. The flexible extended CTD folds back on the N-terminus and blocks the self-interaction domain. The recognition of DNA that constitutes a longer continuous interaction interface and is therefore bound with higher affinity (Fig. 3.3) liberates the N-terminus that is now prone to multimerization. The second possibility involves a functional structure in the CTD that becomes induced upon DNA recognition and in turn interacts with the α-helical N-terminus of a different Mst77F molecule. I favor the second model for the following reasons: (1) The stabilizing charged amino acids within the coiled coil are all anionic and homomultimers likely become destabilized. This is probably what is seen for Mst77F in solution where only little dimer can be detected. However, induced folded structures within the cationic CTD that are capable to interact with the N-terminal coiled coil will therefore be stabilized. (2) For histone H1 it could be shown that a change in CTD amino acid composition partially disrupts its function likely through alterations in the induced structures that convey functionality upon DNA recognition. For the Mst77F shuffled CTD protein a decrease in protein induced DNA cross-linking capacity could be demonstrated (Fig. 3.8). The importance of a dedicated amino acid composition for function supports the idea of inducible secondary structures in the Mst77F CTD that interact with the N-terminal part of the protein to form multimers.

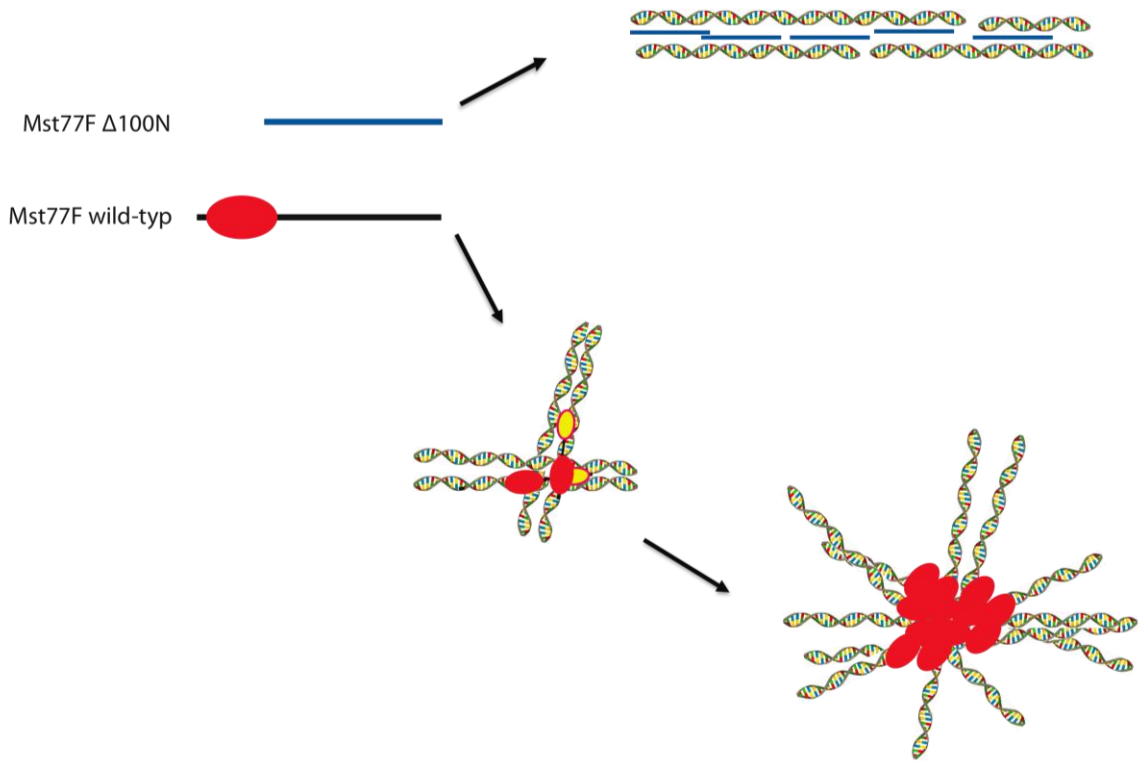


Fig. 4.1 Proposed model of DNA compaction by Mst77F

Mst77F wild-type and Δ100N bind equally well to DNA driven by electrostatic interactions. This interaction is stable and one-dimensional interconnection of individual DNA molecules occurs. In the wild-type protein situation, binding evokes conformational changes of the CTD that now can interact with the N-terminus of other MSt77F molecules. N-terminal coiled coil domain (red sphere); CTD (blue/black line); induced structure in the CTD (yellow sphere with red frame).

