

**Aus dem Rudolf-Virchow-Zentrum für Experimentelle Biomedizin  
der Universität Würzburg**

**Vorstände: Professor Dr. Caroline Kisker &  
Professor Dr. Bernhard Nieswandt**

**Optimierung und Objektivierung der DNA-Biegewinkelmessung zur  
Untersuchung der initialen Schadenserkennung von Glykosylasen im Rahmen  
der Basen-Exzisions-Reparatur**



**Inauguraldissertation  
zur Erlangung der Doktorwürde der  
Medizinischen Fakultät  
der  
Julius-Maximilians-Universität Würzburg**

**vorgelegt von**

**Christian Felix Mehringer  
aus Schweinfurt  
Würzburg, Mai 2020**

**Referentin:** Priv.-Doz. Dr. Ingrid Tessmer

**Korreferentin:** Prof. Dr Antje Gohla

**Dekan:** Prof. Dr. Matthias Frosch

**Tag der mündlichen Prüfung: 02.03.2021**

**Der Promovend ist Zahnarzt**

## **Meinen Eltern**



# Kurzfassung

Im Rahmen dieser Doktorarbeit sollte anknüpfend an die Ergebnisse aus vorangegangenen Untersuchungen der AG Tessmer, das von Büchner *et al.* [1] vorgestellte Modell zur DNA-Schadenserkennung, welches im Speziellen auf Daten zu den Glykosylasen hTDG und hOGG1 basierte, auf seine Allgemeingültigkeit für DNA-Glykosylasen untersucht werden. Das Modell beschreibt den Prozess der Schadenserkennung als eine notwendige Übereinstimmung der passiven Biegung am Schadensort mit dem aktiven Biegungswinkel der schadensspezifischen Glykosylase. Ein wesentlicher Bestandteil dieser Arbeit war zudem die Etablierung einer automatisierten Messsoftware zur objektiven Biegungsmessung an DNA-Strängen in rasterkraftmikroskopischen Aufnahmen. Dies wurde mit verschiedenen Bildverarbeitungsprogrammen sowie einer in MATLAB implementierten Messsoftware erreicht und das Programm zudem auf die Biegungsmessung von proteininduzierten Biegungswinkeln erweitert. Zur Anwendung kam die Methode der automatisierten Biegungsmessung sowohl an rasterkraftmikroskopischen Aufnahmen der Glykosylase MutY gebunden an ungeschädigter DNA als auch an Aufnahmen von DNA mit und ohne Basenschaden. Neben oxoG:A und G:A, den spezifischen MutY-Zielschäden, wurden auch andere Basenschäden wie beispielsweise oxoG:C und ethenoA:T vermessen und zudem die von der Glykosylase MutY an ungeschädigter DNA induzierte Biegung mit den Biegungswinkeln der jeweiligen Zielschäden verglichen. Die Übereinstimmung in den Konformationen der Zielschäden und der Reparaturkomplexe auch für die Glykosylase MutY (wie bereits für hTDG und hOGG1 in oben genannter Arbeit gezeigt) erlauben ein verbessertes Verständnis der Schadenssuche und -erkennung durch DNA-Glykosylasen, indem sie die Allgemeingültigkeit einer Biegungsenergie-basierten initialen Schadenserkennung durch DNA-Glykosylasen unterstützen. Die etablierte Messsoftware kann zukünftig an weiteren DNA-Schäden und den entsprechenden Protein-DNA-Komplexen ihre Anwendung finden und kann somit durch die effektive Gewinnung objektiver Daten in großer Menge zur Stützung des Modells beitragen.

**Optimisation and standardisation of DNA bend angle  
measurements as application of automated DNA bend angle  
measurements to initial damage detection of base excision  
repair glycosylases**

The focus of this thesis was to test the general applicability of a model for initial lesion detection by base excision repair (BER) glycosylases. This thesis built on previous results from the Tessmer laboratory on the human base excision repair (BER) glycosylases hTDG and hOGG1 (Büchner *et al.* [1]). Based on this work, a model for initial lesion detection by glycosylases had been proposed that describes the process of damage recognition as a necessary match of the passive bending at the point of damage with the active bending by the damage-specific glycosylase. An essential component of this work was also the establishment of an automated measurement software for objective bend angle measurements on DNA strands in atomic force microscopy (AFM) images. This was achieved with various image processing programs and a custom written MATLAB software. In addition, the procedure was extended to the measurement of DNA bend angles in protein-DNA complexes. In particular, the automated bend angle analysis was applied to AFM images of the glycosylase MutY bound to non-specific DNA and MutY target lesions (oxoG:A and G:A), as well as other DNA damages (oxoG:C and ethenoA:T). In the analyses, DNA bending induced by MutY in undamaged DNA was measured and compared to bending at the respective target damage. Similarities in the conformations of target damage and repair complexes also for this additional glycosylase (as already shown for hTDG and hOGG1 in above mentioned work) allow an improved understanding of DNA glycosylase damage search and recognition by supporting the general validity of bending energy-based initial damage detection by DNA glycosylases. In addition, the established measurement software can also be used to measure DNA bending by other protein systems in an unbiased manner and on a high-throughput scale. The software thus contributes to the effective acquisition of objective data.

# Inhaltsverzeichnis

<b>Kurzfassung.....</b>	<b>IV</b>
<b>Abbildungsverzeichnis .....</b>	<b>XI</b>
<b>Tabellenverzeichnis .....</b>	<b>XVII</b>
<b>Abkürzungsverzeichnis .....</b>	<b>XVIII</b>
<b>1 Einleitung.....</b>	<b>1</b>
1.1 DNA .....	1
1.2 DNA-Schäden .....	2
1.3 Basen-Exzisionsreparatur .....	3
1.4 Glykosylasen.....	6
1.4.1 hOGG1 .....	7
1.4.2 MutY.....	11
1.4.3 hAAG.....	14
1.5 Vorausgegangene Arbeiten und das Modell zur initialen Schadenserkennung der BER.....	17
1.6 Zielsetzung und Fragestellung .....	22
<b>2 Material und Methoden .....</b>	<b>23</b>
2.1 Material .....	23
2.1.1 Geräte und Verbrauchsmaterialien.....	23
2.1.2 Software .....	24
2.1.3 Internetbasierte Anwendungsprogramme .....	25
2.1.4 Komplettsysteme (Kits) .....	25
2.1.5 Chemikalien.....	25
2.1.6 Proteine.....	26

2.1.7	Puffer.....	26
2.1.8	Oligonukleotide.....	27
2.2	Methoden.....	28
2.2.1	DNA-Präparation für die Rasterkraftmikroskopie.....	28
2.2.2	Rasterkraftmikroskopie.....	31
2.2.2.1	Funktionsweise der Rasterkraftmikroskopie .....	31
2.2.2.2	AFM-Probenherstellung und -Messung .....	33
2.2.2.3	Bildprozessierung .....	37
2.2.3	Manuelle DNA-Längencharakterisierung.....	37
2.2.4	DNA-Biegewinkelmessung an der Schadensposition.....	38
2.2.5	DNA-Höhe und Breite.....	40
2.2.6	Proteinvolumen .....	41
2.2.7	Standardisierte und objektive Biegewinkelmessungen an DNA-Schäden .....	43
2.2.8	Standardisierte und objektive Biegewinkelmessungen an Protein-DNA-Komplexen .....	45
<b>3</b>	<b>Ergebnisse.....</b>	<b>47</b>
3.1	Rasterkraftmikroskopische Aufnahmen .....	47
3.2	Standardisierte und objektive Biegewinkelmessungen an DNA-Schäden.....	52
3.2.1	DNA-Höhe und Breite.....	52
3.2.2	Bildprozessierung.....	54
3.2.3	Formatänderung.....	55
3.2.4	Anwendung eines Median-Filters .....	55
3.2.5	Threshold .....	57
3.2.6	Skelettierung der DNA.....	62



3.2.6.1	Optimierung des Skelettierungsprozesses .....	68
3.2.7	Automatisierte Biege winkelmessung bei 50 % DNA-Länge .....	70
3.2.7.1	Vergleich verschiedener Segmentlängen .....	73
3.3	Biege winkel von DNA-Strängen .....	76
3.3.1	DNA-Biege winkel von ungeschädigter DNA .....	78
3.3.2	DNA-Biege winkel von oxoG:A .....	80
3.3.3	DNA-Biege winkel von G:A .....	82
3.3.4	DNA-Biege winkel von oxoG:C .....	84
3.3.5	DNA-Biege winkel von εA:T .....	86
3.4	Standardisierte und objektive Biege winkelmessungen an Protein-DNA-Komplexen .....	87
3.4.1	Bildprozessierung .....	88
3.4.2	Particle Remover .....	90
3.4.3	Proteinentfernung mit einem Threshold-Intervall .....	93
3.4.4	Shape Filter .....	94
3.4.5	Invertieren und Filtern .....	96
3.4.6	Skelettierung .....	97
3.4.7	Proteinpositionsbestimmung .....	97
3.4.8	DNA-Biege winkelmessung an Protein-DNA-Komplexen .....	100
3.4.8.1	Verbindung unterbrochener Skelettierungen .....	101
3.4.8.2	Molekülfilter .....	101
3.4.8.3	Protein-Radius .....	102
3.4.8.4	Messung des Biege winkels an MutY-DNA-Komplexen .....	104
3.5	Biege winkel von Protein-DNA-Komplexen .....	111
<b>4</b>	<b>Diskussion .....</b>	<b>115</b>
4.1	Diskussion der Methode .....	115

4.1.1	Voraussetzungen für eine optimale Auswertbarkeit der AFM-Aufnahmen.....	115
4.1.1.1	Bildverbessernde Maßnahmen am Rasterkraftmikroskop .....	115
4.1.1.2	DNA-Höhen- und -Breitenveränderung in AFM-Aufnahmen.....	116
4.1.2	DNA-Längeneffekte.....	117
4.1.2.1	Einfluss auf die Schadenslokalisierung.....	117
4.1.2.2	Verkürzung der DNA durch die Probendeposition .....	119
4.1.2.3	Veränderung der DNA-Länge während des manuellen Messverfahrens.....	120
4.1.2.4	Veränderung der DNA-Länge während des automatisierten Messverfahrens.....	121
4.1.3	Möglichkeiten der Schadenslokalisations-Kontrolle.....	122
4.1.4	DNA-Biegewinkeleffekte.....	124
4.1.4.1	Mögliche Einflussfaktoren bei der Probenherstellung.....	124
4.1.4.2	Beeinflussung des Biegewinkels bei der manuellen Auswertung .....	126
4.1.4.3	Beeinflussung des Biegewinkels bei der automatisierten Auswertung .....	126
4.1.5	Proteinerfassung .....	135
4.2	Diskussion der DNA-Biegewinkel.....	137
4.2.1	Ungeschädigte DNA.....	139
4.2.2	MutY-Schadenssubstrat: oxoG:A .....	140
4.2.3	MutY-Schadenssubstrat: G:A.....	141
4.2.4	hOGG1 Schadenssubstrat: oxoG:C .....	143
4.2.5	hAAG Schadenssubstrat: εA:T .....	145
4.2.6	Diskussion der DNA-Biegewinkel an Proteinen.....	146
4.3	Vergleich mit anderen publizierten Programmen .....	148

4.4	Biologische Relevanz.....	152
<b>5</b>	<b>Ausblick.....</b>	<b>156</b>
<b>6</b>	<b>Zusammenfassung .....</b>	<b>158</b>
<b>7</b>	<b>Literaturverzeichnis.....</b>	<b>159</b>
<b>8</b>	<b>Appendix.....</b>	<b>172</b>
8.1	Bildvorbereitung.....	172
8.1.1	DNA.....	172
8.1.2	DNA & Protein.....	174
8.2	Workflow-MATLAB.....	179
8.2.1	DNA.....	179
8.2.2	DNA & Protein.....	180
8.3	MATLAB-Programme.....	181
8.4	DNA-Schadenssubstrate und Biegewinkel.....	218
8.5	Eigene Ergebnisse mit der manuellen Messmethode.....	220
8.5.1	ungeschädigte DNA .....	220
8.5.2	oxoG:A .....	220
8.5.3	G:A.....	221
8.5.4	oxoG:C.....	221
8.5.5	$\epsilon$ A:T.....	222
8.5.6	MutY-nspDNA .....	222
8.6	Programmevaluation.....	223

# Abbildungsverzeichnis

Abbildung 1:	Struktur der DNA.....	1
Abbildung 2:	Übersicht der Basisexzisionsreparatur (BER) .....	5
Abbildung 3:	Schadenssuche der Glykosylase .....	7
Abbildung 4:	Entstehung und Struktur des oxoG-Schadens .....	8
Abbildung 5:	DNA-Reparatur und -Replikation des oxoG-Basenschadens....	9
Abbildung 6:	Biegungsaktivität von hOGG1 .....	11
Abbildung 7:	Konformationen der oxoG:A-Fehlpaarung .....	12
Abbildung 8:	Konformationen der G:A-Fehlpaarung .....	12
Abbildung 9:	Strukturformeln von hAAG-spezifischen DNA-Läsionen .....	15
Abbildung 10:	Struktureller Vergleich der Watson-Crick-Basenpaarung A:T mit dem $\epsilon$ A:T-Basenpaar .....	16
Abbildung 11:	Modell zur initialen Schadenserkenkung durch Glykosylasen	17
Abbildung 12:	Watson-Crick-Basenpaar G:C und Wobble-Basenpaar G:U ...	18
Abbildung 13:	DNA-Biegewinkel an G:U mit und ohne hTDG.....	19
Abbildung 14:	DNA-Biegewinkel von ungeschädigter DNA mit und ohne hTDG .....	20
Abbildung 15:	Schematische Darstellung der DNA-Substratpräparation .....	30
Abbildung 16:	Schematischer Aufbau eines Rasterkraftmikroskops.....	32
Abbildung 17:	Arbeitsschritte der Probenherstellung .....	36
Abbildung 18:	Rasterkraftmikroskopische Aufnahme.....	36
Abbildung 19:	Manuelles Messverfahren zur Längenbestimmung.....	38
Abbildung 20:	Manuelles Messverfahren zur Biegewinkelbestimmung.....	38
Abbildung 21:	Messung der DNA-Breite mit dem Section-Tool.....	40
Abbildung 22:	Exemplarische Proteinvolumenverteilung .....	42

Abbildung 23: Schematische Darstellung der Tangenten-, Sekanten-Messmethode.....	44
Abbildung 24: Arbeitsablauf bei der automatisierten Ermittlung von Biegewinkeln an DNA .....	45
Abbildung 25 : Arbeitsablauf bei der automatisierten Ermittlung von Biegewinkeln an Protein-DNA-Komplexen.....	46
Abbildung 26: Beispiel einer AFM-Aufnahme von DNA .....	48
Abbildung 27: Beispiel einer AFM-Aufnahme von Proteinen und Proteinvolumen (MutY) .....	48
Abbildung 28: Proteinvolumen von MutY, gebunden an ungeschädigter DNA .....	49
Abbildung 29: Längen- und Biegungswinkelverteilung für oxoG:A-DNA aus Messungen verschiedener Experimentatoren .....	50
Abbildung 30: Längen- und Biegungswinkelverteilung für oxoG:C-DNA aus Messungen verschiedener Experimentatoren .....	51
Abbildung 31: Bestimmung der Parameter Höhe und Breite von DNA .....	52
Abbildung 32: Auswirkung der DNA-Höhe auf die Auswertbarkeit mittels Software.....	53
Abbildung 33: Bearbeitung mit GIMP.....	54
Abbildung 34: Umwandlung des RGB- in ein 8 bit-Format.....	55
Abbildung 35: Effekt des Median-Filters .....	56
Abbildung 36: Thresholdbestimmung für FIESTA.....	57
Abbildung 37: Auto-Threshold.....	58
Abbildung 38: Problem der Thresholdbestimmung .....	59
Abbildung 39: Vergleich unterschiedlich hoher Offset-Einstellungen.....	61
Abbildung 40: Schematische Darstellung der Skelettierung .....	62
Abbildung 41: Messung der DNA-Breite in ImageJ.....	63

Abbildung 42: Messung der DNA-Breite mit dem Section-Tool.....	63
Abbildung 43: Messung der DNA-Breite mit „Line Scan“ in FIESTA.....	64
Abbildung 44: Bedienungsoberfläche von FIESTA .....	66
Abbildung 45: DNA-Tracking mit FIESTA .....	67
Abbildung 46: Segmentierung bei kleiner Halbwertsbreite.....	68
Abbildung 47: Effekt der Halbwertsbreite auf die Länge .....	69
Abbildung 48: Problemstellung bei gebogener DNA .....	70
Abbildung 49: DNA-Längen-Interpolation mittels Polynomzug.....	71
Abbildung 50: Vergleich der Filamentlängen vor und nach der Interpolation .....	71
Abbildung 51: Vergleichende Darstellung der Tangenten- und Sekantenmethode .....	72
Abbildung 52: MATLAB-Ergebnis .....	73
Abbildung 53: Skelettierung der DNA mit daran gebundenem Protein .....	74
Abbildung 54: Vergleich der verschiedenen Segmentlängen.....	75
Abbildung 55: Längenverteilung von nsp DNA.....	78
Abbildung 56: Biegewinkelverteilung von nsp DNA (505 bp) mit der Tangentenmethode .....	79
Abbildung 57: Biegewinkelverteilung von nsp DNA (505 bp) mit der Sekantenmethode .....	79
Abbildung 58: Biegewinkelverteilung von nsp DNA (2224 bp) mit der Tangentenmethode .....	80
Abbildung 59: Längenverteilung von oxoG:A.....	81
Abbildung 60: Biegewinkelverteilung von oxoG:A mit der Tangentenmethode .....	81
Abbildung 61: Biegewinkelverteilung von oxoG:A mit der Sekantenmethode .....	82

Abbildung 62:	Längenverteilung von G:A.....	83
Abbildung 63:	Biegewinkelverteilung von G:A mit der Tangentenmethode .....	83
Abbildung 64:	Biegewinkelverteilung von G:A mit der Sekantenmethode .....	84
Abbildung 65:	Längenverteilung von oxoG:C.....	84
Abbildung 66:	Biegewinkelverteilung von oxoG:C mit der Tangentenmethode .....	85
Abbildung 67:	Biegewinkelverteilung von oxoG:C mit der Sekantenmethode .....	85
Abbildung 68:	Längenverteilung von $\epsilon$ A:T .....	86
Abbildung 69:	Biegewinkelverteilung von $\epsilon$ A:T mit der Tangentenmethode .....	86
Abbildung 70:	Biegewinkelverteilung von $\epsilon$ A:T mit der Sekantenmethode .....	87
Abbildung 71:	Auswertung ohne Protein-Filterung.....	88
Abbildung 72:	Anwendung des Yen-Thresholds .....	89
Abbildung 73:	Particle Remover.....	91
Abbildung 74:	FIESTA-Ergebnis für eine Probe von MutY-Komplexen mit oxoG:A-DNA.....	92
Abbildung 75:	Ergebnisse der Biegewinkel an der 50 %-Position.....	93
Abbildung 76:	Darstellung mit einem Threshold-Intervall.....	94
Abbildung 77:	Shape Filter.....	95
Abbildung 78:	Invertierung und Gauß-Filter .....	96
Abbildung 79:	Ergebnis des FIESTA-Trackings .....	97
Abbildung 80:	Intermodes-Threshold .....	98
Abbildung 81:	Schwierigkeit beim Intermodes-Threshold .....	98

Abbildung 82: Intermoden-Threshold und Ergebnis des „3D Object - counters“ .....	99
Abbildung 83: Test zur Proteinauswertung an oxoG:A-MutY-Komplexen ....	100
Abbildung 84: Tracking gebogener und proteinbeladener DNA .....	101
Abbildung 85: Notwendigkeit eines Molekülfilters .....	102
Abbildung 86: Definition der Proteinbindung an DNA .....	103
Abbildung 87: Darstellung der Filamente und Proteinpositionen.....	105
Abbildung 88: Überlagerung der Proteinposition mit skelettierten DNA-Strängen für die MATLAB-Winkelmessung.....	106
Abbildung 89: Auswertung des MATLAB-Ergebnisses .....	106
Abbildung 90: Limitation des Yen- Threshold.....	107
Abbildung 91: Limitation des Shape Filters durch Definition des Thresholdes .....	108
Abbildung 92: Verbesserung der Erfassung von Protein-DNA-Komplexen..	109
Abbildung 93: Proteinposition kleiner als 25 nm .....	109
Abbildung 94: Gemeinsame Darstellung von Proteindaten mit DNA-Filamenten .....	110
Abbildung 95: Anpassung der Proteinerkennung.....	110
Abbildung 96: Proteinvolumenverteilung von MutY.....	111
Abbildung 97: Automatische Biegewinkelmessung an der Proteinposition mit der Tangentenmethode .....	113
Abbildung 98: Automatische Biegewinkelmessung an der Proteinposition mit der Sekantenmethode .....	113
Abbildung 99: Biegewinkelverteilung der automatisierten Biegewinkelmessung ohne Berücksichtigung des Proteinvolumens .....	114
Abbildung 100: Manuelle Messung des Biegewinkels an MutY-DNA-Komplexen mit ungeschädigter DNA .....	114



Abbildung 101: Darstellung der auflösungsbedingten Längenverkürzung .....	120
Abbildung 102: Geometrische Darstellung der Beziehung Tangente – Sekante.....	129
Abbildung 103: Limitation manueller Biegewinkelmessung an einem DNA- Strang .....	131
Abbildung 104: Bestimmung des Höhenmaximums mittels einer Gauß- Anpassung .....	132
Abbildung 105: Beispiele für die automatische Biegewinkelmessung an oxoG:A .....	133
Abbildung 106: Dimensionsvergleich T-Antigen zu MutY .....	135
Abbildung 107: Gegenüberstellung der Biegewinkelergebnisse .....	138
Abbildung 108: Vergleich der Biegewinkel von MutY-nspDNA-Komplexen ...	147
Abbildung 109: Ergebnis der manuellen Biegewinkelmessung an ungeschädigter DNA .....	220
Abbildung 110: Ergebnis der manuellen Biegewinkelmessung an oxoG:A....	220
Abbildung 111: Ergebnis der manuellen Biegewinkelmessung an G:A.....	221
Abbildung 112: Ergebnis der manuellen Biegewinkelmessung an oxoG:C....	221
Abbildung 113: Ergebnis der manuellen Biegewinkelmessung an εA:T.....	222
Abbildung 114: Manuelle Messung des Biegewinkels an MutY-DNA- Komplexen mit ungeschädigter DNA .....	222
Abbildung 115: Möglicher Einfluss des Depositionsverfahrens.....	223

# Tabellenverzeichnis

Tabelle 1: DNA-Substrate.....	34
Tabelle 2: Proteine .....	34
Tabelle 3: Protein-DNA-Substrate .....	34
Tabelle 4: Übersicht der DNA-Biegewinkel.....	77
Tabelle 5: Schnittstellen der DNA-Stränge.....	118
Tabelle 6: Übersicht der Biegewinkel .....	137
Tabelle 7: Biologische Einordnung der Messergebnisse .....	155
Tabelle 8: DNA-Schadenssubstrate .....	218
Tabelle 9: Zusammenfassung der Ergebnisse einzelner DNA-Biegewinkel...	219

# Abkürzungsverzeichnis

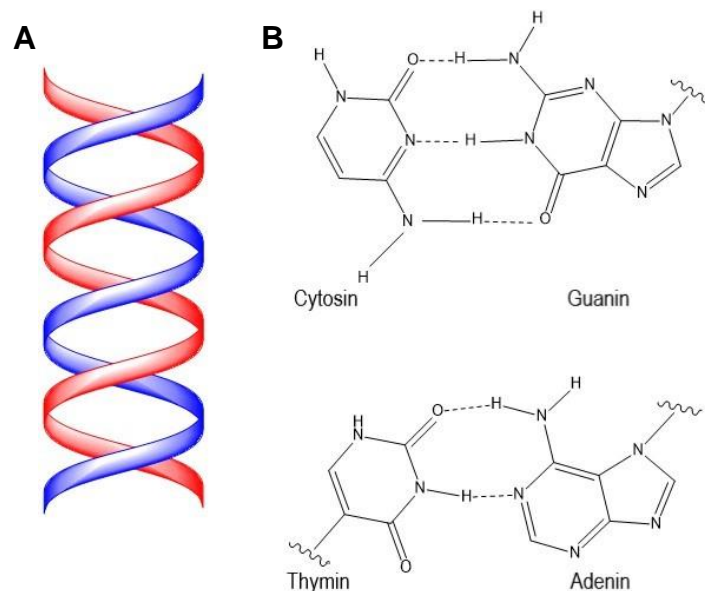
.ibw	Igor Pro Wave-Datei
.mat	Dateiformat von MATLAB (MathWorks)
.tiff	Tagged Image File Format
.txt	Text-Datei
~	ungefähr; circa
±	(sigma) entspricht der Standardabweichung
°C	Grad Celsius
A	Adenin
AAG	3-Alkyladenin-DNA-Glykosylase
AFM	Rasterkraftmikroskopie (engl. atomic force microscopy)
AP-Stelle	apurinische, bzw. apyrimidinische Stelle
ATP	Adenosintriphosphat
BER	Basenexzisionsreparatur
bp	Basenpaare
C	Cytosin
C	Konzentration
cm	Zentimeter
cm <sup>2</sup>	Quadratcentimeter
CTN	C-terminale Domäne
dest.	destilliert
dH <sub>2</sub> O	gefiltertes, deionisiertes Reinstwasser
DNA	Desoxyribonukleinsäure
DSB	Doppelstrangbruch (engl. double strand break)
dsDNA	Doppelstrang-Desoxyribonukleinsäure
EC	Exzisionskomplex (engl. excision complex)
engl.	englisch
<i>et al.</i>	lateinisch: <i>et alias</i> ; und andere
Etheno-A (εA)	1,N <sup>6</sup> -Etheno-Desoxyadenosin
ε <sub>λ</sub>	Extinktionskoeffizient
G	Guanin
h	Stunde(n)
HCl	Chlorwasserstoff
HEPES	2-(4-(2-Hydroxyethyl)-1-piperazinyl)-ethansulfonsäure
HhH	Helix-Haarnadel-Helix (engl.: helix-hairpin-helix)
hOGG1	Humane 8-Oxoguanin-DNA-Glykosylase
I	Intensität
IC	Abfragekomplex (engl. interrogation complex)
kb	Kilobasen
kcal	Kilokalorien

LP	lange Korrektur (engl.: long patch)
M	Molar (mol/l)
Mg	Milligramm
MgCl <sub>2</sub>	Magnesiumchlorid
min	Minute(n)
ml	Milliliter
mM	Millimolar
MPG	N-Methylpurin-DNA-Glykosylase
MutY	Adenin-DNA-Glykosylase
NaCl	Natriumchlorid
NEB	New England Biolabs
nm	Nanometer
nM	Nanomolar
NMR	Kernspinresonanzspektroskopie (engl. nuclear magnetic resonance)
nsp DNA	unspezifische DNA ~ ungeschädigte DNA
nt	Nukleotide
NTD	N-terminale Domäne
oxoG	7,8-Dihydro-8-oxoguanin
r	Abstand
RMS	Durchschnittliche quadratische Rauheit (root mean square)
ROS	reaktive Sauerstoffspezies (engl. reactive oxygen species)
SC	Suchkomplex (engl. search complex)
SP	kurze Korrektur (engl.: short patch)
SPM	Rastersondenmikroskopie (engl. scanning probe microscopy)
SSB	Einzelstrangbruch (engl. single strand break)
ssDNA	Einzelstrang-Desoxyribonukleinsäure (engl. single strand DNA)
STM	Rastertunnelmikroskopie (engl. scanning tunneling microscopy)
T	Thymin
Tyr	Tyrosin
U	Uracil
UDG	Uracil-DNA-Glykosylase
USA	United States of America
UV	Ultraviolettstrahlung
λ	Wellenlänge
μl	Mikroliter
μM	Mikromolar

# 1 Einleitung

## 1.1 DNA

Die Erbinformation eukaryotischer Lebewesen wird im Zellkern auf den Chromosomen gespeichert. Die Chromosomen bestehen allgemein aus vielen verschiedenen einzelnen Nukleotiden, die sich zu Polynukleotiden verbinden, und dann als Nukleinsäuren bezeichnet werden. Man unterscheidet hierbei die Desoxyribonukleinsäure (DNS; engl. DNA) und die Ribonukleinsäuren (RNS; engl. RNA). Während die DNA als der eigentliche Träger der Erbinformation fungiert, ist die RNA dagegen vor allem der Informationsträger für die Genexpression. Durch die Forschungsarbeit von Watson und Crick [2] ist heute bekannt, dass die DNA in Form einer Doppelhelix vorliegt, und entsprechend der Chargaff-Regel immer eine Purinbase mit einer Pyrimidinbase bindet. So bildet Adenin (A) mit Thymin (T) zwei Wasserstoffbrückenbindungen aus und Guanin (G) mit Cytosin (C) drei Wasserstoffbrückenbindungen (Abbildung 1). Durch die antiparallele Anordnung der Einzelstränge ist die komplementäre Basenpaar-Bildung erst möglich, was die Grundlage jeglicher Vorgänge der Genetik bildet [2, 3].



**Abbildung 1: Struktur der DNA**

- A** Modell einer B- förmigen DNA-Doppelhelix-Struktur  
**B** Watson-Crick-Basenpaarungen zwischen den Purinbasen Adenin (A) und Guanin (G) mit den Pyrimidinbasen Thymin (T) und Cytosin (C). Die Wasserstoffbrückenbindungen sind gestrichelt eingezeichnet.

### 1.2 DNA-Schäden

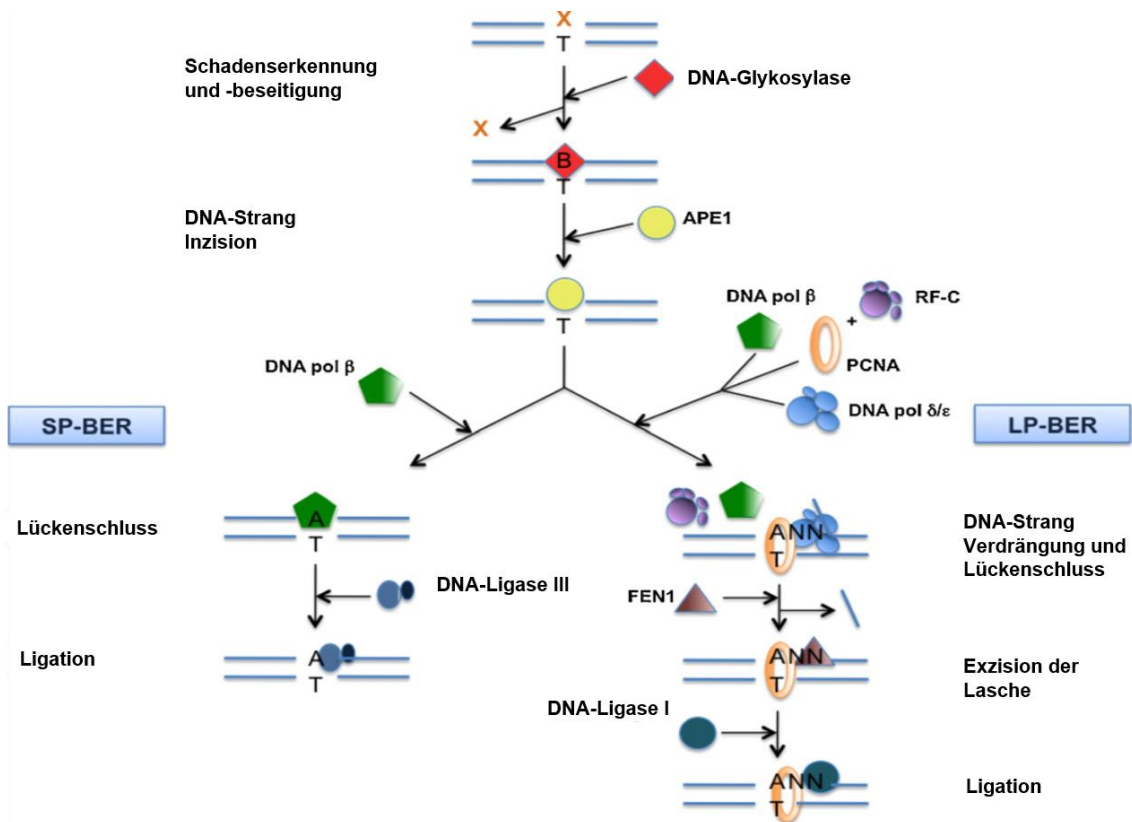
Ein DNA-Schaden bezeichnet eine Veränderung im chemischen Aufbau der DNA, der durch DNA-Reparatur rückgängig gemacht werden muss, um eine Mutation zu verhindern. Pro Tag entstehen durchschnittlich 60.000 DNA-Schäden in einer Säugetierzelle [4]. Schadensauslöser sind beispielsweise energiereiche Strahlung, wie UV- [5, 6], Röntgen- [7-9] und Gammastrahlung [10]. Zum anderen sind auch die Oxidation durch Stoffwechselprodukte [11], die Hydrolyse und die Wirkung mutagener Substanzen, welche zu Alkylierungsreaktionen [12-14] führen können, an Veränderungen im Aufbau der DNA beteiligt. Häufig kommt es hierbei auch zu Einzel (SSBs, engl. single strand breaks)- [15] oder Doppelstrangbrüchen (DSBs, engl. double strand breaks) [16] und weiteren Schäden, die eine ordnungsgemäße Replikation oder Transkription verhindern [17]. Evolutionär haben sich deshalb verschiedene Systeme zur Reparatur der oben erläuterten Schadensbilder entwickelt. Diese führen im Menschen täglich ungefähr  $10^{16}$ - $10^{18}$  Reparaturprozesse durch [18]. Im Rahmen dieser Arbeit liegt das Interesse auf der Basen-Exzisionsreparatur, die zur Gruppe der Exzisionsreparaturmechanismen gezählt wird.

### 1.3 Basen-Exzisionsreparatur

Die Entdeckung des Basen-Exzisions-Reparaturweges geht auf Tomas Lindahl zurück. Bei der Suche nach einem neuen Uracil-Reparaturenzym entdeckte er im Jahr 1974 die Uracil-DNA-Glykosylase (UDG) und beschrieb daraufhin auch als Erster den gesamten Reparaturprozess der Basen-Exzisionsreparatur (BER) [19]. Da der BER eine wichtige Rolle als Reparaturmechanismus von DNA-Schäden zugeschrieben wird, unterliegt dieser Mechanismus intensiver Forschung. So wurden bereits eine Vielzahl an umfangreichen Übersichtsarbeiten [20-22], wie beispielsweise die von Krokan und Bjørås [23] oder Friedman und Stivers [24] veröffentlicht. Zielschäden der BER entstehen durch verschiedene chemische Reaktionen wie Oxidation, Alkylierung oder Desaminierung der DNA [20, 25, 26]. Im Wesentlichen führen diese Modifikationen aber nur zu geringen Verzerrungen im Aufbau des DNA-Gerüsts [23, 27-29]. Die DNA-Glykosylasen sind hierbei die initialen Enzyme der BER [26]. Sie binden an die kleine Furche der DNA [30], wobei eine Glykosylase jeweils eine Gruppe von spezifischen, meist strukturell verwandten Basenschäden, bzw. Basenfehlpaarungen erkennt [29]. Das DNA-Rückgrat wird von der Schadenseite aus durch die Glykosylase gebogen, bzw. „gekinkt“ (von engl. kink, der Knick), um so die aus der großen Furche „geflippte“, bzw. herausgeschwenkte Base in ihrem katalytischen Zentrum fixieren zu können [26, 30, 31]. Die Glykosylasen lassen sich in mono- und bifunktionelle Glykosylasen unterteilen [22, 23]. Die monofunktionellen Glykosylasen besitzen eine reine Glykosylase-Funktion. So wird nur die Bindung zwischen der Base und dem Desoxyribose-Zuckerring gespalten und eine abasische Stelle (engl. apurinic (AP)-site) erzeugt. Zum Schneiden des Rückgrates folgt auf eine monofunktionelle Glykosylase die AP-Endonuklease APE1 [32, 33]. Bifunktionelle Glykosylasen zeichnen sich durch eine zusätzliche  $\beta$ -Lyase- oder eine  $\beta/\delta$ -Lyase-Funktion aus, die auch das dazugehörige Zucker-Phosphatketten-Rückgrat schneidet [34]. Obwohl bifunktionelle Glykosylasen das DNA-Rückgrat selbst schneiden, ist APE1 hier für die Vorbereitung der Schnittstellen für den späteren Lückenschluss durch eine DNA-Polymerase notwendig [35]. Im Anschluss an die Glykosylase teilt sich der Reparaturweg in eine sog. „Short Patch“(SP)-BER, die weitaus häufiger vertreten ist, und die „Long Patch“ (LP)-

BER [36]. Im Gegensatz zur „Short Patch“-BER mit einer einzelnen Nukleotid-Lücke (sog. single nucleotide gap) werden in der „Long Patch“-BER zwei bis sechs Nukleotide ersetzt [37-39]. Nachdem die schadhafte Base entfernt und das DNA-Rückgrat getrennt ist, füllt eine DNA-Polymerase die entstandene Lücke. Im Falle der „Short Patch“-BER fügt die DNA-Polymerase  $\beta$  (Pol  $\beta$ ) ein neues komplementäres Nukleotid, in der „Long Patch“-BER fügt dagegen der Komplex aus dem Protein PCNA (engl. Proliferating-Cell-Nuclear-Antigen) und Pol  $\delta/\epsilon$  mehrere komplementäre Nukleotide ein [39, 40]. Im Anschluss verbindet eine der beiden DNA-Ligasen LIG1 (LP-BER) oder LIG3 (SP-BER) die Zucker-Phosphat-Enden [22, 41]. Die Basen-Exzisionsreparatur gilt allgemein als ein essenzieller DNA-Schadensreparaturweg, um einer Tumorentstehung, einem vorzeitigen Altern und einer neuronalen Degeneration entgegen zu wirken [42-45]. Zunehmend rückt aber auch der Einfluss der BER auf das erworbene Immunsystem [46], die Epigenetik [47, 48] und den Stoffwechsel [49] in den Fokus der Forschung. Bemerkenswert ist die Tatsache, dass der Ausfall einer einzelnen Glykosylase meist keine ausgeprägten phänotypischen Folgen hat, wie dies von einer gestörten Nukleotid-Exzisionsreparatur oder Doppelstrangbruch-Reparatur bekannt ist [23]. Die Zielschäden von verwandten Glykosylasen, wie beispielsweise NTHL1 und NEIL1(2), sind sich häufig sehr ähnlich bzw. identisch [44, 50]. Eine Ausnahme stellt hierbei die Glykosylase hTDG dar, da diese neben der Basen-Reparatur auch für die Aktivierung von DNA mittels Demethylierung benötigt wird [51, 52]. Ähnlich verhält es sich mit der Glykosylase hOGG1, deren Ausfall den Glukosestoffwechsel beeinträchtigt, dies aber keine direkte Auswirkung auf die Ausprägung des Phänotyps hat [49]. Eine Übersicht über den Basenreparaturprozess findet sich nachfolgend in Abbildung 2.





**Abbildung 2: Übersicht der Basisexzisionsreparatur (BER)**

Die Basisexzisionsreparatur (BER) spaltet sich nach der initialen Basenexzision durch eine Glykosylase in zwei Reparaturwege: den „Short-Patch“ (SP)- oder den „Long-Patch“ (LP)-Reparaturweg. Eine monofunktionelle Glykosylase benötigt für das Schneiden des Phosphatrückgrates zusätzlich die Endonuklease APE1. Bifunktionelle Glykosylasen besitzen dagegen eine im Enzym integrierte Endonuklease-Funktion. Die folgenden Reparaturprozesse gleichen sich in den weiteren Kernschritten: eine Polymerase synthetisiert das (die) zu ersetzende(n) Nucleotid(e), nur im LP-Reparaturweg ist zusätzlich die Flap-Endonuklease FEN1 für die Entfernung der Klappe (engl. flap) nötig. Eine Ligase wie LIG1 oder LIG3 beendet den Reparaturprozess durch Ligation des Phosphat-Endes mit der Desoxyribose am DNA-Rückgrat.

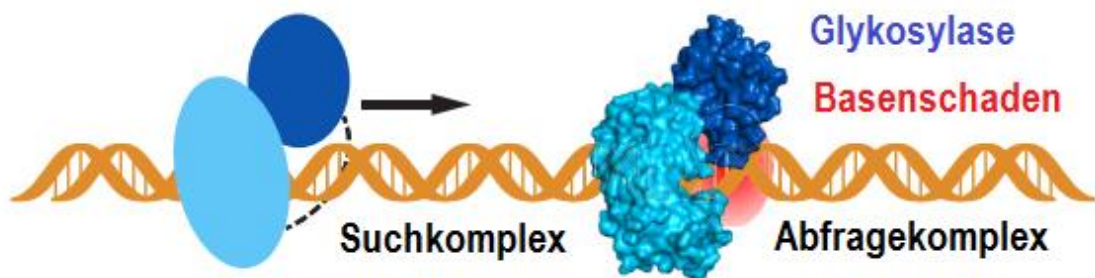
modifiziert übernommen aus [41]

## 1.4 Glykosylasen

Im Allgemeinen lassen sich die verschiedenen bisher bekannten Glykosylasen gemäß ihrem strukturellen Aufbau in sechs Superfamilien einordnen. Zwei große Superfamilien sind die HhH-DNA-Glykosylasen (engl. helix-hairpin-helix) und die H2TH-Glykosylasen (engl. helix-two-turn-helix). Daneben werden noch die Superfamilien von hAAG, UDG, EndoV und die HLR-Glykosylasen (engl. HEAT-like repeat) unterschieden [20]. Je nach vorliegendem Basenschaden existieren schadensspezifische DNA-Glykosylasen, die im Folgenden die BER-Reparaturkaskade initiieren. Während die monofunktionellen Glykosylasen für den Glykosylase-Prozess ein Wassermolekül als Nukleophil verwenden, um die N-glykosidische Basen-Zuckerbindung zu schneiden und so eine abasische Stelle (AP-Site) zu generieren, nutzen bifunktionelle Glykosylasen eine Aminogruppe. So entsteht bei der Deprotonierung eines Lysin-Restes durch Aspartat als Zwischenprodukt eine Schiff'sche Base. In der darauffolgenden  $\beta$ -Elimination wird dieses Zwischenprodukt zur Spaltung der 3'-Phosphodiester-Bindung genutzt [33, 53]. Zusätzlich tragen die Glykosylasen zu Kommunikationszwecken häufig noch spezielle endständige Proteinstrukturen. So befindet sich beispielsweise im aktiven Zentrum einiger Vertreter der Superfamilie der HhH-DNA-Glykosylasen ein  $[\text{Fe}_4\text{S}_4]^{2+}$ -Cofaktor [54]. Ein anderes Beispiel für eine zusätzliche Domäne ist der Zinkfingerkomplex in MutM (der prokaryotischen Variante von hOGG1) zur Fixierung des „geflippten“ Basenzustandes [55].

Die jeweilige Glykosylase erkennt ihr Substrat durch drei aufeinander folgende Konformationszustände: Such-, Abfrage- und Exzisionskomplex. Im Suchkomplex (SC: engl. search complex) bildet sich ein Komplex aus DNA und DNA-Glykosylase, der sich anschließend am DNA-Doppelstrang entlang bewegt und nach möglichen Basendefekten sucht. Die Schadenssuche wird in Abbildung 3 verdeutlicht. In der Literatur wird der Suchvorgang mit zwei Hauptmodi, dem lose verbundenen, "gleitend" assoziativen Modus und dem „hüpfenden“ dissoziativen Modus, beschrieben [24, 56-59]. Durch das Binden der Glykosylasen an die DNA wird das DNA-Rückgrat Glykosylase-spezifisch gebogen. Der für die jeweilige Glykosylase spezifische Winkel kann zwischen  $22^\circ$  und  $70^\circ$  betragen [60]. Die Aufgabe des Abfragekomplexes (IC: engl. interrogation complex) ist die

Untersuchung der DNA zur Erkennung des Schadenssubstrats, um anschließend im Exzisionskomplex (EC: engl. excision complex) die geflippte Base, die im aktiven Zentrum gebunden ist, zu entfernen [1, 24]. Es ist allerdings bisher unklar, ob Glykosylasen die DNA aktiv biegen, um die Basen extrahelikale zu flippen [30], oder die Glykosylasen die Base in einem spontan extrahelikalen Zustand binden [61]. Im Folgenden werden die für die Arbeit relevanten Glykosylasen vorgestellt. Exemplarisch wird anhand der Glykosylase hOGG1 die Schadenssuche detailliert erklärt, wobei sich der Prozess der Schadenssuche allgemein auf andere Glykosylasen übertragen lässt.



**Abbildung 3: Schadenssuche der Glykosylase**

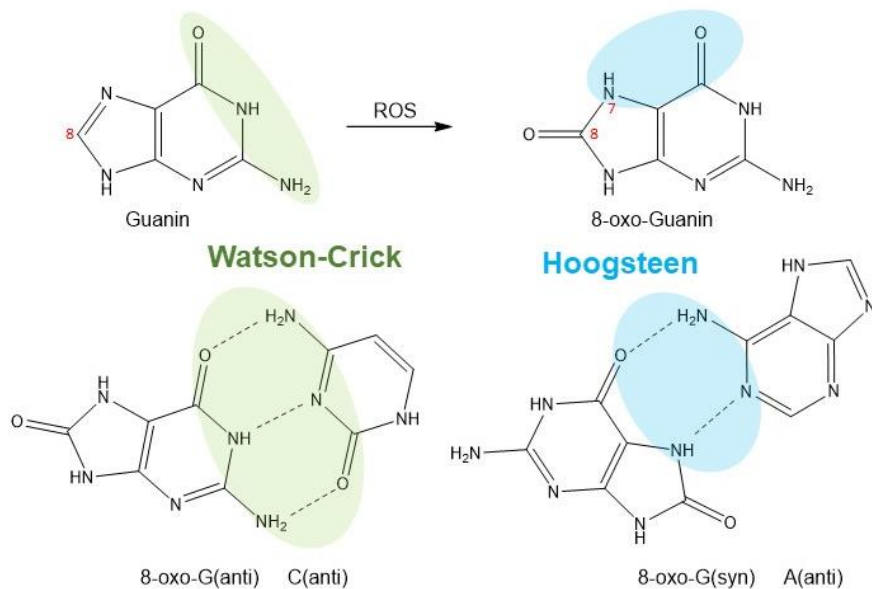
Die Glykosylase (z.B. MutY: N-terminale Domäne in violett und C-terminale Domäne in blau) sucht entlang der DNA nach einer spezifischen Läsion. Es wird sowohl der passive Suchkomplex, als auch der aktive Abfragekomplex gezeigt

modifiziert übernommen aus [62]

### 1.4.1 hOGG1

Die Glykosylase hOGG1 (humane Oxoguanin Glykosylase 1) ist für die Beseitigung des oxidativen DNA-Basenschadens 8-Oxo-7,8-Dihydroguanin (oxoG) zuständig, der mit über  $10^3$  Ereignissen pro gesunde Zelle und Tag als der häufigste DNA-Schaden gilt [41, 63]. Die molekulare Masse beträgt ca. 38 kDa [64, 65]. Die Oxidierungsschäden werden durch freie Sauerstoffradikale (ROS, engl. reactive oxygen species), als ständig anfallende Nebenprodukte des zellulären Metabolismus erzeugt [27, 41, 66, 67]. Daneben führt auch ionisierende Strahlung [23, 68] und UV-Strahlung [69] zur Bildung von ROS und bewirkt somit eine Schädigung [70] beispielsweise in Form von Einzelstrangbrüchen (SSBs, engl. single strand breaks). Als weitere Schädigung energiereicher Strahlung können Doppelstrangbrüche (DSBs, engl. double strand breaks) [71] und Thymin-Dimerisierung entstehen [72]. Das durch ROS entstandene Hydroxylradikal ( $\cdot\text{OH}$ ) oxidiert außerdem an Guanin vorrangig das C8. Das dabei entstehende radikalische

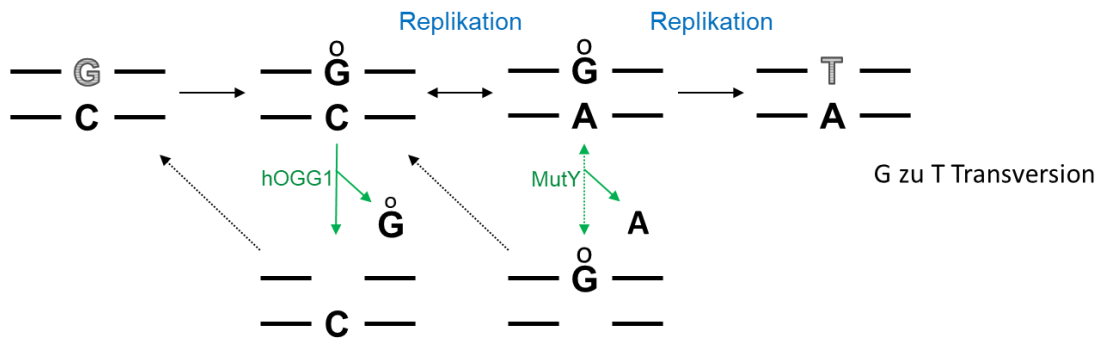
Zwischenprodukt kann zum einen durch weitere Oxidation zum stark mutagenen 8-Oxo-7,8-Dihydroguanin (oxoG) reagieren [73] oder aber auch zum ringoffenen 2,6-Diamino-4-hydroxy-5-Formamidopyrimidin (FapyG) reduziert werden [74]. Die Oxidation hat zunächst, wie in Abbildung 4 zu sehen ist, keinen direkten Einfluss auf die Watson-Crick Basenpaarung von oxoG und Cytosin (C) mit drei Wasserstoffbrückenbindungen, sofern sich beide Basen in der anti-Konformation befinden. Aufgrund von repulsiven Wechselwirkungen kann das oxoG aber auch in die syn-Konformation rotieren, die dann eine stabile Hoogsteen-Bindung mit Adenin (A) begünstigt [75]. Dies hat nach zwei Replikationsrunden in der DNA eine Guanin zu Thymin-Transversion zur Folge [76, 77] (siehe Abbildung 5).



**Abbildung 4: Entstehung und Struktur des oxoG-Schadens**

Die primäre oxoG-Läsion entsteht durch den Angriff einer reaktiven Sauerstoffspezies (ROS) an dem C8 des Guanins. Allerdings kommt es hierbei primär zu keiner Veränderung der Watson-Crick-Oberfläche des Guanins. Folglich ist die Watson-Crick-Paarung von oxoG:C entsprechend G:C mit drei Wasserstoffbrückenbindungen weiterhin möglich. Jedoch ist für die oxoG-Base nun neben der Watson-Crick-Basenpaarung von oxoG(anti) mit Cytosin(anti), auch eine zweite Bindungskonformation möglich. OxoG kann nach einer Rotation in die oxoG(syn)-Konformation zusätzlich mit Adenin(anti) ein stabiles Hoogsteen-Basenpaar bilden. Dies ist ein Zielschaden der Glykosylase MutY.

modifiziert übernommen aus [78]



**Abbildung 5: DNA-Reparatur und -Replikation des oxoG-Basenschadens**

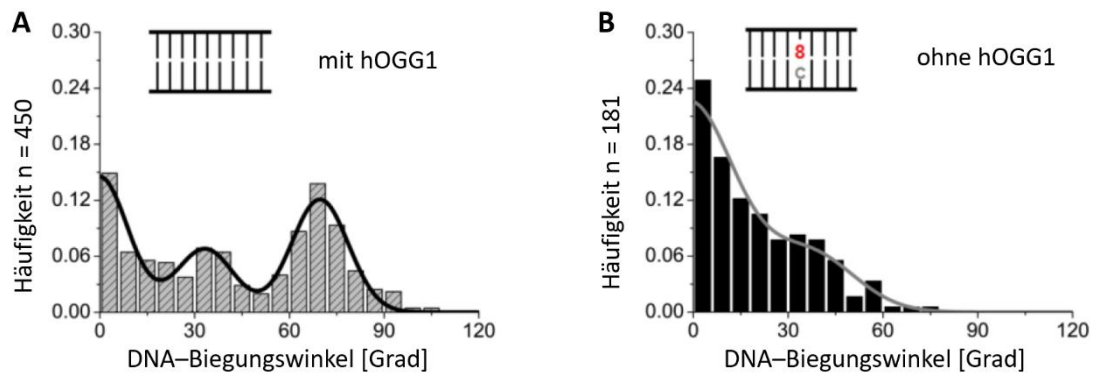
Der durch ROS entstandene Basenschaden oxoG wird im Idealfall durch die Glykosylase hOGG1 erkannt. Nach Entfernung des oxidierten Guanins durch hOGG1 verbleibt eine abasische Stelle. Im weiteren Reparaturmechanismus wird dann von einer Polymerase durch die Insertion eines Guanins wieder die physiologische G:C-Basenpaarung hergestellt. Findet keine adäquate Schadensreparatur durch hOGG1 statt und der oxoG-Strang wird repliziert, so paart die Läsion vorzugsweise mit einem Adenin zu einem oxoG:A-Intermediat. Dies stellt ein Schadenssubstrat für die Glykosylase MutY dar. Nachdem das Adenin entfernt ist, ergänzen wieder Polymerasen die fehlende Base. Es können zwei unterschiedliche Basen eingebaut werden; zum einen Cytosin, wobei sich dann oxoG:C, der Zielschaden von hOGG1 ergibt oder Adenin, wobei wieder oxoG:A entsteht. Wird der oxoG:A Schaden nicht durch die Glykosylase MutY erkannt, kommt es im nächsten Replikationsschritt zu einer Punktmutation durch eine G-C zu T-A-Transversion.

modifiziert übernommen aus [79]

Die bifunktionelle Glykosylase hOGG1 besitzt neben ihrer Glykosylasefunktion zur Basenentfernung auch eine  $\beta$ -Lyase-Aktivität, die das DNA-Rückgrat schneidet [23, 66, 80]. Prinzipiell wird das oxoG über den SP-Reparaturweg entfernt, im Falle eines Pol  $\beta$ -Defektes (SP-Reparatur) kann dies auch durch den LP-Reparaturweg erfolgen [36]. Die Glykosylase hOGG1 zählt, aufgrund des Vorkommens einer Helix-hairpin-Helix-Struktur, zur HhH-Superfamilie. Daneben besitzt diese auch eine Glycin/Prolin-reiche Schleife und ein Aspartat-HhH-GPD-Motiv [23, 29, 66]. Die Glykosylase hOGG1 (und in Prokaryonten die Formamidopyrimidin-DNA-Glykosylase FPG bzw. MutM [63, 81]) repariert sowohl oxoG, das mit Cytosin gepaart ist, als auch das strukturverwandte ringoffene 2,6-Diamino-4-hydroxy-5-Formamidopyrimidin (FapyG) [66]. Die Glykosylase hOGG1 sucht diffusionsgetrieben und eindimensional gleitend die DNA ab [66]. Wird oxoG:C als Schaden erkannt, kommt es zum Flippen der Base in die aktive Tasche des Enzyms, um anschließend die N-glykosidische Basenbindung hydrolytisch zu spalten. Das Enzym durchläuft hierbei drei Konformationsänderungen. Nach dem Erkennen und Herausflippen von oxoG wird ein Bestandteil der Enzymstruktur in den entstandenen Freiraum der DNA inseriert, um das oxoG korrekt für den

Katalase-Vorgang in der aktiven Tasche zu positionieren [82]. Das Flippen erfolgt vermutlich über die große Furche, da hOGG1 entlang der kleinen Furche gleitend nach Schäden sucht. Im Komplex aus Glykosylase und DNA-Strang wird die DNA dabei um 70° vom Protein gebogen [1, 26, 66, 83]. Um einen etwaigen Schaden vor der Replikation reparieren zu können, muss jedoch die gesamte DNA sehr zügig untersucht werden, was das Ausschwenken jeder einzelnen Base in die Bindungstasche unmöglich macht [78]. In der richtungsweisenden Arbeit von Büchner *et al.* [1] zur initialen Schadenserkenkung von Glykosylasen wurden bei der Interaktion der Glykosylase hOGG1 mit ungeschädigter DNA verschiedene intrinsische Biegungswinkel beobachtet. Eine Übereinstimmung mit den DNA-Biegungen am oxoG-Schaden weist auf eine Optimierung der Glykosylase für ihren Zielschaden und daraus resultierende initiale Erkennung hin (siehe Abbildung 6). Des Weiteren wurde von Li *et al.* [78] im initialen Suchvorgang von hOGG1 ein schneller und energieeffizienter Reaktionsweg gefunden, der nur die Eversion der schadhafte Base aufgrund der Ausbildung bestimmter Wasserstoffbrückenbindungen zwischen hOGG1 und oxoG erlaubt. Die Ausstülpung von Guanin wird in diesem mehrschrittigen Prozess schon frühzeitig verhindert.

Für die Basenexzision besitzt hOGG1 als Nukleophil ein Lysin und im katalytischen Zentrum ein Aspartat [84]. Im weiteren Prozess der BER wird die erzeugte Nukleotidlücke mit Guanin geschlossen und so wieder die korrekte DNA-Sequenz hergestellt. Findet keine adäquate Schadensreparatur des oxoG:C-Schadens statt und wird der oxoG-Strang repliziert, paart die Läsion vorzugsweise mit einem Adenin zu einem oxoG:A-Intermediat. Der hierbei entstandene Schaden ist das Substrat der Glykosylase MutY. Scheitert auch der Reparaturprozess von MutY führt die Replikation des A-Strangs im oxoG:A-Intermediat zu einer G:C- zu T:A-Transversionsmutation. Der Prozess der punktuellen Mutation durch eine Transversion ist in Abbildung 5 zu sehen. Diese Punktmutation ist gehäuft in den Zellen einer ganzen Reihe von Karzinomen, wie dem Ovar-, Brust-, Lungen-, Nieren- und kolorektalen Karzinom zu finden [85]. Die Transversion ist eine von verschiedenen Punktmutationen im Tumorsuppressor-Gen p53 [86]. Auch für das Cockayne-Syndrom wird ein Zusammenhang mit der Glykosylase hOOG1 beschrieben [87, 88].



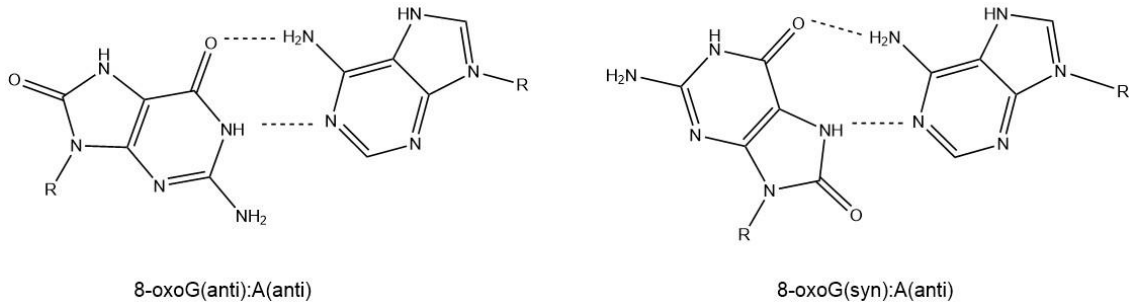
**Abbildung 6: Biegungsaktivität von hOGG1**

- A:** In den Untersuchungen von Büchner *et al.* [1] zur initialen Schadenserkennung von Glykosylasen wurden bei der Interaktion von hOGG1 und ungeschädigter DNA drei intrinsische Biegungswinkel ( $0^\circ$  (44%),  $30^\circ$  (20%) und  $70^\circ$  (36%)) beobachtet.
- B:** Im Diagramm überwiegt ohne den Einfluss der Glykosylase hOGG1 an der Schadensstelle oxoG:A ein intrinsischer DNA-Biegungswinkel von  $0^\circ$  (78%), zusätzlich findet sich bei  $30^\circ$  (22%) eine kleine Biegungsspezies.

modifiziert übernommen aus [1]

### 1.4.2 MutY

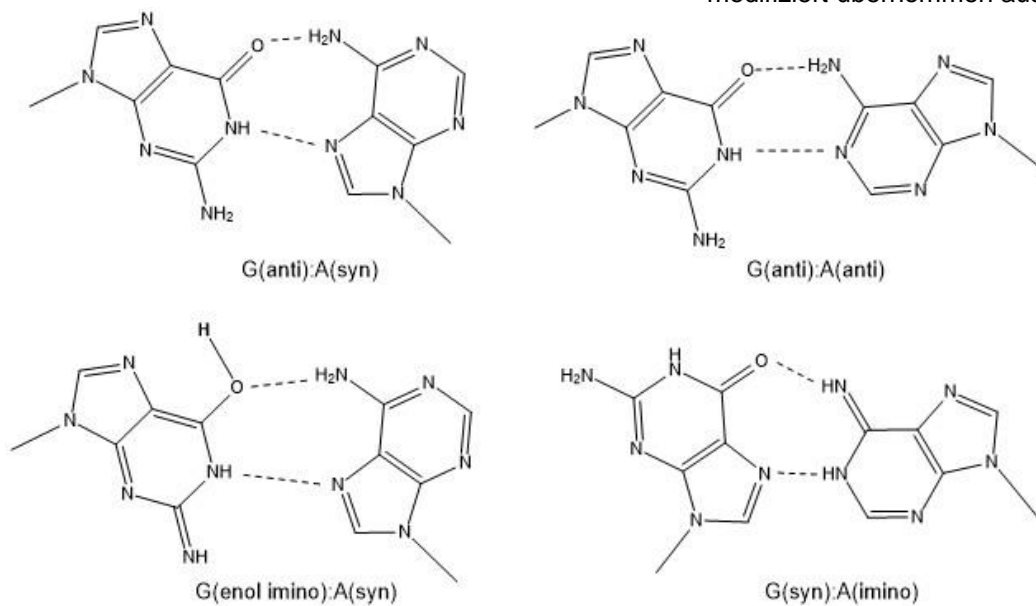
Die bakterielle Adenin-DNA-Glykosylase MutY (*E. coli*) mit dem menschlichen Äquivalent MutYH ist eine monofunktionelle Glykosylase [66, 89], die für die Entfernung eines Adenins, gepaart mit einem 7,8-Dihydro-8-Oxoguanin (oxoG) oder Guanin (G), zuständig ist [23, 28]. Ihre molekulare Masse beträgt etwa 39 kDa [90]. Die Zielschäden von MutY sind oxoG:A und G:A. Die Entstehung des Schadens oxoG:A hängt mit der im vorherigen Kapitel erwähnten Glykosylase hOGG1 und deren Schadenssubstrat oxoG:C zusammen. Das durch ROS oxidierte Guanin (oxoG) kann neben einer Watson-Crick-Basenpaarung mit Cytosin, zusätzlich eine Hoogsteen-Basenpaarung mit Adenin eingehen. Die Entstehung des oxoG-Schadens ist in Abbildung 4 dargestellt. Das Schadenssubstrat G:A entsteht aus oxoG:A im Zusammenhang mit hOGG1 durch die Reparatur des oxoG. Brown *et al.* [91] beschreiben verschiedene verzerrte Konformationen ohne eine energetisch bevorzugte zu benennen. In Abbildung 7 und Abbildung 8 werden die verschiedenen Konformationen als Strukturformel dargestellt.



**Abbildung 7: Konformationen der oxoG:A-Fehlpaarung**

Das Schadenssubstrat oxoG:A wird in der anti-anti- und in der syn-anti-Konformation gezeigt. Die gestrichelten Linien stellen die Ausbildung der Wasserstoffbrücken dar. Die stabile Hoogsteen-Konformation in oxoG(syn):A(anti) ist thermodynamisch begünstigt.

modifiziert übernommen aus [62]



**Abbildung 8: Konformationen der G:A-Fehlpaarung**

In der Abbildung werden die bekannten Konformationen für die Guanin-Adenin-Basenpaarung veranschaulicht.

modifiziert übernommen aus [91]

MutY besitzt an der N-terminalen Domäne (NTD) eine Helix-hairpin-Helix-Struktur [92], weshalb es zu der HhH-Superfamilie der DNA-Glykosylasen gezählt wird. Daneben findet man im katalytischen Zentrum eine Glycin/Prolin-reiche Schleife, sowie einen Aspartat-Rest, der von katalytischer Bedeutung ist [93]. Außerdem besitzt diese Glykosylase den Cofaktor  $[4\text{Fe-4S}]^{2+}$ . Dagegen hat die C-terminale Domäne (CTD) die Erkennung von oxoG zur Aufgabe und entspricht im Aufbau einer 8-oxo-dGTPase [94].



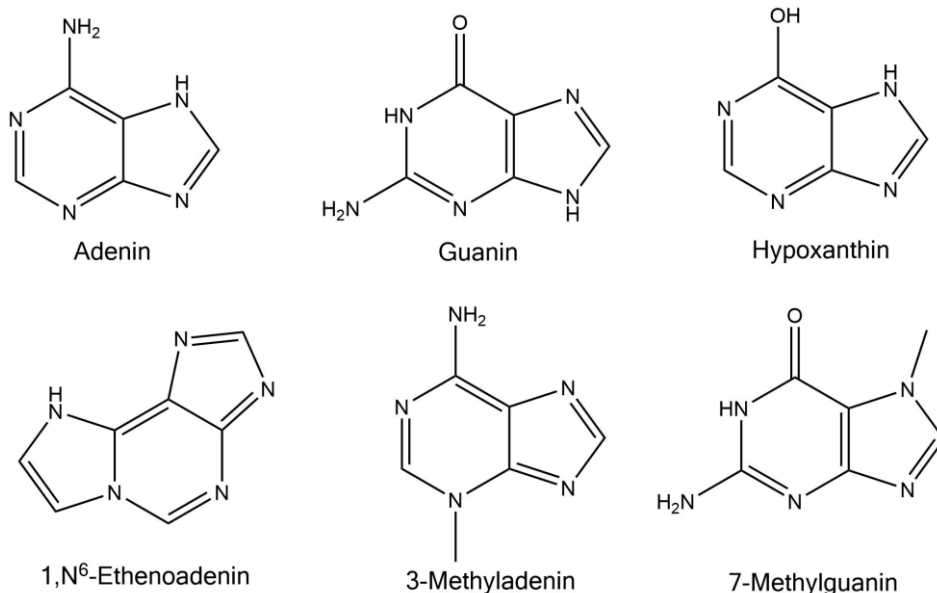
Die Wechselwirkungen zwischen der DNA und der Glykosylase MutY ähneln denen von hOGG1 [66]. Während in diesem Fall das Adenin aus der DNA-Doppelhelix in die aktive Tasche von MutY geschwenkt und der DNA-Strang um 55° gebogen wird, verbleibt oxoG intrahelikal und bildet Wasserstoffbrückenbindungen mit der NTD und der CTD von MutY aus [93]. Ein HhH-Motiv der NTD bindet dabei an die kleine Furche der DNA, und eine N-terminale Schleifen-Region interagiert innerhalb der großen Furche mit oxoG:A [62]. Eine Lysin-reiche Domäne bewirkt zunächst zusammen mit dem katalytischen Sechs-Helix-Fass-Strukturelement durch elektrostatische Wechselwirkungen mit der DNA die Entspiralisierung der Doppelhelix. Zusätzlich wird ein Tyrosin-Rest (Tyr 88) zwischen den oxoG-Schaden und dessen 5'-Nachbar in die Doppelhelix eingeschoben und die DNA geweitet, sodass im Folgenden ein Glycin-Rest (Gln 48) den dabei entstandenen Raum besetzen kann. Auf diese Weise wird die Doppelhelix destabilisiert [66]. Das aus oxoG und Adenin gebildete Hoogsteen-Face wird durch ausgebildete Wasserstoffbrückenbindungen zwischen oxoG und einem Serin-Rest (Ser 308) erkannt und dient somit auch der korrekten Schadenserkennung. Zusätzlich muss oxoG eine 180°-Drehung von der syn- zur anti-Konformation vollziehen, wenn es an MutY gebunden ist, was zu sterischen Abstoßungen zwischen Adenin und oxoG führt und folglich die Extrusion von Adenin in die extrahelikale Bindungstasche begünstigt [93]. Wichtig für die Erkennung des richtigen Schadenssubstrates und einer weiteren Destabilisierung der Doppelhelix am Schadensort ist die Tatsache, dass sich in der oxoG(syn):A(anti)-Konformation die Aminogruppe von oxoG auf der Seite der großen Furche befindet, wohingegen dieselbe Aminogruppe von Guanin in der G:C-Paarung in der kleinen Furche zu liegen kommt. Dies ist ein wesentlicher struktureller Unterschied zwischen dem geschädigten Basenpaar (oxoG:A) und dem Watson-Crick-Basenpaar (G:C) [95]. Nachdem sich das Adenin korrekt in der Bindungstasche der Glykosylase befindet, entfernt MutY im SP-Reparaturweg das Adenin von dem oxoG:A-Paar. Im Zuge des Reparaturweges wird dieses durch ein Cytosin ersetzt. Die Reparatur des entstandene oxoG:C ist im Folgenden wieder die Aufgabe der DNA-Glykosylase hOGG1 (MutM). Findet vor der Replikation keine adäquate Schadensbehebung durch MutY statt, kommt es häufig zu einer Punktmutation in Form einer

G:C- zu A:T-Transversion (Abbildung 5). Dies kann sich durch verstärktes Auftreten von kolorektalen Karzinomen [96] oder die Beeinträchtigung des APC-Tumorsuppressors äußern [97, 98].

### 1.4.3 hAAG

Die monofunktionale humane 3-Methyladenin-DNA-Glykosylase (Alkyladenin-Glykosylase; hAAG, 33 kDa [99-101]) hat die höchste Affinität zu dem Schaden 3-Methyladenin [22], weshalb sie auch als Methylpurin-DNA-Glykosylase (MPG [102]) bezeichnet wird. Außerdem ist sie die bisher einzige in menschlichen Zellen bekannte DNA-Glykosylase, die Alkylierungsschäden reparieren kann. Zu diesen zählen neben 3-Methyladenin (3MeA), auch 1,N<sup>6</sup>-Ethenoadenin ( $\epsilon$ A), Hypoxanthin (Hx) und 7-Methylguanin (7MeG) (Abbildung 9) [25, 103-105]. So entsteht Hypoxanthin beispielsweise durch die Reaktion von Adenin mit reaktiven Stickstoffverbindungen (engl. reactive nitrogen species). Da Hypoxanthin im Unterschied zu Adenin anstelle von Thymin mit Cytosin paart, führt die Adenin-Desaminierung zu einer mutagenen Transversion [104]. 7-Methylguanin ist auf eine Methylierungsreaktion durch Nitrosamine am N<sup>7</sup> des Guanins zurückzuführen [104, 106]. Als das eigentliche physiologische Schadenssubstrat wird allerdings 1,N<sup>6</sup>-Ethenoadenin ( $\epsilon$ A) angesehen, da hierzu die höchste Affinität besteht [107]. Das 1,N<sup>6</sup>-Ethenoadenin ( $\epsilon$ A) paart zunächst mit Thymin. Bei einer ausbleibenden Reparatur kann  $\epsilon$ A durch alle drei möglichen Basen ersetzt werden, wobei die A zu G-Transversion bevorzugt entsteht [108]. Gemäß den Untersuchungen von Guliaev *et al.* [109] existieren für das Schadenssubstrat  $\epsilon$ A:T keine Wasserstoffbrückenbindungen (siehe Abbildung 10). Aktiv ist hAAG ausschließlich in doppelsträngiger DNA, da zu beiden Strängen der DNA-Helix eine Bindung besteht. Im Vergleich zu anderen Glykosylasen besitzt hAAG eine relativ breite Substratspezifität [23]. Die verschiedenen DNA-Schäden erzeugen jeweils keine oder eine kaum merkliche DNA-Helix-Verzerrung [110]. Dies erschwert eine einfache Schadenserkenkung anhand von DNA-Strukturveränderungen. Erkennt hAAG einen spezifischen Schaden, so wird die geschädigte Base in die aktive Tasche geflippt und das Nukleotid über die Ausbildung von Bindungen zwischen der Proteinstruktur und den dem Schaden benachbarten Phosphaten im katalytischen Zentrum fixiert [105]. Dies geschieht, indem ein Tyrosin-Rest an einer  $\beta$ -

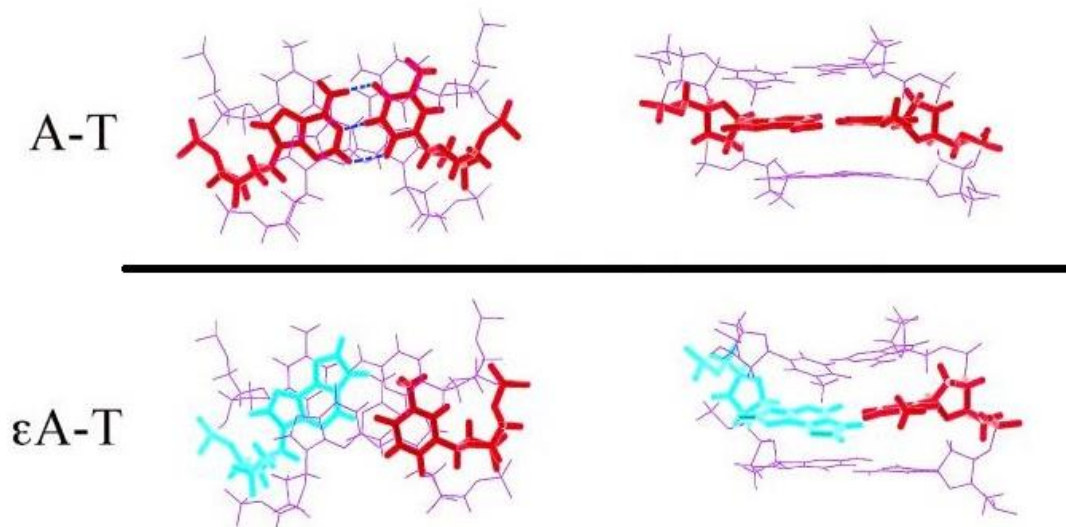
Haarnadel über die kleine Furche in den entstandenen Freiraum greift. Die Insektion der Tyrosin-Seitenkette (Tyr-162) in die zwischen den Basen entstandene Lücke führt an dieser Stelle zu einer Biegung (Kink) der B-förmigen Helix um 22° vom Protein weg (bezüglich der acht zentralen Basenpaare) [105]. Auf diese Weise wird der Freiraum blockiert und die DNA-Helix stabilisiert. [107, 111]. An der Tyrosin-Seitenkette Tyr<sup>162</sup> befinden sich auch das Methionin Met<sup>164</sup> und das Tyrosin Tyr<sup>165</sup>. Diese beiden Aminosäuren weiten außerdem auf der 3'-Seite der schadhaften Base die kleine Furche auf und helfen so, den DNA-Strang fortlaufend in diese Richtung abzusuchen. Tyr<sup>165</sup> drückt auf den Zuckerring von T<sup>9</sup>, destabilisiert so die Bindung zu T<sup>8</sup>, was schließlich die für den Biegeprozess benötigte Energie verringert [105]. Zusätzlich wird während der katalytischen Reaktion in der aktiven Tasche ein Wasser-Molekül genutzt, dass als Nukleophil agiert und die umliegenden sauren Aminosäuren aktiviert, um die N-glykosidische Bindung zur alkylierten Base zu spalten [105]. Ein Genpolymorphismus der 3-Methyladenin-DNA-Glykosylase steht im Zusammenhang mit der rheumatoiden Arthritis [112], ebenso ist ein gehäuftes Vorkommen von 1,N<sup>6</sup>-Ethenoadenin (εA), dem Hauptschaden von hAAG, durch oxidativen Stress mit dem Auftreten von Adenokarzinomen der Lunge vergesellschaftet [113].



**Abbildung 9: Strukturformeln von hAAG-spezifischen DNA-Läsionen**

Gezeigt sind die Strukturformeln spezifischer DNA-Läsionen der DNA-Glykosylase hAAG, einschließlich der beiden unmodifizierten Purine Adenin und Guanin.

modifiziert übernommen aus [103]



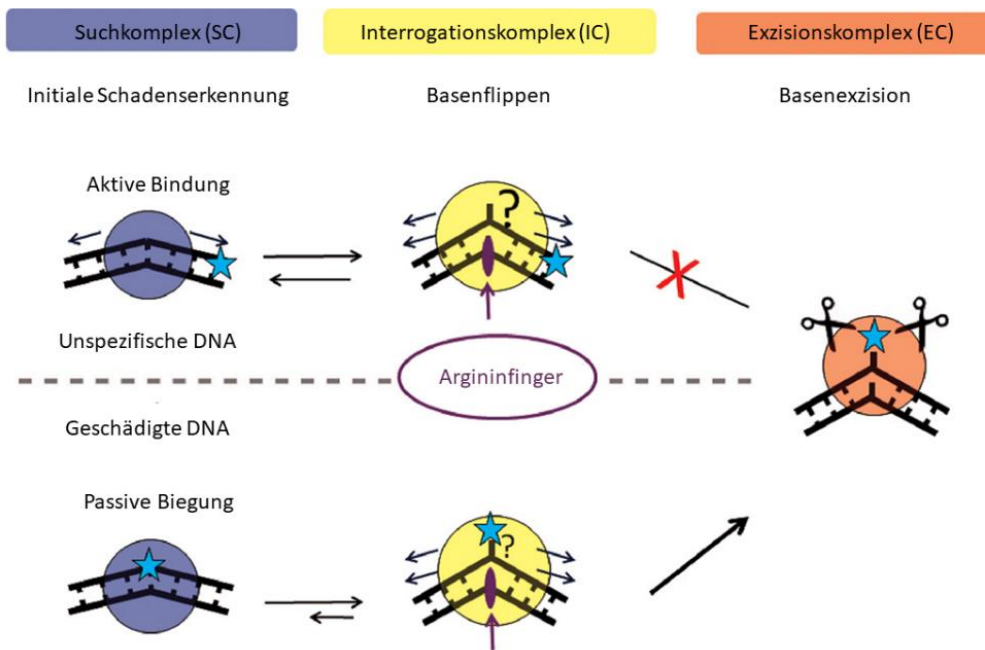
**Abbildung 10: Struktureller Vergleich der Watson-Crick-Basenpaarung A:T mit dem εA:T-Basenpaar**

In der Abbildung ist oberhalb das Watson-Crick-Basenpaar aus Adenin und Thymin zu sehen. Zwischen den beiden Basen kommt es zu der Ausbildung von drei Wasserstoffbrückenbindungen (blau gestrichelte Linien). Dagegen kommen in der unten dargestellten 1,N<sup>6</sup>-Ethenoadenin-Thymin-Paarung keine Wasserstoffbrücken vor. Aufgrund der erhöhten Flexibilität entsteht ein sog. Wobble-Basenpaar.

modifiziert übernommen aus [109]

## 1.5 Vorausgegangene Arbeiten und das Modell zur initialen Schadenserkenkung der BER

Die Doktorarbeit fußt auf den Forschungsergebnissen der Arbeitsgruppe Tessmer des Rudolf-Virchow-Zentrums Würzburg. So veröffentlichten Büchner *et al.* [1] für die humane Thymin-DNA-Glykosylase (hTDG) ein Modell zur Schadenssuche und -erkennung:

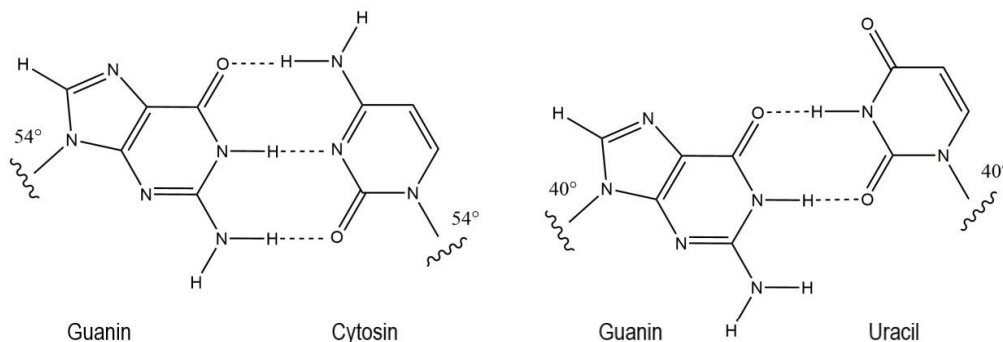


**Abbildung 11: Modell zur initialen Schadenserkennung durch Glykosylasen**

Das Schadenssuchmodell von Büchner *et al.* [1] aus der AG Tessmer zeigt am Beispiel der Glykosylase hTDG eine Möglichkeit der DNA-Läsionssuche und -erkennung durch einen Suchkomplex (SC), Abfragungskomplex (IC) und Exzisionskomplex (EC). Die Glykosylase sucht die DNA nach möglichen Basenschäden ab, die sich durch Stellen mit erhöhter intrinsischer Flexibilität darstellen [24]. Hierbei wechselt die Konformation stets zwischen einem SC- und einem IC-Zustand. Während die SC-Konformation die schadensfreie DNA aktiv krümmt (die schwarzen Pfeile deuten diese aktive Kräfteausübung an), sorgt das Wobble-Basenpaar (hier mit einem blauen Stern gekennzeichnet), das durch die instabile Paarung von G:T und G:U entsteht, für eine zusätzlich erhöhte Flexibilität im DNA-Gerüst. So kann sich dies für den SC-Komplex als ein erleichtertes oder gar passives Biegen (ohne Energieaufwand) bemerkbar machen und so zu einem längeren Verweilen am potenziellen Schadensort führen. Das längere Verweilen führt zu einer höheren Wahrscheinlichkeit, dass der Aufenthalt gleichzeitig mit einem Basenatmungsvorgang stattfindet, indem die Basen durch thermische Energie aus der DNA-Helix herausschwenken [114]. Durch das sog. „Phosphat-Kneifen“ (engl.: phosphate pinching) der Glykosylase wird die DNA-Struktur nochmals stärker gebogen, was wiederum das spontane Basenflippen begünstigt [66]. Verlässt die Base ihren ursprünglichen Platz in der DNA-Helix, wird dies im Falle von hTDG zusätzlich durch einen Arginin-Rest („Arginin-Finger“) stabilisiert, der den Freiraum einnimmt. Zuletzt wird in der EC-Konformation die Richtigkeit der gebundenen Base nochmals kontrolliert. Denn in der Bindungstasche der Glykosylasen kann die Basenexzision nur am spezifischen Schaden katalysiert werden.

modifiziert übernommen aus [1]

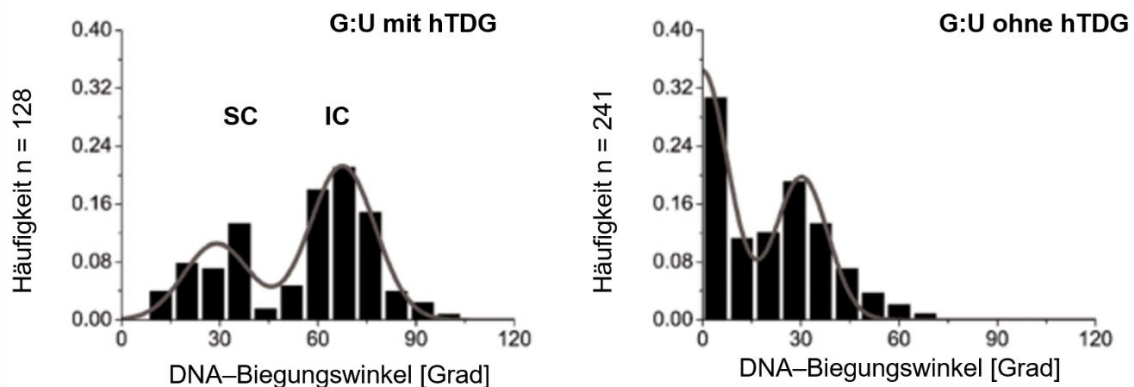
Aus der Beobachtung, dass die Glykosylase hTDG die gleichen Biegewinkel an geschädigter sowie ungeschädigter DNA induziert, und dass der kleinere Biegewinkel mit der Biegung am von hTDG erkannten Schaden (G:T oder G:U) übereinstimmt, ist das Modell einer dreigeteilten Schadenserkenkung, die auf schadensspezifische Konformationen oder Flexibilitäten beruht, durch die Glykosylasen entstanden (Abbildung 11). Entsprechend den verschiedenen Biegunszuständen werden die DNA-Protein-Komplexe in einen Suchkomplex (engl. search complex, SC) mit einem kleineren Biegewinkel und einen Abfragekomplex (engl. interrogation complex, IC) mit einem entsprechend größerem Biegewinkel eingeteilt. Die Basen-Exzision erfolgt in der Konformation des sog. „Exzisions-Komplexes“ (engl. excision complex, EC) [24]. Eine erhöhte Elastizität oder veränderte Konformation am DNA-Schaden führt zu einer passiven Biegung der DNA durch die Glykosylase an der Schadensstelle. Im Gegensatz hierzu ist bei der ungeschädigten DNA mit einer stabilen Watson-Crick-Basenpaarung für die Erzeugung einer Biegung die Aufwendung einer aktiven Kraft nötig. Im Suchkomplex (SC) gleitet die Glykosylase hTDG auf der Suche nach dem spezifischen Schaden stetig über die DNA-Stränge. Spezifische Schadenssubstrate von hTDG sind die mutagenen und cytotoxischen Guanin-Thymin und Guanin-Uracil Basenpaare (blauer Stern in Abbildung 11). Im Gegensatz zu einer Watson-Crick Basenpaarung mit Cytosin und drei Wasserstoffbrückenbindungen bildet Guanin mit Thymin nur zwei Wasserstoffbrückenbindungen aus, die zusätzlich in einer verzerrten DNA-Struktur resultieren, und ein sogenanntes Wobble-Basenpaar bilden [115-117] (Abbildung 12).



