



Mechanistic studies of protein-DNA interactions by single molecule atomic force microscopy

Mechanistische Untersuchungen von protein-DNA-Wechselwirkungen
mittels Einzelmolekül-Rasterkraftmikroskopie

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg

Section Biomedicine

Submitted by

Disha Mohan Bangalore

from

Bangalore, India

Würzburg 2021



Submitted on:

Office stamp

Members of the Thesis Committee

Chairperson: Prof. Dr. Markus Sauer

Primary Supervisor: Dr. Ingrid Tessmer

Supervisor (Second): Prof. Dr. Katrin Heinze

Supervisor (Third): Prof. Dr. Elmar Wolf

Date of Public Defense:

Date of Receipt of Certificates:

“Somewhere, something incredible is waiting to be known.”

-Carl Sagan, astronomer of the people.

Abstract

Protein-DNA interactions are central to many biological processes and form the bedrock of gene transcription, DNA replication, and DNA repair processes. Many proteins recognize specific sequences in DNA- a restriction enzyme must only cut at the correct sequence and a transcription factor should bind at its consensus sequence. Some proteins are designed to bind to specific structural or chemical features in DNA, such as DNA repair proteins and some DNA modifying enzymes. Target-specific DNA binding proteins initially bind to non-specific DNA and then search for their target sites through different types of diffusion mechanisms. Atomic force microscopy (AFM) is a single-molecule technique that is specifically well-suited to resolve the distinct states of target-specific as well as nonspecific protein-DNA interactions that are vital for a deeper insight into the target site search mechanisms of these enzymes. In this thesis, protein systems involved in epigenetic regulation, base excision repair (BER), and transcription are investigated by single-molecule AFM analyses complemented by biochemical and biophysical experiments.

The first chapter of this thesis narrates the establishment of a novel, user-unbiased MatLab-based tool for automated DNA bend angle measurements on AFM data. This tool has then been employed to study the initial lesion detection step of several DNA glycosylases. These results promoted a model describing the altered plasticities of DNA at the target lesions of DNA glycosylases as the fundamental mechanism for their enhanced efficiency of lesion detection.

In the second chapter of this thesis, the novel automated tool has been further extended to provide protein binding positions on the DNA along with corresponding DNA bend angles and applied to the study of DNMT3A DNA methyltransferase. These AFM studies revealed preferential co-methylation at specific, defined distances between two CpG sites by the enzyme and when combined with biochemical analyses and structural modelling supported novel modes of CpG co-methylation by DNMT3A.

In the third chapter of this thesis, the role of 8-oxo-guanine glycosylase (hOGG1) in Myc-mediated transcription initiation has been investigated. AFM analyses revealed that in the presence of oxidative damage in DNA, Myc is recruited to its target site (E-box) by hOGG1 through direct protein-protein interactions, specifically under oxidizing conditions. Intriguingly, oxidation of hOGG1 was further observed to result in dimerization of hOGG1, which may also play a role in the mechanism of transcription regulation by hOGG1 under oxidative stress.

Zusammenfassung

Protein-DNA-Wechselwirkungen sind für viele biologische Prozesse von zentraler Bedeutung und bilden die Grundlage der Gentranskription, der DNA-Replikation und der DNA-Reparaturprozesse. Viele Proteine erkennen bestimmte Basen-Sequenzen in der DNA - ein Restriktionsenzym darf nur an der richtigen Sequenz schneiden, und ein Transkriptionsfaktor sollte an seine Konsensussequenz binden. Einige Proteine sind darauf ausgelegt, an bestimmte strukturelle oder chemische Merkmale der DNA zu binden, wie z. B. DNA-Reparaturproteine und verschiedene DNA-modifizierende Enzyme. Zielspezifische DNA-bindende Proteine binden zunächst an unspezifische DNA und suchen dann durch verschiedene Arten von Diffusionsmechanismen nach ihren Zielstellen in der DNA. AFM ist eine Einzelmolekültechnik, die besonders gut geeignet ist, um die verschiedenen Zustände sowohl der spezifisch gebundenen als auch unspezifischen Protein-DNA-Wechselwirkungen aufzulösen, die für einen tieferen Einblick in die Mechanismen der Zielstellensuche unerlässlich sind. In dieser Arbeit werden Proteinsysteme, die an der epigenetischen Regulation, der Basenexzisionsreparatur (BER) und der Transkription beteiligt sind, durch Einzelmolekül-AFM-Analysen untersucht, und diese Studien werden durch biochemische und biophysikalische Experimente komplementiert.

Das erste Kapitel dieser Arbeit beschreibt die Etablierung eines neuartigen, Benutzer-unabhängigen MatLab-basierten Ansatzes zur automatisierten Messung von DNA-Biegungswinkeln an AFM-Daten. Dieses Tool wurde dann eingesetzt, um den initialen Schritt der DNA-Schadenserkennung verschiedener DNA-Glykosylasen zu untersuchen. Diese Daten unterstützten ein Modell, das die erhöhte Effizienz der DNA-Schadenserkennung durch Glykosylasen basierend auf veränderten DNA-Flexibilitäten an den Stellen ihrer Zielschäden beschreibt.

Im zweiten Kapitel dieser Arbeit wurde dieses neuartige automatisierte Tool erweitert, um Proteinbindungspositionen auf der DNA zusammen mit den entsprechenden DNA-Biegungswinkeln zu ermitteln, und auf die Untersuchung der DNMT3A DNA-Methyltransferase mittels AFM angewendet. Diese AFM-Studien ergaben eine bevorzugte Co-Methylierung in spezifischen, definierten Abständen zwischen zwei CpG-Stellen durch das Enzym und unterstützten in Kombination mit biochemischen Analysen und Strukturmodellierung ein neues Modell der CpG-Co-Methylierung durch DNMT3A.

Im dritten Kapitel dieser Arbeit wurde die Rolle der 8-Oxo-Guanin-Glycosylase (hOGG1) in der Transkriptionsinitiation durch den Transkriptionsfaktor Myc untersucht. AFM-Analysen ergaben, dass in Anwesenheit oxidativer Schäden in DNA, Myc durch hOGG1 über direkte Protein-Protein-Wechselwirkungen an seine Zielstelle (E-box) rekrutiert wird, insbesondere unter oxidierenden Bedingungen. Interessanterweise wurde weiterhin beobachtet, dass die Oxidation von hOGG1 selbst zu einer Dimerisierung von hOGG1 führt, die auch eine Rolle im Mechanismus der Transkriptionsregulation durch hOGG1 unter oxidativem Stress spielen könnte.

Table of contents

Abstract	5
Zusammenfassung	7
Table of Contents	9
1. General Introduction	15
2. Materials and Methods	17
2.1 Materials.....	17
2.1.1 Instruments and consumables.....	17
2.1.2 Chemicals, reagents, and media.....	22
2.1.3 Resins.....	23
2.1.4 Cloning material, enzymes, and recombinant proteins.....	23
2.1.5 Bacterial strains and plasmids.....	25
2.1.6 Oligonucleotides.....	25
2.1.7 Antibodies and enzymes for protein purification and protein ladder...28	
2.1.8 Computer software, databases, servers, and deposited scripts.....	29
2.2 Methods.....	31
2.2.1 Protein preparation.....	31
2.2.1.a Cloning of protein constructs.....	31
<i>PCR amplification</i>	31
<i>Cloning strategy using SLIC method</i>	31
<i>Site-directed mutagenesis and gene modification</i>	32
<i>DNA analysis by gel electrophoresis and ultraviolet-visible spectroscopy</i>	32
<i>Chemical transformation</i>	32

<i>Plasmid isolation</i>	33
2.2.1.b Recombinant protein expression.....	33
2.2.1.c Protein purification.....	33
<i>Cell lysis</i>	33
<i>Chromatography</i>	33
2.2.2 Protein analysis.....	40
<i>SDS-polyacrylamide gel electrophoresis</i>	40
<i>UV-Vis spectrophotometry</i>	40
<i>Analytical size exclusion chromatography</i>	41
<i>Circular dichroism spectroscopy</i>	41
<i>SDS-PAGE studies to investigate hOGG1 dimerization</i>	41
<i>Sedimentation velocity analytical ultracentrifugation</i>	42
<i>Solvent accessibility determination for hOGG1</i>	42
<i>Crystallization of oxidized hOGG1</i>	43
<i>Mass spectrometry</i>	43
<i>Western blot experiments to identify dimerization for</i> <i>hOGG1_{C28A}, hOGG1_{C241A} and hOGG1_{C253A}</i>	44
2.2.3 Atomic force microscopy.....	45
2.2.3.1 DNA substrate preparation for AFM imaging.....	45
<i>Control assays for DNA substrate preparation</i>	45
<i>E-box cloning strategy for DNA substrate preparation</i>	47
2.2.3.2 AFM sample preparation.....	49
<i>Lesion substrates and glycosylases-DNA complexes</i>	49
<i>Dual CpG site containing substrates and DNMT3A-3L</i>	

	<i>DNA methyltransferases</i>	49
	<i>hOGG1, Myc, Max, LSD1 complexes, and DNA substrates</i>	49
2.2.3.3	Sample deposition, Image collection and processing.....	50
2.2.3.4	AFM imaging and processing.....	50
	<i>Principle</i>	50
	<i>Imaging conditions and image processing</i>	51
2.2.3.5	AFM data analysis.....	52
a.	Automated AFM DNA bending analysis at target lesion sites.....	52
	<i>Pre-processing steps for automated DNA bend angle analysis at DNA target lesion sites</i>	52
	<i>MatLab routine for DNA bending analysis at lesion target sites located at 50% DNA length</i>	52
	<i>Persistence length measurements</i>	54
b.	Automated DNA bending analysis at protein positions.....	54
	<i>Pre-processing steps for automated DNA bend angle analysis of protein-DNA complexes</i>	54
	<i>MatLab routine for DNA bending analysis at non-specifically bound protein positions</i>	55
	<i>MatLab routine for DNA bending analysis of proteins (DNMT3A-3L complexes) bound at 50% position on the DNA substrates</i>	56
	<i>Manual DNA bend angle measurements</i>	56

c.	Binding specificity analysis.....	56
	<i>Manual protein position measurements</i>	56
	<i>Automated protein position measurements</i>	57
d.	Stoichiometric analysis.....	58
2.2.4	Biophysical and Biochemical assays for protein-protein and protein-DNA interactions analyses.....	59
	<i>Ensemble FRET experiments</i>	59
	<i>hOGG1 gel-based activity assay</i>	60
	<i>Electrophoretic Mobility Shift Assays</i>	61
	<i>Fluorescence polarization measurements</i>	62
	<i>Native agarose gel electrophoresis</i>	63
	<i>In vitro pull-down assays</i>	63
2.2.5	<i>In silico</i> DNA curvature analysis.....	64
3.	Investigating initial lesion detection strategies of DNA glycosylases by automated AFM DNA bending analyses.....	65
3.1	Introduction.....	65
3.2	Results.....	72
3.2.1	MatLab analysis routine for characterization of glycosylase target lesions.....	72
3.2.2	MatLab analysis routine for characterization of glycosylase-nonspecific DNA complexes.....	75
3.3	Discussion.....	79
3.3.1	Method evaluation.....	79
3.3.2	Characterization of BER target lesions in the absence of glycosylases and glycosylase complexes with undamaged DNA.....	80

	<i>hTDG-undamaged DNA complex and G:U target lesion</i>	80
	<i>hOGG1-undamaged DNA complex and oxoG target lesion</i>	82
	<i>MutY-undamaged DNA complex and oxoG:A and G:A target lesions</i>	83
	<i>hAAG-undamaged DNA complex and ethenoadenine target lesion</i>	84
3.3.3	Model of initial lesion detection by glycosylases.....	85
4.	Single molecule AFM studies reveal dual CpG site binding mechanisms of DNMT3A-3L DNA methyltransferase	87
4.1	Introduction.....	87
4.2	Results.....	92
4.2.1	Co-methylation studies of CpG sites at varying distances.....	92
4.2.2	Sequence preferences by DNMT3A and DNMT3A-3L complexes.....	94
4.2.3	Single molecule AFM analysis to characterize modes of binding of DNMT3A-3L to CpG sites placed at variable distances.....	95
4.2.4	Modelling DNMT3A-3L complexes to understand the different modes of co-methylation.....	100
4.3	Discussion.....	103
5.	Investigations functions of hOGG1 in regulation of Myc-mediated transcription	108
5.1	Introduction.....	108
5.2	Results.....	115
5.2.1	hOGG1 oligomerization is altered under oxidative conditions.....	115
5.2.2	hOGG1 alters Myc binding to target sites.....	123

5.3 Discussion.....	134
5.4 Future perspectives.....	139
6. Conclusions.....	140
Appendix.....	142
Appendix I.....	142
Appendix II.....	169
Appendix III.....	173
Literature.....	192
List of Figures.....	205
List of Tables.....	210
List of Abbreviations.....	212
List of Publications.....	219
List of congress contributions.....	219
Curriculum Vitae.....	221
Acknowledgements.....	225
Affidavit.....	227

1. General Introduction

A wide variety of proteins interacting with DNA play significant roles in many pivotal biological processes like gene transcription, replication, recombination, and DNA repair^{1,2}. In this thesis, I have attempted to understand protein-DNA interactions at a single-molecule level using atomic force microscopy (AFM) imaging supported by ensemble, biochemical and biophysical techniques. Protein-DNA systems involved in base excision repair (DNA glycosylases) and DNA methylation (DNA methyltransferases), as well as interactions between a DNA repair protein and a transcription factor have been studied and are presented in different chapters of this thesis.

Atomic force microscopy emerged as a vital technique out of research in the field of nanotechnology and belongs to the scanning probe microscopy (SPM) family³. It is a powerful tool for the study of protein-DNA interactions, and in fact one of only few methods to visualize protein-DNA interactions at the single-molecule level. AFM thus complements techniques like X-ray crystallography, which provides sub-molecular resolution, and optical microscopy, which is restricted to lower spatial resolution (few hundred nanometers) but provides high temporal resolution. As a single-molecule method, AFM is uniquely suited to resolve heterogeneity in protein-DNA complexes and reveal several distinct states present among them, which otherwise are concealed in bulk techniques. Stoichiometric analyses of protein complexes, binding specificities of proteins on DNA and DNA bend angle measurements on AFM data were major approaches used to derive mechanistic information on the different protein systems studied in this thesis. Techniques like electrophoretic mobility shift assay (EMSA), fluorescence polarization, crystallography, analytical ultracentrifugation (AUC), native agarose gel electrophoresis (NAGE), and fluorescence resonance energy transfer (FRET) that provide ensemble characteristics of biomolecular interactions have been employed in this thesis to support AFM results.

The aims of this thesis were to characterize protein-DNA and protein-protein interactions to gain deeper insight into the target site search mechanisms of DNA glycosylases, to investigate the role of a specific DNA glycosylase in transcription regulation, and to understand distinct co-methylation preferences by DNMT3A methyltransferase.

The first section of this thesis aims to investigate the initial lesion detection step by several DNA glycosylases that bind to structurally different target base lesions. To achieve this, the study further aims to establish a novel, user-unbiased MatLab-based tool for automated DNA bend angle measurements on AFM data.

The second chapter of this thesis aims to understand how DNMT3A DNA methyltransferase can overcome the innate structural preference for co-methylation at dual CpG sites with 12 bp spacing. Towards this aim, manual and automated measurements on the AFM data are performed along with biochemical and structural modelling studies.

The final chapter of this thesis aims to investigate the functions of human 8-oxo-guanine glycosylase (hOGG1) in Myc-mediated transcription. hOGG1 is one of the DNA glycosylases studied in the first chapter of this thesis. Apart from its prime role in base excision repair, it is also believed to be involved in transcription regulation. With this objective, the study employs extensive AFM analyses in conjunction with biochemical and biophysical techniques.

2. Materials and Methods

2.1 Materials

2.1.1 Instruments and consumables

The following list contains the instruments used in this thesis (Table. 2.1).

Table 2.1: Instruments

Instrument	Model	Supplier
Atomic Force Microscope	Molecular force probe (MFP) 3D-Bio AFM	Asylum Research
Autoclave	Systec V-150	Systec
Agarose gel electrophoresis system	Mini-Sub® Cell GT System	Bio-Rad Laboratories
Analytical Ultracentrifuge (AUC)	Optima XL-1	Beckman Coulter
AUC sample cell	Double-sector Epon Charcoal-Filled Centerpiece, Sapphire windows	Beckman Coulter
AUC-Rotor	An-50 Ti	Beckman Coulter
Analytical size exclusion chromatography (SEC) columns	Superose™ 6 increase 10/300 GL Superdex™ 200 10/300 GL (SD 200 10/300) Superdex™ 75 10/300 GL (SD75 10/300)	GE Healthcare
Balances	XS 105 Dual Range (0-1 g) XS 6002S Dual Range (0.5-500g)	Mettler Toledo
Block thermostat	Rotilabo® block thermostat H 250	Carl Roth
Bio-Photometer	6131	Eppendorf
CD Spectropolarimeter	J-810	Jasco

CD cuvette	Cylindrical absorption cuvette, path length 1 mm	Hellma Analytics
Cell disruption system	M-110P	Microfluidics
Centrifuges	5417 R 5424 5804 R 5430 R	Eppendorf
Centrifuges	Avanti J-26 XP Avanti J-HC	Beckman Coulter
Columns for affinity chromatography	Econo-Column 1,5 x 15 cm Econo-Column 2,5 x 20 cm	Biorad
Columns for Ion exchange chromatography	MonoQ® 10/100 GL	GE Healthcare
Electrophoresis	Mini-PROTEAN Tetra Cell	Bio-Rad Laboratories
Electrophoresis power Supply	PowerPac™ Basic	Bio-Rad Laboratories
Fast protein liquid chromatography (FPLC) systems (Protein purification)	ÄKTA™ pure 25 ÄKTA™ avant 25 ÄKTA™ purifier 10	GE Healthcare
Fluorescence spectrometer	Fluoromax4	Horiba Jobin Yvon
Fluorescence cuvette	Ultra-micro cuvettes (30µL)	Hellma Analytics
Fraction collector	Frac-950	GE Healthcare
Gel-drying device	GelAir Gel Dryer	Bio-Rad Laboratories
Imaging System	Odyssey	LI-COR Biosciences
Imaging System	ChemiDoc™ MP Imaging System	Bio-Rad Laboratories
Imaging System	PharosFX imager	Bio-Rad Laboratories
Imaging system	Infinity	Vilber
Incubator	B15 Compact Incubator	Heraeus
Ice machine	94774	Ziegra Eismaschinen
Liquid handling robot	NT8® robot	FORMULATRIX®

Liquid handling robot	LISSY 2002	Zinsser Analytic
Laminar flow hood Class II	-	Thermo Fisher Scientific
Magnetic stirrer	MR 3002	Heidolph Instruments
Microplate reader	CLARIOstar®	BMG LABTECH
Microscope camera	AxioCam MRc	ZEISS
Microscope light source	KL 2500 LCD	ZEISS
Microscope light source	CL 1500 Eco	ZEISS
Microscopes	SteREO Discovery. V12, STEMI 2000-c	ZEISS
Microwave	-	Privileg
Mini Trans-Blot cell	-	Bio-Rad Laboratories
Magnetic stirrer	MR 3002	Heidolph
Monitor pH/C-900 (pH, conductivity temperature) Monitor UV900 (absorption)	-	GE Healthcare
Polymerase chain reaction (PCR)-cyclers	Mastercycler® EPgradient S Mastercycler® proS	Eppendorf
pH meter	BlueLine 14pH	SCHOTT
Pipettes	XLS+LTS PIPET 0.1-2 µL XLS+LTS PIPET 0.5-10 µL XLS+LTS PIPET 2-20 µL XLS+LTS PIPET 20-200 µL XLS+LTS PIPET 100-1000 µL	Mettler-Toledo
Preparative SEC FPLC column	HiLoad™ 16/600 Superdex™ 200 pg (SD 200 16/600) HiLoad™ 16/600 Superdex™ 75 pg (SD75 16/600)	GE Healthcare
Protein Crystallization Imager	ROCKIMAGER®	FORMULATRIX

Sealing robot	RoboSeal	HJ-BIOANALYTIC
Shaking incubators	ISF-1-W ISF-1-X LT-X	Kühner
Spectrophotometer	NanoDrop ND 1000	Peqlab
Sample loading pump	P960	GE Healthcare
Thermomixer	Thermomix comfort	Eppendorf
Ultrapure water system	TKA GenPure	Thermo Fisher Scientific
UV imaging system	Gel Doc™ XR System	Bio-Rad Laboratories
Vortex mixer	Vortex-Genie 2	Scientific Industries

The list below excludes general glass and plastic containers and items.

Table 2.2: Special consumables

Type	Model	Supplier
24-well crystallization plates	Crystalgen SuperClear™ Plate	Jena Bioscience
384-well microplate	Black bottom, non-binding	Greiner Bio-One
96-well crystallization plates	Crystalquick™ 1 square well, flat bottom, low profile	Greiner Bio-One
AFM cantilevers	OMCL-AC240TSA, noncontact/ tapping mode silicon probes	Olympus
Blotting paper	-	Roth
Centrifugal concentrator	Amicon® Ultra-0.5, 4 and 15 mL	Merck Millipore
Centrifuge tube	Cellstar® centrifuge tube- 15 and 50 mL	Greiner Bio-One
Cover slides	Circular, Siliconised, 22 mm	Jena Bioscience
Cuvettes	Rotilabo® -single-use	Carl Roth
CD cuvette	Cylindrical absorption cuvette, path length 1 mm	Hellma Analytics
Concentrators	MWCO 10,000 Vivaspin 20 (30 ml) MWCO 30,000 Vivaspin 20 (30 ml)	Sartorius Stedim Biotech
Fluoromax cuvette	105.202	Hellma Analytics

Dialysis membranes	Spectra/Por®	Spectrum Laboratories
Filter paper	-	Sartorius
Gloves	Nitril gloves	Star Lab
Mica	Grade V-5, 75 x 25 x 0.15 mm	SPI supplies
Microcon centrifugal filters for DNA concentration	10 kDa MWCO (500 µl) 100 kDa MWCO (500 µl)	Millipore
Microcon centrifugal filters for buffer exchange of protein	10 kDa MWCO (500 µl)	Millipore
Microscope slides (76 x 26 x 1 mm)	-	Hartenstein
Nitrocellulose membrane	-	Amersham
Optical quality sealing foil	VIEWseal™	Greiner Bio-One
Parafilm® M	2 in. x 250 ft	Sigma-Aldrich
Polymerase chain reaction (PCR) tubes	Multiply®-Pro cup 0.2 mL, Multiply®-µStrip 0.2 mL chain, 8-Lid chain, flat	Sarstedt
Pipette tips	Pipette tips- 10, 200, 1000 µL	Rainin
Polyvinylidene difluoride (PVDF) membrane	-	Amersham
Precast SDS-Gels	Mini-PROTEAN TGX 4-20% gels	Bio-Rad Laboratories
PageRuler prestained protein ladder	-	Fermentas
Pipette tips 2µL	-	Eppendorf
Reaction tubes	SafeSeal tube- 0.5, 1.5 clear and brown, 2 mL	Sarstedt
Sterile filter	Acrodisc® sterile filter for syringe- 0.22 and 0.45 µm	Pall
Syringes	Omnifix® syringes- 1,5,10, and 20 mL	B. Braun

2.1.2 Chemicals, reagents, and media

The following list contains the chemicals used in this thesis (Table. 2.3). All buffers and solutions were prepared with ultrapure water generated by a TKA GenPure system.

Table 2.3: Chemicals, reagents, and media

Substance	Supplier
2'-Deoxythymidine 5'-triphosphate (dTTP), sodium salt solution	New England Biolabs
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Carl Roth
Acetic acid	Carl Roth
AlexaFluor 488	Invitrogen
2-Propanol	Carl Roth
Acrylamide/Bisacrylamide (37.5:1)	Carl Roth
Acrylamide/Bisacrylamide (29:1)	Carl Roth
Agar	Carl Roth
Agarose NEEO ultra quality	Carl Roth
Protein grade Agarose	Carl Roth
Ammonium persulphate (APS)	Carl Roth
Ampicillin (Amp) sodium salt	Carl Roth
Bovine serum albumin (BSA)	Sigma-Aldrich
Bromophenol blue sodium salt	Carl Roth
Boric acid	Carl Roth
Chloramphenicol (Cam)	Carl Roth
cOmplete™, EDTA-free Protease Inhibitor Cocktail	Sigma-Aldrich (Roche)
Coomassie Brilliant Blue G-250	Carl Roth
Dithiothreitol (DTT)	Carl Roth
Dimethyl sulfoxide (DMSO)	Carl Roth
Ethanol	Carl Roth
Ethylenediaminetetraacetic acid (EDTA)	Carl Roth
Glycerol	Carl Roth
Glycine	Carl Roth
Glutathione	Carl Roth
Hydrogen peroxide (H ₂ O ₂)	Carl Roth
Hydrochloric acid (HCl)	Carl Roth
Imidazole	Carl Roth
Isopropyl-β-D-thiogalactopyranoside (IPTG)	Carl Roth
Kanamycin sulfate (Kan)	Carl Roth

Magnesium chloride hexahydrate (MgCl ₂)	Carl Roth
Methanol	Carl Roth
Lysogeny broth medium (LB)	Carl Roth
Magnesium acetate	Carl Roth
Nickel (II) sulphate hexahydrate	Sigma-Aldrich
Orange G	Sigma-Aldrich
Polyethylene glycol 4000 (PEG4000)	Sigma-Aldrich
Ponceau S	Sigma-Aldrich
Sodium acetate	Carl Roth
Sodium borohydride	Sigma-Aldrich
Sodium chloride (NaCl)	Carl Roth
Sodium dodecyl sulfate (SDS)	Carl Roth
Sodium hydroxide (NaOH)	Carl Roth
Streptomycin sulfate (Strep)	Carl Roth
Terrific Broth (TB) medium	Carl Roth
Tetramethylethylenediamine (TEMED)	Carl Roth
Tris-(2-carboxyethyl)-phosphine (TCEP)	Carl Roth
Tris-(hydroxymethyl)-aminomethane (Tris)	Carl Roth
β-Mercaptoethanol	Applichem

2.1.3 Resins

Table 2.4: Resins

Type	Model	Supplier
Affinity matrix for maltose-amylose isolation	Amylose resin	New England Biolabs
Affinity matrix for GST tag purification	GST beads	Thermo Fisher Scientific
Immobilized metal-ion affinity chromatography resin	Protino® Ni-IDA	MACHEREY-NAGEL

2.1.4 Cloning material, enzymes, and recombinant proteins

Table 2.5: Cloning material, enzymes, and recombinant proteins

Designation	Supplier
2'- Deoxyadenosine 5'- triphosphate (dATP), sodium salt solution	Thermo Fisher Scientific
2'- Deoxycytidine 5'- triphosphate (dCTP), sodium salt solution	Jena Biosciences

2'- Deoxyguanosine 5'- triphosphate (dGTP), sodium salt solution	Thermo Fisher Scientific
2'- Deoxythymidine 5'- triphosphate (dATP), sodium salt solution	Thermo Fisher Scientific
BSA	New England Biolabs
<i>DpnI</i>	New England Biolabs
<i>Nt.BstNBI</i>	New England Biolabs
<i>XhoI</i>	New England Biolabs
<i>NsiI</i>	New England Biolabs
<i>PstI</i>	New England Biolabs
<i>BglII</i>	New England Biolabs
<i>NdeI</i>	New England Biolabs
<i>BsaXI</i>	New England Biolabs
GeneRuler™ 1 kb and 100 bp DNA Ladders	Thermo Fisher Scientific
HF buffer (PCR)	New England Biolabs
GC buffer (PCR)	New England Biolabs
Midori green Advance DNA stain	Biozym Scientific
Midori green Direct	NIPPON Genetics
NEBuffer™ 2 (Cloning)	New England Biolabs
NEBuffer™ 3.1 (DNA substrate preparation)	New England Biolabs
Nucleospin Gel and PCR cleanup kit	Macherey-Nagel
Nucleospin Plasmid kit	Macherey-Nagel
Phusion® high fidelity DNA Polymarse	Thermo Fisher Scientific
Standard Taq Reaction Buffer	New England Biolabs
T4 DNA Polymerase	New England Biolabs
Taq DNA Polymerase	New England Biolabs
T4 polnucleotide kinase	New England Biolabs
T4 DNA ligase	New England Biolabs
T4 DNA ligase buffer	New England Biolabs
ΔN hAAG	New England Biolabs
MutY	Trevigen
hOGG1	MyBioSource, Origene
hOGG1 cDNA	Dharmacon
GST tagged full length Myc (GST-Myc ^{FL})	Antibodies (https://www.antibodies.com/)
Full length Myc	AG Geyer, University of Bonn

Abbreviations: FL- Full length, cDNA- complementary DNA

2.1.5 Bacterial strains and plasmids

Table 2.6: Bacterial strains

Organism	Strain	Usage	Supplier
<i>Escherichia coli</i> (<i>E. coli</i>)	DH5 α	Cloning, plasmid amplification	Invitrogen
<i>E. coli</i>	BL21star (DE3)	Protein expression	Invitrogen
<i>E. coli</i>	BL21star (DE3)	Protein expression	-
<i>E. coli</i>	Rosetta2 (DE3)	Protein expression	Novagen

Table 2.7: Plasmids for protein expression and AFM DNA substrate preparation

Construct	Vector	Host	Setup	Resistance	Author/Supplier
hOGG1 _{wt} , hOGG1 _{K249Q} , hOGG1 _{C28A} , hOGG1 _{C241A} , hOGG1 _{C253A}	pETM-14	Bacterial	6xHis-3C-POI	Kanamycin	EMBL, Hamburg
Myc _{C1-163}	pETM-41	Bacterial	MBP-TEV-6xHis-POI	Kanamycin	EMBL, Hamburg
Lysine specific demethylase 1 (LSD1 _{FL})	pCDF-Duet1	Bacterial	Thioredoxin (trx)-6xHis-[3C]-POI	Streptomycin	Radhika Nair, RVZ Würzburg
Myc-associated factor X full length (Max _{FL})	pGEX 4T1	Bacterial	GST-Thr-POI	Ampicillin	Prof. Elmar Wolf, Biozentrum Würzburg
-	pUC19N	Bacterial	/	Ampicillin	S. Wilson, NIEHS, USA

Abbreviations: POI: protein of interest, TEV: TEV protease site, Trx: Thioredoxin, Thr: Thrombin cleavage site, EMBL: European Molecular Biology Laboratory, MBP- Maltose Binding Protein, GST- Glutathione S-Transferase, FL- Full length, WT- Wild type.

2.1.6 Oligonucleotides

All primers listed here were obtained from Sigma-Aldrich®. Abbreviations are given at the end of the Table. 2.8.

Table 2.8: List of primers used for generation of bacterial expression constructs

Name	Sequence (5'- 3')
hOGG1 _{wt} fw	TTCCAGGGGCCCCATGCCTGCCCCGCGC
hOGG1 _{wt} rv	GGATCCGGTACCTCATTAGTCCTAGCCTTCCGGCCCTTTGG
pETM14 hOGG1 _{wt} fw	TAGGACTAATGAGGTACCGGATCCGAA
pETM14 hOGG1 _{wt} rv	CATGGGCCCCTGGAACAGAAC
hOGG1 _{K249Q} fw	TGGGCACCCAAGTGGCTGAC
hOGG1 _{K249Q} rv	AGTCAGCCACTTGGGTGCCCA
oxoG E-box fw	GCATGCCTC(oxoG)AGTCTAGAGGCACGTGATCTGATCCTC TAGAGTGCACC
E-box fw	GTCTAGAGGCACGTGATCTGATCC
E-box rv	GGATCAGATCACGTGCCTCTAGAC
Myc ₁₋₁₆₃ fw	CAGGGCGCCCCGTATTTCTACTGC
Myc ₁₋₁₆₃ rv	GGATCCGGTACCTCATTAGTCCTAGGCGCTCAGATCCTGCA
pETM41 fw Myc ₁₋₁₆₃	TAGGACTAATGAGGTACCGGATCCGA
pETM41 rv Myc ₁₋₁₆₃	CGGGGCGCCCTGAAAATAAAGATTCTCGC
hOGG1 _{C28A} fw	CATCCCGGCTCCTCGCTCT
hOGG1 _{C28A} rv	GAGCGAGGAGCCGGGATG
hOGG1 _{C241A} fw	AGGCCCTCGCTATCCTGCC
hOGG1 _{C241A} rv	GGCAGGATAGCGAGGGCCT
hOGG1 _{C253A} fw	AGGTGGCTGACGCTATCTGCCT
hOGG1 _{C253A} rv	AGGCAGATAGCGTCAGCCACCT

Abbreviations: forward primer (fw), reverse primer (rv), 8-oxo-guanine (oxoG), and enhancer box (E-box).

DNA substrates designed for AFM studies, Fluorescence polarization experiments, Fluorescence Resonance Energy Transfer (FRET), hOGG1 activity assays, Electromobility shift assays (EMSA) are given in the Table. 2.9 below.

Table 2.9: List of oligonucleotides used in AFM, EMSA, ensemble FRET measurements, fluorescence polarization experiments, and hOGG1 activity assays

DNA substrate	DNA sequence (5'- 3')
Original bottom (AFM and activity assays)	GGT CGA CTC TAG AGG ATC AGA TCT GGT ACC TCT AGA CTC GAG GCA TGC
Cy3 bottom (EMSA and fluorescence polarization)	(Cy3) GGT CGA CTC TAG AGG ATC AGA TCT GGT ACC TCT AGA CTC GAG GCA TGC
AF647 bottom (Fluorescence polarization)	(AF647) GGT CGA CTC TAG AGG ATC AGA TCT GGT ACC TCT AGA CTC GAG GCA TGC
G: U top (AFM)	[Phos]GCA TGC CT (dU) GAG TCT AGA GGT ACC AGA TCT GAT CCT CTA GAG TCG ACC
oxoG:C top (AFM, EMSA and fluorescence polarization)	[Phos]GCA TGC CTC (oxoG) AG TCT AGA GGT ACC AGA TCT GAT CCT CTA GAG TCG ACC
G: A mismatch top (AFM)	[Phos]GCA TGC CTC GAG GCT AGA GGT ACC AGA TCT GAT CCT CTA GAG TCG ACC
oxoG:A mismatch top (AFM)	[Phos]GCA TGC C(oxoG)C GAG TCT AGA GGT ACC AGA TCT GAT CCT CTA GAG TCG ACC
Bottom for oxoG:A (AFM)	GGT CGA CTC TAG AGG ATC AGA TCT GGT ACC TCT AGA CTA GAG GCA TGC
ethenoA:T top (AFM)	[Phos]GCA TGC CTC G(ethenoA)G TCT AGA GGT ACC AGA TCT GAT CCT CTA GAG TCG ACC
Original top with E-box (EMSA and AFM)	[Phos]GCA TGC CTC GAG TCT AGA GGC ACG TGA TCT GAT CCT CTA GAG TCG ACC

Bottom strand with E-box (EMSA and AFM)	GGT CGA CTC TAG AGG ATC AGA TCA CGT GCC TCT AGA CTC GAG GCA TGC
oxoG:C – E-box top (AFM)	[Phos]GCA TGC CTC (oxoG)AG TCT AGA GGC ACG TGA TCT GAT CCT CTA GAG TCG ACC
Cy3 E-box comp bottom (EMSA)	(Cy3) GGT CGA CTC TAG AGG ATC AGA TCA GAT CAC GTG CCT CTA GAC TCG AGG CAT GC
Original top (EMSA, FRET and fluorescence polarization)	GCA TGC CTC GAG TCT AGA GGT ACC AGA TCT GAT CCT CTA GAG TCG ACC
Cy3 oxoG top incision (Activity assay)	(Cy3) GCA TGC CTC GAG TCT AGA GGT ACC A(oxoG)A TCT GAT CCT CTA GAG TCG ACC
oxoG in E-box top (EMSA)	(Cy3) GCA TGC CTC GAG TCT AGA GGC AC (oxoG) TGA TCT GAT CCT CTA GAG TCG ACC
Bottom strand (FRET)	(Cy3) CCT CTA GAC TCG AGG CAT GC (Cy5)

Abbreviations: Alexafluor 647 (AF647), Cyanine (Cy), Phosphate (Phos), 8-oxo-guanine (oxoG), enhancer box (E-box), and ethenoadenine (ethenoA). E-box sequence (CACGTG) is represented in blue color.

2.1.7 Antibodies and enzymes for protein purification, and protein ladder

Table 2.10: Antibodies

Antibody	Catalog number	Supplier
Mouse Anti-GST IgG	-	Covance
Qdot™ 605 goat F(ab') ₂ anti-mouse IgG conjugate (H+L)	Q11001MP	Invitrogen
Mouse Anti-His IgG	SAB1305538	Sigma Aldrich
Horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG	115-035-008	Dianova

Table 2.11: List of enzymes used for protein purification, recombinant proteins, and protein ladder

Name	Supplier
DNaseI	AppliChem
HRV-14, 3C protease	In-house production
Lysozyme	Carl Roth
PageRuler™ Prestained Protein Ladder	Thermo Fisher Scientific
Protease Inhibitor Tablets	Roche

2.1.8 Computer software, databases, servers, and deposited scripts

Table 2.12: List of software, servers and databases used for this study

Software	Model/Version	Author/Reference
Aekta Unicorn 5	Chromatography	GE Healthcare
Spectra Manager	CD-Spectroscopy	Jasco
Fluoromax 4	Ensemble FRET measurements	Horiba Jobin Yvon
GIMP 2.10.24	Cropping AFM images	Spencer Kimball, Peter Mattis
Igor Pro for MFP3D	AFM imaging	Asylum Research
Image J Fiji	Preprocessing for MatLab analysis	Wayne Rasband, NIH, USA
Image SXM 2.02	AFM volume analysis	Steve Barrett
Nanodrop spectrophotometer Software V3	Protein and DNA concentration estimation	Peqlab
OriginPro 9.0	Statistical analysis	Origin Lab Corporation
PyMOL 1.1 eval	Structural analysis	DeLano, 2002
Quantity One (Molecular Imager) 4.6.3	Image acquisition for gels	Biorad
FIESTA 1.05.0005	Skeletonization of DNA	B-cube, Dresden
MatLab R2019b	Position and bend angle analysis	Mathworks
AxioVision	Microscopy imaging for crystals	ZEISS
Cleario 1.2.0	Row to column transformation of the CLARIOstar data	Wolfgang Kölmel, RVZ Würzburg
ExpASy ProtParam	Protein parameters	Gasteiger, et al., 2005

GENTle 1.9.4 free software	DNA, Protein sequence handling generation of primers	Magnus Manske, University Cologne
MARS data analysis software	Fluorescence anisotropy data analysis	BMG Labtech
Microsoft Office 365 Excel, Word	Fluorescence anisotropy data analysis AFM data analysis Word processing	Microsoft Office
ODYSSEY	Imaging software	LI-COR
PDBePISA	Calculation of solvent accessibility of residues and protein- protein interfaces	Krissinel and Henrick, 2007
Protein Data Bank (PDB)	Protein structures	Rcsb.org
PubMed (NCBI)	Literature research	ncbi.nlm.nih.gov/pubmed/
UNICORN	Aekta control and data analysis	GE Healthcare
Uniprot	Information about proteins	Uniprot.org

Table 2.13: Deposited MatLab scripts on Open Science Framework (OSF)

Name	Link to Scripts
MatLab script for bend angle analysis at DNA lesions and DNA-bound protein complexes	https://osf.io/yhwuc/
Modified MatLab script including protein position measurements	https://osf.io/76e9s/

Step-By-Step guide for automated DNA bend angle analyses at DNA lesions and at protein complexes have been deposited on OSF (Appendix, Fig. x3.1, x3.2, and x3.3).

2.2 Methods

2.2.1 Protein preparation

2.2.1.a Cloning of protein constructs

PCR amplification. Sequence and Ligation Independent Cloning (SLIC)⁴, which uses *in vitro* homologous DNA recombination and single-strand annealing, was employed to clone all proteins in this study. Both inserts and vectors were PCR amplified by using a standard PCR reaction protocol for Phusion Polymerase. Primers for the PCR reactions were designed as shown in Table. 2.8 with overlapping 5' overhangs, in both insert and vector based primers. PCR reactions were set up in a total volume of 50 μ L. The PCR reaction setup for SLIC cloning and site directed mutagenesis is shown in the Table. 2.14. The insert was amplified from hOGG1 cDNA (for hOGG1 clones) and pGEX 4T1 vector (a kind gift from Prof. Elmar Wolf, Biozentrum, University of Würzburg) containing Myc gene (for Myc₁₋₁₆₃ clone). PCR conditions are shown in the Table. 2.15.

Table 2.14: PCR reaction set up for SLIC cloning and site directed mutagenesis

PCR component	Concentration
Forward primer	1 μ M
Reverse primer	1 μ M
dNTPs	100 μ M
Phusion® polymerase	1 Unit
HF buffer	1x
Template DNA	~100 ng
Ultrapure water	Make up to 50 μ l

Table 2.15: PCR conditions used for gene amplification and gene modification.

Step	Temperature ($^{\circ}$ C)	Time (seconds)
Initial denaturation	98	30
25 Cycles	98	10
	55-65	30
	72	30 sec/kb
Final extension	72	120
Hold	4	∞

Cloning strategy using SLIC method. SLIC was used for sub-cloning of the GOI (Gene Of Interest) into different vectors. Two pairs of primers were designed- one for the linearization of the vector and the other for the amplification of the GOI), with 9-12

nucleotides annealing overhangs. After PCR amplification, the products were treated with restriction enzyme *DpnI* overnight at ambient temperature to digest the template DNA and increase ligation efficiency. Further, the PCR products were purified using the Nucleospin PCR Clean-up kit (Macherey-Nagel) and the DNA concentration was measured using a spectrophotometer (Nanodrop ND1000, Peqlab). Lastly, a 5 μ l aliquot of the final product was transformed into chemically competent *E. coli* DH5 α cells. The final concentrations of PCR products were estimated to be 123 ng/ μ l (hOGG1_{wt}) and 106 ng/ μ l (Myc₁₋₁₆₃).

Site-directed mutagenesis (SDM) and gene modification. For introducing single point mutations (SNPs), a standard Phusion polymerase PCR protocol was used. Briefly, SDM was performed by a double primer method where both forward and reverse primers were designed to contain the mutation to obtain an error-free mutant. The PCR reaction set up is shown in the Table. 2.14. PCR conditions used are described in the Table. 2.15. Further steps from overnight *DpnI* digestion of the PCR products to chemical transformation remained the same as described in the above section. The final concentrations of PCR products were estimated to be 155 ng/ μ l (hOGG1_{K249Q}), 98 ng/ μ l (hOGG1_{C28A}), 139 ng/ μ l (hOGG1_{C241A}), and 85 ng/ μ l (hOGG1_{C253A}).

DNA analysis by gel electrophoresis and ultraviolet-visible spectroscopy. DNA quality and composition were assessed by DNA agarose gel electrophoresis. The gels contained 1% (w/v) NEEO ultra-quality agarose, 1x TAE buffer and Midori Green Advance (3 μ L/ 50 mL gel). DNA samples were mixed with 6x DNA loading buffer (final concentration: 1x) and subjected to gel electrophoresis in 1x TAE buffer for 30-45 min at a voltage of 100 V. DNA fragments were visualized with the electrophoresis gel imaging cabinet Universal Hood II (Biorad) using a laser to excite the fluorescence of Midori Green, which had bound to DNA. DNA length was estimated by comparison with a DNA ladder (GeneRuler 1 kbp). DNA concentrations were determined by ultraviolet-visible (UV-VIS) spectroscopy using a spectrophotometer (Nanodrop ND1000) and an extinction coefficient $\epsilon_{\text{DNA}} (260 \text{ nm})$ of 0.02 mL/ μ g \cdot cm (as suggested by the Nanodrop V3 software) and a path length of 1 cm.

Chemical transformation. Aliquots of 50 μ L chemically competent *E. coli* DH5 α cells (for plasmid isolation and DNA substrate preparation, Table. 2.6) or *E. coli* expression strains (for protein expression, Table. 2.6) were incubated with 10-100 ng of the DNA containing the gene of interest on ice for 30 minutes. Afterward the cells were subjected to a 90 s heat shock at 42°C in a thermomixer. Subsequently, the cells were incubated on ice for three minutes, before 400 μ L LB medium were added, and the bacteria were shaken at 200 rpm and a temperature of 37°C for 60 minutes. Afterward an aliquot was added to either 5 mL, 10 mL, 50 mL, or 100 mL LB medium in appropriately sized flasks with the required antibiotics (Table. 2.7). The volume of LB medium was chosen depending on the purpose as follows: Purification of DNA (5, 10, and 100 mL), or precultures for expression tests (50 mL) or for obtaining protein for scaled-up purification (100 mL). All samples were incubated overnight at 37 °C in a small-scale shaker with a shake speed of 200 rpm. In

case of aiming to obtain single colonies for analysis of plasmid sequences, a 100 μ L aliquot was added to an LB-agar plate containing the appropriate antibiotic(s) for selection.

Plasmid isolation. Single colonies obtained after chemical transformation were transferred to LB-medium with the appropriate antibiotics. The cultures were shaken overnight at 200 rpm at 37°C and centrifuged at 4,000 g for 10 min at 4°C. The cell pellet was subjected to DNA isolation protocol following the manufacturer instructions of the NucleoBond Plasmid Kit (Macherey-Nagel). The resulting DNA was sent for sequencing with specific or standard primers to the Microsynth SeqLab (<https://srvweb.microsynth.ch>). The sequencing results for all constructs are shown in the Fig. x5.2 in the appendix section.

2.2.1.b Recombinant protein expression

The pre-culture cell suspension was used to scale up the culture, using a 1:100 dilution, into 2 L of LB medium in 5 L flasks for large-scale protein production. The cultures were incubated at 37°C for 4-5 hours or until an OD_{600nm} of ~ 0.8 was reached. Expression of protein was then induced with IPTG (0.5 M final concentration). Post-induction temperature for all proteins was 15°C. Depending on the protein, each construct had its own conditions for optimum production of protein, which are summarized in Table. 2.16.

Table 2.16: Expression strains, antibiotics, and times after induction with IPTG for different proteins

Construct	Vector/Antibiotic	<i>E. coli</i> strain	Duration (h) of induction
hOGG1 _{wt} and mutants	pETM-14/Kan	BL21(DE3)	16-18h
MyC ₁₋₁₆₃	pETM-41/Kan	BL21 (DE3)	18-20h
LSD1 _{FL}	pCDF-Duet1/Strep	Rosetta (DE3)	15-18h
MaX _{FL}	pGEX 4T1/Amp	BL21 (DE3)	15-16h

2.2.1.c Protein purification

Cell lysis. The cell pellet from 8 L culture was resuspended at 4°C in 200 mL lysis buffer (for buffers, see Table. 2.17) supplemented with one Tablet of the Roche EDTA-free cOmplete protease inhibitor cocktail, 300 μ l DNase I solution (~250 UI/ μ L), 75 mg of PMSF and lysozyme (1mg/ml). Cells were lysed in two cycles using a mechanical cell disruptor at ~1,500 bar and the lysate was cleared by centrifugation (1 h at 35,000 g, at 4°C).

Chromatography. The supernatants with the N-terminal 6xHis-tagged protein or GST-tagged protein were applied twice to a gravity flow column containing (4-7 g) Protino® Ni-IDA resin or GST beads (5 g) respectively. This was followed by a single washing step with 30 ml of high salt wash buffer (Table. 2.17). The subsequent protein elution (40 ml) was performed using two steps, the first and second elution buffers containing 500 mM

imidazole (20 ml) and 750 mM imidazole (20 ml) for the 6x-His tagged proteins. In the case of Max, with the N-terminal GST tag, the elution was performed in two steps, first with 15 mM glutathione (15 ml) followed by 25 mM glutathione (15 ml). Affinity chromatography was performed at 4°C, collecting the column flow through of each step on ice. Analysis of aliquots representing the different purification steps was carried out by SDS polyacrylamide gel electrophoresis (SDS-PAGE). The elution fractions containing the protein of interest were pooled and dialyzed against specific dialysis buffers to remove imidazole or glutathione (Table. 2.17). Proteins were concentrated for further steps of purification. Proteins requiring only one step (affinity chromatography) purification were concentrated for flash-freezing and further storage.

Table 2.17: Buffers for cell lysis, affinity chromatography, ion exchange chromatography, and size exclusion chromatography

Name	hOGG1 _{wt} , hOGG1 _{K249Q} , hOGG1 _{C28A} , hOGG1 _{C241A} , hOGG1 _{C253A}	Myc ₁₋₁₆₃	Max	LSD1
Lysis buffer	20 mM Tris-HCl, pH 7.5, 300 mM NaCl, 5 mM imidazole, 10% glycerol	20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 1 mM TCEP, 10% glycerol	20 mM HEPES, pH 7.9, 500 mM NaCl, 1 mM β-ME, 10% glycerol	50 mM HEPES, pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% glycerol
Wash buffer	20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 5 mM imidazole, 10% glycerol	20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1 mM TCEP, 10% glycerol	20 mM HEPES, pH 7.9, 500 mM NaCl, 1 mM β-ME, 10% glycerol	50 mM HEPES, pH 8.0, 500 mM NaCl, 1 mM TCEP, 5% glycerol
Elution buffer 1	20 mM Tris-HCl, pH 7.5, 300 mM NaCl, 500 mM imidazole, 1	20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 500 mM imidazole, 1	20 mM HEPES, pH 7.9, 500 mM NaCl, 15 mM glutathione, 1	50 mM HEPES, pH 8.0, 300 mM NaCl, 500 mM imidazole, 1

	imidazole, 10% glycerol	mM TCEP, 10% glycerol	mM β -ME, 10% glycerol	mM TCEP, 5% glycerol
Elution buffer 2	20 mM Tris- HCl, pH 7.5, 300 mM NaCl, 750 mM imidazole, 10% glycerol	20 mM Tris- HCl, pH 8.0, 300 mM NaCl, 750 mM imidazole 1 mM TCEP, 10% glycerol	20 mM HEPES, pH 7.9, 500 mM NaCl, 25 mM glutathione, 1 mM β -ME, 10% glycerol	50 mM HEPES, pH 8.0, 300 mM NaCl, 750 mM imidazole, 1 mM TCEP, 5% glycerol
Dialysis buffer	20 mM Tris- HCl, pH 7.5, 200 mM NaCl, 10% glycerol	20 mM Tris- HCl, pH 8.0, 200 mM NaCl, 1 mM TCEP, 5% glycerol	20 mM HEPES, pH 7.9, 250 mM NaCl, 5% glycerol	50 mM HEPES, pH 8.0, 150 mM NaCl, 1 mM TCEP, 5% glycerol
Equilibration buffer	-	20 mM Tris- HCl, pH 8.0, 1 mM TCEP, 5% glycerol	-	50 mM HEPES, pH 8.0, 1 mM TCEP, 5% glycerol
Buffer A	-	20 mM Tris- HCl, pH 8.0, 1 mM TCEP, 5% glycerol	-	50 mM HEPES, pH 8.0, 1 mM TCEP, 5% glycerol
Buffer B	-	20 mM Tris- HCl, pH 8.0, 1 M NaCl, 1 mM TCEP, 5% glycerol	-	50 mM HEPES, pH 8.0, 1 M NaCl, 1 mM TCEP, 5% glycerol
SEC buffer	20 mM Tris- HCl, pH 7.5, 200 mM NaCl, 5% glycerol, 0.5 mM DTT	20 mM Tris- HCl, pH 8.0, 200 mM NaCl, 5% glycerol, 0.5 mM DTT	-	50 mM HEPES, pH 8.0, 150 mM NaCl, 5% glycerol, 0.5 mM DTT

Myc¹⁻¹⁶³ and LSD1 were purified using an additional anion exchange chromatography step (column MonoQ 10/100 GL) before the final SEC step. The column was pre-equilibrated with buffer A and eluted with a 30-column volume (CV) linear gradient of buffer A to 30% buffer B concentration in the case of Myc¹⁻¹⁶³ and to 40% buffer B concentration for LSD1. Finally, proteins were equilibrated with a low salt equilibration buffer to obtain a salt concentration of ~90 mM. The proteins were then applied to the ion exchange column pre-equilibrated with buffer A and eluted with a 20 CV linear gradient of buffer A to 100% buffer B. Fractions containing the protein were pooled, concentrated, and further purified by SEC (SD 200 16/600 GL). Pure protein fractions were then pooled, concentrated, and stored at -80°C after flash freezing. Max was quite pure as seen from the SDS-PAGE analysis of affinity chromatography elutions (Fig. 2.4). It was further concentrated, flash-frozen and stored at -80°C. hOGG1_{wt} and hOGG1_{K249Q} mutant were subjected to SEC immediately after dialysis (Fig. 2.1 and 2.2). The fractions were analyzed by SDS-PAGE and those with a purity > 95% were pooled and stored at -80°C after flash freezing. Purification steps for individual proteins are described in the below sections.

- i. hOGG1_{wt}.* After successful cloning of hOGG1_{wt} gene in pETM14 plasmid was confirmed by sequencing for positive colonies, they were selected for protein expression and purification. After immobilized metal affinity chromatography, all fractions were loaded onto an SDS-PAGE gel and elutions containing hOGG1_{wt} were pooled for size exclusion chromatography (SEC) (Fig. 2.1). After SEC, hOGG1_{wt} was concentrated, flash frozen in liquid nitrogen and stored at a concentration of around 5 mg/ml until further use. Batch II (20 mg/ml) and Batch III (18 mg/ml) protein purifications were performed similarly. Experiments with hOGG1 from Batch I are shown in the chapter 5, section 5.2.

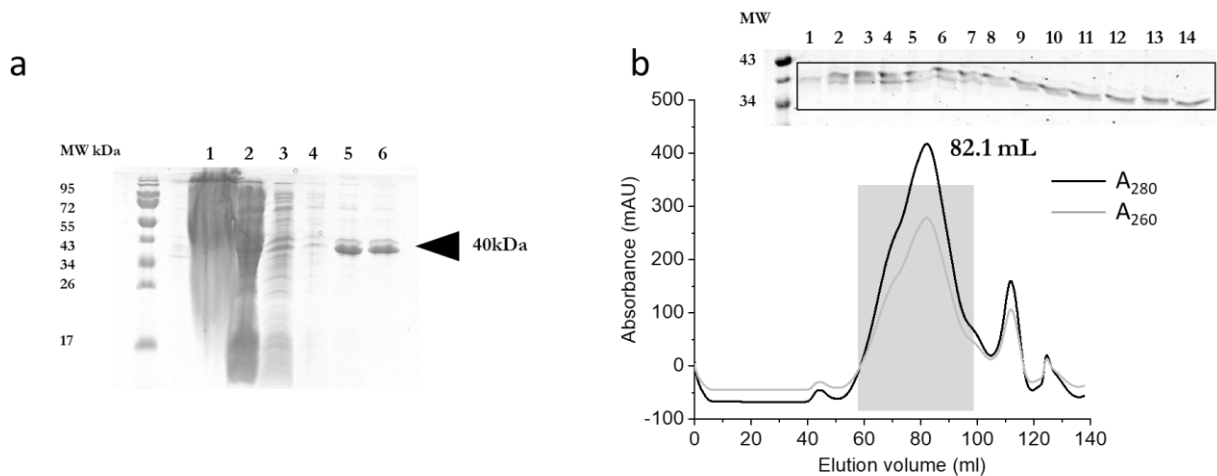


Figure 2.1: Immobilized metal affinity chromatography (IMAC) and SEC purification of hOGG1_{wt}. (a) SDS gel profile showing fractions from IMAC purification of hOGG1_{wt}. Samples from left to right: (1) Insoluble fraction, (2) Flow through, (3) and (4) Washes with lysis buffer, (5) 500 mM Imidazole elution, and (6) 750 mM Imidazole elution. Black arrow indicates the purified protein. (b) Size exclusion chromatography of hOGG1_{wt}. Sample purity was analyzed by SDS-PAGE (15% gel) and fractions 1-14 were pooled to use in all experiments. Fractions from the shaded grey area were used for SDS-PAGE analysis.

- ii. *hOGG1*_{K249Q}.** After successful mutagenesis of *hOGG1*_{wt} gene in pETM14 plasmid to incorporate K249Q mutation was confirmed by sequencing for positive colonies, they were selected for protein expression and purification. After immobilized metal affinity chromatography, all fractions were loaded onto an SDS-PAGE gel and elutions containing *hOGG1*_{K249Q} were pooled for size exclusion chromatography (SEC) (Fig. 2.2). After SEC, *hOGG1*_{K249Q} was concentrated, flash frozen in liquid nitrogen and stored at a concentration of 6 mg/ml until further use.

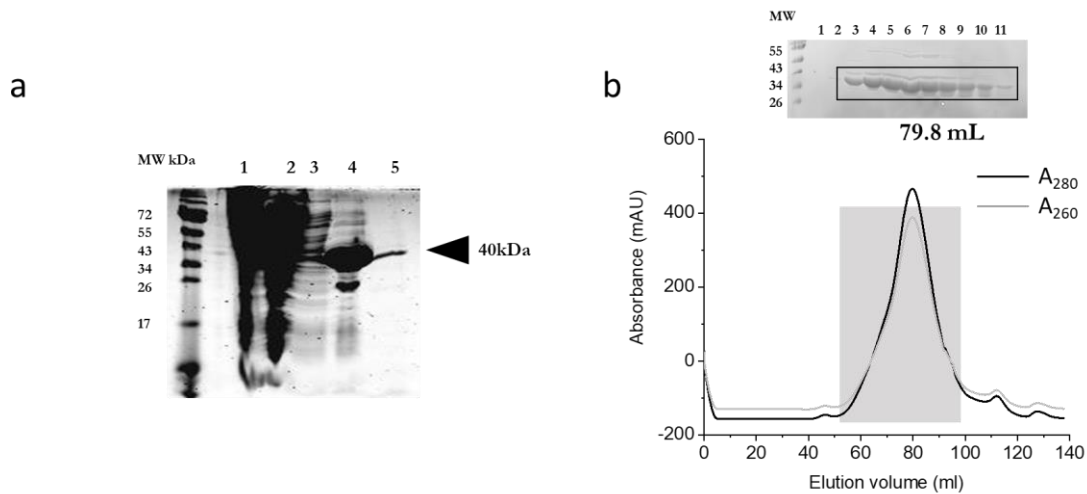


Figure 2.2: IMAC and SEC purification of *hOGG1*_{K249Q}. (a) SDS gel profile showing fractions from IMAC purification of *hOGG1*_{K249Q}. Samples from left to right: (1) Insoluble fraction, (2) Flow through, (3) Wash with lysis buffer, (4) 500 mM Imidazole elution, and (5) 750 mM Imidazole elution. Black arrow indicates the purified protein. (b) Size exclusion chromatography of *hOGG1*_{K249Q}. Sample purity was analyzed by SDS-PAGE (15% gel) and fractions 3-11 were pooled to use in the future experiments. Fractions from the shaded grey area were used for SDS-PAGE analysis.

- iii. *LSD1*.** The *LSD1* construct was obtained from Radhika Karal Nair from Prof. Caroline Kisker's group at the RVZ. *LSD1* full length protein is cloned in pCDF vector and contains a 6X-His tag at the N-terminus.

LSD1 was expressed in Rosetta (DE3) cells (Table. 2.16). After metal affinity chromatography, elutions with *LSD1* were pooled and further purified by ion exchange chromatography and SEC (Fig. 2.3). *LSD1* was concentrated, flash frozen in liquid nitrogen and stored at a concentration of 15 mg/ml until future experiments. For the Pull-down assays, the 6X-His tag was removed for *LSD1*.

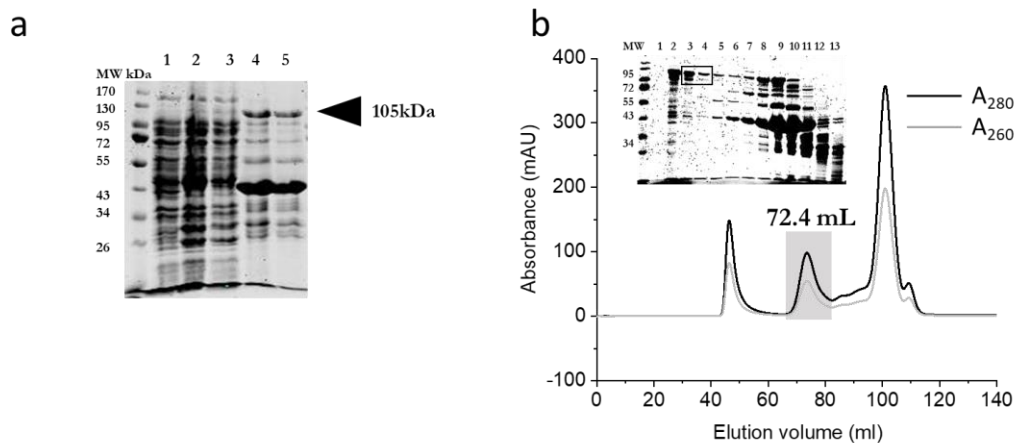


Figure 2.3: IMAC and SEC purification of LSD1. (a) SDS gel profile showing fractions from IMAC purification of LSD1. Samples from left to right: (1) Insoluble fraction, (2) Flow through, (3) Wash with lysis buffer, (4) 500 mM Imidazole elution, and (5) 750 mM Imidazole elution. Black arrow indicates LSD1 in the elution fractions. (b) Size exclusion chromatography of LSD1. Sample purity was analyzed by SDS-PAGE (15% gel) and fraction 3 and 4 was concentrated for further use. Fractions from the shaded grey area were used for SDS-PAGE analysis.

iv. Max. The Max construct was obtained from Prof. Elmar Wolf's lab. It is cloned in the plasmid pGEX 4T1 containing the GST tag. Max was expressed in BL21(DE3) cells. One step affinity purification was performed by using immobilized GST beads⁵. Pure fractions were pooled, dialyzed, concentrated, flash frozen in liquid nitrogen and stored at a concentration of 28.5 mg/ml until further use. Fig. 2.4 shows SDS PAGE gel of Max affinity purification.

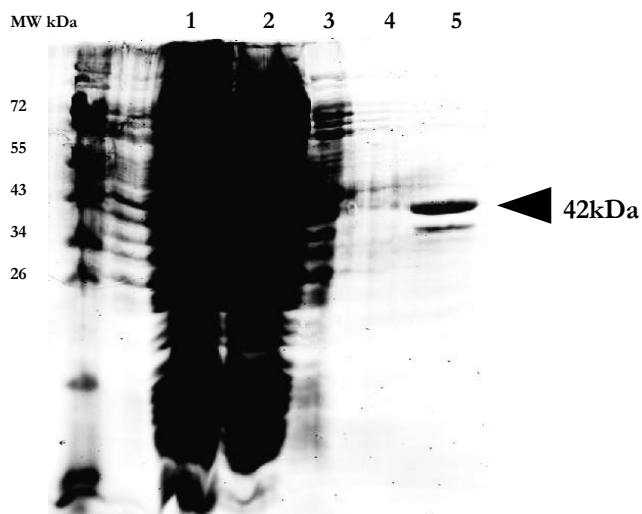


Figure 2.4: SDS gel profile showing fractions from IMAC purification of Max. Samples from left to right: (1) Insoluble fraction, (2) Flow through, (3) Wash with lysis buffer, (4) 15 mM glutathione elution, and (5) 30 mM glutathione elution. Black arrow indicates the purified protein.

- v. **c-Myc¹⁻¹⁶³.** c-Myc stands for cellular Myc and will be referred to as Myc in the next sections. Myc_{CFL} protein posed many hurdles during expression and purification steps. Expression levels were extremely basal even after many attempts by changing expression strains, IPTG concentrations, and induction temperatures. Extremely low concentrations of Myc were obtained from affinity chromatography step. Continuation with further purification steps led to complete loss of the protein. C-terminus domain of Myc is a basic helix-loop-helix leucine zipper domain known to interact with DNA. However, N-terminal domain of Myc could be stably expressed⁶ (the clone was a gift from Prof. Elmar Wolf, Biozentrum Wuerzburg). Myc₁₋₁₆₃ was used in the native polyacrylamide gel electrophoresis and in vitro pull-down experiments instead of full-length Myc. After successful cloning of Myc₁₋₁₆₃ gene in pETM41 plasmid (containing an MBP tag) was confirmed by sequencing for positive colonies, they were selected for protein expression and purification. Fig. 2.5 shows the size exclusion chromatogram and subsequent SDS PAGE gel for Myc₁₋₁₆₃ purification. Pure fractions were pooled, concentrated, flash frozen in liquid nitrogen and stored at a concentration of 1.3 mg/ml until further use.

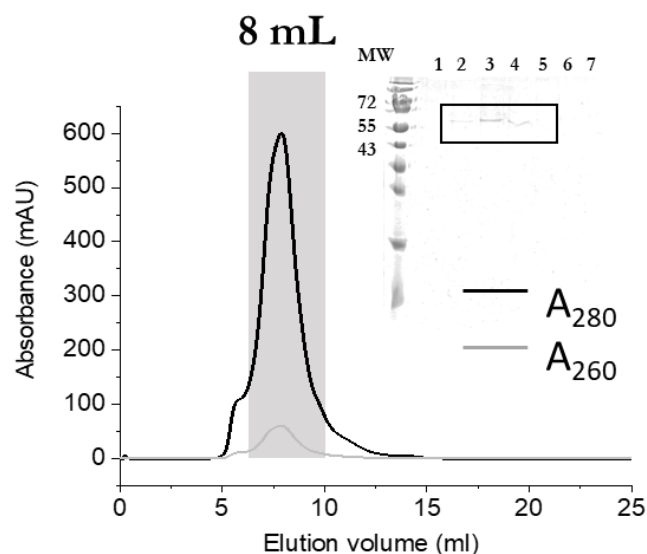


Figure 2.5: Size exclusion chromatography of Myc₁₋₁₆₃. Sample purity was analyzed by SDS-PAGE (15% gel) and fraction 2-5 were concentrated for further use. Fractions from the shaded grey area were used for SDS-PAGE analysis.

All chromatograms were exported to MS excel and were plotted using Origin Pro software. All proteins were purified to $\geq 95\%$ purity as judged by 15% SDS-PAGE. Full length Myc was obtained from AG Geyer, ISB Bonn. In addition, GST tagged full length Myc was purchased from Antibodies (<https://www.antibodies.com/>). The purification profiles for hOGG1_{C28A}, hOGG1_{C241A} and hOGG1_{C253A} and different batches of hOGG1_{wt} purification are shown in the appendix section, Fig. x5.1. The purification steps for all constructs of hOGG1 were the same. Different batches of

hOGG1_{wt} showed different equilibria between monomer and dimer forms regardless of same buffer conditions throughout purifications. hOGG1_{wt} will be referred to as hOGG1 in the next sections.

2.2.2 Protein analysis

SDS-polyacrylamide gel electrophoresis. For the analysis of proteins by SDS-PAGE 15% acrylamide gels were used. The 15 % acrylamide solution for the running gel was made with a 30% acrylamide solution (with 37.5:1 acrylamide:bis-acrylamide ratio), 1.5 M Tris pH 8.8 and 10% SDS. The 5 % acrylamide solution for the stacking gel was made with 30% acrylamide solution, 1.0 M Tris pH 6.8 and 10% SDS. As catalyzers of the polymerization reaction 6 μ L TEMED (Sigma-Aldrich) and 50 μ L of 10% (w/v) ammonium persulfate solution were used for every 5 mL of polyacrylamide gel made.

The gel was set up with a wing clamp assembly (Bio-Rad). Gels were run at 200 V for 40 minutes at room temperature. For protein detection, a Coomassie G 250 staining solution (80 mg of Coomassie G 250 in 1 L of ddH₂O acidified with 3 mL of 37% HCl) was used. After repeated heating in ddH₂O solution in a microwave at 800 W for 60 sec and a subsequent 10-minute incubation in Coomassie G250, destaining was done in water with an initial heating step in the microwave. Proteins bands were analyzed using the protein ladder.

UV/Vis spectrophotometry. DNA and protein concentrations were measured spectrophotometrically using the NanoDrop™ 1000 instrument. Prior to each measurement, a blank was performed utilizing the corresponding buffer of the sample. DNA absorbance was measured at 260 and 280 nm and the purity of the samples was determined by the ratio of the absorbances at 260 nm and 280 nm. For protein concentrations (mg ml⁻¹), the absorbance at 280 nm with a path length of 1 cm was measured and divided by the calculated extinction coefficients at 280 nm for each protein, assuming all cysteine residues being in the reduced state (No disulfide bonds). Extinction coefficients and molecular weights of each protein construct were obtained from the ExPASy ProtParam website and are listed in the Table. 2.18.

Table 2.18: Extinction coefficients and molecular weights of the purified proteins

Protein	Extinction coefficient ($10^3 \text{ M}^{-1}\text{cm}^{-1}$) assuming all Cys residues are in the reduced state	Molecular weight (kDa)
6xHis-hOGG1	68.4	40
GST-Max	26.4	44
MBP-Myc ₁₋₁₆₃	87.8	62
6xHis-LSD1	80.8	105

Analytical size exclusion chromatography. Analytical size exclusion (gel filtration) chromatography (aSEC) was performed to determine dimerization of hOGG1 under oxidizing and reducing conditions using hOGG1 gel filtration buffer containing either 5 μM H_2O_2 or 5 mM DTT for respective runs. For oxidative condition, hOGG1 was treated with H_2O_2 for 3 h and dialyzed overnight and hydrogen peroxide was removed by dialysis. Oxidized hOGG1 was concentrated and stored at -80°C . This oxidized hOGG1 was used for the analytical run. SD200 pg 16/600 gl column was equilibrated with 3 CV of gel filtration buffer prior to hOGG1 injection. 30 μM hOGG1 was injected for each condition after incubating with DTT for 30 min on ice. For ambient condition, hOGG1 was directly injected without any incubation. Samples were subjected to centrifugation at 10,000 g prior to injection and analytical runs were performed at 0.5 ml/min. Dimerization was detected on the chromatograms by following the absorbance at 280 nm as a function of time. The chromatograms were exported to MS excel and were plotted in Origin Pro software (Fig. 5.7 and Appendix Fig. x5.6).

Circular dichroism spectroscopy. To confirm the correct folding of hOGG1 and its mutants, CD measurements were performed with a JASCO (J-810) spectropolarimeter at 20°C . Far UV-spectra from 190 to 260 nm were recorded at a scanning speed of 20 nm/min with a response time of 1 s, a bandwidth of 2 nm, and a path length of 0.1 cm (Hellma Analytics). hOGG1 samples were diluted in Sorenson buffer (71.5 ml of 0.133 M Na_2HPO_4 and 28.5 ml of 0.133M KH_2PO_4 , pH 7.2) to 2.5 μM . hOGG1 was oxidized in presence of 5 μM H_2O_2 and reduced in presence of 5 mM DTT for CD experiments. For analysis, the buffer spectrum was subtracted as reference from protein spectra. The spectra were exported to MS excel and were plotted using Origin Pro software (Fig. 5.4b).

SDS-PAGE studies to investigate hOGG1 dimerization. hOGG1 was diluted to 5 μM in gel filtration buffer for hOGG1. hOGG1 was oxidized in presence of 5 μM H_2O_2 for oxidizing samples and reduced in presence of 5 mM DTT for reducing samples prior to mixing them with 1x SDS sample buffer without β -mercaptoethanol. Samples were denatured for 5 min at 95°C and then 5 μl of the samples were run on a self-made 15% SDS gel (For gel composition, see first section under 2.2.2) for ~50 min at 250 V at room

temperature until the loading dye reached the bottom of the gel. The samples were loaded along with a protein marker for better visualizing both monomer and dimer form of hOGG1. For protein detection, a Coomassie G 250 staining solution (80 mg of Coomassie G 250 in 1 L of ddH₂O acidified with 3 mL of 37% HCl) was used. After repeated heating in ddH₂O solution in a microwave at 800 W for 60 sec and a subsequent 10-minute incubation in Coomassie G250, destaining was done in water with an initial heating step in the microwave. Proteins bands were analyzed using the protein ladder. Experiments were performed in triplicate (Fig. 5.4 and Appendix Fig. x5.3, x5.7a).

Sedimentation velocity analytical ultracentrifugation. Sedimentation velocity analytical ultracentrifugation (SV-AUC) experiments and data analysis were performed in-house. hOGG1 was treated with 100 μ M H₂O₂ and 5 mM DTT for 15 min on ice. For SV-AUC measurements, protein solutions were diluted in hOGG1 gel filtration buffer to a concentration of 100 μ M. 400 μ l of the protein sample as well as the same amount of storage buffer for reference were pipetted into standard double-sector charcoal-filled Epon centerpieces equipped with sapphire windows (Beckman Coulter). The centrifuge used for performing SV-AUC experiments was a Beckman Coulter analytical ultracentrifuge Optima XL-I with an eight-hole An-50 Ti rotor at 40,000 rpm and at ambient temperature overnight. Data were collected in continuous mode at a step size of 6 min using interference optical detection at a wavelength of 280 nm. Using the software SEDFIT, continuous sedimentation co-efficient distributions were obtained as solutions to the Lamm equation $c(s)$ ⁷.

Analysis was performed with regularization at confidence levels 0.74, choosing the floating frictional ratio (f/f_0), time-independent noise, baseline, and meniscus position. The resulting fits to the data displayed RMSD values of ≤ 0.01 . Best fits were achieved with frictional ratios f/f_0 of ~ 1.83 for hOGG1 under ambient conditions, ~ 1.96 for hOGG1 under oxidizing conditions and ~ 1.68 for hOGG1 under reducing conditions. Results were exported to Origin Pro software to obtain the final sedimentation co-efficient values for each sample.

Solvent accessibility determination for hOGG1. The crystal structure of hOGG1 (PDB id: 1ebm) was used to determine interface areas, interface residues and accessible and buried surface area of these residues in an artificially generated hOGG1 dimer by PDBePISA (Proteins, Interfaces, Structures and Assemblies) server. The PDB id of the crystal structure of hOGG1 was entered into the Structure Analysis section of PDBePISA server. After the processing time, Interfaces, Monomers and Assemblies can be accessed by clicking on corresponding buttons. Under the Interfaces section, several hits are generated based on the Complexation Significance Score, which indicates how significant for assembly formation the interface is. The score is defined as a maximal fraction of the total free energy of binding that belongs to the interface in the stable assemblies. The first hit was selected, and details button was clicked. The next tab depicts the inaccessible and solvent accessible residues in navy blue and sky blue colours respectively with values for

accessible surface area (ASA) and buried surface area (BSA). The cysteine residues with the highest ASA values (C28, C241, C253) were selected for further mutational studies.

Crystallization of oxidized hOGG1. To visualize hOGG1 dimer, purified hOGG1 was subjected to oxidation in presence of 5 μM H_2O_2 in the hOGG1 gel filtration buffer at 4°C and used for crystallization trials. Oxidized hOGG1 was initially screened at concentrations of 5, 10 and 20 mg/ml employing the Index screen by sitting drop method using the 96 well plates (0.3 μl protein solution and 0.3 μl precipitant solution were mixed and equilibrated over 40 μl precipitant solution). Precipitation of protein was observed in the drops with 10 and 20 μM concentrations of protein. In the next step, crystallization trays were set up at hOGG1 concentration of 15 mg/ml using sitting drop methods with screens mentioned in the Table. 2.19. Microcrystals obtained in the nucleix suite (duration of 75 days) were optimized by hanging drop method by setting up a grid condition around mother liquor (0.05 M HEPES sodium salt pH 7.0, 0.1 M Potassium chloride, 0.01 M Calcium chloride and 10% (v/v) PEG 400). However, crystallization trials were unsuccessful so far.

Table 2.19: List of crystallization screens used in this study

Name	Supplier of original formulation
Crystal Screen™ 1+2	Hampton Research
Index	Hampton Research
JCSG+	Molecular Dimensions
Nucleix Suite	Qiagen
Wizard 1+2	Emerald BioSystems
Wizard 3+4	Emerald BioSystems

Mass spectrometry. Gel electrophoresis and Mass spectrophotometry were performed by Dr. Stephanie Lamer (AG Schlosser, RVZ, University of Wuerzburg). hOGG1 was treated with acetone and light NEM labeling of free cysteine residues was performed. N-Ethylmaleimide (NEM) is commonly used to block the free cysteine residues in proteins⁸. Another acetone wash was performed to get rid of the light NEM. Subsequently, the sample was treated with TCEP to free cysteines that were involved in disulfide bond formation. These cysteines were labelled with heavy NEM, and SDS-PAGE analysis was performed for hOGG1 under ambient conditions and the band was excised and treated with trypsin and elastase and mass spectrometry was performed to investigate the cysteine(s) residue involved in disulfide bond formation.

Western blot experiments to identify dimerization for hOGG1_{C28A}, hOGG1_{C241A} and hOGG1_{C253A}. To verify whether hOGG1_{C28A} could dimerize or not, 8 μ M purified hOGG1_{C28A}, hOGG1_{C241A} and hOGG1_{C253A} were oxidized and reduced by adding 5 μ M H₂O₂ and 5mM DTT respectively. A non-reducing 15% SDS-PAGE gel was run for these samples until the dye front reached the bottom of the gel.

For blotting, nitrocellulose membranes were used, which were first put into transfer buffer (25 mM Tris/HCl pH 8.3 190 mM glycine, 20% ethanol) for 5 minutes, while the SDS-PAGE was ongoing. Next, the gel together with the pre-activated membrane were assembled in the Mini Trans-Blot cell. The transfer of the proteins from the gel to the membrane was achieved by applying an electric current of 300 mA for 1 hour at 4°C.

To block the nitrocellulose membrane, it was incubated for 1 hour in TBS (50 mM Tris/HCl pH 7.5 150 mM NaCl)-Albumin Fraction V 2.5% (w/v) followed by washing with TBS-T (50 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.05% (w/v) Tween-20). Afterward, the membrane was incubated overnight with the primary antibody at 4°C (usually diluted 1:1000 in TBS buffer following the manufacturer's instructions). The primary antibody was an anti-His antibody (**SAB1305538, Sigma-Aldrich**).

The next day, two 5 min rinsing steps with TBS were conducted followed by two rinsing steps with TBS-T of equal duration. Subsequently, the blots were incubated with the secondary antibody (usually diluted 1:10000 in TBS). As secondary antibodies, goat anti-mouse IgG was used for the detection of monomer and dimer bands. After incubation with the secondary antibody, the blots were washed three times with TBS-T, 10 min each and the blots were developed by chemiluminescence detection using the kit containing NBT/BCIP and bands were detected over color development. The blot was visualized for dimer bands in the Vilber imaging system.