4.4 The biological role of the S149T mutant

β_2-tubulin is the major β-tubulin isoform in *Drosophila* spermatids. It is involved in nuclear shaping (Fuller et al., 1987; Kemphues et al., 1982). The Mst77F mutant ms(3)nc3 has been identified as none-complementing mutation in β_2-tubulin interaction screens. The protein is characterized by a single amino acid substitution replacing serine at position 149 by threonine leading to defective nuclear shaping during postmeiotic spermatid maturation. However, the spaciotemporal expression pattern of Mst77F ms(3)nc3 is normal. Importantly, in young elongating nuclei β_2-tubulin localizes to the perinuclear space with accumulated nuclear Mst77F beneath this area (Rathke et al., 2010). Within this mutant background the DNA condenses normally in the transition from histone based towards protamine organisation. Mst77F

ms(3)nc3 is associated with the DNA but the spermatids do not undergo elongation. In good aggrement with this *in vivo* observation, Mst77F S149T behaved similarly as the wild-type protein in DNA binding and aggregation assays (data not shown) confirming the intact protein DNA interaction interface and condensation function *in vitro* (Fig 3.2). The CTD of Mst77F is likely intrinsically unstructured and shows physiochemical similarities towards the histone H1 CTD. Along with the proposed DNA binding and structural functions, the histone H1 CTD exhibits interface functions in histone H1 protein-protein interactions (Kim K, 2008; Widlak P, 2005). Interestingly, β_2-tubulin is proposed to be unstable and in spermatids of sea urchins axonems (microtubule structures of the sperm tail) of the a tubulin stabilizing function of histone H1 could be shown suggesting a similar function of Mst77F in the process of nuclear shaping during spermatid development (Multigner et al., 1992). Supportive data comes from the observation that β_2-tubulin is less abundant in later stages of spermatid maturation. The stabilization likely would involve steady physical interaction with β_2-tubulin and an interaction interface involving serine 149. The mutation towards threonine would abolish the interaction leading to tubulin degradation. However, this hypothesis suggests a dual function of Mst77F: (a) DNA condensation and (b) nuclear shaping by tubulin stabilization. Whether this involves an orchestrated dual binding of Mst77F to DNA and to β_2-tubulin is elusive. Additional experiments will have to clearify if both proteins interact in the first place.

4.5 Mst77F association with chromatin in vivo

Mst77Fs nuclear localization *in vivo* is detectable at the onset of the young elongating nuclei stage. At this time histone proteins are abundant and might interfere with Mst77Fs DNA association. As expected Mst77F also bound recombinant chromatin *in vitro*. In my experiments this binding was conceptually different from histone H1 binding since no association preference for the nucleosome dyad axis at the entry/exit point of the DNA could be found in MNase protection assays. However, Mst77F bound nucleosome core particles suggesting that the association is random and also driven by electrostatic mechanisms. *In vivo* this might be different due to chaperone proteins that guide Mst77F deposition similar to what

could be shown for histone H1 (Finn et al., 2008). Though, the differences between Mst77F and histone H1 in their molecular organsization translate into distinct functionality in interactions with DNA *in vitro*. The diverging interaction mode with recombinant chromatin might also reflect specialized functions in the context of chromatin. In *mus musculus* and *homo sapiens* the Hils1 protein is proposed to be a spermatid specific linker histone-like protein. The spaciotemporal expression pattern of Mst77F and Hils1 is related but not coherent. Hils1 disappears at the stage of elongating nuclei whereas Mst77F is a component of mature spermatids. Unfortunately, also only very little is known about Hils1 but *in vitro* it also binds mononucleosomes in a H1 like manner and unlike Mst77F produces chromatosome stops (Yan et al., 2003). On the assumption that related processes are guided by similar mechanisms it is unlikely that Mst77F and Hils1 represent homologs.

The distinct binding mode from H1 *in vitro* and the global colocalisation with the core histones *in vivo* functionally might reflect a Mst77F role in the histone eviction processes. Mst77F first localizes to chromatin concomitantly with histone H4 hyper-acetylation in young elongating nuclei. Hyper-acetylation is proposed to be necessary but not sufficient for histone eviction (Rathke et al., 2007). *In vitro* Mst77F was not capable to evict histone proteins from unmodified recombinant chromatin raising the possibility that acetylation might be a prerequisite. However, the direct eviction of histone proteins by Mst77F is rather unlikely since most of the histones are removed in later stages of spermatid maturation between the early and late canoe stage. Conceivable is a scenario that involves Mst77F as a recruiting molecule. This function could involve the N-terminal coiled coil domain as well as the suggested induced structure in the CTD as interaction domains.

