Sequence-specific Methyltransferase-Induced Labeling of DNA (SMILing DNA)

Dissertation

zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
des Fachbereichs Chemie der Universität Dortmund

Angefertigt am Max-Planck-Institut für molekulare Physiologie in Dortmund

eingereicht von
Dipl.-Chem. Goran Pljevaljčić
aus Dortmund

Dortmund, Mai 2002

Die vorliegende Arbeit wurde in der Zeit von März 1999 bis Mai 2002 am Max-Planck-Institut für molekulare Physiologie in Dortmund unter der Anleitung von Prof. Dr. Elmar Weinhold und Dr. Axel J. Scheidig in der Abteilung Physikalische Biochemie von Prof. Dr. Roger S. Goody durchgeführt.

1. Gutachter: Prof. Dr. Roger S. Goody

2. Gutachter: Prof. Dr. Wolfgang Kreiser

Hiermit versichere ich an Eides Statt, dass ich die vorliegende Arbeit selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe.

Dortmund,

Contents

Abbreviations

Å	Ångström (1 Å = 10^{-10} m)
AdoHcy	*S*-Adenosylhomocysteine
AdoMet	*S*-Adenosyl-L-methionine
AP	Alkaline phosphatase
APS	Ammoniumperoxysulfate
AZ	*N*-adenosylaziridine
BAZ	Biotinylated *N*-adenosylaziridine
B-Factor	Temperature factor
Bp	Base pair
Cy5AZ	*N*-adenosylaziridine decorated with Cy5
DAZ	Dansylated *N*-adenosylaziridine
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDIA	*N*-Ethyldiisopropylamine
EDTA	Ethylenediaminetetraacetate
E. coli	*Escherichia coli*
ESI	Electospray ionisation
F_{calc}	Calculated structure factor amplitudes
Fmoc	9-fluorenylmethoxycarbonyl
F_{obs}	Observed structure factor amplitudes
FRET	Fluorescence resonance energy transfer
H	Hours
HEPES	2-[4-(2-Hydroxyethyl)-1-piperazinyl]-ethane sulfonic acid
IPA	2-Propanol
K_D	Dissociation constant
min	Minute
MALDI	Matrix-assisted laser desorption ionisation
MANT	*N*-Methylanthraniloyl

MTase	Methyltransferase
M·*Bcn*IB	N4-cytosine DNA MTase from *Bacillus centrosporus*
M·*Hha*I	C5-cytosine DNA MTase from *Haemophilus haemolyticus*
M·*Taq*I	N6-adenine DNA MTase from *Thermus aquaticus*
MS	Mass spectrum
NHS	*N*-succinimidyl
NMR	Nuclear magnetic resonance
NVOC	6-Nitroveratryloxycarbonyl
OAc	Acetate
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PNA	Peptide nucleic acid
R_f	Retention factor
R-factor	Residual factor
RNA	Ribonucleic acid
RP-HPLC	Reversed-phase high performance liquid chromatography
R_t	Retention time
SDS	Sodium dodecyl sulfate
SMILing	Sequence-specific Methyltransferase-Induced Labeling
TEA	Triethylamine
TEMED	*N*,*N*,*N*',*N*'-Tetramethylethylenediamine
THAP	2,4,6-Trihydroxyacetophenone
TLC	Thin layer chromatography
TFO	Triplex forming oligonucleotide
TMS	Tetramethylsilane
Tris	Tris(hydoxymethyl)amino methane
λ_{Ex}	Excitation wavelength
λ_{Em}	Emission wavelength

Abbreviations for the amino acids and the nucleobases

Amino acids	**Three letter code**	**One letter code**
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glycine	Gly	G
Glutamine	Gln	Q
Glutamic acid	Gly	E
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Nucleobase	**Three letter code**	**One letter code**
Adenine	Ade	A
Cytosine	Cyt	C
Guanine	Gua	G
Thymine	Thy	T

CHAPTER 0

Summary, Conclusions and Outlook (german)

0 Summary, Conclusions and Outlook (german)

Die sequenzspezifische Markierung von DNA ist von besonderer Bedeutung in der Biotechnologie und der Bioanalytik. Sie stellt eine große Herausforderung dar, da definierte DNA-Sequenzen erkannt und anschließend kovalent modifiziert werden müssen. Verschiedene Ansätze zur sequenzspezifischen Erkennung von DNA, wie die Triplehelix-Bildung mit Oligonukleotiden oder PNA, die spezifische Erkennung der kleinen Furche mit Polyamiden und die Erkennung von DNA mit entwickelten Zink-Finger-Proteinen sind an einem Punkt angelangt, an dem gegebene Sequenzen erkannt werden können. Allerdings werden diese Systeme nur in Ausnahmefällen mit reaktiven Gruppen versehen und zur sequenzspezifischen DNA-Markierung eingesetzt. Mit der Entwicklung eines neuen Aziridin Kofaktors für die N6-Adenin-DNA-Methyltransferase (MTase) M·*Taq*I wurde bereits die kovalente sequenzspezifische Markierung von DNA ermöglicht. Dieses „**S**equenzspezifische **M**ethyltransferasen **I**nduzierte ***L**abel**ing***“ von DNA (SMILing DNA) sollte anhand aller drei Klassen von DNA MTasen demonstriert werden. Es wurde gezeigt, dass nicht nur M·*Taq*I sondern auch die C5-Cytosin-DNA-MTase M·*Hha*I und die N4-Cytosin-DNA-MTase M·*Bcn*IB die Kupplung von *N*-Adenosylaziridin (AZ) mit DNA katalysieren. Alle diese Reaktionen sind sequenzspezifisch und belegen die generelle Anwendbarkeit der SMILing-Technologie (Abb. 0.1).

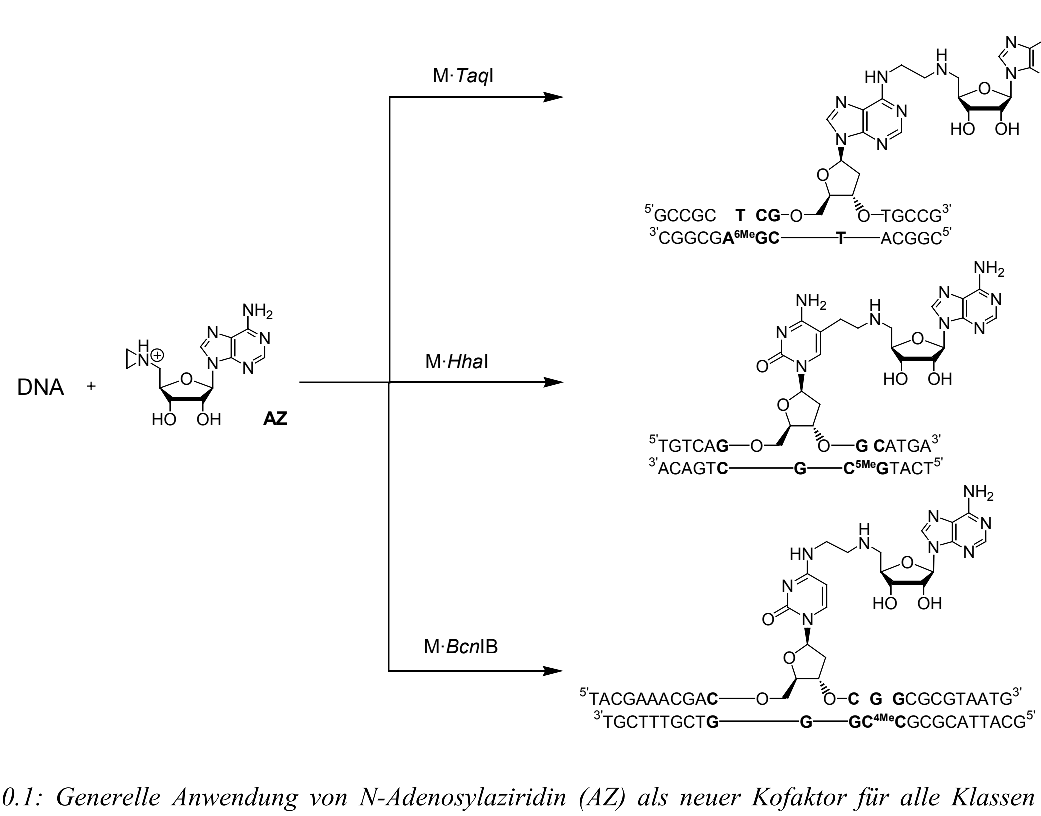


Abb. 0.1: Generelle Anwendung von N-Adenosylaziridin (AZ) als neuer Kofaktor für alle Klassen von DNA MTasen.

Eine Optimierung der Reaktionsgeschwindigkeit der M·*Taq*I-katalysierten Kupplungsreaktion zwischen DNA und AZ wurde durch Variation des pH-Wertes der Reaktionslösung erreicht. Es konnte gezeigt werden, dass die Reaktionsgeschwindigkeit mit der Abnahme des pH-Wertes zunimmt. Das kann mit einer Aktivierung des Aziridinringes durch Protonierung des Aziridinstickstoffs erklärt werden. Eine Modifikation von AZ an der 8-Position scheint keinen Einfluss auf die maximale Reaktionsgeschwindigkeit der Kupplungsreaktion zu haben. Des weiteren konnte gezeigt werden, dass das starke Nukleophil DTT bis zu einer Konzentration von 1 mM keinen messbaren Einfluss auf die Kupplungsreaktion hat.

Zusätzlich konnte die sequenzspezifische Markierung von Plasmid-DNA mit dem dansylierten Aziridinkofaktor DAZ in einer M·*Taq*I-katalysierten Kupplungsreaktion demonstriert werden (Abb. 0.2, A). Weiterhin wurde gezeigt, dass M·*Hha*I die Kupplung des dansylierten Kofaktors DAZ mit Plasmid-DNA ebenfalls katalysiert (Abb. 0.2, B). Dies veranschaulicht exemplarisch, dass N8-modifizierte Aziridinkofaktoren auch mit der Cytosin-spezifischen DNA MTase M·*Hha*I zur Markierung von DNA verwendet werden können.

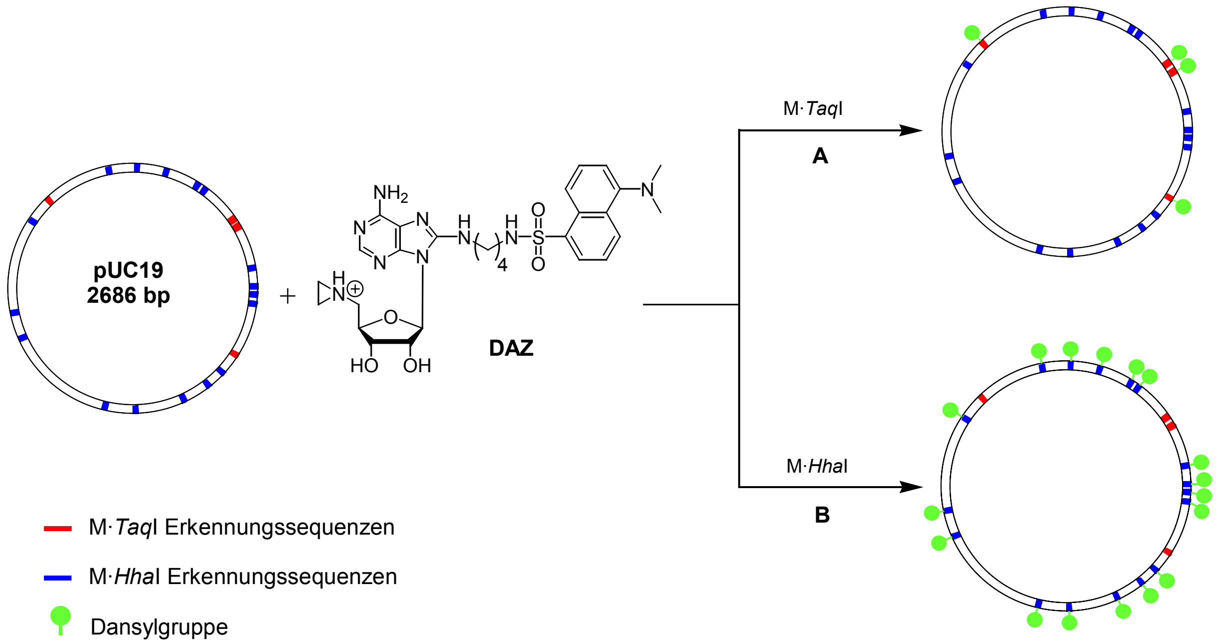


*Abb. 0.2: Sequenzspezifische Fluoreszenzmarkierung von Plasmid-DNA mit dem dansylierten AZ Kofaktor DAZ. Plasmid-DNA Markierung mittels M·*TaqI *und M·*HhaI*. Die M·*TaqI*-katalysierte Kupplung von DAZ an kurze Duplex-Oligonukleotide wurde schon im Detail analysiert (Pljevaljčić, 1999).*

Ein genereller Syntheseweg, unter Verwendung einer photolabilen Schutzgruppe, ermöglichte die einfache Zugänglichkeit zu unterschiedlich modifizierten Kofaktoren. Damit ist der Weg zu unterschiedlich dekorierten Kofaktoren geebnet. Die neue Syntheseroute wurde zur

Synthese eines biotinylierten Aziridinkofaktors (BAZ) und eines mit dem Fluorophor Cy5 modifizierten Aziridinkofaktors (Cy5AZ) verwendet. Beide Kofaktoren binden mit einer Affinität, die vergleichbar ist mit der Affinität des natürlichen Kofaktors AdoMet, an die DNA-MTase M·*Taq*I. Der biotinylierte Aziridinkofaktor BAZ konnte in einer M·*Taq*I-katalysierten Reaktion sequenzspezifisch auf kurze als auch auf lange DNA übertragen werden (Abb. 0.3). Die sequenzspezifische Biotinylierung erfolgt unter kompletten Umsatz des DNA-Eduktes. Besonders interessant an dieser Reaktion ist, dass auch indirekte Markierungssysteme, wie das hier verwendete Biotin-Streptavidin-System, mit der SMILing-Technologie kombiniert werden können.

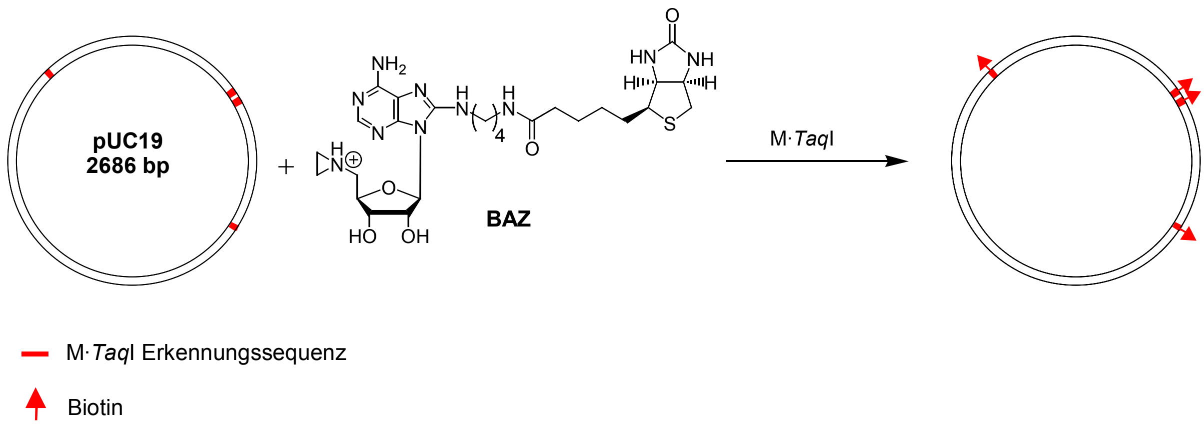


Abb. 0.3: Sequenzspezifische Biotinylierung von Plasmid-DNA mit dem biotinyliertem AZ Kofaktor BAZ. Die Biotinylierung ist auch mit kurzen Duplex-Oligonucleotiden gelungen.

Eine Kristallstruktur des Komplexes ausgehend von M·*Taq*I, DNA und BAZ konnte belegen, dass wie erwartet die exozyklische Aminogruppe von Adenin, in der Erkennungssequenz von M·*Taq*I, mit dem biotinylierten Aziridinkofaktor kovalent verbunden ist, was die Sequenzspezifität der Reaktion eindeutig belegt. Des weiteren konnte durch interne Funktionalisierung von DNA mittels SMILing ein System zur Selektion von Desoxyribozymen entwickelt werden. In Zusammenarbeit mit der Arbeitsgruppe von Dr. Andres Jäschke (FU Berlin), konnte gezeigt werden, dass ein Pool von DNA-Sequenzen mit Biotin markiert werden kann. Diese Methode stellt meiner Kenntnis nach den einfachsten Weg zur spezifischen internen Funktionalisierung von DNA dar. Außerdem konnte die biotinylierte DNA als Templat in der PCR eingesetzt werden. Somit sollte es prinzipiell möglich sein intern funktionalisierte DNA-Pools für die Selektion neuartiger, modifizierter Desoxyribozyme, mit zusätzlichen Funktionalitäten, einzusetzten.

In Anbetracht der Tatsache, dass es eine Vielzahl von verschiedenen DNA-MTasen gibt, erscheint es möglich eine grosse Zahl von unterschiedlichen DNA-Sequenzen zu markieren.

Da die Aziridinkofaktoren allgemein für MTasen entwickelt wurden und da die Kofaktorbindungstasche von MTasen konserviert ist, erscheint es wahrscheinlich, dass die neue Kofaktorfamilie auch für MTasen von anderen Biomolekülen geeignet ist.
Allerdings wurde die Modifizierung von AZ an der 8-Position speziell für M·*Taq*I gewählt. Auch wenn dabei darauf geachtet wurde, dass eine Vielzahl weiterer MTasen diese an der 8-Position modifizierten Kofaktoren verwenden könnten, gibt es trotzdem MTasen, die z.B. eine Modifikation an der 7-Position bevorzugen sollten (siehe Tabelle 3.1b). Für diesen Fall muss eine neue Syntheseroute zu an der 7-Position modifizierten Aziridinkofaktoren entwickelt werden. Damit würde auch das SMILing mit diesen Enzymen ermöglicht werden. Des weiteren könnte die Reaktionsgeschwindigkeit der Markierungsreaktion erheblich verbessert werden. Die größte Reaktionsgeschwindigkeitskonstante, die für die M·*Taq*I-katalysierten Kupplungsreaktionen zwischen DNA und AZ erzielt werden konnte, beträgt 0,6 h^{-1}. Damit ist die Reaktionsgeschwindigkeit der Kupplungsreaktion um zwei Größenordnungen kleiner als die Geschwindigkeit der Reaktion mit dem natürlichen Kofaktor AdoMet. Deshalb muss weitere Arbeit in die Erhöhung der Reaktionsgeschwindigkeit gesteckt werden. Ein Ansatzpunkt könnte eine Modifizierung der Aziridineinheit der Kofaktoren sein. Bei der Entwicklung der Kofaktoren wurde die Methioninseitenkette des natürlichen Kofaktors AdoMet aus Gründen der einfacheren Synthese vernachlässigt. Die Aminosäureseitenkette ist für die Funktion von AdoMet jedoch sehr wichtig, was auch daran ersichtlich wird, dass 5’-deoxy-5’-dimethylthioadenosin ein sehr schlechter Kofaktor ist. Die Methioninseitenkette scheint die Aufgabe zu haben, die Methylgruppe des Kofaktors in Richtung des nukleophilen Akzeptors zu positionieren. Die Aziridinkofaktoren besitzten jedoch keine Seitenkette an der Aziridineinheit, die den Aziridinring in eine für die Reaktion günstige Position orientieren könnte. Die Betrachtung der Kristallstruktur des Komplexes aus M·*Taq*I, DNA und BAZ erlaubt es, einen Kofaktor vorzuschlagen, der am Aziridindreiring eine Hydroxymethyl-Gruppe trägt, die diese Aufgabe erfüllen könnte. Die Hydroxyl-Gruppe könnte eine neue Wasserstoffbrücke zum Protein ausbilden und damit den Aziridinring fixieren, so das die Reaktionsgeschwindigkeit aufgrund einer eingeschränkten Flexibilität des Aziridinrings erhöht werden könnte. Weiterhin könnte eine Aktivierung des Aziridinringes durch Erhöhung der Elektrophilie erreicht werden. Dazu könnte eine elektronenziehende Gruppe, wie z.B. eine Cyano Gruppe, an den Ring angebracht werden. Eine solche Modifikation könnte allerdings auch den Aziridinring soweit aktivieren, dass nicht katalysierte Hintergrundreaktionen mit DNA, die mit den bisherigen Aziridinkofaktoren nicht beobachtet wurden, auftreten.

Außerdem wäre zur Etablierung der SMILing-Technologie wichtig, die Nützlichkeit der Aziridinkofaktoren mit anderen MTasen zu untersuchen. Dazu gehören nicht nur weitere DNA-MTasen sondern auch MTasen anderer Substrate, wie z.B. Protein-MTasen oder RNA-MTasen. Damit würde der Weg zu einem SMILing von weiteren Biopolymeren eröffnet und das Anwendungsspektrum von SMILing enorm erweitert sein.

Eine weitere Frage die in dieser Arbeit beantwortet wurde, ist die Anwendbarkeit von 2-Aminopurin-haltiger DNA zur Untersuchung von Enzymen die einen Basenausklapp-mechanismus besitzen. Bis jetzt ist nicht geklärt worden, ob die beobachtete 2-Aminopurin-Fluoreszenzänderung in der Tat auf ein Herausklappen des Fluorophors aus der DNA und exakt den gleichen Basenausklappmechanismus zurückgeführt werden kann. Im Fall von M·*Taq*I zum Beispiel, erscheint es anhand der Kristallstruktur des Komplexes aus M·*Taq*I, DNA und einem Kofaktoranalog nicht möglich, dass die Position des ausgeklappten Adenins von einem 2-Aminopurin-Rest eingenommen werden kann, da die Aminogruppe von 2-Aminopurin zu einer sterischen Überlappung mit einer Aminosäureseitenkette des Enzyms führen würde. Deshalb wurde die Kristallstruktur von M·*Taq*I, DNA mit 2-Aminopurin an der Zielposition und dem Kofaktoranalog AETA bestimmt. Die Kristallstruktur zeigt, dass sich der 2-Aminopurin-Rest tatsächlich in einer extrahelikalen Position befindet, jedoch im Vergleich zur natürlichen Base leicht im aktiven Zentrum verschoben ist. Damit konnte erstmalig mit einer Kristallstruktur gezeigt werden, dass DNA, die 2-Aminopurin an der Zielposition enthält, eine gute Sonde für die Detektion und Beobachtung des durch DNA-modifizierende Enzyme verursachten Basenausklappens ist.

Zusammenfassend kann gesagt werden, dass SMILing DNA großes Potential für die Biotechnologie, die medizinische Diagnostik (Mutationsdetektion) und die Nanotechnologie besitzt. Weitere Untersuchungen werden zeigen, ob SMILing DNA auf SMILing von anderen Biopolymeren erweitert werden kann.

CHAPTER 1

Introduction

1 Introduction

Life of all species is governed by processes at the molecular level. The cell is the molecular factory of life and contains vital biomolecules such as nucleic acids and proteins. Despite the wealth of information acquired over the last decades regarding the processes taking place in the cellular world, an enormous amount of work still has to be accomplished in order to understand the detailed rules of life. A major task in investigating this world of microstructures is to make them visible, thus providing insight into structures and mechanisms. Different methodologies have been developed and an immense amount of structural information has been accumulated using light microscopy, electron microscopy, X-ray crystallography and NMR spectroscopy. X-Ray crystallography, which enables us to determine the structures of biomolecules at atomic resolution, is contributing greatly to our understanding of biomolecular structures and their mode of interaction. Since 1953, when Watson and Crick published the first model of DNA structure (Watson & Crick, 1953), X-ray crystallography has made enormous advances, culminating in the determination of the crystal structure of such large complexes of proteins and nucleic acids such as the ribosome (Clemons *et al.*, 1999; Ban *et al.*, 2000; Schluenzen *et al.*, 2000). Additionally, advanced NMR techniques are accelerating this collection of high resolution structural information, and the completion of the different genome sequencing projects emphasizes further that the need for even more structural information is essential for the understanding of cellular function.