**Abbildung 12: Watson-Crick-Basenpaar G:C und Wobble-Basenpaar G:U**

In der Watson-Crick Basenpaarung existieren drei Wasserstoffbrücken zwischen Guanin und Cytosin. Guanin kann auch mit Uracil unter Ausbildung von nur zwei Wasserstoffbrückenbindungen eine Basenpaarung eingehen. Dies hat allerdings eine Verzerrung des DNA-Rückgrates zur Folge.  
 modifiziert übernommen aus [115]

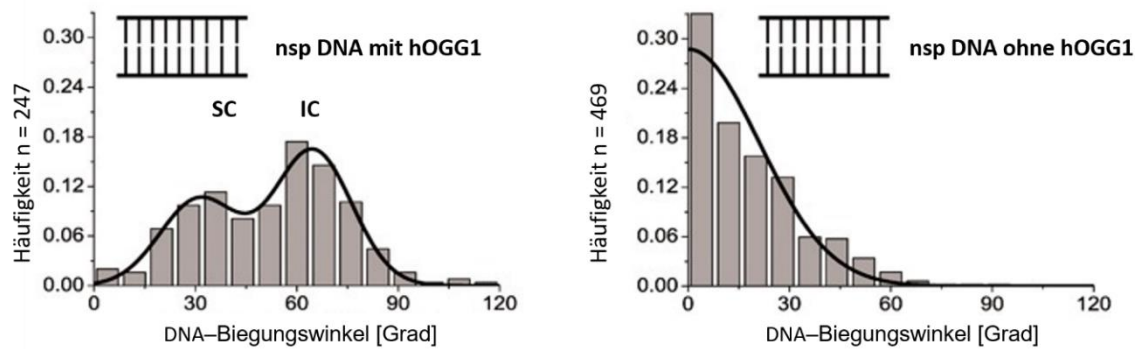
Die Instabilität im Gerüst der DNA-Doppelhelix erzeugte in rasterkraftmikroskopischen Aufnahmen einen DNA-Biegungswinkel von  $30^\circ \pm 8^\circ$  (Abbildung 13) [1]. Im Vergleich dazu ergab die Messung der Biegewinkelverteilung an ungeschädigter DNA einen Wert von  $0^\circ$  (Abbildung 14). Dies zeigte eindeutig, dass schadensfreie DNA prinzipiell keine Biegung im Gerüstaufbau aufwies [118-122]. Bei der Untersuchung der Wechselwirkungen der Glykosylase hTDG mit ihrem Schadenssubstrat konnten zwei Winkel erkannt werden (Abbildung 13). Es bestand ein Gleichgewicht zwischen einem Biegewinkel von  $29^\circ \pm 10^\circ$  und  $68^\circ \pm 10^\circ$ . Bei weiterführenden Versuchen zeigte die Inkubation von ungeschädigter DNA mit hTDG überraschender Weise auch eine Biegung von  $34^\circ \pm 13^\circ$  und  $65^\circ \pm 13^\circ$  (Abbildung 14). Die Zielstruktur ist G:U bzw. G:T (mit einem blauen Stern gekennzeichnet). Da der SC-Komplex beim Versuch die DNA zu biegen ständig aktiv eine zusätzliche Kraft ausübt, wird an ungeschädigter DNA, die ungebogen vorliegt, ein Biegungswinkel von ca.  $30^\circ$  induziert [1].



**Abbildung 13: DNA-Biegewinkel an G:U mit und ohne hTDG**

Die Abbildungen zeigen die Verteilung der DNA-Biegewinkel von hTDG an G:U-Fehlpaarungen mit einem Maximum bei  $29^\circ \pm 10^\circ$  und  $68^\circ \pm 10^\circ$  in rasterkraftmikroskopischen Aufnahmen. In Abwesenheit von Proteinen ergab die Verteilung der intrinsischen DNA-Biegewinkel an der G:U-Fehlpaarung einen sehr ähnlichen DNA-Biegungswinkel von  $30^\circ \pm 8^\circ$ . Zusätzlich lag eine gerade Spezies bei  $0^\circ \pm 8^\circ$  vor.

modifiziert übernommen aus [1]



**Abbildung 14: DNA-Biegewinkel von ungeschädigter DNA mit und ohne hTDG**

Gezeigt wird die Verteilung der DNA-Biegewinkel von hTDG an ungeschädigte DNA gebunden. Es finden sich Gauß-Maxima bei  $34^\circ \pm 13^\circ$  und  $65^\circ \pm 13^\circ$ . Der intrinsische DNA-Biegewinkel von ungeschädigter DNA beträgt  $0^\circ \pm 21^\circ$ .

modifiziert übernommen aus [1]

Trifft die Glykosylase während des Suchvorganges auf einen DNA-Schaden, so ist keine zusätzliche Kraft für die Biegung nötig. Aufgrund der veränderten mechanischen Eigenschaften kommt es zu einer passiven Biegung. Durch den geringeren Energieaufwand verbleibt die Glykosylase länger an dieser Position, was die Wahrscheinlichkeit erhöht, dass sie am Schaden gebunden ist, während dieser eine Basenatmung erfährt [1]. Unter Basenatmung wird das spontane Basen-Flippen, also Herausschwenken einzelner Basen aus der DNA-Doppelhelix durch thermische Energie verstanden [57, 61, 114]. Der Vorgang der Basenatmung kommt in ungeschädigter DNA regelmäßig in kleiner Frequenz vor, jedoch in geschädigter DNA deutlich häufiger [24]. Dieser spontan geflippte Zustand wird dann im Interrogationskomplex stabilisiert. In der IC-Konformation wird zusätzlich das „Phosphat-Kneifen“ (engl. „phosphate pinching“) genutzt, um den Energiebedarf des Basenflippens weiter herabzusetzen [1, 32, 53, 123, 124]. Durch das „Kneifen“ in das DNA-Rückgrat werden benachbarte Phosphatreste näher zueinander gezogen und so ein Flippen der Base nochmals erleichtert. Zur Stabilisierung der extrahelikalen Basen-Position fügt hTDG einen Arginin-Finger (Arginin 275) über die kleine Furche in den entstandenen Freiraum der Doppelhelix [20, 125]. Passt die rotierte Base korrekt in das katalytisch aktive Zentrum von hTDG, kann im Exzisions-Komplex (EC) die Spaltung der N-glykosidischen Bindung zwischen der Base und dem Zuckerring erfolgen [123, 125].



Von besonderem Interesse für die vorliegende Arbeit ist der Zusammenhang zwischen der SC-Konformation und dem hierbei erzeugten DNA-Biegewinkel in AFM-Abbildungen. Gemäß dem Modell von Büchner *et al.* [1] ist die Übereinstimmung des von der Glykosylase im SC-Komplex induzierten aktiven Biegewinkels mit der passiven Biegung des DNA-Stranges am Schadensort der Schlüssel zur effektiven Schadenserkenkung. Ansonsten wäre es für die Glykosylasen weitgehend unmöglich die komplette DNA in der kurzen Zeit zwischen zwei Replikationsprozessen durch das Flippen einer jeden Base in die Bindungstasche nach speziellen Basenschäden abzusuchen [78]. Der in der AFM-Abbildung dargestellte Biegewinkel muss nicht zwingend die wirkliche Konformation reflektieren. Dieser kann auch auf eine veränderte Flexibilität am Schadensort hinweisen. Grundsätzlich ist aber eine Korrelation zwischen der spezifischen Konformation oder Elastizität am Zielschaden und der benötigten Energie des Basenflippens notwendig.

Die Kernaussage des Modells ist folgende: Die natürliche Biegung schadenstragender DNA muss mit der aktiven Biegung durch die spezifischen Glykosylasen, die diese am Schadensort im SC-Protein-DNA-Komplex induziert, übereinstimmen. Dies führt zu einer energetisch günstigeren Konstellation, was einen verlängerten Aufenthalt am Schadensort und somit einer erhöhten Wahrscheinlichkeit für eine Schadensexzision bedingt.

Dieses Modell beruht auf Untersuchungen zu den Glykosylasen hTDG und hOGG1, und soll nun auf die Allgemeingültigkeit für weitere DNA-Glykosylasen untersucht werden, um so möglichst die grundsätzliche Schadenserkenkung aller Glykosylasen im wichtigen BER-Prozess zu verstehen. In der Masterarbeit von Lisa Hirsch (2017) aus der AG Tessmer wurden bereits die Biegungswinkel von weiteren DNA-Schäden ( $\epsilon$ A:T, oxoG:A, oxoG:C, G:A) im Vergleich zu ungeschädigter DNA (nsp DNA) untersucht.

## 1.6 Zielsetzung und Fragestellung

Um in der enormen Menge an ungeschädigter DNA einen Schaden zu finden, ist eine möglichst schnelle und effektive, aber auch energiesparende Methode nötig [24]. Für diesen Vorgang wurde von Büchner *et al.* [1] ein Modell für die Vorerkennung von Schäden durch passende mechanische oder konformationsbezogene Eigenschaften der DNA am Schaden postuliert. Das Augenmerk liegt auf der Kongruenz des durch den DNA-Schaden verursachten statischen Biegewinkels in AFM-Abbildungen und dem Biegewinkel im SC-Komplex der entsprechenden Glykosylase mit (ungeschädigter oder geschädigter) DNA. Die initiale Schadenserkennung erfolgt demnach durch eine Verminderung der für die DNA-Biegung aufzubringenden Kraft (passives Biegen der schadhaften DNA mit geringem Energieaufwand *versus* aktives Biegen ungeschädigter DNA). Die vorliegende Arbeit soll an die Ergebnisse vorausgegangener Arbeiten anknüpfen, wobei zur Objektivierung der Winkelmessung ein auf MATLAB-basierendes Programm etabliert wurde. Mit diesem wurde die Biegung der DNA-Schäden  $\epsilon$ A:T, oxoG:C, oxoG:A, G:A und ungeschädigter DNA erneut bestimmt. Aufbauend auf den Ergebnissen für die Glykosylasen hTDG und hOGG1 wurden die Untersuchungen in dieser Arbeit auf die DNA-Glykosylase MutY erweitert. Um eine mögliche allgemeine Gültigkeit zu eruieren, soll in dieser Arbeit zunächst der Einfluss von verschiedenen Basenschäden auf die DNA-Biegung untersucht werden, um sie anschließend mit dem gemessenen Winkel des Protein-DNA-Komplexes zu vergleichen. Die Arbeit soll damit die Modellvorstellung der Biegewinkel-Kongruenz in der SC-Konformation für eine weitere Glykosylase bestätigen. Ziel ist die Entschlüsselung des Einflusses von schadensbedingten DNA-Biegungen oder Elastizitätsveränderungen für die Erkennung des Schadenssubstrates durch DNA-Glykosylasen. Zusammenfassend sollte in dieser Arbeit geklärt werden, inwieweit spezifische DNA-Biegewinkel an BER-Schäden in AFM-Aufnahmen den spezifischen DNA-Biegewinkeln des jeweiligen Glykosylase-DNA-Komplexes entsprechen und ob das vorgestellte Modell auf weitere DNA-Schäden übertragbar ist. Insbesondere soll zur Objektivierung der Winkelmessung eine computer-gestützte automatisierte Messmethode eingeführt und evaluiert werden.

## 2 Material und Methoden

### 2.1 Material

#### 2.1.1 Geräte und Verbrauchsmaterialien

Bezeichnung	Hersteller
AFM: Molecular Force Probe MFP-3D™	Oxford Instruments Asylum Research, Inc. (Santa Barbara, Kalifornien, USA)
Autoklav	Systec GmbH (Linden, Hessen, Deutschland)
Cantilever (Micro Cantilever) OMCL-AC240TSG	Olympus Europa SE & Co. KG (Hamburg, Deutschland)
Filterpapier: Rotilabo Rundfilter Typ 111A	Carl Roth GmbH & Co. KG (Karlsruhe, Baden-Württemberg, Deutschland)
Gelelektrophorese-Kammer/-Apparatur für Agarose-Gelelektrophorese; PowerPac™ Basic Power Supply	BioRad-Laboratories, Inc. (Hercules, Kalifornien, USA)
Gelgieß-Apparatur für Agarosegele	BioRad-Laboratories, Inc. (Hercules, Kalifornien, USA)
Glimmer (Mica Sheets)	Spi Supplies (West Chester, Pennsylvania, USA)
Handschuhe: StarGuard Comfort Large	Starlab International GmbH (Hamburg, Deutschland)
8er-Kamm, 15er-Kamm für Agarose-Gelelektrophorese	BioRad-Laboratories, Inc. (Hercules, Kalifornien, USA)
Mikrowelle	Siemens; SEG Hausgeräte GmbH (München, Bayern, Deutschland)
Nanodrop Spectrophotometer ND-1000	PeqLab Biotechnologie GmbH (Erlangen, Bayern, Deutschland)
Objektträger	Paul Marienfeld GmbH & Co. KG (Lauda-Königshofen, Baden-Württemberg, Deutschland)
Petrischalen	Greiner Bio-One International GmbH (Kremsmünster, Österreich)

## 2 Material und Methoden

Bezeichnung	Hersteller
Pipetten und Spitzen	Mettler-Toledo Ramm, LCC (Oakland, Kalifornien, USA) Eppendorf AG (Hamburg, Deutschland)
Tesafilm	Tesa SE (Norderstedt, Schleswig-Holstein, Deutschland)
Tischzentrifuge Eppendorf 5415D	Eppendorf AG (Hamburg, Deutschland)
Plastikrohre/Tubes 2 mL; 1,5 mL; 0,5 mL	Sarstedt AG & Co. (Nümbrecht, Nordrhein- Westfalen, Deutschland)
Vortexmischer: VortexGenie2	Scientific Industries, Inc. (Bohemia, New York, USA)
Wasseraufbereitungssystem TKA X- CAD	TKA Wasseraufbereitungssysteme GmbH (Niederelbert, Rheinland-Pfalz, Deutsch- land)

### 2.1.2 Software

Bezeichnung	Hersteller
ChemDraw Prof. 17.0	PerkinElmer, Inc. (Waltham, USA)
FIESTA 1.05.0005	B CUBE Center of Innovation Competence (Dresden, Deutschland) [126]
GIMP 2	GIMP TEAM (Cold Spring, NY, USA)
MATLAB 2014	MathWorks, Inc. (Natick, Massachusetts, USA)
IgorPro 6.20	WaveMetrics, Inc. (Portland, Oregon, USA)
ImageJ 1.50e	Open Source Java Processing Image Program (Wayne Rasband, National Institutes of Health, USA)
MFP-3D	Oxford Instruments Asylum Research, Inc. (Santa Bar- bara, Kalifornien, USA)
Office Professional 2010	Microsoft Corporation (Redmond, USA)
Origin Pro 2017 G 32bit	OriginLab Corporation (Northampton, Massachusetts, USA)
Nanodrop ND-1000 V3.8.1	PeqLab Biotechnologie GmbH (Erlangen, Bayern, Deutschland)
GeoGebra 5.0.573.0	IGI (Linz, Schweiz) [127]

### 2.1.3 Internetbasierte Anwendungsprogramme

Bezeichnung	Internetadresse
Expasy Bioinformatics Resource Portal	<a href="https://www.expasy.org/">https://www.expasy.org/</a>
NEB Cutter V2.0 Tool	<a href="http://nc2.neb.com/NEBcutter2/">http://nc2.neb.com/NEBcutter2/</a>
Uniprot	<a href="http://www.uniprot.org/">http://www.uniprot.org/</a>

### 2.1.4 Komplettsysteme (Kits)

Bezeichnung	Hersteller
NucleoSpin Extract II Kit	Macherey-Nagel GmbH & Co. KG (Düren, Nordrhein-Westfalen, Deutschland)

### 2.1.5 Chemikalien

Bezeichnung	Hersteller
Agarose NEEQ Ultra-Qualität	Carl Roth GmbH & Co. KG (Karlsruhe, Baden-Württemberg, Deutschland)
Ethanol	Carl Roth GmbH & Co. KG (Karlsruhe, Baden-Württemberg, Deutschland)
HEPES/HCl	Carl Roth GmbH & Co. KG (Karlsruhe, Baden-Württemberg, Deutschland)
Magnesium-Acetat	Carl Roth GmbH & Co. KG (Karlsruhe, Baden-Württemberg, Deutschland)
Natrium-Acetat	Carl Roth GmbH & Co. KG (Karlsruhe, Baden-Württemberg, Deutschland)
Natriumchlorid	Carl Roth GmbH & Co. KG (Karlsruhe, Baden-Württemberg, Deutschland)

### 2.1.6 Proteine

Bezeichnung	Konzentration	Hersteller
BsaXI	2.000 Units/mL	NEB, Inc. (Ipswich, Massachusetts, USA)
NdeI	20.000 Units/mL	NEB, Inc. (Ipswich, Massachusetts, USA)
Nt.BstNBI	10.000 Units/mL	NEB, Inc. (Ipswich, Massachusetts, USA)
T4 DNA-Ligase	400.000 Units/mL	NEB, Inc. (Ipswich, Massachusetts, USA)
MutY	5.000 Units/mL 1000nM	Bio-Techne (Minneapolis, Minnesota, USA)
hAAG	10,000 Units/ml 800nM	NEB, Inc. (Ipswich, Massachusetts, USA)
UDG	5.000 Units/mL 1900nM	NEB, Inc. (Ipswich, Massachusetts, USA)

### 2.1.7 Puffer

Bezeichnung	Zusammensetzung
AFM-Depositionspuffer	25 mM HEPES/HCl pH 7,5 bei 25 °C, 25 mM Natrium-Acetat, 10 mM Magnesium-Acetat

### 2.1.8 Oligonukleotide

Bezeichnung	Beschreibung	Hersteller
Bottom-Oligo	5'-GGT CGA CTC TAG AGG ATC AGA TCT GGT ACC TCTAGACTC GAG GCA TGC-3'; 100 µM	Sigma-Aldrich Corporation (St. Louis, Missouri, USA)
εA(A)-Oligo	5'-GCA TGC CTC G(εA)G TCT AGA GGT ACC AGA TCTGAT CCT CTA GAG TCGACC-3'; 5'-Ende phosphoryliert; 100 µM	Gene Link, Inc. (Orlando, Florida, USA)
G:A-Oligo	5'-GCA TGC C(C)GAG TCTAGA GGT ACC AGA TCT GATCCT CTA GAG TCG ACC-3'; 5'-Ende phosphoryliert; 100 µM	Sigma-Aldrich Corporation (St. Louis, Missouri, USA)
oxoG:A-Oligo	5'-GCA TGC C(8-oxoG)C GAGTCT AGA GGT ACC AGA TCTGAT CCT CTA GAG TCG ACC-3'; 5'-Ende phosphoryliert; 100 µM	Gene Link, Inc. (Orlando, Florida, USA)
oxoG:C-Oligo	5'-GCA TGC CTC (8-oxoG)AGTCT AGA GGT ACC AGA TCTGAT CCT CTA GAG TCG ACC-3'; 5'-Ende phosphoryliert; 100 µM	Midland Certified Reagent Co (Midland, Texas, USA)
pUC19N-Plasmid	2729 bp	S. Wilson, National Institute of Environmental Health Sciences (Durham, North Carolina, USA)

## 2.2 Methoden

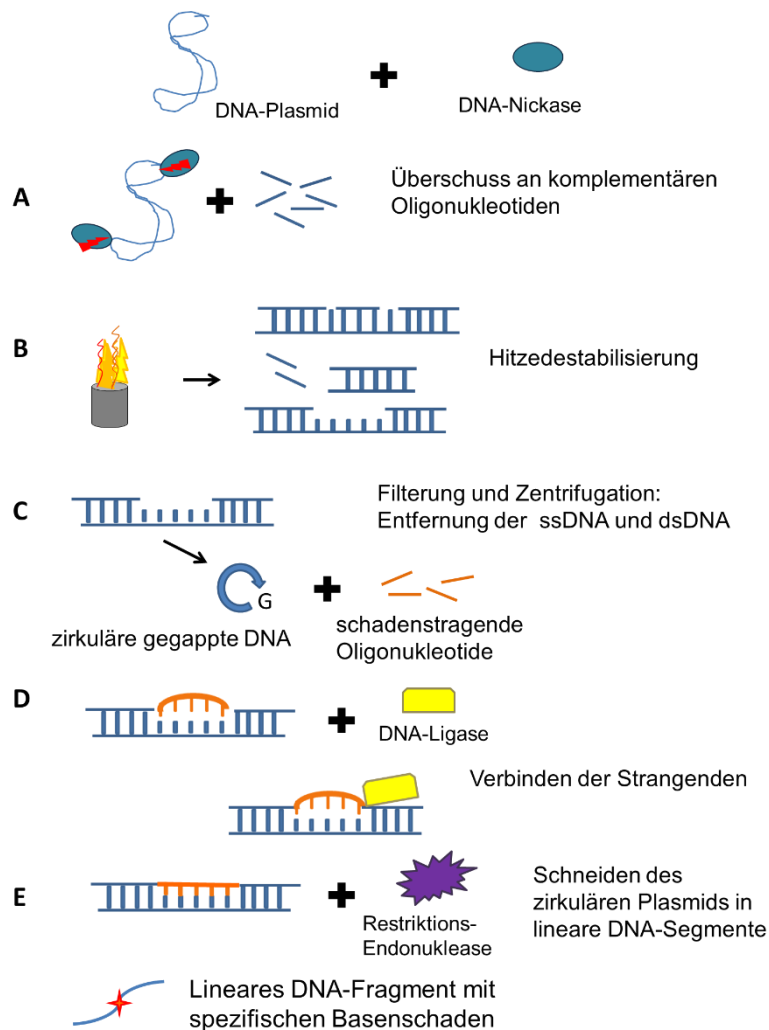
### 2.2.1 DNA-Präparation für die Rasterkraftmikroskopie

Die mit dem Rasterkraftmikroskop aufgenommenen DNA-Fragmente wurden von Lisa Hirsch im Rahmen ihrer Masterarbeit hergestellt. Dabei wandte sie das von der Arbeitsgruppe Tessmer vorgestellte Verfahren zur Präparation von linearem doppelsträngigen DNA-Substrat aus dem pUC19N Plasmid an [128]. Der Vektor pUC19N [1, 128, 129] besteht aus zirkulärer doppelsträngiger DNA und wurde zuvor in *Escherichia coli* transformiert. Das Plasmid pUC19N hat eine Länge von 2729 bp, wobei sich eng beieinanderliegend drei Stellen (422, 440, 470) befinden, die von einem modifizierten Restriktionsenzym, der Nickase Nt.BstNBI, geschnitten werden können. Auf diese Weise entstanden an den genannten Stellen Einzelstrangbrüche (engl. nicks), wobei es im nächsten Schritt durch Erhitzen zu einem Ausschmelzen der ssDNA-Fragments zwischen den Schnittstellen mit einer Gesamtlänge von 48 Nukleotiden (engl. nucleotides, nt) kam. Die herausgelösten Fragmente wurden durch Zugabe von entsprechenden komplementären (Bottom)-Oligonukleotiden im Abkühlprozess zu kurzen dsDNA-Stücken verbunden (engl. annealing). Sowohl die überschüssigen ssDNA-Stücke, als auch die entstandenen dsDNA-Stücke wurden abfiltriert. Übrig blieb die zirkuläre DNA mit der entstandenen Einzelstrang-Lücke (engl. gap). Die erzeugte Lücke wurde nun dazu genutzt, um durch Zugabe von modifizierter, zu der Lücke komplementärer, ssDNA an der gewünschten Stelle einen Schaden zu generieren. Eine Auflistung aller verwendeter Schadens-Oligonukleotide ist in Kapitel 2.1 (Tabelle 2.1.8) zu finden. Das überschüssig zugefügte Schadens-Oligo-Nukleotid wurde durch Inkubation mit der T4 DNA-Ligase in die zirkuläre Plasmid-DNA eingebunden. Im Anschluss erfolgte noch die Linearisierung der zirkulären Plasmid-DNA durch Inkubation mit den zwei Restriktionsenzymen NdeI und BsaXI. Diese Enzyme generierten zwei verschieden lange DNA-Fragmente, nämlich ein 505 bp langes, schadentragendes Fragment, das den eingefügten Schaden bei 50 % seiner Länge enthält, ebenso wie ein 2224 bp langes ungeschädigtes Fragment. Am Ende dieses Kapitels befindet sich zum besseren Verständnis ein Schema der DNA-Präparation (Abbildung 15). Die verschiedenen DNA-Fragmente wurden



mittels einer Agarose-Gelelektrophorese getrennt, und im Anschluss daran mit dem NucleoSpin-Kit aufgereinigt. Zur Konzentrationsbestimmung der DNA-Substrate wurde das NanoDrop™ Spektralphotometer genutzt. Kontrollassays während des Herstellungsprozesses der DNA-Substrate gewährleisteten die korrekte Schadensintegration und Ligation [128].

Alle in der vorliegenden Arbeit verwendeten Schadenssubstrate wurden gemäß dem oben genannten Präparationsverfahren hergestellt. Die Länge der untersuchten DNA-Stränge betrug durch die Verwendung der beiden oben erläuterten Restriktionsenzyme immer 505 bp, was einer Länge von ca. 172 nm entsprach. Der Schaden befand sich durch die Wahl der entsprechenden Restriktionsenzyme jeweils bei 50 % der DNA-Länge. Mit dem Rasterkraftmikroskop wurden ungeschädigte DNA und DNA-Stränge mit den Schäden oxoG:A, G:A, oxoG:C und  $\epsilon$ A:T (Tabelle 1) untersucht.



**Abbildung 15: Schematische Darstellung der DNA-Substratpräparation**

Ziel ist die Herstellung interner DNA-Modifikationen durch die Verwendung einer sog. Nicking-Endonuklease. **A** Die Plasmid-DNA wird durch Inkubation mit einem Nicking-Restriktionsenzym an einem der beiden DNA-Stränge geschnitten. **B** Es wird ein Überschuss an DNA-Oligonukleotiden zugegeben, die komplementär zu der kurzen von der Nickase geschnittenen ssDNA-Strecke sein müssen. Durch Erhitzen (engl. annealing) der Probe wird das kurze ssDNA-Stück aus dem ursprünglichen DNA-Molekül herausgeschmolzen, und während des Abkühlvorganges verbindet sich das herausgelöste ssDNA-Fragment vorzugsweise mit dem im Überschuss zugegebenen komplementären Oligonukleotid. Die entstanden dsDNA-Oligonukleotide als auch die überschüssigen ssDNA-Oligonukleotide werden im folgenden Filtrationsschritt entfernt, sodass nur noch die lückenhaltige zirkuläre DNA verbleibt. **C** Um spezifische Zielstrukturen einzubauen, ist die Insertion eines modifizierten Oligonukleotids in die Lückenregion notwendig. Das im Überschuss zugegebenen Oligonukleotid bindet mit der entsprechenden komplementären Lücke in der DNA. **D** Um den Einbau zu vervollständigen müssen auch die Bindungen am DNA-Rückgrat wiederhergestellt werden. Durch die Inkubation mit einer DNA-Ligase wird das neu inserierte DNA-Oligonukleotid mit der ursprünglichen DNA verbunden. **E** Im letzten Schritt werden durch den Verdau mit den entsprechenden Restriktionsenzymen DNA-Fragmente mit der gewünschten Länge erzeugt. So entsteht aus einem langen ringförmigen DNA-Substrat ein 505 bp langes doppelsträngiges DNA-Stück, dass bei 50 % der Länge den spezifischen Basenschaden enthält.

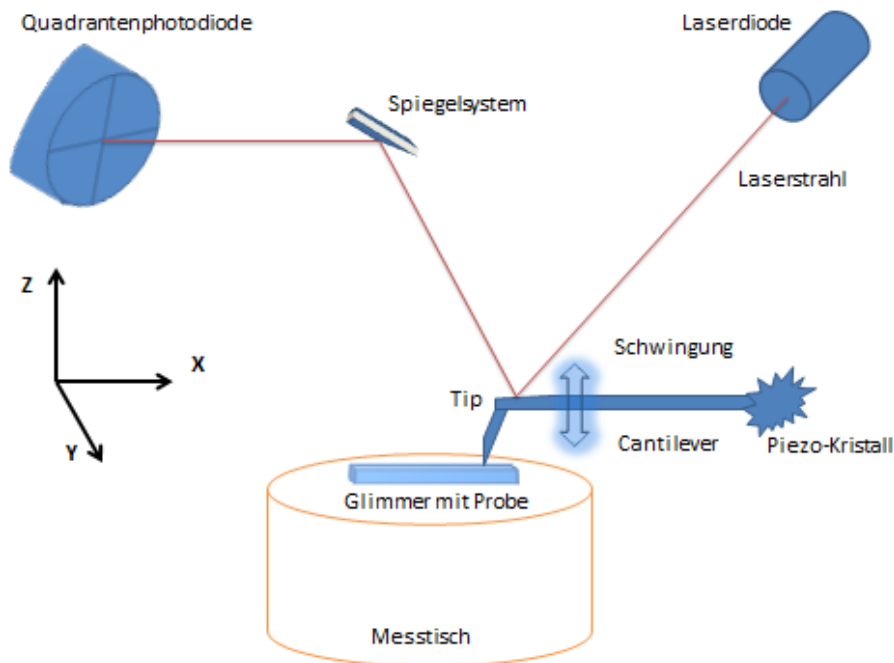
modifiziert übernommen aus [128]

## 2.2.2 Rasterkraftmikroskopie

### 2.2.2.1 Funktionsweise der Rasterkraftmikroskopie

Die Rasterkraftmikroskopie (engl. atomic force microscopy, AFM) ist eine spezielle Form der Rastersondenmikroskopie (engl. scanning probe microscopy, SPM), wobei während eines Scan-Vorganges (engl. scan = abrastern) die wechselwirkenden Kräfte zwischen einer Sonde und der Oberfläche der zu untersuchenden Probe gemessen werden [130]. Das Rasterkraftmikroskop wurde 1986 von Gerd Binnig, Calvin Quate und Christoph Gerber [131] als Weiterentwicklung des Rastertunnelmikroskops [132] vorgestellt. Während des Rastervorganges wird eine nanoskalige Messsonde (engl. tip), welche an einem schwingungsfähigen Hebelarm (engl. cantilever) befestigt ist, rasterartig in Lesezeilen über die Probe geführt [133]. Auf atomarer Basis kommt es zwischen dem Tip und der Probenoberfläche zu Wechselwirkungen, wie den Van-der-Waals-Kräften, elektrostatischen Kräften und Pauli-Repulsion entsprechend des Lennard-Jones-Potentials, welche zu einer Auslenkung des Hebelarms führen, die mit Hilfe eines optischen Systems detektiert werden können (Abbildung 16) [134]. Ein Laserstrahl wird auf den Cantilever zentriert und weiter über Reflexion an diesem mit einer Quadrantenphotodiode detektiert [135]. Die unterschiedliche Oberflächenbeschaffenheit führt zur Auslenkung des Cantilevers und wird durch die daraus resultierende Reflexion des Laserstrahls vom Auswertungsprogramm als Höhe Z punktweise gemessen [136]. Beim zeilenweisen Ablesen in X- und Y-Richtung entsteht so ein dreidimensionales Bild von der Proben-Topographie. Die Betriebsarten der abbildenden Rasterkraftmikroskopie lassen sich in zwei Hauptgruppen, nämlich die statischen und dynamischen Modi aufteilen, wobei sich die dynamische Aufnahmeart zusätzlich in eine Betriebsart mit sog. Amplitudenmodulation (AM) oder Frequenzmodulation (FM) trennen lässt [134]. Im statischen Modus ruht der Tip und gibt durch seine wechselnde Verbiegung während des Scanvorganges die Oberflächeneigenschaften der Probe wider [134]. Im dynamischen Modus wird dagegen ein Piezo-Kristall genutzt, der den Tip zu einer periodischen Oszillation anregt. Insgesamt haben sich drei geläufige Abbildungstechniken etabliert: der Kontakt-Modus mit konstanter Höhe oder alternativ mit konstanter Kraft (statischer Modus), der intermittierende Modus, welcher häufig

auch als repulsiver Modus oder Tapping-Modus [137] bezeichnet wird, und der Nicht-Kontakt-Modus beziehungsweise attraktive Modus. Während im Kontakt-Modus die Messsonde stets statisch die Oberfläche berührt, oszilliert der Tip in den anderen beiden Modi und hat im Falle des intermittierenden Modus nur am Amplitudenmaximum leichten Probenkontakt, im Nicht-Kontakt-Modus hingegen hat er überhaupt keinen Kontakt zur Oberfläche. Zur hochauflösenden Aufnahme von DNA und DNA-Protein-Komplexen eignet sich besonders der repulsive Modus, weshalb alle Aufnahmen im Rahmen dieser Arbeit ausschließlich auf diese Weise angefertigt wurden.



**Abbildung 16: Schematischer Aufbau eines Rasterkraftmikroskops**

Die Rasterkraftmikroskopie dient zur Messung der wechselwirkenden Kräfte zwischen Probe und der Messsondenspitze (Tip) auf atomarer Basis. Auf dem schwingungsgedämpften Messtisch befindet sich ein ultraglatte Glimmer mit enthaltener Messprobe. Der durch den Piezo-Kristall in Schwingung versetzte Cantilever detektiert zeilenweise durch Abstoßung und Anziehung zwischen den Atomen der Probe und des nanoskaligen Tips die unterschiedliche Höhengeometrie der Probe. Um die Informationen überhaupt messen zu können, wird ein von einer Laserdiode generierter Laserstrahl genutzt, und dieser wird dabei möglichst auf die Spitze des Cantilevers ausgerichtet, wo sich auch der Tip befindet. In der Ausgangslage wird der Laserstrahl mit einem Spiegelsystem auf das Zentrum der Quadrantenphotodiode gelenkt, da der Cantilever nun entsprechend der Höhentopographie ausgelenkt wird, wird folglich auch der Laserstrahl in unterschiedliche Richtungen reflektiert. Auf diese Weise kann die Information über die Anordnung der verschiedenen Photodioden in elektrische Signale übersetzt und von der Mikroskopsoftware verarbeitet werden, um mit den Informationen aus der X-, Y- und Z-Position des Tips ein dreidimensionales Bild zu erzeugen.

### 2.2.2.2 AFM-Probenherstellung und -Messung

Zur optimalen Darstellung von DNA und DNA-Protein-Komplexen mit dem Rasterkraftmikroskop bestehen hohe Anforderungen an die Trägersubstanz. Diese muss hierfür sehr flach, glatt und vor allem frei von jeglichen Verunreinigungen sein. Geläufig sind hierbei die Verwendung von Silizium oder Glimmer, da beide eine ähnlich geringe Oberflächenrauigkeit aufweisen (Glimmer 0.05 nm (RMS) [138]; Silizium <0.1 nm (RMS) [139]). Allerdings hat sich das Glimmer-Schichtsilikat besonders wegen der einfachen Präparation mit Hilfe eines Tesafilm<sup>®</sup> zum Erhalt einer glatten, flachen und sauberen Oberfläche bewährt. Das Vorgehen war wie folgt: Mit Hilfe eines Skalpell wurde das Ausgangsmaterial auf einer festen Unterlage in ca. 1 cm<sup>2</sup> große Glimmer-Träger zerschnitten, ein Brechen war zu vermeiden, da es den gleichmäßigen Schichtaufbau hätte nachhaltig beschädigen können. Um eine möglichst atomar glatte Fläche zu generieren, sollte dann die obere Glimmerschicht mit einem zuvor aufgeklebten Tesafilm<sup>®</sup> durch zügiges Abziehen entfernt werden. Auf dem Tesafilm<sup>®</sup> sollte eine dünne, möglichst einschichtige Glimmerschicht haften bleiben. Diese wurde optisch auf Brüche und Rissbildung in der Oberfläche untersucht. Sobald eine makroskopisch einwandfreie Oberfläche erzeugt wurde, konnte der Glimmer für die Probendeposition genutzt werden. Um eine dauerhafte Fixation der DNA auf dem Glimmer-Substrat zu erreichen, wurden die Proben in einem Magnesiumionen-haltigen Puffer (AFM-Puffer, 25 mM HEPES/HCL, pH 7,5 bei 25 °C, 25 mM Natrium-Acetat, 10 mM Magnesium-Acetat) auf das Glimmer aufgebracht. Das zweiwertige Magnesiumkation dient als Verbindung zwischen der negativ geladenen Glimmeroberfläche und der ebenfalls negativ geladenen DNA [140].

Die Deposition wurde standardisiert in AFM-Pufferlösung auf ein Volumen von 20 µl gebracht und auf das vorbereitete Glimmersubstrat aufgetragen. Ziel war stets eine optimale Bedeckung der Glimmeroberfläche mit der DNA-Probe. Dies bedeutet, dass die Oberfläche eine ausreichend große Dichte der zu untersuchenden Substanz besitzen sollte, da so mit einem Minimum an Aufnahmen ausreichend viele Messdaten erzielt werden konnten. Der Abstand zwischen den einzelnen Strukturen sollte aber dennoch so groß sein, dass diese noch eindeutig als einzelne Strukturen erkennbar waren und ein Überlappen möglichst

vermieden wurde. In der Auswertung wurden sich überlappenden DNA-Stränge ausgeschlossen. Mit der Faustregel: 1nM/1000 bp in der Depositionslösung konnte begonnen werden, um sich dann schrittweise an eine für das jeweilige Substrat optimale Depositionskonzentration heranzutasten. Ungeschädigte DNA mit einer Länge von 2224 bp wurde bei einer Konzentration von 1 nM deponiert, spezifische DNA-Substrate mit einer Länge von 505 bp bei 4 nM (Tabelle 1). Bei der alleinigen Abbildung von Proteinen wurden abhängig vom Protein (UDG, hAAG, MutY) 2 nM bis 40 nM deponiert (Tabelle 2). Sollten die Protein-DNA-Interaktionen untersucht werden, so wurden die Glykosylasen mit dem jeweiligen DNA-Substrat für jeweils 30 Minuten bei Raumtemperatur in AFM-Puffer inkubiert. So wurde ausreichend Zeit für die Einstellung eines chemischen Gleichgewichtes gegeben. Gemäß der jeweiligen Bindungsaffinität ( $K_D$ ) des Proteins zu seinem Schadenssubstrat oder zu ungeschädigter DNA musste die Depositionskonzentration angepasst werden (Tabelle 3).

**Tabelle 1: DNA-Substrate**

DNA-Substrat	Länge (bp)	Depositionskonzentration
nsp DNA	2224 bp	1 nM
nsp DNA, oxoG:A, G:A, εA:T, oxoG:C,	505 bp	4 nM

**Tabelle 2: Proteine**

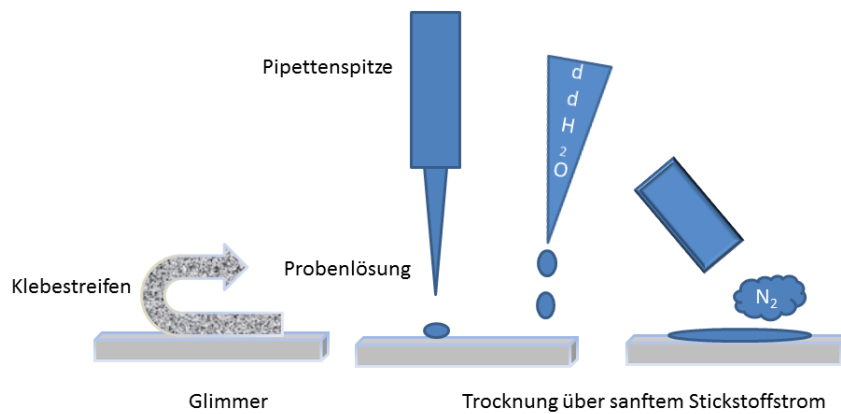
Protein-Substrat	Molekulare Masse (kDa)	Depositionskonzentration
UDG	25,66	10 nM
hAAG	32,18	2 nM
MutY	39,15	37,5 nM

**Tabelle 3: Protein-DNA-Substrate**

Protein-DNA-Substrat	Bindungskoeffizient ( $K_D$ )	Depositionskonzentration
MutY-nspDNA (505 bp)	150 nM ds nspDNA [141]	50 nM – 4 nM
MutY-nspDNA (2224 bp)	150 nM ds nspDNA [141]	50 nM – 1 nM

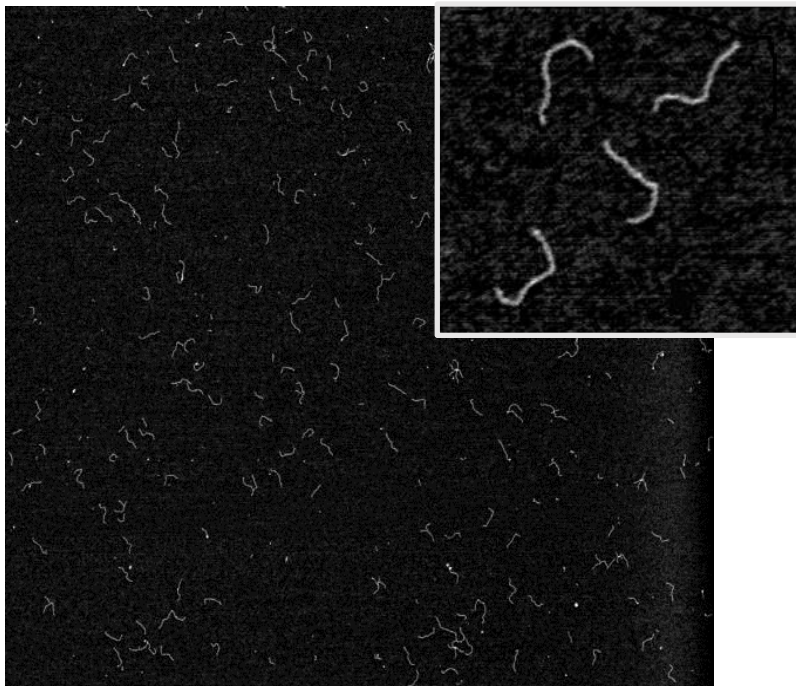
Unmittelbar vor der Inkubation wurde das Probenröhrchen in eine Tischzentrifuge gesetzt, diese ließ man kurz anlaufen. Dadurch sollten entstandene Tropfen vom Rand des Röhrchens in die Lösung befördert und eine optimale Durchmischung gewährleistet werden. Nach der Deposition von 20 µl Probelösung wurde der Glimmer unverzüglich mit deionisiertem, gefiltertem Wasser (ddH<sub>2</sub>O) fünfmalig abgewaschen. Anschließend wurde überflüssiges Wasser vom Randbereich und der Unterseite mit Löschpapier aufgenommen. Schlussendlich erfolgte die Trocknung der Ober- und Unterseite im Kreuzgang mit einem leichten Stickstoffgasstrom. Die Probendeposition ist in Abbildung 17 dargestellt. Mit Tesafilm® auf einem Objektträger befestigt, wurde die Probe im AFM positioniert.

Im Laser-Ausrichtungsschritt musste der Laserstrahl korrekt auf das Zentrum der Quadrantenphotodiode eingestellt werden. Nachdem die Ablenkung (engl. deflection) des Laserstrahls einen leicht negativen Wert angenommen hatte, brachte man den Tip mit „Engage“ (engl. für in Kontakt bringen) mit der Probe in Kontakt. Für Messungen wurden AFM-Messsonden mit einer Resonanzfrequenz von ~70 kHz und einer Federkonstante von ~2 N/m verwendet. Um im repulsiven Modus zu arbeiten, betrug die Oszillationsamplitude optimalerweise ein Volt. Für Orientierungsaufnahmen wurde eine Scan-Geschwindigkeit von 5 µm/s bei einer Bildgröße von 2 µm x 2 µm und 512 x 512 Pixel (Auflösung 3,91 nm/Pixel) eingestellt. Für die Datenaufnahme betrug die Scan-Geschwindigkeit 2,5 µm/s. Die Bildgrößen variierten zwischen 1 µm x 1 µm mit 512 x 512 Pixel, 2 µm x 2 µm mit 1024 x 1024 Pixel, 4 µm x 4 µm mit 2048 x 2048 Pixel, 6 µm x 6 µm mit 3072 x 3072 Pixel und 8 µm x 8 µm bei 4096 x 4096 Pixel. Dies ergab für die Datengewinnung in allen Formaten stets eine Pixelauflösung von 1,95 nm/Pixel. Beispielfähig wird in Abbildung 18 eine rasterkraftmikroskopische Aufnahme von oxoG:A mit einer Größe von 8 µm x 8 µm dargestellt.



**Abbildung 17: Arbeitsschritte der Probenherstellung**

Für gute Aufnahmebedingungen war eine äußerst glatte und völlig schmutzfreie Glimmer-Oberfläche obligat. Indem man einen Tesafilm® zügig und gleichmäßig abzog, entfernte man möglichst nur eine einzelne Glimmerschicht und hatte so eine perfekte Oberflächenqualität geschaffen. Nachdem man die Probe mittels Pipette applizierte, wurde das Glimmer mit Reinstwasser gewaschen und unter einem sanften Stickstoff-Gasstrom getrocknet.



**Abbildung 18: Rasterkraftmikroskopische Aufnahme**

Dargestellt ist exemplarisch eine AFM-Aufnahme von DNA mit dem oxoG:A-Schaden. Der Schaden befindet sich bei 50 % des 505 bp langen DNA-Stranges. Zur detaillierten Darstellung befindet sich am rechten Bildrand zusätzlich ein vergrößerter Bildausschnitt. Bildgröße:  $8 \mu m \times 8 \mu m$

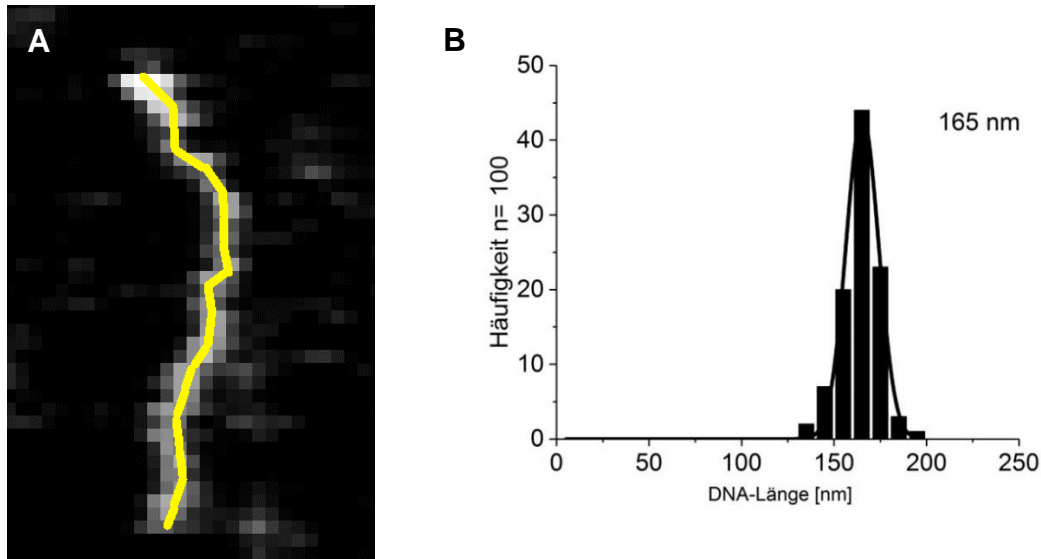


### 2.2.2.3 Bildprozessierung

Alle Aufnahmen wurden im Dateiformat [.ibw] mit dem Modify-Tool der Software MFP-3D/IgorPro (6.20) bearbeitet. Aufnahmebedingte Hintergrundunebenheiten wurden mit der Funktion "plane fit dritter Ordnung XY" minimiert. Mit der Funktion "flatten dritter Ordnung" wurden die Höhen der einzelnen Scanlinien abgeglichen. Mit „Draw Lines“ und „Erase“ konnten einzelne Rausch-Linien beseitigt werden. Mit der Funktion "Exclude Points" konnten beim Flatten-Vorgang ausgewählte größere Probenverschmutzungen von diesem Abgleichvorgang ausgenommen werden, was zu einem homogeneren Bildhintergrund nach der Flatten-Applikation führte. Abschließend wurden die prozessierten Bilder für die folgende Auswertung als [.tiff]-Datei mit einer Pixelauflösung von 1,95 nm/Pixel exportiert.

### 2.2.3 Manuelle DNA-Längencharakterisierung

Das Bildbearbeitungsprogramm ImageJ wurde zur Vermessung der abgebildeten DNA-Strukturen genutzt. Zu Beginn wurde die richtige Skalierung entsprechend der Größe der AFM-Aufnahme eingestellt. Mit der „Draw“-Funktion erfolgte die Nummerierung einzelner DNA-Strukturen, um im Anschluss mit der Freihand-Funktion „Freehand-Tool“ die DNA-Länge zu ermitteln. Vom Bildrand abgeschnittene, sich überlagernde und verschmutzte bzw. versalzte DNA, die sich als Konglomerat aus DNA und großen, weißen (signalintensiven) Fremdkörpern darstellten, wurden hierbei ausgeschlossen. Die Substrate wurden jeweils dreimalig mittig mit dem Freihand-Tool durchfahren, um die mittlere Länge zu ermitteln. Da die Länge der in dieser Arbeit verwendeten DNA-Substrate 505 bp betrug und ein Basenpaar durchschnittlich 0,34 nm [142] lang ist, ergibt sich daraus eine theoretische Länge von 171,7 nm für die verwendeten DNA-Schadens-Substrate. Mit dem Wissen um die theoretische Länge und der Messung der tatsächlichen Länge von ca. 50 DNA-Strängen konnten nach der Auswertung mit Origin und Ermittlung der Standardabweichung zu kurze DNA-Stränge ausgeschlossen werden und die Qualität der DNA-Substratpräparation beurteilt werden (Abbildung 19). Die gemessene Länge sollte mit der errechneten Länge grob ( $\leq 10\%$ ) übereinstimmen, um DNA mit Strangbrüchen auszuschließen, die die relative Position des Schadens (bei 50% DNA-Länge) verändert hätten. Im Weiteren wäre sonst keine spezifische Winkelmessung am Schadensort möglich gewesen.



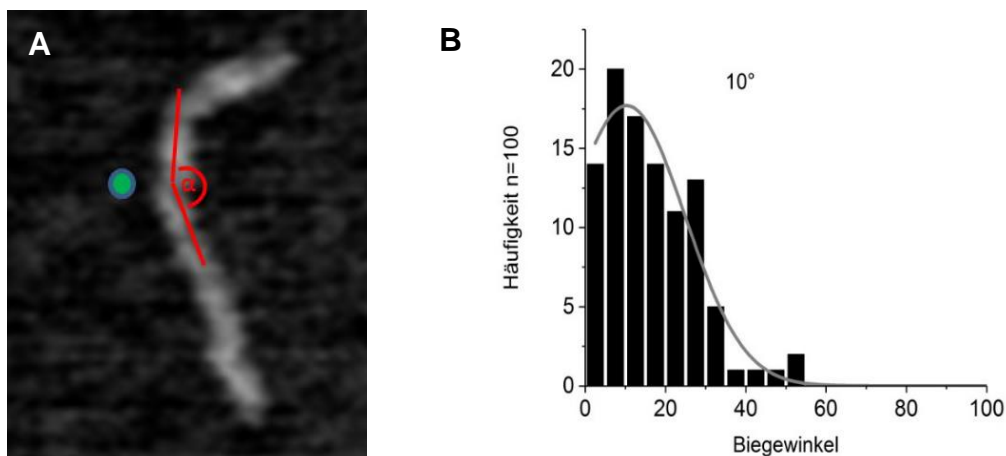
**Abbildung 19: Manuelles Messverfahren zur Längenbestimmung**

**A:** Vermessung der DNA mit dem Freihand-Werkzeug von ImageJ.

**B:** Exemplarische Längenverteilung der manuellen Messmethode.

## 2.2.4 DNA-Biegewinkelmessung an der Schadensposition

Zur Ermittlung des Mittelpunktes wurde von beiden Enden jeweils dreimal die Länge zur Mitte gemessen und mit dem Zeichen-Werkzeug durch einen Punkt markiert (Abbildung 20).



**Abbildung 20: Manuelles Messverfahren zur Biegewinkelbestimmung**

**A:** Nachdem mit dem Freihand-Messwerkzeug von ImageJ durch wiederholtes Messen von beiden Seiten die Mitte des DNA-Stranges ermittelt wurde (grüner Punkt), konnte der intrinsische Biegungswinkel an diesem Punkt mit Hilfe des Winkelmesswerkzeuges bestimmt werden. Hierzu wurde jeweils eine Gerade möglichst mittig durch die zwei gedachten DNA-Schenkel an der Mitte des DNA-Abschnittes gelegt.

**B:** Exemplarische Biegewinkelverteilung der manuellen Messmethode

Die Winkelmessung erfolgte gemäß der „manual tangent overlay“-Methode [1, 118, 121, 143, 144] mit dem Winkel-Werkzeug in ImageJ, indem man möglichst mittig zwei Tangenten durch die zwei gedachten DNA-Schenkel rechts und links der bei 50 % befindlichen Schadensposition legte. Auch das Winkelmessen wurde dreimalig durchgeführt. Der Biegewinkel  $\beta$  an der Schadensposition der DNA wurde aus dem mittleren Wert  $\alpha$ , der drei Winkelmessungen (Formel 1) errechnet.

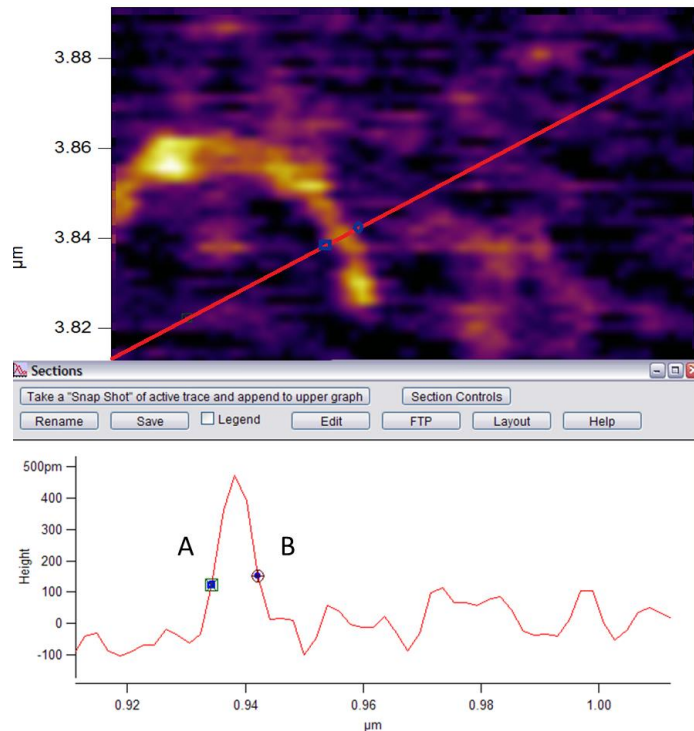
$$\beta = 180 - \alpha \quad \text{Formel 1}$$

Zur statistischen Auswertung und graphischen Darstellung der zuvor ermittelten DNA-Längen und Biegewinkel am Schadensort wurde das statistische Analyse- und Darstellungsprogramm Origin genutzt. Nach Einpflegen der Messdaten wurden diese entsprechend ihrer Häufigkeit gezählt und als Histogramm ausgegeben. Mit einer Gauß-Anpassung konnte anschließend das jeweilige Maximum analysiert werden, wobei die Gauß-Verteilung der DNA-Biegewinkel in der Literatur vielfach bestätigt wurde [145-148]. Für DNA-Längen wurde eine Bin-Größe (Balken, engl. bin) von 10 nm und für die Biegungswinkel eine Bin-Größe von 3,5° verwendet. Die Bin-Größe legt die Einteilung der verschiedenen Werte in die Balken eines Balkendiagrammes fest. Die Bin-Größe der Diagramme sollte sich an der Quadratwurzel der Datenanzahl orientieren, hiervon wurde aber bei der Biegewinkelverteilung abgewichen, da nur durch ein mäßiges „Überbinnen“ die verschiedenen Biegewinkelspezies darzustellen waren. Durch die Anwendung einer Gauß-Anpassung (Formel 2) an die Häufigkeitsverteilung ergibt sich der DNA-Biegewinkel am jeweiligen Maximum mit einer Halbwertsbreite entsprechend der Gaußschen Anpassung. Die Halbwertsbreite ( $w$ ) entspricht zwei Standardabweichungen ( $2\sigma$ ). Die DNA-Länge wird hier stets mit doppelter Standardabweichung ( $2\sigma$ ), die Biegewinkel mit der errechneten einfachen Standardabweichung ( $1\sigma$ ) angegeben.

$$y(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-x_G}{\sigma}\right)^2} \quad \text{Formel 2}$$

### 2.2.5 DNA-Höhe und Breite

Für die Software-basierte automatische Auswertung (siehe Kapitel 2.2.7) wurden die DNA-Höhen und Breiten benötigt. Mit dem Section-Tool der MFP3D-Software konnten die AFM-Aufnahmen im [.ibw]-Format an Hand des Höhen-Oszillogramms vermessen werden. Für die Breitenbestimmung, zu sehen in Abbildung 21, wurden die Punkte A und B so im Oszillogramm platziert, dass die Breite des DNA-Querschnitts möglichst gut wiedergegeben wurde. Zur Ermittlung der DNA-Höhe wurde mit dem Cursor ein Punkt A auf das dem DNA-Strang entsprechende Amplitudenmaximum gesetzt, der Punkt B auf den mittleren Hintergrund gelegt. Das Programm gab dann die Höhen-Differenz  $\Delta y$  von A und B aus, die der DNA-Höhe entsprach.

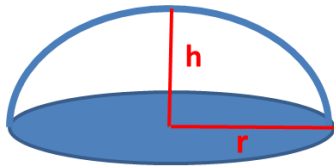


**Abbildung 21: Messung der DNA-Breite mit dem Section-Tool**

Mit dem Cursor wurden die Punkte A und B auf der Höhenkurve (in der Abbildung unten) platziert. Der Abstand beider Punkte in x-Richtung entspricht der DNA-Breite.

### 2.2.6 Proteinvolumen

Zur Charakterisierung und Überprüfung der DNA-gebundenen Proteinspezies wurde deren Volumen bestimmt. Ähnlich zum Vorgehen bei der DNA-Breitenmessung wurde senkrecht zum DNA-Strang die Breite des Proteinpeaks auf der DNA mit den Messpunkten A und B gemessen. Die höchste Erhebung an der DNA-Protein-Position wurde, wie für die DNA-Höhe oben beschrieben, relativ zum mittleren Hintergrund gemessen. Das Volumen wurde entsprechend der Gleichung des Volumens einer Kugelkappe (Formel 3) errechnet [149].



$$V = \frac{\pi}{6 \cdot h} (3r^2 + h^2)$$

**Formel 3**

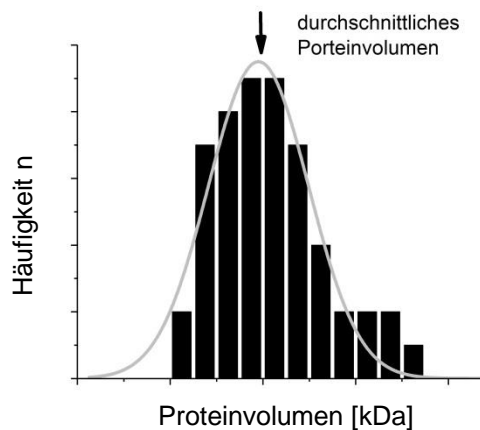
Für die Umrechnung des Protein-Volumens in die entsprechende molekulare Masse fand die Formel 4 aus der Veröffentlichung von Roth *et al.* [150] ihre Anwendung. Die Grundlage der Gleichung war eine Eichgerade der Publikation zur Ermittlung der Relation zwischen dem AFM-Volumen und der molekularen Masse.

$$M_w = \frac{V + 5,9}{1,2}$$

**Formel 4**

Die Verwendung der Software Image SXM war eine weitere Möglichkeit zur Volumen-Bestimmung von einzelnen Proteinen. Nach Öffnen des AFM-Bildes im [.ibw]-Format, wurde entsprechend der Aufnahmeeinstellungen des Bildes die Einheit (nm) gewählt und mit der „Analyse-Funktion“ ein Höhen-Histogramm der Abbildung erstellt. In diesem Histogramm wurde eine Gauß-Kurvenanpassung durchgeführt, um die mittlere Hintergrundhöhe zu definieren. Die auf diese Weise ermittelte Hintergrund-Höhe konnte als Off-Set genutzt werden. So wurde die Höhe des Proteinpeaks über dem Hintergrund und die Proteinpeak-Grundfläche auf dem Hintergrund-Niveau beim Volumen-Messvorgang berücksichtigt. Störende Fragmente und Hintergrundrauschen um die Proteine konnten mit der Radier-Funktion (Erase-Funktion) aus dem Bild gelöscht werden, um so die Ausgangsdaten für die Auswertung zu verbessern. Die Proteine wurden vom

Programm automatisch nummeriert und die Messdaten in Form von Grundfläche und mittlerer Höhe entsprechend der Nummerierung in einer [.txt]-Tabelle ausgegeben. Die Daten wurden getrennt nach Grundfläche, mittlerer Höhe und errechnetem Volumen (Produkt aus Grundfläche und mittlerer Höhe) in einer Excel-Tabelle gespeichert, um schließlich mit Origin dargestellt und statistisch analysiert zu werden (Abbildung 22). Eine Umrechnung des Protein-Volumens in die Molekülmasse erfolgte dann nach der oben eingeführten Umrechnungsgleichung (Formel 4). Die Ergebnisse aus den Messungen mit dem Section-Tool der AFM-Software und den Messungen mit der Image SXM-Software sowie den beiden mathematischen Modellen (Kugelkappe und Zylinder mittlerer Höhe) lieferten vergleichbare Ergebnisse. Die Ergebnisse der Modelle waren folglich vergleichbar.



**Abbildung 22: Exemplarische Proteinolumenverteilung**

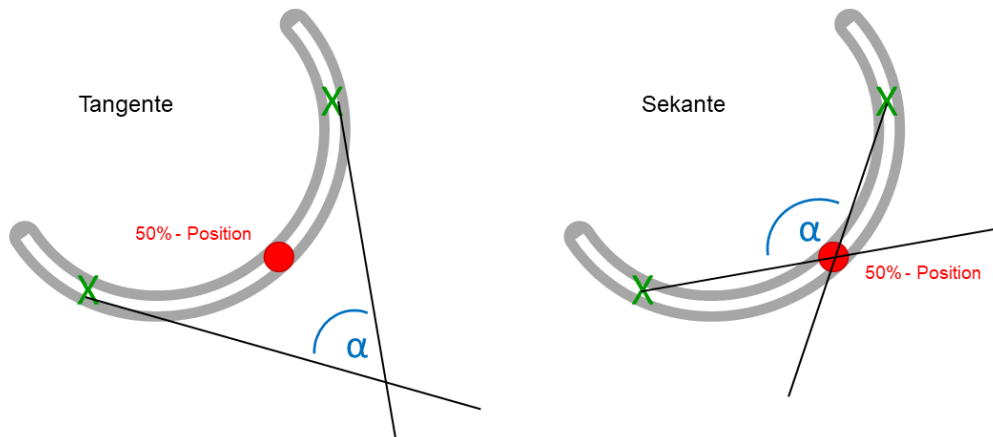
Die Abbildung stellt eine exemplarische Proteinolumenverteilung dar. Der Pfeil verdeutlicht das ermittelte durchschnittliche Proteinolumen.

### 2.2.7 Standardisierte und objektive Biegewinkelmessungen an DNA-Schäden

Nachfolgend wird die Methode der automatisierten Biegewinkelmessung kurz eingeführt, wobei die detaillierte Beschreibung und Etablierung der Methode ein Bestandteil des Ergebnis-Kapitels 3.2 ist.

Für die Auswertung der Aufnahme waren mehrere Vorbereitungen mit anderen Programmen nötig. Die Bild-Prozessierung begann mit dem Schneiden auf ein quadratisches Format mit dem Bildbearbeitungsprogramm GIMP. Darauf folgten weitere Bearbeitungsschritte mit dem Bildverarbeitungsprogramm ImageJ: So erfolgte eine Formatänderung vom RGB-Format in das 8 bit-Format. Es war zudem eine Median-Filterung nötig, die das Hintergrundrauschen minimierte. Ebenso wurde durch die Verwendung eines Thresholds und die Festlegung einer geeigneten Hörschwelle die Filterung des Hintergrundes für die spätere Auswertung in FIESTA bestimmt. Dies waren alle Vorbereitungen für die eigentliche DNA-Skelettierung mit dem Programm FIESTA. FIESTA skelettierte entsprechend den gewählten Einstellungen die DNA-Stränge und speicherte die Koordinaten der erzeugten DNA-Filamente in einer Text-Datei. Mit Hilfe einer speziell für die Auswertung von Biegeungswinkeln an DNA-Filamenten von Hannah Heil (AG Heinze, Rudolf-Virchow-Zentrum Würzburg) entwickelten MATLAB basierenden Software konnten die skelettierten DNA-Fragmente dann analysiert werden. Es standen zwei verschiedene MATLAB-Programmversionen zur Verfügung. Diese unterschieden sich im Programmablauf einzig in der Methode der Winkelmessung. Für die beiden Messmethoden wurden zwei verschiedene geometrische Konstruktionen verwendet. Zum einen wurde der Winkel durch zwei Sekanten, zum anderen durch zwei Tangenten gebildet. Während allgemein in der Geometrie eine Tangente (lat. *tangere*: berühren) eine Gerade darstellt, die eine Kurve nur punktuell berührt, so schneidet eine Sekante (lat. *secare*: schneiden) die Kurve an zwei Punkten (Schematische Darstellung in Abbildung 23). In der MATLAB-Programmierung wurden bei der Sekantenmethode (DNAsegmentAnalysis\_SE-CANT\_20190115) zwei Sekanten an der gewählten Segmentlänge durch die DNA-Strangmitte gelegt. Diese beiden Sekanten schnitten sich im zuvor ermittelten Mittelpunkt des DNA-Stranges. Hingegen nutzte die Tangentenmethode

(DNAsegmentAnalysis\_TANGENT\_20190115) zwei Geraden, die bei der entsprechenden Segmentlänge der von FIESTA ermittelten DNA-Strangmitte nur anlagen. Da die Biegung des DNA-Strangs an der gewählten Segmentlänge die Steigung der Tangente bestimmte, schnitten sich die Tangenten im Gegensatz zu den Sekanten nicht zwangsläufig in der ermittelten 50 %-Position und bei entsprechend größerer Biegung auch außerhalb der DNA.



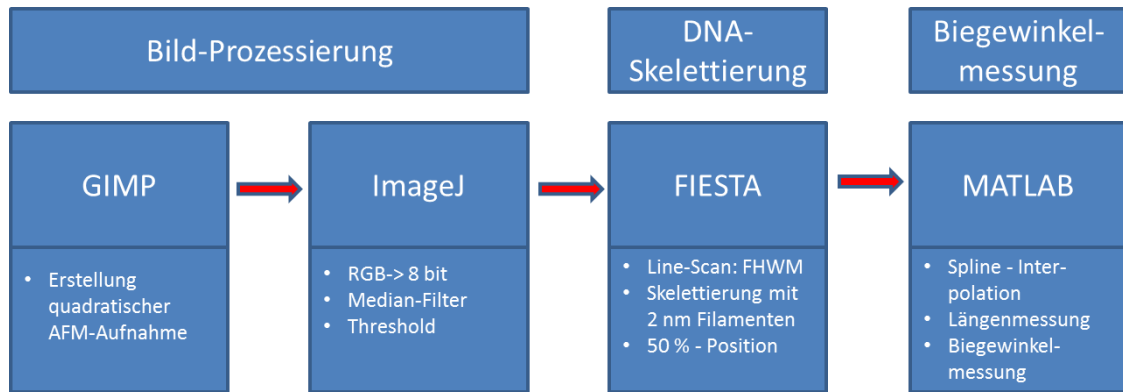
**Abbildung 23: Schematische Darstellung der Tangenten-, Sekanten-Messmethode**

Die graue Linie symbolisiert den DNA-Strang, die weiße Linie die von Fiestra erzeugte Filamentierungslinie. Der rote Punkt zeigt die 50 %-Position der Länge des DNA-Stranges. Die grünen Kreuze entsprechen der gewählten Segmentlänge (im Falle dieser Arbeit 4 nm, 8 nm). Die Tangenten liegen der Filamentierungslinie nur im Bereich des grünen Kreuzes an. Die lokale Biegung am grünen Kreuz gibt die jeweilige Steigung der Tangenten vor. Folglich stimmt der Schnittpunkt der beiden Tangenten nicht mit dem Punkt der 50 %-Position überein. Hingegen verlaufen die Sekanten immer durch die zwei Punkte, nämlich die gewählte Segmentlänge und die 50 %-Position.

Im Anhang befinden sich die entsprechenden MATLAB-Programmierungen und eine detaillierte Bedienungsanleitung. Außerdem wird das automatische Messprogramm online bei „Open Science Framework“ unter der folgenden Adresse zur Verfügung gestellt: <https://osf.io/yhwuc/>

In Abbildung 24 wird der Arbeitsablauf der automatisierten Biege winkelmessung an spezifischen DNA-Positionen veranschaulicht. Dieser Arbeitsablauf wurde für die Biege winkelmessung von singulärer DNA optimiert, die bei 50 % der jeweils 505 bp langen DNA einen Basen-Schaden enthielt. Dieser Basenschaden war für die verschiedenen Glykosylasen spezifisch.





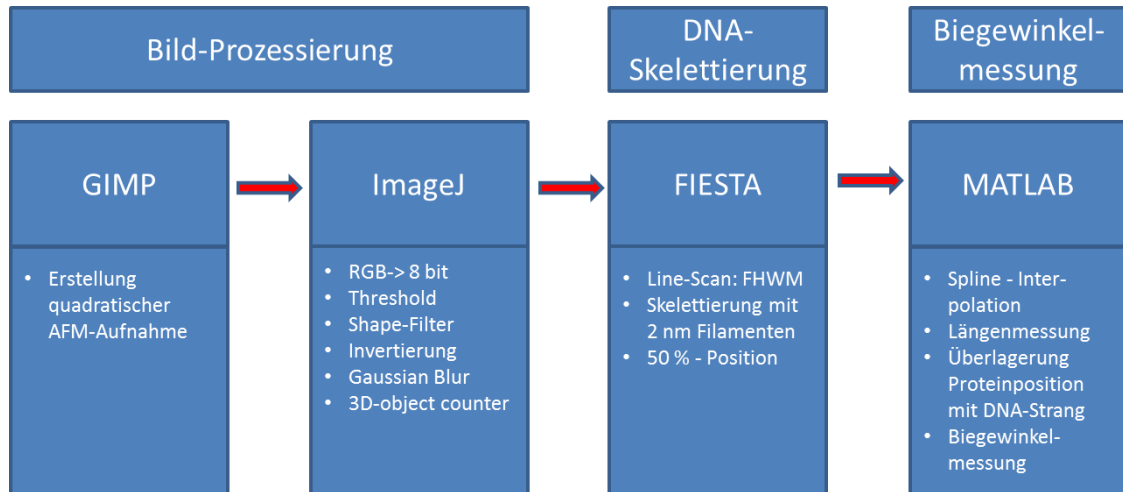
**Abbildung 24: Arbeitsablauf bei der automatisierten Ermittlung von Biegewinkeln an DNA**

Zu Beginn der Auswertung wird die Aufnahme mit dem Bildbearbeitungsprogramm GIMP zugeschnitten, anschließend wird die Aufnahme mit dem Bildverarbeitungsprogramm ImageJ für die Analyse mit FIESTA vorbereitet. Nach einer Formatänderung (RGB in 8 bit) und einer Median-Filterung wird ein Threshold bestimmt. FIESTA skelettisiert die DNA-Stränge mit 2 nm langen Filamenten und stellt einen Datensatz mit den Filamentdaten und der 50 %-Position her. Die in diesen Datensätzen beschriebenen Filamente werden nach einer Spline-Interpolation durch eine speziell für die Messung von Biegewinkeln an DNA programmierte MATLAB-Anwendung ausgewertet. Das Ergebnis wird als [.mat]-Datensatz mit Längen- und Biegewinkelmessung ausgegeben.

### 2.2.8 Standardisierte und objektive Biegewinkel-messungen an Protein-DNA-Komplexen

Wie in der Einleitung beschrieben, sollte zusätzlich auch der von den entsprechenden Glykosylasen an der DNA erzeugte Biegewinkel ermittelt werden. Für diesen Zweck waren mehrere Anpassungen des bestehenden Programmablaufes und zusätzliche Programmfunktionen nötig. So musste in ImageJ auf die Median-Filterung verzichtet werden, stattdessen war ein Shape-Filter (engl. shape; Form) für die separate Darstellung von DNA ohne die Proteine nötig. Für die spätere Auswertung des Shape-Filter Ergebnisses in FIESTA waren zusätzliche Bearbeitungsschritte in ImageJ notwendig. Die Shape-Filter-Darstellung wurde invertiert und ein Gaußscher Verwischungsfilter (Gaussian Blur) angewendet. Parallel wurden die Proteine mit der Programmfunktion des „3D object counters“ in ImageJ vermessen und in Form von Koordinaten lokalisiert. Die Skelettierung mit FIESTA wurde identisch wie oben beschrieben durchgeführt, allerdings nur an der proteinfreien Darstellung des Shape-Filters. Die Adaption der MATLAB-Software war größeren Umfanges. Die Position der Proteine wurde mit den Filamenten überlagert und anschließend an der Position der gebundenen Proteine die

Biegewinkel gemessen. Die genaue Etablierung dieser Methode und der verschiedenen Arbeitsschritte wird detailliert in den Ergebnissen unter 3.4 beschrieben. Eine Übersicht der Arbeitsschritte zeigt die nachfolgende Abbildung 25.



**Abbildung 25 : Arbeitsablauf bei der automatisierten Ermittlung von Biegewinkeln an Protein-DNA-Komplexen**

Zu Beginn der Auswertung wird die Aufnahme mit dem Bildbearbeitungsprogramm GIMP zugeschnitten, anschließend wird die Aufnahme mit dem Bildverarbeitungsprogramm ImageJ für die Analyse mit FIESTA vorbereitet. Nach einer Formatänderung (RGB in 8 bit) wird ein Threshold bestimmt, der Shape-Filter und der „3D object counter“ verwendet. Das Ergebnisbild des Shape-Filters wird in ImageJ noch invertiert und mit einem Gaußschen Verwischungsfilter (Gaussian Blur) bearbeitet. FIESTA skelettiert die DNA-Stränge und erstellt einen Datensatz mit den Filamentdaten. Die in diesen Datensätzen beschriebenen Filamente werden durch eine speziell für die Messung von Biegewinkeln an Protein-DNA-Komplexen programmierte MATLAB-Anwendung ausgewertet. MATLAB führt eine Spline-Interpolation durch und überlagert die Protein-Position mit dem DNA-Strang. Das Ergebnis aus gemessener Länge und Biegewinkel wird wiederum als [.mat]-Datensatz ausgegeben.

## 3 Ergebnisse

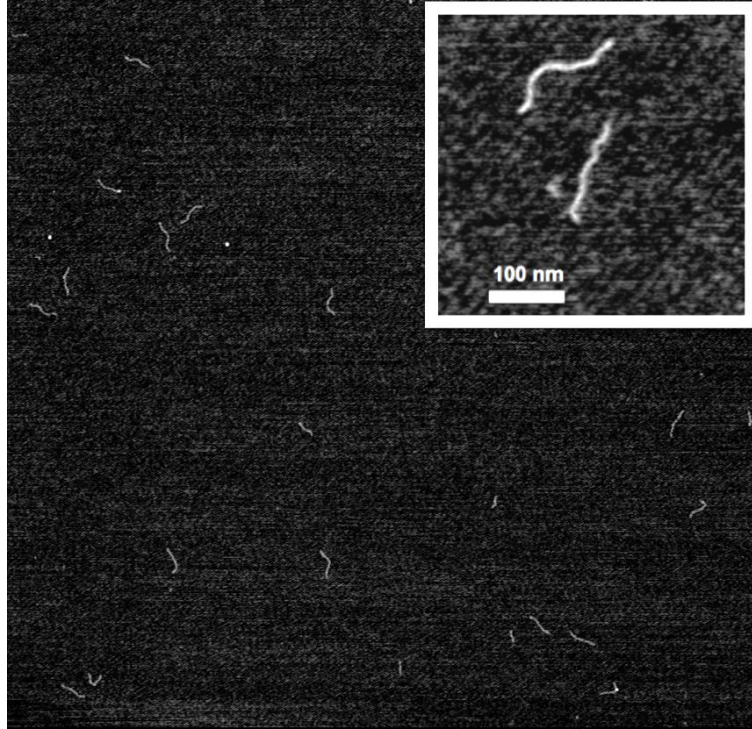
### 3.1 Rasterkraftmikroskopische Aufnahmen

Mit dem Rasterkraftmikroskop wurden verschiedene Glykosylasen und deren spezifische DNA-Substrate abgebildet. Alle zur Deposition verwendeten DNA-Schadenssubstrate wurden gemäß dem in den Methoden erläuterten Vorgehen von der Arbeitsgemeinschaftskollegin Lisa Hirsch im Rahmen ihrer Masterarbeit hergestellt. Die Schadensloci und deren Sequenzkontexte wurden in dem Kapitel Material in der Tabelle 2.1.8 aufgelistet. Die einzelnen Präparationsschritte sind in dem Kapitel 2.2.1 beschrieben. Das Vorgehen entspricht der Methodik, die in der Veröffentlichung von Büchner und Tessmer [128] zur DNA-Präparation für AFM-Studien vorgestellt wurde.

Zuerst wurden sowohl die DNA-Substrate (Abbildung 26), als auch die Proteine (Abbildung 27) einzeln aufgenommen. Im Anschluss wurden die verschiedenen Glykosylasen mit der den jeweiligen Zielschaden tragenden DNA und auch zusätzlich mit schadensfreier DNA inkubiert und aufgenommen (Abbildung 28). Abgebildet wurden die DNA-Substrate oxoG:A und  $\epsilon$ A:T, beide jeweils mit einer Länge von 505 bp. Die Abbildung ungeschädigter DNA erfolgte sowohl mit einer Länge von 2224 bp, als auch mit einer Länge von 505 bp. Zusätzlich wurden die Glykosylasen hAAG und MutY jeweils einzeln aufgenommen. Das genaue Vorgehen der Depositionen kann den Methoden entnommen werden.

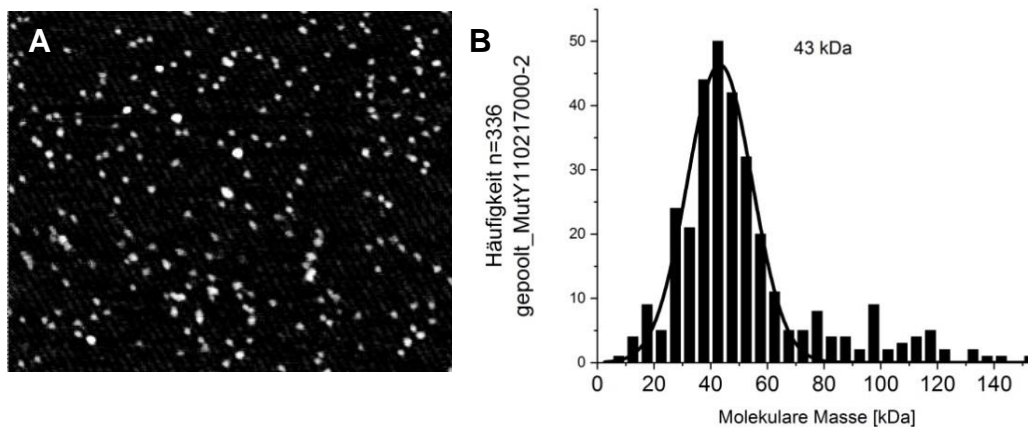
Zur Identifizierung der gebundenen Proteine und deren Unterscheidung von Verschmutzungen, wurde das Proteinvolumen der Proteine sowohl in einer Aufnahme, die nur die Proteinspezies enthielt, als auch in Aufnahmen, in denen die Proteine an DNA gebunden waren, ermittelt. Das Proteinvolumen konnte anhand der Formel 4 in dessen ungefähre Molekülmasse umgerechnet werden und schließlich mit der aus der Literatur bekannten Masse verglichen werden. Eine Volumenbestimmung von einzelnen Proteinen wurde bevorzugt mit der Image SXM-Software durchgeführt (Abbildung 27 B). Im Falle von an DNA gebundenen Proteinen wurde hingegen das Section-Tool (Abbildung 28 B) verwendet. Hier mussten die Höhe und der Durchmesser des Proteins gemessen werden.

Anhand von Formel 3 zur Bestimmung des Volumens eines Kugelausschnittes konnte wiederum das Volumen festgestellt werden. Zusätzlich wurde das Volumen gebundener Proteine mit dem Volumen von freien Proteinen verglichen.



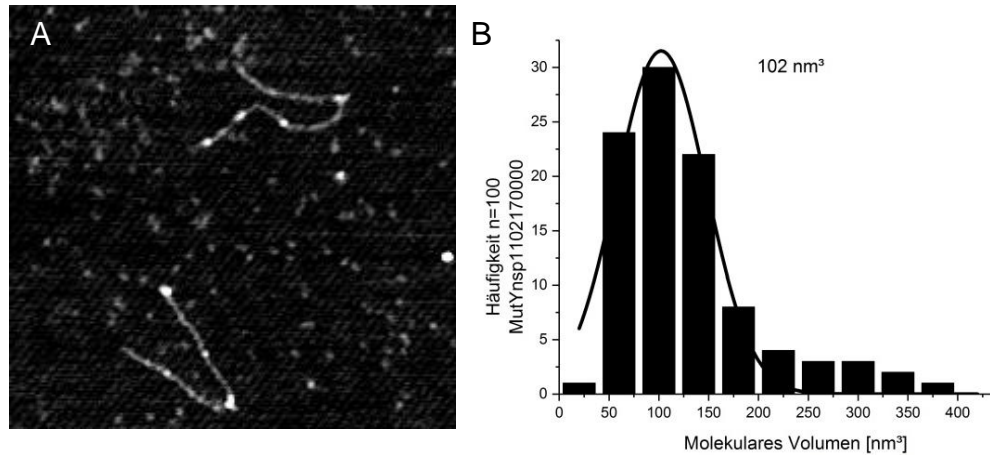
**Abbildung 26: Beispiel einer AFM-Aufnahme von DNA**

Die Aufnahme zeigt beispielhaft eine rasterkraftmikroskopische Aufnahme von DNA mit dem Basenschaden oxoG:A bei einer Länge von 505 bp. Rechts befindet sich zusätzlich ein vergrößerter Ausschnitt.



**Abbildung 27: Beispiel einer AFM-Aufnahme von Proteinen und Proteinvolumen (MutY)**

**A:** Abgebildet ist eine rasterkraftmikroskopische Aufnahme der Glykosylase MutY  
**B:** Die Auswertung des Proteinvolumens von MutY aus Abbildung 27 A (Experiment 02.11.2017) mit Image SXM erbrachte vereinigt eine mittlere molekulare Masse von 43 kDa.