In early canoe stage nuclei the chromatin has been primed for histone eviction by global ubiquitinylation and acetylation and DNA bound Mst77F could function as a platform for executing factors. This would involve enzymes that introduce DNA nicks and chromatin remodelling comlexes which facilitate histone expulsion (Rathke et al., 2007). For histone H1 it could be shown that it recruits apoptosis specific nucleases via its CTD and stimulates DNA cleavage in programmed cell death (Widlak P, 2005). A similar function of Mst77F in the process of histone removal by recruiting DNA nicking enzymes is therefore conceivable and would make Mst77F essential for the transition from nucleo-histone towards protamine organization.

Furthermore, it is a generally accepted paradigm that transcription is silenced during postmeiotic spermatid maturation in *Drosophila* whereas in mammals transcription has been reported. However, recent studies challenged that point of view by providing evidence for solid ongoing transcription. Importantly, this transcription was found to occur during chromatin remodeling (Barreau et al., 2008; Vibranovski et al., 2010). The functional implications of these transcripts and their products are largely unknown. However, a repressive function of Mst77F on transcription could be demonstrated *in vitro.* It might be that Mst77F localization "marks" regions of silent chromatin whereas Mst77F free regions are prone to transcription during remodeling processes. Whether Mst77F exerts its repressive effect on its own or represents the platform for interacting proteins that in turn execute their repressive function is unclear. A genome wide Mst77F localization study and correlation of these data with mRNA expression profiles could give answers to this question.

The apparently oversimplified *in vitro* situation where Mst77F strongly condenses DNA upon recognition is not found *in vivo.* The initial chromatin association in early canoe shaped nuclei does not instantly cause condensation of the genome. According to the model that I proposed in the previous paragraph, Mst77F might be structurally inactivated - that is the incapability to multimerize. This effect might be due to an interaction partner (like a chaperone) that binds to the induced structure in the CTD upon DNA recognition or stays associated also after deposition. Again, for histone H1 proteins interaction partner dependent functions could be demonstrated and this might also hold true for Mst77F (Happel and Doenecke, 2009). Upon histone expulsion the inhibiting factor dissociates, Mst77F becomes "activated" and replaces the histones as a genome architectural protein. The reason why Mst77F initially binds DNA/chromatin in an "inactive" conformation or becomes "inactivated" might be due to additional functions prior to DNA condensation (see above: nuclear shaping – interaction with tubulin; transcriptional inhibition etc.). Concomitantly also transition proteins and little later protamines associate with the nascent DNA structures. The functional boundaries between Mst77F, transition proteins and protamine are totally unclear. Additionally, transition proteins as well as protamines are dispensible in Drosophila leaving Mst77F as the major structural genome organizing protein.

4.6 Outlook

To understand the complex process of spermatogenesis, in particular the transition from the nucleo-histone towards protamine organisation of the genome, it will be necessary to integrate the function of orchestrated processes like histone modifications, histone eviction and degradation, DNA breaks and condensation, nuclear shaping, variant structural proteins etc.. One piece in this puzzle is Mst77F that is suggested to take part in condensation processes, though is also implicated in nuclear shaping mediated by microtubules. The spaciotemporal expression pattern that shows Mst77F association with the nucleo-histone genome configuration questions the simple function of a genome organizer suggested by my *in vitro* experiments. Similar to histone H1 proteins the Mst77F function might be versatile and different functions may depend on the respective interaction partner or posttranslational modification (Happel and Doenecke, 2009). On the basis of my *in vitro* data and the accessible information from other studies I consider Mst77F to be different from histone H1 proteins even though they share physiochemical parameters and intrinsically unstructured domains that are vital in varying processes. To date no Mst77F homolog in mammals could be faithfully identified. The fact that the *Drosophila* genome condenses approximately 10 fold stronger than its mammalian counterpart raises the possibility that a homolog does not exist.

Further experiments addressing interaction partners, global genome localization and Mst77Fs proposed role in nuclear shaping have to be carried out to understand its exact role in *Drosophila* spermatogenesis. Eventually a null mutant will answer the question if Mst77F is really essential for DNA condensation *in vivo*.

Summary

The mechanisms of spermatogenesis, in particular the process of postmeiotic spermatid maturation, are little understood. During this developmental stage massive molecular and morphological rearrangements occur and the major events are conserved across higher eukaryots. This involves the condensation of the DNA after histone removal by two major classes of proteins: (a) transition proteins and (b) protamines. Beyond these proteins nothing is known about additional factors that directly impact DNA condensation. In *Drosophila* Mst77F has been implicated in the DNA condensation process on the basis of immunocytology experiments and its distant sequence homology towards histone H1 family proteins and protamines. Even though Mst77F clearly localizes inside the nucleus and is a component of the condensed DNA complex in differentiated sperm cells, a direct interaction with DNA, the prerequisite for condensation, could not be demonstrated.