Besides questions concerning the structure and function of cellular machineries, mechanistic aspects are of major interest. In some cases, these questions have been addressed using kinetic crystallography by employing caged compounds as molecular triggers (Schlichting & Goody, 1997). More frequently, in view of the limits of kinetic crystallography, other methods have been applied to understand biomolecular mechanisms. In this respect intrinsic physical properties of the biopolymers or specifically conjugated reporter molecules, such as fluorescent dyes, have been utilized for visualization of and monitoring catalytic activity. The attached reporter molecule must be chosen such as it does not interfere with the structure of a biopolymer or hinder the interaction with other molecules. Molecules modified in such a manner are then suitable for *in vitro* and *in vivo* monitoring of biomolecular events. Two highlights of this visualization technique were the observations of myosin movement relative to actin (Finer *et al.*, 1994) and the rotation of F1-ATPase (Noji *et al.*, 1997).

Hence, specific labeling of proteins, nucleic acids or other biomolecules is of particular interest to scientists working in biophysics, biochemistry and molecular biology. Bearing in mind the size of biopolymers and the different reactivities of their monomeric units, this is a challenging task. In addition to the labeling of enzymes, there is great interest in the labeling of DNA, the cellular carrier of genetic information. A specific challenge is the sequence-specific labeling of DNA.

1.1 Sequence-specific recognition and labeling of DNA

DNA carries the genetic information of life and thus is responsible for the unity between members of successive generations. The special double helical structure of DNA enables an organism to preserve and pass a stable pool of information from one generation to the next allowing the organism to maintain its identity relatively unchanged over millions of years. DNA comprises two complementary polynucleotide molecules which are wound around each other in an antiparallel fashion forming a double helix. The four nucleobases adenine (A), thymine (T), guanine (G) and cytosine (C) are stacked perpendicularly to the helical axis and point to each other forming the so called Watson-Crick base pairs (bp) in which A forms hydrogen bonds with T and G with C. Additional stability of the biopolymer is achieved by base stacking of adjacent bases. The twist of the backbone defines a major and a minor groove whereby the sugar-phosphate backbone defines the diameter of the grooves and the edges of the bases act as a floor. DNA exists in diverse double helical conformations like the A-, B- and Z-form DNA. Different parameters like water content, salt concentration and even sequence of the nucleobases define the final three-dimensional shape of DNA (Fig. 1.1). However, the B-form is the most relevant one in biological systems and will be considered for further discussions. The information content of DNA is encoded in the sequence of the four nucleobases. Evolution has produced specialized enzymes which can copy the DNA or transfer the stored information into other biopolymers. DNA polymerases replicate DNA by using one polynucleotide strand as a template for the synthesis of the complementary daughter DNA strand. RNA polymerases transcribe DNA into mRNA which is then translated on the ribosome into a protein. In viruses, reverse transcriptases form DNA from RNA. This flow of biological information is expressed in Scheme 1.1.

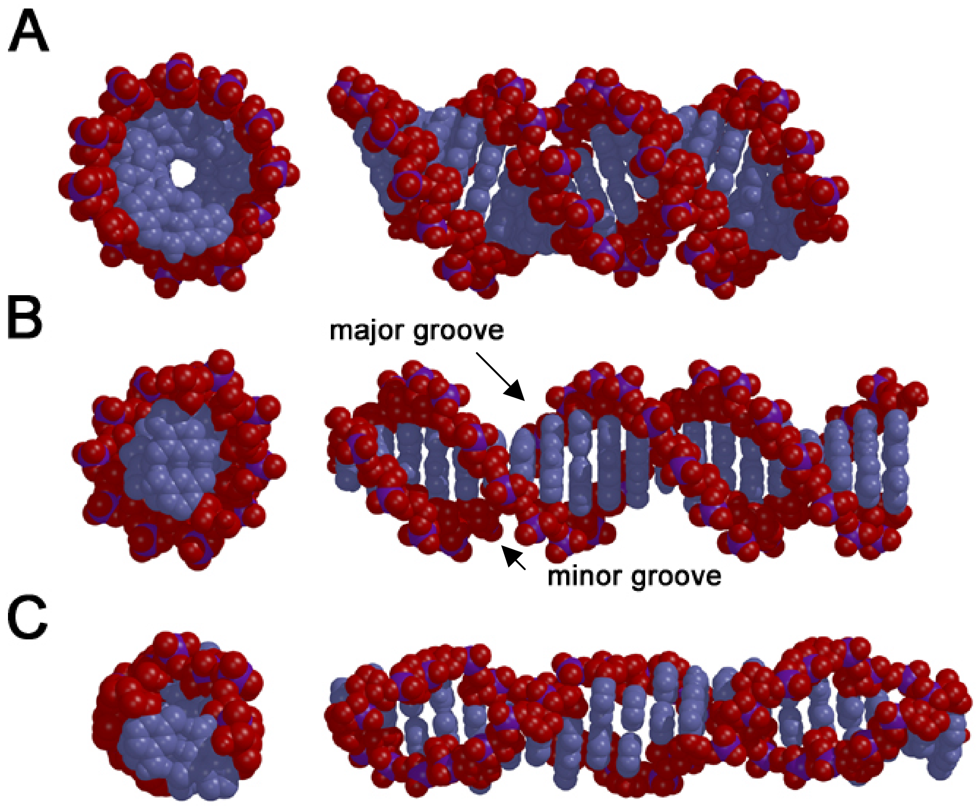


Fig. 1.1: Three ideal structures of A-DNA (A), B-DNA (B) and Z-DNA (C). The structures were generated using the program NAMOT and are composed of 20 bp in each case. Red spheres denote the sugar phosphate backbone in which phosphorous is highlighted in magenta. The bases are colored in light blue.

Scheme 1.1: Flow of biological information.

Throughout all organisms there exists a single genetic code which correlates the sequence of the nucleobases with the polypeptide sequence. A corresponding code for the recognition of DNA sequence by proteins has not yet been found. Specific DNA sequences are obviously recognized by different proteins in an idiosyncratic manner. Thus, a sequence-specific recognition of DNA is essential for reading out information which demonstrates the importance of sequence-specificity.

In order to label DNA covalently in a sequence-specific manner it is necessary first to recognize DNA specifically. This sequence-specificity is mostly obtained by specific interactions with the nucleobases because they define this unique storage system.

1.1.1 Binding of small molecules to DNA

Generally, four different binding modes of small molecules with duplex DNA can be distinguished. Firstly, non specific binding to the phosphate backbone *via* electrostatic interaction. Secondly, specific binding of molecules to either groove of DNA by hydrogen bonding to the nucleobases. Thirdly, specific intercalation of aromatic planar molecules between two base pairs *via* Van der Waals interactions. A fourth mode is a specific binding to DNA as a result of a special DNA conformation caused by the nucleobase sequence.
Since it is known that information in DNA is stored in its base sequence, it is of great interest to target sequences specifically, thus, the latter three binding modes are of interest. In order to target a unique site within the human genome a minimum of 15-16 contiguous base pairs have to be addressed (Dervan, 1986), which demonstrates that the size of the targeted sequence is essential for specificity. Several approaches have been examined to address the question of sequence-specific recognition of DNA by small molecules. The most important of those will be shortly presented.
An obvious approach to sequence targeting is to make use of the common Watson-Crick base pairing rules. For that purpose novel nucleic acids have been designed which contain the natural nucleobases but consist of a modified backbone. The peptide nucleic acids (PNAs), consisting of a peptide backbone, are of particular interest as they form stable hybrids with DNA and RNA. As the hybrids show even greater stability than homogenous DNA duplexes these novel nucleic acid analogues may seem interesting for strand exchange and thus for specific recognition of DNA. However, the natural DNA structure is disrupted by this way of targeting. In order to avoid drastic structural changes of DNA, other methods have to be applied.
In the early 1960s Lerman found that planar aromatic molecules can intercalate DNA (Lerman, 1961) and a great effort has been made to find intercalators which specifically bind DNA for a range of different purposes. Classical intercalators place themselves between two base pairs lengthening the B-form DNA and showing a preference for 5’-pyrimidine-purine-3’ steps (Braña *et al.*, 2001). The fluorescent intercalator ethidium bromide (**1**) for example is still widely used for visualization of DNA in agarose gels (Fig. 1.1.1a). The fluorescence of the ethidium compound is even increased while stacked within DNA, however ethidium bromide doesn’t show a sequence-specificity.

A **B**

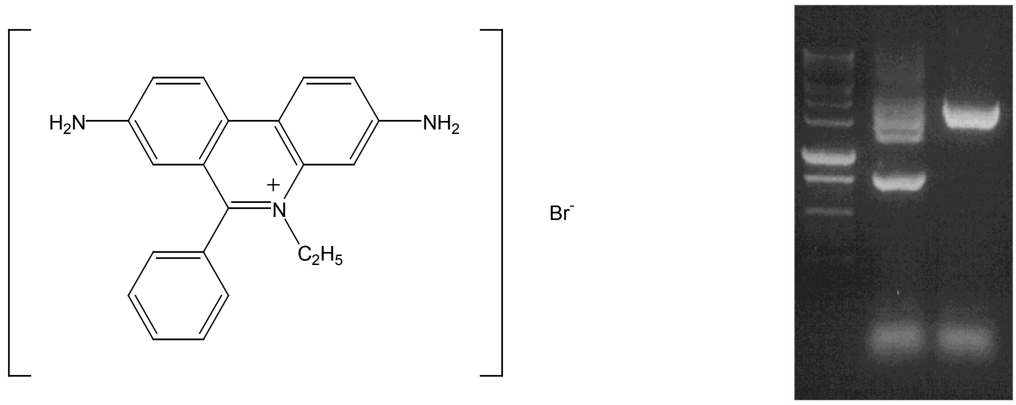


*Fig. 1.1.1a: Classical DNA intercalator ethidium bromide (**A**) and typical agarose gel after electrophoresis and staining of DNA with ethidium bromide (**B**).*

Intercalating agents inhibit polymerases and topoisomerases and since the first antibiotics and antitumor agents useful for chemotherapy were found among them, the search for new intercalators has made great progress. By creating *bis*- and *poly*-intercalators connected by different linkers specific sequences could be targeted. Bailly and co-workers have shown that nitro- and amino-substituted elinafide, a bisnaphthylimide, with antitumor activity binds sequence-specifically to DNA (Bailly *et al.*, 1996). A recent example of a sequence specific intercalator was reported by the Iverson group. The authors present a peptide linked *bis*-intercalator recognizing d(GGPyPuCC)$_2$ (Guelev *et al.*, 2001), where Py stands for pyrimidine and Pu for purine. The specificity is obtained by a peptide linker which lies in the major groove. Iverson and co-workers previously described an altered specificity by modifying the peptide linker (Guelev *et al.*, 2000) opening possibilities to target different specific sequences. Utilizing a fluorescent intercalator, one could imagine sequence-specific labeling, which however, considerably modifies the structure of B-form DNA and is not covalent. Furthermore, the binding constants of intercalators are in the high micromolar range which represents a rather weak binding compared to protein-DNA interactions. Nevertheless, the binding of the *bis*-intercalator presented by the Iverson group is obviously due to peptide linker-major groove contacts. This suggests that a single intercalator can be used to recognize two base pairs but rather that recognition of multiple base pairs is obtained by specific molecule-groove interactions.

The property of Watson-Crick base pairs to form additional hydrogen bonds in either groove of the DNA double helix (Hélène & Lancelot, 1982) and the discovery by Rich and co-workers that nucleic acids are able to form triple helices (Felsenfeld *et al.*, 1957) was used for sequence-specific targeting. Independently, the groups of Dervan in Pasadena (California) and

Hélène in Paris (France) targeted double stranded DNA with oligonucleotides which bind in the major groove (Moser & Dervan, 1987; Le Doan *et al.*, 1987). These triple helix forming oligonucleotides (TFO) bind preferentially to polypurine·polypyrimidine sequences. Besides the usual Watson-Crick base pairing within the double helix now a third strand binds *via* Hoogsteen hydrogen bonding in the major groove of DNA. The bases of the third strand combine in the so-called pyrimidine motif in which T binds to an AT-bp and a protonated C binds to a GC-bp (Fig. 1.1.1b). The orientation of the third strand is parallel to the purine rich strand as denoted by the plus minus symbols.

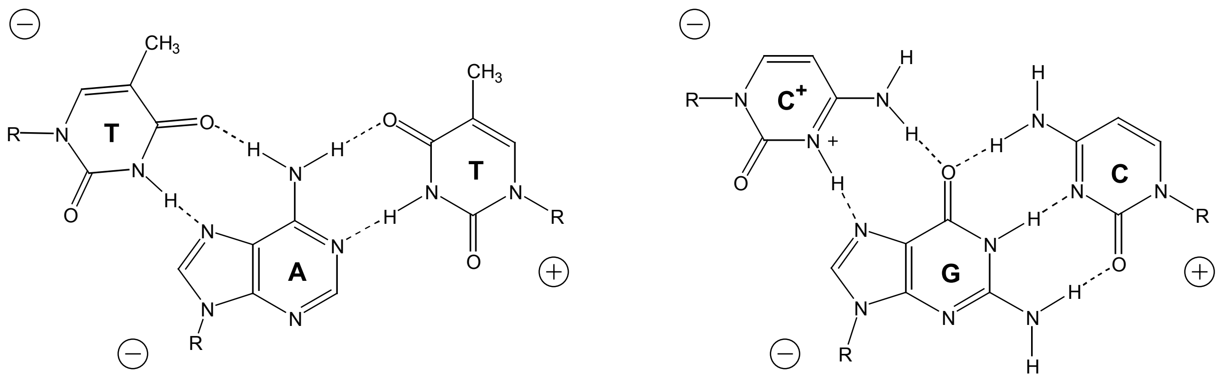


Fig. 1.1.1b: Hydrogen bond patterns formed within the pyrimidine motif.

Additionally to the pyrimidine motif, Beal *et al.* found the purine motif in which A binds to an AT-bp and G binds to a GC-bp (Fig. 1.1.1c). In the purine motif the third strand binds anti-parallel to the purine rich strand *via* reverse-Hoogsteen hydrogen bonding (Beal & Dervan, 1991).

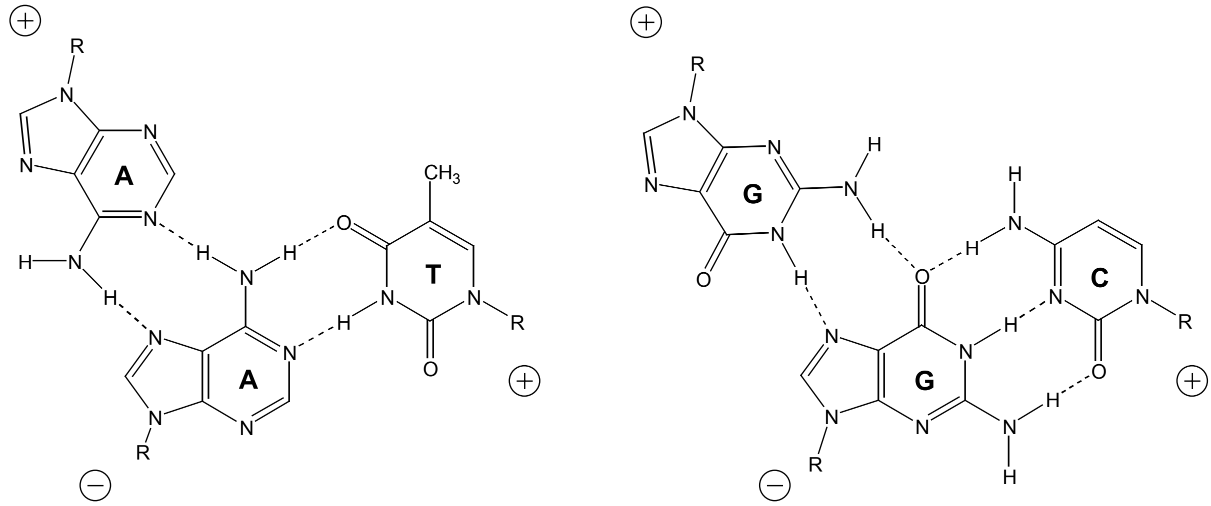


Fig. 1.1.1c: Hydrogen bond patterns formed within the purine motif.

Finally, there is a mixed pyrimidine-purine motif in which GGC and TAT triples are formed. However, in spite of the different motifs the TFO approach remains limited to purine rich sequences. Further, the formed triple helices are not very stable to temperature and pH changes. These problems were partly overcome by introducing other nucleobases in the TFO's like 5-methylcytosine instead of cytosine (Lee *et al.*, 1984). Further stabilization of the triple helix can be obtained by the use of ribonucleotides instead of deoxyribonucleotides for the TFO (Escudé *et al.*, 1992; Shimizu *et al.*, 1992) or the use of intercalators which specially bind to triple helices (Escudé *et al.*, 1998). The main purpose of all these triple helical structures was to inhibit protein-DNA interactions, especially on promoter sequences as proposed in the anti-gene strategy (Hélène, 1991). This was intended to be an alternative to the anti-sense strategy (Hélène & Toulmè, 1990). However, results from the Dervan group with bifunctional molecules consisting of a TFO as sequence recognition element and an attached DNA alkylating agent such as *N*-bromoacetyl, which alkylates the N7-position of guanine in a sequence-specific manner (Povsic & Dervan, 1990; Povsic *et al.*, 1992), suggest that TFO's can be used as labeling tools. Also the recently presented results from Parel *et al.* who used the fluorescent nucleobase 2-aminopurine within a TFO (Parel & Leumann, 2001), demonstrate that this approach can be used to specifically label purine rich sequences. However, the complete TFO represents a very large label and thus the anti gene strategy proposed by Hélène is probably a more suitable application of TFO's, unless new strategies are developed.

A very stable form of triplexes can be obtained with PNA. Interestingly, PNA does not preferentially form classical triplexes with double-stranded DNA as described above but a triplex invasion complex (2PNA/DNA; Nielsen & Egholm, 1999). Here a homo-purine, double-stranded DNA can be targeted most effectively with *bis*-PNA connected with an ethylene glycol type linker. The homo-purine strand of DNA forms Watson-Crick hydrogen bonds with one PNA strand and Hoogsteen hydrogen bonds with the other. The other DNA strand does not form any hydrogen bonds in the targeted sequence. Therefore, a triplex invasion complex can only be formed after opening of the DNA double helix. Once formed, the complex does not dissociate for hours which emphasizes its stability. The chemical structure of a PNA molecule and a triplex invasion complex are shown in Fig. 1.1.1d.

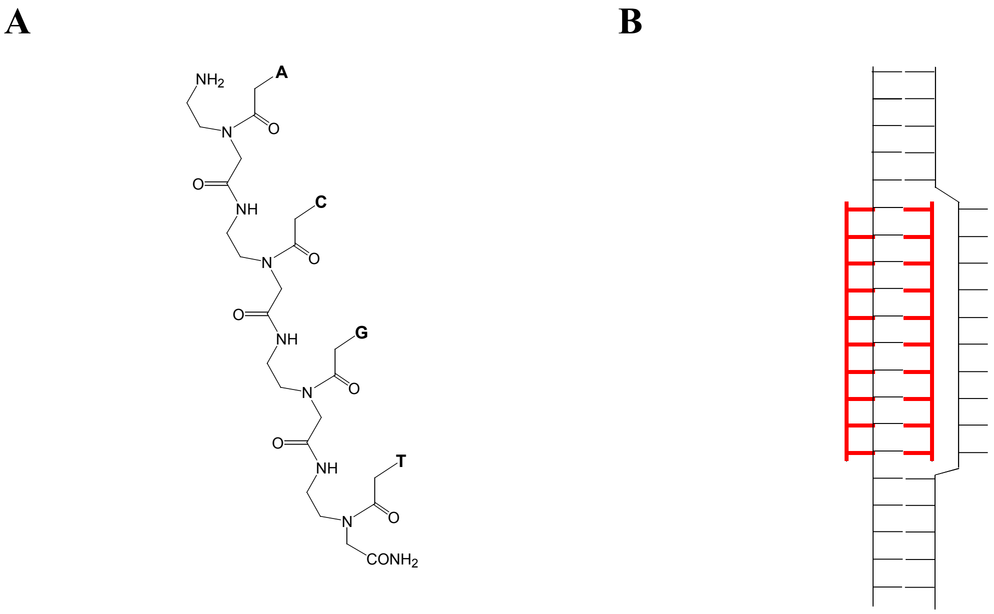


*Fig. 1.1.1d: Chemical structure of a PNA molecule (**A**) and a schematic representation of a triplex invasion complex (**B**). The DNA molecule is black, the PNA molecule is red.*

In addition to the major groove of DNA, the minor groove can also be targeted, as the nucleobases exhibit free valences therein. The possibilities of the different base pairs to form additional hydrogen bonds in the minor groove are shown in Fig. 1.1.1e.

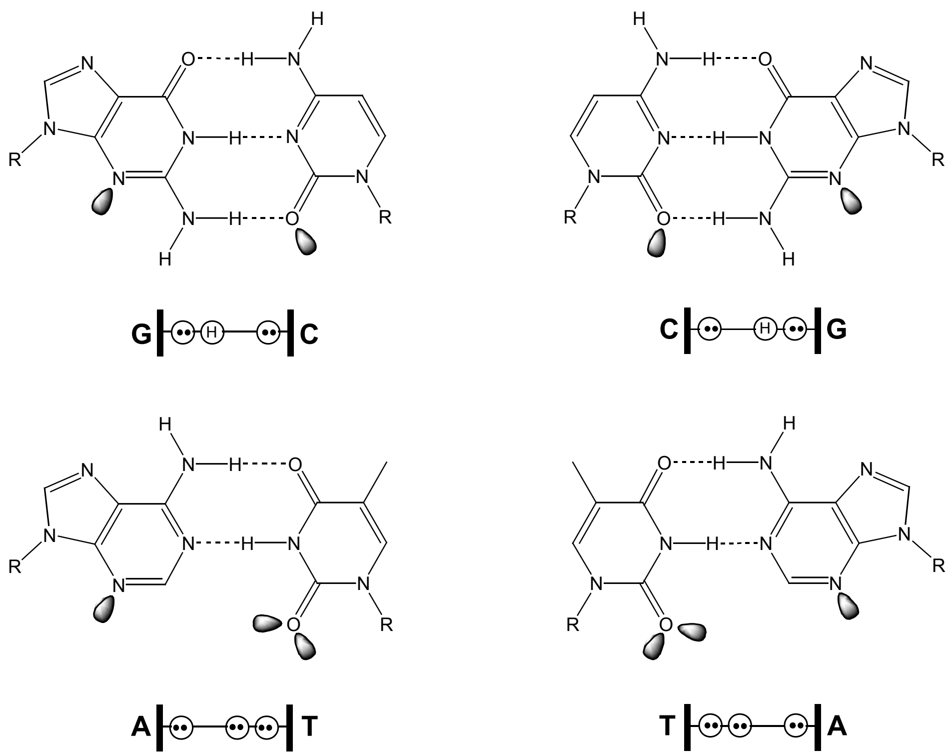


Fig. 1.1.1e: Free valences of Watson-Crick base pairs in the minor groove (adapted from Dervan & Bürli, 1999)

The nucleobase pairs either display one hydrogen bond donor (2-H of guanine) and two acceptors (GC and CG pair) or three hydrogen bond acceptors (AT and TA pair). Several natural products like chromomycin, actinomycin D or distamycin bind in the minor groove. Only distamycin and the structurally related netropsin, two oligopeptides containing *N*-methylpyrrole (Py) units, consist of repeating building blocks which are fairly simple to mimic (Fig. 1.1.1f, A). This led the group around Dervan to investigate the so-called lexitropsins (information reading oligopeptides). The convex structure of distamycin and related molecules fits nicely in the concave minor groove suggesting a structure based recognition (Fig. 1.1.1f, B; Coll *et al.*, 1987). However, NMR studies on distamycin in complex with DNA, revealed that distamycin can bind in the minor groove of DNA with either one or two molecules (Pelton & Wemmer, 1989; Chen *et al.*, 1994). The two distamycin molecules are arranged in an anti-parallel fashion and pairwise arrangement of the Py units allows recognition of a specific base sequence.