2.2.3 Atomic force microscopy

2.2.3.1 DNA substrate preparation for AFM imaging

All the AFM substrates used in this study were derived from the circular dsDNA 2,729 bp long pUC19N plasmid, a modified pUC19 plasmid⁹ to contain three nickase *Nt.BstNBI* sites in close proximity. Lesion containing substrates were prepared by first nicking pUC19N plasmid with *Nt.BstNBI* enzyme at 55°C for 2 h 30 min. This was followed by heat inactivation step performed at 85°C for half hour. The nicked 48 bases long DNA was melted out of the plasmid by performing repeated heating at 68°C for 20 min followed by centrifugation steps using a 100kDa molecular weight cut-off filter microcon at 10,000 x g at room temperature up to 8 cycles. All these steps were performed in the presence of an excess (10 x) counter oligonucleotide- bottom 48 nt long strand complementary to the top strand between the two *Nt.BstNBI* sites (Table. 2.9). This process is referred to as DNA gapping. During this approach, the nicked 48 nt long top strand is melted out of the helix to anneal to the excess complementary strand present in the mixture and is removed by centrifugation to obtain a clean gapped pUC19N plasmid for further steps. Lesion containing oligonucleotide (Table. 2.9) was annealed into this 48 nt long ssDNA gap at 45°C for 2 h by providing it in 20-fold excess concentration. The mixture was then incubated overnight in the presence of T4 DNA ligase at ambient temperature to seal the nicks present at the gaps between the inserted sequence and the plasmid at the two ends (5' and 3'). The ligated product was treated with the enzymes *NdeI* and *BsaXI* to perform restriction digestion resulting in the 505 bp long DNA substrate containing the lesion of interest at 50% total length. Restriction digestion generates 2,224 bp long non-specific substrate along with the short lesion containing substrate, which is used in the AFM studies of glycosylase-DNA complexes. The digested products were run on a 1% agarose gel with an appropriate DNA ladder (See section 2.2.1a, for agarose gel electrophoresis procedure) and both 505 bp and 2,224 bp long substrates were obtained by gel extraction. Concentrations were estimated using a Nanodrop spectrophotometer (See section 2.2.1a, for UV/Vis spectrophotometry). The DNA substrate preparation is shown schematically in the Fig. 2.6.

Control assays for DNA substrate preparation. To confirm successful completion of gapping, insertion and ligation steps, appropriate control assays were performed at each stage (Fig. 2.7). To confirm successful gapping, a control assay based on a restriction enzyme, *XhoI* was performed. The restriction site for *XhoI* lies in the dsDNA region between the two nicks. Hence if the plasmid is completely gapped (ssDNA gap), incubation with *XhoI* does not incise the DNA and hence shorter cleaved fragments are not observed on the gel (Fig. 2.11a) After the ligation step, an additional assay was performed based on the restriction enzymes *NsiI* and *PstI*. *NsiI* has a restriction site at the 5' end of the nick and *PstI* at the 3' end of the nick. Successful digestion by these enzymes suggests complete ligation of the inserted lesion containing substrate into the plasmid. All

the control assays were performed and run on a 1% agarose gel as described in the section 2.2.1a.

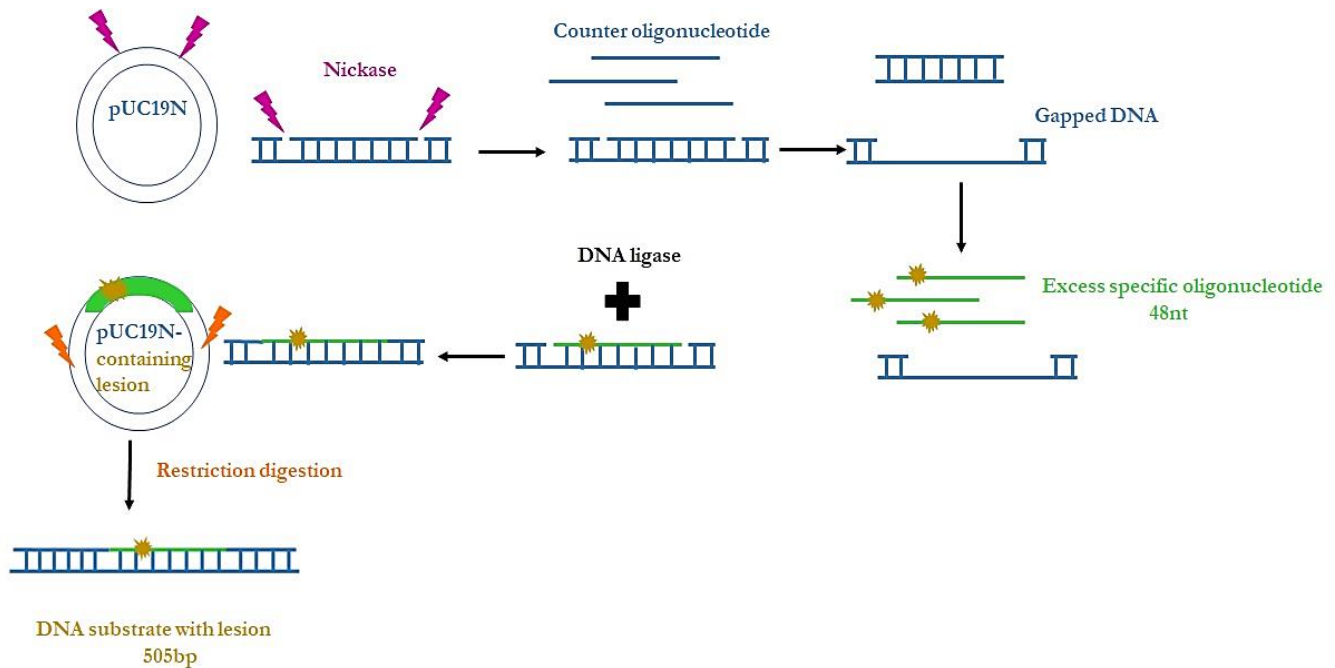


Figure 2.6: Schematic representation of the DNA substrate preparation for AFM studies. Circular dsDNA pUC19N plasmid is treated with a nickase to create two proximally located nicks (indicated by magenta lightning bolt symbols). During the gapping step, the nicked ssDNA is melted out of the double helix in presence of an excess complementary oligonucleotide. The melted-out strand forms base pairs with the counter oligonucleotide and is removed by centrifugation. In the next step, the gapped DNA is incubated with an excess of counter oligonucleotide (green) containing the lesion of choice (yellow) and the 5' and 3' nicks are sealed in the presence of T4 DNA ligase. The ligated product is subjected to restriction digestion with enzymes whose restriction sites (orange lightning bolts) are present far away (to obtain a substrate of ~500 bp length) on either side of the sealed nicks. The final DNA substrate is 505 bp long containing a centrally located lesion (50% position). Control assays are performed at regular intervals during the substrate preparation to ensure completion of each step (Fig. 2.7).

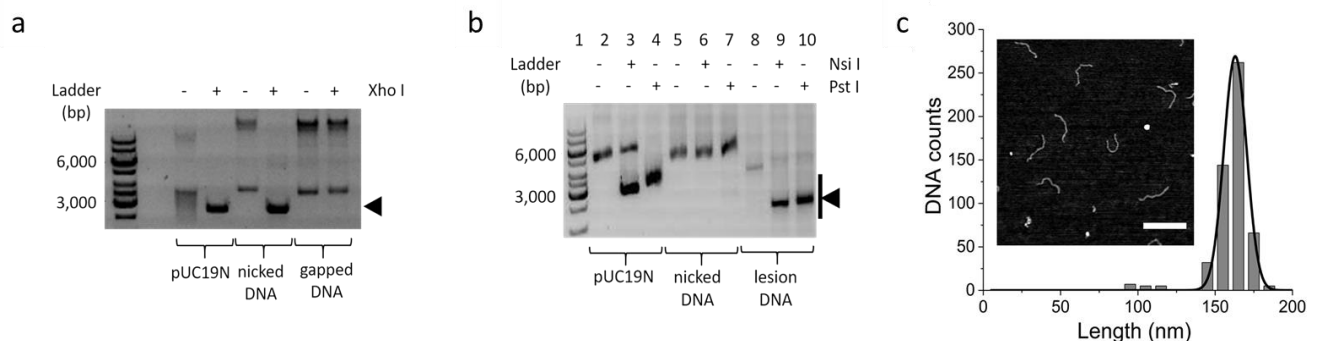


Figure 2.7: Characterization of the prepared DNA substrates for AFM imaging. (a) Gapping assay that confirms the 48 nt stretch has been removed after nicking pUC19N at the two Nt.BstNBI sites. Black arrow indicates the Xho I digestion product. Original pUC19N plasmid undergoes digestion and acts as a positive

control in the control assays. Further, the presence of 5' and 3' nicks do not affect *XhoI* digestion as confirmed by the additional control. The lack of incision by *XhoI* (black arrow) in the pUC19N after gapping (last lane) confirms complete gapping. **(b)** Ligation assay to confirm 5' and 3' ligation of 48 nt insert containing the lesion. Restriction sites of *NsiI* and *PstI* are located 5' and 3' ligation sites respectively. Incision of the ligated plasmid by both the enzymes confirms complete ligation (black arrow, lanes 8-10). Both controls, intact and the nicked pUC19N plasmid were included. *NsiI* and *PstI* digested products were observed for the intact plasmid (lanes 3 and 4) while they were not seen in the presence of nicks (lanes 6 and 7). **(c)** DNA length distribution (from MatLab measurements) of final 505 bp long product with lesion at 50% of the DNA length (inset: exemplary AFM image, scale bar 200 nm). The prepared DNA substrates were observed to be clean in the AFM images without any contamination. These substrates were further used for automated DNA bending analyses.

Enhancer-box (E-box) cloning strategy for DNA substrate preparation. In the published studies for Nuclear Factor kappa-B (NFκB) mediated transcription, it was reported that hOGG1 bound at its target lesion, oxoG ~8 nt upstream (of the NFκB target site) enhanced the recruitment of NFκB to its consensus motif¹⁰. I used the same strategy (oxoG and E-box spaced 8 bp apart) to investigate if hOGG1 influenced Myc recruitment to E-box consensus motif.

To perform AFM studies with DNA substrates containing an E-box or an oxoG lesion as well as an E-box (Table. 2.9), pUC19N plasmid was used. E-box sequence (CACGTG) was inserted into pUC19N plasmid using PCR amplification by Phusion polymerase. The PCR reaction set up and conditions are mentioned in Table. 2.14 and 2.15 respectively. The cloning strategy adopted to clone E-box (Fig. 2.8) was the same as SDM (Section 2.2.1a). The complete E-box sequence was present in both forward and reverse primers (Table. 2.8). Further steps from overnight *DpnI* digestion of the PCR products to chemical transformation remained the same as described in the section 2.2.1a. pUC19N_E-box plasmid was used further for AFM DNA substrate preparation and by procedures explained in the section 2.2.3.1. Insertion of the E-box was confirmed by sequencing (Appendix, Fig. x5.2).

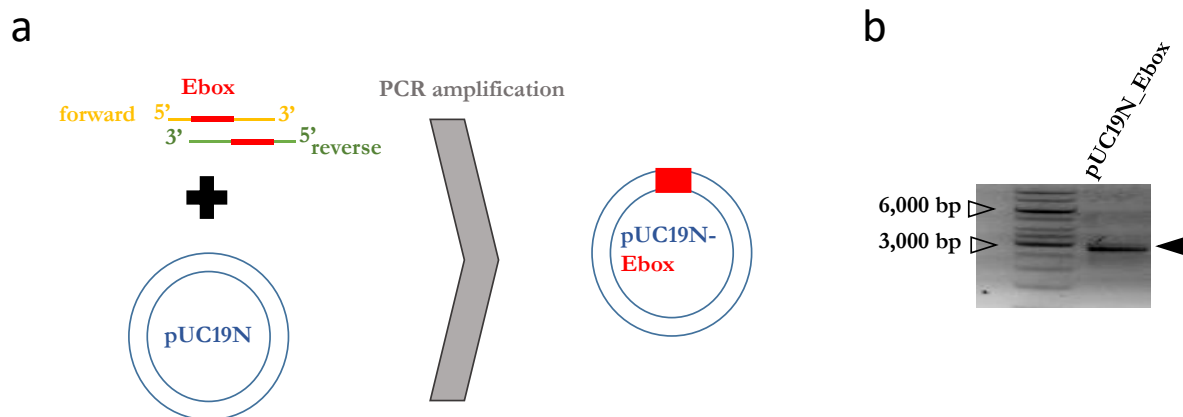


Figure 2.8: Cloning strategy for E-box motif in pUC19N plasmid. (a) E-box sequence (CACGTG) (in red) was included in both forward (yellow) and reverse (green) primers. PCR amplification of pUC19N plasmid (blue concentric circles) was performed in the presence of these two primers for successful

incorporation of the E-box in pUC19N plasmid (represented by the red rectangle in the PUC19N plasmid). **(b)** Agarose gel electrophoresis showing E-box containing pUC19N PCR product. Black filled arrow shows the PCR product for pUC19N_E-box. Black empty arrows denote the sizes 6,000 and 3,000 bp in the DNA ladder.

For preparing substrates containing both oxoG lesion and E-box motif, pUC19N_E-box plasmid was used as the starting material. The subsequent steps were same as described in the section 2.2.1a. For better visualization, the whole procedure of generating the substrate with both oxoG lesion and E-box motif has been depicted in the Fig. 2.9. The inserted oligonucleotide contains an oxoG lesion (hOGG1 target site) and an E-box motif (Myc target site) spaced 8 nts apart based on a previous study¹⁰.

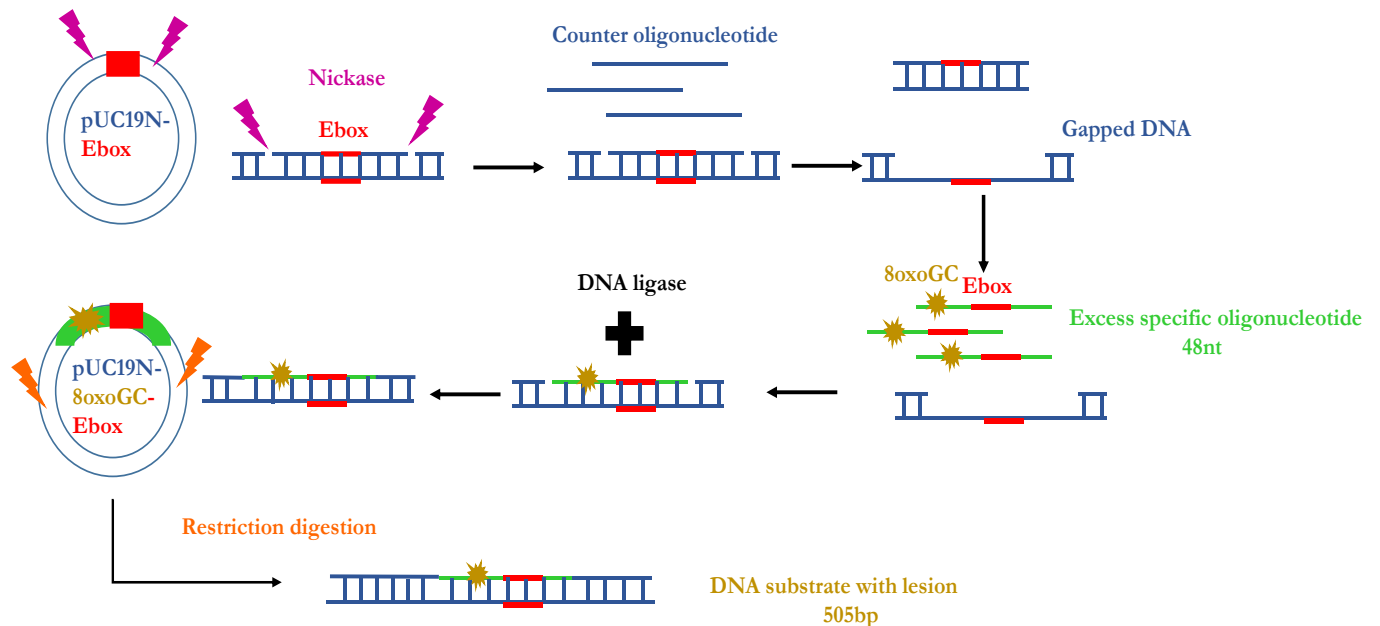


Figure 2.9: Schematic representation of oxoG-Ebox DNA substrate preparation. Circular dsDNA pUC19N plasmid containing E-box (shown in red) is treated with a nickase to create two proximally located nicks (indicated by magenta lightning bolt symbols). During the gapping step, the nicked ssDNA is melted out of the double helix in presence of an excess complementary oligonucleotide. The melted-out strand forms base pairs with the counter oligonucleotide and is removed by centrifugation. The counter oligonucleotide also contains the palindromic E-box sequence for perfect complementarity during gapping. In the next step, the gapped DNA is incubated with an excess of counter oligonucleotide (green) containing the lesion of choice (here, oxoG) (yellow) placed upstream (here, 8 nts) of the E-box motif (in red) and the 5' and 3' nicks are sealed in the presence of T4 DNA ligase. The ligated product is subjected to restriction digestion with enzymes whose restriction sites (orange lightning bolts) are present far away (to obtain a substrate of ~500 bp length) on either side of the sealed nicks. The final DNA substrate is 505 bp long containing a centrally located lesion (oxoG) and E-box motif spaced 8 bp apart (50% DNA length). Control assays are performed at regular intervals during the substrate preparation to ensure completion of each step (Fig. 2.7). Presence of E-box does not affect the restriction sites of the enzymes used in the control assays. As an exemplary, oxoG lesion with E-box containing substrate preparation is shown here. Instead of nicking and gapping native pUC19N plasmid, pUC19N_E-box plasmid has been used here.

2.2.3.2 AFM sample preparation

Lesion substrates and glycosylase-DNA complexes. 20 µl was used as the standard volume for all incubations and depositions onto freshly stripped mica substrate. Both DNA dilutions and protein-DNA incubations were performed at ambient temperature for 15 min. Incubations were made in the AFM buffer containing 25 mM HEPES/ HCl, pH 7.5 at 25°C, 25 mM sodium acetate, 10 mM magnesium acetate. At pH 7.5, mica is negatively charged. The divalent magnesium ions in the incubation AFM buffer helps in the stable chelation of negatively charged DNA to the negatively charged mica surface. DNA deposition concentrations were 4 nM (for 505 bp substrates) and 0.5 nM (2,224 bp undamaged DNA substrates). Glycosylase protein concentrations were varied from 10 to 300 nM based on their binding affinities to DNA.

Dual CpG site containing substrates and DNMT3A-3L DNA methyltransferases. DNA substrates D6, D9, and D12 and the readily assembled DNMT3A-3L protein complexes were obtained from AG Jeltsch, IBTB Stuttgart (Appendix, Fig.x4.1 and Table. x4.1). Protein complexes were incubated with DNA substrates (4 nM) at a concentration of 100 nM in the AFM buffer (for composition, see above) at room temperature for 20 min. Incubation volume was kept constant at 20 µl for all depositions. Free DNA substrates (4 nM) were deposited on to freshly cleaved mica and imaged for control measurements. Free proteins were deposited at a concentration of 7 nM to obtain control measurements.

hOGG1, Myc, Max, LSD1 complexes, and DNA substrates. hOGG1 was oxidized in the presence of 5 µM H₂O₂ and dialysed for AFM experiments as described before in section 2.2.2 (under analytical size exclusion chromatography method). For reducing conditions, hOGG1 was reduced in the presence of 5 mM hOGG1. AFM sample volumes were 20 µl in all experiments. For AFM experiments with hOGG1 and Myc-Max complexes on oxoG DNA, all incubations were performed in the AFM buffer with oxidized and reduced hOGG1 (150 nM) incubated with 30 nM Myc and 3 nM oxoG substrate for 15 min at ambient temperature. Free Myc and Max control measurements were made by depositing 10 nM protein each.

For QD-conjugated Myc-Max complexes with hOGG1 on DNA, Myc was incubated with Anti-GST monoclonal antibody in the molar ratio of 1:1 under ambient temperature for 30 min (concentration, 1 µM). Subsequently, this mixture was incubated with Qdot™ 605 goat F(ab')₂ anti-mouse IgG conjugate (H+L) for 30 min to label Myc with the quantum dot (QD) (ratio 1:1, concentration 500 nM). QD labelled Myc was then incubated with Max in the molar ratio of 1:1 for 10 min to form a stable complex at a concentration of 20 nM. For the final deposition, 2 nM of QD-MycMax complex was incubated with 150 nM of oxidized/reduced hOGG1 in the presence of 3 nM oxoG-Ebox substrate with both present at 50% position. As controls, QD-Myc with DNA and QD-MycMax with DNA incubations were also prepared for depositions at same concentrations.

For hOGG1 and LSD1 on DNA, 1 μM hOGG1 was crosslinked with 1 μM LSD1 and 20 nM DNA using 0.1% glutaraldehyde in the AFM buffer for 15 min at ambient temperature. This mixture was diluted to 10 nM protein for depositions.

2.2.3.3 Sample deposition, Image collection and processing

The samples are deposited on to clean, flat mica surface using a pipette without touching the surface of the mica. In the next step, they are allowed to spread evenly across the mica surface for 1 min, rinsed with deionized ultrapure water, dried by a gentle nitrogen stream, and fixed on a microscope slide for AFM imaging.

2.2.3.4 AFM imaging and processing

Principle. A 3D topography image is generated based on the movement of a sharp tip attached to the end of a cantilever scanning the surface of a sample (Fig. 2.10). To obtain very high resolution of biomolecules imaged, a very sharp tip with diameter typically less than ~ 10 nm is used¹¹. Tips are generally made of silicon. Imaging is performed based on the non-covalent forces like long-range electrostatic interactions, very short-range Pauli repulsion (when the tip is deflected away from the sample), and short-range van-der-Waals forces (attractive force) between the tip and the sample. A focused laser beam is impinged on the cantilever to translate the deflection to a detectable signal. The deflected laser beam coming from the cantilever is reflected onto a mirror and the reflected signal then serves as an input to a position sensitive split photodiode detector¹². The deflection of the cantilever leads to a change in the position of the laser beam on the split photodiode. The optical signal is then converted to a differential electronic signal. The interaction between the tip and the sample is modulated by a feedback controller, which maintains a constant distance between them. The feedback system also provides the height information of the specimen (sample) for each pixel and forms the basis of the final 3D topography map of the sample. The movement of the tip in the z-axis is regulated by a piezo scanner connected to the AFM scanning head holding the cantilever. By measuring the z-displacement of the tip at every xy-position, a 3D topography map of the specimen is generated by the imaging software.

AFM imaging can be performed by three common modes: contact, tapping (intermittent contact) and non-contact mode.

Tapping mode is the most preferred mode of imaging of biological specimens. As the name suggests, the tip intermittently touches the surface of the sample by vibrating the cantilever. This mode of imaging is minimally invasive hence mitigating damage to the biological sample. The cantilever is oscillated with a pre-defined frequency (near its resonance frequency) and amplitude with a piezo actuator that is connected to a sine wave generator. Upon contact with the surface of the sample, the oscillation amplitude of the cantilever is interrupted. The feedback system then restores the amplitude by upregulating the tip z-direction and simultaneously reads out sample height information, which is translated to a final 3D topography map by the computer¹³.

Imaging conditions and image processing. Imaging was performed using a Molecular Force Probe 3D (MFP-3D) AFM. Images were collected with a scan speed of $2.5 \mu\text{m/s}$ in the tapping mode oscillation using AC240 AFM cantilevers with resonance frequency of $\sim 70\text{kHz}$ and spring constant of $\sim 2 \text{ N/m}$. AFM images were collected with a resolution of 1.95 nm/pixel and were plane fitted and flattened to 3rd order using the MFP Igor Pro software. Finally, they were exported in .tiff format with the image resolution that was used while collecting them.

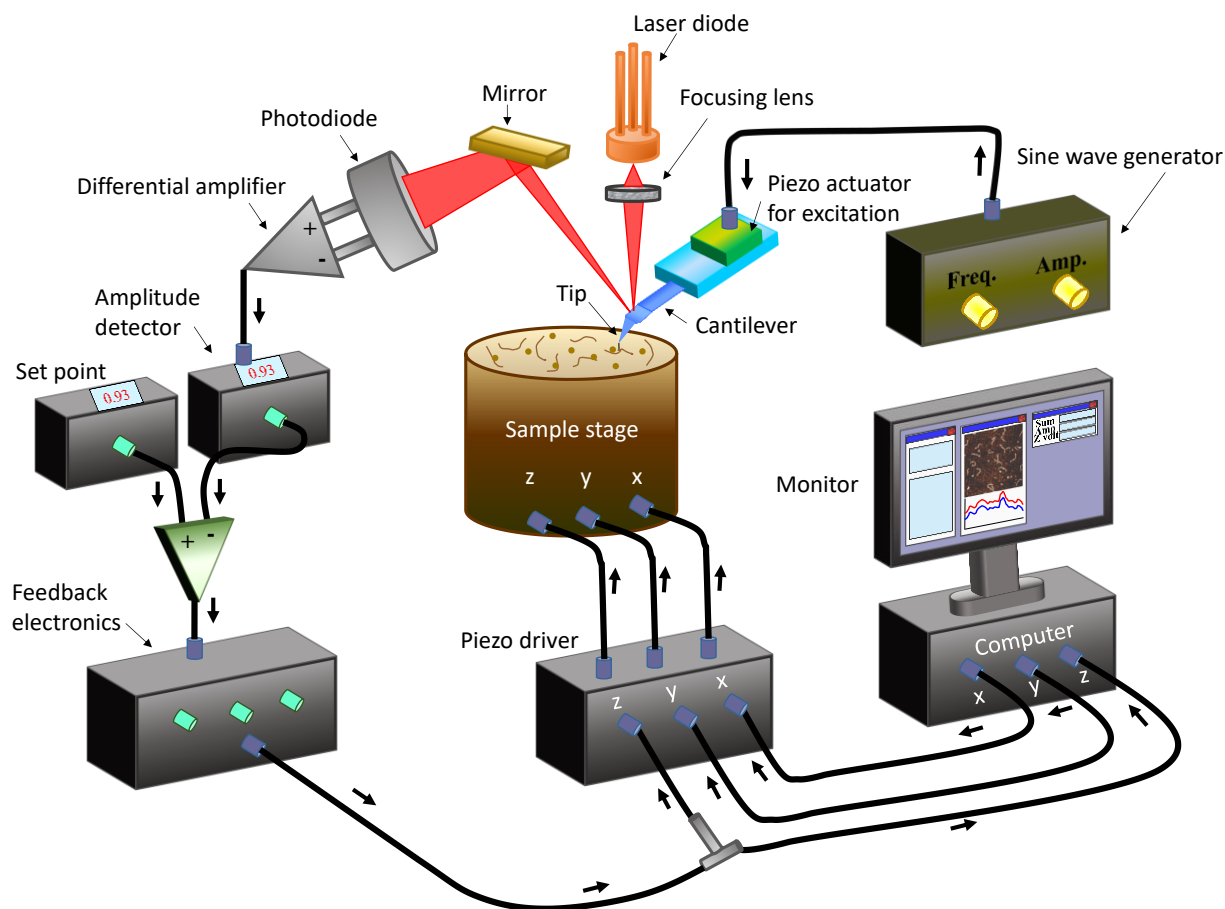


Figure 2.10: Schematic of an AFM set up. The sample to be imaged is placed on the sample stage. A sharp tip attached to a cantilever scans the sample by tapping mode. Cantilever deflection is detected by the shift in the position of a laser beam reflected from the cantilever onto a position sensitive photodiode, and translated into sample height information. A feedback loop regulates the distance between the tip and the sample surface and simultaneously records the height information of the sample at every x/y-position. Using the height information, the imaging software generates a 3D topography image of the sample. Details are described in the text section above (adapted with permission from Ando et al, 2013, and Bangalore and Tessmer, 2018).

2.2.3.5 AFM data analysis

2.2.3.5a Automated AFM DNA bending analysis at target lesion sites

As described under DNA substrate preparation section, 505 bp long DNA substrates contain target lesions at 50% DNA lengths. Bend angles at these sites are measured automatically by MatLab.

Pre-processing steps for automated DNA bend angle analysis at DNA target lesion sites.

All software settings are described in Table. x3.1 and schematic of the procedure is shown in Fig. 3.6. The exported .tiff AFM images are cropped to original sizes using GIMP software and are exported using LZW option to retain the original resolution. The cropped images are imported to ImageJ software, converted to 8-bit format and a median filter is applied over two pixels. Yen threshold is applied to select continuous DNA strands without any interruptions while still excluding the non-specific background contamination signals. The threshold value is set such that the entire DNA filaments are selected for further analyses.

For bend angle measurements by MatLab tool, DNA strands in the pre-processed images must be skeletonized using FIESTA software. The binary image from the ImageJ software is loaded into FIESTA software and the yen threshold value noted in the ImageJ software is entered along with full width half maximum (FWHM) of various DNA strands measured by drawing section lines perpendicular to the DNA backbone. Further, the average FWHM value is calculated and entered with other configuration settings before running the program. FIESTA skeletonizes the DNA using 2 nm rigid lines through the center of the DNA backbone. In the next step, segmented track data output file is opened to discard interrupted filament tracks, DNA filaments with wrongly traced DNA backbone, and ambiguous bundles of DNA filaments by selecting them manually using the cursor and deleting. The selected tracks are then saved in the .mat file format ready to be loaded into MatLab tool.

MatLab routine for DNA bending analysis at lesion target sites located at 50% DNA length. After starting MatLab software, the working folder is opened for saving all results. The same folder contains the DNA bend angle analysis script. The DNA skeletonization information (.mat file) from FIESTA is uploaded to MatLab. MatLab performs the first step of discarding DNA filaments with wrong lengths. This action is essential to correctly localize the lesion position by MatLab else the bend angles from the non-specific sites can contaminate the results. This procedure discards the broken filaments. MatLab prompts the user to enter the theoretical length of the substrate (here, for 505 bp long substrate, it is 172 nm assuming rise of 0.34 nm/bp in the B-form DNA). B-form DNA possess minor vibrations in the backbone that shorten the contour lengths in the AFM images. These undulations are restricted by the resolution limit of AFM. Hence, to account for these effects on the DNA, DNA lengths are allowed to vary between two standard

deviations from the Gaussian centre of the fitted DNA length distribution (150 to 180 nm). Using automated spline interpolation, 2 nm rigid backbone fragments are converted to 0.1 nm for the DNA filaments with correct lengths. MatLab determines the bend angles at the 50% target lesion by placing two intersecting tangents from a query point distance (here, 8 nm) to the site of lesion. DNA bend angle is defined by the deviation from the straight DNA backbone conformation (180° - the angle measured by MatLab). DNA bend angle distributions were plotted and fit with single or multiple gaussian curves in Origin Pro software. Individual fits to these multimodal distributions are shown in Fig. x3.10. DNA bend angle values are obtained from the centers of these single or multiple Gaussian curves (Table. 3.1).

Automated bend angle analyses using MatLab tool requires certain AFM quality criteria to be met. Images with uneven background generate false results. Hence, such bad quality images were avoided for processing. I estimated the heights of the MatLab selected and non-selected DNA and found that the minimum height requirement is ≥ 0.5 nm (Fig. x3.5a). However, only images with an even background were used and confirmed that only particular bend angle states were not favoured by MatLab tool by manually comparing bend angles of selected and non-selected DNA filaments (Fig. x3.5c). Images with less than 70% selection were excluded from analyses for obtaining reliable results.

In addition to image quality, bend angles obtained can vary depending on the tangent or secant method of measurement. Exemplarily for undamaged DNA, both methods were tested. The query point distances were 8 nm in both cases. The bend angle measurements were $\sim 8^\circ$ and 0° from tangent and secant geometry methods respectively (Fig. x3.6a). Another test was conducted where the bend angles from non-specifically bound glycosylase complexes were compared with the crystal structures of the same glycosylase bound to undamaged and lesion DNA. The comparison showed that similar bend angles as seen in the crystal structures (IC complexes) of all four glycosylases tested here with tangent line geometry using 8 nm query distance (Fig. x3.6b). Additionally, the manually measured bend angles for oxoG:A shown exemplarily (Fig. x3.6c) matched well with the angles measured by tangent method rather than the secant method, both at 8 nm query point distance. To assess the relationship between the tangent and the secant line geometry, oxoG:A bend angles were measured by both methods using different query point distances. Results obtained from secant geometry at 8 nm query point distance were identical to those from tangent geometry at 4 nm query point distance suggesting secant method is also valid but at a different query point distance (Fig. x3.6d). However, tangent line geometry has been adopted in the MatLab tool.

For better comparison of different bend angle states between lesion sites and at non-specifically bound glycosylase-DNA complexes, the query point distance needs to be fixed. Query point distance is measured from the site of lesion (50% DNA length) or bound protein to the two points on either side of the target site along with tangents are traced. The intersection of these two tangents gives rise to an angle measured by the MatLab

tool. 180°- this angle gives the DNA bend angle. In these studies, a query point distance of 8 nm was used. With increasing protein sizes, the query point distance needs to be increased during the analysis using the tool. Query point distances were varied from 4 nm to 15 nm to understand their effects on the DNA bend angles measured. 4 nm query point distance gives rise to very noisy results since only 2 pixels (pixel resolution in these studies is 1.95 nm/pixel) are considered for measurements. On the other hand, 15 nm query point distance provided broader bend angle distributions than usual indicating contributions from non-specific DNA at larger distances along the tangents. Since all the glycosylases have a size of less than 7 nm, the bend angle measurements were optimized by using 8 nm query point distances to avoid any artefacts (Fig. x3.7).

Persistence length measurements. AFM imaging is performed in air using the air-dried samples. To ensure that the deposited molecules behave similarly to the hydrated samples and are well equilibrated, DNA end-to-end lengths (R) were measured by the MatLab tool. Worm like chain (WLC) model is gives the DNA persistence length L_p , using the equation, $L_p: \langle R^2 \rangle_{2D} = 4 L_p L_c \{1 - (2L_p/L_c) (1 - e^{-L_c/2L_p})\}$. L_c is the DNA contour length. In the experiments for 505 bp lesion substrates, $\langle R^2 \rangle_{2D} = (17,214 \pm 422) \text{ nm}^2$ was obtained for the given buffer conditions. The corresponding value of R for 505 bp long undamaged DNA was ~ 131 nm. This translates to a persistence length of ~ 45 nm. The persistence lengths of all 505 bp long DNA substrates fall in the range of 40-45 nm (Appendix, Fig. x3.13 and table. x3.3).

2.2.3.5b Automated DNA bending analysis at protein positions

MatLab based automated tool in this study was used to measure non-specifically bound glycosylase-DNA complexes and DNMT3A-3L complexes at specific dual CpG sites (50% position) of ~ 400 bp long substrates.

Pre-processing steps for automated DNA bend angle analysis of protein-DNA complexes. The schematic workflow is described in Fig. 3.8 and all software settings for MatLab and FIESTA¹⁴ are provided in the Table. x3.2 in the appendix section. The .tiff images from MFP3D software are imported into GIMP software to crop the images to their exact size (with only the image pixels) by excluding outer marked regions of the image. The cropped image is exported using the LZW option to retain the resolution of the image. The cropped image is then imported into ImageJ¹⁵ and converted to an 8-bit image first and a median filter is applied over two pixels. Further, Yen threshold is applied to create a binary image. The threshold value is selected in a way to include only the DNA and DNA-bound protein molecules and excludes the free proteins in the images. In the first step to select only the DNA filaments, shape filter plugin is applied with an elongation parameter of 0.75-1 and a perimeter between 90- ∞ ($60-\infty$ for extremely curved filaments) pixels to exclude protein molecules and select only DNA. Gaussian blur is then applied over two pixels as a final step. The steps described here constitute image pre-processing.

Parallely, cropped images are also processed in ImageJ for selection and localization of DNA bound protein molecules. For this purpose, intermodes threshold filter is applied to select only the protein peaks on the DNA by manually adjusting the threshold values. Using 3D object counter, the co-ordinates of the obtained protein peaks are saved in the .txt format which is required for further MatLab routine.

For bend angle measurements by MatLab tool, DNA strands in the pre-processed images must be skeletonized using FIESTA software. The Gaussian blurred binary image from the ImageJ software is loaded into FIESTA software and the yen threshold value noted in the ImageJ software is entered. For successful skeletonization, full width half maximum (FWHM) of various DNA strands is measured by drawing section lines perpendicular to the DNA backbone. Further, the average FWHM value is calculated and entered with other configuration settings before running the program. FIESTA skeletonizes the DNA using 2 nm rigid lines through the center of the DNA backbone. In the next step, segmented track data output file is opened to discard wrongly connected filament tracks and ambiguous bundles of DNA filaments by selecting them manually using the cursor. The selected tracks are then saved in the .mat file format ready to be loaded in the MatLab tool.

Highly bent DNA strands (DNA kinks) may not be segmented during skeletonization by FIESTA. This leads to loss of important data from the images. However, these broken tracks can be connected by the MatLab tool by entering the total number of DNA filament pairs to be connected and the filament identifiers obtained from FIESTA of these respective pairs. The schematic of the procedure is shown in appendix, Fig. x3.4.

MatLab routine for DNA bending analysis at non-specifically bound protein positions. After starting MatLab software, the working folder is opened for saving all results. The same folder contains the DNA bend angle analysis script. The .mat file with the DNA skeletonization data and the .txt file with the protein co-ordinates are opened in MatLab. To better describe the continuous DNA structure, DNA skeleton segment lengths are reduced to 0.1 nm by automatic spline interpolation step. First, MatLab prompts the user to enter the area filter to the protein positions to avoid protein aggregates and wrong localization data. The cut-off levels range between 0.0000001-0.0001 μm^2 with an option to go up to 0.001 μm^2 for larger protein systems. The free protein radii can be measured for each protein (proteins vary in size) and DNA widths can be estimated (using freehand or rigid line tool) in ImageJ software This information is necessary to correctly localize the DNA bound proteins. Bound proteins are defined by the sum of protein radii and the DNA width. If the value is higher than the cut-off value, these protein peaks are excluded from further analyses. As a further caution, only proteins with a spacing of ≥ 50 nm from each other on the DNA are considered for analyses to avoid influence of closely spaced protein molecules on the DNA curvature. MatLab uses tangent geometry to measure the bend angle by intersecting two tangents from a user defined query distances from the DNA backbone at protein positions. The query point distances need to be sufficiently longer than the sizes of the protein molecules to avoid false placement of the tangents due to

unknown DNA topography beneath the bound protein molecules. All the glycosylases employed in this study have radii ranging between ~3-7 nm. Hence for the analyses presented here, a query point distance of 8 nm was used (higher than the size of the glycosylases). DNA bend angle distributions were plotted, and Gaussian fitting was performed in Origin Pro (Originlab Corporation, Northampton, USA) using the bend angle values from MatLab for individual glycosylase complexes. Centers of the Gaussian fits to the revealed the DNA bend angles. Bimodal and multimodal distributions were obtained for different glycosylase-DNA complexes (Fig. 3.10 and Fig. x3.9). Automated bend angle analyses for the non-specifically bound hTDG and hOGG1 were performed with the existing AFM images¹⁶. N-terminally truncated hAAG (ΔN_{72} hAAG) and MutY were bought from New England Biolabs and Trevigen respectively. The deletion of the N-terminus in hAAG did not affect DNA bending in these studies (Appendix, Fig. x3.12).

MatLab routine for DNA bending analysis of proteins (DNMT3A-3L complexes) bound at 50% position on the DNA substrates. The MatLab tool that was developed for analyses in the previous section¹⁷ was further extended to enable measurement of protein binding positions and corresponding bend angles at these sites¹⁸. The expanded MatLab script is deposited at Open Science Foundation (OSF) (<https://osf.io/76e9s/>) and is available for public use. The image pre-processing steps are as described above for non-specifically bound proteins (See appendix, Fig. x3.3).

During the MatLab run, 10 nm was entered as the cut-off radii for DNMT3A-3L complexes. MatLab selects only the DNA skeleton lines with correct length for further analysis (to localize CpG sites present at 50% positions). Theoretical DNA length for the substrates with dual CpG sites is ~139 nm (~400 bp, assuming 0.34 nm/bp rise). Theoretical DNA lengths for different substrates were calculated and entered with a flexibility of two standard deviations to account for DNA undulations. Since the two DNA ends are indistinguishable, MatLab measures the protein positions from the closer end and hence position distributions are plotted to 50% DNA length only. MatLab provides the DNA bend angles corresponding to the bound protein positions together in an output table. In addition, the output image is numbered for the protein peaks which is helpful in volume, height, and length measurements of these complexes.

Manual DNA bend angle measurements. DNA bend angles at protein bound positions or at 50% target sites were measured using the angle tool in Image J software (Fig. 3.6a). Two straight lines are drawn through the center of DNA backbone with an angle at the required position. This angle is ϕ , the DNA bend angle is derived as $\theta = 180^\circ - \phi$.

2.2.3.5c Binding specificity analysis

Manual protein position measurements. Protein positions were measured manually using Image J tool with freehand lines laid through the center of the DNA backbone to the center of the bound protein from either of the two ends of the DNA substrate. This length

obtained is divided by the length of the entire DNA substrate and converted to a percentage value to obtain percent positions.

For the quantum dot-AFM (QD-AFM) data analyses, all positions measured manually were normalized to the total number of DNA filaments present in all images for each incubation. This allowed to interpret the number of binding events of QDMyc-Max complexes at 50% position in presence or absence of hOGG1. All experiments were performed in triplicate (Appendix, x5.15). All the distributions were plotted and fitted using Gauss function in Origin Pro software. Binding specificities (S) were obtained from the ratio of total occupancy for the specific site (Eg: oxoG) defined by A_{sp} , area under the Gaussian fit and that of the non-specific DNA sites (A_{nsp} , area of background from 2 to 50% total DNA length)¹⁹ (see Fig. 2.11) using Eq. 1-

$$S = N \cdot \left(\frac{A_{sp}}{A_{nsp}} \right) + 1 \quad (1)$$

where, N is the number of total available binding sites (number of base pairs, $N = 505 \text{ bp} - 2 \times 2\% \text{ at ends} = 485 \text{ bp}$). Average specificities (\pm SD) were calculated from triplicate experiments and p-values were derived using two-tailed student's t test.

Automated protein position measurements. The positions of DNMT3A-3L complexes on the dual CpG site containing substrates were measured using the extended MatLab tool. The output table contains protein positions and corresponding DNA bend angles. Protein position distributions were plotted and fitted using Gauss function in Origin Pro software. Binding specificity calculations were performed by subdividing the DNA in 20 segments (each segment accounts for 5% DNA length, 2.5% on either side from the 50% position). The binding probability in each of these DNA segments (N_{obs}) was compared with equal probability of binding to each of these DNA segments by using $N_{exp} = N_{total}/20$. The binding specificity (S) was calculated in terms of observed binding events over expected binding events in each DNA fragment for all substrates using the formula N_{obs}/N_{exp} .

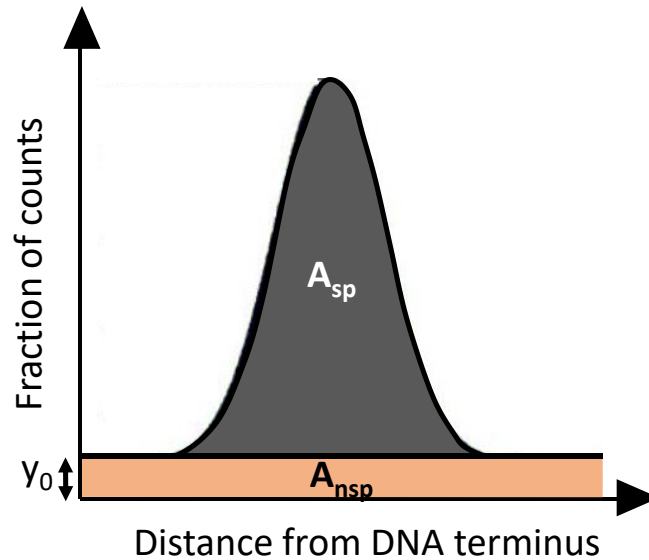


Figure 2.11: Schematic of a position distribution of a protein binding to a specific target site. Using the areas under the Gaussian fit to the distribution, A_{sp} (in dark grey) and A_{nsp} (in pink) and the height of the background (y_0), protein binding specificities are calculated by equation 1.

2.2.3.5d Stoichiometric analysis

Density slice option in the ImageSXM (S. Barrett, University of Liverpool) software was used to analyze volumes of DNMT3A-3L complexes on the DNA. Area and height information was obtained from the software to calculate volumes in MS excel. Volume is obtained by subtracting the background noise height from the height of the protein peaks and multiplying this value with the area of the complexes. Histograms were plotted with these volumes and were fit using Gaussian curves in Origin Pro software. Volumes (V) of free proteins can be translated to molecular weights (MW) using the equation- $MW = (V + 5.9)/1.2$ and the volume calibration curve of the AFM setup used in this study²⁰. Lengths and heights of DNMT3A-3L complexes bound to DNA were measured manually at 50% DNA lengths using ImageJ and ImageSXM software programs respectively. For the volumes of protein-DNA complexes, the values may not linearly correspond to the original molecular weights of the proteins because of volume contributions from dsDNA and this might pose difficulty to apply the linear relationship between the AFM volumes and molecular weights²¹. Hence, only volumes are reported in all the studies in this thesis.

2.2.4 Biophysical and Biochemical assays for protein-protein and protein-DNA interaction analyses

Ensemble FRET experiments. Ensemble FRET measurements were performed to verify the average bend angle values obtained for different glycosylase-DNA complexes. The 48 nt long non-specific DNA used as a counter oligonucleotide during the gapping step was annealed to a 20 nt long bottom strand containing Cy3 FRET donor at 5' end and Cy5 FRET acceptor at 3' end (Table. 2.9) in 15:1 ratio to confirm all fluorophore containing oligonucleotides were in the double stranded DNA (dsDNA) format. As controls, top strand annealed with bottom strand containing only Cy3 (5') and only Cy5 fluorophores (3') were used to correct for background signals. For the undamaged DNA only control measurements, 50 nM FRET DNA substrate was excited at Cy3 (FRET donor) excitation wavelength, 509 nm and at Cy5 (FRET acceptor) excitation wavelength, 649 nm and emission wavelength were recorded in AFM buffer in a 50 μ l quartz cuvette (Hellma Analytics). For the FRET measurements with glycosylases, 1 μ M of each protein concentration was applied subsequently to the DNA and the samples were incubated at ambient temperature for 15 min. For each of these glycosylases, FRET acceptor emission was detected prior to and after protein addition using a spectrofluorometer (Fluoromax 4 series, Jobin Yvon, Horiba Scientific) (Fig. x3.8). Cy3 emission maxima of the Cy3 only labelled DNA samples in the presence or absence of glycosylases were normalized to the Cy3 emission maxima of FRET DNA substrate samples with or without glycosylases. By correcting for signals coming from Cy3, buffer, DNA, and individual proteins in the samples, Cy5 acceptor emission spectra were obtained. The emission spectra recorded at direct excitation of Cy5 acceptor is used for further analyses. The Cy5 acceptor emission spectra at Cy3 excitation for FRET substrate samples with or without proteins were normalized to Cy5 emission spectra at Cy5 excitation for the same corresponding samples to correct for varying fluorophore concentrations. All experiments were performed in triplicate. The intensity spectra recorded were used to calculate FRET efficiencies, E_{FRET} and average DNA bend angles subsequently.

FRET efficiency is given by the equation (2),

$$E = \frac{I_{AD} \cdot \epsilon_{AA} - I_{AA} \cdot \epsilon_{AD}}{I_{AA} \cdot \epsilon_{DD}} \quad (2)$$

where $I_{AD@509 \text{ nm}}$ = maximum intensity of acceptor (Cy5) at donor (Cy3) excitation, $I_{AA@649 \text{ nm}}$ = maximum intensity of acceptor at acceptor excitation, $\epsilon_{AA@649 \text{ nm}}$ = extinction coefficient of acceptor at acceptor excitation wavelength ($250,000 \text{ M}^{-1} \text{ cm}^{-1}$), $\epsilon_{AD@509 \text{ nm}}$ = extinction coefficient of acceptor at donor excitation wavelength ($3,079 \text{ M}^{-1} \text{ cm}^{-1}$), and $\epsilon_{DD@509 \text{ nm}}$ = extinction coefficient of donor at donor excitation wavelength ($71,769 \text{ M}^{-1} \text{ cm}^{-1}$).

To relation between FRET efficiency and FRET donor-acceptor distances is given by the equation (3).

$$r_{D-A} = R_0 \left(\frac{1}{E} - 1 \right)^{\frac{1}{6}} \quad (3)$$

where r_{D-A} is the distance between FRET donor-acceptor pair (here, Cy3 and Cy5) and R_0 = Förster radius (5.6 nm for Cy3/Cy5 for freely rotating dyes and thus orientation factor $\kappa^2 = 2/3$)²². The extent of DNA bending dictates the distance r_{D-A} between the donor-acceptor pair.

The distance r_{D-A} is then used to calculate DNA bend angles using cosine law defined by the equation (4)-

$$\cos \Phi = \frac{r_{D-A}^2 - b^2 - c^2}{-2bc} \quad (4)$$

After determining Φ , DNA bend angle (Θ) is calculated using the equation $\Theta = 180 - \Phi$. For the Cy3-Cy5 donor-acceptor pair, the total maximum distance was estimated to be 7.17 nm ($b=c= 7.17$ nm assuming all glycosylases bind preferentially in the middle of the 20 bp long labelled bottom strand).

hOGG1 gel-based activity assay. For hOGG1 activity assay, 48 nt long 5' Cy3 labelled oxoG containing top strand is annealed with non-specific bottom strand in 1:1 molar ratio to obtain a 5 μ M stock solution. For annealing, both top and bottom oligonucleotides were incubated at 95°C for 10 min on a thermomixer and subjected to gradual cooling to 15°C to ensure optimal dsDNA formation. In-house purified hOGG1 and hOGG1_{K249Q} were used for all activity assay experiments. For all assays with oxidizing conditions, hOGG1 was subjected to oxidation in presence of 5 μ M H₂O₂ in the hOGG1 gel filtration buffer for 3 h at 4°C and dialyzed overnight against the same buffer without H₂O₂ and concentrated, flash frozen and stored at -80°C. For all incubations with reducing conditions, buffer containing 5 mM DTT was used. The reactions in the assay contained hOGG1 under oxidizing and reducing conditions, buffer, and DNA substrate. Activity assay buffer composition was 25 mM Tris-HCl, pH 7.6, 100 mM NaCl. The final concentration of hOGG1 and DNA substrate were 1 μ M and 20 nM respectively. The final volume of each reaction was 10 μ l. The assay was performed at 37°C for time points 5, 15, 30 and 45 minutes in a dark place to avoid photobleaching of the fluorophore and the abasic site generated by hOGG1 was cleaved by incubating with 0.2 N NaOH for 5 min and the reaction was stopped by heating to 95°C for 5 min and adding Orange G dye containing

urea (composition for a volume of 100 ml: 2 ml 50X TAE, 0.15g Orange G dye, 34 ml glycerol, and 64 ml 11M urea) ²³. All the reactions were loaded into the wells of a self-made 15% denaturing Urea-PAGE gel (7 M Urea) in 1X TBE (Acrylamide: Bisacrylamide in the ratio 37.5:1). The gel composition has been shown below in the Table. 2.20. The electrophoresis was performed at 4°C for 120 min at 100 volts. Along with assay reactions, a DNA ladder with 13-bp, 20-bp and 48-bp substrates was loaded to better visualize the cleaved product. The gel was visualized in the Pharos FX imager using Cy3 detection in the low intensity mode to get maximum signal. The cleaved products and the uncut bands were analyzed for intensities in ImageJ. For this purpose, Images were imported to ImageJ and converted to 8-bit. Subsequently, a rectangle box was localized to the band and using “Gels” option in the Analyze section of the software, the intensities were plotted. Using the line tool, the area under the curve for the intensity peaks were marked. These areas were selected using the wand tool and the intensity values were obtained in the results window. The intensity values measured for uncut and cut bands were corrected for background gray values. The intensity of the cut band was divided by the sum of the intensities of both cut and uncut bands to obtain percent product formation. These values were plotted against incubation times (10, 20 and 30 min) for the triplicate experiments and an average (\pm SD) over these three time points were taken and were plotted using histograms in Origin Pro software (Appendix, Fig. x5.8). p-values were calculated using two-tailed student’s t-test.

Table 2.20: Denaturing 15% Urea-PAGE gel composition

Component	Quantity (for 50 ml)
40% Acrylamide (29:1)	18.75 ml
Urea	24 g
10% APS	200 μ l
TEMED	20 μ l
10X TBE	5
dH ₂ O	Rest of the volume

Electrophoretic Mobility Shift Assays. For EMSA assays, oxoG lesion containing top strand and undamaged top strand were annealed with Cy3 labelled bottom strand in the molar ratio of 1:1 respectively.

The buffer used in all the EMSA experiments contained 20 mM Tris-HCl, pH 7.6, 200 mM NaCl. The concentrations of hOGG1 and LSD1 were 1 μ M each and Myc and Max were

100 nM each respectively. For annealing, both top and bottom oligonucleotides were incubated at 95°C for 10 min on a thermomixer and subjected to gradual cooling to 15°C to ensure optimal dsDNA formation. For all assays with oxidizing conditions, hOGG1 was subjected to oxidation in presence of 5 μ M H₂O₂ in the hOGG1 gel filtration buffer for 3 h at 4°C and dialysed overnight against the same buffer without H₂O₂ and concentrated, flash frozen and stored at -80°C. For all incubations with reducing conditions, buffer containing 5 mM DTT was used. The final volume of each reaction was 10 μ l. All EMSA reactions were incubated at room temperature for 30 min in a dark place to avoid photobleaching of the fluorophore. After 30 min, Orange G dye (composition for a volume of 100 ml: 2 ml 50X TAE, 0.15g Orange G dye, and 60 ml glycerol) was added to each reaction and loaded into the wells of a self-made 15% Native PAGE gel (Acrylamide: Bisacrylamide in the ratio 29:1) in 1X TBE. The gel composition has been shown below in the Table. 2.21. The electrophoresis was performed at 4°C for 240- 260 min at 100 volts. The gel was visualized in the Pharos FX imager using Cy3 detection in the low intensity mode to get maximum signal. All experiments were performed in triplicate.

Table 2.21: 15% Native-PAGE gel composition

Component	Quantity (for 15 ml)
Acrylamide 30 (39.5:1)	6.25 ml
10% APS	80 μ l
TEMED	8 μ l
10X TBE	1.5 ml
dH ₂ O	Rest of the volume

Fluorescence polarization measurements. For fluorescence polarization studies, oxoG lesion containing top strand and undamaged top strand were annealed with Alexafluor 647 labelled bottom strand in the molar ratio of 1:1 respectively. For assessing the binding affinity values, 5 nM AF647 (5') labelled oxoG and undamaged 48 nt long substrates were incubated with hOGG1 concentration starting from 8 μ M hOGG1 and serially diluted to ~0 μ M in hOGG1 gel filtration buffer. For protein-protein interaction studies, hOGG1 was labelled with Ni-NTA dye with AlexaFluor 488 by incubating both in the molar ratio of 1:3 in the hOGG1 gel filtration buffer for 30 min in the dark at ambient temperature. For removing residual dye, a 10kDa MWCO filter was washed with buffer and centrifuged 10 min at 10,000xg at 4°C in the washed filter. The filter was set upside down on a new 1.5 ml tube and hOGG1 was eluted into this tube by centrifugation at 1,000xg for 3 min. This process was repeated for three times. To check if the labelling was in the 1:1 ratio,

nanodrop spectrophotometer was used. To calculate the ratio of hOGG1 and Alexafluor 488, Lambert Beer's law was used (Eq. 5)-

$$c = \frac{A}{\epsilon \times L} \quad (5)$$

Where c is the concentration, A is the absorption, ϵ is the molar attenuation coefficient and L is the optical path length. Proteins have their maximal absorption at the 280 nm wavelength. The ratio of concentration of hOGG1 to Alexafluor 488 gave a value of ~ 1 indicating all hOGG1 molecules were labelled. For assessing the binding affinity values between proteins, 5 nM labelled hOGG1 was incubated with Myc₁₋₁₆₃ and LSD1 separately with concentrations starting from 40 μ M (Myc) and 60 μ M (LSD1) and serially diluted both proteins to ~ 0 μ M in separate incubations with hOGG1 in the EMSA buffer. The basic principle of fluorescence polarization is that the degree of polarization of a fluorophore is inversely related to its molecular rotation²⁴. Fluorescence polarization is measured as the difference of emitted parallel and perpendicular intensities to the excitation plane normalized by the total fluorescence emission intensity. When two biomolecules interact, the rotation of the fluorophore (attached to one of the molecules under study) diminishes when compared to the free molecule. This gives rise to a change in parallel and perpendicular emitted intensities. The polarization values are recorded from clariostar (with gain= 1700; AF647- excitation: 650 nm, emission: 671 nm, AF488- excitation: 599 nm, emission: 520 nm, 120 cycles at 1s/ cycle) and plotted against the ligand (unlabelled molecule) concentration (logarithmic scale) (Appendix, Fig. x5.5 and x5.10). The individual polarization curves were fit by Hill equation using the non-linear fit (NL-Fit) option in Origin Pro software. R^2 values were ≥ 0.95 . The average binding affinities, K_D (\pm SD) were calculated based on the triplicate values (Table. x5.1 and x5.2) for each sample. p-values were calculated using a two-tailed student's t-test.

Native agarose gel electrophoresis. For protein-protein interaction analysis, hOGG1, Myc₁₋₁₆₃ and LSD1 were incubated in the molar ratio of 1:1:1 respectively in the buffer containing 20 mM Tris-HCl pH 7.6, 200 mM NaCl for 30 min on ice. 1% agarose solution was prepared in 0.5X TBE was casted, and electrophoresis was performed in Mini-Sub® Cell GT System for 5 h at 4°C. Gel was stained overnight using Coomassie G 250 staining solution (80 mg of Coomassie G 250 in 1 L of ddH₂O acidified with 3 mL of 37% HCl). Destaining was done in water for 12 h the next day and the gels were visualized using Vilber imaging system to identify the interactions based on gel shifts. Experiments were performed in duplicate (Fig. 5.12 and Appendix, Fig. x5.10).

In vitro pull-down assays. For visualizing direct protein-protein interactions, 20 μ l of Protino® Ni-IDA beads were washed three times with 100 μ l buffer containing 20 mM Tris-HCl, pH 7.6, 200 mM NaCl. 1.3 μ M of the bait protein containing 6x-His tag was incubated with the beads for 5 min to allow it to bind effectively. This mixture was incubated with a 30-fold excess prey protein (40 μ M) without the 6x-His tag for 20 min at ambient

temperature with constant mixing. As controls, beads without bait protein were incubated with prey protein. Subsequently, three 30 μ l washes were given to all the incubations including controls. 10 μ l of both the beads and washes were loaded onto a 15% SDS-PAGE gel after adding 1x SDS sample buffer with β -mercaptoethanol. Samples were denatured for 5 min at 95°C and then 5 μ l of the samples were run on a self-made 15% SDS gel for ~50 min at 250 V at room temperature until the loading dye reached the bottom of the gel. For protein detection, a Coomassie G 250 staining solution was used. After repeated heating in ddH₂O solution in a microwave at 800 W for 60 sec and a subsequent 10-minute incubation in Coomassie G250, destaining was done in water with an initial heating step in the microwave. After complete destaining procedure, protein bands are visualized using Vilber imaging system. Experiments were performed in triplicate (Fig. 5.11 and Appendix, Fig. x5.9).

2.2.5 *In silico* DNA curvature analysis

To obtain DNA bending of the 48 bp long undamaged substrate, the sequence (original top, Table. 2.9) was uploaded to the online DNA Curvature Python Analysis tool (<https://www.lfd.uci.edu/~gohlke/dnacurve/>). The output file consisted of PDB coordinates. PyMOL was used to visualize the DNA bending and to measure the DNA bend angle at 50% DNA length.

3. Investigating initial lesion detection strategies of DNA glycosylases by automated AFM DNA bending analyses

3.1 Introduction

The genome is constantly under attack leading to spontaneous introduction of modifications in its chemistry and structure. It is estimated that the human genome undergoes up to 10^4 - 10^5 lesions every day in a mammalian cell²⁵⁻²⁹. Base excision repair is the predominant pathway to correct base lesions caused by oxidation, alkylation, and deamination³⁰⁻³² (Fig. 3.1).

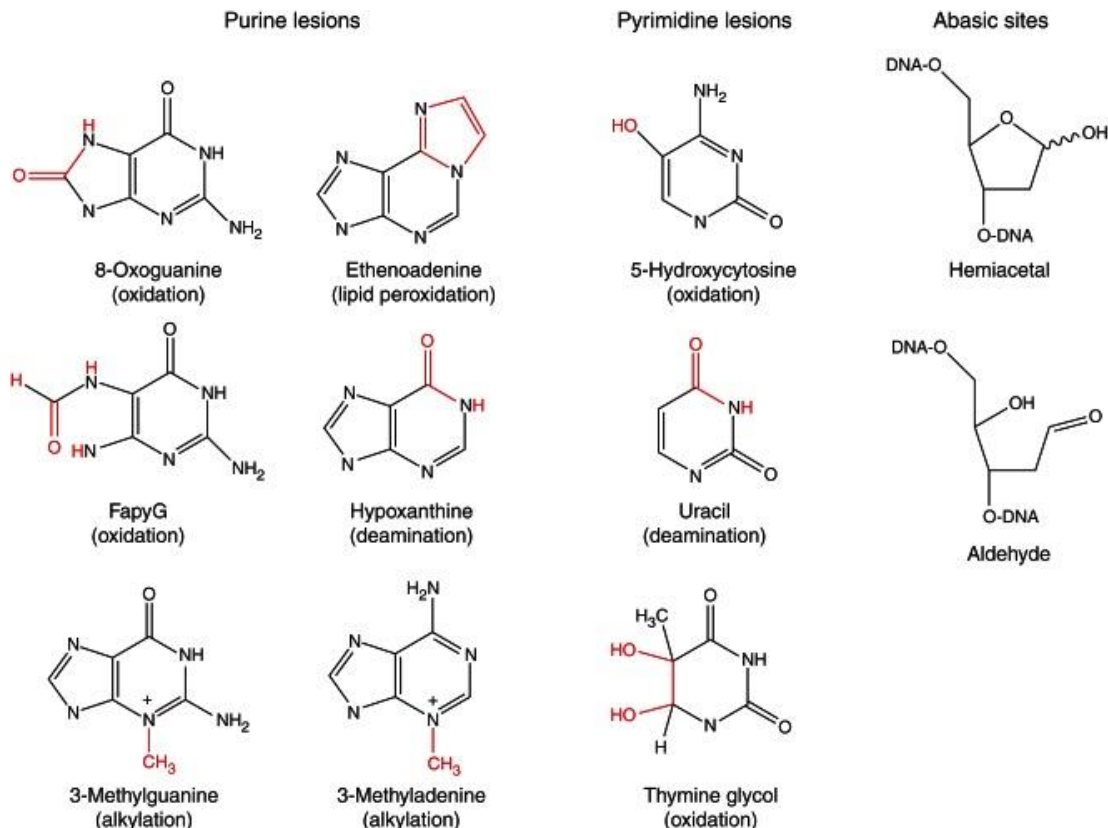


Figure 3.1: Overview of common base lesions and abasic sites. Figure taken with permission from Krokan and Bjørås, 2013.

Most of the base lesions hardly cause any strong distortion in the double helical structure of DNA and are highly innocuous and unobtrusive. They mildly alter the hydrogen bonding pattern with minor distortions in the geometry of the DNA backbone, which go unnoticed by the replicating polymerases leading to transition or transversion mutations and causing cytotoxic effects on the cell. An exemplary of a transversion mutation is shown in Fig. 3.2. Hence, these base lesions with such high mutagenic and cytotoxic potential are a threat to genomic integrity and need to be corrected.

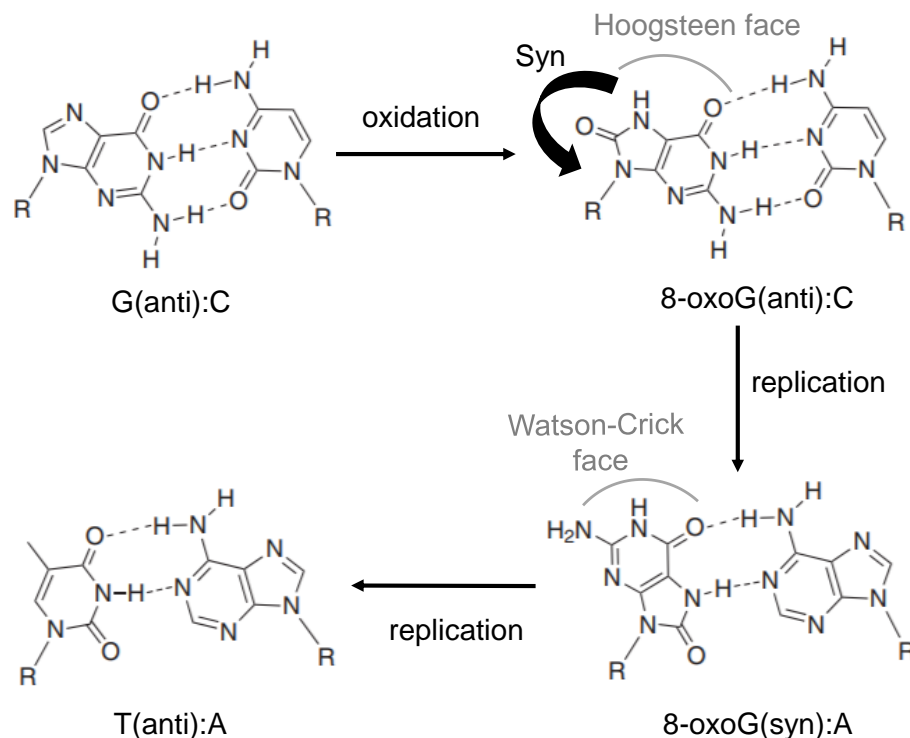


Figure 3.2: Schematic of G:C to T:A transversion mutation. 8-oxoguanine (8-oxoG) is the product of guanine (G) oxidation. 8-oxoG differs from G by an oxo group at C8 position and an NH at N7 position. The presence of an extra oxygen atom does not cause major alterations in the hydrogen bonding pattern of G (anti) to cytosine (C). However, when 8-oxoG is present in the syn-conformation with Hoogsteen face available for hydrogen bonding, it is mispaired with adenine (A) by polymerases because the Hoogsteen face of 8-oxoG resembles Watson-Crick face of thymine (T). In the subsequent round of replication, T is incorporated by polymerases opposite A leading to G:C to T:A transversion mutation.

Lesion detection and elimination of these lesions is performed by the initiating enzymes of base excision repair (BER), the DNA glycosylases. They are highly accurate and efficient in identifying and excising base lesions. Hereby, (almost) all glycosylases employ a common mode of damage recognition: flipping the damaged base into its catalytic active site pocket for sensitive recognition of these inconspicuous lesions and perform slow hydrolysis of the N-glycosidic bond between the damaged base and the phosphate DNA backbone. Action of monofunctional glycosylases on base lesions results in the formation of an apurinic/apyrimidic site (AP-site), which further acts as a substrate for an AP-endonuclease (APE1 in humans) (Fig. 3.3). Activity of APE1 results in the formation of a single strand nick in the DNA 5' to the AP site leaving a 3' hydroxyl group and a 5' deoxyribose phosphate (dRP) group. Bifunctional glycosylases exhibit additional AP lyase activity themselves, forming a 3' α , β -unsaturated aldehyde (PUA) with a 5' phosphate present at the other end of the nick (Fig. 3.3). The resulting substrates are further processed by a DNA polymerase, Pol β and XRCC1 to fill the gap caused by the missing base in single nucleotide BER (SN-BER) or the whole missing strand is resynthesized and FEN1 and Ligase I complete the process in long patch BER (LP-BER)³³ (Fig. 3.3).

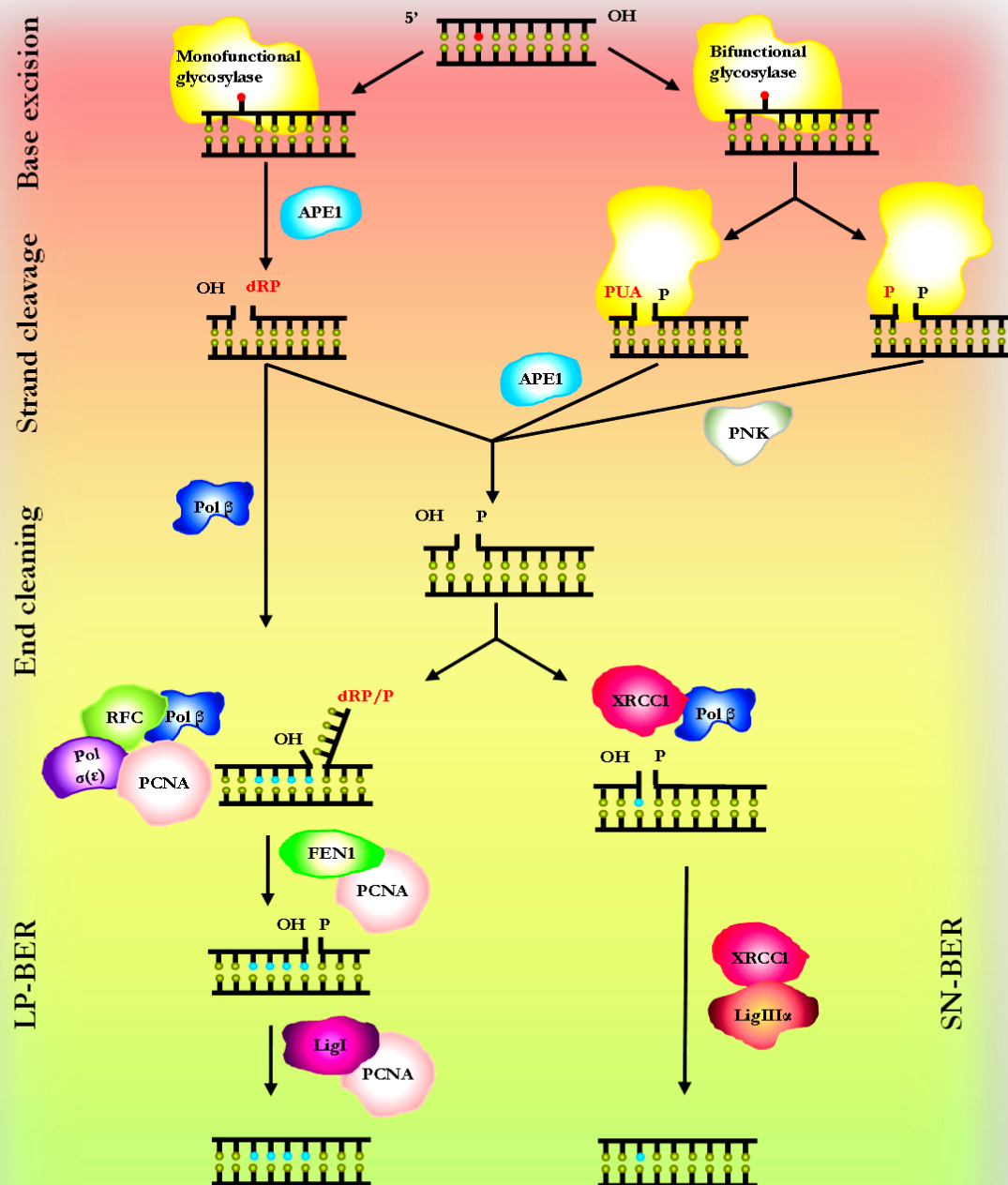


Figure 3.3: Overview of base excision repair (BER) pathway. DNA glycosylases perform initial lesion search and excision leaving behind an abasic site in the DNA. In the case of monofunctional glycosylases, the AP site is further processed by apurinic/apyrimidic endonuclease (APE1) while bifunctional glycosylases themselves perform AP lyase activity on the modified DNA ends (dRP: deoxyribose phosphate, P: phosphate, PUA: 3'-phosphor- α , β -unsaturated aldehyde). The nucleotide gap is filled by Pol β in the single nucleotide BER (SN-BER) and sealed by Ligase III α /XRCC1. In the long patch BER (LP-BER), a stretch of DNA nucleotides is synthesized by a strand displacing polymerase machinery. Finally, flap endonuclease

(FEN1), and Ligase I process the resulting single-strand DNA (ssDNA) flap and seal the gaps in the DNA. Figure adapted with permission from Hegde et al, 2008.

Glycosylases face an uphill task of identifying small non-bulky lesions among the excess of undamaged DNA in the cells. Interrogation of each base in the DNA would be highly inefficient and energy consuming during lesion search by a glycosylase. Glycosylases are known to perform lesion search independent of any usage of biochemical energy. Instead, they have adopted strategies to identify lesions while they are still embedded inside the DNA double helix. Lesion detection by a simple one-dimensional sliding across DNA seems unlikely for an efficient search, and in fact, glycosylases have been shown to perform three-dimensional facilitated diffusive search for damages, termed hopping, in addition to sliding, across the sea of undamaged bases³⁴. Glycosylases have evolved to exploit the altered DNA stability or flexibility at the site of lesion to perform an efficient search during initial lesion detection.

Each glycosylase is designed for the search, detection, and excision of a limited number of base lesions. DNA glycosylases are believed to employ both active and passive mechanisms during initial lesion detection¹⁶. This study aims to investigate the passive mechanism of lesion search by these enzymes. Double stranded DNA (dsDNA) undergoes local conformational modulations referred to as thermal base breathing. dsDNA when heated undergoes transition to a single stranded DNA (ssDNA). The midpoint of this transition is defined as DNA thermal melting temperature (T_m). Base breathing occurs below T_m where the inter-strand hydrogen bonds momentarily break and form between the bases and base unstacking takes place³⁵. The rate of disturbances in base stacking and base pairing is low and forms the major contributing reasons for base breathing of duplex DNA. The A:T and the G:C base pairs undergo base breathing at rates of >1,000 /sec and >100 /sec, respectively^{36,37}. Both base pairing and stacking are affected in the presence of lesions³⁸. Hence, the lesions themselves promote a “passive” mechanism for their recognition by locally destabilizing the DNA. Base breathing involves a fraction of $\sim 10^{-6}$ of the total dsDNA at a given time³⁵. This transient strand separation in dsDNA gives an increased chance for single strand DNA binding targets such as transcription factors and base excision repair proteins to bind. The probability of binding is further increased for repair proteins due to higher base fluctuations occurring at the lesions.

Base lesion detection and excision requires three forms of the glycosylase complexes. Lesion search among vast undamaged bases requires the enzyme to be in a special form which is referred to as Search complex (SC) conformation. The second step is to successfully detect the base lesion using the altered structural or mechanical property of DNA at this site. The enzyme undergoes transition for this purpose, from the SC conformation to an interrogation mode, referred to as Interrogation complex (IC) conformation. This transition step is characterized by the altered properties of DNA at the sites of lesion. Base lesions, however, do not significantly perturb DNA while they are still embedded in the double helix. This study thus aims to understand this enigmatic step of

lesion search and detection. Finally, a transition from IC conformation to an Excision complex (EC) conformation is dictated by a complex reaction co-ordinate for base flipping into the the enzyme active site and excision by hydrolysis of the glycosidic bond (Fig. 3.4).

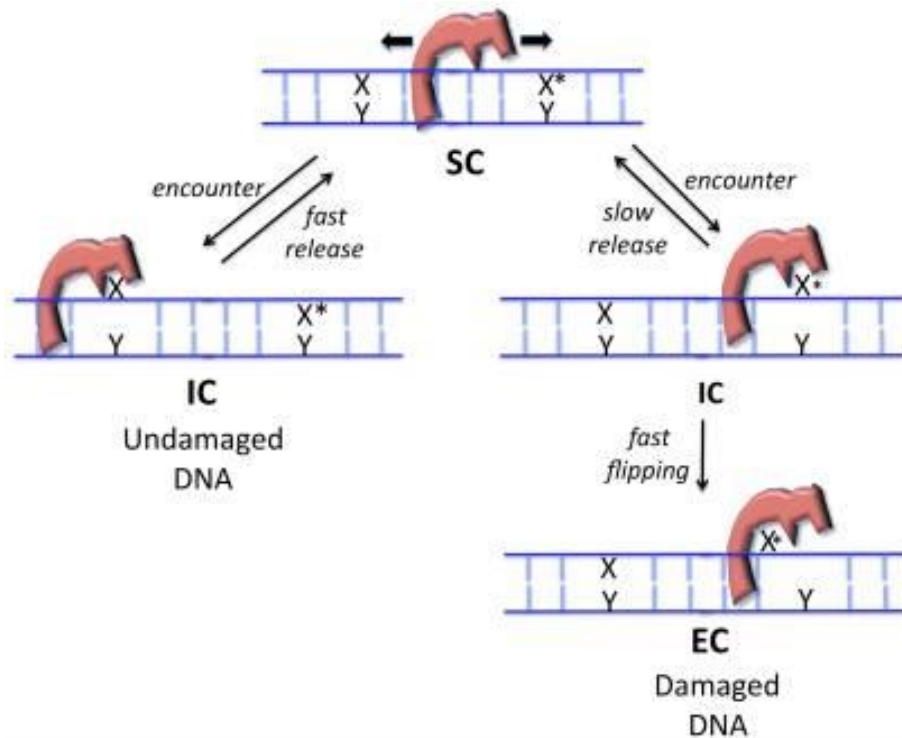


Figure 3.4: Lesion recognition by different glycosylase-DNA complexes. In the SC conformation, enzyme binds to non-specific DNA and performs lesion search by sliding and hopping along the DNA (indicated by black arrows). In the IC conformation, glycosylase performs extrahelical or intrahelical interrogation of lesions and undamaged bases. The EC conformation is specific to damaged bases, where the lesion is sequestered into the active site. Figure taken with permission from Stivers et al, 2010.

DNA glycosylases are generally small proteins with sizes ranging from ~20- 50 kDa. In general, glycosylases are characterized by a helix-hairpin-helix (HhH) motif^{39,40} required for DNA binding in the minor groove and a hydrophobic finger, hairpin or a wedge amino acid, which is inserted in the place of the base lesion to stabilize DNA duplex^{16,41,42}. Glycosylases perform a pinch-push-pull strategy resulting in phosphate pinching which further destabilizes the DNA^{43,44}. These actions by glycosylases constitute the “active” mechanism of lesion detection. The combination of “passive” destabilization of the DNA by the lesions themselves and “active” destabilization by the glycosylases finally results in preferential flipping of a damaged base into the catalytic active site pocket for excision (defined as excision complex (EC) conformation). The wedge residue that is inserted in the DNA backbone in the place of lesion stabilizes the whole glycosylase EC conformation¹⁶. It has been shown that the wedge residue is responsible for glycosylase pausing for interrogation and DNA kinking at damaged bases⁴¹. Wedge residue is used by the glycosylases to test the flexibility and robustness of the base pairs in the DNA.

Mutation of this key residue to alanine led to faster diffusion rates compared to wild type glycosylases⁴¹. Uracil DNA glycosylase (UDG) is a homologue of thymine DNA glycosylase (TDG) and targets Uracil in the G:U wobble base pair. An exemplary of phosphate pinching performed by this enzyme complex is depicted in Fig. 3.5.