On the basis of these findings my hypothesis was that Mst77F is a structural component of spermatid DNA that contributes to the condensation process. To address this hypothesis I conducted bioinformatic analysis of the Mst77F protein sequence to identify a putative DNA interaction domain. I found a highly charged, intrinsically unstructured C-terminal domain. This domain shows similarity towards the C-terminal domain of linker histone H1 that could be shown to bind DNA/chromatin. Qualitative and quantitativ binding experiments with recombinant Mst77F mutants and short DNAs corroborated the *in silico* predicted DNA interaction interface. The Mst77F DNA interaction was further characterized in detail by equilibrium binding experiments that identified an ionic interaction mode and provided the thermodynamic parameters. Besides the binding to DNA the proposed Mst77F function in DNA condensation processes was investigated. Therefore, Mst77F complexes with short DNAs were analysed by different biochemical assay systems and compared to a proposed homolog, histone hH1.4. I found Mst77F specific DNA aggregation properties that rely on the N-terminal domain of the protein. These observed differences were subsequently

further characterized by Atomic Force Microscopy with long DNAs. Based on the recorded structural differences protein multimerization experiments were conducted and the following model of the Mst77F-DNA interaction was proposed:
The Mst77F C-terminal domain constitutes a flexible, extended domain in solution. In this conformation protein-protein interactions via the N-terminus are rare. Upon recognition of DNA by the C-terminal domain, Mst77F undergoes a conformational change that increases its affinity for self-interactions dramatically. The C-terminal domain on its own can aggregate multiple individual DNA molecules in a one-dimensional manner. However, spreading of aggregation towards highly condensed structures is mediated by cooperative protein multimerization.

Additionally, the functional consequences of the formed Mst77F DNA complexes were tested in an *in vitro* transcription experiment. Lastly, since Mst77F arises when the genome is still in histone-based configuration my aim was to investigate binding-, structural- and functional effects also in the context of recombinant chromatin.

The present work presents the first comprehensive *in vitro* study on a DNA architecture protein presumably involved in DNA condensation during postmeiotic spermatid maturation in *Drosophila.* The insights obtained by this study could help to improve the understanding of spermatogenesis, in particular observed differences between mice and flies in DNA condensation.

Bibliography

Combet, C., Blanchet, C., Geourjon, C., and Deleage, G. (2000). NPS@: network protein sequence analysis. Trends in biochemical sciences *25*, 147-150.

Crane-Robinson, C. (1997). Where is the globular domain of linker histone located on the nucleosome? Trends in biochemical sciences *22*, 75-77.

Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. Journal of molecular biology *319*, 1097-1113.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic acids research *11*, 1475-1489.

Dootz, R., Toma, A.C., and Pfohl, T. (2011). Structural and dynamic properties of linker histone H1 binding to DNA. Biomicrofluidics *5*, 24104.

Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005a). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics *21*, 3433-3434.

Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005b). The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. Journal of molecular biology *347*, 827-839.

Dow, L.K., Jones, D.N., Wolfe, S.A., Verdine, G.L., and Churchill, M.E. (2000). Structural studies of the high mobility group globular domain and basic tail of HMG-D bound to disulfide cross-linked DNA. Biochemistry *39*, 9725-9736.

Dragan, A.I., Klass, J., Read, C., Churchill, M.E., Crane-Robinson, C., and Privalov, P.L. (2003). DNA binding of a non-sequence-specific HMG-D protein is entropy driven with a substantial non-electrostatic contribution. Journal of molecular biology *331*, 795-813.

Dunker, A.K., and Obradovic, Z. (2001). The protein trinity--linking function and disorder. Nature biotechnology *19*, 805-806.

Ekman, D., Light, S., Bjorklund, A.K., and Elofsson, A. (2006). What properties characterize the hub proteins of the protein-protein interaction network of Saccharomyces cerevisiae? Genome biology *7*, R45.

Fan, L., and Roberts, V.A. (2006). Complex of linker histone H5 with the nucleosome and its implications for chromatin packing. Proceedings of the National Academy of Sciences of the United States of America *103*, 8384-8389.

Fan, Y., Nikitina, T., Zhao, J., Fleury, T.J., Bhattacharyya, R., Bouhassira, E.E., Stein, A., Woodcock, C.L., and Skoultchi, A.I. (2005). Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. Cell *123*, 1199-1212.

Fang, H., Clark, D.J., and Hayes, J.J. (2012). DNA and nucleosomes direct distinct folding of a linker histone H1 C-terminal domain. Nucleic acids research *40*, 1475-1484.