A **B**

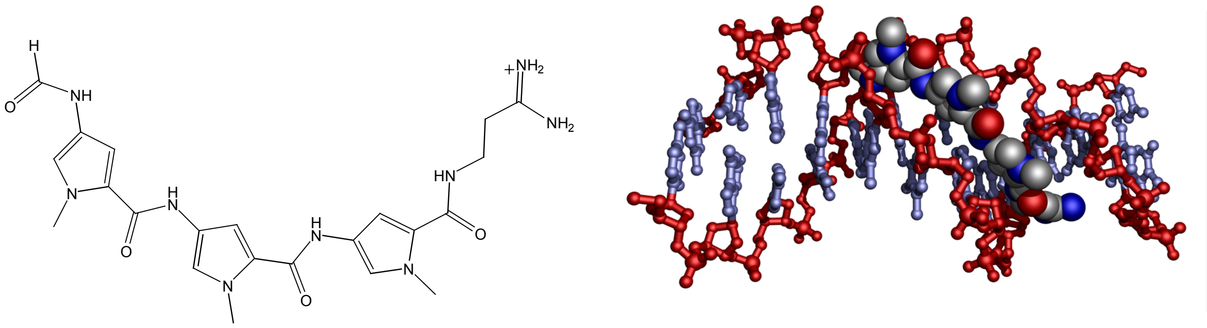


Fig. 1.1.1f: Distamycin (A) and crystal structure of distamycin in complex with DNA (B).

Based on these observations, Dervan and co-workers developed the hairpin polyamides which bind specifically the DNA minor groove. Mimicking distamycin, the hairpin polyamides are composed of Py units connected *via* amide bonds. Two polyamide oligomers are connected by γ-aminobutyric acid (Mrksich *et al.*, 1994). This arrangement allows two Py units to bind pairwise to one base pair in the minor groove. In addition to the Py units present in the natural lexitropsins, Dervan *et al.* introduced the aromatic amino acids *N*-methylimidazole (Im) and 3-hydroxy-1-methylpyrrole (Hp) units in order to be able to distinguish all four possible base pair combinations (White *et al.*, 1998). Extending this approach, the Dervan group introduced β-alanine instead of a Py/Py pair at appropriate positions to adjust the curvature of the polyamides to the minor groove curvature (Trauger *et al.*, 1996; Wang *et al.*, 2001) in order to recognize larger DNA sequences. A model of the binding pattern of hairpin polyamides to

DNA is shown in Fig. 1.1.1g. The carboxamide hydrogens make specific contacts to N3 of A and O2 of T, thus, recognizing AT and TA base pairs, respectively. The nitrogen of *N*-methylimidazole allows interaction with the 2-NH_2 of guanosine, thus, recognizing GC base pairs. The 3-hydroxy-1-methylpyrrole unit distinguishes an AT from a TA base pair because the additional hydroxyl group favours T over A for steric reasons and forms an additional hydrogen bond to O2 of T. Therefore, the appropriate pair-wise arrangement of the different aromatic amino acids allows recognition of specific sequences. This approach to sequence-specific recognition of DNA *via* the minor groove represents a powerful tool to target specifically a given DNA sequence, especially when the hairpin polyamides are combined with triple helix forming oligonucleotides extending its versatility to target both the minor and the major groove (Szewczyk *et al.*, 1996a; Szewczyk *et al.*, 1996b). Similar to the TFO approach, the polyamide approach can be used to target special sequences of promoters. However, the latter is more suitable for targeting the most important transcription factor recognition sequences (Dervan, 2001). In terms of labeling, the attachment of an alkylating agent to hairpin polyamides, leads to sequence specific alkylation of the N3-position of adenine (Tao *et al.*, 1999; Wurtz & Dervan, 2000). However, like the TFO, the polyamides represent quite a big label. In addition, they bind in the minor groove of DNA and thus may seem more suitable for blocking DNA binding proteins than solely for labeling. An advantage of both polyamides and oligonucleotides in comparison to the intercalating agents, is their high binding affinity to DNA which is in the nanomolar to picomolar range.

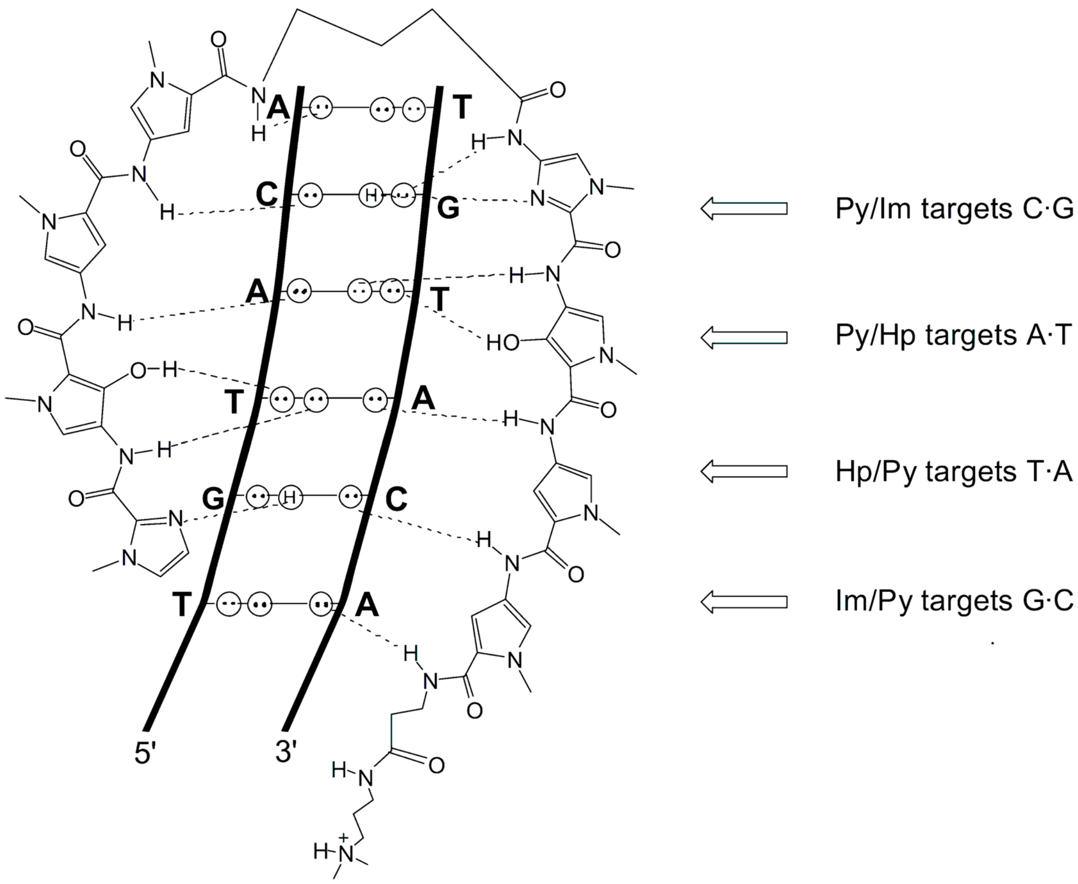


Fig. 1.1.1g: Hydrogen bonding pattern of a hairpin polyamide to a DNA hexamer (adapted from Dervan & Bürli, 1999).

Although enormous progress has been made in the field of sequence-specific recognition of DNA during recent decades, chemists are just starting to understand and make use of the possibilities provided by these approaches. However, none of the methods for sequence-specific recognition is yet suitable for sequence-specific labeling of DNA with small reporter groups.

1.1.2 Labling of DNA

In principle, modified oligodeoxynucleotides are prepared by either chemical or enzymatic methods. The standard chemical method to synthesize specifically labeled short oligonucleotides is the phosphoramidite solid-phase synthesis of DNA designed by Letsinger and improved by Caruthers (Caruthers *et al.*, 1987). The method is used in state-of-the-art DNA synthesizers and can be applied for the incorporation of modified phosphoramidites into short oligonucleotides. The modified phoshoramidite can either already contain the reporter group or the phosphoramidite consists of a linker with a terminal functional group which can be modified with different reporter groups postsynthetically. The latter procedure is still a popular route which is revealed by a recent publication by Sheppard and co-worker (Dey & Sheppard, 2001). Other chemical modifications of nucleic acids can be achieved, by bisulfite-catalyzed transaminations of the N4-amino group of cytosines, bromination of the C5-position of pyrimidines and the C8-position of purines followed by reactions with derivatized amines or the photochemical reaction of azides like photobiotin with nucleotides, to name just a few (Kessler, 1992). Apart from the solid-phase synthesis of DNA which is limited to 50-100 monomeric units in reasonable yields, the chemical methods are not sequence specific.

Enzymatic labeling methods make use of DNA-polymerases or terminal transferases. Homogenous labeling can be achieved by incorporation of modified nucleoside triphosphates (NTP) using Klenow polymerase (random primed labeling), *Escherichia coli* DNA polymerase I (nick translation) or *Taq* DNA polymerase (Polymerase chain reaction, PCR). In the latter case, the modified nucleotides can either be incorporated directly into DNA or incorporated within chemically synthesized primers. The use of terminal transferases for labeling can result only in end-labeled DNA. The trivial prerequisite for all these enzymatic labeling methods is that the modified NTP is recognized and incorporated by the enzymes. Unfortunately, none of these methods are sequence specific. The only possibility to label large DNA fragments specifically would be a tedious combination of solid-phase synthesis of a short modified oligonucleotide which could be incorporated enzymatically into DNA using ligases. However, none of these different labeling methods allow a simple sequence-specific labeling of large DNA.

1.2 DNA binding proteins

Nature has evolved many different proteins which interact with DNA. Essential processes require the recognition of specific DNA sequences by a variety of proteins. Many of those exert their effect by binding to appropriate locations and in some cases changing the DNA conformation. Others, like polymerases, nucleases, glycosylases, integrases, recombinases or DNA methyltransferases covalently modify DNA. All of them specifically recognize DNA by different protein-DNA interactions. Most of the proteins interact *via* the major groove with DNA because the major groove provides more space than the minor groove to the incoming protein and the free hydrogen bond valences within the major groove make an unequivocal differentiation between the different base pairs possible. The best characterized structural motifs of proteins for specific DNA recognition, the helix-turn-helix motif, the zinc finger motif and the leucine zipper motif, will be briefly presented.

The Helix-Turn-Helix motif (Fig.1.2, A) primarily found as a common recognition motif of procaryotic transcription activator/repressor proteins comprises two parallel α-helices which are connected by a β-turn region containing four amino acids of which glycine is in the second position. The loop region is not directly involved in DNA recogntion but either stabilizes the two helices relative to each other allowing interaction between the two or supports the formation of homo and heterodimers, respectively (Li *et al.*, 1995; Wilson *et al.*, 1995). The DNA recognition is achieved by one of the two α-helices which binds the major groove interacting sequence-specifically with the nucleobases by amino acid-nucleobase contacts.

The zinc finger motif (Fig.1.2, B), a predominant eucaryotic DNA binding motif, was discovered in *Xenopus* transcription factor IIIA (Klug & Rhodes, 1987). TFIIIA contains several zinc finger motifs which consist of two invariant Cys residues, two invariant His residues and several conserved hydrophobic residues. The invariant Cys and His residues coordinate tetrahedrally a Zn^{2+} ion. In some zinc fingers the invariant His residues are exchanged for two Cys residues and the Zn^{2+} ion is coordinated by four Cys residues. The motif is stabilized by the Zn coordination and a hydrophobic pocket formed inside the motif. Similar to the helix-turn-helix motif, the recognition of DNA occurs *via* insertion of an α–helix into the major groove of DNA. Usually, a zinc finger recognizes three base pairs where contacts with arginine and guanine residues seem to be essential (Pavletich & Pabo, 1991).

Monomeric *leucine zipper* (Fig.1.2, C) proteins comprise of two domains. One leucine rich α-helical domain and another mainly containing basic amino acid residues responsible for DNA binding. The second domain forms an α-helix when bound to DNA (Weiss *et al.*, 1990). A special feature of leucine zipper proteins is their ability to dimerize *via* the leucine-containing helices forming a coiled coil structure. The leucine residues of both helices interact with each other by hydrophobic interactions. The recognition of DNA occurs *via* the major groove.

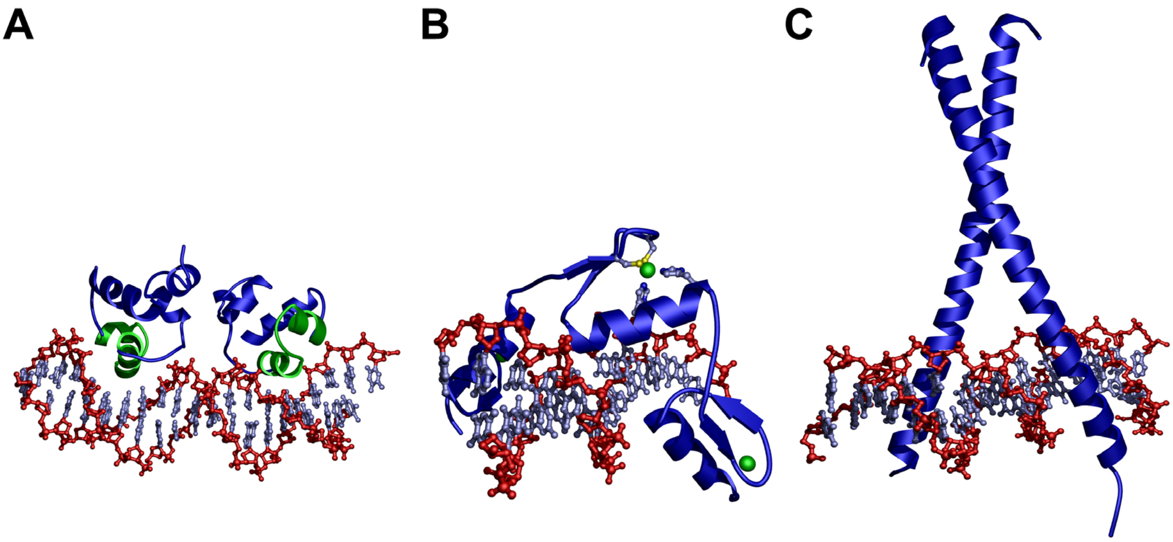


Fig. 1.2: Examples for the three best characterized DNA binding motifs.

A: Crystal structure of the 434-repressor in complex with DNA (Aggarwal et al., 1988). The Helix-Turn-Helix motif is presented in green.

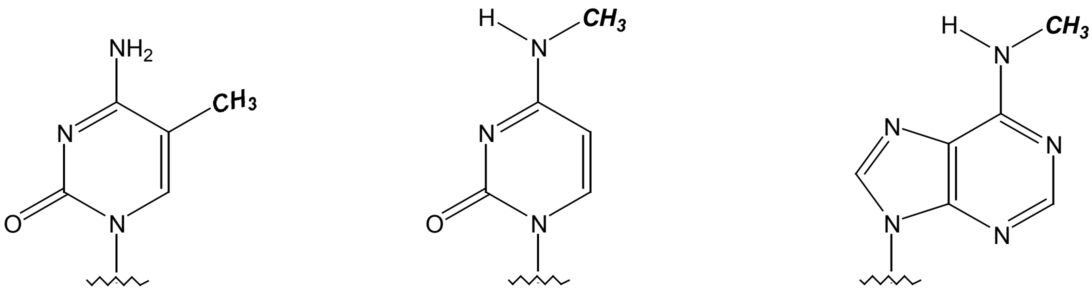
B: Crystal structure of Zif268 in complex with DNA (Pavletich & Pabo, 1991). There are three zinc finger motifs of which the inner one is shown. The Zn ions are presented in green.

C: Crystal structure of GCN4 leucine zipper (blue) in complex with DNA (Ellenberger et al., 1992).

Beside these motifs, many other ways for protein-DNA interactions have been discovered and with each solved protein-DNA complex structure the discovery of novel interactions is highly likely. In summary, there is no easy way to predict DNA sequence recognition by amino acid sequence.

1.2.1 DNA Methyltransferases

In addition to the prevalent four bases within DNA, adenine, cytosine, guanine and thymine, one also finds *N*6-methyladenine, *N*4-methylcytosine and *C*5-methylcytosine (Fig. 1.2.1), which increase the information content of DNA.



*C*5-Methylcytosine *N*4-Methylcytosine *N*6-Methyladenine

Fig. 1.2.1: Methylated bases within DNA.

These methylated bases arise from the action of another DNA binding protein, the DNA methyltransferases (MTases). DNA MTases catalyze the transfer of the activated methyl group of *S*-adenosyl-L-methionine (AdoMet) to either cytosine or adenine within non- or hemimethylated, double-stranded DNA. One distinguishes between *denovo* and maintainance methylation. The transfer of the methyl group to the nucleobase is performed in a sequence-specific fashion. DNA MTases are catagorized into two classes according to the atom that is modified. The N-MTases catalyze the methylation of the exocyclic amino group of either adenine or cytosine and the C-MTases catalyze the methylation of the C5 position of cytosine. In eucaryotes, generally C5-methylated cytosines are found whereas in prokaryotes *N*6-methyladenine and *N*4-methylcytosine are found additionally. The biological role of these methylated bases inside DNA is under intensive investigation and the functions elucidated so far exemplify the importance of DNA methylation in regulatory processes. In prokaryotes DNA MTases are primarily part of restriction-modification systems, where they are capable of protecting the host genome against cleavage by the endogenous restriction endonucleases. Furthermore, they play a crucial role in repair of base mismatches after DNA replication. By distinguishing between methylated parental strand and newly synthesized daughter strand the organisms are capable of repairing the DNA in favor of the parental strand. In eucaryotes methylation of cytosines, particularly of so called CpG islands, has diverse functions and was found recently to play a pivotal role in different human genetic diseases. The best characterized diseases are ICF (immunodeficiency, centromeric instability and facial anomalies) syndrome, the Rett syndrome and the fragile X syndrome. It was shown that a mutation in the DNA MTase 3B (DNMT3B) gene is responsible for the ICF syndrome (Hansen *et al.*, 1999; Okano *et al.*, 1999; Xu *et al.*, 1999). The Rett syndrome was shown to result from a mutation in the MECP2 gene, which encodes a protein that binds to methylated DNA (Amir *et al.*, 1999; Wan *et al.*, 1999). The fragile X syndrome, the most common form of an inherited mental retardation, arises from silencing of the FMR1 gene which is due to an aberrant *denovo* methylation of an upstream CpG island (Kremer *et al.*, 1991; Oberle *et al.*, 1991). In addition, there is evidence that DNA methylation has evolved as a genome defence mechanism. Apparently, methylation of transposons and retrovirus sequences within the mammalian genome effects their transcriptional silencing (Bestor & Tycko, 1996). Besides the recently gained knowledge about DNA methylation in diseases, DNA methylation was already associated with the phenomenon of genomic imprinting (Tycko, 1997). An immense amount of experimental results also suggests that methylation plays a major role in cell differentiation (Li *et al.*, 1992). Furthermore, there is a controversial issue about the role of hyper- and hypo-methylation in carcinogenesis (Chen *et al.*, 1998).

C5-DNA MTases are a homologous class of enzymes in which 10 conserved motifs could be identified (Lauster *et al.*, 1989; Posfai *et al.*, 1989; Kumar *et al.*, 1994). They are numbered from I to X utilizing the nomenclature of Posfai *et al.* (1989). Sequence comparison of N-DNA MTases with the C-DNA MTases and first three dimensional structural information have shown that both classes are structurally similar and that nine motifs, corresponding to motifs I-VIII and X defined for C-DNA MTases, can be assigned to the N-DNA MTases (Malone *et al.*, 1995). The motifs I, II, III and X are involved in cofactor binding and motifs IV-VIII have catalytic functions. Interestingly, the cofactor binding site is conserved among all DNA MTases and even other AdoMet-dependent MTases like RNA, protein or small molecule MTases (Schluckebier *et al.*, 1995; Djordjevic & Stock, 1997; Gong *et al.*, 1997). The N-DNA MTases are further classified in three groups based on the sequential order of the conserved motifs. An MTase belongs to the α-group when the order of the motifs corresponds to cofactor binding region, DNA recognition region and catalytic region. MTases of the β−group show the sequential order: catalytic region, DNA recognition region and cofactor binding region. The sequential order in the γ-group MTases is: the cofactor binding region, catalytic region and DNA recognition region.

1.2.2 DNA MTase from *Haemophilus haemolyticus* (M·*Hha*I)

The *Hha*I DNA MTase is part of the restriction-modification system of *Haemophilus haemolyticus* (Roberts *et al.*, 1976). It recognizes the sequence 5’-G**C**GC-3’ and catalyzes the methylation of the 5-position of the second cytosine residue. The gene of the MTase was cloned, sequenced and overexpressed in *E. coli* (Caserta *et al.* 1987; Wu & Santi, 1988). The enzyme consists of 327 amino acids and has a molecular weight of 37 kD.

The reaction mechanism of C5-DNA MTases is well understood. Wu & Santi (1987) were the first to propose that the mechanism of C5-DNA MTases is similar to that of thymidylate synthase where an enzyme cysteine thiolate attaches covalently to the 6-position of uracil thereby pushing the electrons to the 5-position which then can attack the methylene group of the methylene-THF cofactor. Such a mechanism was modified for C5-DNA MTases by Chen *et al.* (1993) who suggested that an enamine intermediate is favoured (Fig. 1.2.2a) instead of an anion intermediate as in the thymidylate synthase case.

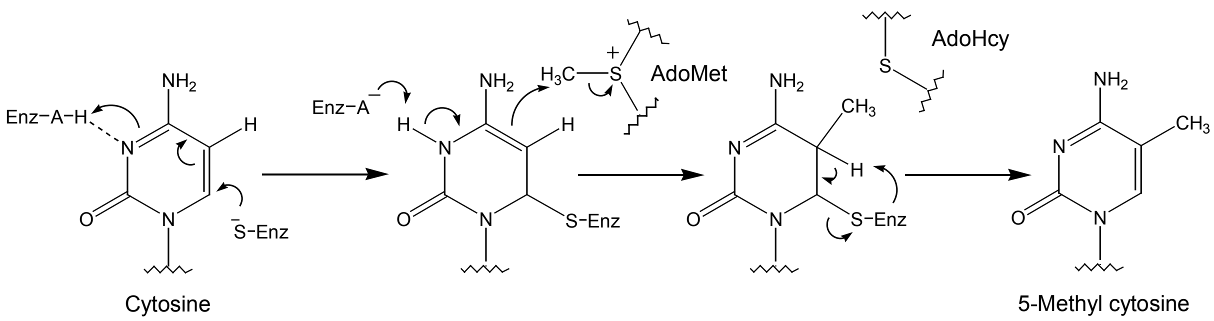


Fig. 1.2.2a: Mechanism of the C-DNA MTases catalyzed methylation reaction originally proposed by Santi et al.(1987) and modified by Chen et al. (1993).