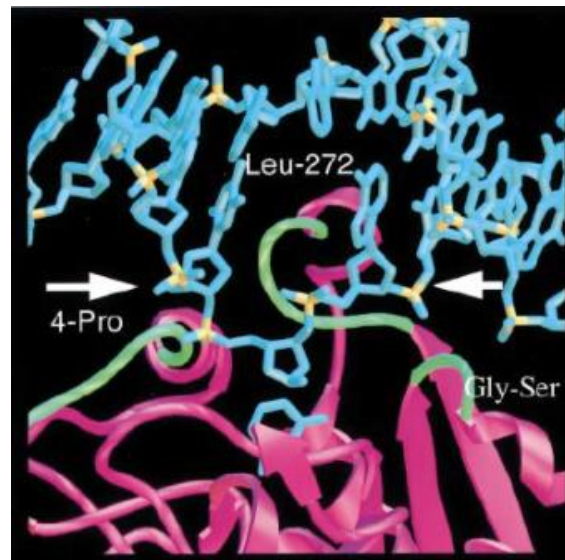


Figure 3.5: Phosphate pinch by UDG. Three conserved Ser-Pro loop motifs (in green) in the UDG enzyme compress the flipped uracil containing strand in DNA backbone (in blue) at the phosphates present 5' and 3' (indicated by white arrows) of uracil. Figure taken with permission from Parikh et al, 1998.

Previous results from single molecule AFM studies of human thymine DNA glycosylase (hTDG) and human oxo-guanine DNA glycosylase (hOGG1) bound to undamaged DNA showed a dynamic equilibrium during lesion search between the interrogation complex (IC) complex (where glycosylase kinks the DNA and attempts to flip the base into its catalytic active site pocket) and the search complex (SC) conformation¹⁶. This study thus proposed continuous lesion interrogation of undamaged DNA by these two DNA glycosylases in a model describing the initial lesion detection strategies employed by glycosylases. Similar DNA bend angles in the search complex (SC) conformation and the innate bending seen at the respective target lesions in the absence of the glycosylases led to the proposal that the precise degree of altered stability and modified structural and mechanical properties of the DNA at a target lesion assists glycosylases in lesion detection. The search complex conformation observed for both glycosylases thus represents passive bending at the respective target lesion since the bend angles match that of lesions in the absence of proteins. Both glycosylases then further attempt to evert the bases by strongly kinking both undamaged and lesion DNA by an active, energy consuming mechanism. This distinct energy requirements for glycosylases to bend the damaged *versus* undamaged DNA serve as an initial lesion detection signal in the proposed model. Less energy requirement to bend the DNA at the target lesion compared to undamaged sites then further advocates longer glycosylase residence times and

pausing at the lesions. During these enhanced residence times, increased destabilization at target lesions and concomitant higher base breathing frequency further leads to preferential formation of the flipped out base state, which is then inserted into the catalytic active site pocket of the enzyme for excision (compared to undamaged bases).

In this thesis, this model was expanded to different, structurally diverse glycosylases and an experimenter independent, automated MatLab based tool was developed for investigation of DNA bend angles from AFM images that allowed for high-throughput analyses. As controls, hTDG and hOGG1 were also included in these studies. The results from the automated tool demonstrated similar bend angle states for these enzymes as obtained from the previous manual analyses. In addition, the automated AFM analyses were further validated by ensemble fluorescence resonance energy transfer (FRET) measurements as well as FRET simulations. The automated MatLab tool was then used to measure DNA bend angles from AFM images of DNA substrates at various target lesions in the absence of protein, and at glycosylase complexes with nonspecific (undamaged) DNA. In addition to testing hTDG- and hOGG1-DNA complexes, the tool was applied to human alkyl adenine glycosylase (hAAG) and *E. coli* adenine DNA glycosylase (MutY). The MatLab tool has been deposited and is freely available for public use at Open Science Framework at <https://osf.io/yhwuc/>.

This work has been published in Scientific Reports in 2020 [Bangalore and Heil *et al*, Automated AFM analysis of DNA bending reveals initial lesion sensing strategies of DNA glycosylases, *Sci Rep*, 2020, doi: <https://doi.org/10.1038/s41598-020-72102-7>] and was a collaborative effort combining AFM analyses and ensemble FRET studies, development of MatLab tool, and FRET simulation data. MatLab tool and FRET simulation data were contributed by AG Heinze, RVZ Würzburg.

3.2 Results

The aim of this study is to compare DNA bend angles of the nonspecific glycosylase-DNA search complexes with the innate bending of the corresponding target lesions. DNA bend angles were measured using the novel MatLab tool. The following sections describe the establishment of the MatLab routine used in this study.

3.2.1 MatLab analysis routine for characterization of glycosylase target lesions

An advantage of AFM lies in not only detecting the invariable distortions but also changes in the elasticity of DNA manifested as variations in the DNA bend angle distributions. All DNA substrates were designed such that they contain the required lesion at 50% of total DNA length for unambiguous localization of the lesion (see Methods, Fig. 2.6). In the first step, DNA strands in the images are selected by thresholding in ImageJ and are skeletonized in FIESTA (Fig. 3.6b). The selected DNA filaments are then fed into the MatLab routine, where the total lengths of the filaments (contour lengths) are measured. For measuring DNA bend angles, only strands with correct lengths (~ 170 nm for ~ 500 bp DNA considering 0.34 nm/bp⁴⁵) are used. This excludes the broken or overlapping DNA filaments to avoid false mapping of the centers of DNA filaments (Fig. 3.6d). The MatLab tool measures the bend angles at target lesions using tangent geometry with varying query point distances (here, 8 nm). The individual steps are also described in the Methods section 2.2.3.5a, and complete instructions for the usage of the software and configuration settings are given in the appendix section (Fig. x3.1 and Table. x3.1) and are available at <https://osf.io/yhwuc/> (OSF). The DNA bend angle is defined as the deviation from the straight backbone of the DNA (180° -measured angle). As a first control, the MatLab analysis was applied to undamaged (non-specific) B-form DNA. The DNA bend angle was measured at the 50% position on the DNA filaments (Fig. 3.6e). A Gaussian fit to the bend angle distribution showed a maximum at $\sim 8^\circ$ ($\sim 50^\circ$ width) which matches well with the reported values of ~ 2 - 15° ⁴⁶⁻⁴⁸. The undamaged DNA oligonucleotide sequence used for AFM substrate generation was further fed into an online DNA curvature analysis python extension tool (C. Gohlke, <https://www.lfd.uci.edu/~gohlke/dnacurve/>), which gave a bending of $\sim 6^\circ$ (shown in inset, Fig. 3.6e). As an additional control, a B-form nicked DNA substrate was also imaged that contained a centrally located single strand cut (nick) at the same position as the glycosylase target lesions of the other substrates, and the same sequence as the undamaged DNA. At the 50% position, the nicked DNA showed a maximum at a DNA bend angle of 0° (unbent DNA), which contained two unbent conformers, $\sim 30\%$ with a narrow Gaussian width of $\sim 10^\circ$ and $\sim 70\%$ with a broad Gaussian width of $\sim 70^\circ$. The very broad distribution indicates the high elasticity of the nicked substrate compared to the intact B-form DNA likely due to loosening of the DNA double helical parameters caused by the break in the DNA backbone at the site of nick (Fig. 3.11a).

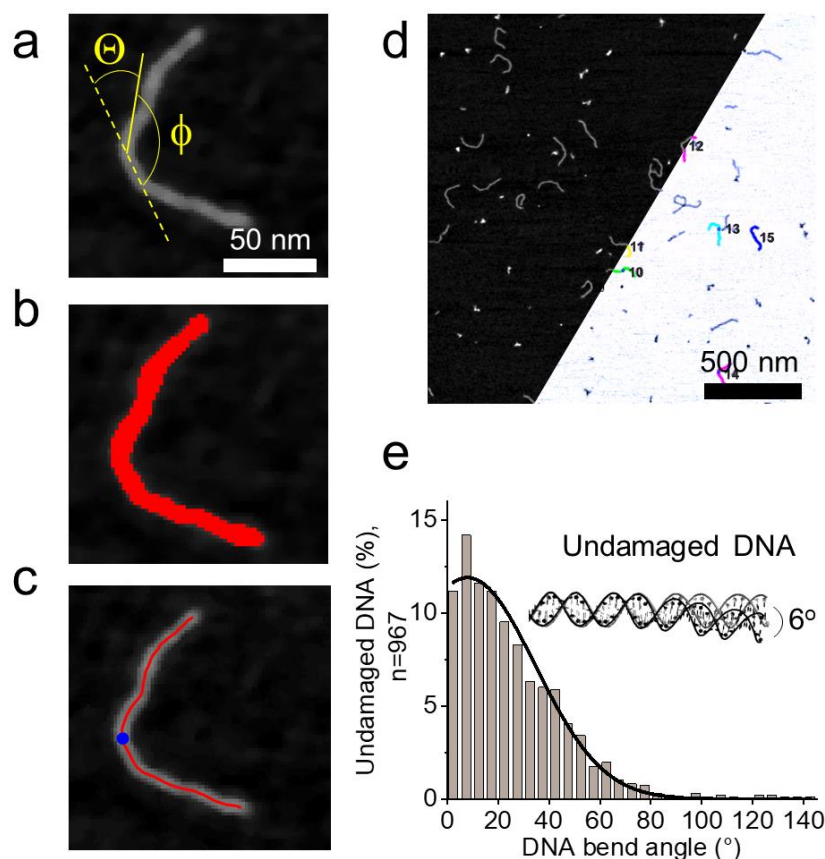


Figure 3.6: Schematic of individual steps of automated DNA bend angle measurement at target lesions. (a) Manual DNA bend angle measurement at the lesion located at 50% of the total DNA length. The DNA bend angle θ is measured as $\theta = 180^\circ - \phi$. Scale bar: 50 nm. (b) By thresholding in ImageJ, DNA filaments are selected. (c) Skeletonization of the DNA filaments by placing 2 nm rigid segment lines along the center of the DNA backbone. Skeletonized DNA filaments represent the DNA contour lengths. The blue filled circle represents the 50% position on the DNA. (d) AFM image used by MatLab for DNA bend angle measurements. Upper half: original AFM image. Bottom half: Output MatLab images show the colored DNA filaments of correct length used for analyses. MatLab uses tangent geometry by laying two tangents on either side of the 50% position of DNA length and measuring the DNA bend angle at the intersection of both the tangents (as shown in (a)). Scale bar: 500 nm. (e) The measured DNA bend angles for undamaged DNA show a maximum at $\sim 8^\circ$, consistent with $\sim 6^\circ$ predicted by the online curvature tool for this sequence (inset). Depending on the number of maxima observed, a (single or multiple) Gaussian fit to the DNA bend angle distribution indicates the different conformational state(s) as the center(s) of the Gaussian peak(s). A shift in the bend angle states for the lesion containing substrates compared to the undamaged DNA signifies enhanced DNA bending (shift to higher bend angles) or loosening of DNA double helical constraints (lower bend angles) by the presence of the lesion.

Target lesions of DNA glycosylases, hOGG1 (oxoG:C), MutY (oxoG:A and G:A), hTDG (G:U) and hAAG (ethenoA) were analyzed for bend angles. All lesions showed distinct bend angle distributions with multiple conformations and widths. Multimodal Gaussian fits to all these states of different lesions revealed specific bend angles and are summarized in the table 3.1 (see section 3.3.2 for details).

Lesion	Bend angles at lesions $\pm 2 \sigma$ width (percent population)		
Undamaged	-	$8^\circ \pm 53^\circ$ (100%)	-
oxoG	-	$10^\circ \pm 35^\circ$ (71%)	$48^\circ \pm 35^\circ$ (29%)
oxoG:A	$0^\circ \pm 11^\circ$ (25%)	$15^\circ \pm 14^\circ$ (21%)	$30^\circ \pm 43^\circ$ (54%)
G:A	$0^\circ \pm 10^\circ$ (17%)	$15^\circ \pm 15^\circ$ (20%)	$30^\circ \pm 46^\circ$ (63%)
G:U	$4^\circ \pm 20^\circ$ (56%)	$23^\circ \pm 5^\circ$ (10%)	$37^\circ \pm 25^\circ$ (34%)
ethenoA	$0^\circ \pm 15^\circ$ (24%)	$8^\circ \pm 56^\circ$ (76%)	-
Nick	$0^\circ \pm 10^\circ$ (27%) $0^\circ \pm 70^\circ$ (73%)	-	-

Table 3.1: DNA bend angle statistics for BER target lesions obtained from automated AFM analyses. DNA bend angles were determined as the centers of Gaussian fits to the DNA bend angle distributions obtained from automated AFM bend angle analyses of DNA containing a specific lesion at 50% DNA length.

The bend angles obtained from MatLab tool for target lesions were validated using the manual bend angle measurements. The DNA bend angle distribution for G:U mismatch from MatLab when plotted at a bin size as previous manual analyses¹⁶, showed similar bend angle states (Fig. 3.7). The two bend angle states $\sim 20^\circ$ and $\sim 40^\circ$ (Fig. 3.7c) of G:U lesion obtained by automated analyses give an average of $\sim 30^\circ$ (Fig. 3.7b) due to a lower resolution at a higher bin size of $\sim 10^\circ$.

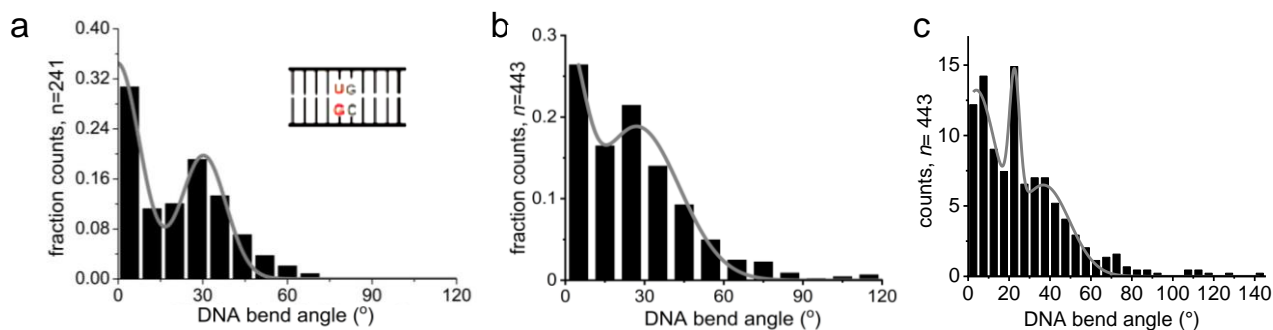


Figure 3.7: Comparison of manual (a) and automated (b, c) measurements of DNA bending at G:U mismatch sites shown exemplarily. Bin size of $\sim 10^\circ$ (b) and $\sim 3.5^\circ$ (c) show the different bend angle states. Finer binning resolves the $\sim 30^\circ$ bend angle (b) state into $\sim 20^\circ$ and $\sim 40^\circ$ angles (c). Manual bend angle distribution is obtained with permission from Buechner et al. (2015) NAR 43: 2716-2729.

3.2.2 MatLab analysis routine for characterization of glycosylase-nonspecific DNA complexes

To assess the bend angles of glycosylase lesion search complexes and investigate if these indeed mirror the bending shown by the corresponding target lesions, the MatLab pipeline was applied to measure DNA bend angles in AFM images at glycosylase complexes on undamaged DNA (Fig. 3.8).

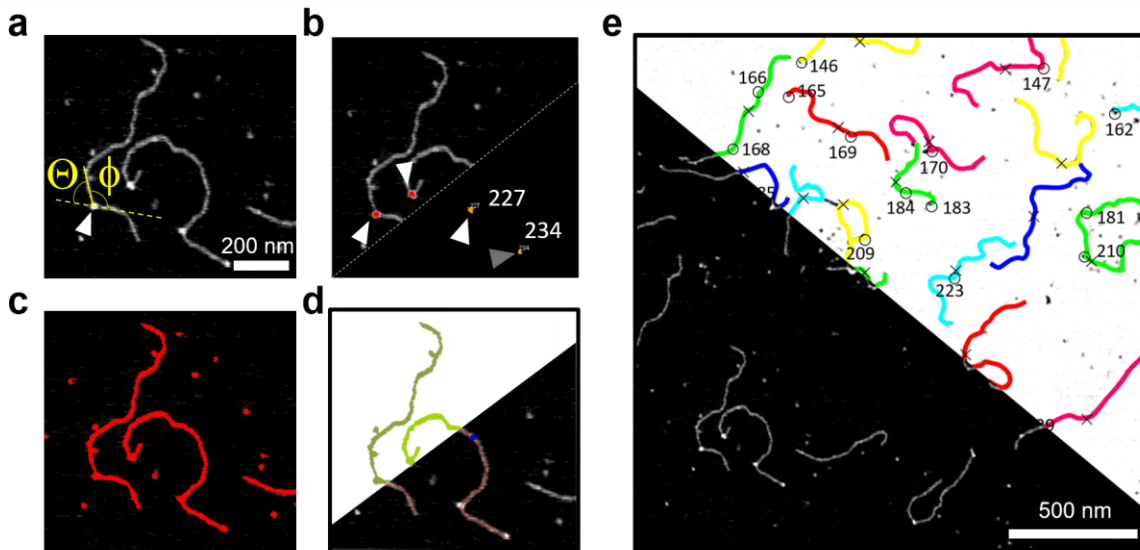


Figure 3.8: Schematic of individual steps of automated DNA bend angle analyses at protein bound sites on undamaged DNA. (a) Manual bend angle measurement at the protein bound sites on the DNA. The intersection of the two tangents laid across the protein bound site (white arrow) along the DNA backbone provides the angle ϕ . DNA bend angle θ is obtained by the equation $\theta = 180^\circ - \phi$. DNA bend angle is defined by the deviation from the straight conformation of the DNA backbone (yellow dashed line). Scale bar: 200 nm (b) Thresholding (upper half) and counting the protein peaks using 3D object counter (lower half), protein peaks bound to DNA are selected (white arrows). Protein molecules bound at the ends of DNA filaments are not considered for analysis (grey arrows). (c) Thresholding and (d) selecting DNA filaments (upper half) using shape filter plugin in ImageJ. Selected DNA filaments are skeletonized in FIESTA (lower half). Red lines represent skeletonized DNA backbone. (e) The skeletonized DNA filaments are fed into MatLab together with protein position coordinates. The output MatLab images are overlaid on the original images. Protein bound positions are highlighted by open circles and multiple-colored lines are the DNA filaments. The x marks denote themed points of the DNA filaments. Scale bar: 500 nm. The angle ϕ between tangent lines to points on the DNA at a user defined query distance (here: 8 nm) from the protein positions is automatically measured for all DNA bound protein complexes, and the corresponding DNA bend angles θ are returned by the tool.

In this routine, two parallel steps are employed to pre-process the images for separate selection of DNA and bound proteins. Tracing the DNA backbone and locating the protein peaks thus form the essential steps prior to DNA bend angle analysis by MatLab. Proteins bound to DNA have increased heights when compared to free protein molecules. This feature is exploited by the threshold filter (specifically, intermodes threshold filter) and the 3D object counter is then applied in ImageJ to select only the DNA bound protein

molecules. In parallel, using the Yen threshold filter and elongation parameter in the shape filter plugin in the ImageJ software, DNA strands can be selected (Fig. 3.8c). This process excludes free and bound protein molecules. In the next step, 2 nm rigid line fragments are assigned along the center of the DNA backbone to skeletonize the DNA strands as above for DNA bend angle analyses at lesions. In the next step, the obtained protein positions are overlaid with the skeletonized DNA lines to enable MatLab to measure the bend angles at the protein co-ordinates. For this, the MatLab software places tangent lines along the DNA skeleton around the protein peaks and the angle between them is measured (Fig. 3.8a) as described above for measurements of bend angles at the 50% position on DNA substrates. All configuration settings for FIESTA and MatLab softwares are specified in the appendix section (Table. x3.2) and each step of the approach is described in the Methods section. An exemplary case has further been provided in the appendix section, Fig. x3.2. Detailed instructions and an exemplary image are also available at <https://osf.io/yhwuc/> (OSF).

The DNA bend angle distributions for nonspecific glycosylase-DNA complexes indicated SC and IC conformations with distinct bend angle states. Multimodal Gaussian fits to all these states revealed specific bend angles and are summarized in the table 3.2.

The bend angles for the IC conformations of the glycosylase-DNA complexes are validated against the bend angles from crystal structures showing IC/EC complexes of respective glycosylases (see section 3.3.2 for details).

Furthermore, the results obtained from the automated AFM analysis were validated by ensemble fluorescence energy transfer (FRET) measurements performed in solution, as well as FRET simulations (For details refer appendix, Fig. x3.11). The principle of the ensemble FRET measurements is described in Fig. 3.9a. FRET efficiencies calculated using intensity ratios of acceptor emission at donor and acceptor excitation was used to calculate average (\pm SD) DNA bend angles from triplicate measurements (Fig. 3.9b). Average bend angle values for glycosylase-DNA complexes calculated using the bend angle states and their populations from automated AFM analyses were similar to those obtained from ensemble FRET measurements and FRET simulation studies (Table. 3.2).

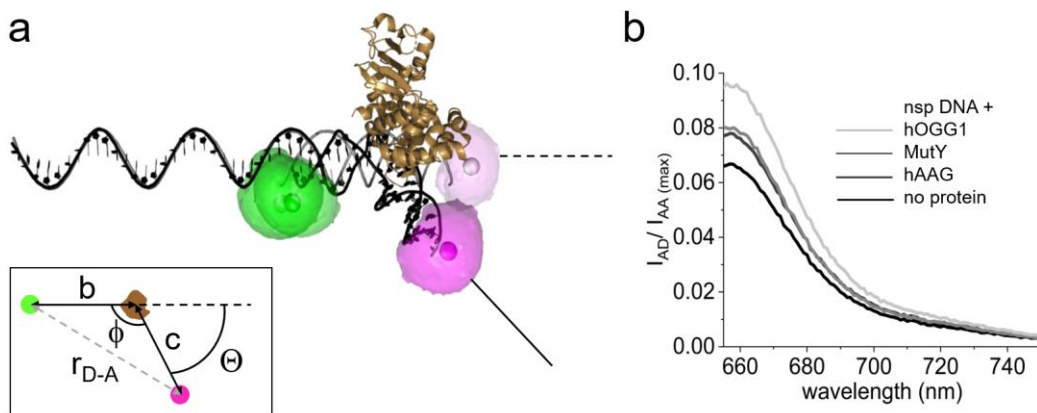


Figure 3.9: Ensemble FRET measurements of average DNA bend angles for glycosylase-DNA complexes. (a) Principle of the intensity-based ensemble FRET measurements. FRET measurements were made using short undamaged DNA substrates (20 bp with a 28 nt overhang) that contained a FRET donor, Cy3 (green), at one end of the dsDNA stretch, and a FRET acceptor, Cy5 (magenta), at the other end. Measurements were carried out in the absence or presence of different glycosylases (here hOGG1, shown in gold, PDB id: 1EBM). High FRET is indicated by bright Cy5 and faded Cy3 molecules and low FRET by bright Cy3 and faded Cy5 molecules moved away from each other. With the Förster distance R_0 of the donor-acceptor pair, the distance r_{D-A} between donor and acceptor (grey dashed line) corresponds to the measured FRET efficiency E : $r_{D-A} = R_0 (1/E - 1)^{1/6}$. From r_{D-A} , the angle Φ introduced in the DNA by protein binding is calculated using the cosine law, $\cos \Phi = (r_{D-A}^2 - b^2 - c^2) / (-2bc)$ and finally the DNA bend angle Θ (deviation from straight DNA backbone) is calculated using $\Theta = 180 - \Phi$ (inset). (b) Cy5 (acceptor) emission upon Cy3 (donor) excitation is normalized to maximum acceptor emission at acceptor excitation for the FRET substrates in the absence and presence of different glycosylases. Full spectra are presented in the appendix section (Fig. x3.8). Intensity ratios at donor and acceptor excitation formed the basis of FRET efficiency calculations and were further used for calculating DNA bend angles for each glycosylase-DNA complex.

	DNA bend angles from AFM $\pm 2\sigma$ width (relative population)	Average from AFM angles weighted by populations	$E_{\text{FRET}} \pm \text{SD}$ from ensemble measurements	DNA bend angle $\pm \text{SD}$ from FRET	$E_{\text{FRET}} \pm \text{SD}$ From simulations	DNA bend angle $\pm \text{SD}$ from simulations
Non-specific (nsp) DNA	8°±53° (100%)	8°	0.191±0.002	7°±4°	0.186±0.001	6.6°±0.2°
nsp DNA + hTDG	30°±25° (51%)	41°	n.d	n.d	n.d	n.d
	69°±25° (49%)					
nsp DNA + hOGG1	0°±20° (36%)	50°	0.291±0.007	49°±2°	0.235±0.035	41.5°±5.7°
	35°±15° (17%)					
	67°±20° (47%)					
nsp DNA + MutY	15°±30° (46%)	34°	0.233±0.019	34°±8°	0.224±0.019	32.2°±2.7°
	50°±30° (54%)					
nsp DNA + hAAG	0°±10° (27%)	33°	0.230±0.005	33°±2°	0.209±0.011	25.9°±1.4°
	16°±15° (28%)					
	44°±30° (45%)					

Table 3.2: DNA bend angle statistics obtained from AFM, ensemble FRET measurements and FRET simulations for non-specific DNA with and without bound glycosylases. DNA bend angles were obtained from the centers of Gaussian fits to the bend angle distributions for non-specific DNA (50% position) or of glycosylase-DNA complexes from automated AFM analyses. FRET efficiency data shown here are averages from triplicate measurements. FRET simulations were based on the bend angle states and their corresponding populations from AFM experiments, and DNA bend angles were subsequently calculated from the obtained average FRET efficiencies. Abbreviation: n.d- not determined, SD- standard deviation.

3.3 Discussion

3.3.1 Method evaluation

Single molecule techniques offer invaluable advantages for the study of transient and highly dynamic protein-nonspecific DNA interactions, because they allow to resolve individual states that go unnoticed by bulk biochemical and biophysical techniques. In the work presented here, single molecule atomic force microscopy has been employed to resolve the multiple, distinct states involved in glycosylase interactions with undamaged DNA and the multiple, distinct states of their respective target lesions. Although they provide much higher, near-atomic resolution, structural techniques such as x-ray crystallography and cryo-electron microscopy require highly ordered complexes with homogeneity and hence are not suitable for studying distinct states of nonspecific protein-DNA complexes. Single-molecule FRET (smFRET) has the potential to resolve individual conformers in protein interactions in solution with high spatial and time resolution⁴⁹. Although smFRET is a powerful method to study dynamic protein-DNA interactions, it fails to detect individual states of bending when the protein-DNA complexes under study have a wide dynamic range. In addition, some of the more subtle conformational changes in protein-DNA systems may not exhibit detectable changes in FRET, for example change in DNA bending from 8° curvature of B-form DNA to 0° bend angle at a nick. In fact, multiple efforts by the collaboration partners (AG Heinze, RVZ Würzburg) to perform smFRET using wide-field total internal reflection fluorescence (TIRF) microscopy and fluorescence correlation spectroscopy (FCS) were unsuccessful and provided only ambiguous results. Instead, ensemble FRET can easily be applied to these samples and has been applied here. However, ensemble FRET can only provide average DNA bend angles. These averages were strongly consistent with the average bend angles calculated from bend angle states and populations from automated AFM analyses and served to confirm the accuracy of the novel MatLab based bend angle analyses. Although ensemble FRET is fast and easy, it cannot provide insight into the individual contributing conformers that are essential for understanding the mechanism behind glycosylase-DNA interactions. FRET efficiencies are influenced by the binding positions of the glycosylases on the DNA. Using smFRET, it is difficult to visualize these binding positions and hence cannot give information of the individual states. However, merit of AFM is that it allows visualization of the complexes and hence is suited for separation of distinct states. FRET simulations were therefore performed using the DNA bend angles and the populations of glycosylase-undamaged DNA complexes for each state (For details refer appendix, Fig. x3.11). Results from these simulations agreed with those from ensemble FRET measurements and strongly supported the automated AFM analyses.

For AFM imaging, DNA and protein molecules are deposited onto mica in the presence of cations like Mg²⁺. This process is reported to enhance the DNA elasticity on a solid substrate than compared to in solution⁵⁰. Since the DNA elasticity is already increased at the sites of base lesions, deposition onto a solid surface may further help in resolving the

minor differences observed to study such distinct conformational states of these innocuous lesions.

The major limitation of manual AFM analyses lies in the number of data points, which leads to low resolution in the resulting DNA bend angle distributions. This in turn does not allow the resolution of more subtle conformational variations, such as the distinction between straight DNA backbone and bend angles near zero-degree DNA bending (Fig. 3.7b and 3.7c). AFM combined with the presented automated AFM analysis tool has the capacity to provide large amounts of data, which then allows finer binning in the resulting bend angle distributions and thus the resolution of small differences in bend angle states for protein-DNA complexes and specific DNA sites, especially in the low bend angle regime.

3.3.2 Characterization of BER target lesions in the absence of glycosylases and glycosylase complexes with undamaged DNA

Along with measuring DNA bending of non-specifically bound hTDG and hOGG1 complexes, the MatLab routine was also applied to characterize complexes of two other glycosylases, MutY and hAAG. As a negative control, experiments with DNA substrate in the absence of glycosylases was also included, which did not show any elevated peaks or blobs on the DNA filaments confirming the substrates to be non-contaminated. The DNA bend angles in glycosylase complexes were compared with bending seen at the corresponding DNA target lesions.

hTDG-undamaged DNA complex and G:U target lesion

The DNA bend angle distribution for non-specifically bound hTDG shows a bimodal behaviour with two bend angle states, a mildly bent $\sim 30^\circ$ and a strongly kinked $\sim 70^\circ$ state (Fig. 3.10a). Bimodal DNA bend angle distribution for hTDG was demonstrated earlier by manual measurements¹⁶ and the results from MatLab stand validated against these earlier measurements. The strong kinking observed is interpreted as the hTDG interrogation complex (IC) conformation, in which the enzyme attempts to flip the interrogated base into the catalytic active site pocket of hTDG by phosphate pinching^{16,51}. The 30° bend angle state represents the search complex (SC) conformation of hTDG as described in the earlier studies¹⁶. Similar hTDG IC and SC populations seen from the automated DNA bend angle distributions are also consistent with the previous manual analyses¹⁶. These data suggest that target lesion testing by hTDG is a dynamic process during lesion detection and validate the novel automated analysis tool.

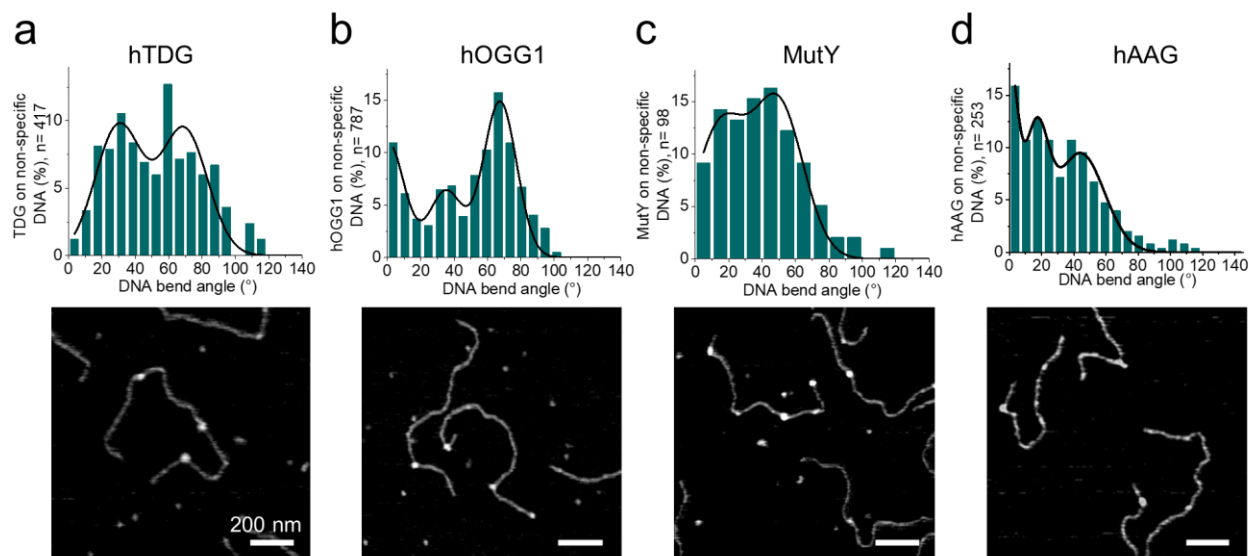


Figure 3.10: Characterization of non-specifically bound DNA glycosylase complex conformations using MatLab based automated DNA bend angle analyses. Top: DNA bend angle distributions with Gaussian fits and Bottom: exemplary AFM images of (a) hTDG, (b) hOGG1, (c) MutY, and (d) hAAG, bound at non-specific sites on DNA. Individual Gaussian fits for the multimodal distributions for each glycosylase system are shown in the appendix section (Fig. x3.9). Scale bar: 200 nm.

Target lesions of hTDG are the G:T and G:U mismatches. G:U and G:T mispairs arise from deamination of unmethylated and methylated cytosine bases in the DNA leading to the formation of wobble base pairs with low thermodynamic stability and significant alterations in the DNA double helix^{37,52} (inset in Fig. 3.11b). The angles between each glycosidic bond and the line segments connecting each C1' atom in the DNA backbone with G:U mismatch are different from the natural B-form DNA⁵³. Previous manual bend angle analyses of AFM data had demonstrated DNA bend angles of $\sim 30^\circ$ at G:U and G:T sites¹⁶. The novel, automated DNA bend angle analysis approach revealed bend angles of $\sim 20^\circ$ and $\sim 40^\circ$ for the G:U mispair (Fig 3.11b). The data also showed a dominant maximum at $\sim 4^\circ$ (56% of total population), which had been also present in the manual analyses (0°) but could not be resolved due to higher bin size ($>4^\circ$) used because of limited data points and difficulty in manually measuring small bend angles (small bend angles are measured as 0°) (Fig. 3.7). In the manual analyses, a DNA substrate with the G:U lesion located at 46% of total DNA length was used. Hence, DNA bend angles were measured at two sites (both 46% from each end of the DNA substrate) which resulted in non-specific DNA background being introduced from one of the two sites in the bend angle data¹⁶. Furthermore, there is no background contamination from the non-specific DNA at 50% positions in the new substrate confirming that $\sim 0^\circ$ observed is due to the lesion which when plotted at a finer bin size of $\sim 3.5^\circ$ showed a mild bending of $\sim 4^\circ$ (Fig. 3.11b). These were the major differences between the automated and manual bend angle

measurements. Bend angle of hTDG SC conformation ($\sim 30^\circ$) is consistent with G:U bend angle ($\sim 30^\circ$, bin size $\sim 10^\circ$).

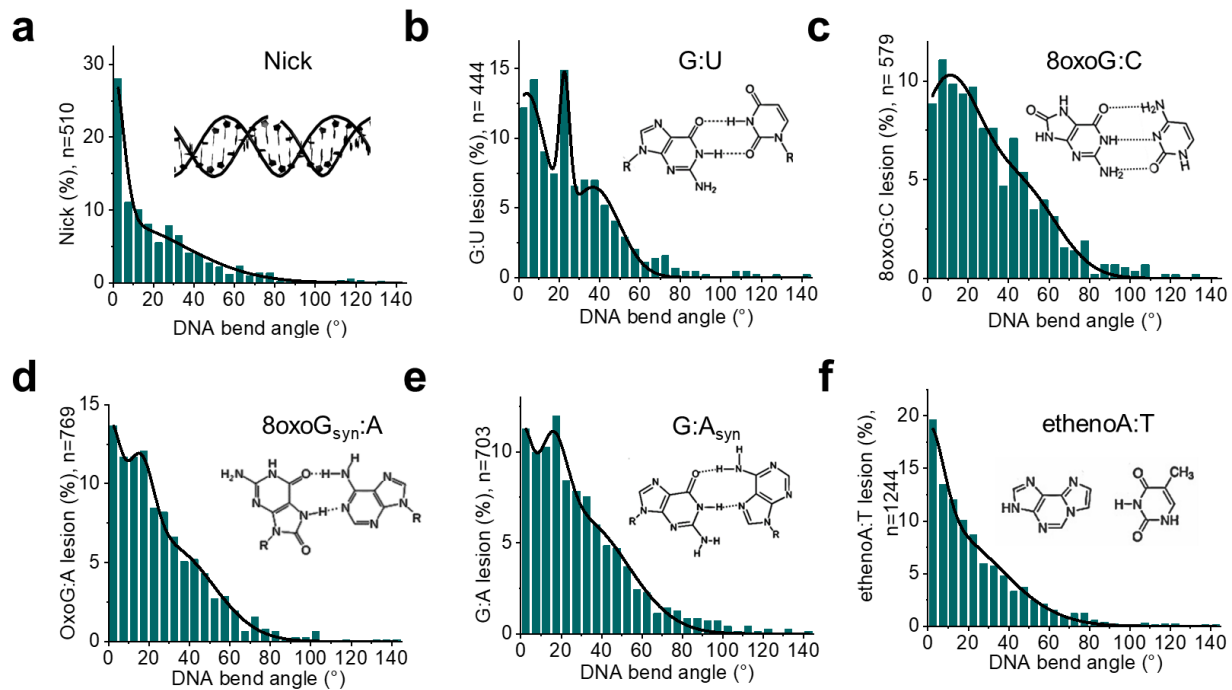


Figure 3.11: Characterization of target lesions of BER glycosylases. DNA bend angle distributions for the DNA substrates with (a) a single strand nick, (b) a G:U mismatch, a target lesion of hTDG; (c) oxoG, a target lesion of hOGG1; (d) an oxoG:A mismatch and (e) a G:A mismatch, both target lesions of MutY, and (f) an ethenoA lesion, a substrate of hAAG. The structures of the individual lesion base (mis-)pairs are shown as insets in the corresponding bend angle distributions. The individual Gaussian fits of each of these multimodal Gaussians are shown in the appendix (Fig. x3.10).

hOGG1-undamaged DNA complex and oxoG target lesion

Automated DNA bend angle analyses of hOGG1 bound at non-specific DNA sites showed a multimodal distribution with three distinct states with bend angles of $\sim 0^\circ$, $\sim 35^\circ$ and $\sim 70^\circ$, consistent with previous manual analyses^{16,54} (Fig 3.10b). Bend angles of $\sim 0^\circ$ and $\sim 70^\circ$ were dominant in the distribution with 36% and 47% of all complexes, respectively, and represent the search complex (SC) and interrogation complex (IC) conformations, respectively, as already established in previous analyses¹⁶ (Fig. 3.10b). The IC state $\sim 70^\circ$ bend angle can be validated very well by crystal structures of hOGG1 bound to an oxoG lesion or undamaged DNA, showing bend angles of $\sim 70^\circ$ and $\sim 80^\circ$, respectively^{55,56}. The intermediate with a bend angle of $\sim 35^\circ$ was observed in the present as well as former DNA bend angle distributions from AFM analyses¹⁶ and was considerably less populated (17% of complexes).

The major DNA lesion substrate of hOGG1 is 7,8-dihydro-8-oxoguanine (oxoG). When cells are continuously exposed to free radical oxygen species, guanine, due to its lowest oxidation potential among all DNA bases, undergoes spontaneous oxidation at its O8 position, which leads to the formation of oxoG⁵⁷. OxoG exhibits double helical constraints only slightly different from natural B-form DNA and does not cause any major base pair destabilization in the DNA^{58,59} (inset in Fig. 3.11c). OxoG retains the flat nature of guanine and the Watson-Crick edge of the base is undisturbed hence causing very little distortions in the DNA^{60,61}. OxoG is found to be thermodynamically moderately unstable and to have an increased flexible nature^{62,63}. Consequently, oxoG can adopt both the *syn*- and *anti*-conformation and exists as a dynamic equilibrium between these two states. In the *syn*-conformation, oxoguanine presents its Hoogsteen face for hydrogen bonding like thymine and can thus be mispaired with adenine during replication and this leads to high mutagenicity^{64,65} (Fig. 3.2). In the *anti*-conformation, oxoG presents its Watson-Crick face to the complementary cytosine base leading to the formation of oxoG:C lesion (Fig. 3.2). This is consistent with the slightly broader bend angle distribution for the oxoG:C lesion compared to undamaged DNA due to slightly increased DNA elasticity or flexibility at the lesion^{62,63}. Besides this, bend angle distributions for oxoG:C lesion and undamaged DNA do not reveal any other major differences, suggesting no major distortions of the DNA by this lesion, consistent with previous studies^{58,59,63,65}. The bend angle distributions reveal two states with Gaussian maxima at $\sim 10^\circ$ and $\sim 48^\circ$, both with Gaussian widths of $\sim 35^\circ$ (Fig. 3.11c).

MutY-undamaged DNA complex and oxoG:A and G:A target lesions

MutY targets oxoG:A mispairs and is an unusual glycosylase because it flips out and removes the adenine base, not the damaged oxoG base from the mispair. Automated bend angle analysis of MutY-undamaged DNA complexes revealed Gaussian maxima at $\sim 15^\circ$ and $\sim 50^\circ$ (Fig. 3.10c). Like hTDG, the SC and the IC species in the bimodal distribution for MutY-DNA complexes show approximately equal populations. The larger (50°) bend angle is consistent with the bend angle of $\sim 55^\circ$ for DNA in the crystal structure of MutY bound to an oxoG:A lesion⁶⁶. Strong bending in the IC ($\sim 50^\circ$) may result from the conformational shift for the oxoG base in the MutY bound state observed in the crystal structure. OxoG preferentially adopts the *syn*-conformation since the *anti*-conformation is structurally and energetically unfavourable^{63,65}. While oxoG is in its *syn*-conformation when paired with adenine, it shifts to the *anti*-conformation when bound to MutY⁶⁶. This suggests that oxoG undergoes a 180° turn about its glycosidic bond, which drives the adenine out of the DNA double helix into the catalytic active site pocket of MutY⁶⁶. The $\sim 15^\circ$ species in the AFM bend angle distribution may correspond to the mildly bent search complex (SC) conformation of MutY-undamaged DNA complex. To investigate this, I will in the following sections again compare this bend angle state with bend angle distributions of MutY target lesions.

Both oxoG:A and G:A are target lesions of MutY. OxoG:A represents a highly distorted Hoogsteen pair (- 15 kcal/mol)⁶⁷. Structural studies, however, have suggested that oxoG:A introduces only mild destabilization in the double helical DNA, similar to oxoG:C⁶⁵. Repair of oxoG by hOGG1 when mispaired with adenine leads to the formation of a G:A mispair. In this mispair, guanine is in its *anti*-conformation and can mispair with the *syn*- as well as *anti*-conformation of adenine^{63,68–70} (inset, Fig 3.11d and e). Both forms exhibit low thermal stability⁶³ and introduce large distortions in the DNA⁷¹.

DNA bend angle distributions of oxoG:A and G:A show minor but significant differences compared to undamaged DNA. Bend angle distributions show overall three conformers with ~0°, 15°, and 30° (Fig. 3.11d and e). Most strikingly, both oxoG:A and G:A have a distinct conformational state with a Gaussian maxima of ~15° that is not present for undamaged DNA (Fig. 3.11d and e), indicating altered structure from natural B-form DNA. The broader distribution for the higher bend angle states especially for G:A (>60% of conformers at 30° ± 46°) are consistent with enhanced instability and deformation of the DNA double helix due to the mismatch. G:A has a higher capacity of helix disruption compared to oxoG:A, consistent with the slightly broader distribution for G:A⁶⁶. The straight species observed (0°, ~20% of total conformers for both G:A and oxoG:A) may again indicate the release of double helical constraints similar to nicked DNA. Importantly, the clear maximum at ~15° (~20% of conformers) for both oxoG:A and G:A is unique to the target lesions of MutY and is comparable to bending in the SC conformation of MutY.

hAAG-undamaged DNA complex and ethenoadenine target lesion

Most glycosylases contain a helix-hairpin-helix (HhH) motif that is involved in strong DNA kinking. hAAG is the only exception in its family of glycosylases without an HhH motif⁷².

The DNA bend angle distribution from automated AFM analyses of hAAG-nonspecific DNA complexes shows Gaussian maxima at ~0°, ~20° and ~45° (Fig. 3.10d). The 20° bend angle state is again consistent with bending observed in the crystal structure of hAAG bound to a target lesion (~20°) representing its IC/EC state⁷³.

Ethenoadenine lacks the ability to form a hydrogen bond with its complementary base (here: thymine) and the DNA at this lesion hence displays high flexibility when compared to intact, undamaged B-form DNA^{74,75} (inset in Fig. 3.11f). The DNA bend angle distributions from the AFM analyses revealed distinct differences between the ethenoadenine lesion and undamaged DNA, most strikingly the release of ~8° bending in the undamaged substrate to ~0° at the site of the lesion (~24% of conformers), similar as seen for the nicked substrate (Fig. 3.11a). An additional state with a broad width (~60°) at ~8° can further be observed from the Gaussian fits. The highly enhanced width of the 8° state and the shift to the 0° state indicated are consistent with high flexibility at the ethenoadenine lesion site compared to undamaged DNA (Fig. 3.11f). Bend angle of hAAG SC conformation (~0°) is consistent with ethenoA bend angle (~0°).

The automated AFM analyses also revealed some interesting aspects. G:U and oxoG:A lesions show additional 0° bent states in the absence of glycosylases, which may be due to reported destabilization at these lesions^{37,52,63,67} (Fig. 3.11b and d). This unbent species maybe be stabilized by the binding of respective glycosylases (hTDG and MutY) as seen from the DNA bend angle distributions (Fig. 3.10a and c).

DNA bend angle distribution of hOGG1-undamaged DNA complex revealed two additional states, 0° and ~35° apart from IC conformation (~70°) (Fig. 3.10b). The bend angle states of oxoG:C lesions (~10° and ~50°) are not observed in the bend angle distribution for hOGG1- undamaged DNA complexes (Fig. 3.11c). Since flexibility at an oxoG:C lesion is only slightly different from undamaged DNA, it can be easily bent or unbent by the hOGG1. These additional states (0° and ~35°) might suggest an interrogation mechanism to correctly identify such subtlety in flexibility at an oxoG:C lesion.

For hAAG bound to undamaged DNA, apart from the SC (0°) and IC (~20°) conformations, an additional bent state at ~45° was observed from the DNA bend angle distribution (Fig. 3.10d). This is a novel observation, and this strongly bent species was validated by ensemble FRET measurements and FRET simulations. It has been reported that the catalytic active site pocket of hAAG is not fully formed in the early stages of interrogation steps by the crystal structure and this is further confirmed by different electrostatic properties observed for the enzyme surface during initial 1D lesion search and subsequent IC conformation⁷³. hAAG might exploit the additional, stronger bending as an energetic test to correctly detect its highly flexible target lesion ethenoadenine.

3.3.3. Model of initial lesion detection by glycosylases

DNA glycosylases are known to bind DNA at non-specific sites initially and scan along the DNA until they find their target lesion³⁶. During the scanning process, they constantly transition between the search complex conformation (SC) and the interrogation complex conformation (IC)¹⁶. Accordingly, the AFM analyses provided IC conformations for four distinct glycosylases bound non-specifically to undamaged DNA that showed DNA bending consistent with the corresponding crystal structures (~70° for hTDG⁵¹, ~70-80° for hOGG1^{55,56}, ~50° for MutY⁶⁶ and ~20° for hAAG⁷³). Moreover, SC conformations for hTDG, hAAG and MutY showed similar bend angles as their target lesions in the absence of protein (~30° for hTDG and G:U, ~15° for MutY and oxoG:A/G:A, 0° for hAAG and ethenoA). These results suggest an initial lesion detection model for glycosylases based on the energetics of DNA bending and suitability of various glycosylases to their respective target lesion characteristics (Fig. 3.12).

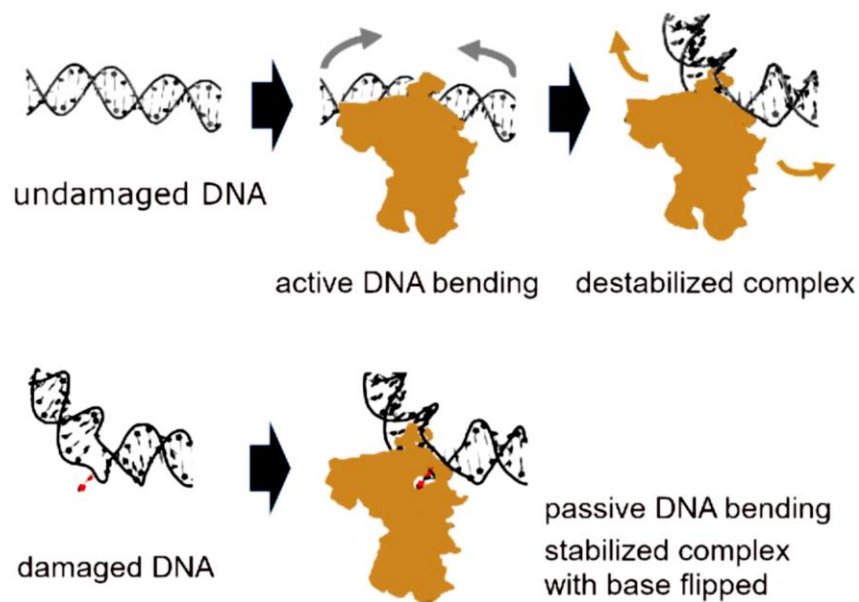


Figure 3.12: Schematic of the initial lesion sensing strategy employed by DNA glycosylases.

Introduction of a base lesion in the DNA double helix disrupts the base pairing and hence the DNA structure. For better understanding, the base lesion (in red color) is flipped out of the double helix due to increased frequency of base breathing. The MatLab based DNA bend angle distributions for glycosylase-undamaged DNA complexes and base lesions showed that bending in the search complex (SC) conformation matches the innate bending at corresponding lesions. This will result in passive lesion detection by glycosylases with binding to the lesion being energetically favorable compared to binding to undamaged DNA. In contrast, glycosylases actively (energy consuming) bend the undamaged DNA (grey arrows). The difference in the energy requirements to perform DNA bending at lesions and at non-specific sites results in the dissociation of the glycosylase (beige arrows) from the undamaged DNA whereas it leads to a strong binding at the site of lesion. Additionally, the enhanced pausing of the glycosylases at the target lesions coinciding with the lowered energy barrier for base breathing at these sites enhances the formation of the interrogation complex (IC) conformation characterized by base flipping into the catalytic active site pocket for base excision. Undamaged DNA is shown in grey color and damaged DNA in black color for easier visualization.

Overall, the results presented in this chapter lend support to the earlier proposed initial lesion detection model for DNA glycosylases by demonstrating similar bend angles for SC conformations of glycosylase-DNA complexes and their unbound target lesions¹⁶. The MatLab tool presented here can be applied to other protein-DNA systems with similar constraints.

4. Single molecule AFM studies reveal dual CpG site binding mechanisms of DNMT3A-3L DNA methyltransferase

4.1 Introduction

DNA methylation has important roles from genetic imprinting to gene regulation and chromatin structure modulation^{76–78}. The human genome consists of $\sim 10^7$ CpG dinucleotides, both in non-methylated as well as methylated states⁷⁹. DNA methylation is mainly catalyzed by a group of enzymes called DNA methyltransferases (DNMTs). They catalyze the methylation of cytosine bases by transferring the methyl group from S-adenosyl methionine (SAM) to the carbon on the fifth position of cytosine to form 5-Methylcytosine (5mC)⁸⁰ (Fig. 4.1). In humans, there are mainly three DNMTs in action: DNMT1, DNMT3A and DNMT3B⁸¹.

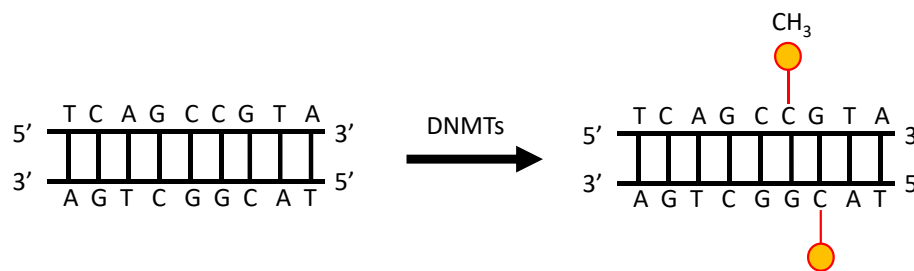


Figure 4.1: DNA methylation catalyzed by DNA methyltransferases. While DNMT3A and DNMT3B are involved in *de novo* methylation, DNMT1 is responsible for the maintenance of methylation patterns. The methyl group is transferred from S-adenosyl methionine to the sixth carbon on the cytosine base in the DNA by these methyltransferases.

The two main enzymes involved in *de novo* DNA methylation are DNMT3A and DNMT3B methyltransferases⁸². Both paralogs catalyze cytosine methylation on DNA preferably in CpG islands^{83,84} and set up various methylation patterns during gametogenesis and post-implantation development^{85,86}. On the other hand, DNMT1 is involved in the maintenance of methylation during DNA replication to inherit methylation patterns to the newly generated offspring DNA strand⁸⁰.

DNMTs catalyze the methyl transfer by initially attacking the C6-position of the target cytosine base by the active site cysteine residue leading to the formation of a transient enzyme-DNA covalent complex, which is resolved later during the methyl transfer⁸⁴. It has been hypothesized based on earlier kinetic studies that the energy for the base pair and base stack destruction for the embedded cytosine may arise from changes in free energy during interactions between DNMT molecules and DNA⁸⁷. DNMTs employ a strategy of base flipping like glycosylases⁸⁸. DNMT3A and DNMT3B differ by very few amino acids, however, both the enzymes consist of a C-terminal methyltransferase (MTase) domain, proline-tryptophan-tryptophan-proline (PWWP) regulatory domain and an ATRX-DNMT3-DNMT3L-type (ADD) zinc finger domain^{83,89,90} (Fig. 4.2a). DNMT3A plays a role in the development of mammals^{85,86} and most importantly in carcinogenesis^{91,92}. The DNMT3-

like protein (DNMT3L) is an inactive homologue of DNMT3A, which has been found to act as a co-factor and alleviate the activity of DNMT3A enzymes through direct interactions^{84,93}. DNMT3L in conjunction with DNMT3A is involved in setting up methylation patterns in the genome^{93–96}. The vacant position in place of the flipped out cytosine in the DNA double helix is stabilized by a residue in the catalytic loop of DNMT3A enzyme⁹⁰. DNMT3L interacts with the C-terminal portion of the catalytic loop in DNMT3A, and this might play a role in stabilizing its conformation and hence stimulating its catalytic activity⁹⁷ (Fig. 4.2b).

Structural studies with dsDNA generated by annealing a central CpG containing strand to a 2'- deoxy-Zebularine (Z) containing strand showed that DNMT3A-3L complexes flip the zebularine base out of the DNA double helix into the catalytic sites to perform methylation^{98,99}. Cytosine is replaced with zebularine to induce formation of a stable covalent complex, which could be crystallized easily because of homogeneity. The presence of natural cytosine residues instead of zebularines would result in transient complexes, which cannot be crystallized^{90,99,100}. In the crystal structure, a DNMT3A-3L heterotetramer is bound to a 25 bp long dsDNA containing two ZpG sites placed at a distance of 14 bp (this corresponds to 12 bp between two CpG sites) and flip both Z bases out of the helix into the two active sites of centrally located DNMT3A subunits^{90,99}. DNMT3L interacts with the catalytically active C-terminus of the DNMT3A MTase domain⁹⁷ forming a 3L-3A-3A-3L butterfly-like heterotetrameric structure with dimensions $\sim 160 \times 60 \times 50 \text{ \AA}^3$ (Fig. 4.2b). The first interface between the two-central adjacent DNMT3A subunits is symmetric and is characterized by the presence of two arginine and aspartate residues (RD interface) that are contributed by the two subunits. This interface is responsible for DNMT3A oligomerization and creates a DNA binding cleft^{97,101}. DNMT3L lacks the RD interface and hence blocks the co-operative binding for subsequent multimerization of DNMT3A subunits¹⁸. The second interface is between DNMT3A and DNMT3L, made by the four stacked phenylalanine residues termed the FF interface (hydrophobic in nature) (Fig. 4.2b). These phenylalanine residues from DNMT3L interact with the catalytic residues in DNMT3A, suggesting a source for the stimulatory effect of DNMT3L on DNMT3A. The FF interface has a high structural symmetry suggesting that such an interface can be formed by even two DNMT3A subunits. The enzyme complex interacts with the DNA using the RD interface, a disordered loop from the target recognition domain (TRD), and the loop containing the catalytic residues⁹⁰ (Fig. 4.2b).

Two DNMT3A molecules are placed on the DNA in such a manner that the two active sites can co-methylate two suitably spaced CpGs simultaneously. Due to structural limitations, both methylation events occur on the complementary DNA strands (Fig. 4.2b). The outermost DNMT3L subunits are easily replaceable with two additional DNMT3A subunits forming the smallest catalytically active homotetramer with the same arrangement as DNMT3A-3L complexes¹⁰². Analytical ultracentrifugation and size exclusion chromatography studies have demonstrated homotetramer formation by the DNMT3A catalytic domain as well as heterotetramer formation by the DNMT3A catalytic

domain and the DNMT3L C-terminal domain^{101–103}. Previously, atomic force microscopy imaging studies showed that both proteins multimerize on DNA at higher protein concentrations forming large protein-DNA fibers^{100,101,104,105}. Larger complexes containing only DNMT3A were able to interact with two separate DNA molecules. However, this was not true for DNMT3A-3L¹⁰⁴.

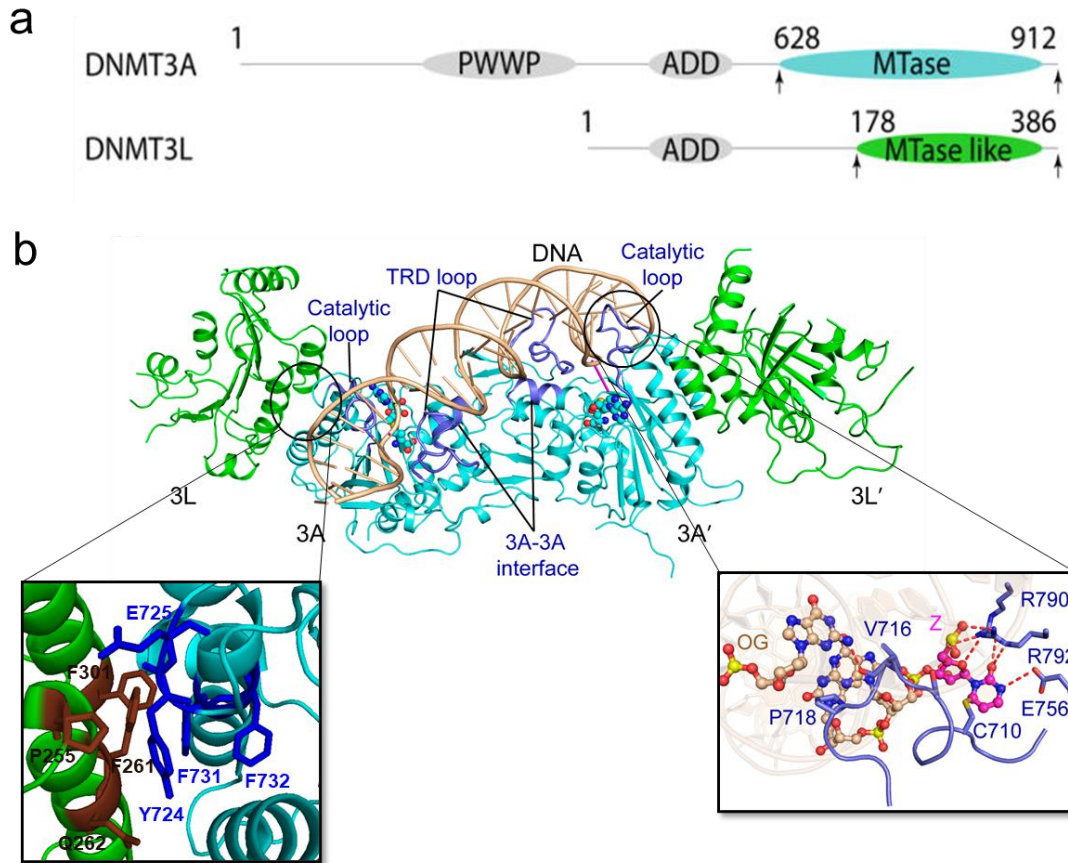


Figure 4.2: Structure of DNMT3A-3L heterotetramer with DNA containing two ZpG sites. (a) Domain architecture of DNMT3A and DNMT3L subunits. The small black arrows represent the C-terminal MTase and MTase-like domains in DNMT3A and DNMT3L, respectively. DNMT3A and DNMT3L contain a common ADD domain motif. DNMT3A contains an additional N-terminal PWWP motif. Figure from Zhang et al, 2018. **(b)** Butterfly-like structure of DNMT3A-3L heterotetramer bound to dual ZpG site containing 25 mer DNA. DNMT3L subunits (in green) interact with the central catalytically active DNMT3A subunits (in cyan). The flipped out zebularine residues are shown in purple. The inset shown on the left panel of the structure indicate the interactions between DNMT3L subunit and the catalytic loop of DNMT3A subunit. The four phenyl alanine residues forming the FF interface are also indicated. In the inset on the right panel show the interactions between the catalytic loop (violet) of DNMT3A subunit and DNA (beige). The hydrogen bonds are indicated as dashed lines. Two CpGs in the DNA must be suitably spaced to fit the distance between the active site centers of two DNMT3A subunits. Interactions between DNA and DNMT3A take place via the RD interface, a disordered loop from the target recognition domain (TRD), and the loop containing the catalytic residues. Figure modified with permission from Ren et al, 2018.

Apo⁹⁷ and DNA bound DNMT3A-3L structures^{90,99} do not exhibit major conformational alterations in the proteins. However, DNMT3A introduced ~40° bending in the 25 mer

DNA with the two ZpG sites bound to the two active sites of the DNMT3A subunits⁹⁰. In contrast, the shorter 10mer dsDNA containing two ZpG sites bound to a single DNMT3A-3L heterotetramer was unbound⁹⁰. The distance between the two ZpG sites do not match with the distance between the two active sites in DNMT3A subunits⁹⁰. Hence, as observed from the crystal structures⁹⁰ and the kinetic studies¹⁰⁶, a spacing of 12 bp between the two ZpG sites is strongly favored by the DNMT3A-3L heterotetramer and forms the basis for this study.

Although, kinetic and structural studies provided useful insight into the mode of interactions between DNMT3A-3L complexes and its target sites, these involved artificial covalent trapping of the enzyme-DNA complex, which may provide biased results and shift the conformational equilibria of these complexes. Meanwhile, it was also observed that the CpG site neighboring sequences had a strong effect on the activity of DNMTs^{89,107–109}. It is well known that the sequence composition of the DNA has a strong effect on DNA curvature^{110–112} (see also chapter 3). Subsequently, sequence preferences could influence DNA bending necessitated for DNMT3A- dual CpG site interactions.

To investigate the different co-methylation preferences shown by DNMT3A-3L heterotetramers, I have applied single molecule AFM experiments using the purified DNMT3A-3L heterotetramer, containing the catalytic domain of DNMT3A and C-terminal domain of DNMT3L, and ~400 bp long DNA substrates containing pairs of CpG sites in the center (50% position) with variable spacing between them. Initially, next generation sequencing (NGS) coupled Deep-Enzymology was applied to reveal preferences for different CpG site spacings in DNMT3A and DNMT3A-3L mediated co-methylation. In addition, the effects of flanking sequences on the co-methylation efficiency were tested using kinetic studies. Biochemical and kinetic studies revealed different modes of co-methylation performed by DNMT3A and DNMT3A-3L complexes at the CpG sites and preferred spacings of CpG sites by 6, 9, and 12 bp. These three substrates were used for single molecule AFM characterization of co-methylation modes. With the high-resolution capacity of AFM, it is possible to distinguish between DNMT3A-3L-DNA complexes at target sites, here dual CpG sites placed at 50% of total DNA length (~400 bp) with 6, 9, and 12 bp spacings and non-specifically bound complexes (anywhere except 50% position). AFM thus enables the resolution of different conformational states of the protein-DNA complexes at target sites and non-specific sites at the single molecule level^{2,17,113}. For all AFM analyses of DNMT3A-3L-DNA complexes, the automated, high-throughput, and user independent MatLab based tool introduced in chapter 3 of this thesis was used. This tool has been further expanded for these studies to automatically detect the protein binding positions on the DNA along with the corresponding DNA bend angles in the protein-DNA complexes. In addition to positional and bending analyses, volumes of the DNMT3A-3L complexes bound to the three different substrates (with 6, 9, and 12 bp spacings between CpG sites) were also measured. Surprisingly, DNMT3A-3L complexes bound to the three different DNA substrates revealed significant differences. The DNA bending obtained for DNMT3A-3L bound to CpG sites placed 12 bp apart agrees well with

the reported value from the crystal structure (with the same spacing between the CpG sites)⁹⁰. DNMT3A-3L complexes at dual CpG sites in all three substrates indicate almost equal populations of two types of oligomers with volumes, $\sim 100 \text{ nm}^3$ and $\sim 200 \text{ nm}^3$. However, volume analyses for the 6 bp and 9 bp substrates indicate $\sim 200 \text{ nm}^3$ species predominantly. Subsequent analyses of the heights and lengths of protein complexes on both substrates revealed significant differences between them. In the case of 6bp substrate, two DNMT3A-3L heterotetramers were arranged in a side-by-side fashion and in the case of 9 bp substrate, two heterotetramers were arranged more compactly and indicated towards tetramer swap on the DNA. Moreover, structural modelling of DNMT3A-3L complexes with DNA supported data from AFM and provided an explanation for the different conformational properties of the complexes of two heterotetramers bound to CpGs with spacings of 6 bp or 9 bp. Altogether, these results elucidate how DNMT3A and DNMT3A-3L complexes can perform co-methylation *in vivo* and set up different methylation patterns without the limitation of 12 bp distance between dual CpG sites.

This work has been published in *Nucleic Acids Research* in 2021 [Emperle and Bangalore *et al*, Structural and biochemical insight into the mechanism of dual CpG site binding and methylation by the DNMT3A DNA methyltransferase, *Nucleic Acids Research*, 2021, doi: <https://doi.org/10.1093/nar/qkab600>] and was a collaborative effort combining AFM analyses and kinetic co-methylation data. Kinetic studies and structural modeling were contributed by AG Jeltsch, IBTB Stuttgart.

4.2 Results

4.2.1 Co-methylation studies of CpG sites at varying distances

To analyze the activity of DNMT3A and DNMT3A-3L complexes on different DNA substrates, a NGS coupled Deep-Enzymology approach was applied^{89,114}. The results were helpful in appreciating the collaborative scheme of the two DNMT3A subunits of DNMT3A-3L heterotetramers in methylating the substrate. Partially randomized sequence containing DNA substrates were used to explore the impact of distance between the CpG sites (2-15 bp) (Appendix, table. x4.1) and the neighboring sequences on the binding and mode of methylation by DNMT3A-3L complexes^{89,108,109,114}. The two CpG sites were embedded in the central 23 bp stretch with common neutral randomized sequence (Appendix, table. x4.1) and the mixed pool of dsDNA substrates was subjected to methylation by DNMT3A and DNMT3A-3L assemblies. Hairpin-bisulfite conversion coupled to NGS was used to assess methylation on both DNA strands. Percent methylation increased from 1% to 11.6% by adding increasing concentration of DNMT3L to DNMT3A as was expected¹¹⁵ (Fig. 4.3a). Co-methylation levels were ~0% in the absence of enzymes (Fig. 4.3a).

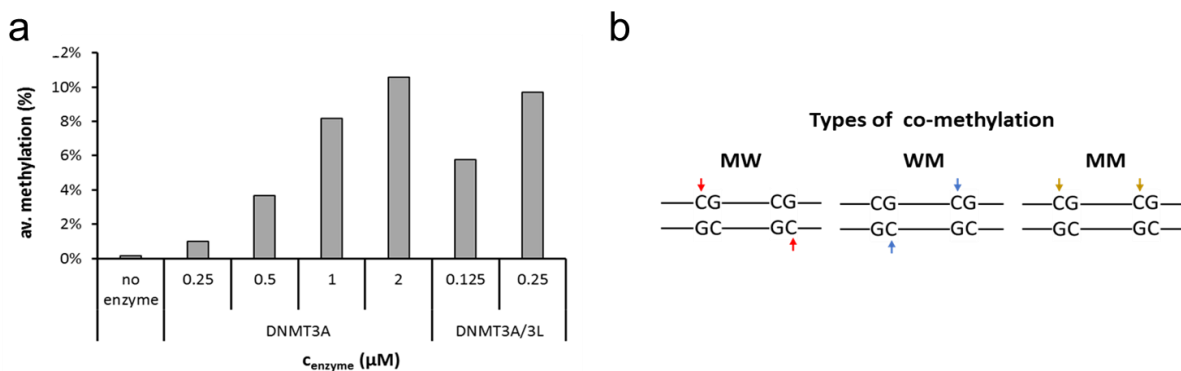


Figure 4.3: Co-methylation of dual CpG sites spaced at different distances by DNMT3A and DNMT3A-3L complexes. (a) Overall average methylation levels were assessed for different enzyme concentrations along with the no-enzyme control. Here the average methylation levels for all four cytosines in the two CpG sites is averaged over all distances. **(b)** Different modes of co-methylation observed in this study. Colored arrows indicate the site of methylation on each substrate. Methylation in the upper DNA strand is indicated by 'M', methylation in the lower DNA strand by 'W'. Note that mechanistically MM and WW methylation are essentially the same. Here all events are considered as MM co-methylation.

In total, three variations of co-methylation were observed based on the order of strand methylation: MW co-methylation, WM co-methylation and MM or WW co-methylation. Upper strand methylation is denoted as 'M', lower strand methylation is denoted as 'W'. A methylation event in the upper strand followed by a second methylation event on the lower strand is denoted by 'MW'. Similarly, first methylation event on the lower strand followed by second methylation event on the upper strand is denoted by 'WM'. Two methylation events occurring either on the upper strand or on the lower strand are denoted

by 'MM' or 'WW' respectively (Fig. 4.3b). MM and WW co-methylation percentages were combined (presented here as MM co-methylation data) since both are symmetrically related and essentially mean two methylation events on the same DNA strand. However, WM and MW modes of co-methylation are not the same.

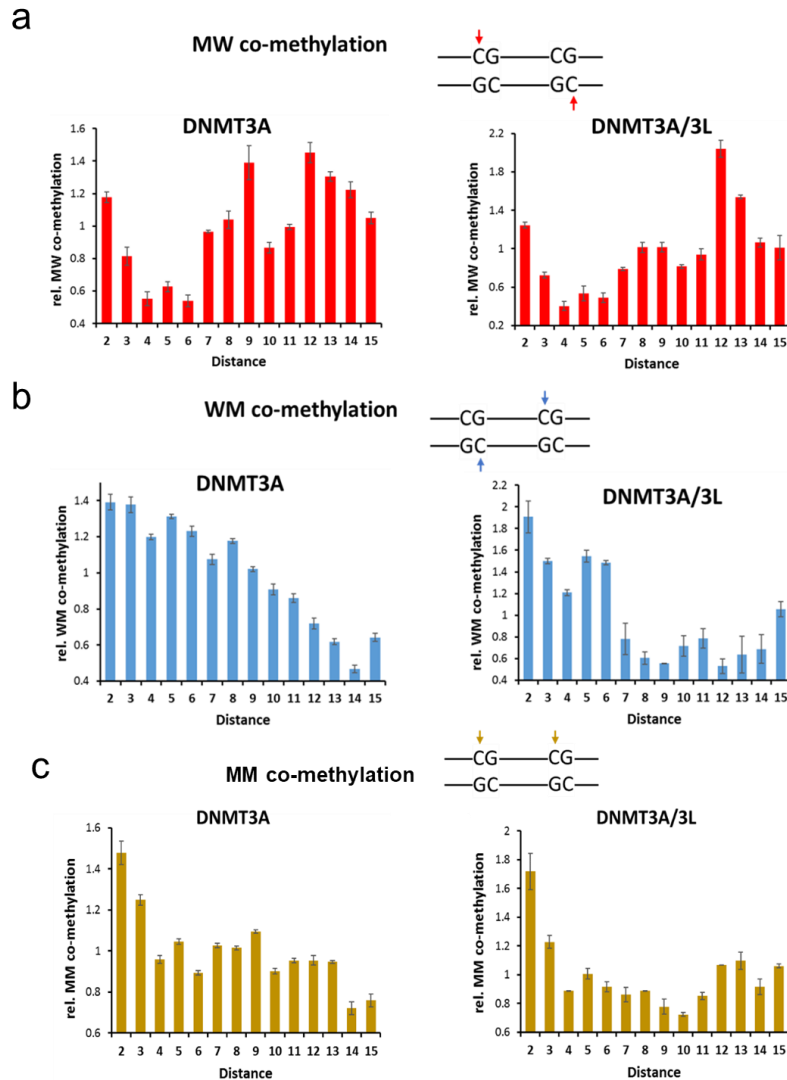


Figure 4.4: Co-methylation studies with DNMT3A and DNMT3A-3L complexes at CpG sites placed at variable distances in three different modes of co-methylation. Co-methylation levels of CpG sites placed at variable distances for DNMT3A (left panels) and for DNMT3A-3L complexes (right panels) for **(a)** MW mode, **(b)** WM mode, and **(c)** MM mode. Indicated values represent the averages from two separate datasets with error bars obtained after normalization.

After obtaining co-methylation profiles for different enzyme concentrations, the average co-methylation percentages were normalized to assess relative levels of co-methylation with error ranges (\pm SD) from multiple experiments. It was observed that DNMT3A-3L could co-methylate CpG sites with 12 bp spacing in the MW mode, but it could not do so

in WM or MM fashion^{90,106} (Fig. 4.4). Methylation levels were high for CpG sites at 2 bp distance for all modes of co-methylation (Fig. 4.4). Co-methylation percentages were high for 9 bp and 12 bp distances between CpG sites for MW co-methylation mode by DNMT3A (Fig. 4.4a). For DNMT3A-3L, the peak at 12 bp was further enhanced and the peaks at 2 bp and 9 bp were diminished (Fig. 4.4a). On the other hand, for WM co-methylation, a peak was observed for 5-6 bp in the case of DNMT3A-3L in addition to the 2 bp peak for both DNMT3A and DNMT3A-3L (Fig. 4.4b). For MM co-methylation, DNMT3A or DNMT3A-3L failed to exhibit any peaks at other distances apart from 2 bp spacing (Fig. 4.4c).

4.2.2 Sequence preferences by DNMT3A and DNMT3A-3L complexes

Modulation of activity of DNMT enzymes has been shown to be affected by the CpG site flanking sequences^{116,117}. Enhanced MW co-methylation by DNMT3A and DNMT3A-3L was seen at 12 bp spacing (Fig. 4.4a). To understand the preferential methylation in MW mode by the enzymes at these different spacings, the impact of immediate flanking sequences between two CpG sites placed 12 bp apart on the efficiency of co-methylation was assessed. From the crystal structure, it is clear that DNA bending is essential for co-methylation of dual CpG sites placed 12 bp apart and bending/curvature is dependent on the nucleotide composition of the DNA strands between the two CpG sites¹¹⁸. To investigate the sequence effects on the degree of co-methylation, the sequence in the center (position 5-8 placed between two CpG sites) of the DNA substrates was varied (Fig. 4.5a). Sequence preferences persisted for the DNMT3A and DNMT3A-3L complexes with high efficiency of co-methylation seen for thymine enriched upper DNA strand highlighting importance of the presence of highly flexible A-tracts between the two CpG sites for easier bending by protein complexes (Fig. 4.5b). Additionally, efficiently co-methylated substrates displayed a depletion of guanine bases in the upper strand between CpG sites (Fig. 4.5b). GC rich sequences are known to have higher melting temperatures and binding energies than A tracts or AT rich sequences¹¹⁹. The underlying reason for this behaviour is the presence of three hydrogen bonds between guanine and cytosine as opposed to two hydrogen bonds between adenine and thymine. Additionally, alternating AT-rich sequences between GC-rich sequences are known to help in the opening of the double helix and hence the attachment of DNA-binding proteins¹²⁰. It has also been demonstrated that G-tracts with regularly spaced A tracts display enhanced bending¹²¹. Additionally, the degree of co-methylation by DNMT3A or DNMT3A-3L is very high at short distances (2-3 bp) and keeps decreasing with increasing distance. This is explained by short distances readily accessible by these enzyme molecules on the DNA without performing major movements.

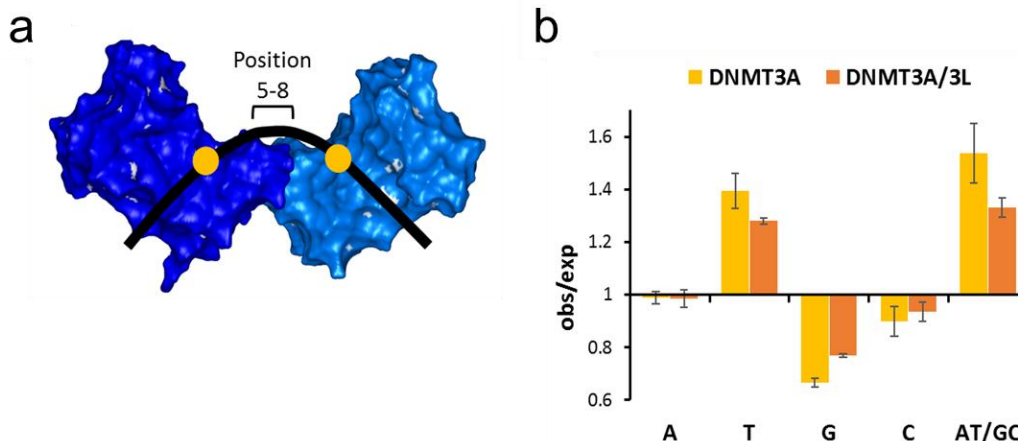


Figure 4.5: Sequence composition of the central region between dual CpG sites placed 12 bp apart influences MW co-methylation. (a) Schematic depicting the positions 5-8 considered for sequence estimation at the center of bending. (b) Enrichment and depletion averages of bases in co-methylated substrates by DNMT3A and DNMT3A-3L. Standard deviations are shown by the error bars.