Finn, R.M., Browne, K., Hodgson, K.C., and Ausio, J. (2008). sNASP, a histone H1-specific eukaryotic chaperone dimer that facilitates chromatin assembly. Biophysical journal *95*, 1314-1325.

Fuller, M.T. (1993). The development of Drosophila (Cold Spring Harbor, NY: Cold Spring Harbor Press).

Fuller, M.T., Caulton, J.H., Hutchens, J.A., Kaufman, T.C., and Raff, E.C. (1987). Genetic analysis of microtubule structure: a beta-tubulin mutation causes the formation of aberrant microtubules in vivo and in vitro. The Journal of cell biology *104*, 385-394.

Gallagher, S.R. (2006). One-dimensional SDS gel electrophoresis of proteins. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] *Chapter 10*, Unit 10 12A.

Garner, M.M., and Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. Nucleic acids research *9*, 3047-3060.

Gasteiger E., H.C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. (2005). Protein Identification and Analysis Tools on the ExPASy Server (Humana Press).

Gould-Somero, M.a.H., L. (1974). The timing of RNA synthesis for spermiogenesis in organ cultures of Drosophila melanogaster testes. Wilhelm Roux Arch, 133–148.

Grant, P.A. (2001). A tale of histone modifications. Genome biology *2*, REVIEWS0003.

Griffith, F. (1928). The Significance of Pneumococcal Types. The Journal of hygiene *27*, 113-159.

Gunasekaran, K., Tsai, C.J., Kumar, S., Zanuy, D., and Nussinov, R. (2003). Extended disordered proteins: targeting function with less scaffold. Trends in biochemical sciences *28*, 81-85.

Hacques, M.F., Muller, S., De Murcia, G., Van Regenmortel, M.H., and Marion, C. (1990). Accessibility and structural role of histone domains in chromatin. biophysical and immunochemical studies of progressive digestion with immobilized proteases. Journal of biomolecular structure & dynamics *8*, 619-641.

Hansen, J.C., Lu, X., Ross, E.D., and Woody, R.W. (2006). Intrinsic protein disorder, amino acid composition, and histone terminal domains. The Journal of biological chemistry *281*, 1853-1856.

Happel, N., and Doenecke, D. (2009). Histone H1 and its isoforms: contribution to chromatin structure and function. Gene *431*, 1-12.

Hartl, D.L., Jones, Elizabeth W. (2001). Genetics: Analysis of Genes and Genomes, Vol fifth Edition.

Hayashihara, K., Uchiyama, S., Shimamoto, S., Kobayashi, S., Tomschik, M., Wakamatsu, H., No, D., Sugahara, H., Hori, N., Noda, M.*, et al.* (2010). The middle region of an HP1-binding protein, HP1-BP74, associates with linker DNA at the entry/exit site of nucleosomal DNA. The Journal of biological chemistry *285*, 6498-6507.

Hirose, Y., and Manley, J.L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. Nature *395*, 93-96.

Honig, B., Sharp, K., and Gilson, M. (1989). Electrostatic interactions in proteins. Progress in clinical and biological research *289*, 65-74.

Huntley, M.A., and Golding, G.B. (2002). Simple sequences are rare in the Protein Data Bank. Proteins *48*, 134-140.

Jacobs, S.A., W. Fischle, and S. Khorasanizadeh (2004). Assays for the determination of
structure and dynamics of the interaction of the chromodomain with histone peptides, Vol 376.

Jean O.Thomas, C.R.a.J.T.F. (1991). Cooperative binding of the globular domains of histones
H1 and H5 to DNA. Nucleic acids research *20*, 187-194.

Jedrusik, M.A., and Schulze, E. (2001). A single histone H1 isoform (H1.1) is essential for chromatin silencing and germline development in Caenorhabditis elegans. Development *128*, 1069-1080.

Jelesarov, I., and Bosshard, H.R. (1999). Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. Journal of molecular recognition : JMR *12*, 3-18.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Johnson, W.C., Jr. (1990). Protein secondary structure and circular dichroism: a practical guide. Proteins *7*, 205-214.

Jones, S., Shanahan, H.P., Berman, H.M., and Thornton, J.M. (2003). Using electrostatic potentials to predict DNA-binding sites on DNA-binding proteins. Nucleic acids research *31*, 7189-7198.

Kemphues, K.J., Kaufman, T.C., Raff, R.A., and Raff, E.C. (1982). The testis-specific beta-tubulin subunit in Drosophila melanogaster has multiple functions in spermatogenesis. Cell *31*, 655-670.