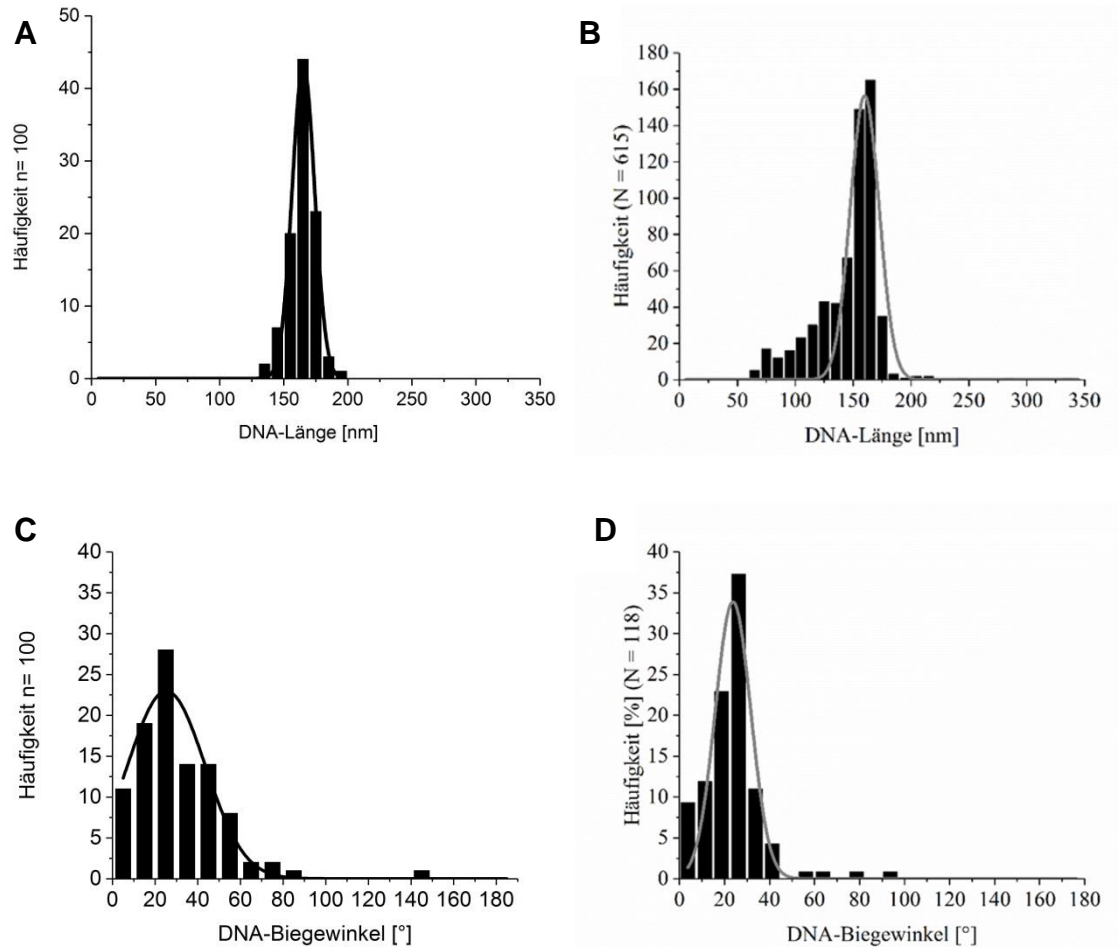


**Abbildung 28: Proteinvolumen von MutY, gebunden an ungeschädigter DNA**

**A:** Der Bildausschnitt zeigt eine rasterkraftmikroskopische Aufnahme von ungeschädigter DNA mit MutY-DNA-Komplexen

**B:** Die Auswertung des Proteinvolumen von MutY vom Experiment am 02.11.2017 mit Hilfe des Section Tools ergab ein mittleres molekulares Volumen von 102 nm<sup>3</sup>

Neben der Volumenmessung an Proteinen wurde mit dem etablierten manuellen Verfahren, was im Kapitel Methoden unter 2.2.3 beschrieben ist, die Länge der DNA und der Biegewinkel an der 50 %-Position gemessen. Verschiedene Rasterkraft-Aufnahmen von bereits ausgewerteten Bildern wurden zur Objektivierung der Messergebnisse und für den Vergleich der vom Experimentator abhängigen Messabweichungen nochmals mit dem etablierten manuellen Verfahren in ImageJ vermessen und anschließend die jeweiligen Ergebnisse verglichen. Am Beispiel der Messung derselben AFM-Daten für oxoG:A durch verschiedene Experimentatoren in Abbildung 29 konnte eine Übereinstimmung der DNA-Längen und Biegewinkel erkannt werden.



**Abbildung 29: Längen- und Biegungswinkelverteilung für oxoG:A-DNA aus Messungen verschiedener Experimentatoren**

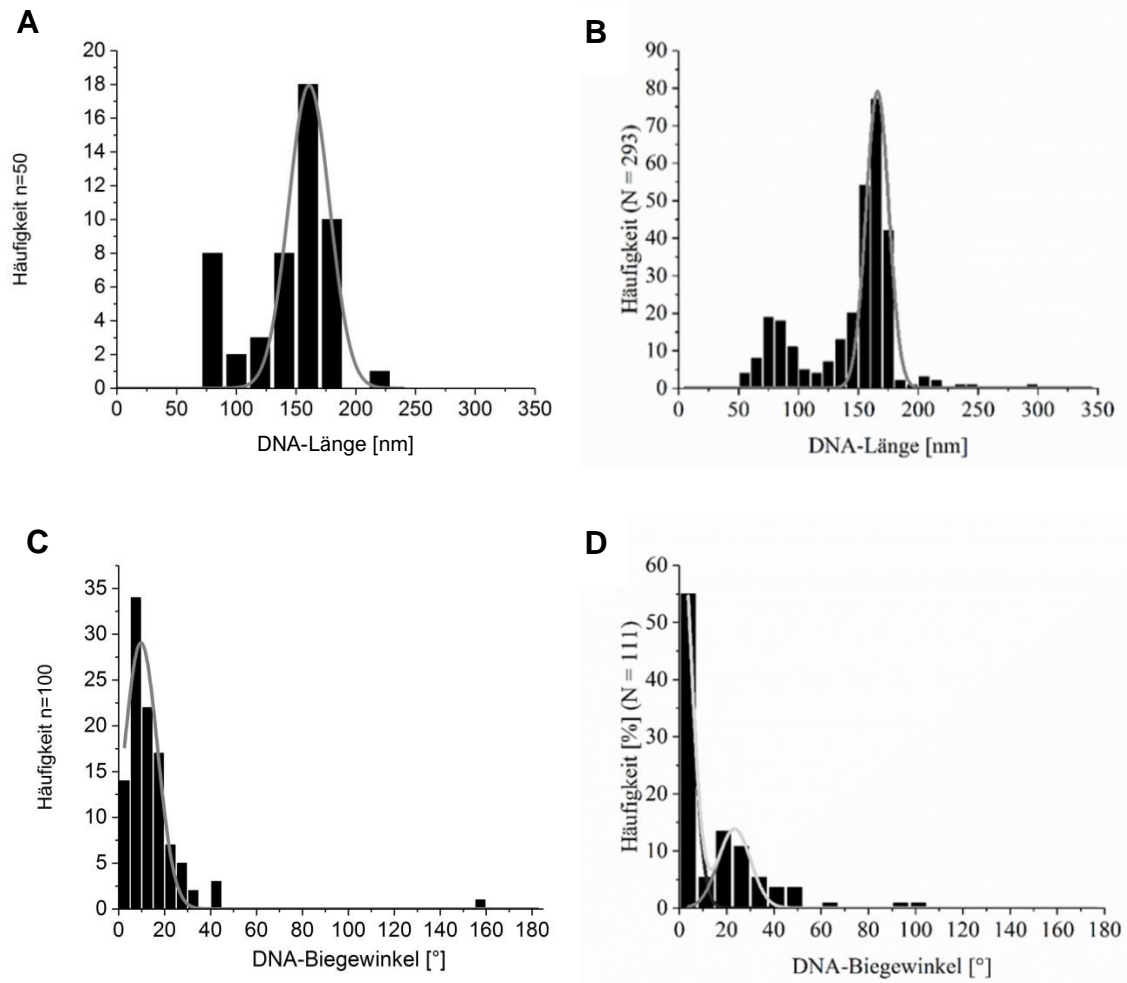
**A:** eigene Messergebnisse der Länge von oxoG:A:  $164 \text{ nm} \pm 17 \text{ nm}$  ( $2 \sigma$ )

**B:** Längenmessung von Lisa Hirsch für oxoG:A:  $159 \text{ nm} \pm 25 \text{ nm}$  ( $2 \sigma$ )

**C:** eigene Messergebnisse des Biegewinkels von oxoG:A:  $25^\circ \pm 9^\circ$  ( $1 \sigma$ )

**D:** Biegungswinkelmessung von Lisa Hirsch an oxoG:A:  $24^\circ \pm 8^\circ$  ( $1 \sigma$ )

Während es bei Messungen an oxoG:A zu vergleichbaren Ergebnissen kam, differierten die Werte teilweise, wie zum Beispiel bei oxoG:C zwischen Messungen verschiedener Experimentatoren (Abbildung 30) sowie bei einer wiederholten Messung von oxoG:A. Die Messungen schienen von der Darstellungsqualität und besonders von der Länge der Winkeltangenten abhängig zu sein. Die manuell ermittelten Werte differierten deutlich zwischen verschiedenen Experimentatoren und Experimenten, sodass die Einführung eines automatischen Experimentator-unabhängigen Auswertungsprogrammes für objektivere Messergebnisse gerade im Bereich kleiner Biegewinkel unerlässlich war. Ein solches Programm wurde im Rahmen dieser Arbeit etabliert und evaluiert, wie im nächsten Kapitel 3.2 beschrieben.



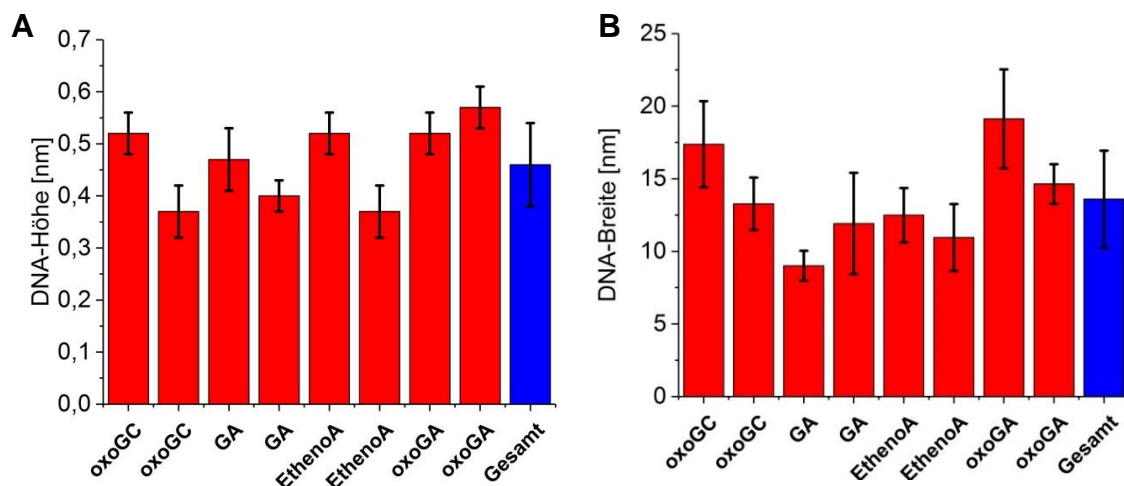
**Abbildung 30: Längen- und Biegungswinkelverteilung für oxoG:C-DNA aus Messungen verschiedener Experimentatoren**

- A:** eigene Messergebnisse der Länge von oxoG:C:  $160 \text{ nm} \pm 34 \text{ nm}$  ( $2 \sigma$ )
- B:** Längenmessung von Lisa Hirsch für oxoG:C:  $166 \text{ nm} \pm 19 \text{ nm}$  ( $2 \sigma$ )
- C:** eigene Messergebnisse des Biegewinkels von oxoG:C:  $10^\circ \pm 7^\circ$  ( $1 \sigma$ )
- D:** Biegungswinkelmessung von Lisa Hirsch an oxoG:C:  $0^\circ \pm 5^\circ$  und  $23^\circ \pm 7^\circ$  ( $1 \sigma$ ) aus Doppel-Gauß-Fit

## 3.2 Standardisierte und objektive Biegewinkelmessungen an DNA-Schäden

### 3.2.1 DNA-Höhe und Breite

Bei Programmeinführung mussten anfangs verschiedene Probenparameter abgeschätzt werden. Zwei wichtige Parameter, die es zu Beginn zu ermitteln galt, waren die mittlere Höhe und Breite der DNA. Die Messung dieser Größen erfolgte mit dem Section Tool der MFP-3D-Software entsprechend dem in den Methoden beschriebenen Vorgehen. Von den vier verschiedenen verwendeten Schadenssubstraten wurden je zwei Aufnahmen untersucht. Dabei wurden die Parameter in jeder Aufnahme an fünf verschiedenen DNA-Strängen jeweils doppelt vermessen. Bei den Messungen ergab sich hierbei für die DNA eine durchschnittliche Breite von 13 nm  $\pm$  2 nm sowie eine Höhe von 0,46 nm  $\pm$  0,08 nm. Die Abbildung 31 zeigt die Ergebnisse für die verschiedenen vermessenen Datensätze (rot) und für die Gesamtwerte (blau). Im Zusammenhang mit Kapitel 3.2.5 Threshold werden diese Werte wiederaufgegriffen.

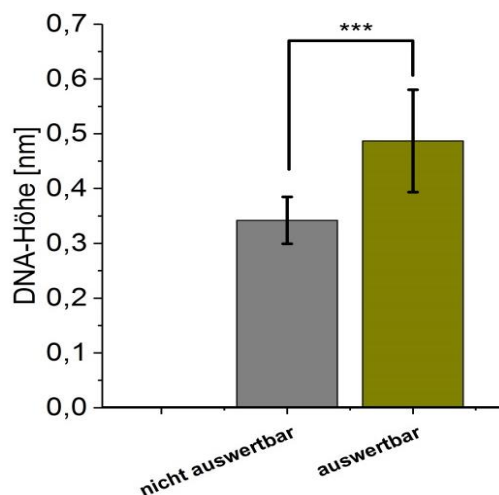


**Abbildung 31: Bestimmung der Parameter Höhe und Breite von DNA**

Die Ermittlung der Höhen- und Breitenparameter des DNA-Stranges war für eine korrekte Software-Abstimmung notwendig. Von jedem Schadenssubstrat wurden je zwei Aufnahmen untersucht. So wurden die Parameter in jeder Aufnahme an fünf verschiedenen DNA-Strängen jeweils doppelt gemessen. Bei der Messung mit dem Section-Tool ergab die DNA-Höhe (A) im Mittel einen Wert von 0,46 nm  $\pm$  0,8 nm und die Breite (B) einen mittleren Wert von 13 nm  $\pm$  2 nm.



Bei der Etablierung der Auswertungssoftware zeigte sich, dass verschiedene AFM-Aufnahmen nicht mit dem Programm auswertbar waren. Um zu ermitteln, ob die Bildqualität dafür verantwortlich war, wurden die Bilder durch die jeweilige DNA-Höhe charakterisiert, da diese die Schärfe der Messsonde widerspiegelt. Es zeigte sich, dass sich die Höhe der DNA-Stränge von jedem der mit der Software auswertbaren Bildern signifikant von den Werten von nicht auswertbaren Bildern unterschied (Abbildung 32). So wurde mit dem Section-Tool in jeder Aufnahme an jeweils drei verschiedenen DNA-Strängen die Höhe gemessen und daraus der Mittelwert und die Standardabweichung ermittelt. Zusätzlich wurde die Höhe der DNA an jeder Messposition jeweils dreifach gemessen und gemittelt. Für nicht mit dem Programm auswertbare Aufnahmen ergab sich eine durchschnittliche Höhe von 0,34 nm mit einer Standardabweichung von 0,04 nm. Dagegen ergab die Messung der auswertbaren Bilder eine durchschnittliche Höhe von 0,49 nm mit einer Standardabweichung von 0,1 nm. Für die Berechnung der Signifikanz der Abweichung zwischen diesen beiden Größen mit dem Student t-Test ergab sich ein t-Wert von 4,162, was einem p-Wert von 0,00007419 entsprach und somit eine hochsignifikante Abweichung darstellte. Folglich wurden für die Anwendung der Software Aufnahmen benötigt, die durch die Verwendung eines scharfen Tips eine möglichst große DNA-Höhe und Bildqualität wiedergeben.

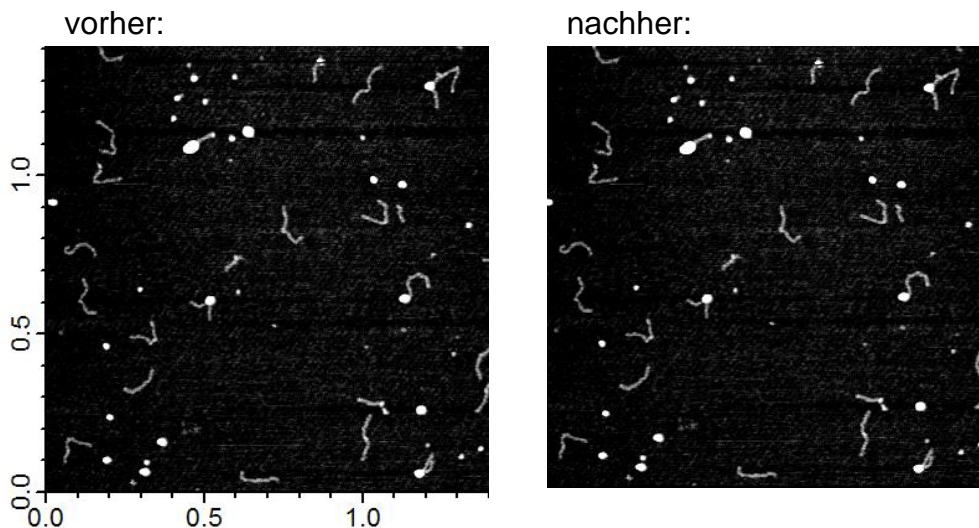


**Abbildung 32: Auswirkung der DNA-Höhe auf die Auswertbarkeit mittels Software**

Bei der Messung der DNA-Höhe ergab sich für nicht auswertbare Aufnahmen ein durchschnittlicher Wert von 0,34 nm  $\pm$ 0,04 nm, für auswertbare Aufnahmen ein durchschnittlicher Wert von 0,49 nm  $\pm$ 0,1 nm. Bei einem t-Wert von 4,162 und einem p-Wert von 0,00007419 war die Abweichung hochsignifikant.

### 3.2.2 Bildprozessierung

Da die Tracking-Software FIESTA ausschließlich quadratische Bilder, also gleiche Pixelanzahl in Länge und Breite, verarbeiten kann, wurden mit der Software GIMP die Aufnahmen gemäß der entsprechenden Bildgröße geschnitten (Abbildung 33). Skalierungsgeraden und Koordinatenangaben wurden dabei mit einem Schneidewerkzeug entfernt und das Bild schließlich mit LZW-Kompression als [.tiff]-Datei exportiert. Mit der Einführung des Programmes sollte die Messung der Biegewinkel sowohl an DNA ohne Proteinkomplexe als auch DNA mit Proteinen automatisiert werden. Zu Beginn der Programmeinführung konzentrierte man sich zur Vereinfachung vorerst ausschließlich auf die Auswertung von Biegewinkeln an DNA-Schäden in Abwesenheit von Proteinen.



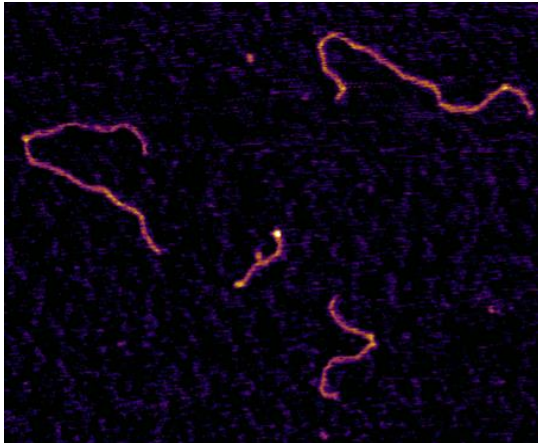
**Abbildung 33: Bearbeitung mit GIMP**

Mit dem „Auswahl-“ und „Schneidewerkzeug“ des Bildbearbeitungsprogramms GIMP wurden alle Abbildungen pixelgenau gemäß dem Aufnahmeformat auf ein quadratisches Format geschnitten, Skalierungsgerade und Lineal entfernt und mit LZW-Kompression als [.tiff]-Datei exportiert. Im Fall dieser Arbeit betrug die Pixelauflösung finaler Tiff-Bilder immer 1.95 nm/Pixel (Abbildungen zwischen 2  $\mu\text{m}$  x 2  $\mu\text{m}$  und 8  $\mu\text{m}$  x 8  $\mu\text{m}$  mit einem Vielfachen von 512 Pixeln).

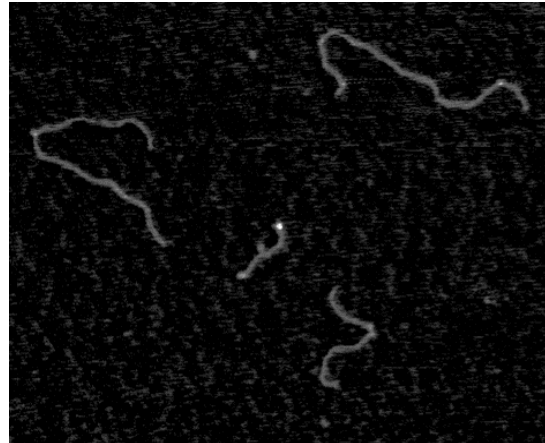
### 3.2.3 Formatänderung

Das dimensionsgerecht vorbereitete Bild wurde in ImageJ über [Type] vom RGB-Format zum 8 bit-Format umformatiert (Abbildung 34). Dies erwies sich als notwendig, da die weiteren Arbeitsschritte nur mit einer Darstellung der Aufnahme in Graustufen möglich waren.

vorher:



nachher:



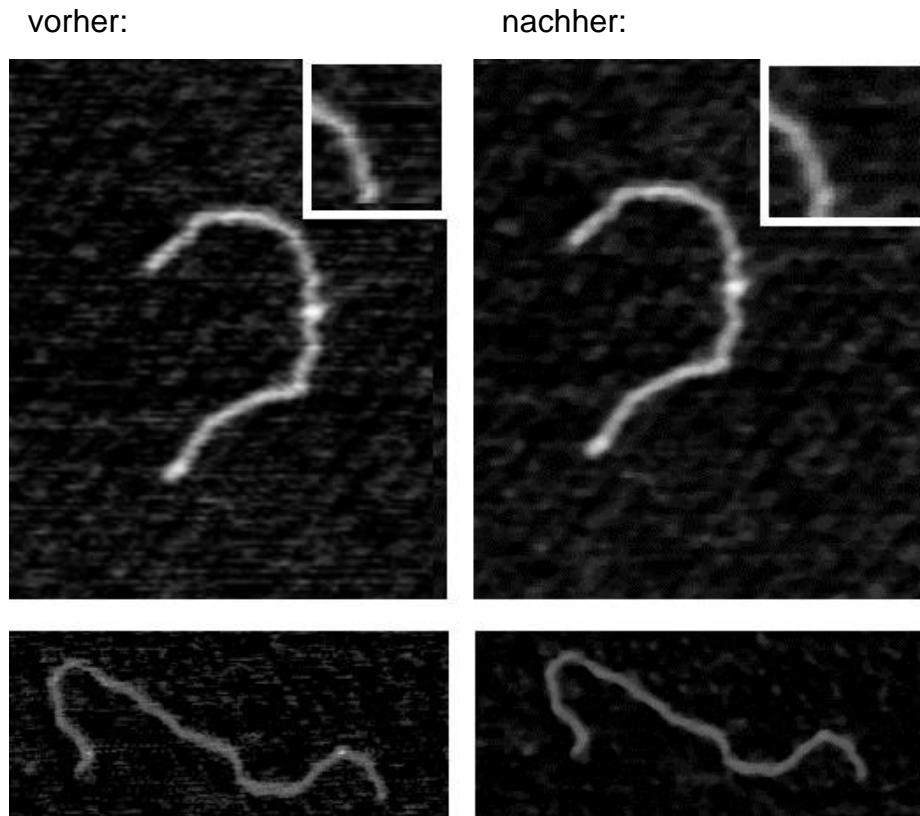
**Abbildung 34: Umwandlung des RGB- in ein 8 bit-Format**

Die Formatumwandlung war eine Notwendigkeit, um die folgende Bildverarbeitung durchführen zu können.

### 3.2.4 Anwendung eines Median-Filters

Der Filter „Median 2px“ im Programm ImageJ sorgte für eine Reduzierung des Hintergrundrauschens, da aus den Helligkeitswerten jeweils zweier Pixel der mediane Wert errechnet wurde. So wurden die DNA-Stränge deutlich intensiver, die Begrenzung aber auch weicher dargestellt. Für die weitere Bearbeitung mit FIESTA war dies von großem Vorteil, da so das Hintergrundrauschen sehr gut über den medianen Wert des Hintergrunds beseitigt wurde. Auch für die Filamentierung in FIESTA war der mediane Wert von Vorteil, da die DNA signalintensiver und weniger scharf begrenzt war. Wenn die Aufnahme ohne Median-Filter von FIESTA bearbeitet wurde, störte das Hintergrundrauschen. Zum einen wurde es als eigenständiges Molekül oder Filament erkannt, zum anderen überlagerte sich das Hintergrundrauschen mit dem DNA-Strang. Aufgrund dessen wich FIESTA teilweise von der Strangmitte zu der Überlagerung ab oder die DNA wurde in selteneren Fällen sogar als zwei getrennte Stränge filamentiert.

Vorgehen: Mit einem zwei Pixel Median-Filter (unter [Process]; [Filters] zu finden) wurde das Hintergrundrauschen vermindert (Abbildung 35), damit FIESTA im Folgenden die DNA-Strukturen besser erkennen und eine Filamentierung durchführen konnte. Auf diese Weise bearbeitet, wurde das Bild separat für den FIESTA-Arbeitsgang gespeichert (Bildname.8bit-median-2px.tiff).

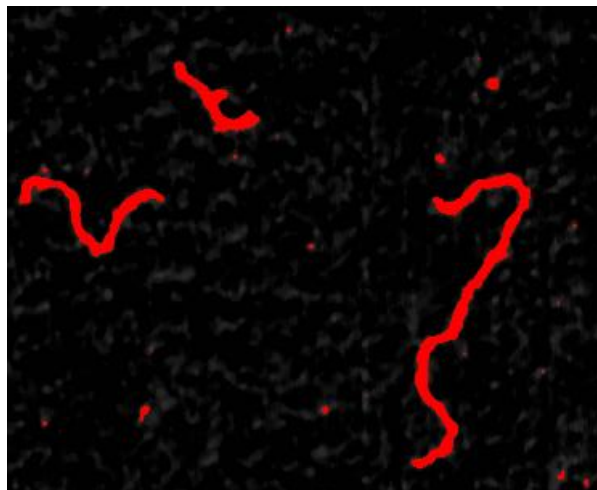


**Abbildung 35: Effekt des Median-Filters**

Um das aufnahmebedingte Hintergrundrauschen zu reduzieren, wurde ein Median-Filter von zwei Pixel verwendet. Ohne die Verwendung dieses Filters war zwar die Halbwertsbreite deutlich kleiner, aber der Skelettierungsvorgang mit FIESTA ergab immer in viele kleine Filamente geteilte DNA-Stränge. Für eine gute Skelettierung war deshalb eine mediane Filterung unverzichtbar.

### 3.2.5 Threshold

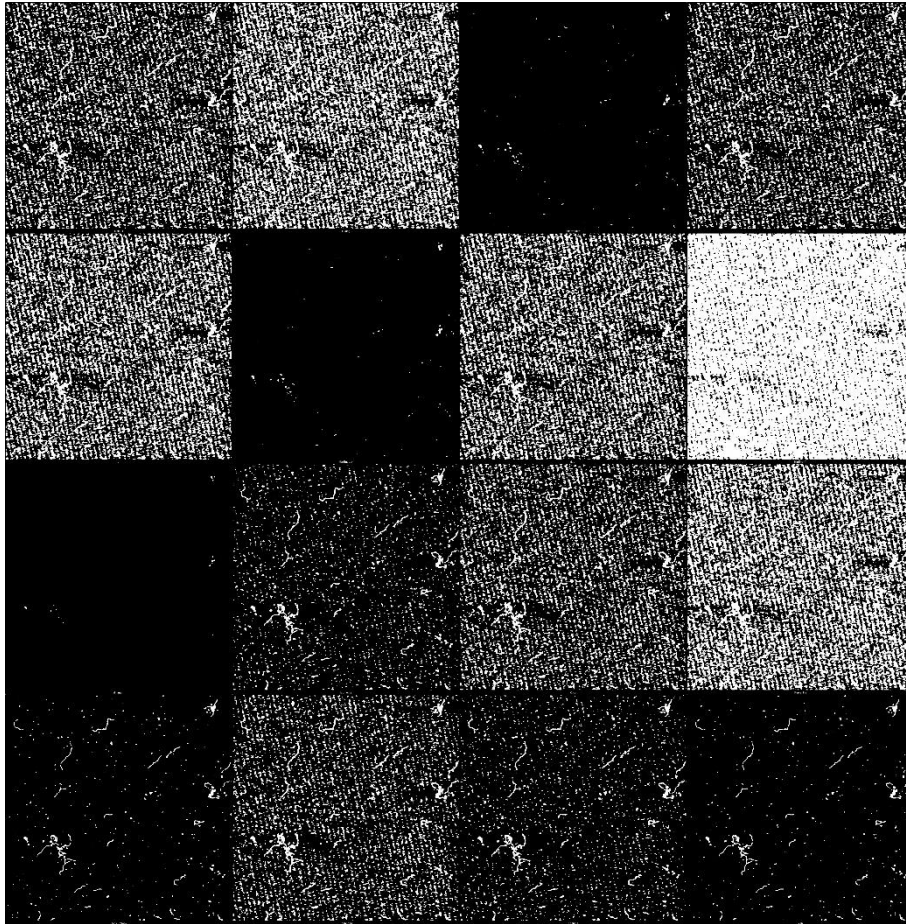
Damit möglichst nur DNA-Stränge und keine Hintergrundartefakte beim Trackingvorgang ausgewertet wurden, was die Rechenzeit verlängert und auch die Tracking-Ergebnisse von FIESTA unbrauchbar gemacht hätte, musste für die Auswertung in FIESTA eine Höhen-Schwelle (engl. threshold) mit Hilfe von ImageJ festgelegt werden (Abbildung 36). Der Threshold dient als eine selektiv gewählte Höhe, die den Hintergrund bestimmt. Alles was eine niedrigere Höhe als den Schwellenwert besitzt, wird ausgeblendet und nicht dargestellt. Initial diente die DNA-Höhenbestimmung aus Kapitel 3.2.1 hier als Anhaltspunkt. Mit ImageJ wurde nur der geeignete Threshold für die spätere FIESTA-Skelettierung bestimmt, dieser jedoch in ImageJ nicht angewandt, da man sonst mit der Anwendung des Thresholdes ein binäres Bild erzeugte, was für FIESTA nicht auswertbar war.



**Abbildung 36:** Thresholdbestimmung für FIESTA

Die Abbildung zeigt die Wahl einer geeigneten Hintergrundschwelle in ImageJ für die darauffolgende Skelettierung mit FIESTA. Alle rot markierten Strukturen liegen über der gewählten Threshold-Höhe. Der Wert wurde vermerkt, aber der entsprechende Threshold nicht in ImageJ, sondern erst anschließend in FIESTA angewandt.

Als Entscheidungshilfe für die Wahl des richtigen Thresholdes konnte der Auto-Threshold genutzt werden (Abbildung 37). Der Auto-Threshold ist eine für das Programm ImageJ entwickelte Zusatzfunktion, die automatisch alle in ImageJ verfügbaren Threshold-Typen auf das ausgewählte Bild anwendet. In unserem Fall hatte sich „Yen“, „Otsu“, „MaxEntropy“, und „Intermodes“ bewährt, wobei meist „Yen“ die beste Höhenschwelle für die DNA-Darstellung setzte und deshalb für alle Aufnahmen seine Verwendung fand.



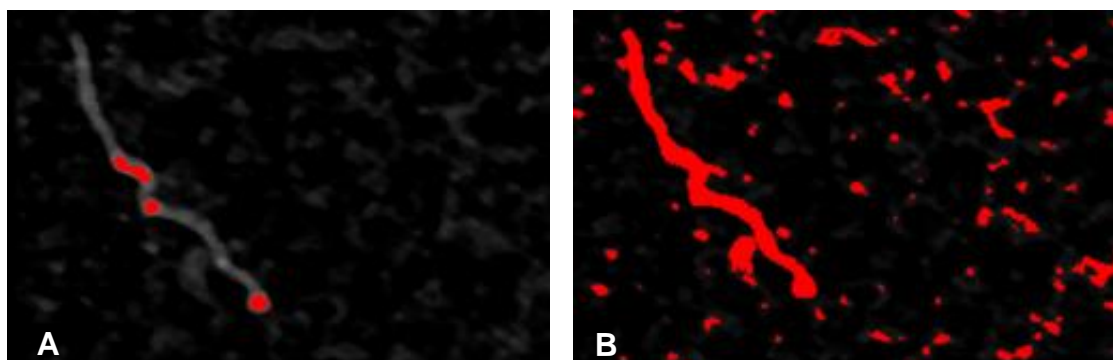
**Abbildung 37: Auto-Threshold**

Die Funktion des Auto-Threshold war für die Überprüfung der Eignung verschiedener Threshold-Einstellungen zur Bildauswertung sinnvoll. Aufzählung der verschiedenen Threshold-Modi von oben links nach unten rechts: Default, Huang, Intermodos, IsoData, Li, MaxEntropy, Mean, MinError, Minimum, Moments, Otsu, Percentile, RenyiEntropy, Shanbhag, Triangle, Yen. Der Yen-Threshold zeigte für die DNA-Auswertung die besten Ergebnisse.

Trotzdem war für einige Bilder eine leichte, manuelle Anpassung der Minimumhöhe (Cutoff) nötig. Dies resultierte teilweise aus nicht optimalen Aufnahmebedingungen, wie einer unebenen Glimmer-Oberfläche oder einem höheren Grad an Verunreinigungen auf der Oberfläche. Ein anderer Störfaktor war die Breite der abgebildeten DNA, die durch den Verschleiß des Tips bei Gebrauch zunahm (siehe auch Kapitel 3.2.1). Um den Aspekt der abnehmenden Tip-Schärfe und der DNA-Breite auf das Messergebnis zu untersuchen, wurden bei der Festlegung der Halbwertsbreite von jedem Bild mehrere DNA-Breiten vermessen.

Vorgehen: Es wurde über [Image]; [Adjust]; [Threshold]; [Yen] (Abbildung 72) eine Einstellungsschwelle gewählt, die 1,51 % der Datenpunkte in der Abbildung darstellte. Dabei wurden 98,49 % aller Bilddaten mit einer Signalintensität unterhalb der Schwellengrenze nicht dargestellt. Dies hatte zur Folge, dass sehr kleine Partikel ausgeblendet und das Hintergrundrauschen effektiv reduziert werden konnte. Der in ImageJ bestimmte Threshold wurde allerdings nicht tatsächlich in ImageJ angewendet, sondern erst bei der FIESTA-Skelettierung genutzt. Die maximale Signalintensität wird durch 255 Einheiten definiert, wobei 0 keinem Signal und 255 der maximalen Signalintensität entspricht. Bei der Anwendung des Thresholdes in ImageJ wird ein binäres Bild erzeugt. Ein binäres Bild kann aber nicht mehr mit FIESTA ausgewertet werden.

Allerdings traten beim Setzen eines geeigneten Thresholdes bei manchen Aufnahmen auch Schwierigkeiten auf, die eine sinnvolle Auswertung im Weiteren verhinderten. Das Problem schien hier nicht unbedingt die Breite der DNA zu sein, sie spielte zwar eine Rolle, in dem sie die Wahl des richtigen Thresholdes erschwerte, aber Ausschlusskriterium war meist viel mehr eine flache, breite DNA mit vielen Schmutzpartikeln. Dadurch war es unmöglich das Höhenminimum so zu wählen, dass die DNA komplett erhalten blieb und gleichzeitig das Hintergrundrauschen verschwand. Um die DNA in ihrer kompletten Länge markieren zu können, hätte der Threshold so tief sitzen müssen, dass er mit den Verschmutzungen, beziehungsweise mit dem Hintergrund verschmolzen wäre (Abbildung 38).

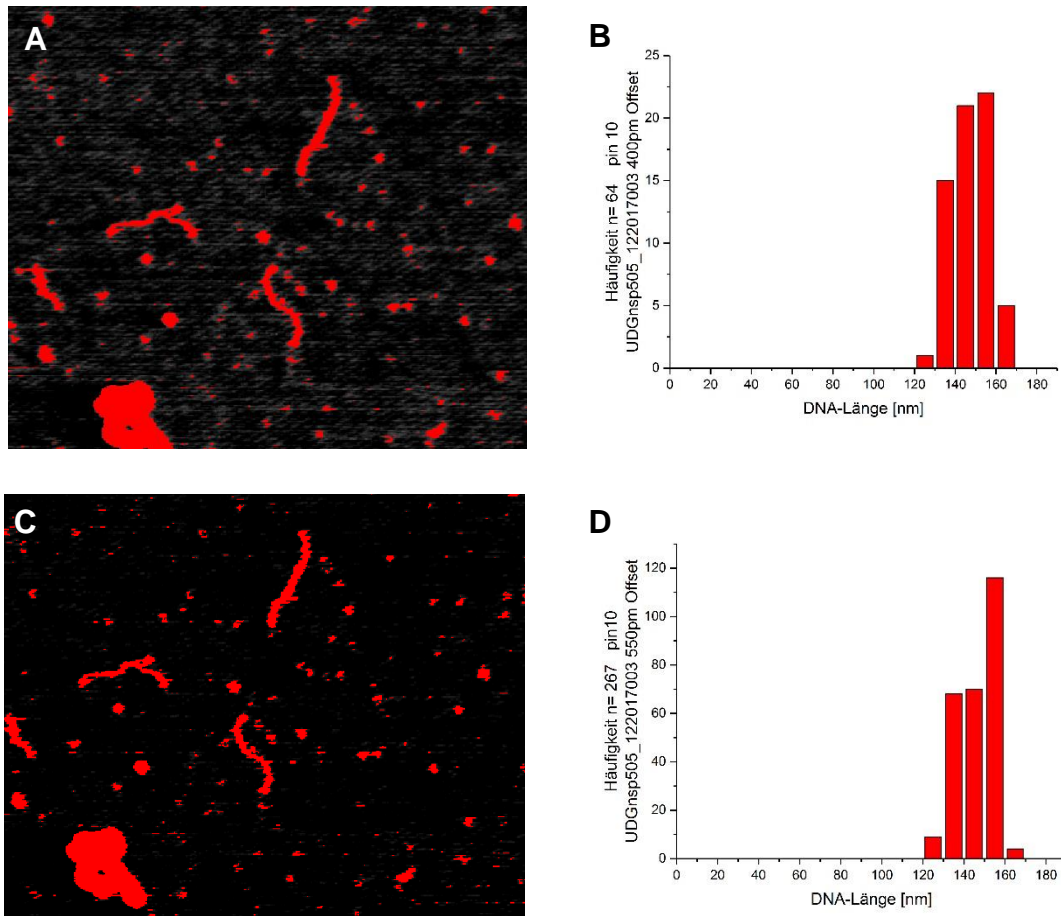


**Abbildung 38: Problem der Thresholdbestimmung**

In der Abbildung wird das Problem der passenden Hintergrundhöhe ersichtlich. Unter **A** ist der Yen-Threshold mit 0,01 % zu sehen, **B** zeigt den Versuch des manuellen Senkens der Höhe für eine Auswertung. Dieser führt aber zu keinem verwertbaren Ergebnis. In der Folge war ein solches Bild nicht auswertbar.

Um das oben geschilderte Problem der Auswertbarkeit qualitativ schlechterer Aufnahmen zu evaluieren, sollte als mögliche Ursache ein beim [.tiff]-Export zu hoch festgesetzter Hintergrund ausgeschlossen werden. Bereits bei dem Export der Aufnahme vom [.ibw]-Format in das [.tiff]-Format konnte der Hintergrund variiert werden, um eine optimale DNA-Darstellung zu ermöglichen. Um diesen Einfluss zu untersuchen, wurden Aufnahmen mit unterschiedlicher Hintergrundhöhe aus der AFM-Software exportiert und die Auswirkung auf die Verwendung des Thresholdes in ImageJ analysiert. Trotz der Verwendung eines unterschiedlich hohen Hintergrundes während des [.tiff]-Exportes, war es nicht möglich einen geeigneten Threshold zu definieren, der zuverlässig ausschließlich die DNA markierte. Bei einem Cutoff von 400 pm betrug der Yen-Threshold 72, aber trotz manueller Anpassung war nur eine sehr inhomogene Darstellung möglich, sodass eine Auswertung deutlich erschwert war. Bei einem Cutoff von 550 pm lag der Yen-Threshold bei 30, eine Auswertung war zwar besser möglich, aber die Hintergrundbeschaffenheit erschwerte eine Auswertung trotzdem. Der Hintergrund erschien deutlich dunkler, was auch durch den etwa um 50 % geringeren automatischen Threshold-Wert verdeutlicht wurde, trotzdem waren die DNA-Strukturen nicht deutlich definierter darzustellen (Abbildung 39). Hieraus ließ sich folgern, dass die Auswertung der Aufnahmen mit starkem Hintergrundrauschen durch Veränderung des Offset-Wertes bei dem [.tiff]-Export nicht verbessert werden konnte.



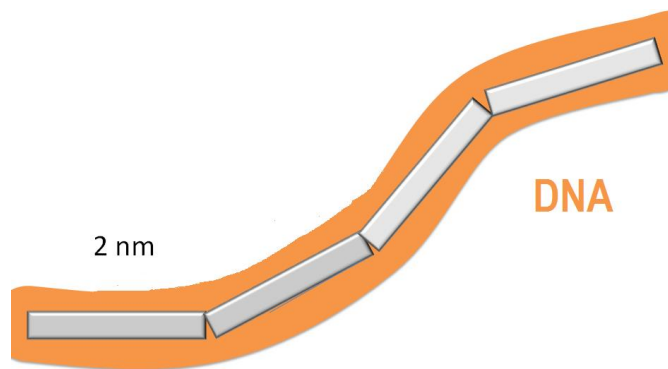


**Abbildung 39: Vergleich unterschiedlich hoher Offset-Einstellungen**

Die Darstellung der DNA bei 550 pm (**C**), statt mit 400 pm (**A**) war nicht deutlich schärfer, bzw. kontrastreicher, allerdings bezweckte der dunklere Hintergrund bei der automatischen Auswertung eine größere Ausbeute an skelettierbarer DNA im Vergleich zu dem niedrigeren Threshold, was sich deutlich in der Anzahl der Häufigkeiten (400 pm Cutoff: n = 64 (**B**); 550 pm Cutoff: n = 267(**D**)) bemerkbar machte. Der Einfluss auf die ermittelten Längen war gering.

### 3.2.6 Skelettierung der DNA

Das Programm FIESTA (*Fluorescent Image Evaluation Software for Tracking and Analysis*) ist eine auf MATLAB basierende Software von Ruhnow *et al.* [126] zur Analyse von bewegten Fluoreszenz-Aufnahmen. FIESTA (1.05.0005) segmentiert die DNA-Stränge basierend auf einem Gauß-Fit und dessen Halbwertsbreite (engl. full width at half maximum, FWHM). Zur Segmentierung werden einzelne Filamente mit einer Länge von 2 nm durch das errechnete Signalmaximum (die Mitte des DNA-Stranges) gelegt (Abbildung 40).

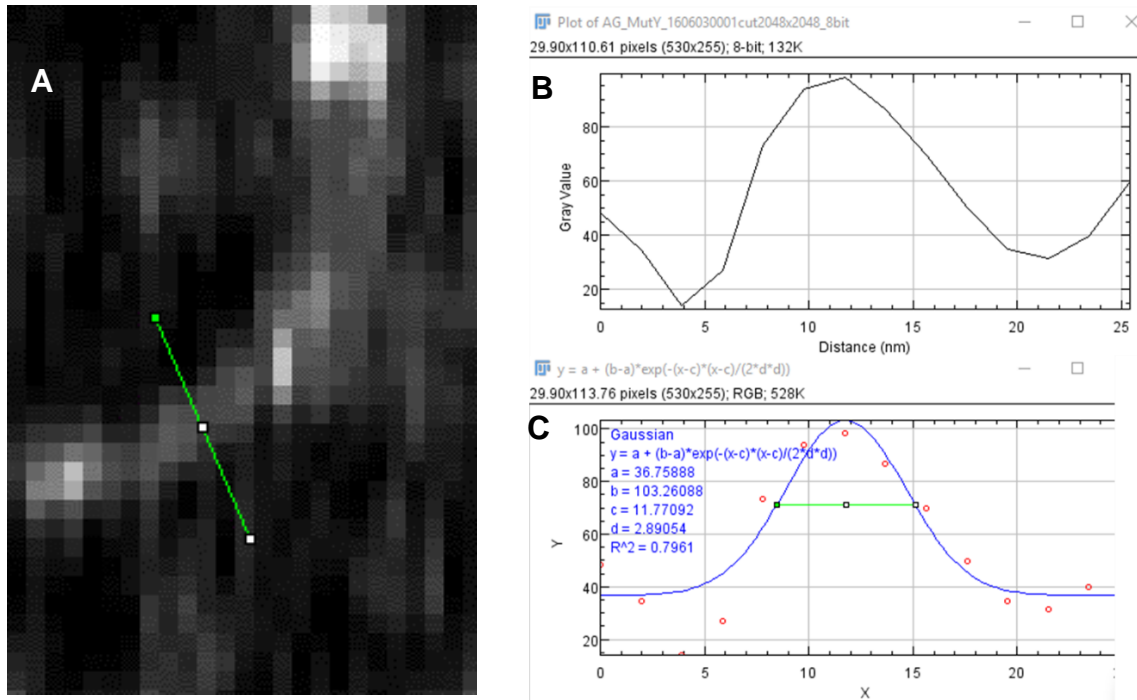


**Abbildung 40: Schematische Darstellung der Skelettierung**

Die DNA (orange) wird durch FIESTA mit 2 nm langen Segmenten (grau) skelettiert.

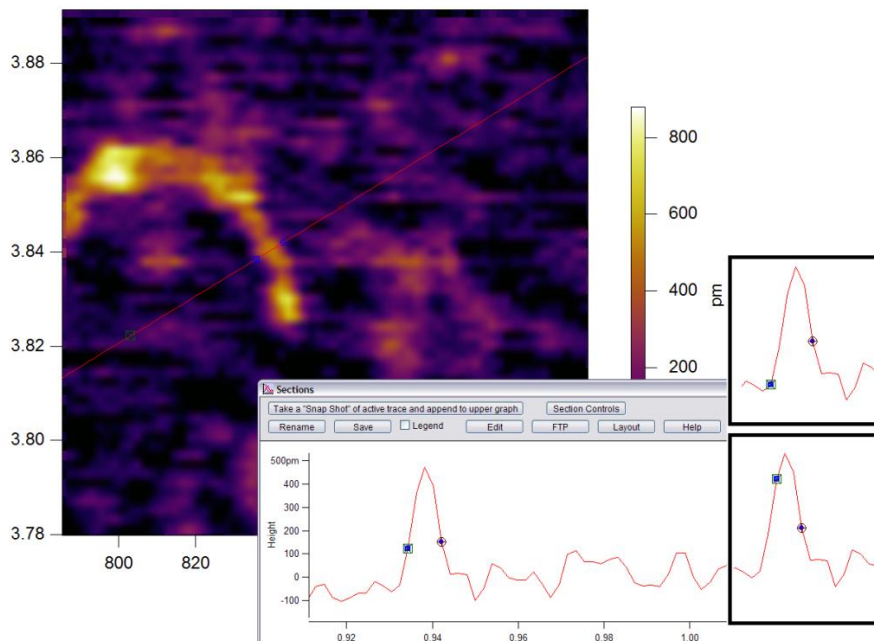
Zur Abschätzung der korrekten DNA-Halbwertsbreite wurden drei verschiedene Messmethoden miteinander verglichen. Die DNA-Breite wurde an der gleichen Stelle mit ImageJ (Abbildung 41), dem Section Tool der MFP Software (Abbildung 42) und dem „Line Scan“-Werkzeug von FIESTA (Abbildung 43) gemessen. Für alle drei Methoden ergaben sich vergleichbare Ergebnisse.

### 3 Ergebnisse



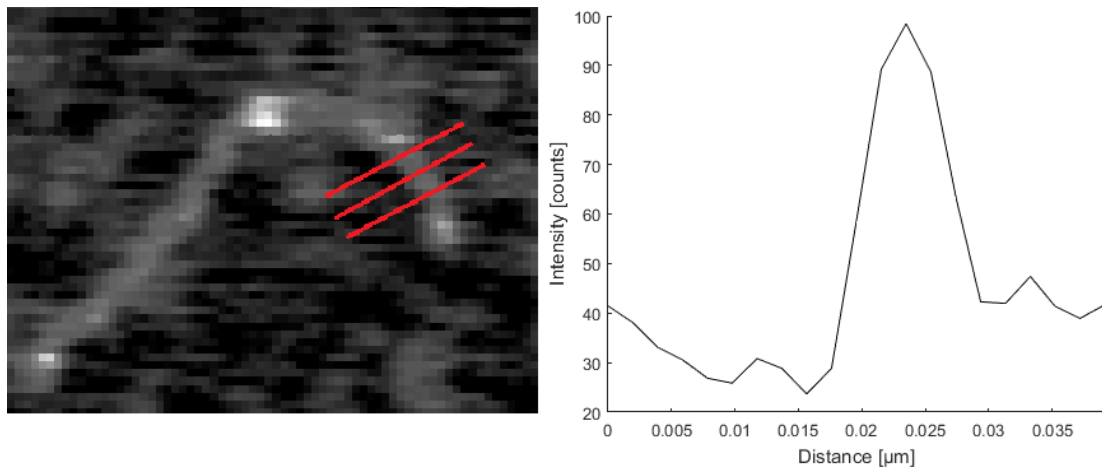
**Abbildung 41: Messung der DNA-Breite in ImageJ**

Die Messung der DNA-Breite erfolgte gemäß dem in den Methoden erläuterten Vorgehen. In **A** wurde die grüne Messlinie senkrecht zur DNA angelegt. Das Messergebnis des lokalen Höhenprofils ist in **B** zu sehen. **C** zeigt die Gaußanpassung des Messergebnisses aus **B**. Das Ergebnis der mittleren Breite [nm] nach der Gaußanpassung (grüne Messlinie in **C**) beträgt 7 nm.



**Abbildung 42: Messung der DNA-Breite mit dem Section-Tool**

Auch mit dem Section-Tool der Igor Pro 2 AFM Software wurde der Durchmesser an identischer Stelle gemessen. Allerdings bestand die Schwierigkeit darin, dass die Messpunkte nur in einem gewissen Mess-Intervall gesetzt werden konnten, das durch die Pixelauflösung bestimmt wurde. Es ergab sich eine DNA-Breite von 6 nm.



**Abbildung 43: Messung der DNA-Breite mit „Line Scan“ in FIESTA**

Zum Vergleich der verschiedenen Messmethoden wurde die DNA-Breite auch in FIESTA mit dem „Line Scan“ ermittelt. FIESTA erstellte ein Höhenhistogramm und errechnete mit einer Gauß-Anpassung eine Halbwertsbreite von 6 nm.

Die Halbwertsbreite wurde bei der automatischen Auswertung grundsätzlich mit dem „Line Scan“ in FIESTA bestimmt. Im Folgenden werden die für die Skeletierung in FIESTA benötigten Arbeitsschritte einzeln aufgeführt.

Vorgehen: Zunächst wurde das zuvor in ImageJ modifizierte Bild (Formatänderung und Medianfilterung) in FIESTA geladen (Abbildung 35). Nach dem Laden der Aufnahme wurden von FIESTA zwei verschiedene Parameter abgefragt. Für die Eingabe [Time difference] musste in unserem Fall keine Angabe gemacht werden, da es sich jeweils um ein einzelnes Bild handelte. Für die Bildpunktgröße [Pixel size] wurde je nach Bildgröße und Pixelauflösung beispielsweise 1,95 nm/Pixel für ein 8 µm x 8 µm großer Bildfläche mit 4096 x 4096 Pixeln eingetragen. Anschließend wurde am linksseitigen Bedienfeld von FIESTA auf den Off-Set-Mode gewechselt (nicht in Abbildung 44 dargestellt) und hierbei der in ImageJ mit dem Yen-Threshold ermittelte Wert eingestellt. Nach dem Zurückwechseln in den Grauskala-Modus wurde unter dem Menüpunkt [Tools] die Funktion [Line Scan] aktiviert und hiermit eine Linie senkrecht zur DNA-Achse eingezeichnet, um die Halbwertsbreite [FWHM] der DNA zu ermitteln (Abbildung 43). Nach dem Befehl [Estimate FWHM] und Anklicken des erzeugten Plottes mit dem Cursor wurde der gemessene FWHM-Wert ausgegeben. Dieser wurde durch [Apply to configuration] direkt in die Konfigurationseinstellungen von FIESTA für

die folgende Auswertung übernommen. Die allgemeinen Tracking-Einstellungen werden in der FIESTA-Bedienungsoberfläche „Configuration“ vorgenommen (Abbildung 44). So konnten unter [Check configuration] neben der mit FIESTA ermittelten Halbwertsbreite [FWHM] alle weiteren Einstellungen für die Segmentierungsaufgabe festgelegt werden (Abbildung 44): Die Angaben unter „General options“ wurden teilweise schon in den vorausgegangenen Arbeitsschritten festgelegt. So waren bei den [Stack Options] neben der [Pixel size] auch der Threshold und diesen betreffende Werte wie die [FWHM] voreingestellt. Als Darstellungsschwelle [Threshold] wurde auf den zu Beginn in ImageJ ermittelten Wert zurückgegriffen [average before]. Einzig bei [Tracking] war eine Einstellung vorzunehmen. Es sollten nur Filamente und keine Moleküle analysiert werden, weshalb [Tracking] von [only filaments] aktiviert wurde. Im Bereich „Tracking options“ wurden bei „Connecting“ und „Tracks“ keine Veränderungen der voreingestellten Werte vorgenommen, da diese Parameter nur für die Vermessung von Bewegungen bei Fluoreszenzserienbildern, wofür FIESTA ursprünglich konzipiert war, einer Einstellung bedurften. Als Referenzpunkt [Reference point] für die Skelettierung wurde die Bestimmung der Strangmitte [Center point] gewählt. FIESTA markierte den Mittelpunkt mit einem blauen Punkt (siehe Abbildung 45). Die Länge der Segmente, die den DNA-Strang skelettieren, war variabel. Um die Länge der einzelnen Segmente möglichst kurz zu wählen und so eine möglichst detaillierte Filamentierung zu gewährleisten, wurde bei der Einstellung der [Tracking options]: [Experimental options]; [Track especially curved filaments]; [box size] [25%] gewählt. Durch diese Einstellungsoption konnte die Größe der Anpassungsbox von  $8\sigma$  (100%) auf  $2\sigma$  (25%) reduziert werden. Dies erhöhte die Anzahl der angepassten Segmente entlang der Kontur des Filaments, indem es die Segmente auf eine Länge von 2 nm verkürzte.

**General options**

Stack: StackType: none

Stack Options: Pixel size: 1.95 nm Time difference: 0 ms Last Frame: 0 Parallel Processing: none

Tracking: Only Molecules:  Only Filaments:   Include fit data (Warning: approx. 50x-100x data size)

Threshold: Area: 30 pixel Height: 2 Fit(CoD): 0 FWHM(Est.): 500 nm Border Margin: 2 pixel

Filter for thresholding (will be used for relative/constant intensity thresholding):  none  average before  smooth after  filter background

---

**Tracking options**

**Molecules**

Connecting: max. Velocity: 1000 nm/s Verification Steps: 2

Weights: Position: 30 % Direction: 30 % Speed: 20 % Length: 20 %

Disregard filaments at the edge of the field of view

Reference Point:  Start Point  Center Point  End Point

Experimental Options: The following options are experimental and require more testing!

Correct for focus drift - different reference points for filament tips

Track especially curved filaments - reduce fit box size by: 25 % (25% - 100%)

Track dynamic filaments

---

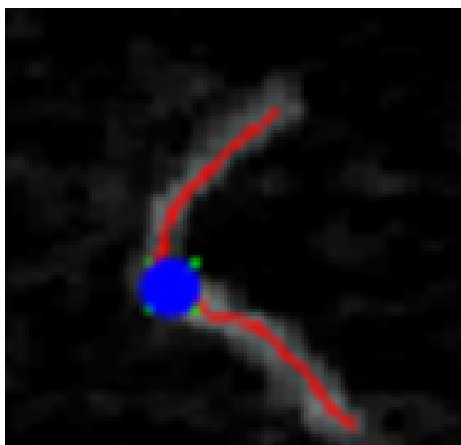
**Filaments**

Tracks: Minimum Length: 4 frames Maximum Break: 2 frames Maximum Angle: 45 °

**Abbildung 44: Bedienungsoberfläche von FIESTA**

Im Bedienfeld „FIESTA-Configurations“ wurden die Parameter für den Skelettierungsvorgang der DNA eingestellt.

FIESTA führte daraufhin an den abgebildeten DNA-Strukturen für jedes Segment eine 2D Gauß-Anpassung durch und legte die Skelettierungslinie mittig in die DNA-Fragmente. Wurden alle Einstellungen vorgenommen, konnten diese über [Apply] angewendet werden. Mit [Add stack to local queue] und nach Bestätigung der Warnmeldung über eine große Datenmenge mit [Add anyway] wurde die Arbeitsanweisung mit der gewählten Konfiguration zur Arbeitsmappe hinzugefügt. Nach Klicken von [Analyse local queue] begann FIESTA dann die lokale Arbeitsmappe gemäß den zuvor gewählten Tracking-Einstellungen abzuarbeiten. Die Tracking-Daten und Bilder wurden im [.mat]-Format innerhalb desselben Dateipfades, wie das Ausgangsbild gespeichert. Nach Beendigung der Auswertung wurden diese Dateien dann der Reihe nach mit [Data], [Load Tracks] und [Load Objects] geöffnet. Zur Darstellung der Ergebnisse wurde im Work-Panel unter [Data] und [Filaments] die Optionen [Show all filaments] und [Show filament positions] aktiviert. Nun erschienen im FIESTA-Bild die Trackinglinien und die Mittelpunkte der skelettierten DNA-Stränge (blaue Punkte). Um die Auswertung von MATLAB zu beschleunigen und zu vereinfachen, wurden Artefakte und überlagerte DNA-Stränge nach manuellem Markieren aus der Tracking-Liste gelöscht und die selektierten FIESTA-Tracks unter [Data], [Save all tracks as] mit demselben Dateinamen durch den Zusatz „selected“ nochmals abgespeichert (z.B. GAnew\_201706140002cut4096xy-median2px.mat als GAnew\_201706140002-cut4096xy-median2px.selected.mat).

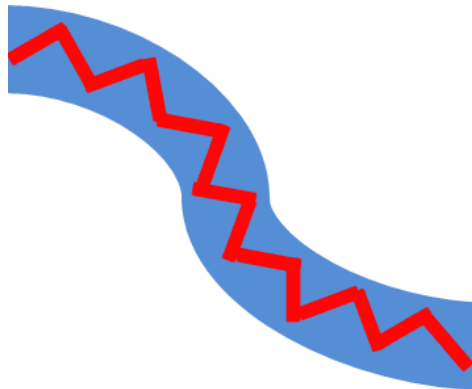


**Abbildung 45: DNA-Tracking mit FIESTA**

FIESTA skelettierte den DNA-Strang mit 2 nm langen Filamenten, wobei gemäß der Halbwertsbreite eine Linie durch die Mitte des Stranges gelegt wurde. Entsprechend der Einstellung „center point“ markierte FIESTA zusätzlich den Mittelpunkt durch einen blauen Punkt.

### 3.2.6.1 Optimierung des Skelettierungsprozesses

Bei Programmeinführung musste zunächst das Optimierungspotenzial der verschiedenen Parameter untersucht werden. So musste evaluiert werden, inwiefern das Ergebnis der Halbwertsbreitenbestimmung und das Ergebnis der Skelettierung miteinander korrelierten. Im Falle einer Verkleinerung des erwarteten FWHM kam es zu einer deutlich fehlerhaften Filamentierung. Ursächlich für diesen Effekt war, dass bei dieser Einstellung das Skelett nicht mehr zwingend mittig im DNA-Strang zu liegen kam, sondern zwischen den beiden Außenseiten eine leicht pendelnde Form annahm und mehr Segmente in die Aufnahme der DNA gelegt werden konnten.

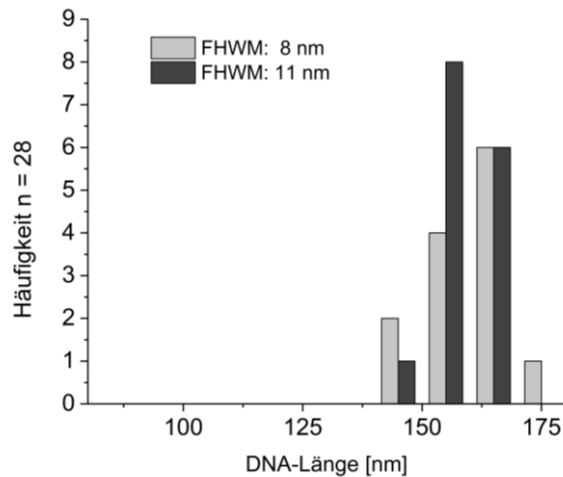


**Abbildung 46: Segmentierung bei kleiner Halbwertsbreite**

Wurde die Halbwertsbreite zu klein gewählt kam es zu einer schwingenden Segmentierung.

Der Versuch, ein Bild mit zwei geringgradig unterschiedlichen Halbwertsbreiten (FWHM von 8 und 11) auszuwerten, zeigte, dass die von FIESTA ermittelte Halbwertsbreite bereits zu optimalen Ergebnissen führte und evtl. maximal um einen Betrag von +2 geändert werden sollte. Der Einfluss der Auswertung mit unterschiedlicher Halbwertsbreite auf die ermittelte DNA-Länge ist in Abbildung 47 zu sehen.





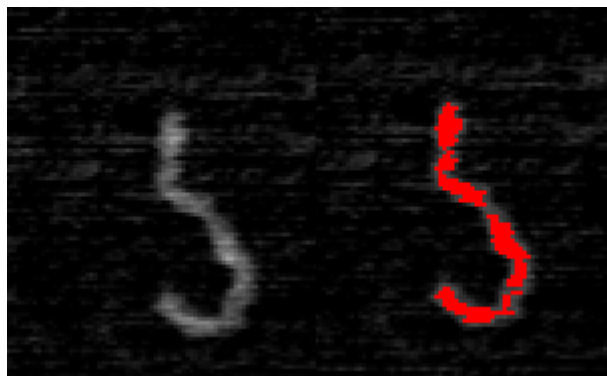
**Abbildung 47: Effekt der Halbwertsbreite auf die Länge**

Die kleinere Halbwertsbreite von 8 nm ergab insgesamt minimal längere DNA-Filamente als die 11 nm große Halbwertsbreite.

Die Wichtigkeit bei der DNA-Auswertung auf eine korrekte Länge zu achten, rührte daher, da nur so sichergestellt werden konnte, dass sich der zu untersuchende Basenschaden auch wirklich bei 50 % der Länge befand. Eine Winkelmessung am nicht Schadensort sollte so möglichst ausgeschlossen werden. Für die adäquate DNA-Länge von 505 bp, die rechnerisch einer Länge von 171 nm entsprach, wurde von FIESTA durchschnittlich stets eine Länge von 150 nm ermittelt und hierfür eine Längensreuung von ca.  $2\sigma$  in Form von  $\pm 20$  nm gemessen.

Ein weiteres Problem ergab sich bei stark gebogener DNA, die als jeweils zwei separate Stränge erkannt wurde. Ebenso wurden verschmutzte DNA-Stränge an der Verschmutzungsstelle als getrennte Filamente erkannt. Am DNA-Strang befindliche endständige Strukturen, die entweder Protein-Agglomerate oder Verschmutzungen darstellten, wurden von FIESTA häufig als eigenständiges Filament oder Molekül erkannt und mussten aussortiert werden. Um die Schwierigkeiten der Skelettierung zu beheben, wurde versucht bei stark gebogenen DNA-Strängen einen anderen Threshold zu wählen. Um die Eignung aller verfügbaren Threshold-Modi zu testen, wurde in ImageJ erneut die Funktion Auto-Threshold aus Abbildung 37 (Kapitel 3.2.5) verwendet. Außerdem wurde versucht, den Threshold-Wert generell zehn Einheiten unter den Yen-Threshold zu legen, um so die Kontinuität der DNA-Stränge zu verbessern. Allerdings führte dies, wie

bereits zuvor in Kapitel 3.2.5 beschrieben, zur stärkeren Verschmelzung des Hintergrundrauschens mit der DNA. Erhöhte man den Threshold, konnten einige DNA-Stränge beim Skelettierungsprozess nicht mehr zusammenhängend erkannt werden, da sich der Durchmesser an Abschnitten der DNA so stark verringerte, dass diese nicht mehr als ein einziges Filament dargestellt wurden. Auch dies stellte somit keine befriedigende Lösung für eine bessere Auswertung dar (Abbildung 48). Aus diesem Grund waren für die hier vorgestellte automatisierte Winkelmessung AFM-Aufnahmen mit relativ geringem Hintergrundrauschen und scharfen, hohen DNA-Strukturen von Vorteil.



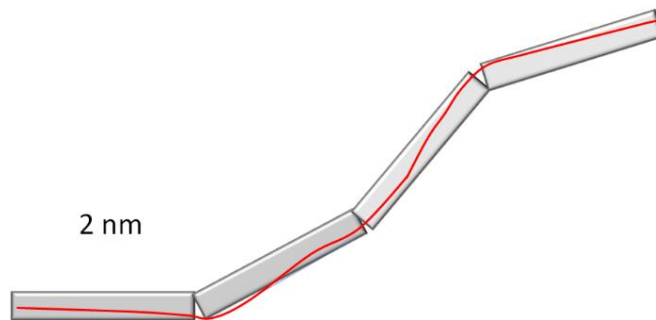
**Abbildung 48: Problemstellung bei gebogener DNA**

Bei gebogenen DNA-Strukturen konnte bei der Wahl des adäquaten Thresholds häufig gerade im Bereich der Biegung keine durchgehende Markierung (rot) erreicht werden.

### 3.2.7 Automatisierte Biegewinkelmessung bei 50 % DNA-Länge

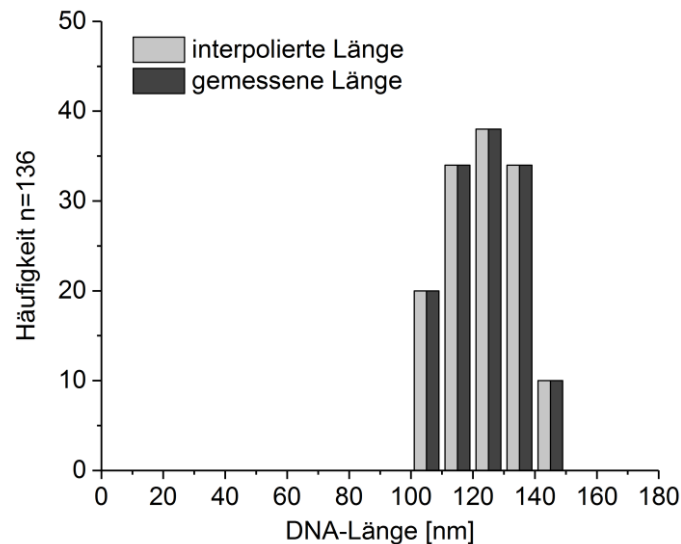
Zur Auswertung der mit FIESTA erzeugten Skelettierungsdaten diente das MATLAB-Programm `DNASegmentAnalysis_20171211.m` (Hannah Heil). Es wurden der Reihe nach die FIESTA-Tracks und die selektierten FIESTA-Tracks geladen, zusätzlich wurden die Datenpunkte mit der AFM-Aufnahme überlagert. MATLAB erstellte ein Längenhistogramm aus den FIESTA-Daten und fragte anschließend nach der zu erwarteten DNA-Länge [expected DNA-length] und der gewünschten Segmentierungslänge [Curvature Query Points]. Entsprechend der erwarteten Länge wurden die gesamten Filamentdaten gefiltert, sodass bei der weiteren Auswertung nur Filamente mit einer Abweichung von weniger als zwei Standardabweichungen von der erwarteten Länge berücksichtigt wurden. MATLAB

fürte zusätzlich an der FIESTA-Skelettierung (2 nm lange Filamente) eine Polynomzug (Spline)-Interpolation von 0,1 nm durch, um skelettierungsbedingte „Kanten“ zu egalisieren. Schematisch ist dies in Abbildung 49 dargestellt. Erwartungsgemäß waren die Längen der DNA-Stränge nach der Interpolation leicht vergrößert, was sich aber maximal auf die erste Dezimale auswirkte und deshalb vernachlässigt werden konnte, da sich kein relevanter Einfluss auf die Ergebnisse der Längenmessung ergab. Ein Vergleich der Filamentlängen vor und nach der Interpolation wird in Abbildung 50 gezeigt.



**Abbildung 49: DNA-Längen-Interpolation mittels Polynomzug**

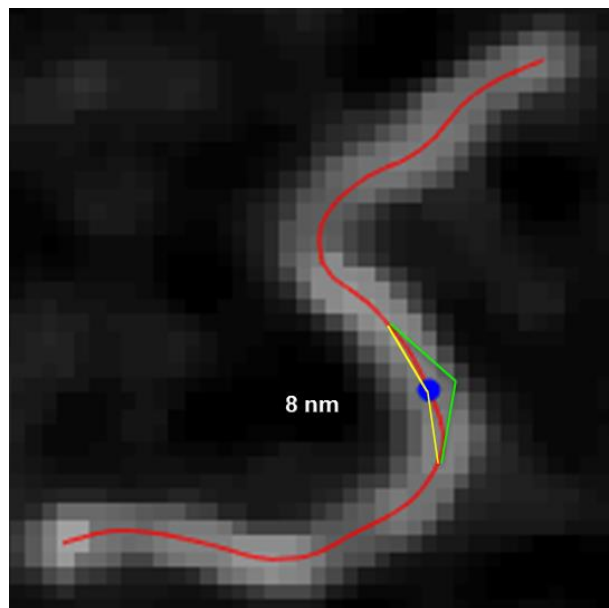
Die von FIESTA erzeugten Filamente bestanden aus 2 nm langen Segmenten (grau), welche in MATLAB durch einen Polynomzug mit einer Länge von 0,1 nm interpoliert wurden (rote Linie).



**Abbildung 50: Vergleich der Filamentlängen vor und nach der Interpolation**

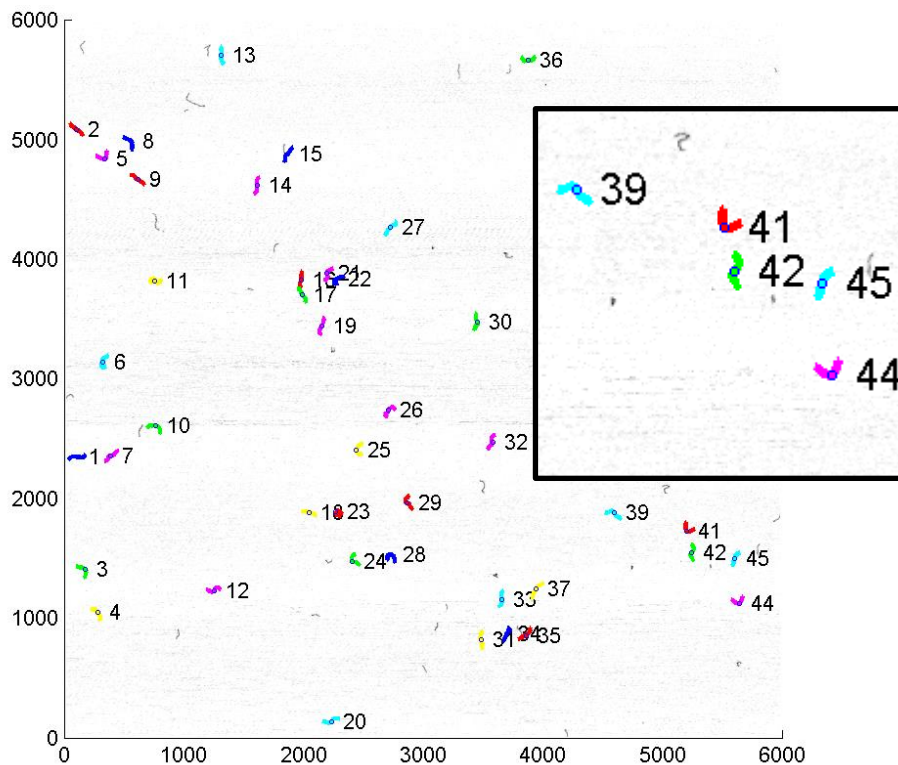
Gezeigt ist in grau die Länge der Filamente gemäß der FIESTA-Skelettierung und in schwarz die Länge der Filamente nach der MATLAB-Interpolation. Der Unterschied ist marginal und beschränkt sich maximal auf die erste Dezimale.

MATLAB nutzte dann die Skelettierungsdaten von FIESTA, um an der DNA-Mitte, das heißt bei 50 % der DNA-Länge, den aus den Tangenten oder Sekanten gebildeten Winkel zu messen (Abbildung 51). Die Länge der Winkeltangenten oder -sekanten konnte durch Eingabe der Abfragepunkte (engl. query points) frei gewählt werden. Für die Biegewinkelmessungen an DNA wurden in dieser Arbeit verschiedene Segmentlängen gewählt. Das Programm berücksichtigte bei der Segmentierungs-Berechnung keine DNA-Stränge, die kürzer als 120 nm waren, oder nicht selektierte Tracks. Mit [run] wurde die Analyse gestartet. Nach Beendigung der Auswertung wurde von MATLAB für jede gewählte Segmentlänge je ein DNA-Winkel-Histogramm und eine [.txt]-Datei ausgegeben, welche die entsprechenden Messergebnisse beinhaltete. MATLAB erstellte zur Veranschaulichung zusätzlich einen Objektplan der ausgewerteten DNA-Stränge (Abbildung 52). Die [.txt]-Datei konnte mit dem Import-Tool in Origin eingepflegt werden. Zusätzlich war zur Biegewinkelanalyse in Origin noch die Subtraktion  $\beta = 180 - \alpha$  nötig.



**Abbildung 51: Vergleichende Darstellung der Tangenten- und Sekantenmethode**

Veranschaulicht werden die Effekte unterschiedlicher Segmentlängen sowie die Unterschiede zwischen der Tangenten- und Sekantenmethode. Gezeigt wird die Originalaufnahme eines DNA-Stranges nach der FIESTA-Skelettierung mit der vom Programm generierten roten Filamentierungslinie, sowie blauem DNA-Mittelpunkt. Während die Tangenten (grüne Linien) die Filamentierungslinie (rot) jeweils nur berühren, schneiden die Sekanten (gelbe Linien) die Filamentierungslinie an je zwei Stellen, nämlich im gewählten Segmentabstand und zusätzlich im Mittelpunkt (blau). Die Tangenten schneiden sich je nach Steigung der gewählten Segmentlänge außerhalb des DNA-Mittelpunktes. Lediglich im Falle einer geraden DNA würde der Schnittpunkt mit dem DNA-Mittelpunkt übereinstimmen.



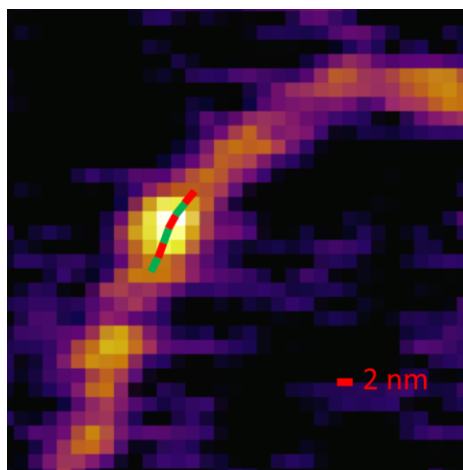
**Abbildung 52: MATLAB-Ergebnis**

MATLAB erstellte neben einer Ergebnistabelle als [.mat]-Datei auch einen Objektplan der ausgewerteten DNA-Stränge. Die DNA wurde jeweils bunt mit der entsprechenden Filamentnummer aus FIESTA versehen und der Mittelpunkt durch einen schwarzen Kreis markiert.

### 3.2.7.1 Vergleich verschiedener Segmentlängen

Das Programm FIESTA segmentiert die DNA-Stränge, indem es einen Gauß-Fit durchführt und nach Bestimmung der FWHM einzelne Filamente mit einer Länge von 2 nm in das errechnete Signalmaximum legt (Abbildung 53). Bei Einführung der automatisierten Methode wurde mit dem Programm ImageJ die typische Messgeradenlänge für die manuelle DNA-Biegewinkelmessungen abgeschätzt. Basierend auf diesen Ergebnissen wurde neben kleinen Längen (2 nm, 4 nm) und der manuellen Messung ähnlichen Längen (6 nm, 8 nm) auch große Tangentenlängen (15 nm und 20 nm) mit der automatisierten Methode getestet. Bei einer Auflösung von 1,95 nm/Pixel entsprach eine Tangentenlänge von 2 nm gerade der Größe eines Pixels. Die Biegewinkelmessung mit einer Tangentenlänge entsprechend der Bildauflösung konnte mit einer Verteilung gegen 0° nur Hintergrundrauschen darstellen. Auch bei 4 nm langen Tangenten wurde noch mit einem erhöhten Beitrag durch Hintergrundrauschen gerechnet. Im Gegensatz dazu

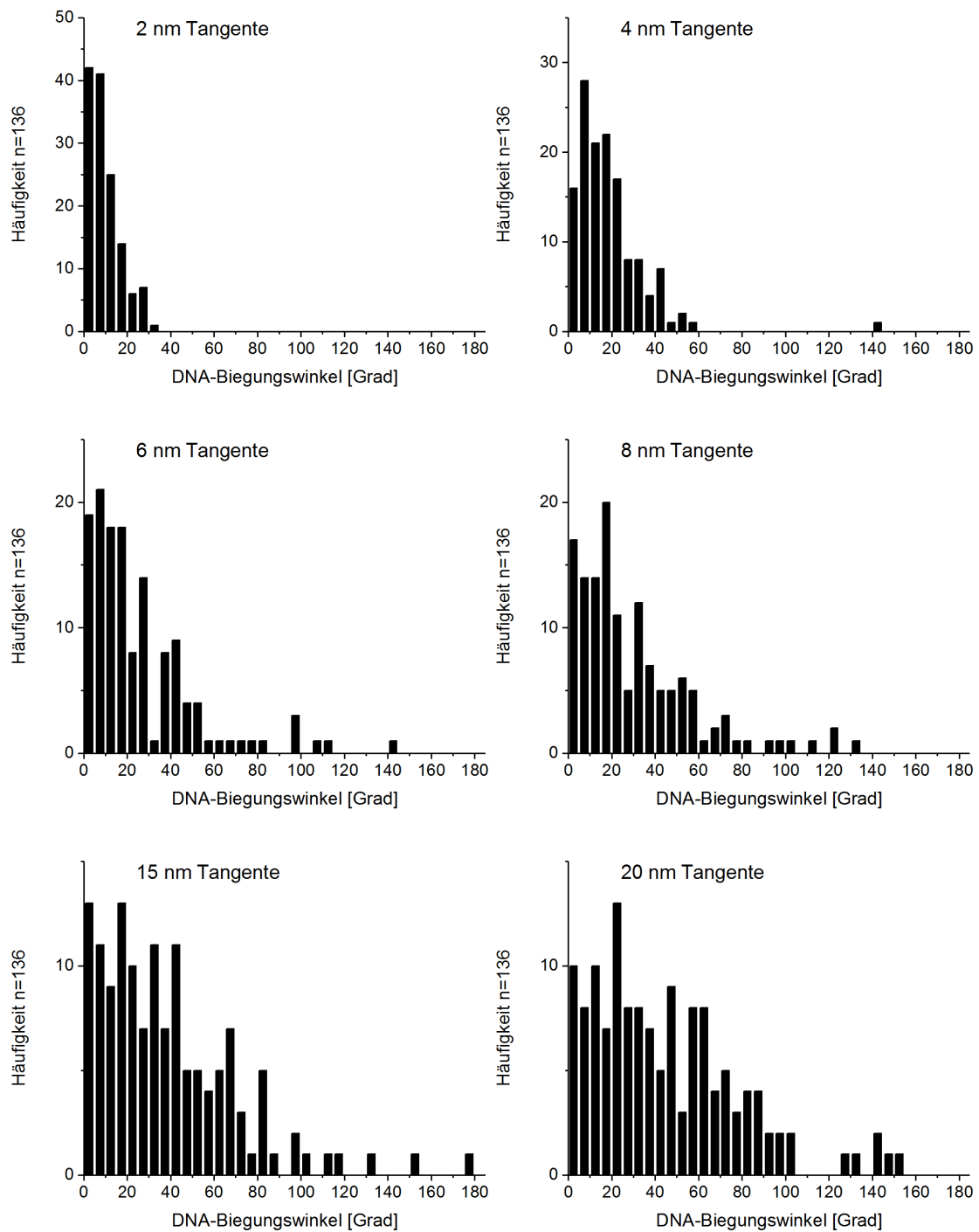
zeigte sich bei einer Tangentenlänge von 15 nm und 20 nm eine sehr breite Verteilung, sodass die lokale Biegung am DNA-Schaden durch die Überlagerung von sequenzabhängigen Biegungen der ungeschädigten DNA unkenntlich wurde. Da die automatisierte Messmethode auch für die Vermessung Protein-induzierter DNA-Biegewinkel ihre Anwendung finden sollte, wurden zur Bestimmung der hierfür benötigten Segmentlänge der Proteindurchmesser von freien Protein-Peaks und an DNA gebundenen Peaks gemessen. Es ergab sich exemplarisch für die Glykosylase MutY ein Proteindurchmesser von 6 nm bis 12 nm, wobei der Durchschnitt bei 7,2 nm lag. Allerdings variierte dieser Wert von Bild zu Bild, da der Proteindurchmesser vom Verschleiß der Messsonde und dem somit größer werdenden Messsondendurchmesser abhing. Während eine Länge von 4 nm als zu klein erachtet wurde, und eine 6 nm lange Messgerade den Protein-Peak nur geringfügig überragte, erschien eine Messgeradenlänge von 8 nm am zuverlässigsten. Im Folgenden wurde die Segmentlänge entsprechend sowohl bei der Winkelmessung an alleiniger DNA als auch an DNA-Protein-Komplexen für eine bessere Vergleichbarkeit einheitlich auf 8 nm Länge festgelegt. In Abbildung 54 werden Biegewinkelmessungen am Basenschaden G:A mit verschiedenen langen Messgeraden (2 nm, 4 nm, 6 nm, 8 nm, 15 nm, 20 nm) dargestellt. Aufgrund der geringeren Anzahl (Häufigkeit  $n = 136$ ) an Daten wurde die Bin-Größe von 3,5 auf 5 erhöht.



**Abbildung 53: Skelettierung der DNA mit daran gebundenem Protein**

Die Glykosylase MutY war an ungeschädigter DNA gebunden. Die Filamente mit einer Länge von 2 nm wurden farbig eingezeichnet (grün und rot). Je nach Proteindurchmesser, erschien eine Segmentlänge von 4 nm als unsicher, da die Segmente hierbei kaum über den Proteindurchmesser reichten. Eine Segmentlänge von 8 nm war also vorzuziehen.

### 3 Ergebnisse



**Abbildung 54: Vergleich der verschiedenen Segmentlängen**

Die Biegewinkelverteilung des Basenschadens G:A stellt beispielhaft die Verwendung von verschiedenen langen Segmenten in der automatisierten Methode dar. Bei der Verwendung einer Messgeradenlänge entsprechend der Auflösung von 2 nm konnte keine schadensspezifische Biegung erfasst werden. Auch bei 4 nm Länge musste mit einer Beeinflussung durch Hintergrundrauschen gerechnet werden. Für die Messung von spezifischen Biegewinkeln an Basenschäden und den von Proteinen an DNA erzeugten Verzerrungen wurde eine Tangentenlänge von 8 nm favorisiert. Allgemein konnte eine Verbreiterung der Biegewinkelverteilung mit zunehmender Messgeradenlänge erkannt werden. Dies konnte gerade bei einer Länge von 15 nm und 20 nm festgestellt werden.