4.2.3 Single molecule AFM analysis to characterize modes of binding of DNMT3A-3L to CpG sites placed at variable distances

Based on the results from co-methylation studies, DNMT3A-3L interactions with DNA substrates containing two CpG sites placed at different distances were characterized using AFM analyses. Since DNMT3A shows co-operative binding on DNA, these were eliminated from AFM studies and only DNMT3A-3L heterotetrameric complexes were used. DNA substrates of ~400 bp length containing a pair of CpG sites in the center (at 50% of total DNA length) with spacing of 6, 9, or 12 bp were designed for AFM studies (Appendix, Fig. x4.1a). The variable region between the two CpG sites in the substrates used in this study had a well-balanced mix of AT and GC sequences to nullify any effects of the sequence (Appendix, Fig. x4.1b). The substrates were designed to contain CpG free regions surrounding the central dual CpG sites. However, they contained single CpG sites at ~25-30% of total length. Single molecule AFM imaging is very well-suited to differentiate the DNMT3A-3L complexes specifically bound to the double CpG site at the 50% position from complexes bound at a single CpG site (at ~25-30%) and non-specific DNA (other positions). The substrates in these studies are denoted as D6, D9, and D12 for 6, 9, and 12 bp distances, respectively. D6 was designed based on the preferential WM co-methylation observed for DNMT3A-3L complexes from the kinetic studies (Fig. 4.4b). D9 was chosen based on preferential MW co-methylation seen for DNMT3A and slightly weaker but significant for DNMT3A-3L complexes. The substrate with 12 bp distance between dual CpG sites was designed based on the previous structural data⁹⁰ and the observed preferential MW co-methylation by DNMT3A and DNMT3A-3L complexes. Low protein concentrations were applied in these studies to avoid the formation of protein-DNA fibers as previously observed^{101,102,105}. In the AFM data

analyses, the automated MatLab tool for DNA bend angle analyses was employed that has been developed for the glycosylase studies (see chapter 3). In addition to measuring DNA bend angles and DNA lengths, the tool has further been expanded for the studies on DNMT3A-3L to measure the positions of bound protein complexes on the DNA. The expanded tool provides the corresponding bend angles for each binding position together with the corresponding DNA total length to exclude DNA substrates with incorrect lengths, for which the 50% position would not accurately reflect the position of the dual CpG sites, from further analyses. The output image from MatLab thus contains all DNA strands in the AFM images and bound protein peaks numbered and marked, which then allows to measure the volumes, lengths, and heights of the individual complexes (correlated with their positions and DNA bend angles) using other softwares. Thus, DNMT3A-3L complexes on DNA containing dual CpG sites at variable distances were analyzed for positions, DNA bend angles, volumes, heights, and lengths. DNA length distributions for all three substrates D6, D9 and D12 obtained from MatLab analyses showed lengths consistent with the theoretical lengths of ~400 bp DNA (~139 nm based on 0.34 nm/bp) and are presented in the appendix section (Fig. x4.3). The free unbound DNA lengths were >122 nm for all substrates. Slight shortening in the contour lengths of DNA may be due to DNA undulations (see chapter 3). The DNA lengths of all three substrates in the presence of proteins agreed with the lengths obtained in the absence of the protein complexes confirming that the proteins binding to the DNA does not lead to their shortening (Appendix, Fig. x4.3).

D9 and D12 exhibited strong binding preferences by DNMT3A-3L complexes at the 50% position suggesting high specificities and affinities to the dual CpG sites at 9 bp and 12 bp distances (Fig. 4.6). The preference of binding at a site was defined using the observed number of binding events (N_{obs}) obtained from the heights of the bins in the position distributions and the expected number of events (N_{exp}) obtained from each binning interval when equal probability of binding at all positions on the DNA (no specificity) is assumed. The ratio of observed to expected numbers (obs/exp) of binding events at the 45- 50% positions of the D12 substrate was 8.6, with 43% of total binding events located at these dual CpG sites. For the D9 substrate, the peak at 50% DNA length was broader compared to the D12 substrate and a lower preference of binding was seen by these complexes at the central CpG sites (obs/exp = 3.5 at 50% position). However, if binding events from 40-50% were considered to be centrally bound complexes due to the broad position distribution, the area under the curve still accounted for 52% of all binding events. In addition, a second binding preference at 25-30% DNA length was noticeable in the case of D9 substrates, which corresponds to binding at the transition of CpG free flanking regions to average DNA (Appendix section, Fig. x4.1). Importantly, this position is well resolved from the specific sites at 50% DNA length in the substrates (Appendix, Fig. x4.1b). For the D6 substrate, the peak at 25-30% DNA lengths was even of similar size as the peak at 50% position (two CpG sites with 6 bp spacing) highlighting similar affinities towards the dual CpG sites (obs/exp = 3.1 at 50% DNA length, 16% of total binding

events) and the single CpG sites (25-30% DNA length). Due to the higher affinities for D9 and D12 at 50% position, this peak is comparably dwarfed for these substrates (Fig. 4.6).

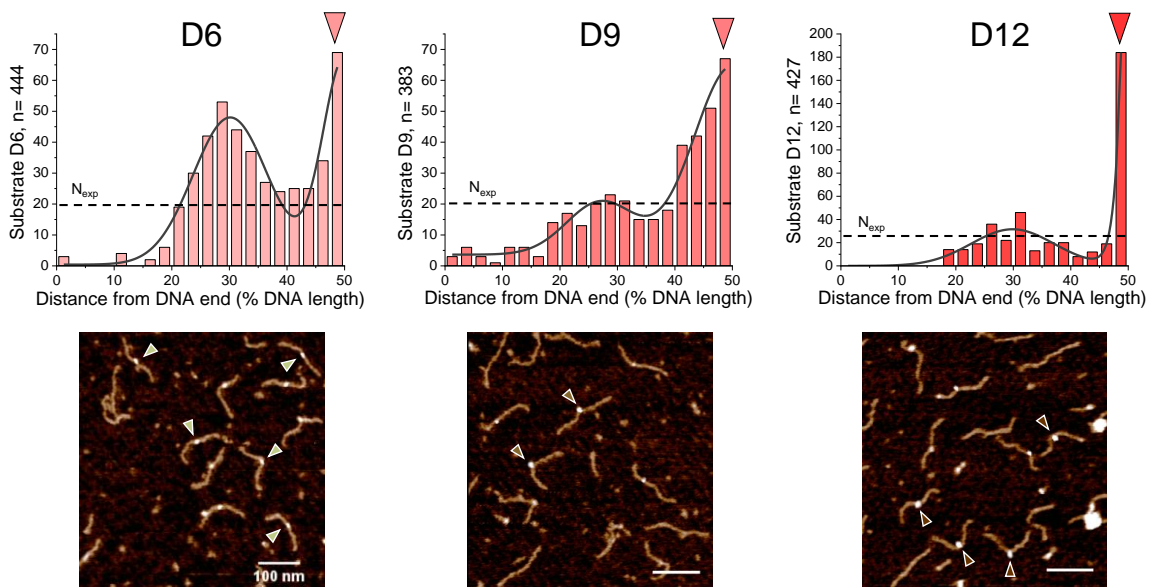


Figure 4.6: DNMT3A-3L binding specificities at the dual CpG sites in the center separated by 6, 9, and 12 bp obtained from single-molecule AFM studies. The top panels indicate position distributions of protein complexes bound to DNA substrates containing two CpGs with different spacings at 50% of DNA length: 6 bp (D6), 9 bp (D9), 12 bp (D12). All three substrates demonstrated an additional peak apart from 50% peak; hence positions were fit with two Gaussians forming a bi-modal distribution. N_{exp} is indicated by horizontal dashed lines and represents the expected number of binding events occurring in each bin in the distribution assuming equal probability of binding at all positions along the DNA length. Bottom panels show exemplary AFM images indicating complexes (indicated by triangles) bound at the center of D6, D9, and D12 substrates. Scale bar: 100 nm.

Next, to determine the number of individual heterotetramers contained in the individual complexes at 50% positions of all three substrates, protein volumes were measured from the AFM images (Fig. 4.6). The volumes at dual CpG sites (50% DNA length) with a spacing of 6 and 9 bp indicated that two DNMT3A-3L heterotetramers (~ 200 kDa) were bound at this site (volumes corresponding to ~ 200 nm³), while two Gaussian maxima were observed in the bimodal distribution for the volumes of DNMT3A-3L complexes bound at 50% DNA length in D12 substrate (volumes corresponding to ~ 200 nm³ as two heterotetramers and ~ 100 nm³ as a single heterotetramer). As controls, the volumes of the complexes bound between 0-40% DNA length corresponding to non-specific DNA were also determined (Fig. 4.7a). Volumes on average DNA indicated that predominantly single heterotetramers of DNMT3A-3L are bound to these regions since average DNA still

contains random single CpG sites and non-specific sites. As a further control, volumes of unbound DNMT3A-3L complexes were measured (see appendix, text x4.1).

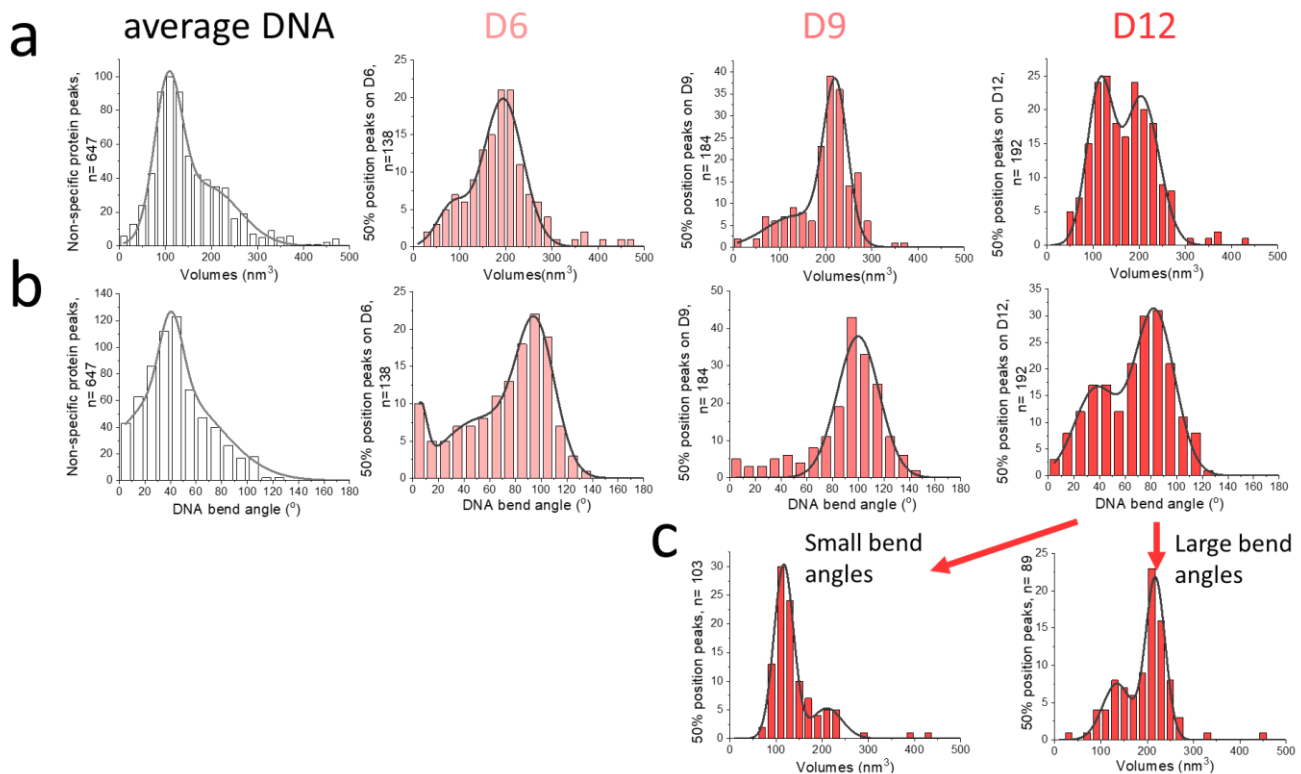


Figure 4.7: AFM analyses to characterize the DNMT3A-3L complexes bound to dual CpG containing DNA substrates placed at different distances. AFM distributions showing (a) volumes and (b) DNA bending angles of DNMT3A-3L complexes. Volumes and bending angles were measured for complexes at 50% positions on substrates containing dual CpG sites separated by 6 bp (D6), 9 bp (D9), and 12 bp (D12). In addition, non-specific or average DNA has been included as a control (left panel). Binding positions between DNA ends and 40% DNA length were considered as average DNA. (c) Separation of volumes for mildly bent species (~40°, left) and strongly bent species (~80°, right) for the D12 substrate.

Next, I analyzed the DNA bend angles introduced by the DNMT3A-3L complexes bound at the central CpG sites (50% DNA length) and at the average regions of the D6, D9 and D12 substrates (Fig. 4.7b), correlated with the obtained binding preferences and protein volumes. From the bend angle distributions of the D12 substrate, 40° bending and an additional 80° bending were observed at the dual CpG sites. Separation of these two different bend angle species showed volumes consistent with a single DNMT3A-3L heterotetramer for the 40° bent species (~100 nm³) as seen in the crystal structure⁹⁰, while the stronger bending arises from double DNMT3A-3L heterotetramers (Fig. 4.7c). Presumably, each heterotetramer in these complexes bends the DNA by 40° to give the total 80° bending. Control measurements on DNMT3A-3L complexes on average DNA showed a broad DNA bend angle distribution with a Gaussian maximum also at ~40° (Fig. 4.7b). Interestingly, DNA bend angle distributions of DNMT3A-3L complexes bound at the central dual CpG site of D6 and D9 substrates demonstrated much stronger bending in

these complexes by $\sim 100^\circ$ (Fig. 4.7b). As further controls, measurements of DNA bend angles at the centers (50% positions) of D6, D9 and D12 substrates in the absence of protein revealed a straight DNA conformation (Gaussian fits centered at 0°) at these positions (Appendix, Fig. x4.2b). Therefore, AFM data shows that the DNA is actively bent by the bound DNMT3A-3L complexes.

For the larger complexes ($\sim 200 \text{ nm}^3$) corresponding to two DNMT3A-3L heterotetramers bound at the central dual CpG sites of D6 and D9 substrates, which showed stronger DNA bending by $\sim 100^\circ$, the lengths and heights of complexes on the DNA were measured. D6 and D9 substrates revealed heights of centrally bound DNMT3A-3L complexes of $\sim 0.5 \text{ nm}$ and $\sim 1.1 \text{ nm}$, respectively, and lengths of $\sim 20 \text{ nm}$ and $\sim 10 \text{ nm}$, respectively (Fig. 4.8b). These values convey that double heterotetramers bound at the central dual CpG sites with 6 bp spacing have a vertically flat and horizontally elongated shape along the DNA while the complexes on D9 are more tightly arranged with enhanced heights (Fig. 4.8a). As controls, the heights, and lengths of the smaller species of single heterotetramers (volume of $\sim 100 \text{ nm}^3$) bound at the 50% position of D6 and D9 substrates were also determined (Appendix, Fig. x4.4). All these complexes showed lengths of $\sim 7 \text{ nm}$, shorter than the length of two heterotetramers together for the D9 substrate. Their heights were $\sim 0.5 \text{ nm}$, comparable to the height of the flat, double heterotetrameric conformers on the D6 substrate, supporting a side-by-side arrangement of these complexes on this substrate (see also section 4.2.4 below).

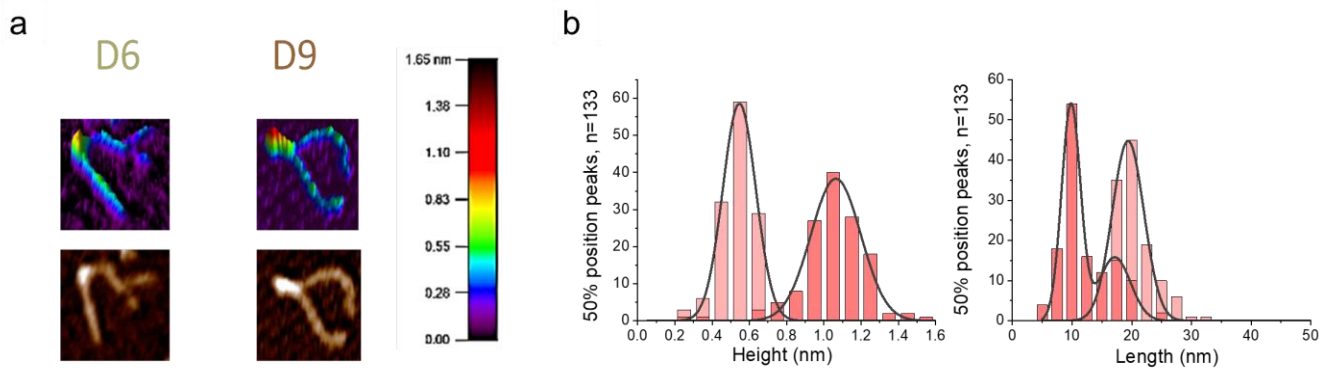


Figure 4.8: AFM based characterization of double heterotetrameric DNMT3A-3L complexes with CpG containing DNA substrates separated by 6 bp and 9 bp. (a) Exemplary AFM images of double heterotetrameric DNMT3A-3L complexes bound to dual CpG sites in D6 and D9 substrates. The volumes of the complexes shown here are 206 nm^3 (D6) and 217 nm^3 (D9), consistent with the size of a double heterotetramer. The corresponding heights of these complexes are 0.69 nm and 1.18 nm and their lengths on the DNA are 18.9 nm and 9.4 nm (for D6 and D9, respectively). Images are shown in 3D (top) and 2D top view (bottom) for better visualization of the unique differences. Areas of these images are $\sim 130 \times 130 \text{ nm}^2$. The scale bar at the right indicates the heights in the 3D images in units of nm. **(b)** Average heights (left) and lengths (right) of double heterotetrameric DNMT3A-3L complexes bound at the central dual CpG sites separated by 6 bp (D6 substrate, beige) and 9 bp (D9 substrate, pale brown). The Gaussian fit centers

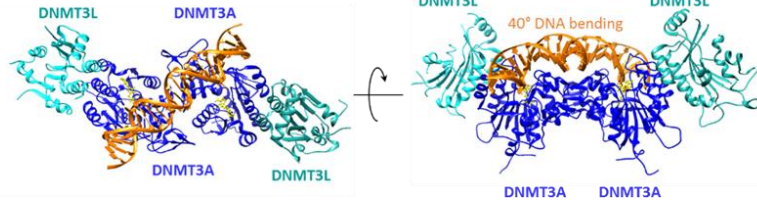
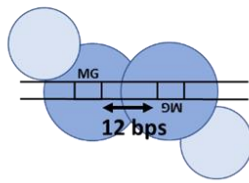
indicate average heights of complexes on CpG pairs of ~0.6 nm for D6 and ~1.1 nm for D9 and average lengths of complexes of ~20 nm for D6 and ~10 nm for D9.

4.2.4 Modelling DNMT3A-3L complexes to understand the different modes of co-methylation

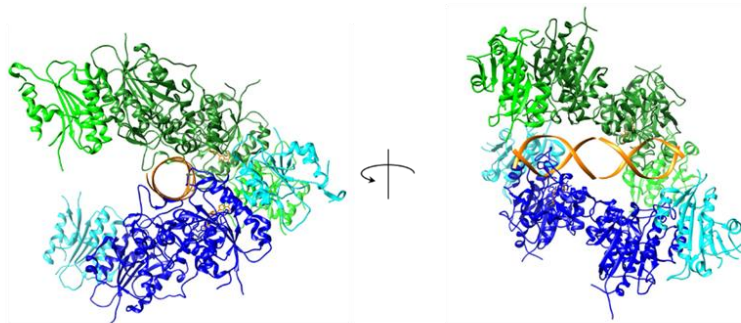
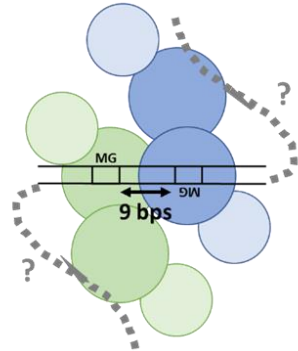
From the co-methylation assays and AFM analyses, it was observed that MW co-methylation was preferred at a distance of 12 bp and a single or double DNMT3A-3L heterotetramers were predominantly bound at the dual CpG site in D12 substrate. These results are consistent with the co-crystal structure of a single DNMT3A-3L heterotetramer bound to dual zebularine (Z) containing DNA⁹⁰ together with additional biochemical data¹²². However, to explain co-methylation seen at other distances apart from 12 bp, the DNMT3A-3L would have to undergo significant conformational transitions. Previous crystal structures^{90,97,123} of DNMT3A-3L indicate low conformational flexibility and it seems far-fetched for a single tetramer to prefer co-methylation of CpG sites at a distance of 9 bp as observed for DNMT3A as well as DNMT3A-3L here. Based on the results of MW and WM co-methylation preferences at different distances for dual CpG sites and structural and stoichiometric inferences from the AFM data, different orientations of two heterotetrameric DNMT3A-3L were modelled for the substrates containing two CpG sites at distances of 6, 9, and 12 bp (Fig. 4.9). Modelling results were obtained by DNMT3A-3L structures bound to a long DNA with two CpG sites placed at 12 bp distance (PDB id: 6BRR)⁹⁰. Single DNMT3A-3L heterotetramer bound to dual CpG sites placed 12 bp apart is explained by the crystal structure^{90,99} (Fig. 4.9a). To visualize two heterotetramers at the dual CpG, two DNMT3A-3L bound DNA duplexes (from PDB id: 6BRR) were connected without any steric clashes, which led to a side-by-side arrangement of two heterotetramers (Fig. 4.11).

For the WM co-methylation on the 6 bp substrate, two adjacent DNMT3A tetramers each bound to one of the two CpG sites is a plausible explanation. Modelling results indicated that both complexes could approach each other up to 7 bp as the possible distance between CpG sites without major steric clashes. Likely, spacing between CpG sites can be a little shorter with additional conformational modulations in the complexes. The co-methylation studies shown above also support this short distance between the two CpG sites. From former data, the interface between the two tetramers is formed by the loop region (R831-K855) which was demonstrated to be necessary for co-operative binding of DNMT3A on DNA^{101,102,105}. In this model, the formation of double heterotetramers results in long elongated complexes in an extended conformation along the DNA (Fig. 4.9c). This model is completely corroborated by the AFM results showing large volumes with long lengths and low heights of these complexes together with strong DNA bending.

a Tetramer structure



b Model: Tetramer swap



c Model: Side-by-side

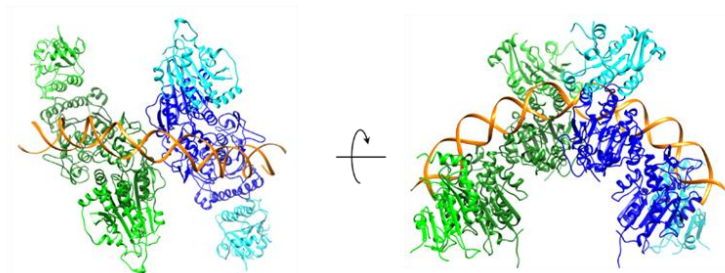
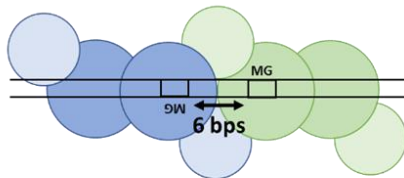


Figure 4.9: Models describing MW and WM mode of co-methylation at distances of 6, 9, and 12 bps. **(a)** Model for MW co-methylation at dual CpG sites with 12 bp spacing. The two DNMT3A subunits at the center in the heterotetramer are colored blue, DNMT3L is colored cyan, DNA orange and AdoHcy (S-adenosyl L-homocysteine) is represented in yellow. MW mode of co-methylation at 12 bp distance agrees with the crystal structure of DNMT3A-3L (PDB id: 6BRR). **(b)** Model for MW co-methylation at dual CpG sites with 9 bp spacing. MW mode of co-methylation can occur at shorter distance compared to 12 bp if two adjacent tetramers interact with the DNA via only one DNMT3A subunit leading to a tetramer-swap of the DNA. The model was formulated by using two structures of DNMT3A-3L complexes with two short DNA substrates each bound to one DNMT3A subunit (PDB id: 6F57). The DNA molecules of both tetramers were joined creating a distance with of 9 bp spacing between the CpG sites. **(c)** Model for WM co-methylation at dual CpG sites with 6 bp spacing. WM mode of co-methylation can be explained by two side-by-side DNMT3A tetramers interacting with each other on the DNA. The model was formulated by approaching two DNMT3A-3L structures with long DNA substrates bound to both DNMT3A subunits (PDB id: 6BRR) as close as to a spacing of 7 bp between the two CpG sites. With additional conformational adjustments in the protein complexes, this model can be extended to 6 bp. The zoomed view of the model is shown in appendix figure x4.5.

For the observed MW co-methylation for the 9 bp substrate, the DNMT3A-3L structure with two short 10 mer DNA substrates with each containing a CpG site bound to one DNMT3A subunit in a heterotetrameric complex was considered (PDB id: 6F57)⁹⁰. Both DNA molecules can be merged generating one single DNA strand with two CpG sites each interacting with one DNMT3A subunit coming from the two adjacent DNMT3A-3L tetramers. The two additional DNMT3A subunits in each heterotetramer might interact with the flanking DNA sequence leading to a tetramer swap on the DNA at the RD interfaces (Fig. 4.9b). From the modelling results, it was observed that two tetramers could approach each other without any steric clashes up to ~8-9 bp forming highly condensed complexes. These results are consistent with and strongly supported by the obtained AFM results showing high and dense complexes with volumes corresponding to two tetramers for 9 bp spacing of CpG sites. While strong bending is observed from the AFM analyses for DNMT3A-3L complexes on the D9 substrate, the model does not show bending. However, there are two additional DNMT3A subunits in each tetramer that are unbound to the central CpG sites. Hypothetically, these unbound DNMT3A subunits could bind the flanking DNA sequences resulting in strong DNA bending and additional condensation of the structure. When one of the additional free DNMT3A subunit grabs the flanking DNA sequence, the centroid of the whole DNMT3A-3L complex is shifted towards the nearer end of the DNA (towards the side from which the flanking sequence is grabbed). This is consistent with the broad position distribution observed for the D9 substrate (Fig. 4.6). The strong DNA bending of 100° also suggests that only one arm of the flanking DNA is grabbed (Fig. 4.7b). Ideally, when both the sequences are grabbed on the opposite sides, the DNA bend angles would get nullified and width of the peak at 50% DNA length in the position distribution would be narrower. However, this strongly suggests that only one flanking DNA sequence is grabbed by the entire complex.

4.3 Discussion

The human genome consists of randomly placed dual CpG sites with varying distances and isolated CpG sites. *De novo* methylation at these sites is performed by heterotetrameric DNMT3A-3L complexes or by DNMT3A homotetramers. A homotetramer of DNMT3A or a heterotetramer of DNMT3A and DNMT3L form the basic smallest catalytically active unit consisting of the two DNMT3A subunits centrally present, which have been shown to methylate CpG sites preferentially on opposite strands at a distance of 12 bp⁹⁰. DNMT3L stimulates DNMT3A activity by direct interactions. The aim here was to understand how this enzyme complex can comparably methylate natural DNA with single CpG sites and randomly spaced dual CpG sites without the preference for 12 bp distance. Towards this aim, co-methylation preferences by DNMT3A and DNMT3A-3L of CpG sites placed at varying distances from each other were studied. AFM studies were then performed to understand different modes of co-methylation by DNMT3A-3L complexes and their conformational rearrangements necessary for adjusting to different dual CpG site spacings.

As noted from the NGS coupled Deep-Enzymology studies, dual CpG sites are co-methylated in three different modes, either co-methylation of two CpG sites on the same DNA strand top or bottom (MM or WW co-methylation) or on opposite DNA strands (MW and WM co-methylation). In MW mode of co-methylation, site 1 is targeted by the enzyme on in the upper strand and site 2 in the lower strand, while in WM co-methylation, site 1 in the lower and site 2 in the upper strands are targeted (Fig. 4.3b). MW and WM modes of co-methylation are asymmetrical and therefore clearly require different DNMT3A-3L complex conformations for co-methylation. From the kinetic studies, peaks at unique distances were obtained for different modes of co-methylation by DNMT3A and DNMT3A-3L complexes. The short distances (1-2 bp) can be understood by considering a sliding mechanism of a single enzyme complex or single DNMT3A subunit along the DNA turnover after turnover. On the other hand, co-methylation peaks observed at 2 bp distance between the CpG sites in the MW and WM mode of co-methylation cannot be explained by this model since the enzyme complex cannot stay on one strand performing only sliding but needs to perform methylation on the opposite strand, too. These observations can, however, be reconciled by faster diffusive processes such as 'hopping' (reassociation to the same strand or a different strand) or 'intersegmental transfer' (reassociation to a different strand or part of genome). During the hopping process, it is possible that the CpG site is transferred from one of the centrally located DNMT3A subunits to the other leading to the switch of the methylated DNA strand. DNA binding proteins are well known to exhibit one-dimensional diffusion on DNA by sliding and hopping to locate their innate target sites after initially binding to non-specific DNA^{124,125} (see chapter 3). DNA binding proteins find their target sites mainly depending on two criteria: how long a protein spends performing 1D diffusion along the DNA (mean sliding time (t_1)) and how rapid is its movement on the DNA (determined by its sliding diffusion

co-efficient(D_1). However, hopping has been observed up to several hundreds of nanometres¹²⁶. In summary, both the facilitated diffusion processes, sliding and hopping are efficient at short distances as they are diffusional processes and decline with the square of the distance (Fig. 4.10). Therefore, these phenomena cannot support co-methylation preferences observed at very large distances.

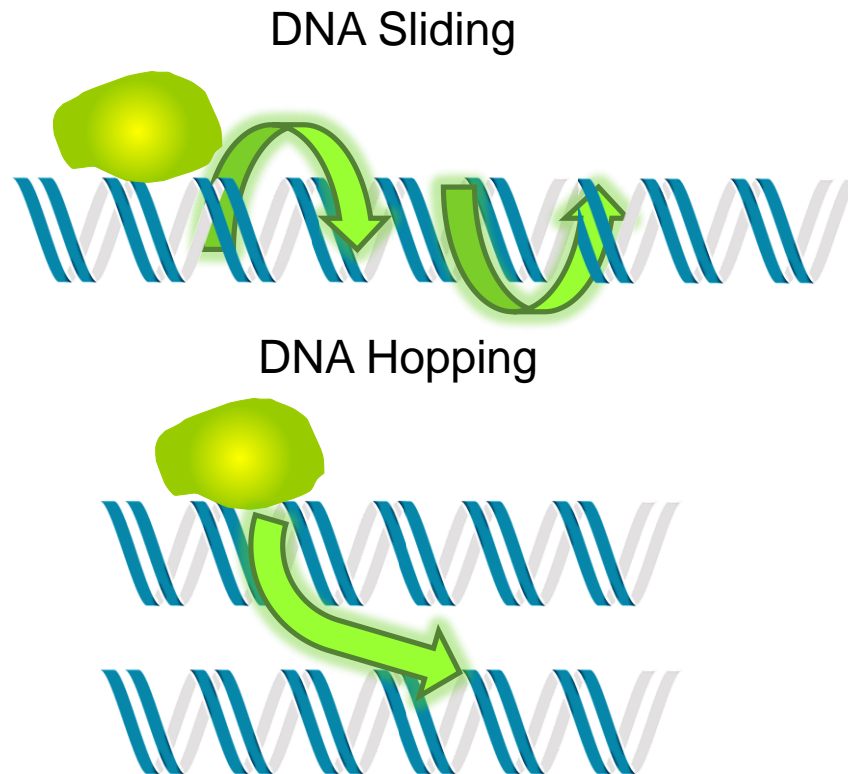


Figure 4.10: Modes of diffusion behavior of DNA binding proteins on the genome for target search.
Figure adapted with permission from Halford and Marko, 2004.

For both MW and WM co-methylation modes, preferences were observed at larger distances of 6, 9 and 12 bp. To investigate the structural basis for these preferences, in this thesis single molecule AFM imaging studies were performed with purified DNMT3A-3L complexes and DNA substrates with 6, 9, and 12 bp distancing between dual CpG sites. As previously observed from the crystal structure and biochemical data^{90,122}, these AFM analyses distinctly showed increased preference for binding to CpG sites separated by 12 bp (Fig. 4.6).

Along with the most preferred distance of 12 bp, the co-methylation data surprisingly uncovered preferences of 8-9 bp for MW and 5-6 bp for WM co-methylation modes. In agreement with these results, the AFM analyses indicated increased binding preference by DNMT3A-3L at the center of D12 and D9 substrates where the dual CpG sites are located. Though the peak at 50% DNA length was present for the D6 substrate, the population was lower compared to the other two substrates consistent with lower

preference for binding to CpG pairs with 6 bp distance. An additional protein peak was observed at around 25-30% of the DNA length in all three substrates. This position marks the transition from the central (102 bp) CpG free region common to all AFM substrates surrounding the central pair of CpG sites to the rest of the non-specific DNA. The non-specific DNA still contains low number of CpG sites at different distances. This observation also supports DNMT3A-3L sliding and hopping along the DNA in search of CpG target sites after initial binding to the CpG free part of the DNA substrate. If a DNMT3A-3L complex binds the CpG free DNA it can move in two ways: It either reaches the double CpG site at the center of the DNA substrate, or it binds to the first CpG site present at the transition of the CpG free central region and the average DNA at ~25-30% of the DNA length giving a plausible reason why many protein complexes are enriched in this region. In particular, two pairs of CpG sites with a spacing of 13 bp at 23% and 28% of the DNA length could be contributing to the 25-30% peak (Appendix, Fig. 4.1b).

Additionally, AFM results showed significant differences in the structural characterization of the DNMT3A-3L complexes bound at the central CpG sites for the DNA substrates with 6, 9, and 12 bp spacing (Fig. 4.7). AFM analyses of the DNMT3A-3L complexes bound to the central CpG sites in D12 substrate revealed two species, a single DNMT3A-3L tetramer with a DNA bending approximately 40° as observed in the crystal structure⁹⁰ (Fig. 4.7c) and a double DNMT3A-3L heterotetramer bending the DNA by $\sim 80^\circ$. Strikingly, 40° bending was also seen for complexes bound at single CpG sites and average DNA (Fig. 4.7). Together with the fact that no bending of CpG sites in the absence of protein was observed, these data imply that the DNMT3A-3L complex directs the DNA bending through protein-DNA interactions. DNA bending has been reported in many DNA binding protein systems and has been proposed as an energetic test to recognize their target sites based on the altered flexibility of target sites^{127,128}. This applies particularly to the DNA repair proteins whose target sites generally distort or destabilize the DNA duplex^{17,21,129,130}. It has been previously shown that DNA bending plays a pivotal role in the recognition of CpG sites by the DNMT3A-3L complex¹³¹. To correlate DNA bending, extent of co-methylation of CpG sites and the DNA sequence, the sequence of the region flanking the central CpG sites was analyzed here (Appendix, Fig. 4.1), and it was observed that a high AT/GC ratio and presence of A-tracts increased MW co-methylation by DNMT3A and DNMT3A-3L. It is known that high AT/GC ratio supports the DNA to bend more easily, and A-tracts are famous for inducing intrinsic bending in the DNA^{118,132}.

In parallel, AFM data analyses also indicated a second class of complexes bound to the D12 substrate showing volumes about twice that of the single tetramer and DNA bending of $\sim 80^\circ$ (Fig. 4.7c). It has been earlier demonstrated that DNMT3A-3L binds co-operatively to DNA^{101,102,105,122}. This implies that the possibility of a second tetramer binding next to the first one is high (Fig. 4.11). Hence, the large volumes/large bending observed in the AFM data for D12 bound complexes likely correspond to two tetramers binding in a side-by-side arrangement on the DNA. In this model, the DNA bending would be additive,

exactly as observed here. On the other hand, AFM data analyses of DNMT3A-3L complexes bound to D6 and D9 substrates revealed predominantly large volumes corresponding to preferential binding of two DNMT3A-3L heterotetramers at the dual CpG sites with 6 bp and 9 bp spacing (Fig. 4.7a). Modelling results demonstrated that WM co-methylation is possible in a 6 bp distance between the CpG sites by orienting two adjacent DNMT3A tetramers in a way where each DNMT3A subunit interacts with the DNA. This resulting side-by-side dimer of DNMT3A-3L tetramers has a rather flat elongated shape along the DNA in an extended conformation. This is clearly verified by the AFM analyses that show long and low, flat complexes on the DNA. These complexes display strong DNA bending by $\sim 100^\circ$. The side-by-side arrangement proposed for complexes formed on D6 substrate is similar to the large complexes formed on D12 substrate, which also show strong DNA bending. However, the interaction pattern between the DNA and protein complex is different in each case. In the D12 substrate, one tetramer interacts with both CpG sites, which are optimally spaced within the tetramer, while the second tetramer is co-operatively bound to the first and interacts with the non-specific region of the DNA neighbouring the first heterotetramer-DNA complex (Fig. 4.11). In contrast, in the D6 substrate, each tetramer interacts with one CpG site with one of its DNMT3A subunits and with non-specific DNA with the other one (Fig. 4.9c). The enhanced DNA bending of 100° (more than the sum of DNA bending induced by two adjacent tetramers) in the complexes on the D6 substrate may be arising from collateral conformational readjustments in the complexes.

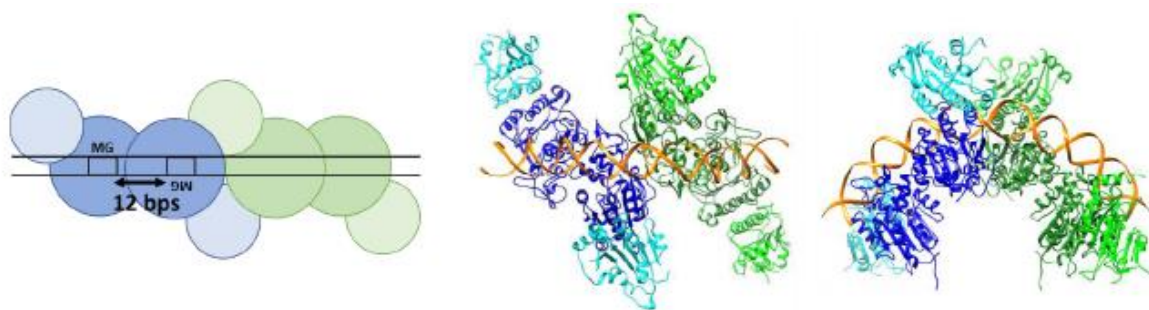


Figure 4.11: Two DNMT3A-3L heterotetramer complexes binding in a side-by-side fashion on the D12 substrate. Two binding patterns can occur for the DNMT3A-3L double heterotetramer: the first where each CpG site is bound to one DNMT3A molecule from each tetramer (Fig. 4.9c) and the second where both CpG sites are bound by two DNMT3A subunits in the first tetramer leaving the second tetramer to bind non-specifically (shown above).

Now to appreciate the mechanism of co-methylation of dual CpG sites placed 9 bp apart by the dimers of DNMT3A-3L tetramers, a structural model was fashioned where two DNMT3A-3L tetramers can bind to the DNA with only one DNMT3A subunit from each tetramer interacting with a CpG site (Fig. 4.9b). This type of arrangement leads to a tetramer swap on the interacting DNA and results in very compact complexes, as corroborated by the AFM heights and lengths of protein complexes. These complexes were observed to induce very strong bending in the DNA by $\sim 100^\circ$. In contrast, the two

short DNA molecules in the model could be connected to give two DNMT3A-3L heterotetramers bound to two CpG sites with 9 bp spacing without DNA bending. However, as can be seen in Figure 4.9b, in this model, each DNMT3A-3L tetramer contains one additional DNMT3A subunit that is not interacting with the central DNA containing the two CpG sites. These subunits are orphaned and may be able to bind the flanking free DNA ends forming loops. Either the left or the right DNA flank may fold back and form ancillary interactions with the free DNMT3A subunits leading to asymmetric additional DNA binding with sufficient conformational alterations to ensure no successive interaction with the second flank. These structural rearrangements could provide a plausible explanation for the strong DNA bending seen from the AFM data together with the tight packing of these complexes. An additional effect of this proposed interaction with the flanking DNA could be a shift of the centroid of the complex by a few base pairs (For the D9 substrate, shift from a 50% position to 40% position would indicate ~40 bp which would translate to ~13.5 nm considering 0.34 nm/bp) (refer to appendix, Fig. 4.1a). Such positional variability may explain the markedly broader position distribution of these complexes at the dual CpG sites in the D9 substrate (width ~10% SD for D9 as opposed to ~2% SD for D6 and D12). In addition, these models also support the differences in co-methylation for DNMT3A and DNMT3A-3L. In DNMT3A complexes, the additional DNMT3A subunits replace DNMT3L at the edges⁹⁰ and are available for additional DNA interaction (which is not the case for DNMT3L), which may explain the higher MW co-methylation peak for CpG sites at 9 bp distance for DNMT3A than DNMT3A-3L. This ancillary binding of DNMT3A-3L complexes might further stabilize the whole structure and result in increased preferences for MW co-methylation of CpG sites with this spacing (9 bp). Although the modelling results for the D9 substrate are convincing, agree with the additional interactions of the DNMT3A-3L complex with DNA, and are supported by all the biochemical and structural data, more details need to be explored experimentally to supplement these models.

Overall, the combined biochemical and single molecule AFM studies presented within this thesis uncovered novel arrangements of dimers of tetrameric DNMT3A or DNMT3A-3L on CpG containing DNA, including both side-by-side and tetramer-swap modes dependent on the distances between the CpG sites. This flexibility in the higher order structures of DNMT3A-3L complexes results in a relaxation of the previously thought stringent distance requirement of 12 bp between two CpG sites. This adaptability of the DNMT3A and DNMT3A-3L complexes is essential for *in vivo* global *de novo* DNA methylation and in the generation of imprints and in human development^{82,94,95}.

5. Investigating functions of hOGG1 in regulation of Myc-mediated transcription

5.1 Introduction

Among a plethora of DNA damage, oxidative modifications are the most prominent type caused by the reactive oxygen species (ROS) generated when the cell faces oxidative stress. Due to its lowest oxidation potential among all DNA bases, guanine most easily undergoes oxidation¹³³ to form 7,8-dihydro-8-oxoguanine (oxoG) lesions^{134,135}. oxoG lesions differ from guanine by a single atom and do not perturb the DNA double helical structure¹⁷ (see also Chapter 3). OxoG is recognized and excised predominantly by the 8-oxoguanine DNA glycosylase (hOGG1 in humans) in the base excision repair (BER) pathway^{30,64}. In BER, lesion detection by glycosylases involves binding to non-specific regions on the DNA, moving on the DNA in 1D diffusion or hopping modes while continuously probing for the presence of their target lesions, and finally flipping of the damaged base into the enzyme active site pockets (see also Chapter 3)^{16,17,136–138}. Glycosylase activity then results in the formation of an abasic (AP) site, which is processed by apurinic/ apyrimidic endonuclease I (APE1 in humans) producing a nick in the DNA backbone.

While the pivotal role of hOGG1 is in the repair of oxoG lesions in DNA and thus maintaining genomic integrity through its glycosylase activity, there has been increasing evidence of its importance in auxiliary pathways such as cell metabolism, inflammatory processes, or innate immune response^{139–143}. For example, it has been suggested that hOGG1 interaction with oxoG lesions in DNA enhances transcription of NFκB target genes, *CXCLs*, *ILs* and *TNFα* that code for inflammatory cytokines and chemokines¹⁰. The promoter regions of several such target genes contain continuous runs of guanine bases. In fact, 72% of the human gene promoters are GC-rich¹⁴⁴. Of all four canonical bases in the DNA, guanine is the most electron rich and sensitive to oxidation. When guanine bases π stack upon each other, the 5' most guanine has the highest sensitivity to oxidation due to a process governed by long-distance electron tunneling through the hydrophobic core of duplex DNA^{145,146}. Analysis of the distribution of oxoG lesions in DNA by next generation sequencing and oxoG-sequencing experiments revealed their confinement predominantly to promoter regions of the target genes of several major transcription factors^{147–150}. Such DNA modifications in the promoter regions may act as a transient mark to which hOGG1 may then bind to either block or activate transcription initiation¹⁵¹. In fact, enhancement of Myc-mediated transcription was observed in the presence of oxoG lesions and hOGG1 on the promoter regions of *Ncl* and *Cad* genes.

Several studies have reported a role of hOGG1 in the regulation of transcription of certain important genes. For instance, it has been demonstrated that the presence of oxoG lesions ~8-11 bp upstream of the NFκB binding consensus sequence, hOGG1 enhanced NFκB occupancy on the DNA and augmented the mRNA levels of its target genes^{142,152}.

Moreover, it was shown that *ogg*^{-/-} mice express lower levels of proinflammatory cytokines/chemokines such as TNF α , IL-4, IL-6, IL-10, IL-12, and IL-17 decreasing the homing of inflammatory cells to the site of wounds¹⁵³. In addition, and in contrast to this enhancement of transcription factors binding to their target sequences, the conventional view that oxidative DNA damage may be detrimental to cellular processes was also shown to be true in several cases. When oxoG lesions were located in the cognate recognition elements of transcription factors like nuclear factor kappa-light-chain enhancer of activated B cells (NF κ B)¹⁵⁴, specificity protein 1 (Sp1)¹⁵⁵ or cAMP responsive element binding protein 1 (CREB)¹⁵⁶, their binding affinity was found to be lowered extensively.

Pro-inflammatory genes are transcribed when a cell is under oxidative stress. Myc is known as a master regulator of a varied range of genes involved in apoptosis, inflammation processes¹⁵⁰, cell proliferation, metabolism and tumorigenesis¹⁵⁷. There are three forms of Myc proteins found in vertebrates- cellular-Myc (c-Myc or Myc), Myc found in neuroblastomas (N-Myc) and small cell lung carcinomas (L-Myc)^{158,159}. c-Myc will be designated as Myc in this thesis. Myc binds to its cognate recognition sequence in the presence of its binding partner, Myc associated factor (Max). The Myc-Max complex is a heterodimer which recognizes and binds the CACGTG consensus sequence called E-box motif¹⁶⁰⁻¹⁶².

As Myc is involved to a major extent in regulating inflammatory processes, which are generated by oxidative stress conditions and enhancement of Myc target gene transcription was observed in the presence of hOGG1, in this thesis, I aim to investigate the influence of hOGG1 on the recruitment of Myc to its consensus sequence, E-box, which then modulates transcription of target genes. To this end, I will first introduce an overview of possible scenarios of how Myc-mediated transcription may be regulated by oxoG lesions and/or hOGG1. A schematic of the different models is shown in Fig 5.1. These schematic models can be divided into scenarios that either involve hOGG1 catalytic activity, or those that do not (see Fig.5.1).

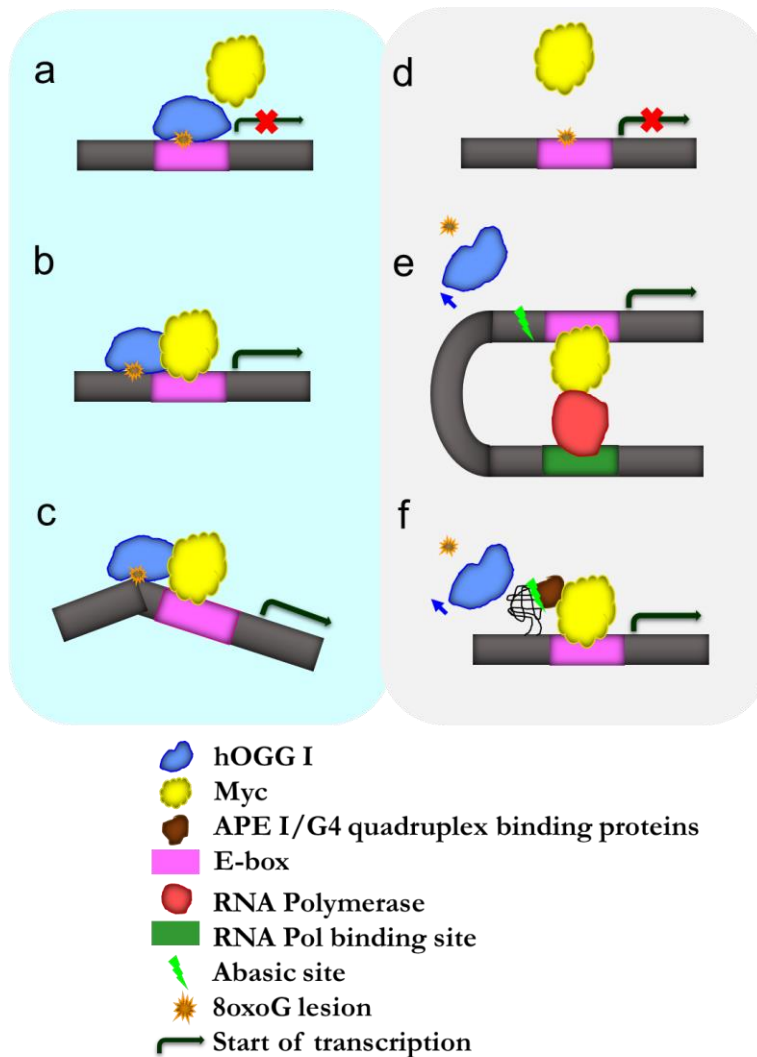


Figure 5.1: Schematic models describing various possible scenarios of Myc-mediated transcription regulation by hOGG1 under oxidative conditions. Oxidative conditions introduce oxoG lesions in the DNA. **(a)** hOGG1 (blue) bound to the oxoG lesion (orange) in the E-box motif (magenta) abrogates Myc (yellow) binding to its cognate CACGTG sequence (E-box) thus leading to transcription inhibition. **(b)** hOGG1 bound to the oxoG lesion sitting at a suitable distance upstream of E-box augments the recruitment of Myc to its consensus sequence through direct protein-protein interactions. **(c)** Enhancement of downstream target gene transcription may result from structural modifications in the DNA induced by protein-protein interactions. In this scenario, recruitment may still take place without direct protein-protein interactions. **(d)** The presence of an oxoG lesion alone in the E-box motif inhibits Myc binding to this region. This model depends on hOGG1 catalytic activity to remove the obstacle and enable Myc binding to its target site. **(e)** hOGG1-initiated base excision repair leads to the formation of an abasic site (AP site, green lightning bolt) and eventually a strand break, which releases DNA double helix constraints and allows DNA conformational changes to take place bringing the transcription machinery together from larger distances on DNA¹⁶³. **(f)** A conformational change induced by hOGG1 catalytic activity enhances transcription. The AP site generated by hOGG1 activity in the G rich sequences in the promoter regions leads to the formation of a G4 quadruplex, which in turn allows to APE1 binding that may then recruit Myc. Along with APE1,

components of transcription machinery may also bind to G4 quadruplex and assist in transcription initiation¹⁴⁶.

In the first two scenarios, either an oxoG lesion alone or hOGG1 bound to an oxoG lesion within the E-box motif poses an obstacle to Myc binding. In this scenario, either the hOGG1-lesion complex (Fig. 5.1a) or the lesion (Fig. 5.1d) inhibits Myc binding to its consensus recognition sequence and either attenuates or abrogates transcription of downstream target genes. Evidence for such mechanisms has been demonstrated by previous studies for the transcription factors Sp1, NFκB, and CREB^{154,155,164}.

Secondly, Myc-mediated transcription may be enhanced by direct protein-protein interactions between Myc and hOGG1, which is bound to an oxoG lesion at an optimal distance from the Myc consensus binding sequence (Fig. 5.1b). In this scenario, either hOGG1 and Myc may interact prior to DNA binding and arrive at the promoter region as a preformed complex, or Myc may be recruited by hOGG1 already bound to the oxoG lesion, to enhance recruitment of Myc to its cognate E-box motif by direct protein-protein interactions. For example, it has been demonstrated that hOGG1 bound to an oxoG lesion present ~8-11 bp upstream of the NFκB binding site augmented NFκB occupancy on the DNA¹⁰. Moreover, direct interactions have previously been demonstrated between hOGG1 and Sp1¹³⁹.

hOGG1 binding to oxoG lesions upstream of E-box can bend the DNA (see Chapter 3) and induce architectural changes in the DNA allowing the recruitment of Myc to its target site (Fig. 5.1c). Such a mechanism has been proposed for NFκB and transcription initiation complex recruitment^{163,165}. This scenario may involve direct protein-protein interactions between hOGG1 and Myc or not. In another case, hOGG1 catalytic activity leads to the formation of single strand breaks eventually in the DNA which induces conformational changes in the chromatin. As observed in chapter 3, nicks in the DNA show higher bendability and their presence lead to opening of chromatin which may allow components of transcription machinery present at large distances on the DNA to come closer and form a complex to initiate transcription¹⁶³ (Fig. 5.1e). Direct protein-protein interactions are also possible between APE1 bound to AP site (product of hOGG1 activity) and Myc either with or without DNA bending like proposed for hOGG1 (Fig. 5.1b and c)¹⁵⁰. These are alternative scenarios where transcription is enhanced as an indirect consequence of hOGG1 activity.

In the final scenario, the abasic site (AP site) generated by glycosylase activity of hOGG1 is responsible for transcription regulation. The AP site is the substrate for apurinic or apyrimidic endonuclease (APE1) of the BER pathway. The AP site together with APE1 have been implicated in transcription regulation in the case of target genes containing G4 quadruplexes in the upstream regulatory regions^{163,166,167}. It has been shown that *VEGF* expression is enhanced by the presence of oxoG lesions in the promoter containing guanine-rich potential G4 quadruplex forming sequence (PQS)¹⁶⁸. PQS contains up to five runs of 3-5 guanine bases with gaps of 2-3 other DNA bases. hOGG1 initiates BER of

oxoG in the PQS, which leads to the formation of an AP site. The AP site destabilizes the duplex structure of B-form DNA resulting in the melting of the complementary strands and the formation of a G4 quadruplex. Due to destabilization, AP site is present in the 5th spare G track of the G4, which forms a loop like structure and is accessible for binding by APE1. APE1 binds the AP site present in the single stranded DNA (ssDNA) loop very tightly and its endonuclease activity is stalled, resulting in extended binding time of APE1 at this site^{146,168,169}. APE1 may then serve to directly recruit transcription factors and/or RNA polymerase to the DNA^{146,170} (Fig. 5.1f). Other transcription factors like Myc-associated zinc-finger protein (Maz) and heterogenous nuclear ribonucleoprotein A1 bind to G4 quadruplexes and recruit RNA polymerase for enhanced transcription¹⁷¹. In addition, CHIP and quantitative RT-PCR studies have shown that the repair of oxoG lesions by hOGG1 in the promoter PQS regions of *bcl2*, *Myc*, and *KRAS* genes modulate transcription^{166,172,173}. A schematic describing the process is shown in Fig. 5.2.

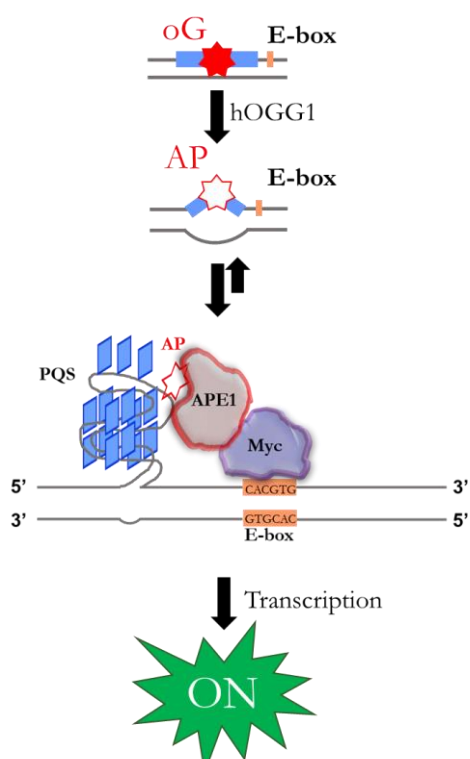


Figure 5.2: Schematic model of the initiation of G4 quadruplex formation by an oxoG lesion in a target gene promoter containing a potential G4 quadruplex forming sequence (PQS). Repair of the oxoG lesion by hOGG1 leads to the formation an AP site. Due to the absence of a base, DNA at this site is destabilized, which leads to a structural switch from duplex DNA to G4 quadruplex. AP site is located in the 5th spare G track (loop region) due to inability to form hydrogen bonds with other guanine bases in the core quadruplex structure. APE1 performs endonuclease function when its target site is present in ssDNA leading to its prolonged residence time at this site. APE1 then eventually recruits transcription factor/ RNA polymerase allowing an enhanced transcription of downstream target genes (Fig. 5.1f). In addition, other G4 quadruplex binding proteins also augment transcription^{146,171}.

Another interesting feature that has been proposed to govern Myc binding to E-boxes is the involvement of the histone modifier Lysine Specific Demethylase (LSD1). Methylated lysine residues on histones, particularly H3K4me3, are targeted and demethylated by LSD1, to open condensed chromatin. LSD1 has been shown to be involved in the transcription regulation of androgen and estrogen receptor target genes¹⁷⁴. Interaction between Myc and LSD1 has been demonstrated earlier in the context of activation of transcription of the Myc target genes *Ncl* and *Cad*¹⁵⁰. Quantitative ChIP studies have shown that hOGG1, LSD1 and Myc were all present on the same chromatin regions¹⁵⁰. hOGG1 was not present on the DNA when either LSD1 or hOGG1 were silenced, and transcription of *Ncl* and *Cad* genes was inhibited under these conditions. LSD1 contains a Flavin Adenine Dinucleotide (FAD) molecule, which when reduced during histone demethylation activity leads to the formation of hydrogen peroxide (H₂O₂). Hence H₂O₂ is generated during demethylation and is believed to lead to local oxidation in the vicinity of upstream promoter regions of Myc target genes resulting in the local formation of 8-oxo-guanine lesions, the substrate of hOGG1. From these studies, it was hypothesized that LSD1 may recruit hOGG1 to oxoG lesions in the proximity of E-box motifs. Figure 5.3 shows a schematic of the involved events.

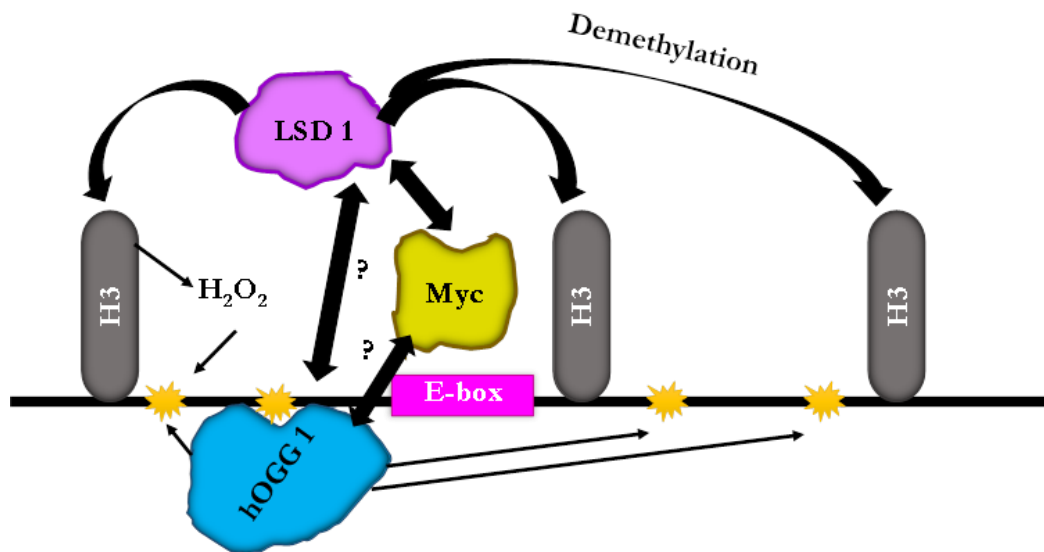


Figure 5.3: Role of DNA oxidation and hOGG1 in Myc-mediated transcription regulation. H₂O₂ is produced locally by LSD1 activity as a by-product during demethylation of lysine residues on the histones. This leads to the formation of oxoG lesions (orange) on the DNA, specifically near the promoter regions of Myc-target genes. hOGG1 is recruited to initiate repair of the oxoG lesions. In the model shown here, hOGG1 enhances the recruitment of Myc to its cognate recognition DNA consensus sequence, CACGTG (termed E-box motif) and subsequent transcription of downstream target genes. The question marks indicate possible interactions between all three proteins.

Furthermore, it has been hypothesized that upon oxidative stress in the cells, along with the formation of oxoG lesions in the DNA, hOGG1 is also oxidatively modified¹⁶⁵. Under

the reducing conditions in the cell, hOGG1 actively performs oxoG lesion repair. It has been suggested that oxidation of hOGG1 abrogates its catalytic activity by cysteine modifications, leading to prolonged residence time at the lesion and once redox conditions in the cell are re-established, hOGG1 regains its activity¹⁶⁵. Another study showed that the oxidative stress induced in the presence of cadmium demolished hOGG1 glycosylase activity irreversibly *in vitro* and reversibly *in vivo*¹⁷⁵. In this thesis, hOGG1 dimer formation upon oxidation was observed by several methods. Interestingly, however, oxidative modifications of hOGG1 did not affect its activity.

In this thesis, I aim to investigate direct interactions of hOGG1, Myc and LSD1 in the absence of presence of an oxoG lesion in the DNA by atomic force microscopy (AFM), electromobility shift assay (EMSA), fluorescence polarization, *in vitro* pull-down experiments, and native agarose gel electrophoresis (NAGE). In addition, AFM studies were performed with pre-oxidized or reduced hOGG1 to unambiguously demonstrate the effect of oxidation of hOGG1 itself on the recruitment of Myc. In summary, results from this study help to alleviate our understanding of the role of hOGG1 in Myc-mediated transcription.

5.2 Results

5.2.1 hOGG1 oligomerization is altered under oxidative conditions

In this thesis, during several experiments it was observed that two distinct oligomeric states existed for hOGG1, depending on oxidative or reducing conditions. The results shown in Figure 5.4 demonstrate the observed two states of the protein. hOGG1 either pre-treated with H₂O₂ or incubated in DTT containing buffer was loaded under ambient (buffer containing no reducing agent) conditions onto a 15% SDS PAGE gel. Inferring from the protein ladder next to the samples, it was observed that hOGG1 behaved as a dimer under oxidizing conditions and as a monomer when present in DTT containing buffer (Fig. 5.4a, left panel). Under ambient conditions, 50% of both species were observed. Interestingly, oxidation of hOGG1 was reversible (Fig. 5.4a, right panel). CD spectroscopy of hOGG1 confirmed that the secondary structure of hOGG1 was not altered under the oxidative and reductive conditions (Fig 5.4b).

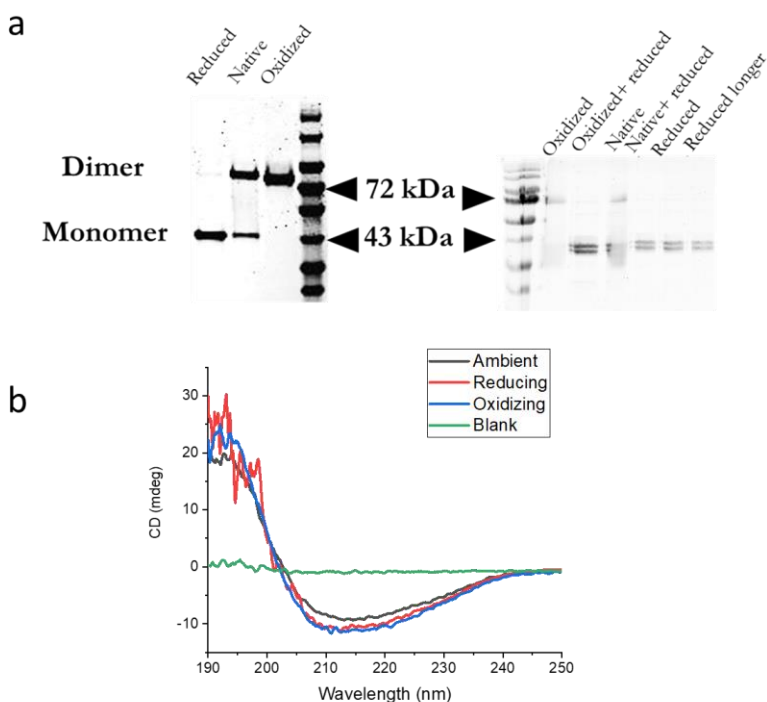


Figure 5.4: Analysis and characterization of oligomeric states of hOGG1 by SDS-PAGE and circular dichroism experiments. (a) Left panel: Non-reducing SDS-PAGE analysis showed a change in the molecular weight of hOGG1 upon oxidation to ~80 kDa representing a hOGG1 a dimer. Under reducing conditions, hOGG1 (~40 kDa) exists as a monomer and an equilibrium between monomer and dimer species was seen under ambient conditions. Experiments were performed in triplicate (see appendix for individual experiments, Fig. x5.3). Right panel: Non-reducing SDS-PAGE analysis showed that oxidation of hOGG1 is reversible. hOGG1 forms a complete dimer when treated with H₂O₂ (Lane 1). When oxidized hOGG1 is treated with the reducing agent (DTT), hOGG1 forms a monomer (Lane 2). hOGG1 under ambient and reducing conditions are included as controls (Lane 3-6). (b) Circular dichroism studies show that the secondary/tertiary structure of hOGG1 was unaltered under extreme redox conditions (DTT: 5 mM and H₂O₂: 100 μ M).

AFM experiments on pre-oxidized hOGG1 or hOGG1 in reducing (non-oxidizing) environment incubated with oxoG lesion containing DNA substrate showed volumes of oxidized hOGG1 and reduced hOGG1 bound to DNA that corresponded to dimeric and monomeric hOGG1, respectively (top and bottom panels, Fig. 5.5a). The results from AFM and SDS PAGE experiments were further corroborated with EMSA experiments where oxidized and reduced forms of hOGG1 were incubated with oxoG lesion containing DNA or undamaged DNA (Fig. 5.5b). The gels show the intermediate shift representing the monomer species exclusively under reducing conditions (grey arrow in the bottom panel of Fig. 5.5b). This species was not observed for undamaged DNA, but only on the oxoG lesion containing DNA implying that the monomeric form of hOGG1 is stabilized at the lesion. hOGG1 binding to oxoG DNA under oxidizing conditions or to undamaged DNA under oxidizing or reducing conditions required higher protein concentrations and bound as a higher order oligomer. The exact oligomeric state cannot be resolved here because the protein-DNA complex bands remain in the gel pockets. Consistently, AFM volume analyses also suggest that under reducing conditions hOGG1 predominantly binds to oxoG as a monomer, while under oxidizing conditions it is present as a dimer.

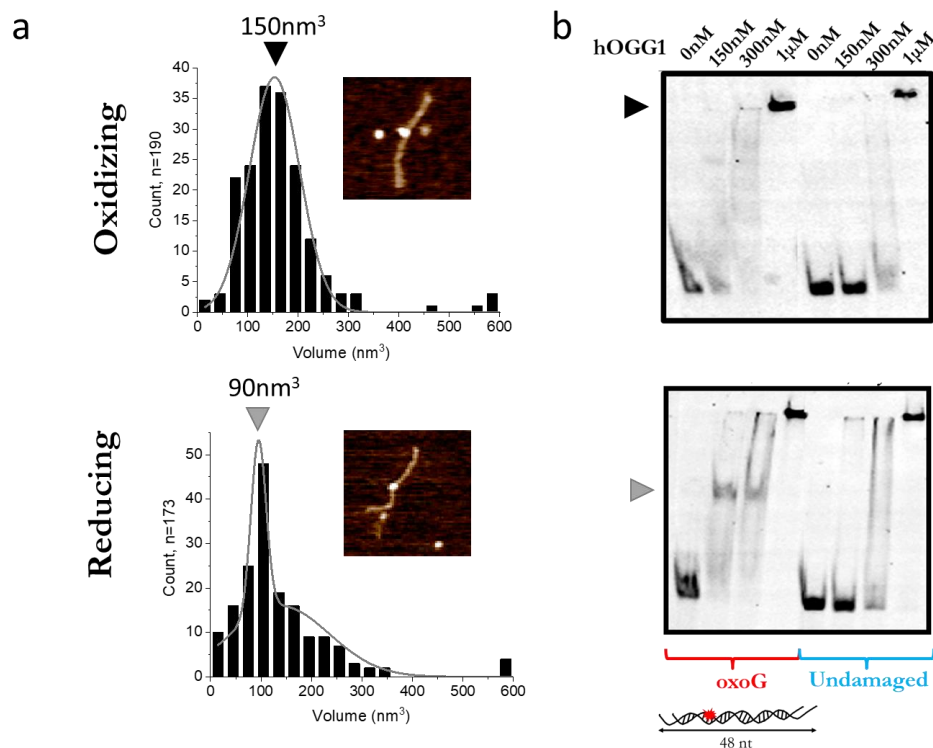


Figure 5.5: hOGG1 oligomerization studies using AFM volume analysis and electrophoretic mobility shift assays. (a) AFM volume analyses of hOGG1 complexes bound at an oxoG lesion (located at 50% of the DNA substrate length) show dimer (~150 nm³, black arrow) and monomer (~90 nm³, grey arrow) species of hOGG1 under oxidizing (top panel) and reducing conditions (bottom panel) respectively. The insets show exemplary hOGG1 complexes bound at oxoG lesion (50% position) under oxidizing conditions and reducing conditions. Individual experiments are shown in Fig. x5.4 in the appendix section. (b) EMSA with 48 nt long substrates (containing oxoG lesion and undamaged) and oxidized/reduced hOGG1. An intermediate shift is observed only under reducing conditions on oxoG lesion DNA representing a monomer of hOGG1 (grey arrow). However, this species is absent under oxidative conditions and hOGG1 forms a dimer (or a higher order oligomeric species) indicated by the super shift (black arrow). A schematic of the DNA substrate containing oxoG lesion used for EMSAs is shown below the gels.

Consistent with the EMSA results (Fig. 5.5b), the specificities for oxoG binding calculated from the AFM position distributions revealed a higher specificity towards oxoG lesions for monomeric hOGG1 under reducing than for the dimer of pre-oxidized hOGG1 (Fig. 5.6a). Specificities of 320 and 785 were obtained for pooled data from three independent experiments for pre-oxidized hOGG1 and under non-oxidative conditions, respectively. In addition, different affinities for reduced and oxidized hOGG1 (wt as well as catalytically inactive hOGG1_{K249Q}) towards an oxoG lesion or undamaged DNA were also supported by fluorescence polarization experiments (Fig. 5.6b). AlexaFluor 647 (AF647) labelled DNA, with or without an oxoG lesion was incubated with hOGG1 and fluorescence polarization values were recorded using a plate reader. The AF647 polarization values obtained for each assay were plotted against increasing concentration of the ligand (hOGG1_{wt} or hOGG1_{K249Q}). K_D values were obtained by fitting these curves using the Hill equation in Origin Pro software. The final plots show K_D values for each interaction under

the two distinct redox conditions (Fig 5.6b). Measurements were made in triplicate. These data showed higher affinities towards the oxoG lesion under reducing conditions for both, hOGG1_{wt} and hOGG1_{K249Q}. At undamaged sites marginal or no difference was seen in affinity between oxidizing and reducing conditions for hOGG1 and hOGG1_{K249Q}.

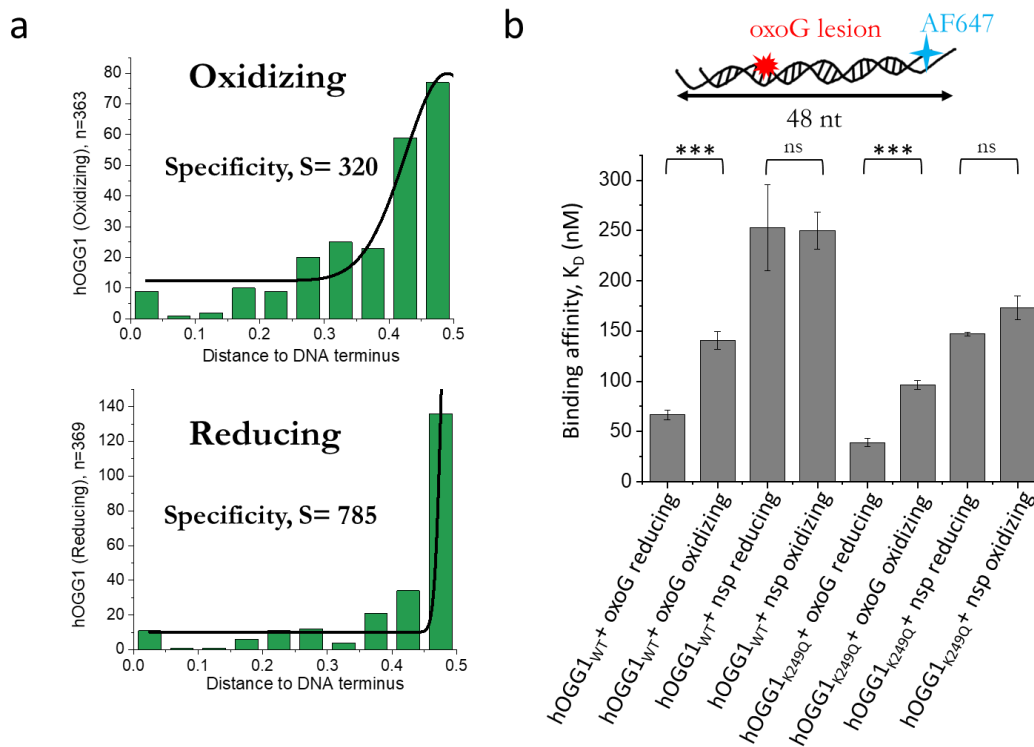


Figure 5.6: hOGG1 shows enhanced specificity and affinity towards oxoG lesion under reducing conditions. (a) The specificity fits to the position distributions reveal that hOGG1 has higher recognition ($S \sim 785$) towards oxoG lesion under reducing rather than oxidizing conditions ($S \sim 320$). Values are obtained from the triplicate. Individual experiments are shown in appendix section, Fig. x5.4. (b) Fluorescence polarization measurements show higher K_D values for interactions of hOGG1_{wt} and hOGG1_{K249Q} with oxoG lesion containing substrate under reducing compared to oxidizing conditions (hOGG1_{wt} oxidizing: 140.4 ± 9.3 nM, hOGG1_{wt} reducing: 66.5 ± 4.7 nM, hOGG1_{K249Q} oxidizing: 96.7 ± 3.8 nM, hOGG1_{K249Q} reducing: 39.1 ± 5.0 nM). hOGG1_{wt} and hOGG1_{K249Q} interacting with undamaged DNA show non-significant differences in affinities between oxidizing and reducing conditions (hOGG1_{wt} oxidizing: 249.8 ± 18.0 nM, hOGG1_{wt} reducing: 252.1 ± 43.2 nM, hOGG1_{K249Q} oxidizing: 172.8 ± 10.1 nM, hOGG1_{K249Q} reducing: 146.8 ± 2.1 nM. Refer to appendix section Fig. x5.5 and Table x5.1 for exemplary polarization curves and binding affinity values from individual experiments of the triplicate). The top panel shows the DNA substrate used for fluorescence polarization studies. The oxoG lesion containing or undamaged top strand was annealed separately with AF647 labelled bottom strand. (Significance values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, n.s- non-significant. p values were obtained by a two-tailed student t -test. p -values for affinities observed between oxidizing and reducing conditions: hOGG1_{wt} bound to oxoG: 0.000112 (bars 1 and 2), hOGG1_{wt} bound to non-specific or undamaged (nsp) DNA: 0.466 (bars 3 and 4), hOGG1_{K249Q} with oxoG: 0.00004 (bars 5 and 6), and hOGG1_{K249Q} with nsp: 0.52 (bars 7 and 8).

To further investigate the above seen oligomeric changes in hOGG1, two other independent methods were used. Using sedimentation velocity analytical ultracentrifugation (SV-AUC) under oxidizing conditions, an additional species with a

sedimentation co-efficient of 4 S was seen in addition to a monomeric species (2 S) for 100 μ M of hOGG1, which completely absent under reducing conditions (Fig. 5.7a). Consistent with the data presented above, under ambient conditions, dimer formation was also observed but in a lower number compared to oxidizing conditions. However, the dimeric state showed only low population compared to the monomeric state in the AUC experiments, in contrast to the results seen in the experiments described before. In addition, analytical size exclusion chromatography of hOGG1 under oxidative condition showed an equilibrium between monomeric and dimeric hOGG1 that was shifted towards the lower elution volume compared to under reducing conditions. These results indicate a higher ratio of hOGG1 dimer to monomer under oxidizing compared to reducing conditions (Fig. 5.7b). Additionally, a different ratio of monomeric to dimeric states was also observed for a lower concentration of hOGG1 applied to the analytical column (Appendix, Fig. x5.6).

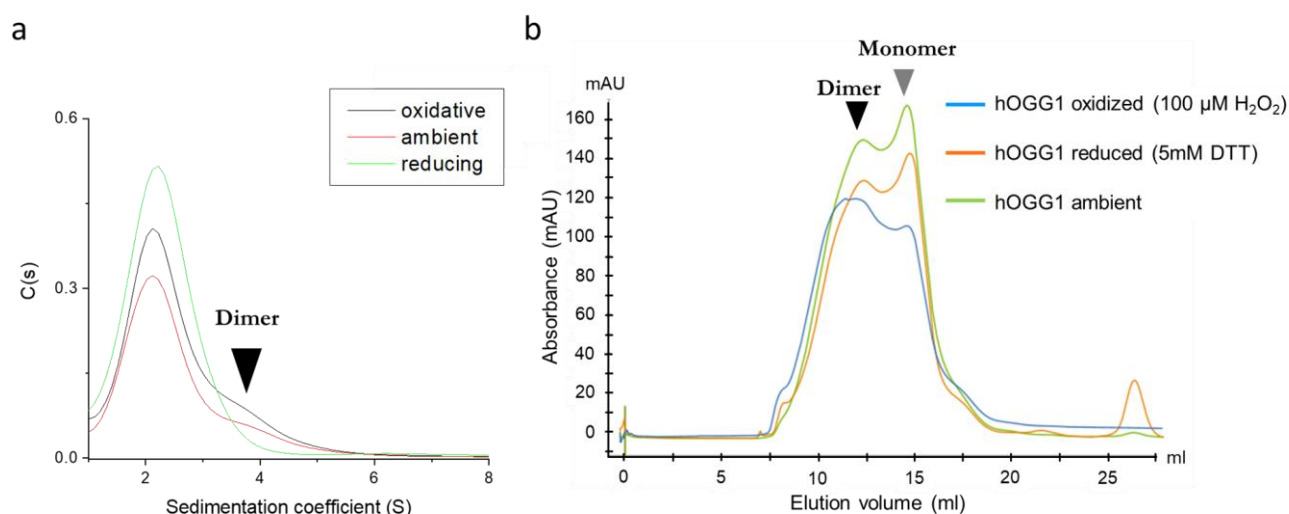


Figure 5.7: SV-AUC and analytical SEC experiments reveal an equilibrium between the monomer and dimer forms of hOGG1. (a) 100 μ M hOGG1 was applied to AUC experiments. Dimer formation (4S) is observed under oxidizing as well as under ambient conditions and is completely abolished under reducing conditions. Dimer formation is thus enhanced under oxidative condition. The predominant state of hOGG1 is found to be a monomer (2S) under all environments in the AUC measurements. (b) 30 μ M of hOGG1 was applied to analytical size exclusion column. Under oxidizing as well as reducing conditions, an equilibrium between a monomeric (grey arrow) and a dimeric state (black arrow) was observed. Under oxidative condition, the equilibrium was shifted towards the elution volume of the hOGG1 dimer (12.1 mL). Ambient and reducing conditions show similar populations of monomer and dimer, with slightly more monomer (elution volume: 14.8 mL) than dimer.