Kim K, C.J., Heo K, Kim H, Levens D, Kohno K, Johnson EM, Brock HW, An W. (2008). Isolation and characterization of a novel H1.2 complex that acts as a repressor of p53-mediated transcription. The Journal of biological chemistry.

Klass, J., Murphy, F.V.t., Fouts, S., Serenil, M., Changela, A., Siple, J., and Churchill, M.E. (2003). The role of intercalating residues in chromosomal high-mobility-group protein DNA binding, bending and specificity. Nucleic acids research *31*, 2852-2864.

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

Kouzarides, T. (2007). SnapShot: Histone-modifying enzymes. Cell *131*, 822.

Leuba, S.H., and Bustamante, C. (1999). Analysis of chromatin by scanning force microscopy. Methods Mol Biol *119*, 143-160.

Lis, J.T., and Schleif, R. (1975). Size fractionation of double-stranded DNA by precipitation with polyethylene glycol. Nucleic acids research *2*, 383-389.

Lowary, P.T., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. Journal of molecular biology *276*, 19-42.

Lu, X., Hamkalo, B., Parseghian, M.H., and Hansen, J.C. (2009a). Chromatin condensing functions of the linker histone C-terminal domain are mediated by specific amino acid composition and intrinsic protein disorder. Biochemistry *48*, 164-172.

Lu, X., and Hansen, J.C. (2004). Identification of specific functional subdomains within the linker histone H10 C-terminal domain. The Journal of biological chemistry *279*, 8701-8707.

Lu, X., Wontakal, S.N., Emelyanov, A.V., Morcillo, P., Konev, A.Y., Fyodorov, D.V., and Skoultchi, A.I. (2009b). Linker histone H1 is essential for Drosophila development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure. Genes & development *23*, 452-465.

Lucius, H., Haberland, A., Zaitsev, S., Dalluge, R., Schneider, M., and Bottger, M. (2001). Structure of transfection-active histone H1/DNA complexes. Molecular biology reports *28*, 157-165.

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

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

Lupas, A., Van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. Science *252*, 1162-1164.

Maresca, T.J., Freedman, B.S., and Heald, R. (2005). Histone H1 is essential for mitotic chromosome architecture and segregation in Xenopus laevis egg extracts. The Journal of cell biology *169*, 859-869.

Mason, J.M., and Arndt, K.M. (2004). Coiled coil domains: stability, specificity, and biological implications. Chembiochem : a European journal of chemical biology *5*, 170-176.

McKearin, D. (1997). The Drosophila fusome, organelle biogenesis and germ cell differentiation: if you build it. BioEssays : news and reviews in molecular, cellular and developmental biology *19*, 147-152.

Meyer, T.S.L., B. L. (1965). Use of Coomassie brilliant blue R250 for the electrophoresis of microgram quantities of parotid saliva proteins on acrylamide-gel strips. Biochimica et Biophysica Acta 107 (1): 144–145.

Mitchell, P.J., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science *245*, 371-378.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor symposia on quantitative biology *51 Pt 1*, 263-273.

Multigner, L., Gagnon, J., Van Dorsselaer, A., and Job, D. (1992). Stabilization of sea urchin flagellar microtubules by histone H1. Nature *360*, 33-39.

Nelson, H.C., Finch, J.T., Luisi, B.F., and Klug, A. (1987). The structure of an oligo(dA).oligo(dT) tract and its biological implications. Nature *330*, 221-226.

Ner, S.S., Blank, T., Perez-Paralle, M.L., Grigliatti, T.A., Becker, P.B., and Travers, A.A. (2001). HMG-D and histone H1 interplay during chromatin assembly and early embryogenesis. The Journal of biological chemistry *276*, 37569-37576.

Nikitina, T., Shi, X., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., and Woodcock, C.L. (2007). Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. Molecular and cellular biology *27*, 864-877.

Nishikawa, K., and Scheraga, H.A. (1976). Geometrical criteria for formation of coiled-coil structures of polypeptide chains. Macromolecules *9*, 395-407.

Nishioka, K., Rice, J.C., Sarma, K., Erdjument-Bromage, H., Werner, J., Wang, Y., Chuikov, S., Valenzuela, P., Tempst, P., Steward, R., *et al.* (2002). PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. Molecular cell *9*, 1201-1213.

O'Hare, P., and Williams, G. (1992). Structural studies of the acidic transactivation domain of the Vmw65 protein of herpes simplex virus using 1H NMR. Biochemistry *31*, 4150-4156.

Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C.J., and Dunker, A.K. (2003). Predicting intrinsic disorder from amino acid sequence. Proteins *53 Suppl 6*, 566-572.