### 3.3 Biegewinkel von DNA-Strängen

Nach der Etablierung des Verfahrens zur automatischen Biegewinkelmessung wurde dieses Verfahren sowohl im Tangenten- als auch im Sekantenverfahren an verschiedenen DNA-Substraten angewandt. Vermessen wurden verschiedene 505 bp lange DNA-Schadenssubstrate, die zur Erleichterung des Messens ihren Schaden jeweils bei 50 % ihrer Länge besaßen. Außerdem wurde zum Vergleich auch ungeschädigte DNA mit der gleichen Länge verwendet. Eine Klassenbreite (bin-size) von  $3,5^\circ$  erwies sich als sehr aufschlussreich, da durch das mäßige „Überbinnen“ auch kleine Biegewinkel ( $<10^\circ$ ) gut dargestellt werden konnten. Wie in Kapitel 3.2.7.1 erläutert, werden im Folgenden nur die Biegewinkelverteilungen für 4 nm und 8 nm lange Segmente gezeigt. Eine Zusammenstellung aller DNA-Biegewinkel befindet sich in der folgenden Tabelle 9. Neben den mit der automatischen Messsoftware ermittelten Biegewinkelwerte sind auch die entsprechenden manuell gemessenen Ergebnisse aufgeführt.



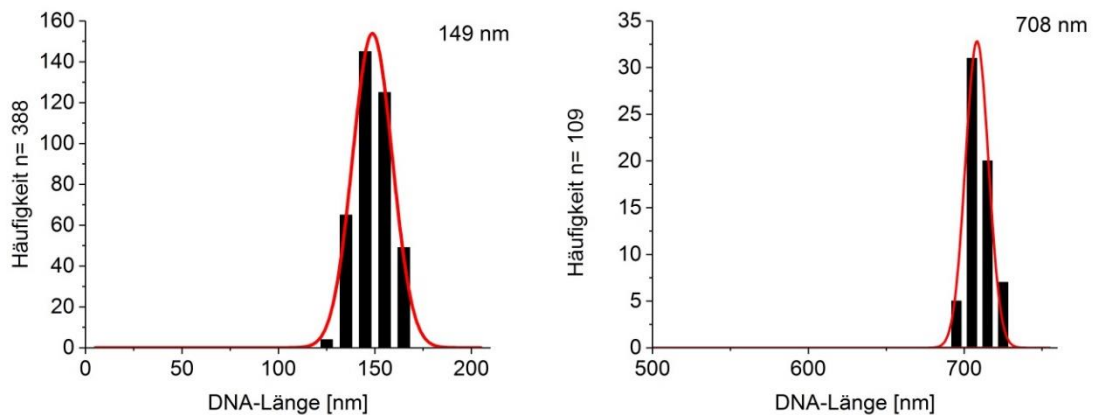
### 3 Ergebnisse

**Tabelle 4: Übersicht der DNA-Biegewinkel**

		Messverfahren				
		manuell		automatisierte Messsoftware (T = Tangente; S = Sekante)		
		Lisa Hirsch	Eigene Arbeit	4 nm	8 nm	
Substrat	Länge	Biegewinkel ( $\pm 1\sigma$ )				
nsp DNA	2224 bp	$0^\circ \pm 9^\circ$		T	$6^\circ \pm 13^\circ$	$6^\circ \pm 21^\circ$
		(n = 506)		(n = 109)		
	505 bp	$0^\circ \pm 5^\circ$	$0^\circ \pm 6^\circ$	T	$5^\circ \pm 16^\circ$	$5^\circ \pm 15^\circ$ $37^\circ \pm 14^\circ$
		(n = 271)	(n = 50)	S	$0^\circ \pm 10^\circ$	$0^\circ \pm 17^\circ$
		(n = 432)				
G:A	505 bp	$22^\circ \pm 16^\circ$	$10^\circ \pm 15^\circ$	T	$6^\circ \pm 14^\circ$	$2^\circ \pm 64^\circ$ $15^\circ \pm 4^\circ$
				S	$0^\circ \pm 8^\circ$	$5^\circ \pm 15^\circ$
		(n = 300)	(n = 100)	(n = 630)		
oxoG:A	505 bp	$24^\circ \pm 13^\circ$	$15^\circ \pm 9^\circ$	T	$5^\circ \pm 13^\circ$	$0^\circ \pm 3^\circ$ $15^\circ \pm 18^\circ$
				S	$0^\circ \pm 9^\circ$	$5^\circ \pm 14^\circ$
		(n = 669)	(n = 100)	(n = 877)		
oxoG:C	505 bp	$0^\circ \pm 5^\circ$	$10^\circ \pm 7^\circ$	T	$11^\circ \pm 14^\circ$	$0^\circ \pm 3^\circ$ $27^\circ \pm 5^\circ$ $11^\circ \pm 4^\circ$ $30^\circ \pm 15^\circ$
		$18^\circ \pm 13^\circ$		S	$0^\circ \pm 10^\circ$	$5^\circ \pm 9^\circ$ $10^\circ \pm 8^\circ$ $27^\circ \pm 9^\circ$
		(n = 288)	(n = 100)	(n = 62)		
				(n = 62)		
$\epsilon$ A:T	505 bp	$15^\circ \pm 10^\circ$	$13^\circ \pm 8^\circ$	T	$5^\circ \pm 20^\circ$	$0^\circ \pm 26^\circ$
				S	$0^\circ \pm 10^\circ$	$0^\circ \pm 17^\circ$
		(n = 304)	(n = 50)	(n = 648)		

### 3.3.1 DNA-Biegewinkel von ungeschädigter DNA

Es handelte sich um die Darstellung von vereinigten Messergebnissen für ungeschädigte DNA. Ausgewertet wurden drei Experimente. Drei Aufnahmen vom 26.04.2017, drei Aufnahmen vom 29.03.2017 und zwei Aufnahmen vom 09.12.2017. Die Länge der 505 bp langen DNA betrug  $149 \text{ nm} \pm 21 \text{ nm}$ , die Länge 2224 bp langer DNA ergab  $708 \text{ nm} \pm 14 \text{ nm}$ . Für die Biegewinkelmessung an ungeschädigter 505 bp langer DNA ergab sich mit dem Tangenten-Messverfahren ein Biegungswinkel von  $5^\circ \pm 16^\circ$  bei einer Segmentlänge von 4 nm. Bei 8 nm Segmentlänge zeigte sich sowohl ein Wert von  $5^\circ \pm 15^\circ$  als auch einen Wert von  $37^\circ \pm 14^\circ$ . Im Gegensatz zu den Messergebnissen der Tangente ergab sich bei der Auswertung mit der Sekantenmethode sowohl für 4 nm ( $0^\circ \pm 10^\circ$ ) lange, als auch für 8 nm ( $0^\circ \pm 17^\circ$ ) lange Segmente ein Biegewinkel von  $0^\circ$ . Zusätzlich wurde an der 50 %-Position von ungeschädigter 2224 bp langer DNA ein Biegungswinkel von jeweils  $6^\circ$  (4 nm  $6^\circ \pm 13^\circ$  und 8 nm  $6^\circ \pm 21^\circ$ ) ermittelt.



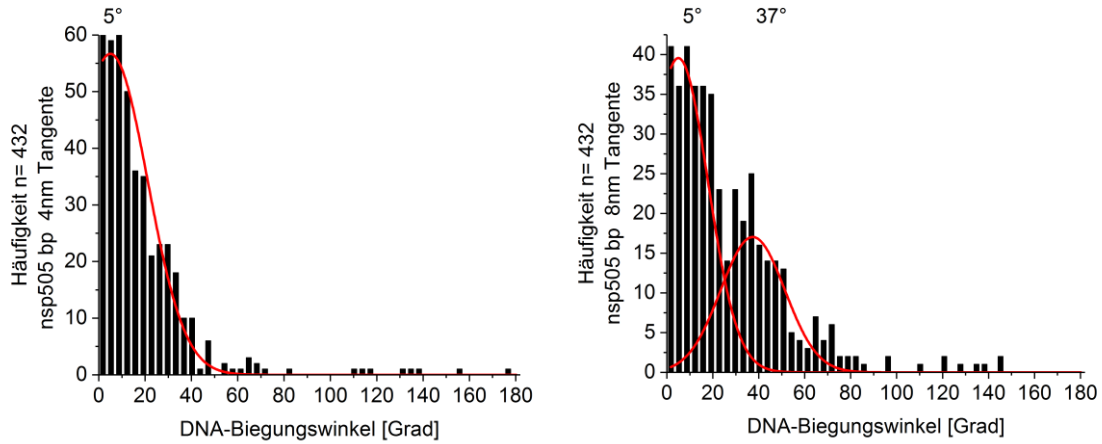
Länge von 505 bp nsp DNA

Länge von 2224 bp nsp DNA

**Abbildung 55: Längenverteilung von nsp DNA**

Die Abbildung verdeutlicht die gemessene Länge von 505 bp ( $149 \text{ nm} \pm 21 \text{ nm}$ ) und 2224 bp ( $708 \text{ nm} \pm 14 \text{ nm}$ ) langen ungeschädigten DNA-Strängen.

### 3 Ergebnisse

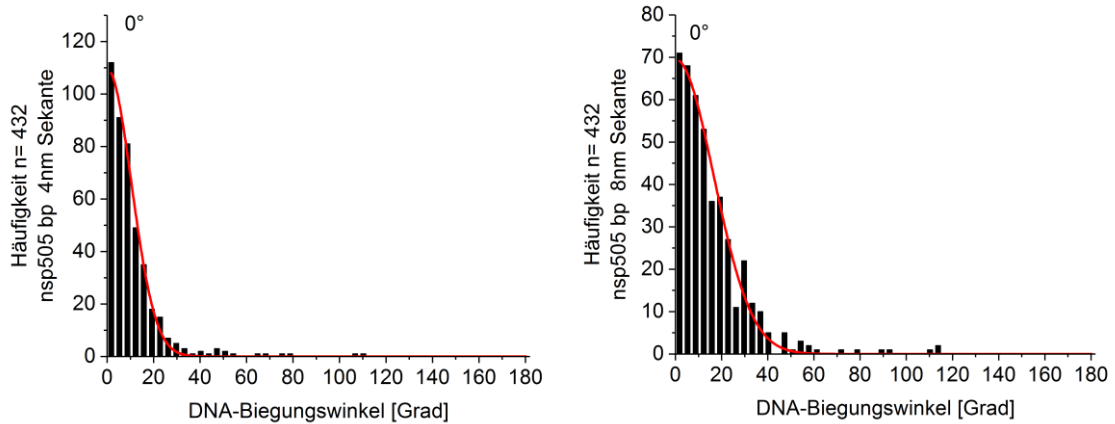


nsp DNA 4 nm (505 bp)

nsp DNA 8 nm (505 bp)

**Abbildung 56: Biegewinkelverteilung von nsp DNA (505 bp) mit der Tangentenmethode**

Die Auswertung bei 4 nm Segmentlänge ergab einen Biegungswinkel von  $5^\circ \pm 16^\circ$ , bei 8 nm Segmentlänge  $5^\circ \pm 15^\circ$  und  $37^\circ \pm 14^\circ$ . Die gemessenen Biegewinkel weichen sowohl bei einer Segmentlänge von 4 nm als auch bei 8 nm Segmentlänge deutlich von  $0^\circ$  ab.

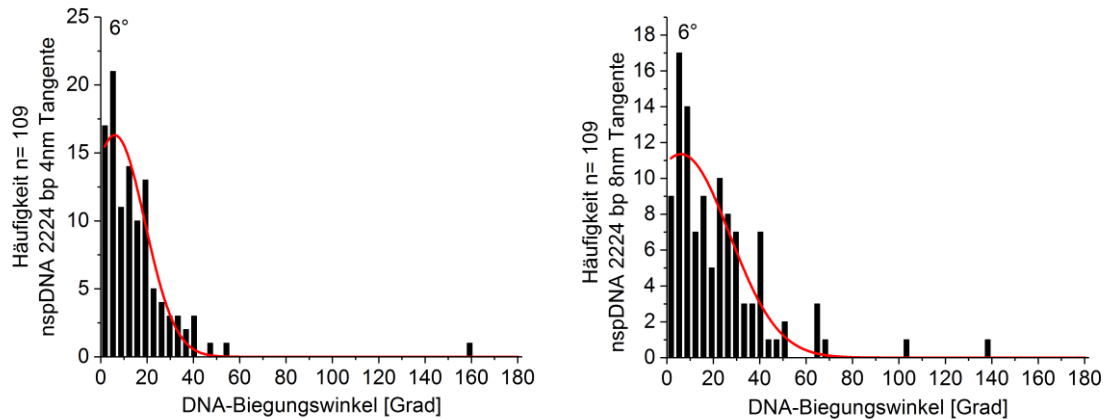


nsp DNA 4 nm (505 bp)

nsp DNA 8 nm (505 bp)

**Abbildung 57: Biegewinkelverteilung von nsp DNA (505 bp) mit der Sekantenmethode**

Die Auswertung der ungeschädigten 505 bp langen DNA mit der Sekantenmethode zeigte bei 4 nm langen Segmenten einen Biegungswinkel von  $0^\circ \pm 10^\circ$  und bei 8 nm langen Segmenten einen Winkel von  $0^\circ \pm 17^\circ$ .



nsp DNA 4 nm (2224 bp)

nsp DNA 8 nm (2224 bp)

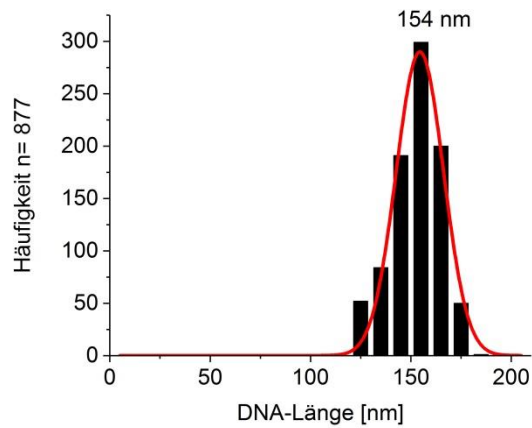
**Abbildung 58: Biegewinkelverteilung von nsp DNA (2224 bp) mit der Tangentenmethode**

Die 2224 bp lange ungeschädigte DNA wies eine signifikante Biegung von  $6^\circ$  sowohl bei 4 nm als auch bei 8 nm Segmentlänge  $\pm 13^\circ$  auf.

### 3.3.2 DNA-Biegewinkel von oxoG:A

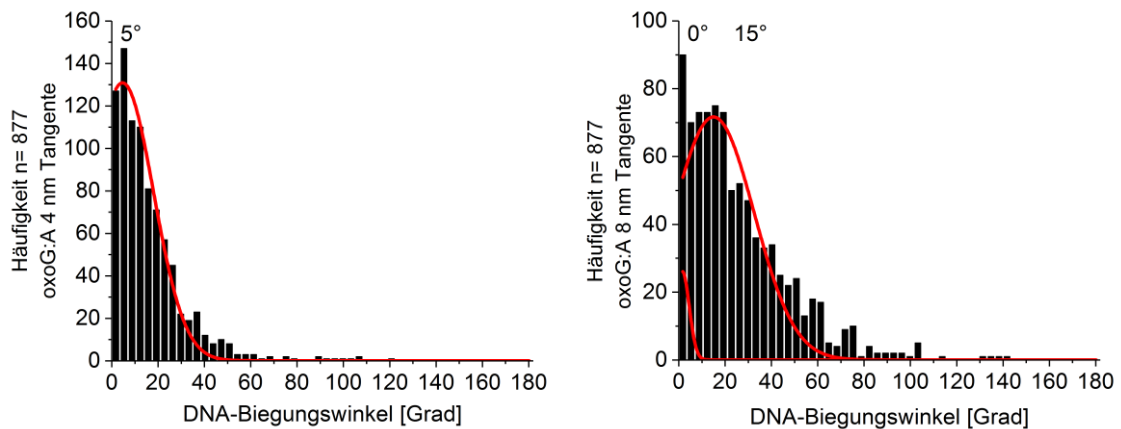
Ausgewertet wurden vier Experimente mit drei Aufnahmen vom 14.09.2016, vier Aufnahmen vom 22.03.2017, einer Aufnahme vom 06.06.2017 und drei Aufnahmen vom 12.05.2017. Die Länge von oxoG:A schadenstragenden DNA-Strängen betrug durchschnittlich  $154 \text{ nm} \pm 24 \text{ nm}$ . Die Messung des Biegewinkels von oxoG:A, einem Schadenssubstrat von MutY, mit der Tangentenmethode ergab bei 4 nm langen Segmenten einen Wert von  $5^\circ \pm 13^\circ$  und bei 8 nm langen Segmenten zwei weitere Winkelgrößen, und zwar sowohl bei  $0^\circ \pm 3^\circ$  als auch  $15^\circ \pm 18^\circ$ . Bei Verwendung der Sekantenmethode wurde bei 4 nm langen Segmenten ein Biegewinkel von  $0^\circ \pm 9^\circ$  gemessen. Bei 8 nm langen Sekanten vergrößerte sich der Biegewinkel auf  $5^\circ \pm 14^\circ$ .

### 3 Ergebnisse



**Abbildung 59: Längenverteilung von oxoG:A**

Die durchschnittliche Länge von oxoG:A betrug  $154 \text{ nm} \pm 24 \text{ nm}$ .

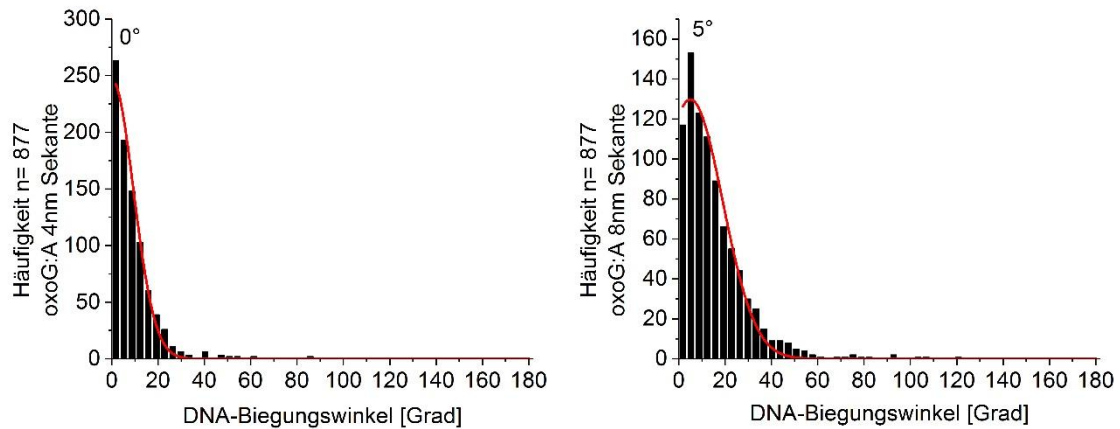


oxoG:A 4 nm

oxoG:A 8 nm

**Abbildung 60: Biegewinkelverteilung von oxoG:A mit der Tangentenmethode**

Der Biegewinkel 4 nm langer Segmente hatte einen Wert von  $5^\circ \pm 13^\circ$ . Bei 8 nm Segmenten zeigte sich eine Biegung von  $0^\circ \pm 3^\circ$  und  $15^\circ \pm 18^\circ$ .



oxoG:A 4 nm

oxoG:A 8 nm

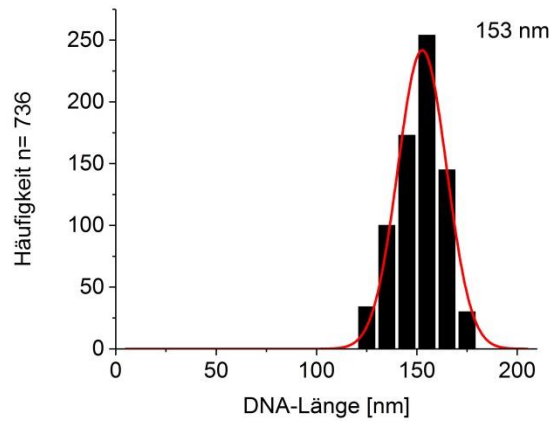
**Abbildung 61: Biegewinkelverteilung von oxoG:A mit der Sekantenmethode**

Der Biegungswinkel bei 4 nm langen Sekanten betrug  $0^\circ \pm 9^\circ$ . Für die Segmentlänge von 8 nm ergab sich ein Biegungswinkel von  $5^\circ \pm 14^\circ$ .

### 3.3.3 DNA-Biegewinkel von G:A

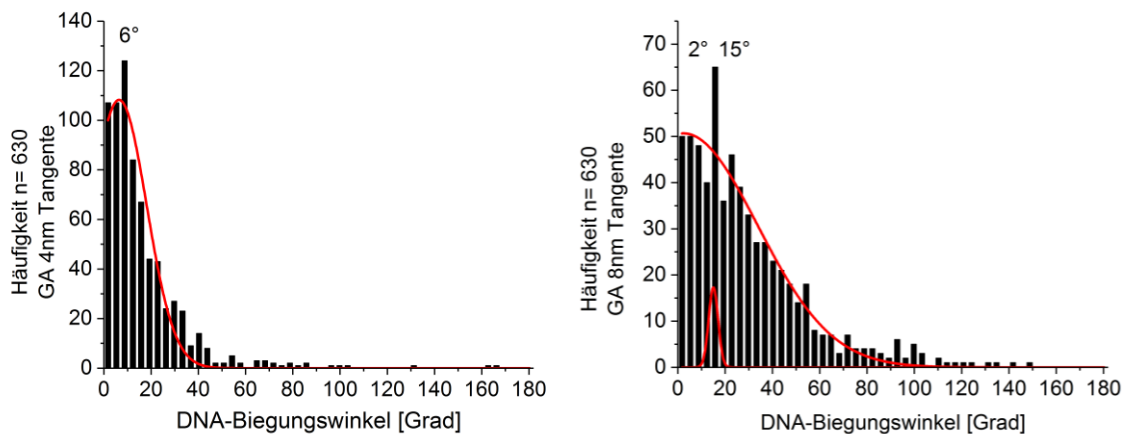
Ausgewertet wurden zwei Experimente mit je zwei Aufnahmen vom 22.03.2017 und vom 12.06.2017. Die Auswertung 505 bp langer DNA mit dem Schaden G:A aus dem Experiment vom 12.06.2017 von Lisa Hirsch ergab eine durchschnittliche Länge von  $153 \text{ nm} \pm 24 \text{ nm}$ . Der Biegungswinkel bei 4 nm langen Segmenten betrug mit der Tangentenmethode vereinigt  $6^\circ \pm 12^\circ$ . Für 8 nm Segmentlänge ergab sich ein Biegungswinkel von  $2^\circ \pm 32^\circ$  mit einem zusätzlichen scharfen Maximum bei  $15^\circ \pm 4^\circ$  und einer angedeuteten Spezies bei  $\sim 30^\circ$ . Die Vermessung der identischen AFM-Aufnahmen mit der Sekantenmethode zeigte bei 4 nm langen Segmenten einen Biegewinkel von  $0^\circ \pm 8^\circ$  und bei 8 nm langen Segmenten einen Winkel von  $5^\circ \pm 15^\circ$ .

### 3 Ergebnisse



**Abbildung 62: Längenverteilung von G:A**

Die Auswertung von DNA mit dem Schaden G:A ergab eine durchschnittliche Länge von 153 nm  $\pm$ 24 nm.

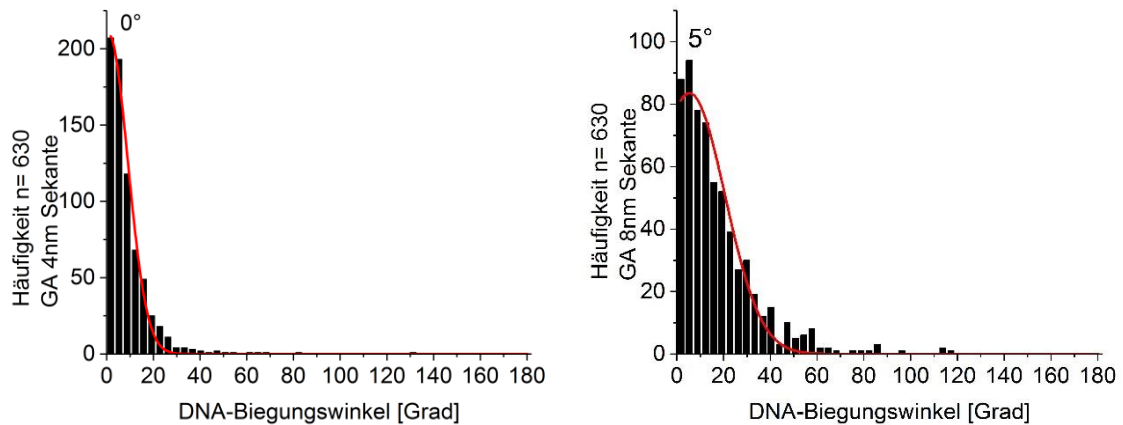


G:A 4 nm

G:A 8 nm

**Abbildung 63: Biegewinkelverteilung von G:A mit der Tangentenmethode**

Der Biegungswinkel bei 4 nm langen Segmenten betrug vereint  $6^\circ \pm 12^\circ$ . Für eine Segmentlänge von 8 nm ergab sich ein Biegungswinkel von  $2^\circ \pm 32^\circ$  mit einem zusätzlichen scharfen Maximum bei  $15^\circ \pm 4^\circ$  und einer andeuteten Spezies bei  $\sim 30^\circ$ .



G:A 4 nm

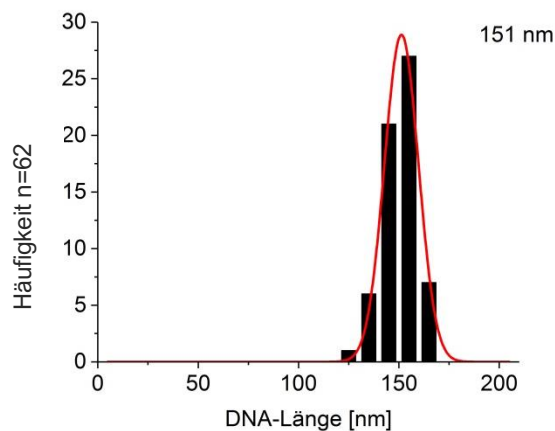
G:A 8 nm

**Abbildung 64: Biegewinkelverteilung von G:A mit der Sekantenmethode**

Der Biegungswinkel bei 4 nm langen Sekantensegmenten betrug vereint  $0^\circ \pm 8^\circ$ . Für die Segmentlänge von 8 nm ergab sich ein Biegungswinkel von  $5^\circ \pm 15^\circ$ .

### 3.3.4 DNA-Biegewinkel von oxoG:C

Ausgewertet wurden drei Aufnahmen von einem Experiment am 16.05.2017. Die Länge von oxoG:C schadenstragenden DNA-Strängen betrug durchschnittlich  $151 \text{ nm} \pm 17 \text{ nm}$ . Die Messung des Biegewinkels gemäß Tangentenmethode ergab bei 4 nm langen Segmenten einen Wert von  $11^\circ \pm 14^\circ$  und  $27^\circ \pm 5^\circ$  und bei 8 nm langen Segmenten drei Biegungsmaxima bei  $0^\circ \pm 3^\circ$ ,  $11^\circ \pm 4^\circ$  und  $30^\circ \pm 15^\circ$ . Bei Anwendung der Sekante als Messmethode deuteten sich ähnliche Biegewinkel an. Bei 4 nm Segmentlänge ein Winkel von  $0^\circ \pm 5^\circ$  und  $10^\circ \pm 8^\circ$  und bei 8 nm Segmentlänge ein Winkel von  $5^\circ \pm 8^\circ$  und  $27^\circ \pm 9^\circ$ .

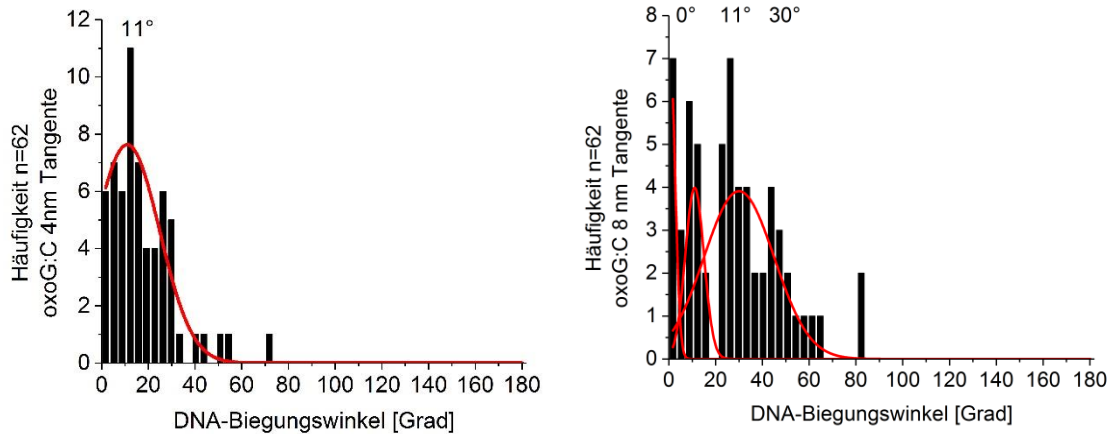


**Abbildung 65: Längenverteilung von oxoG:C**

Bei oxoG:C betrug die mittlere Länge  $151 \text{ nm} \pm 17 \text{ nm}$ .



### 3 Ergebnisse

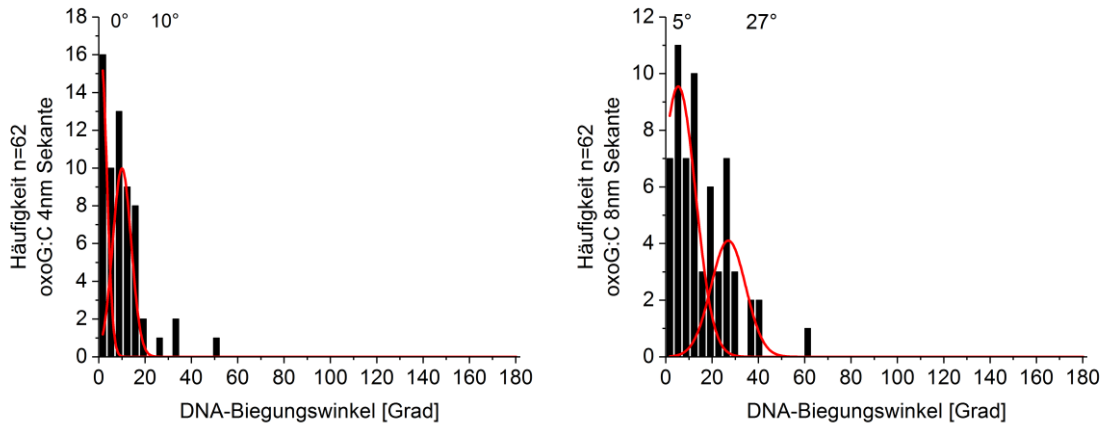


oxoG:C 4 nm

oxoG:C 8 nm

**Abbildung 66: Biegewinkelverteilung von oxoG:C mit der Tangentenmethode**

Für oxoG:C, das Schadenssubstrat von hOGG1, zeigte sich bei einer Segmentlänge von 4 nm eine Biegung der DNA am Schadensort von  $11^\circ \pm 14^\circ$  und  $27^\circ \pm 5^\circ$ . Für 8 nm Segmentlänge zeigen sich drei Biegungsmaxima bei ungefähr  $0^\circ \pm 3^\circ$ ,  $11^\circ \pm 4^\circ$  und  $30^\circ \pm 15^\circ$ . Aufgrund der geringen Datenmenge war eine sinnvollere Gauß-Anpassung nicht möglich.



oxoG:C 4 nm

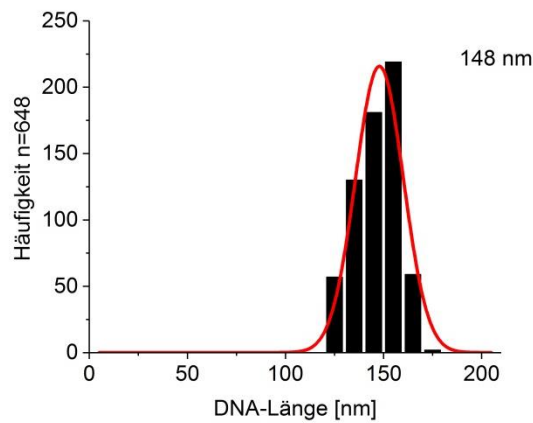
oxoG:C 8 nm

**Abbildung 67: Biegewinkelverteilung von oxoG:C mit der Sekantenmethode**

Der Biegungswinkel bei 4 nm langen Sekantensegmenten betrug  $0^\circ \pm 5^\circ$  und  $10^\circ \pm 8^\circ$ . Für die Segmentlänge von 8 nm ergab sich ein Biegungswinkel von  $5^\circ \pm 8^\circ$  und  $27^\circ \pm 9^\circ$ .

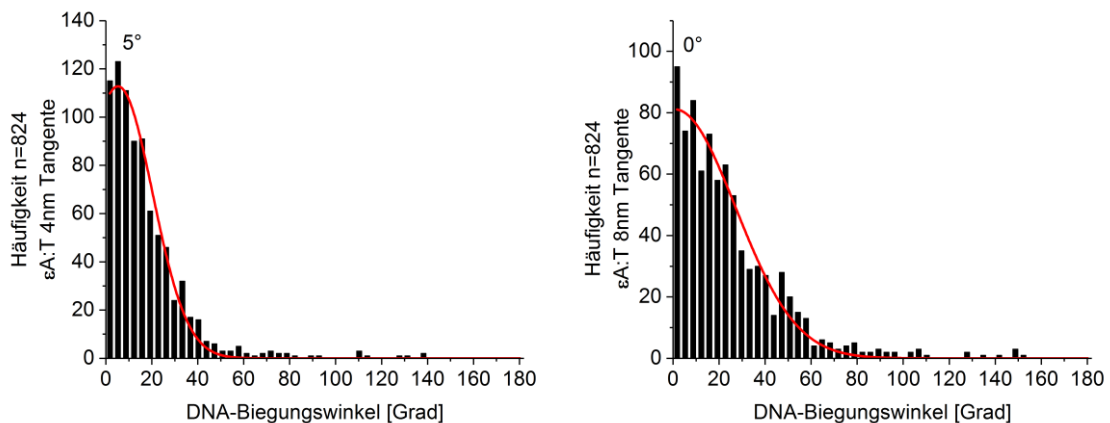
### 3.3.5 DNA-Biegewinkel von $\epsilon$ A:T

Ausgewertet wurden zwei Experimente mit zwei Aufnahmen vom 21.04.2017 und drei Aufnahmen vom 02.06.2017. Die Auswertung 505 bp langer DNA mit dem Schaden  $\epsilon$ A:T ergab vereint eine durchschnittliche Länge von  $148 \text{ nm} \pm 24 \text{ nm}$ . Bei der Tangentenmethode zeigte sich für die 4 nm langen Segmente ein durchschnittlicher Biegungswinkel von  $5^\circ \pm 15^\circ$  und für die 8 nm langen Segmenten ein durchschnittlicher Biegungswinkel von  $0^\circ \pm 26^\circ$  mit zwei angedeuteten Spezies ( $\sim 20^\circ$  und  $\sim 45^\circ$ ). Hingegen glichen sich die gemessenen Biegewinkel bei der Sekantenmethode. Sowohl für 4 nm lange Segmente ergaben sich  $0^\circ \pm 10^\circ$  als auch für 8 nm lange Segmente  $0^\circ \pm 17^\circ$ .



**Abbildung 68: Längenverteilung von  $\epsilon$ A:T**

Bei der Auswertung von  $\epsilon$ A:T wurde eine durchschnittliche Länge von  $148 \text{ nm} \pm 24 \text{ nm}$  ermittelt.

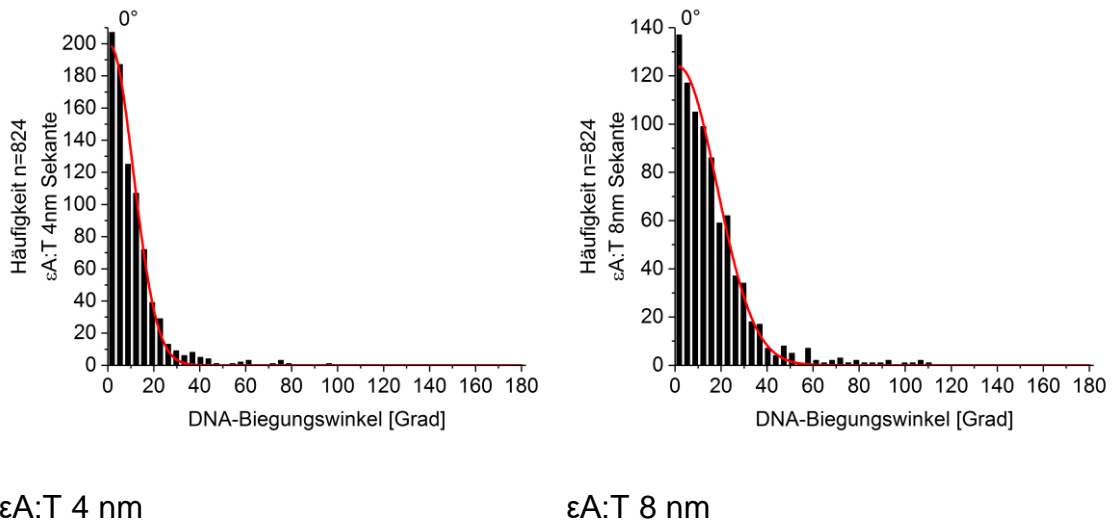


$\epsilon$ A:T 4 nm

$\epsilon$ A:T 8 nm

**Abbildung 69: Biegewinkelverteilung von  $\epsilon$ A:T mit der Tangentenmethode**

Für das Schadenssubstrat  $\epsilon$ A:T wurde bei 4 nm Segmentlänge ein Biegungswinkel von  $5^\circ \pm 15^\circ$  gemessen, bei 8 nm Segmentlänge ein Wert von  $0^\circ \pm 26^\circ$ .

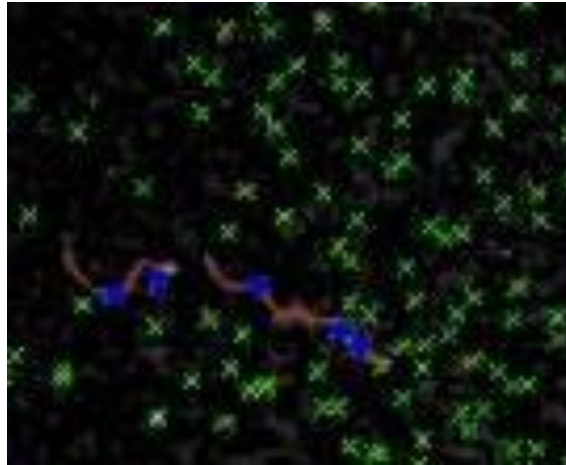


**Abbildung 70: Biegewinkelverteilung von  $\epsilon$ A:T mit der Sekantenmethode**

Der Biegungswinkel bei 4 nm langen Sekantensegmenten betrug vereint  $0^\circ \pm 10^\circ$ . Für die Segmentlänge von 8 nm ergab sich ein Biegungswinkel von  $0^\circ \pm 17^\circ$ .

### 3.4 Standardisierte und objektive Biegewinkelmessungen an Protein-DNA-Komplexen

Für die Ermittlung des Biegewinkels der DNA an der Position eines daran gebundenen Proteins war, im Vergleich zu der Biegewinkelbestimmung an der 50 %-Position der DNA-Länge, eine neue Anpassung des Programmablaufes nötig. Anstatt der von FIESTA ermittelten Mitte des DNA-Stranges wurde die Position des Proteinpeaks in Form von Koordinaten benötigt. Aufgrund der schwierigen Skelettierung der DNA in Anwesenheit von Proteinen (Abbildung 71), mussten diese temporär entfernt werden, was zum einen mit dem Particle Remover sowie einer Proteinausblendung mit dem Threshold-Intervall versucht wurde, aber erst zufriedenstellend mit dem Shape Filter erreicht werden konnte. Der Skelettierungsprozess in FIESTA konnte somit ähnlich dem Verfahren zur Skelettierung proteinfreier DNA durchgeführt werden. Zusätzlich musste die Position der Proteine separat ermittelt und diese in MATLAB anschließend mit den Daten aus der Skelettierung überlagert werden, um an dieser Stelle den Biegewinkel bestimmen zu können.

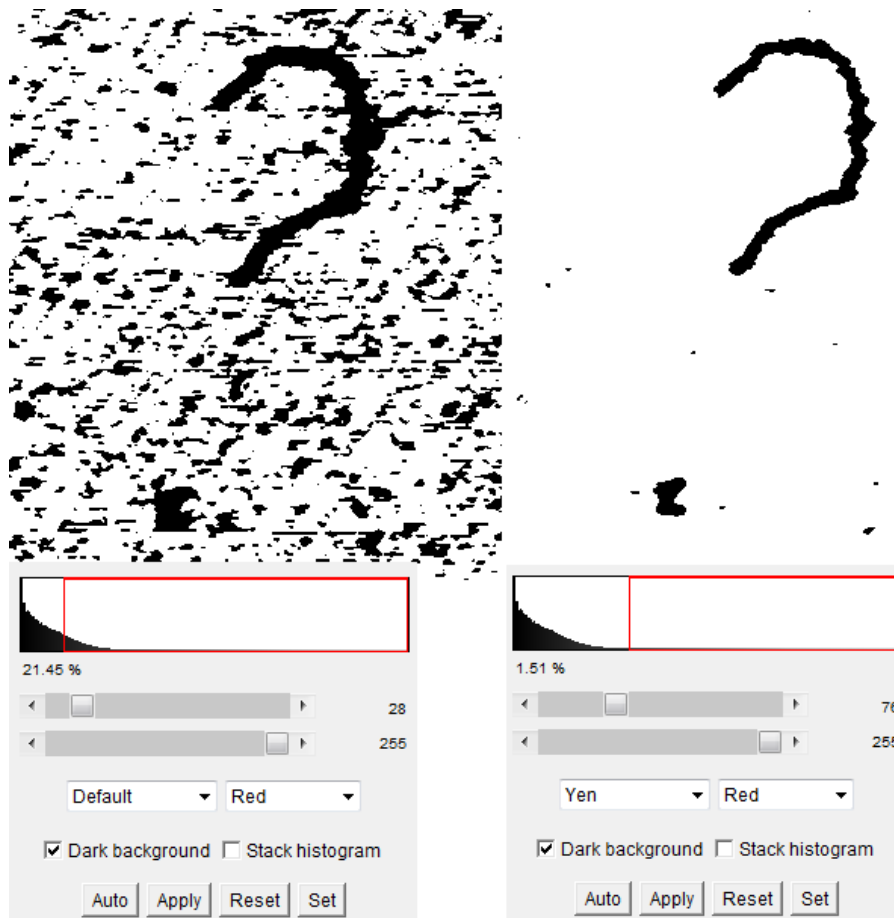


### **Abbildung 71: Auswertung ohne Protein-Filterung**

Eine direkte Auswertung der Aufnahme in FIESTA, ohne Verwendung des Protein-Filters, ausschließlich im 8 bit-Format und Median-Filter, schied aus, da es keine zufriedenstellende Threshold-Einstellung gab. Entweder wurden die DNA-Stränge sehr stark zerteilt (mehrere Mittelpunkte bzw. blaue Punkte) und verkürzt, oder das starke Hintergrundrauschen stellte ein weiteres Problem dar. Diese Störpartikel wurden vom Programm FIESTA fälschlicherweise als DNA-Mittelpunkte blau markiert und mussten anschließend manuell gelöscht werden. Nach dem Löschvorgang zeigten sich diese Partikel als grüne Kreuze.

### **3.4.1 Bildprozessierung**

Entsprechend dem Vorgehen bei der Analyse von Aufnahmen ohne Proteine, wurde das Bild in GIMP auf die tatsächliche Aufnahmegröße geschnitten und im Anschluss in ImageJ geöffnet. Über [Type] wurde die jeweilige Aufnahme wieder vom RGB in das 8 bit-Format umformatiert. Analog zu dem Vorgehen bei proteinfreier DNA wurde in ImageJ der passende Threshold-Wert für jede Aufnahme mit dem Yen-Threshold bestimmt, wobei geringgradige manuelle Anpassungen manchmal notwendig waren. Da sowohl der Particle Remover als auch der Shape Filter ausschließlich bei binären Bildern funktionierten, musste zu deren Erzeugung der Threshold in ImageJ nicht nur bestimmt, sondern auch angewandt werden.

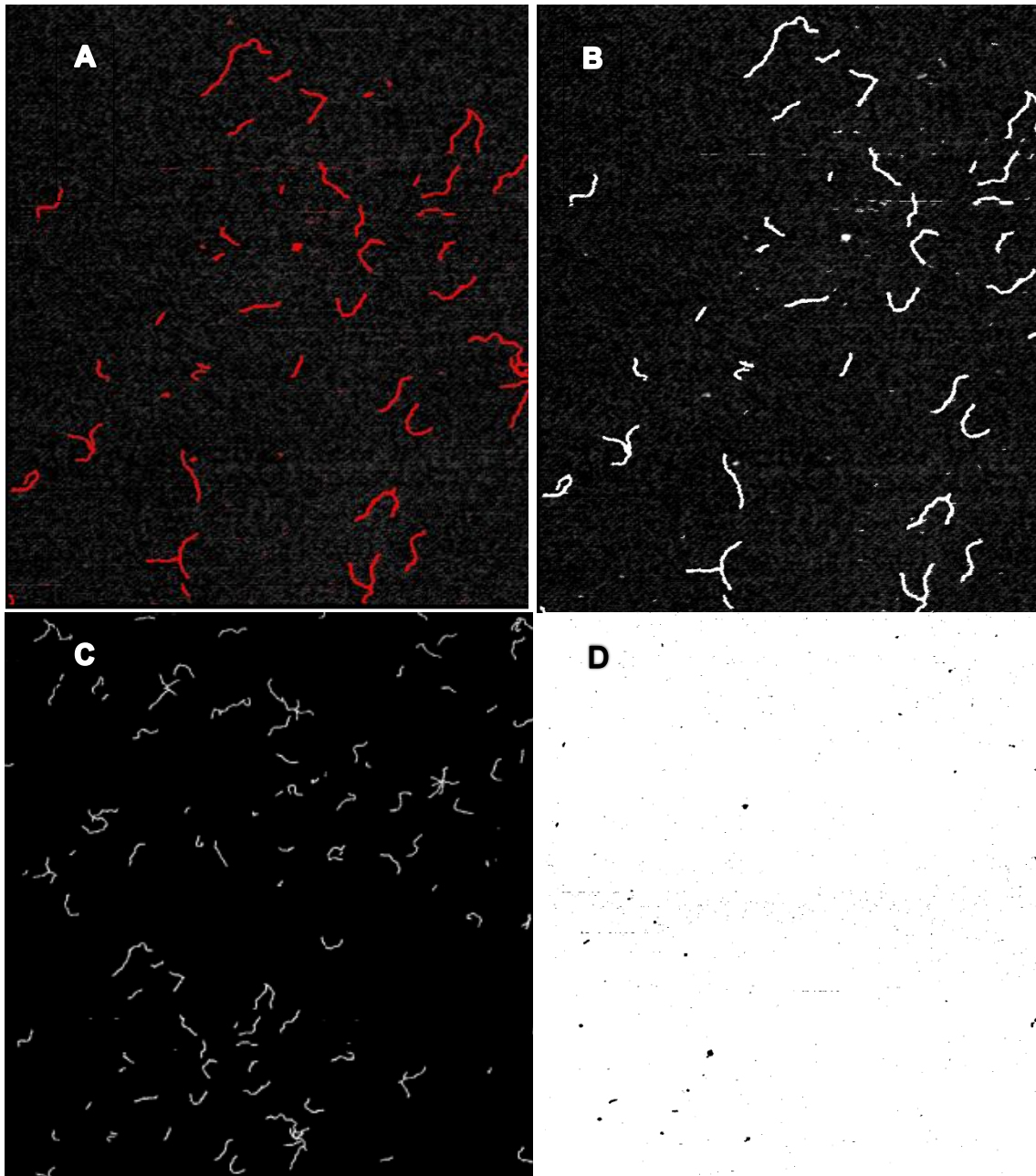


**Abbildung 72: Anwendung des Yen-Thresholds**

Die Anwendung eines Thresholdes erzeugte ein binäres Bild (links). Rechts erfolgte die Anpassung des Yen-Thresholds zu Optimierung der DNA-Auswertung. Der Threshold diente zur Minimierung des Hintergrundrauschens. So legte „Yen“ das Höhenminimum auf einen Wert, der 1,51 % aller Datenpunkte berücksichtigte.

### 3.4.2 Particle Remover

Der Particle Remover war ein erweiterter „Particle Analyzer“. Die Funktion bestand darin, entsprechend der Auswahlkriterien Partikel farblich zu markieren, und eine Kopie des Ausgangsbildes ohne diese markierten Partikel zu erstellen. Dies war eine Möglichkeit vorhandene Proteine (neben der DNA liegend) für eine separate Auswertung der DNA mit FIESTA aus der Aufnahme zu entfernen. Nach dem Skelettieren mit FIESTA wurden die Filamentierungsdaten gemeinsam mit den Proteinpositionen in MATLAB korreliert, und die DNA-Biegewinkel an den Proteinpositionen ausgewertet. Die Aufnahmen wurden wie bereits zuvor etabliert in das 8 bit-Format umgewandelt, ein Median Filter von 2 Pixeln angewandt und der Yen-Threshold gesetzt. Im Menü des Particle Removers wurde für den Parameter „Circularity“ 0 – 0,5 verwendet, was bewirkte, dass vornehmlich runde Strukturen wie die Proteine ausgefiltert wurden und die längliche DNA zurückblieb. Als Ausgabeoption wurde die „Objects map“ gewählt. Als Ergebnis wurde eine Tabelle mit Größe und Position der gefilterten Proteinpartikel ausgegeben. Zudem erhielt man mehrere Bilder (Abbildung 73): Entsprechend dem Ausgangsbild mit markiertem Threshold (A), allerdings anstelle der roten Threshold-Markierung, jetzt in weißer Farbmarkierung (B), die sog. „Count mask“, die die ausgefilterten DNA-Strukturen zeigte (C), und das Ergebnisbild (D) des Particle Removers mit den punktförmigen Proteinpartikeln. Die „Count mask“ sollte schließlich zur Auswertung in FIESTA benutzt werden.

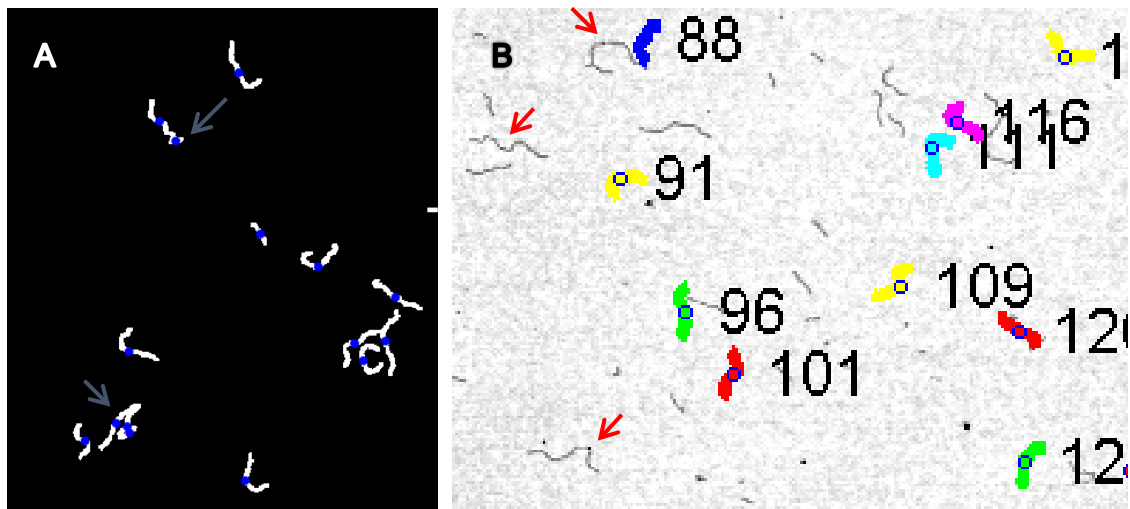


**Abbildung 73: Particle Remover**

- A:** Ausgangsbild im 8 bit-Format, mit 2 Pixel Median-Filter und Yen-Threshold.
- B:** Der Particle Remover zeigt die markierten Strukturen als intensiv weiß und entspricht der vorherigen Threshold-Einstellung.
- C:** Die Count Mask ist ein temporär erstelltes Bild, welches sich allerdings nicht für die Auswertung speichern lässt. Diese wäre allerdings eine mögliche Grundlage für die FIESTA-Auswertung gewesen.
- D:** Das Ergebnisbild (Objects map) des Particle Removers zeigt die entfernten/zurückgebliebenen Partikel.

Die „Count mask“ des Particle Removers stellte einen guten Ansatz zur DNA-Auswertung mit FIESTA dar, ließ sich allerdings zunächst nicht speichern oder exportieren. Dieses Problem konnte behoben werden, indem, ohne den

Threshold wirklich anzuwenden, die Bilder nur mit diesem markiert wurden, und anschließend das oben gezeigte Bild C (Abbildung 73) ausgegeben wurde. Dieses konnte nun gespeichert und mit FIESTA ausgewertet werden. Auf diese Weise wurden zwar kleine Partikel entfernt, aber störende Proteinpeaks auf bzw. an der DNA blieben erhalten. Außerdem wurde stark gebogene DNA weiterhin als getrennte Filamente skelettiert, am Ende befindliche Peaks als einzelne Filamente gewertet und die DNA teils verkürzt erfasst.



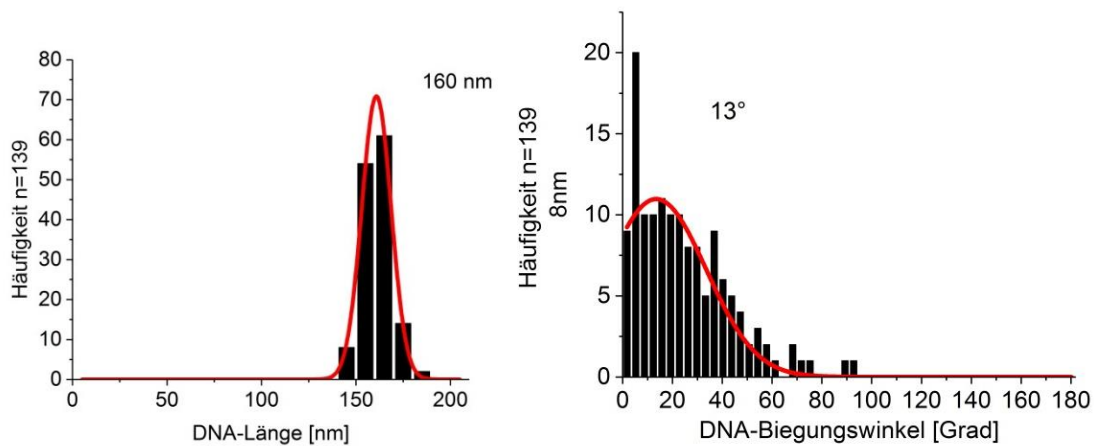
**Abbildung 74:** FIESTA-Ergebnis für eine Probe von MutY-Komplexen mit oxoG:A-DNA

- A:** Ergebnis der FIESTA-Auswertung: Problematische Teilung einzelner DNA-Stränge (blaue Pfeile)
- B:** Ergebnis der MATLAB-Auswertung: Viele DNA-Stränge wurden nicht erkannt (nicht farbig markierte DNA), was gerade bei mit Protein besetzter DNA häufiger auftrat (rote Pfeile). Farbige und nummerierte Stränge stellen von MATLAB ausgewertete DNA dar.

Zu Testzwecken wurde ein mit dem Particle Remover bearbeitetes Bild (Abbildung 74) mit dem MATLAB Programm für die Biegewinkelbestimmung an der 50 %-Position ausgewertet. Bei einer Länge von 160 nm ergab sich für das oxoG:A-Schadenssubstrat mit der Tangentenmethode bei 8 nm Segmentlänge ein Biegewinkel von  $13^\circ$ . Das Ergebnis ist konsistent zu den vorherigen Messungen (Kapitel 3.3.2). Dieser Test zeigte, dass die mit dem Particle Remover bearbeiteten Bilder prinzipiell mit dem Messprogramm in MATLAB auswertbar waren. Ein Vorteil des Particle Removers war die Tatsache, dass für die spätere Auswertung separat jeweils eine Aufnahme mit ausschließlich DNA und eine Aufnahme mit ausschließlich Proteinen erzeugt wurde. Aufgrund der oben genannten Probleme bei der Auswertung wurde im Weiteren nach besseren



Möglichkeiten für die Auswertung von DNA-Protein-Komplexen gesucht. Allerdings lieferte das Programm die Grundlage für die Methodik der Biegewinkelbestimmung an DNA-Protein-Komplexen. Die Skelettierung der DNA sollte an DNA-Strängen, die von den jeweiligen Proteinen separiert wurden, durchgeführt werden. Die Zusammenführung der zugehörigen Protein-Daten sollte erst im Schritt der Biegewinkelmessung in MATLAB geschehen.



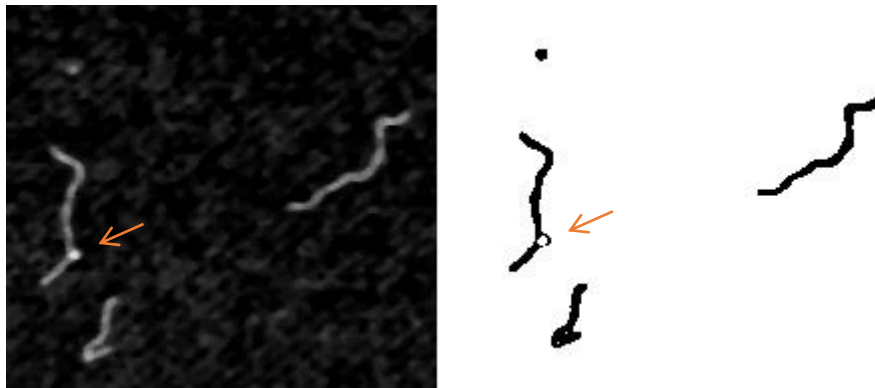
**Abbildung 75: Ergebnisse der Biegewinkel an der 50 %-Position**

Um die Auswertbarkeit von mit dem Particle Remover bearbeiteten Bildern zu testen, wurde eine Aufnahme von MutY und oxoG:A-DNA mit dem MATLAB-Programm zur Winkelmessung an der 50 %-Position genutzt. Die Abbildungen zeigen die Länge der DNA und einen Biege­winkel von  $13^\circ$  für 8 nm lange Segmente.

#### 3.4.3 Proteinentfernung mit einem Threshold-Intervall

Zusätzlich wurde durch Anpassen eines Threshold-Intervalls an die Minimum- und Maximum-Höhe der DNA versucht, die Protein-Peaks aus der Aufnahme zu entfernen. In ImageJ wurde über den Yen-Threshold wieder die DNA-Höhe er­mittelt und als Minimum gesetzt. So wurde der Hintergrund entfernt. Es verblie­ben noch die DNA und die Proteine. Als Maximum-Threshold wurde in ImageJ ebenfalls der Intermodes-Threshold gewählt, und dieser schrittweise erniedrigt, bis alle Proteine zuverlässig markiert wurden, jedoch keine DNA. Dieser Wert wurde ebenso wie der zuvor ermittelte Yen-Threshold vermerkt. Durch Anwen­dung eines Darstellungsintervalls in ImageJ aus den beiden obigen Threshold- werten wurden nur noch die DNA-Stränge dargestellt (Abbildung 76), während die Proteine dadurch ausgeblendet wurden. Anstatt dieser waren häufig aber

noch kreisförmige schwarze Umrandungen („weiße Löcher“) zu sehen, welche die DNA berührten. Ein solcher DNA-Strang war für eine Skelettierung mit FIESTA nicht zu gebrauchen. Zwar bot ImageJ die Möglichkeit, die entstanden „Löcher“ aufzufüllen, allerdings konnte dies das initiale Problem der Proteinfiltration nicht lösen. Die auf- oder angelagerten Protein-Peaks störten das Tracking in FIESTA weiterhin, da sich der Umriss des DNA-Protein-Komplexes nicht veränderte, und ein einzelnes DNA-Filament als getrennte Stränge dargestellt wurde. Aus diesem Grund wurde dieser Lösungsansatz nicht weiterverfolgt.



**Abbildung 76: Darstellung mit einem Threshold-Intervall**

Durch die Wahl eines geeigneten unteren und oberen Threshold-Wertes wurde versucht, ausschließlich den DNA-Strang, ohne das gebundene Protein, darzustellen. Der untere Wert orientierte sich am Yen-Threshold, wohingegen der obere Wert kleiner als der Intermodes-Threshold gewählt werden musste. Es entstanden „weiße Löcher“ an der Proteinposition (roter Pfeil).

### 3.4.4 Shape Filter

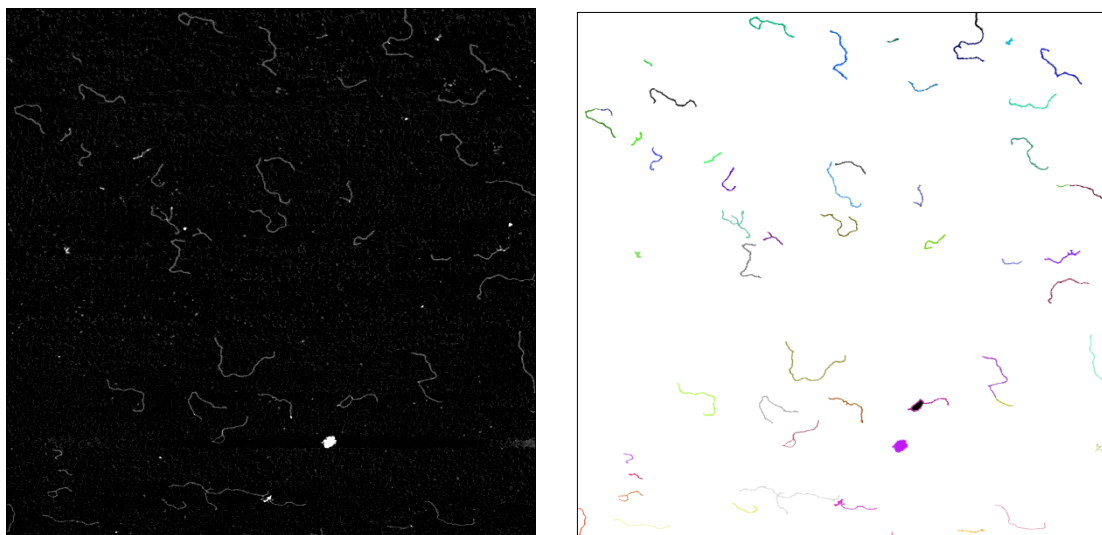
Einen neuen Ansatz für die Auswertung von DNA-Protein-Komplexen lieferte die Veröffentlichung „JBlob: An ImageJ Library for Connected Component Analysis and Shape Analysis“ von T. Wagner und H.-G. Lipinski [151]. Die Aufnahmen wurden in das 8 bit-Format umgewandelt und der Yen-Threshold gesetzt. Der Shape Filter erwies sich als sehr vielversprechendes ImageJ-Plugin, da so ein Bild ganz speziell nach jeglicher Art von Formen gefiltert werden konnte. Zuerst wurde versucht, den Parameter Elongation für die DNA-Stränge anzupassen. Die Elongation war gemäß Formel 5 definiert, wobei a der kleineren Achse der angepassten Ellipse und b der größeren Achse der angepassten Ellipse entsprach.

$$1 - \frac{a}{b}$$

**Formel 5**

In einem Bereich von 0,9 bis 1 wurden nur sehr gerade DNA-Stränge dargestellt. Bei der Einstellung 0,75 bis 1 wurde die vorkommende DNA zwar ausreichend gut dargestellt, zudem aber auch viele kleine Partikel. Beim Test der verschiedenen Parametereinstellungen hat sich der Parameter Perimeter als die effizienteste Filter-Funktion herausgestellt. Hierbei wurden im Bereich von „90 bis unendlich Pixel“ ausschließlich nur noch „längliche“ DNA und keine „runden“ Proteine oder sonstige Verschmutzungen mehr angezeigt. War die DNA bezüglich Länge und Biegung sehr inhomogen, musste der kleinere Wert schrittweise auf „60 bis unendlich Pixel“ gesenkt werden. So wurden auch kurze DNA-Stränge, die stark gebogen waren und ein sehr großes Protein trugen, wie dies beispielsweise bei dem Testprotein T-Antigen (~200 kDa bei Monomeren) der Fall war, noch in die Auswahl einbezogen. Durch die Entfernung des Hintergrundrauschens und freier Proteine verblieb nur die DNA. Die an die DNA gebundenen Proteine blieben zwar in der Darstellung erhalten, allerdings kam es zu keiner Beeinträchtigung der Skelettierung bzw. der Biegewinkelmessung.

Vorgehen: Der Shape Filter entfernte mit der Einstellung des Parameters Perimeter auf einen Bereich von „90 bis unendlich Pixel“ alle Partikel bis auf die länglichen DNA-Stränge. Auf diese Weise konnte das Hintergrundrauschen sehr gut reduziert werden.



**Abbildung 77: Shape Filter**

Nachdem die Aufnahme vom RGB (rot-grün-blau) in das 8 bit-Format umgewandelt wurde, konnten die DNA-Stränge durch die Verwendung des Shape Filters aus der Aufnahme extrahiert werden. So wurde die Auswertung mit FIESTA ermöglicht.

### 3.4.5 Invertieren und Filtern

Um das gefilterte Bild von FIESTA bearbeiten zu lassen, musste das Bild in ImageJ invertiert und ein Gaußscher Verwischungs-Filter genannt „Gaussian Blur“ angewandt werden (Abbildung 78).



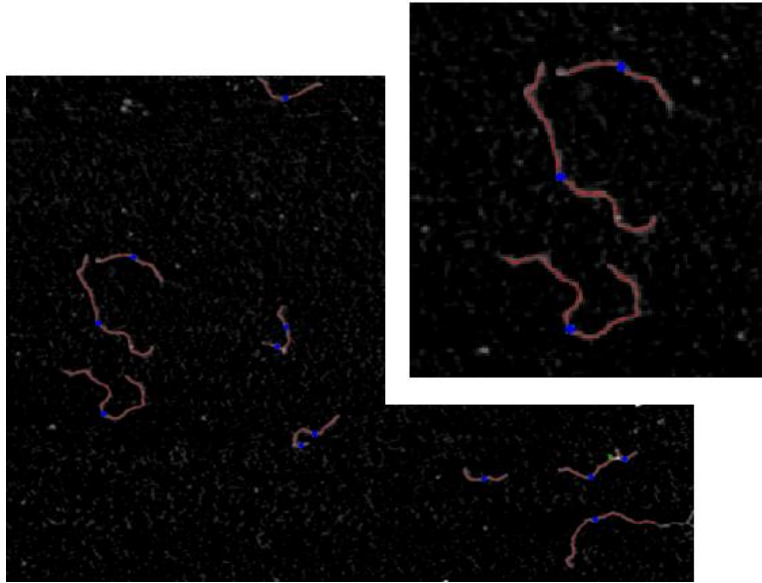
**Abbildung 78: Invertierung und Gauß-Filter**

Die Abbildung zeigt einen Bildausschnitt nach Invertierung des Shape Filter-Ergebnisbildes und anschließender Verwendung des „Gaussian Blur“-Filters. Die DNA stellt sich nun leicht verwischt dar.

Da alle in FIESTA geöffneten Bilder invertiert dargestellt wurden, mussten die Ergebnisbilder des Shape Filters in ImageJ auch zusätzlich invertiert werden. Unterblieb das Invertieren wurde die DNA in FIESTA, statt als weißer Strang mit entsprechender Höhe auf schwarzem Untergrund, als schwarze Linie auf weißem Hintergrund dargestellt. Für FIESTA stellte die DNA so eine Vertiefung der Oberfläche dar, weshalb das Skelettieren des gesamten Untergrundes als ein zu großes Objekt zum Absturz des Programmes führte. Nach der Invertierung des Shape Filter-Ergebnisbildes war dieses von FIESTA aber noch immer nicht auswertbar, da es sich um ein binäres Bildformat handelte. Ursächlich hierfür war, dass FIESTA während der Ermittlung der Halbwertsbreite (FWHM) und dem Skelettierungsvorgang immer eine Gauß-Anpassung der Daten durchführte. Durch die Verwendung des Verwischungsfilters „Gaussian Blur“ an je zwei Pixeln konnte das Bild so bearbeitet werden, dass eine Auswertung mit FIESTA möglich war. Die Halbwertsbreite verdoppelte sich bei diesem Verwischungsvorgang allerdings von durchschnittlich 13 nm bei rein Median-gefilterten Bildern, auf fast 25 nm bei den Gauß-modifizierten Bildern. Nach diesen Bearbeitungsschritten wurde das Bild als [.tiff]-Datei für den FIESTA-Arbeitsgang gespeichert.

### 3.4.6 Skelettierung

Die Auswertung in FIESTA (Abbildung 79) unterschied sich von der DNA-Skelettierung in Abwesenheit von Proteinen nur insofern, dass anstelle des Median-gefilterten Bildes das mit dem Shape Filter modifizierte, invertierte und Gauß-modifizierte Bild verwendet wurde.

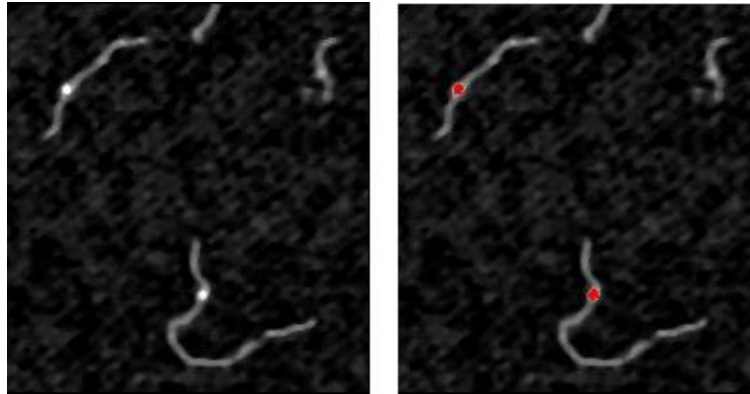


**Abbildung 79: Ergebnis des FIESTA-Trackings**

Die gefilterte DNA wird durch eine rote Filamentlinie dargestellt. Der blaue Punkt markiert die Mitte eines jeden Filaments, allerdings wird der Mittelpunkt für die Messungen an den Proteinpeaks nicht zwingend benötigt.

### 3.4.7 Proteinpositionsbestimmung

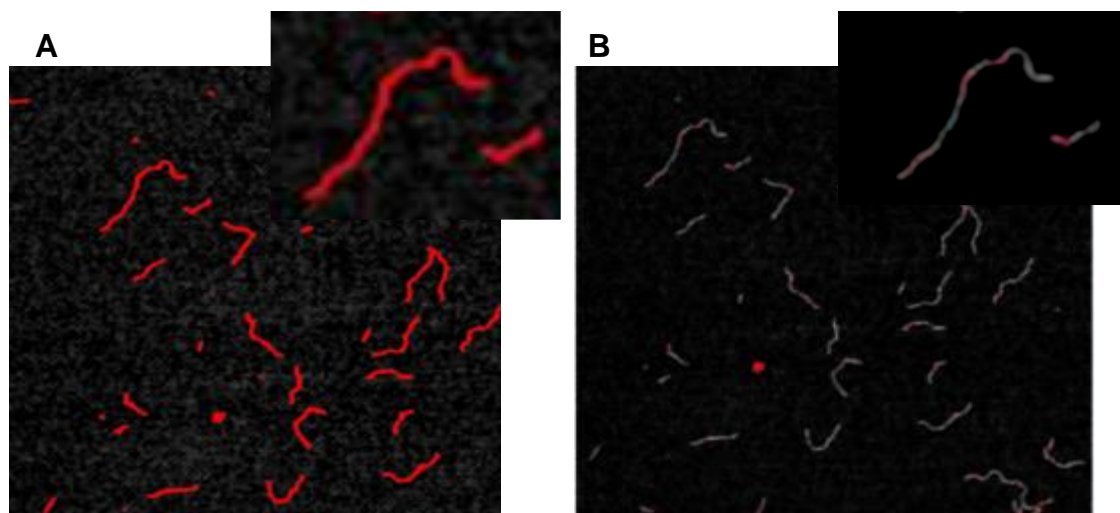
Im Weiteren war es für die Bestimmung der Protein-Peak-Position wichtig, dass das Ausgangsbild erneut im 8 bit-Format in ImageJ geöffnet wurde, und unter [Analyse] und [Set Scale] die Pixelauflösung und Bildgröße eingestellt wurde. Für die selektive Darstellung der Proteine wurde der Intermoden-Threshold genutzt (Abbildung 80), und im Folgenden mit dem „3D Object counter“ die genaue Bildposition der Proteinpeaks in Form von Koordinaten erhalten. Durch die Wahl eines geeigneten Thresholds sollten gerade alle Protein-Moleküle vom Thershold sicher erfasst werden (Abbildung 80). Es durften aber keine DNA-Stränge markiert werden, da dies die Ausgangslage der Proteinerfassung darstellte.



**Abbildung 80: Intermoden-Threshold**

Mit dem Intermoden-Threshold wurden alle signalintensiveren Strukturen rot markiert. Ziel war es die Hintergrundhöhe so zu wählen, dass ausschließlich Proteinen und keine höheren Anteile der DNA erfasst wurden.

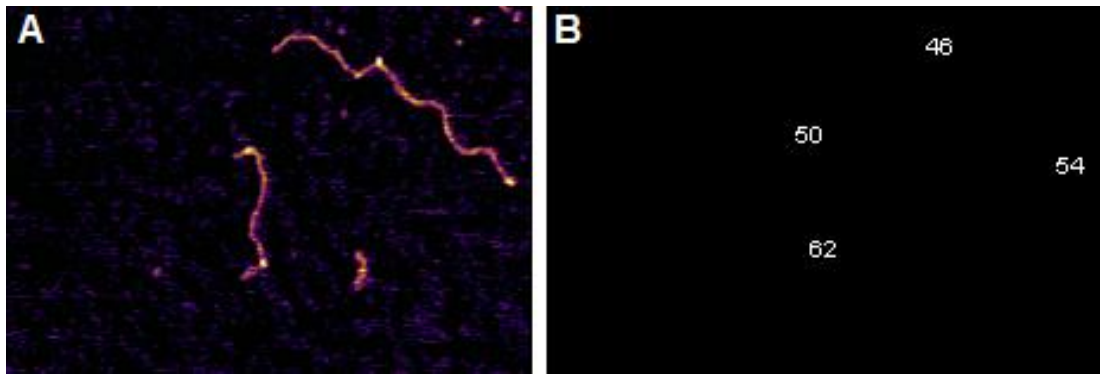
Auch bei der Anwendung des Intermoden-Threshold war, ähnlich zum Yen-Threshold eine optimale Darstellung der Proteinstrukturen nicht möglich, wenn neben den Proteinen die kompletten DNA-Stränge mitmarkiert wurde (Abbildung 81). Hierfür musste der Threshold teilweise manuell angepasst werden. Die manuelle Erhöhung des Hintergrundes musste in diesem Fall sehr behutsam erfolgen, da sich die Höhen von DNA und Proteinen bei kleinen Molekülen wie den Glykosylasen meist nur geringfügig unterschieden.



**Abbildung 81: Schwierigkeit beim Intermoden-Threshold**

- A** Mit dem Intermoden-Threshold wurden alle Strukturen rot markiert, nicht nur die höher liegenden Moleküle. Der Threshold war zu tief gesetzt
- B** Trotz manueller Anpassung der Hintergrundhöhe wurden, neben den Proteinen, teilweise immer noch auch höhere Anteile der DNA erfasst.

Der zuvor ermittelte Threshold wurde auch im „3D Object counter“ in ImageJ eingestellt und für die Peakgröße als Minimum ein Wert von fünf Pixeln eingegeben. Die Datei mit den ermittelten Protein-Daten wurde als [.txt]-Datei gespeichert, ebenso wie die erstellten Objektkarten (Objects map) als [.tiff]-Datei (Abbildung 82).



**Abbildung 82: Intermodales-Schwellenwert und Ergebnis des „3D Object counters“**

- A:** Der Intermodales-Schwellenwert wurde genutzt, um die Proteine zu lokalisieren. Anschließend wurden mit dem „3D Object counter“ die Protein-Koordinaten ermittelt (MutY-nspAC1605100001HtRM0)
- B:** Die Ausgabe der Ergebnisse des „3D Object counters“ erfolgte als [.txt]-Datei und je nach Wahl aus den verschiedenen Plänen (Maps) als Objekt- oder Centroid-Darstellung (Centroid = geometrischer Schwerpunkt) (hier ist die Objektdarstellung gezeigt).

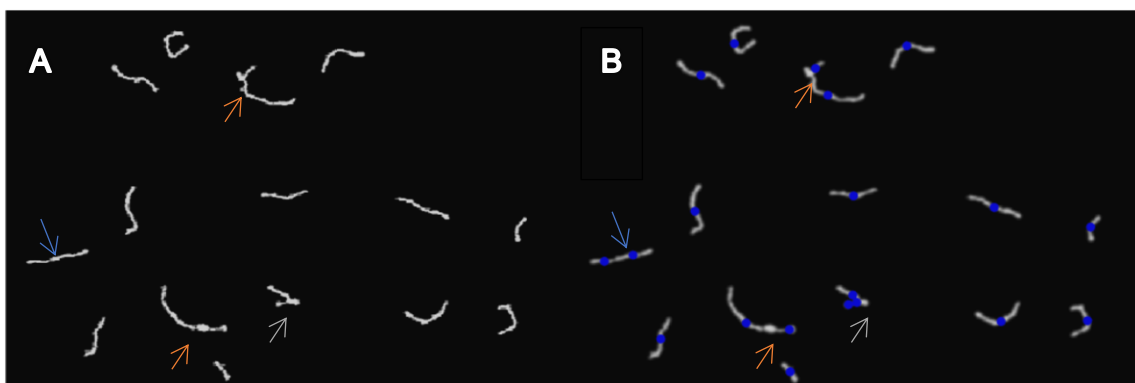
Bei der Proteinauswertung mit dem „3D Object counter“ war es wichtig, einen Grenzwert für die Flächengröße festzulegen. Der Minimalwert war wichtig, um Partikel, die kleiner als Proteine waren, auszuschließen. Diese kleineren Strukturen im Bild entstanden entweder durch Verschmutzungen, wenn der DNA-Strang leicht von der Glimmeroberfläche aufgeworfen war oder wenn zwei DNA-Stränge übereinander lagen. Auch gerade am DNA-Ende kam entwundene dsDNA gehäuft vor, die oft erhöhte Strukturen aufwies und deshalb vom „3D Object counter“ erfasst wurde.

Die Wahl eines Maximalwertes beschleunigte die Auswertung und klammerte grobe Verschmutzungen aus. Dies war gerade bei den sehr kleinen Protein-DNA-Komplexen wichtig, da bei einem niedrig gewählten Threshold oder Aufnahmen schlechterer Darstellung fälschlicherweise zu viele Partikel auf der DNA mit dem „3D Object counter“ erfasst wurden. Daher ist es wichtig, möglichst verschmutzungsfreie Protein-DNA-Proben für die Abbildung mit dem Rasterkraftmikroskop zu verwenden.

### 3.4.8 DNA-Biegewinkelmessung an Protein-DNA-Komplexen

Für die Messung des DNA-Biegewinkels an gebundenen Proteinen wurde prinzipiell das Programm „DNAsegmentAnalysisatProteinPositionKinked\_20180403.m“ verwendet. Es standen entsprechend der Biegewinkelmessung an alleiniger DNA zwei bis auf die Methode der Biegewinkelkonstruktion identische Messprogramme zur Verfügung. Für die Winkelbestimmung mittels zweier Sekanten wurde das Programm „DNAsegmentAnalysisatProteinPositionKinked\_SECANT\_20190116.m“ genutzt. Wobei für die Tangentenmethode „DNAsegmentAnalysisatProteinPositionKinked\_TANGENT\_20190116.m“ angewandt wurde.

Für die Biegewinkelbestimmung von DNA-Protein-Komplexen war eine schrittweise Anpassung der Programme an neu auftretende Herausforderungen nötig. Zu Beginn der Programmentwicklung der Protein-DNA-Auswertung wurde gerade DNA mit gebundenen Proteinen als zwei separate Filamente erkannt (Abbildung 83). Ziel war es diesen Fehler zu beheben. Da sich weder durch Anpassung der Halbwertsbreite, Setzen eines geeigneten Threshold-Intervalls auf DNA-Höhe noch durch den Particle Remover oder sonstige Plugins ein Erfolg verzeichnen ließ, wurde versucht dieses Problem im Zuge der MATLAB-Auswertung zu korrigieren.



**Abbildung 83: Test zur Proteinauswertung an oxoG:A-MutY-Komplexen**

**A:** Gezeigt wird das Ausgabebild des Shape Filters nach Invertierung und Gauß-Verwischung.

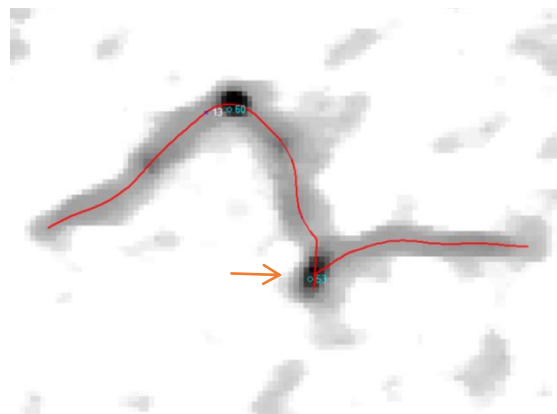
**B:** Identischer Bildausschnitt wie A nach der FIESTA-Skelettierung.

Die FIESTA-Filamentierung trennte proteinbesetzte DNA-Stränge (roter Pfeil), ebenso stark gebogene DNA (grüner Pfeil) und DNA-Stränge, die zwar verbunden, aber unterschiedlich signalintensiv, bzw. hoch waren (blauer Pfeil). Die blauen Punkte markieren die von FIESTA ermittelten Mittelpunkte der DNA-Fragmente.



### 3.4.8.1 Verbindung unterbrochener Skelettierungen

Ein wichtiger Fortschritt der MATLAB-Optimierung war, dass nun auch stark gebogene DNA, die von FIESTA als zwei separate Filamente verarbeitet wurde, gezielt in MATLAB verbunden werden konnte. Für das Problem wurde von Hannah Heil die Option des selektiven Verbindens der beiden getrennten Filamente in MATLAB implementiert (DNAsegmentAnalysisatProtein-Position-Kinked\_20180403.m). Über eine Abfragemaske konnte nun die Gesamtzahl der zu verbindenden Filamente genannt werden, um dann anschließend die jeweils zu verbindenden Filamente einzeln, durch ein Komma getrennt, aufzuzählen. MATLAB verband automatisch die mit dem geringsten Abstand zueinander befindlichen Filament-Enden mit einer möglichst kurzen Linie (Abbildung 84). Auch an der eingefügten Verbindungslinie wurde die gleiche Polynom-Interpolation, die bei den restlichen Filamenten angewandt wurde, durchgeführt.



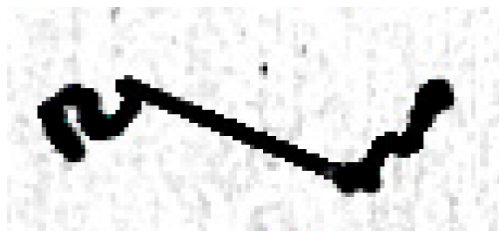
**Abbildung 84: Tracking gebogener und proteinbeladener DNA**

Durch „DNAsegmentAnalysisatProteinPositionKinked\_20180403.m“ bestand die Option des selektiven Verbindens der getrennten DNA an Stellen starker Biegung. Die selektive Verbindung zeigt sich im Bereich des spitzen Aufeinandertreffens der beiden Filamentierungslinien (roter Pfeil).

### 3.4.8.2 Molekülfilter

Allerdings hatte die Verbindung der gewünschten Filamente nicht immer zufriedenstellend funktioniert. Es wurden trotz richtiger Angabe der Filamentnummern reproduzierbar auch falsche Filamente (ca. 2% der Fälle) verbunden (Abbildung 85). Es musste davon ausgegangen werden, dass bei Vorhandensein von großen rundlichen Strukturen (von FIESTA als Molekül bezeichnet) diese von FIESTA als Objekt erkannt wurden, obwohl bei der Einstellung „tracking only

filaments“ keine solche Moleküle erkannt werden sollten. Aus diesem Grund wurde ein Molekülfilter in MATLAB implementiert, da vorhandene Moleküle zusammen mit den Filamenten in derselben Ergebnistabelle aufgeführt wurden, und so die spätere Filamentzuordnung durch MATLAB störten. Der programmierte Filter durchsuchte fortan die Ergebnistabelle vor der Auswertung nach „Molekülen“ und entfernte diese aus der Objektliste. Als Moleküle wurden Objekte definiert, die eine kleinere Länge als 2 nm aufwiesen.