Previously it has been hypothesized that hOGG1 undergoes cysteine modifications upon oxidation¹⁶⁵. Further aim of this study aim was to identify the cysteine residue that might be responsible for disulfide bond formation leading to dimerization. To this end, solvent accessibility analysis was performed. The crystal structure of hOGG1 bound to an oxoG lesion in DNA (PDB id: 1ebm) was uploaded to the PDB-PISA server to obtain the solvent parameters of the interfaces and assemblies in a potential hOGG1 dimer. Cys28, Cys 241

and Cys253 residues showed highest solvent accessibility (indicated in Fig. 5.8a). Therefore, these three hOGG1 mutants (C28A, C241A and C253A) were cloned, expressed, and purified (Refer appendix Fig. x5.1 and Fig. x5.2 for details). A non-reducing SDS-PAGE based dimerization assay (see appendix Fig. x5.6 for details) and subsequent western blot analysis (Fig. 5.8b) was carried out for all three cysteine mutants to assess dimerization and hOGG1_{wt} was included as a positive control. Out of these variants, only hOGG1_{C28A} did not show dimerization upon oxidation indicating that the C28 residue might be responsible for disulfide bond formation in the hOGG1 dimer. SDS-PAGE gel revealed presence of additional bands above hOGG1 monomer band for all three cysteine mutants (Appendix, Fig. x5.7a). These were confirmed to be contaminant proteins from expression strains by mass spectrometry (Appendix, Fig. x5.7b). To identify the cysteine residue in hOGG1 responsible for dimerization, mass spectrometry analysis of the dimer band of hOGG1 was performed. Interestingly, the results suggested similar probabilities of oxidation and disulfide bond formation for all cysteine residues present in hOGG1 (Fig. 5.8c). These experiments provided ambiguous results.

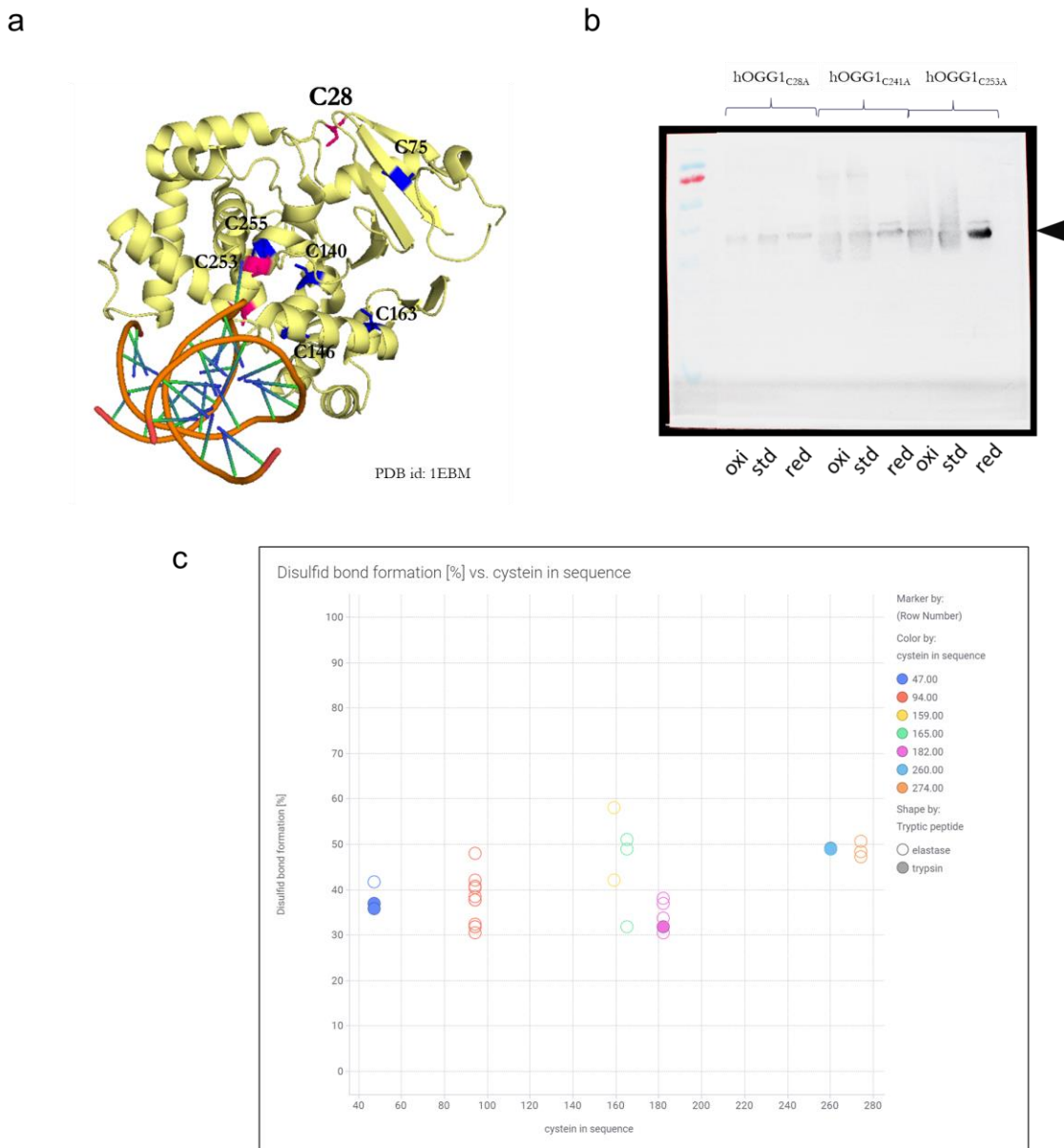


Figure 5.8: Western blot and Mass spectrometry analysis of hOGG1 to identify cysteine residues involved in dimerization. (a) PDB structure of hOGG1 bound to oxoG lesion displaying all seven cysteines, C28, C75, C140, C146, C163, C253 and C255. Based on the solvent accessibility values from PDB-ePISA server, C28, C146 and C253 were expected to be most susceptible to disulfide bond formation. (b) Western blot analysis of all three hOGG1 cysteine mutants. hOGG1_{C28A} did not dimerize indicating C28 as the residue responsible for disulfide bond formation. Black arrow indicates the bands seen at the monomeric size of hOGG1 (~40 kDa). (c) Based on mass spectrometry (MS) results after light and heavy isotope labelling of cysteines in hOGG1, it was observed from the identified cysteines in the peptides obtained after hOGG1 digestion with trypsin and elastase, that they all have comparable percentage of disulfide bond formation.

Nevertheless, initial crystallization experiments were performed on oxidized hOGG1 with the aim to resolve the structural basis of hOGG1 dimerization. Although microcrystals were observed in the Nucleix suite after 75 days at 20°C (Fig. 5.9), optimization of the

microcrystals using the hanging drop method was performed. However, all attempts to obtain bigger crystals were unsuccessful. Crystals were extremely difficult to be fished out of mother liquor for diffraction. Difficulties in obtaining crystals of the hOGG1 dimer are consistent with the mass spectrometry results above, which indicated all cysteines to be involved in disulfide bond formation. This would lead to heterogeneity and be unfavorable for crystal formation.

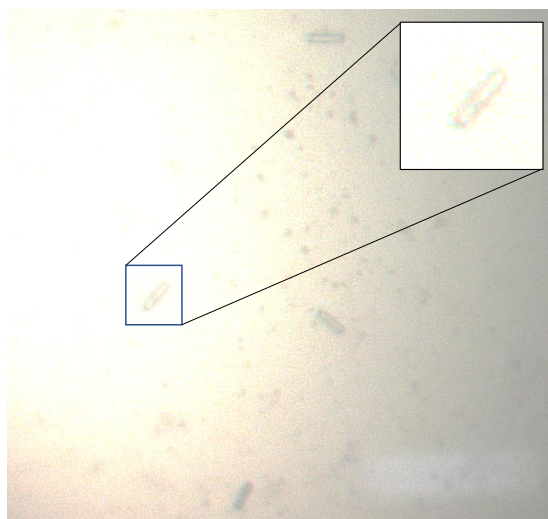


Figure 5.9: Microcrystals of oxidized hOGG1 obtained in the nucleix suite. These initial rod-shaped crystals were obtained at 20°C after 75 days.

To investigate the effect of oxidative and reductive environments on hOGG1 catalytic activity, hOGG1 was incubated with Cy3 labelled oxoG lesion containing DNA at 37°C and the abasic site left by the glycosylase was nicked by adding NaOH. The reaction was stopped by heating to 95°C for 5 min and adding orange G dye. Analysis of product formation revealed similar amounts of cleaved products under oxidizing and reducing conditions (Fig. 5.10a) suggesting that hOGG1 glycosylase activity is unaffected by the redox conditions. As a control, the catalytically inactive hOGG1_{K249Q} variant was used, which did not display any activity (Fig. 5.10b). Importantly, these data demonstrate that hOGG1 catalytic activity is not affected by oxidizing conditions, in contrast to what had been proposed in the literature¹⁶⁵. (Fig. 5.10c). Kinetics of these data are not shown here since intensities of all the cleaved bands from 10 min and onwards do not show any increase in intensities.

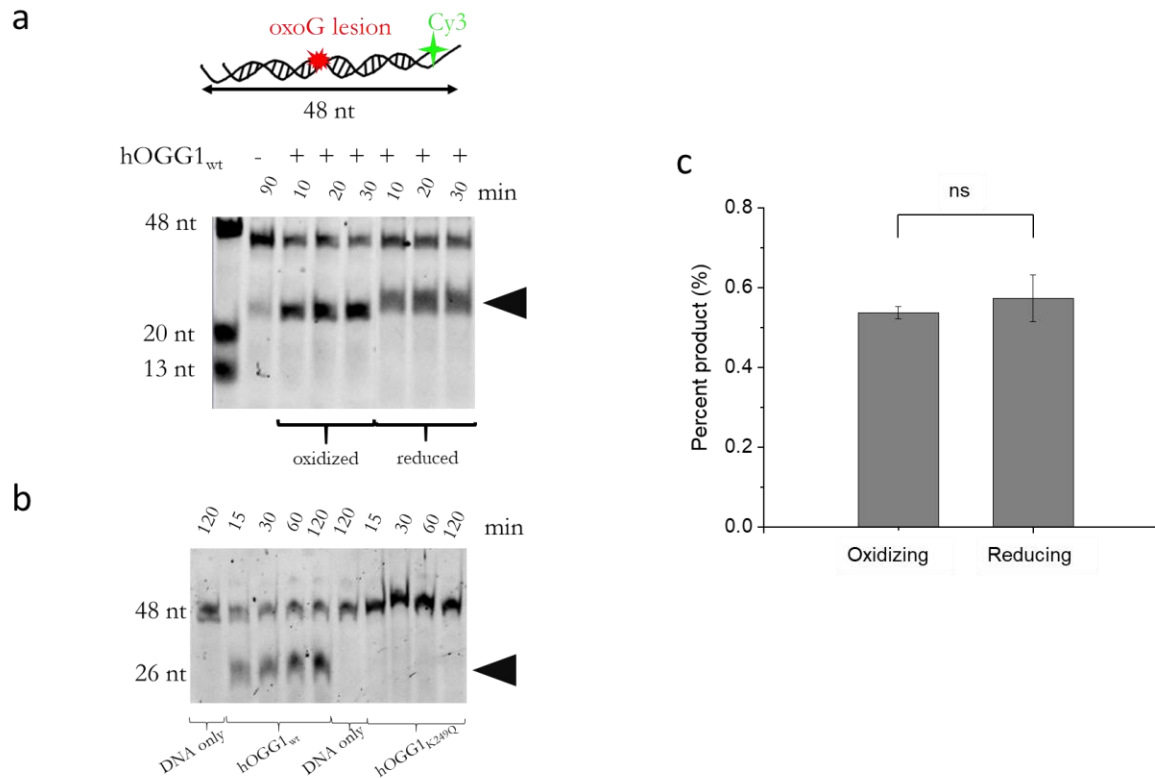


Figure 5.10: Extreme redox conditions do not alter hOGG1 activity. (a) hOGG1_{wt} shows activity under both oxidative and reductive environments (refer to Appendix, Fig. x5.7 for duplicate and triplicate experiments). Black arrow shows the cleaved DNA due to hOGG1 activity. Cy3 labelled, oxoG containing DNA substrate used in the assay is shown schematically at the top. (b) 15% Urea gel showing the cleaved products when Cy3 labelled oxoG lesion containing DNA is incubated with hOGG1_{wt} for increasing time as indicated. At the same time, lanes with the catalytically inactive hOGG1_{K249Q} variant did not show bands at the height of cleaved DNA demonstrating its inactivity. (c) Percent cleaved DNA was calculated for three time points after correcting for the background grey values and the average (\pm SD) percentage values were plotted for both oxidizing and reducing conditions. Gels from duplicate and triplicate experiments are shown in the appendix section, Fig. x5.8.

5.2.2 hOGG1 alters Myc binding to target sites

Myc is known to be a “Master” transcription factor that binds ~25,000 sites in the human genome¹⁷⁶. Myc binds to a promoter element of its target genes called E-box¹⁷⁷. It has been shown that transcription regulation by Myc is also influenced by LSD1 histone modifier¹⁷⁸. Histone demethylation by LSD1 leads to opening of condensed chromatin, which in turn allows easier access and promotes Myc binding to its target sites. It has also been hypothesized that histone demethylation by LSD1 leads to local DNA oxidation resulting in the formation of oxoG lesions, which then lead to the recruitment of the DNA glycosylase hOGG1, which may be involved in enhancing Myc recruitment to its binding site^{150,179–181}. In this context, LSD1 has been also shown to interact with Myc¹⁵⁰. It has

also been demonstrated that hOGG1 was present in the upstream promoter regions of *Ncl* and *Cad* genes, two targets of Myc¹⁵⁰. The premise of this study was to investigate the effect(s) of hOGG1 on Myc binding to its target recognition sequence to understand the roles of hOGG1 in regulation of Myc induced transcription.

There are several different possibilities for how hOGG1 may influence Myc binding to its cognate DNA element, the E-box motif, as described in the introduction section (Fig. 5.1). The first two models are trivial. An oxoG lesion in the E-box motif may block Myc binding (Fig. 5.1d), which can be restored by repair activity of hOGG1. In the second scenario, hOGG1 staying bound to oxoG lesion may also itself block Myc binding (Fig. 5.1a). These models are not related to oxoG introduction in DNA by LSD1 activity and do not explain enhancement of Myc recruitment by LSD1.

I want to focus here on the role of hOGG1 in LSD1 induced enhancement of Myc binding to its cognate E box recognition site. To investigate the model of direct protein-protein interactions, I performed biochemical and biophysical experiments to investigate interactions between the three components, - Myc, hOGG1 and LSD1. Direct protein-protein interactions were initially confirmed by *in vitro* pull-down assays. GST-tagged Myc_{CFL} was purchased to perform these experiments. For interactions of hOGG1 and LSD1, 6x His-tagged hOGG1 was used as the bait protein. For interactions of Myc with LSD1 and Myc with hOGG1, GST-tagged Myc was used as the bait protein. The proteins used as prey in these experiments did not contain any tags. Under both oxidizing and reducing conditions, Myc_{CFL} was seen to directly interact with hOGG1 as it was pulled down (Fig. 5.11a and b). Interactions were also observed for the hOGG1-LSD1 samples when 6x His-tagged hOGG1 was pulled down using Ni-IDA beads. Myc- LSD1 interaction was also observed when GST-tagged Myc was pulled down using GST beads corroborating previous reports¹⁵⁰. These experiments show for the first time that hOGG1 and Myc as well as hOGG1 and LSD1 directly interact with each other to form complexes. In addition, similar complex formation was seen both under oxidizing and reducing environments.

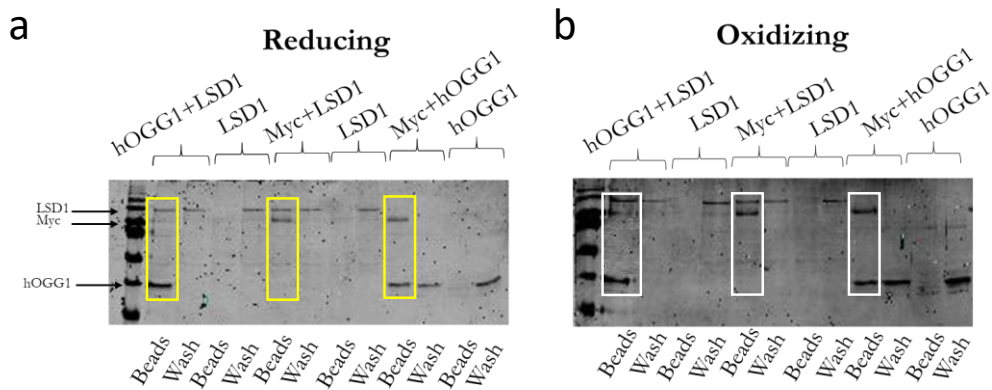


Figure 5.11: In vitro pull-down Assays to demonstrate direct protein-protein interactions under reducing (a) and oxidizing (b) conditions. Untagged LSD1 and hOGG1 were pulled down by either his tagged hOGG1 or Myc. The SDS-PAGE analysis of the pull-down assays revealed interactions of hOGG1 with Myc, hOGG1 with LSD1, and LSD1 with Myc under reducing as well as oxidative stress conditions. The yellow and white boxes indicate the lanes of pulled down species under reducing (a) and oxidizing (b) conditions, respectively. Duplicate and triplicate experiments are shown in the Appendix section (Fig. x5.9).

Myc_{FL} posed difficulties in purification and could not be purified successfully. C-terminal domain of Myc is necessary for interaction with Max and DNA binding. N-terminal domains are involved in the regulation of Myc and interaction with other proteins like Fbw7 E3 ligase and SPT5^{6,182}. Hence, Myc₁₋₁₆₃ construct is used in native agarose gel electrophoresis studies instead of Myc_{FL}. To directly visualize direct interactions of hOGG1 with Myc, Myc with LSD1, and hOGG1 with LSD1, proteins were incubated in equimolar ratios at room temperature for 30 min. Native agarose gel electrophoresis assays showed shifts for each of these mixed protein samples, relative to lanes containing the single proteins (Fig. 5.12). Furthermore, an additional shift was observed for samples containing all three components (Fig. 5.12, lane 7) suggesting that LSD1, hOGG1, and Myc form a triple complex. All interactions were reproduced for oxidizing and reducing environments and for the wild-type and the catalytically inactive hOGG1 mutant (hOGG1_{K249Q}) (refer appendix, Fig. x5.10) showing that hOGG1 irrespective of its repair activity is able to interact with the other two protein components in this system. Discrepancies in gel shifts was observed due to different running behavior of gels caused by different chambers used. Further optimization is required to obtain consistent results.

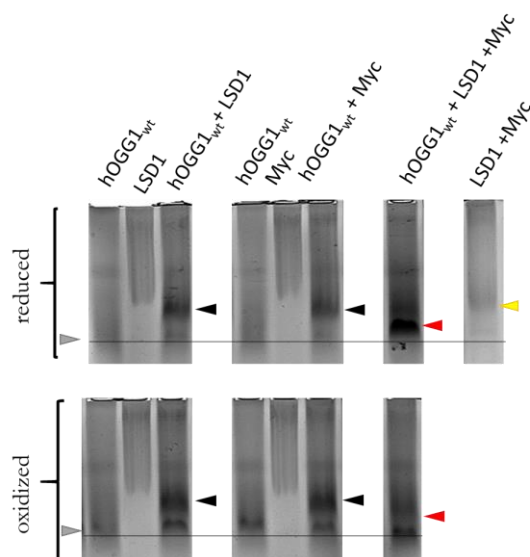


Figure 5.12: Native agarose gel electrophoresis (NAGE) experiments demonstrating direct protein-protein interactions by hOGG1_{wt} with Myc₁₋₁₆₃ and LSD1. A shift is observed in the presence of hOGG1_{wt} with either Myc₁₋₁₆₃ or LSD1 under both oxidizing and reducing conditions relative to the lanes with only hOGG1 present. Individual proteins were loaded alongside as controls. As a control, the previously reported interaction between LSD1 and Myc₁₋₁₆₃ was also tested. Black arrows represent the shift for dual protein complexes of hOGG1 with other two components, the grey arrow shows the band height of hOGG1 alone, red arrow indicates the triple complex and yellow arrow indicates the height of Myc-LSD1 complex. Black line drawn shows the differences between the heights of hOGG1 alone and other dual and triple complexes.

To quantify the qualitative results obtained from NAGE assays, fluorescence polarization experiments were performed. Novel interactions of hOGG1 with Myc and hOGG1 and LSD1 were tested for binding affinities in these experiments. For these measurements, hOGG1 was labelled with AlexaFluor 488 (AF488) and Myc₁₋₁₆₃ construct was used. Change in the polarization of the fluorophore on hOGG1 was measured in these experiments. The measurements were made in triplicate. The polarization values obtained were plotted against increasing concentration of the ligand (Myc or LSD1). K_D values were obtained by fitting these curves using the Hill equation in *Origin Pro*. Exemplary curves are shown in the Appendix section (Fig. x5.11 and Table. x5.2). Fig. 5.13 shows the K_D values for each interaction under the two distinct redox conditions. hOGG1 shows an increased affinity for Myc under oxidizing conditions (~ 700 nM versus ~ 1.1 μ M under reducing conditions). In contrast, the affinity between hOGG1 and LSD1 is slightly, but significantly decreased under oxidative stress (~ 860 nM versus ~ 750 nM under reducing conditions). These data support enhanced recruitment of Myc by hOGG1 under oxidative stress conditions, and suggest that under these conditions, an interaction between hOGG1 and LSD1 would also be weakened, potentially leading to dissociation of hOGG1 from LSD1.

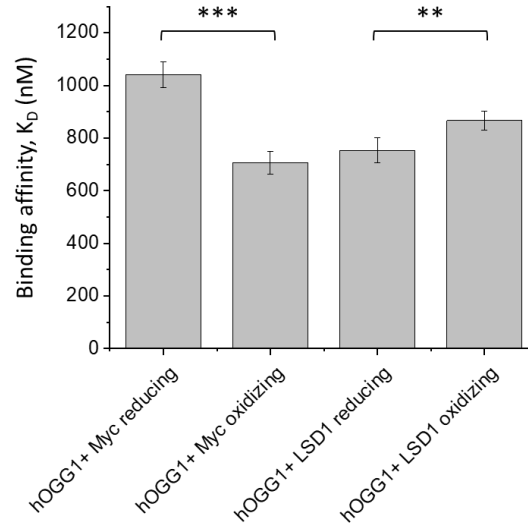


Fig 5.13: Bar plot showing affinities for interactions between hOGG1 and Myc or LSD1 under reducing and oxidizing conditions. Under oxidizing conditions, the interactions of hOGG1 and Myc and hOGG1 and LSD1 show K_D values of 706 ± 43 nM and 867 ± 34 nM, respectively. Under reducing conditions, K_D values for hOGG1 with Myc and LSD1 interactions are 1065 ± 69 nM and 754 ± 47 nM, respectively. Significance values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, n.s- non-significant. p -values were obtained by performing a two-tailed student t-test on the pooled data from triplicate experiments. p -value for hOGG1 with Myc sample under oxidizing and reducing condition was 0.000688 and similarly for hOGG1 with LSD1 sample was 0.01539. Exemplary polarization curves and binding affinities from individual experiments of the triplicate are shown in the Appendix (Fig. x5.11 and Table x5.2).

To investigate recruitment of Myc to DNA by hOGG1, EMSA was performed for hOGG1 and Myc in the presence of oxoG lesion containing DNA. The DNA was fluorescently (Cy3) labelled for these experiments. Experiments were performed with both hOGG1_{wt} and hOGG1_{K249Q}. Gel shifts clearly showed different heights for Myc-DNA, hOGG1-DNA, and hOGG1-Myc-DNA complexes, indicating recruitment of Myc to DNA by hOGG1 (Fig 5.14). hOGG1-Myc-complexes were shifted in the case of hOGG1_{wt} and hOGG1_{K249Q} indicating both hOGG1 monomer and dimer/ higher order oligomer are capable of interactions with Myc (Fig. 5.14). This was observed to be true under both oxidizing and reducing conditions.

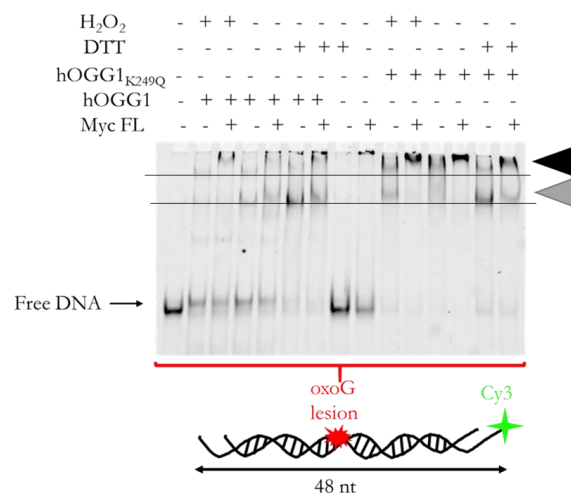
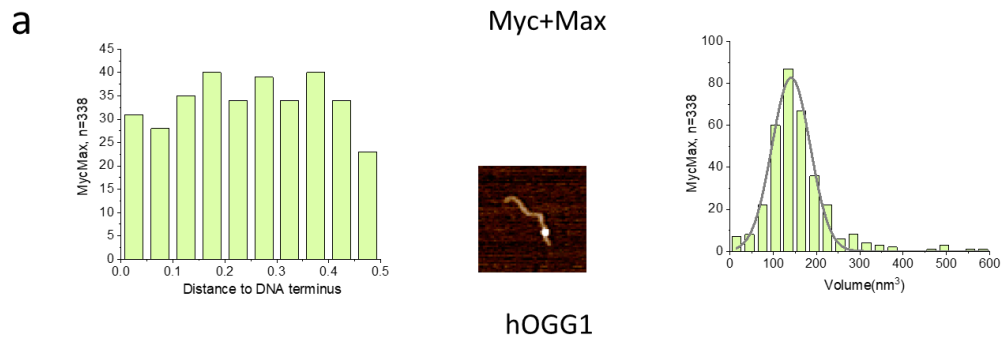
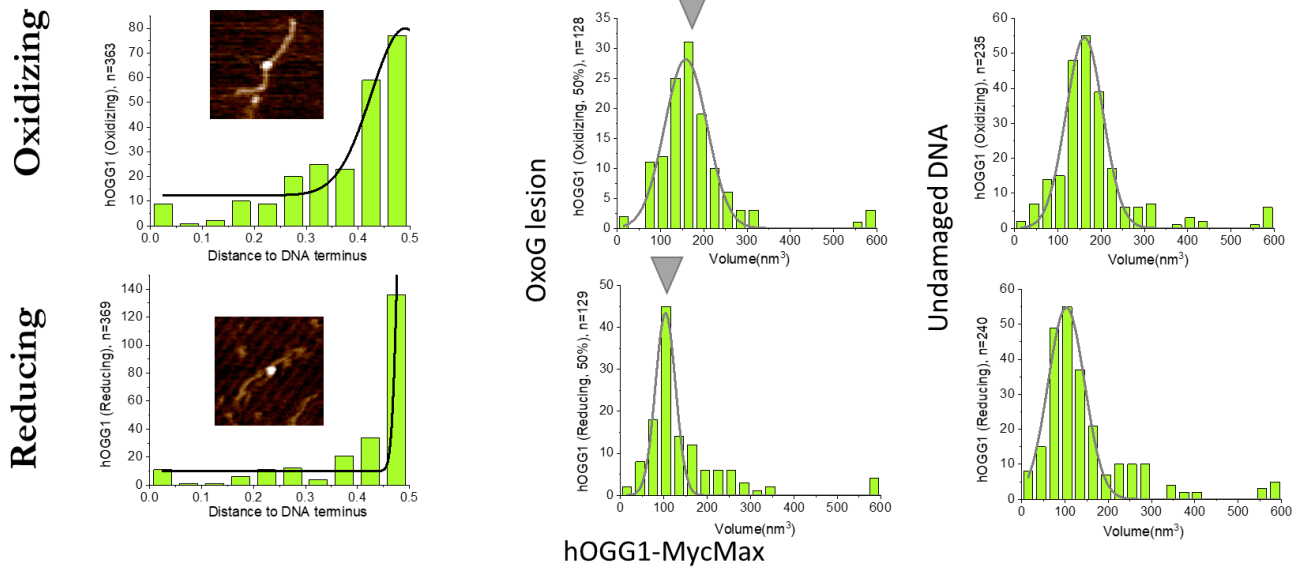


Figure 5.14: EMSA studies of hOGG1 interaction with Myc on DNA containing an oxoG lesion. A shift is clearly seen when Myc is incubated with hOGG1- oxoG containing DNA complex under oxidative as well as reducing conditions, for both hOGG1_{wt} and catalytically inactive hOGG1_{K249Q}. The shifts for monomeric hOGG1-Myc-DNA complexes are indicated by grey arrow and black arrow shows the shift for dimeric/ higher order oligomeric hOGG1-Myc-DNA complexes.

To visualize these complexes on DNA and to confirm if the higher order oligomer band seen in EMSA is indeed a dimer species (black arrow, Fig. 5.14), AFM was applied to study interactions of hOGG1 and Myc on DNA containing an oxoG lesion (at 50% of DNA substrate length) under oxidizing and reducing conditions. Incubations were carried out in the presence of Max, the interaction partner of Myc to provide native conditions. Volume analyses of the DNA bound protein complexes showed higher volumes for incubations with both hOGG1 and Myc compared to control experiments with the individual proteins (only hOGG1 or only Myc) (Fig. 5.15). This increase in volume was seen in both oxidizing and reducing settings. However, due to the monomeric state of hOGG1 under reducing conditions, the shift to higher volumes was more evident under reducing conditions. These experiments were also carried out with the catalytically inactive mutant of hOGG1 (hOGG1_{K249Q}). Similar results were obtained indicating interactions between hOGG1_{K249Q} and MycMax complexes under both conditions (Appendix, Fig. x5.13). Specificities for proteins on oxoG containing DNA under different conditions were, hOGG1_{wt}: ~323 (oxidizing), ~785 (reducing), hOGG1_{K249Q}: ~276 (oxidizing), ~560 (reducing), hOGG1_{wt}-MycMax: ~655 (oxidizing), ~481 (reducing) and hOGG1_{K249Q}-MycMax: ~362 (oxidizing), ~498 (reducing) (See Fig. 5.15 and appendix Fig. x5.13).



b



c

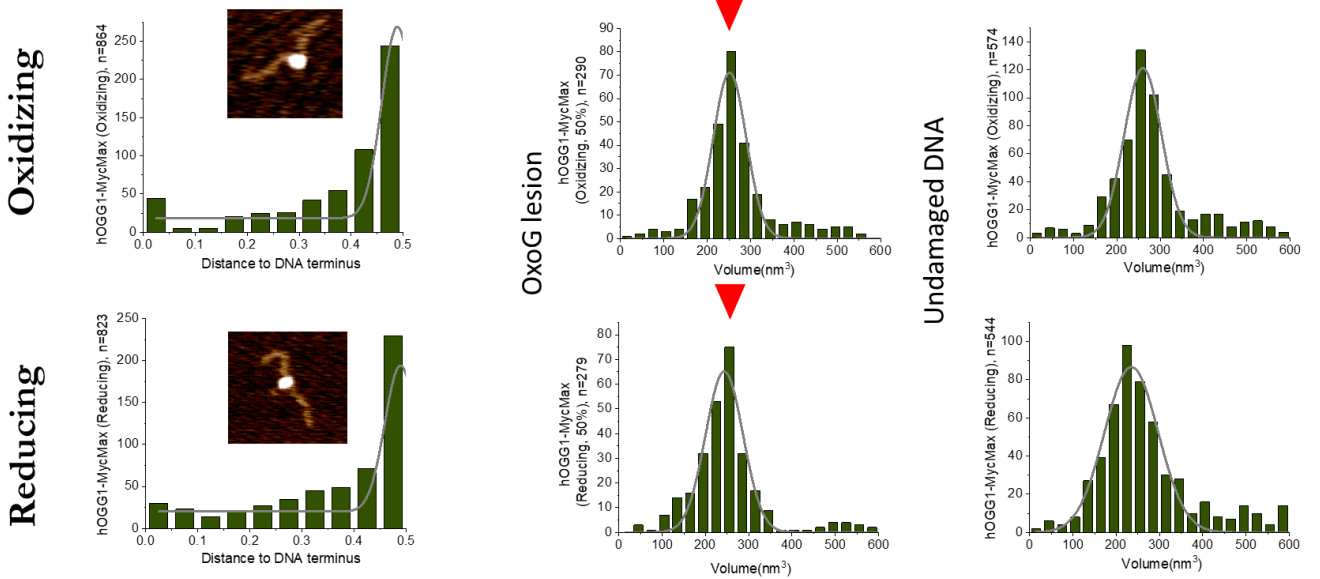


Figure 5.15: AFM studies of hOGG1 and Myc/Max interaction on DNA. (a) Position and volume distributions for Myc-Max on oxoG lesion containing DNA. A Gaussian fit to the volume distribution shows a volume of $\sim 140 \text{ nm}^3$, consistent with a Myc-Max heterodimer. The volumes of the free (non-DNA-bound) Myc ($\sim 48 \text{ kDa}$) and Max ($\sim 18 \text{ kDa}$) proteins were $63 \pm 12 \text{ nm}^3$ and $31 \pm 7 \text{ nm}^3$, respectively (refer to Appendix Fig. x5.12, experiment 3). Myc-Max shows no binding preference for oxoG (at 50% position). The DNA substrate does not contain an E-box for Myc/Max binding. (b) Position (left) and volume distributions for hOGG1 on oxoG DNA (with oxoG at 50% of the DNA length). The middle panel shows hOGG1 volumes for complexes bound at the oxoG lesion, the right panel shows volumes on undamaged DNA sites. The top and bottom rows represent experiments under oxidizing and reducing conditions, respectively. hOGG1 showed volumes of $\sim 160 \text{ nm}^3$ under oxidizing and $\sim 100 \text{ nm}^3$ under reducing conditions (indicated by grey arrows). (c) Position (left) and volume distributions for hOGG1+Myc+Max complexes. The middle column shows complex volumes at oxoG lesion positions, the right panel at undamaged DNA sites. Top and bottom row again represent oxidizing and reducing conditions, respectively. The volumes for hOGG1+Myc+Max complexes are shifted to higher volumes compared to those of Myc+Max alone (a) or hOGG1 alone (b), ranging from $\sim 200 \text{ nm}^3$ to $\sim 300 \text{ nm}^3$ (red arrows) at oxoG lesions and undamaged sites under both oxidizing and reducing conditions, suggesting interaction between hOGG1 and Myc/Max. Data have been pooled from triplicate experiments. Individual experiments are shown in the Appendix section (Fig. x5.12).

AFM analyses were also applied to investigate hOGG1 interaction with LSD1 on DNA. Volume analyses on these crosslinked samples in the presence of oxoG lesion containing DNA also showed interaction between hOGG1 and LSD1 (Appendix, Fig. x5.14).

To examine enhancement of Myc-Max binding to E-boxes in the presence of hOGG1 upstream of the E-box, AFM was applied to samples of hOGG1, Max, and Myc conjugated with quantum dots 605 (QD, see Methods chapter 2.2.3.2 for details of the conjugation procedure). The DNA substrate in these experiments contained an oxoG lesion 8 bp upstream of an E-box motif at the 50% position (oxoG at 49.8% and the E-box at 51.6% DNA length, Fig. 5.16a). The exact positions will not be resolvable within the resolution of the AFM experiments and will both appear as near 50% of DNA length. However, the distinct QD label on Myc allows to pin-point enhancement of its recruitment along with Max at E-box (50% position) by hOGG1. In addition, Myc-Max in the previous experiments did not show any preference towards oxoG lesion (Fig. 5.15a). This indicates that in the present experiments Myc-Max complexes are likely bound to E-box and not oxoG lesion at 50% position. Consistent with the requirement for Max^{5,162}, (QD-)Myc alone did not show stable binding to the E-box (Fig. 5.16b). In the presence of Max, (QD-)Myc bound to the Ebox with moderate specificity (183 ± 43 (reducing) and 245 ± 32 (oxidizing)). AFM position distributions of QD containing complexes on DNA in the presence of hOGG1 distinctly showed enhanced (QD-)Myc-Max binding at the lesion/E-box position (at $\sim 50\%$ DNA length), particularly under oxidizing conditions (Fig. 5.16c, $S = 407 \pm 40$). Under reducing conditions, a moderate enhancement of (QD-)Myc at the lesion/E-box was also seen (Fig. 5.16d, $S = 255 \pm 46$). Experiments were performed in triplicate (Appendix, Fig. x5.15). These results support stimulation of Myc binding to its recognition sequence in DNA by hOGG1 when the cell is under oxidative attack.

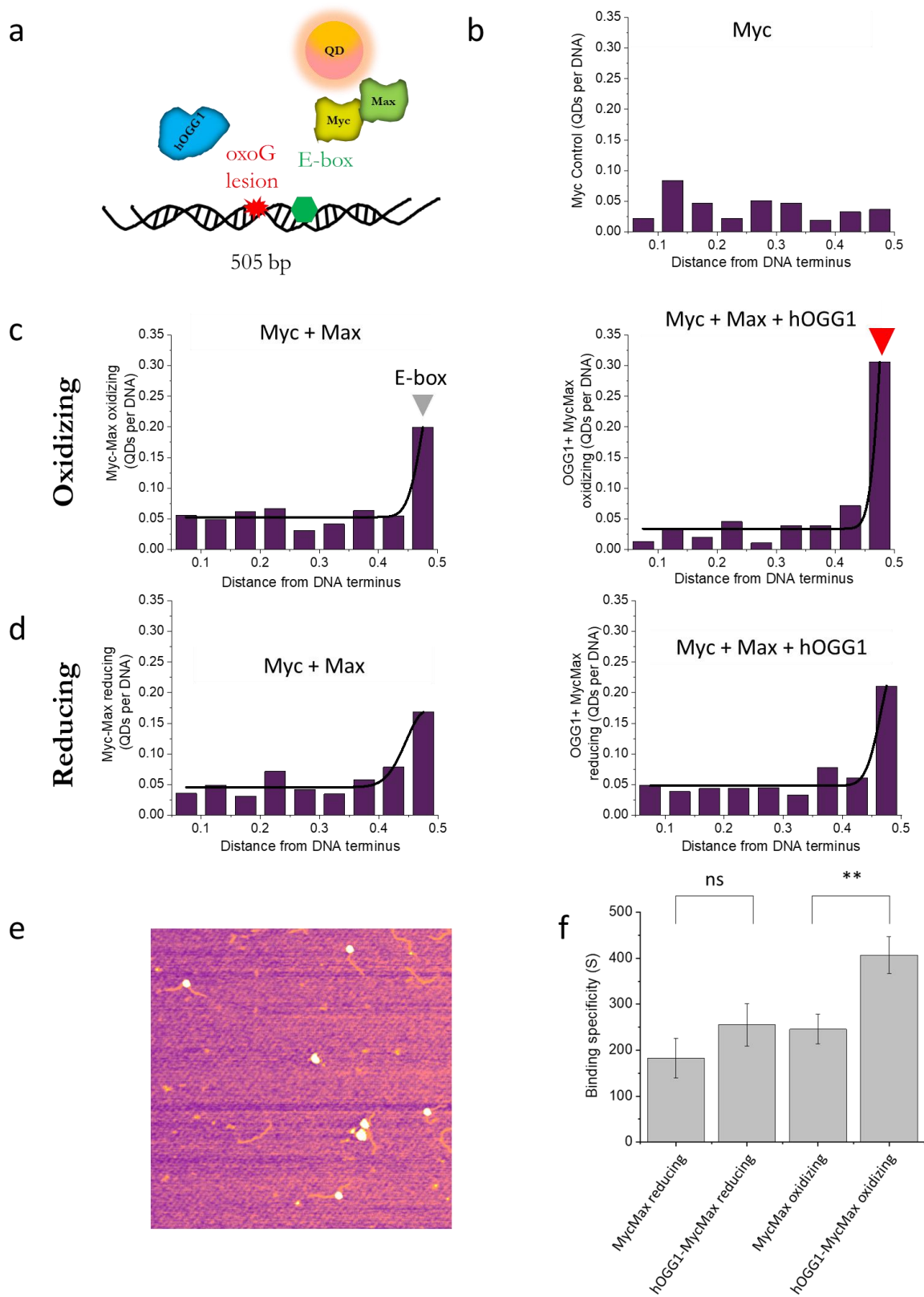


Figure 5.16: Quantum Dot- AFM experiments to investigate enhanced binding of Myc-Max to E-box in the presence of hOGG1. (a) The schematic depicts the principle of the experiment. Myc is conjugated to a quantum dot and then incubated with Max, hOGG1 and oxoG- Ebox containing DNA substrate. (b) Position distribution of the Myc-only control shows no binding specificity to E-box in the absence of Max. Positional analyses for QD-Myc-Max (left panels) and QD-Myc-Max-hOGG1 (right panels) complexes under oxidizing (c) and reducing (d) conditions. Particularly under oxidizing conditions, hOGG1 binding to oxoG lesion upstream of the Ebox motif enhances Myc-Max recruitment at Ebox. Binding specificities were obtained from the Gaussian fits to the position distributions. Specificity values are calculated for each experiment of the triplicate (refer to appendix Fig. x5.15). The distributions shown here are from pooled data. (e) An exemplary AFM image displaying QD-Myc-Max complexes with hOGG1 at 50% position of the entire DNA substrate. (f) Average specificities (\pm SD) calculated for individual samples were: 183 ± 43 (reducing) and 245 ± 32 (oxidizing) for Myc/Max and 255 ± 46 (reducing) and 407 ± 40 (oxidizing) for hOGG1+Myc/Max. Average binding specificities (S) (\pm SD) for each condition from individual experiments. (Significance values: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, n.s- non-significant). P-values were obtained by performing two-tailed student t-test from the triplicate experiments. p-value for Myc-Max sample and hOGG1-MycMax sample under oxidizing conditions was 0.02 and under reducing conditions, it was 0.13. Specificities for hOGG1 and hOGG1-MycMax complexes are indicated by grey and red arrows respectively.

To test the scenario in which transcription factor recruitment takes place due to the DNA conformational changes induced by hOGG1, DNA bend angle analyses were performed (using the MatLab tool introduced in Chapter 3). hOGG1 and hOGG1-Myc complexes bound to oxoG lesion containing DNA revealed differences in bend angle states between these different types of complexes. These experiments were performed under non-reducing (oxidative) conditions. Similar to hOGG1 bound to undamaged DNA (described in Chapter 3), hOGG1 shows 0° , $\sim 33^\circ$ and $\sim 69^\circ$ bend angle states also at oxoG lesions (Fig. 5.17a). Interestingly, the intermediate species with bending by $\sim 30^\circ$ is strongly enhanced at the lesion compared to on undamaged DNA (compare panels (a) and (b) in Fig. 5.17), suggesting a role of this conformation in lesion recognition. hOGG1-Myc complexes present at oxoG also showed a major bend angle state with similar bending ($\sim 23^\circ$, Fig. 5.17c) and a pronounced reduction of the $\sim 70^\circ$ state of the hOGG1 IC complex. On undamaged DNA, hOGG1-Myc complex conformations were characterized by bend angles similar to those of hOGG1 alone at an oxoG lesion (0° , $\sim 37^\circ$ and $\sim 71^\circ$, Fig. 5.17d). Myc-Max complexes showed DNA bending of $\sim 12^\circ$ at non-specific sites (Fig 5.17e). Interestingly, this bending differs drastically from that previously reported for Myc-Max bound at their E-box target site ($\sim 80^\circ$)¹⁸³. Importantly, these data point towards conformational changes in the DNA that may be responsible for enhanced binding of Myc at the target sites.

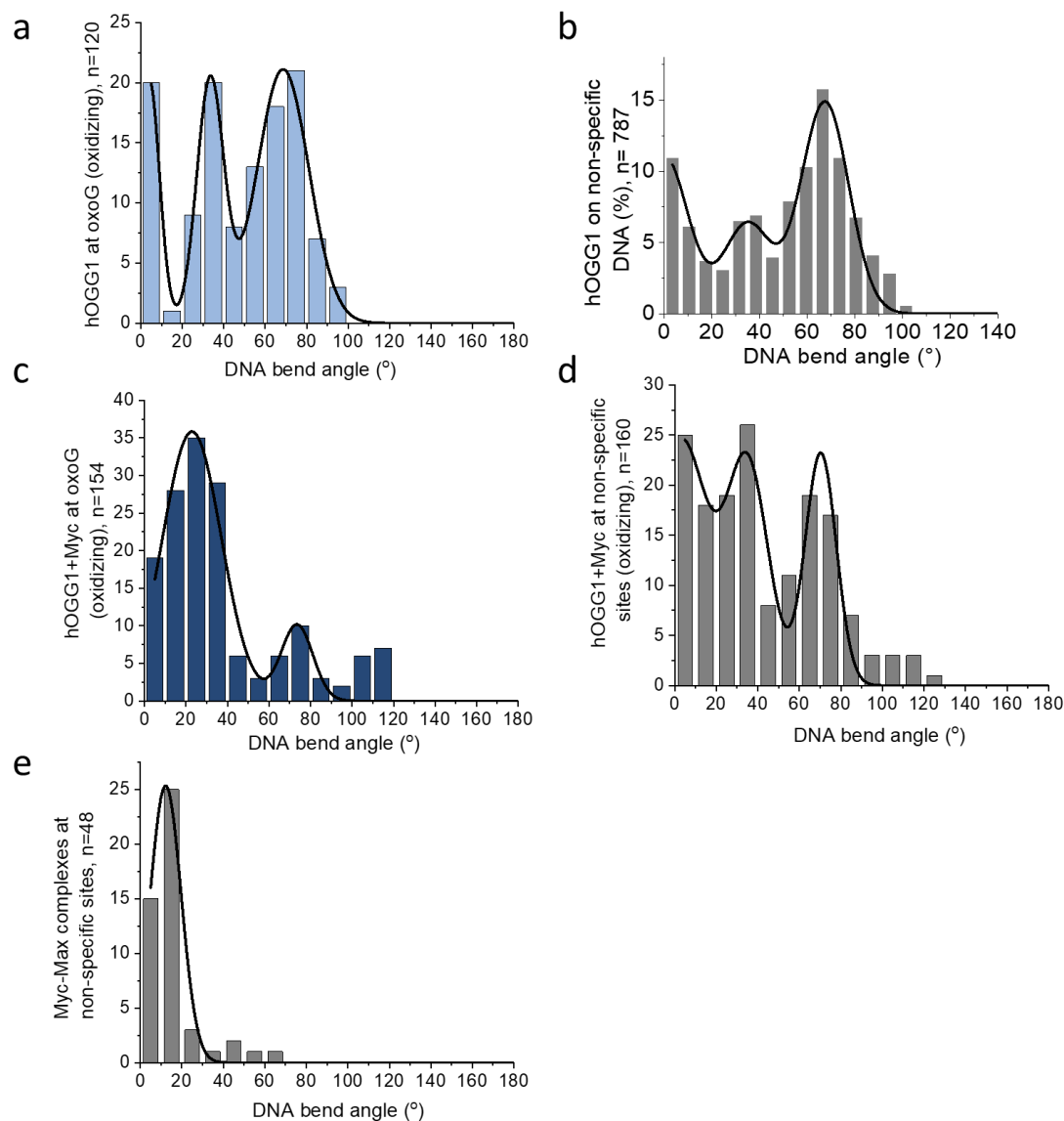


Figure 5.17: AFM bend angle analysis of hOGG1 and hOGG1-Myc complexes at oxoG lesion. (a) A multimodal Gaussian fit to the DNA bend angle distribution for hOGG1 at oxoG lesion revealed 0° , $\sim 33^\circ$ and $\sim 69^\circ$ bend angle states. (b) Gaussian fits to the DNA bend angle distribution of hOGG1 on undamaged DNA revealed three states with bend angles 0° , $\sim 35^\circ$ and $\sim 70^\circ$ (Data from section 3.1)¹⁷ as also observed with the oxoG lesion containing substrate (a) but with different population of states. (c) Bend angle distribution for hOGG1-Myc complexes at oxoG lesion exhibit $\sim 23^\circ$ and a minor species with $\sim 74^\circ$ DNA bending. (d) Gaussian fits to the DNA bend angle distribution of hOGG1-Myc complexes at undamaged sites revealed three states with bend angles 0° , $\sim 37^\circ$ and $\sim 71^\circ$ suggesting that binding of Myc to hOGG1 bound at non-specific sites does not influence the conformation of the hOGG1-DNA complex. (e) Gaussian fit to the DNA bend angle distribution of Myc-Max complexes at non-specific sites showed a bend angle of $\sim 12^\circ$. The bend angle distributions shown here are from pooled data (refer to Appendix Fig. x5.16 for the individual experiments).

5.3 Discussion

One of the most frequent and cardinal DNA lesions, 7,8-dihydro-8-oxoguanine (8-oxo-G) is caused by oxidative stress to the cells¹⁸⁴ and has been implicated in the malfunctioning of several physiological processes, ageing¹⁴¹, carcinogenesis¹⁸⁵, and some of the neurodegenerative diseases¹⁸⁶. It is a major underlying basis for mutagenesis and instability in the genome¹⁸⁷. Recognition and excision of 8-oxo-G lesions is performed by DNA glycosylases. These DNA glycosylases are the initiating enzymes in the base excision repair (BER) pathway^{185,30,64}. One of the main DNA glycosylases to repair 8-oxo-G lesions in humans is the 8-oxoguanine glycosylase 1 (hOGG1). Several studies in recent times have shed light on additional non-canonical activity of hOGG1 in transcription regulation of genes engaged in, for example, the innate immune response¹⁸⁸, metabolic processes^{189,190}, and apoptosis^{191,192}. It has been shown for transcription factor NFκB, that hOGG1 alters interactions with its cognate binding motif¹⁵² and that hOGG1 binding in promoter regions leads to the increased transcription of NFκB target genes¹⁵². Earlier chromatin immunoprecipitation (ChIP) studies and quantitative RT-PCR experiments have presented evidence of involvement of hOGG1 in modulation of transcription factors including Myc^{150,193}.

Another aspect of oxidative exposure of cells is the damage to other macromolecules such as proteins and RNA apart from DNA. In a previous report, it was demonstrated that 8-oxo-G glycosylase activity of hOGG1 was inhibited indirectly and reversibly *in vivo* by exposing it to toxic concentrations of cadmium¹⁷⁵. From this study and other reviews in the past¹⁶⁵, it has been hypothesized that this abolition of glycosylase activity is a consequence of cysteine modifications in hOGG1. Importantly, it has been hypothesized that these cysteine modifications are reversed to reduced state once the redox conditions in the cell are re-established¹⁷⁵. In this context, it has further been proposed that during oxidative stress, formation of 8-oxo-G lesions is the primary effect while hOGG1 cysteine modification is a secondary effect and that these modifications lead to the inactivity of hOGG1¹⁶⁵.

The SDS-PAGE and AFM experiments presented in this Chapter clearly show that hOGG1 exists in two oligomeric forms, as a monomer and a dimer, and that the equilibrium between these states depends on the redox environment. Importantly, the oxidized dimeric hOGG1 could be reversed back to the monomeric form in the presence of DTT. Interestingly, the patient-derived hOGG1 Ser326Cys mutant has previously been shown to be dimeric, dimerization coming from the additional, incorrect cysteine residue^{194,195}. This mutant was also shown to have reduced glycosylase activity compared to hOGG1 wt. The evidence presented here for dimerization of wild type hOGG1 under oxidative conditions is, however, completely novel. Sequencing results (Appendix, Fig. x5.2) confirmed that the wild type hOGG1 cloned in these studies did not contain any mutations. However, hOGG1 dimerization by oxidation requires further investigation to understand its molecular details.

Since the Ser326Cys mutant dimer (above) was found to be less active in oxoG repair, it was surprising that the results from activity assays on wild type (WT) hOGG1, in this thesis, demonstrated similar enzyme activity under oxidative (dimeric hOGG1) and reducing (monomeric hOGG1) under equilibrium conditions (after ≥ 10 -minute incubation). However, AFM data showed reduced recognition of oxoG lesions by hOGG1 under oxidative compared to reducing conditions. Consistent with these results, fluorescence polarization studies showed weaker binding of hOGG1 to oxoG containing DNA (but not to undamaged DNA) under oxidizing compared to reducing conditions. These data suggest that the reduced, monomeric form of hOGG1 indeed binds to oxoG lesions more stably than the oxidized, dimeric form.

Myc is known to govern transcription of a wide variety of genes involved in apoptosis, cell proliferation, metabolism and tumorigenesis¹⁵⁷. In particular, Myc is also involved in transcription of many pro-inflammatory genes which in turn involves oxidative stress in the cell¹⁵⁰. Myc in complex with its binding partner Max complex recognizes and binds the CACGTG consensus sequence called E-box motif¹⁶⁰⁻¹⁶². Binding of Myc to E-boxes is helped by demethylation of histone H3K4me3 to open the chromatin. This process is achieved by recruiting a Lysine Specific Demethylase, LSD1 (also known as KDM1). LSD1 has been shown to be involved in the transcription regulation of androgen and estrogen receptor target genes¹⁷⁴. Interaction between Myc and LSD1 has also been demonstrated in the context of activation of the Myc target genes *Ncl* and *Cad* genes¹⁵⁰. It has been shown by quantitative ChIP studies that hOGG1, LSD1 and Myc were all present on the same chromatin regions¹⁵⁰. Upon LSD1 silencing, transcription of Myc target genes was highly attenuated. LSD1 contains a Flavin Adenine Dinucleotide (FAD) molecule, which when reduced during histone demethylation activity leads to the formation of hydrogen peroxide (H₂O₂). Hence H₂O₂ is generated during demethylation and is believed to lead to local oxidation in the vicinity of upstream promoter regions of Myc target genes resulting in the local formation of 8-oxo-guanine (oxoG) lesions, the substrate of hOGG1. From these studies, it had been hypothesized that LSD1 may recruit hOGG1 to the lesion in the proximity of E-box motifs, and that hOGG1 may then serve to enhance Myc binding to the E-box motif.

This thesis aimed to investigate the interactions between LSD1, hOGG1, and Myc, and to resolve the mechanism by which Myc may be recruited to its target E-box by hOGG1. Protein-protein interactions in this system were studied by a range of methods. NAGE and *in vitro* pull-down experiments clearly showed that hOGG1 interacts with Myc (in the absence of DNA) regardless of reducing or oxidative environment. These data suggest that hOGG1 and Myc transcription factor can pre-form a complex that may then bind on the DNA. Additionally, hOGG1 and LSD1 also showed interactions (also in the absence of DNA), and finally, all three proteins showed triple-complex formation, suggesting all three proteins, hOGG1, LSD1 and Myc may be recruited together to the DNA. EMSA experiments demonstrated that both hOGG1 wild type and the catalytically inactive hOGG1 interact with Myc on oxoG containing DNA under both oxidizing and reducing

conditions. In fact, complex formation under oxidative conditions for hOGG1_{wt} and for inactive hOGG1_{K249Q} variant was enhanced. Furthermore, fluorescence polarization studies demonstrated enhanced affinity of hOGG1-Myc complexes under oxidizing ($K_D=706\pm 43$ nM) compared to reducing conditions ($K_D=1065\pm 69$ nM), supporting Myc recruitment by hOGG1 favourably under oxidative environment. The interaction between hOGG1 and LSD1 remained largely same under both oxidative and reductive environments although a marginally weaker affinity was observed under oxidative environment. This decrease in affinity between LSD1 and hOGG1 ($K_D=867\pm 34$ nM) may ease dissociation of hOGG1 from LSD1, so that after LSD1 demethylation activity, LSD1 may thus dissociate from the complex to release hOGG1-Myc(-Max) to bind to lesions in DNA upstream of a Myc target gene.

Specificities were higher for hOGG1-MycMax samples, particularly under oxidative conditions for both hOGG1_{wt} and hOGG1_{K249Q} mutant (Fig. 5.15 and appendix Fig. x5.13). This suggests that hOGG1 activity is not necessary for enhanced recruitment of Myc. QD-AFM data further strongly supported increased Myc-Max binding to the E-box motif in the presence of oxidized OGG1 and an oxoG lesion upstream close to the E-box (Fig. 5.16). These results support enhanced recruitment of Myc to its target genes by hOGG1 under oxidative conditions *in vivo*.

Towards the aim of understanding the enhanced transcription factor recruitment brought about by the DNA conformational changes induced by hOGG1, DNA bend angles were measured from AFM data of hOGG1 and hOGG1-Myc complexes bound at oxoG lesions and on undamaged DNA using the automated MatLab tool (see chapter 3). These analyses demonstrated distinct bend angle states at oxoG lesions when hOGG1 interacts with Myc (different to those of hOGG1 or Myc alone). Furthermore, DNA bend angles in hOGG1-Myc complexes at non-specific DNA sites were comparable to those of hOGG1 alone. These results indicate that oxoG-lesion bound hOGG1 might be necessary for inducing conformational changes in the DNA leading to enhanced transcription factor binding and gene expression.

In summary, in this thesis, I have demonstrated the effects of oxidative modifications in DNA (8-oxo-G) and of oxidative modifications of hOGG1 on Myc-mediated transcription by several *in vitro* biochemical and biophysical methods. In the introduction, I presented several possible scenarios of how hOGG1 may aid enhancement of Myc transcription factor binding and as a result increase transcription of its target genes. Based on the data presented in this chapter, a model for the mechanism of Myc recruitment to its target genes by hOGG1 evolves (Fig. 5.18). hOGG1, Myc-Max and LSD1 may arrive at the target genes as a complex together. LSD1 recognizes the methyl groups on lysine residue K4 on histones H3 and induces chromatin opening. Upon demethylation of H3K4 by LSD1, H₂O₂ is generated as a by-product as described¹⁵⁰. This leads to local oxidation of the DNA resulting in oxoG lesion formation and LSD1 dissociation from the hOGG1-MycMax complex. Many genes have GC rich sequences in the upstream promoter and regulatory

regions and since guanine bases have the lowest redox potential, they easily get oxidized to oxo-guanine. The regulatory regions contain the consensus sequences for transcription factors, here specifically, the E-box consensus sequence for Myc-Max heterodimer. Demethylation activity by LSD1 along with oxidizing the DNA may also oxidize hOGG1 which consequently may form a dimer that more loosely binds to the lesion. The more loosely bound complex may then dissociate from the oxoG lesion to scan the DNA in complex with Myc-Max. Together with the conformational changes observed at the oxoG lesion for the hOGG1-Myc complex under non-reducing (oxidizing) conditions, the more loosely bound dimer may then allow Myc-Max to recognize the E-box while still being bound to the hOGG1 dimer, and thus deliver Myc to its target site. The QD-AFM data that showed enhanced Myc-Max recruitment to the Ebox by hOGG1 under oxidative conditions are consistent with this model. The conclusions from the presented experiments I will now summarize in a model of hOGG1 induced enhancement of Myc transcription initiation under oxidative stress (Fig. 5.18).

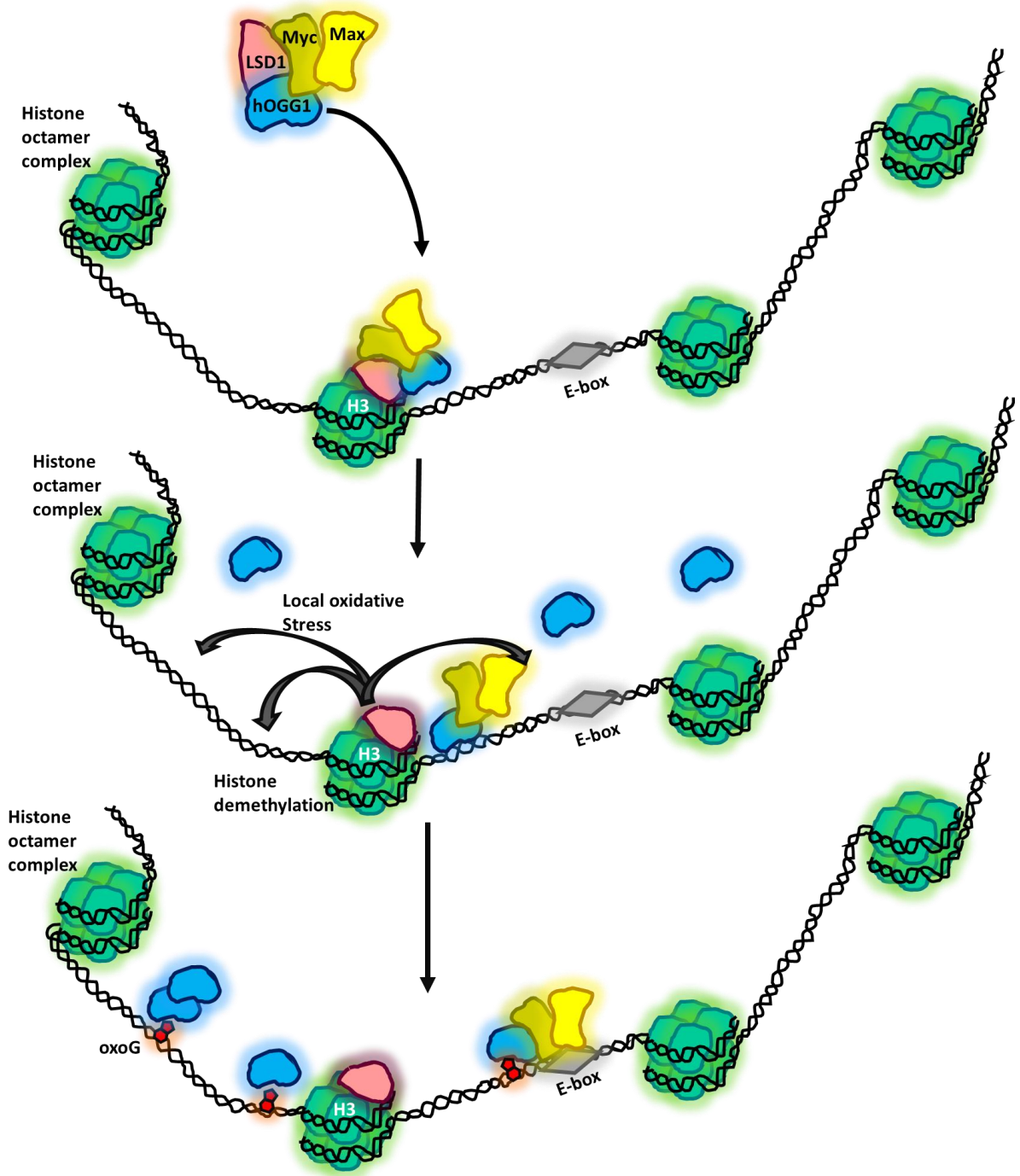


Figure 5.18: Schematic model showing the role of DNA oxidation and hOGG1 involvement in Myc-driven transcription. The proteins hOGG1 (blue), Myc-Max (dark and light yellow) and LSD1 (pink) arrive as a triple complex. The whole complex binds to the histone protein H3 (teal blue, other histones shown in green) via interaction mediated by LSD1. Upon H3K4 demethylation by LSD1, hydrogen peroxide is generated by its FAD reaction in the local environment (dark grey arrows), which leads to the formation of oxoG lesions (red) in the vicinity of promoter regions of Myc target genes. Oxidative environment further leads to the oxidation of hOGG1, which induces dissociation of hOGG1-Myc-Max complex from LSD1 and loading on to the oxoG lesions in the DNA. Conformational changes in the hOGG1-Myc-Max-DNA complex then lead to dissociation of the complex from the lesion and translocation downstream on the DNA, and facilitate loading of Myc-Max complex at the cognate E-box motif present downstream to the lesion.

5.4 Future perspectives

While several aspects of enhanced Myc recruitment to E-box have been tested in this thesis, scenarios involving oxoG lesion in the E-box motif with or without hOGG1 bound need to be investigated. These events are trivial and may only lead to transcription inhibition or negative gene regulation. To this end, EMSA and AFM studies could be performed with Myc-Max complex and a DNA substrate designed to contain oxoG lesion in the E-box motif in the presence or absence of hOGG1.

Another possibility is oxoG lesion formation in the G-rich promoter regions by activity of LSD1. To test this scenario involving G4 quadruplex formation induced by hOGG1 activity, a DNA substrate containing 5 continuous G runs could be designed such that it forms a G4 quadruplex containing a 5th spare G-track upstream of an E-box motif. An AP site could be incorporated in the 5th G run such that it is present in the spare loop region upon quadruplex formation (Fig. 5.19). Recruitment of Myc-Max to E-box in the presence of APE1 bound to this AP site (Fig. 5.1f) could be tested using AFM.

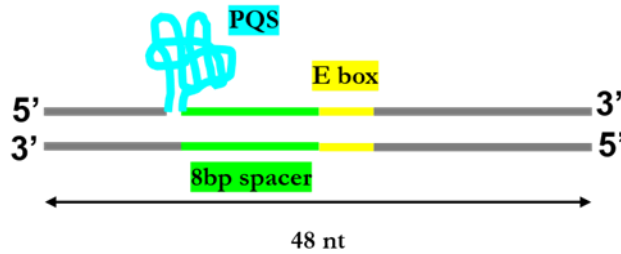


Figure 5.19: Schematic of G4 quadruplex and E-box containing oligonucleotide. A 48 nt long DNA oligonucleotide containing potential G4-quadruplex forming sequence (PQS, cyan) that eventually assembles into the tertiary structure (G4 quadruplex) and an E-box motif (yellow) 8 nt downstream of the PQS. This oligo could be used to generate 505 bp long AFM substrate using methods described in chapter 2.2.3.1.

6. Conclusions

Data obtained in this thesis provide a deeper insight into mechanistic processes of protein-DNA interactions involved in base excision repair (DNA glycosylases), transcription initiation by the specific DNA glycosylase hOGG1, and epigenetic modification (DNA methyltransferases- DNMTs). AFM imaging was the centrally used technique in this thesis to reveal changes in DNA bending and protein stoichiometries in these systems, and to study target site search and recognition by different protein systems. In addition, biochemical and biophysical experiments were performed to complement the data obtained from single-molecule AFM imaging.

AFM imaging in my studies has provided substantial detail on the target site search and recognition by structurally distinct DNA glycosylases. In this context, a novel, experimenter independent, MatLab-based, automated approach for DNA bending and positional analyses has been established, which further augments the power of AFM to study protein-DNA interactions at the single-molecule level. This method has been applied to the study of DNA interactions both by glycosylases and by a DNA methyltransferase. The MatLab-based tool boosts the analyses of minute distortions in the DNA substrates by base lesions or protein interactions particularly in the low bend angle regime. DNA bend angles obtained from the automated approach were able to support a model of initial lesion sensing strategies of DNA glycosylases based on altered mechanical properties of DNA at glycosylase target lesions. In addition, positional and DNA bend angle data obtained with the novel tool, combined with AFM volume analyses, provided novel insight into co-methylation strategies of DNMT3A DNA methyltransferase. The tool can also be applied to a range of protein-DNA systems with similar constraints and may thus benefit wider scientific community.

A major focus of my thesis was on the role of the specific DNA glycosylase hOGG1 in Myc-mediated transcription. AFM imaging demonstrated enhanced recruitment of Myc to its target site, the E-box by hOGG1 in the presence of an oxoG lesion upstream of the E-box motif, particularly under oxidizing conditions. AFM volume analyses of the DNA bound protein peaks showed direct interactions between hOGG1 and Myc, as also supported by additional biochemical studies. AFM bend angle analyses revealed conformational changes between hOGG1 and hOGG1-Myc complexes at oxoG lesions, which may aid in the loading of Myc on its downstream E-box binding motif. In addition, biochemical and biophysical assays as well as AFM demonstrated direct protein-protein interactions between hOGG1 and LSD1, a histone demethylase that generates local oxidative stress in the cells and showed increased affinity between hOGG1 and Myc and decreased affinities between hOGG1 and LSD1 under oxidizing conditions. From these data, model of the role of hOGG1 in Myc-induced gene transcription has been developed.

Overall, this thesis provides a comprehensive study of structural and functional characterization of a wide range of protein-DNA systems involved in DNA repair, transcription, and DNA modification by combining high resolution, single-molecule AFM imaging analyses with several biochemical and biophysical methods.

Appendix

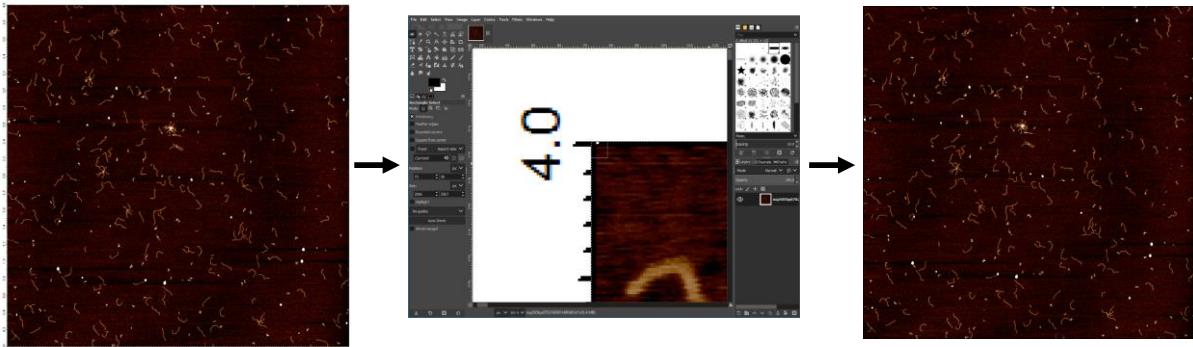
Appendix I

Figure x3.1: Step-By-Step guide for automated bend angle analysis at 50% lesion site.

Step 1: Produce .tif image

a) Image DNA with specific target lesion introduced at 50% DNA length by AFM. Process raw AFM image (plane fit and flatten to 3rd order) in AFM format (.ibw for MFP-3D/Igor-Pro, pixel resolution 1.95 nm/pixel). Export to .tif format (pixel resolution 1.95 nm/pixel) in AFM software.

b) Crop image in .tif format (remove frame and axes) in GIMP software.

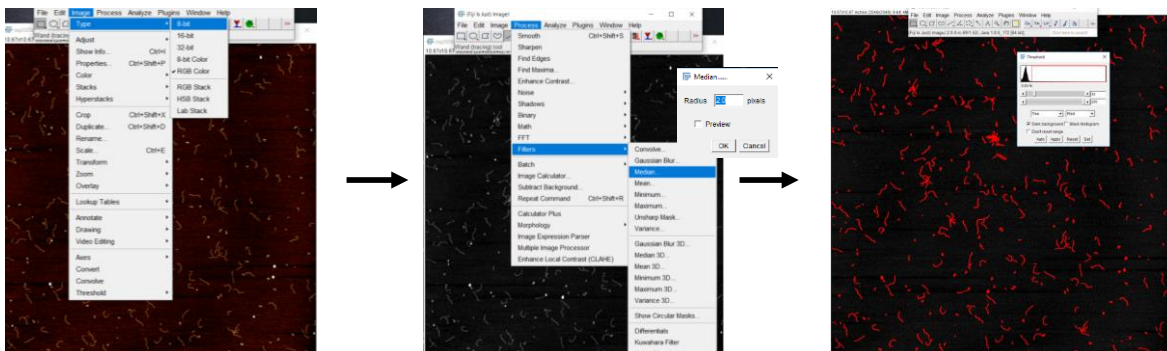


Step 2: Create DNA-only image by thresholding

a) Upload the cropped image to ImageJ software and change the format from RGB color to 8-bit (under Image> Type).

b) Choose Median filter (under Process) and set radius to 2 pixels. Press OK.

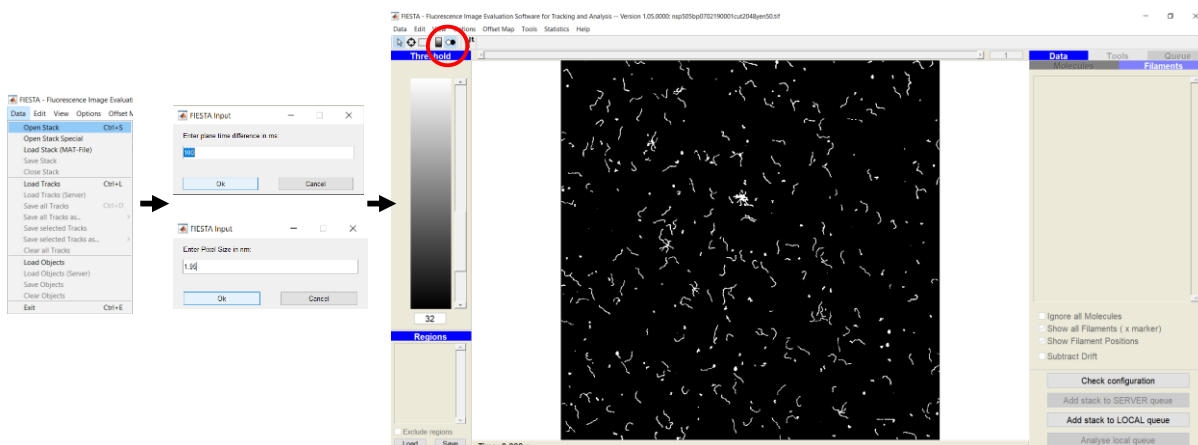
c) Choose Image > Adjust > Threshold and adjust Yen threshold to select DNA (red) over the background. Close threshold tab and save the image in .tif format. Note threshold value (here 32) for further processing.



Step 3: Track DNA filaments in FIESTA I : thresholding

a) Open saved image from ImageJ in FIESTA tracking software by choosing "Open Stack" (under Data) and enter plane time difference 100 ms and pixel size (here 1.95 nm)

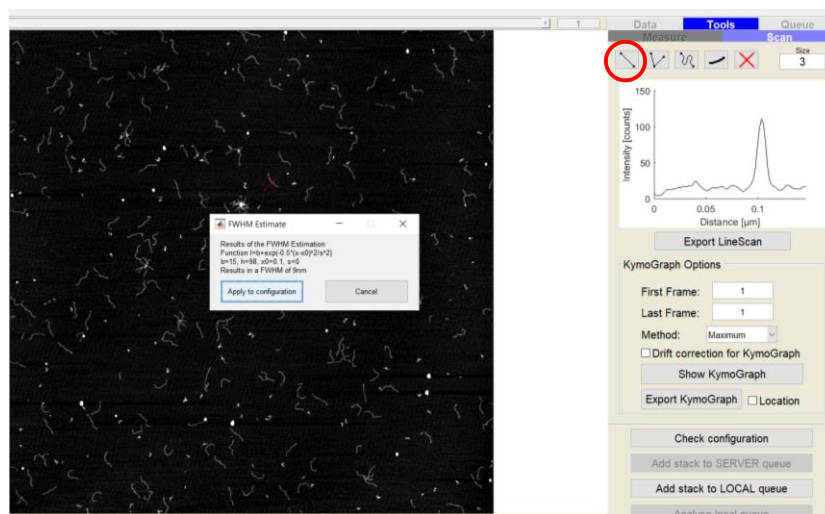
b) Change from normal to threshold mode (in Data: top left corner, circled in red). Enter the threshold value from ImageJ in left panel as shown here. Switch back to grayscale mode.



Step 4: Track DNA filaments in FIESTA II : full width at half maximum (FWHM) estimate

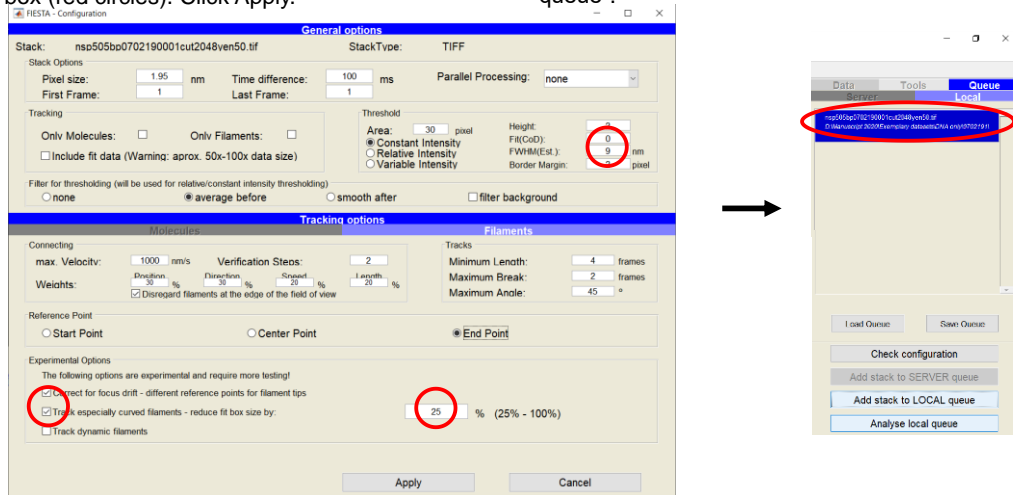
a) In Tools > Scan (right panel) choose straight line symbol (red circle). Draw a line across a DNA filament (red line in image) to obtain cross section. Right click on the cross section to get estimate FWHM (here: 9 nm).

b) Measure FWHM values of 4-5 DNA filaments from different areas of the image. Choose cancel. Delete the cross section by right mouse click on the cross section.



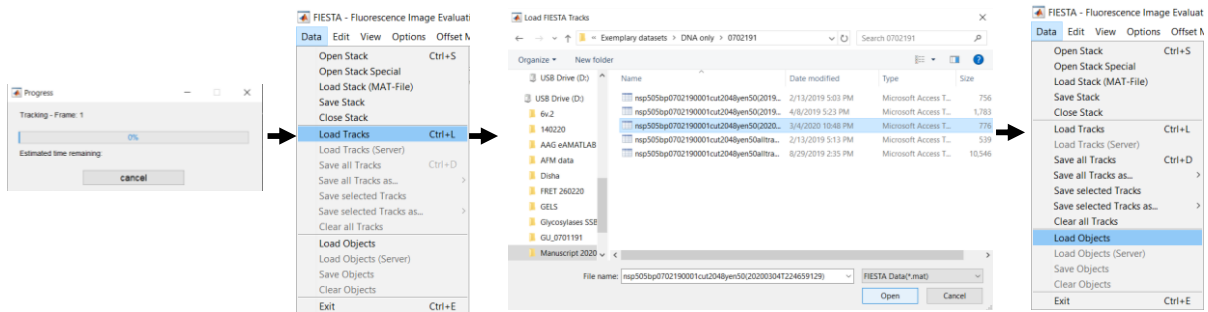
Step 5: Track DNA filaments in FIESTA III : configurations

- a) Choose Configuration. In General options, enter the average FWHM value from 4-5 filaments (red circle). In tracking options, choose Filaments, tick “track especially curved filaments - reduce fit box size by” and enter 25% in the box (red circles). Click Apply.
- b) Then click “Add stack to LOCAL queue” and confirm “add anyway”. FIESTA will add the job to the local queue (shown in red oval in right panel). This will take a few minutes. Click on this job and press “Analyse local queue”.



Step 6: Tracking DNA filaments with FIESTA IV : loading tracks & objects

- a) FIESTA runs the job for a few minutes (far left panel) and generates a .mat file in the image folder. Now go to Data > Load Tracks and load the generated .mat file.
- b) Again go to Data > Load Objects and repeat the procedure with the same .mat file.