Olivieri, G., and Olivieri, A. (1965). Autoradiographic study of nucleic acid synthesis during spermatogenesis in Drosophila melanogaster. Mutation research *2*, 366-380.

Patterton, H.G., Landel, C.C., Landsman, D., Peterson, C.L., and Simpson, R.T. (1998). The biochemical and phenotypic characterization of Hho1p, the putative linker histone H1 of Saccharomyces cerevisiae. The Journal of biological chemistry *273*, 7268-7276.

Phair, R.D., Scaffidi, P., Elbi, C., Vecerova, J., Dey, A., Ozato, K., Brown, D.T., Hager, G., Bustin, M., and Misteli, T. (2004). Global nature of dynamic protein-chromatin interactions in vivo: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. Molecular and cellular biology *24*, 6393-6402.

Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. Genes & development *20*, 2922-2936.

Ponte, I., Vila, R., and Suau, P. (2003). Sequence complexity of histone H1 subtypes. Molecular biology and evolution *20*, 371-380.

Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L., and Sweet, R.M. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature *362*, 219-223.

Rasband, W.S. (1997 - 2011). ImageJ.

Rathke, C., Baarends, W.M., Jayaramaiah-Raja, S., Bartkuhn, M., Renkawitz, R., and Renkawitz-Pohl, R. (2007). Transition from a nucleosome-based to a protamine-based chromatin configuration during spermiogenesis in Drosophila. Journal of cell science *120*, 1689-1700.

Rathke, C., Barckmann, B., Burkhard, S., Jayaramaiah-Raja, S., Roote, J., and Renkawitz-Pohl, R. (2010). Distinct functions of Mst77F and protamines in nuclear shaping and chromatin condensation during Drosophila spermiogenesis. European journal of cell biology *89*, 326-338.

Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., and Dunker, A.K. (2001). Sequence complexity of disordered protein. Proteins *42*, 38-48.

Roque, A., Iloro, I., Ponte, I., Arrondo, J.L., and Suau, P. (2005). DNA-induced secondary structure of the carboxyl-terminal domain of histone H1. The Journal of biological chemistry *280*, 32141-32147.

Roque, A., Ponte, I., Arrondo, J.L., and Suau, P. (2008). Phosphorylation of the carboxy-terminal domain of histone H1: effects on secondary structure and DNA condensation. Nucleic acids research *36*, 4719-4726.

Russell, S.R., and Kaiser, K. (1993). Drosophila melanogaster male germ line-specific transcripts with autosomal and Y-linked genes. Genetics *134*, 293-308.

Sadowski, I., Ma, J., Triezenberg, S., and Ptashne, M. (1988). GAL4-VP16 is an unusually potent transcriptional activator. Nature *335*, 563-564.

Sambrook, J., and Russell, D.W. (2001). Molecular cloning: A Laboratory Manual 3rd Edition (Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press).

Sandhu, K.S., and Dash, D. (2007). Dynamic alpha-helices: conformations that do not conform. Proteins *68*, 109-122.

Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. Proceedings of the National Academy of Sciences of the United States of America *95*, 5857-5864.

Schwarz, P.M., Felthauser, A., Fletcher, T.M., and Hansen, J.C. (1996). Reversible oligonucleosome self-association: dependence on divalent cations and core histone tail domains. Biochemistry *35*, 4009-4015.

Shazman, S., Celniker, G., Haber, O., Glaser, F., and Mandel-Gutfreund, Y. (2007). Patch Finder Plus (PFplus): a web server for extracting and displaying positive electrostatic patches on protein surfaces. Nucleic acids research *35*, W526-530.

Shen, X., Yu, L., Weir, J.W., and Gorovsky, M.A. (1995). Linker histones are not essential and affect chromatin condensation in vivo. Cell *82*, 47-56.

Shewmaker, F., Ross, E.D., Tycko, R., and Wickner, R.B. (2008). Amyloids of shuffled prion domains that form prions have a parallel in-register beta-sheet structure. Biochemistry *47*, 4000-4007.

Sinz, A. (2003). Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. Journal of mass spectrometry : JMS *38*, 1225-1237.

Stasevich, T.J., Mueller, F., Brown, D.T., and McNally, J.G. (2010). Dissecting the binding mechanism of the linker histone in live cells: an integrated FRAP analysis. The EMBO journal *29*, 1225-1234.

Subirana, J.A. (1990). Analysis of the charge distribution in the C-terminal region of histone H1 as related to its interaction with DNA. Biopolymers *29*, 1351-1357.