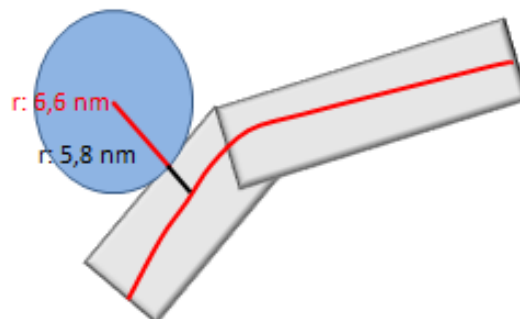
**Abbildung 85: Notwendigkeit eines Molekülfilters**

Die Funktion eines Molekülfilters war nötig, da FIESTA trotz der Einstellung „tracking only filaments“ teilweise große rundliche Strukturen als Moleküle erkannte. Da die Objekt-Datensätze für Moleküle und Filamente in einer einzigen Ergebnis-Tabelle zusammen aufgeführt wurden, war die Reihenfolge der Filament-Objekte und so auch eine Zuordnung der einzelnen Filamente in MATLAB nicht immer möglich.

#### **3.4.8.3 Protein-Radius**

Für die Protein-Filament-Kolokalisation war es wichtig festzulegen, wie groß der Proteindurchmesser bzw. -Radius war, denn nach der Skelettierung durch FIESTA standen in MATLAB ausschließlich die Filamentdaten der einzelnen DNA-Stränge zur Verfügung. Diese Filamente bestanden lediglich aus einer ein-dimensionalen Linie in der Mitte der DNA-Breite, die den Verlauf des DNA-Stranges beschrieb. Es war somit nicht, wie in der Aufnahme dargestellt, der komplette Umfang der DNA verfügbar. Auch die Proteine wurden nur durch die Koordinaten des Proteinmittelpunktes dargestellt. Somit war es für die Definition, ob ein Protein an DNA gebunden ist oder nicht, wichtig den Durchmesser der DNA sowie der Proteine zu bestimmen und diesen bei der Wahl des akzeptablen Abstandes zwischen der DNA-Linie und des Proteinmittelpunktes einzubeziehen. Auf diese Weise konnte durch Angabe eines maximalen Abstandes zwischen dem skelettierten DNA-Strang und der Koordinaten des Proteinmittelpunktes der Zustand „Protein an DNA gebunden“ genau definiert werden. MATLAB ermittelte bei dem Berechnungsvorgang für jedes Protein, den geringsten Abstand zwischen

Proteinmittelpunkt und Segmentlinie. Überschritt der errechnete Wert den zuvor als maximalen Abstand definierten Wert, wurde an dieser Stelle kein Biegewinkel ermittelt. Die Wahl des Proteinradius war von der durchschnittlichen Proteingröße, die entsprechend auch mit der Fläche korrelierte, abhängig. Hierzu wurde beispielhaft für die Glykosylase MutY der Protein-Durchmesser zum einen an freien Proteinen, zum anderen an Protein-DNA-Komplexen ermittelt. Für zwei verschiedene Glykosylasen (MutY, UDG) wurde in unabhängigen Experimenten neben dem durchschnittlichen Durchmesser von freien Proteinen und an DNA gebundenen Proteinen auch der DNA-Durchmesser selbst ermittelt. Für eine qualitativ gute Aufnahme des DNA-Substrats G:A, inkubiert mit MutY und aufgenommen mit einem scharfen Tip, betrug der durchschnittliche Durchmesser freier Proteine 13,2 nm und an DNA gebundener Proteine 14 nm. Als DNA-Querschnitt wurde 11,6 nm gemessen. Hieraus ergaben sich MutY-Proteinradien von ca. 7 nm und DNA-Radien von ca. 6 nm. Entsprechend wurde für MutY ein Maximalwert von 13 nm für die erlaubte Distanz (um noch als an die DNA-gebunden zu gelten) zwischen den Protein- und DNA-Mittelpunkt gewählt (siehe Abbildung 86). Für eine weitere Aufnahme von ungeschädigter DNA im Komplex mit UDG wurde ein DNA-Querschnitt von 16,6 nm gemessen, was auf eine eher breite Tip-Spitze hindeutete. Der durchschnittliche Durchmesser freier Proteine betrug in dieser Aufnahme 18,1 nm, und an DNA gebundener Proteine 18,8 nm. Für UDG zeigte sich in den AFM-Aufnahmen ein Radius von ca. 9,4 nm, was zusammen mit DNA-Radien in dieser Aufnahme von ca. 8,3 nm eine maximal tolerable DNA-Protein-Distanz von ca. 18 nm ergab.

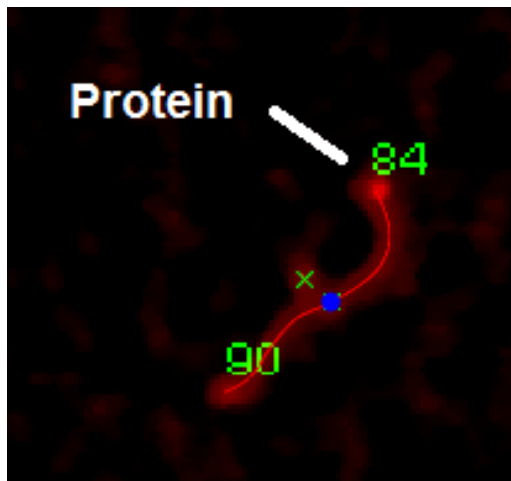


**Abbildung 86: Definition der Proteinbindung an DNA**

Als Regelung für MutY, ab wann ein Protein (blau) an DNA (graue Filamente) gebunden ist, wurde eine maximale Entfernung Proteinmittelpunkt zu Filamentlinie (rot) von 13 nm angenommen.

### 3.4.8.4 Messung des Biegewinkels an MutY-DNA-Komplexen

Die Proteinauswertung wurde optimiert am Beispiel der Glykosylase MutY im Komplex mit 2224 bp langer unbeschädigter DNA. Verglichen wurde die Auswertung von MATLAB mit bereits bestehenden Ergebnissen aus der AG Tessmer, welche manuell ermittelt worden waren. Problematisch war die Tatsache, dass sich das Volumen und somit auch die Höhe dieser sehr kleinen Proteine (33 – 39 kDa) nur in geringem Maße von der Höhe der DNA unterschied. War die DNA aufgeworfen und nicht völlig flach am Glimmerhintergrund gebunden, so ähnelten sich die beiden Höhen. Für nahezu jedes Bild musste eine manuelle Nachjustierung des Intermoden-Thresholds durchgeführt werden, um eine möglichst gute DNA-Protein-Diskriminierung zu erreichen. Für deutlich größere Proteine sollte die Auswertung einfacher und somit schneller ablaufen, da hier ein automatischer Threshold viel zuverlässiger arbeiten kann. Wurde in dem Programm FIESTA sowohl das Ausgangsbild nach der notwendigen RGB-Format in 8 bit-Format-Umwandlung und der Median 2 px-Filterung, als auch die ausgegebene „Objects map“ des „3D Object counter“ geladen, konnte man diese in der Überlagerung aus DNA-Filament und nummeriertem Protein-Peak optimal betrachten (Abbildung 87). Dies war eine Kontrollmöglichkeit, um die korrekte Skelettierung und die richtige Proteinzuzuordnung zu überprüfen und fehlerhafte Ergebnisse zu verwerfen. Außerdem erleichterte diese Darstellungsweise die Einstellung des Flächen-Grenzwertes für die MATLAB-Analyse. Anhand der Nummerierung wurde für optisch richtig erscheinende Proteinpeaks die zugehörigen Flächen-Angaben aus der „3D Object counter“-Tabelle abgelesen und als Anhaltspunkt genutzt.

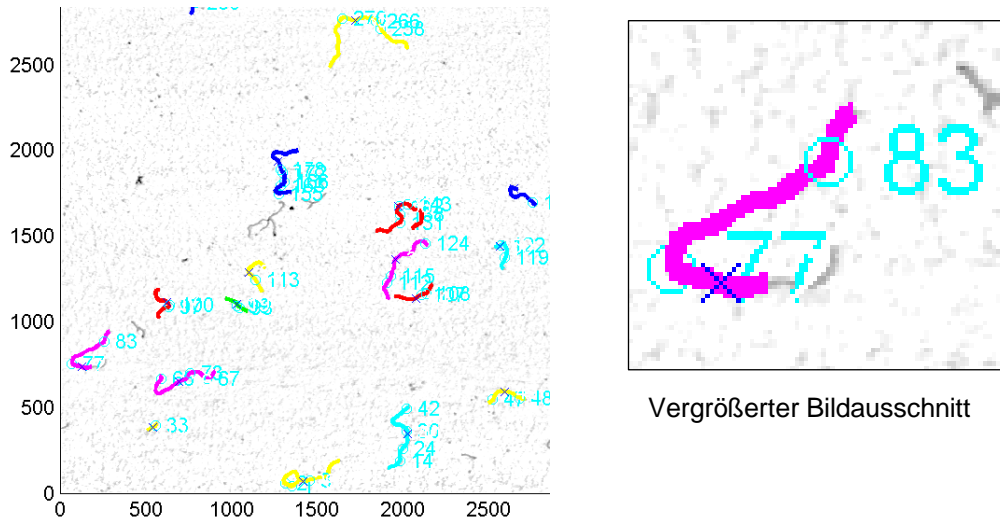


**Abbildung 87: Darstellung der Filamente und Proteinpositionen**

Dargestellt ist ein DNA-Strang (in FIESTA) jeweils mit ermittelten Filamenten (rote Linie) und darüber eine Projektion der Proteinposition (grüne Nummerierung) aus der Objects map des „3D object counters“ (ImageJ). Die blauen Punkte zeigen die von FIESTA bestimmte Mitte des entsprechenden Filaments an. Diese Darstellung eignete sich gut, um die korrekte Skelettierung und den Sitz der Proteine zu überprüfen. Ein endständiges Protein, wie Nr. 84, konnte auf diese Weise von der Auswertung ausgeschlossen werden. Das grüne Kreuz stellt ein von der Auswertung ausgeschlossenes Protein in FIESTA dar.

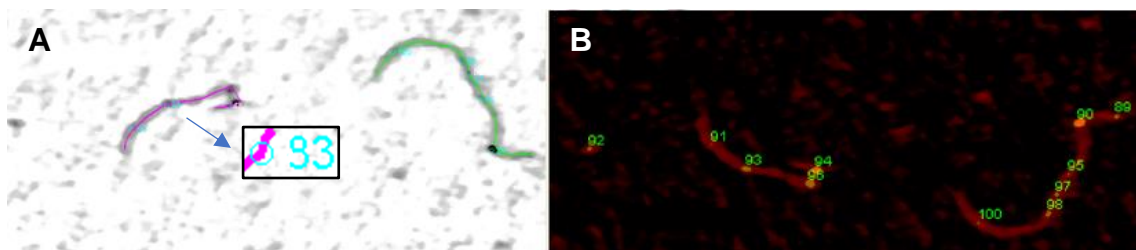
Nach Eingabe der zu messenden Abfragepunkte (8 nm Segmentlänge) wurde die Interpolation gestartet. Daraufhin wurde die [.txt]-Datei mit den Protein-Koordinaten geladen und entsprechend dem gewählten Proteinradius mit den DNA-Filamenten korreliert. Am Schnittpunkt des Proteinradius mit dem DNA-Filament wurde der Winkel gemäß den gewählten Abfragepunkten (Tangenten mit 8 nm Abstand) errechnet (Abbildung 88). Damit ein Biegewinkel an den Protein-DNA-Komplexen gemessen wurde, musste der Abstand zwischen den an die DNA gebundenen Proteinen mindestens 50 nm betragen. So konnte eine mögliche Beeinflussung der Biegewinkel durch benachbarte Protein-DNA-Komplexe ausgeschlossen werden.

Für die spätere Auswertung wurde das Ausgangsbild geladen, und darin die DNA-Stränge gemäß FIESTA nummeriert. Die Proteinposition wurde durch einen Kreis markiert und die entsprechende Nummer des „3D Object counter“ angezeigt (Abbildung 89). Dies erlaubte eine Zuordnung des Biegewinkels am Protein mit dem jeweiligen Proteinvolumen. Die Arbeitsdaten von MATLAB wurden in den Abbildungsordner gespeichert.



**Abbildung 88: Überlagerung der Proteinposition mit skelettierten DNA-Strängen für die MATLAB-Winkelmessung**

MATLAB zeigte bei der Protein-Biegewinkelbestimmung im Objektplan die ausgewerteten DNA-Stränge bunt gefärbt. Das gebundene Protein wurde jeweils durch einen Kreis und mit der „3D Object counter“-Nummerierung in türkise markiert. Wurden zu viele Strukturen als Proteine erkannt, kam es zu einer Überlagerung der türkisen Beschriftungen.



**Abbildung 89: Auswertung des MATLAB-Ergebnisses**

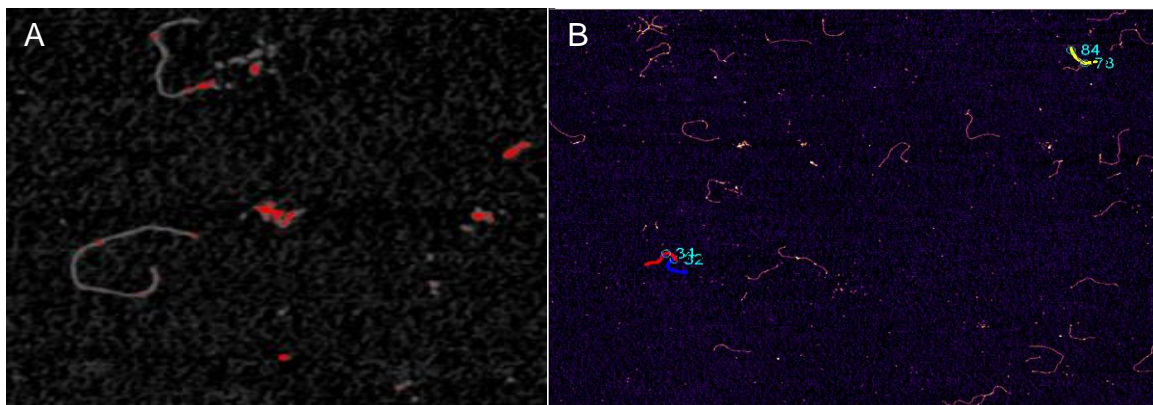
- A:** Skelettierte Filamente (rot und grün) mit eingezeichneter Proteinposition (türkise Kreise; siehe Vergrößerung) in MATLAB.
- B:** Für die Überprüfung der Proteinbestimmung wurden die Filament- und Proteindaten zusammen mit dem Ausgangsbild in FIESTA geladen.

Weil das Programm auch für die Biegewinkelmessung an Aufnahmen weiterer Proteinsysteme genutzt werden sollte, waren zusätzliche Anpassungen nötig. Es lagen die manuell gemessenen Winkelergebnisse mit den dazugehörigen Proteinvolumina vor. Bei den vorliegenden Aufnahmen handelte es sich allerdings an Stelle von 505 bp „kurze“ DNA-Stränge um 2224 bp „lange“ ungeschädigte DNA-Stränge. Die verschiedenen Anpassungen für den speziellen Fall der Winkelbestimmung an den langen ungeschädigten DNA-Strängen ergaben das Programm „DNasegmentAnalysisatProteinPositionKinked\_20180425\_allfilaments.m“. Hierbei wurden alle DNA-Stränge, gleich welcher Länge ausgewertet, da bei dieser langen ungeschädigten DNA die Längen der einzelnen DNA-

Stränge, im Vergleich zur deutlich homogeneren 505 bp langen DNA, teilweise stark variierten. Die Einstellung, dass alle DNA-Stränge gleich welcher Länge von MATLAB ausgewertet wurden, war wichtig, da MATLAB in der eigentlichen Programmierung im ersten Auswertungsteilschritt die Längen aller FIESTA-getrackten DNA-Strukturen gemäß einer Gauß-Verteilung überprüfte und im weiteren Auswertungsprozess lediglich noch Filamente mit einer Länge  $\pm 20$  nm um das Maximum der Gauß-Verteilung berücksichtigte. Auf eine Längenfilterung konnte bei ungeschädigter DNA verzichtet werden, da für ungeschädigte DNA der Ort der Proteinbindung unbedeutend war. Dagegen sollte der Biegewinkel bei geschädigter DNA am Schadensort gemessen werden, weshalb nur an der 50 %-Position gemessen wurde und diese deshalb exakt bekannt sein musste.

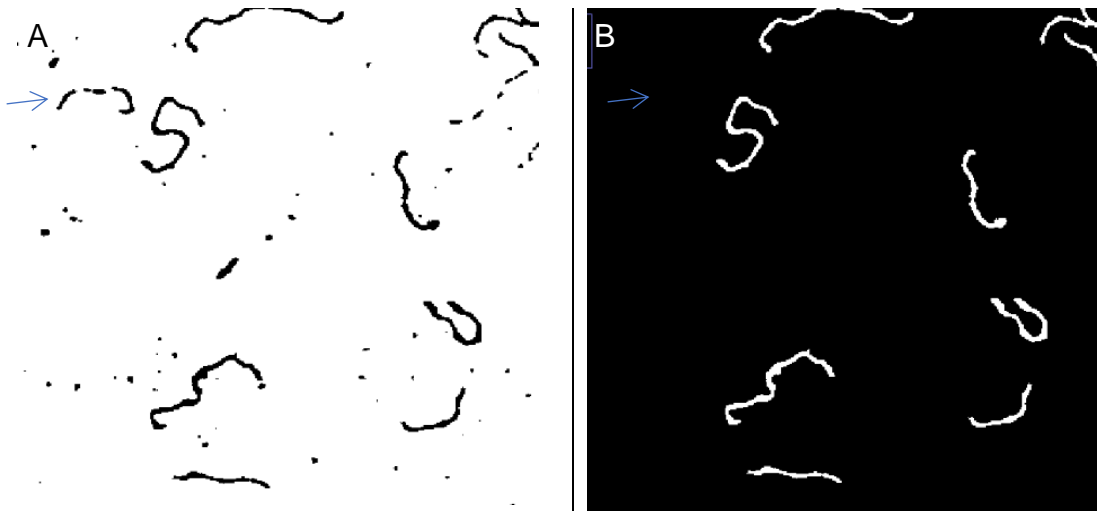
Die Optimierung für die Auswertung der vorhandenen Aufnahmen erforderte folgende Anpassungen von MATLAB:

Die erste Modifikation betraf den Threshold, da der sonst verwendete Yen-Threshold die deutlich längere DNA nicht mehr zufriedenstellend darstellte. So musste der Threshold jeweils aufwendig manuell auf das einzelne Bild adaptiert werden (Abbildung 90). Eine nicht ideale Threshold-Einstellung führte zu einer Zerteilung der langen DNA-Strukturen, sodass der nachfolgend angewandte Shape Filter die geteilten DNA-Stücke ausfilterte (Abbildung 91).



**Abbildung 90: Limitation des Yen-Threshold**

Bei Anwendung des Yen-Thresholds (A) wurden von MATLAB (B) nur drei relativ kurze Filamente mit vier daran gebundenen Proteinen erkannt. Die DNA war teilweise unterschiedlich lang und sehr inhomogen, sodass sich kein homogenes Höhenprofil abbildete. Deshalb gestaltete sich auch die Festlegung eines passenden Thresholds schwieriger, als dies bei der kürzeren und homogeneren DNA der Fall war. Häufig war deshalb eine Modifikation des Yen-Thresholdes notwendig.

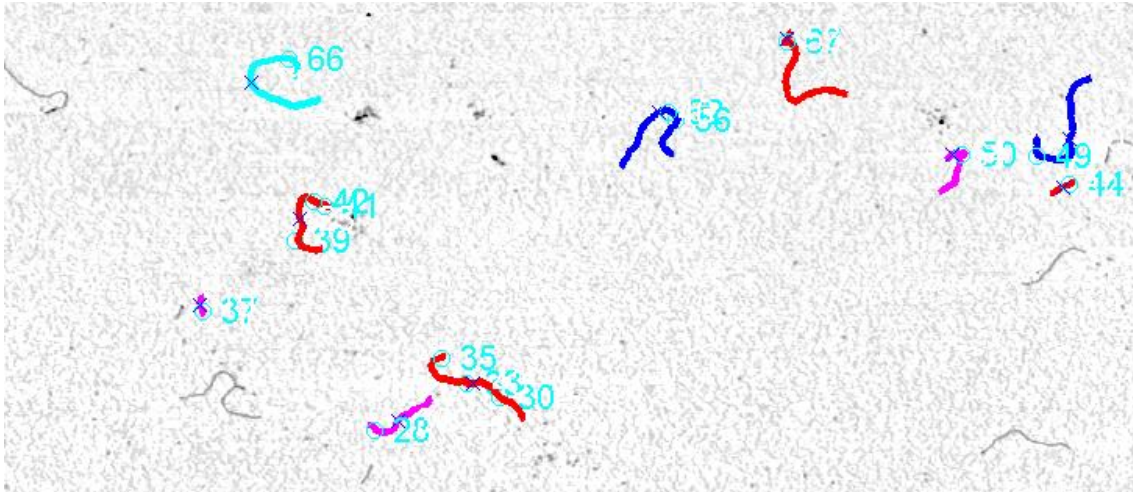


**Abbildung 91: Limitation des Shape Filters durch Definition des Thresholds**

**A** zeigt beispielhaft die DNA-Strukturen nach Festlegen eines Thresholds. Hierbei deutet der blaue Pfeil exemplarisch auf ein zerteilte DNA-Struktur. **B** zeigt denselben Bildausschnitt. Der Shape Filter entfernte die zerteilten DNA-Stränge (blauer Pfeil), wobei diese auch im Weiteren schwierig auszuwerten gewesen wären, da stets nur zwei Filamente manuell verbunden werden konnten. Eine Möglichkeit wäre, den Perimeter von 90 auf bis zu 60 zu erniedrigen. So wurden aber auch nicht mehr alle störenden Partikel aussortiert.

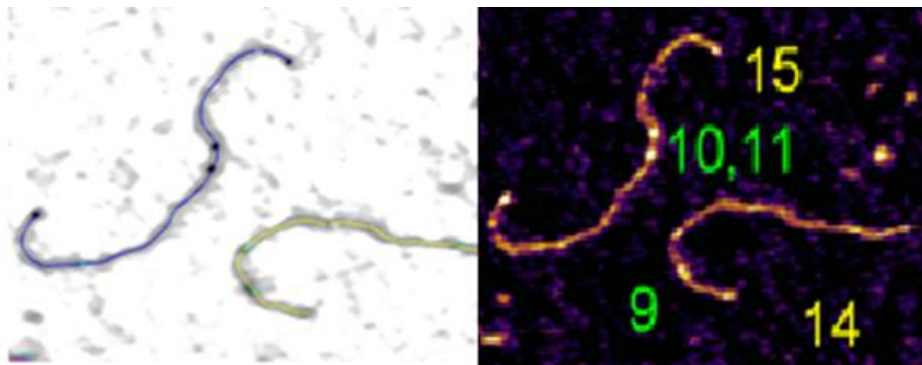
Außerdem wurde bei der Programmeinführung ein Mindestabstand von 50 nm zwischen den Proteinen definiert. Dieser Abstand war aus vorherigen Arbeiten [1] etabliert. Da die manuelle Auswertung für diese Aufnahmen bereits vorlag, sollten für eine bessere Vergleichbarkeit der Methoden bei der automatisierten Messung möglichst viele Messpositionen ermittelt werden, die mit den Messpositionen der manuellen Ergebnisse übereinstimmten. Bei der Wahl eines geeigneten Thresholds erwies sich die gleichzeitige Erfassung der relativ kleinen Glykosylasen und aufgeworfener DNA-Stellen, die bei „langen“ DNA-Strängen vermehrt auftraten, als problematisch. Da bei den manuellen Messungen bereits alle Proteine ausgeschlossen wurden, die näher als 50 nm zueinander auf der DNA gebunden waren, wurde für die Auswertung (DNAsegmentAnalysisatProteinPositionKinked-20180417\_allfilaments.m) speziell dieser Bilder der Abstand zusätzlich auf 25 nm verringert (Abbildung 92). So sollte vermieden werden, dass erhöhte DNA-Bereiche, die fälschlicherweise als Protein-Peaks auf der DNA erkannt wurden, zu einem Unterschreiten der 50 nm Schwelle führten. Anderenfalls wäre eine Biegewinkelmessung an Protein-DNA-Komplexen durch die 50 nm Schwelle ungewollt verhindert worden (Abbildung 93).





**Abbildung 92: Verbesserung der Erfassung von Protein-DNA-Komplexen**

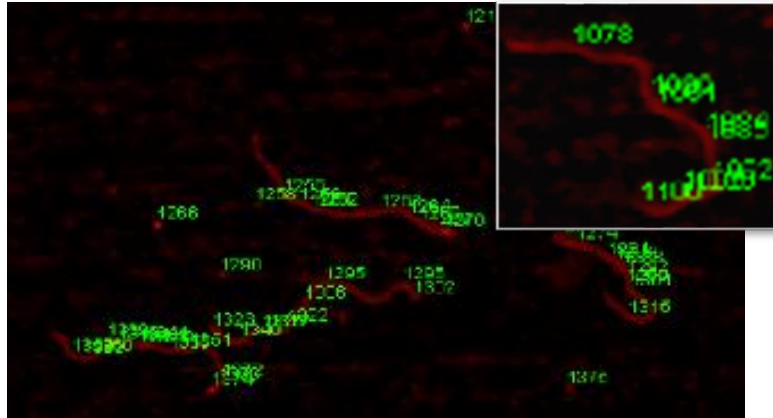
Um die Erfassung von Proteinen auf den DNA-Strängen zu verbessern, wurde der Mindestabstand zusätzlich auf 25 nm reduziert. Das MATLAB-Ergebnisbild zeigt eine bessere Erfassung. Jedoch wurden noch immer einige Proteine nicht erkannt.



**Abbildung 93: Proteinposition kleiner als 25 nm**

Bei der Einstellung von 25 bis 700 nm<sup>2</sup> wurden theoretisch alle Proteine erfasst, außer der Mindestabstand von 25 nm wurde zusätzlich unterschritten.

Ein ähnliches Problem ergab sich bei der Detektion der Proteine und deren Position. Bei Einstellung eines geeigneten Thresholds musste bei der 2224 bp langen DNA häufig in Kauf genommen werden, dass auch Teile des DNA-Stranges, bzw. Schmutzpartikel markiert und somit vom „3D Object counter“ erfasst wurden. Wurde nun der Minimum-Grenzwert im „3D Object counter“ zu niedrig gewählt, ergab sich ein vergleichbares Problem. Es wurden wieder zu viele DNA-Peaks fälschlich als Protein-DNA-Peaks erkannt. In Abbildung 94 wird das zuvor geschilderte Problem demonstriert, was sich in der Unmenge an teilweise überlappenden Proteinkennzeichnungen an den einzelnen DNA-Strängen zeigt. Folglich erfüllten viele korrekte Protein-DNA-Peaks den Mindestabstand von 25 nm nicht mehr und wurden bei der MATLAB-Auswertung aussortiert.



**Abbildung 94: Gemeinsame Darstellung von Proteindaten mit DNA-Filamenten**

In FIESTA wurden DNA und Protein-Peaks übereinander dargestellt (xy2px und objects map). Es wurden deutlich zu viele Peaks fälschlich als Protein-Peaks erkannt, und dadurch der tolerierte Mindestabstand zwischen einzelnen Proteinen-Peaks deutlich unterschritten. Deshalb wurde der Threshold bei der Proteindarstellung erhöht und der Größen-Filter des „3D Object counters“ vergrößert, um somit mehr falsche Peak-Positionen auszufiltern und einen ausreichenden Mindestabstand zwischen den Proteinpositionen sicherzustellen.

Um die Auswertung möglichst vieler Protein-DNA-Peaks zu ermöglichen, musste bei der Proteinerkennung der Minimum-Grenzwert auf jedes Bild optimal angepasst werden. Beispielhaft ist in Abbildung 95 ein gutes MATLAB-Ergebnis zu sehen.

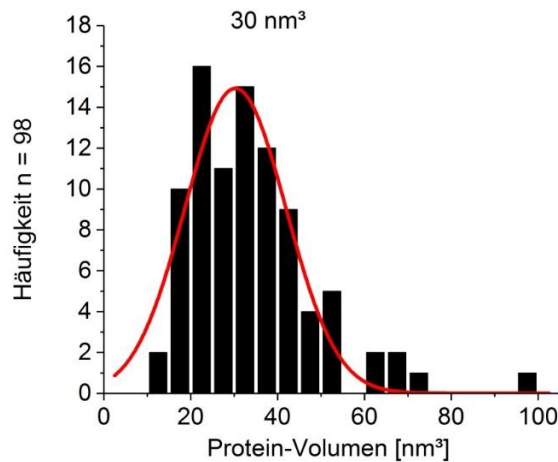


**Abbildung 95: Anpassung der Proteinerkennung**

MATLAB-Ergebnis nach Auswertung der Aufnahme mit „DNAsegmentAnalysisatProteinPosition-Kinked\_20180425\_allfilaments.m“. Die Proteine sind als türkise Kreise mit der dazugehörigen Nummerierung auf den verschiedenfarbigen DNA-Strängen dargestellt. Die große Zahl an Proteinpositionen auf der DNA konnte hierbei durch einen optimierten Minimum-Grenzwert verringert werden.

### 3.5 Biegewinkel von Protein-DNA-Komplexen

Nach Anpassung des MATLAB-Programms für die Erfassung des Biegewinkels an der Position des Protein-Peaks, anstatt der 50 %-Position, wurden Aufnahmen von MutY gebunden an ungeschädigte DNA aus einem vorigen Experiment der AG Tessmer genutzt. Die Glykosylase MutY (*E. coli*) besitzt eine Masse von 39 kDa. Nach Umrechnung gemäß der Formel 4 entspräche dies einem Volumen von 42,1 nm<sup>3</sup>. In Abbildung 96 zeigte sich das Maximum der Proteinvolumen-Verteilung von MutY bei 30 nm<sup>3</sup>. Deshalb erfolgte die Messung der Biegewinkel ausschließlich an DNA gebundenen Proteinen mit einem Volumen zwischen 20 nm<sup>3</sup> und 40 nm<sup>3</sup>. Die Abweichung vom theoretisch errechneten Wert kann durch die Bildqualität oder die Protein-Konformationen bedingt sein. An das Ende des DNA-Stranges gebundene Proteine wurden ebenso nicht ausgewertet, wie benachbarte Proteine, die weniger als 50 nm (bzw. 25 nm im speziellen Fall; siehe 3.4.8.4) voneinander trennten. Auf diese Weise sollte eine mögliche gegenseitige Beeinflussung auf die DNA-Biegung ausgeschlossen werden.



**Abbildung 96: Proteinvolumenverteilung von MutY**

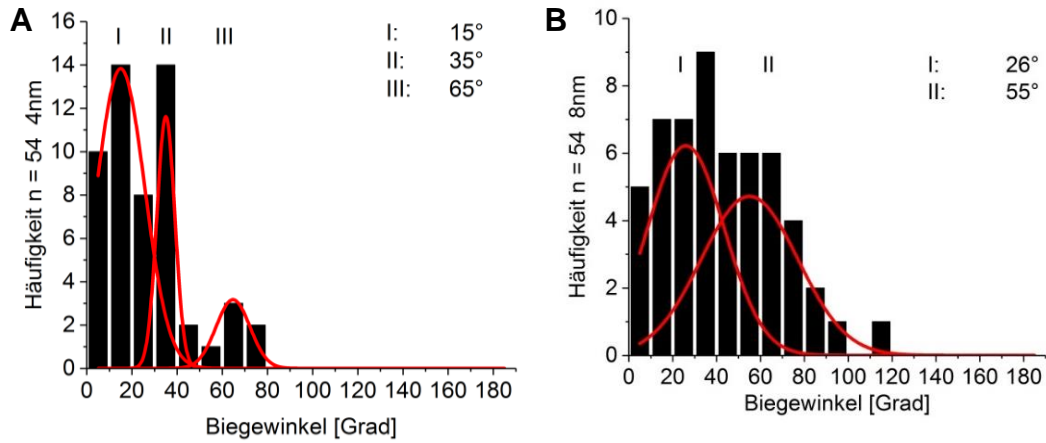
Die Abbildung zeigt die Proteinvolumenverteilung von MutY gebunden an ungeschädigte DNA mit einer Länge von 2224 bp. Ein Monomer von MutY war bei dem Maximum von 30 nm<sup>3</sup> ±23 nm<sup>3</sup> zu sehen. Für die Auswertung der induzierten Biegewinkel wurde ein Intervall zwischen 20 nm<sup>3</sup> und 40 nm<sup>3</sup> ausgewählt.

Die automatisierte Auswertung mittels MATLAB ergab je nach gewählter Tangentenlänge unterschiedliche Biegewinkel, wie schon für Messungen an DNA in Abwesenheit von Proteinen gezeigt. Aufgrund der relativ geringen Datenanzahl im Vergleich zu der Biegewinkelmessung an DNA-Schäden konnte die Bin-Größe von  $3,5^\circ$  nicht verwendet werden. Stattdessen wurde eine Bin-Größe von  $10^\circ$  vergleichbar der DNA-Längen- bzw. Proteinvolumen-Darstellung gewählt. In Abbildung 97 zeigten sich mit der Tangentenmethode bei einer Segmentlänge von 4 nm drei verschiedene Biegewinkel, bei  $15^\circ \pm 11^\circ$ ,  $35^\circ \pm 4^\circ$  und  $65^\circ \pm 8^\circ$ . Dagegen dominierten bei 8 nm Segmentlänge Biegewinkel von  $26^\circ \pm 18^\circ$  und  $55^\circ \pm 22^\circ$  die Messung.

Bei Wiederholung der Messung der DNA-Biegewinkel an MutY-Proteinkomplexen mit der Sekantenmethode wurde für 4 nm lange Segmente ein Wert von  $13^\circ$  ermittelt. Für die Verwendung von 8 nm langen Segmenten zeigten sich Biegewinkelwerte von  $16^\circ \pm 8^\circ$ ,  $35^\circ \pm 5^\circ$  und  $60^\circ \pm 6^\circ$  (Abbildung 98).

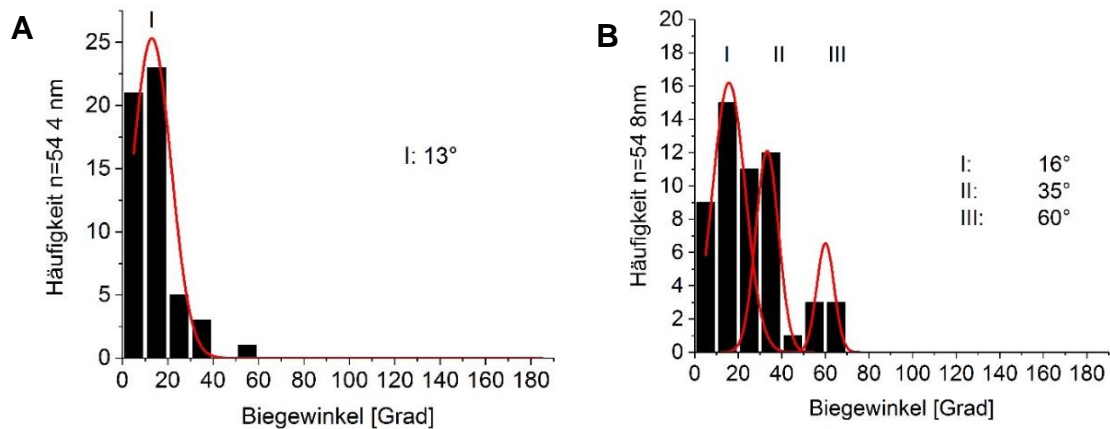
Bei Darstellung der Biegewinkel aller Proteine ungeachtet ihres Volumens mit der Tangentenmethode in Abbildung 99 zeigt sich für die Tangentenmethode ein deutliches Maximum bei  $13^\circ \pm 12^\circ$  und  $46^\circ \pm 18^\circ$  und für die Sekantenmethode entsprechend bei  $19^\circ \pm 13^\circ$  und  $60^\circ \pm 4^\circ$ .

Für den Vergleich der manuellen Biegewinkelmessung gegenüber dem automatisierten Verfahren wurde die Auswirkung von MutY auf die Biegung von ungeschädigter DNA zusätzlich in eigenen Aufnahmen manuell untersucht. Hierfür wurde die Glykosylase MutY mit einer 2224 bp langen ungeschädigten DNA inkubiert. Bei der eigenen manuellen Auswertung mit ImageJ zeigt die Abbildung 100 zwei deutliche Biegungsspezies bei  $15^\circ \pm 13^\circ$  und  $52^\circ \pm 14^\circ$ .



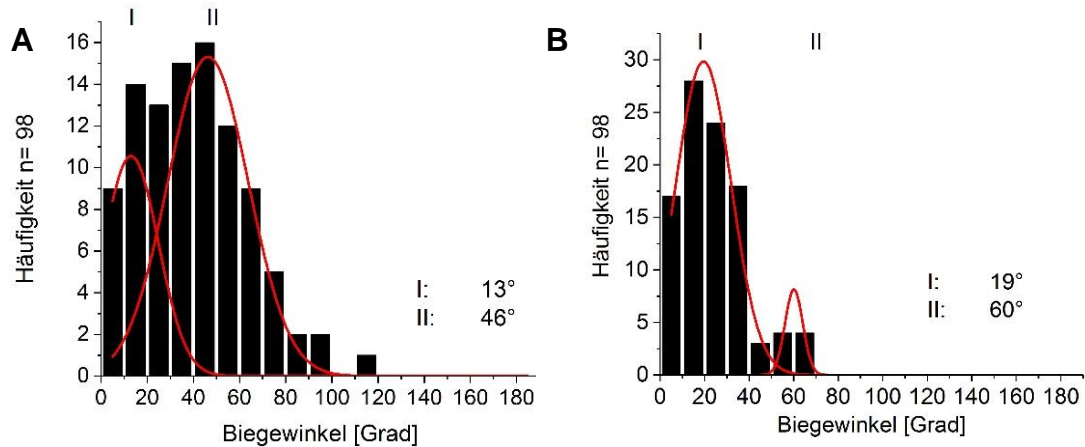
**Abbildung 97: Automatische Biege­winkel­messung an der Proteinposition mit der Tangentenmethode**

Dargestellt sind die Mess­er­geb­nisse der MutY-induzierten Biege­winkel an ungeschädigter DNA. Um die Ergebnisse zu vergleichen wurden dieselben Mess­kri­te­rien wie bei der manuellen Messung angewandt. Die am Protein ermittelten Biege­winkel entsprechend der Nummerierung immer der manuellen Mess­po­si­tion zugeordnet und verglichen. Unter **A** ist die Biege­winkel­ver­tei­lung bei 4 nm Segmentlänge zu sehen. Deutlich erkennbar waren insgesamt drei Biege­winkel von  $15^\circ \pm 11^\circ$ ,  $35^\circ \pm 4^\circ$  und  $65^\circ \pm 8^\circ$ . **B** zeigt bei einer Segmentlänge von 8 nm Biege­winkel von  $26^\circ \pm 18^\circ$  und  $55^\circ \pm 22^\circ$ .



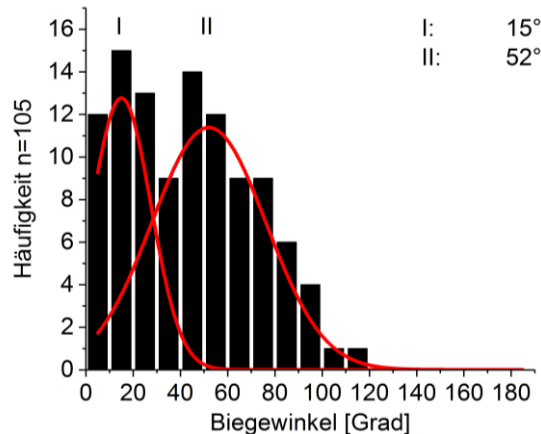
**Abbildung 98: Automatische Biege­winkel­messung an der Proteinposition mit der Sekantenmethode**

Entsprechend dem Vorgehen bei der Tangentenmethode wurden die Biege­winkel­messungen der MutY-induzierten Biege­winkel an ungeschädigter DNA mit der Sekantenmethode wiederholt. **A** zeigt die Biege­winkel­ver­tei­lung bei 4 nm Segmentlänge mit einem Wert von  $13^\circ \pm 9^\circ$ . Unter **B** finden sich die Mess­er­geb­nisse bei einer Segmentlänge von 8 nm mit drei verschiedenen Biege­winkel­n von  $16^\circ \pm 8^\circ$ ,  $35^\circ \pm 5^\circ$  und  $60^\circ \pm 6^\circ$ .



**Abbildung 99: Biege winkelverteilung der automatisierten Biege winkelmessung ohne Berücksichtigung des Proteinvolumens**

Gezeigt ist die Biege winkelmessung an MutY-DNA-Komplexen ohne Berücksichtigung der Proteinvolumina, die jeweils mit einer Segmentlänge von 8 nm gemessenen wurde. Es ergaben sich mit der Tangentenmethode (A) zwei Biegeungsspezies bei  $13^\circ \pm 12^\circ$  und  $46^\circ \pm 18^\circ$ . Entsprechende Biege winkel mit der Sekantenmethode (B) zeigten sich bei  $19^\circ \pm 13^\circ$  und  $60^\circ \pm 4^\circ$ .



**Abbildung 100: Manuelle Messung des Biege winkels an MutY-DNA-Komplexen mit ungeschädigter DNA**

Gezeigt ist die manuelle Messung von MutY-DNA-Komplexen mit ungeschädigter DNA (2224 bp), wobei das Proteinvolumen nicht berücksichtigt wurde. Es lassen sich zwei Biegeungsspezies bei  $15^\circ \pm 13^\circ$  und  $52^\circ \pm 14^\circ$  erkennen.

## 4 Diskussion

### 4.1 Diskussion der Methode

Bei der Vermessung digitaler Bilder muss man sich stets dessen bewusst sein, dass diese nur eine Annäherung an die wirklich vorhandene Situation darstellen. So unterschieden schon Dorst und Smeulders [152] zwischen digitalen und digitalisierten Linien. Im Falle dieser Doktorarbeit entsprechen die rasterkraftmikroskopischen Aufnahmen der DNA digitalisierten Linien, da jede Linie eine Annäherung an einen wirklich vorhandenen DNA-Strang darstellt. Hingegen ist eine digitale Linie nach deren Definition nur eine Aneinanderreihung von digitalen Pixeln, die nicht auf einem wirklich vorhandenen Original basiert. Für die Vermessung der DNA in AFM-Aufnahmen dieser Arbeit wurden die acht Freiheitsgrade der Längenmessung durch den Freeman-Code [153, 154] zugrunde gelegt, auch deshalb müssen die hier gewonnenen Daten stets als Annäherung an wirklich vorhandene DNA-Maße betrachtet werden. Die nun folgende Diskussion soll die angewandte Methode kritisch beurteilen, ihre Vorzüge darstellen und rückblickend auf mögliche Verbesserungen hinweisen.

#### 4.1.1 Voraussetzungen für eine optimale Auswertbarkeit der AFM-Aufnahmen

##### 4.1.1.1 Bildverbessernde Maßnahmen am Rasterkraftmikroskop

Für die hochauflösende Darstellung von DNA und DNA-Proteinkomplexen waren optimale Aufnahmebedingungen notwendig. So sollte das Substrat für die Probedeposition möglichst glatt sein und die Spitze des Tips, der AFM-Sonde, möglichst spitz. Ein scharfer Tip war entscheidend, da die Breite des Tips die gemessene Höhe und Breite der DNA vorgibt. Dabei ist die maximale Auflösung jedoch stets durch den Tip-Konvolutions-Effekt begrenzt [155]. Ein spitzer Tip hat eine durchschnittliche Breite von  $\leq 10$  nm [156]. In diesem Fall konnte mit dem Rasterkraftmikroskop eine maximale Auflösung erzielt werden. Während der Aufnahme verschliss der Tip durch den Abbildungsvorgang allerdings beständig. So wurde die Spitze des Tips stumpf und folglich die abgebildeten Strukturen breiter. Eine unnötig verbreiterte Aufnahme der DNA erschwerte die korrekte Winkelmessung

bei der manuellen Messmethode, da das Platzieren der Winkeltangenten weniger präzise möglich war. Hier war die Nutzung einer automatisierten Messmethode von Vorteil. Allerdings war auch für die automatisierte Winkelmessung eine gute Bildqualität mit möglichst scharfer Darstellung der abzubildenden Objekte wichtig. Dies wurde gewährleistet, indem jede Probendeposition erst mit einem gebrauchten und somit stumpfen Tip als Übersichtsbild mit schneller Rasterrate aufgenommen wurde, bevor die eigentliche Aufnahme eines passenden Depositionsabschnittes mit einem neuen und folglich scharfen Tip durchgeführt wurde. Dieses Vorgehen wurde genutzt, um möglichst reine Abschnitte mit hoher DNA-Dichte sehr detailreich aufzunehmen.

### **4.1.1.2 DNA-Höhen- und -Breitenveränderung in AFM-Aufnahmen**

Die jeweilige Höhe und Breite der DNA in Rasterkraftaufnahmen stellte sehr wichtige Werte für die automatisierte Auswertung dar. Wie bereits ausführlich in Kapitel 3.2.5 beschrieben, wurde ein geeigneter Schwellenwert (Threshold) festgelegt. Diese Maßnahme war nötig, um die DNA-Strukturen in den Rasterkraftaufnahmen für das Tracking-Programm FIESTA zu definieren. Die definierte Mindesthöhe bei Verwendung eines Thresholds war Faktoren wie der Rauigkeit des Glimmersubstrates und dem Bildrauschen geschuldet. War die Höhendifferenz zwischen DNA-Strang und Hintergrund zu gering, so ließ sich die DNA-Struktur nicht mehr ausreichend scharf definiert darstellen. Dies wurde unter anderem durch eine schlechte Aufnahmequalität oder verschmutzte bzw. kontaminierte Depositionen verursacht. In Kapitel 3.2.1 wurde bereits beschrieben, dass die Höhe der DNA-Stränge für eine Auswertung durch die automatische Messmethode in rasterkraftmikroskopischen Aufnahmen durchschnittlich 0,49 nm betragen sollte. Wie dort in Abbildung 32 dargestellt, besteht ein signifikanter Zusammenhang zwischen durchschnittlicher DNA-Höhe und der Auswertbarkeit der jeweiligen AFM-Aufnahme mit der automatisierten Messsoftware.

In der vorliegenden Arbeit betrug die durchschnittlich gemessene DNA-Höhe 0,46 nm  $\pm$ 0,08 nm und die durchschnittliche DNA-Breite 13 nm  $\pm$ 2 nm (Abbildung 31). Ähnliche Werte fanden sich auch in der Literatur. So beschrieben Rivetti *et al.* [143] für die Vermessung der DNA in rasterkraftmikroskopischen Aufnahmen



für Höhe und Breite des DNA-Polymers Werte von durchschnittlich  $0,55 \text{ nm} \pm 0,15 \text{ nm}$  bzw.  $12 \text{ nm} \pm 3 \text{ nm}$ . Bezanilla *et al.* [157] gaben für die Höhe der DNA  $0,43 \text{ nm} \pm 0,08 \text{ nm}$  und für die Breite  $11,2 \text{ nm} \pm 1,8 \text{ nm}$  an. Die Ergebnisse der Messungen an DNA in AFM-Aufnahmen weichen somit vom allgemein bekannten Wert für den Durchmesser der DNA ab. In der Literatur sind für den Durchmesser B-förmiger DNA  $20,4 \text{ \AA}$  angegeben, was etwa  $2 \text{ nm}$  entspricht [158, 159]. Die Abweichungen von diesem Wert, also die Zunahme der DNA-Breite und Verringerung der DNA-Höhe, sind für rasterkraftmikroskopische Aufnahmen bekannt und wurden in der Literatur vielfach diskutiert [160]. Als Ursache für diese Abweichung des Durchmessers wird zum einen der Verbreiterungseffekt der Cantileverspitze aufgeführt und zum anderen werden für die geringere DNA-Höhe Wechselwirkungen der DNA sowohl mit der Cantileverspitze als auch mit der Glimmeroberfläche angegeben [140, 143, 161-163]. Die Verbreiterung der aufgenommenen DNA ist als Tip-Konvolutions-Effekt bekannt und wurde bereits in Kapitel 4.1.1.1 erwähnt [164]. Die verringerte Höhe ist eine Folge der mechanischen Kompression während des Abbildungsvorgangs im Tapping-Modus. Außerdem erscheint die Partikel-Höhe wegen des überlagerten Flüssigkeitsfilms in den getrockneten Proben etwas verringert.

### 4.1.2 DNA-Längeneffekte

#### 4.1.2.1 Einfluss auf die Schadenslokalisierung

Für eine schadensspezifische Winkelmessung musste sichergestellt werden, dass sich der Schaden bei genau 50 % der DNA-Länge befand. Hierfür wurde anhand der bei der DNA-Präparation genutzten Basensequenzen und Restriktionsschnittstellen die genaue Lage des Schadens ermittelt. Da bei der Herstellung der schadenstragenden DNA-Stränge immer die gleichen Restriktionsenzyme verwendet wurden, besaßen die DNA-Stränge auch immer dieselbe Länge. Durch das Wissen um die genauen Stellen der Schadensinsertion und die Verwendung der immer gleichen Restriktionsenzyme konnte die Lage des Schadens basengenau errechnet werden. Für die Insertion der DNA-Schäden wurden jedoch je nach Basenschaden geringfügig verschiedene Oligonukleotide verwendet. Die genaue Basenabfolge der inserierten Oligonukleotide sind dem Kapitel

2.1.8 zu entnehmen. Bei der Herstellung der Basenschäden kam es somit abhängig von den ligierten Oligonukleotiden zu geringen Verschiebungen des Basenschadens um wenige Basenpaare im Bereich der 50 %-Position. Bedingt durch diesen Umstand bestand zwischen den unterschiedlichen Schäden eine sehr geringe Varianz der Schadenslage von 0,15 bis 1,55 nm. Durch die Verwendung der Restriktionsenzyme NdeI und BsaXI ergab sich eine konstante DNA-Länge von 505 bp, wobei sich die Mitte bei 202,5 bp befand. Metrisch entsprach eine Länge von 505 bp 171,5 nm bzw. 85,75 nm für die halbe Länge (1 bp entspricht 0,34 nm). In der folgenden Tabelle 5 ist für jedes der untersuchten Schadenssubstrate die genaue Schadenslage in Bezug auf die Gesamtlänge aufgelistet. Die Länge wurde immer von 5' nach 3' gezählt. Es zeigte sich eine geringe, mit der Messmethode nicht auflösbare Abweichung von der 50 %-Position.

**Tabelle 5: Schnittstellen der DNA-Stränge**

Schadenssubstrat	Restriktionsenzyme	Schadenslage [bp/ nm/ %]	Abweichung vom Mittelpunkt [ $\Delta$ nm]
oxoG:C	NdeI / BsaXI	254 / 86,4 / 50,3	0,51
G:A	NdeI / BsaXI	257 / 87,4 / 50,9	1,55
oxoG:A	NdeI / BsaXI	252 / 85,7 / 50,0	0,15
$\epsilon$ A:T	NdeI / BsaXI	255 / 86,7 / 50,5	0,85

Bei den in dieser Arbeit verwendeten Messmethoden zur Biegewinkelbestimmung an AFM-Aufnahmen, ist aus den zuvor erläuterten Gründen die Kenntnis der DNA-Länge und der daraus zu ermittelnden Basenschadenslage von sehr großer Bedeutung für die Aussagekraft der Ergebnisse. Die ausgewerteten Aufnahmen sollten daher eine bezüglich der Länge möglichst homogene DNA-Spezies aufzeigen. Dabei sollte die Länge der dargestellten Filamente möglichst der errechneten DNA-Länge entsprechen. Durch die Darstellung eines Längenhistogramms in MATLAB und die Bestimmung einer Ziellänge, von welcher die Länge der ausgewerteten DNA-Stränge nur um  $\pm 20$  nm abweichen durfte, wurde sichergestellt, dass das ausgewertete DNA-Filamente die korrekte Ziellänge hatte und somit der gemessene Biegewinkel an der 50 %-Position dem Winkel am DNA-Schaden entsprach.

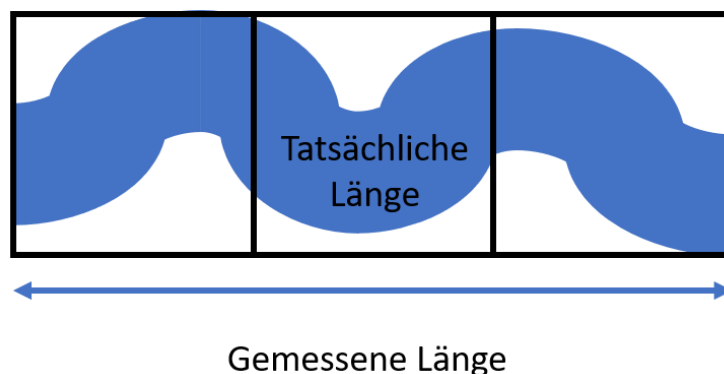
### 4.1.2.2 Verkürzung der DNA durch die Probendeposition

Es kommen mehrere mögliche Fehlerquellen für Abweichungen der DNA-Länge in Frage. Lagerungsbedingte Schäden oder Strangbrüche am Schadens-Locus durch die bestehende Instabilität dieser DNA-Region könnten mögliche Ursachen für eine DNA-Verkürzung darstellen. Brüche am Schadensort würden sich allerdings aufgrund der speziellen Schadenslage bei 50 % durch 86 nm lange DNA-Stränge bemerkbar machen. Durch die mit der Biegewinkelbestimmung einhergehende Längenmessung konnten die zuvor beschriebenen Fehlerquellen ausgeschlossen werden. Auch eine Beschädigung der DNA während des Gefriervorganges ist durch die Kryokonservierung mit flüssigem Stickstoff und die Lagerung bei -80 °C als unwahrscheinlich anzusehen, da es sich in diesem Fall um 505 bp lange DNA handelte und die Gefahr einer solchen Schädigung erst ab Längen von 2000 bp beschrieben wurde [165].

Auch die Depositionsweise könnte zu einem Bruch oder einer Verformung der DNA-Helix führen, wenn der Glimmer mit zu starkem Luftdruck getrocknet wurde [166]. Die B-Form der DNA Helix könnte durch die Trocknung und die Wechselwirkung mit bivalenten Metallionen, wie  $Mg^{2+}$ , bei der Bindung an die Glimmeroberfläche gestört werden [167]. Zudem könnten intramolekulare Wechselwirkungen mit Faltungsvorgängen des DNA-Stranges einhergehen, was die Genauigkeit der Längenmessung beeinflusst [168]. Rivetti *et al.* [143] haben jedoch festgestellt, dass  $Mg^{2+}$ -Ionen nur so schwache Wechselwirkungen zwischen der DNA und der Glimmer-Oberfläche erzeugen, dass sich die DNA ausreichend equilibrieren kann und es folglich zu keinen statischen Veränderungen kommt. Neben der Beeinträchtigung der B-Form und der mechanischen Flexibilität, deren Ausmaß sich in der Biegung der DNA bei der Aufnahme mit dem Rasterkraftmikroskop zeigte, führten auch die trockenen Aufnahmebedingungen zu einer Verkürzung der DNA-Länge [169].

#### 4.1.2.3 Veränderung der DNA-Länge während des manuellen Messverfahrens

Allgemein wurde eine auflösungsbedingte Verkürzung der DNA von bis zu 10 % als tolerabel angesehen. In der Literatur sind für die Vermessung von DNA mit dem Rasterkraftmikroskop Abweichungen zwischen der theoretischen und der gemessenen Länge von bis zu 8 % bekannt [143, 170]. Eine Ursache für diese Längendifferenz war auflösungsbedingt. Minimale Biegungen, die sich innerhalb eines Pixels befanden, waren durch die Auflösung rasterkraftmikroskopischer Aufnahmen von 1,95 nm/Pixel nicht aufzulösen [143] (Abbildung 101). Neben der Pixelauflösung war der Tip-Durchmesser ( $\leq 10$  nm [156]) ein wichtiger Faktor. Bei der manuellen Messmethode war methodisch bedingt zusätzlich mit einer Messungenauigkeit von mindestens einem Pixel je DNA-Strang-Ende zu rechnen. Bei einer Auflösung von 1,95 nm/Pixel entsprach eine Abweichung von einem Pixel bei einer theoretischen Gesamtlänge von 171 nm etwa einer messbedingten Abweichung von ca. 2 %.



**Abbildung 101: Darstellung der auflösungsbedingten Längenverkürzung**

Die Abbildung soll beispielhaft das Prinzip der Auflösungsverkürzung darstellen. Gezeigt werden drei Pixel (schwarze Rechtecke) mit DNA-Strang (blau). Feine Biegungen innerhalb der Auflösung sind nicht darstellbar. (Die Abbildung ist nicht maßstabsgerecht!)

#### **4.1.2.4 Veränderung der DNA-Länge während des automatisierten Messverfahrens**

##### *FIESTA*

Die Skelettierung der DNA mit FIESTA durch kurze Segmente (2 nm lang) entlang des errechneten Signalmaximums für jeden DNA-Strang war ein wichtiger Schritt für die objektive Charakterisierung der aufgenommenen DNA. Diese automatische Skelettierung durch FIESTA eignete sich als objektive Methode zur Längenmessung. Da FIESTA abhängig von der gewählten Halbwertsbreite immer gleich weit in die Enden der DNA-Stränge skelettierte, wurde auch die Länge standardisiert errechnet. Jedoch zeigte sich, dass FIESTA den Filamentierungsvorgang nicht vollständig bis in die DNA-Enden durchführte. Bei der darauffolgenden Auswertung der DNA-Länge wurde von MATLAB durchschnittlich ein Wert um 150 nm gemessen, obwohl die DNA-Stränge eine theoretische Länge von 171 nm besitzen müssten. Dies entspricht einer verkürzten Darstellung von ca. 12 % im Vergleich zu 10 % Verkürzung bei der manuellen Messmethode. Diese Abweichung zwischen gemessener und errechneter Länge muss als zusätzlicher methodischer Fehler von FIESTA angesehen werden. Allgemein bestand bei zu gering definierter Halbwertsbreite die Möglichkeit, dass die DNA durch FIESTA zu kurz skelettiert wurde und somit MATLAB zu kurze Längenergebnisse ausgab. Wie bereits im vorhergehenden Kapitel (4.1.2.3) beschrieben, wurden minimale Biegungen während des Skelettierungsvorganges auflösungsbedingt auch durch FIESTA nicht berücksichtigt. Hinzu kommt, dass es im Bereich von starken Biegungen durch die 2 nm großen Skelettierungsfilamente zu einer gewissen Verrundung und somit zu weiteren Längeneinbußen kam. Die Schadenslage bei 50 % der DNA-Länge ist für die automatische Auswertung sehr von Vorteil, denn gerade durch diesen Umstand hatte eine methodisch zu kurze Messung an den DNA-Enden auf die 50 %-Position keinen Einfluss.

### *MATLAB*

Bei zusätzlichem Auftreten von ausgeprägten Hintergrundsignalen, wurde die Auswertung der Aufnahme ebenfalls beeinflusst, da die Partikel teilweise mit der DNA im Bild verschmolzen und folglich fälschlicherweise in die Auswertung mit einbezogen wurden, was zu einer scheinbaren Verlängerung der DNA führte. Um den Einfluss der geschilderten Fehlerquellen auf die Auswertung zu minimieren, wurden ausschließlich diejenigen DNA-Stränge ausgewertet, welche sich in einem Längenbereich von  $\pm 20$  nm um das Häufigkeitsmaximum befanden. So konnten alle eindeutig zu kurze und zu lange DNA-Stränge von der Auswertung ausgeschlossen werden. Bei der Software-gestützten automatisierten Auswertung wurde von MATLAB die zu erwartende Länge abgefragt und ein Längenhistogramm erstellt. Die Daten aller segmentierter DNA-Stränge wurden nach Filamenten gefiltert, die weniger als zwei Standardabweichungen von der erwarteten Länge abwichen. Für die folgende Biegewinkelmessung wurden somit nur diese Filamente berücksichtigt. Auf diese Weise konnte sichergestellt werden, dass die 50 %-Länge korrekt bestimmt wurde und diese folglich der Lokalisation des Schadens entsprach, was eine korrekte Auswertung der Biegeungswinkel ermöglichte.

### **4.1.3 Möglichkeiten der Schadenslokalisations-Kontrolle**

Bei rasterkraftmikroskopischen Darstellungen ist der Basenschaden an der 50 %-Position nicht direkt sichtbar, jedoch gewährleisteten Kontrollassays während der Herstellung der DNA-Substrate die korrekte Schadensintegration. Ein Ansatz für die direkte Kontrolle der Schadenslokalisierung könnte die Fluoreszenz-Markierung des Basenschadens sein [171, 172]. Dies wäre eine zusätzliche Kontrollmöglichkeit der richtigen Messstelle unabhängig von der korrekten Länge und deren 50 %-Position. In diesem Fall könnte man komplementär zur geschädigten Base ein Nukleosidanalogen einfügen, welches dann mit Hilfe eines Fluorophors markiert werden kann. Jedoch kann auch diese Modifikation der Basenabfolge zu neuen sterischen Veränderungen der DNA führen und somit den Biegeungswinkel beeinflussen. Deshalb muss auch die Umsetzung dieser Methode kritisch

betrachtet werden [173], gerade wenn spezifische DNA-Schäden untersucht werden sollen.

Bei Untersuchung von Protein-DNA-Wechselwirkungen wäre die zusätzliche Markierung der gebundenen Glykosylasen eine Option. Die Fluoreszenzmarkierung der Glykosylasen könnte vor der Inkubation mit der DNA stattfinden, wobei die Bindungseigenschaften der Glykosylasen nicht beeinträchtigt werden sollten, oder der Fluoreszenzfarbstoff bindet in einer Kopplungsreaktion an die bereits an DNA-Schäden fixierten Glykosylasen. Milles und Lemke [174] stellten in ihrer Publikation eine breite Übersicht verschiedener gängiger Möglichkeiten der Fluoreszenzmarkierung vor. Dabei erscheint die Verwendung der relativ kleinen Maleimid- und Succinimidylester, die an Cystein- bzw. Lysinreste binden, als eine geeignete Möglichkeit zur Markierung von Glykosylasen. Einen möglichen Ansatz für die Fluoreszenz-Markierung von Einzelmolekülen zeigten Lee *et al.* [175] für ROS-induzierte Schäden. Fronczek *et al.* [138] stellten mit der Verknüpfung der beiden Methoden von AFM und FIONA (engl. Fluorescence Imaging with One-Nanometer Accuracy; Fluoreszenzbildgebung mit Nanometerpräzision) eine Methode vor, die mit einer Genauigkeit von bis zu 8 nm verschiedene Proteine bestimmen und lokalisieren konnte. Ein weiteres Beispiel für die Kombination aus Rasterkraftmikroskopie und Fluoreszenz-Messungen zeigten Hards *et al.* [171]. Die Fluoreszenzmarkierung von Glykosylasen könnte eine Volumensmessung für die Proteinbestimmung überflüssig machen und bei Bindung an den Basenschaden die Schadenslokalisierung zusätzlich zu den Kontrollassays verifizieren.

#### 4.1.4 DNA-Biegewinkeleffekte

##### 4.1.4.1 Mögliche Einflussfaktoren bei der Probenherstellung

Ein großer Vorteil der Rasterkraftmikroskopie gegenüber anderen Messmethoden ist, dass jeder DNA-Strang einzeln gemessen wird und so über die Verteilung der Biegewinkel mehr Informationen gewonnen werden können. Eine eher breite Verteilung im Histogramm spiegelt eine erhöhte Flexibilität am Schadensort wider, während eine Biegewinkelverteilung mit mehreren Biegungsmaxima den Rückschluss auf verschiedene Biegungszustände erlaubt [176]. Durch die Aufnahme mit dem Rasterkraftmikroskop wird ein dreidimensionales DNA-Polymer als zweidimensionale Struktur abgebildet. Die Höheninformation wird entweder durch Grauwerte oder einen Farbcode dargestellt. Die in AFM-Aufnahmen gemessenen Winkel stellen nicht zwingend die tatsächliche Biegung der DNA dar. Eine mögliche Ursache für die dargestellte Biegung ist eine erhöhte Flexibilität am Schadensort, wie in der Einleitung bereits erläutert. So führen die in der Watson-Crick-Basenpaarung ausgebildeten Wasserstoffbrückenbindungen zu einer sehr stabilen DNA-Doppelhelix, die sich weniger verbiegen kann, bzw. kaum flexibel ist. Ein Maß für die Steifigkeit einer Polymerkette ist die Persistenzlänge, die bei DNA von der Salzkonzentration abhängig ist [177]. So können positiv geladene Oberflächen, welche für die Fixierung von Proben in der Rasterkraftmikroskopie ihre Anwendung finden, eine Erhöhung der Flexibilität von DNA bedingen [178]. Bei intakter doppelsträngiger DNA (dsDNA) beträgt die Persistenzlänge ~50 nm [143, 178] im Vergleich zu ~1 nm bei einzelsträngiger DNA (ssDNA) [179]. Ein Grund für die Beeinflussung der Flexibilität von DNA und damit einhergehend der gemessenen Biegewinkel könnte folglich in der Art der Proben deposition gesehen werden [180]. Ist die Basenpaarung gestört, kann es auch zur Bildung eines Wobble-Basenpaars mit stark erhöhter Elastizität in der Doppelhelix kommen. Bei der Deposition von schadhafter DNA könnte sich allgemein die Elastizität am Schaden bei der Bindung des negativ geladenen DNA-Rückgrates an die negativ geladene Glimmroberfläche über die bivalenten  $Mg^{2+}$ -Ionen durch eine Verbiegung äußern [178, 181-185]. Rouzina und Bloomfield [167] beschrieben in ihrem mathematischen Simulationsmodell eine elektrostatisch bedingte Biegung von B-förmiger DNA durch multivalente Ionen wie  $Mg^{2+}$



von 20° – 40° in Lösung. Auch durch die Trocknung der DNA-Suspension unter dem Stickstoffgasstrom könnte die Biegung der DNA prinzipiell artifiziell erzeugt worden sein. Gerade vom Strukturdefekt begünstigt, wäre eine Biegung an der 50 %-Position wahrscheinlicher als für ungeschädigte DNA. Dem Abwasch-Vorgang ist ein ähnlicher Effekt für artifizielle Biegungen beizumessen [121]. Ein möglicher Ansatz zur Untersuchung dieser Überlegung ist, die Biegung von geschädigter DNA an der 50 %-Position und an einer ungeschädigten Position zu vergleichen. So könnte man einen möglichen Einfluss des Depositionsverfahrens genauer untersuchen. Nach Schulz *et al.* [147] haben der Abwasch- und Trocknungsvorgang jedoch nur einen geringen Einfluss auf den Biegungswinkel, wobei aber in ungeschädigter DNA eine geringgradige Biegung induziert wurde. Die mögliche Beeinflussung der Biegewinkel durch das Depositionsverfahren könnte neben einer konkreten Biegung auch in einer Verbreiterung der Biegewinkelverteilung resultieren [143]. Daneben können Abweichungen in den Biegewinkelergebnissen der AFM-Aufnahmen auch mit unterschiedlichen DNA-Sequenzabfolgen (Wechselwirkungen mit den benachbarten Basen) und Puffereigenschaften begründet werden [186, 187]. Um den Einfluss der Experimentator-abhängigen Deposition zu untersuchen, wurden in der vorliegenden Arbeit zusätzlich zu eigenen Proben auch Aufnahmen von Lisa Hirsch vom automatisierten Messprogramm ausgewertet. Es wurden beispielsweise bei oxoG:A-geschädigter DNA ähnliche Biegewinkel für die beiden unabhängig produzierten Proben gemessen, so dass der Einfluss als nicht erheblich angesehen wurde (Abbildung 115). Interessant wäre auch die kurze und lange (ungeschädigte) DNA aus Plasmiden innerhalb eines Präparationsprozesses zu vermessen und nach vorhandenen Abweichungen der Biegewinkel zu untersuchen, um so mögliche sequenz- und präparationsbedingte Unterschiede zu erkennen. Eine Untersuchung des Sequenzeinflusses wurde durch die AG Tessmer bereits durchgeführt. So wurde der Biegewinkel an ungeschädigter DNA in 50 nm-Abständen gemessen. Es ergab sich eine für ungeschädigte DNA übliche Biegewinkelverteilung, die mit der Biegung an ungeschädigter DNA bei 50% der 505 bp Fragmente übereinstimmte. Auf ähnliche Weise könnte die Überprüfung auf die anderen Präparationsschritte ausgeweitet werden, um weitere potenzielle Einflüsse zu bewerten.

### 4.1.4.2 Beeinflussung des Biegewinkels bei der manuellen Auswertung

Eine Unsicherheit der manuellen Messmethode in ImageJ war die Tatsache, dass keine definierte Vergrößerung für die DNA und keine eindeutige Länge sowie Breite für die Tangenten festgelegt war. Demnach würden bei geringer Vergrößerung die Tangenten tendenziell länger ausfallen, als dies bei einer stärkeren Vergrößerung der Fall wäre. Auch die Qualität und Auflösung einer Aufnahme könnte die manuelle Biegewinkelmessung beeinflussen. Theoretisch kann man die Biegewinkel auch nur an zwei Pixeln an der 50 %-Position messen. Allein die Variabilität in der Tangentenlänge und die daraus resultierende unterschiedliche Lage der Tangenten im DNA-Strang machte die manuelle Messmethode angreifbar [176]. Auch Rivetti *et al.* [155] diskutierten, dass die nicht ausreichend definierte Positionierung der Tangente zu Experimentator-abhängigen Ergebnissen führen kann. Sinnvoll wäre es unter diesem Gesichtspunkt auch die manuelle Messung bei definierter Vergrößerung und Tangentenlänge zu wiederholen. Hierbei könnte die DNA-Breite zur Eichung der Länge und Breite der Messlinie verwendet werden. Aus diesem Grund ist im Rahmen der automatisierten Messmethode die Erzeugung einer Filamentierungslinie, die sich an dem Maximum der Gaußverteilung der DNA-Breite orientiert, ein wichtiger erster Schritt für eine objektive Messung der Biegungswinkel bei 50 % der DNA-Länge (DNA-Schaden).

### 4.1.4.3 Beeinflussung des Biegewinkels bei der automatisierten Auswertung

Die einzelnen DNA-Stränge wurden durch FIESTA, wie bereits im Kapitel 2.2.7 beschrieben, in 2 nm große Filamente skelettiert und die DNA-Stränge im Anschluss als Polymere (aus 2 nm Filamenten) gespeichert. Anschließend hat MATLAB zusätzlich eine Polynomzuanpassung an den Filamentierungslinien durchgeführt. Für die Messung der Biegewinkel wurden in MATLAB zwei verschiedene Messansätze (Sekante und Tangente) verglichen. Der Abstand der Messpunkte zu der 50 %-Position bzw. dem Protein konnte durch die Länge der Messgeraden variabel definiert werden. Für die beiden Methoden wurde jeweils eine Länge von 4 nm und 8 nm eingesetzt und miteinander verglichen.

### *Einfluss der Messgeradenlänge*

Die Auflösung der Rasterkraftaufnahmen betrug, wie im Methodenkapitel 2.2.2.2 beschrieben, 1,95 nm/Pixel. Gemäß dieser Auflösung wurde bei der Winkelmessung eine Länge zwischen zwei (4 nm) und vier (8 nm) Pixel für die Messgeraden verwendet, da sich die Messgeraden in diesem Fall an der Pixelgröße orientierten. Zwei Messgeraden mit jeweils einer Länge von zwei Pixeln suggerierten theoretisch eine sehr genaue Messung. Mögliche kleine Biegungen innerhalb der 2 nm Auflösung könnten allerdings trotzdem nicht aufgelöst werden, zudem könnten durch den Trackingvorgang oder die Polynomzugfunktion artifiziell erzeugte Fehler einen größeren Einfluss auf die Biegewinkelergebnisse haben, im Vergleich zu 8 nm bzw. 4 Pixel langen Messgeraden. Durch die Auflösung von 1,95 nm/Pixel (1 bp entspricht 0,34 nm) war es ohnehin nicht möglich eine lokale Biegung zwischen zwei oder drei Basen zu erfassen. Eine Messgeradenlänge von 2 nm eignete sich folglich nicht zur Biegewinkelmessung. Ziel war es eine durch den Basenschaden verursachte Krümmung (20 bp) in den Rasterkraftaufnahmen zu ermitteln. Goodsell und Dickerson [188] unterschieden beispielsweise zwischen einer lokalen Biegung (2-3 bp; darstellbar durch NMR und Röntgenkristallographie) und einer Krümmung der DNA ( $\leq 20$  bp; darstellbar durch Gelelektrophorese). Wie in Kapitel 3.2.7.1 bereits dargestellt, lieferten die 4 nm und 6 nm langen Messgeraden vergleichbare Biegewinkelergebnisse. Mit zunehmender Länge der Messgeraden (15 nm und 20 nm) nahm die Breite der Biegewinkelverteilung zu, da die vom Basenschaden ausgehende Biegung zunehmend durch die Biegung von ungeschädigter DNA überlagert wurde. Chen *et al.* [122] verwendeten in Untersuchungen zum Biegeverhalten von hOGG1 für die Vermessung von oxoG-geschädigter DNA 5 nm lange Messtangenten, da ein hOGG1-oxoG:C-Komplex auf der DNA diese Größe einnahm. Der für gebundene MutY-Proteine bestimmte Durchmesser betrug durchschnittlich 14 nm. Deshalb war im Fall dieser Arbeit die Segmentlänge von 8 nm für die automatisierte Biegewinkelmessung an DNA-Protein-Komplexen von MutY bei Berücksichtigung des Proteindurchmessers ideal gewählt, da die Messgeraden beidseitig aus dem Protein herausragen sollten. Außerdem wurde durch die Verwendung der 8 nm langen Messgeraden sichergestellt, dass das Programm universell für die

Biegewinkel-Messung an verschiedenen DNA-Glykosylase-Komplexen verwendet werden konnte. Da zudem die Ergebnisse mit 8 nm langen Messgeraden den manuellen Werten und den in der Literatur aufgeführten Werten am nächsten kamen, wurden die Biegewinkel der 8 nm Tangentenmessung besonders berücksichtigt.

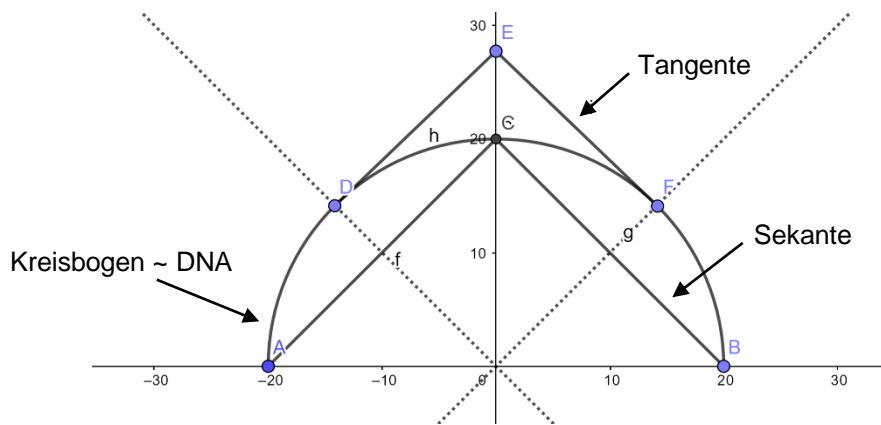
### *Vergleich der Sekante und Tangente*

Die Intention der beiden verschiedenen Messverfahren zur Winkelbestimmung war einen Ansatz zu finden, der dem manuellen Messverfahren entsprechende Werte ermittelte. Die Werte wurden außerdem mit den in der Literatur angegebenen Biegewinkeln verglichen. Die manuelle Vermessung mit der Tangentenmethode (engl. manual tangent overlay) ist eine geläufige Methode für die Biegewinkelbestimmung von DNA-Molekülen in rasterkraftmikroskopischen Aufnahmen [1, 118, 119, 121, 143, 144, 176, 189].

Theoretisch war eine Tangentenmessung der manuellen Methode am ähnlichsten. Jedoch war bei einer automatischen Biegewinkelmessung nahe des Auflösungsmaximums (sehr große Vergrößerung) nicht klar, ob die Tangente auch unter diesen Bedingungen die beste Messmethode darstellte. Unter der Annahme, dass bei der manuellen Vermessung durch eine geringere Bildvergrößerung längere Tangenten verwendet wurden, kann gefolgert werden, dass die Messgeraden initial den DNA-Strang schneiden. Per Definition entspricht dies einer Sekante, da die Messgeraden eine Kurve in zwei Punkten schneiden. Folglich entspräche die manuelle Tangentenmethode („manual tangent overlay“) in Kombination mit langen Messgeraden bei genauer Betrachtung eigentlich einer Sekante. Jedoch lieferte die automatische Messung mit der Sekantenmethode für die verschiedenen DNA-Substrate Werte, die mit den aus der Literatur bekannten Biegewinkelwerten nicht übereinstimmten. Deshalb wurde die Tangente für die automatische Biegewinkelmessung an DNA als der bessere Messansatz bewertet.

Bei Vergleich der beiden Messansätze wurden für sehr kurze Messgeraden von beispielsweise 4 nm (entspricht jeweils zwei Pixel) relativ ähnliche Ergebnisse erwartet. Allerdings war eine generelle Übereinstimmung der

Biegewinkelverteilung bei den Ergebnissen von 8 nm Sekante und 4 nm Tangente zu beobachten. Beispielsweise wurde für oxoG:A mit der Tangentenmethode bei einer Segmentlänge von 4 nm am Schadensort eine Biegung der DNA von  $5^\circ \pm 14^\circ$  gemessen. Die Sekantenmethode zeigte bei einer Segmentlänge von 8 nm ebenso einen Biegewinkel von  $5^\circ \pm 15^\circ$  (Abbildung 60 und Abbildung 61). Diese Beobachtung lässt sich durch die Ähnlichkeit der Steigungen beider Geraden erklären, was nachfolgend in Abbildung 102 veranschaulicht ist.



**Abbildung 102: Geometrische Darstellung der Beziehung Tangente – Sekante**

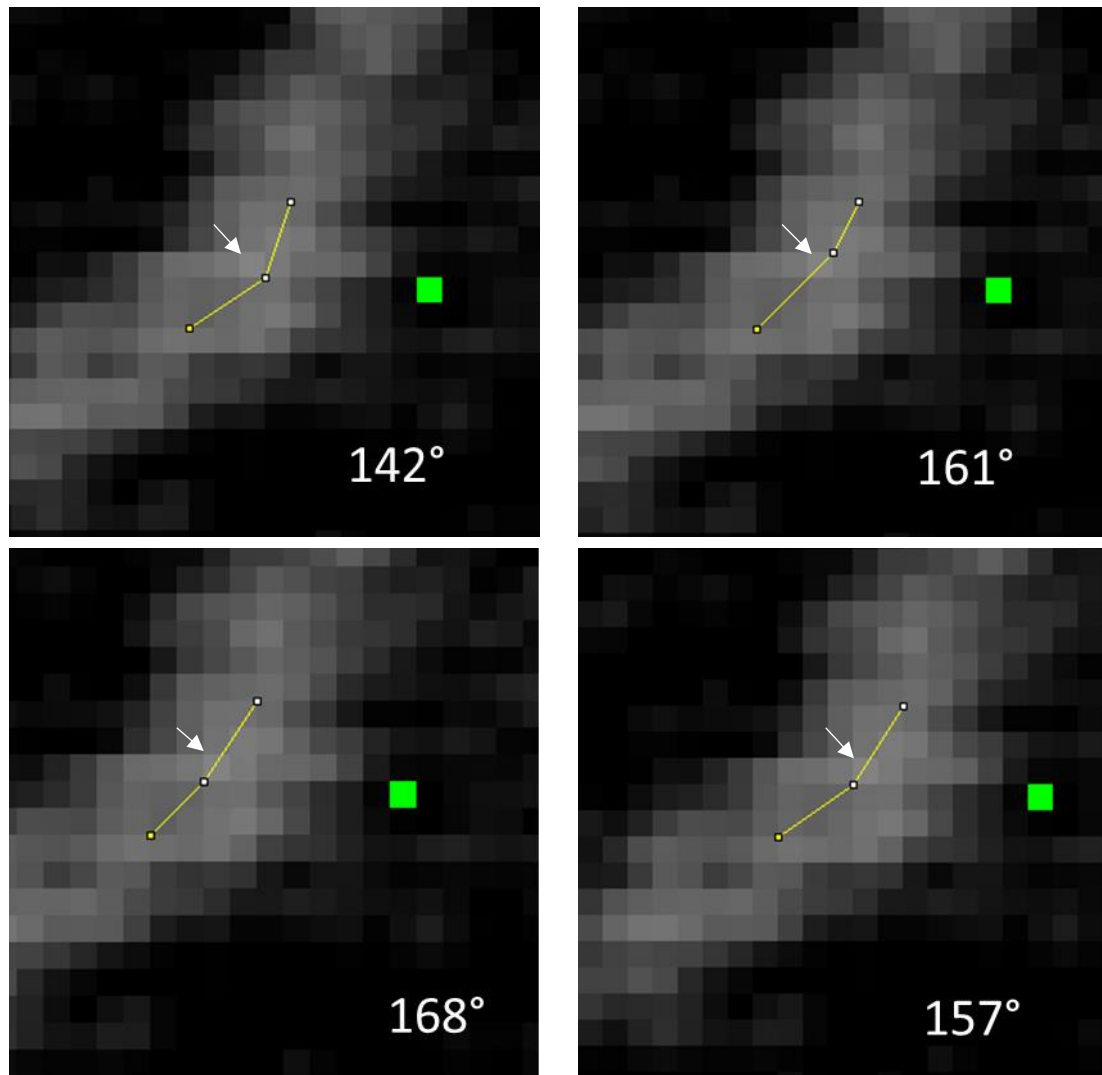
Die Abbildung zeigt schematisch anhand eines Halbkreises die Lagebeziehung zwischen Tangenten und Sekanten bei der Biegewinkelbestimmung. Hierbei soll speziell auf die vergleichbare Steigung einer Tangente (4 nm) und einer Sekante doppelter Länge (8 nm) (Kreisbogenlänge) eingegangen werden.