By using 5-fluoro-2-deoxycytosine instead of the target cytosine Osterman *et al.* (1988) could trap the proposed covalent intermediate. The catalytic cysteine residue could be identified to be part of motif IV containing the invariant PC dipeptide. Additional evidence of the cysteine thiolate attack was provided with the solution of a co-crystal structure of M·*Hha*I with cognate DNA and cofactor (Klimasauskas *et al.*, 1994). Another striking feature observed in the crystal structure of Klimasauskas *et al.* (1994) was a novel mode of DNA distortion. The cytosine normally positioned in the center of the DNA double helix is flipped out from the DNA by 180° and positioned in the catalytic site of the enzyme (Fig. 1.2.2b). The place of the cytosine is taken up by a protein loop thus stabilizing the B-form DNA. This phenomenal base flipping mechanism was observed for the first time and due to the conserved sequences among C-DNA MTases base flipping was proposed for this class of enzymes. The crystal structure of another C-DNA MTase M·*Hae*III also shows a flipped out target base (Reinisch *et al.*, 1995) which supports the previous assumption. Additionally, evidence for base flipping of DNA repair enzymes (Vassylyev *et al.*, 1995) suggests that most DNA modifying enzymes acting on nucleobases flip out a base to gain access to their site of action.

A **B**

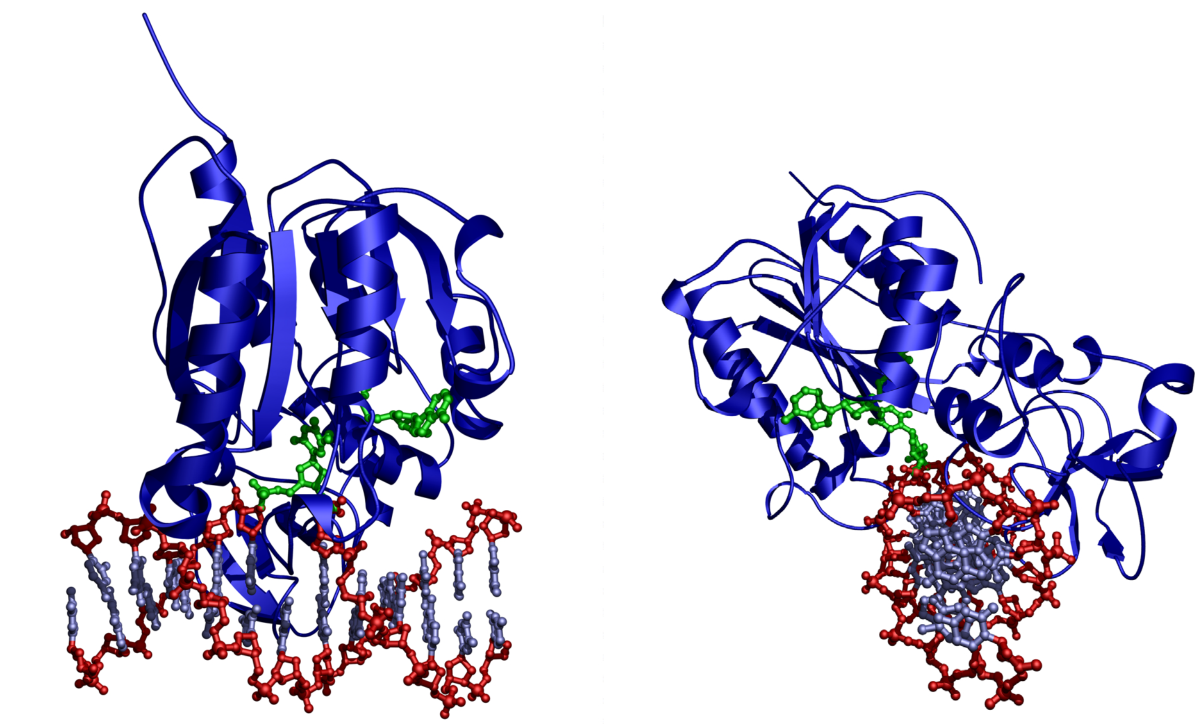


Fig.1.2.2b: Crystal structure of M·HhaI in complex with cognate DNA and cofactor AdoHcy (A). The same crystal structure rotated by 90° (B). The protein is represented in blue, DNA backbone in red, nucleobases in light blue and extrahelical cytosine and AdoHcy in green.

1.2.3 DNA MTase from *Thermus aquaticus* (M·*Taq*I)

M·*Taq*I is part of the restriction-modification system of *Thermus aquaticus*. The enzyme is a *N*6-adenine DNA MTase catalyzing the methylation of adenine within the 5'-TCG**A**-3' sequence of double-stranded DNA. The optimal temperature for maximal activity of M·*Taq*I is at 65 °C. Its gene was cloned, sequenced (Slatko *et al.*, 1987; Barany *et al.*, 1992) and overexpressed in *E. coli* (Labahn *et al.*, 1994). M·*Taq*I contains 421 amino acids and has a molecular weight of 47 kD. According to the arrangement of the nine conserved motifs the enzyme belongs to the γ-class of N-MTases.

The crystal structure of M·*Taq*I in complex with AdoMet (Labahn *et al.*, 1994) illustrates that the MTase consists of two domains which are connected by a loop. The catalytic domain contains the cofactor binding site and all nine conserved motifs. The two domains span a groove which was proposed to be the DNA binding site based on its size and positive surface potential. A base flipping mechanism similar to that of M·*Hha*I was also proposed for M·*Taq*I. Evidence came from different biochemical investigations such as a 2-aminopurine base flipping assay, photochemical crosslink studies and a permanganate assay (Holz *et al.*, 1998; Holz *et al.*, 1999; Serva *et al.*, 1998). This assumption was verified by a recent co-crystal structure of M·*Taq*I in complex with a cofactor analogue and a 10-mer substrate DNA (Goedecke *et al.*, 2001). The structure shows the flipped out target adenine and a new mode of protein-nucleic acid interaction (Fig. 1.2.3). Stabilization of the extrahelical adenine is obtained by partial occupation of the abandoned space by the orphaned thymine. The structure represents the first co-crystal structure of a N-DNA MTase with substrate DNA and gives the first insight into the mechanism of DNA N-methylation. In contrast to chemical methylation of adenine where the N1 position is methylated in the first place and the final product is generated after Dimroth rearrangement (Haines *et al.*, 1964), the enzyme directly methylates the 6-amino group of adenine. As suggested earlier (Malone *et al.*, 1995) the motif IV forms hydrogen bonds to the exocyclic amino group of adenine positioning it for an in-line attack on the methyl group of AdoMet. The structure answers some basic questions about the mechanism of DNA adenine N6-methylation whereas questions about the base flipping mechanism couldn't be answered and more biochemical investigations have to be performed.

A **B**

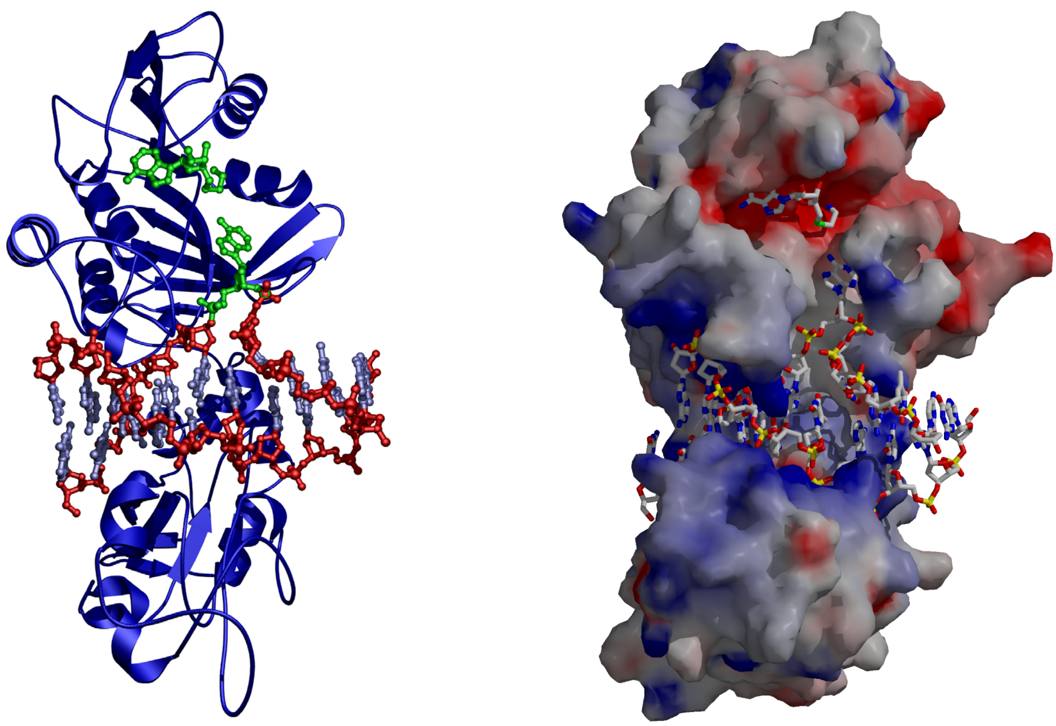


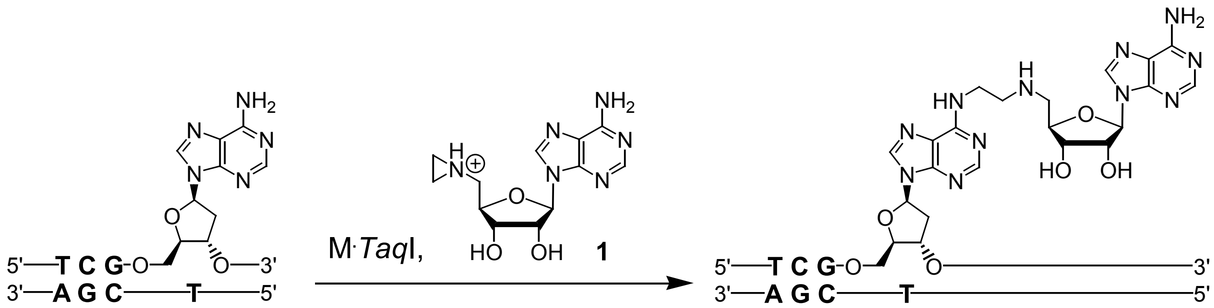
*Fig.1.2.3: Crystal structure of M·*TaqI* in complex with cognate DNA and a cofactor analogue AETA. (A) Representation of secondary structure elements (blue). Extrahelical adenine and cofactor analogue are represented in green. DNA backbone in red and other nucleobases in light blue. (B) Surface potential representation (red denotes negative electrostatic potential, blue positive electrostatic potential and white is neutral). DNA and cofactor are drawn in ball and stick representation.*

1.2.4 DNA MTase from *Bacillus centrosporus* (M·*Bcn*IB)

The DNA MTase from *Bacillus centrosporus* M•*Bcn*I was the first N4-methylcytosine DNA MTase to be discovered (Janulaitis *et al.*, 1983). It is part of a restriction-modification system and recognizes the sequence 5'-C**C**(C/G)GG-3' in which the second cytosine becomes methylated. The gene of M•*Bcn*I was cloned, sequenced and overexpressed in *E. coli* (Janulaitis *et al.*, 1982; Povilionis *et al.*, 1988). Two co-existing MTases, M•*Bcn*IA and M•*Bcn*IB, were found in the restriction-modification system, where M•*Bcn*IA catalyzes the methylation of both single- and double-stranded DNA whereas M·*Bcn*IB only catalyzes the methylation of double-stranded DNA (Merkiene *et al.* 1998). M•*Bcn*IB consists of 321 amino acids with a molecular weight of 37 kD. No three-dimensional structural information about the enzyme is available at the moment.

1.3 A Novel cofactor for the DNA MTase M·*Taq*I

As sequence-specific recognition is a very challenging task for chemists and the study of protein-DNA interactions does not disclose a general feature, one can think about utilizing the specificity of an enzyme to achieve the goal of sequence-specific DNA labeling. All DNA MTases utilize the common methyl group donor AdoMet as a cofactor for DNA methylation and achieve sequence-specific labeling with a single carbon unit. Obviously, the methyl group is not an attractive reporter group if one wants to avoid the application of radioactivity. Thus, new cofactors for DNA MTases which could transfer larger groups onto the DNA could be very interesting for labeling DNA sequence-specifically. Weinhold and coworkers introduced such a novel cofactor for the DNA MTase M·*Taq*I and showed for the first time that a DNA MTase is able to transfer more than just a single carbon unit (methyl group), namely the cofactor itself, onto DNA. The authors synthesized a cofactor analogue which carries instead of the methylsulfonium moiety in AdoMet an aziridine group. The new cofactor analogue *N*-adenosylaziridine (AZ, **1**) could be coupled in a M·*Taq*I-catalyzed reaction to the exocylic amino group of adenine within the recognition sequence of a 14-base pair duplex oligonucleotide (Pignot *et al.*, 1998, Scheme 1.3).



Scheme 1.3: Sequence-specific M·TaqI-catalyzed coupling of the novel cofactor AZ with DNA.

The authors suggest that the newly developed cofactor could potentially be used as a delivery system for different reporter molecules. For this purpose, they suggest an attachment of reporter groups to the 8-position of AZ (**1**).

1.4 SMILing DNA

The approach developed in the group of Weinhold was expanded during my diploma research project. The suggestions made in the publication about the new cofactor were investigated and a cofactor analogue which carries a fluorescent dye attached to the 8-position (Fig. 1.4) was synthesized. The dansylated cofactor analogue (**2**, DAZ) binds to M·*Taq*I (K_D = 0.55 µM) even stronger than the natural cofactor AdoMet (K_D = 2 µM) and is in analogy to the reaction presented by Pignot *et al.* (1998) coupled by M·*Taq*I to the target adenine within a 14-base pair duplex oligonucleotide (Pljevaljčić, 1999).

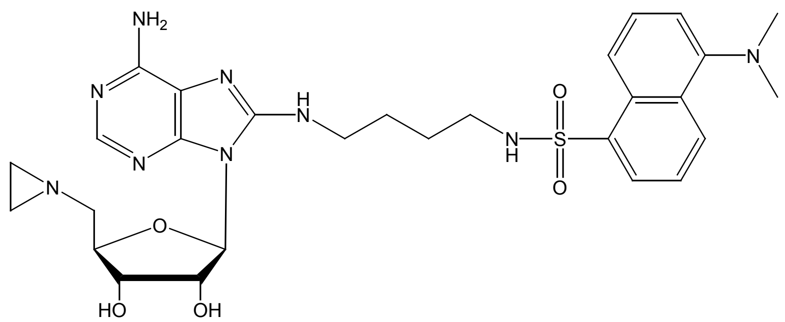


*Fig. 1.4: Dansylated cofactor analogue DAZ **(2)**.*

Thus, a new method for sequence-specific labeling of DNA using M·*Taq*I as catalyst was developed. In addition, this method has the potential to be a new general tool for sequence-specific DNA labeling.

CHAPTER 2

Aims of the project

2 Aims of the project

The aim of this project is to establish the general applicability of Sequence-specific Methyltransferase-Induced Labeling (SMILing). The technique was already introduced during my diploma thesis where fluorescence labeling of DNA with the dansylated aziridine cofactor DAZ was achieved. First of all, the coupling of *N*-adenosylaziridine (AZ) with DNA using the N6-adenine DNA MTase M·*Taq*I as catalyst should be optimized with respect to the pH value of the reaction buffer. Furthermore, the effect of the strong nucleophile DTT in solution should be investigated. Since, SMILing DNA was only shown for short duplex deoxyoligonucleotides the technique should be used to label plasmid DNA with the dansylated cofactor analogue DAZ.

Expanding this approach, short duplex deoxyoligonucleotides should be coupled with AZ using different DNA MTases. In order to show the general applicability of SMILing, representatives of the other two classes of DNA MTases should be used, namely, the C5-cytosine DNA MTase M·*Hha*I and the N4-cytosine DNA MTase M·*Bcn*IB. Additionally, the fluorescent cofactor analogue DAZ should be used with M·*Hha*I for the labeling of plasmid DNA to prove that the modification at the 8-position of the DAZ cofactor does not hinder the usage of the novel cofactor with other DNA MTases. Since the labeling of DNA with C5-cytosine DNA MTases is of special medical interest, as C5-cytosine methylation plays a major role in mammalian organisms, the reaction of M·*Hha*I should be optimized further regarding the pH value.

The development of a general synthetic route to 8-modified cofactor analogues for general usage in the SMILing technology should contribute to a quick extension of the possibilities using SMILing. Special focus should be directed to the synthesis of a biotinylated cofactor analogue (BAZ) which can be used for SMILing. Biotin plays such an important role because the very strong binding of streptavidin-conjugates to biotin allows the detection of labeled species in the picomolar range. First of all, a biotinylated cofactor analogue should be used for the M·*Taq*I-catalyzed labeling of a short duplex oligonucleotide as a prove of principle. A crystal structure of M·*Taq*I, DNA and BAZ should reveal the specific labeling of DNA. Further, the labeling of plasmid DNA should be achieved with this cofactor analogue. The specificity of the reaction and the amount of labeled DNA should be determined with the biotinylated cofactor analogue.

Additionally, as an example SMILing DNA should be used for the selection of deoxyribozymes by internal functionalization. In collaboration with the group of Dr. Andres

Jäschke (FU Berlin, Germany), a pool of DNA sequences should be labeled with biotin. The labeled DNA sequences should be amplified by PCR. This would reveal that the PCR reaction can be applied on DNA modified by SMILing and at the same time would show that SMILing can be used for the internal functionalization of DNA. Such functionalized DNA would then be suitable for the selection of DNA catalyzed chemical reactions.

Finally, crystallographic methods should be applied to determine a ternary complex of M·*Taq*I, DNA containing 2-aminopurine and a cofactor analogue. With this, first structural information would be obtained about a flipped out 2-aminopurine base, which provides an assay for base flipping enzymes. Thus, a statement about the quality of the base flipping assays using 2-aminopurine could be made. These base flipping assays could appear important in the search for new DNA MTases which again are essential for expanding SMILing DNA.

CHAPTER 3

Results and Discussion

3 Results and Discussion

3.1 Coupling of novel cofactors with DNA using different DNA methyltransferases

A review of the development of the novel cofactors is essential for the understanding of the ideas behind the SMILing technique. The new cofactors are based on the natural cofactor *S*-adenosyl-L-methionine (**3**, AdoMet). The cofactors were developed using the N6-adenine DNA MTase M·*Taq*I as a model system. Since the cofactor binding site of DNA MTases is conserved among DNA MTases (Schluckebier *et al.*, 1995), for a general application it was necessary to maintain the important features of the cofactor. The adenosyl moiety, for example, was found to be fundamental for binding, since it acts as a molecular anchor for cofactor binding by DNA MTases (Pignot *et al.*, 2000). The amino acid side chain can, however, be neglected for binding. The basic idea leading to the novel cofactors is illustrated in Fig. 3.1. The reaction mechanism of M·*Taq*I for the naturally occurring methyl group transfer involves a nucleophilic attack of the exocyclic amino group of the target adenine to the methyl group of **3**. The amino group is represented in Fig. 3.1 only as a nucleophile (Nu). The activated methyl group of **3** is transferred to the amino group and the methylation reaction is completed by release of a proton from the methyl ammonium group.

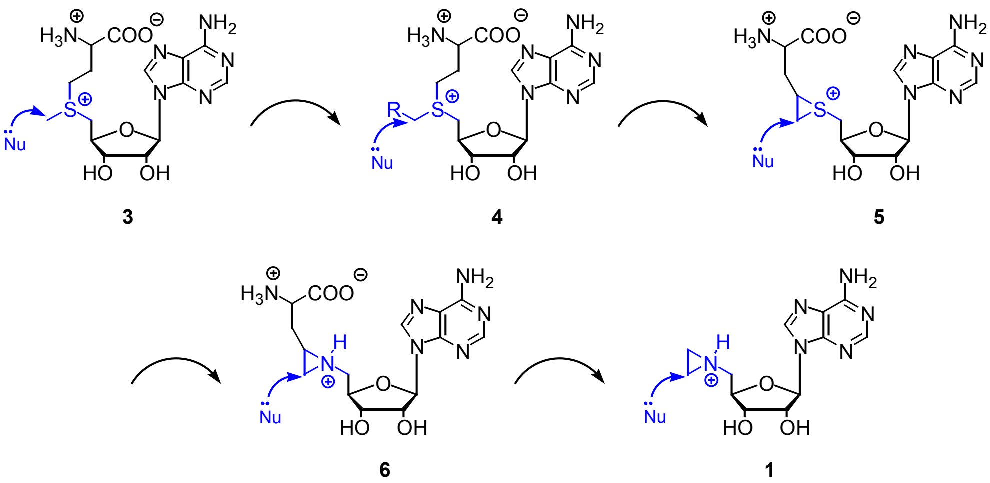


Fig. 3.1: Idea towards the N-adenosylaziridine cofactor.

It might be expected that use of a cofactor analogue **4** with for example, an ethyl group at the sulfonium center would lead to ethyl group transfer to the nucleophile in a M·*Taq*I-catalyzed reaction. However, *S*-adenosyl-L-ethionine is known to be an inhibitor of DNA MTases (Cox & Irving, 1977). This can be understood on the basis of inspection of the crystal structure of M·*Taq*I in complex with AdoMet. Around the methyl group of the cofactor, there is not enough space to accommodate additional atoms, which makes this approach to transferring more than a methyl group impossible. In order to circumvent spatial limitations, a connection from the methyl group to Cγ of the side chain would lead to the thiiranium compound **5**. Such a cofactor for M·*Taq*I could be completely transferred *via* a ring opening reaction to the target nucleophile. However, thiiranium compounds are unstable in nucleophilic solvents (Smit *et al.*, 1975). For this reason the more stable aziridine moiety as in **6** was chosen. This modification was also used with aziridino-glutamate to act as an alkylating agent in an enzyme-catalyzed reaction (Tanner & Miao, 1994). In order to simplify the synthesis, the amino acid side chain of AdoMet was not included and *N*-adenosylaziridine (AZ) **1** was chosen as the cofactor analogue. As previously mentioned, Pignot and co-workers showed that **1** acts as a cofactor for M·*Taq*I and that it could be specifically coupled in an enzyme-catalyzed reaction to a short duplex oligonucleotide (Pignot *et al.*, 1998). A fluorescent derivative carrying a tethered dansyl fluorophore at the 8-position of AZ was shown to couple to the same duplex oligonucleotide (Pljevaljčić, 1999). This was the beginning of the Sequence-specific Methyltransferase-Induced Labeling technique refered to as SMILing DNA. Herein, the 8-position of the cofactor plays a major role in the design of these new cofactors. The crystal structure of M·*Taq*I in complex with AdoMet reveals that the 8-position points towards the solvent and interference of an attached fluorophore at this position with the protein is not very likely. Since SMILing should be generally applicable for most DNA MTases this prerequisite has to be fullfilled also for other enzymes. Actually, the known crystal structures of various DNA MTases in complex with their cofactors show that the 8-position of the cofactor is accessible from the solvent (Table 3.1a).

Table 3.1a: Different types of DNA MTases.

Name	Substrate	Complex	8-Position of the cofactor	Reference
M·*Taq*I	N6-adenine	AdoMet	accessible	Labahn *et al.*, 1994
DpnM	N6-adenine	AdoMet	accessible	Tran *et al.*, 1998
M·*Rsr*I	N6-adenine	MTA[b]	accessible	Scavetta *et al.*, 2000
M·*Hha*I	C5-cytosine	AdoHcy[a] and DNA	accessible	Klimasauskas *et al.*, 1994
M·*Pvu*II	N4-cytosine	AdoMet	accessible	Gong *et al.*, 1997

[a] *S*-Adenosyl-L-homocysteine; [b] 5'-deoxy-5'-methylthioadenosine.

As the cofactor binding pocket is conserved among DNA MTases (Schluckebier *et al.*, 1995), it might be expected that most DNA MTases accept the novel cofactors. Interestingly, the cofactor binding site is even conserved among other AdoMet-dependent MTases. Therefore, the novel cofactors could also be suitable for labeling other substrates. A selection of different AdoMet-dependent MTases with known crystal structures is listed in Table 3.1b.