Sunil Jayaramaiah Raja, R.R.-P. (2005). Replacement by Drosophila melanogaster Protamines and Mst77F of Histones during Chromatin Condensation in Late Spermatids and Role of Sesame in the Removal of These Proteins from the Male Pronucleus†. MCB *25*, 6165-6177.

Syed, S.H., Goutte-Gattat, D., Becker, N., Meyer, S., Shukla, M.S., Hayes, J.J., Everaers, R., Angelov, D., Bednar, J., and Dimitrov, S. (2010). Single-base resolution mapping of H1-nucleosome interactions and 3D organization of the nucleosome.

Proceedings of the National Academy of Sciences of the United States of America *107*, 9620-9625.

Tatham, A.S., and Shewry, P.R. (2000). Elastomeric proteins: biological roles, structures and mechanisms. Trends in biochemical sciences *25*, 567-571.

Tompa, P. (2002). Intrinsically unstructured proteins. Trends in biochemical sciences *27*, 527-533.

Uversky, V.N. (2002). Natively unfolded proteins: a point where biology waits for physics. Protein science : a publication of the Protein Society *11*, 739-756.

van Holde, K., and Zlatanova, J. (1996). What determines the folding of the chromatin fiber? Proceedings of the National Academy of Sciences of the United States of America *93*, 10548-10555.

Verdaguer, N., Perello, M., Palau, J., and Subirana, J.A. (1993). Helical structure of basic proteins from spermatozoa. Comparison with model peptides. European journal of biochemistry / FEBS *214*, 879-887.

Vibranovski, M.D., Chalopin, D.S., Lopes, H.F., Long, M., and Karr, T.L. (2010). Direct evidence for postmeiotic transcription during Drosophila melanogaster spermatogenesis. Genetics *186*, 431-433.

Vibranovski, M.D., Lopes, H.F., Karr, T.L., and Long, M. (2009). Stage-specific expression profiling of Drosophila spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed genes. PLoS genetics *5*, e1000731.

Vila, R., Ponte, I., Collado, M., Arrondo, J.L., Jimenez, M.A., Rico, M., and Suau, P. (2001). DNA-induced alpha-helical structure in the NH2-terminal domain of histone H1. The Journal of biological chemistry *276*, 46429-46435.

Vila, R., Ponte, I., Jimenez, M.A., Rico, M., and Suau, P. (2002). An inducible helix-Gly-Gly-helix motif in the N-terminal domain of histone H1e: a CD and NMR study. Protein science : a publication of the Protein Society *11*, 214-220.

Vogel, B., Loschberger, A., Sauer, M., and Hock, R. (2011). Cross-linking of DNA through HMGA1 suggests a DNA scaffold. Nucleic acids research *39*, 7124-7133.

Vucetic, S., Brown, C.J., Dunker, A.K., and Obradovic, Z. (2003). Flavors of protein disorder. Proteins *52*, 573-584.

Wanker, E.E., Sun, Y., Savitz, A.J., and Meyer, D.I. (1995). Functional characterization of the 180-kD ribosome receptor in vivo. The Journal of cell biology *130*, 29-39.

Widlak P, K.M., Parseghian MH, Lu X, Hansen JC, Garrard WT. (2005). The histone H1 C-terminal domain binds to the apoptotic nuclease, DNA fragmentation factor (DFF40/CAD) and stimulates DNA cleavage. Biochemistry.

Winter, R.B., Berg, O.G., and von Hippel, P.H. (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor--operator interaction: kinetic measurements and conclusions. Biochemistry *20*, 6961-6977.

Wootton JC, F.S. (1996). Analysis of compositionally biased regions in sequence databases, Vol 266.

Wouters-Tyrou, D., Martinage, A., Chevaillier, P., and Sautiere, P. (1998). Nuclear basic proteins in spermiogenesis. Biochimie *80*, 117-128.

Xiao, B., Jing, C., Kelly, G., Walker, P.A., Muskett, F.W., Frenkiel, T.A., Martin, S.R., Sarma, K., Reinberg, D., Gamblin, S.J.*, et al.* (2005). Specificity and mechanism of the histone methyltransferase Pr-Set7. Genes & development *19*, 1444-1454.

Yan, W., Ma, L., Burns, K.H., and Matzuk, M.M. (2003). HILS1 is a spermatid-specific linker histone H1-like protein implicated in chromatin remodeling during mammalian spermiogenesis. Proceedings of the National Academy of Sciences of the United States of America *100*, 10546-10551.

Yoon, C., Prive, G.G., Goodsell, D.S., and Dickerson, R.E. (1988). Structure of an alternating-B DNA helix and its relationship to A-tract DNA. Proceedings of the National Academy of Sciences of the United States of America *85*, 6332-6336.