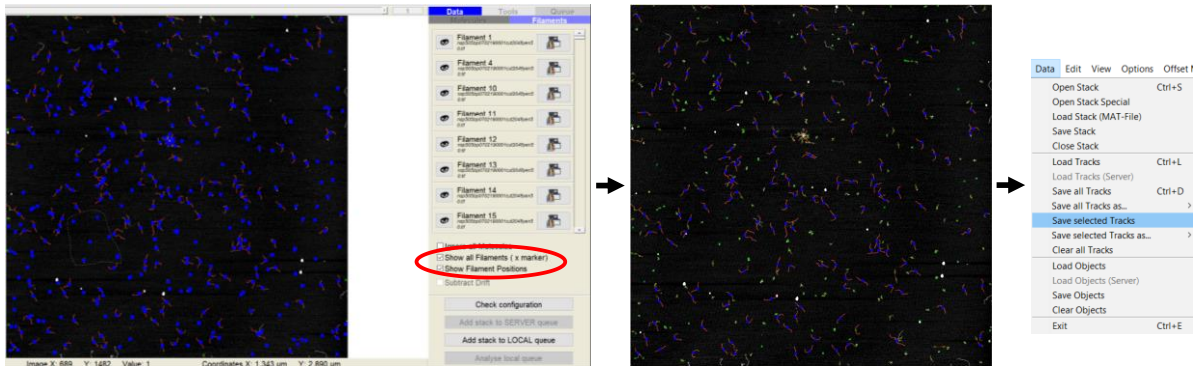


Step 7: Tracking DNA filaments with FIESTA V : sorting filaments

a) Once tracks and objects have been uploaded, go to Data > Filaments (right panel) and tick "Show all Filaments" and "Show Filament Positions" (red oval in panel). The screen will appear as shown below.

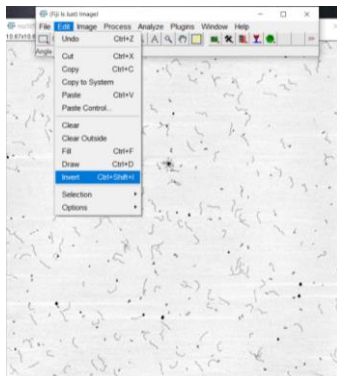
b) To exclude blue dots that mark other image features than filaments from further analysis, use left mouse button and select them. Press delete. The image will look as shown below.

c) Go to Data > Save selected Tracks and save the .mat file with all selected filaments.

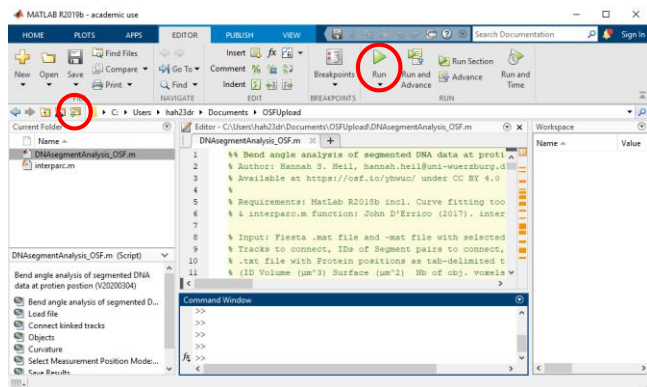


Step 8: DNA bend angle analysis in MatLab

a) In ImageJ, invert the cropped image after converting it to 8 bit as shown below.



b) Place interparc.m script, all .mat files from FIESTA, and cropped inverted image in the same folder. Open MatLab and upload the folder with all data (circled in red) as shown below. Now click on green run button (circled in red) to run MatLab DNASegmentAnalysis.m.

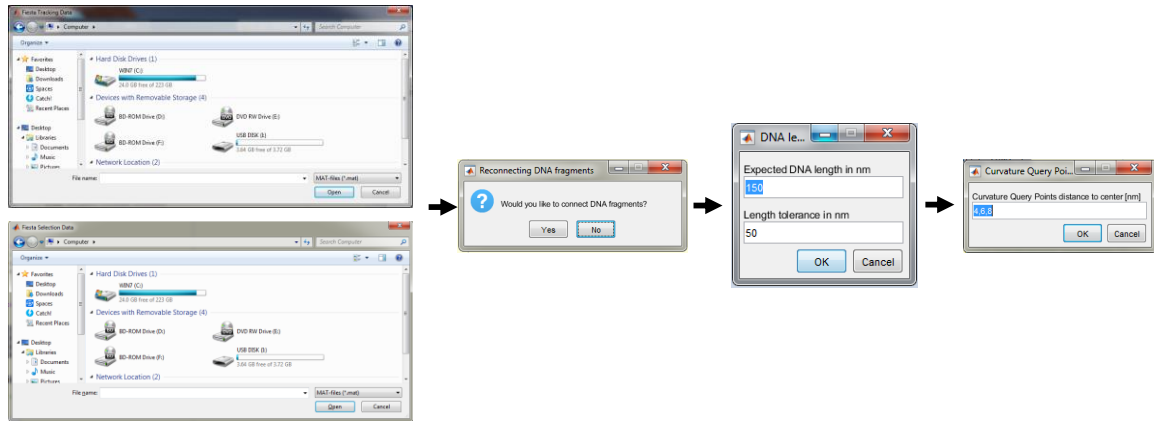


Step 9: DNA bend angle analysis in MatLab II

a) Upload the .mat files from FIESTA, first the FIESTA tracking data and second the selection .mat file.

b) A tab appears asking if any filaments are to be connected as shown below. Choose No. In the next tab, enter the expected DNA length in nm (here 170 nm) and tolerance (here 50 nm). Click OK.

c) Choose curvature query distance in points (here 4, 6, 8 nm) and press OK.



Step 10: DNA bend angle analysis in MatLab III: bend angle measurement at 50% DNA length

a) A tab appears asking the position of the angle measurement. Choose "at 50 %" and click OK.

b) Now MatLab prompts you to upload the cropped inverted image. You can overwrite the existing image with the MatLab output image. MatLab then prompts you to give the name for the output image.

c) The MatLab script provides DNA lengths, measured angles at 50% DNA length, and DNA bend angles (180° – measured angle) for the selected DNA filaments. Save the workspace again as a .mat file (hit "Speichern"). The MatLab output figure shows the selected numbered filaments in color.

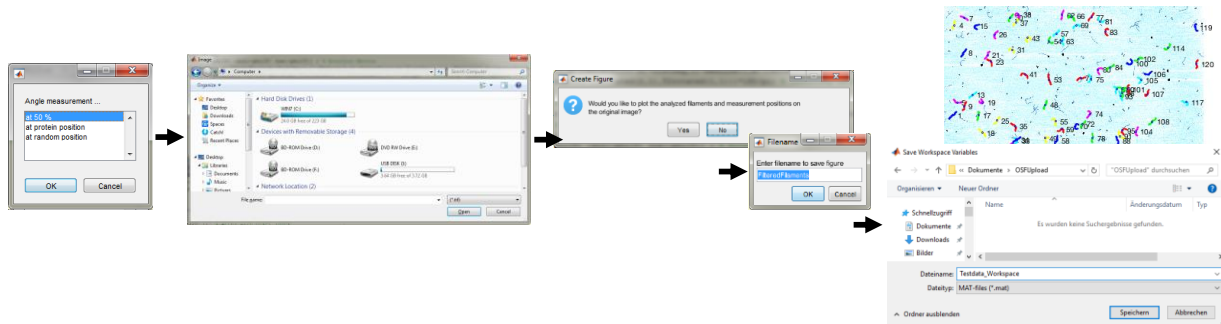
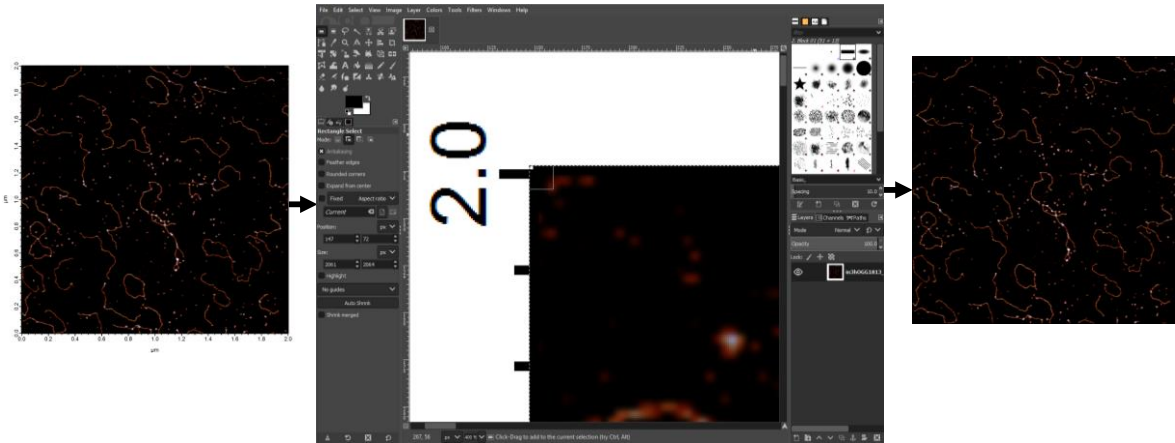


Figure x3.2: Step-By-Step guide for automated bend angle analysis for non-specifically bound protein-DNA complexes.

Step 1: .tif image

a) Raw AFM image in .ibw format is preprocessed to .tif format in *MFP 3D IgorPro* software. The pixel resolution here is 1.95 nm/pixel.

b) The image in .tif format is cropped to the actual size in GIMP software to remove axes and labels.

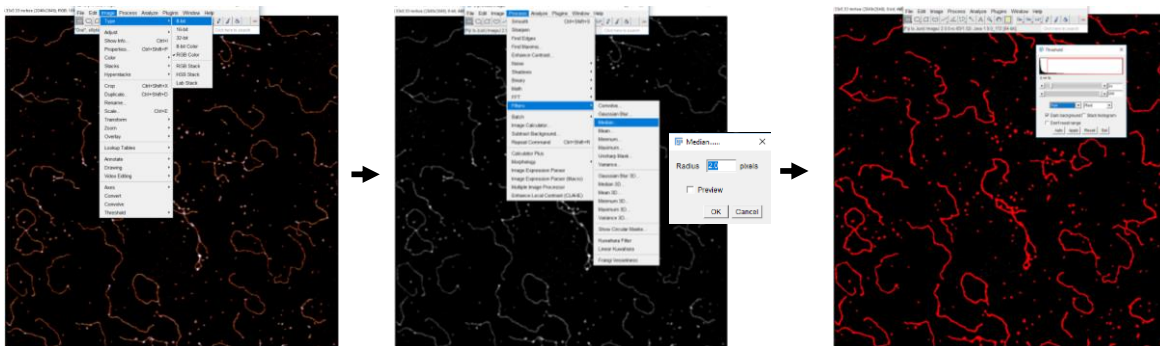


Step 2: Create DNA-only image – Threshold DNA

a) Upload the cropped image to ImageJ software and change the format from RGB color to 8-bit (Under Image> Type option).

b) Then choose Median filter under Process option and set radius to 2 pixels and press OK.

c) Again choose Adjust > Threshold under Image option and adjust Yen threshold to select DNA leaving out the background and free protein. Note the threshold value and press Apply button.

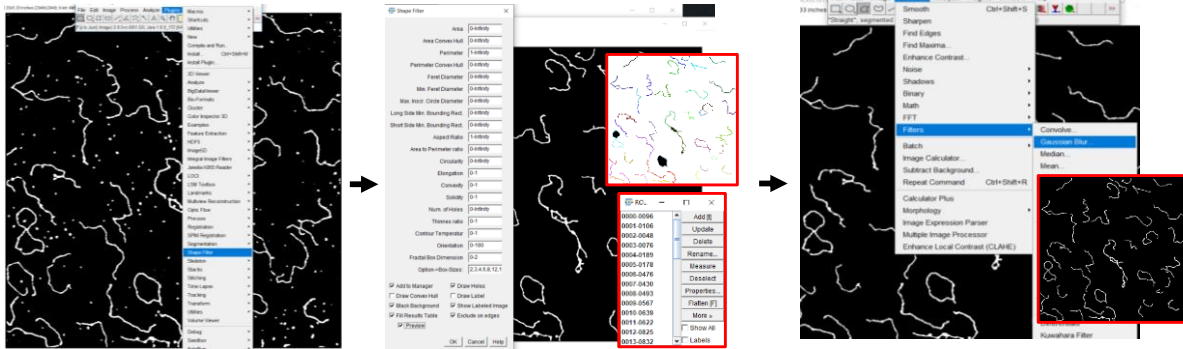


Step 3: Create DNA-only image – Shape filter

a) Go to Plugins > Shape filter option, set the perimeter from 90- ∞ and tick he boxes as shown below and press OK.

b) Save the generated image with colored DNA filaments and the table (outlined in red) as shown below (middle).

c) Go to Process > Filters > Gaussian Blur and choose 2 pixels as radius and press OK. Image will appear as outlined in red below (right).

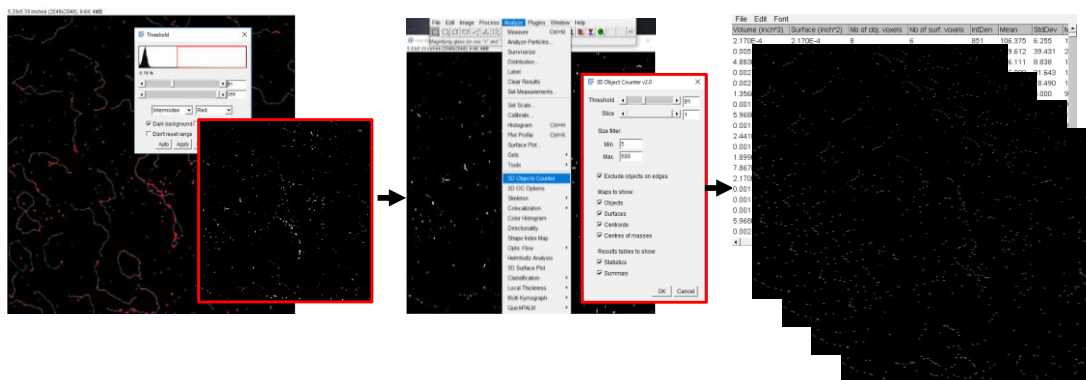


Step 4: Identify protein positions

a) To choose only protein positions, open the cropped image in ImageJ again and convert to 8-bit and apply Median filter. Go to Image > Threshold. Choose Intermodes to get protein positions. Note the value of threshold and press Apply button. The image will appear outlined in red (left).

b) Go to Analyze > 3D Objects Counter and type in the threshold value and set the size filter with min 5 and max 500 and press OK (outlined in red).

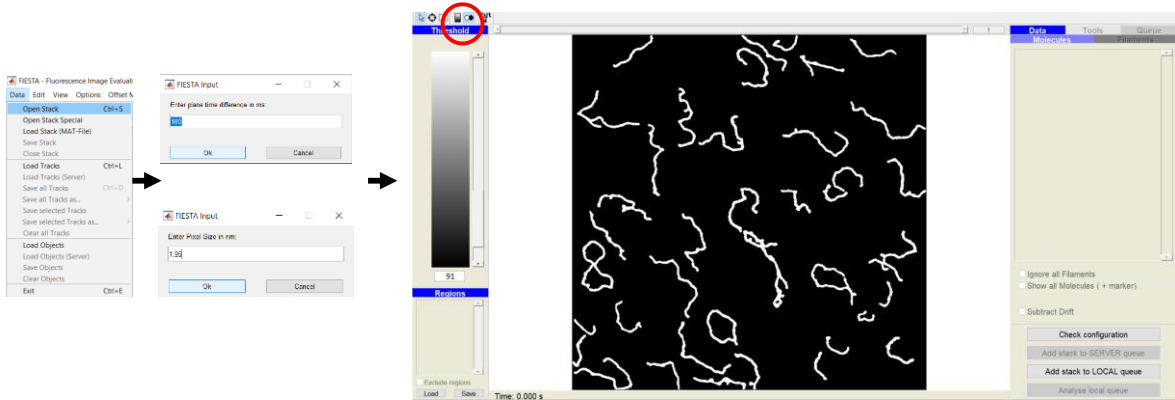
c) A table with protein position co-ordinates and maps chosen in the 3D object counter tab will be displayed as shown on right. Save the table and all generated images.



Step 5: Tracking DNA filaments with FIESTA – Threshold

a) Open the saved image from ImageJ in FIESTA tracking software by choosing open stack under data option and enter plane time difference as 100 ms and here pixel size as 1.95 nm

b) Change from normal to threshold mode at top left corner (circled in red). Enter the threshold value as shown here and switch back to normal mode.

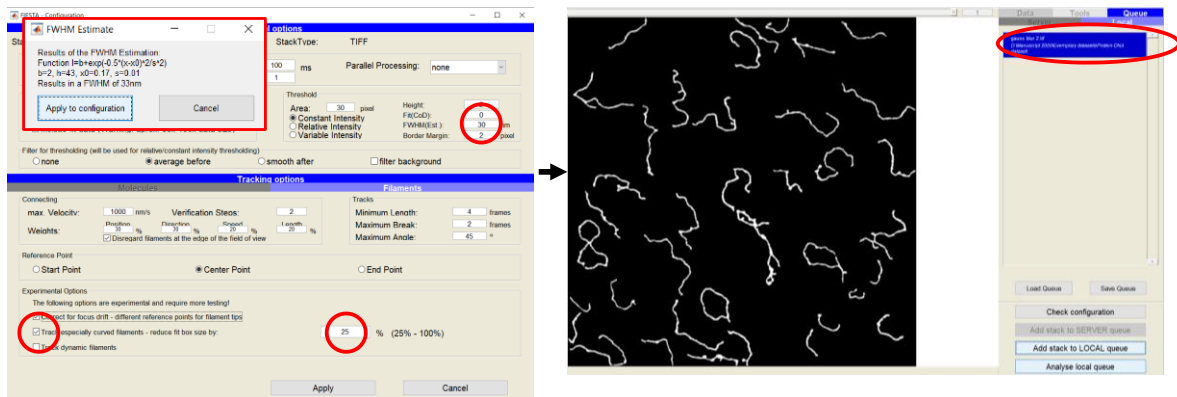


Step 6: Tracking DNA filaments with FIESTA – Configurations

a) Choose Tools > Scan option and press the button with straight line symbol. Draw a crosssection across a DNA filament and right click to choose estimate FWHM (outlined in red). Measure FWHM values of 4-5 DNA filaments from different areas of the image. Choose cancel and delete the crosssection by a right clicking on it. Here the average FWHM is 30 nm.

b) Now click on Check configuration button and enter estimated average FWHM value from 4-5 filaments measured under General options. Under tracking options, go to Filaments and tick track especially curved filaments- reduce fit box size by: enter 25% in the box and click apply button.

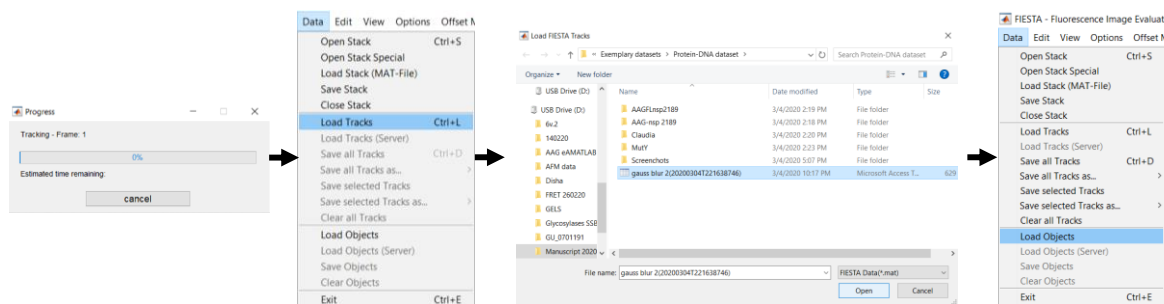
c) Then click on Add stack to LOCAL queue and press add anyway button. FIESTA will add the job in few minutes to the local queue (shown in red oval). Click on this job on the right panel and press Analyse local queue button.



Step 7: Tracking DNA filaments with FIESTA – Load tracks & objects

a) FIESTA runs the job for a few minutes (far left panel) and generates a .mat file in the same folder of the image. Now go to Data > Load tracks option and load the generated .mat file.

b) Again go to data > Load objects and repeat with the same .mat file.



Step 8: Tracking DNA filaments with FIESTA – Sorting filaments

a) Once tracks and objects have been uploaded go to Data > Filaments on the right panel in FIESTA and tick Show all filaments and Show filament positions (shown in red oval). The screen will appear as shown below.

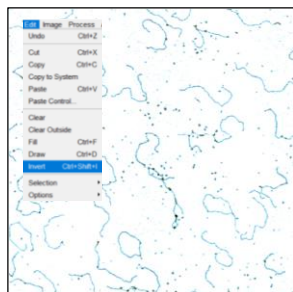
b) To choose only the filaments and exclude all other blue dots from further analysis, use left button on the mouse and select them and press delete. The image will look as shown in the middle after selection. Some filaments might be broken as shown in the inset outlined in red. Note the filament numbers as shown to be connected.

c) Go to Data > Save selected tracks and save the .mat file with all the selected filaments.

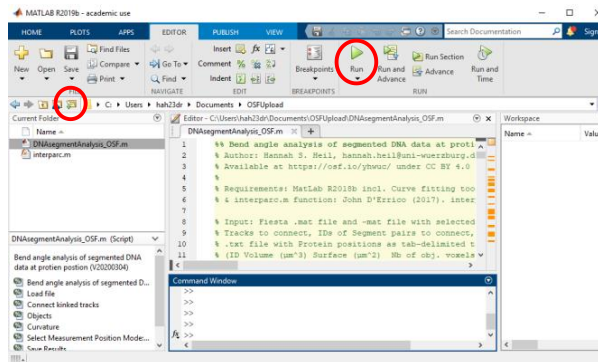


Step 9 : MatLab based DNA bend angle analysis – Run

a) In ImageJ, invert the cropped image after converting it to 8 bit as shown below (left).



b) Now to run MatLab DNASegmentAnalysis.m script, place this script, interparc.m script and all your mat files from FIESTA and cropped inverted image in the same folder. Open MatLab and upload the required folder with all data (circled in red) as shown below (right). Now click on run button in green (circled in red).

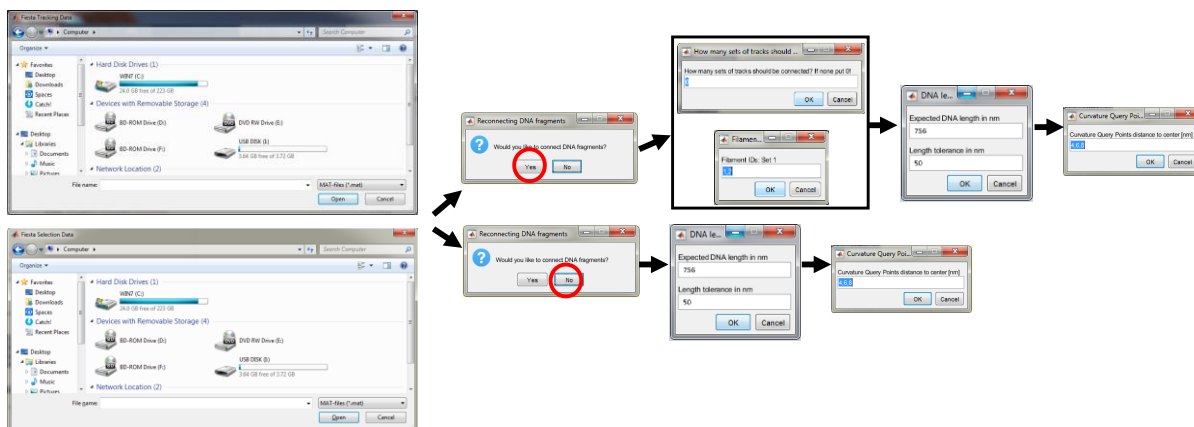


Step 10: MatLab based DNA bend angle analysis – Input

a) Upload the .mat files from FIESTA, first the FIESTA tracking data and second the selection .mat file.

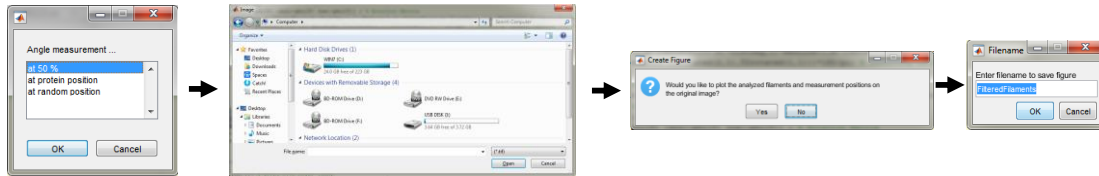
b) A tab appears asking if any filaments are to be connected as shown below. Choose Yes/No depending on your images. If you press yes button, you can first enter the number of filament sets to be connected and then enter filament IDs before you continue to enter expected DNA length in nm (here 756 nm) and tolerance (here 50 nm).

c) Next choose 4, 6, 8 nm curvature query distance points and press OK.



Step 11: MatLab based DNA bend angle analysis – Analysis mode “Angle measurement at protein position”

- a) A tab appears asking the position of the angle measurement. Choose “at protein position” and click OK.
- b) Now MatLab prompts you to upload the cropped inverted image. You can overwrite the existing image with the MatLab output image.
- c) MatLab then prompts you to give the name for the output image. Press OK.



Step 12: MatLab based DNA bend angle analysis – Load protein positions

- a) MatLab prompts you to upload .txt file with protein positions from 3D object counter,
- b) Once you upload the co-ordinates, choose protein are cutoff, protein radius and minimum protein-protein distance and once you press OK
- c) MatLab script is executed and you can save the workspace again as a .mat file. The output figure from MatLab gives the selected filaments for measurement in color as shown below with protein positions numbered in cyan color and DNA filaments in black.

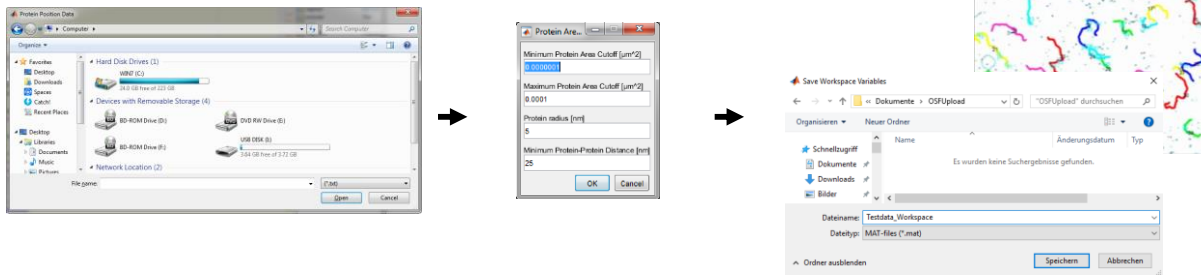
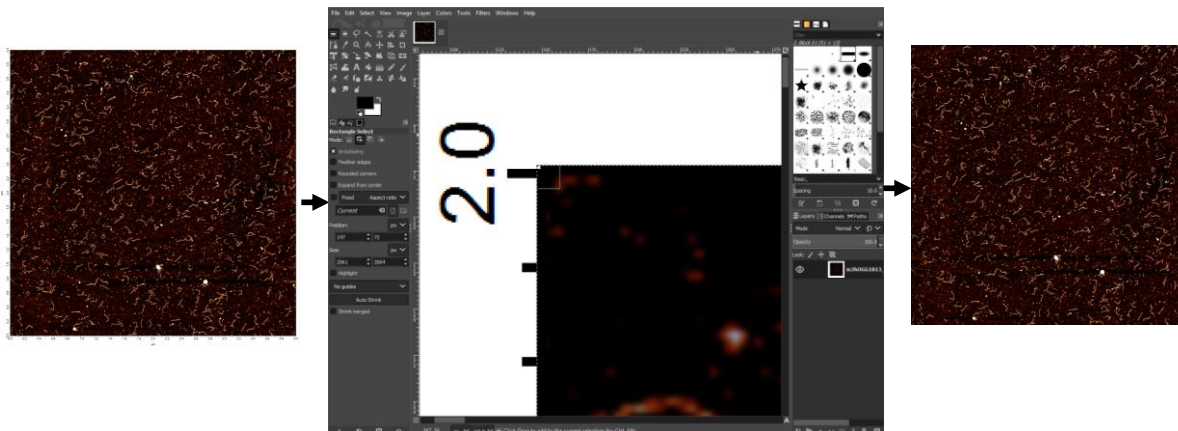


Figure x3.3: Step-By-Step guide for automated bend angle analysis for protein-DNA complexes at 50% DNA length.

Step 1: .tif image

a) Raw AFM image in .ibw format is preprocessed to .tif format in *MFP 3D IgorPro* software. The pixel resolution here is 1.95 nm/pixel.

b) The image in .tif format is cropped to the actual size in GIMP software to remove axes and labels.

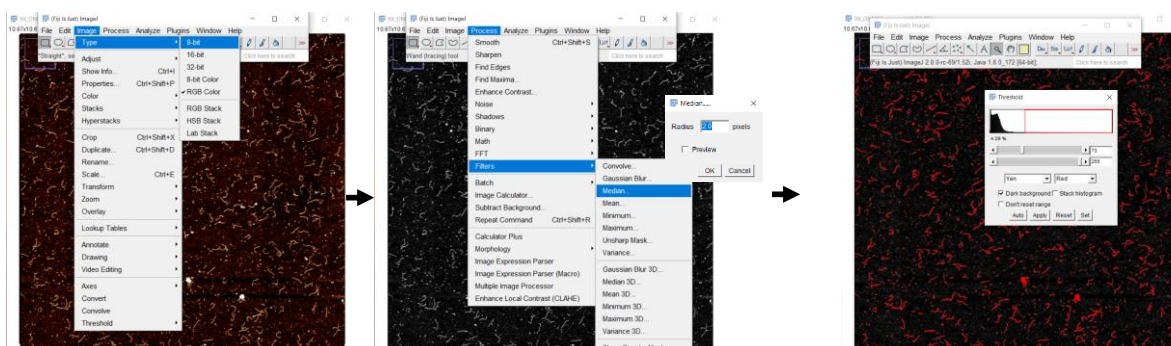


Step 2: Create DNA-only image – Threshold DNA

a) Upload the cropped image to ImageJ software and change the format from RGB color to 8-bit (Under Image> Type option).

b) Then choose Median filter under Process option and set radius to 2 pixels and press OK.

c) Again choose Adjust > Threshold under Image option and adjust Yen threshold to select DNA leaving out the background and free protein. Note the threshold value and press Apply button.

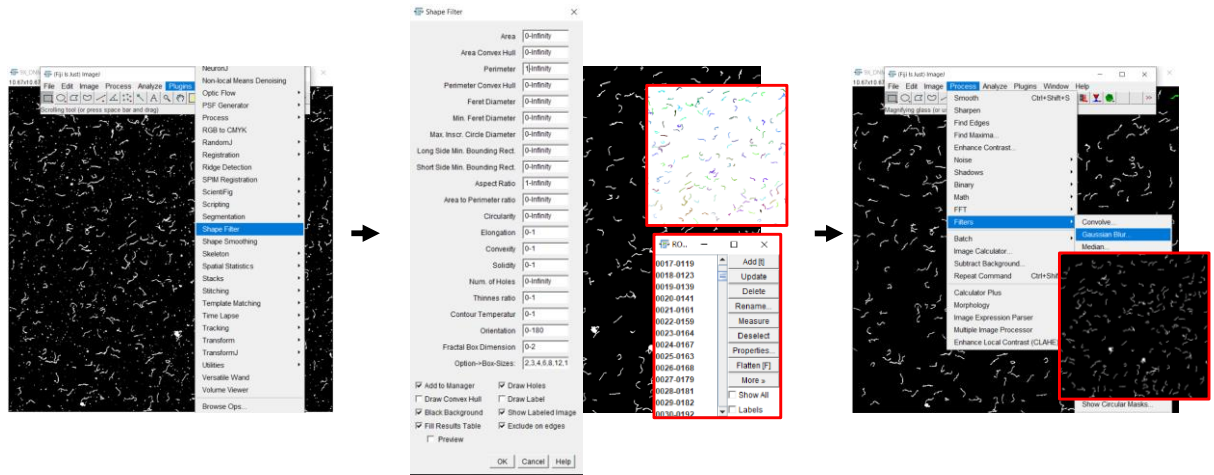


Step 3: Create DNA-only image – Shape filter

a) Go to Plugins > Shape filter option, set the perimeter from 90- ∞ and tick the boxes as shown below and press OK.

b) Save the generated image with colored DNA filaments and the table (outlined in red) as shown below (middle).

c) Go to Process > Filters > Gaussian Blur and choose 2 pixels as radius and press OK. Image will appear as outlined in red below (right).

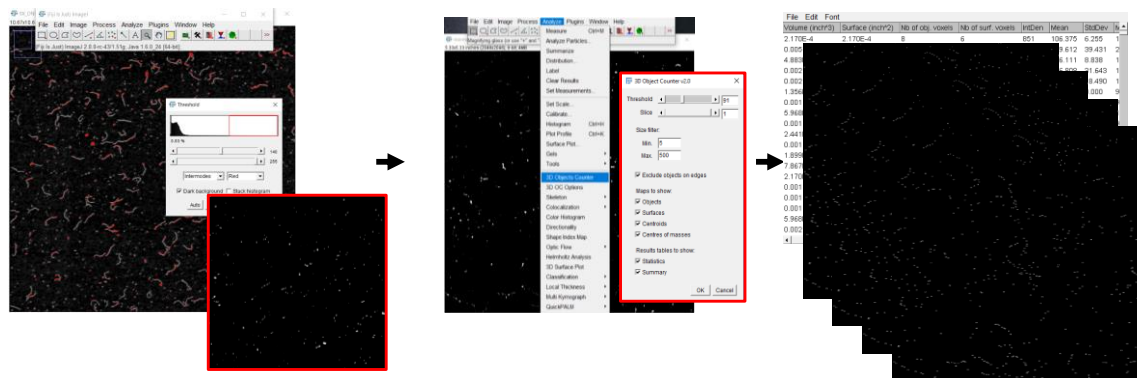


Step 4: Identify protein positions

a) To choose only protein positions, open the cropped image in ImageJ again and convert to 8-bit and apply Median filter. Go to Image > Threshold. Choose Intermodes to get protein positions. Note the value of threshold and press Apply button. The image will appear outlined in red (left).

b) Go to Analyze > 3D Objects Counter and type in the threshold value and set the size filter with min 5 and max 500 and press OK (outlined in red).

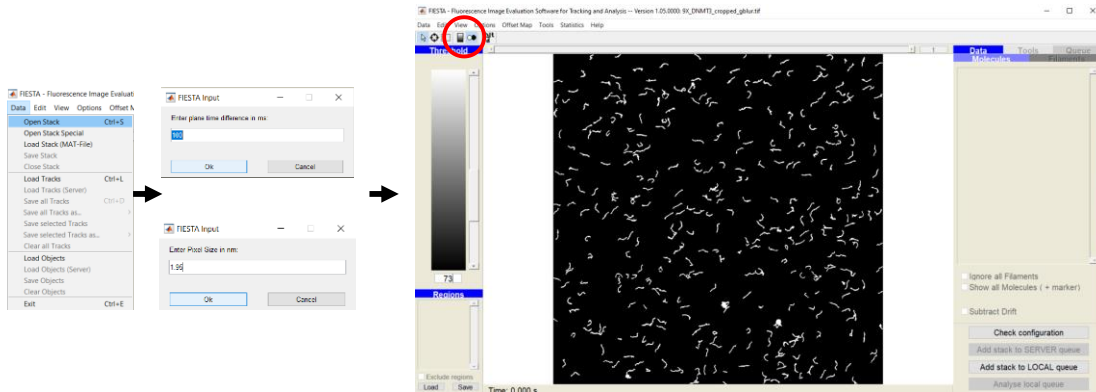
c) A table with protein position co-ordinates and maps chosen in the 3D object counter tab will be displayed as shown on right. Save the table and all generated images.



Step 5: Tracking DNA filaments with FIESTA – Threshold

a) Open the saved image from ImageJ in FIESTA tracking software by choosing open stack under data option and enter plane time difference as 100 ms and here pixel size as 1.95 nm

b) Change from normal to threshold mode at top left corner (circled in red). Enter the threshold value as shown here and switch back to normal mode.

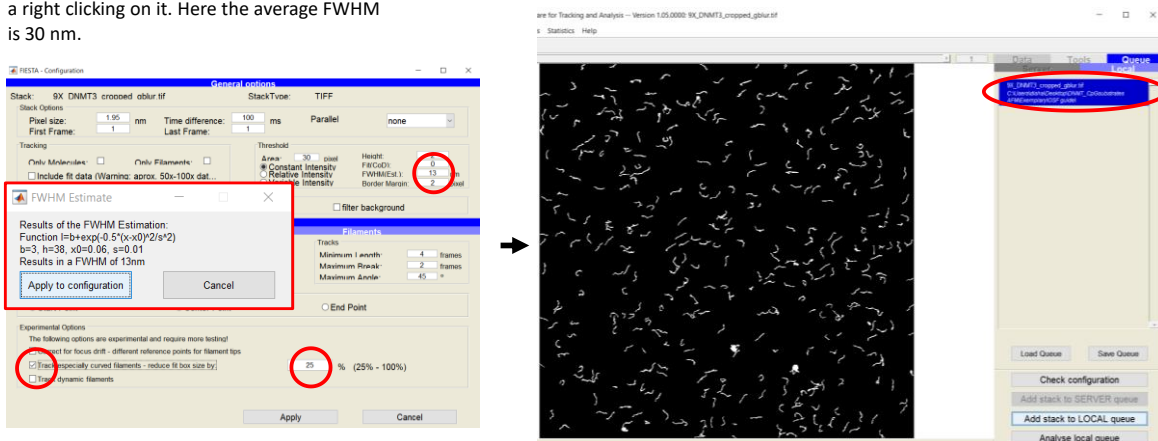


Step 6: Tracking DNA filaments with FIESTA – Configurations

a) Choose Tools > Scan option and press the button with straight line symbol. Draw a crosssection across a DNA filament and right click to choose estimate FWHM (outlined in red). Measure FWHM values of 4-5 DNA filaments from different areas of the image. Choose cancel and delete the crosssection by a right clicking on it. Here the average FWHM is 30 nm.

b) Now click on Check configuration button and enter estimated average FWHM value from 4-5 filaments measured under General options. Under tracking options, go to Filaments and tick track especially curved filaments- reduce fit box size by: enter 25% in the box and click apply button.

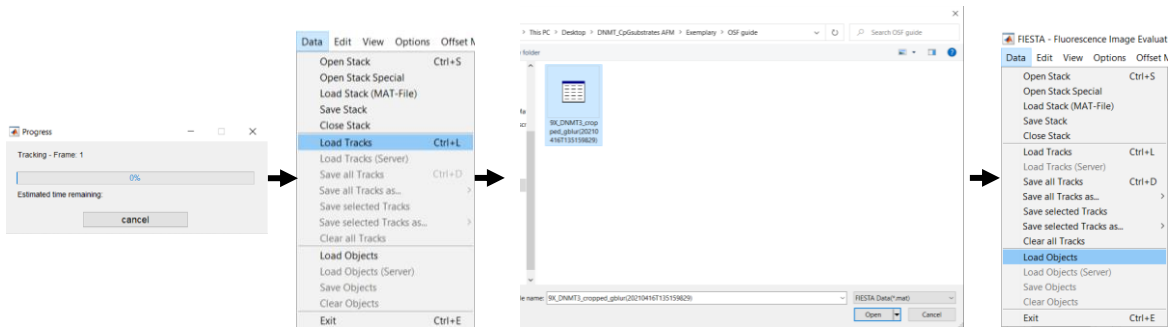
c) Then click on Add stack to LOCAL queue and press add anyway button. FIESTA will add the job in few minutes to the local queue (shown in red oval). Click on this job on the right panel and press Analyse local queue button.



Step 7: Tracking DNA filaments with FIESTA – Load tracks & objects

a) FIESTA runs the job for a few minutes (far left panel) and generates a .mat file in the same folder of the image. Now go to Data > Load tracks option and load the generated .mat file.

b) Again go to data > Load objects and repeat with the same .mat file.

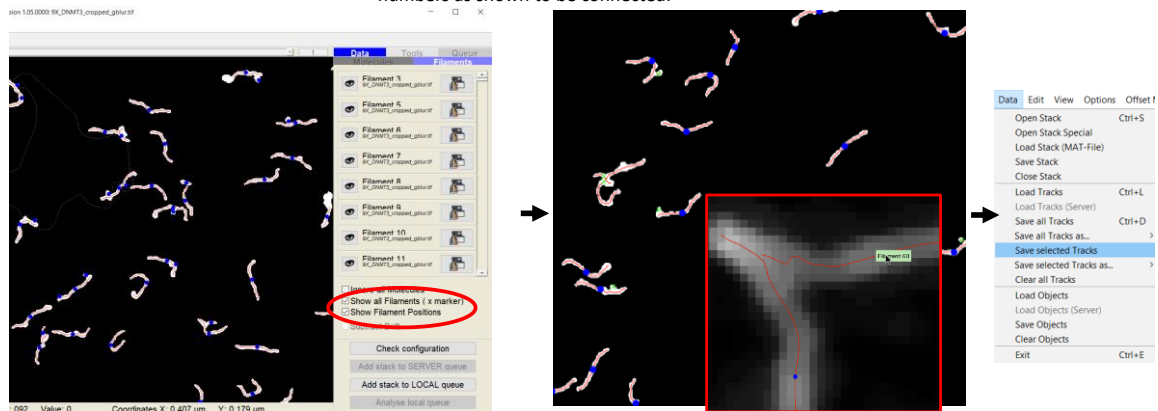


Step 8: Tracking DNA filaments with FIESTA – Sorting filaments

a) Once tracks and objects have been uploaded go to Data > Filaments on the right panel in FIESTA and tick Show all filaments and Show filament positions (shown in red oval). The screen will appear as shown below.

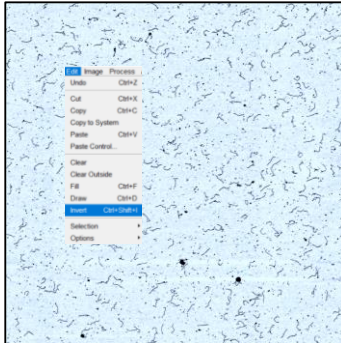
b) To choose only the filaments and exclude all other blue dots from further analysis, use left button on the mouse and select them and press delete. The image will look as shown in the middle after selection. Some filaments might be broken as shown in the inset outlined in red. Note the filament numbers as shown to be connected.

c) Go to Data > Save selected tracks and save the .mat file with all the selected filaments.

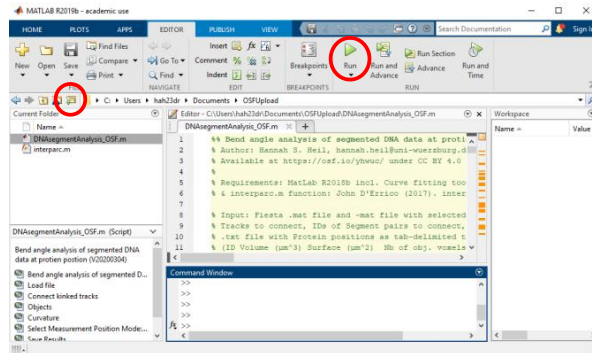


Step 9 : MatLab based DNA bend angle analysis – Run

a) In ImageJ, invert the cropped image after converting it to 8 bit as shown below (left).



b) Now to run MatLab DNASegmentAnalysis.m script, place this script, interparc.m script and all your mat files from FIESTA and cropped inverted image in the same folder. Open MatLab and upload the required folder with all data (circled in red) as shown below (right). Now click on run button in green (circled in red).

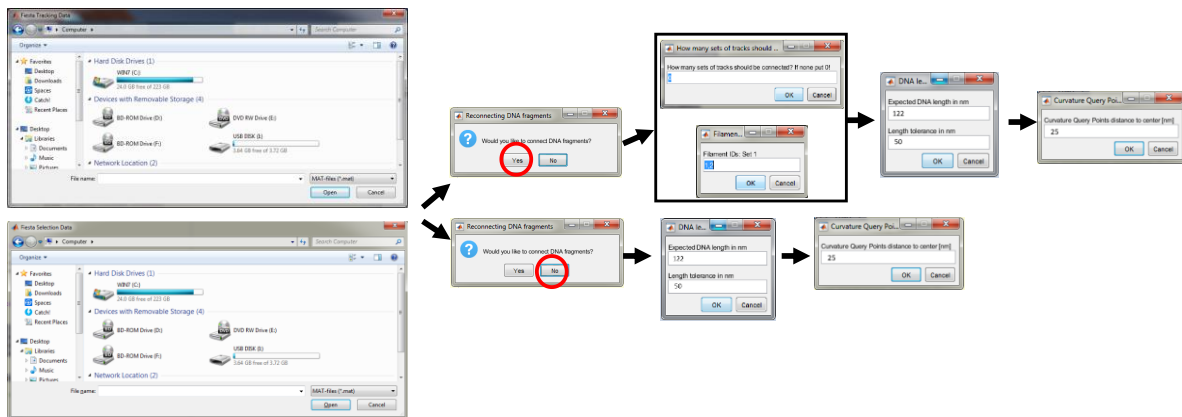


Step 10: MatLab based DNA bend angle analysis – Input

a) Upload the .mat files from FIESTA, first the FIESTA tracking data and second the selection .mat file.

b) A tab appears asking if any filaments are to be connected as shown below. Choose Yes/No depending on your images. If you press yes button, you can first enter the number of filament sets to be connected and then enter filament IDs before you continue to enter expected DNA length in nm (here 122 nm) and tolerance (here 50 nm).

c) Next choose 25 nm curvature query distance points and press OK.

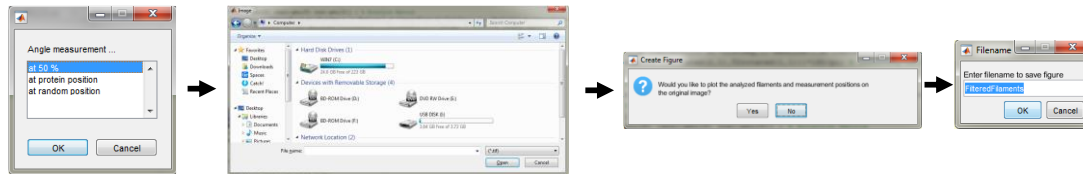


Step 11: MatLab based DNA bend angle analysis – Analysis mode “Angle measurement at protein position”

a) A tab appears asking the position of the angle measurement. Choose “at protein position” and click OK.

b) Now MatLab prompts you to upload the cropped inverted image. You can overwrite the existing image with the MatLab output image.

c) MatLab then prompts you to give the name for the output image. Press OK.



Step 12: MatLab based DNA bend angle analysis – Load protein positions

a) MatLab prompts you to upload .txt file with protein positions from 3D object counter,

b) Once you upload the co-ordinates, choose protein are cutoff, protein radius and minimum protein-protein distance and once you press OK

c) MatLab script is executed and you can save the workspace again as a .mat file along with the results table. The output figure from MatLab gives the selected filaments for measurement in color as shown below with protein positions numbered in cyan color and DNA filaments in black.

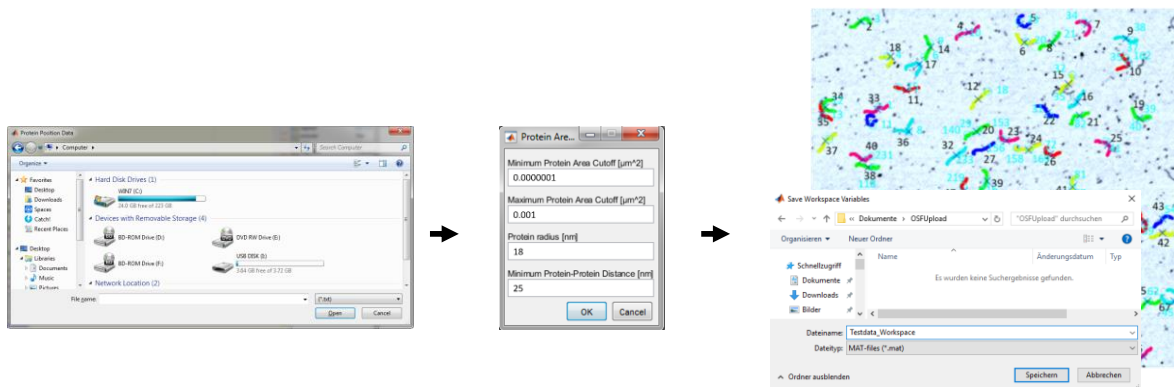


Table x3.1: Configuration settings of FIESTA and MatLab software programs used for the automated DNA bend angle measurements at lesions. For the parameters not mentioned here, default values were used.

Software	Entries	Values
FIESTA	Pixel size in nm	Enter pixel resolution (e.g. here 1.95 nm/pixel)
	Box size [25%-100%]	25%
	Track especially curved filaments	Select
	Threshold	threshold value from ImageJ (see Image pre-processing in Methods)
	Estimated FWHM	DNA FWHM value obtained by section tool in FIESTA (e.g. here 9 nm)
	Show all filaments and filament positions	Select
MATLAB	Expected DNA length in nm	Enter approximate length of DNA (theoretical length $0.34 \text{ nm/bp} \times N$, N is the number of base pairs)
	Curvature query points distance to center	Enter desired segment length in nm (e.g. here 8nm, see Choice of query point distance in Methods)
	Angle measurement	at 50% (see DNA preparation in Methods)

Table x3.2: Configuration settings of FIESTA and MatLab software programs for automated bend angle measurements for protein-DNA complexes. For the parameters not mentioned here, default values were used.

Software	Entries	Values
FIESTA	Pixel size in nm	Enter pixel resolution (e.g. here 1.95 nm/pixel)
	Box size [25%-100%]	25%
	Track especially curved filaments	Select
	Threshold	threshold value from ImageJ (see Image pre-processing in Methods)
	Estimated FWHM	DNA FWHM value obtained by section tool in FIESTA (e.g. here 9 nm)
	Show all filaments and filament positions	Select
MATLAB	Expected DNA length in nm	Enter approximate length of DNA (theoretical length $0.34 \text{ nm/bp} \times N$, N is the number of base pairs)
	Curvature query points distance to center	Enter desired segment length in nm (e.g. here 8nm, see Choice of query point distance in Methods)
	Number of broken filament sets and filament IDs	For connecting disrupted DNA skeleton lines, enter the number of filament sets and their IDs obtained from FIESTA (see Connecting kinked DNA filaments in Automated analysis of DNA bend angles at protein positions, Methods)
	Angle measurement	at protein position

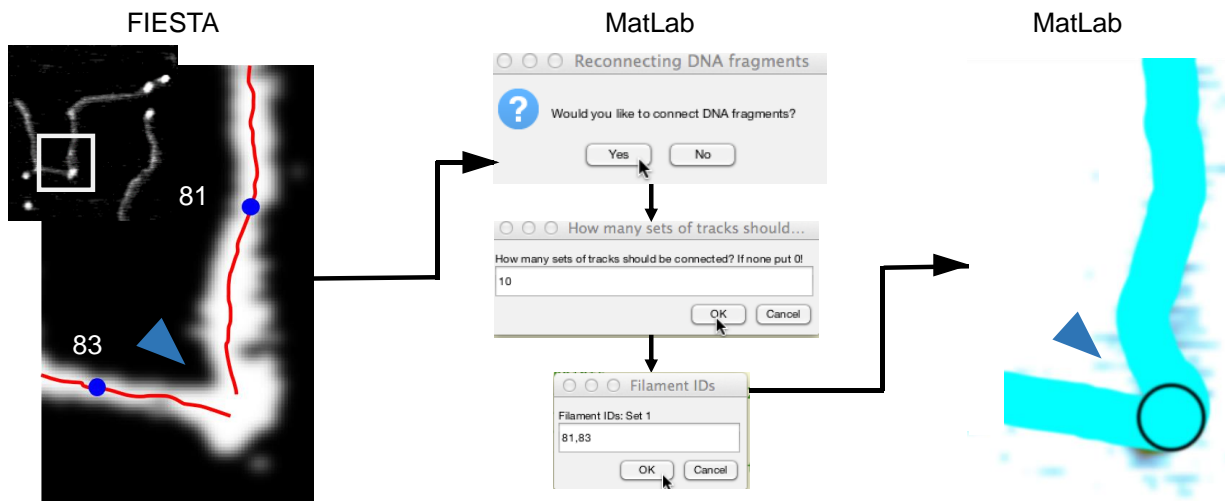


Figure x3.4: Schematic showing the procedure to connect kinked filaments during MatLab analysis. Extreme left: DNA skeletonization procedure (red line represents the DNA skeleton) in FIESTA is generally interrupted at the kinked positions of the protein bound sites (depicted by the blue arrow). The inset depicts the original AFM image with area selection showing the kinked DNA at the protein-DNA complex site. These interruptions in the DNA skeleton lines lead to the formation of two separate filaments (here 81 and 83) of shorter length. Shortening of the substrate length prompts MatLab to not select these strands even though they are valid protein-DNA complexes. This results in the loss of higher bend angles for the lesions and the protein-DNA complexes. Middle: MatLab prompts to enter the total number of filaments to be connected, and their identifiers obtained from FIESTA skeletonization process. Extreme right: The filament ends are then connected by MatLab, resulting in a complete DNA strand as shown here (empty black circle marks the sealed region).

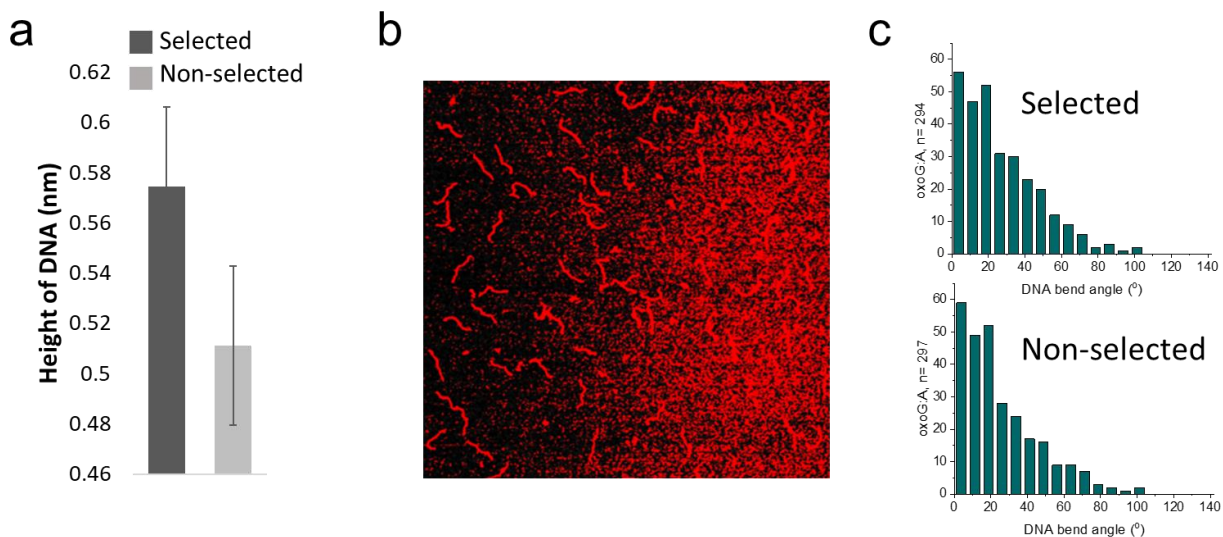


Figure x3.5: AFM image quality criteria for performing automated MatLab analyses. (a) Sufficiently sharp AFM imaging probes result in images with DNA heights of ≥ 0.5 nm and are essential to obtain continuous, unbroken DNA filaments during the FIESTA skeletonisation process, without the interference of background noise. Correct DNA lengths can be obtained from only continuously skeletonized DNA and thus be selected for MatLab analyses. Manual analyses of the revealed increased heights of the selected DNA versus non-selected DNA. (b) Similarly, the substrate background in the images must not have differing heights at different positions. Uneven background results in inconsistent selection of DNA filaments. Height cut-off marking by density slice (in red) shows the uneven background in the AFM images. These images must be avoided for automated MatLab analyses (c) Even though the requirements of sufficient image quality (a and b) are met, not all DNA fragments in the images are selected. Moreover, we confirmed that the selection of DNAs was not biased towards specific bend angle states. Manual measurements of DNA bend angles for MatLab selected (top) and non-selected (bottom) DNA filaments show comparable bend angles (exemplary: oxoG:A substrate), indicating that a sub-optimal image quality leads to the loss of DNA selection and hence data, but not wrong results. Scale bar: 200 nm.

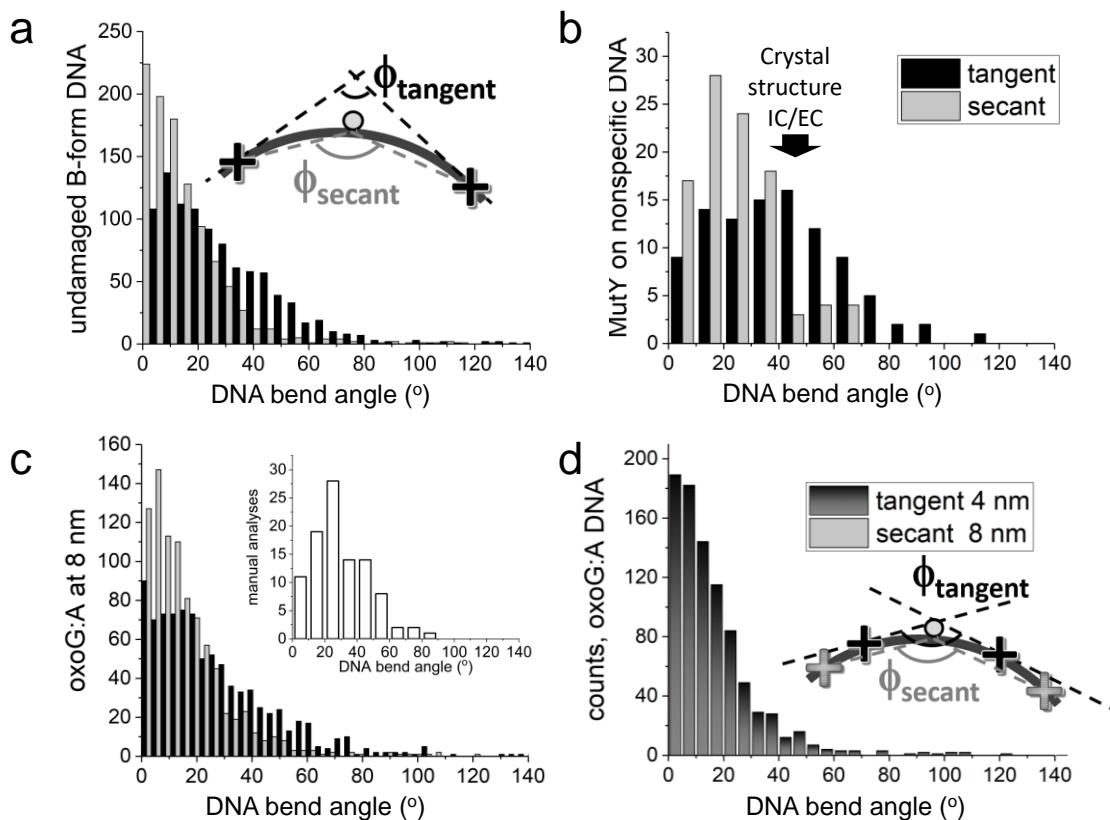


Figure x3.6: Comparison of Tangent versus secant method for measuring DNA bend angles. In the manual analysis by tangent overlay approach, the angle between straight lines through the DNA backbone on either side of the target site is estimated to calculate the DNA bend angle. Here to understand whether MatLab uses the tangent or the secant geometry more closely to determine DNA bend angles, we applied both methods to various test datasets. Black represents bend angles measured by tangent geometry and grey represents secant geometry. (a) Results with tangents rather than secant lines at 8 nm query point distance from the 50% position compare well with the in-silico predicted reference bend angle value for the undamaged DNA substrate (6°). The inset depicts the principle behind tangent and secant approach. (b) As a further control, we compared the DNA bend angles in IC/EC glycosylase complexes known from crystal structures with our obtained bend angles. For example, MutY bound to its target lesions (PDB id: 1RRQ) has been shown to bend the DNA by $\sim 55^\circ$, comparable to bend angles obtained for the IC conformation from automated MatLab analyses ($n=98$) only by tangent approach, but not the secant method (at 8 nm query point distance). (c) Exemplary bend angle distributions determined with tangent and secant method at 8 nm distance from the 50% position for oxoG:A lesion. Bend angles measured by tangent rather than secant approach are consistent with those obtained from manual measurements (wider binning is used in the inset due to low number of data points). (d) Understanding relation between tangent and secant approaches. Bend angles obtained with the tangent method using 4 nm query point distance and the secant method at 8 nm distance from the 50% position (exemplarily shown for oxoG:A DNA) overlay exactly. A schematic of the tangent (black, 4 nm query point distance) and secant (grey, 8 nm query point distance) geometries is shown in the inset.

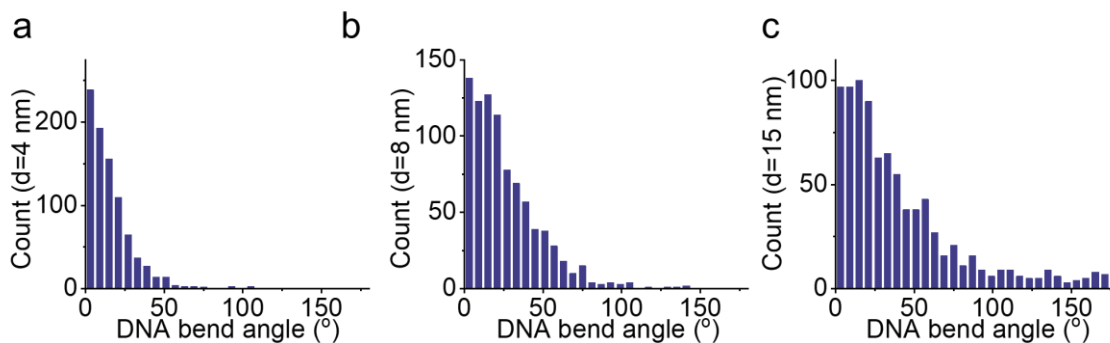


Figure x3.7: Query point distances of the tangents influence the DNA bend angle measurements.

Tangent overlay by manual method is a well-approved procedure for the determination of DNA bend angles from AFM images. In this approach, lines are laid along the middle of the DNA backbone on either side of the target position, for example using the angle tool in ImageJ, and from the angle measured at the intersection of these lines the DNA bend angle is obtained (180° -this angle). However, the specific line lengths and with that the specific distances from the target site that are probed (query points) are not fixed and are subject to vary significantly between experimenters based on for example, extent of zooming the image or different image sizes. In contrast, our automated MatLab approach regulates the query point distance. User can enter required value of the query point distance when the MatLab prompts. DNA bend angles were determined by automated MatLab analyses at 4 nm (a), 8 nm (b), and 15 nm (c) from the target position at 50% DNA length (exemplarily: oxoG:A lesion substrate, $n=877$). The pixel resolution of the images used in this study is ~ 2 nm. With 4 nm query point distance, only two pixels are included in each direction from the target site in the analysis resulting in the introduction of noise in the bend angles finally resulting in a very narrow distribution with a maximum at 0° , relatively independent of DNA lesion type or protein binding. At the other end of the spectrum, measuring at query point distances of 15 nm, the local effects of DNA distortion are washed out by the contributions from adjacent non-specific DNA undulations on either side of the target site. Slightly larger bin size (bin size 6°) has been used to observe bend angles for 15 nm query point distance ($N=\sqrt{n}$, N number bins, n number data points). 8 nm query point distance has been used in our all our studies.

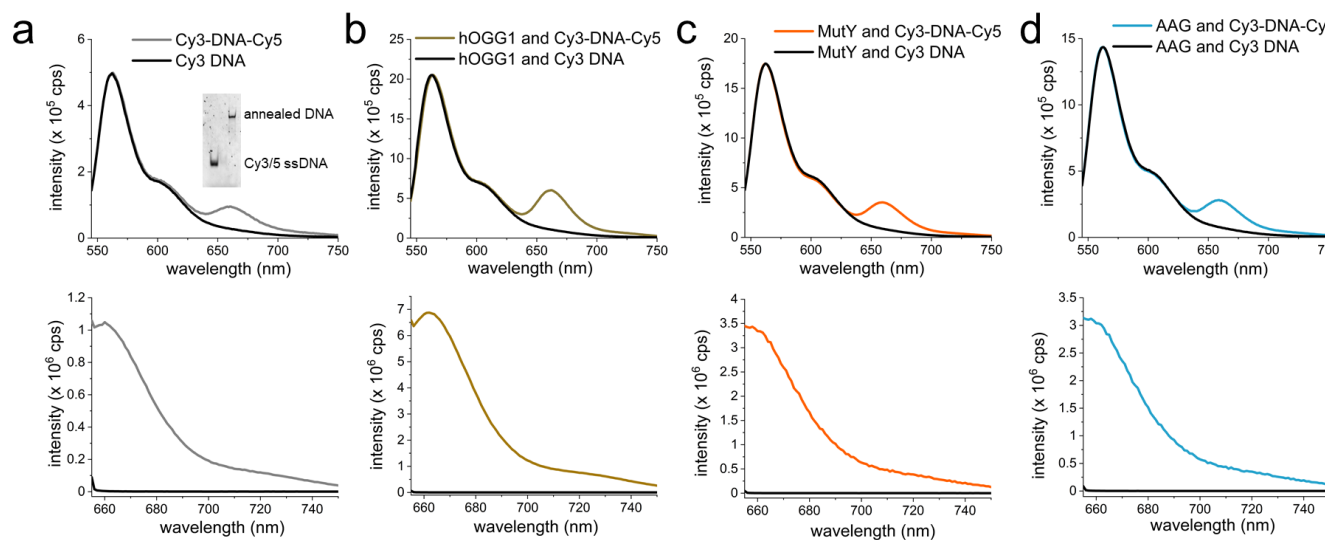


Figure x3.8: Ensemble FRET measurements to obtain average DNA bend angles in hOGG1-DNA, MutY-DNA and hAAG-DNA complexes. Intensity spectra for the emission of undamaged Cy3- and Cy5-labelled DNA (FRET substrate) (a) in the absence of protein (grey), and in the presence of specific glycosylases (b) hOGG1 (gold), (c) MutY (red), and (d) hAAG (blue). Top: emission spectra of the acceptor at Cy3 (donor) excitation; bottom: emission spectra of the acceptor at direct Cy5 (acceptor) excitation. Cy3-only-labelled DNA \pm protein was used for the correction of background signals (indicated by black curves in a-d); these signals were subtracted from the individual emission curves for the glycosylases measured with the FRET DNA substrate for FRET efficiency (E_{FRET}) calculations. The inset in (a) shows PAGE gel confirming that the Cy3- and Cy5-labelled ssDNA strand was fully annealed to the non-labelled, 48 nt undamaged top strand, and the FRET substrate was subsequently present only in double stranded DNA (dsDNA) form with Cy3 and Cy5 at the ends.

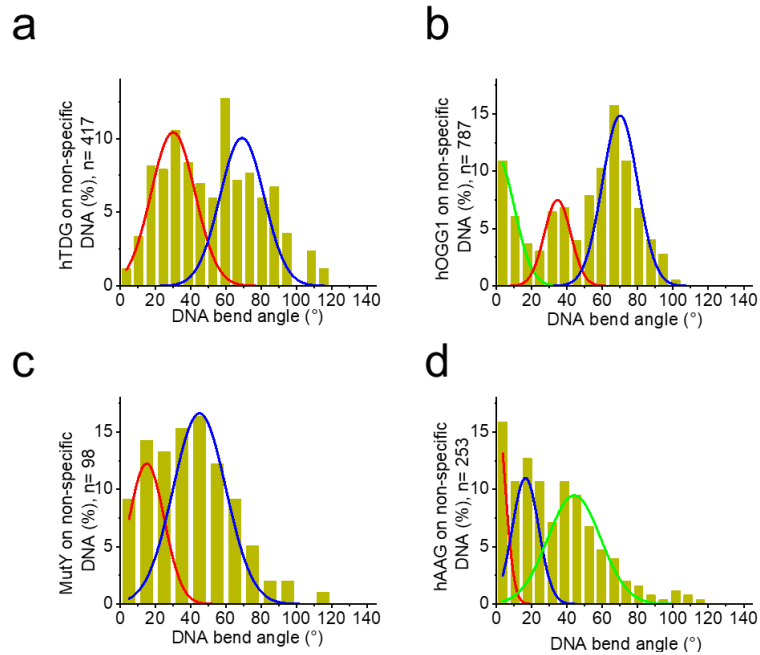


Figure x3.9: Individual Gaussian fits shown in the multimodal fits for DNA-glycosylase complexes. DNA bend angle distributions for glycosylase complexes with undamaged DNA for (a) hTDG, (b) hOGG1, (c) MutY, and (d) hAAG.

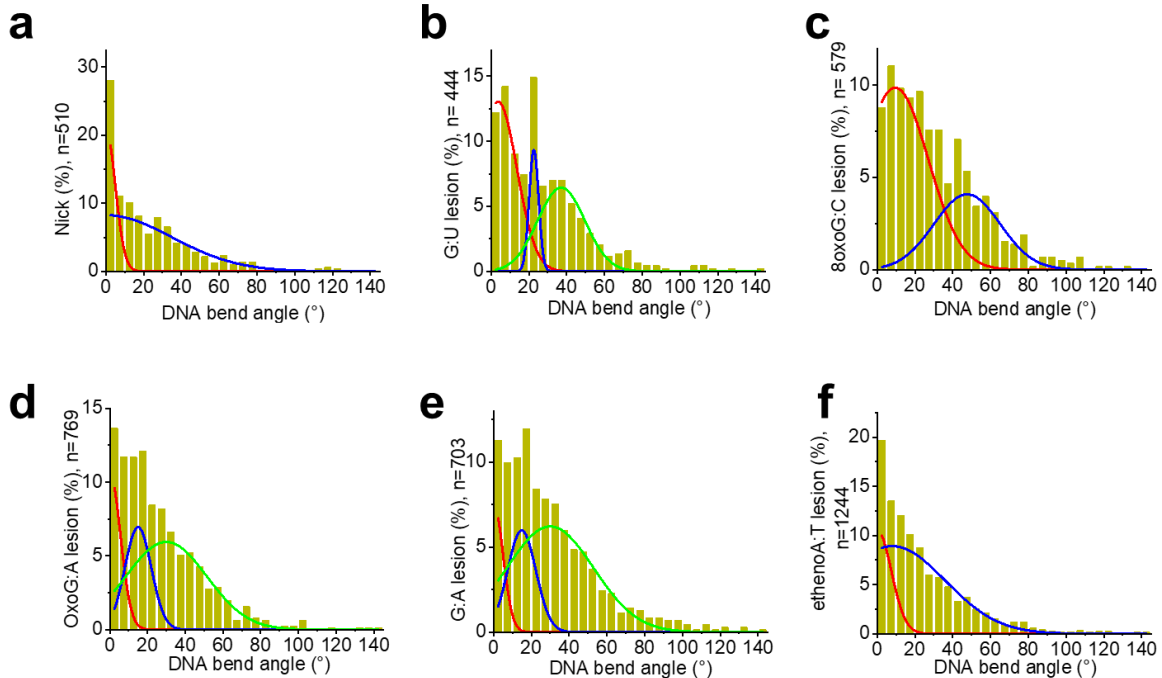


Figure x3.10: Individual Gaussian fits shown in the multimodal fits for lesion containing substrates in the absence of glycosylases. Bend angle distributions at DNA lesions at (a) DNA nick, (b) G:U mismatch, (c) 8oxoG lesion, (d) 8oxoG:A lesion/mismatch, (e) G:A mismatch, and (f) ethenoA lesion.

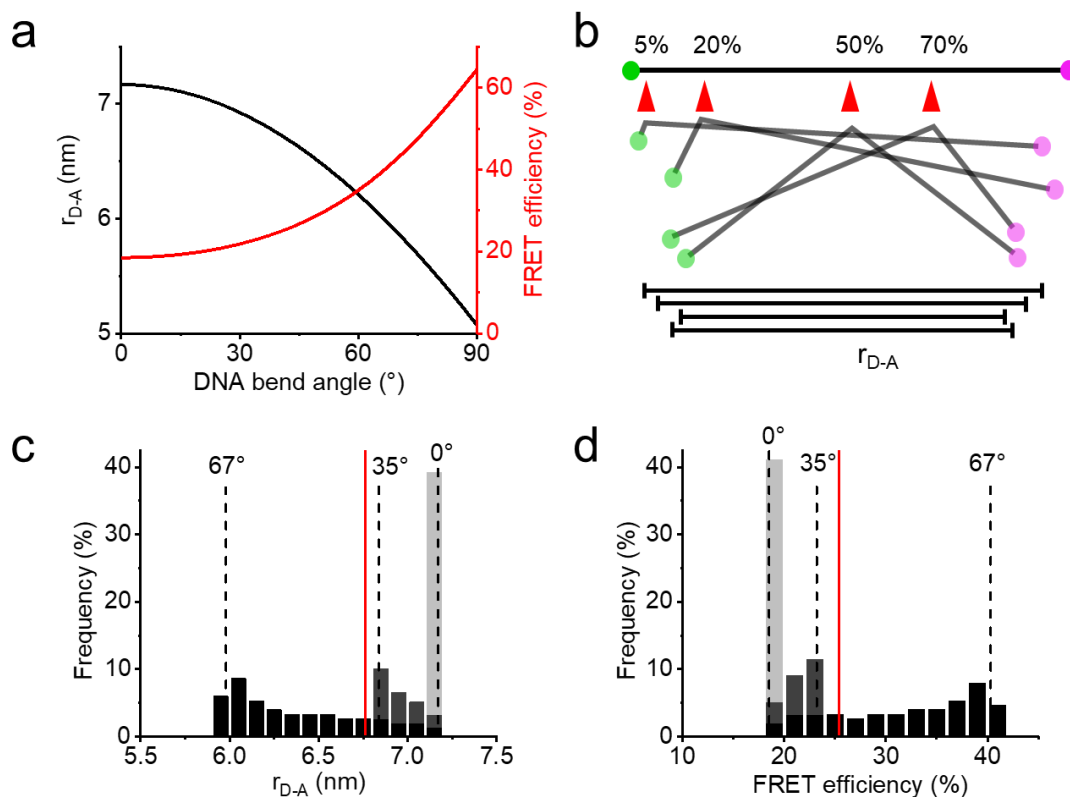


Figure x3.11: FRET simulations to validate average AFM bend angles for glycosylase-DNA complexes. (a) For a given bend angle at 50% position, the plot depicts the resulting donor-acceptor distance r_{D-A} and FRET efficiency E_{FRET} between donor (Cy3) and acceptor (Cy5) pair used in our ensemble FRET measurements. (b) At fixed bend angle and varying binding positions (red arrows) on the DNA, donor-acceptor distance r_{D-A} and hence FRET efficiencies (indicated as percentages) for the given FRET pair (Cy3- green, Cy5- magenta) used in our studies. (c,d) Stacked histograms of obtained donor-acceptor distance, r_{D-A} (c) and resulting E_{FRET} (d), shown for hOGG1 as an example. DNA bend angles as well as their respective frequencies as determined from AFM analyses were defined at each equally probable position along the DNA filament (red arrows in (b)). The values expected for hOGG1 for the respective bend angles (67° black, 35° grey, and 0° light grey) and a bend position at 50 % between donor and acceptor are indicated (dashed black lines). Red lines show the average DNA bend angles, which agree well with our average bend angles from ensemble intensity-based FRET experiments. These data hence support that the DNA bend angles obtained by AFM and DNA bending in solution are highly similar. Figure obtained with permission from Bangalore and Heil et al, 2020.

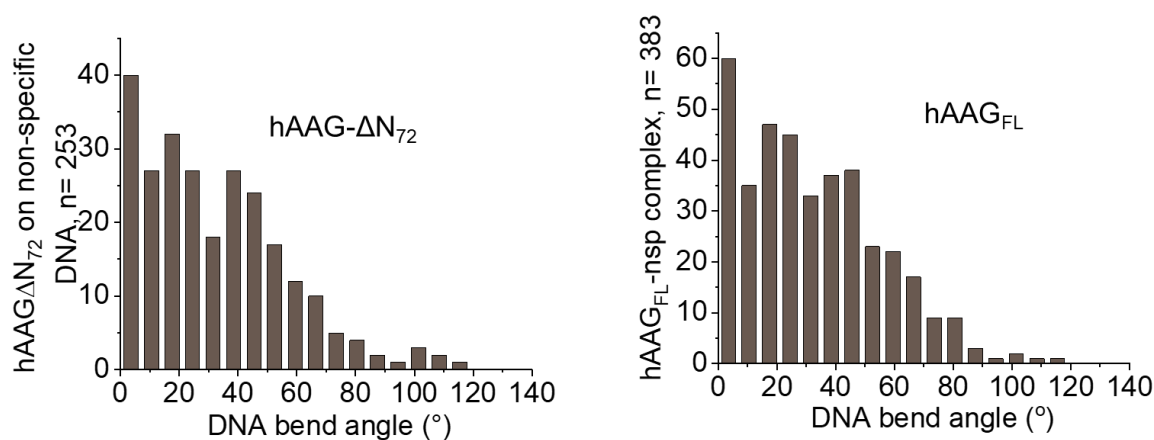


Figure x3.12: DNA bending obtained by full length hAAG (right)(b) versus the N-terminally truncated variant used in these studies (left)(a). N-terminal truncation is by 72 amino acids.

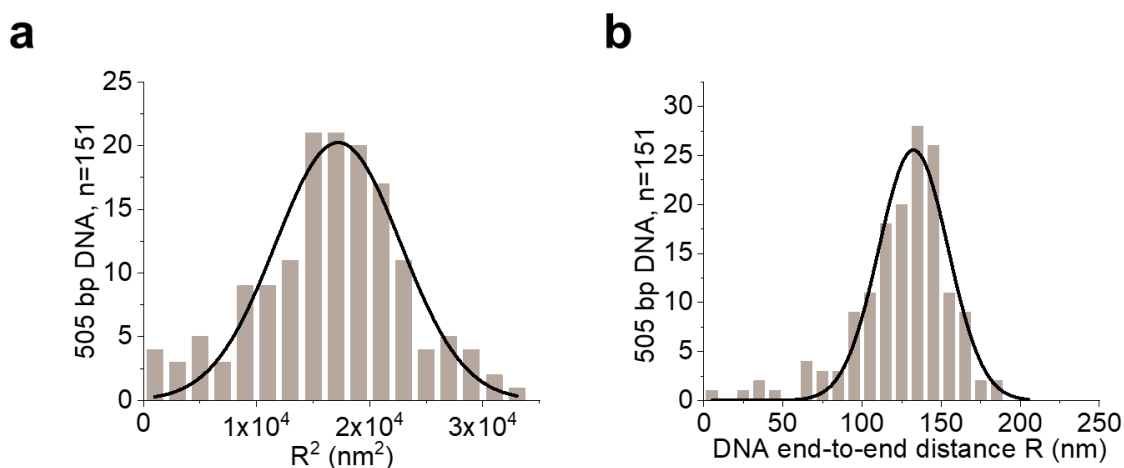


Figure x3.13: End-to-end distance estimation on undamaged 505 bp long DNA substrate. Using worm like chain (WLC) model, end-to-end distances (R) are calculated from DNA persistence lengths L_P : $\langle R^2 \rangle_{2D} = 4 L_P L_C \{1 - (2L_P/L_C)(1 - e^{-L_C/2L_P})\}$. The contour length of the 505bp long substrates used in this study is $L_C \approx 172$ nm. For this data (with buffer conditions 25 mM Na⁺, 10 mM Mg²⁺ and DNA depositions on mica), I obtained $2D = (17,214 \pm 422)$ nm² and an average end-to-end distance $R \approx 131$ nm for the undamaged DNA substrate. The corresponding persistence length L_P is ~ 45 nm, consistent with 2D equilibrated DNA structures in the AFM images. Persistence lengths of all other 505bp long lesion containing substrates ranged between 40 and 45 nm (Table. x3.3).