Table. 3.1b: Different AdoMet dependent MTases

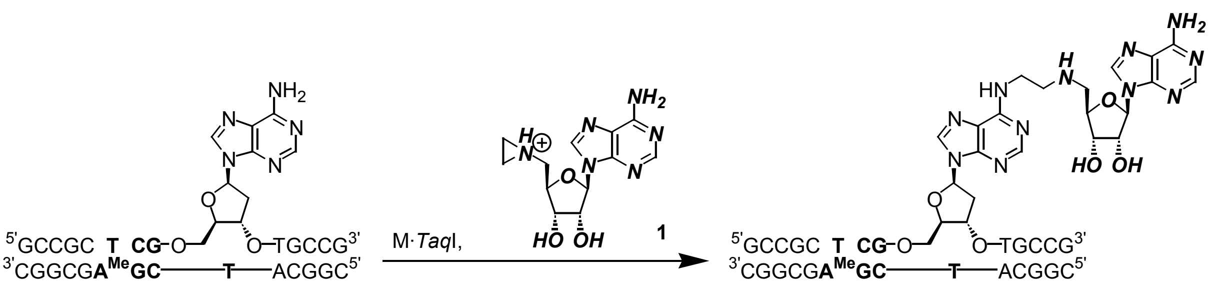
Name	Substrate	Complex	Accesible position of the cofactor	Reference
RNA MTases				
VP39	2'-OH of m7GpppG/A	AdoMet	8-position	Hodel *et al.*, 1996
ErmC'	N6-adenine in 23S rRNA	AdoMet	7-position	Bussiere *et al.*, 1998
Protein MTases				
CheR	glutamic acid	AdoHcy	8-position	Djordjevic & Stock, 1997
Small molecule MTases				
COMT	catechol	AdoMet	None	Vidgren *et al.*, 1994

Table 3.1b demonstrates that various AdoMet-dependent MTases for different substrates possess an accessible 8-position. In other cases, the 7-position seems to be more suitable. Therefore, the SMILing technique should not only be applicable to the N6-adenine DNA MTase M·*Taq*I but also to other DNA MTases, and even other classes of AdoMet-dependent MTases should accept the novel cofactors. This would imply a possible extension of SMILing DNA to SMILing of many other biomolecules, since AdoMet-dependent MTases exist for

different kinds of biomolecules ranging from small molecules to proteins, RNA and DNA. Here, the general applicability of the aziridine cofactors was investigated for SMILing DNA. For this purpose the C5-cytosine DNA MTase M·*Hha*I and the N4-cytosine DNA MTase M·*Bcn*IB were used as representatives of the other two classes of DNA MTases. Since SMILing DNA was only demonstrated for short duplex oligonucleotides, the ability of M·*Taq*I to catalyze the labeling of large DNA fragments was investigated. Furthermore, the labeling reaction was optimized with respect to the reaction rate by investigating the pH dependence of the reaction. In addition, the influence of strong nucleophiles in the reaction buffer, such as DTT or β–mercaptoethanol, on the reaction was investigated. Special attention was given to the development of a biotinylated cofactor, since the biotin label is very attractive due to diverse and highly sensitive detection possibilities.

3.1.1 Coupling of *N*-adenosylaziridine with DNA using M·*Taq*I

The new cofactor analogue AZ **(1)** was shown to be sequence-specifically coupled to a short duplex oligonucleotide by M·*Taq*I (Scheme 3.1.1). The reaction was carried out at pH 6 because only a protonated aziridine can act as an alkylating agent. The pK_a value of the aziridine nitrogen of alkylated aziridines in aqueous solution is 8-9 (Buist & Lucas, 1957). Assuming that the pK_a value is not changed within the protein environment, the aziridine nitrogen should be protonated at pH 6 and the aziridine ring activated for a nucleophilic attack. Furthermore, in the publication by Pignot *et al.* (1998), the reaction buffer did not contain any strong nucleophiles like DTT or β-mercaptoethanol in order to prevent an attack of these compounds on the cofactor, thus slowing down or even inhibit the coupling reaction.



7·8 **7^AZ·8**

Scheme 3.1.1: M·TaqI-catalyzed coupling of a 14mer duplex oligonucleotide with N-adenosylaziridine.

Therefore, the influence of DTT as nucleophile in the reaction buffer and the pH value of the reaction buffer on the reaction rate were investigated. The 14mer duplex oligonucleotide **7·8** was modified in a M·*Taq*I-catalyzed coupling reaction with AZ (**1**) to give the modified duplex **7^{AZ}·8**. The reaction was carried out under single turnover conditions at 37 °C. This temperature does not correspond to the activity maximum of M·*Taq*I which is at 65 °C. However, since the melting point of the used short duplex oligonucleotide **7·8** should be below 65 °C, the reaction had to be carried out at a temperature at which the DNA exists in its double-stranded form and is not already melted. In the coupling reaction the product **7^{AZ}·8** is not directly released. As previously described for the M·*Taq*I-catalyzed coupling of **7·8** with DAZ (see Pljevaljčić, 1999), also in this case a complex of protein and modified DNA is primarily formed from which the DNA product can be released by degradation of the protein with proteinase K. Since aziridines are known to react readily with DNA (Tomasz, 1995), a control experiment without enzyme was carried out. In this control experiment no product formation was observed which indicates that the uncatalyzed reaction is slow compared to the M·*Taq*I-catalyzed reaction. Additionally, the reaction was carried out with and without DTT (1 mM) in the reaction buffer. Both reactions proceed equally well yielding product **7^{AZ}·8**. This indicates that 1 mM DTT does not influence the course of the reaction.

Optimization of the reaction rate was carried out by performing reactions at different pH values. The reaction progress was analyzed by anion exchange chromatography and the amount of product DNA quantified by integration of the product peak. A plot of product DNA against time is shown in Fig. 3.1.1 on the left for the reaction at pH 6. The data points were fitted with the software GraFit using a single exponential function. The determined rate constants in dependence of the pH are shown in Fig. 3.1.1 on the right. It can be clearly seen that with increasing pH the reaction rate decreases. This supports the assumption that the reaction is faster at acidic pH due to an activation of the aziridine moiety by protonation of the ring nitrogen. It should be noted that further lowering the pH value leads to precipitation of M·*Taq*I.

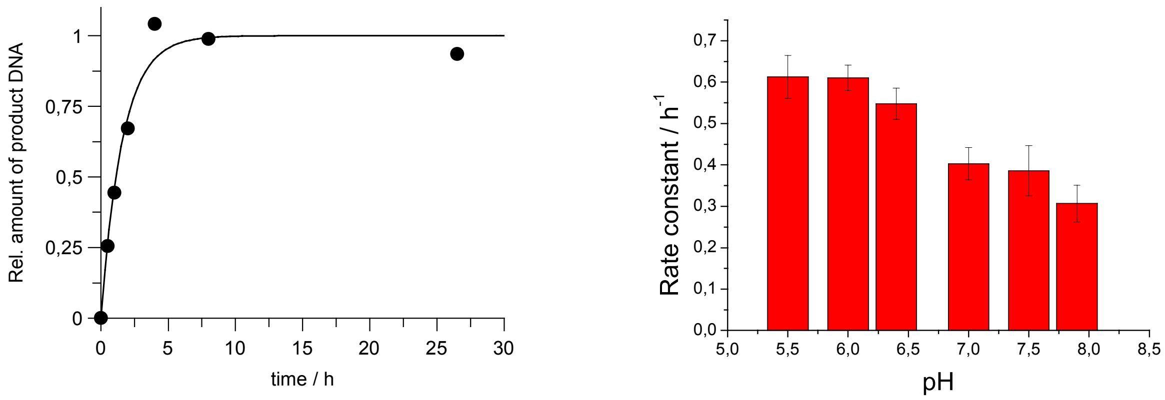
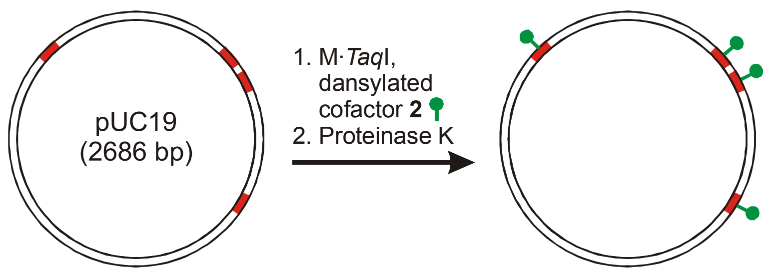


Fig. 3.1.1: Plot of the amount of modified DNA against time at pH 6 (left). The data points were fitted using the software GraFit applying a single exponential function. Diagram of the pH dependence of the reaction rates for the M·TaqI-catalyzed coupling reaction (right).

The actual value of the labeling reaction rate, 0.63 ± 0.09 h^{-1} at pH 6 is very slow compared to the natural methylation reaction rate. For the methylation reaction a turnover rate of 0.1 s^{-1} at 37 °C was determined (Wölcke, 1998). Thus, the DNA MTase M·*Taq*I is a considerably slow enzyme compared to other enzymes which have turnover numbers in the range of 10-10^3 s^{-1} (Fersht, 1999), and DNA labeling with M·*Taq*I is about 600 times slower than the reaction with the natural cofactor AdoMet.

3.1.2 Labeling of plasmid DNA using M·*Taq*I

The coupling reaction of AZ with short duplex oligonucleotides occurs with the highest reaction rate under slightly acidic conditions. Using these conditions the coupling reaction between the fluorescent cofactor **2** (DAZ) and plasmid DNA (pUC19) was investigated. The plasmid pUC19 is a circular DNA molecule with 2686 base pairs (bp). It contains four recognition sites for M·*Taq*I. A M·*Taq*I-catalyzed reaction as illustrated in Scheme 3.1.2 is expected to occur.



Scheme 3.1.2: Expected labeling reaction of plasmid pUC19.
Red bars denote the recognition sequence of M·TaqI 5'-TCGA-3'.

The reaction of pUC19, M·*Taq*I and DAZ was carried out at 65 °C, the activity maximum of M·*Taq*I. The progress of the reaction was monitored by anion exchange chromatography in which the absorption of the plasmid at 260 nm and the fluorescence of the dansyl fluorophore (excitation at 330 nm; emission at 551 nm) were monitored. The anion exchange chromatograms in Fig. 3.1.2a. demonstrate that pUC19 elutes with a retention time of 21.9 min and that the amount of plasmid remains constant during the time course.

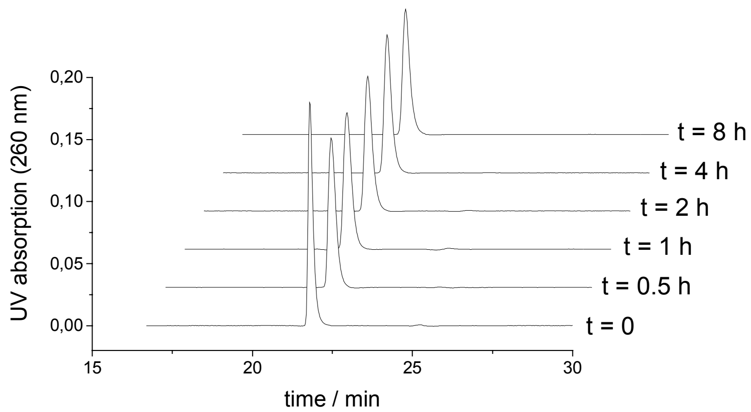


Fig. 3.1.2a: M·TaqI-catalyzed SMILing reaction of pUC19 with DAZ. Anion exchange chromatograms after indicated reaction times with detection at 260 nm.

Therefore, the reaction does not lead to degradation of DNA. Fig. 3.1.2b illustrates the corresponding chromatograms in which the fluorescence of the dansyl fluorophore is recorded.

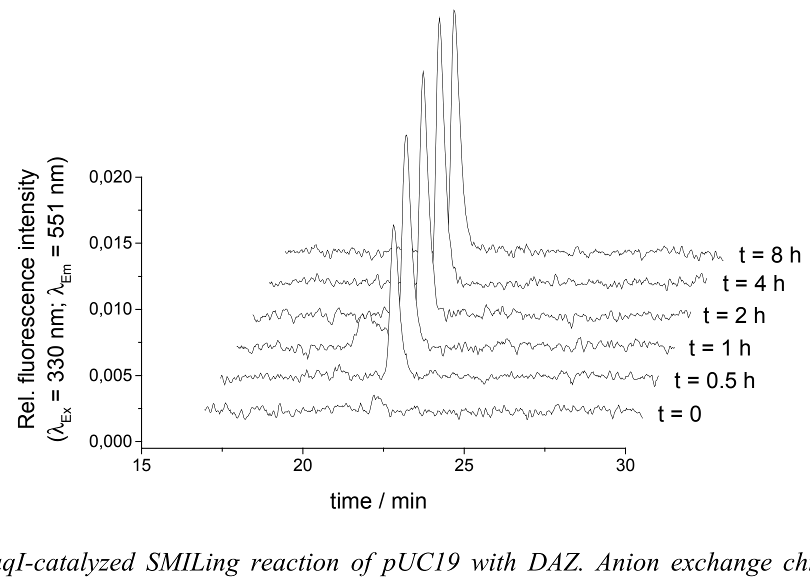


Fig. 3.1.2b: M·TaqI-catalyzed SMILing reaction of pUC19 with DAZ. Anion exchange chromatograms after different incubation times with fluorescence detection (excitation at 330 nm; emission at 551 nm). The time axis is corrected for the delay time between the UV and fluorescence detector.

Obviously, the dansyl fluorescence can be measured at the same retention time as the DNA (compare with Fig. 3.1.2a). The amount of fluorescence increases with time while at time zero no fluorescence can be detected at the retention time of DNA. This demonstrates labeling of the pUC19 plasmid with the dansyl fluorophore. The reaction is completed after 4 h. The rate of the reaction based on fluorescence signal is 2.2 ± 0.2 h^{-1}. This is about 3.5 times faster than the coupling reaction with the short duplex oligonucleotides and AZ. The simplest explanation for this increase in rate is that the reaction was carried out at the temperature optimum of M·*Taq*I (65 °C) whereas the reaction with the short duplex oligonucleotides was performed at 37 °C. This experiment demonstrates that even large DNA, namely pUC19, can be labeled using the DAZ cofactor. As a control, the reaction was carried out without M·*Taq*I in order to demonstrate that the observed reaction is enzyme-catalyzed. While the time trace of the anion exchange chromatograms for the UV-absorption at 260 nm equals the time trace of the reaction with enzyme, the time trace for the observed fluorescence shows no dansyl fluorescence at all (Fig. 3.1.2c).

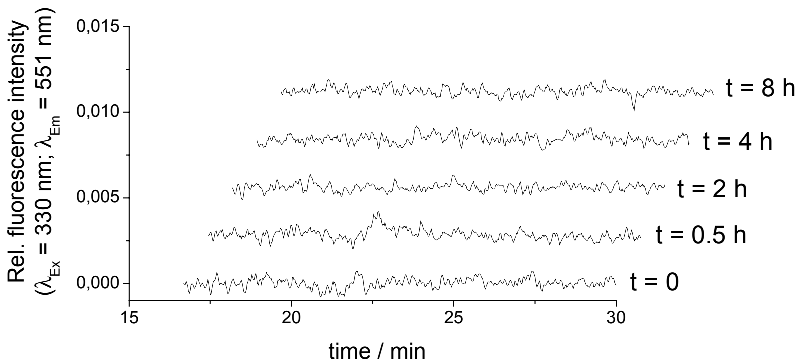


Fig. 3.1.2c: Control experiment with pUC19 and DAZ in the absence of M·TaqI. Anion exchange chromatograms after different incubation times with fluorescence detection (excitation at 330 nm; emission at 551 nm). The time axis is corrected for the delay time between the UV and fluorescence detector.

This control experiment reveals that no background reaction of the cofactor with DNA can be detected. Therefore, labeling of the plasmid DNA is definitely enzyme-catalyzed. However, the experiment does not clarify whether the reaction occurs quantitatively and sequence-specifically like the SMILing reaction of short duplex oligonucleotides does (Pljevaljčić, 1999). The question about quantity of the labeling reaction is quite difficult to answer with this type of experiment as there is no easy possibility to calibrate the observed fluorescence signal. Therefore, it is not possible to make a statement about the amount of fluorophore attached to the DNA. Other types of experiments have to be performed to answer this question (see below). The question concerning the specificity of the enzyme-catalyzed reaction was addressed by performing two parallel reactions. The pUC19 plasmid was either treated with M·*Taq*I and the natural cofactor AdoMet or only with AdoMet. The plasmids were purified with the QiagenPCR purification Kit. One would expect in the first case that all M·*Taq*I recognition sites are methylated due to the natural methylation reaction and in the second case that no methylation took place. When the purified plasmids are now both treated in a second step with M·*Taq*I and the dansylated cofactor DAZ the first one should not be labeled as the M·*Taq*I recognition sites are protected by methyl groups whereas the second plasmid should be labeled as it was previously observed. In Fig. 3.1.2d the time traces of anion exchange chromatograms for the methylated plasmid are shown. Only the anion exchange chromatograms in which the fluorescence was recorded are presented as the chromatograms for the UV-absorption at 260 nm are identical to the previously shown ones.

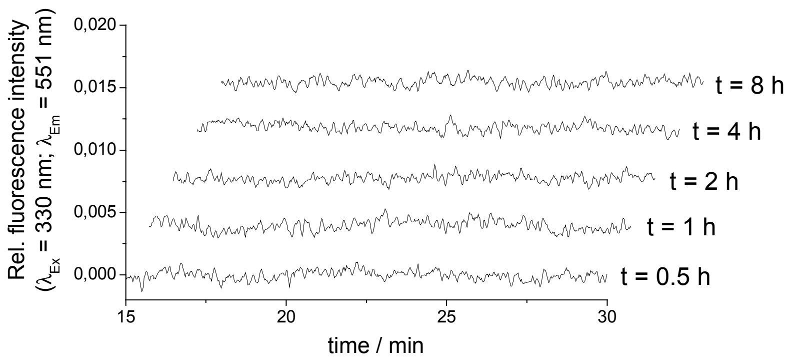


Fig. 3.1.2d: Anion exchange chromatograms of previously methylated pUC19. The dansyl fluorescence (excitation at 330 nm; emission at 551 nm) is recorded. The time axis is corrected for the delay time between the UV and fluorescence detector.

As expected no fluorescence signal can be detected because the labeling reaction is not expected to occur at methylated M·*Taq*I recognition sites. In contrast to this, the non-methylated pUC19 becomes labeled with the dansyl fluorophore as described above. The time traces of the anion exchange chromatograms of the non-methylated pUC19 is shown in Fig. 3.1.2e. The amount of observed fluorescence equals that previously observed.

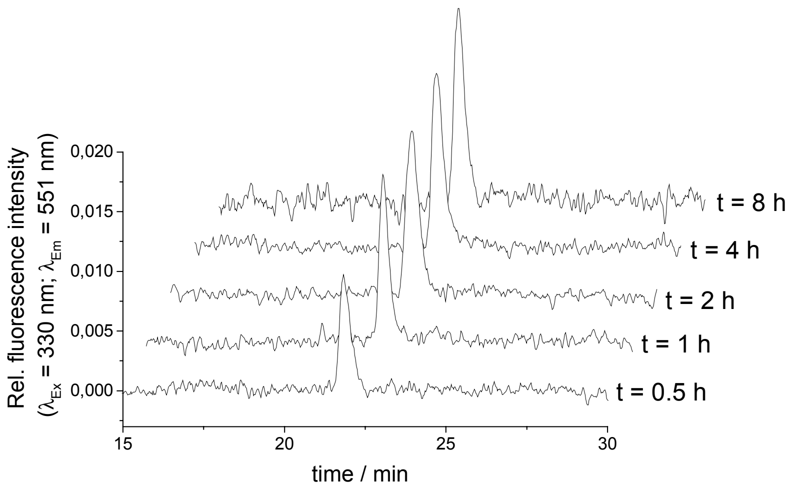


Fig. 3.1.2e: Anion exchange chromatograms of labeling non-methylated pUC19. The dansyl fluorescence (excitation at 330 nm; emission at 551 nm) is recorded. The time axis is corrected for the delay time between the UV and fluorescence detector.

This first control experiment demonstrates that methylated sites cannot be labeled with DAZ and that the sequence-specificity of M·*Taq*I is not altered with the unnatural cofactor DAZ. Supposing that M·*Taq*I performs the methylation reaction sequence-specifically the labeling reaction is also sequence-specific. The experiment does not give information about the sequence which becomes methylated and labeled, respectively. However, experiments from Friederich (2001) suggest that the sequence specificity of M·*Taq*I is comparable to that of restriction endonucleases. According to that M·*Taq*I methylates 5'-TCGA-3' about 10^5 times better than the second best recognized site 5'-TCTA-3'. Therefore, it can be assumed that the labeling reaction is sequence-specific as well. To summarize this set of experiments, Sequence-specific Methyltransferase-Induced Labeling of DNA (SMILing DNA) is also applicable to plasmid DNA using M·*Taq*I as a model system.

3.1.3 Coupling of *N*-adenosylaziridine with DNA using M·*Hha*I

The AZ cofactor for SMILing DNA with M·*Taq*I was designed based on the structure of the natural cofactor and should also work with other DNA MTases. Therefore, the coupling reaction of AZ with a short duplex oligonucleotide was carried out using the C5-cytosine DNA MTase from *Haemophilus haemolyticus* (M·*Hha*I). Cofactor-free enzyme was kindly provided by the group of Dr. Saulius Klimasauskas, Vilnius (Lithuania). First of all, the affinity of AZ to M·*Hha*I was determined. For that purpose a binding assay was carried out described by Pignot (1999). M·*Hha*I contains a single tryptophan residue (W41) in the cofactor binding site. Its fluorescence is strongly reduced upon cofactor binding. A titration of M·*Hha*I with AZ is shown in Fig. 3.1.3a.

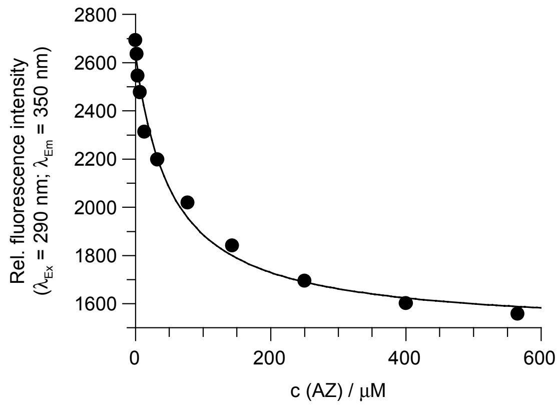
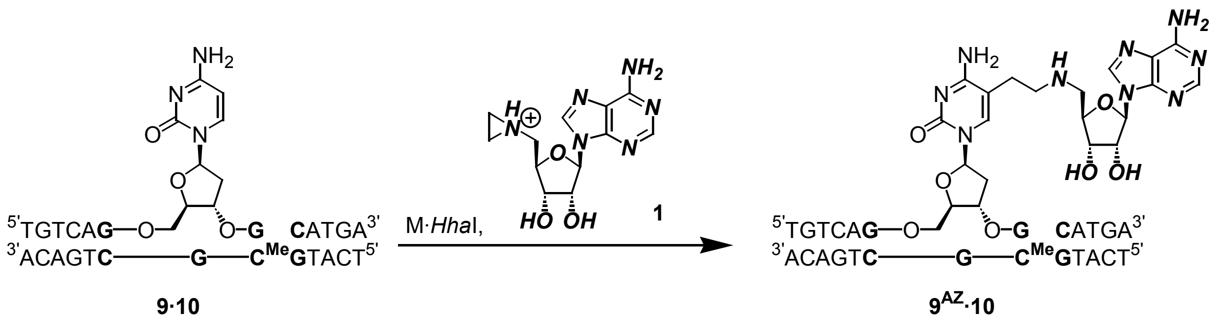


Fig. 3.1.3a: Plot of the tryptophan fluorescence versus AZ concentration. The data points were fitted with the software GraFit applying the quadratic equation for one binding site. The concentration of M·HhaI was 1 µM.