#### *Vorzüge der automatisierten gegenüber der manuellen Messmethode*

Es sollte zunächst beachtet werden, dass auch die automatisierte Messung mittels Tangenten und die manuelle Messung von DNA-Biegewinkeln teilweise gering differierende Ergebnisse lieferte. Zusätzlich zu den zuvor diskutierten Aspekten der manuellen Messmethode könnten weitere Faktoren eine mögliche Ursache hierfür darstellen. Wie bereits in Kapitel 4.1.4.2 erwähnt, zeigte sich, dass die Positionierung der Tangente innerhalb der manuellen Messmethode nur unzureichend definiert ist [155]. So stellt die Variabilität der Messgeradenlängen eine mögliche Ursache der unterschiedlichen Messergebnisse dar. Abhängig von der Bildqualität kann die Bildvergrößerung und folglich auch die Messgeradenlänge variieren. Während bei der Auswertung von AFM-Aufnahmen guter Bildqualität tendenziell eine geringe Vergrößerung ausreichend war, war bei verminderter Bildqualität eine stärkere Vergrößerung nötig. Hinzu kommt, dass der

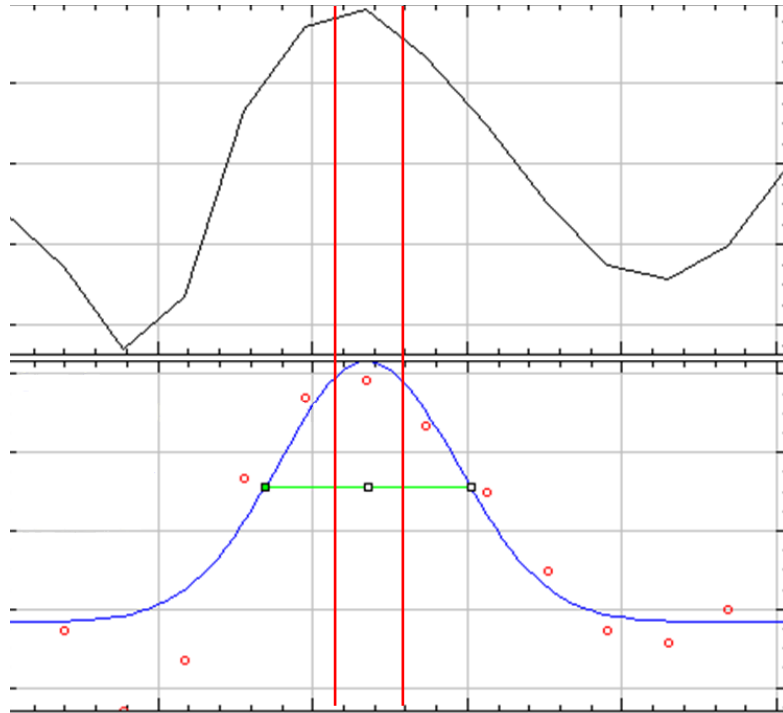
weiteren Verlauf des DNA-Stranges die Tangentenpositionierung möglicherweise unterbewusst beeinflusste. Weiterhin bedingt ImageJ Abweichungen der Biege-  
winkelergebnisse zwischen beiden Messverfahren. So ist die Positionierung des  
Tangenten-Scheitelpunktes am Schadensort in ImageJ nur an den Pixelgren-  
zen/-ecken möglich (Abbildung 103). Außerdem unterliegt der vom Experimenta-  
tor eingezeichnete Mittelpunkt des DNA-Stranges (Abbildung 103, grünes Pixel)  
bei der manuellen Methode einer großen Variation. Hierbei ist die automatische  
Methode durch FIESTA/MATLAB präziser, da die Halbwertsbreite des DNA-  
Stranges mittels einer Gauß-Anpassung bestimmt wird. Das errechnete Höhen-  
maximum dient der exakten Konstruktion des DNA-Scheitelpunktes für die Bie-  
gewinkelmessung (Abbildung 104). Die Darstellung der DNA in Form von seg-  
mentierten Filamenten durch FIESTA ermöglicht eine weitaus bessere Auflösung  
im Vergleich zur AFM-bedingten Pixelauflösung von 1,95 nm/Pixel und erlaubt so  
auch eine genau errechnete und pixelunabhängige 50 %-Positionierung des  
Scheitelpunktes. Auch die Variabilität der Ergebnisse in Abhängigkeit vom Expe-  
rimentator lässt sich somit deutlich minimieren. Gering differierende Biege-  
winkel-  
ergebnisse zwischen automatischer und manueller Messmethode könnten durch  
die aufgeführten Faktoren begründet werden. Weiterhin besteht erst durch die  
automatisierte Messmethode die Möglichkeit Biege-  
winkel differenzierter abzubilden. Die verfeinerte Darstellung der Biege-  
winkelverteilung ermöglicht die ver-  
schiedenen vorkommenden Konformationen eines DNA-Schadens besser ana-  
lysieren zu können.

Anhand der Biege-  
winkel-Verteilung von oxoG:A sollen exemplarisch Ergebnisse  
der 8 nm Tangentenmethode mit dem manuellen Verfahren verglichen werden.  
Beispielhaft werden in der nachfolgenden Abbildung 105 einzelne FIESTA-ske-  
lettierte DNA-Stränge aus der MATLAB-Auswertung gezeigt. Auffällig war, dass  
gerade bei den Beispielen zu 0° in der manuellen Messung tendenziell größere  
Biege-  
winkel gemessen wurden. In allen anderen Bereichen (13°, 25°, 40°) wur-  
den die Biege-  
winkel bei der manuellen Messung tendenziell unterschätzt.



**Abbildung 103: Limitation manueller Biegewinkelmessung an einem DNA-Strang**

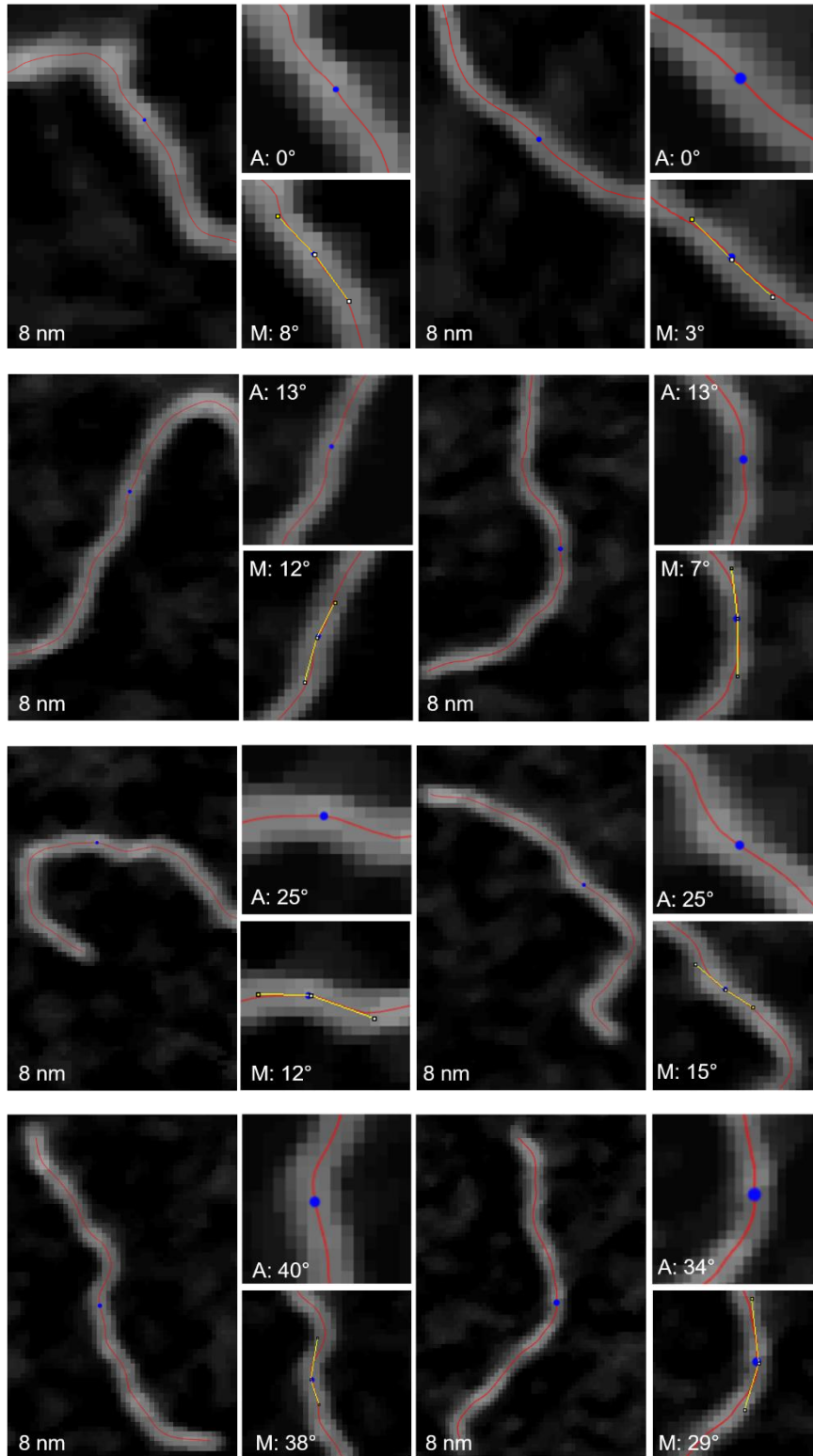
Die vier Abbildungen zeigen denselben DNA-Ausschnitt mitsamt dessen Mittelpunktmarkierung (grünes Quadrat). Beispielhaft soll hier die Limitation der manuellen Messmethode mit ImageJ gezeigt werden. Die Positionierung der drei Winkelmesspunkte ist nur im Bereich der Schnittpunkte von jeweils vier Pixeln möglich (= Eckpunkte eines Pixels). Die beiden oberen Bilder verdeutlichen, dass der Scheitelpunkt nicht direkt mittig in dem hellsten/höchsten Pixel (weißer Pfeil) positioniert werden kann. Stattdessen ist nur eine Platzierung an den vier Eckpunkten dieses Pixels möglich. Die Auswirkung dieser Tatsache zeigt sich in den verschiedenen Biegewinkelergebnissen ( $141^\circ$  bzw.  $161^\circ$ ). Ebenso kann die Positionierung der Tangentenendpunkte bei identischem Mittelpunkt durchaus variieren ( $168^\circ$  bzw.  $157^\circ$ ). Dies zeigen die beiden unteren Bilder. Auch hier besteht die Problematik der Eckpunktplatzierung.



**Abbildung 104: Bestimmung des Höhenmaximums mittels einer Gauß-Anpassung**

Das Bild zeigt ein Höhenhistogramm eines DNA-Querschnittes in ImageJ. Im oberen Anteil sieht man den tatsächlich gemessenen Durchschnitt im Bild (schwarz), darunter die dazugehörige Gauß-Anpassung der Daten (blau). Die roten Kreise geben den Höhenwert des jeweiligen Pixels wieder. Die horizontale grüne Linie zeigt die Halbwertsbreite an (FWHM), die zwei senkrechten roten Geraden demonstrieren die Höhen, die dem höchsten Pixel benachbart sind. Aufgrund der Beschreibung im Text ist es bei der Positionierung der Messschenkel nicht möglich das absolute Maximum zu markieren, sondern es wurde stets ein Grenzbereich zwischen zwei Pixeln als Scheitelpunkt gewählt. Zudem liegt das Maximum der Gauß-Anpassung über dem höchsten tatsächlichen Wert und stimmt nicht exakt mit dessen Lage bezüglich der y-Achse überein. Auf diese Weise ermöglicht die Gauß-Anpassung eine verbesserte Messung ohne limitierende Pixelauflösung.





**Abbildung 105: Beispiele für die automatische Biegewinkelmessung an oxoG:A**

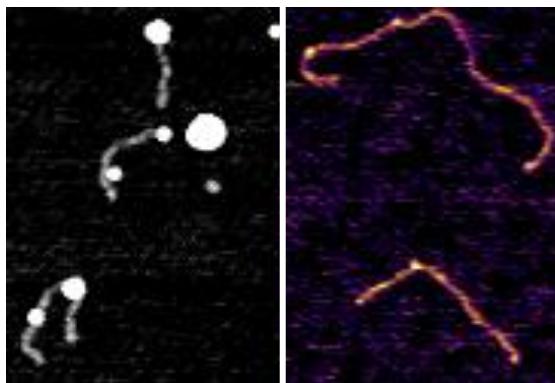
Beispielhaft sind DNA-Stränge mit oxoG:A-Schaden und FIESTA-Skelettierung gezeigt. A zeigt stets eine Vergrößerung der 50 %-Position mit Angabe des automatischen Messergebnisses der 8 nm Tangente. Darunter zeigt M denselben DNA-Strang mit manuell eingefügten Tangenten in ImageJ sowie den manuell ermittelten Biegewinkel.

### *Weiterführende Evaluationsvorschläge der automatisierten Biegewinkelmessung*

Um die korrekte Funktionsweise des MATLAB-Programms bei der Biegewinkelmessung zu kontrollieren, könnte sich eine Parabel-Funktion eignen. Die Parabel ließe sich durch eine einfache mathematische Funktion beschreiben. Anhand dieser Formel ließe sich der passende Graph mit einer Geometrie-Software wie beispielsweise „GeoGebra“ zeichnen. Die Abbildung müsste dann gespeichert und mit derselben Vorgehensweise, wie bei der DNA-Auswertung beschrieben, ausgewertet werden. Der Vorteil einer Parabelfunktion liegt in der relativ einfachen Darstellung und der Möglichkeit, anhand der Funktion zu jedem beliebigen Punkt die entsprechende Biegung errechnen zu können. Auf diese Weise könnte der am Graphen gemessene Winkel mit dem theoretischen Winkel verglichen werden. Einen anderen Ansatz der Programmvalidierung stellt die Auswertung gemäß dem Monte Carlo Prinzip simulierter wurmkettenartige DNA dar [185]. Gemäß Rivetti *et al.* [143] liegt die DNA bei der Probendeposition für AFM-Aufnahmen unter Einsatz von  $Mg^{2+}$  auf der Glimmoberfläche equilibriert vor und kann so durch ein 2D wurmkettenartiges (WLC) Modell beschrieben werden. Weiterhin wäre es interessant die sequenzspezifische Biegung von in dieser Arbeit vermessenen DNA-Substraten zu berechnen und mit den Ergebnissen der automatischen Messsoftware zu vergleichen. Das von Goodsell und Dickerson [188] entwickelte Analyseprogramm „BEND“ könnte beispielsweise für die Berechnung der Biegewinkel mit verschiedenen Simulationsmodellen genutzt werden. Bei der Programmvalidierung sollte zusätzlich zur Winkelbestimmung auch die unvollständige Filamentierung untersucht werden. Teilweise war die nachträgliche manuelle Verbindung getrennter DNA-Filamente auch an der Position von gebundenen Proteinen (vor allem großer Proteine) notwendig. In diesem Fall sollte ausgeschlossen werden, dass der Biegewinkel an einer künstlich errechneten Filamentlinie ermittelt wird. Da fraglich ist, ob diese errechnete Verbindungslinie immer den tatsächlichen DNA-Verlauf beschreibt und außerdem von der Länge der Linie abhängig ist. Es sollte untersucht werden, inwieweit der Bedarf besteht, für die manuelle Verbindung zweier getrennter Filamente einen Grenzwert zu definieren. Ebenso wäre eine Verbesserung der Filamentierung von FIESTA wünschenswert, um die getrennte Skelettierung zu verhindern.

### 4.1.5 Proteinerfassung

Für die Biegewinkelmessung an Protein-DNA-Komplexen ist die exakte Lokalisierung der Proteine entscheidend. Die Protein-Erfassung bei einer großen Proteinspezies gestaltete sich einfacher, da sich diese deutlicher von der DNA-Höhe unterschied. Beispielhaft wurde das, im Vergleich zu MutY, relativ große Protein T-Antigen (~200 kDa bei Monomeren) mit der automatisierten Messmethode vermessen (Abbildung 106). So war auch die Wahl eines höheren Hintergrund-Schwellenwertes (Threshold) bei der Proteinerkennung möglich und der deutlichere Kontrast zwischen den Proteinen und der DNA hatte eine verbesserte Proteinerkennung zur Folge. Gerade bei kleinen Proteinen, wie den Glykosylasen, kam es häufig zu einer überlagerten Darstellung der Höhen von Proteinen und DNA-Strängen, was eine manuelle Anpassung an die jeweiligen Abbildungsparameter unumgänglich machte. Somit musste für jedes Bild/Experiment eine Optimierung bezüglich des Threshold-Wertes für DNA- und Protein-Höhen durchgeführt werden, um eine möglichst optimale Einstellung für die separate DNA- und Protein-Darstellung zu finden. Für eine Optimierung der Höhenwerte hatte es sich als sehr nützlich erwiesen das AFM-Bild und die „Objects map“ in FIESTA zu laden und die entsprechenden Proteine mit den Kennnummern zu analysieren. In der Statistiktabelle des „3D Object Counters“ konnte die Fläche für die Protein-Objekte entnommen werden. Diese Größenangaben konnten bei der Wahl der Grenzwerte für MATLAB oder als Anhaltspunkt für eine zweite verfeinerte Auswertung mit dem „3D Object Counter“ genutzt werden.



**Abbildung 106: Dimensionsvergleich T-Antigen zu MutY**

Beispielhaft wurden neben den Protein-DNA-Komplexen von MutY, auch T-Antigen (~200 kDa bei Monomeren) mit der automatischen Messsoftware ausgewertet. Die T-Antigen Proteine (links) hoben sich im Vergleich zu MutY (rechts) deutlicher von der DNA ab und waren somit einfacher zu lokalisieren.

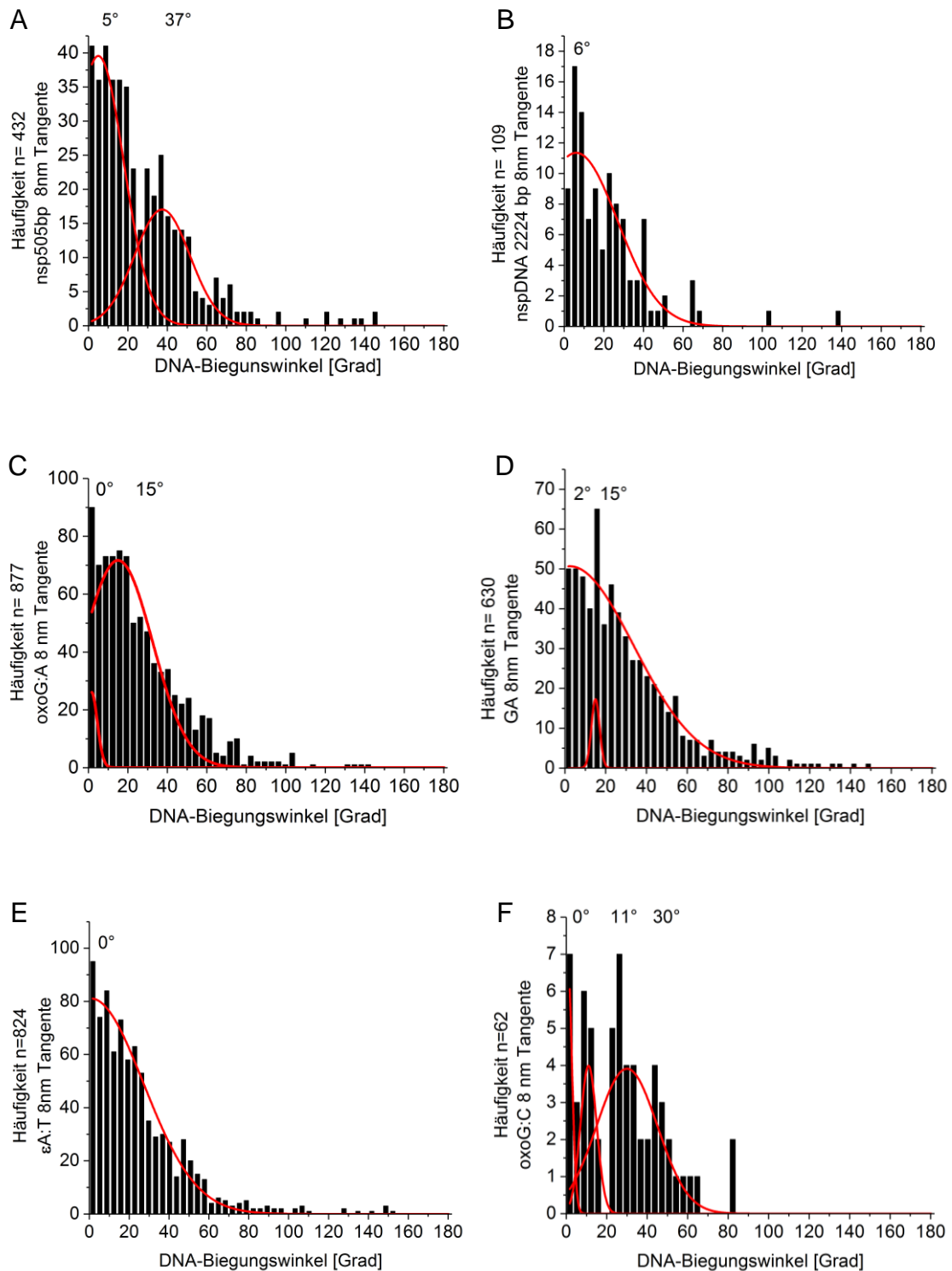
Bei der Programminitialisierung für die Biege winkelmessung an Protein-DNA-Komplexen bereitete die separate Darstellung der DNA ohne die gebundenen Proteine zunächst Schwierigkeiten. Die verschiedenen Ansätze sind im Kapitel Ergebnisse von 3.4.2 bis 3.4.4 beschrieben. Der Versuch aus Kapitel 3.4.3 nur die DNA mit zwei verschiedenen, sehr schmal definierten Threshold-Werten separat ohne Proteine darzustellen, lies an der Protein-Position sehr häufig weiße „Löcher“ mit schwarzer Umrandung entstehen. Daraus ergab sich die Fragestellung, ob das an die DNA gebundene Protein wirklich zu einer deutlichen Erhöhung der DNA an dieser Stelle führt, oder ob es aufgrund des Bindungsmodus des Proteins nicht nur zu einer marginalen Erhöhung, bzw. zu einer seitlichen Verlagerung des Proteins an der DNA kommt. Aus all den oben aufgeführten Versuchen und dem Problem der DNA-Protein-Diskriminierung konnte gefolgert werden, dass sich der größte Teil der Proteine nicht direkt auf dem DNA-Strang befindet, sondern vielmehr daneben, was die Threshold-Bestimmung erschwerte. Auch Schulz *et al.* [147] beschrieben in ihren AFM-Untersuchungen zu Protein-DNA-Komplexen in Flüssigkeit ein Überwiegen der lateral gebundenen Proteine (0,75-fach mehr) gegenüber den auf der DNA sitzenden Proteine. So wäre auch die fehlerhafte Entfernung der Peaks von der DNA mit einem sehr schmalen Höhen-Intervall zu erklären. Es ist anzunehmen, dass die große Häufigkeit von Proteinen, die seitlich an die DNA angelagert sind, in den AFM-Aufnahmen nicht zufällig ist. Da die Glykosylasen an eine bestimmte Stelle der DNA, die kleine Furche, binden, besteht die Möglichkeit, dass die Methode der DNA Fixierung auf dem Glimmer die Position der Glykosylasen beeinflusst oder die Proteine möglicherweise durch den Glimmer durch supramolekulare Wechselwirkungen auch angezogen werden. Zusätzlich besteht die Gefahr, dass so eine weitere Biegung der DNA induziert werden könnte. Außerdem wäre denkbar, dass aufgrund der Schwerkraft ein großer Teil der gebundenen Proteine der Glimmer-Oberfläche anliegt. Die aufgeführten Einflussmöglichkeiten auf die Proteinlage könnten in einer nachhaltigen Verzerrung und so in einer artifiziell erzeugten Biegung der DNA in der rasterkraftmikroskopischen Aufnahme resultieren. Der Einfluss der Probendeposition auf die DNA-Länge und Biegung wurde zuvor im Kapitel 4.1.4.1 diskutiert.

## 4.2 Diskussion der DNA-Biegewinkel

In der Diskussion sollen die manuellen Biegewinkel von Lisa Hirsch mit den jeweiligen Ergebnissen der automatisierten Software verglichen werden und mit den aus der Literatur bekannten Biegewinkeln bewertet werden. Die durch verschiedene Basenschäden induzierten DNA-Biegungen können neben der Rasterkraftmikroskopie noch mit weiteren Methoden ermittelt werden. Während die Rasterkraftmikroskopie eine relativ direkte Möglichkeit der Messung von DNA-Biegungen darstellt, sind aus der Literatur weitere Methoden bekannt: der Förster Resonanz Energie Transfer (FRET) [190], die Röntgenstrukturanalyse [191, 192], die Gelelektrophorese [193-197], die Elektronenmikroskopie [198, 199] und die Kernspinresonanzspektroskopie (NMR, englisch: nuclear magnetic resonance) [200, 201]. Wie bereits in der methodischen Diskussion erläutert, ist die Tangentenmethode mit 8 nm Segmentlänge für die Vermessung von DNA-Biegewinkeln am geeignetsten, weshalb in der Ergebnisdiskussion der Fokus auf den Ergebnissen dieser Methode liegt. In der folgenden Diskussion sollen die Messergebnisse dieser Arbeit mit bekannten Biegewinkeln aus der Literatur verglichen werden. Zum besseren Vergleich werden nachfolgend die verschiedenen Biegewinkelergebnisse gegenübergestellt. Die mit der 8 nm Tangenten-Methode vermessenen DNA-Substrate (ungeschädigte DNA, oxoG:A, G:A, εA:T und oxoxG:C) finden sich in Tabelle 6 und Abbildung 107.

**Tabelle 6: Übersicht der Biegewinkel**

	Substrat					
	nsp DNA		oxoG:A	G:A	oxoG:C	εA:T
Länge	505 bp	2224 bp	505 bp	505 bp	505 bp	505 bp
Biegewinkel	5° ±15°	6° ±21°	0° ± 3°	2° ±32°	0° ± 3°	0° ±26°
	37° ±14°		15° ±18°	15° ± 4°	11° ± 4° 30° ±15°	



**Abbildung 107: Gegenüberstellung der Biegewinkelergebnisse**

Die aufgeführten Biegewinkelergebnisse der verschiedenen DNA-Substrate wurden bereits im Ergebniskapitel dargestellt. In dieser Abbildung sollen ausschließlich die Ergebnisse mit der Tangentenmethode und 8 nm Länge zu Vergleichszwecken gegenübergestellt werden. Um im Histogramm auch kleine ( $<10^\circ$ ) Winkel auflösen zu können, war ein mäßiges „Überbinen“ nötig.

(A: nsp 505 bp, B: nsp 2224 bp, C: oxoG:A, D: G:A, E:  $\epsilon$ A:T, F: oxoG:C).

### 4.2.1 Ungeschädigte DNA

Unter Verwendung der Tangentenmethode ergaben sich für die 505 bp lange ungeschädigte DNA ( $n = 432$ ) mit einer durchschnittlichen DNA-Länge von  $149 \text{ nm} \pm 21 \text{ nm}$  mit 8 nm Tangentenlänge zwei Maxima:  $5^\circ$  und  $37^\circ$ . Bei der Messung von ungeschädigter 2224 bp langer DNA ( $n = 109$ ) mit einer durchschnittlichen DNA-Länge von  $708 \text{ nm} \pm 14 \text{ nm}$  wurde mit der Tangentenmethode (8 nm) eine Biegung von  $6^\circ$  gemessen. Insgesamt kann bei dieser Verteilung mit relativ geringer Datenmenge aber auch eine generelle Tendenz gegen  $0^\circ$  interpretiert werden, die durch eine größere Anzahl an Datenpunkten ( $n \geq 1000$  Datenpunkte) unterstützt werden könnte.

Die Ergebnisse der automatisierten Tangentenmessung weichen von denen in Büchner *et al.* [1] und dem von Lisa Hirsch manuell ermittelten Biegungswinkel von  $0^\circ$  für ungeschädigte DNA ab: Lisa Hirsch ermittelte sowohl bei 505 bp langer DNA  $0^\circ \pm 5^\circ$  als auch bei 2224 bp langer DNA  $0^\circ \pm 9^\circ$ , Büchner *et al.* [1] ermittelten für ungeschädigte DNA einen Biegungswinkel von  $0^\circ \pm 18^\circ$ . Allerdings wurde in beiden Arbeiten ein breiteres Bin-Maß ( $7^\circ$  statt  $3,5^\circ$ ) verwendet, da weniger Daten zur Verfügung standen. Dies erlaubte keine Auflösung von Biegungen  $< 7^\circ$ . Die eigene manuelle Messung mit ImageJ ergab einen Biegungswinkel von  $0^\circ \pm 6^\circ$  (siehe Appendix Abbildung 109). Auch in der Primärliteratur wird für ungeschädigte DNA mit stabilen Watson-Crick-Basenpaaren keine Biegung ( $0^\circ$ ) aus AFM-Aufnahmen berichtet [93, 118-122, 202]. Beispielsweise gaben Chen *et al.* [122] für die Biegungswinkelmessung an 6 nm langen Fragmenten ungeschädigter DNA eine Gaußsche Biegungswinkelverteilung mit einem Maximum bei  $0^\circ$  an. Die Standardabweichung von  $19^\circ$  wurde hier auf die intrinsische Flexibilität der Moleküle zurückgeführt. Allerdings ist aus der Literatur mithilfe anderer Messmethoden und Simulationen bekannt, dass auch ungeschädigte DNA nicht vollkommen gerade ist, sondern eine sequenzabhängige globale Biegung von etwa  $5^\circ - 15^\circ$  vorliegen kann [203]. Beispielsweise beschrieben Garcia *et al.* [197] bei Untersuchungen mit dem Rasterkraftmikroskop und der Affinitätselektrophorese (EMSA) für ungeschädigte DNA einen Biegungswinkel von  $4^\circ \pm 23^\circ$ , weitere Werte von  $6^\circ$  (elektrophoretische Mobilitätsmodulationen) bis zu  $15^\circ$  sind ebenfalls bekannt [204, 205]. Für rasterkraftmikroskopische Aufnahmen diskutierten Faas *et al.*

[185], dass die Bindung der DNA an eine geladene Fläche eine statische Biegung von ca.  $5^\circ$  pro Dinukleotidschritt hervorrufen kann, wobei es zu keiner Beeinträchtigung der Dynamik im DNA-Rückgrat kommt. Auch Hardin *et al.* [144] beschrieben einen geringen Einfluss des Depositionspuffers als mögliche Ursache für das zusätzliche Vorkommen von geringen Biegewinkeln an ungeschädigter DNA in rasterkraftmikroskopischen Aufnahmen. Weiterhin gaben Rouzina und Bloomfield [167] in ihren Untersuchungen für Wechselwirkungen zwischen  $Mg^{2+}$ -Kationen und B-förmiger DNA einen Biegewinkel von mindestens  $5,5^\circ$  an. Auch dies stellt eine mögliche Erklärung für den Biegewinkel von  $5^\circ$  von ungeschädigter DNA dar. Zudem beschrieben Rouzina und Bloomfield [167], dass multivalente Ionen teilweise starke Biegungen ( $20^\circ - 40^\circ$ ) in B-förmige DNA induzieren können. Dies lässt sich als Hinweis für den Biegewinkel von  $37^\circ$  aufführen. Allerdings wurde dieser Wert für andere Messmethoden beschrieben und eine Übertragung auf die AFM-Technik sollte kritisch betrachtet werden. Die Verteilung des zugehörigen Histogramms lässt aber auch die Vermutung zu, dass sich die Verteilungslücke zwischen  $5^\circ$  und  $37^\circ$  durch eine größere Anzahl von Daten schließen könnte. Allerdings verdeutlicht die Übereinstimmung der Literaturwerte für die Biegung ungeschädigter DNA mit dem in der vorliegenden Arbeit gemessenen Biegewinkel an ungeschädigter DNA von  $5^\circ$  die Qualität des Messverfahrens.

### 4.2.2 MutY-Schadenssubstrat: oxoG:A

Für oxoG:A, einem Schadenssubstrat von MutY, zeigten sich bei einer Tangentenlänge von 8 nm und einer durchschnittlichen DNA-Länge von  $154 \text{ nm} \pm 24 \text{ nm}$  zwei Biegeungswinkel der DNA am Schadensort von  $0^\circ \pm 3^\circ$  und  $15^\circ \pm 18^\circ$  mit einer angedeuteten Spezies bei  $30^\circ$ . Die manuellen Winkelmessungen von Lisa Hirsch ergaben für oxoG:A einen durchschnittlichen Wert von  $24^\circ \pm 13^\circ$ . Die eigenen manuellen Messungen mit ImageJ ergaben einen Biegeungswinkel von  $15^\circ \pm 9^\circ$  (siehe Appendix Abbildung 110). Dieser Wert deckt sich mit den Ergebnissen der automatisierten Messsoftware.

Wie bereits in der Einleitung beschrieben, bindet oxoG bevorzugt mit Cytosin über eine Watson-Crick-Basenpaarung, allerdings ist auch die Bindung mit Adenin über eine Hoogsteen-Bindung möglich [66]. Das Adenin wird während



Replikationsprozessen eingefügt [206]. Die Abbildung 7 visualisiert die beiden Bindungskonformationen. Cheng *et al.* [207] haben in Computersimulationen von oxoG:A in der anti-anti-Konformation ein spontanes Ausschwenken der Base aus der Doppelhelix, und nach einer extrahelikalen 180° Drehung, ein Zurückschwenken in die syn-anti-Konformation beobachtet. Ebenso haben Cheng *et al.* [207] für oxoG:A die Ausbildung einer Hoogsteen-Bindung in syn-anti-Konformation im Gegensatz zur anti-anti-Form der Basenpaare als die thermodynamisch bevorzugte Anordnung beschrieben. Für oxoG:A in Lösung beschrieben sie eine Biegung um 15° zum benachbarten Basenpaar. Da auch bei der automatischen Auswertung die Biegungsspezies bei 15° dominierte, könnte dies der syn-anti-Konformation (Hoogsteen-Bindung) zugeordnet werden. Die Biegungsspezies bei 0° entspricht möglicherweise der thermodynamisch weniger günstigen anti-anti-Konformation.

### 4.2.3 MutY-Schadenssubstrat: G:A

Die Auswertung des weiteren MutY-Schadenssubstrates G:A ergab mit FIESTA-MATLAB bei einer Tangentenlänge von 8 nm und einer durchschnittlichen DNA-Länge von 153 nm  $\pm$ 24 nm am Schaden einen Biegungswinkel von 2°  $\pm$ 32° mit einer separaten Biegungsspezies bei 15°  $\pm$ 4°. Auch ein Winkel von ca. 27° deutete sich in der Biegewinkelverteilung als kleinere Biegungsspezies an. Die manuellen Winkelmessungen von Lisa Hirsch ergaben für G:A einen durchschnittlichen Biegungswinkel von 22°  $\pm$ 16°. Die eigenen manuellen Messungen mit ImageJ ergaben 10°  $\pm$ 15° (Abbildung 111). Die Biegungswinkel der beiden MutY-Substrate G:A und oxoG:A stimmten sowohl bei den manuellen Ergebnissen von Lisa Hirsch (22° und 24°), als auch bei den Ergebnissen der automatisierten Auswertung (0° und 2°; 15° und 15°) im Rahmen der Messgenauigkeit überein. Dieses Ergebnis bestätigt die Annahme, dass verschiedene Zielstrukturen einer Glykosylase ähnliche Biegungswinkel aufweisen.

In der Literatur wurde prinzipiell eine Biegung für den G:A-Basenschaden beschrieben. Für die G:A-Fehlpaarung sind aus NMR [208, 209] und Röntgenstrukturanalysen [210] Veränderungen der Doppelhelix-Konformation durch eine lokale Beeinflussung der Breite der kleinen Furche bekannt. Cheng *et al.* [207]

beschrieben in ihren Untersuchungen wechselnde Konformationszustände mit einem bis zu  $60^\circ$  ansteigenden lokalen Biegewinkel an der G:A-Fehlpaarung. Weiterhin wurden von Cheng *et al.* [207] anhand eines Simulationsmodells ein lokaler Biegewinkel von  $27^\circ \pm 17^\circ$  bei dem Konformationswechsel von G(anti):A(anti) zu G(anti):A(syn) beobachtet. Bei Betrachtung der automatisierten Biegewinkelverteilung von G:A kann ebenfalls eine relativ breite Biegewinkelverteilung mit einem deutlichen Abfall der Häufigkeit bis zu  $60^\circ$  beobachtet werden. Zudem legt die relativ breite Biegewinkelverteilung das Vorhandensein verschiedener Konformationen nahe, wie in der Literatur beschrieben (Strukturformeln siehe Abbildung 8), [91, 211]. So könnten die verschiedenen Maxima in der Biegewinkelverteilung (ein ungebogener Zustand und weitere Zustände mit Biegungen von  $15^\circ$ ,  $27^\circ$  bis hin zu  $60^\circ$ ) den verschiedenen Konformationen entsprechen. Anknüpfend an die Diskussion von oxoG:A könnten die Konformationszustände von G:A den Biegewinkelmaxima wie folgt zugeordnet werden: G(anti):A(anti) bei  $0^\circ$ , G(syn):A(anti) bei  $15^\circ$  und G(anti):A(syn) bei  $27^\circ$ . Entsprechend den Beobachtungen von Cheng *et al.* [207] ist dem Biegewinkelmaximum bei ca.  $27^\circ$  die G(anti):A(syn)-Konformation zuzuordnen, ebenso wie  $0^\circ$  der G(anti):A(anti)-Konformation. Bei einem neutralem pH-Wert liegen G(anti):A(anti) und G(anti):A(syn) im gleichen Verhältnis vor. Bei einem niedrigen pH-Wert bis 5,5 dominiert jedoch die G(syn):A(anti)-Konformation [207, 211, 212], da G(syn):A(anti) durch zusätzliche Protonierung des Adenins auch zur Ausbildung von zwei Wasserstoffbrückenbindungen befähigt ist [187]. Da diese stabile Bindungskonformation nur pH-abhängig ausgebildet werden kann und der in dieser Arbeit verwendete Depositionspuffer einen pH-Wert von 7,5 aufwies, ist die relativ kleine, aber sehr scharfe Spezies bei  $15^\circ$  zu erklären. Beachtet werden muss hier jedoch, dass die Konformation der G:A-Fehlpaarung durch lokale Ladungszustände beeinflusst werden kann. Neben dem pH-Wert nehmen so auch die Wechselwirkungen mit Nachbarbasenpaaren und dem Phosphatkettenrückgrat einen relevanten Einfluss auf die ausgebildete Konformation der Fehlpaarung [211]. Vor dem Hintergrund der unterschiedlich legierten Schadensstellen (bis 5 bp abweichend) kann auch der Einfluss von Dipol-Wechselwirkungen des jeweiligen Schadenssubstrates mit benachbarten Basen auf die DNA-Biegung nicht ausgeschlossen werden

[187]. Allgemein ist festzustellen, dass die oxoG:A-Fehlpaarung stabiler ist als die G:A-Fehlpaarung, da oxoG:A eine stabile Hoogsteen-Bindung ausbilden kann, und wie im vorherigen Kapitel beschrieben thermodynamisch bedingt bevorzugt in dieser Konformation (bei 15°) vorliegt [93, 207].

### 4.2.4 hOGG1 Schadenssubstrat: oxoG:C

Für eine Tangentenlänge von 8 nm besaß die DNA am Schadensort oxoG:C bei einer durchschnittlichen DNA-Länge von 151 nm  $\pm$ 17 nm einen Biegewinkel von 0° und zwei weitere Winkel bei 11° und 30°. Auf Grund der geringen Datenanzahl ( $n = 62$ ) muss die Gauß-Anpassung eher kritisch betrachtet werden. Die eigenen manuellen Messungen mit ImageJ ergaben einen Biegewinkel von 10°  $\pm$ 7° (Abbildung 112). Lisa Hirsch ermittelte manuell aus verschiedenen Experimenten einen durchschnittlichen Wert von 0°  $\pm$ 5°, mit einer zusätzlichen deutlich kleineren Spezies bei 18°  $\pm$ 13°. Büchner *et al.* [1] haben für oxoG:C ebenfalls zwei Biegewinkelspezies beschrieben (-2°  $\pm$ 15° und 36°  $\pm$ 15°). Die manuellen Messergebnisse und die Ergebnisse der Software gleichen sich im Rahmen der Messgenauigkeit.

Für den Basenschaden oxoG:C konnte in anderen wissenschaftlichen Untersuchungen keine signifikante Biegung der DNA-Helix nachgewiesen werden. So beschrieben Lipscomb *et al.* [213] in ihren Röntgenstrukturanalysen für die Basenpaarung von oxoG:C eine Watson-Crick-Basenpaarung und kaum vorkommende Biegungen, weshalb diese Basenpaarung der normalen B-förmigen DNA entspricht (Abbildung 4). Auch durch NMR-Untersuchungen sind diese Struktureigenschaften für oxoG:C bekannt [214]. Computersimulationen für G:C und oxoG:C zeigten ebenfalls keine deutlichen Konformationsveränderungen [215]. Nach Cheng *et al.* [207] wäre auch eine Basenöffnung unwahrscheinlich. OxoG bevorzugt die Bindung zu Cytosin gemäß der klassische Watson-Crick Basenpaarung und ähnelt so sehr stark dem ungeschädigten Watson-Crick-Paar G:C [213, 214, 216]. Nach Reynisson und Steenken [217] ist die Bindungsenergie von oxoG:C vergleichbar mit G:C, wobei die exotherme Bindungsenergie für G:C -25 kcal/mol beträgt. Beide Basenpaarungen bilden in der Watson-Crick-Basenpaarung drei Wasserstoffbrückenbindungen aus und liegen in der anti-anti-

Konformation vor. Im Hoogsteen-Paar liegt oxoG in der syn-Konformation vor und bildet zwei Wasserstoffbrücken zu Cytosin in der anti-Konformation aus, wobei hier die Bindungsenergie ca. -10 kcal/mol beträgt [216]. Außerdem beschrieben Reynisson und Steenken [216] in ihren mathematischen Simulationen für oxoG in Bezug auf die Nukleotide einen Biegewinkel von ungefähr 30°.

Chen *et al.* [122] ermittelten in ihren AFM-Aufnahmen an oxoG:C-schadenstragender DNA einen Biegewinkel von 0° ±19°. Ähnlich zu den Ergebnissen von Büchner *et al.* [1] zeigte sich auch bei den Untersuchungen von Chen *et al.* [122] an den Komplexen aus hOGG1 und ungeschädigter DNA ein deutlicher Biegungswinkel von 71°, zusätzlich zwei kleine Biegungsspezies bei ~0° und ~30°. Die Übereinstimmung zwischen den Biegewinkeln an oxoG und im hOGG1-Komplex mit ungeschädigter DNA wurde als Indiz für ein aktives Biegen von hOGG1 an ungeschädigter DNA und passives Biegen am Zielschaden oxoG im Suchkomplex (SC) gewertet [122]. In Röntgenstrukturanalysen zeigt der hOGG1-oxoG-DNA-Komplex einen deutlichen Biegewinkel von ca. 70° am Schadensort, entsprechend der ca. 70° Biegung aus AFM-Messungen [66, 218].

Bei Betrachtung der Ergebnisse mit 8 nm Tangenten fallen zwei deutliche Biegungsspezies (11° und 30°) auf. Dies deckt sich mit der aus der Literatur beschriebenen Dominanz der Watson-Crick-Basenpaarung gegenüber der Hoogsteen-Bindung. Die Watson-Crick-Basenpaarung von oxoG:C ist vergleichbar mit der ungeschädigter DNA, für die keine deutliche Biegung zu erwarten ist; für die Hoogsteen-Basenpaarung dagegen schon. Die kleine separate Spezies bei 11° könnte man auch zu einer Gesamtverteilung gegen 0° zählen und die Verteilungslücke durch eine zu geringe Anzahl an Datenpunkten (n = 62) rechtfertigen. Wenn man wie bereits bei der ungeschädigten DNA erwähnt, für die Watson-Crick-Basenpaarung von einer Biegung von bis zu 15° ausgeht, könnten die beiden Spezies (0° und 11°) diesem Zustand zugeordnet werden. Vergleicht man das Ergebnis von oxoG:C (11°) mit ungeschädigter DNA (5°) könnte der etwas vergrößerte Biegewinkel bei oxoG:C auch auf eine durch den Basenschaden verursachte geringere Stabilität hinweisen. Der Biegewinkel von 27° bzw. 30° könnte folglich das Vorhandensein des Hoogsteen-Basenpaares repräsentieren. Aufgrund der ungünstigen energetischen Verhältnisse im Hoogsteen-Basenpaar

sollte die Watson-Crick-Basenpaarung überwiegen, (große Fläche der Gaußkurve an  $0^\circ - 11^\circ$  versus der bei  $\sim 30^\circ$ , siehe Abbildung 107). Büchner *et al.* [1] beschrieben zwei deutliche Biegespezies ( $0^\circ$  und  $\sim 60^\circ$ ), wobei der kleinere Biegewinkel in der Häufigkeit deutlich überwog, was die oben erwähnte Konformationsverteilung unterstützt. Die Spezies mit  $30^\circ$ -Biegung würde mit dem Biegewinkel von Reynisson und Steenken [217] sowie Chen *et al.* [122] von ca.  $30^\circ$  für oxoG:C übereinstimmen. Die breite Verteilung der  $30^\circ$ -Biegung im Histogramm spiegelt auch die erhöhte Flexibilität am potenziellen Hoogsteen-Basenpaar wider.

Zusammenfassend können die kleinen Biegewinkel ( $0^\circ$  bis  $11^\circ$ ) der Watson-Crick-Basenpaarung oxoG(anti):A(anti) und der Biegewinkel von  $\sim 30^\circ$  dem Hoogsteen-Basenpaar oxoG(syn):A(anti) zugerechnet werden. Insgesamt ist zu bemerken, dass für dieses Schadenssubstrat nur 62 Datenpunkte für die automatische Auswertung zur Verfügung standen. Dies schränkt die Aussagekraft der so generierten Biegewinkel ein, lässt aber erste Vermutungen zu.

### 4.2.5 hAAG Schadenssubstrat: $\epsilon$ A:T

Die Auswertung von  $\epsilon$ A:T ergab bei einer durchschnittlichen DNA-Länge von 148 nm  $\pm 24$  nm für die 8 nm langen Tangenten einen durchschnittlichen Bieigungswinkel von  $0^\circ \pm 26^\circ$ . Von Lisa Hirsch wurde aus manuellen Messungen ein Bieigungswinkel von  $15^\circ \pm 10^\circ$  angegeben. Eigene manuelle Messergebnisse ergaben einen Bieigungswinkel von  $13^\circ \pm 8^\circ$  (Abbildung 113).

Die von hAAG spezifischen Schäden verursachen kaum Helixverzerrungen [219]. Singer *et al.* [220, 221] beschrieben für  $\epsilon$ A:T die Entstehung von „sperrigen Läsionen“ (engl. bulky lesions). Aus der Literatur sind keine Bindungen zu einer komplementären Base über Wasserstoffbrückenbindungen bekannt, wobei  $\epsilon$ A:T in anti-anti-Konformation vorliegt [222]. Hieraus lässt sich folgern, dass es zu keiner zusätzlichen Verzerrung der DNA, die durch den Basenschaden begünstigt wird, kommt. Da für diesen DNA-Schaden keine Biegung angenommen wird, ist eine Verteilung um  $0^\circ$  zu erwarten. Die Bieigungswinkelverteilung gegen  $0^\circ$  von  $\epsilon$ A:T könnte durch die im Vergleich zu der ungeschädigten DNA ( $5^\circ$ ) erhöhte Flexibilität am Schadensort durch das Fehlen

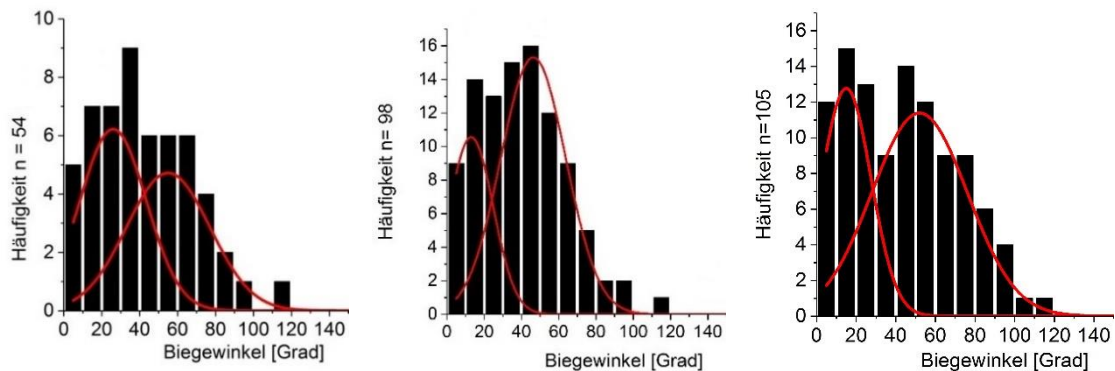
der Wasserstoffbrückenbindungen verursacht werden. Die hohe Flexibilität am Schadensort durch die fehlende Basenpaarung sollte zudem in einer verhältnismäßig breiten Biegewinkelverteilung resultieren. Quantitativ überwiegen auch Biegewinkel bis  $\sim 25^\circ$ , was die zuvor erwähnte Flexibilität von  $\epsilon A:T$  widerspiegeln könnte. Auch bei den eigenen manuellen Messergebnissen ließ sich eine Verteilung der Biegewinkel bis  $\sim 25^\circ$  erkennen (Abbildung 113).

Die B-förmige DNA wird über die zentralen 8 Basenpaare im Exzisions- oder Interrogations-Komplex von hAAG um  $22^\circ$  gebogen [105, 223]. Die in der Biegewinkelverteilung angedeutete Spezies bei  $\sim 25^\circ$  würde somit mit dem IC übereinstimmen. Das Überwiegen der Biegewinkel von bis zu  $25^\circ$  in Abwesenheit von hAAG lassen die Vermutung zu, dass sich die Flexibilität am Schaden durch fehlende Wasserstoffbrückenbindungen bis zu der Größe des Biegewinkels für den hAAG-DNA-Komplex erstreckt.

### **4.2.6 Diskussion der DNA-Biegewinkel an Proteinen**

Die Biegewinkelmessung an Protein-DNA-Komplexen von MutY an ungeschädigter DNA erfolgte selektiv mit der Tangentenmethode durch jeweils 8 nm lange Messgeraden an Proteinpeaks, die das proteinspezifische Volumen aufwiesen. Der mit der Tangentenmethode gemessene Winkel von  $\sim 50^\circ$  für den Schadenserkennungskomplex ist vergleichbar mit dem durch Röntgenkristallographie ermittelten Biegungswinkel für am Schaden gebundenes MutY [93]. Im vorausgegangenen methodischen Kapitel 4.1.4.3 der Diskussion wurde schon allgemein auf die verschiedenen Segmentlängen eingegangen. Die Segmentlänge von 8 nm war für die Biegewinkelmessungen an Proteinpositionen aufgrund des Proteindurchmessers notwendig. Bei Darstellung der Biegewinkel aller Proteine, ungeachtet ihres Volumens, zeigte sich ein deutliches Maximum bei  $46^\circ$ . Dieser Wert ist ebenfalls nahe an dem publizierten Biegewinkel von MutY. Auch bei der manuellen Messung von MutY, gebunden an 2224 bp lange ungeschädigte DNA und nach spezifischen Proteinvolumen gefiltert, war eine dominante Biegungsspezies bei  $52^\circ$  feststellbar. Zusätzlich enthielten sowohl die automatischen als auch die manuellen Verteilungen eine Biegungsspezies bei ca.  $15^\circ$ . Dieser Wert stimmt mit den für die MutY-Substrate oxoG:A und G:A gemessenen

Biegewinkeln überein (jeweils  $\sim 15^\circ$ ). Weiterhin ist in der MutY-nsp-DNA-Biege-winkel-Verteilung der für ungeschädigte DNA typische Biegewinkel ( $5^\circ$ ) nicht er-kennbar, was ein klares Zeichen für die aktive Biegung von MutY ist. Hieraus könnte sich eine Stabilisierung der SC- und IC-Zustände ableiten lassen ( $\sim 15^\circ$  und  $\sim 50^\circ$ ). Diese kann sowohl bei der automatischen, nach Proteinvolumen ge-filterten Biegewinkelverteilung, bei der ungefilterten, als auch bei der eigenen ma-nuellen Messung deutlich erkannt werden. Zu Vergleichszwecken sind in der nachfolgenden Abbildung 108 nochmals die zuvor diskutierten Ergebnisse dar-gestellt.



**Abbildung 108: Vergleich der Biegewinkel von MutY-nspDNA-Komplexen**

- A:** automatische Biegewinkelmessung an MutY-nspDNA-Komplexen mit der 8 nm Tangentenmethode nach Volumenfilterung (Intervall: 20 – 40  $\text{cm}^3$ ) ( $25^\circ$ ,  $55^\circ$ ).  
**B:** automatische Biegewinkelmessung an MutY-nspDNA-Komplexen mit der 8 nm Tangentenmethode ohne Volumenfilterung ( $13^\circ$ ,  $46^\circ$ ).  
**C:** manuelle Biegewinkelmessung an MutY-nspDNA-Komplexen ohne Volumenfilterung ( $15^\circ$ ,  $52^\circ$ ).

Die Biegewinkelverteilung der MutY-nspDNA-Komplexe stellt sich relativ breit dar, was durch eine geringe Datenmenge und die größer gewählte „Bin“-Breite bedingt ist. Rivetti *et al.* [155] diskutierten, dass es bei der Biegewinkelmessung an Protein-DNA-Komplexen zu einer Verbreiterung der Biegewinkelverteilung kommt, bedingt durch die natürliche Flexibilität der DNA [148] und dem zuvor erwähnten Tip-Konvolutionseffekt [140]. Auch Schulz *et al.* [147] beschrieben für die Biegewinkelmessung von Proteinen an ungeschädigter DNA eine breitere Verteilung als für Protein-DNA-Komplexe an Protein-spezifischer DNA. Eine schmalere Biegewinkelverteilung am spezifischen Schaden lässt sich durch die stabilere Konformation dieses Protein-DNA-Komplexes erklären. Zwischen den nach

Volumen-gefilterten Biegewinkelergebnissen und den nicht gefilterten Ergebnissen besteht eine kleine Abweichung, die aber im Rahmen der Standardabweichung kein Gewicht hat. Eine Erklärung für den geringen Unterschied zwischen nach Volumen-gefilterten Biegewinkelergebnissen und den nicht gefilterten Ergebnissen sollte weniger in einer fehlerhaften Volumenmessung gesucht werden. Im Falle der Biegewinkelmessung an MutY-nspDNA-Komplexen ist die wahrscheinlichere Ursache für die Unterschiede der Biegewinkelergebnisse durch die geringe Datenmenge von 54 Datenpunkten nach der Volumenfilterung im Vergleich zu 98 Datenpunkten vor der Filterung zu sehen. Die Volumenfilterung ist für die korrekte Biegewinkelmessung an Protein-DNA-Komplexen unerlässlich, da hierdurch der ungeschädigte Hintergrund (Kontaminationen des Probenmaterials) reduziert wird. Außerdem ist die Volumenfilterung ein wichtiges Werkzeug, um die Messung ausschließlich an der gewünschten Protein-Konformation durchzuführen. Da viele Proteine zu der Ausbildung von Oligomeren neigen, ist auch dies ein Grund für eine Volumenfilterung.

### **4.3 Vergleich mit anderen publizierten Programmen**

In der Literatur finden sich bereits Publikationen, die sich mit der Vermessung von Biegungen an DNA-Strängen in rasterkraftmikroskopischen Aufnahmen beschäftigen. Im Folgenden soll die Methode zur automatisierten Biegewinkelmessung mit anderen Ansätzen verglichen werden.

Ein freiverfügbares Programm zur Vermessung von DNA-Strängen ist das 2015 von Sanchez und Wyman [224] vorgestellte SFMetrics. Die SFMetrics-Software wurde zur schnellen Vermessung von Biomolekülen, wie DNA oder Proteinen, entwickelt. Es handelt sich um ein halbautomatisches Analyseprogramm für die Längen-, Volumen- und Biegewinkelbestimmung. Auch diese Software ist auf MATLAB basierend, kann aber als alleinstehendes Programm, abgekoppelt von MATLAB, verwendet werden. Hierfür muss eine gesonderte MATLAB-Zusatzdatei (MCR-Datei: MATLAB Compiler Runtime; eigenständiger Datensatz mit MATLAB Anwendungen und Bibliotheken) heruntergeladen werden. Neben den in [.tiff]-Dateien umgewandelten Aufnahmen, können die rasterkraftmikroskopischen Aufnahmen auch direkt im Aufnahmeformat des Rasterkraftmikroskops



(z.B. [.ibw]-Format) ausgewertet werden. Ein Vorteil des Programmes ist neben der vielfältigen Funktionsweise und der Verarbeitung von verschiedenen Aufnahme-Formaten, auch die Unterstützung bei der Auswertung großer Datenmengen durch die halbautomatische Methode. So sind nicht mehrere verschiedene Programme für die Vermessung notwendig, wie in unserem Fall ImageJ, FIESTA und MATLAB. Geplant war zum Vergleich einige Aufnahmen dieser Arbeit exemplarisch mit der SFMetrics-Software zu vermessen. Allerdings war die Auswertung von AFM-Aufnahmen mit diesem Programm nicht zufriedenstellend. Trotz der Informationen aus der Publikation und der zusätzlich beigefügten Anleitung war die Auswertung mit den unterschiedlichsten Bildern, Bildformaten und Parametern nicht umsetzbar. Das SFMetrics-Programm konnte weder eigenständig noch von Mikroskopie-Experten des Rudolf-Virchow-Zentrums erfolgreich auf die rasterkraftmikroskopischen Aufnahmen angewandt werden. Einige der MATLAB-Werkzeuge und -Bibliotheken des Programmes scheinen seit der Publikation nicht mehr ausreichend aktualisiert worden zu sein. Es konnten keine Vergleichsmessungen angestellt werden, jedoch stellt die Publikation einen interessanten Ansatz für die automatisierte Biege winkelmessung von DNA dar. Außerdem hat sie dazu beigetragen den Fokus dieser Arbeit auf die Etablierung einer automatischen Messsoftware auszurichten.

Der Fokus der Arbeit von Hardin *et al.* [144] lag auf dem Vergleich verschiedener Methoden für die Objektivierung der Biege winkelmessung an DNA-Strängen aus rasterkraftmikroskopischen Aufnahmen. Hardin *et al.* [144] haben vier verschiedene Messmethoden zur Biege winkelbestimmung von DNA-Proteinkomplexen (DNA gebunden an drei verschiedene Topoisomerase-Typen) untersucht. Die manuelle Biege winkel-Messung wurde ebenfalls mit dem Winkelmesswerkzeug von ImageJ gemäß der „manual tangent overlay“-Methode durchgeführt. Die automatische Messmethode wurde gemäß dem manuellen Tangentenverfahren in ein Messprogramm (LabVIEW) implementiert. Über verschiedene Bildbearbeitungsschritte konnte in diesen Untersuchungen mit dem Programm LabVIEW Biege winkelmessungen an den DNA-Proteinkomplexen durchgeführt werden. Alternativ zur Biege winkelmessung mit Messgeraden wurden auch FRET- und modifizierte EED-Messungen (end-to-end-distance-measurement), die von Dame *et*

*al.* [176] entwickelt wurden, verwendet. Beispielsweise wurden an den DNA-Proteinkomplexen aus drei verschiedenen gebundenen Topoisomerasen mit der manuellen Tangentenmethode Biegewinkel von  $66^\circ$ ,  $80^\circ$  und  $87^\circ$ , mit der automatisierten Tangentenmethode Werte von  $100^\circ$ ,  $102^\circ$  und  $103^\circ$  beschrieben. Die Ergebnisse der automatisierten Methode zeigten somit eine viel geringere Variabilität. Die Werte wurden ebenso mit den Ergebnissen von FRET und EED verglichen. Für die FRET-Messungen wurden Biegewinkel von  $126^\circ$  bis  $140^\circ$ , für die EED-Messungen Werte von  $84^\circ$  bis  $123^\circ$  aufgeführt. Außerdem wurde mit denselben Messmethoden ungeschädigte DNA vermessen. Es kam zu einer guten Übereinstimmung zwischen den gemessenen Biegewinkeln aus der automatischen Biegewinkelmessung und den FRET- und EED-Ergebnissen. Allerdings muss beachtet werden, dass die Ergebnisse verschiedenartiger Messmethoden nicht direkt miteinander verglichen werden können (FRET- und AFM-Puffer waren unterschiedlich zusammengesetzt). Interessant an der Arbeit von Hardin *et al.* [144] ist, dass auch Protein-DNA-Komplexe mit der „manual tangent overlay“-Methode und einer automatisierten Tangenten-Methode vermessen wurden. Es zeigten sich innerhalb der Tangenten-Methoden weitestgehend übereinstimmende Biegungen. Weiterhin wurden FRET und EED-Messungen durchgeführt. Mittlerweile wurden FRET-Messungen in der AG Tessmer durchgeführt, die die Ergebnisse dieser Arbeit validieren konnten.

Faas *et al.* [185] untersuchten in ihrer Publikation die Auswirkungen auf die strukturellen Eigenschaften von an Oberflächen gebundener DNA. Dabei wurde der Einfluss des Depositionsverfahrens auf die Eigenschaften der DNA-Moleküle insbesondere der Persistenzlänge, der globalen Biegung sowie damit einhergehend der Flexibilität untersucht. Zur Vermessung von DNA entwickelten sie ein MATLAB-basierendes Programm, was auf EM- und AFM-Aufnahmen angewandt wurde, um die globale DNA-Biegung eines DNA-Strangs anhand des Abstandes zwischen den beiden DNA-Enden zu berechnen. Es ergab sich eine (möglicherweise depositionsbedingte) statische Biegung der DNA von  $3,7^\circ - 7,2^\circ$ , was einer durchschnittlichen statischen Biegung von  $5^\circ$  entsprach. Während in den Messungen der Publikation der Einfluss der Deposition bzw. der Glimmeroberfläche auf die globale DNA-Krümmung durch die veränderte Elastizität

(Persistenzlänge) untersucht wurde, ist der Fokus dieser Arbeit der Einfluss verschiedener Schäden auf die lokale DNA-Biegung im Vergleich zu ungeschädigter DNA. Das Ergebnis von Faas *et al.* stimmt allerdings mit den Beobachtungen aus den Biegewinkelmessungen dieser Arbeit gut überein. So wurde an ungeschädigter DNA auch ein lokaler Biegewinkel von  $\sim 5^\circ$  gemessen.

Gegenüber den oben vorgestellten Programmen wies das hier etablierte Vorgehen entscheidende Vorteile auf. Während die anderen Programme für die Vermessung ungeschädigter DNA konzipiert wurden, zeigte sich die Biegewinkelmessung an einem spezifischen DNA-Schaden komplexer, da der Schaden vor der eigentlichen Biegewinkelmessung lokalisiert werden musste. Die Methode dieser Arbeit verfolgte hierbei einen relativ einfachen Ansatz, um den Biegewinkel direkt an spezifischen Schadensstellen zu messen (Implementierung der Messung bei 50 % DNA-Länge). Da sich der Schaden präparationsbedingt immer an der 50 %-Position der DNA-Länge befand, entfiel eine aufwändige Schadenslokalisierung. Wie aus den Ergebnissen mit der automatischen Messsoftware hervorgeht, eignet sich das Tangentenverfahren sehr gut zur Bestimmung des Biegewinkels an der Schadensstelle. Ein Vorteil der MATLAB-Programmierung ist zudem die Möglichkeit, die Parameter an neue Aufgaben anpassen zu können. So kann sowohl die Messmethode als Tangente oder Sekante durchgeführt werden, als auch die Länge der Messgeraden variiert werden. Die Variation der Tangentenlänge ermöglicht auch die Analyse von deutlich größeren Protein-DNA-Komplexen. Ein entscheidender Vorteil des hier etablierten Programms ist die Möglichkeit, dass im Vergleich mit den zuvor aufgeführten Programmen der Biegewinkel nicht nur an DNA gemessen werden kann, sondern durch die Kombination mit der zuvor ermittelten Proteinlokalisierung („3D Object Counter“) auch Biegewinkel an Protein-DNA-Komplexen automatisch bestimmt werden können. Dies erwies sich für die objektive Untersuchung des Biegeverhaltens von Proteinen als äußerst wichtig.

## 4.4 Biologische Relevanz

Das Modell von Büchner *et al.* [1] (in Kapitel 1.5 beschrieben) stellt eine mögliche Erklärung für die bisher noch immer nicht vollständig geklärte Schadenserkennung durch Glykosylasen im initialen Schritt der BER dar. Die Anforderungen an die Glykosylasen als Initiator der BER sind: Einen Basenschaden in kurzer Zeit, sehr effizient und mit geringem Energieaufwand aus der immens großen Menge an DNA sicher zu identifizieren. Da es der Aufwendung von Energie bedarf, eine Base extrahelikal zu flippen, muss es für die Glykosylasen eine schnelle Möglichkeit geben, während des thermodynamischen Gleitens entlang des DNA-Stranges einen potentiell passenden Basenschaden zu identifizieren, ohne jede Base in das katalytische Zentrum schwenken zu müssen und dort die strukturelle Übereinstimmung zwischen Schaden und aktivem Zentrum zu prüfen [34, 114]. Jede einzelne Base extrahelikal zu schwenken, würde viel Zeit und Energie in Anspruch nehmen [78]. Das Modell zur Schadenssuche basiert auf den drei verschiedenen Konformationen: Suchkomplex (SC), Abfragekomplex (IC) und Exzisionskomplex (EC) (Abbildung 11) [1, 24]. In der Suchkomplex-Konformation gleitet die Glykosylase entlang der DNA [58]. Ein Basenschaden kann durch eine Veränderung der Basenpaarung zu einer Deformation der Doppelhelix-Struktur führen und/oder eine erhöhte Flexibilität am Schadensort verursachen [24]. Während der Suchkomplex für eine Biegung ungeschädigter DNA einen höheren Energiebetrag aufwenden muss, unterstützt die erhöhte intrinsische Flexibilität am Basenschaden den Biegevorgang [225]. Durch diesen erleichterten und energiesparenden Biegevorgang kommt es zu einer längeren Verweildauer am Schadensort [1]. Da die Glykosylasen permanent zwischen der SC- und IC-Konformation wechseln, führt dies zu einer erhöhten Wahrscheinlichkeit für ein spontanes Flippen der Base durch Basenatmung [57, 61, 114] während der Dauer der Glykosylase-Bindung. Durch „Phosphate-Pinching“ („Kneifen in das Phosphatrückgrat“) kann das Basenflippen weiterhin verstärkt werden [32, 66, 123, 124]. Durch das Einschwenken eines Strukturelementes wie eines Arginin-Fingers (hTDG spezifisch) oder alternativer Proteinstrukturen in den entstandenen Freiraum der DNA-Helix kann die extrudierte Base im aktiven Zentrum der Glykosylase zusätzlich stabilisiert werden [20, 125]. Büchner *et al.* [1] hatten für

ungeschädigte DNA in Abwesenheit der Thymin-DNA-Glykosylase (hTDG) einen Biegewinkel von  $0^\circ$ , in Anwesenheit von hTDG zwei enzymespezifische Biegewinkel festgestellt. Der kleinere Biegewinkel demonstrierte den Suchkomplex und der größere von beiden den Abfragekomplex. Bei DNA, die einen hTDG-spezifischen Basenschaden enthielt, wurden am Basenschaden ein Biegewinkel, der dem Glykosylase-spezifischen Suchkomplex entsprach, auch ohne die Anwesenheit der Glykosylase gemessen. In Anwesenheit der Glykosylase wurden entsprechend den Ergebnissen an ungeschädigter DNA die beiden Winkel für den SC- und IC-Komplex auch am hTDG-spezifischen Basenschaden gefunden [1]. Sowohl Chen *et al.* [122] als auch Büchner *et al.* [1] konnten ebenfalls für die Glykosylase hOGG1 Biegewinkel beobachten, die das hier vorgestellte Modell untermauern. So beschrieben diese Publikationen Populationen bei  $\sim 0^\circ$  und  $\sim 30^\circ$  (SC) sowie eine Population Biegewinkel von  $\sim 70^\circ$  (IC) [122,1]. Der Biegewinkel von ca.  $70^\circ$  stimmt mit den Literaturangaben für die Kristallstruktur (IC/EC) des hOGG1-oxoG-DNA-Komplexes überein [26].

Eine Übertragung des Modells von Büchner *et al.* [1] auf die Ergebnisse dieser Arbeit soll im Folgenden durchgeführt werden. Gemäß dem Modell sollten für Glykosylase-Komplexe an ungeschädigter DNA zwei verschiedengroße Biegewinkelspezies (SC und IC) messbar sein. Bei Betrachtung der Biegewinkelverteilungen für MutY-nsp-DNA-Komplexe zeigten sich sowohl für die automatische Biegewinkelmessung mit 8 nm Tangenten, als auch für die manuelle Messmethode zwei eindeutig trennbare Biegungswinkel ( $\sim 15^\circ$  und  $\sim 50^\circ$ , siehe Abbildung 108). Wie zuvor im Kapitel 4.2.6 diskutiert, entspricht der größere Biegewinkel dem aus der Literatur bekannten Wert für die Kristallstruktur des MutY-oxoG:A-Komplex von  $55^\circ$  [93]. Dies kann dem Abfragekomplex (IC) zugeordnet werden, da auch in diesen röntgenkristallografischen Untersuchungen die Base extraheikal in der Bindungstasche von MutY fixiert ist. Der kleinere Biegewinkel ( $\sim 15^\circ$ ) kann als Suchkomplex interpretiert werden (siehe Abbildung 108).

Die Übereinstimmung der gemessenen Biegewinkel für die DNA-Schadenssubstrate von MutY (G:A und oxoG:A) mit den durch MutY an ungeschädigter DNA hervorgerufenen Biegungen ist für das Modell von Büchner *et al.* [1] entscheidend. Die Biegewinkelbestimmungen der beiden MutY-Substrate G:A und

oxoG:A ergaben sowohl bei den manuellen Ergebnissen von Lisa Hirsch (22° und 24°), als auch bei den Ergebnissen der automatisierten Auswertung (15° und 15°) ähnliche Konformationen. Die Winkel aus manuellen und automatisierten Messungen unterschieden sich zwar um ca. 10°, stimmten aber innerhalb einer Messmethode überein. Diese Ergebnisse sind konsistent mit der Vorstellung, dass die verschiedenen Zielstrukturen einer Glykosylase auch ähnliche Biege- winkel aufweisen. Bei der automatischen Winkelmessung mit der Tangentenmethode ergab sich außerdem für beide Substrate (G:A und oxoG:A) zusätzlich ein Winkel von  $\sim 0^\circ$  (2° und 0°). Dieses Ergebnis zeigt, dass unabhängig von der Messmethode ein identischer und folglich glykosylasenspezifischer Biege- winkel für die verwandten Schadenssubstrate von MutY besteht. Die beiden MutY-Scha- denssubstrate besitzen folglich eine vergleichbare schadensbedingte Flexibilität.

Gemäß dem Modell kommt es zu einer Übereinstimmung zwischen den Biege- winkeln am Schadenssubstrat und der durch den MutY-Komplex an ungeschä- digter DNA verursachten Biegung. Die automatisierte Messung mit der Tangen- tenmethode an MutY-Komplexen mit ungeschädigter DNA ergab die beiden Win- kel von  $\sim 15^\circ$  und  $50^\circ$ . Auch bei eigenen manuellen Messungen ergaben sich an denselben Aufnahmen  $15^\circ$  und  $52^\circ$ . Gemäß dem Modell sollte der Biege- winkel MutY-spezifischer Schäden mit dem des Suchkomplexes von MutY übereinstim- men. Dies traf folglich auf die Biege- winkelverteilung von MutY mit ungeschädigter DNA zu. Der Winkel von ca.  $15^\circ$  kann folglich dem SC-Komplex und der Winkel von ca.  $50^\circ$  dem IC-Komplex zugeordnet werden. Weiterhin ist in der MutY-nsp- DNA-Biege- winkel-Verteilung der für ungeschädigte DNA beobachtete Biege- winkel ( $5^\circ$ ) nicht erkennbar, was ein klares Zeichen für die aktive Biegung von MutY ist. Aus der Biege- winkelverteilung lässt sich eine Stabilisierung der SC- und IC- Zustände durch MutY ableiten. Die große Ähnlichkeit der Biege- winkel zwischen den MutY spezifischen DNA-Schäden in Abwesenheit der Glykosylase und dem SC-Komplex von MutY-DNA-Komplexen an ungeschädigter DNA bestätigen das Modell von Büchner *et al.* [1] zur initialen Schadensdetektion durch Glykosylasen. Zudem kommt es hier zu einer guten Übereinstimmung zwischen den aus der Röntgenstrukturanalyse recherchierten Winkeln von MutY-DNA-Komplexen und den in rasterkraftmikroskopischen Aufnahmen ermittelten Werten. Auch die

Tatsache, dass sich die Biegewinkel von oxoG:C (hOGG1) und εA:T (hAAG) von den beiden oben genannten ähnlichen Biegewinkeln von G:A und oxoG:A für MutY unterscheiden, spricht für das Modell der AG Tessmer, da nicht nur ähnliche Zielschäden einer Glykosylase vergleichbare Biegewinkel besitzen, sondern auch die Substrate anderer Glykosylasen verschiedene Biegewinkel zeigen. Nachfolgend stellt die Tabelle 7 eine Übersicht der Ergebnisse zur Glykosylase MutY und deren Schadenssubstrate sowie ihrer biologischen Einordnung dar.

**Tabelle 7: Biologische Einordnung der Messergebnisse**

Schadenssubstrat von MutY		8 nm	Interpretation
oxoG:A	anti – anti	0°	Watson-Crick
	syn – anti	15°	Hoogsten-Face
	anti – syn	30°	
G:A	anti – anti	0°	Watson-Crick
	syn – anti	15°	(Hoogsten-Face)
	anti – syn	27°	
<b>automatische Auswertung an MutY-Komplexen mit ungeschädigter DNA</b>			
ohne Filterung nach Proteinvolumen		14°	SC-Komplex
		46°	IC-Komplex

Zusammenfassend wurde im Rahmen dieser Doktorarbeit eine Methode zur objektiven automatisierten DNA-Biegewinkelmessung etabliert und evaluiert. Diese wurde erfolgreich auf die Untersuchung von DNA-Schäden und Glykosylase-DNA-Komplexen angewendet.

## 5 Ausblick

Das Ziel dieser Arbeit – die Etablierung einer automatisierten Messsoftware zur objektiven Biegewinkelmessung – wurde mit verschiedenen Bildverarbeitungsprogrammen und einer MATLAB-Messsoftware erreicht. Die automatisierte Messmethode für DNA-Biegewinkel hat sich als vorteilhaft erwiesen und bietet die Grundlage für eine schnelle standardisierte Auswertung. Die mit der automatischen Messmethode gewonnenen Ergebnisse stützen das Modell von Büchner *et al.* [1] und tragen so zu einem besseren Verständnis der initialen Schadenserkennung von Glykosylasen bei. Als neues Ziel für zuverlässige und signifikante Informationen kann die Gewinnung einer größeren Datenmenge mit Hilfe der Messsoftware definiert werden. Basierend auf der hier vorgestellten Methode wurden mittlerweile die Versuche von Büchner *et al.* [1] an der Glykosylase hTDG erneut automatisiert ausgewertet und eine Übereinstimmung mit den publizierten Daten beobachtet. Die hier beschriebenen Untersuchungen resultierten in einem Manuskript, das sich aktuell unter Begutachtung befindet.

Zudem wäre es sehr sinnvoll, die Reproduzierbarkeit und die Validität der automatischen Winkelbestimmung an einem Testmodul durchzuführen. Hierfür könnte sich beispielsweise die Funktion einer Parabel mit definiertem Biegewinkel eignen.

Interessant wären weiterhin Biegewinkelmessungen an Protein-DNA-Komplexen von MutY mit schadensspezifischer DNA wie G:A oder oxoG:A zu wiederholen. Dies wäre mit dem beschriebenen Auswertungsverfahren gut zu bewerkstelligen, da sich der Basenschaden der DNA präparationsbedingt bei 50 % der DNA-Länge befindet und FIESTA die 50 %-Länge an jedem DNA-Filament für die Winkelmessung durch MATLAB markiert (siehe Abbildung 45). Problematisch in dieser Hinsicht könnte nur der Fall einer getrennten Skelettierung durch FIESTA werden, falls der DNA-Strang am Protein-Locus in der Skelettierung unterbrochen wird, wie häufig beobachtet. Durch eine manuelle Auswahl können die zwei Filamente in MATLAB rechnerisch verbunden werden. Allerdings erfolgt die Biegewinkelmessung an der 50 %-Position auf diese Weise nicht an der FIESTA-Skelettierung, sondern an der Verbindungslinie. Bei der Biegewinkelmessung an



DNA kam es aufgrund der kurzen Distanz zu keiner nennenswerten Beeinträchtigung. Es sollte geprüft werden, in wie weit dies auf Protein-DNA-Komplexe übertragbar ist.

Für die Biege winkelmessung von Protein-DNA-Komplexe am spezifischen Schaden wäre es eventuell hilfreich eine Zusatzoption in MATLAB zu integrieren, die im Ausgabebild die 50 %-Position eines jeden DNA-Stranges markiert und misst. Ein anderer Ansatz dies zu vereinfachen wäre, die Auswertung in MATLAB so zu modifizieren, dass die Biege winkelmessung ausschließlich an Protein-DNA-Komplexen durchgeführt wird, die sich bei 50 % der DNA-Länge befinden. Interessant wäre es auch die verschiedenen Glykosylasen nicht nur mit den jeweiligen spezifischen Schäden zu inkubieren, sondern auch mit spezifischen Schäden anderer Glykosylasen. Durch die Lage der Schäden an der 50 %-Position der DNA-Länge könnten gemäß der Positionsverteilung der Proteine Rückschlüsse auf die mögliche Erkennung von Fremdschäden durch diese Glykosylasen angestellt werden.

## 6 Zusammenfassung

Ziel der Arbeit war, mit Hilfe des von Büchner *et al.* [1] vorgestellten Modells die Vorgänge der initialen Schadenserkennung durch Glykosylasen im BER-Prozess genauer zu untersuchen. Grundlage des Modells der AG Tessmer war die Beobachtung, dass in rasterkraftmikroskopischen Aufnahmen der Biegungswinkel der DNA, der durch einen Basenschaden am Schadensort induziert wird, sehr ähnlich zu dem gemessenen Biegewinkel des Glykosylase-DNA-Komplexes an ungeschädigter DNA ist. Da sich bei der manuellen Biegewinkelmessung mit ImageJ bei unterschiedlichen Experimentatoren differierende Messwerte ergaben, wurde nach einer möglichen Methode zur Objektivierung und Vereinfachung der Biegewinkelmessung an DNA und DNA-Protein-Komplexen gesucht. So lag der Fokus dieser Arbeit größtenteils auf der Entwicklung eines standardisierten und objektiven Messverfahrens. In Zusammenarbeit mit Hannah Heil (AG Heinze) wurde diese Methode entwickelt. Durch vorbereitende Bildbearbeitungsschritte in ImageJ, die Skelettierung in FIESTA und schließlich die Biegewinkelmessung mittels 8 nm langer Tangenten durch MATLAB konnte ein solches Verfahren etabliert werden. Die ermittelten Werte für die Biegung ungeschädigter DNA wichen von den Werten aus Vorgängerarbeiten leicht ab, stimmten allerdings mit in der Literatur zitierten Werten zur DNA-Biegung überein. Ähnlich verhielt es sich auch mit den Biegewinkeln von geschädigter DNA. Auch die für Protein-DNA-Komplexe von MutY gemessenen Interrogationskomplex-Biegewinkel deckten sich gut mit den Werten aus der Röntgen-Strukturanalyse. Sie zeigten einen stark gebogenen Komplex (EC) mit  $\sim 50^\circ$  DNA-Biegung und einen weniger stark gebogenen Komplex ( $\sim 15^\circ$ ), der dem Suchkomplex (SC) zugeordnet werden konnte. Damit konnte das Spektrum an unterschiedlichen Glykosylase-DNA-Komplexen, die das Modell der initialen Schadenserkennung durch Glykosylasen unterstützen, auf MutY erweitert werden (zusätzlich zu hTDG und hOGG1 [1]). Ein langfristiges Ziel ist es, mithilfe des hier auf seine Eignung geprüften Programmes im Rahmen anderer themenbezogener Arbeiten viele weitere objektive und konsistente Daten zu gewinnen. Auf diese Weise soll ein wichtiger Beitrag zum besseren Verständnis der DNA-Schadenserkennung geleistet werden.

## 7 Literaturverzeichnis

1. 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 Research, 2015. 43(5): S. 2716-2729.
2. Watson, J.D., Crick, F.H., *Genetical implications of the structure of deoxyribonucleic acid*. Nature, 1953. 171(4361): S. 964-967.
3. Netzker, R., *Zellzyklus und molekulare Genetik*, aus *Biochemie - Duale Reihe*, Rassow, J., Netzker, R., Hauser, K., Deutzmann, R., Hrsgg. 2016, Thieme: Stuttgart. S. 417-419.
4. Bernstein, C., Bernstein, H., *Epigenetic Reduction of DNA Repair in Progression to Cancer*, aus *Advances in DNA Repair*, Chen, C.C., Hrsg. 2015, InTech. S. 171.
5. Yu, S.-L., Lee, S.-K., *Ultraviolet radiation: DNA damage, repair, and human disorders*. Molecular & Cellular Toxicology, 2017. 13(1): S. 21-28.
6. Hollaender, A., Greenstein, J.P., Jenrette, W.V., *Effects of Ultraviolet Radiation on Sodium Thymonucleate*. Journal of the National Cancer Institute, 1941. 2(1): S. 23-28.
7. Rothkamm, K., Löbrich, M., *Evidence for a lack of DNA double-strand break repair in human cells exposed to very low x-ray doses*. Proceedings of the National Academy of Sciences, 2003. 100(9): S. 5057-5062.
8. Muller, H.J., *Artificial transmutation of the gene*. Science, 1927. 66(1699): S. 84-87.
9. Taylor, B., Greenstein, J.P., Hollaender, A., *Effects of X-Radiation on Thymus Nucleic Acid*. Science, 1947. 105(2723): S. 263-264.
10. Cox, R.A., Overend, W.G., Peacocke, A.R., Wilson, S., *The action of gamma-rays on sodium deoxyribonucleate in solution*. Proceedings of the Royal Society of London B: Biological Science, 1958. 149(937): S. 511-533.
11. Halliwell, B., Aruoma, O.I., *DNA damage by oxygen-derived species. Its mechanism and measurement in mammalian systems*. FEBS Letters, 1991. 281(1,2): S. 9-19.
12. Kondo, N., Takahashi, A., Ono, K., Ohnishi, T., *DNA Damage Induced by Alkylating Agents and Repair Pathways*. Journal of Nucleic Acids, 2010. 2010.
13. Drablos, F. et al., *Alkylation damage in DNA and RNA--repair mechanisms and medical significance*. DNA Repair, 2004. 3(11): S. 1389-1407.
14. Karran, P., Hjelmgren, T., Lindahl, T., *Induction of a DNA glycosylase for N-methylated purines is part of the adaptive response to alkylating agents*. Nature, 1982. 296(5859): S. 770-773.
15. Caldecott, K.W., *Single-strand break repair and genetic disease*. Nature Reviews Genetics, 2008. 9: S. 619.
16. Mehta, A., Haber, J.E., *Sources of DNA double-strand breaks and models of recombinational DNA repair*. Cold Spring Harbor Perspectives in Biology, 2014. 6(9): S. a016428.
17. Roos, W.P., Kaina, B., *DNA damage-induced cell death: From specific DNA lesions to the DNA damage response and apoptosis*. Cancer Letters, 2013. 332(2): S. 237-248.
18. Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A., Ellenberger, T., *DNA repair and mutagenesis*. 2 ed. 2006: American Society for Microbiology Press. 9-17.

19. Lindahl, T., *An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues*. Proceedings of the National Academy of Sciences, 1974. 71(9): S. 3649-3653.
20. Brooks, S.C., Adhikary, S., Rubinson, E.H., Eichman, B.F., *Recent advances in the structural mechanisms of DNA glycosylases*. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2013. 1834(1): S. 247-271.
21. McCullough, A.K., Dodson, M., Lloyd, R.S., *Initiation of base excision repair: glycosylase mechanisms and structures*. Annual Review of Biochemistry, 1999. 68(1): S. 255-285.
22. David, S.S., Williams, S.D., *Chemistry of Glycosylases and Endonucleases Involved in Base-Excision Repair*. Chemical Reviews, 1998. 98(3): S. 1221-1262.
23. Krokan, H.E., Bjoras, M., *Base excision repair*. Cold Spring Harbor Perspectives in Biology, 2013. 5(4): S. a012583.
24. Friedman, J.I., Stivers, J.T., *Detection of damaged DNA bases by DNA glycosylase enzymes*. Biochemistry, 2010. 49(24): S. 4957-4967.
25. Singer, B., Hang, B., *What Structural Features Determine Repair Enzyme Specificity and Mechanism in Chemically Modified DNA?* Chemical Research in Toxicology, 1997. 10(7): S. 713-732.
26. Stivers, J.T., Jiang, Y.L., *A mechanistic perspective on the chemistry of DNA repair glycosylases*. Chemical Reviews, 2003. 103(7): S. 2729-2759.
27. Lindahl, T., *Instability and decay of the primary structure of DNA*. Nature, 1993. 362(6422): S. 709-715.
28. David, S.S., O'shea, V.L., Kundu, S., *Base-excision repair of oxidative DNA damage*. Nature, 2007. 447(7147): S. 941.
29. Krokan, H.E., Standal, R., Slupphaug, G., *DNA glycosylases in the base excision repair of DNA*. Biochemical Journal, 1997. 325 (1): S. 1-16.
30. Huffman, J.L., Sundheim, O., Tainer, J.A., *DNA base damage recognition and removal: New twists and grooves*. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2005. 577(1): S. 55-76.
31. Slupphaug, G., Mol, C.D., Kavli, B., Arvai, A.S., Krokan, H.E., Tainer, J.A., *A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA*. Nature, 1996. 384(6604): S. 87-92.
32. Parikh, S.S., Mol, C.D., Slupphaug, G., Bharati, S., Krokan, H.E., Tainer, J.A., *Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA*. The EMBO Journal, 1998. 17(17): S. 5214-5226.
33. Sun, B., Latham, K.A., Dodson, M., Lloyd, R.S., *Studies on the Catalytic Mechanism of Five DNA Glycosylases PROBING FOR ENZYME-DNA IMINO INTERMEDIATES*. Journal of Biological Chemistry, 1995. 270(33): S. 19501-19508.
34. Fromme, J.C., Verdine, G.L., *Base Excision Repair*, aus *Advances in Protein Chemistry*. 2004, Academic Press. S. 1-41.
35. Carey, D.C., Strauss, P.R., *Human apurinic/aprimidinic endonuclease is processive*. Biochemistry, 1999. 38(50): S. 16553-16560.
36. Fortini, P., Parlanti, E., Sidorkina, O.M., Laval, J., Dogliotti, E., *The type of DNA glycosylase determines the base excision repair pathway in mammalian cells*. Journal of Biological Chemistry, 1999. 274(21): S. 15230-15236.
37. Robertson, A.B., Klungland, A., Rognes, T., Leiros, I., *DNA repair in mammalian cells: Base excision repair: the long and short of it*. Cellular and Molecular Life Sciences, 2009. 66(6): S. 981-993.