Substrate	R² (nm²)	R (nm)	L_p (nm)
nick	16,270 ± 435	128 ± 2	41 ± 3
G:U	16,656 ± 602	129 ± 2	42 ± 3
8oxoG:C	16,562 ± 306	129 ± 2	42 ± 3
oxoG:A	16,853 ± 556	130 ± 2	43 ± 3
G:A	16,267 ± 277	128 ± 2	41 ± 3
ethenoA:T	16,087 ± 330	127 ± 2	40 ± 3
undamaged B-form DNA	17,214 ± 422	131 ± 2	45 ± 3

Table x3.3: End-to-end distances R and persistence lengths L_P of undamaged and lesion containing 505 bp DNA substrates used in this study. Square roots of the Gaussian centers for R^2 distributions gave end-to-end distances R . All DNA substrates show comparable end-to-end distances that are consistent with the fully 2D equilibrated B-form DNA (L_P 40-50 nm).

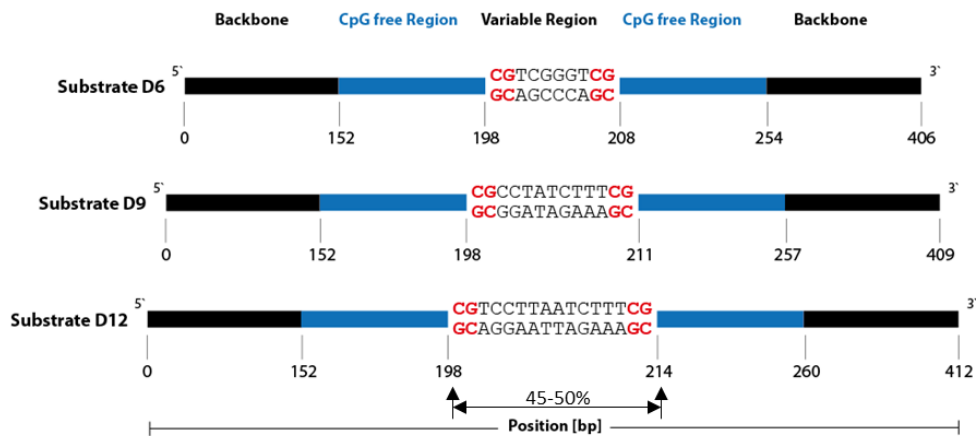
Appendix II

Text

x4.1 Free DNMT3A-3L showed volumes of $\sim 70 \text{ nm}^3$. Using the calibration for the AFM set up used in this study¹⁹, $\sim 70 \text{ nm}^3$ roughly translates to $\sim 60 \text{ kDa}$ which, is the size of DNMT3A-3L heterodimer. At low concentrations (7 nM) and in the absence of DNA, DNMT3A does not show multimerization hence explaining the low volumes observed.

For the volumes of protein-DNA complexes, the higher values observed is the result of volume contributions from dsDNA²¹. This might pose difficulty to apply the linear relationship between the AFM volumes and molecular weights. Hence volumes have not been translated to molecular weights in this study.

a



b

Sequence of substrate D12

```

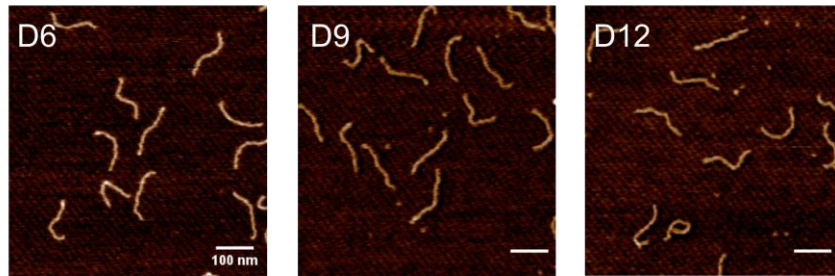
TTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGAT 50
AACAAATTTACACAGGAAACAGCTATGACCATGATTACCGCCAAGCTTGGT 100
ACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCC 150
TTAAAAGGAGGCCCGCATTAGAGTCTCTCTGTTTGATGGAATTGCAACG 200
TCCTTAATCTTTCGGGTACTTGGTGTCTCAAGGAGTTGGGTATTAAAGTG 250
GAAAAGTACTAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCCGC 300
TCGAGCATGCATCTAGAGGGCCCAATTCGCCCTATAGTGAGTCGTATTAC 350
AATTCACTGGCCGTCGTTTTACAACCGTCGTGACTGGGAAAACCCTGGCGT 400
TACCCAACTTAA

```

Fig x4.1: Schematic showing DNA substrate preparation for AFM experiments. a) D6, D9 and D12 substrates (400 bp long with additional variable spacing in the middle shown contain dual CpG sites (colored in red) with variable distances of 6, 9, and 12 bp respectively. D12 substrate shows the 45-50% position falls between the two CpG sites (corresponds to $\sim 20 \text{ bp}$). **b)** Entire sequence of the D12 substrate used in the AFM studies. The central pair of CpG sites is displayed in yellow, the sequence in between the CpG sites is displayed in grey and CpG free region is colored in blue. The sequence outside the CpG free regions contains random CpG sites. Black triangles indicate the two sites (25% and 28% DNA lengths) with dual CpG sites placed 13 bp apart. This explains the peak at 25-30% DNA length seen in the position

distributions. The long CpG free region allows distinction species bound at 50% positions and elsewhere on the DNA.

a



b

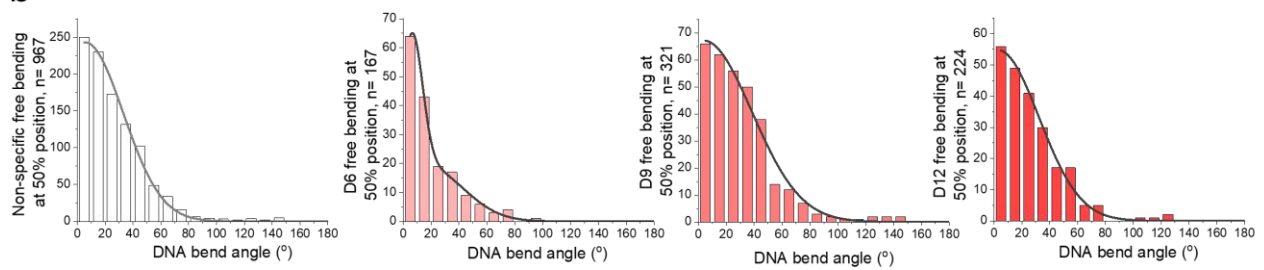


Figure x4.2: AFM analyses for the DNA substrates with CpG sites placed at different distances in the absence of protein. a) AFM images of free D6, D9 and D12 DNA substrates. b) DNA bend angle distribution of free DNA substrates. The D6, D9, and D12 bend angle distributions are shown as shaded histograms at 50% of the DNA length in the absence of protein. The white bend angle distribution is for non-specific sites obtained from chapter 3.

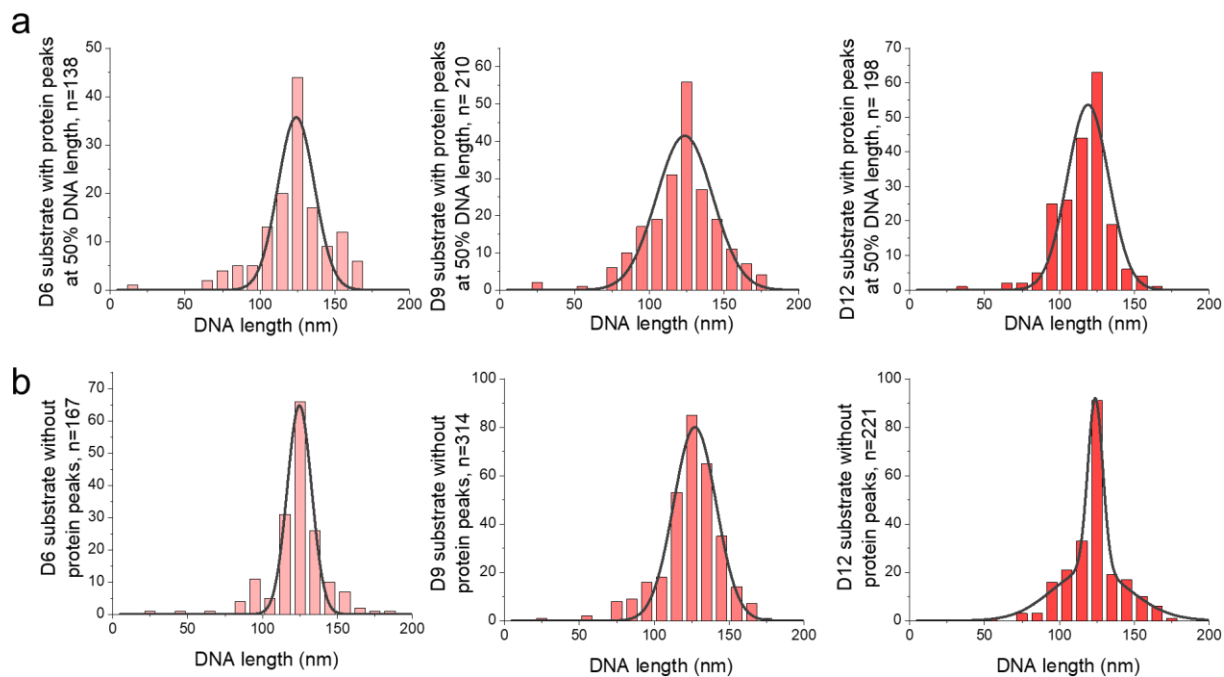


Figure x4.3: Histograms showing DNA length distributions for D6, D9 and D12 substrates obtained from MatLab routine. a) in the presence of DNMT3A-3L complexes at 50% positions b) in the absence of proteins on the DNA for all the three DNA substrates D6, D9, and D12 (left to right)

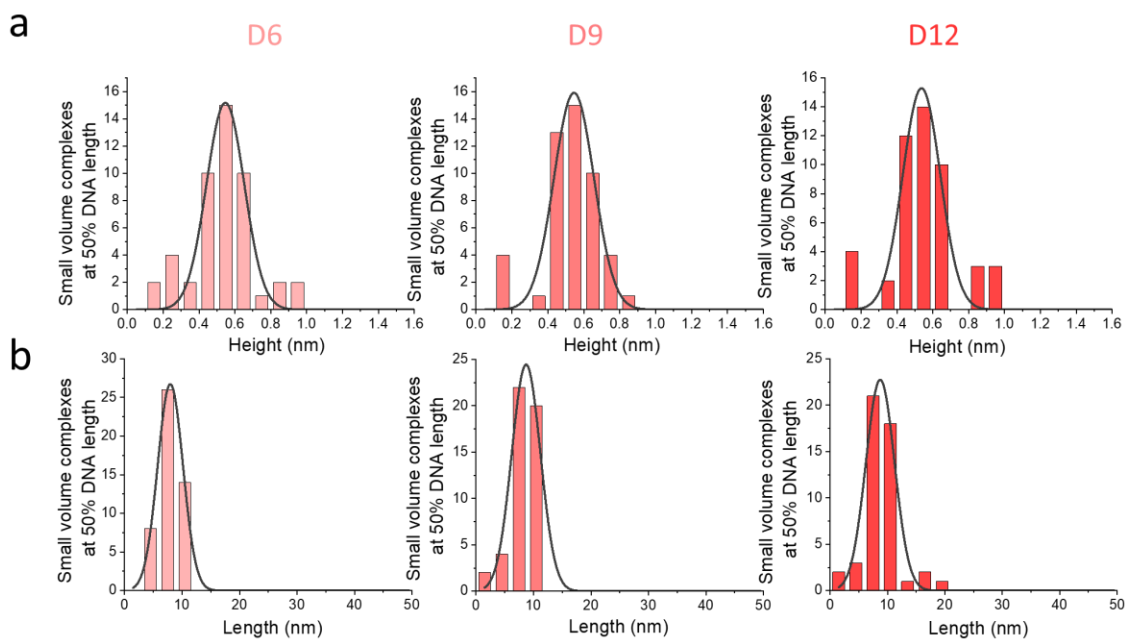
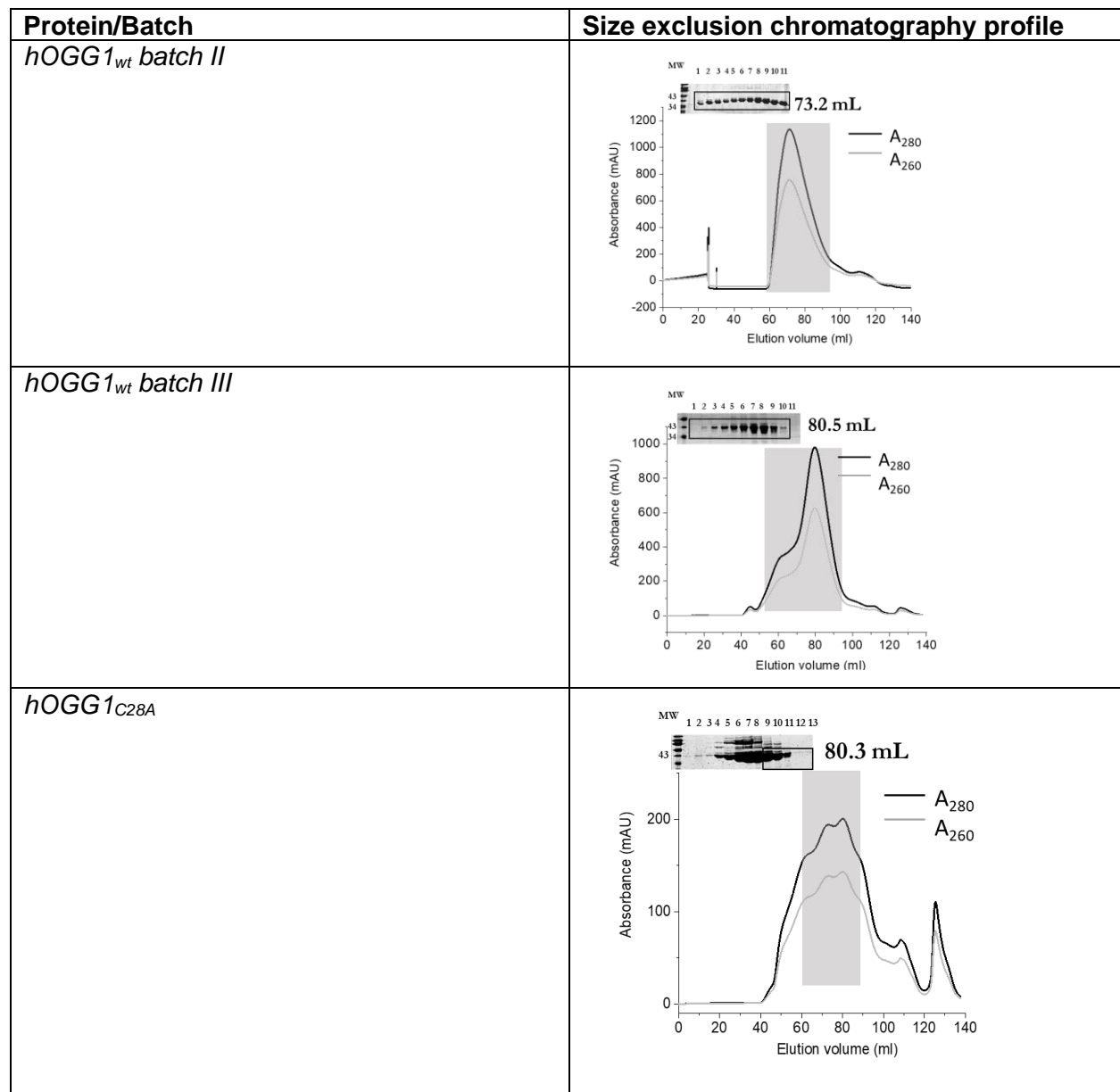


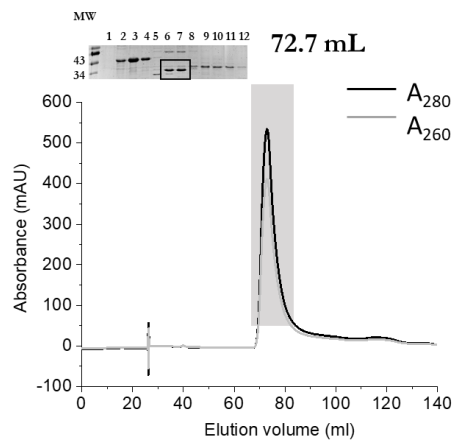
Figure x4.4: Heights and lengths of single heterotetrameric complexes of DNMT3A/3L (based on volumes) bound to CpG pairs with 6, 9, and 12 bp spacing. Heights and lengths were measured using ImageSXM and ImageJ software programs respectively.

Appendix III

Figure x5.1: Size exclusion profiles and SDS-PAGE analyses of hOGG1_{C28A}, hOGG1_{C241A}, hOGG1_{C253A} and batches II and III of hOGG1_{wt}. Elution volumes are mentioned in the chromatograms. Grey shaded area denotes the fractions selected for subsequent SDS-PAGE analysis. Insets in the SEC chromatograms show SDS-PAGE analysis of the corresponding SEC fractions. Molecular weight (MW) protein ladder indicates 43 kDa and 34 kDa sizes. hOGG1 is a ~40 kDa protein. Black rectangle box indicates the fractions pooled and concentrated for further experiments.



*hOGG1*_{C241A}



*hOGG1*_{C253A}

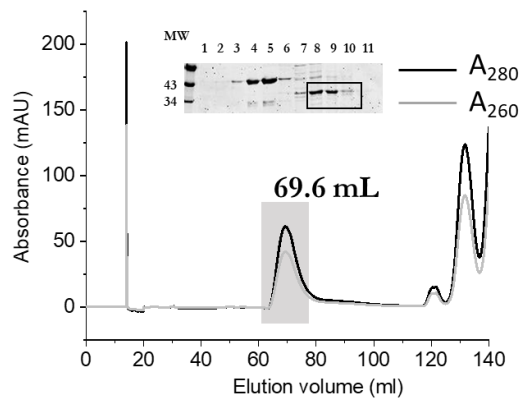
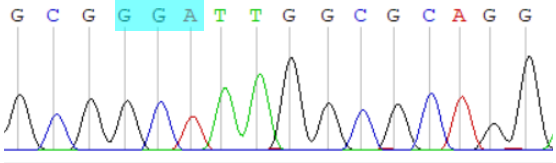
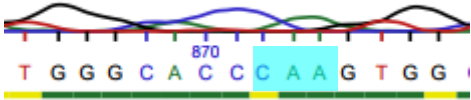


Figure x5.2: Sequencing results for hOGG1_{C28A}, hOGG1_{C241A}, hOGG1_{C253A}, hOGG1_{wt}, Myc₁₋₁₆₃, E-box in pUC19N, and hOGG1_{K249Q}. Mutation at 326 position in wild-type hOGG1 results in the formation of a dimer¹⁹⁵. However, for hOGG1_{wt}, region marked in cyan color shows that at 326 position, serine residue (GGA codes for serine) is present and not cysteine mutation hence confirming that the clone is indeed wild type hOGG1. For the rest of the clones, cyan colored regions show either the inserted gene/sequence (in the case of wild type) or the mutation (in the case of mutants).

Gene/ Insert	Sequencing result
hOGG1 _{wt}	
Myc ₁₋₁₆₃	<p>>MycM41_8rev_T7term</p> <p>NAKCKCGRWTTCGGATCMGGTACCTCAWTAKTMCATGGYKCCCTGAAAANKAAGATTWCW SCTCATGGGGWTGWGTTATYGYTAYTGTGGTTGTTGTTCCGAGMTCSAATKAKTYGMS GWCKTTCCKKDCYATCGACAGTCYGCACCCGCTGGCRGHGTGATCMCCGMAGTAS CAMGGCAKACCAGAAAGYGGACATCTGCRSRATGTTSGGATGRITYCRMCTTCKGRGC</p>
E-box in pUC19N	<p>>pEbox3R_M13rev-49</p> <p>NNNNCCAGCTTGCNNGCCTGCAGGTCGACTCTAGAGGATCAGATCACGTGCCTCTAGAC TCGAGGCATGCATTGACTCGATCCCCGGTACCGAGCTCGAATTCACTGGCCGTGTTT TACAACGTGTGACTGGGAAAACCTGGCGTTACCCAACTAATCGCCTTGACAGCATTC</p>
hOGG1 _{K249Q}	
hOGG1 _{C28A}	<p>>C28A 1F_T7</p> <p>NNNRNGWAATTCCTCTAGAATAATTTGTTTAACTTTAAGAAGGAGATATACCATGAAA CATCACCATCACCATCACTCCGCGGGTCTGGAAGTCTGTTCAGGGGGCCCATGCCTGCC CGCGCGCTTCTGCCAGGGCCATGGGGCATCGTACTCTAGCCTCCACTCCTGCCCTGTGG GCCTCCATCCCGGCTCCTCGCTCCATCCCGGCTCCTCGTCCATCCCGGCTCCTCGCTCC ATCCCGGCTCCTCGCTCCATCCCGGCTCCTCGTCCATCCCGGCTCCTCGTCCATCCCGG GCTCCTCGTCCATCCCGGCTCCTCGTCCATCCCGGCTCCTCGTCCATCCCGGCTCCTCG CGCTCCATCCCGGCTCCTCGTCCATCCCGGCTCCTCGTCCATCCCGGCTCCTCGCTCC ATCCCGGCTCCTCGTCCATCCCGGCTCCTCGTCCATCCCGGCTCCTCGCTCCTCATCC</p>
hOGG1 _{C241A}	<p>>C241A 2F_T7</p> <p>NNNRNGWAATTCCTCTAGAATAATTTGTTTAACTTTAAGAAGGAGATATACCATGAAA CATCACCATCACCATCACTCCGCGGGTCTGGAAGTCTGTTCAGGGGGCCCATGCCTGCC CGCGCGCTTCTGCCAGGGCCATGGGGCATCGTACTCTAGCCTCCACTCCTGCCCTGTGG GCCTCCATCCCGGCTCCTCGCTCCATCCCGGCTCCTCGTCCATCCCGGCTCCTCGCTCC TCTTCCGGTGGAGGACAAAGTCTCTGCACACTGGAGTGGTACTAGCGGATCAAGTA TGGACACTGACTCAGACTGAGGAGCAGCTCCACTGCCTGTGTACCGAGGAGCAAGAGC CAGGCTAGCAGGCCACACAGAGCAGGCTGGAGGCCGTGGCAGTACTTCCAGCTAGAT GTACCTCGCTCAACTGTATCACCACTGGGGTCCCGTGGACTCCACTTCCAGAGGTG GCTCGAATTTCCAGAGTGTGSACTCTGCGACAGAGCCCATGATGCTTTCTCT TTTATCTGTTCTCCACAAACACATCCCGGCTCCTCGTCCATCCCGGCTCCTCGCTGTC CAGGCTTTGACCTCGGCTCATCCAGCTTGTATGATGATCACTACCATGGCTTCCCGAGC CTGCAGGCCCTGGCTGGCCAGAGGTGGAGGCTCATCTCAGGAAGCTGGGCTGGGCTAT CGTGCCCTTACGTGAGTCCAGTGCOCGAGCCATCTGGAAGAACAGGGCCGGCTAGCC TGGCTGCAGCAGTACGAGAGTCTCATATGAGGAGGSCCAGAGGCCCTCGCTATCCCTG CCAGGCCCTCCGATCTCCAGGGCCCTCGCTATCTCGCAGGCCCTCGCTATCTCGCA</p>

<i>hOGG1</i> _{C253A}	<pre> >C253A_4F_T7 NNWVACATTCCTCTAGAAFAATTTTGTTFACCTTTRAGAAGGAGATATACCATGAAC ATCACCATACACATCACTCCGCGGGTCTGGAGTTCTGTTCACGGGGCCCATGCCTGGCC GGCGGTTCTGGCCAGGGCGATGGGGCATGTACTGTAGCTCCACTCCCTGCTGGG CTCCATCCCGTGCCTCGCTCTGAGCTGGCCTGGAACCTGGTTCTGCCTCTGGACAAT CTTCCGGTGGAGGGAGCAAAGTCTGCACACTGGAGTGGTGTACTAGCGGATCAAGTAT GGACTGACTCAGACTGAGGAGCAGCTCCACTGCACTGTGTACCGAGGACACAGAGCC AGCTAGCAGGCCACACACAGAGCTGGAGGCTCCCGCAGTACTTCAGCTAGATG TTACCTGGCTCAACTGTATCACCACTGGGGTCCCGTGGACTCCCACTTCCAAAGAGTGG CTCAGAAATCCAAAGTGTGCACACTGTGGGACAAAGCCCATCCGATGCTTTTCTCTT TTATCTGTCTCCACCAACAACATCGCCGCATCACTGGCATGGTGGAGCGGCTGTGCC AGGCTTTGGACCTCGGCTCATCCAGCTTGATGATGCACCTACCATGGCTTCCCGAGCC TGCAGGCTGTGCTGGCCAGAGTGGAGCTCATCTCAGAGACTGGCCGGCTATC GTGCCGTTACCTGAGTGCACCTGCCCGAGCCATCCTGGAGAACAGGCGGGCTAGCT GGCTCAGCAGCTACGAGTCTCTATATGAGGAGGCCCAAGGCCCTGTGCATCTCCG CTGGAGTGGCCACCAAGTGTGCTGACGATCTGCTAGTGGCTGACCTATCTGCCTA </pre>
-------------------------------	---

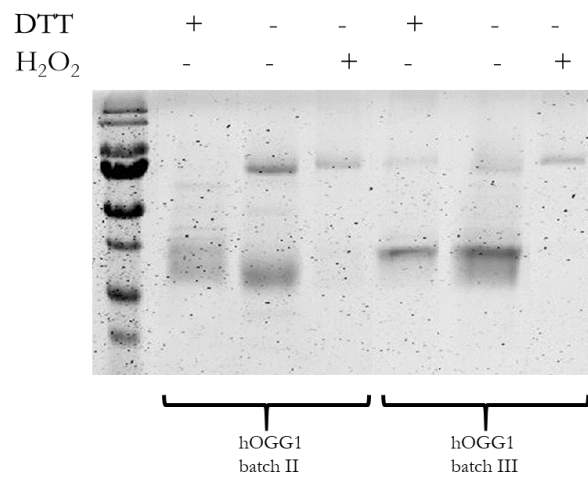


Figure x5.3: Duplicate and triplicate non-reducing SDS-PAGE based hOGG1 dimerization experiments. Batch II and batch III of hOGG1_{wt} purifications showed dimer formation upon oxidation and monomer formation when treated with DTT. Batch II showed higher ratio of dimer species in original purification compared to batch III. Hence under ambient conditions, dimer band is more intense for batch II in comparison to batch III (Fig. x5.1).

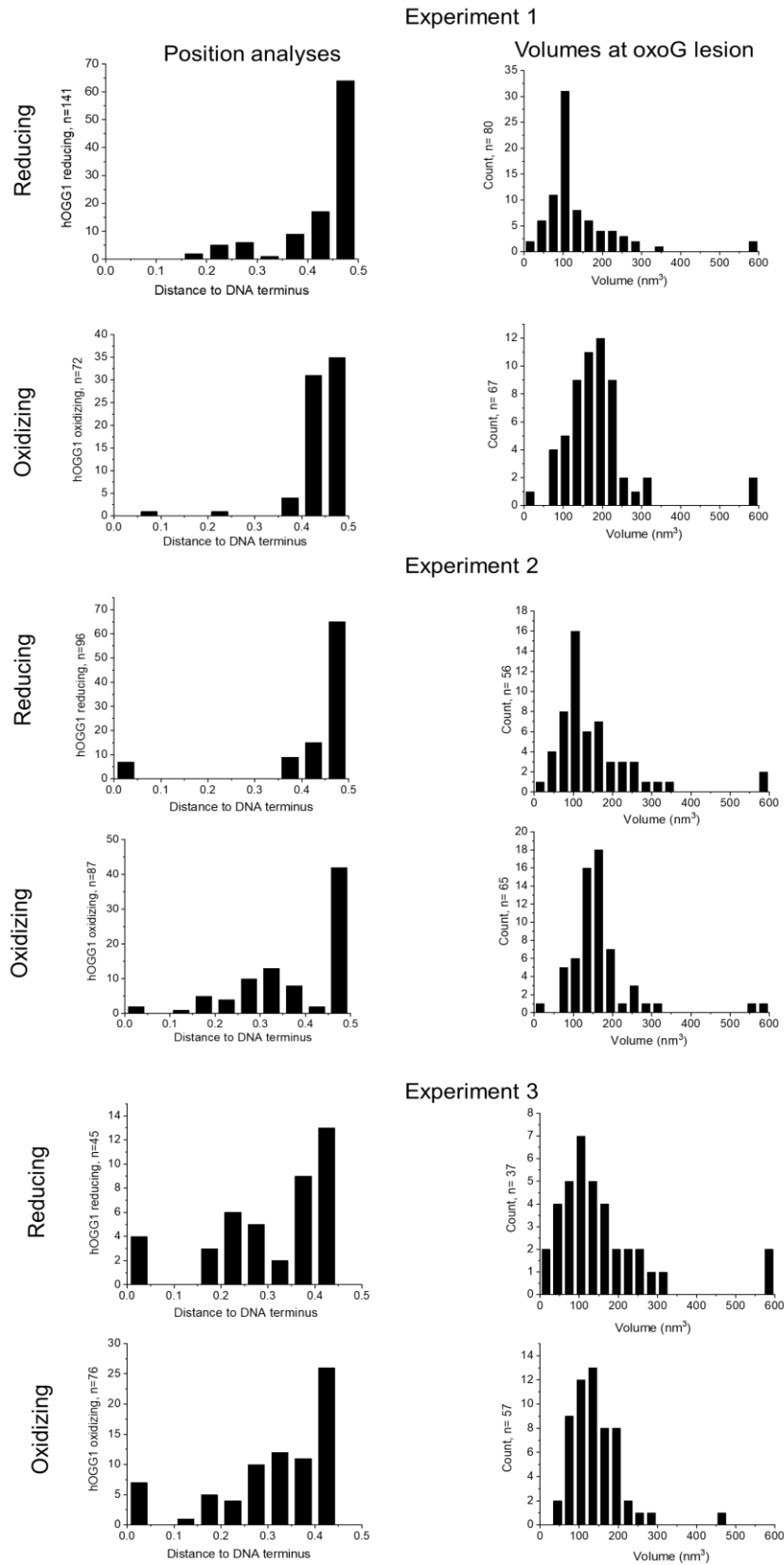


Figure x5.4: Individual AFM experiments of hOGG1 with oxoG lesion containing substrate. Pooled data is shown in Fig. 5.5 and Fig. 5.6.

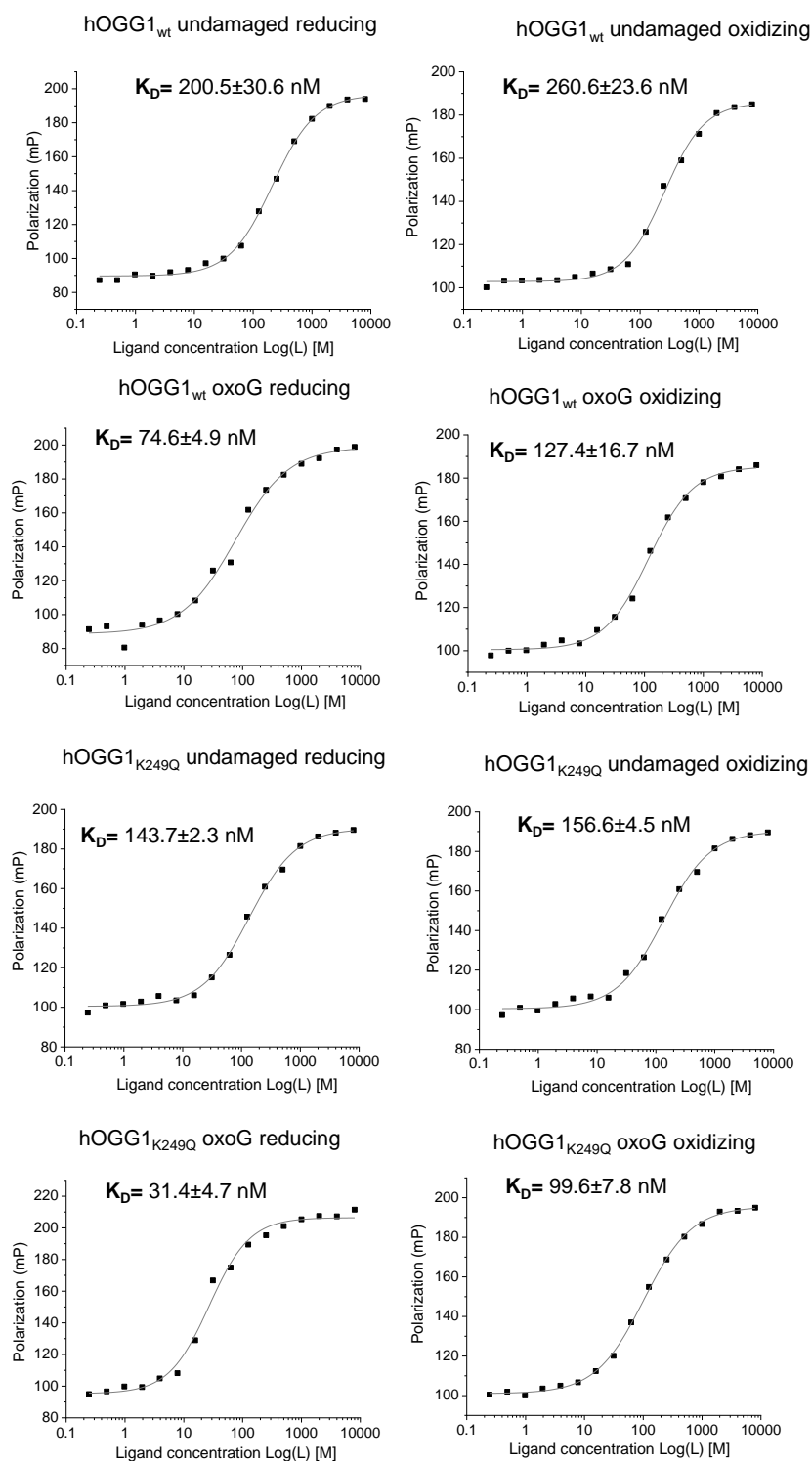


Figure x5.5: Exemplary individual fluorescence polarization curves from triplicate measurements of $hOGG1_{wt}$ and $hOGG1_{K249Q}$ with oxoG and undamaged substrates under oxidizing and reducing conditions. The binding affinities (K_D) of proteins for each measurement are indicated in the plots. The values from the triplicate measurements were used to determine the average binding affinities, $K_D(\pm SD)$. The individual and average $K_D(\pm SD)$ for all measurements are mentioned in the Table x5.1.

Table x5.1: K_D values (\pm SD) from triplicate measurements of hOGG1_{wt} and hOGG1_{K249Q} with oxoG and undamaged substrates under oxidizing and reducing conditions.

Incubation sample	~ K_D values (nM) with SD			
	Experiment 1	Experiment 2	Experiment 3	Triplicate pooled
hOGG1 oxoG reducing	62.4 \pm 8.5	63.7 \pm 12.6	74.6 \pm 4.9	66.8 \pm 4.7
hOGG1 oxoG oxidizing	127.4 \pm 16.7	139.8 \pm 15.9	153.7 \pm 5.8	140.4 \pm 9.3
hOGG1 undamaged reducing	200.5 \pm 30.6	319.5 \pm 51.8	235.9 \pm 19.3	252.1 \pm 43.2
hOGG1 undamaged oxidizing	220.6 \pm 20.1	267.8 \pm 8.4	260.6 \pm 23.6	249.8 \pm 18.0
hOGG1 K249Q oxoG reducing	31.4 \pm 4.7	45.5 \pm 2.1	40.5 \pm 8.6	39.1 \pm 5.0
hOGG1 K249Q oxoG oxidizing	90.6 \pm 3.3	99.9 \pm 6.7	99.6 \pm 7.8	96.7 \pm 3.8
hOGG1 K249Q undamaged reducing	146.6 \pm 6.5	149.8 \pm 5.9	143.7 \pm 2.3	146.8 \pm 2.1
hOGG1 K249Q undamaged oxidizing	156.6 \pm 4.5	178.6 \pm 20.7	183.3 \pm 15.2	172.8 \pm 10.1

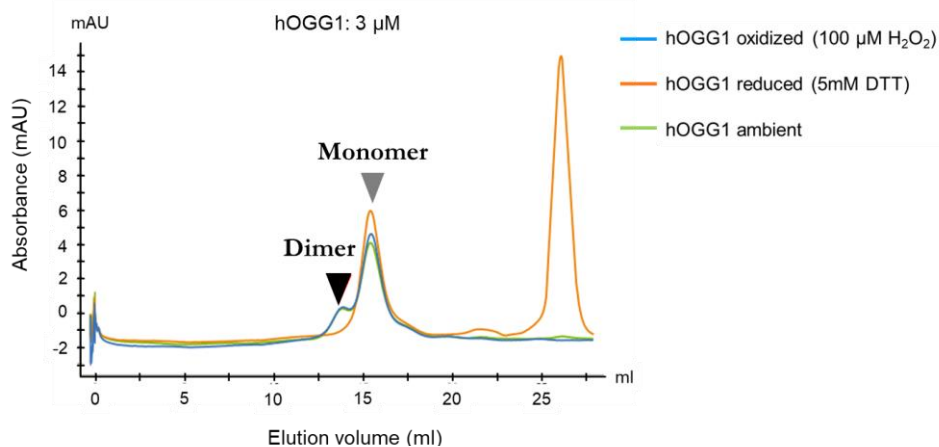


Figure x5.6: aSEC experiment with 3 μ M hOGG1 from batch I. Under reducing conditions and lower concentration (see Fig. 5.7b for aSEC experiments with 30 μ M hOGG1), dimer species (black arrow) is not observed. Under oxidizing and ambient conditions, lower ratio of dimer to monomer species is seen for 3 μ M hOGG1 compared to 30 μ M hOGG1 (Fig. 5.7b).

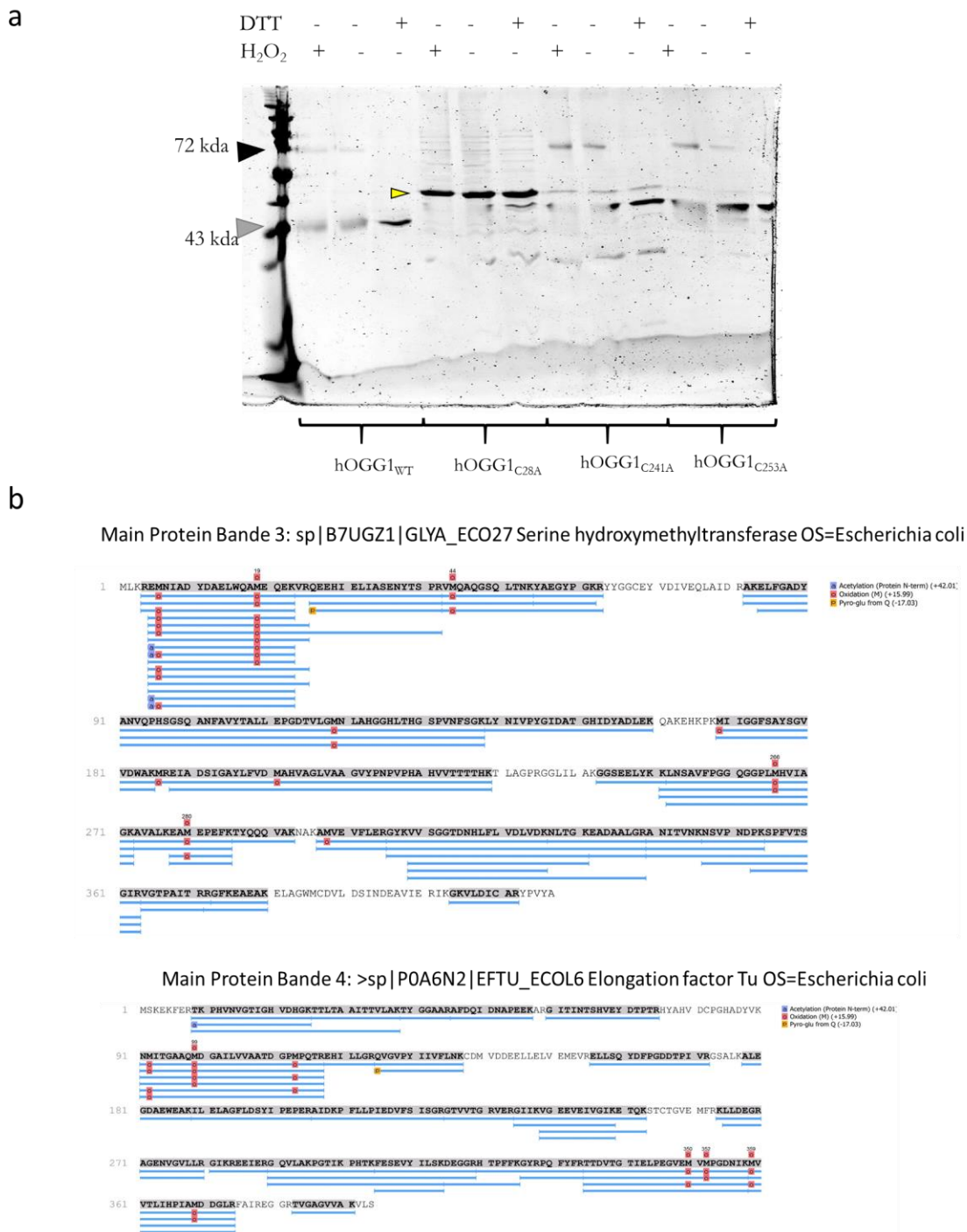


Figure x5.7: Non-reducing SDS-PAGE based dimerization experiments for hOGG1 cysteine mutants. (a) Wild type hOGG1, hOGG1_{C241A} and hOGG1_{C253A} showed dimer formation (black arrow) under oxidative conditions. However, hOGG1_{C28A} did not dimerize. Monomer species are indicated by grey arrow. All hOGG1 cysteine mutants had an additional band (yellow arrow) above monomer species. (b) Mass spectrometry of these bands revealed them to be contaminant proteins from expression cells, namely GLYA_ECO27 and EFTU_ECOL6. The MS results for these bands are shown above. Western blot and MS results for the dimer band of hOGG1_{wt} are shown in Fig. 5.8.

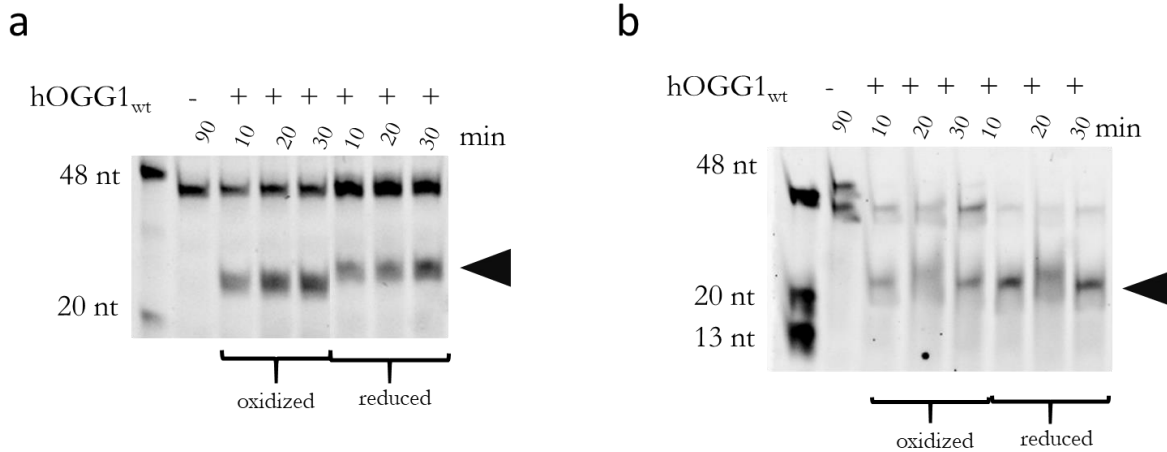


Figure x5.8: Duplicate (a) and triplicate (b) experiments showing gel-based hOGG1 activity assays to demonstrate hOGG1 activity under oxidative and reductive environments. Black arrows indicate the cleaved products by hOGG1. The cut and uncut (48 bp) gel bands were measured for their intensities. After correcting for background grey values, percent product was taken as average (\pm SD) over first three time points and plotted as a histogram (Refer to Fig. 5.10).

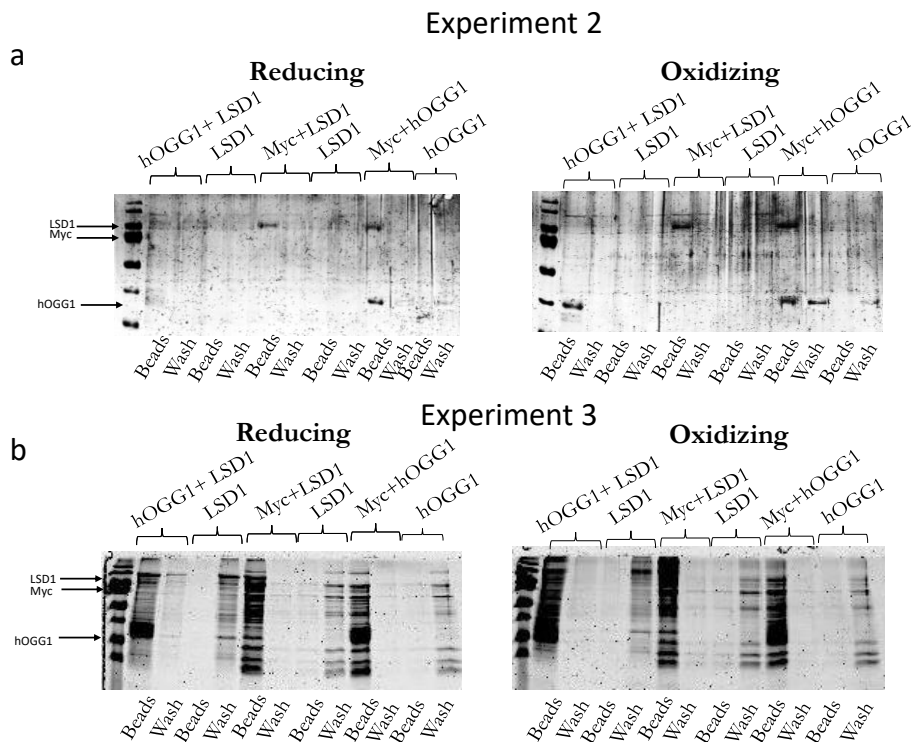


Figure x5.9: Duplicate (a) and triplicate (b) experiment showing in vitro pull-down studies to demonstrate interactions between hOGG1, Myc and LSD1 under oxidizing and reducing conditions.

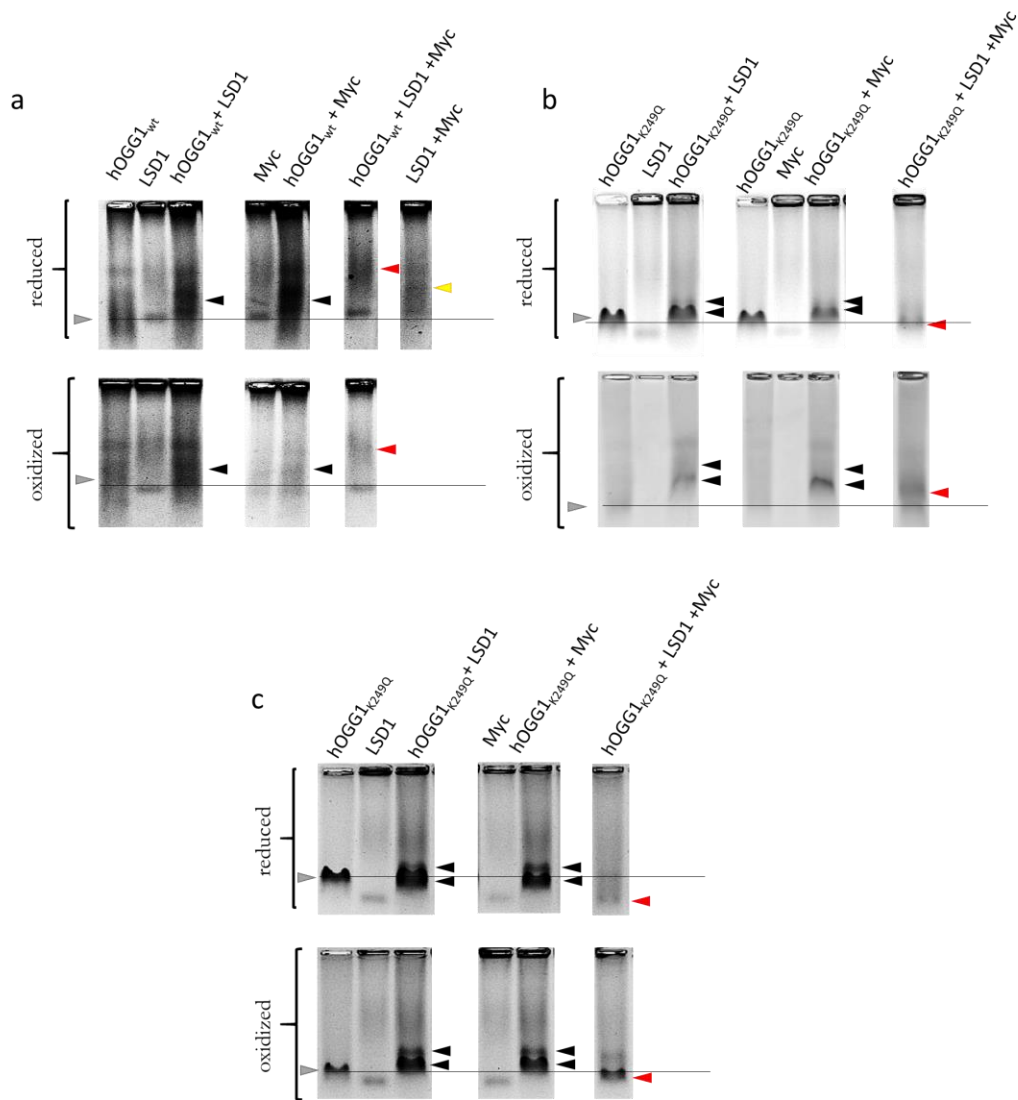


Figure x5.10: Duplicate experiment showing Native PAGE analysis to demonstrate interactions between hOGG1, Myc and LSD1 under oxidizing and reducing conditions. (a) NAGE assays with hOGG1_{wt}. Grey arrow indicates height of hOGG1 alone. Black arrows represent the shift for dual protein complexes of hOGG1 with other two components, the grey arrow shows the band height of hOGG1 alone, red arrow indicates the triple complex and yellow arrow indicates the height of Myc-LSD1 complex. Black line drawn shows the differences between the heights of hOGG1 alone and other dual and triple complexes. (b) and (c) NAGE assays with hOGG1_{K249Q}. Discrepancies in the gel shifts were observed to different running behavior of the gels (different chambers were used for running gels and thus the running behavior was different).

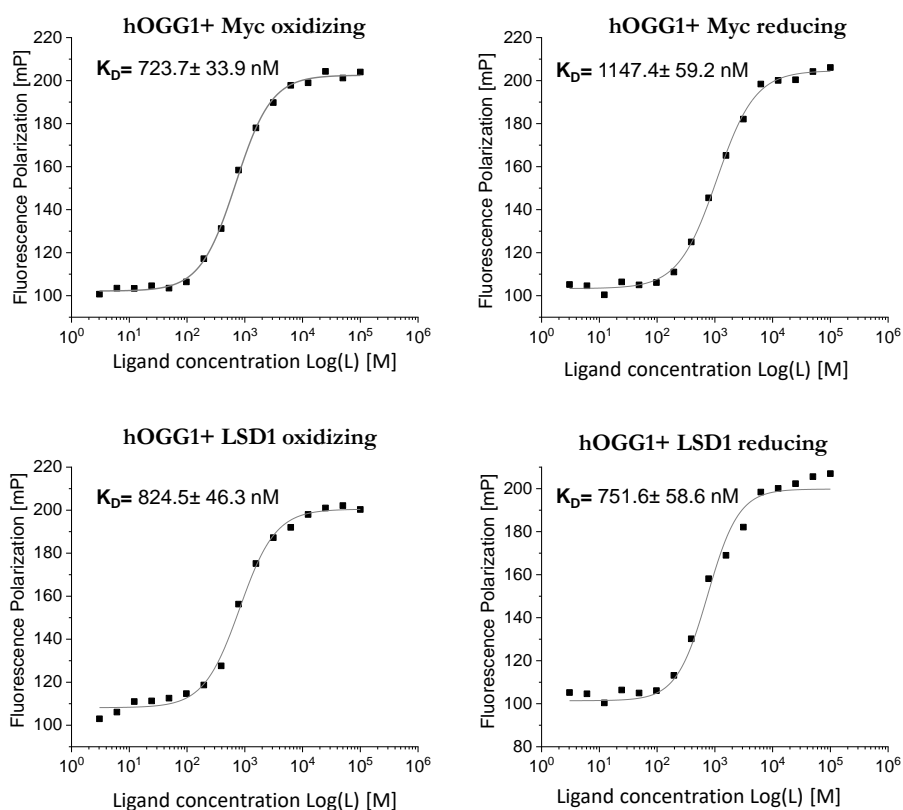
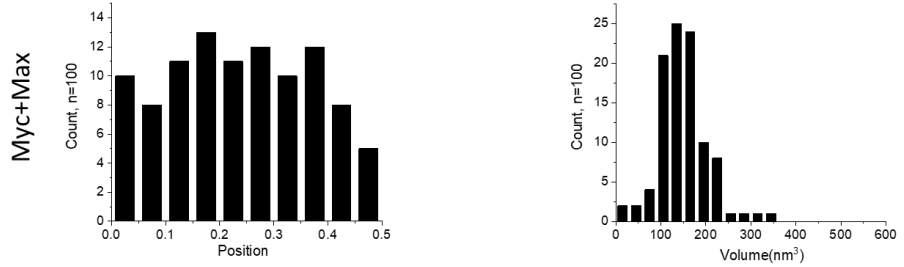


Figure x5.11: Exemplary individual fluorescence polarization curves from triplicate measurements of $hOGG1_{wt}$ with Myc and $hOGG1_{wt}$ with LSD1 under oxidizing and reducing conditions. The binding affinities (K_D) for each measurement are indicated in the plots. The values from the triplicate measurements were used to determine the average binding affinities, $K_D(\pm SD)$. The individual and average $K_D(\pm SD)$ for all measurements are mentioned in the Table x5.2.

Table x5.2: K_D values ($\pm SD$) from triplicate measurements of $hOGG1_{wt}$ with Myc and $hOGG1_{wt}$ with LSD1 under oxidizing and reducing conditions.

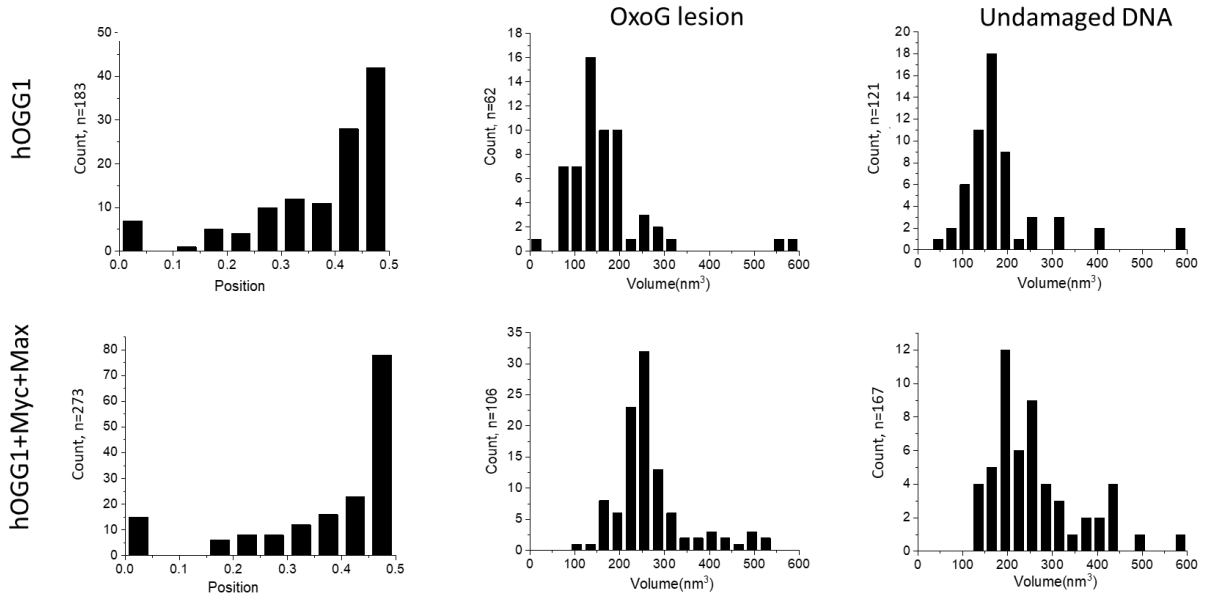
Protein sample	~ K_D values (nM) with SD			
	Experiment 1	Experiment 2	Experiment 3	Triplicate pooled
hOGG1+ Myc reducing	1069.3 \pm 78.6	978.5 \pm 36.5	1147.4 \pm 59.2	1065.0 \pm 69.0
hOGG1+ Myc oxidizing	646.6 \pm 51.4	746.8 \pm 72.8	723.7 \pm 33.9	705.7 \pm 42.8
hOGG1+ LSD1 reducing	812.6 \pm 23.9	751.6 \pm 58.6	696.7 \pm 70.2	753.5 \pm 47.4
hOGG1+ LSD1 oxidizing	907.7 \pm 25.4	869.8 \pm 35.9	824.5 \pm 46.3	867.2 \pm 34.0

a Experiment I



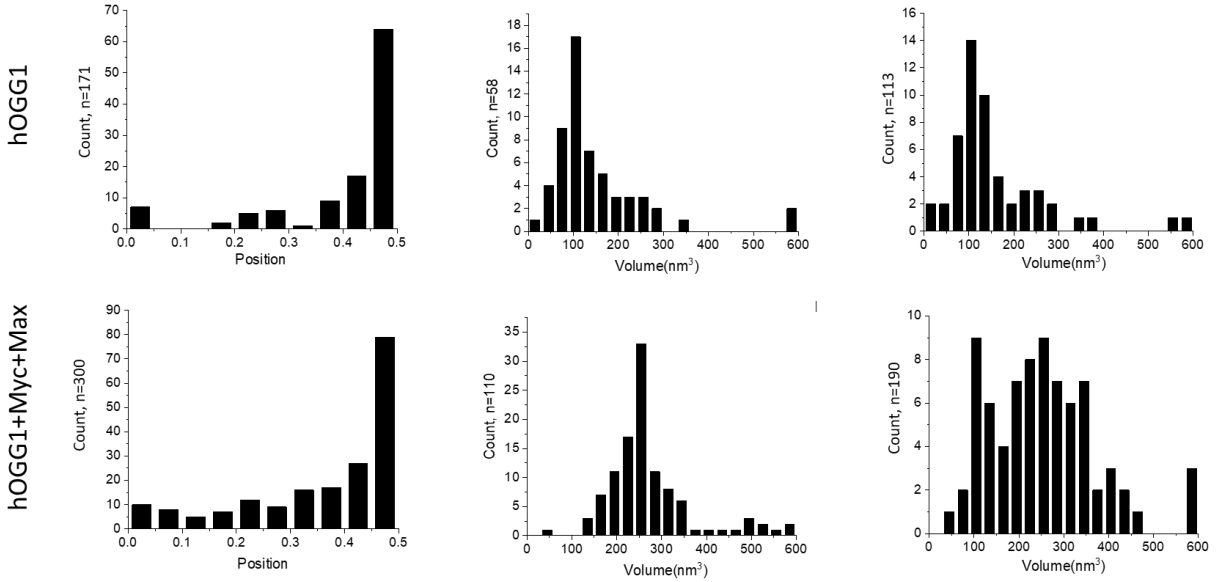
b

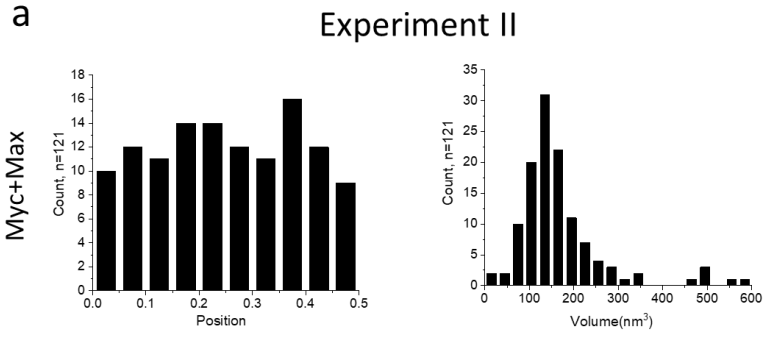
Oxidizing



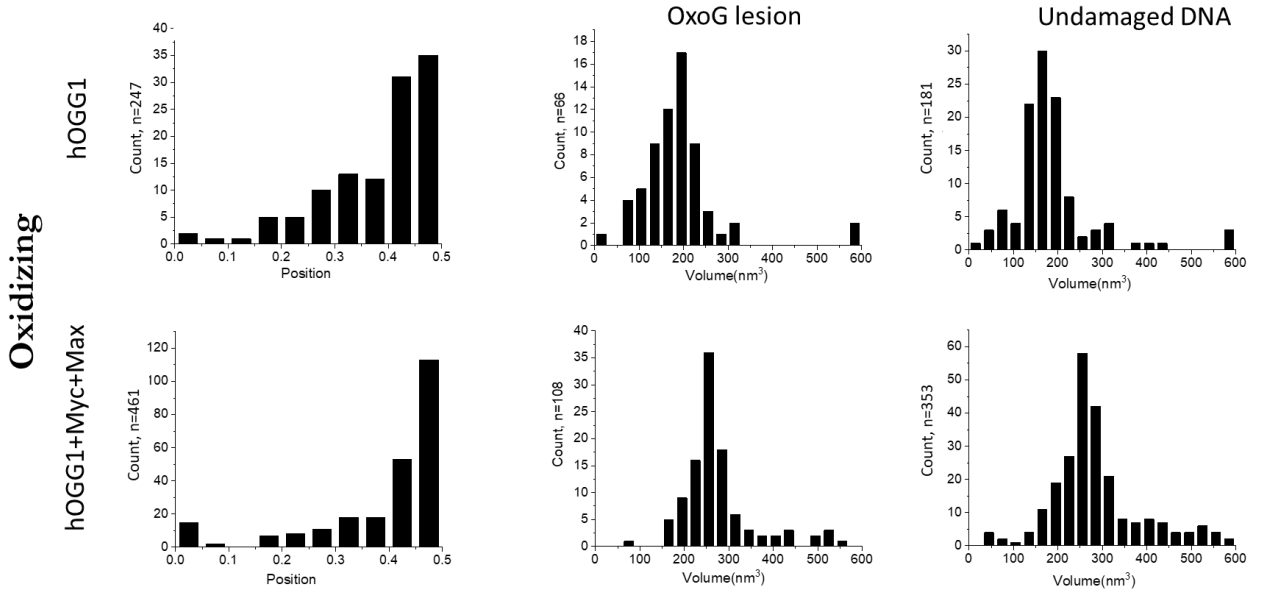
c

Reducing

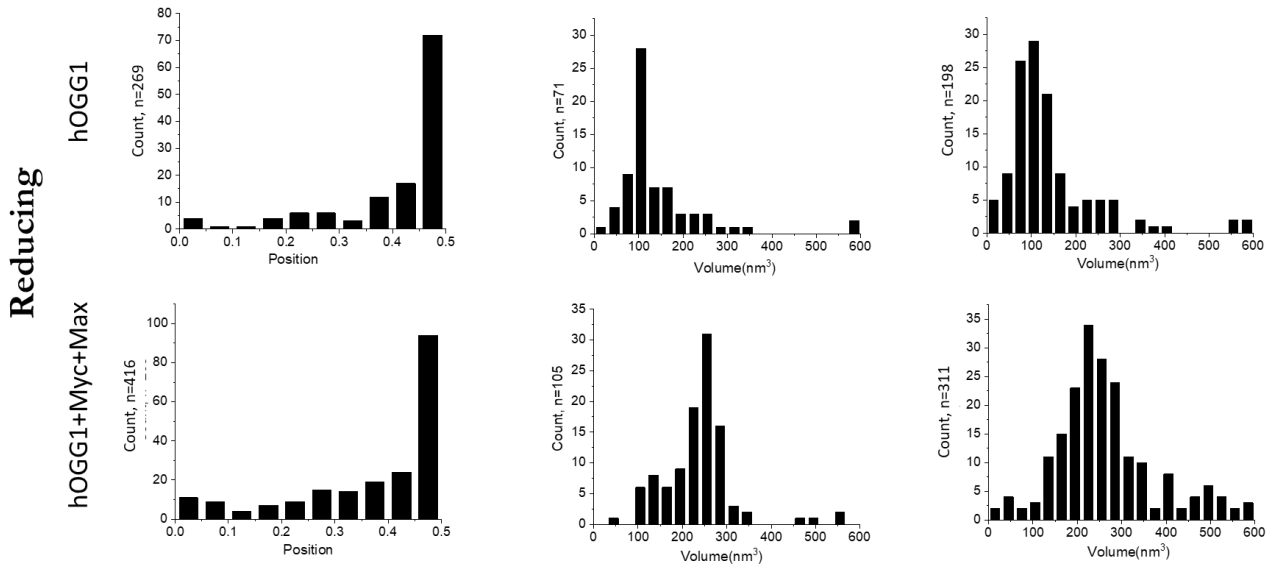




b



c



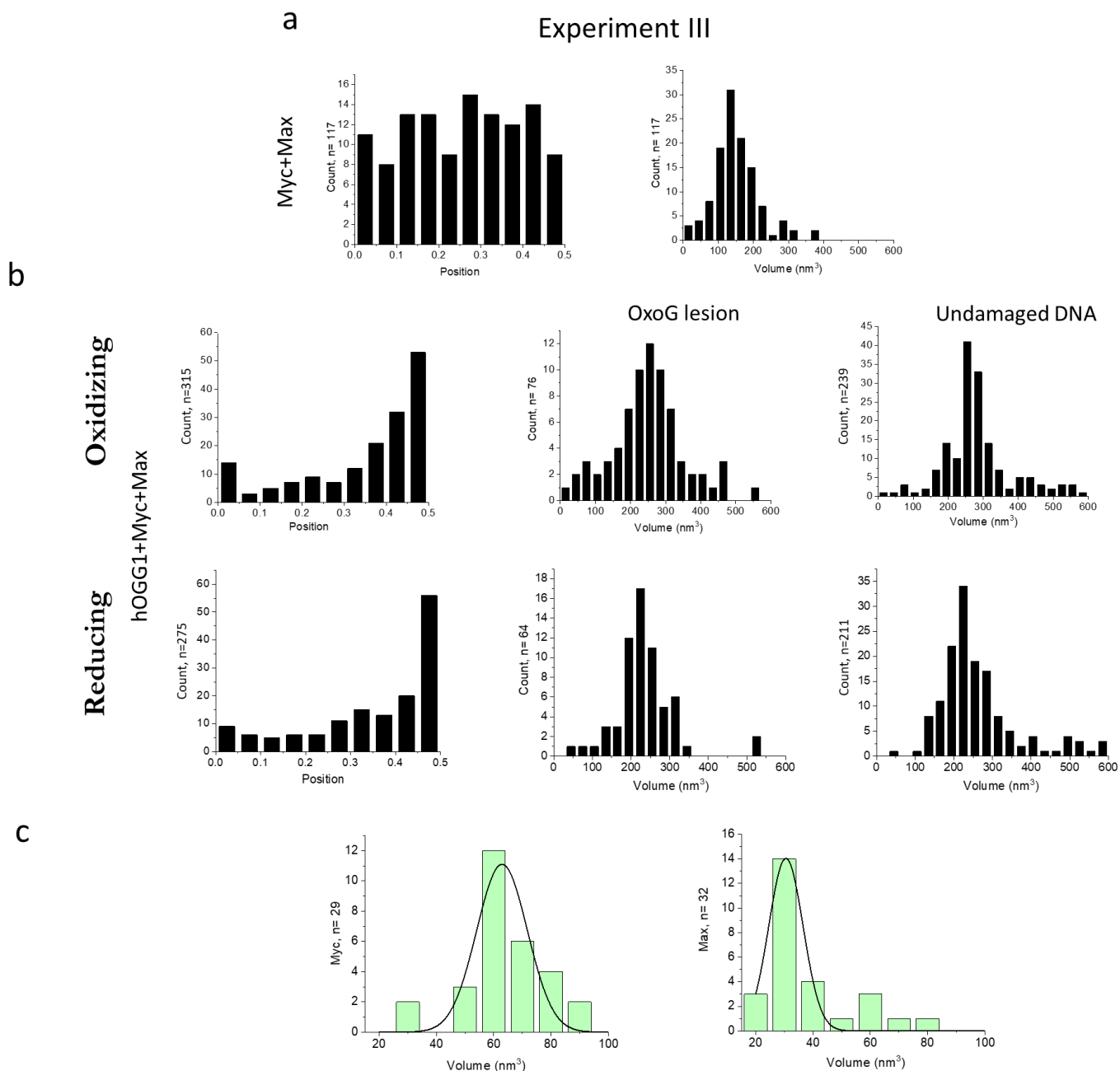


Figure x5.12: Individual AFM experiments of Myc recruitment by hOGG1 binding to oxoG lesion. Pooled results for the triplicate have been shown in the Fig. 5.15. MycMax control samples do not show any specificity towards oxoG (50% position). Positions and volumes for hOGG1-MycMax complexes were combined from the above three experiments and data is shown in Fig. 5.15. hOGG1 control samples were obtained from experiment I and II only to avoid redundancy. Results were reproducible in all three experiments and were therefore pooled. The volumes of the free (non-DNA-bound) Myc (~48 kDa) and Max (~18 kDa) proteins were $63 \pm 12 \text{ nm}^3$ and $31 \pm 7 \text{ nm}^3$, respectively (experiment 3, c).

Experiment I

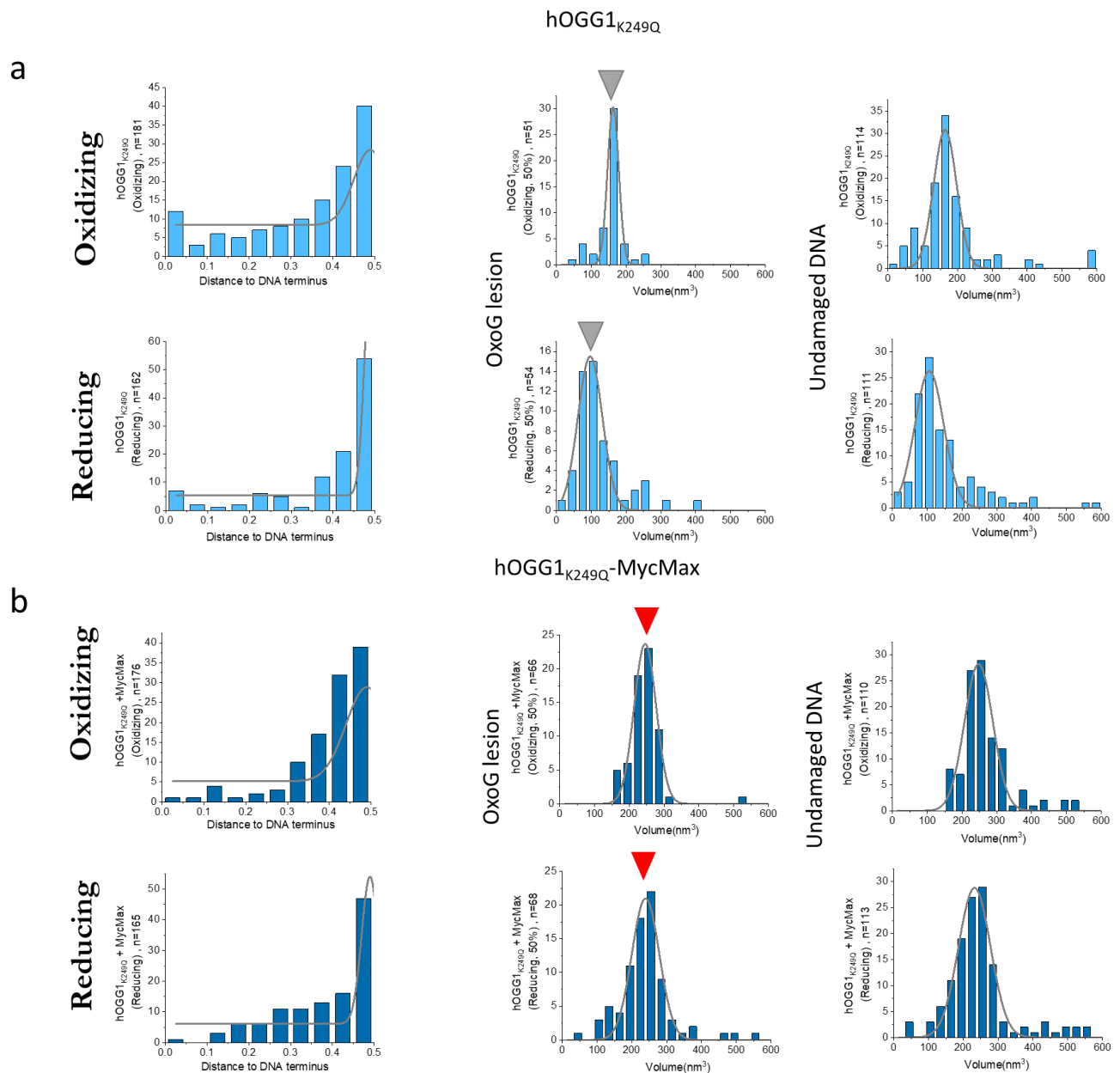


Figure x5.13: hOGG1_{K249Q} and Myc/Max interaction on DNA. (a) Position (left) and volume distributions for hOGG1 on oxoG DNA (with oxoG at 50% of the DNA length). The middle panel shows hOGG1 volumes for complexes bound at the oxoG lesion, the right panel shows volumes on undamaged DNA sites. The top and bottom rows represent experiments under oxidizing and reducing conditions, respectively. hOGG1 showed volumes of ~162 nm³ under oxidizing and ~105 nm³ under reducing conditions (indicated by grey arrows). **(b)** Position (left) and volume distributions for hOGG1_{K249Q}+Myc+Max complexes. The middle column shows complex volumes at oxoG lesion positions, the right panel at undamaged DNA sites. Top and bottom row again represent oxidizing and reducing conditions, respectively. The volumes for hOGG1_{K249Q}+Myc+Max complexes are shifted to higher volumes compared to those of Myc+Max alone (Fig. 5.15a) or hOGG1_{K249Q} alone (a), ranging from ~200 nm³ to ~260 nm³ (red arrows) at oxoG lesions and

undamaged sites under both oxidizing and reducing conditions, suggesting interaction between hOGG1_{K249Q} and Mvc/Max.