The data points were fitted to the quadratic equation for one binding site using the program GraFit and a binding constant of 53 ± 10 µM was obtained. In comparison to M·*Taq*I, AZ binds four times more strongly to M·*Hha*I. Compared to the natural cofactor AdoMet however, the AZ cofactor binds 50 times more weakly. As AZ binds to M·*Hha*I with an acceptable binding constant the reaction was carried out in analogy to the M·*Taq*I-catalyzed coupling reaction using a short duplex oligonucleotide (Scheme 3.1.3).



*Scheme 3.1.3: M·HhaI-catalyzed reaction between the duplex **9·10** and AZ **1**.*

In this case the 13mer duplex oligonucleotide **9·10** was chosen as substrate. The DNA contains the recognition sequence of M·*Hha*I, 5'-G**C**GC-3', in which the second cytosine is the target base. In the lower strand **10** the target base is already methylated and thus not available for further modification. The reaction was carried out under single turnover conditions at 37 °C. This time the reaction was carried out at pH 7.4 as this proves to be the pH value at which the protein is most stable. Continuous aliquots were taken out from the reaction solution at distinct time intervals and the aliquots were treated for 1 h with proteinase K at 37 °C. The solution was injected onto an anion exchange HPLC column and the UV absorption was recorded at 260 nm. The obtained anion exchange chromatograms after different incubation times are shown in Fig. 3.1.3b.

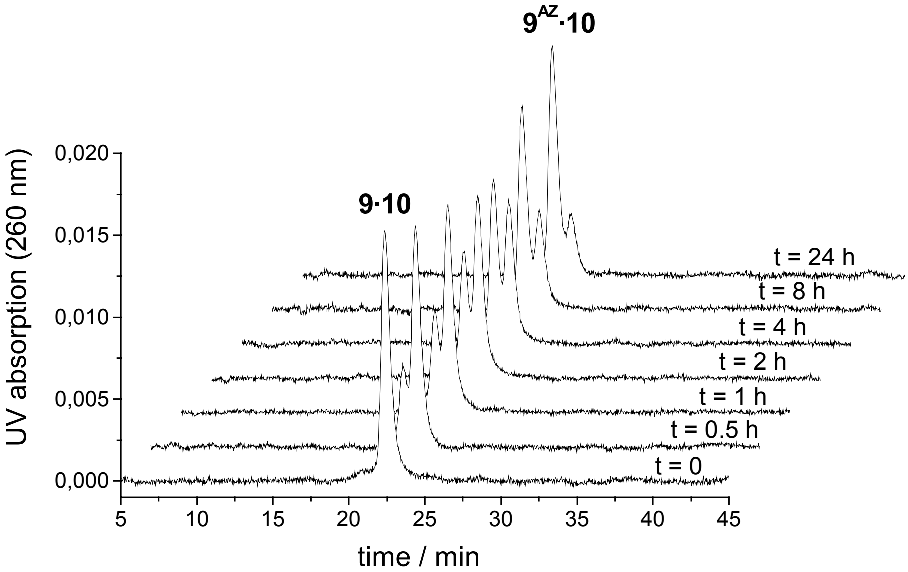


*Fig. 3.1.3b: M·HhaI-catalyzed coupling reaction between **9·10** and AZ **1** at 37 °C. Anion exchange chromatograms with recorded UV absorption at 260 nm.*

The anion exchange chromatogram at the beginning of the reaction shows the double-stranded oligonucleotide **9·10** with a retention time of 22.4 min. As the reaction progresses the signal of **9·10** decays and a signal of the newly formed compound **9^{AZ}·10** (R_t 21.4 min) increases. The retention time of **9^{AZ}·10** is slightly smaller than that of the starting material. After 24 h **9·10** is almost completely converted into **9^{AZ}·10**. In a control experiment without enzyme **9·10** is not converted at all. This indicates that the reaction is enzyme-catalyzed. When the reaction solution is not treated with proteinase K, a complex of protein, DNA and cofactor is observed analogous to the M·*Taq*I-catalyzed reaction. Also the elution behavior of the reaction product on the anion exchange column resembles the behavior already observed in the M·*Taq*I-catalyzed reaction of AZ with DNA. These are two hints which already suggest that a M·*Hha*I-catalyzed coupling reaction of DNA with AZ occurs. In this reaction, the enzyme needs 24 h to transform most of **9·10** into **9^{AZ}·10** which is at least twice as long as M·*Taq*I needs for the reaction. However, a new preparation of M·*Hha*I which was kindly provided by Egle Merkiene (Klimasauskas Laboratory) almost completely converts the DNA into the product in only 30 min. The anion exchange chromatograms of the DNA before and after the reaction are shown in Fig. 3.1.3c.

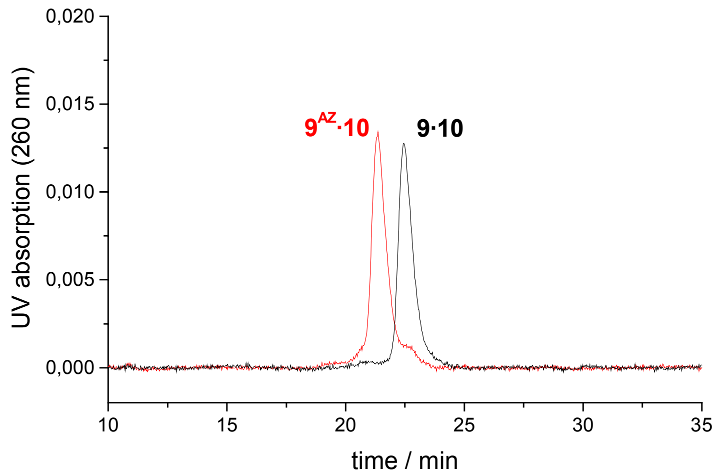


Fig. 3.1.3c: Anion exchange chromatograms of DNA before and after the M·HhaI-catalyzed coupling reaction with AZ at 37 °C.

The course of the reaction using this new enzyme preparation could not easily be monitored by HPLC analysis. The reaction however is much faster than the M·*Taq*I-catalyzed reaction. In order to analyze the reaction product with mass spectrometry, **9^{AZ}·10** was isolated from the anion exchange column and desalted using the Seppack C18 columns from Waters. Prior to measurement by MALDI mass spectrometry the sample was further purified using the ZipTip C18 pipette tips. The observed spectrum is shown in Fig. 3.1.3d. The mass spectrum was calibrated using the single-stranded oligonucleotide **9**.

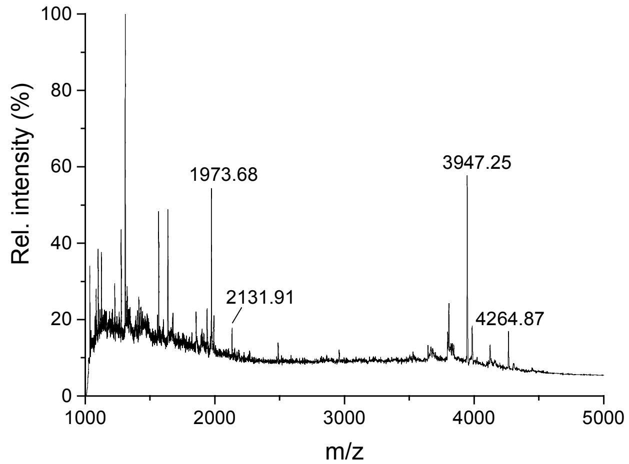


*Fig. 3.1.3d: MALDI mass spectrum of the modified duplex ologonucleotide **9^{AZ}·10**.*

The mass spectrum shows the signals for strand **10** and the modified strand **9^{AZ}**. The signal at 1,973.68 corresponds to double negative **10** (calc. 1,971.68 [**10**-2H]$^{2-}$) and the signal at 2,131.91 corresponds to double negative **9^{AZ}** (calc. 2,129.91 [**9^{AZ}**-2H]$^{2-}$). A zoom into the spectrum at higher m/z values is shown in Fig. 3.1.3e.

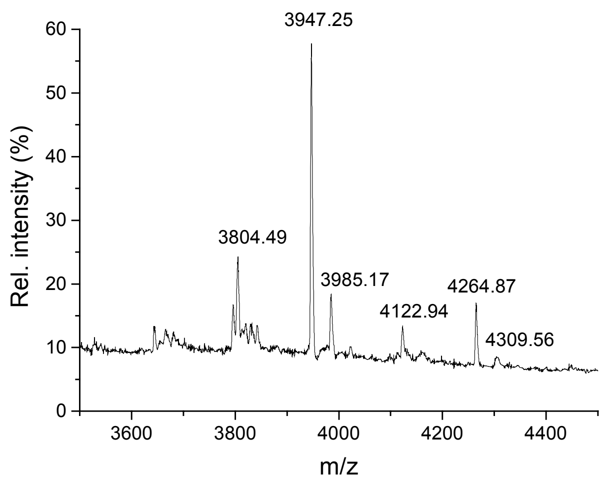


Fig. 3.1.3e: Zoom into the full MALDI mass spectrum shown in Fig. 3.1.3d.

The mass spectrum shows a signal at 3,947.25 which corresponds to strand **10** (calc. 3,945.71 [**10**-H]$^-$). The peak at 4,264.87 corresponds to strand **9** modified with AZ (calc. 4,262.82 [**9**^{Az}-H]$^-$). Additionally, signals of potassium adducts of each oligonucleotide can be seen (signals at 3,985.17 and 4,309.56). The signals at 3,804.49 and 4,122.94 could not be assigned but arise probably from a degradation of the oligonucleotides. In summary, the mass spectrum reveals that the cofactor analogue AZ becomes coupled to strand **9**. Unfortunately, the mass spectrum does not allow a statement about the base or even the sequence which was modified. However, since the reaction was carried out with the C5-cytosine DNA MTase M·*Hha*I it is most likely that the C5-position of the target cytosine was modified with AZ. This is also in agreement with the observation that no signal corresponding to the mass of a modified lower strand **10** (calc. 4,237.83 [**10**^{AZ}-H]$^-$) is present. A further experiment was performed which should allow a statement about the modified nucleobase and its position. The same coupling reaction as described for duplex oligonucleotide **9·10** was carried out with a duplex oligonucleotide which contains an A-T base pair instead of the C-G-base pair at target position. With this duplex oligonucleotide no coupling reaction occurs. Therefore, it is most likely that the target cytosine becomes modified in a M·*Hha*I-catalyzed coupling reaction and not another nucleobase. Thus, the sequence-specific labeling of DNA using a representative of the second class of DNA MTases, the C5-cytosine DNA MTase M·*Hha*I, could be demonstrated. The performed experiments however, cannot prove that the C5-position of cytosine was modified. Since M·*Hha*I is a C5-cytosine DNA MTase it is very unlikely that another position of the base becomes modified; especially because the reaction mechanism suggested by Santi *et al.* (1987) proceeds *via* an activated cytosine at the C5-position.

3.1.4 Single-turnover experiments at different pH-Values

The M·*Taq*I-catalyzed coupling reaction is pH sensitive. Reactions with M·*Hha*I were carried out at different pH values in order to investigate if the reaction rate can be further increased. The pH dependency should be possible to monitor even with an enzyme preparation of low specific activity. The experiment was carried out with the enzyme preparation which took about 24 h for conversion because the reaction with the new enzyme preparation is too fast to follow with HPLC. The reaction was again monitored by anion exchange chromatography and the amount of product formed with time is shown for pH 7, as an example in Fig. 3.1.4 on the left. The dependence of the reaction rate on the pH is plotted in Fig. 3.1.4 on the right.

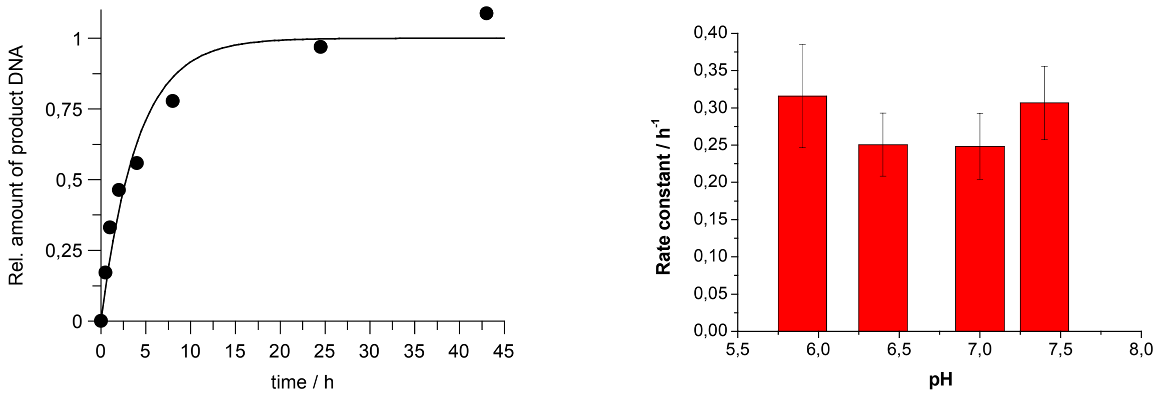


Fig. 3.1.4: Plot of the amount modified DNA formed with time at pH 7 (left). The data points were fitted with a single exponential function using the software GraFit. Diagram of the pH dependence of the reaction rate for the M·HhaI-catalyzed reaction (right).

In contrast to the M·*Taq*I-catalyzed coupling reaction, the reaction with M·*Hha*I is independent of the pH value within the investigated pH interval. Therefore, an optimizatin of the reaction rate could not be achieved by simply increasing the pH value.

3.1.5 Labeling of plasmid DNA using M·*Hha*I

The important result that the coupling reaction of DNA with AZ is not limited to M·*Taq*I but also possible with M·*Hha*I encourages the investigation of SMILing of plasmid DNA with DAZ and other DNA MTases like M·*Hha*I. As substrate the already mentioned plasmid pUC19 was utilized which contains 17 recognition sites for M·*Hha*I. The reaction was carried out at 37 °C and monitored again by anion exchange chromatography. Fig. 3.1.5a shows two anion exchange chromatograms in which the absorption at 260 nm was recorded.

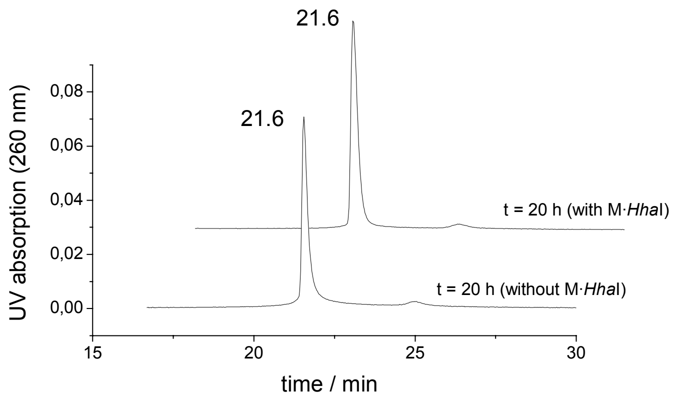


Fig. 3.1.5a: Anion exchange chromatograms of the experiments with pUC19 and DAZ at 37 °C. The upper trace shows the reaction with M·HhaI-catalysis and the lower without. In both cases the absorption at 260 nm was recorded.

After 20 hours reaction time the same amount of pUC19 plasmid, eluting after 21.6 min, is present in the experiment either with or without enzyme. This illustrates that the DNA is not degraded during the reaction. By observing the dansyl fluorescence, one can see that the plasmid becomes labeled only in the case where enzyme is present (Fig. 3.1.5b). This demonstrates that the reaction is enzyme-catalyzed and that an uncatalyzed reaction does not occur.

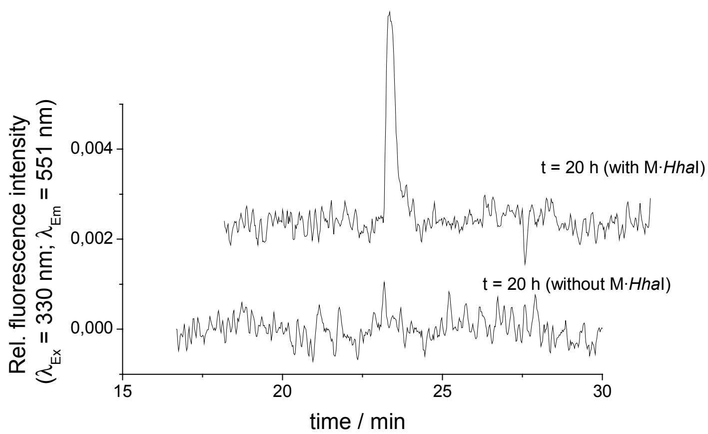
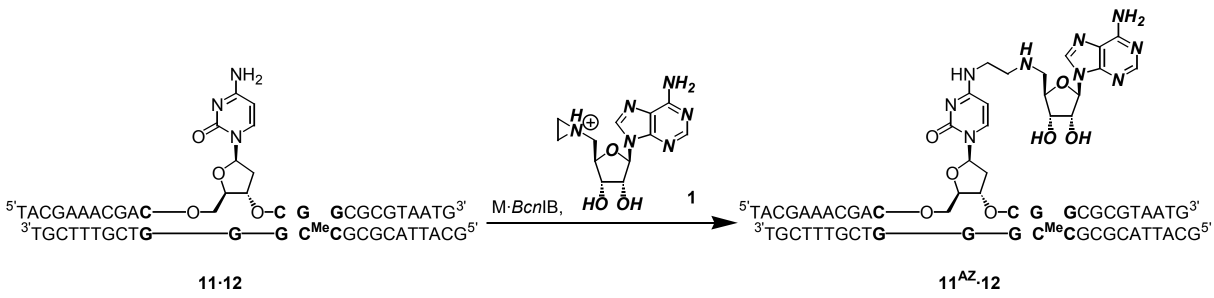


Fig. 3.1.5b: Anion exchange chromatograms of the experiments with pUC19 and DAZ at 37 °C. The upper trace shows the reaction with M·HhaI-catalysis and the lower without. In both cases the dansyl fluorescence (excitation at 330 nm; emission at 551 nm) was recorded. The time axis is corrected for the delay time between the UV and fluorescence detector.

Thus, the general applicability of the newly developed cofactor family for SMILing of small and large DNA fragments could be demonstrated. Owing to the conserved binding pocket of DNA MTases it is expected that the new cofactors can also be used in combination with other DNA MTases expanding the range of labeled sequences.

3.1.6 Coupling of *N*-adenosylaziridine with DNA using M·*Bcn*IB

For the sake of completeness a representative of the third type of DNA MTases which methylates the N4-position of cytosines, was tested in a coupling reaction of a short duplex oligonucleotide with AZ. The enzyme was obtained from the lab of Dr. Saulius Klimasauskas, Vilnius (Lithuania). It is a N4-cytosine DNA MTase from *Bacillus centrosporus* (M·*Bcn*IB) which methylates double-stranded DNA only. The coupling reaction was performed using the duplex oligonucleotide **11·12**. The enzyme recognizes the sequence 5'-C**C**(C/G)GG-3' in which the second cytosine is methylated on its exocyclic amino group. With AZ it is expected that the coupling reaction occurs at the same position (Scheme 3.1.6).



*Scheme 3.1.6: M·BcnIB-catalyzed coupling reaction between **11·12** and AZ **1**.*

The reaction was carried out at 37 °C at pH 6.9. Aliquots were taken from the reaction mixture at distinct time intervals and treated with proteinase K at 37 °C for 1 h. The aliquots were analyzed by anion exchange chromatography. The observed anion exchange chromatograms are presented in Fig. 3.1.6a.

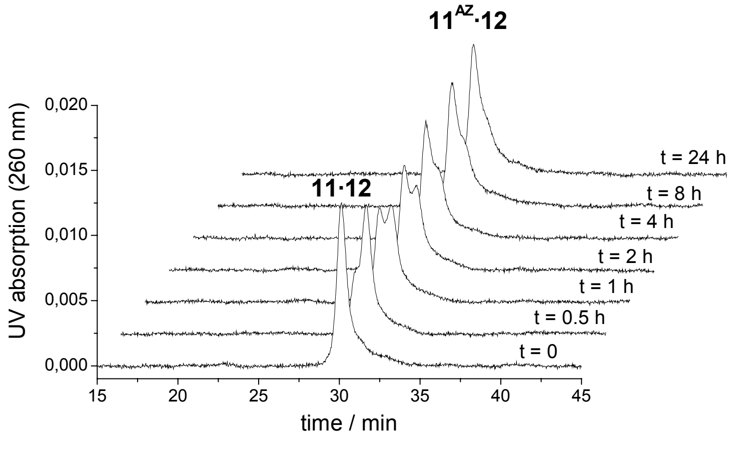


*Fig. 3.1.6a: M·BcnIB-catalyzed coupling reaction between **11·12** and AZ **1** at 37 °C. The anion exchange chromatograms were recorded at 260 nm.*

At t = 0 the duplex oligonucleotide **11·12** elutes with a retention time of 30.2 min. With increasing reaction time a new compound **11^{AZ}·12** is formed with a slightly smaller retention time (29.3 min). After about 24 h almost all of the starting material becomes converted to the new compound. Fig. 3.1.6b shows the two anion exchange chromatograms before and after the coupling reaction. A control experiment in the absence of enzyme does not show any product formation.

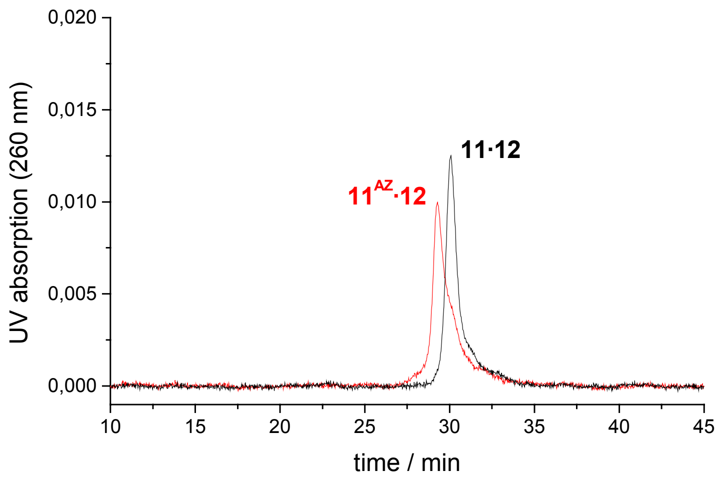


Fig. 3.1.6b: Anion exchange chromatograms of DNA before and after the M·BcnIB-catalyzed coupling reaction with AZ at 37 °C.