38. Klungland, A., Lindahl, T., *Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirement for DNase IV (FEN1)*. The EMBO Journal, 1997. 16(11): S. 3341-3348.
39. Fortini, P., Pascucci, B., Parlanti, E., Sobol, R.W., Wilson, S.H., Dogliotti, E., *Different DNA polymerases are involved in the short- and long-patch base excision repair in mammalian cells*. Biochemistry, 1998. 37(11): S. 3575-3580.
40. Singhal, R.K., Prasad, R., Wilson, S.H., *DNA polymerase beta conducts the gap-filling step in uracil-initiated base excision repair in a bovine testis nuclear extract*. Journal of Biological Chemistry, 1995. 270(2): S. 949-957.
41. van Loon, B., Markkanen, E., Hubscher, U., *Oxygen as a friend and enemy: How to combat the mutational potential of 8-oxo-guanine*. DNA Repair, 2010. 9(6): S. 604-616.
42. Xu, G., Herzig, M., Rotrekl, V., Walter, C.A., *Base excision repair, aging and health span*. Mechanisms of Ageing and Development, 2008. 129(7-8): S. 366-382.
43. Wallace, S.S., *Base excision repair: a critical player in many games*. DNA Repair, 2014. 19: S. 14-26.
44. Wallace, S.S., Murphy, D.L., Sweasy, J.B., *Base excision repair and cancer*. Cancer Letters, 2012. 327(1-2): S. 73-89.
45. Dizdaroglu, M., *Oxidatively induced DNA damage and its repair in cancer*. Mutation Research - Reviews in Mutation Research, 2015. 763: S. 212-245.
46. Stavnezer, J., *Complex regulation and function of activation-induced cytidine deaminase*. Trends in Immunology, 2011. 32(5): S. 194-201.
47. Bellacosa, A., Drohat, A.C., *Role of base excision repair in maintaining the genetic and epigenetic integrity of CpG sites*. DNA Repair, 2015. 32: S. 33-42.
48. Nabel, C.S., Manning, S.A., Kohli, R.M., *The curious chemical biology of cytosine: deamination, methylation, and oxidation as modulators of genomic potential*. ACS Chemical Biology, 2012. 7(1): S. 20-30.
49. Scheffler, K. et al., *8-oxoguanine DNA glycosylase (Ogg1) controls hepatic gluconeogenesis*. DNA Repair, 2018. 61: S. 56-62.
50. Liu, M., Doublé, S., Wallace, S.S., *Neil3, the final frontier for the DNA glycosylases that recognize oxidative damage*. Mutation Research, 2013. 0: S. 4-11.
51. Xu, J. et al., *Thymine DNA Glycosylase (TDG) is involved in the pathogenesis of intestinal tumors with reduced APC expression*. Oncotarget, 2017. 8(52): S. 89988-89997.
52. Cortazar, D. et al., *Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability*. Nature, 2011. 470(7334): S. 419-423.
53. Schormann, N., Ricciardi, R., Chattopadhyay, D., *Uracil-DNA glycosylases- structural and functional perspectives on an essential family of DNA repair enzymes*. Protein Science, 2014. 23(12): S. 1667-1685.
54. Fromme, J.C., Verdine, G.L., *Structure of a trapped endonuclease III-DNA covalent intermediate*. The EMBO Journal, 2003. 22(13): S. 3461-3471.
55. Sugahara, M. et al., *Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, Thermus thermophilus HB8*. The EMBO Journal, 2000. 19(15): S. 3857-3869.
56. Cravens, S.L., Hobson, M., Stivers, J.T., *Electrostatic properties of complexes along a DNA glycosylase damage search pathway*. Biochemistry, 2014. 53(48): S. 7680-7692.
57. Porecha, R.H., Stivers, J.T., *Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils*. Proceedings of the National Academy of Sciences of the United States of America, 2008. 105(31): S. 10791-10796.

58. Blainey, P.C., van Oijen, A.M., Banerjee, A., Verdine, G.L., Xie, X.S., *A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA*. Proceedings of the National Academy of Sciences, 2006. 103(15): S. 5752-5757.
59. Schonhofs, J.D., Stivers, J.T., *Timing facilitated site transfer of an enzyme on DNA*. Nature Chemical Biology, 2012. 8(2): S. 205.
60. Olmon, E.D., Delaney, S., *Differential Ability of Five DNA Glycosylases to Recognize and Repair Damage on Nucleosomal DNA*. ACS Chemical Biology, 2017. 12(3): S. 692-701.
61. Stivers, J.T., *Extrahelical Damaged Base Recognition by DNA Glycosylase Enzymes*. Chemistry, 2008. 14(3): S. 786-793.
62. Wang, L., Chakravarthy, S., Verdine, G.L., *Structural Basis for the Lesion-scanning Mechanism of the MutY DNA Glycosylase*. Journal of Biological Chemistry, 2017. 292(12): S. 5007-5017.
63. Ba, X., Boldogh, I., *8-Oxoguanine DNA glycosylase 1: Beyond repair of the oxidatively modified base lesions*. Redox Biology, 2018. 14: S. 669-678.
64. Hegde, M.L., Hazra, T.K., Mitra, S., *Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells*. Cell Research, 2008. 18(1): S. 27-47.
65. UniProt Consortium. *UniProtKB - O15527 (OGG1\_HUMAN)*. (2019); Verfügbar unter: <https://www.uniprot.org/uniprot/O15527> [29.07.2019].
66. Bruner, S.D., Norman, D.P., Verdine, G.L., *Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA*. Nature, 2000. 403(6772): S. 859-866.
67. Dizdaroglu, M., *Chemical determination of free radical-induced damage to DNA*. Free Radical Biology and Medicine, 1991. 10(3-4): S. 225-242.
68. Dedon, P.C., *The chemical toxicology of 2-deoxyribose oxidation in DNA*. Chemical Research in Toxicology, 2008. 21(1): S. 206-219.
69. Cadet, J., Douki, T., *Oxidatively generated damage to DNA by UVA radiation in cells and human skin*. Journal of Investigative Dermatology, 2011. 131(5): S. 1005-1007.
70. Cadet, J., Sage, E., Douki, T., *Ultraviolet radiation-mediated damage to cellular DNA*. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2005. 571(1): S. 3-17.
71. Ward, J.F., *Some biochemical consequences of the spatial distribution of ionizing radiation-produced free radicals*. Radiation Research, 1981. 86(2): S. 185-195.
72. Yagura, T., Makita, K., Yamamoto, H., Menck, C.F.M., Schuch, A.P., *Biological Sensors for Solar Ultraviolet Radiation*. Sensors, 2011. 11(4): S. 4277-4294.
73. Grollman, A.P., Moriya, M., *Mutagenesis by 8-oxoguanine: an enemy within*. Trends in Genetics, 1993. 9(7): S. 246-249.
74. Cadet, J., Wagner, J.R., Shafirovich, V., Geacintov, N.E., *One-electron oxidation reactions of purine and pyrimidine bases in cellular DNA*. International Journal of Radiation Biology, 2014. 90(6): S. 423-432.
75. Kouchakdjian, M. et al., *NMR structural studies of the ionizing radiation adduct 7-hydro-8-oxodeoxyguanosine (8-oxo-7H-dG) opposite deoxyadenosine in a DNA duplex. 8-Oxo-7H-dG(syn).dA(anti) alignment at lesion site*. Biochemistry, 1991. 30(5): S. 1403-1412.
76. McCulloch, S.D., Kokoska, R.J., Garg, P., Burgers, P.M., Kunkel, T.A., *The efficiency and fidelity of 8-oxo-guanine bypass by DNA polymerases delta and eta*. Nucleic Acids Research, 2009. 37(9): S. 2830-2840.

77. Michaels, M.L., Miller, J.H., *The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine)*. Journal of Bacteriology, 1992. 174(20): S. 6321-6325.
78. Li, H. et al., *DNA Deformation-Coupled Recognition of 8-Oxoguanine: Conformational Kinetic Gating in Human DNA Glycosylase*. Journal of the American Chemical Society, 2017. 139(7): S. 2682-2692.
79. Beard, W.A., Batra, V.K., Wilson, S.H., *DNA polymerase structure-based insight on the mutagenic properties of 8-oxoguanine*. Mutation Research, 2010. 703(1): S. 18-23.
80. Aburatani, H. et al., *Cloning and characterization of mammalian 8-hydroxyguanine-specific DNA glycosylase/apurinic, apyrimidinic lyase, a functional mutM homologue*. Cancer Research, 1997. 57(11): S. 2151-2156.
81. Tchou, J. et al., *8-oxoguanine (8-hydroxyguanine) DNA glycosylase and its substrate specificity*. Proceedings of the National Academy of Sciences of the United States of America, 1991. 88(11): S. 4690-4694.
82. Stivers, J.T., *Site-specific DNA damage recognition by enzyme-induced base flipping*. Progress in Nucleic Acid Research and Molecular Biology, 2004. 77: S. 37-65.
83. Dalhus, B., Laerdahl, J.K., Backe, P.H., Bjoras, M., *DNA base repair--recognition and initiation of catalysis*. FEMS Microbiology Reviews, 2009. 33(6): S. 1044-1078.
84. Norman, D.P., Chung, S.J., Verdine, G.L., *Structural and biochemical exploration of a critical amino acid in human 8-oxoguanine glycosylase*. Biochemistry, 2003. 42(6): S. 1564-1572.
85. Greenman, C. et al., *Patterns of somatic mutation in human cancer genomes*. Nature, 2007. 446: S. 153.
86. Hollstein, M. et al., *Somatic point mutations in the p53 gene of human tumors and cell lines: updated compilation*. Nucleic Acids Research, 1996. 24(1): S. 141-146.
87. Tuo, J., Chen, C., Zeng, X., Christiansen, M., Bohr, V.A., *Functional crosstalk between hOgg1 and the helicase domain of Cockayne syndrome group B protein*. DNA Repair, 2002. 1(11): S. 913-927.
88. Stevnsner, T. et al., *Mitochondrial repair of 8-oxoguanine is deficient in Cockayne syndrome group B*. Oncogene, 2002. 21(57): S. 8675-8682.
89. Williams, S.D., David, S.S., *Evidence that MutY is a monofunctional glycosylase capable of forming a covalent Schiff base intermediate with substrate DNA*. Nucleic Acids Research, 1998. 26(22): S. 5123-5133.
90. UniProt Consortium. *UniProtKB - P17802 (MUTY\_ECOLI)*. (2019); Verfügbar unter: <https://www.uniprot.org/uniprot/P17802> [29.07.2019].
91. 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*. Proceedings of the National Academy of Sciences of the United States of America, 1986. 83(8): S. 2402-2406.
92. Guan, Y. et al., *MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily*. Nature Structural & Molecular Biology, 1998. 5(12): S. 1058-1064.
93. 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*. Nature, 2004. 427(6975): S. 652-656.
94. Noll, D.M., Gogos, A., Granek, J.A., Clarke, N.D., *The C-terminal domain of the adenine-DNA glycosylase MutY confers specificity for 8-oxoguanine.adenine mispairs and may have evolved from MutT, an 8-oxo-dGTPase*. Biochemistry, 1999. 38(20): S. 6374-6379.

95. Manlove, K.R., Cassirer, E.F., Plowright, R.K., Cross, P.C., Hudson, P.J., *Contact and contagion: Probability of transmission given contact varies with demographic state in bighorn sheep*. *Journal of Animal Ecology*, 2017. 86(4): S. 908-920.
96. Al-Tassan, N. et al., *Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors*. *Nature genetics*, 2002. 30(2): S. 227-232.
97. Aceto, G.M. et al., *Correlation between mutations and mRNA expression of APC and MUTYH genes: new insight into hereditary colorectal polyposis predisposition*. *Journal of Experimental & Clinical Cancer Research*, 2015. 34: S. 131.
98. Weren, R.D., Ligtenberg, M.J., Geurts van Kessel, A., De Voer, R.M., Hoogerbrugge, N., Kuiper, R.P., *NTHL1 and MUTYH polyposis syndromes: two sides of the same coin?* *The Journal of Pathology*, 2018. 244(2): S. 135-142.
99. Engelward, B.P. et al., *Base excision repair deficient mice lacking the Aag alkyladenine DNA glycosylase*. *Proceedings of the National Academy of Sciences*, 1997. 94(24): S. 13087-13092.
100. Samson, L., Derfler, B., Boosalis, M., Call, K., *Cloning and characterization of a 3-methyladenine DNA glycosylase cDNA from human cells whose gene maps to chromosome 16*. *Proceedings of the National Academy of Sciences of the United States of America*, 1991. 88(20): S. 9127-9131.
101. UniProt Consortium. *UniProtKB - P29372 (3MG\_HUMAN)*. (2019); Verfügbar unter: <https://www.uniprot.org/uniprot/P29372> [29.07.2019].
102. Chakravarti, D., Ibeanu, G.C., Tano, K., Mitra, S., *Cloning and expression in Escherichia coli of a human cDNA encoding the DNA repair protein N-methylpurine-DNA glycosylase*. *Journal of Biological Chemistry*, 1991. 266(24): S. 15710-15715.
103. O'Brien, P.J., Ellenberger, T., *Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase*. *The Journal of Biological Chemistry*, 2004. 279(11): S. 9750-9757.
104. Lenz, S.A.P., Wetmore, S.D., *QM/MM Study of the Reaction Catalyzed by Alkyladenine DNA Glycosylase: Examination of the Substrate Specificity of a DNA Repair Enzyme*. *The Journal of Physical Chemistry B*, 2017. 121(49): S. 11096-11108.
105. Lau, A.Y., Scharer, 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*, 1998. 95(2): S. 249-258.
106. Boysen, G., Pachkowski, B.F., Nakamura, J., Swenberg, J.A., *The formation and biological significance of N7-guanine adducts*. *Mutation Research*, 2009. 678(2): S. 76-94.
107. Kartalou, M., Samson, L.D., Essigmann, J.M., *Cisplatin adducts inhibit 1,N(6)-ethenoadenine repair by interacting with the human 3-methyladenine DNA glycosylase*. *Biochemistry*, 2000. 39(27): S. 8032-8038.
108. Basu, A.K., Wood, M.L., Niedernhofer, L.J., Ramos, L.A., Essigmann, J.M., *Mutagenic and genotoxic effects of three vinyl chloride-induced DNA lesions: 1,N6-ethenoadenine, 3,N4-ethenocytosine, and 4-amino-5-(imidazol-2-yl)imidazole*. *Biochemistry*, 1993. 32(47): S. 12793-12801.
109. Guliaev, A.B., Hang, B., Singer, B., *Structural insights by molecular dynamics simulations into differential repair efficiency for ethano-A versus etheno-A adducts by the human alkylpurine-DNA N-glycosylase*. *Nucleic Acids Research*, 2002. 30(17): S. 3778-3787.
110. Ezaz-Nikpay, K., Verdine, G.L., *The effects of N7-methylguanine on duplex DNA structure*. *Chemistry & Biology*, 1994. 1(4): S. 235-240.



111. Lavery, R., Sklenar, H., *Defining the structure of irregular nucleic acids: conventions and principles*. Journal of Biomolecular Structure and Dynamics, 1989. 6(4): S. 655-667.
112. Huang, C.M., Chen, S.Y., Huang, P.H., Tsai, F.J., *Effect of MPG gene rs2858056 polymorphism, copy number variation, and level of serum MPG protein on the risk for rheumatoid arthritis*. PLOS ONE, 2015. 10(3): DOI: 10.1371/journal.pone.0120699. [10.08.2019]
113. Speina, E. et al., *Decreased repair activities of 1,N(6)-ethenoadenine and 3,N(4)-ethenocytosine in lung adenocarcinoma patients*. Cancer Research, 2003. 63(15): S. 4351-4357.
114. Verdine, G.L., Bruner, S.D., *How do DNA repair proteins locate damaged bases in the genome?* Chemistry & Biology, 1997. 4(5): S. 329-334.
115. Varani, G., McClain, W.H., *The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems*. EMBO Reports, 2000. 1(1): S. 18-23.
116. Moe, J.G., Russu, I.M., *Kinetics and energetics of base-pair opening in 5'-d (CGCGAATTCGCG)-3' and a substituted dodecamer containing G. cntdot. T mismatches*. Biochemistry, 1992. 31(36): S. 8421-8428.
117. Hunter, W.N., Brown, T., Kneale, G., Anand, N.N., Rabinovich, D., Kennard, O., *The structure of guanosine-thymidine mismatches in B-DNA at 2.5-Å resolution*. Journal of Biological Chemistry, 1987. 262(21): S. 9962-9970.
118. Buechner, C.N., Heil, K., Michels, G., Carell, T., Kisker, C., Tessmer, I., *Strand-specific recognition of DNA damages by XPD provides insights into nucleotide excision repair substrate versatility*. Journal of Biological Chemistry, 2014. 289(6): S. 3613-3624.
119. Tessmer, I., Melikishvili, M., Fried, M.G., *Cooperative cluster formation, DNA bending and base-flipping by O6-alkylguanine-DNA alkyltransferase*. Nucleic Acids Research, 2012. 40(17): S. 8296-8308.
120. Wang, H. et al., *DNA bending and unbending by MutS govern mismatch recognition and specificity*. Proceedings of the National Academy of Sciences, 2003. 100(25): S. 14822-14827.
121. Rees, W.A., Keller, R.W., Vesenska, J.P., Yang, G., Bustamante, C., *Evidence of DNA bending in transcription complexes imaged by scanning force microscopy*. Science, 1993. 260(5114): S. 1646-1649.
122. Chen, L., Haushalter, K.A., Lieber, C.M., Verdine, G.L., *Direct visualization of a DNA glycosylase searching for damage*. Chemistry & Biology, 2002. 9(3): S. 345-350.
123. 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*. Proceedings of the National Academy of Sciences, 2008. 105(26): S. 8890-8895.
124. Barrett, T.E. et al., *Crystal structure of a G: T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions*. Cell, 1998. 92(1): S. 117-129.
125. Maiti, A., Morgan, M.T., Drohat, A.C., *Role of two strictly conserved residues in nucleotide flipping and N-glycosylic bond cleavage by human thymine DNA glycosylase*. Journal of Biological Chemistry, 2009. 284(52): S. 36680-36688.
126. Ruhnnow, F., Zwicker, D., Diez, S., *Tracking single particles and elongated filaments with nanometer precision*. Biophysical Journal, 2011. 100(11): S. 2820-2828.
127. Hohenwarter, M., *GeoGebra-didaktische Materialien und Anwendungen für den Mathematikunterricht*, 2006. Dissertation, Universität Salzburg

128. Buechner, C.N., Tessmer, I., *DNA substrate preparation for atomic force microscopy studies of protein-DNA interactions*. Journal of Molecular Recognition, 2013. 26(12): S. 605-617.
129. Yanisch-Perron, C., Vieira, J., Messing, J., *Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors*. Gene, 1985. 33(1): S. 103-119.
130. Butt, H.-J., Cappella, B., Kappl, M., *Force measurements with the atomic force microscope: Technique, interpretation and applications*. Surface Science Reports, 2005. 59(1): S. 1-152.
131. Binnig, G., Quate, C.F., Gerber, C., *Atomic force microscope*. Physical Review Letters, 1986. 56(9): S. 930-933.
132. Binnig, G., Rohrer, H., Gerber, C., Weibel, E., *Surface Studies by Scanning Tunneling Microscopy*. Physical Review Letters, 1982. 49(1): S. 57-61.
133. Meyer, G., Amer, N.M., *Novel optical approach to atomic force microscopy*. Applied Physics Letters, 1988. 53(12): S. 1045-1047.
134. Giessibl, F.J., *Advances in atomic force microscopy*. Reviews of Modern Physics, 2003. 75(3): S. 949-983.
135. Weisenhorn, A., Hansma, P., Albrecht, T., Quate, C., *Forces in atomic force microscopy in air and water*. Applied Physics Letters, 1989. 54(26): S. 2651-2653.
136. Hoo, C.M., Starostin, N., West, P., Mecartney, M.L., *A comparison of atomic force microscopy (AFM) and dynamic light scattering (DLS) methods to characterize nanoparticle size distributions*. Journal of Nanoparticle Research, 2008. 10(1): S. 89-96.
137. Round, A.N., Miles, M.J., *Exploring the consequences of attractive and repulsive interaction regimes in tapping mode atomic force microscopy of DNA*. Nanotechnology, 2004. 15(4): S. S176.
138. Fronczek, D.N. et al., *High accuracy FIONA-AFM hybrid imaging*. Ultramicroscopy, 2011. 111(5): S. 350-355.
139. Fang, Y., Hoh, J.H., *Surface-directed DNA condensation in the absence of soluble multivalent cations*. Nucleic Acids Research, 1998. 26(2): S. 588-593.
140. Bustamante, C., Keller, D., Yang, G., *Scanning force microscopy of nucleic acids and nucleoprotein assemblies*. Current Opinion in Structural Biology, 1993. 3(3): S. 363-372.
141. Chepanoske, C.L., Golinelli, M.-P., Williams, S.D., David, S.S., *Positively Charged Residues within the Iron-Sulfur Cluster Loop of E. coli MutY Participate in Damage Recognition and Removal*. Archives of Biochemistry and Biophysics, 2000. 380(1): S. 11-19.
142. Alberts, A., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J., *Molecular Biology of the Cell* Gerland, 1994.
143. Rivetti, C., Guthold, M., Bustamante, C., *Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis*. Journal of Molecular Biology, 1996. 264(5): S. 919-932.
144. Hardin, A.H., Sarkar, S.K., Seol, Y., Liou, G.F., Osheroff, N., Neuman, K.C., *Direct measurement of DNA bending by type IIA topoisomerases: implications for non-equilibrium topology simplification*. Nucleic Acids Research, 2011. 39(13): S. 5729-5743.
145. Cherny, D.I., Striker, G., Subramaniam, V., Jett, S.D., Paleček, E., Jovin, T.M., *DNA bending due to specific p53 and p53 core domain-DNA interactions visualized by electron microscopy*. Journal of Molecular Biology, 1999. 294(4): S. 1015-1026.

146. Le Cam, E. et al., *Conformational analysis of a 139 base-pair DNA fragment containing a single-stranded break and its interaction with human poly (ADP-ribose) polymerase*. *Journal of Molecular Biology*, 1994. 235(3): S. 1062-1071.
147. Schulz, A., Mucke, N., Langowski, J., Rippe, K., *Scanning force microscopy of Escherichia coli RNA polymerase. sigma54 holoenzyme complexes with DNA in buffer and in air*. *Journal of Molecular Biology*, 1998. 283(4): S. 821-836.
148. Bustamante, C., Rivetti, C., *Visualizing protein-nucleic acid interactions on a large scale with the scanning force microscope*. *Annual Review of Biophysics and Biomolecular Structure*, 1996. 25(1): S. 395-429.
149. Fuentes-Perez, M.E., Dillingham, M.S., Moreno-Herrero, F., *AFM volumetric methods for the characterization of proteins and nucleic acids*. *Methods*, 2013. 60(2): S. 113-121.
150. Roth, H.M., Römer, J., Grundler, V., Van Houten, B., Kisker, C., Tessmer, I., *XPB helicase regulates DNA incision by the Thermoplasma acidophilum endonuclease Bax1*. *DNA Repair*, 2012. 11(3): S. 286-293.
151. Wagner, T., Lipinski, H.-G., *IJBlob: an ImageJ library for connected component analysis and shape analysis*. *Journal of Open Research Software*, 2013. 1(1).
152. Dorst, L., Smeulders, A.W.M., *Length estimators for digitized contours*. *Computer Vision, Graphics, and Image Processing*, 1987. 40(3): S. 311-333.
153. Freeman, H., *Computer Processing of Line-Drawing Images*. *ACM Computing Surveys*, 1974. 6(1): S. 57-97.
154. Groen, F.C.A., Verbeek, P.W., *Freeman-code probabilities of object boundary quantized contours*. *Computer Graphics and Image Processing*, 1978. 7(3): S. 391-402.
155. Rivetti, C., Walker, C., Bustamante, C., *Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility*. *Journal of Molecular Biology*, 1998. 280(1): S. 41-59.
156. Maragliano, C., Glia, A., Stefancich, M., Chiesa, M., *Effective AFM cantilever tip size: methods for in-situ determination*. *Measurement Science and Technology*, 2014. 26(1): DOI: 10.1088/0957-0233/26/1/015002. [03.05.2019]
157. Bezanilla, M., Manne, S., Laney, D.E., Lyubchenko, Y.L., Hansma, H.G., *Adsorption of DNA to Mica, Silylated Mica, and Minerals: Characterization by Atomic Force Microscopy*. *Langmuir*, 1995. 11(2): S. 655-659.
158. Arnott, S., Hukins, D.W., *Optimised parameters for A-DNA and B-DNA*. *Biochemical and Biophysical Research Communications*, 1972. 47(6): S. 1504-1509.
159. Mandelkern, M., Elias, J.G., Eden, D., Crothers, D.M., *The dimensions of DNA in solution*. *Journal of Molecular Biology*, 1981. 152(1): S. 153-161.
160. Santos, S., *The Intrinsic Resolution Limit in the Atomic Force Microscope: Implications for Heights of Nano-Scale Features*. *PLOS ONE*, 2011. 6(8): DOI: 10.1371/journal.pone.0023821. [10.09.2019]
161. Moreno-Herrero, F., Colchero, J., Baro, A., *DNA height in scanning force microscopy*. *Ultramicroscopy*, 2003. 96(2): S. 167-174.
162. Müller, D.J., Engel, A., *The height of biomolecules measured with the atomic force microscope depends on electrostatic interactions*. *Biophysical Journal*, 1997. 73(3): S. 1633-1644.
163. Bustamante, C., Erie, D.A., Keller, D., *Biochemical and structural applications of scanning force microscopy*. *Current Opinion in Structural Biology*, 1994. 4(5): S. 750-760.
164. Winzer, A.T., Kraft, C., Bhushan, S., Stepanenko, V., Tessmer, I., *Correcting for AFM tip induced topography convolutions in protein-DNA samples*. *Ultramicroscopy*, 2012. 121: S. 8-15.

165. Klingström, T., Bongcam-Rudloff, E., Pettersson, O.V., *A comprehensive model of DNA fragmentation for the preservation of High Molecular Weight DNA*. *BioRxiv*, 2018: DOI: 10.1101/254276. [02.01.2019]
166. Zhang, J., McCauley, M.J., Maher, L.J., 3rd, Williams, M.C., Israeloff, N.E., *Mechanism of DNA flexibility enhancement by HMGB proteins*. *Nucleic Acids Research*, 2009. 37(4): S. 1107-1114.
167. Rouzina, I., Bloomfield, V.A., *DNA bending by small, mobile multivalent cations*. *Biophysical Journal*, 1998. 74(6): S. 3152-3164.
168. Cherny, D.I., Jovin, T.M., *Electron and scanning force microscopy studies of alterations in supercoiled DNA tertiary structure*. *Journal of Biological Chemistry*, 2001. 313(2): S. 295-307.
169. Murugesapillai, D., McCauley, M.J., Maher, L.J., 3rd, Williams, M.C., *Single-molecule studies of high-mobility group B architectural DNA bending proteins*. *Biophysical Reviews*, 2017. 9(1): S. 17-40.
170. Rivetti, C., Codeluppi, S., *Accurate length determination of DNA molecules visualized by atomic force microscopy: evidence for a partial B- to A-form transition on mica*. *Ultramicroscopy*, 2001. 87(1): S. 55-66.
171. Hards, A., Zhou, C., Seitz, M., Bräuchle, C., Zumbusch, A., *Simultaneous AFM manipulation and fluorescence imaging of single DNA strands*. *ChemPhysChem*, 2005. 6(3): S. 534-540.
172. Yin, Y. et al., *Dynamics of spontaneous flipping of a mismatched base in DNA duplex*. *Proceedings of the National Academy of Sciences*, 2014. 111(22): S. 8043-8048.
173. Kundukad, B., Yan, J., Doyle, P.S., *Effect of YOYO-1 on the mechanical properties of DNA*. *Soft Matter*, 2014. 10(48): S. 9721-9728.
174. Milles, S., Lemke, E.A., *What precision-protein-tuning and nano-resolved single molecule sciences can do for each other*. *BioEssays*, 2013. 35(1): S. 65-74.
175. Lee, J., Kim, Y., Lim, S., Jo, K., *Single-molecule visualization of ROS-induced DNA damage in large DNA molecules*. *Analyst*, 2016. 141(3): S. 847-852.
176. Dame, R.T. et al., *Analysis of scanning force microscopy images of protein-induced DNA bending using simulations*. *Nucleic Acids Research*, 2005. 33(7): DOI: 10.1093/nar/gni073. [15.02.2019]
177. Brunet, A., Tardin, C., Salomé, L., Rousseau, P., Destainville, N., Manghi, M., *Dependence of DNA Persistence Length on Ionic Strength of Solutions with Monovalent and Divalent Salts: A Joint Theory–Experiment Study*. *Macromolecules*, 2015. 48(11): S. 3641-3652.
178. Podestà, A. et al., *Positively charged surfaces increase the flexibility of DNA*. *Biophysical Journal*, 2005. 89(4): S. 2558-2563.
179. Manning, G.S., *The persistence length of DNA is reached from the persistence length of its null isomer through an internal electrostatic stretching force*. *Biophysical Journal*, 2006. 91(10): S. 3607-3616.
180. Murugesapillai, D., Bouaziz, S., Maher, L.J., Israeloff, N.E., Cameron, C.E., Williams, M.C., *Accurate nanoscale flexibility measurement of DNA and DNA-protein complexes by atomic force microscopy in liquid*. *Nanoscale*, 2017. 9(31): S. 11327-11337.
181. Pastré, D. et al., *Adsorption of DNA to mica mediated by divalent counterions: a theoretical and experimental study*. *Biophysical Journal*, 2003. 85(4): S. 2507-2518.
182. McFail-Isom, L., Shui, X., Williams, L.D., *Divalent cations stabilize unstacked conformations of DNA and RNA by interacting with base pi systems*. *Biochemistry*, 1998. 37(49): S. 17105-17111.

183. Chiu, T.K., Dickerson, R.E., *1 Å crystal structures of B-DNA reveal sequence-specific binding and groove-specific bending of DNA by magnesium and calcium.* Journal of Molecular Biology, 2000. 301(4): S. 915-945.
184. Mirzabekov, A.D., Rich, A., *Asymmetric lateral distribution of unshielded phosphate groups in nucleosomal DNA and its role in DNA bending.* Proceedings of the National Academy of Sciences, 1979. 76(3): S. 1118-1121.
185. Faas, F.G.A., Rieger, B., van Vliet, L.J., Cherny, D.I., *DNA deformations near charged surfaces: electron and atomic force microscopy views.* Biophysical Journal, 2009. 97(4): S. 1148-1157.
186. Beckwitt, E.C., Kong, M., Van Houten, B., *Studying protein-DNA interactions using atomic force microscopy.* Seminars in Cell & Developmental Biology, 2018. 73: S. 220-230.
187. Brown, T., Leonard, G.A., Booth, E.D., Chambers, J., *Crystal structure and stability of a DNA duplex containing A(anti) · G(syn) base-pairs.* Journal of Molecular Biology, 1989. 207(2): S. 455-457.
188. Goodsell, D.S., Dickerson, R.E., *Bending and curvature calculations in B-DNA.* Nucleic Acids Research, 1994. 22(24): S. 5497-5503.
189. Janicijevic, A. et al., *DNA bending by the human damage recognition complex XPC-HR23B.* DNA Repair, 2003. 2(3): S. 325-336.
190. Woźniak, A.K., Schröder, G.F., Grubmüller, H., Seidel, C.A.M., Oesterhelt, F., *Single-molecule FRET measures bends and kinks in DNA.* Proceedings of the National Academy of Sciences, 2008. 105(47): S. 18337-18342.
191. Fratini, A.V., Kopka, M.L., Drew, H.R., Dickerson, R.E., *Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG.* Journal of Biological Chemistry, 1982. 257(24): S. 14686-14707.
192. Hizver, J., Rozenberg, H., Frolow, F., Rabinovich, D., Shakked, Z., *DNA bending by an adenine--thymine tract and its role in gene regulation.* Proceedings of the National Academy of Sciences of the United States of America, 2001. 98(15): S. 8490-8495.
193. Thompson, J.F., Landy, A., *Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes.* Nucleic Acids Research, 1988. 16(20): S. 9687-9705.
194. Marini, J.C., Levene, S.D., Crothers, D.M., Englund, P.T., *Bent helical structure in kinetoplast DNA.* Proceedings of the National Academy of Sciences of the United States of America, 1982. 79(24): S. 7664-7668.
195. Stellwagen, N.C., *Anomalous electrophoresis of deoxyribonucleic acid restriction fragments on polyacrylamide gels.* Biochemistry, 1983. 22(26): S. 6186-6193.
196. Stellwagen, N.C., *Electrophoresis of DNA in agarose gels, polyacrylamide gels and in free solution.* Electrophoresis, 2009. 30 Suppl 1(Suppl 1): S. S188-195.
197. Garcia, R.A., Bustamante, C.J., Reich, N.O., *Sequence-specific recognition of cytosine C5 and adenine N6 DNA methyltransferases requires different deformations of DNA.* Proceedings of the National Academy of Sciences of the United States of America, 1996. 93(15): S. 7618-7622.
198. Muzard, G., Theveny, B., Revet, B., *Electron microscopy mapping of pBR322 DNA curvature. Comparison with theoretical models.* The EMBO Journal, 1990. 9(4): S. 1289-1298.
199. Frontali, C. et al., *An absolute method for the determination of the persistence length of native DNA from electron micrographs.* Biopolymers, 1979. 18(6): S. 1353-1373.
200. Ojha, R.P. et al., *DNA bending and sequence-dependent backbone conformation.* European Journal of Biochemistry, 1999. 265(1): S. 35-53.

201. Van de Ven, F.J., Hilbers, C.W., *Nucleic acids and nuclear magnetic resonance*. European Journal of Biochemistry, 1988. 178(1): S. 1-38.
202. Bowman, B.R., Lee, S., Wang, S., Verdine, G.L., *Structure of Escherichia coli AlkA in complex with undamaged DNA*. Journal of Biological Chemistry, 2010. 285(46): S. 35783-35791.
203. Young, M.A., Ravishanker, G., Beveridge, D., Berman, H.M., *Analysis of local helix bending in crystal structures of DNA oligonucleotides and DNA-protein complexes*. Biophysical Journal, 1995. 68(6): S. 2454-2468.
204. Haran, T.E., Kahn, J.D., Crothers, D.M., *Sequence elements responsible for DNA curvature*. Journal of Molecular Biology, 1994. 244(2): S. 135-143.
205. Kahn, J.D., Crothers, D.M., *Measurement of the DNA bend angle induced by the catabolite activator protein using Monte Carlo simulation of cyclization kinetics*. Journal of Molecular Biology, 1998. 276(1): S. 287-309.
206. Shibutani, S., Bodepudi, V., Johnson, F., Grollman, A.P., *Translesional synthesis on DNA templates containing 8-oxo-7, 8-dihydrodeoxyadenosine*. Biochemistry, 1993. 32(17): S. 4615-4621.
207. Cheng, X., Kelso, C., Hornak, V., de los Santos, C., Grollman, A.P., Simmerling, C., *Dynamic Behavior of DNA Base Pairs Containing 8-Oxoguanine*. Journal of the American Chemical Society, 2005. 127(40): S. 13906-13918.
208. Patel, D.J., Kozlowski, S.A., Ikuta, S., Itakura, K., *Deoxyguanosine-deoxyadenosine pairing in the d (CGAGAAATTCGCG) duplex: conformation and dynamics at and adjacent to the dG. cntdot. dA mismatch site*. Biochemistry, 1984. 23(14): S. 3207-3217.
209. Sanchez, A.M., Volk, D.E., Gorenstein, D.G., Lloyd, R.S., *Initiation of repair of A/G mismatches is modulated by sequence context*. DNA Repair, 2003. 2(8): S. 863-878.
210. Hunter, W.N., Brown, T., Kennard, O., *Structural features and hydration of d (CGCGAATTAGCG); a double helix containing two GA mispairs*. Journal of Biomolecular Structure and Dynamics, 1986. 4(2): S. 173-191.
211. Leonard, G.A., Booth, E.D., Brown, T., *Structural and thermodynamic studies on the adenine.guanine mismatch in B-DNA*. Nucleic Acids Research, 1990. 18(19): S. 5617-5623.
212. Gao, X., Patel, D.J., *G(syn).cntdot.A(anti) mismatch formation in DNA dodecamers at acidic pH: pH-dependent conformational transition of G.cntdot.A mispairs detected by proton NMR*. Journal of the American Chemical Society, 1988. 110(15): S. 5178-5182.
213. Lipscomb, L.A. et al., *X-ray structure of a DNA decamer containing 7,8-dihydro-8-oxoguanine*. Proceedings of the National Academy of Sciences, 1995. 92(3): S. 719-723.
214. Oda, Y. et al., *NMR studies of a DNA containing 8-hydroxydeoxyguanosine*. Nucleic Acids Research, 1991. 19(7): S. 1407-1412.
215. Dršata, T.s., Kara, M., Zacharias, M., Lankaš, F., *Effect of 8-oxoguanine on DNA structure and deformability*. The Journal of Physical Chemistry B, 2013. 117(39): S. 11617-11622.
216. Reynisson, J., Steenken, S., *The calculated base pairing energy of 8-oxoguanine in the syn-anti conformation with cytosine, thymine, adenine and guanine*. Journal of Molecular Structure: THEOCHEM, 2005. 723(1-3): S. 29-36.
217. Reynisson, J., Steenken, S., *DNA-base radicals. Their base pairing abilities as calculated by DFT*. Physical Chemistry Chemical Physics, 2002. 4(21): S. 5346-5352.

218. Norman, D.P.G., Bruner, S.D., Verdine, G.L., *Coupling of Substrate Recognition and Catalysis by a Human Base-Excision DNA Repair Protein*. Journal of the American Chemical Society, 2001. 123(2): S. 359-360.
219. Leonard, G.A., McAuley-Hecht, K.E., Gibson, N.J., Brown, T., Watson, W.P., Hunter, W.N., *Guanine-1, N6-Ethenoadenine Base Pairs in the Crystal Structure of d (CGCGAATT (. epsilon. dA) GCG)*. Biochemistry, 1994. 33(16): S. 4755-4761.
220. Singer, B., Abbott, L., Spengler, S., *Assessment of mutagenic efficiency of two carcinogen-modified nucleosides, 1, N 6-ethenodeoxyadenosine and O 4-methyldeoxythymidine, using polymerases of varying fidelity*. Carcinogenesis, 1984. 5(9): S. 1165-1171.
221. Singer, B., Spengler, S., *Replication and transcription of polynucleotides containing ethencytosine, ethenoadenine and their hydrated intermediates*. IARC scientific publications, 1986(70): S. 359-371.
222. Kouchakdjian, M., Eisenberg, M., Yarema, K., Basu, A., Essigmann, J., Patel, D.J., *NMR studies of the exocyclic 1, N4-ethenodeoxyadenosine adduct (. epsilon. dA) opposite thymidine in a DNA duplex. Nonplanar alignment of epsilon. dA (anti) and dT (anti) at the lesion site*. Biochemistry, 1991. 30(7): S. 1820-1828.
223. Wyatt, M.D., Samson, L.D., *Influence of DNA structure on hypoxanthine and 1,N6-ethenoadenine removal by murine 3-methyladenine DNA glycosylase*. Carcinogenesis, 2000. 21(5): S. 901-908.
224. Sanchez, H., Wyman, C., *SFMetrics: an analysis tool for scanning force microscopy images of biomolecules*. BMC Bioinformatics, 2015. 16: S. 27.
225. Stivers, J.T., Pankiewicz, K.W., Watanabe, K.A., *Kinetic mechanism of damage site recognition and uracil flipping by Escherichia coli uracil DNA glycosylase*. Biochemistry, 1999. 38(3): S. 952-963.

# 8 Appendix

## 8.1 Bildvorbereitung

Nachfolgend wird eine detaillierte Anleitung zur Anwendung der hier etablierten automatisierten Messmethode gegeben. Die mit „→“ getrennten Punkte stellen die einzelnen Befehle im jeweiligen Programm dar. *Kursiv* sind konkrete Arbeitsanweisungen.

### 8.1.1 DNA

#### 1. ImageJ/FiJ

- Image → Type → RGB in 8 bit
- Process → Filter → Median → 2 px Median Filter

*Save Picture*

- File → Save As *xy median2px*
- Image → Adjust → Threshold → Yen → Auto →  Dark background

*Auto-Threshold ggf. anpassen und notieren*

#### 2. FIESTA

- Data → Open Stack → *Aufnahme\_xy median2px.tif laden*
- Time *equal*
- Pixel: *Pixel-Size (nm)*
- ○● → *Threshold eingeben aus ImageJ*
- zurück in Grauwertskala
- Tools → Line Scan → *rechts oben* → ●-● *anklicken* → *Linie durch DNA legen*
- *Histogramm Linie anklicken* → Estimate FWHM → *FWHM wird ausgegeben*
- Apply to configuration
- Check configuration
  - FWHM
  - Tracking:  Only Filaments
  - average before
  - Reference point: Center Point
  - Tracking Options
    - Track especially curved filaments



- reduce fit box size by: 25 %
- Apply → Add stack to local queue
  - Warnung! → Add anyway
  - *Auflistung in Local Queue*
- Analyse local queue


Nach Ende der Analyse werden die Ergebnisse in den Ursprungsordner als [.mat]-File gespeichert

- *Öffnen der Ergebnisse*
  - Data → Load Tracks
  - Data → Load Objects
  - *rechts: Filaments*
    - Show all Filaments
    - Show Filament Position
    - Ignore all Molecules
      - *durch Anklicken des Objekts oder des Filaments wird dieses in einem separaten Fenster geöffnet und alle Informationen angezeigt*
      - *zusätzlich kann jedes Filament markiert und auch gelöscht werden; Objekte können so nicht gelöscht werden!*

Nach der Auswahl der richtigen Filamente als *.selected.mat* abspeichern

- Data → Save all Tracks → *.selected.mat*

### 3. MATLAB

- MatLab öffnen und den Ordner mit dem SegmentAnalysis-Programm öffnen
-  Browse for folder → *auswählen* → Select a new folder
- Doppelklick auf MatLabScript → Script mit Programmierung wird geladen *rechts befindet sich der Workspace*
- Run *klicken*
- open Fiesta output data: *FIESTA.mat* öffnen
- open Fiesta output data: *FIESTA selected.mat* öffnen
- DNA length: Längenhistogramm wird erstellt und ausgegeben
  - Expected Length: *z.B. 150 nm*
- *Aufnahme als 8bit.tiff öffnen*
  - Bild mit Filamenten in den zuvor gewählten Programm-Ordner unter dem Zusatz *filtered filaments* speichern.



- Wahl der Curvature Query Points (2,4,10,15,20,50)  
⇒ Auswertung startet

Histogramm mit Winkelverteilung für jede Segmentlänge wird erstellt und automatisch gespeichert

Daten der Auswertung werden vorerst in den Programm-Ordner gespeichert.

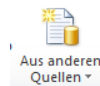
*Einfügen der Daten in Origin oder Excel*

#### 4. Origin

- *Import-Assistent* 
- *Datei auswählen und Tabelle erstellen*
- *zusätzliche Spalte einfügen* 
- Col (180 – Col(X))

#### 5. Excel

- Daten → Aus anderen Quellen
- Von XML-Dateien importieren
- Datenquelle auswählen



### 8.1.2 DNA & Protein

#### 1. ImageJ/ FiJ

##### A: DNA

- Image → Type → RGB in 8 bit

*Save Picture*

- File → Save As *8bit.tif*
- Image → Adjust → Threshold → Yen → Auto →  Dark background

*Auto-Threshold ggf. anpassen und notieren*

- Apply Threshold

*Save Picture*

- File → Save As *8bityen\_thersold.tif*
  - Plugins → Shape Filter
    - Perimeter: 90-Infinity
    - Add to manager
    - Black Background
    - Fill Results Table
    - Draw Holes
    - Show Labeled Image
    - Exclude on Edges
    - *Ausgabe des "gefilterten Bildes"*

Save Picture: *AG\_MutY\_160603001cut2048x2048\_shape\_fliter.tif*
  - Edit → Invert
  - Process → Filters → Gaussian blur → 2px
- Save Picture: *AG\_MutY\_160603001cut2048x2048\_shape\_fliter\_invert\_gauss.tif*

## B: Protein

- Image → Type → RGB in 8 bit
- Analyse → Set Scale → *Bildgröße in Pixel und nm eintragen, Einheit definieren*
- Image → Adjust → Threshold → Intermodes → Auto →  Dark background

*Auto-Threshold ggf. anpassen und notieren*

- Apply Threshold

*Save Picture*

- File → Save As *8bit\_intermodes.tif*
- Analyse → 3 D object counter
  - Threshold *wählen* → *zuvor ermittelten Threshold eintragen*
  - Size filter Min  Max
  - *Muss jeweils entsprechend der Molekülgröße, bzw. der zu vermessenden Partikel für jedes Bild neu angepasst werden. Bildqualität und Darstellungshöhe variieren*
  - Exclude objects on edges
  - Maps:  Objects

- Centroids
- Center of mass
- Statistics
- Summary

*Maps werden je in separatem Fenster ausgegeben, jedes Fenster separat abspeichern*

Statistics-Daten aus 3D object counter-Statistics unter *Statistics for MutY* speichern

## 2. FIESTA

- Data → Open Stack → *Aufnahme\_xymedian2px.tif* laden
- Time *equal*
- Pixel: *Pixel-Size (nm)*
- ○● → *Thershold eingeben aus ImageJ*
- zurück in Grauwertskala
- Tools → Line Scan → *rechts oben* → ●-● *anklicken* → *Linie durch DNA legen*
- ***Histogramm Linie anklicken*** → Estimate FWHM → ***FWHM wird ausgegeben***
- Apply to configuration
- Check configuration
  - FWHM
  - Tracking:  Only Filaments
  - average before
  - Reference point: Center Point
  - Tracking Options
    - Track especially curved flimanets
    - reduce fit box size by: **25%**
- Apply → Add stack to local queue
  - Warnung! → Add anyway
  - *Auflistung in Local Queue*
- **Analyse local queue**


Nach Ende der Analyse werden die Ergebnisse in den Ursprungsordner als .mat – File gespeichert

- *Öffnen der Ergebnisse*
  - Data → Load Tracks
  - Data → Load Objects
  - *rechts: Filaments*
    - Show all Filaments
    - Show Filament Position
    - Ignore all Molecules
      - *durch Anklicken des Objekts oder des Filaments wird dieses in einem separaten Fenster geöffnet und alle Informationen angezeigt*
      - *zusätzlich kann jedes Filament durch markiert und auch gelöscht werden; Objekte können so nicht gelöscht werden!*

Nach der Auswahl der richtigen Filamente als *.selected.mat* abspeichern

- Data → Save all Tracks → *.selected.mat*

### 3. MATLAB

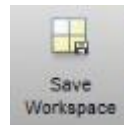
- MATLAB öffnen und den Ordner mit dem SegmentAnalysis-Programm öffnen
-  Browse for folder → *auswählen* → Select a new folder
- Doppelklick auf MATLABScript → Script mit Programmierung wird geladen  
*rechts befindet sich der Workspace*
- Run *klicken*
- open Fiesta output data: *FIESTA.mat* öffnen
- open Fiesta output data: *FIESTA selected.mat* öffnen
- DNA length: Längenhistogramm wird erstellt und ausgegeben
  - Expected Length: z.B. *150 nm*
  - Wahl der Curvature Query Points (2,4,10,15,20,50)  
⇒ Auswertung startet
- Laden der Proteinpositionen*
- Load → Statistics for *MutY...*  
*Auswahl des Min CutOff & Max CutOff*
  - Minimum Protein Area Cutoff (nm<sup>2</sup>): *1*
  - Maximum Protein Area Cutoff (nm<sup>2</sup>): *500*
- *Protein-Radius: 13 nm*
- *Aufnahme als 8bit.tiff öffnen*

Histogramm mit Winkelverteilung für jede Segmentlänge wird erstellt und automatisch gespeichert

- Save Figure:

Bild mit Filamenten und Proteinen in den zuvor gewählten Programm-Ordner speichern



- Ergebnistabelle (Results table) wird automatisch gespeichert unter *Statistics for UDGnsp505.workspace*
- Zusätzlich den kompletten Workspace speichern unter der Rubrik **Workspace**



Daten der Auswertung werden vorerst in den Programm-Ordner gespeichert.

*Einfügen der Daten in Origin oder Excel*

#### 4. Origin

- *Import-Assistent* 
- *Datei auswählen und Tabelle erstellen*
- *zusätzliche Spalte einfügen* 
- Col (180 – Col(X))

#### 5. Excel

- Daten → Aus anderen Quellen
- Von XML-Dateien importieren
- Datenquelle auswählen



## 8.2 Workflow-MATLAB

### 8.2.1 DNA

Aufgabenstellung:

- nur die ausgewählten, von FIESTA gut skelettierten, Filamente sollten ausgewertet werden
- die Filamente dürfen maximal  $\pm 20$  nm von der erwarteten Länge abweichen
- durch die Filamente werden gemäß eines Polynomzuges 0,1 nm lange Segmente gelegt
- Bestimmung des Krümmungswinkel bei 50 % der Länge des DNA-Strangs entsprechend dem Abfragepunktintervall(4, 6, 8 nm)

Parameter/ Eingaben

- Ergebnisse von FIESTA als .mat-Datei, inklusive den ausgewählten Filamenten (selected)
- erwartete Länge der DNA-Stränge
- Ausgangsbild als .tif-Datei
- Abfragepunkte: 4nm, 6 nm, 8 nm zum Mittelpunkt der DNA-Länge

Ergebnisse:

- .mat-Datei mit den Workspace-Variablen
- zwei Bilder mit den analysierten Filamenten als .fig- und .png-Datei
- Bilder mit den Winkel-Histogrammen als .fig- und .png-Datei, je nach gewählter Segmentlänge
- Ergebnisse als .txt-Datei  
(Spalte 1: Filament-ID; Spalte 2: Biegewinkel bei 4 nm; Spalte 3: Biegewinkel bei 6 nm; Spalte 4: Biegewinkel bei 8 nm; Spalte 5: Länge des Filaments in nm; Spalte 6: Fehler der Länge; Spalte 7: interpolierte Länge des Filaments in nm)

## 8.2.2 DNA & Protein

Aufgabenstellung:

- nur die ausgewählten, von FIESTA gut skelettierten Filamente sollten ausgewertet werden
- die Filamente dürfen maximal  $\pm 20$  nm von der erwarteten Länge abweichen
- durch die Filamente werden gemäß eines Polynomzuges 0,1 nm lange Segmente gelegt
- Laden und Vergleichen der Protein-Position mit der Lage der DNA - Stränge
- Bestimmung des Krümmungswinkel an Protein-Position auf dem DNA – Strang entsprechend dem Abfragepunktintervall (4, 6, 8 nm)

Parameter/ Eingaben

- Ergebnisse von FIESTA als .mat – Datei, inklusive den ausgewählten Filamenten (selected)
- erwartete Länge der DNA-Stränge
- Protein-Position aus 3 D object counter als .txt-Datei
- Ausgangsbild als .tif-Datei
- Abfragepunkte: 4 nm, 6 nm, 8 nm zur Protein-Position

Ergebnisse:

- .mat - Datei mit den workspace Variablen
- zwei Bilder mit den analysierten Filamenten als .fig- und .png-Datei
- Ergebnisse als .txt-Datei  
(Spalte 1: Filament-ID; Spalte 2: Protein-ID, Spalte 3: Länge des Filaments in nm; Spalte 4: interpolierte Länge des Filaments in nm; Spalte 5: Protein – Fläche in nm<sup>2</sup>, Spalte 6: Abstand Protein zu Filament in nm; Spalte 7: Biegewinkel bei 4 nm; Spalte 8: Biegewinkel bei 6 nm; Spalte 9: Biegewinkel bei 8 nm)



## 8.3 MATLAB-Programme

### DNA mit spezifischen Basenschaden an der 50 %-Position

Programm dient zur Auswertung von DNA-Strängen mit spezifischem Schaden an der 50 %-Position der Gesamtlänge

```
% Analysis of segmented DNA data
% 2017-11-27 Hannah Heil (V20171205)
% Input: Fiesta .mat file and -mat file with selected tracks, expected DNA length
in nm, .tif-Image,
% QueryPoint distance to center in nm

% Output:
% 1. .fig & .png-Image with filtered Filaments
% 2. .fig & .png-Image with angle-histogramms
% 3. Results as .txt (Column 1: Filament index, following columns:
% Curvature Angle at querypoints, last 3 columns: trackt filament lenght, error,
filament lenght after interpolation in nm)
% 4. workspace variables as .mat

% Short Summary:
% * Segmented filament data is filtered to only consider
% filaments with a deviation of less than two standarddeviations from the
% expected length.
% * The filtered filaments are segmented with equidistant
% segments by spline interpolation
% * Determine curvature angle in the QueryPoint interval by calculating the
% angle phil of the tangent one with x-axis via atan2, rotate 2nd tangent
% by -phil. This way angle of rotated 2nd tangent to x-axis is curvature
% angle.

%% Load file
clear all
close all

disp('Open Fiesta Output Data')
% Set path directory via standard GUI interface
[FileName, PathName] = uigetfile('*.mat','Fiesta Tracking Data');
fullpath = fullfile(PathName,FileName); %
Create full path adress
load(fullpath);
disp(FileName)
OrgFileName=fullpath;
clear fullpath FileName PathName;

disp('Open Selection Data')
% Set path directory via standard GUI interface
[FileName, PathName] = uigetfile('*.mat','Fiesta Selection Data'); %
fullpath = fullfile(PathName,FileName);
Create full path adress
load(fullpath);
disp(FileName)
clear fullpath FileName PathName;

% create array with selected filament index
SelectionArray=[];

for n=1: size(Filament,2)
    SelectionArray(1,n)=str2num(Filament(1,n).Name(10:end));
```

```

end
clear n

% Filter Selected Objects
ObjectsOriginalData=Objects{1,1}.data;
Objects{1,1}.data={}; % empty Objects.data
ObjectsOriginalBackground=Objects{1,1}.background;
Objects{1,1}.background=[]; % empty Objects.background
ObjectsOriginalCenter_x=Objects{1,1}.center_x;
Objects{1,1}.center_x=[]; % empty Objects.center_x
ObjectsOriginalCenter_y=Objects{1,1}.center_y;
Objects{1,1}.center_y=[]; % empty Objects.center_y
ObjectsOriginalCom_x=Objects{1,1}.com_x;
Objects{1,1}.com_x=[]; % empty Objects.com_x
ObjectsOriginalCom_y=Objects{1,1}.com_y;
Objects{1,1}.com_y=[]; % empty Objects.com_y
ObjectsOriginalHeight=Objects{1,1}.height;
Objects{1,1}.height=[]; % empty Objects.height
ObjectsOriginalLength=Objects{1,1}.length;
Objects{1,1}.length=[]; % empty Objects.length
ObjectsOriginalOrientation=Objects{1,1}.orientation;
Objects{1,1}.orientation=[]; % empty Objects.orientation
ObjectsOriginalPoints=Objects{1,1}.points;
Objects{1,1}.points={}; % empty Objects.points
ObjectsOriginalWidth=Objects{1,1}.width;
Objects{1,1}.width=[]; % empty Objects.width

for n=1:size(SelectionArray,2)
    Objects{1,1}.data{1,n}=ObjectsOriginalData{1,SelectionArray(1,n)};
    Objects{1,1}.background(:,n)=ObjectsOriginalBackground(:,SelectionAr-
ray(1,n));
    Objects{1,1}.center_x(:,n)=ObjectsOriginalCenter_x(:,SelectionArray(1,n));
    Objects{1,1}.center_y(:,n)=ObjectsOriginalCenter_y(:,SelectionArray(1,n));
    Objects{1,1}.com_x(:,n)=ObjectsOriginalCom_x(:,SelectionArray(1,n));
    Objects{1,1}.com_y(:,n)=ObjectsOriginalCom_y(:,SelectionArray(1,n));
    Objects{1,1}.height(:,n)=ObjectsOriginalHeight(:,SelectionArray(1,n));
    Objects{1,1}.length(:,n)=ObjectsOriginalLength(:,SelectionArray(1,n));
    Objects{1,1}.orientation(:,n)=ObjectsOriginalOrientation(:,SelectionAr-
ray(1,n));
    Objects{1,1}.points{1,n}=ObjectsOriginalPoints(:,SelectionArray(1,n));
    Objects{1,1}.width(:,n)=ObjectsOriginalWidth(:,SelectionArray(1,n));
end
clear n

%% Objects
%Filter Objects by length
%plot length histogram to select expected length
disp('Expected DNA length in nm')
prompt = {'Expected DNA length in nm'};
dlg_title = 'DNA length [nm]';
num_lines = 1;
def = {'150'};
info = inputdlg(prompt,dlg_title,num_lines,def);
expectedDNAlength=str2num(info{1});
CutoffLength=floor(expectedDNAlength*2/30)*10; % Lower length threshold
MaxiumDNALength=ceil(max(Objects{1, 1}.length(1,:))/10)*10;
LengthData=Objects{1, 1}.length(1,:);

%fit gaussian to length distribution of all length bigger than half of
%expected length and filter all objects with length deviation bigger than
%2xstd

%Histogram of Length Data
defbincenters=CutoffLength+5:10:MaxiumDNALength+5;

```

```

defbinedges=(linspace(CutoffLength,MaxiumDNALength,((MaxiumDNALength-Cut-
offLength)/10)+1)); % define bin edges, first entry is left edge of first bin,
last entry is right edge of last bin
[LegthHisto,ind] = histc(LengthData,defbinedges); % LengthHisto: number of
counts in the bins defined by defbinedges; ind: bin identifier for each
LengthData entry

GaussFitLengthdistribution=fit(defbincenters',LegthHisto','gauss1'); % Fit with
gaussian function
SigmaLength=GaussFitLengthdistribution.cl/sqrt(2);
MeanLength=GaussFitLengthdistribution.bl;
% Histogram with Gaussian Fit
figure(1)
clf
hold on
bar(defbinedges,LegthHisto,'histc')
plot(GaussFitLengthdistribution)

%Filter Objects with Mean-2std>L>Mean+2std

% qualityApprovedLength =( Objects{1, 1}.length(1,:) > MeanLength-(2*Sig-
maLength)) &...
% ( Objects{1, 1}.length(1,:) < MeanLength+(2*SigmaLength));
qualityApprovedLength =( Objects{1, 1}.length(1,:) > MeanLength-(20)) &...
( Objects{1, 1}.length(1,:) < MeanLength+(20));

% Apply Filter
LengthFiltered=[];
DataFiltered={};
LengthFiltered = Objects{1, 1}.length(:,qualityApprovedLength);
CenterFiltered(:,1)=Objects{1, 1}.center_x(:,qualityApprovedLength);
CenterFiltered(:,2)=Objects{1, 1}.center_y(:,qualityApprovedLength);
counter=1;
for n=1:size(LengthData,2)
    if qualityApprovedLength(n)==1
        DataFiltered{1,counter}=Objects{1, 1}.data{1,n};
        counter=counter+1;
    end
end
clear n
% Plot Objects on image that passed filter conditions
disp('Open Image')
% Set path directory via standard GUI interface
[FileName, PathName] = uigetfile('*.tif','Image');
fullpath = fullfile(PathName,FileName); %
Create full path adress
Image=imread(fullpath); % open source image
disp(FileName)

clear fullpath FileName PathName;

Xarray=1:1:size(Image,1); % Array with pixel index
Yarray=1:1:size(Image,2);
Xarray=Xarray*Config.PixSize; % Array with nm index
Yarray=Yarray*Config.PixSize;
C = {'b+','r+','g+','y+','m+','c+','m+','b+','r+','g+','y+','m+','
','c+','m+','b+','r+','g+','y+','m+','c+','m+','b+','r+','g+','y+','
','m+','c+','b+','r+','g+','y+','m+','c+','b+','r+','g+','y+','m+','
','c+','b+','r+','g+','y+','m+','c+','b+','r+','g+','y+','m+','c+','
','b+','r+','g+','y+','m+','c+','b+','r+','g+','y+','m+','c+','b+','
','r+','g+','y+','m+','c+','b+','r+','g+','y+','m+','c+','b+','r+','
','g+','y+','m+','c+','b+','r+','g+','y+','m+','c+','b+','r+','g+','
','y+','m+','c+','b+','r+','g+','y+','m+','c+','b+','r+','g+','y+','
','m+','c+','b+','r+','g+','y+','m+','c+','b+','r+','g+','y+','m+','
','c+','k+','b+','r+','g+','y+','m+','c+','b+','r+','g+','y+','m+'}

```



```

    plot(DataFiltered{1,n}(:,1),DataFiltered{1,n}(:,2),C{n},'Lin-
ewidth',2,'MarkerSize',0.5) % plot filament
    scatter(CenterFiltered(n,1),CenterFiltered(n,2),5,'bo') % draw circle at
center position
    textstr=[' ' num2str(n)];
    text(double(CenterFiltered(n,1)),double(CenterFiltered(n,2)),textstr) % add
index to each filament
end
disp('Enter filename to save figure') % ask for filename for image
prompt = {'Enter filename to save figure'};
dlg_title = 'Filename';
num_lines = 1;
def = {'FilteredFilaments'};
ImageFileNameDLG = inputdlg(prompt,dlg_title,num_lines,def);
ImageFileNameFig=[ImageFileNameDLG{1} '.fig'];
ImageFileNamepng=[ImageFileNameDLG{1} '.png'];
saveas(1,ImageFileNameFig)
saveas(1,ImageFileNamepng)

%% Curvature
% At which distances to the center should the curvature be evaluated?

disp('Curvature Query Points')
prompt = {'Curvature Query Points'};
dlg_title = 'Distance to center [nm]';
num_lines = 1;
def = {'2,4,10,15,20,50'};
info = inputdlg(prompt,dlg_title,num_lines,def);
QueryPoints=str2num(info{1});

% How many segments are needed to hit center and have eqidistant segments
% of ~0.1 nm length: NInterp
NInterp=round(((LengthFiltered(1,:)/0.1)-1)/2)*2+1;

% Interpolate
tic
NewLine={};
CenterPoint=[];
for n=1:size(LengthFiltered,2)
NewLine{1,n}=interparc(NInterp(n),DataFiltered{1,n}(:,1),Data-
Filtered{1,n}(:,2),'spline');
NewLine{1,n}(1,3)=0;
    NewLine{1,n}(1,4)=0;
    for m=2:size(NewLine{1,n},1)
        NewLine{1,n}(m,3)=sqrt((NewLine{1,n}(m,1)-NewLine{1,n}(m-1,1))^2+(New-
Line{1,n}(m,2)-NewLine{1,n}(m-1,2))^2); % actual length of each segment
        NewLine{1,n}(m,4)=NewLine{1,n}(m-1,4)+NewLine{1,n}(m,3); % actual total
length along path after spline interpolation
    end
    LengthFiltered(3,n)=NewLine{1,n}(end,4); % actual total path length after
spline interpolation
    CenterPoint(n)=ceil(size(NewLine{1,n},1)/2); % Index of CenterPoint of each
path
end
toc
clear n m

% Get angle between tangents at a distance from the center set as the Querypoints
alpha={};

for n=1:size(QueryPoints,2)
    DistSeg=QueryPoints(n)/0.1;
    for m=1:size(LengthFiltered,2)
        % Get angle phil to x axis of first tangent prior to center point

```

```

    alpha{1,n}(m,1)=(atan2((NewLine{1,m}(CenterPoint(m)-DistSeg+1,2)-New-
Line{1,m}(CenterPoint(m)-DistSeg-1,2)),(NewLine{1,m}(CenterPoint(m)-
DistSeg+1,1)-NewLine{1,m}(CenterPoint(m)-DistSeg-1,1))))); % atan2(y,x) of tan-
gent at querypoint before center position
    % Rotate 2nd tangent by -phil
    phil=alpha{1,n}(m,1);
    R=[cos(-phil) -sin(-phil) ;sin(-phil) cos(-phil)] ; % Rotation Matrix
    T2rotated=R*[(NewLine{1,m}(CenterPoint(m)+DistSeg-1,1)-New-
Line{1,m}(CenterPoint(m)+DistSeg+1,1));(NewLine{1,m}(CenterPoint(m)+DistSeg-
1,2)-NewLine{1,m}(CenterPoint(m)+DistSeg+1,2))];
    alpha{1,n}(m,2)=abs(atan2(T2rotated(2,1),T2rotated(1,1)))*180/pi; % result-
ing tangent cutting angle for querypoint in degree
end
clear n m

xbins=0:10:180; % define bin edges

for m=1:ceil(size(QueryPoints,2)/4)
    figure(m+1)
    clf
    for n=1:4
        if ((m-1)*4+n)<=size(QueryPoints,2)
            subplot(2,2,n)
            hist(alpha{1,((m-1)*4+n)}(:,2),xbins)
            title(['Curvature angle at ' num2str(QueryPoints(1,((m-1)*4+n))) ' nm to
center'])
            xlim([0 180])
            end
        end
        set(gcf,'units','normalized','outerposition',[0 0 1 1])
        PlotFileNameFig=['QuerypointHistogramm_' num2str(m) '.fig'];
        PlotFileNamepng=['QuerypointHistogramm_' num2str(m) '.png'];
        saveas(m+1,PlotFileNameFig) % save as .fig
        saveas(m+1,PlotFileNamepng) % save as .png
    end
clear n m
%% Summarize Results

n=size(QueryPoints,2);
m=size(LengthFiltered,2);
% write all results into one matrix: Filament index, curvature angle for QP
% 1, ..., last 3 columns: tracked filament length, error, filament length
% after interpolation
Results=zeros(m,n+1);
Results(:,1)=[1:1:m];

for p=1:n
    Results(:,1+p)=alpha{1,p}(:,2);
end
clear p
LengthFilterTransp=LengthFiltered';
for p=1:3 % add length info in last 3 collumns: tracked length, affiliated error,
length after interpolation
    Results(:,end+1)=LengthFilterTransp(:,p);
end
OutputFileName=[OrgFileName(1:end-4) '_SegmentAnalysisResultlts.txt'];
save(OutputFileName,'Results','-ascii') % save results table as txt file
OutputFileNameMat = [OrgFileName(1:end-4) '_Workspace'];
uisave({'Results','Config','DataFiltered','Filament','LengthFil-
tered','LengthData','Molecule','NewLine','Objects','QueryPoints'}, OutputFile-
NameMat); % Save in .mat file

```

## Ungeschädigte gekinkte DNA mit daran gebunden Proteinen

Nach Ermitteln der Längen-Verteilung, werden DNA-Stränge  $\pm 20$  nm der durchschnittlichen Länge erfasst; Mindestabstand 50 nm zwischen den einzelnen Protein-Peaks, stark gebogene/ gekinkte DNA-Stränge können über eine Maske unter Angabe ihrer Nummer selektiv verbunden werden. Allerdings können nur jeweils 2 Segmente verbunden werden

```
%% Analysis of segmented DNA data at protien postion
% 2017-11-27 Hannah Heil (V20171205) -> Analysis of segmented DNA data
% 2018-03-22 modified by Hannah Heil: Angle is measured at protein position
% 2018-04-03 commented, debuged version
% Input: Fiesta .mat file and -mat file with selected tracks, .txt file with
Protein positions expected DNA length in nm, .tif-Image,
% QueryPoint distance to protein position in nm

% Output:
% 1. .fig & .png-Image with analysed Filaments
% 2. Results as .txt (Columns: 1 Filament index, 2 protein index, 3 tracked
filament lenght [nm], 4 filament lenght after interpolation [nm],
% 5 Protein Area [ $\mu$ m], 6 usw Curvature Angle at querypoints
% 3. workspace variables as .mat

% Short Summary:
% * Filament segments of kinked tracks are combined to one
% * Segmented filament data is filtered to only consider
% filaments with a deviation of less than 20 nm from the
% expected length.
% * The filtered filaments are segmented with equidistant
% segments by spline interpolation
% * Protein Postions Filter: Upper and lower area limit, minimum distance
% of 50 nm in between proteins
% * Determine curvature angle in the QueryPoint interval by calculating the
% angle phil of the tangent one with x-axis via atan2, rotate 2nd tangent
% by -phil. This way angle of rotated 2nd tangent to x-axis is curvature
% angle.

%% Load files
clear all
close all

disp('Open Fiesta Output Data')
% Set path directory via standard GUI interface
[FileName, PathName] = uigetfile('*.mat','Fiesta Tracking Data');
fullpath = fullfile(PathName,FileName);
Create full path adress
load(fullpath);
disp(FileName)
OrgFileName=fullpath;
clear fullpath FileName PathName;

disp('Open Selection Data')
% Set path directory via standard GUI interface
[FileName, PathName] = uigetfile('*.mat','Fiesta Selection Data');
fullpath = fullfile(PathName,FileName);
Create full path adress
load(fullpath);
disp(FileName)
clear fullpath FileName PathName;
```

```

% create array with selected filament index
SelectionArray=[];

for n=1: size(Filament,2)
    SelectionArray(1,n)=str2num(Filament(1,n).Name(10:end));
end
clear n

% Filter Selected Objects
ObjectsOriginalData=Objects{1,1}.data;
Objects{1,1}.data={}; % empty Objects.data
ObjectsOriginalBackground=Objects{1,1}.background;
Objects{1,1}.background=[]; % empty Objects.background
ObjectsOriginalCenter_x=Objects{1,1}.center_x;
Objects{1,1}.center_x=[]; % empty Objects.center_x
ObjectsOriginalCenter_y=Objects{1,1}.center_y;
Objects{1,1}.center_y=[]; % empty Objects.center_y
ObjectsOriginalCom_x=Objects{1,1}.com_x;
Objects{1,1}.com_x=[]; % empty Objects.com_x
ObjectsOriginalCom_y=Objects{1,1}.com_y;
Objects{1,1}.com_y=[]; % empty Objects.com_y
ObjectsOriginalHeight=Objects{1,1}.height;
Objects{1,1}.height=[]; % empty Objects.height
ObjectsOriginalLength=Objects{1,1}.length;
Objects{1,1}.length=[]; % empty Objects.length
ObjectsOriginalOrientation=Objects{1,1}.orientation;
Objects{1,1}.orientation=[]; % empty Objects.orientation
ObjectsOriginalPoints=Objects{1,1}.points;
Objects{1,1}.points={}; % empty Objects.points
ObjectsOriginalWidth=Objects{1,1}.width;
Objects{1,1}.width=[]; % empty Objects.width

for n=1:size(SelectionArray,2)
    Objects{1,1}.data{1,n}=ObjectsOriginalData{1,SelectionArray(1,n)};
    Objects{1,1}.background(:,n)=ObjectsOriginalBackground(:,SelectionAr-
ray(1,n));
    Objects{1,1}.center_x(:,n)=ObjectsOriginalCenter_x(:,SelectionArray(1,n));
    Objects{1,1}.center_y(:,n)=ObjectsOriginalCenter_y(:,SelectionArray(1,n));
    Objects{1,1}.com_x(:,n)=ObjectsOriginalCom_x(:,SelectionArray(1,n));
    Objects{1,1}.com_y(:,n)=ObjectsOriginalCom_y(:,SelectionArray(1,n));
    Objects{1,1}.height(:,n)=ObjectsOriginalHeight(:,SelectionArray(1,n));
    Objects{1,1}.length(:,n)=ObjectsOriginalLength(:,SelectionArray(1,n));
    Objects{1,1}.orientation(:,n)=ObjectsOriginalOrientation(:,SelectionAr-
ray(1,n));
    Objects{1,1}.points{1,n}=ObjectsOriginalPoints(:,SelectionArray(1,n));
    Objects{1,1}.width(:,n)=ObjectsOriginalWidth(:,SelectionArray(1,n));
end
clear n

%% Connect kinked tracks
% How many tracks should be connected?
disp('How many sets of tracks should be connected?')
prompt = {'How many sets of tracks should be connected?'};
dlg_title = 'How many sets of tracks should be connected?';
num_lines = 1;
def = {'3'};
info = inputdlg(prompt,dlg_title,num_lines,def);
NTracks=str2num(info{1});
SegID=[];

for n=1:NTracks
    disp('Track Segments')
    prompt = {'Track Segment IDs'};
    dlg_title = 'Track Segment IDs';

```



```

        num_lines = 1;
        def = {'1,2'};
        info = inputdlg(prompt,dlg_title,num_lines,def);
        SegID(:,n)=str2num(info{1})';
end
clear n
% Find Ends
for n=1:NTracks
    a=find(SelectionArray==SegID(1,n));
    b=find(SelectionArray==SegID(2,n));
    Start1=Objects{1, 1}.data{1,a}(1,1:2);
    End1=Objects{1, 1}.data{1,a}(end,1:2);
    Start2=Objects{1, 1}.data{1,b}(1,1:2);
    End2=Objects{1, 1}.data{1,b}(end,1:2);
    EndDist(1,1)=sqrt((Start1(1,1)-Start2(1,1)).^2+(Start1(1,2)-
Start2(1,2)).^2);
    EndDist(1,2)=sqrt((Start1(1,1)-End2(1,1)).^2+(Start1(1,2)-End2(1,2)).^2);
    EndDist(1,3)=sqrt((Start2(1,1)-End1(1,1)).^2+(Start2(1,2)-End1(1,2)).^2);
    EndDist(1,4)=sqrt((End1(1,1)-End2(1,1)).^2+(End1(1,2)-End2(1,2)).^2);
    c=find(EndDist(1,:)==min(EndDist(1,:)));
    ASeg=Objects{1, 1}.data{1,a};
    BSeg=Objects{1, 1}.data{1,b};
    if c==1
        Objects{1, 1}.data{1,a}=cat(1,flipud(ASeg),BSeg);
    elseif c==2
        Objects{1, 1}.data{1,a}=cat(1,flipud(ASeg),flipud(BSeg));
    elseif c==3
        Objects{1, 1}.data{1,a}=cat(1,ASeg,BSeg);
    elseif c==4
        Objects{1, 1}.data{1,a}=cat(1,ASeg,flipud(BSeg));
    end
    % Full Length
    Objects{1, 1}.data{1,a}(1,3)=0;
    for m=2:size(Objects{1, 1}.data{1,a},1)
        Objects{1, 1}.data{1,a}(m,3)=Objects{1, 1}.data{1,a}(m-1,3)+sqrt((Ob-
jects{1, 1}.data{1,a}(m,1)-Objects{1, 1}.data{1,a}(m-1,1)).^2+(Objects{1,
1}.data{1,a}(m,2)-Objects{1,1}.data{1,a}(m-1,2)).^2);
    end
    Objects{1, 1}.length(1,a)=max(Objects{1, 1}.data{1,a}(m,3));
end

%% Filament Objects
%Filter Objects by length
%plot length histogram to select expected length
disp('Expected DNA length in nm')
prompt = {'Expected DNA length in nm'};
dlg_title = 'DNA lenght [nm]';
num_lines = 1;
def = {'150'};
info = inputdlg(prompt,dlg_title,num_lines,def);
expectedDNAlength=str2num(info{1});
CutoffLength=floor(expectedDNAlength*2/30)*10; % Lower length threshold
MaxiumDNALength=ceil(max(Objects{1, 1}.length(1,:))/10)*10;
LengthData=Objects{1, 1}.length(1,:);

%fit gaussian to length distribution of all length bigger than half of
%expected length and filter all objects with length deviation bigger than
%2xstd

%Histogram of Length Data
defbincenters=CutoffLength+5:10:MaxiumDNALength+5;
defbinedges=(linspace(CutoffLength,MaxiumDNALength,((MaxiumDNALength-Cut-
offLength)/10)+1)); % define bin edges, first entry is left edge of first bin,
last entry is right edge of last bin

```

```

[LegthHisto,ind] = histc(LengthData,defbinedges); % LengthHisto: number of
counts in the bins defined by defbinedges; ind: bin identifier for each
LengthData entry

GaussFitLengthdistribution=fit(defbincenters',LegthHisto','gauss1'); % Fit with
gaussian function
SigmaLength=GaussFitLengthdistribution.cl/sqrt(2);
MeanLength=GaussFitLengthdistribution.bl;
% Histogram with Gaussian Fit
figure(1)
clf
hold on
bar(defbinedges,LegthHisto,'histc')
plot(GaussFitLengthdistribution)

%Filter Objects with Mean-2std>L>Mean+2std

% qualityApprovedLength =( Objects{1, 1}.length(1,:) > MeanLength-(2*Sig-
maLength)) &...
% ( Objects{1, 1}.length(1,:) < MeanLength+(2*SigmaLength));
qualityApprovedLength =( Objects{1, 1}.length(1,:) > MeanLength-(50)) &...
( Objects{1, 1}.length(1,:) < MeanLength+(50));

%Ensure that parts of conected filament segments are filtered out
for n=1:NTracks
    b=find(SelectionArray==SegID(2,n));
    qualityApprovedLength(1,b)=0;
end

% Apply Filter
FilamentIndexFiltered=[];
LengthFiltered=[];
DataFiltered={}; % X[px] Y[px] distance[px] width[px] height[counts] back-
ground[counts]
LengthFiltered = Objects{1, 1}.length(:,qualityApprovedLength);
CenterFiltered(:,1)=Objects{1, 1}.center_x(:,qualityApprovedLength);
CenterFiltered(:,2)=Objects{1, 1}.center_y(:,qualityApprovedLength);
FilamentIndexFiltered(:,1)=SelectionArray(1,qualityApprovedLength);
counter=1;
for n=1:size(LengthData,2)
    if qualityApprovedLength(n)==1
        DataFiltered{1,counter}=Objects{1, 1}.data{1,n};
        counter=counter+1;
    end
end
clear n

%% Curvature
% At which distances to the Protein Position should the curvature be evaluated?

disp('Curvature Query Points')
prompt = {'Curvature Query Points'};
dlg_title = 'Distance to center [nm]';
num_lines = 1;
def = {'2,4,10,15,20,50'};
info = inputdlg(prompt,dlg_title,num_lines,def);
QueryPoints=str2num(info{1});

% How many segments are needed to hit center and have eqidistant segments
% of ~0.1 nm length: NInterp
NInterp=round(((LengthFiltered(1,:)/0.1)-1)/2)*2+1;

% Interpolate
tic
NewLine={};

```

```

CenterPoint=[];
for n=1:size(LengthFiltered,2)
NewLine{1,n}=interparc(NInterp(n),DataFiltered{1,n}(:,1),Data-
Filtered{1,n}(:,2),'spline');
NewLine{1,n}(1,3)=0;
    NewLine{1,n}(1,4)=0;
    for m=2:size(NewLine{1,n},1)
        NewLine{1,n}(m,3)=sqrt((NewLine{1,n}(m,1)-NewLine{1,n}(m-1,1))^2+(New-
Line{1,n}(m,2)-NewLine{1,n}(m-1,2))^2);% actual length of each segment
        NewLine{1,n}(m,4)=NewLine{1,n}(m-1,4)+NewLine{1,n}(m,3); % actual total
length along path after spline interpolation
    end
    LengthFiltered(3,n)=NewLine{1,n}(end,4); % actual total path length after
spline interpolation
    CenterPoint(n)=ceil(size(NewLine{1,n},1)/2); % Index of CenterPoint of each
path
end
toc
clear n m

% Load protein positions

disp('Open Protein Positions')
% Set path directory via standard GUI interface
[FileName, PathName] = uigetfile('*.txt','Fiesta Tracking Data');
fullpath = fullfile(PathName,FileName); %
Create full path adress %
ProtDataRaw=dlmread(fullpath,'\t',1,0); %
read data from tab-delimited txt file excluding first line %
disp(FileName) %
matrix-columns: Volume ( $\mu\text{m}^3$ ) Surface ( $\mu\text{m}^2$ ) Nb of obj. voxels Nb of surf.
voxels IntDen Mean StdDev Median Min Max X Y Z Mean dist. to surf.
(nm) SD dist. to surf. (nm) Median dist. to surf. (nm) XM YM ZM BX BY
BZ B-width B-height B-depth
OrgFileName=fullpath;
clear fullpath FileName PathName;

ProtData=ProtDataRaw(:,[1 3 12 13]); % ProtData: Index Area[ $\mu\text{m}^2$ ] X[px] Y[px]

% Discard Objects exceeding Area Threshold
disp('Protein Area Min and Max Limits')
prompt = {'Minimum Protein Area Cutoff [ $\mu\text{m}^2$ '],'Maximum Protein Area Cutoff
[ $\mu\text{m}^2$ ']};
dlg_title = 'Protein Area Cutoff [ $\mu\text{m}^2$ ']';
num_lines = 1;
def = {'0.0000001','0.0001'};
info = inputdlg(prompt,dlg_title,num_lines,def);
MinAreaCutoff=str2num(info{1});
MaxAreaCutoff=str2num(info{2});

qualityApprovedArea =(ProtData(:,2) > MinAreaCutoff)&(ProtData(:,2) < MaxAr-
eaCutoff);
ProtDataFiltered=ProtData(qualityApprovedArea,:);
ProtDataFiltered(:,5)=ProtDataFiltered(:,3)*Config.PixSize; %Xpos [nm]
ProtDataFiltered(:,6)=ProtDataFiltered(:,4)*Config.PixSize; %Ypos [nm]

% Protein radius input in nm
disp('Protein radius [nm]')
prompt = {'Protein radius [nm]'};
dlg_title = 'Protein radius [nm]';
num_lines = 1;
def = {'5'};
info = inputdlg(prompt,dlg_title,num_lines,def);
ProteinRadius=str2num(info{1}); %Protein Radius [nm]

```

```

% Protein radius based on area in nm
% ProteinRadius=sqrt(ProtDataFiltered(:,2)/pi)*1000; %Protein Radius [nm]
ColocMatrix=cell(1,size(LengthFiltered,2)); % for each filament a cell element
contains ID of all colocalizing proteins
DistanceArray=cell(1,size(LengthFiltered,2)); % for each filament a cell ele-
ment contains distances of segment elements to protein position
%qualityApprovedColocProt=zeros(size(ProtDataFiltered,1),1); % Array to filter
proteins without filament colocalization
%qualityApprovedColocFila=zeros(size(LengthFiltered,2),1); % Array to filter
filaments without protein colocalization
CountColocalization=zeros(size(LengthFiltered,2),1); % Counts number of pro-
teins colocalizing with each filament

% Find Proteins colocated with filament
for n=1:size(LengthFiltered,2) % loop all relevant filament
    for m=1:size(ProtDataFiltered,1)
        DistanceArray{1,n}(:,m)=sqrt((NewLine{1,n}(:,1)-ProtData-
Filtered(m,5)).^2+(NewLine{1,n}(:,2)-ProtDataFiltered(m,6)).^2);
        if min(DistanceArray{1,n}(:,m))<=ProteinRadius;
            ColocMatrix{1,n}(1,end+1)=m;
            %qualityApprovedColocProt(m,1)=1;
            %qualityApprovedColocFila(n,1)=1;
            CountColocalization(n,1)=CountColocalization(n,1)+1;
        end
    end
end
clear n m

% Filter Proteins on DNA with neighbouring proteins closer than 50 nm
ProtNew=[];
ProtSort=[];

for n=1:size(LengthFiltered,2) % loop all relevant filament
    Prot=ColocMatrix{1,n};
    if (size(ColocMatrix{1,n},2)==2) && (sqrt((ProtDataFiltered(ColocMa-
trix{1,n}(1,1),5)-ProtDataFiltered(ColocMatrix{1,n}(1,2),5)).^2+(ProtData-
Filtered(ColocMatrix{1,n}(1,1),6)-ProtDataFiltered(ColocMa-
trix{1,n}(1,2),6)).^2)<=50)
        Prot=[];
    elseif size(ColocMatrix{1,n},2)>2
        v=1:1:size(ColocMatrix{1,n},2);
        P = nchoosek(v,2);
        for m=1:size(P,1)
            if sqrt((ProtDataFiltered(ColocMatrix{1,n}(1,P(m,1)),5)-ProtData-
Filtered(ColocMatrix{1,n}(1,P(m,2)),5)).^2+(ProtDataFiltered(ColocMa-
trix{1,n}(1,P(m,1)),6)-ProtDataFiltered(ColocMatrix{1,n}(1,P(m,2)),6)).^2)<50;
                ProtSort=[ProtSort ColocMatrix{1,n}(1,P(m,1)) ColocMa-
trix{1,n}(1,P(m,2))];
            % ProtNew(1,end+1)=ColocMatrix{1,n}(1,P(m,1));
            % ProtNew(1,end+2)=ColocMatrix{1,n}(1,P(m,2));
        end
    end
    qualitySort=ones(1,size(ColocMatrix{1,n},2));
    for p=1:size(ProtSort,2)
        A=find(ColocMatrix{1,n}(1,:)-ProtSort(1,p)==0);
        qualitySort(1,A)=0;
    end
    qualitySort=logical(qualitySort);
    ColocMatrix{1,n}=ColocMatrix{1,n}(1,qualitySort);

% Prot=unique(ProtNew); % remove doublets
% Prot=Prot(any(Prot,1)); %remove zeros
end
end