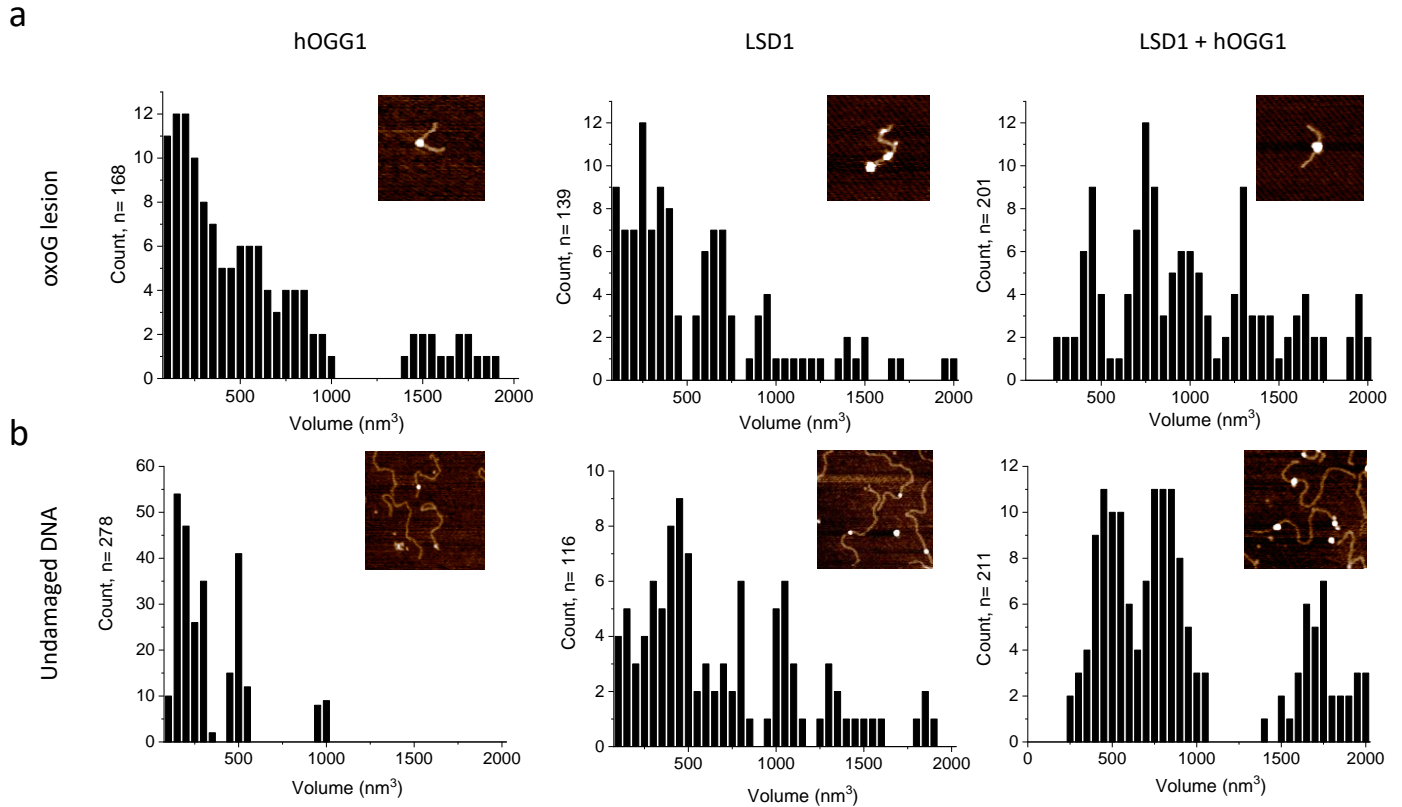
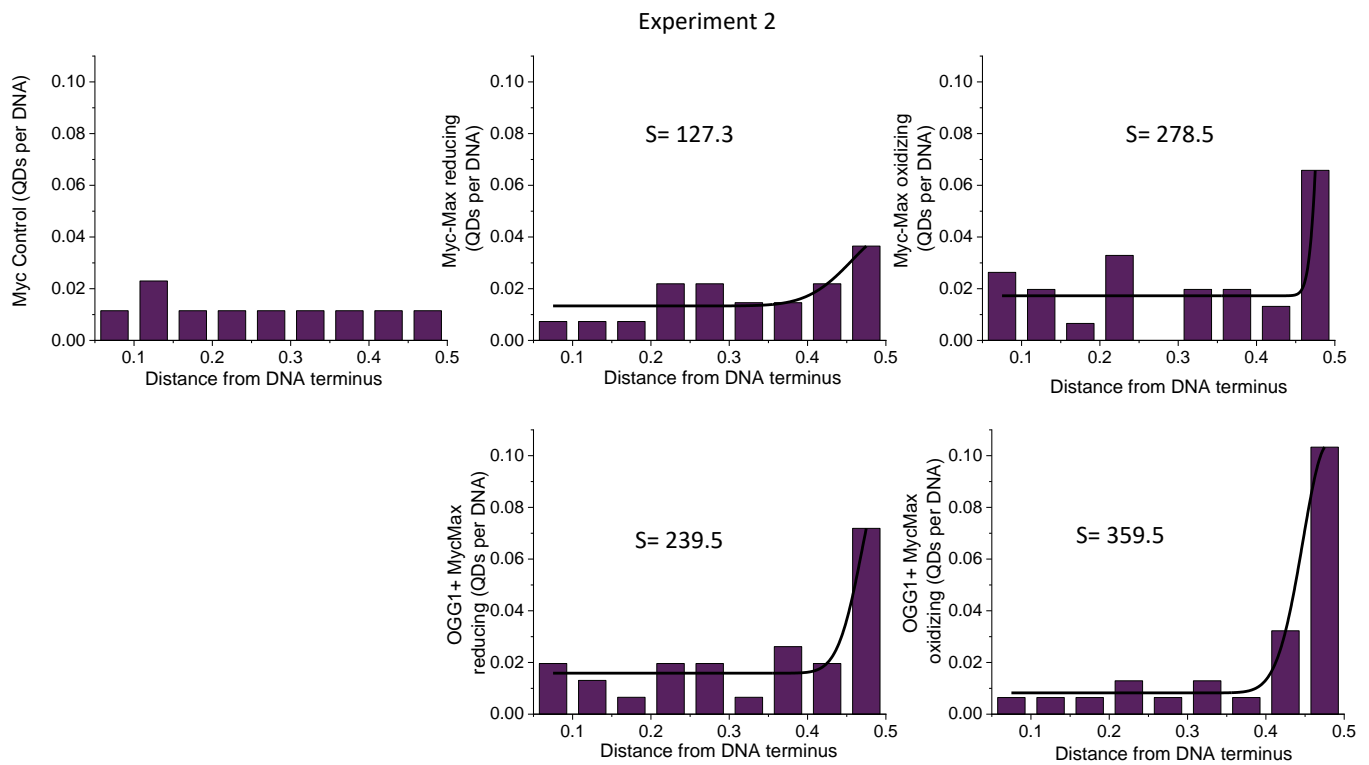
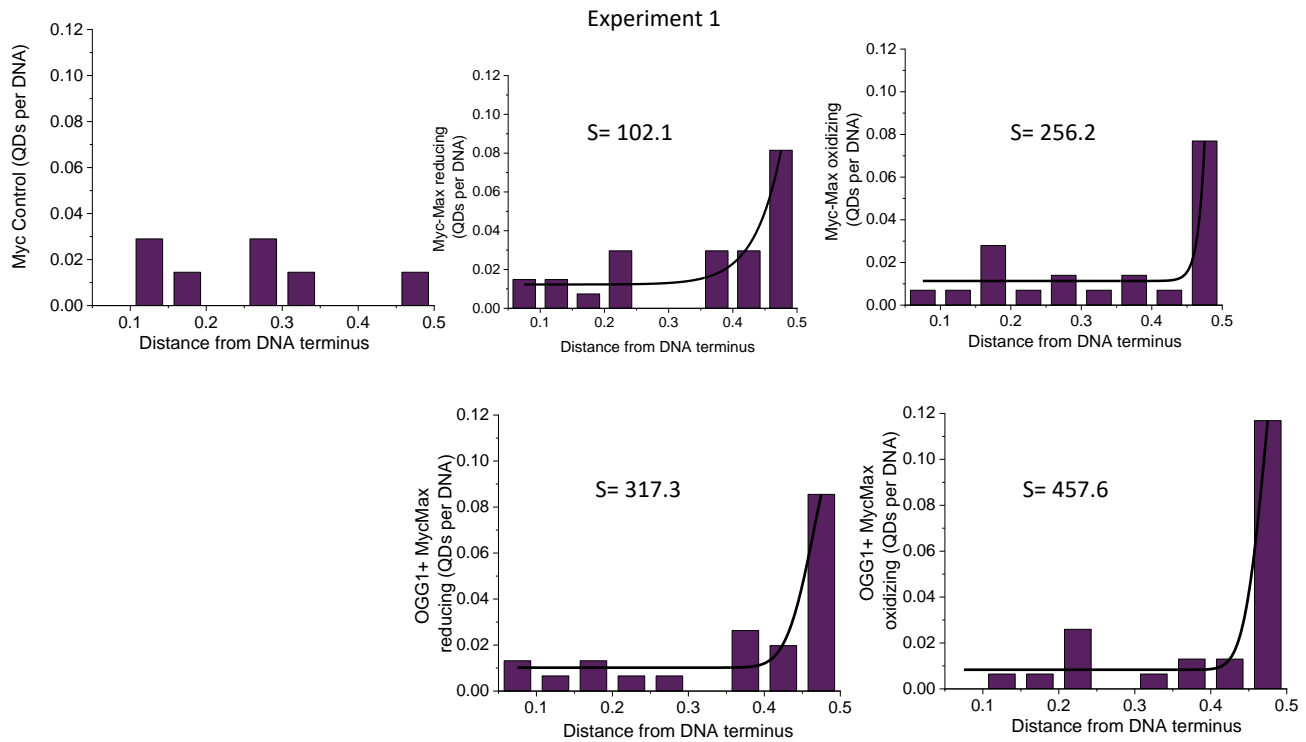


Figure x5.14: hOGG1 and LSD1 interaction on DNA. hOGG1 and LSD1 were incubated together and crosslinked on DNA using 0.1% glutaraldehyde. While crosslinked hOGG1 (in the absence of LSD1) showed a maximum volume between ~150 nm³ and ~500 nm³ at both oxoG lesion (a) and undamaged sites in the DNA (b), LSD1 showed larger volumes between ~250 nm³ and ~400 nm³. In the samples incubated with both hOGG1 and LSD1, higher volumes up to ~750 nm³ were observed supporting an interaction between hOGG1 and LSD1 at the oxoG lesion (a) as well as on undamaged DNA (b). Extremely high volumes are observed in these experiments due to crosslinking.



Experiment 3

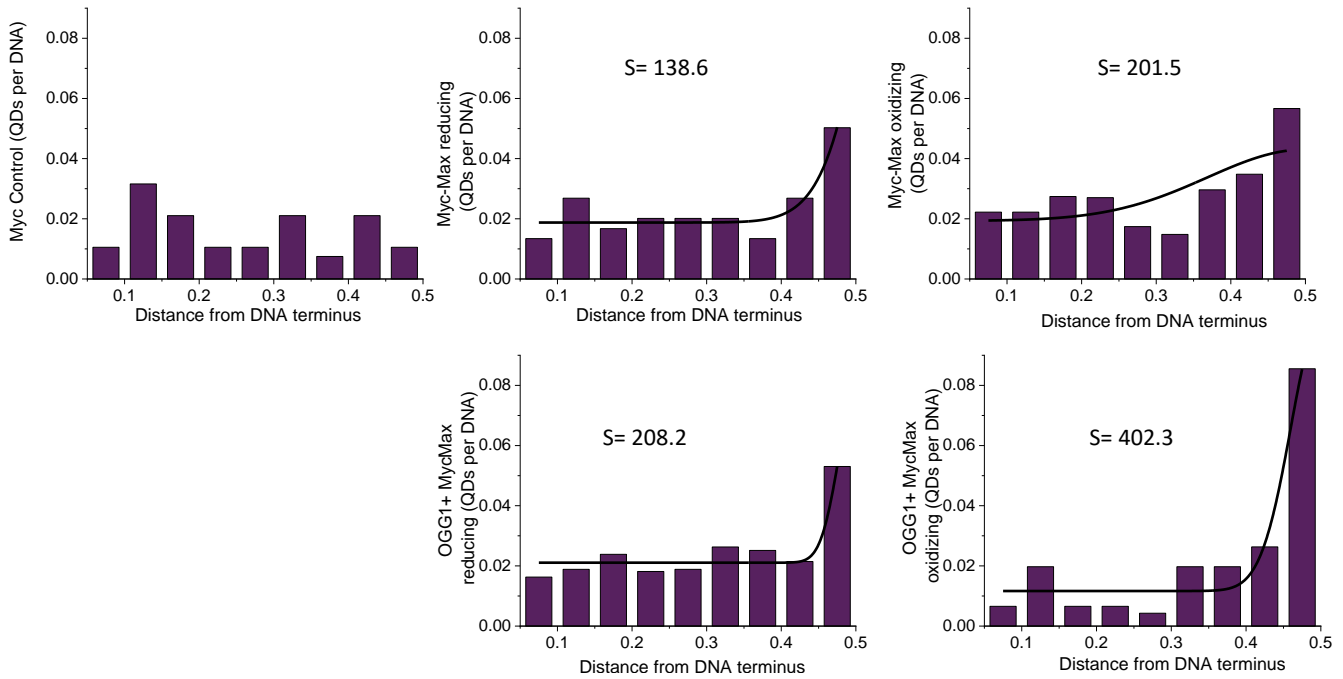


Figure x5.15: Triplicate AFM experiments of hOGG1 mediated enhancement of QD labelled Myc binding to E-box. Position distributions of QD labelled Myc-Max (QD-Myc-Max) complexes bound to E-box in the presence and absence of hOGG1 for all three experiments (pooled data are shown in Fig. 5.16) are shown above. QD-Myc was included as a control. Myc alone does not show any specificity towards E-box. Fits were produced by Gaussian fitting to the distribution. Binding specificities (S) (indicated in the distributions) were calculated as described in the section 2.2.3.5c. These values were used to calculate average specificities ($\pm SD$) (Fig. 5.16f).

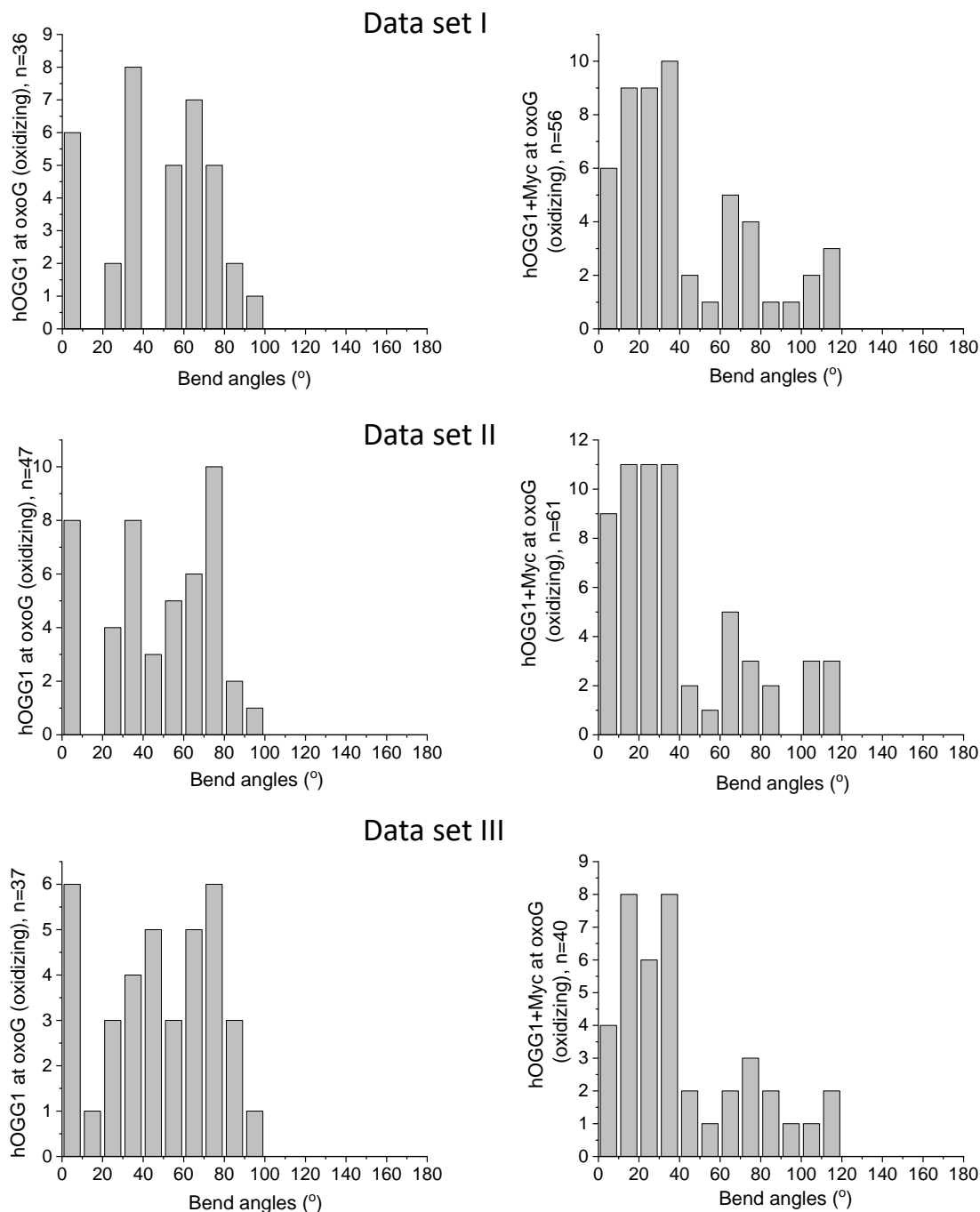


Figure x5.16: Triplicate automated AFM MatLab analyses showing the DNA bend angles for hOGG1 and hOGG1-Myc complexes at oxoG lesion in the oxidative environment. Results were reproducible in all three experiments and were therefore pooled (Fig. 5.17). hOGG1 samples with oxoG lesion containing substrate was included as a control. Change in bend angle states of hOGG1(0°, ~33° and ~69°) at oxoG lesion were clearly seen in the presence of Myc (~23° and ~74°) in all three experiments.

Literature

1. Halford, S. E. & Marko, J. F. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040–3052 (2004).
2. Bangalore, D. M. & Tessmer, I. Unique insight into protein-DNA interactions from single molecule atomic force microscopy. *AIMS Biophys.* **5**, (2018).
3. Binnig, G., Quate, C. F. & Gerber, C. Atomic force microscope. *Phys. Rev. Lett.* **56**, 930–933 (1986).
4. Li, M. Z. & Elledge, S. J. SLIC: A method for sequence- and ligation-independent cloning. *Methods Mol. Biol.* **852**, 51–59 (2012).
5. Farina, A., Faiola, F. & Martinez, E. Reconstitution of an E box-binding Myc:Max complex with recombinant full-length proteins expressed in *Escherichia coli*. *Protein Expr. Purif.* **34**, 215–222 (2004).
6. Baluapuri, A. *et al.* Article MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation Article MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation. 1–14 (2019). doi:10.1016/j.molcel.2019.02.031
7. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
8. Puljung, M. C. & Zagotta, W. N. Labeling of Specific Cysteines in Proteins Using Reversible Metal Protection. *Biophysj* **100**, 2513–2521 (2011).
9. Hou, E. W., Prasad, R., Asagoshi, K., Masaoka, A. & Wilson, S. H. Comparative assessment of plasmid and oligonucleotide DNA substrates in measurement of in vitro base excision repair activity. *Nucleic Acids Res.* **35**, 1–10 (2007).
10. Pan, L. *et al.* OGG1-DNA interactions facilitate NF- κ B binding to DNA targets. *Sci. Rep.* **7**, 1–10 (2017).
11. Wilson, N. R. & MacPherson, J. V. Carbon nanotube tips for atomic force microscopy. *Nat. Nanotechnol.* **4**, 483–491 (2009).
12. Tessmer, I., Kaur, P., Lin, J. & Wang, H. Investigating bioconjugation by atomic force microscopy. *J. Nanobiotechnology* **11**, 1–17 (2013).
13. Ando, T., Uchihashi, T. & Kodera, N. High-speed AFM and applications to biomolecular systems. *Annu. Rev. Biophys.* **42**, 393–414 (2013).
14. Ruhnnow, F., Zwicker, D. & Diez, S. Tracking single particles and elongated filaments with nanometer precision. *Biophys. J.* **100**, 2820–2828 (2011).
15. Schneider, C. A., Rasband, W. S., Eliceiri, K. W. & Instrumentation, C. NIH Image to ImageJ : 25 years of Image Analysis. **9**, 671–675 (2017).
16. Buechner, C. N., Maiti, A., Drohat, A. C. & Tessmer, I. Lesion search and recognition by thymine DNA glycosylase revealed by single molecule imaging.

Nucleic Acids Res. **43**, 2716–2729 (2015).

17. Bangalore, D. M. *et al.* Automated AFM analysis of DNA bending reveals initial lesion sensing strategies of DNA glycosylases. *Sci. Rep.* **10**, (2020).
18. Emperle, M. *et al.* Structural and biochemical insight into the mechanism of dual CpG site binding and methylation by the DNMT3A DNA methyltransferase INTRODUCTION. 1–15 (2021).
19. Rill, N., Mukhortava, A., Lorenz, S. & Tessmer, I. Erratum: Alkyltransferase-like protein clusters scan DNA rapidly over long distances and recruit NER to alkyl-DNA lesions (Proceedings of the National Academy of Sciences of the United States of America (2020) 117 (9318-9328) DOI: 10.1073/pnas.1916860117). *Proc. Natl. Acad. Sci. U. S. A.* **117**, 24001 (2020).
20. Rill, N., Mukhortava, A., Lorenz, S. & Tessmer, I. Alkyltransferase-like protein clusters scan DNA rapidly over long distances and recruit NER to alkyl-DNA lesions. **117**, (2020).
21. Beckwitt, E. C. *et al.* Single molecule analysis reveals monomeric XPA bends DNA and undergoes episodic linear diffusion during damage search. *Nat. Commun.* 1–14 doi:10.1038/s41467-020-15168-1
22. Parks, J. W., Kappel, K., Das, R. & Stone, M. D. Single-molecule FRET-Rosetta reveals RNA structural rearrangements during human telomerase catalysis. *Rna* **23**, 175–188 (2017).
23. Vidal, A. E., Hickson, I. D., Boiteux, S. & Radicella, J. P. Mechanism of stimulation of the DNA glycosylase activity of hOGG1 by the major human AP endonuclease: Bypass of the AP lyase activity step. *Nucleic Acids Res.* **29**, 1285–1292 (2001).
24. Lundblad, J. R., Laurance, M. & Goodman, R. H. Fluorescence of Protein-DNA Interactions Polarization Analysis and Protein-Protein. 607–612
25. Yousefzadeh, M. *et al.* Dna damage—how and why we age? *Elife* **10**, 1–17 (2021).
26. Stratigopoulou, M., van Dam, T. P. & Guikema, J. E. J. Base Excision Repair in the Immune System: Small DNA Lesions With Big Consequences. *Front. Immunol.* **11**, (2020).
27. Jacobs, A. L. & Schär, P. DNA glycosylases: In DNA repair and beyond. *Chromosoma* **121**, 1–20 (2012).
28. Martin, L. J. DNA damage and repair: Relevance to mechanisms of neurodegeneration. *J. Neuropathol. Exp. Neurol.* **67**, 377–387 (2008).
29. Eyler, D. E., Burnham, K. A., Wilson, T. E. & O'Brien, P. J. Mechanisms of glycosylase induced genomic instability. *PLoS One* **12**, 1–21 (2017).
30. Krokan, H. E. & Bjoras, M. Chapter 06: Base Excision Repair. *Cold Spring Harb Perspect Biol.* **5**, a012583 (2013).
31. Robertson, A. B., Klungland, A., Rognes, T. & Leiros, I. Base excision repair: The

- long and short of it. *Cell. Mol. Life Sci.* **66**, 981–993 (2009).
32. Dalhus, B., Laerdahl, J. K., Backe, P. H. & Bjørås, M. DNA base repair - Recognition and initiation of catalysis. *FEMS Microbiol. Rev.* **33**, 1044–1078 (2009).
 33. Hegde, M. L., Hazra, T. K. & Mitra, S. Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells. *Cell Res.* **18**, 27–47 (2008).
 34. Lee, A. J., Warshaw, D. M. & Wallace, S. S. Insights into the glycosylase search for damage from single-molecule fluorescence microscopy. *DNA Repair (Amst.)* **20**, 23–31 (2014).
 35. Jose, D., Datta, K., Johnson, N. P. & Von Hippel, P. H. Spectroscopic studies of position-specific DNA ‘breathing’ fluctuations at replication forks and primer-template junctions. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 4231–4236 (2009).
 36. Friedman, J. I. & Stivers, J. T. Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry* **49**, 4957–4967 (2010).
 37. Moe, J. G. & Russu, I. M. Kinetics and Energetics of Base-Pair Opening in 5'-d(CGCGAATTCGCG)-3' and a Substituted Dodecamer Containing G·T Mismatches. *Biochemistry* **31**, 8421–8428 (1992).
 38. Yakovchuk, P., Protozanova, E. & Frank-kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. **34**, 564–574 (2006).
 39. Denver, D. R., Swenson, S. L. & Lynch, M. An evolutionary analysis of the helix-hairpin-helix superfamily of DNA repair glycosylases. *Mol. Biol. Evol.* **20**, 1603–1611 (2003).
 40. Parikh, S. S., Mol, C. D. & Tainer, J. A. Base excision repair enzyme family portrait: Integrating the structure and chemistry of an entire DNA repair pathway. *Structure* **5**, 1543–1550 (1997).
 41. Nelson, S. R., Dunn, A. R., Kathe, S. D., Warshaw, D. M. & Wallace, S. S. Two glycosylase families diffusively scan DNA using a wedge residue to probe for and identify oxidatively damaged bases. *Proc. Natl. Acad. Sci. U. S. A.* **111**, (2014).
 42. Dunn, A. R., Kad, N. M., Nelson, S. R., Warshaw, D. M. & Wallace, S. S. Single Qdot-labeled glycosylase molecules use a wedge amino acid to probe for lesions while scanning along DNA. *Nucleic Acids Res.* **39**, 7487–7498 (2011).
 43. Wong, I., Lundquist, A. J., Bernards, A. S. & Mosbaugh, D. W. Presteady-state analysis of a single catalytic turnover by *Escherichia coli* uracil-DNA glycosylase reveals a ‘pinch-pull-push’ mechanism. *J. Biol. Chem.* **277**, 19424–19432 (2002).
 44. Werner, R. M. *et al.* Stressing-out DNA? The contribution of serine-phosphodiester interactions in catalysis by uracil DNA glycosylase. *Biochemistry* **39**, 12585–12594 (2000).
 45. Fosado, Y. A. G. *et al.* A single nucleotide resolution model for large-scale simulations of double stranded DNA. *Soft Matter* **12**, 9458–9470 (2016).

46. Zahran, M., Daidone, I., Smith, J. C. & Imhof, P. Mechanism of DNA recognition by the restriction enzyme EcoRV. *J. Mol. Biol.* **401**, 415–432 (2010).
47. Young, M. A., Ravishanker, G., Beveridge, D. L. & Berman, H. M. Analysis of local helix bending in crystal structures of DNA oligonucleotides and DNA-protein complexes. *Biophys. J.* **68**, 2454–2468 (1995).
48. Narayana, N., Gineil, S. L., Russu, I. M. & Berman, H. M. Crystal and Molecular Structure of a DNA Fragment: D(CCGTGAATTCACG). *Biochemistry* **30**, 4449–4455 (1991).
49. Dienerowitz, M. *et al.* Single-molecule FRET dynamics of molecular motors in an ABEL Trap. (2021).
50. Podestà, A. *et al.* Positively charged surfaces increase the flexibility of DNA. *Biophys. J.* **89**, 2558–2563 (2005).
51. Maiti, A., Morgan, M. T., Pozharski, E. & Drohat, A. C. Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8890–8895 (2008).
52. Imhof, P. & Zahran, M. The Effect of a G:T Mismatch on the Dynamics of DNA. *PLoS One* **8**, 28–37 (2013).
53. Mcdowell, J. A. & Turner, D. H. Investigation of the Structural Basis for Thermodynamic Stabilities of Tandem GU Mismatches: Solution Structure of (rGAGGUCUC)₂ by Two-Dimensional NMR and Simulated Annealing †,‡. **2960**, 14077–14089 (1996).
54. Chen, L., Haushalter, K. A., Lieber, C. M. & Verdine, G. L. Direct visualization of a DNA glycosylase searching for damage. *Chem. Biol.* **9**, 345–350 (2002).
55. Bruner, S. D., Norman, D. P. G. & Verdine, G. L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**, 859–866 (2000).
56. Banerjee, A., Yang, W., Karplus, M. & Verdine, G. L. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* **434**, 612–618 (2005).
57. D’Augustin, O., Huet, S., Campalans, A. & Radicella, J. P. Lost in the crowd: How does human 8-oxoguanine DNA glycosylase 1 (OGG1) find 8-oxoguanine in the genome? *Int. J. Mol. Sci.* **21**, 1–18 (2020).
58. Lipscomb, L. A. *et al.* X-ray structure of a DNA decamer containing 7,8-dihydro-8-oxoguanine. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 719–723 (1995).
59. Ikehara, M., Kawase, Y. & Ohtsuka, E. Nmr studies of a dna containing 8-methoxydeoxyguanosine. *Nucleosides and Nucleotides* **11**, 261–272 (1992).
60. McCauley, M. J. *et al.* Quantifying the stability of oxidatively damaged DNA by single-molecule DNA stretching. *Nucleic Acids Res.* **46**, 4033–4043 (2018).
61. Singh, S. K. *et al.* Characterization of DNA with an 8-oxoguanine modification.

Nucleic Acids Res. **39**, 6789–6801 (2011).

62. Malins, D. C., Polissar, N. L., Ostrander, G. K. & Vinson, M. A. Single 8-oxoguanine and 8-oxo-adenine lesions induce marked changes in the backbone structure of a 25-base DNA strand. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 12442–12445 (2000).
63. Cheng, X. *et al.* Dynamic behavior of DNA base pairs containing 8-oxoguanine. *J. Am. Chem. Soc.* **127**, 13906–13918 (2005).
64. David, S. S., O’Shea, V. L. & Kundu, S. Base-excision repair of oxidative DNA damage. *Nature* **447**, 941–950 (2007).
65. Lukin, M. & de los Santos, C. NMR structures of damaged DNA. *Chem. Rev.* **106**, 607–686 (2006).
66. Fromme, J. C., Banerjee, A., Huang, S. J. & Verdine, G. L. Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nat. Mater.* **427**, 652–656 (2004).
67. Reynisson, J. & Steenken, S. The calculated base pairing energy of 8-oxoguanine in the syn-anti conformation with cytosine, thymine, adenine and guanine. *J. Mol. Struct. THEOCHEM* **723**, 29–36 (2005).
68. Allawi, H. T. & SantaLucia, J. Nearest neighbor thermodynamic parameters for internal G-A mismatches in DNA. *Biochemistry* **37**, 2170–2179 (1998).
69. Brown, T., Hunter, W. N., Kneale, G. & Kennard, O. Molecular structure of the G-A base pair in DNA and its implications for the mechanism of transversion mutations. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 2402–2406 (1986).
70. Brovarets’, O. O. & Hovorun, D. M. DPT tautomerisation of the G-Asyn and A*-G*syn DNA mismatches: A QM/QTAIM combined atomistic investigation. *Phys. Chem. Chem. Phys.* **16**, 9074–9085 (2014).
71. Rossetti, G. *et al.* The structural impact of DNA mismatches. *Nucleic Acids Res.* **43**, 4309–4321 (2015).
72. Lau, A. Y., Schäerer, O. D., Samson, L., Verdine, G. L. & Ellenberger, T. Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: Mechanisms for nucleotide flipping and base excision. *Cell* **95**, 249–258 (1998).
73. Setser, J. W., Lingaraju, G. M., Davis, C. A., Samson, L. D. & Drennan, C. L. Searching for DNA lesions: Structural evidence for lower- and higher-affinity DNA binding conformations of human alkyladenine DNA glycosylase. *Biochemistry* **51**, 382–390 (2012).
74. Vallur, A. C., Feller, J. A., Abner, C. W., Tran, R. K. & Bloom, L. B. Effects of hydrogen bonding within a damaged base pair on the activity of wild type and DNA-intercalating mutants of human alkyladenine DNA glycosylase. *J. Biol. Chem.* **277**, 31673–31678 (2002).
75. Wyatt, M. D. & Samson, L. D. Influence of DNA structure on hypoxanthine and 1,N6-ethenoadenine removal by murine 3-methyladenine DNA glycosylase.

Carcinogenesis **21**, 901–908 (2000).

76. Jones, P. A. Functions of DNA methylation : islands , start sites , gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492 (2012).
77. Bergman, Y. & Cedar, H. DNA methylation dynamics in health and disease. **20**, (2013).
78. Schübeler, D. Function and information content of DNA methylation. (2015). doi:10.1038/nature14192
79. Edwards, J. R., Yarychkivska, O., Boulard, M. & Bestor, T. H. DNA methylation and DNA methyltransferases. *Epigenetics and Chromatin* **10**, 1–10 (2017).
80. Moore, L. D., Le, T. & Fan, G. DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23–38 (2013).
81. Zhang, W. & Xu, J. DNA methyltransferases and their roles in tumorigenesis. *Biomark. Res.* 1–8 (2017). doi:10.1186/s40364-017-0081-z
82. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. **99**, 247–257 (1999).
83. Jeltsch, A. & Jurkowska, R. Z. Allosteric control of mammalian DNA methyltransferases – a new regulatory paradigm. **44**, 8556–8575 (2016).
84. Gowher, H., Jeltsch, A., Lafayette, W. & Biochemistry, T. HHS Public Access. **46**, 1191–1202 (2019).
85. Development, M. DNA Methylation Reprogramming during Mammalian Development. (2019). doi:10.3390/genes10040257
86. Chen, Z. & Zhang, Y. Role of Mammalian DNA Methyltransferases in Development. 1–24 (2019).
87. Xu, F. *et al.* Molecular and Enzymatic Profiles of Mammalian DNA Methyltransferases: Structures and Targets for Drugs. *Curr. Med. Chem.* **17**, 4052–4071 (2010).
88. Lin, C. C., Chen, Y. P., Yang, W. Z., Shen, J. C. K. & Yuan, H. S. Structural insights into CpG-specific DNA methylation by human DNA methyltransferase 3B. *Nucleic Acids Res.* **48**, 3949–3961 (2020).
89. Gao, L. *et al.* Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. *Nat. Commun.* 1–14 (2020). doi:10.1038/s41467-020-17109-4
90. Zhang, Z. *et al.* Structural basis for DNMT3A-mediated de novo DNA methylation. *Nat. Publ. Gr.* (2018). doi:10.1038/nature25477
91. Yang, L., Rau, R. & Goodell, M. A. HHS Public Access. **15**, 152–165 (2018).
92. Hamidi, T., Singh, A. K. & Chen, T. Genetic alterations of DNA methylation machinery in human diseases. **7**, 247–265 (2015).

93. Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H. & Tajima, S. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J. Biol. Chem.* **279**, 27816–27823 (2004).
94. Hata, K., Okano, M., Lei, H. & Li, E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. **1993**, 1983–1993 (2002).
95. Kaneda, M., Okano, M., Hata, K. & Sado, T. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. **429**, 2–5 (2004).
96. Xu, G., Lin, C., Bollman, B. & Bestor, T. H. Dnmt3L and the Establishment of Maternal Genomic Imprints. 2–6
97. Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A. & Cheng, X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* **449**, 248–251 (2007).
98. Gerasimait, R., Merkiene, E. & Klimašauskas, S. Direct observation of cytosine flipping and covalent catalysis in a DNA methyltransferase. *Nucleic Acids Res.* **39**, 3771–3780 (2011).
99. Ren, W., Gao, L. & Song, J. Structural Basis of DNMT1 and DNMT3A-Mediated DNA Methylation. (2018). doi:10.3390/genes9120620
100. Zhou, L. *et al.* Zebularine: A novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. *J. Mol. Biol.* **321**, 591–599 (2002).
101. Jurkowska, R. Z. *et al.* Oligomerization and Binding of the Dnmt3a DNA Methyltransferase to Parallel DNA Molecules. **286**, 24200–24207 (2011).
102. Jurkowska, R. Z. *et al.* Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. **36**, 6656–6663 (2008).
103. Nguyen, T. V. *et al.* The R882H DNMT3A hot spot mutation stabilizes the formation of large DNMT3A oligomers with low DNA methyltransferase activity. *J. Biol. Chem.* **294**, 16966–16977 (2019).
104. Karetka, M. S., Botello, Z. M., Ennis, J. J. & Chou, C. Reconstitution and Mechanism of the Stimulation of de Novo Methylation by Human DNMT3L * □. **281**, 25893–25902 (2006).
105. Rajavelu, A., Jurkowska, R. Z. & Jeltsch, A. Function and disruption of DNA Methyltransferase 3a cooperative DNA binding and nucleoprotein filament formation. **40**, 569–580 (2012).
106. Gao, L., Anteneh, H. & Song, J. Dissect the DNMT3A- and DNMT3B-mediated DNA Co-methylation through a Covalent Complex Approach. *J. Mol. Biol.* **432**, 569–575 (2020).
107. Emperle, M. *et al.* The DNMT3A R882H mutation does not cause dominant negative effects in purified mixed DNMT3A / R882H complexes. 1–9 (2018). doi:10.1038/s41598-018-31635-8

108. Dukatz, M. *et al.* Complex DNA sequence readout mechanisms of the DNMT3B DNA methyltransferase. **48**, 11495–11509 (2020).
109. Adam, S. *et al.* DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation. *Nat. Commun.* 1–15 (2020). doi:10.1038/s41467-020-17531-8
110. Johnson, S., Chen, Y. J. & Phillips, R. Poly(dA:dT)-Rich DNAs Are Highly Flexible in the Context of DNA Looping. *PLoS One* **8**, (2013).
111. Dlakic, M. & Harrington, R. E. Bending and torsional flexibility of G/C-rich sequences as determined by cyclization assays. *J. Biol. Chem.* **270**, 29945–29952 (1995).
112. Geggier, S. & Vologodskii, A. Sequence dependence of DNA bending rigidity. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 15421–15426 (2010).
113. Buechner, C. N. & Tessmer, I. DNA substrate preparation for atomic force microscopy studies of protein-DNA interactions. *J. Mol. Recognit.* **26**, 605–617 (2013).
114. Emperle, M. *et al.* Mutations of R882 change flanking sequence preferences of the DNA methyltransferase DNMT3A and cellular methylation patterns. **47**, 11355–11367 (2019).
115. Gowher, H., Liebert, K., Hermann, A., Xu, G. & Jeltsch, A. Mechanism of Stimulation of Catalytic Activity of Dnmt3A and. *J. Biol. Chem.* **280**, 13341–13348 (2005).
116. Handa, V. & Jeltsch, A. Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome. *J. Mol. Biol.* **348**, 1103–1112 (2005).
117. Mallona, I. *et al.* Flanking sequence preference modulates de novo DNA methylation in the mouse genome. *Nucleic Acids Res.* **49**, 145–157 (2021).
118. Allemannl, R. K. & Egli, M. DNA recognition and bending.
119. Veccham, S. P., Lee, J., Mao, Y. & Horn, P. R. A Non-Perturbative Pairwise-Additive Analysis of Charge Transfer Contributions to Intermolecular Interaction Energies.
120. Vlijm, R. & Dekker, C. Counterintuitive DNA Sequence Dependence in Supercoiling-Induced DNA Melting. 1–14 (2015). doi:10.1371/journal.pone.0141576
121. Milton, D. L., Casper, M. L., Wills, N. M. & Gesteland, R. F. Guanine tracts enhance sequence directed DNA bends. **18**, 4–7 (1990).
122. Gao, L., Anteneh, H. & Song, J. HHS Public Access. **432**, 569–575 (2021).
123. Anteneh, H., Fang, J. & Song, J. Structural basis for impairment of DNA. *Nat. Commun.* 1–12 doi:10.1038/s41467-020-16213-9
124. Halford, S. E. & Marko, J. F. How do site-speci c DNA-binding proteins nd

- their targets ? **32**, 3040–3052 (2004).
125. Halford, S. E. & Szczelkun, Æ. M. D. How to get from A to B : strategies for analysing protein motion on DNA. 257–267 (2002). doi:10.1007/s00249-002-0224-4
 126. Schwarz, F. W. *et al.* The dynamics of the monomeric restriction endonuclease BcnI during its interaction with DNA. **45**, 5968–5979 (2017).
 127. Garvie, C. W. & Wolberger, C. Recognition of Specific DNA Sequences Review. **8**, 937–946 (2001).
 128. Manuscript, A. Origins of specificity in protein-DNA recognition. 233–269 (2012). doi:10.1146/annurev-biochem-060408-091030.Origins
 129. Yang, W. Structure and mechanism for DNA lesion recognition. *Cell Res.* **18**, 184–197 (2008).
 130. Leblanc, S. J. *et al.* Coordinated protein and DNA conformational changes govern mismatch repair initiation by MutS. **46**, 10782–10795 (2018).
 131. Pan, H. *et al.* CpG and methylation-dependent DNA binding and dynamics of the methylcytosine binding domain 2 protein at the single-molecule level. **45**, 9164–9177 (2017).
 132. Lundin, M., Nehlin, O. & Ronne, H. Importance of a Flanking AT-Rich Region in Target Site Recognition by the GC Box-Binding Zinc Finger Protein MIG1. **14**, 1979–1985 (1994).
 133. Cadet, J., Douki, T. & Ravanat, J. Free Radical Biology & Medicine Oxidatively generated base damage to cellular DNA. *Free Radic. Biol. Med.* **49**, 9–21 (2010).
 134. Words, K. LONG -DISTANCE ELECTRON TRANSFER THROUGH DNA. 51–70 (2002). doi:10.1146/annurev.biochem.71.083101.134037
 135. Margolin, Y., Cloutier, J., Shafirovich, V., Geacintov, N. E. & Dedon, P. C. Paradoxical hotspots for guanine oxidation by a chemical mediator of inflammation. **2**, 365–366 (2007).
 136. Huffman, J. L., Sundheim, O. & Tainer, J. A. DNA base damage recognition and removal : New twists and grooves. **577**, 55–76 (2005).
 137. Stivers, J. T. & Jiang, Y. L. A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem. Rev.* **103**, 2729–2759 (2003).
 138. Stivers, J. T. Site-Specific DNA Damage Recognition by Enzyme-Induced Base Flipping. **77**, 37–65 (2004).
 139. Ba, X. *et al.* 8-Oxoguanine DNA Glycosylase-1 Augments Proinflammatory Gene Expression by Facilitating the Recruitment of Site-Specific Transcription Factors. *J. Immunol.* **192**, 2384–2394 (2014).
 140. Radák, Z. *et al.* Author ' s Accepted Manuscript. *Free Radic. Biol. Med.* (2015). doi:10.1016/j.freeradbiomed.2015.01.004

141. Radak, Z. & Boldogh, I. 8-oxo-7,8-dihydroguanine Link to gene expression, aging and.PDF. *Free Radic Biol Med.* **49**, 587–596 (2010).
142. Pan, L. *et al.* Oxidized guanine base lesions function in 8-oxoguanine DNA glycosylase-1-mediated epigenetic regulation of nuclear factor κ B-driven gene expression. *J. Biol. Chem.* **291**, 25553–25566 (2016).
143. Fleming, A. M., Zhu, J., Jara-Espejo, M. & Burrows, C. J. Cruciform DNA Sequences in Gene Promoters Can Impact Transcription upon Oxidative Modification of 2-Deoxyguanosine. *Biochemistry* **59**, 2616–2626 (2020).
144. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. (2006).
145. Genereux, J. C. & Barton, J. K. Mechanisms for DNA Charge Transport. 1642–1662 (2010).
146. Fleming, A. M. & Burrows, C. J. Interplay of Guanine Oxidation and G-Quadruplex Folding in Gene Promoters. *J. Am. Chem. Soc.* **142**, 1115–1136 (2020).
147. Pastukh, V. *et al.* An oxidative DNA “damage” and repair mechanism localized in the VEGF promoter is important for hypoxia-induced VEGF mRNA expression. *Am. J. Physiol. - Lung Cell. Mol. Physiol.* **309**, L1367–L1375 (2015).
148. Yoshihara, M. I., Jiang, L. I., Akatsuka, S. H., Suyama, M. I. & Toyokuni, S. H. Genome-wide Profiling of 8-Oxoguanine Reveals Its Association with Spatial Positioning in Nucleus. 603–612 (2014).
149. Ding, Y., Fleming, A. M. & Burrows, C. J. Sequencing the Mouse Genome for the Oxidatively Modified Base 8 - Oxo-7,8-dihydroguanine by OG-Seq. 8–11 (2017). doi:10.1021/jacs.6b12604
150. Amente, S. *et al.* LSD1-mediated demethylation of histone H3 lysine 4 triggers Myc-induced transcription. *Oncogene* **29**, 3691–3702 (2010).
151. Giorgio, M., Dellino, I. G., Gambino, V., Roda, N. & Pelicci, P. G. On the epigenetic role of guanosine oxidation. *Redox Biol.* **29**, 101398 (2020).
152. Pan, L. *et al.* OGG1-DNA interactions facilitate NF- κ B binding to DNA targets. *Sci. Rep.* **7**, 1–10 (2017).
153. Mabley, J. G. *et al.* Regulating Inflammation. *FASEB J.* **18**, 1–18 (2004).
154. Hailer-morrison, M. K., Kotler, J. M., Martin, B. D. & Sugden, K. D. Oxidized Guanine Lesions as Modulators of Gene Transcription . Altered p50 Binding Affinity and Repair Shielding by 7 , 8-Dihydro-8-oxo-2 ' -deoxyguanosine Lesions in the NF- κ B Promoter Element †. 9761–9770 (2003).
155. Ramon, O. *et al.* the binding of the transcription factor Sp1 to its cognate target DNA sequence (GC box) Effects of 8-oxo-7 , 8-dihydro-2 -deoxyguanosine on the Binding of the Transcription Factor Spl to its Cognate Target DNA Sequence (GC box). **5762**, (2016).

156. Moore, S. P. G., Toomire, K. J. & Strauss, P. R. DNA modifications repaired by base excision repair are epigenetic. *DNA Repair (Amst)*. **12**, 1152–1158 (2013).
157. Dang, C. V. MYC, Metabolism, Cell Growth, and Tumorigenesis. 1–15 (2013).
158. Conacci-Sorrell, M., McFerrin, L. & Eisenman, R. N. An overview of MYC and its interactome. *Cold Spring Harb. Perspect. Med.* **4**, 1–24 (2014).
159. Chen, H., Liu, H. & Qing, G. Targeting oncogenic Myc as a strategy for cancer treatment. *Signal Transduct. Target. Ther.* **3**, 1–7 (2018).
160. Guo, J. *et al.* Sequence specificity incompletely defines the genome-wide occupancy of Myc. *Genome Biol.* **15**, 482 (2014).
161. Jonathan Posner and Bradley S. Peterson, J. A. R. 基因的改变 NIH Public Access. *Bone* **23**, 1–7 (2008).
162. Walhout, A. J. M., Gubbels, J. M., Bernardts, R., Van Der Vliet, P. C. & Timmers, H. T. M. C-Myc/Max heterodimers bind cooperatively to the E-box sequences located in the first intron of the rat ornithine decarboxylase (ODC) gene. *Nucleic Acids Res.* **25**, 1493–1501 (1997).
163. Wang, R., Hao, W., Pan, L., Boldogh, I. & Ba, X. The roles of base excision repair enzyme OGG1 in gene expression. *Cell. Mol. Life Sci.* **75**, 3741–3750 (2018).
164. Moore, S. P. G., Kruchten, J., Toomire, K. J. & Strauss, P. R. Transcription factors and DNA repair enzymes compete for damaged promoter sites. *J. Biol. Chem.* **291**, 5452–5460 (2016).
165. Ba, X. & Boldogh, Istvan. 8-Oxoguanine DNA glycosylase 1: Beyond repair of the oxidatively modified base lesions. *Redox Biol.* **14**, 669–678 (2018).
166. Roychoudhury, S. *et al.* Endogenous oxidized DNA bases and APE1 regulate the formation of G-quadruplex structures in the genome. *Proc. Natl. Acad. Sci. U. S. A.* **117**, (2020).
167. Li, M., Dai, N., Wang, D. & Zhong, Z. Distinct APE1 Activities Affect the Regulation of VEGF Transcription Under Hypoxic Conditions. *Comput. Struct. Biotechnol. J.* **17**, 324–332 (2019).
168. Fleming, A. M., Ding, Y. & Burrows, C. J. Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 2604–2609 (2017).
169. Broxson, C., Hayner, J. N., Beckett, J., Bloom, L. B. & Tornaletti, S. Human AP endonuclease inefficiently removes abasic sites within G4 structures compared to duplex DNA. **42**, 7708–7719 (2014).
170. Antoniali, G., Lirussi, L., Poletto, M. & Tell, G. Emerging roles of the nucleolus in regulating the DNA damage response: The noncanonical DNA repair enzyme APE1/Ref-1 as a paradigmatical example. *Antioxidants Redox Signal.* **20**, 621–639 (2014).
171. Cogoi, S., Ferino, A., Miglietta, G., Pedersen, E. B. & Xodo, L. E. The regulatory

- G4 motif of the Kirsten ras (KRAS) gene is sensitive to guanine oxidation: Implications on transcription. *Nucleic Acids Res.* **46**, 661–676 (2018).
172. Fleming, A. M., Ding, Y. & Burrows, C. J. Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair. **1**, 1–18 (2016).
 173. Plavec, J. & Podbev, P. Oxidative lesions modulate G-quadruplex stability and structure in the human BCL2 promoter. *Bielskut e.* **49**, 2346–2356 (2021).
 174. Perillo, B., Tramontano, A., Pezone, A. & Migliaccio, A. LSD1: more than demethylation of histone lysine residues. *Exp. Mol. Med.* **52**, 1936–1947 (2020).
 175. Bravard, A. *et al.* Redox Regulation of Human OGG1 Activity in Response to Cellular Oxidative Stress. *Mol. Cell. Biol.* **26**, 7430–7436 (2006).
 176. Bafico, A., Grumolato, L. & Aaronson, S. A. Oncogenes and Signal Transduction. 17–30
 177. Poole, C. J. & van Riggelen, J. MYC—master regulator of the cancer epigenome and transcriptome. *Genes (Basel)*. **8**, (2017).
 178. Amente, S., Lania, L. & Majello, B. Epigenetic reprogramming of Myc target genes. *Am. J. Cancer Res.* **1**, 413–418 (2011).
 179. Anand, R. & Marmorstein, R. Structure and mechanism of lysine-specific demethylase enzymes. *J. Biol. Chem.* **282**, 35425–35429 (2007).
 180. Lan, F., Nottke, A. C. & Shi, Y. Mechanisms involved in the regulation of histone lysine demethylases. **1**, 316–325 (2008).
 181. Forneris, F. *et al.* Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process. **579**, 2203–2207 (2005).
 182. Yada, M. *et al.* Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7. **23**, 2116–2125 (2004).
 183. Lavigne, P. & Grandbois, M. Direct Visualization of the Binding of c-Myc / Max Heterodimeric b-HLH-LZ to E-Box Sequences on the hTERT Promoter † Re. 10279–10286 (2007).
 184. Larsen, E. *et al.* Repair and mutagenesis at oxidized DNA lesions in the developing brain of wild-type and Ogg1^{-/-} mice. *Oncogene* **25**, 2425–2432 (2006).
 185. Ba, X. *et al.* The role of 8-oxoguanine DNA glycosylase-1 in inflammation. *Int. J. Mol. Sci.* **15**, 16975–16997 (2014).
 186. Ames, B. N., Shigenaga, M. K. & Hagen, T. M. Oxidants, antioxidants, and the degenerative diseases of aging. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 7915–7922 (1993).
 187. Poetsch, A. R. The genomics of oxidative DNA damage, repair, and resulting mutagenesis. *Comput. Struct. Biotechnol. J.* **18**, 207–219 (2020).
 188. Hao, W. *et al.* Effects of the stimuli-dependent enrichment of 8-oxoguanine DNA glycosylase1 on chromatinized DNA. *Redox Biol.* **18**, 43–53 (2018).

189. Sampath, H. *et al.* 8-Oxoguanine DNA Glycosylase (OGG1) Deficiency Increases Susceptibility to Obesity and Metabolic Dysfunction. *PLoS One* **7**, 1–10 (2012).
190. Yuzefovych, L. V *et al.* Alteration of Mitochondrial Function and Insulin Sensitivity in Primary Mouse Skeletal Muscle Cells Isolated From Transgenic and Knockout Mice : Role of. **154**, 2640–2649 (2013).
191. Youn, C. *et al.* Human 8-Oxoguanine DNA Glycosylase Suppresses the Oxidative Stress – Induced Apoptosis through a p53-Mediated Signaling Pathway in Human Fibroblasts. **5**, 1083–1099 (2007).
192. Schooten, V. & Hodges, N. J. of. *Toxicol. Lett.* (2020). doi:10.1016/j.toxlet.2020.06.019
193. Hao, W. *et al.* Enzymatically inactive OGG1 binds to DNA and steers base excision repair toward gene transcription. *FASEB J.* **34**, 7427–7441 (2020).
194. Kaur, M. P. *et al.* Cellular accumulation of Cys326-OGG1 protein complexes under conditions of oxidative stress. *Biochem. Biophys. Res. Commun.* **447**, 12–18 (2014).
195. Hill, J. W. & Evans, M. K. Dimerization and opposite base-dependent catalytic impairment of polymorphic S326C OGG1 glycosylase. *Nucleic Acids Res.* **34**, 1620–1632 (2006).

List of Figures

Figure 2.1: IMAC and SEC purification of hOGG1 _{wt}	36
Figure 2.2: IMAC and SEC purification of hOGG1 _{K249Q}	37
Figure 2.3: IMAC and SEC purification of LSD1.....	38
Figure 2.4: SDS gel profile showing fractions from IMAC purification of Max.....	38
Figure 2.5: Size exclusion chromatography of MyC ₁₋₁₆₃	39
Figure 2.6: Schematic representation of the DNA substrate preparation for AFM studies.....	46
Figure 2.7: Characterization of the prepared DNA substrates for AFM imaging.....	46
Figure 2.8: Cloning strategy for E-box motif in pUC19N plasmid.....	47
Figure 2.9: Schematic representation of oxoG-Ebox DNA substrate preparation.....	48
Figure 2.10: Schematic of an AFM set up.....	51
Figure 2.11: Schematic of a position distribution of a protein binding to a specific target site.....	58
Figure 3.1: Overview of common base lesions and abasic sites.....	65
Figure 3.2: Schematic of G:C to T:A transversion mutation.....	66
Figure 3.3: Overview of base excision repair (BER) pathway.....	67
Figure 3.4: Lesion recognition by different glycosylase-DNA complexes.....	69
Figure 3.5: Phosphate pinch by UDG.....	70
Figure 3.6: Schematic of individual steps of automated DNA bend angle measurement at target lesions.....	73
Figure 3.7: Comparison of manual (a) and automated (b, c) measurements of DNA bending at G:U mismatch sites shown exemplarily.....	74
Figure 3.8: Schematic of individual steps of automated DNA bend angle analyses at protein bound sites on undamaged DNA.....	75
Figure 3.9: Ensemble FRET measurements of average DNA bend angles for glycosylase-DNA complexes.....	77
Figure 3.10: Characterization of non-specifically bound DNA glycosylase complex conformations using MatLab based automated DNA bend angle analyses.....	81
Figure 3.11: Characterization of target lesions of BER glycosylases.....	82
Figure 3.12: Schematic of the initial lesion sensing strategy employed by DNA glycosylases.....	86
Figure 4.1: DNA methylation catalyzed by DNA methyltransferases.....	87
Figure 4.2: Structure of DNMT3A-3L heterotetramer with DNA containing two ZpG sites.....	89

Figure 4.3: Co-methylation of dual CpG sites spaced at different distances by DNMT3A and DNMT3A-3L complexes.....	92
Figure 4.4: Co-methylation studies with DNMT3A and DNMT3A-3L complexes at CpG sites placed at variable distances in three different modes of co-methylation.....	93
Figure 4.5: Sequence composition of the central region between dual CpG sites placed 12 bp apart influences MW co-methylation.....	95
Figure 4.6: DNMT3A-3L binding specificities at the dual CpG sites in the center separated by 6, 9, and 12 bp obtained from single-molecule AFM studies.....	97
Figure 4.7: AFM analyses to characterize the DNMT3A-3L complexes bound to dual CpG containing DNA substrates placed at different distances.....	98
Figure 4.8: AFM based characterization of double heterotetrameric DNMT3A-3L complexes with CpG containing DNA substrates separated by 6 bp and 9 bp.....	99
Figure 4.9: Models describing MW and WM mode of co-methylation at distances of 6, 9, and 12 bps.....	101
Figure 4.10: Modes of diffusion behavior of DNA binding proteins on the genome for target search.....	104
Figure 4.11: Two DNMT3A-3L heterotetramer complexes binding in a side-by-side fashion on the D12 substrate.....	106
Figure 5.1: Schematic models describing various possible scenarios of Myc-mediated transcription regulation by hOGG1 under oxidative conditions.....	110
Figure 5.2: Schematic model of the initiation of G4 quadruplex formation by an oxoG lesion in a target gene promoter containing a potential G4 quadruplex forming sequence (PQS).....	112
Figure 5.3: Role of DNA oxidation and hOGG1 in Myc-mediated transcription regulation.....	113
Figure 5.4: Analysis and characterization of oligomeric states of hOGG1 by SDS-PAGE and circular dichroism experiments.....	115
Figure 5.5: hOGG1 oligomerization studies using AFM volume analysis and electrophoretic mobility shift assays.....	117
Figure 5.6: hOGG1 shows enhanced specificity and affinity towards oxoG lesion under reducing conditions.....	118
Figure 5.7: SV-AUC and analytical SEC experiments reveal an equilibrium between the monomer and dimer forms of hOGG1.....	119

Figure 5.8: Western blot and Mass spectrometry analysis of hOGG1 to identify cysteine residues involved in dimerization.....	121
Figure 5.9: Microcrystals of oxidized hOGG1 obtained in the nucleix suite.....	122
Figure 5.10: Extreme redox conditions do not alter hOGG1 activity.....	123
Figure 5.11: In vitro pull-down Assays to demonstrate direct protein-protein interactions under reducing (a) and oxidizing (b) conditions.....	125
Figure 5.12: Native agarose gel electrophoresis (NAGE) experiments demonstrating direct protein-protein interactions by hOGG1 _{wt} with Myc ₁₋₁₆₃ and LSD1.....	126
Fig 5.13: Bar plot showing affinities for interactions between hOGG1 and Myc or LSD1 under reducing and oxidizing conditions.....	127
Figure 5.14: EMSA studies of hOGG1 interaction with Myc on DNA containing an oxoG lesion.....	128
Figure 5.15: AFM studies of hOGG1 and Myc/Max interaction on DNA.....	129
Figure 5.16: Quantum Dot- AFM experiments to investigate enhanced binding of Myc-Max to E-box in the presence of hOGG1.....	131
Figure 5.17: AFM bend angle analysis of hOGG1 and hOGG1-Myc complexes at oxoG lesion.....	133
Figure 5.18: Schematic model showing the role of DNA oxidation and hOGG1 involvement in Myc-driven transcription.....	138
Figure 5.19: Schematic of G4 quadruplex and E-box containing oligonucleotide.....	139
Figure x3.1: Step-By-Step guide for automated bend angle analysis at 50% lesion site.....	142
Figure x3.2: Step-By-Step guide for automated bend angle analysis for non-specifically bound protein-DNA complexes.....	147
Figure x3.3: Step-By-Step guide for automated bend angle analysis for protein-DNA complexes at 50% DNA length.....	153
Figure x3.4: Schematic showing the procedure to connect kinked filaments during MatLab analysis.....	160
Figure x3.5: AFM image quality criteria for performing automated MatLab analyses.....	161
Figure x3.6: Comparison of Tangent versus secant method for measuring DNA bend angles.....	162

Figure x3.7: Query point distances of the tangents influence the DNA bend angle measurements.....	163
Figure x3.8: Ensemble FRET measurements to obtain average DNA bend angles in hOGG1-DNA, MutY-DNA, and hAAG-DNA complexes.....	164
Figure x3.9: Individual Gaussian fits shown in the multimodal fits for DNA-glycosylase complexes.....	165
Figure x3.10: Individual Gaussian fits shown in the multimodal fits for lesion containing substrates in the absence of glycosylases.....	165
Figure x3.11: FRET simulations to validate average AFM bend angles for glycosylase-DNA complexes.....	166
Figure x3.12: DNA bending obtained by full length hAAG (right)(b) versus the N-terminally truncated variant used in these studies (left)(a).....	167
Figure x3.13: End-to-end distance estimation on undamaged 505 bp long DNA substrate.....	167
Figure x4.1: Schematic showing dual CpG site containing DNA substrates for AFM experiments.....	169
Figure x4.2: AFM analyses for the DNA substrates with CpG sites placed at different distances in the absence of protein.....	170
Figure x4.3: Histograms showing DNA length distributions for D6, D9 and D12 substrates obtained from MatLab routine.....	171
Figure x4.4: Heights and lengths of single heterotetrameric complexes of DNMT3A/3L (based on volumes) bound to CpG pairs with 6, 9, and 12 bp spacing.....	171
Figure x4.5: Detailed view of the side-by-side DNMT3A/3L model bound at CpG pairs spaced at 6 bp distance.....	172
Figure x5.1: Size exclusion profiles and SDS-PAGE analyses of hOGG1 _{C28A} , hOGG1 _{C241A} , hOGG1 _{C253A} and batches II and III of hOGG1 _{wt}	173
Figure x5.2: Sequencing results for hOGG1 _{C28A} , hOGG1 _{C241A} , hOGG1 _{C253A} , hOGG1 _{wt} , Myc ₁₋₁₆₃ , E-box in pUC19N, and hOGG1 _{K249Q}	175
Figure x5.3: Duplicate and triplicate non-reducing SDS-PAGE based hOGG1 dimerization experiments.....	176
Figure x5.4: Individual AFM experiments of hOGG1 with oxoG lesion containing substrate....	177

Figure x5.5: Exemplary individual fluorescence polarization curves from triplicate measurements of hOGG1 _{wt} and hOGG1 _{K249Q} with oxoG and undamaged substrates under oxidizing and reducing conditions.....	178
Figure x5.6: aSEC experiment with 3 μM hOGG1 from batch I.....	179
Figure x5.7: Non-reducing SDS-PAGE based dimerization experiments for hOGG1 cysteine mutants.....	180
Figure x5.8: Duplicate (a) and triplicate (b) experiments showing gel-based hOGG1 activity assays to demonstrate hOGG1 activity under oxidative and reductive environments.....	181
Figure x5.9: Duplicate (a) and triplicate (b) experiment showing in vitro pull-down studies to demonstrate interactions between hOGG1, Myc and LSD1 under oxidizing and reducing conditions.....	181
Figure x5.10: Duplicate experiment showing Native PAGE analysis to demonstrate interactions between hOGG1, Myc and LSD1 under oxidizing and reducing conditions.....	182
Figure x5.11: Exemplary individual fluorescence polarization curves from triplicate measurements of hOGG1 _{wt} with Myc and hOGG1 _{wt} with LSD1 under oxidizing and reducing conditions.....	183
Figure x5.12: Individual AFM experiments of Myc recruitment by hOGG1 binding to oxoG lesion.....	184
Figure x5.13: hOGG1 _{K249Q} and Myc/Max interaction on DNA.....	187
Figure x5.14: hOGG1 and LSD1 interaction on DNA. hOGG1 and LSD1 were incubated together and crosslinked on DNA using 0.1% glutaraldehyde.....	188
Figure x5.15: Triplicate AFM experiments of hOGG1 mediated enhancement of QD labelled Myc binding to E-box.....	189
Figure x5.16: Triplicate automated AFM MatLab analyses showing the DNA bend angles for hOGG1 and hOGG1-Myc complexes at oxoG lesion in the oxidative environment.....	191

List of Tables

Table 2.1: Instruments.....	17
Table 2.2: Special consumables.....	20
Table 2.3: Chemicals, reagents, and media.....	22
Table 2.4: Resins.....	23
Table 2.5: Cloning material, enzymes, and recombinant proteins.....	23
Table 2.6: Bacterial strains.....	25
Table 2.7: Plasmids for protein expression and AFM DNA substrate preparation.....	25
Table 2.8: List of primers used for generation of bacterial expression constructs.....	26
Table 2.9: List of oligonucleotides used in AFM, EMSA, ensemble FRET measurements, fluorescence polarization experiments, and hOGG1 activity assays.....	27
Table 2.10: Antibodies.....	28
Table 2.11: List of enzymes used for protein purification, recombinant proteins, and protein ladder.....	29
Table 2.12: List of software, servers and databases used for this study.....	29
Table 2.13: Deposited MatLab scripts on Open Science Framework (OSF).....	30
Table 2.14: PCR reaction set up for SLIC cloning and site directed mutagenesis.....	31
Table 2.15: PCR conditions used for gene amplification and gene modification.....	31
Table 2.16: Expression strains, antibiotics, and times after induction with IPTG for different proteins.....	33
Table 2.17: Buffers for cell lysis, affinity chromatography, ion exchange chromatography and size exclusion chromatography.....	34
Table 2.18: Extinction coefficients and molecular weights of the purified proteins.....	41
Table 2.19: List of crystallization screens used in this study.....	43
Table 2.20: Denaturing 15% Urea-PAGE gel composition.....	61
Table 2.21: 15% Native-PAGE gel composition.....	62
Table 3.1: DNA bend angle statistics for BER target lesions obtained from automated AFM analyses.....	74

Table 3.2: DNA bend angle statistics obtained from AFM, ensemble FRET measurements and FRET simulations for non-specific DNA with and without bound glycosylases.....	78
Table x3.1: Configuration settings of FIESTA and MatLab software programs used for the automated DNA bend angle measurements at lesions.....	159
Table x3.2: Configuration settings of FIESTA and MatLab software programs for automated bend angle measurements for protein-DNA complexes.....	159
Table x3.3: End-to-end distances R and persistence lengths L_P of undamaged and lesion containing 505 bp DNA substrates used in this study.....	168
Table x4.1: Sequences of the oligonucleotide substrates used in the co-methylation kinetics of DNMT3A DNA methyltransferase.....	172
Table x5.1: KD values (\pm SD) from triplicate measurements of hOGG1 _{wt} and hOGG1 _{K249Q} with oxoG and undamaged substrates under oxidizing and reducing conditions...	179
Table x5.2: KD values (\pm SD) from triplicate measurements of hOGG1 _{wt} with Myc and hOGG1 _{wt} with LSD1 under oxidizing and reducing conditions.....	183

List of Abbreviations

For amino acids, the one or three letter code was used, according to the International Union of Pure and Applied Chemistry (IUPAC) regulations.

Abbreviation	Name
1D	One dimensional
2D	Two dimensional
3D	Three dimensional
5mC	5-Methyl cytosine
A	Adenine
A	Alanine
A ₂₆₀	Absorbance at 260 nm
A ₂₈₀	Absorbance at 280 nm
AAG	Alkyl adenine glycosylase
ADD	ATRX-DNMT3-DNMT3L-type
AF488	Alexafluor 488
AF647	Alexafluor 647
AFM	Atomic Force Microscopy
Amp	Ampicillin
A _{nsp}	area of background
AP site	Apurinic/apyrimidic site/ abasic site
APE1	Apurinic/apyrimidic endonuclease 1
APS	Ammonium persulphate
ASA	Accessible Surface Area
aSEC	Analytical Size Exclusion Chromatography
A _{sp}	area under the Gaussian fit
A-tract	Adenine tract
AUC	Analytical Ultracentrifuge
BCIP	5-bromo-4-chloro-3'- indolyphosphate p-toluidine salt
BER	Base excision repair
bp	Base pairs
BSA	Bovine Serum Albumin
BSA	Buried Surface Area
C	Cysteine
C	Cytosine
C ₂ H ₃ NaO ₂	Sodium acetate
C ₄ H ₆ MgO ₄	Magnesium acetate
Cam	Chloramphenicol
CC	Coiled Coil
CD	Circular Dichroism
cDNA	Complementary Deoxyribonucleic acid
ChIP-Seq	Chromatin Immunoprecipitation coupled Sequencing
cm	Centimeter

c-Myc	Cellular Myc
Co-IP	Co-Immunoprecipitation
CREB	cAMP responsive element binding protein 1
CTD	C-terminal domain
CV	Column Volume
Cy	Cyanine
Cys	Cysteine
D12	Substrate with two CpG placed 12 bps apart
D6	Substrate with two CpG placed 6 bps apart
D9	Substrate with two CpG placed 9 bps apart
dATP	2'-Deoxyadenosine 5'-triphosphate, sodium salt
dCTP	2'- Deoxycytidine 5'- triphosphate
ddH ₂ O	Double distilled water
dGTP	2'-Deoxyguanosine 5'-triphosphate, sodium salt
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DNMT	DNA methyltransferase
DNMT3L	DNMT3-like protein
dRP	5'-deoxyribose phosphate
dsDNA	Double stranded DNA
DTT	Dithiothreitol
dTTP	2'-Deoxythymidine 5'-triphosphate
dU	Deoxy-Uridine
<i>E. coli</i>	Escherichia coli
E-box/ Ebox	Enhancer box
EC	Excision complex
EDTA	Ethylenediaminetetraacetic acid
E _{FRET}	FRET efficiency
EMBL	European Molecular Biology Laboratory
EMSA	Electrophoretic Mobility Shift Assay
ethenoA	Ethenoadenine
ε _{DNA}	Extinction co-efficient of DNA
ε _{prot}	Extinction co-efficient of the protein
FAD	Flavin Adenine Dinucleotide
FCS	Fluorescence correlation spectroscopy
FEN-1	5'-flap endonuclease 1
FeS	Iron-sulfur

FIESTA	Fluorescence Image Evaluation Software for Tracking and Analysis
FL	Full-length
FPLC	Fast protein liquid chromatography
FPS	Fluorescence Positioning System
FRET	Fluorescence Resonance Energy Transfer
Fw	Forward primer
FWHM	Full width half maximum
G	Guanine
GG-NER	Global genome NER
GIMP	GNU Image Manipulation Program
GOI	Gene of interest
GST	Glutathione S-Transferase
G-tract	Guanine tract
H ₂ O ₂	Hydrogen peroxide
H3K4	Histone 3 Lysine 4
hAAG	Human Alkyl adenine DNA glycosylase
HCl	Hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HF	High fidelity
HhH	Helix-Hairpin-Helix
His	Histidine
hOGG1	Human 8-oxoguanine DNA glycosylase
HR	Homologous recombination
HRP	Horseradish peroxidase
HS-AFM	High-speed AFM
hTDG	Human Thymine DNA glycosylase
IBTB	Institute for Biochemistry and Technical Biochemistry
IC	Interrogation complex
IgG	Immunoglobulin G
IMAC	Immobilized Metal Affinity Chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
ISB	Institute of Structural Biology
IUPAC	International Union of Pure and Applied Chemistry
K	Lysine
Kan	Kanamycin sulfate
kb	Kilo base
KCl	Potassium chloride
K _D	Binding affinity
kDa	Kilo dalton
KH ₂ PO ₄	Potassium dihydrogen phosphate
L	Liter
LB-medium	Luria Bertani medium

L _c	Contour length
Lig	DNA ligase
L-Myc	small cell lung carcinomas
L _p	Persistence length
LP-BER	Long Patch BER
LSD1	Lysine specific demethylase 1
LZW	Lempel-Ziv-Welch
M	Molar
MatLab	Matrix Laboratory
Max	Myc associated factor X
Maz	Myc-associated zinc-finger protein
MBP	Maltose binding protein
MFP	Molecular Force Probe
mg	Milligram
Mg ₂ SO ₄	Magnesium sulfate
MgCl ₂	Magnesium chloride hexahydrate
ml	Milliliter
MM	Co-methylation of site 1 and 2 on the same strand
mM	Millimolar
MR	Mismatch repair
mRNA	Messenger RNA
MS	Microsoft
MS	Mass spectrometry
MutY	<i>E. coli</i> adenine DNA glycosylase
MW	Co-methylation of site 1 in the upper and site 2 in the lower strand
MW	Molecular weight
MWCO	Molecular Weight Cut-off
N	Newton
N	total number of binding sites
NaCl	Sodium chloride
NAGE	Native agarose gel electrophoresis
NaOH	Sodium Hydroxide
NBT	nitro-blue tetrazolium chloride
NBT/BCIP	5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
NEIL	Endonuclease VIII-like protein
NEM	N-Ethylmaleimide
NER	Nucleotide excision repair
N _{exp}	Expected number of events
NFκB	Nuclear Factor kappa-B

ng	nano gram
NGS	Next Generation Sequencing
NHEJ	Non-homologous end joining
NL	Non-linear
NL	Non-linear
nm	nanometer
nM	Nanomolar
NMR	Nuclear Magnetic Resonance
N-Myc	Myc found in neuroblastomas
N_{obs}	Observed number of events
Nsp	Non-specific
nt	nucleotides
NTH1	Endonuclease III-like protein 1
N_{total}	Total number of events
°C	Degree Celsius
OD_{600nm}	Optical density at 600 nm
OSF	Open Science Framework
oxoG/ oxo-G/8-oxo-G	8-oxoguanine
P	Phosphate
PAGE	Polyacrylamide gel electrophoresis
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PDI	Protein-DNA Interactions
PEG	Polyethylene glycol
PEG4000	Polyethylene glycol 4000
PISA	Proteins, Interfaces, Structures and Assemblies
PMSF	Phenylmethylsulfonyl fluoride
PNK	Polynucleotide kinase
POI	Protein of Interest
Pol	Polymerase
PQS	Potential G4 quadruplex forming sequence
PTM	Post Translational Modification
PUA	3'Phosphor- α , β - unsaturated aldehyde
p-value	Probability value
PVDF	Polyvinylidene difluoride
PWWP	proline-tryptophan-tryptophan-proline
Q	Glutamine
Qdot/QD	Quantum dot
R	End-to-end distance
R	Arginine
RD	Arginine-Aspartate
RGB	Red Green Blue

RMSD	Root Mean Square Deviation
RNA	Ribonucleic acid
ROS	Reactive Oxygen Species
RS	Restriction Site
RT	Room temperature
RT-PCR	Real Time-Polymerase Chain Reaction
Rv	Reverse primer
RVZ	Rudolf Virchow Zentrum
S	Specificity
SAM	S-adenosyl methionine
SC	Search complex
SD	Superdex
SD	Standard deviation
SDM	Site directed mutagenesis
SDS	Sodium Dodecyl Sulfate
SEC	Size exclusion chromatography
SEM	Standard error of the mean
Ser	Serine
SFM	Scanning Force Microscopy
SLIC	Sequence and Ligation independent cloning
sm	Single molecule
SN-BER	Single-nucleotide BER
SNP	Single point mutation
Sp1	Specificity protein 1
SP-BER	Short-patch BER
SPM	Scanning probe microscopy
ssDNA	Single stranded DNA
STM	Scanning tunneling microscopy
Strep	Streptomycin sulfate
SV	Sedimentation velocity
T	Thymine
TAE	Tris-Acetate-EDTA
TB	Terrific Broth
TBE	Tris-Borate-EDTA
TBST	Tris Buffered Saline Tween
TCEP	Tris-(2-carboxyethyl)-phosphine
TCR	Transcription coupled repair
TDG	Thymine DNA glycosylase
TE buffer	Tris-EDTA
TEMED	Tetramethylethylenediamine
TFIIH	Transcription Factor IIH
Thr	Thrombin
TIRF microscopy	Total internal reflection fluorescence microscopy
T _m	Melting Temperature

TRD	Target Recognition Domain
Tris	Tris-(hydroxymethyl)-aminomethane
trx	Thioredoxin
TSS	Transcription start site
U	Uracil
UDG	Uracil DNA glycosylase
UI	Units
UV	Ultraviolet
UV-VIS	Ultraviolet-Visible
WB	Western Blot
WLC model	Worm Like Chain Model
WM	Co-methylation of site 1 in the lower and site 2 in the upper strand.
wt	Wild type
WT	Wild type
XRCC1	X-ray repair cross-complementing protein 1
Z	2'- deoxy-Zebularine
β ME	β -mercaptoethanol
μ g	Microgram
μ l	Microliter
μ m	micrometer
μ M	Micromolar

List of Publications

1. **Disha M Bangalore***, Hannah S Heil*, Christian F Mehringer, Lisa Hirsch, Katherina Hemmen, Katrin G Heinze, Ingrid Tessmer, Automated AFM analysis of DNA bending reveals initial lesion sensing strategies of DNA glycosylases, *Sci Rep*, 2020, doi: <https://doi.org/10.1038/s41598-020-72102-7>.
2. Max Emperle*, **Disha M Bangalore***, Sabrina Adam, Stefan Kunert, Hannah S Heil, Katrin G Heinze, Pavel Bashtrykov, Ingrid Tessmer, Albert Jeltsch, Structural and biochemical insight into the mechanism of dual CpG site binding and methylation by the DNMT3A DNA methyltransferase, *Nucleic Acids Research*, 2021, doi: <https://doi.org/10.1093/nar/gkab600>.
3. **Disha M Bangalore** and Ingrid Tessmer, Unique insight into protein-DNA interactions from single molecule atomic force microscopy, *AIMS Biophysics*, 2018, doi: [10.3934/biophy.2018.3.194](https://doi.org/10.3934/biophy.2018.3.194). (*Review article*)
4. **Disha M Bangalore** and Ingrid Tessmer, Single molecule studies of hOGG1-Myc interactions in transcription regulation. (*Manuscript in preparation*)

* These authors contributed equally to this work

List of Congress contributions

4th NanoScientific Forum Europe 2021. Poster presentation: Automated AFM analysis of DNA bending reveals initial lesion sensing strategies of DNA glycosylases. (*Upcoming*)

15th International GSLS Student Symposium. Eureka 2020. Poster presentation: Investigating the functions of hOGG1 in transcription regulation. October 2020.

14th International GSLS Student Symposium. Eureka 2019. Poster presentation: Investigating the functions of hOGG1 in transcription regulation. October 2019.

Curriculum Vitae

Acknowledgements

I would like to express my immense gratitude to many people who have encouraged and supported me during the memorable three years of my PhD life.

First, a special thanks to my primary supervisor, Ingrid for coming up with such great projects and giving me the opportunity to work on them. You have been my constant guiding light, always pushing me that one extra mile and trusting in me during the last three eventful years. Your constant advice, constructive criticisms, numerous stimulating discussions, and encouragement during the stressful times have definitely made me a better scientist and allowed me to analyze results beyond the superficial facts.

I am deeply grateful to my second supervisor, Katrin Heinze for providing new ideas and inputs during our collaboration for the project involving automated AFM analyses of the glycosylases. I would also like to thank her for interesting suggestions and discussions during committee meetings that have helped me in my projects.

A great thanks to my third supervisor, Elmar wolf who has been a wonderful mentor, always providing insights into the world of Myc. I would like to thank him for his advice and for providing confidence during our meetings. Thank you for providing Myc and Max constructs which helped me immensely to conduct my studies.

I take this opportunity to also thank our collaborators Albert Jeltsch (University of Stuttgart) for providing me an opportunity to work on the topic of epigenetic regulation which allowed me to explore the world of DNA methyltransferases.

A big thank you to Hannah for developing MatLab scripts (and modifying them as we progressed in the projects) and teaching me the fundamentals of TIRF microscopy. In addition, I would like to thank Christian for teaching me the MatLab workflow initially.

Special thanks to AG Geyer in Bonn for providing full length Myc protein for my studies.

To Max, for providing the DNMT enzymes, CpG containing DNA substrates and for many fruitful discussions.

To Caroline, Hermann, Sebastian, Florian, Jochen, and Dan, thank you for the helpful suggestions during weekly strubi meetings.

I would also like to thank David, Anja, and Mariam for their inputs and interesting discussions on different biology topics during their master's practical courses with me. Felicia, Szabolcs, Sarah, Paula, and Jennifer, thank you for making the long durations in the lab more cheerful!

I would like to thank Teresa Frank, Andrea Schott-Heinzmann, Kathrin Haaf, and Sabine Sattler for their caring support and assistance in all the administrative work. Many thanks to Bernhard Fröhlich, Roland Markert, and Niklas Terhoeven for their constant IT-support and the help with my laptop. Thanks to Monica Kuhn, Nicole Bader, Michelle Endres, and Lars Schönemann for ordering, organizing, sorting, and running the lab.

I want to thank GSLS for arranging great workshops and providing guidance to ensure successful completion of PhD.

To the Strubis, for the wonderful working atmosphere, especially Anabel, Ishu, Aparna, Ayshwarya, Nasir, Susobhan, Ashwin, Ngoc, Theresa, and Jeannette, who became good friends.

Special regards to my old friend Radhika K Nair, who has held me up during my bad situations and constantly provided encouraging words. Thank you also for providing LSD1 constructs for my experiments. I would also like to thank my close friends back home for their constant motivation.

Finally, I would like to thank my family, especially my mom and my sister, who embraced my decision of pursuing PhD abroad, despite knowing the sorrow of being apart. Thank you Amma and Nidhu, for always cheering me up during stressful times. You have been great during last three years!

Saving the best for the last, my husband AJ! A big THANK YOU for selflessly and unconditionally following me on this memorable journey, constantly being my pillar of strength. I can't thank you enough for always being there, through thick and thin. I am glad to have shared the failures and successes of my life in these last few years. I love you and can't wait to start the next chapter of my life with you!

Affidavit

I hereby confirm that my thesis entitled “**Mechanistic studies of protein-DNA interactions by single molecule atomic force microscopy**” is the result of my own work. I did not receive any help or support from commercial consultants. All sources and/or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Würzburg,

(Date)

.....

(Signature)

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass die Dissertation „**Mechanistische Untersuchungen von protein-DNA-Wechselwirkungen mittels Einzelmolekül-Rasterkraftmikroskopie**“ eigenständig, d.h. insbesondere selbstständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Würzburg,

(Datum)

.....

(Unterschrift)