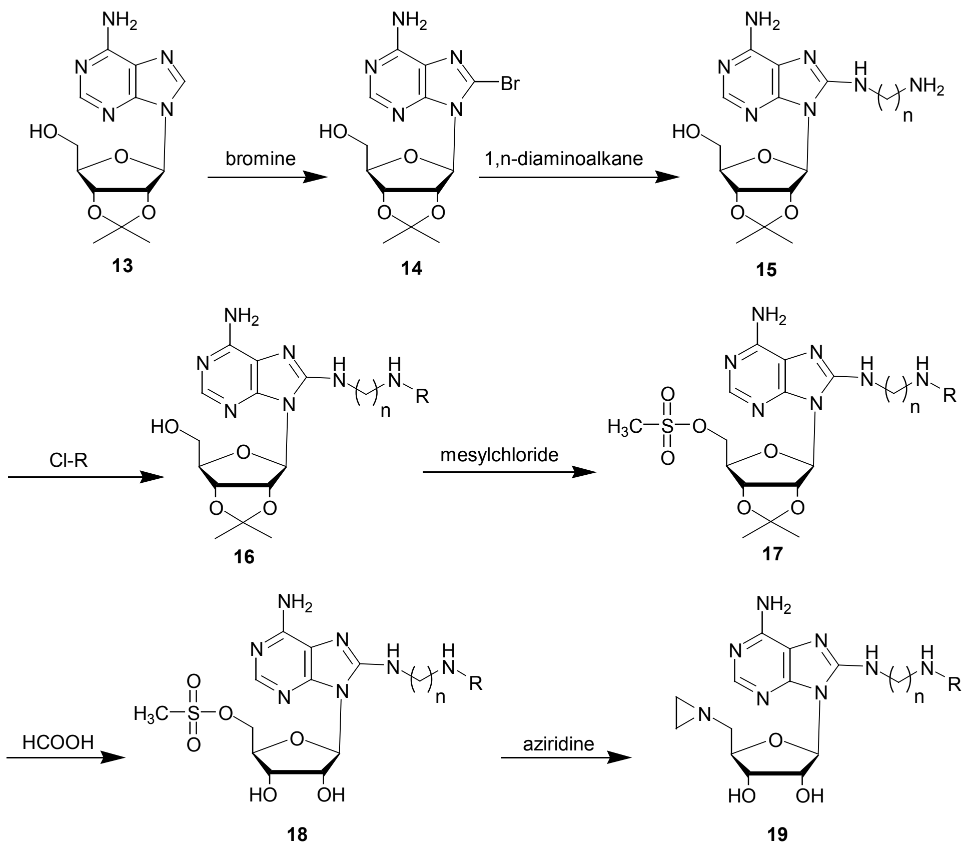
When the reaction solution is not treated with proteinase K prior to HPLC analysis a complex of protein, DNA and cofactor can be observed. This behavior parallels that already described with M·*Taq*I and M·*Hha*I. The major difference results from the larger size of the duplex oligonucleotide which leads to a larger retention time of **11·12** in comparison to the shorter oligonucleotides used with M·*Taq*I and M·*Hha*I. Furthermore, the difference in retention time of modified and unmodified duplex oligonucleotide is not as pronounced as in the former cases. However, the observations strongly suggest that an M·*Bcn*IB-catalyzed coupling reaction of DNA and AZ does occur. Since M·*Bcn*IB is a N4-cytosine DNA MTase it is most likely that the coupling reaction occurs at the exocyclic amino group of the second cytosine within the 5'-C**C**CGG-3' DNA sequence. This is supported by two other experiments. A MALDI mass spectrum of the modified DNA shows the modified and non-modified strand. However, due to the length of the duplex oligonucleotide the intensity of the measured spectrum is much lower than for 14mer oligonucleotides and the signal for the non-modified strand appears slightly above the noise. Furthermore, the coupling reaction of AZ with a duplex oligonucleotide in which the target C-G base pair was exchanged by an A-T base pair does not occur. This suggests that the target cytosine becomes modified in a M·*Bcn*IB-catalyzed coupling reaction and not another nucleobase.

In summary, these experiments demonstrate that the newly developed cofactor AZ can be utilized with all three classes of DNA MTases for SMILing DNA. Regarding the fact that about 3000 procaryotic restriction endonucleases targeting about 250 different DNA sequences are known today (see: http://www.neb.com/rebase) and for each restriction endonuclease there exists a DNA MTase, SMILing DNA can be used to target a huge number of different sequences. To the best of my knowledge, this extremely challenging task cannot be achieved with any other method.

3.2 Synthesis of a cofactor analogue precursor

3.2.1 Strategy for a general synthesis

The synthesis of DAZ, an AZ derivative which is modified at the 8-position with a tethered dansyl fluorophore was already presented in my diploma thesis (Pljevaljčić, 1999). However, the synthetic route by which the novel cofactor was obtained is not suitable for the synthesis of a whole series of cofactors with different reporter groups. Therefore, a general synthesis for AZ derivatives was worked out which allows the attachment of different reporter groups at the end of the synthesis. The synthesis presented in my diploma thesis was modified because the reporter group, the dansyl fluorophore, was introduced into the molecule in an early stage of the synthesis. In the general synthesis a protecting group is first introduced into the molecule. This leads to a cofactor analogue precursor which can be transformed after deprotection into the desired labeled AZ derivatives. The chosen new synthetic strategy for cofactor analogues is illustrated in Scheme 3.2.1a.



R = protecting group; n = 4

Scheme 3.2.1a: Strategy for the synthesis of cofactor analogue precursors.

The crucial considerations in this synthesis are the choice of the tether length at the 8-position and the choice of the protecting group. The linker length is essential for two reasons. Firstly, the linker has to be long enough so that the cofactor analogues can bind to the protein without interference of the reporter group with the protein. The DAZ cofactor contains a linker with four CH_2-units which proved to be long enough to bind tightly to the protein and to allow SMILing DNA. Secondly, the length of the linker has to be suitable for indirect labeling systems. Indirect labeling systems are e.g. the biotin-streptavidin system. In this biotin is used as a modification group which would be attached to the linker. In a second step biotin is detected *via* a strong non covalent interaction with the protein streptavidin which itself is conjugated either with a fluorophore or another protein which is used for signal amplification. This means that biotin has to be far enough from its labeled species so that streptavidin can bind to biotin. Two images of the crystal structure of streptavidin in complex with biotin are shown (Fig. 3.2.1). The left image shows the binding pocket of biotin. In the right image the protein is rotated by 90°.

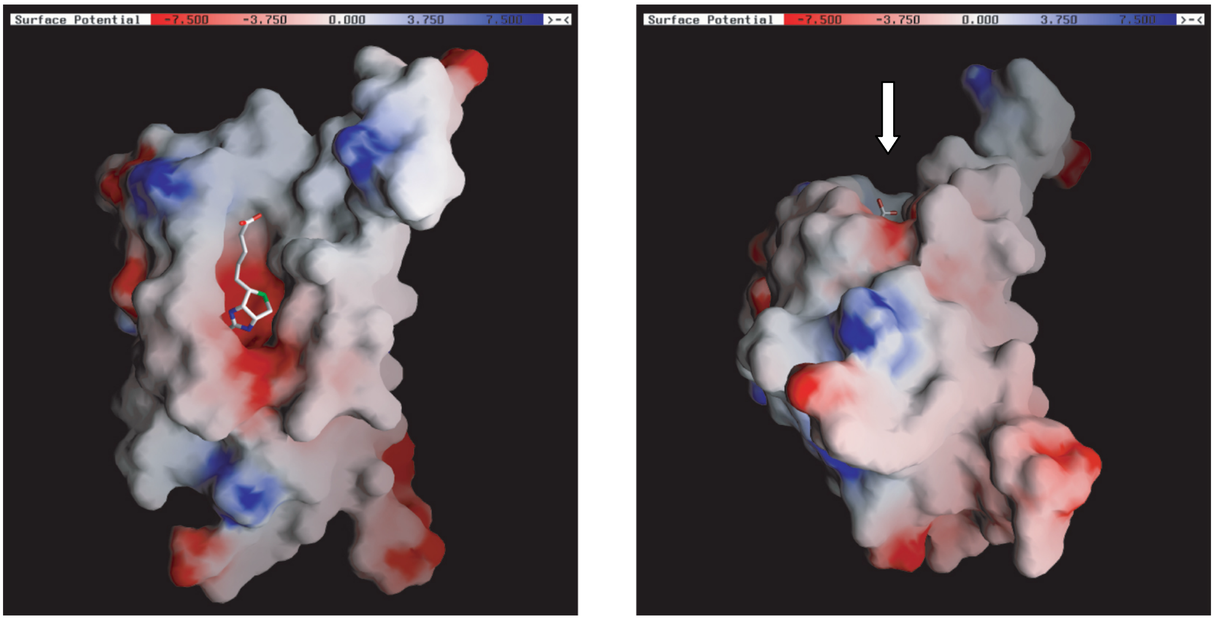
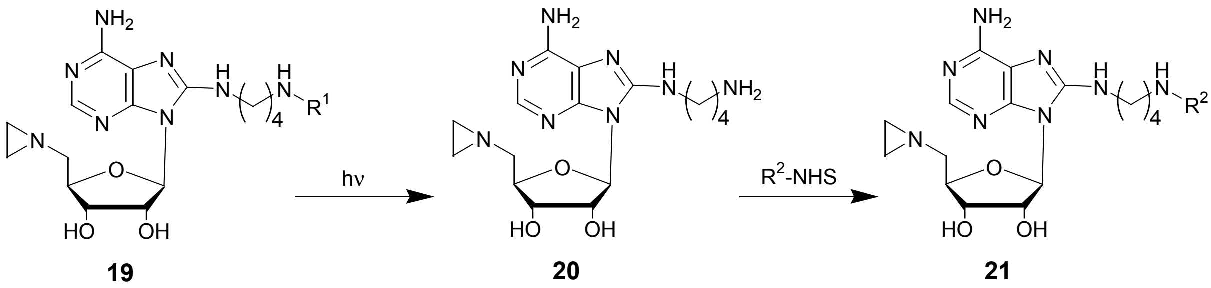


Fig. 3.2.1: Surface potential representation of a streptavidin-biotin complex (red denotes negative electrostatic potential, blue denotes positive electrostatic potential and white is neutral). The binding of biotin (stick representation) is illustrated (left). The carboxylate end of the biotin molecule reaches the surface of the protein (right).

In the image on the right, the carboxylate end of the biotin molecule is visible and reaches the surface of the protein (white arrow). Therefore, a linker with four CH_2-units (length of 3.9 Å), as used for DAZ, attached to that carboxylate end should be sufficient to allow interaction between biotinylated DNA and streptavidin. For this reason, the linker length was chosen to remain the same as used in the case of DAZ. The choice of a suitable protecting group for this synthetic strategy is somewhat more a trial and error procedure. However, the protecting group has to fulfill several prerequisites. One essential step is the removal of the protecting group. As the protecting group has to stay in the molecule until the end of the synthesis, the deprotection strategy must not include deprotection under acidic conditions as the aziridine moiety would be destroyed. Further the protecting group should resist mild basic conditions as the activation of the 5'-hydroxyl group includes basic conditions. Interactions of the protecting group with formic acid or mesyl chloride must be excluded as well. For these reasons different protecting groups which can be deprotected under basic conditions were investigated. *N*-phthalimide, formamide, acetamide, chloroacetamide and 9-fluorenyl-methoxycarbonyl (Fmoc) were tested for suitability. However, none of the chosen protecting groups performed satisfactorily. For this reason, a photolabile protecting group was chosen finally. This group can be removed by light treatment yielding the free amine **20**. In a final step compound **20** can then be coupled with any amine reactive probe to give differently decorated cofactor analogues. Here, the reaction with an NHS-ester is shown as an example yielding the desired product **22** (Scheme 3.2.1b).

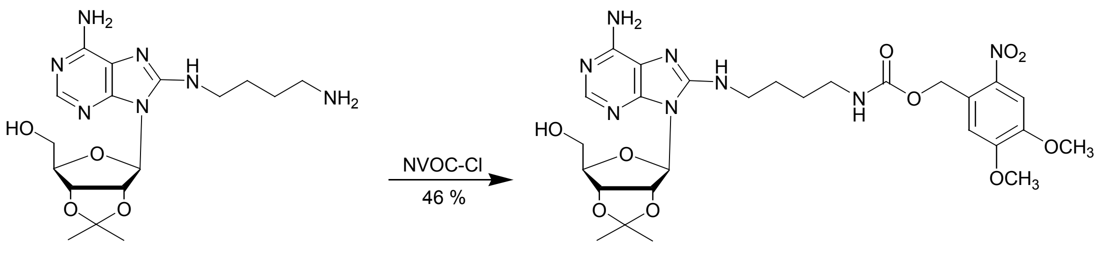


R^1 = NVOC; R^2 = NHS-ester of reporter group

Scheme 3.2.1b: Deprotection of cofactor analogue precursor ***19*** *and introduction of reporter groups to yield desired cofactor analogue* ***21****.*

3.2.2 Synthesis of an adenosine derivative with an NVOC-protected aminolinker at the 8-position

The synthesis starts from the commercially available 2',3'-*O*-isopropylidene adenosine (**13**). As described elsewhere (Pljevaljčić, 1999) the 8-position of **13** is brominated by adding bromine to a solution of **13** in potassium acetate buffer (pH 3.9). The resulting white solid, 8-bromo-2',3'-*O*-isopropylidene adenosine (**14**), is almost quantitatively converted in a nucleophilic substitution with 1,4-diaminobutane into the primary amine **15**. The crude amine is used without further purification for the next step. The third and crucial step of the synthesis is to introduce a protecting group for the primary amine of the linker. 6-Nitroveratryl-chloroformate is added to the amine **15** in dry pyridine at room temperature. The resulting product is purified by silica column chromatography yielding the protected amine **22** (Scheme 3.2.2).



*Scheme 3.2.2: Synthesis of the NVOC-protected amine **22**.*

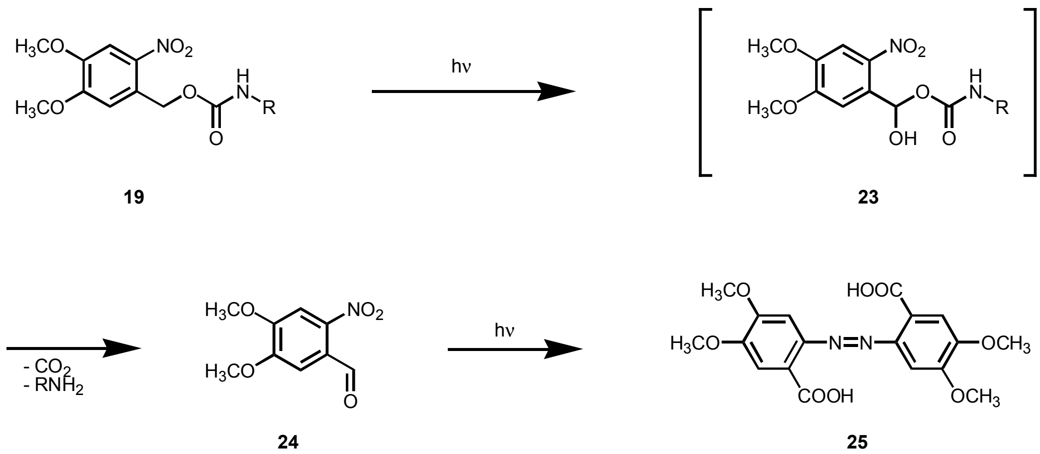
6-Nitroveratryloxycarbonyl (NVOC) was chosen as a protecting group for several reasons. NVOC is photolabile and can be removed at wavelengths above 320 nm. Thus, the nucleobase chromophore (λ_{max} = 280 nm) does not get destroyed by the irradiation. Another advantage of this protecting group is that it doesn't carry any functional groups which could react in the following synthetic steps. Therefore, compound **22** is suitable for usage in the following steps.

3.2.3 Synthesis of a protected *N*-adenosylaziridine derivative

The 5'-hydroxyl group of the NVOC-protected nucleoside **22** was activated using mesylchloride. The reaction was carried out in dichloromethane in the presence of 4-dimethylaminopyridine (DMAP) and triethylamine. A problem in this reaction is the subsequently occurring intramolecular cyclisation of N3 with the activated 5'-position which yields an undesired cyclonucleoside (Townsend, 1988). However, cyclonucleoside formation can be minimized by using short reaction times. Mesylate **17** (R = NVOC) was purified by silica column chromatography at low temperature. The maximum yield of the mesylation is about 55 %. The purified mesylate **17** (R = NVOC) was deprotected almost quantitatively at the 2',3'-hydroxyl groups with formic acid to yield **18** (R = NVOC). Here, the cyclonucleoside formation does not represent a problem as under acidic conditions the N1-position of adenine is protonated, thus reducing the nucleophilicity of N3. Without further purification **18** (R = NVOC) was converted to the protected AZ derivative **19** in an aziridine/base solution. Diisopropylethylamine was chosen as a non-nucleophilic base. The ratio of aziridine to base was chosen to be 3:1 in order to avoid polymerization of aziridine. The cofactor analogue precursor **19** was purified by HPLC chromatography and obtained with a yield of 20 %.

3.2.4 Deprotection of the photolabile NVOC-protecting group

The cofactor analogue precursor **19** was deprotected to yield the free amine which can then be used for reactions with any amine reactive reporter group. The deprotection was carried out in triethanolamine buffer at pH 8 (small amounts of DMSO were used to dissolve **19** completely). The buffer was chosen for various reasons. Triethanolamine has the advantage of being a non-nucleophilic buffer. The pH value was chosen to be slightly basic to avoid activation of the aziridine moiety by protonation and not too basic so as to minimize a side reaction of the liberated aldehyde **24** with the free amine. The proposed mechanism for the deprotection of the NVOC group is illustrated (Scheme 3.2.4).



Scheme 3.2.4: Proposed deprotection mechanism of NVOC-protected amines (Patschornik et al., 1970).

The NVOC-protected amine **19** is proposed to form **23** in a light induced reaction. After liberation of CO_2 the amine is released and benzaldehyde **24** is formed. A dimerization of **24** is proposed to occur in the subsequent light induced reaction yielding the diazo-compound **25**. The formation of the diazo product can be verified by UV-spectroscopy.

In a test reaction, a solution of the precursor **19** was illuminated with laser light at 355 nm. The energy of the laser was chosen to be low (20-30 mJ). UV-spectra of the solution after different amounts of laser pulses are shown (Fig. 3.2.4). Enhancing the energy of the laser light does not lead to a faster deprotection reaction but to a photodestruction of **19**.

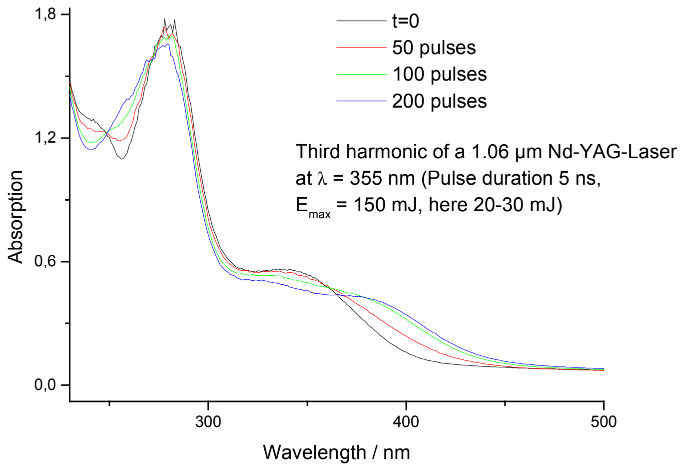
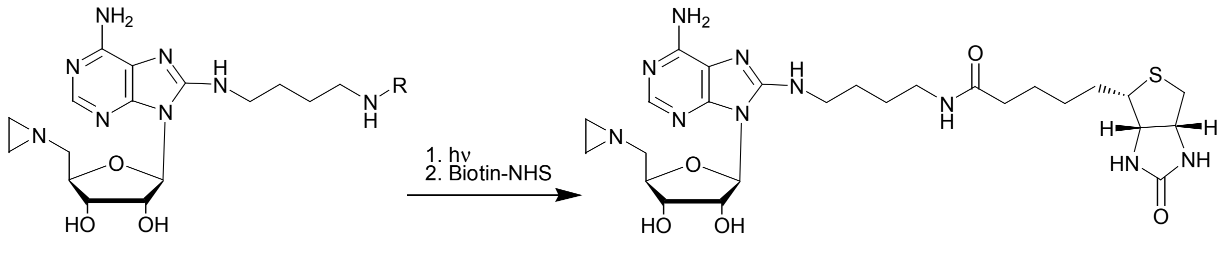


*Fig. 3.2.4: UV-spectra of the deprotection of the N-adenosylaziridine precursor **19** after different laser pulses.*

The curve in black illustrates the initial spectrum. At about 280 nm the absorption of the nucleobase is visible and at 340 nm the NVOC chromophore absorbes. After irradiation at 355 nm the absorption at 340 nm decreases and the absorption at 400 nm increases which indicates deprotection and formation of compound **25**. When the resulting compound is injected onto a reversed phase HPLC column the signal of the initial precursor (R_t = 48.9 min) totally disappears and a new signal with a retention time of 22 min arises. The novel compound was analyzed by mass spectrometry and shown to be the deprotected amine **20**. Experiments on a larger scale were performed on solutions of 100-200 ml. The protected nucleoside **19** was dissolved in several µl DMSO and triethanolamine buffer was added to give a solution of $OD_{340\,nm} = 1$. The solution was irradiated with a mercury lamp until the starting material was converted completely into the free amine **20**.

3.2.5 Synthesis of the biotinylated cofactor analogue

The solution of free amine **20** was treated with biotin-NHS ester to yield the biotinylated cofactor analogue BAZ **27**. As the biotin-NHS ester is not light sensitive deprotection and coupling with biotin can be performed in a one pot reaction (Scheme 3.2.5).

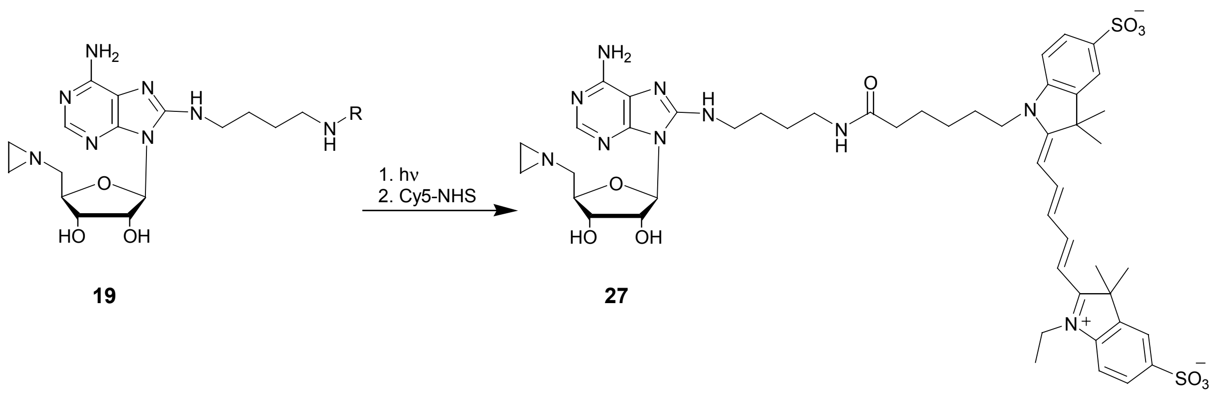


Scheme 3.2.5: Synthesis of the biotinylated cofactor analogue BAZ ***(26)****. R represents the NVOC-protecting group.*

The BAZ cofactor analogue was purified by HPLC chromatography and was suitable for usage in biochemical assays. The yield of **26** with respect to **19** is 45 %.

3.2.6 Synthesis of a fluorescent cofactor analogue

A fluorescent cofactor analogue was synthesized after deprotection of the NVOC group. The Cy5 fluorophore was chosen because of its excellent optical characteristics. The solution of the free amine **20** was lyophilized to a volume of 1 ml and treated with Cy5-NHS ester yielding the Cy5AZ cofactor analogue **27** (Scheme 3.2.6).



Scheme 3.2.6: Synthesis of the fluorescent cofactor analogue Cy5AZ ***(27)****. R represents the NVOC-protecting group.*

The cofactor analogue **27** was purified by reversed phase HPLC and was suitable for biochemical assays. The identity of Cy5AZ was verified by ESI-MS as only small amounts of the cofactor analogue were synthesized due to the high costs of the fluorophore.