```

```

clear A p m n

%qualityApprovedColocProt=logical(qualityApprovedColocProt); % convert into
logical values
%qualityApprovedColocFila=logical(qualityApprovedColocFila);

% Filament-Protein Distance Histogram for each Protein
MinDistanceArray=[]; % Array (row: filament, column: protein)
for n=1:size(LengthFiltered,2) % loop all relevant filament
    MinDistanceArray(n,:)=min(DistanceArray{1,n});
end
clear n
FilProtDistance=[]; % Row 1: Filament Protein Distance in nm, Row 2: Filament
ID, Row 3: Protein ID
for n=1:size(ProtDataFiltered,1) % loop all protein data
    b=find(MinDistanceArray(:,n)==min(MinDistanceArray(:,n)));
    FilProtDistance(1,n)=MinDistanceArray(b,n); % Filament Protein Distance in
nm
    FilProtDistance(2,n)=FilamentIndexFiltered(b,1); % filament ID
    FilProtDistance(3,n)=ProtDataFiltered(b,1); % Protein ID
    FilProtDistance(4,n)=ProtDataFiltered(b,2); % Protein Area  $\mu\text{m}^2$ 
end
clear n

% bins=0:0.5:100;
% [DistHisto,Distind] = histc(FilProtDistance(1,:),bins); % LengthHisto: number
of counts in the bins defined by defbinedges; ind: bin identifier for each
LengthData entry
%
% figure(1)
% clf
% bar(bins,DistHisto,'histc')

% scatter(ProtDataFiltered(:,5),ProtDataFiltered(:,6),5,'co') % draw circle at
center position

% Get angle between tangents at a distance from the center set as the Querypoints
alpha={};
Results=[];
h=1; % counter

for n=1:size(ColocMatrix,2) % loop through filaments
    if size(ColocMatrix{1,n},2)~=0 % if filament colocalizes with protein
        for m=1:size(ColocMatrix{1,n},2) % loop through colocalizing proteins
            DistSegMax=max(QueryPoints)/0.1;
            ProtPosSegm=find(DistanceArray{1,n}(:,ColocMa-
trix{1,n}(1,m))==min(DistanceArray{1,n}(:,ColocMatrix{1,n}(1,m))));
            if (ProtPosSegm-DistSegMax-1)>=1 && (ProtPosSegm+DistSeg-
Max+1<=size(NewLine{1,n},1)) % only include proteins with further away from DNA
Ends than biggest QueryPoint distance
                Results(h,1)=FilamentIndexFiltered(n,1); % filament ID
                Results(h,2)=ProtDataFiltered(ColocMatrix{1,n}(1,m),1); % Protein
ID
                Results(h,3)=LengthFiltered(1,n); % filament length fiesta segmen-
tation in nm
                Results(h,4)=LengthFiltered(3,n); % filament length fine segmenta-
tion in nm
                Results(h,5)=ProtDataFiltered(ColocMatrix{1,n}(1,m),2); % protien
area in  $\mu\text{m}^2$ 
                Results(h,6)=min(DistanceArray{1,n}(:,ColocMatrix{1,n}(1,m))); %
protein filament distance [nm]
                for k=1:size(QueryPoints,2)
                    DistSeg=QueryPoints(k)/0.1;

```

```

        alpha{1,n}(k,1)=(atan2((NewLine{1,n}(ProtPosSegm-
DistSeg+1,2)-NewLine{1,n}(ProtPosSegm-DistSeg-1,2)),(NewLine{1,n}(ProtPosSegm-
DistSeg+1,1)-NewLine{1,n}(ProtPosSegm-DistSeg-1,1)))); % atan2(y,x) of tangent
at querypoint before center position
        % Rotate 2nd tangent by -phil
        phil=alpha{1,n}(k,1);
        R=[cos(-phil) -sin(-phil) ;sin(-phil) cos(-phil)] ; % Rota-
tion Matrix
        T2rotated=R*[(NewLine{1,n}(ProtPosSegm+DistSeg-1,1)-New-
Line{1,n}(ProtPosSegm+DistSeg+1,1));(NewLine{1,n}(ProtPosSegm+DistSeg-1,2)-
NewLine{1,n}(ProtPosSegm+DistSeg+1,2))];
        Results(h,6+k)=abs(atan2(T2rotated(2,1),T2ro-
tated(1,1)))*180/pi; % resulting tangent cutting angle for querypoint in degree
    end
        h=h+1;
    end
end
end
end
end

clear n m k

% Plot Objects on image that were evaluated
disp('Open Image')
% Set path directory via standard GUI interface
[FileName, PathName] = uigetfile('*.tif','Image');
fullpath = fullfile(PathName,FileName); %
Create full path adress
Image=imread(fullpath); % open source image
disp(FileName)

clear fullpath FileName PathName;

Xarray=1:1:size(Image,1); % Array with pixel index
Yarray=1:1:size(Image,2);
Xarray=Xarray*Config.PixSize; % Array with nm index
Yarray=Yarray*Config.PixSize;
C = {'b+','r+','g+','y+','m+','c+','m+','b+','r+','g+','y+','m+','
','c+','m+','b+','r+','g+','y+','m+','c+','m+','b+','r+','g+','y+','
','m+','c+','b+','r+','g+','y+','m+','c+','b+','r+','g+','y+','m+','
','c+','b+','r+','g+','y+','m+','c+','b+','r+','g+','y+','m+','c+','
','b+','r+','g+','y+','m+','c+','b+','r+','g+','y+','m+','c+','b+','
','r+','g+','y+','m+','c+','b+','r+','g+','y+','m+','c+','b+','r+','
','g+','y+','m+','c+','b+','r+','g+','y+','m+','c+','b+','r+','g+','
','y+','m+','c+','b+','r+','g+','y+','m+','c+','b+','r+','g+','y+','
','m+','c+','b+','r+','g+','y+','m+','c+','b+','r+','g+','y+','m+','
','m+','c+','b+','r+','g+','y+','m+','c+','b+','r+','g+','y+','m+','
','b+','r+','g+','y+','m+','c+','k+','b+','r+','g+','y+','m+','c+','
','k+','b+','r+','g+','y+','m+','c+','k+','b+','r+','g+','y+','m+','
','c+','k+','b+','r+','g+','y+','m+','c+','k+','b+','r+','g+','y+','
','m+','c+','k+','b+','r+','g+','y+','m+','c+','k+','b+','r+','g+','
','y+','m+','c+','k+','b+','r+','g+','y+','m+','c+','k+','b+','r+','
','g+','y+','m+','c+','k+','b+','r+','g+','y+','m+','c+','k+','b+','
','r+','g+','y+','m+','c+','k+','b+','r+','g+','y+','m+','c+','k+','
','b+','r+','g+','y+','m+','c+','k+','b+','r+','g+','y+','m+','c+','
','k+','b+','r+','g+','y+','m+','c+','k+','b+','r+','g+','y+','m+','
','c+','k+','b+','r+','g+','y+','m+','c+','m+','k+','b+','r+','g+','
','y+','m+','c+','m+','k+','b+','r+','g+','y+','m+','c+','m+','k+','
','b+','r+','g+','y+','m+','c+','k+','b+','r+','g+','y+','m+','c+'}

```

```

', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-
', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-
', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-
', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-
', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-
', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-
', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-
', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-
', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-
', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-
', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-
', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-
', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-
', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-
', 'm+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'm+-', 'k+-', 'b+-', 'r+-', 'g+-
', 'y+-', 'm+-', 'c+-', 'm+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-
', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-
', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-
', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-
', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-
', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-
', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-
', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-
', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-
', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-
', 'm+-', 'c+-'];

```

figure(2)

```

clf
hold on
imagesc(Xarray,Yarray,Image) % draw image
colormap(flipud(gray)) % color with inverted gray scale
axis square % plot formate square
for n=1:size(ColocMatrix,2) % loop through filaments
    if size(ColocMatrix{1,n},2)~=0 % if filament colocalizes with protein
        plot(DataFiltered{1,n}(:,1),DataFiltered{1,n}(:,2),C{n},'Lin-
eWidth',2,'MarkerSize',0.5) % plot filament
        for m=1:size(1,n),2)
            scatter(ProtDataFiltered(ColocMatrix{1,n}(1,m),5),ProtDataFiltered(Col-
ocMatrix{1,n}(1,m),6),8,'co') % draw circle at center position
        end
        textstr=[' ' num2str(FilamentIndexFiltered(n,1))];
        text(double(CenterFiltered(n,1)),double(CenterFiltered(n,2)),textstr,'Col-
or','w') % add index to each filament
    end
end
disp('Enter filename to save figure') % ask for filename for image
prompt = {'Enter filename to save figure'};
dlg_title = 'Filename';
num_lines = 1;
def = {'FilteredFilaments'};
ImageFileNameDLG = inputdlg(prompt,dlg_title,num_lines,def);

```

```

ImageFileNameFig=[ImageFileNameDLG{1} '.fig'];
ImageFileNamepng=[ImageFileNameDLG{1} '.png'];
saveas(2,ImageFileNameFig)
saveas(2,ImageFileNamepng)

%% Export Results
OutputFileName=[OrgFileName(1:end-4) '_ProtPosDNAAngleAnalysisResults.txt'];
OutputFileName2=[OrgFileName(1:end-4) '_ProtPosDNAAngleAnalysisFilProtDis-
tance.txt'];
save(OutputFileName,'Results','-ascii') % save results table as txt file
save(OutputFileName2,'FilProtDistance','-ascii') % save FilProtDistance table
as txt file
OutputFileNameMat = [OrgFileName(1:end-4) '_Workspace'];
uisave({'Results','Config','DataFiltered','Filament','LengthFil-
tered','LengthData','Molecule','NewLine','Objects','QueryPoints','ProtData-
Filtered','FilProtDistance'}, OutputFileNameMat); % Save in .mat file

```



## Lange ungeschädigte DNA mit daran gebunden Proteinen

Für sehr lange DNA-Stränge, die auch starke Biegungen aufweisen, Mindestabstand 50 nm zwischen den einzelnen Protein-Peaks, keine Längensfilter

```
%% Analysis of segmented DNA data at protien postion
%geändert für MutYnspAC16051 am 16.04.2018
% 2017-11-27 Hannah Heil (V20171205) -> Analysis of segmented DNA data
% 2018-03-22 modified by Hannah Heil: Angle is measured at protein position
% 2018-04-03 commented, debuged version
% Input: Fiesta .mat file and -mat file with selected tracks, .txt file with
Protein positions expected DNA length in nm, .tif-Image,
% QueryPoint distance to protein position in nm

% Output:
% 1. .fig & .png-Image with analysed Filaments
% 2. Results as .txt (Columns: 1 Filament index, 2 protein index, 3 tracked
filament lenght [nm], 4 filament lenght after interpolation [nm],
% 5 Protein Area [ $\mu$ m], 6 usw Curvature Angle at querypoints
% 3. workspace variables as .mat

% Short Summary:
% * Filament segments of kinked tracks are combined to one
% * Segmented filament data is filtered to only consider
% filaments with a deviation of less than 20 nm from the
% expected length.
% * The filtered filaments are segmented with equidistant
% segments by spline interpolation
% * Protein Postions Filter: Upper and lower area limit, minimum distance
% of 50 nm in between proteins
% * Determine curvature angle in the QueryPoint interval by calculating the
% angle phil of the tangent one with x-axis via atan2, rotate 2nd tangent
% by -phil. This way angle of rotated 2nd tangent to x-axis is curvature
% angle.

%% Load files
clear all
close all

disp('Open Fiesta Output Data')
% Set path directory via standard GUI interface
[FileName, PathName] = uigetfile('*.mat','Fiesta Tracking Data');
fullpath = fullfile(PathName,FileName);
Create full path adress
load(fullpath);
disp(FileName)
OrgFileName=fullpath;
clear fullpath FileName PathName;

disp('Open Selection Data')
% Set path directory via standard GUI interface
[FileName, PathName] = uigetfile('*.mat','Fiesta Selection Data');
fullpath = fullfile(PathName,FileName);
Create full path adress
load(fullpath);
disp(FileName)
clear fullpath FileName PathName;

% create array with selected filament index
SelectionArray=[];

for n=1: size(Filament,2)
```

```

        SelectionArray(1,n)=str2num(Filament(1,n).Name(10:end));
end
clear n

% Filter Selected Objects
ObjectsOriginalData=Objects{1,1}.data;
Objects{1,1}.data={}; % empty Objects.data
ObjectsOriginalBackground=Objects{1,1}.background;
Objects{1,1}.background=[]; % empty Objects.background
ObjectsOriginalCenter_x=Objects{1,1}.center_x;
Objects{1,1}.center_x=[]; % empty Objects.center_x
ObjectsOriginalCenter_y=Objects{1,1}.center_y;
Objects{1,1}.center_y=[]; % empty Objects.center_y
ObjectsOriginalCom_x=Objects{1,1}.com_x;
Objects{1,1}.com_x=[]; % empty Objects.com_x
ObjectsOriginalCom_y=Objects{1,1}.com_y;
Objects{1,1}.com_y=[]; % empty Objects.com_y
ObjectsOriginalHeight=Objects{1,1}.height;
Objects{1,1}.height=[]; % empty Objects.height
ObjectsOriginalLength=Objects{1,1}.length;
Objects{1,1}.length=[]; % empty Objects.length
ObjectsOriginalOrientation=Objects{1,1}.orientation;
Objects{1,1}.orientation=[]; % empty Objects.orientation
ObjectsOriginalPoints=Objects{1,1}.points;
Objects{1,1}.points={}; % empty Objects.points
ObjectsOriginalWidth=Objects{1,1}.width;
Objects{1,1}.width=[]; % empty Objects.width

for n=1:size(SelectionArray,2)
    Objects{1,1}.data{1,n}=ObjectsOriginalData{1,SelectionArray(1,n)};
    Objects{1,1}.background(:,n)=ObjectsOriginalBackground(:,SelectionArray(1,n));
    Objects{1,1}.center_x(:,n)=ObjectsOriginalCenter_x(:,SelectionArray(1,n));
    Objects{1,1}.center_y(:,n)=ObjectsOriginalCenter_y(:,SelectionArray(1,n));
    Objects{1,1}.com_x(:,n)=ObjectsOriginalCom_x(:,SelectionArray(1,n));
    Objects{1,1}.com_y(:,n)=ObjectsOriginalCom_y(:,SelectionArray(1,n));
    Objects{1,1}.height(:,n)=ObjectsOriginalHeight(:,SelectionArray(1,n));
    Objects{1,1}.length(:,n)=ObjectsOriginalLength(:,SelectionArray(1,n));
    Objects{1,1}.orientation(:,n)=ObjectsOriginalOrientation(:,SelectionArray(1,n));
    Objects{1,1}.points{1,n}=ObjectsOriginalPoints(:,SelectionArray(1,n));
    Objects{1,1}.width(:,n)=ObjectsOriginalWidth(:,SelectionArray(1,n));
end
clear n

%% Connect kinked tracks
% How many tracks should be connected?
disp('How many sets of tracks should be connected?')
prompt = {'How many sets of tracks should be connected?'};
dlg_title = 'How many sets of tracks should be connected?';
num_lines = 1;
def = {'3'};
info = inputdlg(prompt,dlg_title,num_lines,def);
NTracks=str2num(info{1});
SegID=[];

for n=1:NTracks
    disp('Track Segments')
    prompt = {'Track Segment IDs'};
    dlg_title = 'Track Segment IDs';
    num_lines = 1;
    def = {'1,2'};
    info = inputdlg(prompt,dlg_title,num_lines,def);
    SegID(:,n)=str2num(info{1})';
end

```

```

clear n
% Find Ends
for n=1:NTracks
    a=find(SelectionArray==SegID(1,n));
    b=find(SelectionArray==SegID(2,n));
    Start1=Objects{1, 1}.data{1,a}(1,1:2);
    End1=Objects{1, 1}.data{1,a}(end,1:2);
    Start2=Objects{1, 1}.data{1,b}(1,1:2);
    End2=Objects{1, 1}.data{1,b}(end,1:2);
    EndDist(1,1)=sqrt((Start1(1,1)-Start2(1,1)).^2+(Start1(1,2)-
Start2(1,2)).^2);
    EndDist(1,2)=sqrt((Start1(1,1)-End2(1,1)).^2+(Start1(1,2)-End2(1,2)).^2);
    EndDist(1,3)=sqrt((Start2(1,1)-End1(1,1)).^2+(Start2(1,2)-End1(1,2)).^2);
    EndDist(1,4)=sqrt((End1(1,1)-End2(1,1)).^2+(End1(1,2)-End2(1,2)).^2);
    c=find(EndDist(1,:)==min(EndDist(1,:)));
    ASeg=Objects{1, 1}.data{1,a};
    BSeg=Objects{1, 1}.data{1,b};
    if c==1
        Objects{1, 1}.data{1,a}=cat(1,flipud(ASeg),BSeg);
    elseif c==2
        Objects{1, 1}.data{1,a}=cat(1,flipud(ASeg),flipud(BSeg));
    elseif c==3
        Objects{1, 1}.data{1,a}=cat(1,ASeg,BSeg);
    elseif c==4
        Objects{1, 1}.data{1,a}=cat(1,ASeg,flipud(BSeg));
    end
    % Full Length
    Objects{1, 1}.data{1,a}(1,3)=0;
    for m=2:size(Objects{1, 1}.data{1,a},1)
        Objects{1, 1}.data{1,a}(m,3)=Objects{1, 1}.data{1,a}(m-1,3)+sqrt((Ob-
jects{1, 1}.data{1,a}(m,1)-Objects{1, 1}.data{1,a}(m-1,1)).^2+(Objects{1,
1}.data{1,a}(m,2)-Objects{1,1}.data{1,a}(m-1,2)).^2);
    end
    Objects{1, 1}.length(1,a)=max(Objects{1, 1}.data{1,a}(m,3));
end

%% Filament Objects
%Filter Objects by length
%plot length histogram to select expected length
disp('Expected DNA length in nm')
prompt = {'Expected DNA length in nm'};
dlg_title = 'DNA lenght [nm]';
num_lines = 1;
def = {'300'};
info = inputdlg(prompt,dlg_title,num_lines,def);
expectedDNAlength=str2num(info{1});
CutoffLength=floor(expectedDNAlength*2/30)*10; % Lower length threshold
MaxiumDNAlength=ceil(max(Objects{1, 1}.length(1,:))/10)*10;
LengthData=Objects{1, 1}.length(1,:);

%fit gaussian to length distribution of all length bigger than half of
%expected length and filter all objects with length deviation bigger than
%2xstd

%Histogram of Length Data
defbincenters=CutoffLength+5:10:MaxiumDNAlength+5;
defbinedges=(linspace(CutoffLength,MaxiumDNAlength,((MaxiumDNAlength-Cut-
offLength)/10)+1)); % define bin edges, first entry is left edge of first bin,
last entry is right edge of last bin
[LegthHisto,ind] = histc(LengthData,defbinedges); % LengthHisto: number of
counts in the bins defined by defbinedges; ind: bin identifier for each
LengthData entry

GaussFitLengthdistribution=fit(defbincenters',LegthHisto','gauss1'); % Fit with
gaussian function

```

```

SigmaLength=GaussFitLengthdistribution.c1/sqrt(2);
MeanLength=GaussFitLengthdistribution.b1;
% Histogram with Gaussian Fit
figure(1)
clf
hold on
bar(defbinedges,LegthHisto,'histc')
plot(GaussFitLengthdistribution)

%Filter Objects with Mean-2std>L>Mean+2std

% qualityApprovedLength = ( Objects{1, 1}.length(1,:) > MeanLength-(2*SigmaLength)) &...
% ( Objects{1, 1}.length(1,:) < MeanLength+(2*SigmaLength));
qualityApprovedLength = ( Objects{1, 1}.length(1,:) > MeanLength-(250)) &...
( Objects{1, 1}.length(1,:) < MeanLength+(300));

%Ensure that parts of conected filament segments are filtered out
for n=1:NTracks
    b=find(SelectionArray==SegID(2,n));
    qualityApprovedLength(1,b)=0;
end

% Apply Filter
FilamentIndexFiltered=[];
LengthFiltered=[];
DataFiltered={}; % X[px] Y[px] distance[px] width[px] height[counts] back-
ground[counts]
LengthFiltered = Objects{1, 1}.length(:,qualityApprovedLength);
CenterFiltered(:,1)=Objects{1, 1}.center_x(:,qualityApprovedLength);
CenterFiltered(:,2)=Objects{1, 1}.center_y(:,qualityApprovedLength);
FilamentIndexFiltered(:,1)=SelectionArray(1,qualityApprovedLength);
counter=1;
for n=1:size(LengthData,2)
    if qualityApprovedLength(n)==1
        DataFiltered{1,counter}=Objects{1, 1}.data{1,n};
        counter=counter+1;
    end
end
clear n

%% Curvature
% At which distances to the Protein Position should the curvature be evaluated?

disp('Curvature Query Points')
prompt = {'Curvature Query Points'};
dlg_title = 'Distance to center [nm]';
num_lines = 1;
def = {'2,4,10,15,20,50'};
info = inputdlg(prompt,dlg_title,num_lines,def);
QueryPoints=str2num(info{1});

% How many segments are needed to hit center and have eqidistant segments
% of ~0.1 nm length: NInterp
NInterp=round(((LengthFiltered(1,:)/0.1)-1)/2)*2+1;

% Interpolate
tic
NewLine={};
CenterPoint=[];
for n=1:size(LengthFiltered,2)
    NewLine{1,n}=interparc(NInterp(n),DataFiltered{1,n}(:,1),Data-
Filtered{1,n}(:,2),'spline');
    NewLine{1,n}(1,3)=0;
    NewLine{1,n}(1,4)=0;
end

```

```

    for m=2:size(NewLine{1,n},1)
        NewLine{1,n}(m,3)=sqrt((NewLine{1,n}(m,1)-NewLine{1,n}(m-1,1))^2+(New-
Line{1,n}(m,2)-NewLine{1,n}(m-1,2))^2);% actual length of each segment
        NewLine{1,n}(m,4)=NewLine{1,n}(m-1,4)+NewLine{1,n}(m,3); % actual total
length along path after spline interpolation
    end
    LengthFiltered(3,n)=NewLine{1,n}(end,4); % actual total path length after
spline interpolation
    CenterPoint(n)=ceil(size(NewLine{1,n},1)/2); % Index of CenterPoint of each
path
end
toc
clear n m

% Load protein positions

disp('Open Protein Positions')
% Set path directory via standard GUI interface
[FileName, PathName] = uigetfile('*.txt','Fiesta Tracking Data');
fullpath = fullfile(PathName,FileName); %
Create full path adress %
ProtDataRaw=dlmread(fullpath,'\t',1,0); %
read data from tab-delimited txt file excluding first line %
disp(FileName) %
matrix-columns: Volume (µm^3) Surface (µm^2) Nb of obj. voxels Nb of surf.
voxels IntDen Mean StdDev Median Min Max X Y Z Mean dist. to surf.
(nm) SD dist. to surf. (nm) Median dist. to surf. (nm) XM YM ZM BX BY
BZ B-width B-height B-depth
OrgFileName=fullpath;
clear fullpath FileName PathName;

ProtData=ProtDataRaw(:,[1 3 12 13]); % ProtData: Index Area[µm^2] X[px] Y[px]

% Discard Objects exceeding Area Threshold
disp('Protein Area Min and Max Limits')
prompt = {'Minimum Protein Area Cutoff [µm^2]','Maximum Protein Area Cutoff
[µm^2]'};
dlg_title = 'Protein Area Cutoff [µm^2]';
num_lines = 1;
def = {'0.0000001','0.0001'};
info = inputdlg(prompt,dlg_title,num_lines,def);
MinAreaCutoff=str2num(info{1});
MaxAreaCutoff=str2num(info{2});

qualityApprovedArea =(ProtData(:,2) > MinAreaCutoff)&(ProtData(:,2) < MaxAr-
eaCutoff);
ProtDataFiltered=ProtData(qualityApprovedArea,:);
ProtDataFiltered(:,5)=ProtDataFiltered(:,3)*Config.PixSize; %Xpos [nm]
ProtDataFiltered(:,6)=ProtDataFiltered(:,4)*Config.PixSize; %Ypos [nm]

% Protein radius input in nm
disp('Protein radius [nm]')
prompt = {'Protein radius [nm]'};
dlg_title = 'Protein radius [nm]';
num_lines = 1;
def = {'5'};
info = inputdlg(prompt,dlg_title,num_lines,def);
ProteinRadius=str2num(info{1}); %Protein Radius [nm]

% Protein radius based on area in nm
% ProteinRadius=sqrt(ProtDataFiltered(:,2)/pi)*1000; %Protein Radius [nm]
ColocMatrix=cell(1,size(LengthFiltered,2)); % for each filament a cell element
contains ID of all colocalizing proteins
DistanceArray=cell(1,size(LengthFiltered,2)); % for each filament a cell ele-
ment contains distances of segment elements to protein position

```

```

%qualityApprovedColocProt=zeros(size(ProtDataFiltered,1),1); % Array to filter
proteins without filament colocalization
%qualityApprovedColocFila=zeros(size(LengthFiltered,2),1); % Array to filter
filaments without protein colocalization
CountColocalization=zeros(size(LengthFiltered,2),1); % Counts number of pro-
teins colocalizing with each filament

% Find Proteins colocated with filament
for n=1:size(LengthFiltered,2) % loop all relevant filament
    for m=1:size(ProtDataFiltered,1)
        DistanceArray{1,n}(:,m)=sqrt((NewLine{1,n}(:,1)-ProtData-
Filtered(m,5)).^2+(NewLine{1,n}(:,2)-ProtDataFiltered(m,6)).^2);
        if min(DistanceArray{1,n}(:,m))<=ProteinRadius;
            ColocMatrix{1,n}(1,end+1)=m;
            %qualityApprovedColocProt(m,1)=1;
            %qualityApprovedColocFila(n,1)=1;
            CountColocalization(n,1)=CountColocalization(n,1)+1;
        end
    end
end
clear n m

% Filter Proteins on DNA with neighbouring proteins closer than 50 nm
ProtNew=[];
ProtSort=[];

for n=1:size(LengthFiltered,2) % loop all relevant filament
    Prot=ColocMatrix{1,n};
    if (size(ColocMatrix{1,n},2)==2)&&(sqrt((ProtDataFiltered(ColocMa-
trix{1,n}(1,1),5)-ProtDataFiltered(ColocMatrix{1,n}(1,2),5)).^2+(ProtData-
Filtered(ColocMatrix{1,n}(1,1),6)-ProtDataFiltered(ColocMa-
trix{1,n}(1,2),6)).^2)<=50)
        Prot=[];
    elseif size(ColocMatrix{1,n},2)>2
        v=1:1:size(ColocMatrix{1,n},2);
        P = nchoosek(v,2);
        for m=1:size(P,1)
            if sqrt((ProtDataFiltered(ColocMatrix{1,n}(1,P(m,1)),5)-ProtData-
Filtered(ColocMatrix{1,n}(1,P(m,2)),5)).^2+(ProtDataFiltered(ColocMa-
trix{1,n}(1,P(m,1)),6)-ProtDataFiltered(ColocMatrix{1,n}(1,P(m,2)),6)).^2)<50;
                ProtSort=[ProtSort ColocMatrix{1,n}(1,P(m,1)) ColocMa-
trix{1,n}(1,P(m,2))];
            % ProtNew(1,end+1)=ColocMatrix{1,n}(1,P(m,1));
            % ProtNew(1,end+2)=ColocMatrix{1,n}(1,P(m,2));
        end
    end
    qualitySort=ones(1,size(ColocMatrix{1,n},2));
    for p=1:size(ProtSort,2)
        A=find(ColocMatrix{1,n}(1,:)-ProtSort(1,p)==0);
        qualitySort(1,A)=0;
    end
    qualitySort=logical(qualitySort);
    ColocMatrix{1,n}=ColocMatrix{1,n}(1,qualitySort);

% Prot=unique(ProtNew); % remove doublets
% Prot=Prot(any(Prot,1)); %remove zeros
end
end
clear A p m n

%qualityApprovedColocProt=logical(qualityApprovedColocProt); % convert into
logical values
%qualityApprovedColocFila=logical(qualityApprovedColocFila);

```

```

% Filament-Protein Distance Histogram for each Protein
MinDistanceArray=[]; % Array (row: filament, column: protein)
for n=1:size(LengthFiltered,2) % loop all relevant filament
    MinDistanceArray(n,:)=min(DistanceArray{1,n});
end
clear n
FilProtDistance=[]; % Row 1: Filament Protein Distance in nm, Row 2: Filament
ID, Row 3: Protein ID
for n=1:size(ProtDataFiltered,1) % loop all protein data
    b=find(MinDistanceArray(:,n)==min(MinDistanceArray(:,n)));
    FilProtDistance(1,n)=MinDistanceArray(b,n); % Filament Protein Distance in
nm
    FilProtDistance(2,n)=FilamentIndexFiltered(b,1); % filament ID
    FilProtDistance(3,n)=ProtDataFiltered(b,1); % Protein ID
    FilProtDistance(4,n)=ProtDataFiltered(b,2); % Protein Area  $\mu\text{m}^2$ 
end
clear n

% bins=0:0.5:100;
% [DistHisto,Distind] = histc(FilProtDistance(1,:),bins); % LengthHisto: number
of counts in the bins defined by defbinedges; ind: bin identifier for each
LengthData entry
%
% figure(1)
% clf
% bar(bins,DistHisto,'histc')

% scatter(ProtDataFiltered(:,5),ProtDataFiltered(:,6),5,'co') % draw circle at
center position

% Get angle between tangents at a distance from the center set as the Querypoints
alpha={};
Results=[];
h=1; % counter

for n=1:size(ColocMatrix,2) % loop through filaments
    if size(ColocMatrix{1,n},2)~=0 % if filament colocalizes with protein
        for m=1:size(ColocMatrix{1,n},2) % loop through colocalizing proteins
            DistSegMax=max(QueryPoints)/0.1;
            ProtPosSegm=find(DistanceArray{1,n}(:,ColocMa-
trix{1,n}(1,m))==min(DistanceArray{1,n}(:,ColocMatrix{1,n}(1,m))));
            if (ProtPosSegm-DistSegMax-1)>=1 && (ProtPosSegm+DistSeg-
Max+1<=size(NewLine{1,n},1)) % only include proteins with further away from DNA
Ends than biggest QueryPoint distance
                Results(h,1)=FilamentIndexFiltered(n,1); % filament ID
                Results(h,2)=ProtDataFiltered(ColocMatrix{1,n}(1,m),1); % Protein
ID
                Results(h,3)=LengthFiltered(1,n); % filament length fiesta segmen-
tation in nm
                Results(h,4)=LengthFiltered(3,n); % filament length fine segmenta-
tion in nm
                Results(h,5)=ProtDataFiltered(ColocMatrix{1,n}(1,m),2); % protien
area in  $\mu\text{m}^2$ 
                Results(h,6)=min(DistanceArray{1,n}(:,ColocMatrix{1,n}(1,m))); %
protein filament distance [nm]
                for k=1:size(QueryPoints,2)
                    DistSeg=QueryPoints(k)/0.1;
                    alpha{1,n}(k,1)=(atan2((NewLine{1,n}(ProtPosSegm-
DistSeg+1,2)-NewLine{1,n}(ProtPosSegm-DistSeg-1,2)),(NewLine{1,n}(ProtPosSegm-
DistSeg+1,1)-NewLine{1,n}(ProtPosSegm-DistSeg-1,1)))); % atan2(y,x) of tangent
at querypoint before center position
                    % Rotate 2nd tangent by -phil
                    phil=alpha{1,n}(k,1);

```

```

R=[cos(-phil) -sin(-phil) ;sin(-phil) cos(-phil)] ; % Rota-
tion Matrix
T2rotated=R*[ (NewLine{1,n} (ProtPosSegm+DistSeg-1,1)-New-
Line{1,n} (ProtPosSegm+DistSeg+1,1)); (NewLine{1,n} (ProtPosSegm+DistSeg-1,2)-
NewLine{1,n} (ProtPosSegm+DistSeg+1,2)) ];
Results(h,6+k)=abs (atan2 (T2rotated(2,1),T2ro-
tated(1,1)))*180/pi; % resulting tangent cutting angle for querypoint in degree
end
h=h+1;
end
end
end
end
clear n m k

% Plot Objects on image that were evaluated
disp('Open Image')
% Set path directory via standard GUI interface
[FileName, PathName] = uigetfile('*.tif','Image');
fullpath = fullfile(PathName,FileName); %
Create full path adress
Image=imread(fullpath); % open source image
disp(FileName)

clear fullpath FileName PathName;

Xarray=1:1:size(Image,1); % Array with pixel index
Yarray=1:1:size(Image,2);
Xarray=Xarray*Config.PixSize; % Array with nm index
Yarray=Yarray*Config.PixSize;
C = {'b+-','r+-','g+-','y+-','m+-','c+-','m+-','b+-','r+-','g+-','y+-','m+-',
'c+-','m+-','b+-','r+-','g+-','y+-','m+-','c+-','m+-','b+-','r+-','g+-','y+-',
'm+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-','y+-','m+-',
'c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-',
'b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-',
'r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-',
'g+-','y+-','m+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-',
'y+-','m+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-',
'y+-','m+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-',
'm+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-','y+-',
'm+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-','y+-',
'b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-',
'k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-',
'c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-',
'm+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-',
'y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-',
'b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-',
'k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-',
'c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-',
'm+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-',
'y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-',
'g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-',
'r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-'

```





```

def = {'FilteredFilaments'};
ImageFileNameDLG = inputdlg(prompt,dlg_title,num_lines,def);
ImageFileNameFig=[ImageFileNameDLG{1} '.fig'];
ImageFileNamepng=[ImageFileNameDLG{1} '.png'];
saveas(2,ImageFileNameFig)
saveas(2,ImageFileNamepng)

%% Export Results
OutputFileName=[OrgFileName(1:end-4) '_ProtPosDNAAngleAnalysisResults.txt'];
OutputFileName2=[OrgFileName(1:end-4) '_ProtPosDNAAngleAnalysisFilProtDis-
tance.txt'];
save(OutputFileName,'Results','-ascii') % save results table as txt file
save(OutputFileName2,'FilProtDistance','-ascii') % save FilProtDistance table
as txt file
OutputFileNameMat = [OrgFileName(1:end-4) '_Workspace'];
uisave({'Results','Config','DataFiltered','Filament','LengthFil-
tered','LengthData','Molecule','NewLine','Objects','QueryPoints','ProtData-
Filtered','FilProtDistance'}, OutputFileNameMat); % Save in .mat file

```

## Lange ungeschädigte DNA (speziell adaptiert)

Speziell für die Auswertung der bereits manuell gemessenen Aufnahmen von MutY-nsp DNA (2224 bp lang) angepasst. Es werden alle Filamente erfasst, keine Mindest- oder Maximallänge festgelegt, auch der Mindestabstand der Proteine beträgt nur 25 nm. Diese Anpassung war nötig, um möglichst alle Protein-Peaks, die manuell vermessen wurden, auch mit der automatisierten Messvariante erfassen zu können

```
%% Analysis of segmented DNA data at protien postion
%20180425 Kick out molecule objects from Fiesta Data
%geändert für MutYnspAC16051 am 16.04.2018
% 2017-11-27 Hannah Heil (V20171205) -> Analysis of segmented DNA data
% 2018-03-22 modified by Hannah Heil: Angle is measured at protein position
% 2018-04-03 commented, debuged version
% Input: Fiesta .mat file and -mat file with selected tracks, .txt file with
Protein positions expected DNA length in nm, .tif-Image,
% QueryPoint distance to protein position in nm

% Output:
% 1. .fig & .png-Image with analysed Filaments
% 2. Results as .txt (Columns: 1 Filament index, 2 protein index, 3 tracked
filament lenght [nm], 4 filament lenght after interpolation [nm],
% 5 Protein Area [µm], 6 usw Curvature Angle at querypoints
% 3. workspace variables as .mat

% Short Summary:
% * Filament segments of kinked tracks are combined to one
% * Segmented filament data is filtered to only consider
% filaments with a deviation of less than 20 nm from the
% expected length.
% * The filtered filaments are segmented with equidistant
% segments by spline interpolation
% * Protein Postions Filter: Upper and lower area limit, minimum distance
% of 50 nm in between proteins
% * Determine curvature angle in the QueryPoint interval by calculating the
% angle phil of the tangent one with x-axis via atan2, rotate 2nd tangent
% by -phil. This way angle of rotated 2nd tangent to x-axis is curvature
% angle.

%% Load files
clear all
close all

disp('Open Fiesta Output Data')
% Set path directory via standard GUI interface
[FileName, PathName] = uigetfile('*.*mat','Fiesta Tracking Data');
fullpath = fullfile(PathName,FileName);
Create full path adress
load(fullpath);
disp(FileName)
OrgFileName=fullpath;
clear fullpath FileName PathName;

disp('Open Selection Data')
% Set path directory via standard GUI interface
[FileName, PathName] = uigetfile('*.*mat','Fiesta Selection Data');
```

```

fullpath = fullfile(PathName,FileName); %
Create full path adress
load(fullpath);
disp(FileName)
clear fullpath FileName PathName;

% create array with selected filament index
SelectionArray=[];

for n=1: size(Filament,2)
    SelectionArray(1,n)=str2num(Filament(1,n).Name(10:end));
end
clear n

% Filter Molecules
h=1;
MoleculeFilter=[]; % Indix of Filament Objects

for n=1:size(Objects{1,1}.data,2)
    if length(Objects{1,1}.data{1,n}) >= 2
        MoleculeFilter(h)=n;
        h=h+1;
    end
end

% Filter Selected Objects
ObjectsOriginalData=Objects{1,1}.data;
Objects{1,1}.data={}; % empty Objects.data
ObjectsOriginalBackground=Objects{1,1}.background;
Objects{1,1}.background=[]; % empty Objects.background
ObjectsOriginalCenter_x=Objects{1,1}.center_x;
Objects{1,1}.center_x=[]; % empty Objects.center_x
ObjectsOriginalCenter_y=Objects{1,1}.center_y;
Objects{1,1}.center_y=[]; % empty Objects.center_y
ObjectsOriginalCom_x=Objects{1,1}.com_x;
Objects{1,1}.com_x=[]; % empty Objects.com_x
ObjectsOriginalCom_y=Objects{1,1}.com_y;
Objects{1,1}.com_y=[]; % empty Objects.com_y
ObjectsOriginalHeight=Objects{1,1}.height;
Objects{1,1}.height=[]; % empty Objects.height
ObjectsOriginalLength=Objects{1,1}.length;
Objects{1,1}.length=[]; % empty Objects.length
ObjectsOriginalOrientation=Objects{1,1}.orientation;
Objects{1,1}.orientation=[]; % empty Objects.orientation
ObjectsOriginalPoints=Objects{1,1}.points;
Objects{1,1}.points={}; % empty Objects.points
ObjectsOriginalWidth=Objects{1,1}.width;
Objects{1,1}.width=[]; % empty Objects.width

% Kick out Molecule Data
for n=1:size(MoleculeFilter,2)
    Objects{1,1}.data{1,n}=ObjectsOriginalData{1,MoleculeFilter(1,n)};
    Objects{1,1}.background(:,n)=ObjectsOriginalBackground(:,MoleculeFilter(1,n));
    Objects{1,1}.center_x(:,n)=ObjectsOriginalCenter_x(:,MoleculeFilter(1,n));
    Objects{1,1}.center_y(:,n)=ObjectsOriginalCenter_y(:,MoleculeFilter(1,n));
    Objects{1,1}.com_x(:,n)=ObjectsOriginalCom_x(:,MoleculeFilter(1,n));
    Objects{1,1}.com_y(:,n)=ObjectsOriginalCom_y(:,MoleculeFilter(1,n));
    Objects{1,1}.height(:,n)=ObjectsOriginalHeight(:,MoleculeFilter(1,n));
    Objects{1,1}.length(:,n)=ObjectsOriginalLength(:,MoleculeFilter(1,n));
    Objects{1,1}.orientation(:,n)=ObjectsOriginalOrientation(:,MoleculeFilter(1,n));
    Objects{1,1}.points{1,n}=ObjectsOriginalPoints(:,MoleculeFilter(1,n));
    Objects{1,1}.width(:,n)=ObjectsOriginalWidth(:,MoleculeFilter(1,n));
end

```

```

clear n

% Filter Selected Objects
ObjectsOriginalData=Objects{1,1}.data;
Objects{1,1}.data={}; % empty Objects.data
ObjectsOriginalBackground=Objects{1,1}.background;
Objects{1,1}.background=[]; % empty Objects.background
ObjectsOriginalCenter_x=Objects{1,1}.center_x;
Objects{1,1}.center_x=[]; % empty Objects.center_x
ObjectsOriginalCenter_y=Objects{1,1}.center_y;
Objects{1,1}.center_y=[]; % empty Objects.center_y
ObjectsOriginalCom_x=Objects{1,1}.com_x;
Objects{1,1}.com_x=[]; % empty Objects.com_x
ObjectsOriginalCom_y=Objects{1,1}.com_y;
Objects{1,1}.com_y=[]; % empty Objects.com_y
ObjectsOriginalHeight=Objects{1,1}.height;
Objects{1,1}.height=[]; % empty Objects.height
ObjectsOriginalLength=Objects{1,1}.length;
Objects{1,1}.length=[]; % empty Objects.length
ObjectsOriginalOrientation=Objects{1,1}.orientation;
Objects{1,1}.orientation=[]; % empty Objects.orientation
ObjectsOriginalPoints=Objects{1,1}.points;
Objects{1,1}.points={}; % empty Objects.points
ObjectsOriginalWidth=Objects{1,1}.width;
Objects{1,1}.width=[]; % empty Objects.width

for n=1:size(SelectionArray,2)
    Objects{1,1}.data{1,n}=ObjectsOriginalData{1,SelectionArray(1,n)};
    Objects{1,1}.background(:,n)=ObjectsOriginalBackground(:,SelectionArray(1,n));
    Objects{1,1}.center_x(:,n)=ObjectsOriginalCenter_x(:,SelectionArray(1,n));
    Objects{1,1}.center_y(:,n)=ObjectsOriginalCenter_y(:,SelectionArray(1,n));
    Objects{1,1}.com_x(:,n)=ObjectsOriginalCom_x(:,SelectionArray(1,n));
    Objects{1,1}.com_y(:,n)=ObjectsOriginalCom_y(:,SelectionArray(1,n));
    Objects{1,1}.height(:,n)=ObjectsOriginalHeight(:,SelectionArray(1,n));
    Objects{1,1}.length(:,n)=ObjectsOriginalLength(:,SelectionArray(1,n));
    Objects{1,1}.orientation(:,n)=ObjectsOriginalOrientation(:,SelectionArray(1,n));
    Objects{1,1}.points{1,n}=ObjectsOriginalPoints(:,SelectionArray(1,n));
    Objects{1,1}.width(:,n)=ObjectsOriginalWidth(:,SelectionArray(1,n));
end
clear n
%% Connect kinked tracks
% How many tracks should be connected?
disp('How many sets of tracks should be connected?')
prompt = {'How many sets of tracks should be connected?'};
dlg_title = 'How many sets of tracks should be connected?';
num_lines = 1;
def = {'3'};
info = inputdlg(prompt,dlg_title,num_lines,def);
NTracks=str2num(info{1});
SegID=[];

for n=1:NTracks
    disp('Track Segments')
    prompt = {'Track Segment IDs'};
    dlg_title = 'Track Segment IDs';
    num_lines = 1;
    def = {'1,2'};
    info = inputdlg(prompt,dlg_title,num_lines,def);
    SegID(:,n)=str2num(info{1})';
end
clear n
% Find Ends
for n=1:NTracks

```

```

a=find(SelectionArray==SegID(1,n));
b=find(SelectionArray==SegID(2,n));
Start1=Objects{1, 1}.data{1,a}(1,1:2);
End1=Objects{1, 1}.data{1,a}(end,1:2);
Start2=Objects{1, 1}.data{1,b}(1,1:2);
End2=Objects{1, 1}.data{1,b}(end,1:2);
EndDist(1,1)=sqrt((Start1(1,1)-Start2(1,1)).^2+(Start1(1,2)-
Start2(1,2)).^2);
EndDist(1,2)=sqrt((Start1(1,1)-End2(1,1)).^2+(Start1(1,2)-End2(1,2)).^2);
EndDist(1,3)=sqrt((Start2(1,1)-End1(1,1)).^2+(Start2(1,2)-End1(1,2)).^2);
EndDist(1,4)=sqrt((End1(1,1)-End2(1,1)).^2+(End1(1,2)-End2(1,2)).^2);
c=find(EndDist(1,:)==min(EndDist(1,:)));
ASeg=Objects{1, 1}.data{1,a};
BSeg=Objects{1, 1}.data{1,b};
if c==1
    Objects{1, 1}.data{1,a}=cat(1,flipud(ASeg),BSeg);
elseif c==2
    Objects{1, 1}.data{1,a}=cat(1,flipud(ASeg),flipud(BSeg));
elseif c==3
    Objects{1, 1}.data{1,a}=cat(1,ASeg,BSeg);
elseif c==4
    Objects{1, 1}.data{1,a}=cat(1,ASeg,flipud(BSeg));
end
% Full Length
Objects{1, 1}.data{1,a}(1,3)=0;
for m=2:size(Objects{1, 1}.data{1,a},1)
    Objects{1, 1}.data{1,a}(m,3)=Objects{1, 1}.data{1,a}(m-1,3)+sqrt((Ob-
jects{1, 1}.data{1,a}(m,1)-Objects{1, 1}.data{1,a}(m-1,1)).^2+(Objects{1,
1}.data{1,a}(m,2)-Objects{1,1}.data{1,a}(m-1,2)).^2);
end
Objects{1, 1}.length(1,a)=max(Objects{1, 1}.data{1,a}(m,3));
end

%% Filament Objects
%Filter Objects by length
%plot length histogram to select expected length
% disp('Expected DNA length in nm')
% prompt = {'Expected DNA length in nm'};
% dlg_title = 'DNA lenght [nm]';
% num_lines = 1;
% def = {'300'};
% info = inputdlg(prompt,dlg_title,num_lines,def);
% expectedDNALength=str2num(info{1});
% CutoffLength=floor(expectedDNALength*2/30)*10; % Lower length threshold
% MaxiumDNALength=ceil(max(Objects{1, 1}.length(1,:))/10)*10;
LengthData=Objects{1, 1}.length(1,:);

%fit gaussian to length distribution of all length bigger than half of
%expected length and filter all objects with length deviation bigger than
%2xstd

%Histogram of Length Data
% defbincenters=CutoffLength+5:10:MaxiumDNALength+5;
% defbinedges=(linspace(CutoffLength,MaxiumDNALength, ((MaxiumDNALength-Cut-
offLength)/10)+1)); % define bin edges, first entry is left edge of first bin,
last entry is right edge of last bin
% [LegthHisto,ind] = histc(LengthData,defbinedges); % LengthHisto: number of
counts in the bins defined by defbinedges; ind: bin identifier for each
LengthData entry
%
% GaussFitLengthdistribution=fit(defbincenters',LegthHisto','gauss1'); % Fit
with gaussian function
% SigmaLength=GaussFitLengthdistribution.c1/sqrt(2);
% MeanLength=GaussFitLengthdistribution.b1;
% Histogram with Gaussian Fit

```

```

% figure(1)
% clf
% hold on
% bar(defbinedges,LegthHisto,'histc')
% plot(GaussFitLengthdistribution)

%Filter Objects with Mean-2std>L>Mean+2std

% qualityApprovedLength =( Objects{1, 1}.length(1,:) > MeanLength-(2*SigmaLength)) &...
% ( Objects{1, 1}.length(1,:) < MeanLength+(2*SigmaLength));
qualityApprovedLength =( Objects{1, 1}.length(1,:) > 50);% &...
% ( Objects{1, 1}.length(1,:) < MeanLength+(300));

%Ensure that parts of conected filament segments are filtered out
for n=1:NTracks
    b=find(SelectionArray==SegID(2,n));
    qualityApprovedLength(1,b)=0;
end

% Apply Filter
FilamentIndexFiltered=[];
LengthFiltered=[];
DataFiltered={}; % X[px] Y[px] distance[px] width[px] height[counts] back-
ground[counts]
LengthFiltered = Objects{1, 1}.length(:,qualityApprovedLength);
CenterFiltered(:,1)=Objects{1, 1}.center_x(:,qualityApprovedLength);
CenterFiltered(:,2)=Objects{1, 1}.center_y(:,qualityApprovedLength);
FilamentIndexFiltered(:,1)=SelectionArray(1,qualityApprovedLength);
counter=1;
for n=1:size(LengthData,2)
    if qualityApprovedLength(n)==1
        DataFiltered{1,counter}=Objects{1, 1}.data{1,n};
        counter=counter+1;
    end
end
clear n

%% Curvature
% At which distances to the Protein Position should the curvature be evaluated?

disp('Curvature Query Points')
prompt = {'Curvature Query Points'};
dlg_title = 'Distance to center [nm]';
num_lines = 1;
def = {'4,6,8'};
info = inputdlg(prompt,dlg_title,num_lines,def);
QueryPoints=str2num(info{1});

% How many segments are needed to hit center and have eqidistant segments
% of ~0.1 nm length: NInterp
NInterp=round(((LengthFiltered(1,:)/0.1)-1)/2)*2+1;

% Interpolate
tic
NewLine={};
CenterPoint=[];
for n=1:size(LengthFiltered,2)
NewLine(1,n)=interparc(NInterp(n),DataFiltered{1,n}(:,1),Data-
Filtered{1,n}(:,2),'spline');
NewLine{1,n}(1,3)=0;
    NewLine{1,n}(1,4)=0;
    for m=2:size(NewLine{1,n},1)
        NewLine{1,n}(m,3)=sqrt((NewLine{1,n}(m,1)-NewLine{1,n}(m-1,1))^2+(New-
Line{1,n}(m,2)-NewLine{1,n}(m-1,2))^2);% actual length of each segment
    end
end

```

```

        NewLine{1,n}(m,4)=NewLine{1,n}(m-1,4)+NewLine{1,n}(m,3); % actual total
length along path after spline interpolation
    end
    LengthFiltered(3,n)=NewLine{1,n}(end,4); % actual total path length after
spline interpolation
    CenterPoint(n)=ceil(size(NewLine{1,n},1)/2); % Index of CenterPoint of each
path
end
toc
clear n m

% Load protein positions

disp('Open Protein Positions')
% Set path directory via standard GUI interface
[FileName, PathName] = uigetfile('*.txt','Fiesta Tracking Data');
fullpath = fullfile(PathName,FileName); %
Create full path adress
ProtDataRaw=dlmread(fullpath,'\t',1,0); %
read data from tab-delimited txt file excluding first line
disp(FileName) %
matrix-columns: Volume ( $\mu\text{m}^3$ ) Surface ( $\mu\text{m}^2$ ) Nb of obj. voxels Nb of surf.
voxels IntDen Mean StdDev Median Min Max X Y Z Mean dist. to surf.
(nm) SD dist. to surf. (nm) Median dist. to surf. (nm) XM YM ZM BX BY
BZ B-width B-height B-depth
OrgFileName=fullpath;
clear fullpath FileName PathName;

ProtData=ProtDataRaw(:,[1 3 12 13]); % ProtData: Index Area[ $\mu\text{m}^2$ ] X[px] Y[px]

% Discard Objects exceeding Area Threshold
disp('Protein Area Min and Max Limits')
prompt = {'Minimum Protein Area Cutoff [ $\mu\text{m}^2$ '],'Maximum Protein Area Cutoff
[ $\mu\text{m}^2$ ']};
dlg_title = 'Protein Area Cutoff [ $\mu\text{m}^2$ ]';
num_lines = 1;
def = {'0.0000001','0.0001'};
info = inputdlg(prompt,dlg_title,num_lines,def);
MinAreaCutoff=str2num(info{1});
MaxAreaCutoff=str2num(info{2});

qualityApprovedArea =(ProtData(:,2) > MinAreaCutoff)&(ProtData(:,2) < MaxAr-
eaCutoff);
ProtDataFiltered=ProtData(qualityApprovedArea,:);
ProtDataFiltered(:,5)=ProtDataFiltered(:,3)*Config.PixSize; %Xpos [nm]
ProtDataFiltered(:,6)=ProtDataFiltered(:,4)*Config.PixSize; %Ypos [nm]

% Protein radius input in nm
disp('Protein radius [nm]')
prompt = {'Protein radius [nm]'};
dlg_title = 'Protein radius [nm]';
num_lines = 1;
def = {'5'};
info = inputdlg(prompt,dlg_title,num_lines,def);
ProteinRadius=str2num(info{1}); %Protein Radius [nm]

% Protein radius based on area in nm
% ProteinRadius=sqrt(ProtDataFiltered(:,2)/pi)*1000; %Protein Radius [nm]
ColocMatrix=cell(1,size(LengthFiltered,2)); % for each filament a cell element
contains ID of all colocalizing proteins
DistanceArray=cell(1,size(LengthFiltered,2)); % for each filament a cell ele-
ment contains distances of segment elements to protein position
%qualityApprovedColocProt=zeros(size(ProtDataFiltered,1),1); % Array to filter
proteins without filament colocalization

```



```

%qualityApprovedColocFila=zeros(size(LengthFiltered,2),1); % Array to filter
filaments without protein colocalization
CountColocalization=zeros(size(LengthFiltered,2),1); % Counts number of pro-
teins colocalizing with each filament

% Find Proteins colocated with filament
for n=1:size(LengthFiltered,2) % loop all relevant filament
    for m=1:size(ProtDataFiltered,1)
        DistanceArray{1,n}(:,m)=sqrt((NewLine{1,n}(:,1)-ProtData-
Filtered(m,5)).^2+(NewLine{1,n}(:,2)-ProtDataFiltered(m,6)).^2);
        if min(DistanceArray{1,n}(:,m))<=ProteinRadius;
            ColocMatrix{1,n}(1,end+1)=m;
            %qualityApprovedColocProt(m,1)=1;
            %qualityApprovedColocFila(n,1)=1;
            CountColocalization(n,1)=CountColocalization(n,1)+1;
        end
    end
end
clear n m

% Filter Proteins on DNA with neighbouring proteins closer than 50 nm
ProtNew=[];
ProtSort=[];

for n=1:size(LengthFiltered,2) % loop all relevant filament
    Prot=ColocMatrix{1,n};
    if (size(ColocMatrix{1,n},2)==2)&&(sqrt((ProtDataFiltered(ColocMa-
trix{1,n}(1,1),5)-ProtDataFiltered(ColocMatrix{1,n}(1,2),5)).^2+(ProtData-
Filtered(ColocMatrix{1,n}(1,1),6)-ProtDataFiltered(ColocMa-
trix{1,n}(1,2),6)).^2)<=25)
        Prot=[];
    elseif size(ColocMatrix{1,n},2)>2
        v=1:1:size(ColocMatrix{1,n},2);
        P = nchoosek(v,2);
        for m=1:size(P,1)
            if sqrt((ProtDataFiltered(ColocMatrix{1,n}(1,P(m,1)),5)-ProtData-
Filtered(ColocMatrix{1,n}(1,P(m,2)),5)).^2+(ProtDataFiltered(ColocMa-
trix{1,n}(1,P(m,1)),6)-ProtDataFiltered(ColocMatrix{1,n}(1,P(m,2)),6)).^2)<25;
                ProtSort=[ProtSort ColocMatrix{1,n}(1,P(m,1)) ColocMa-
trix{1,n}(1,P(m,2))];
                % ProtNew(1,end+1)=ColocMatrix{1,n}(1,P(m,1));
                % ProtNew(1,end+2)=ColocMatrix{1,n}(1,P(m,2));
            end
        end
        qualitySort=ones(1,size(ColocMatrix{1,n},2));
        for p=1:size(ProtSort,2)
            A=find(ColocMatrix{1,n}(1,:)-ProtSort(1,p)==0);
            qualitySort(1,A)=0;
        end
        qualitySort=logical(qualitySort);
        ColocMatrix{1,n}=ColocMatrix{1,n}(1,qualitySort);

% Prot=unique(ProtNew); % remove doublets
% Prot=Prot(any(Prot,1)); %remove zeros
    end
end
clear A p m n

%qualityApprovedColocProt=logical(qualityApprovedColocProt); % convert into
logical values
%qualityApprovedColocFila=logical(qualityApprovedColocFila);

% Filament-Protein Distance Histogram for each Protein
MinDistanceArray=[]; % Array (row: filament, column: protein)
for n=1:size(LengthFiltered,2) % loop all relevant filament

```

```

    MinDistanceArray(n,:)=min(DistanceArray{1,n});
end
clear n
FilProtDistance=[]; % Row 1: Filament Protein Distance in nm, Row 2: Filament
ID, Row 3: Protein ID
for n=1:size(ProtDataFiltered,1) % loop all protein data
    b=find(MinDistanceArray(:,n)==min(MinDistanceArray(:,n)));
    FilProtDistance(1,n)=MinDistanceArray(b,n); % Filament Protein Distance in
nm
    FilProtDistance(2,n)=FilamentIndexFiltered(b,1); % filament ID
    FilProtDistance(3,n)=ProtDataFiltered(b,1); % Protein ID
    FilProtDistance(4,n)=ProtDataFiltered(b,2); % Protein Area um^2
end
clear n
% bins=0:0.5:100;
% [DistHisto,Distind] = histc(FilProtDistance(1,:),bins); % LengthHisto: number
of counts in the bins defined by defbinedges; ind: bin identifier for each
LengthData entry
%
% figure(1)
% clf
% bar(bins,DistHisto,'histc')

% scatter(ProtDataFiltered(:,5),ProtDataFiltered(:,6),5,'co') % draw circle at
center position

% Get angle between tangents at a distance from the center set as the Querypoints
alpha={};
Results=[];
h=1; % counter

for n=1:size(ColocMatrix,2) % loop through filaments
    if size(ColocMatrix{1,n},2)~=0 % if filament colocalizes with protein
        for m=1:size(ColocMatrix{1,n},2) % loop through colocalizing proteins
            DistSegMax=max(QueryPoints)/0.1;
            ProtPosSegm=find(DistanceArray{1,n}(:,ColocMa-
trix{1,n}(1,m))==min(DistanceArray{1,n}(:,ColocMatrix{1,n}(1,m))));
            if (ProtPosSegm-DistSegMax-1>=1) && (ProtPosSegm+DistSeg-
Max+1<=size(NewLine{1,n},1)) % only include proteins with further away from DNA
Ends than biggest QueryPoint distance
                Results(h,1)=FilamentIndexFiltered(n,1); % filament ID
                Results(h,2)=ProtDataFiltered(ColocMatrix{1,n}(1,m),1); % Protein
ID
                Results(h,3)=LengthFiltered(1,n); % filament length fiesta segmen-
tation in nm
                Results(h,4)=LengthFiltered(3,n); % filament length fine segmenta-
tion in nm
                Results(h,5)=ProtDataFiltered(ColocMatrix{1,n}(1,m),2); % protien
area in um^2
                Results(h,6)=min(DistanceArray{1,n}(:,ColocMatrix{1,n}(1,m))); %
protein filament distance [nm]
                for k=1:size(QueryPoints,2)
                    DistSeg=QueryPoints(k)/0.1;
                    alpha{1,n}(k,1)=(atan2((NewLine{1,n}(ProtPosSegm-
DistSeg+1,2)-NewLine{1,n}(ProtPosSegm-DistSeg-1,2)),(NewLine{1,n}(ProtPosSegm-
DistSeg+1,1)-NewLine{1,n}(ProtPosSegm-DistSeg-1,1)))); % atan2(y,x) of tangent
at querypoint before center position
                    % Rotate 2nd tangent by -phil
                    phil=alpha{1,n}(k,1);
                    R=[cos(-phil) -sin(-phil) ;sin(-phil) cos(-phil)]; % Rota-
tion Matrix
                    T2rotated=R*[(NewLine{1,n}(ProtPosSegm+DistSeg-1,1)-New-
Line{1,n}(ProtPosSegm+DistSeg+1,1));(NewLine{1,n}(ProtPosSegm+DistSeg-1,2)-
NewLine{1,n}(ProtPosSegm+DistSeg+1,2))];

```

```

Results(h,6+k)=abs(atan2(T2rotated(2,1),T2ro-
tated(1,1)))*180/pi; % resulting tangent cutting angle for querypoint in degree
end
h=h+1;
end
end
end
end

clear n m k

% Plot Objects on image that were evaluated
disp('Open Image')
% Set path directory via standard GUI interface
[FileName, PathName] = uigetfile('*.tif','Image');
fullpath = fullfile(PathName,FileName); %
Create full path adress
Image=imread(fullpath); % open source image
disp(FileName)

clear fullpath FileName PathName;

Xarray=1:1:size(Image,1); % Array with pixel index
Yarray=1:1:size(Image,2);
Xarray=Xarray*Config.PixSize; % Array with nm index
Yarray=Yarray*Config.PixSize;
C = {'b+-','r+-','g+-','y+-','m+-','c+-','m+-','b+-','r+-','g+-','y+-','m+-',
', 'c+-','m+-','b+-','r+-','g+-','y+-','m+-','c+-','m+-','b+-','r+-','g+-','y+-',
', 'm+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-','y+-','m+-',
', 'c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-',
', 'b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-',
', 'r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-',
', 'g+-','y+-','m+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-',
', 'y+-','m+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-','y+-',
', 'm+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-','y+-','m+-',
', 'c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-','y+-','m+-',
', 'c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-',
', 'b+-','r+-','g+-','y+-','m+-','c+-','b+-','r+-','g+-','y+-','m+-','c+-','b+-',
', 'r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-',
', 'b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-',
', 'k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-',
', 'c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-',
', 'm+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-',
', 'y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-',
', 'g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-',
', 'r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-',
', 'b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-',
', 'k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-',
', 'c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-',
', 'm+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-','g+-',
', 'y+-','m+-','c+-','k+-','b+-','r+-','g+-','y+-','m+-','c+-','k+-','b+-','r+-'

```

```

', 'g+-', 'y+-', 'm+-', 'c+-', 'm+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'm+-',
', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'm+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-',
', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-',
', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-',
', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-',
', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-',
', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-',
', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-',
', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-',
', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-',
', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-',
', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-',
', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-',
', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-',
', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-',
', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-',
', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-',
', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-',
', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-',
', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-',
', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-', 'm+-', 'c+-', 'k+-', 'b+-', 'r+-', 'g+-', 'y+-',
', 'm+-', 'c+-'];

```

figure(2)

clf

hold on

imagesc(Xarray,Yarray,Image) % draw image

colormap(flipud(gray)) % color with inverted gray scale

axis square % plot formate square

```

for n=1:size(Results,1) % loop through filaments

```

```

    plot(DataFiltered{1,find(FilamentIndexFiltered==Results(n,1))}(:,1),Data-
Filtered{1,find(FilamentIndexFiltered==Results(n,1))}(:,2),C{n},'Lin-
eWidth',2,'MarkerSize',0.5); % plot filament

```

```

    scatter(ProtDataFiltered(find(ProtDataFiltered(:,1)==Re-
sults(n,2)),5),ProtDataFiltered(find(ProtDataFiltered(:,1)==Re-
sults(n,2)),6),30,'co'); % draw circle at protein position

```

```

    scatter(CenterFiltered(find(FilamentIndexFiltered==Results(n,1)),1),Cen-
terFiltered(find(FilamentIndexFiltered==Results(n,1)),2),30,'bx'); % draw cross
at dna center

```

```

    textstrProt=[' ' num2str(Results(n,2))];

```

```

    text(ProtDataFiltered(find(ProtDataFiltered(:,1)==Re-
sults(n,2)),5),ProtDataFiltered(find(ProtDataFiltered(:,1)==Re-
sults(n,2)),6),textstrProt,'Color','c'); % add index to each protein
    textstr=[' ' num2str(Results(n,1))];

```

```

    text(double(CenterFiltered(find(FilamentIndexFiltered==Re-
sults(n,1)),1)),double(CenterFiltered(find(FilamentIndexFiltered==Re-
sults(n,1)),2)),textstr,'Color','w'); % add index to each filament
end

```

```

disp('Enter filename to save figure') % ask for filename for image

```

```

prompt = {'Enter filename to save figure'};

```

```

dlg_title = 'Filename';

```

```

num_lines = 1; def = {'FilteredFilaments'};

```

```

ImageFileNameDLG = inputdlg(prompt,dlg_title,num_lines,def);

```

```

ImageFileNameFig=[ImageFileNameDLG{1} '.fig'];

```

```

ImageFileNamepng=[ImageFileNameDLG{1} '.png'];

```

```

saveas(2,ImageFileNameFig)

```

```

saveas(2,ImageFileNamepng)

```

```

%% Export Results
OutputFileName=[OrgFileName(1:end-4) '_ProtPosDNAAngleAnalysisResults.txt'];
OutputFileName2=[OrgFileName(1:end-4) '_ProtPosDNAAngleAnalysisFilProtDis-
tance.txt'];
save(OutputFileName,'Results','-ascii') % save results table as txt file
save(OutputFileName2,'FilProtDistance','-ascii') % save FilProtDistance table
as txt file
OutputFileNameMat = [OrgFileName(1:end-4) '_Workspace'];
uisave({'Results','Config','DataFiltered','Filament','LengthFil-
tered','LengthData','Molecule','NewLine','Objects','QueryPoints','ProtData-
Filtered','FilProtDistance','ColocMatrix'}, OutputFileNameMat); % Save in
.mat file

def = {'FilteredFilaments'};
ImageFileNameDLG = inputdlg(prompt,dlg_title,num_lines,def);
ImageFileNameFig=[ImageFileNameDLG{1} '.fig'];
ImageFileNamepng=[ImageFileNameDLG{1} '.png'];
saveas(2,ImageFileNameFig)
saveas(2,ImageFileNamepng)

%% Export Results
OutputFileName=[OrgFileName(1:end-4) '_ProtPosDNAAngleAnalysisResults.txt'];
OutputFileName2=[OrgFileName(1:end-4) '_ProtPosDNAAngleAnalysisFilProtDis-
tance.txt'];
save(OutputFileName,'Results','-ascii') % save results table as txt file
save(OutputFileName2,'FilProtDistance','-ascii') % save FilProtDistance table
as txt file
OutputFileNameMat = [OrgFileName(1:end-4) '_Workspace'];
uisave({'Results','Config','DataFiltered','Filament','LengthFil-
tered','LengthData','Molecule','NewLine','Objects','QueryPoints','ProtData-
Filtered','FilProtDistance','ColocMatrix'}, OutputFileNameMat); % Save in
.mat file

def = {'FilteredFilaments'};
ImageFileNameDLG = inputdlg(prompt,dlg_title,num_lines,def);
ImageFileNameFig=[ImageFileNameDLG{1} '.fig'];
ImageFileNamepng=[ImageFileNameDLG{1} '.png'];
saveas(2,ImageFileNameFig)
saveas(2,ImageFileNamepng)

%% Export Results
OutputFileName=[OrgFileName(1:end-4) '_ProtPosDNAAngleAnalysisResults.txt'];
OutputFileName2=[OrgFileName(1:end-4) '_ProtPosDNAAngleAnalysisFilProtDis-
tance.txt'];
save(OutputFileName,'Results','-ascii') % save results table as txt file
save(OutputFileName2,'FilProtDistance','-ascii') % save FilProtDistance table
as txt file
OutputFileNameMat = [OrgFileName(1:end-4) '_Workspace'];
uisave({'Results','Config','DataFiltered','Filament','LengthFil-
tered','LengthData','Molecule','NewLine','Objects','QueryPoints','ProtData-
Filtered','FilProtDistance','ColocMatrix'}, OutputFileNameMat); % Save in
.mat file

```

## 8.4 DNA-Schadenssubstrate und Biegewinkel

Tabelle 8: DNA-Schadenssubstrate

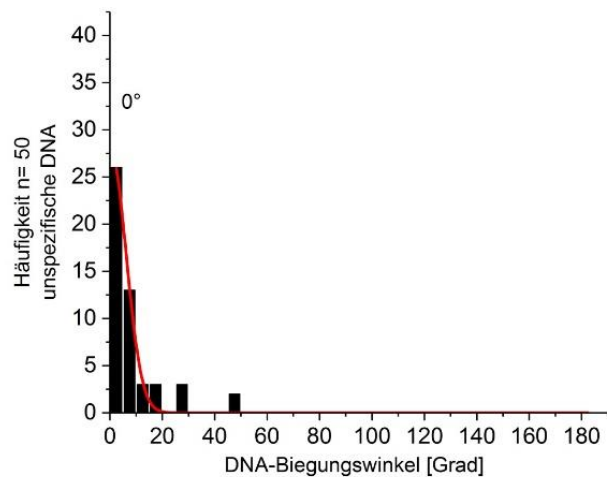
Substrate	Schadensposition	Restriktionsenzyme	DNA-Sequenzen
8oxoG:C	254 bp = 86,36 nm $\Delta$ 0,55 nm (50%)	Ndel/BsaXI	GCA TGC CTC <b>oxo</b> GAG TCT AGA GGT ACC AGA TCT GAT CCT CTA GAG TCG ACC
G:A	257 bp = 87,38 nm $\Delta$ 1,55 nm (51%)	Ndel/BsaXI	GCA TGC CTC GAG <b>G</b> CT AGA GGT ACC AGA TCT GAT CCT CTA GAG TCG ACC
oxoG:A	252 bp = 85,67 nm $\Delta$ 0,15 nm (50%)	Ndel/BsaXI	GCA TGC <b>Coxo</b> GC GAG TCT AGA GGT ACC AGA TCT GAT CCT CTA GAG TCG ACC
ethenoA:T	255bp = 86,7 nm (50%)	Ndel/BsaXI	GCA TGC CTC <b>Getheno</b> AG TCT AGA GGT ACC AGA TCT GAT CCT CTA GAG TCG ACC
nsp DNA			GCA TGC CTC GAG TCT AGA GGT ACC AGA TCT GAT CCT CTA GAG TCG ACC

**Tabelle 9: Zusammenfassung der Ergebnisse einzelner DNA-Biegewinkel**

Substrat	Biegewinkel				
	automatisierte Messsoftware				Messverfahren
	Länge	4 nm	6 nm	8 nm	Tangente/ Sekante
nsp DNA	505 bp	$5^\circ \pm 16^\circ$	$2^\circ \pm 25^\circ$	$5^\circ \pm 15^\circ$	Tangente
				$37^\circ \pm 14^\circ$	
		$0^\circ \pm 10^\circ$	$0^\circ \pm 14^\circ$	$0^\circ \pm 17^\circ$	Sekante
	2224 bp	$6^\circ \pm 13^\circ$	$9^\circ \pm 15^\circ$	$6^\circ \pm 21^\circ$	Tangente
oxoG:A	505 bp	$5^\circ \pm 13^\circ$	$5^\circ \pm 20^\circ$	$0^\circ \pm 3^\circ$	Tangente
				$15^\circ \pm 18^\circ$	
		$0^\circ \pm 9^\circ$	$0^\circ \pm 14^\circ$	$5^\circ \pm 14^\circ$	Sekante
G:A	505 bp	$6^\circ \pm 12^\circ$	$8^\circ \pm 15^\circ$	$2^\circ \pm 32^\circ$	Tangente
				$15^\circ \pm 4^\circ$	
		$0^\circ \pm 8^\circ$	$5^\circ \pm 10^\circ$	$5^\circ \pm 15^\circ$	Sekante
oxoG:C	505 bp	$11^\circ \pm 14^\circ$	$9^\circ \pm 10^\circ$	$0^\circ \pm 3^\circ$	Tangente
		$27^\circ \pm 5^\circ$		$11^\circ \pm 4^\circ$	
				$30^\circ \pm 15^\circ$	
		$0^\circ \pm 5^\circ$	$9^\circ \pm 10^\circ$	$5^\circ \pm 8^\circ$	Sekante
		$10^\circ \pm 8^\circ$		$27^\circ \pm 9$	
$\epsilon$ A:T	505 bp	$5^\circ \pm 15^\circ$	$5^\circ \pm 20^\circ$	$0^\circ \pm 26^\circ$	Tangente
				$0^\circ \pm 17^\circ$	
		$0^\circ \pm 10^\circ$	$0^\circ \pm 14^\circ$	$0^\circ \pm 17^\circ$	Sekante

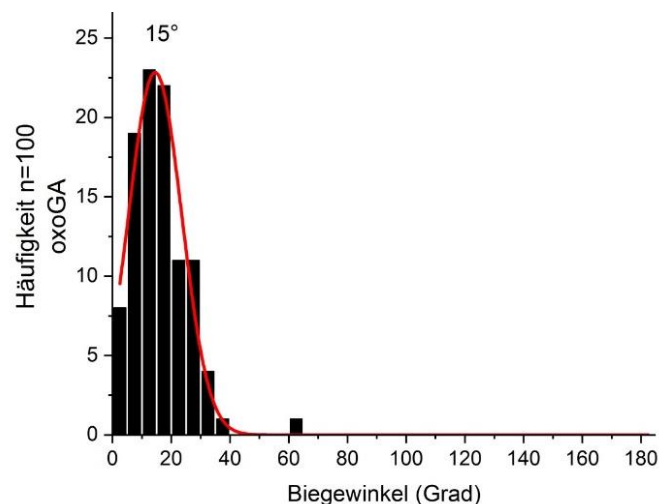
## 8.5 Eigene Ergebnisse mit der manuellen Messmethode

### 8.5.1 ungeschädigte DNA



**Abbildung 109: Ergebnis der manuellen Biegungswinkelmessung an ungeschädigter DNA**  
Bei der manuellen Vermessung von ungeschädigter DNA (nsp505\_1704260001) ergab sich eine Biegung von  $0^\circ \pm 6^\circ$ .

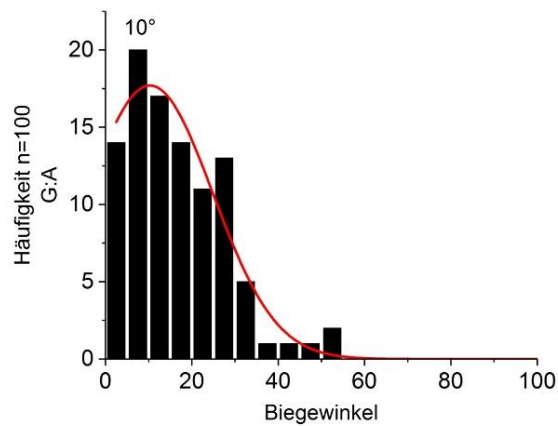
### 8.5.2 oxoG:A



**Abbildung 110: Ergebnis der manuellen Biegungswinkelmessung an oxoG:A**  
Bei der manuellen Vermessung von oxoG:A (oxoGA\_201702230004) ergab sich eine Biegung von  $15^\circ \pm 9^\circ$ .



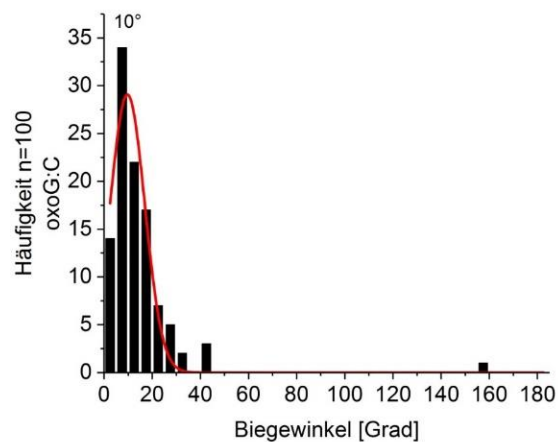
### 8.5.3 G:A



**Abbildung 111: Ergebnis der manuellen Biegewinkelmessung an G:A**

Bei der manuellen Vermessung von G:A (GAnew\_20170614000) ergab sich eine Biegung von  $10^\circ \pm 15^\circ$ .

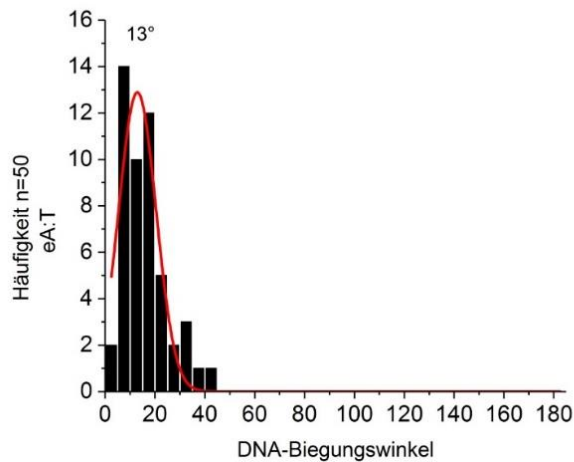
### 8.5.4 oxoG:C



**Abbildung 112: Ergebnis der manuellen Biegewinkelmessung an oxoG:C**

Bei der manuellen Vermessung von oxoG:C (oxoGC\_1705240000-oxoGC\_1705240004) ergab sich eine Biegung von  $10^\circ \pm 7^\circ$ .

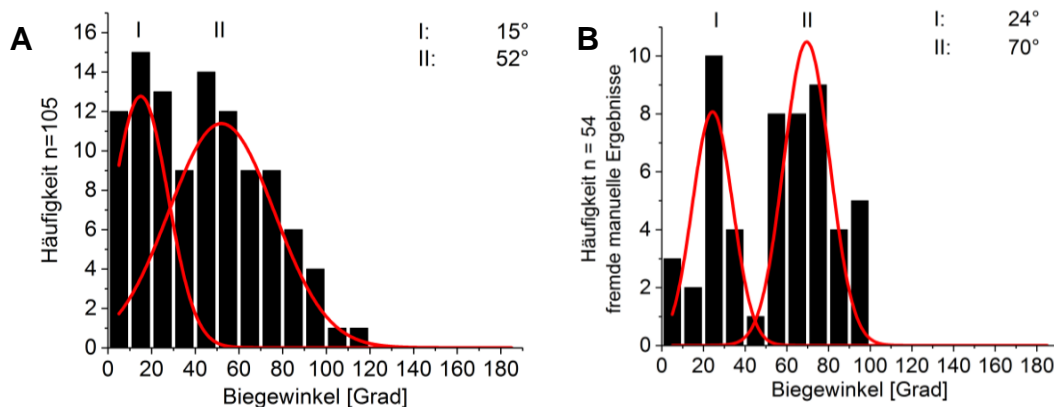
### 8.5.5 $\epsilon$ A:T



**Abbildung 113: Ergebnis der manuellen Biegungswinkelmessung an  $\epsilon$ A:T**

Bei der manuellen Vermessung von  $\epsilon$ A:T (ethenoA-20170421000) ergab sich eine Biegung von  $13^\circ \pm 8^\circ$ .

### 8.5.6 MutY-nspDNA

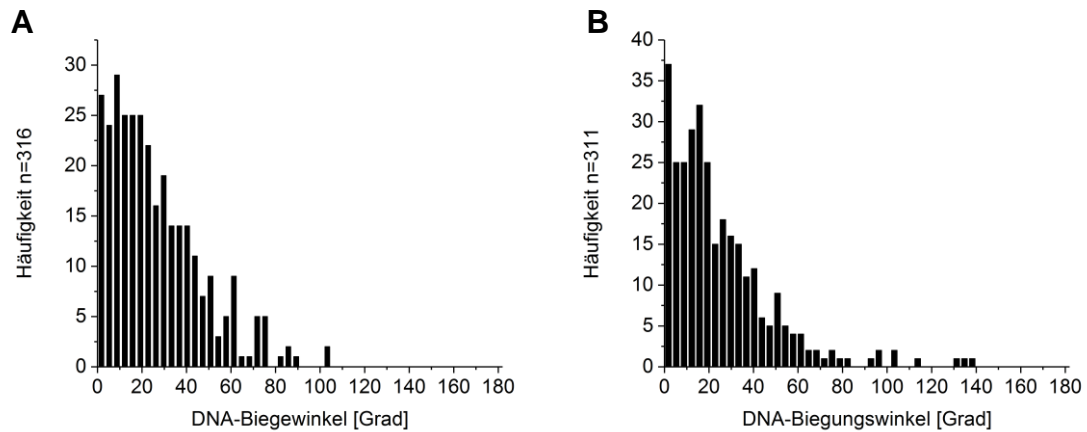


**Abbildung 114: Manuelle Messung des Biegewinkels an MutY-DNA-Komplexen mit ungeschädigter DNA**

**A:** Gezeigt ist die eigene manuelle Messung von MutY-DNA-Komplexen mit ungeschädigter DNA, wobei das Proteinvolumen nicht berücksichtigt wurde. Es lassen sich zwei Biegungsspezies bei  $15^\circ \pm 13^\circ$  und  $52^\circ \pm 14^\circ$  erkennen.

**B:** Die Abbildung zeigt die Messergebnisse anderer Experimentatoren an der Peakposition, wobei nur Proteine mit einem Volumen von  $20 \text{ nm}^3$  bis  $40 \text{ nm}^3$  in die Auswertung einbezogen wurden. Außerdem mussten die Proteine einen Mindestabstand von  $50 \text{ nm}$  zum nächstgelegenen Protein aufweisen und sich nicht am Ende des DNA-Stranges befinden. Es zeigten sich zwei Biegungsspezies, eine kleinere Fraktion mit einem Winkel von  $24^\circ \pm 10^\circ$  und eine Größere bei einem Biegewinkel von  $70^\circ \pm 11^\circ$ . Der Vergleich der vorausgegangenen Messungen aus der AG Tessmer mit den selbst gemessenen Werten verdeutlicht nochmals, wie wichtig eine Standardisierung der Biegungswinkelmessung ist. Die Vermessung der DNA-Biegung mit der Software zur automatisierten Biegungswinkelmessung ist ein Ansatz für eine objektivere Messverfahren.

## 8.6 Programmevaluation



**Abbildung 115: Möglicher Einfluss des Depositionsverfahrens**

Ziel war die Untersuchung einer möglichen Auswirkung des Depositionsverfahrens auf die Biegewinkelmessung. Exemplarisch wurde jeweils ein Experiment mit dem gleichen Basenscha-den oxoG:A (**A**: eigene Deposition; **B**: von Lisa Hirsch deponiert) mit der automatisierten Messsoftware untersucht. Es zeigte sich keine depositionsbedingte Beeinflussung der Biege-winkelergebnisse.

## 9 Eigene Publikationen

Aus dieser Promotion ist die folgende Veröffentlichung hervorgegangen, die sich zur Zeit unter Begutachtung befindet:

### **Automated AFM analysis of DNA bending reveals initial lesion sensing strategies of DNA glycosylases**

Disha M Bangalore<sup>1</sup>, Hannah S Heil<sup>1</sup>, **Christian F Mehringer**, Lisa Hirsch, Katherina Hemmen, Katrin G Heinze\*, and Ingrid Tessmer\*

(<sup>1</sup> These authors contributed equally)

Rudolf Virchow Center for Experimental Biomedicine, Josef-Schneider-Strasse 2, 97080 Würzburg, Germany

# Danksagung

Diese Danksagung gilt all denjenigen, die durch jegliche Unterstützung zum Gelingen dieser Doktorarbeit beigetragen haben

Besonderer Dank gilt Priv.-Doz. Dr. Ingrid Tessmer für die Bereitstellung des interessanten Promotionsthemas und für die Betreuung und Unterstützung bei der Erarbeitung dieser Dissertation.

Für die Übernahme des Korreferats gilt mein Dank Frau Prof. Dr. Antje Gohla.

Für die Übernahme der Berichterstattung danke ich Frau Prof. Dr. Martina Prelog.

Ebenso gilt mein Dank der Arbeitsgruppe Mikroskopie von Frau Prof. Dr. Katrin Heinze des Rudolf-Virchow-Zentrums für Experimentelle Biomedizin für die technische Hilfestellung. Besonders möchte ich hierbei Hannah Heil hervorheben, die durch ihre IT-Fähigkeiten diese Dissertation maßgeblich unterstützt hat.

Besonderer Dank gilt meinen Eltern, die mich während meines Studiums und der Dissertation immer uneingeschränkt unterstützt haben.

Außerdem möchte ich mich bei Sandra für ihre liebevolle Unterstützung und Motivation bedanken.