3.3 Binding of AZ cofactors to M·*Taq*I

The binding of the two newly synthesized cofactors to M·*Taq*I was investigated using two different binding assays. The biotinylated cofactor analogue BAZ does not contain an intrinsic optical property which could be used in a binding assay. Therefore, the binding affinity of BAZ to M·*Taq*I was determined using a competitive assay (Pignot, 1999). M·*Taq*I was preincubated with a fluorescent cofactor analogue, MANT-AdoHcy. The binding constant of MANT-AdoHcy to M·*Taq*I is known. Addition of BAZ to this solution leads to a decrease of the MANT fluorescence as the MANT-AdoHcy is displaced from the binding pocket by BAZ. The fluorescence was measured in a fluorescence resonance energy transfer experiment (FRET) in which the tryptophan residues of the protein are excited and the energy is transferred to the MANT fluorophore whose fluorescence is detected subsequently. The decrease of the MANT fluorescence was observed while the concentration of BAZ was increased (Fig. 3.3a).

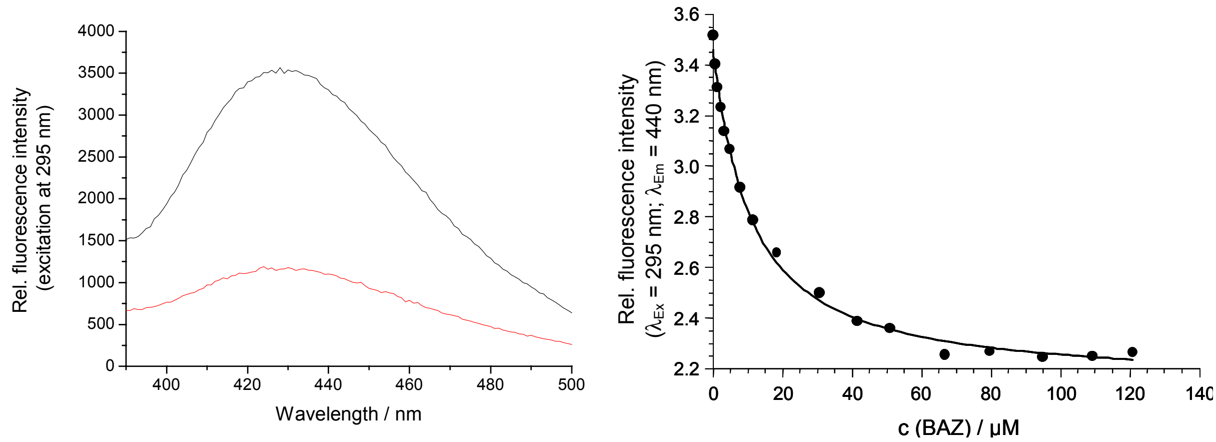


Fig. 3.3a: Competitive binding between BAZ and MANT-AdoHcy to M·TaqI. Emission spectra at the beginning (black) and the end (red) of the titration with BAZ (left). Plot of the measured fluorescence intensity versus the concentration of BAZ (right).

The data points were fitted with the program Scientist and a K_D value of BAZ and M·*Taq*I of 8.2 µM was calculated. Therefore, the cofactor analogue binds tightly to M·*Taq*I. The binding constant is about 4 times higher than the binding of the natural cofactor AdoMet (2 µM) but at least one order of magnitude smaller than the binding of AZ (206 µM) to M·*Taq*I.

The binding of the Cy5Az cofactor analogue was investigated making use of its intrinsic fluorescent properties. The change of the Cy5 fluorescence in dependence of added protein was monitored. The data with the fitted binding curve are shown in Fig. 3.3b.

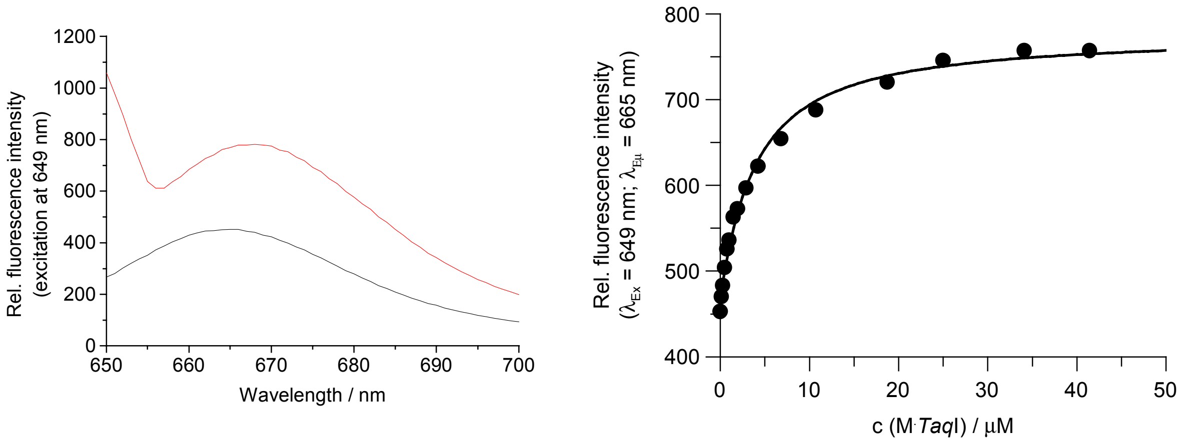
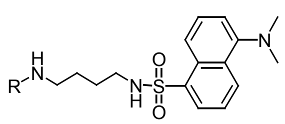
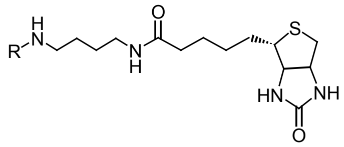
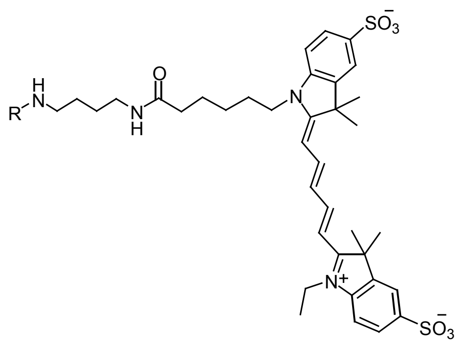


Fig. 3.3b: Cofactor binding of Cy5Az to M·TaqI. Emission spectra at the beginning (black) and the end (red) of the titration (left). Plot of the measured fluorescence intensity versus the concentration of M·TaqI (right). The determined K_D value is 3.4 ± 0.4 µM.

The fitted data result in a K_D of 3.4 ± 0.4 µM. The determined K_D value demonstrates that the binding of Cy5Az to M·*Taq*I is tight and that the large fluorophore does not negatively interfere with the protein. The dissociation constant is in the same range as the binding constant of the natural cofactor AdoMet. As SMILing DNA with a fluorescently labeled cofactor analogue was already performed during my diploma research work the focus was directed to the biotinylated cofactor analogue. SMILing with BAZ was investigated in order to prove that indirect labeling systems are suitable for this labeling technique. A list of the different AZ derivatives with the corresponding binding constants to M·*Taq*I are listed below (Table 3.3).

Table 3.3: Dissociation constants of the different cofactor analogues (the *N*-adenosylaziridinyl moiety is represented as R)

Abbreviation	Side chain at the 8-position	K_D value / µM
AZ	R–H	206.1[a]
DAZ		0.5 ± 0.13[b]
BAZ		8.2
Cy5AZ		3.4 ± 0.4

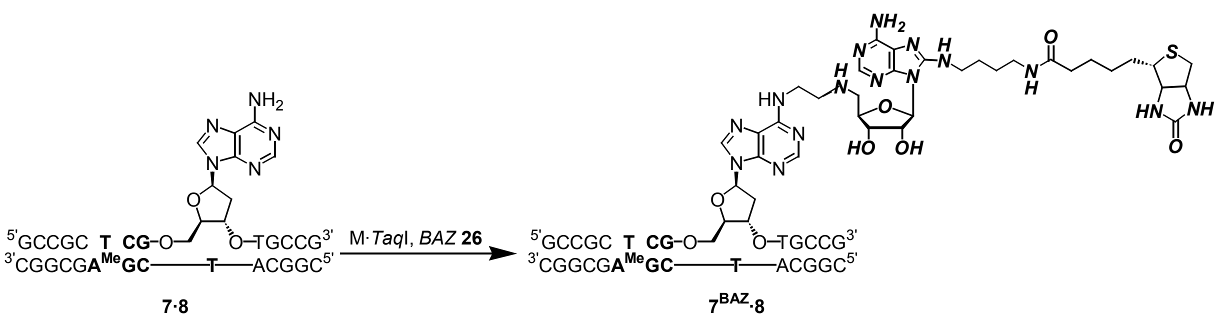
[a]Pignot, 1999.

[b]Pljevaljčić, 1999.

3.4 SMILing of different DNA substrates using BAZ and M·*Taq*I

3.4.1 SMILing of short duplex oligonucleotides

The first experiments to prove SMILing DNA using BAZ were carried out with the short duplex oligonucleotides already used for the prior experiments with M·*Taq*I. The SMILing reaction with the biotinylated cofactor should lead to a biotinylated DNA as illustrated below (Scheme 3.4.1).



*Scheme 3.4.1: M·TaqI-catalyzed biotinylation of **7·8**.*

The coupling reaction was carried out at 37 °C at pH 6. Aliquots were taken from the reaction solution after different reaction times and injected onto an anion exchange column on a HPLC system. The anion exchange chromatograms obtained are shown in Fig. 3.4.1a.

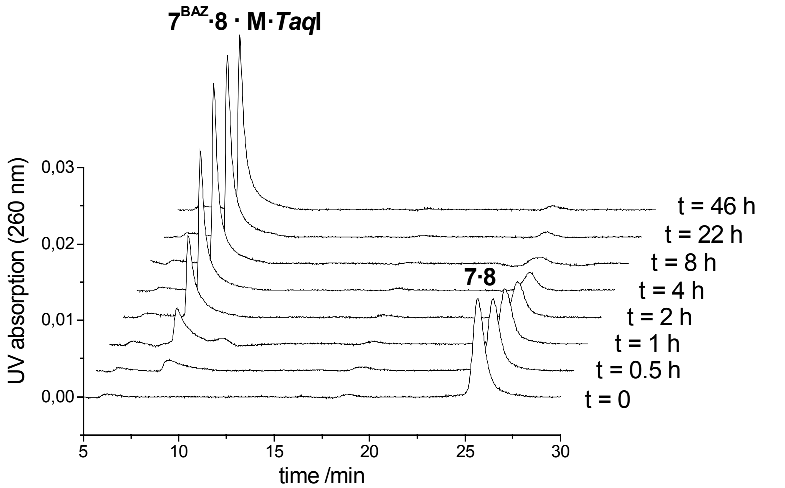


Fig. 3.4.1a: M·TaqI-catalyzed coupling reaction between ***7·8*** *and BAZ* ***26*** *at 37 °C. Anion exchange chromatograms with recorded absorption at 260 nm.*

The initial sample (t = 0) shows only duplex oligonucleotide **7·8** with a retention time of 25.6 min. As the reaction proceeds a product with a retention time of 8.2 min is formed. After 8 h almost all of the DNA is converted into product. UV-spectra of the product suggest, as already observed in the SMILing reaction with DAZ, that the product consists of a complex of modified DNA and protein. In order to release the DNA from the DNA-protein complex, the complex was treated with proteinase K for 1 h at 37 °C. The released DNA $\mathbf{7^{BAZ}{\cdot}8}$ elutes with a smaller retention time than the unmodified DNA. The anion exchange chromatograms of the starting DNA and the product is shown in Fig. 3.4.1b.

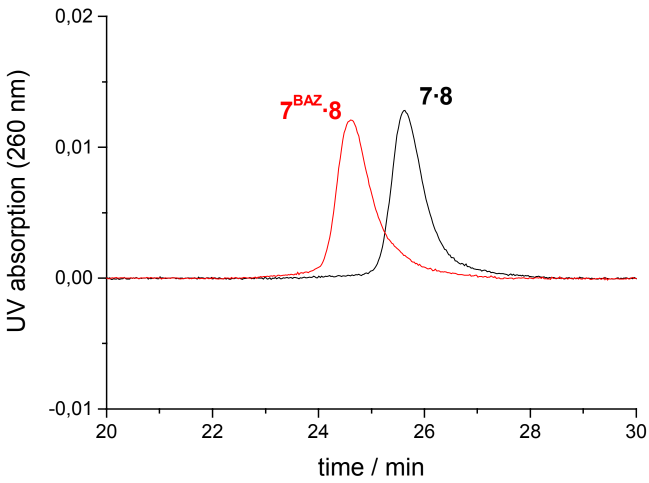


Fig. 3.4.1b: Details of the anion exchange chromatograms of DNA before and after the M·TaqI-catalyzed coupling reaction with BAZ at 37 °C.

The retention time of the biotinylated DNA (24.6 min) is smaller than that of the non-modified DNA. This observation parallels the observation that the retention time of DNA is smaller after the coupling reaction with AZ. However, in the case of SMILing DNA with DAZ the retention time of the modified DNA increased relative to the non-modified DNA. This can be explained with the more hydrophobic character of the dansyl fluorophore after modification with DAZ in comparison to both modifications with AZ and BAZ. The dansyl group therefore interacts more tightly with the anion exchange material and elutes with a bigger retention time. All the observations strongly suggest that the DNA was biotinylated. Confirmation of this assumption should come from mass spectrometry and X-ray crystallography.

The modified DNA $\mathbf{7^{BAZ}\cdot 8}$ was isolated from the anion exchange column and desalted prior to preparation of a MALDI mass spectrum. The obtained spectrum is shown in Fig. 3.4.1c. The mass spectrum was calibrated using the single-stranded oligonucleotide **7**.

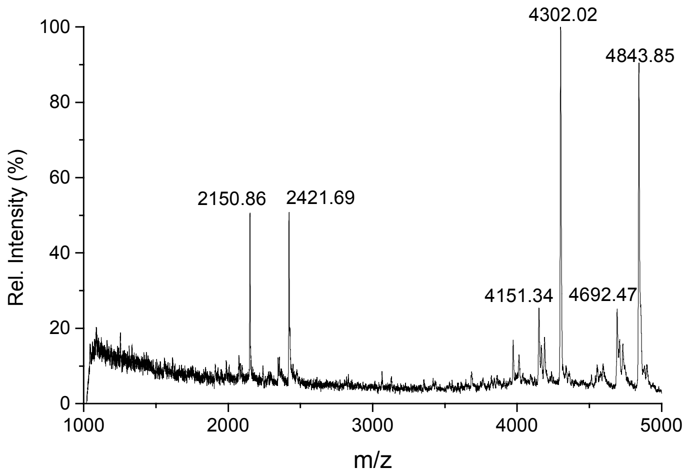


Fig. 3.4.1c: MALDI mass spectrum of the modified duplex oligonucleotide 7^{BAZ}·8.

Four major signals appear in the mass spectrum. The signal at 4,302.0 corresponds to strand **8** (calc. 4,300.8 [**8**-H]$^-$). The double negatively charged strand **8** gives a signal at 2,150.9 (calc. 2,149.9 [**8**-2H]$^{2-}$). The modified strand $\mathbf{7^{BAZ}}$ gives a signal at 4,843.9 (calc. 4,841.0 [$\mathbf{7^{BAZ}}$-H]$^-$). The corresponding double negatively charged species gives a signal at 2,421.7 (calc. 2,420.0 [$\mathbf{7^{BAZ}}$ -2H]$^{2-}$). The two signals at 4,141.3 and 4,692.5 correspond to depurinated DNA strands in which a guanine was released from the original strand. MALDI mass spectrometry clearly confirms that the duplex oligonucleotide **7·8** becomes biotinylated at strand **7** which contains the non-methylated recognition sequence of M·*Taq*I. In order to get information about the base which is modified X-ray crystallography is applied (see chapter 3.5).

3.4.2 Single-turnover experiments at different pH-Values

The influence of the pH value on the reaction rate was investigated for SMILing with BAZ. The data should be consistent with that of the coupling reaction of AZ with DNA under M·*Taq*I-catalysis. The SMILing reaction with BAZ was carried out at different pH values and the reaction rate determined by analysis of the HPLC signals. Fig. 3.4.2 (left) shows a plot of the amount of modified DNA versus reaction time. The observed rate constants were plotted in Fig. 3.4.2 (right) versus the pH values.

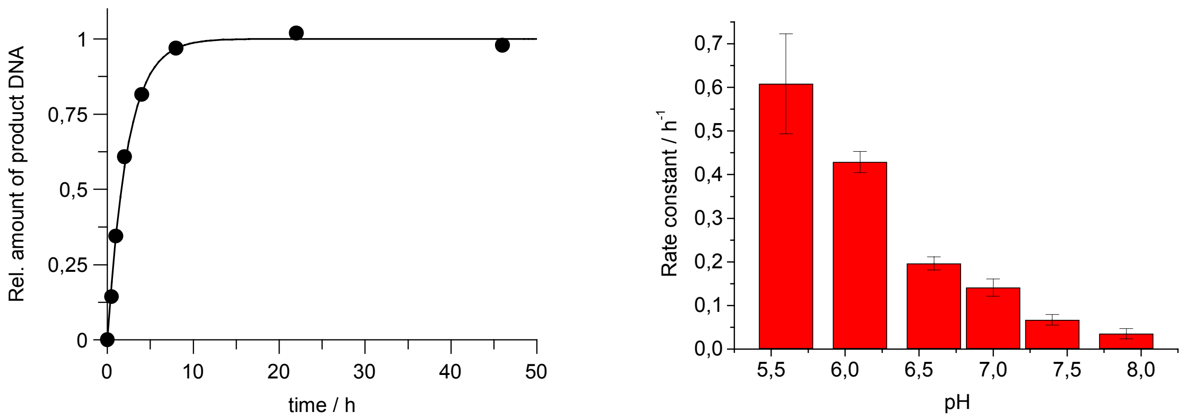


*Fig. 3.4.2: Plot of the amount modified DNA formed with time at pH 6.1 (left). The data points were fitted with a single exponential function using the software GraFit. Diagram of the pH-dependence of the reaction rates for the M·TaqI-catalyzed reaction of BAZ with the duplex oligonucleotide **7·8** (right).*

The plot illustrates that the reaction is pH dependent. The coupling reaction of AZ and BAZ both perform the reaction with approximately the same maximum rate constant which suggests that the modification on the 8-position of the cofactor does not have a significant influence on the maximum reaction velocity. Therefore, in the case of N8-modified cofactor analogues the raction rate can be increased with slightly acidic conditions. However, the binding of BAZ to the protein is about 25 fold tighter than that of AZ which shows an additional advantage of the N8-modified analogues.

3.4.3 SMILing of a 47-mer duplex oligonucleotide

In collaboration with the group of Dr. Andres Jäschke, FU Berlin (Germany), the SMILing technique has been used for the development of a selection system for the isolation of deoxyribozymes. In the group of Dr. Andres Jäschke it was previously shown that terminal functionalized RNA could be used to select ribozymes which catalyze the Diels-Alder reaction (Seelig & Jäschke, 1998). With SMILing DNA an internal functionalization of DNA should be achieved which would expand the diversity of the DNA library for selection experiments. This internally modified DNA should be investigated for the suitability to be amplified using PCR. Such an amplifiable system is a prerequisite for the selection of deoxyribozymes.

Here the principal of such a selection system was investigated using BAZ for internal modification. A 47-mer duplex oligonucleotide (JSQ) with a single recognition site for M·*Taq*I was labeled with biotin using the reaction described in 3.4.1. In JSQ one adenine within the recognition site is methylated, thus allowing only the labeling of the non-methylated strand. The reaction was performed at 65 °C for 4 h. After that the solution was treated with proteinase K for 1 h at 37 °C. In a control experiment the same reaction was performed with a 47-mer oligonucleotide (JSN) without a M·*Taq*I recognition site. Denaturing acrylamide gel electrophoresis was used to analyze the reaction. A fluorescent streptavidin-fluorescein conjugate was utilized to realize a shift of biotinylated oligonucleotide. The gel is shown below (Fig. 3.4.3). The staining of the bands was perfomed using SYBR-Gold.

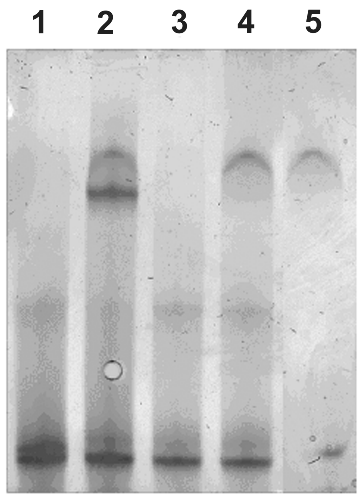
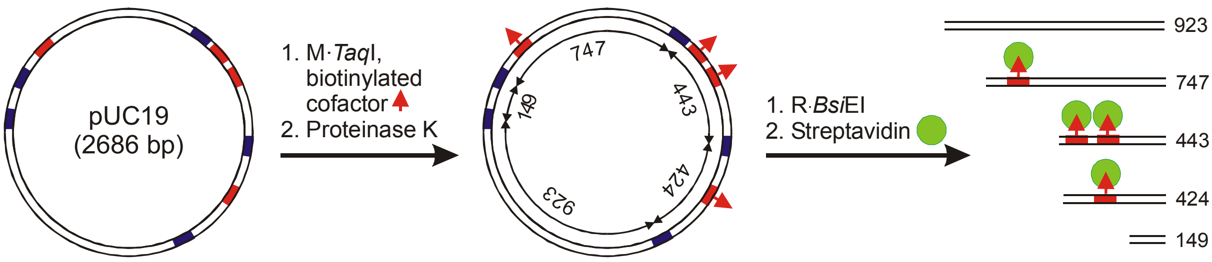


Fig. 3.4.3: Denaturing acrylamide gel illustrating the labeling of a 47-mer duplex oligonucleotide (JSQ) with biotin after treatment with proteinase K. (Lane 1: JSQ; Lane 2: JSQ plus streptavidin-fluorescein; Lane 3: JSN; Lane 4: JSN plus streptavidin-fluorescein; Lane 5: streptavidin-fluorescein).

Principally, in a denaturing acrylamide gel containing urea double-stranded DNA becomes denatured. For biotinylated JSQ one can see that one strand is shifted when streptavidin-fluorescein is added (Lane 2). This indicates that one strand of JSQ is biotinylated. Comparison of the intensities of the two bands allows the assumption that the biotinylation occurs almost quantitatively. In contrast, the DNA without M·*Taq*I site (JSN) is not shifted in the presence of streptavidin-fluorescein, which indicates that SMILing is indeed sequence-specific. Thus, it can be concluded that SMILing of a 47mer duplex oligonucleotide with biotin using M·*Taq*I and BAZ works fine. Most interestingly a subsequent PCR reaction performed in the group of Dr. Jäschke leads to amplification of the biotinylated strand. This shows that biotinylated DNA obtained by SMILing DNA can be used as a template for DNA polymerases. Having performed this labeling successfully, the same reaction was performed with a pool of 47-mer DNA. The DNA containing the 5'-TCGA-3' sequence (expected 10 %) should be labeled. Within the pool 5 % biotinylated DNA was found which demonstrates the labeling of the DNA containing 5'-TCGA-3' sites. These experiments represent a proof of principle for internal functionalization of DNA pools which can potentially be used to establish a selection system of deoxyribozymes.

3.4.4 SMILing of plasmid DNA

The SMILing reaction of large DNA with BAZ was investigated using the plasmid pUC19 (Scheme 3.4.4). The plasmid contains four recognition sites for M·*Taq*I. After treatment of pUC19 with M·*Taq*I and BAZ, one would expect four biotin labels on the plasmid. Cutting the plasmid afterwards with the restriction endonuclease R·*Bsi*EI which contains five restriction sites on pUC19, should lead to five DNA fragments of different size which either do contain the biotin label (fragments 424, 443 and 747 bp) or do not (fragments 149 and 923 bp). After addition of streptavidin the biotinylated DNA fragments should give a gel shift in an agarose gel.



Scheme 3.4.4: M·TaqI-catalyzed labeling of pUC19. Red bars denote M·TaqI recognition sites and blue bars denote R·BsiEI restriction sites.

The agarose gel for this experiment is shown below (Fig. 3.4.4).

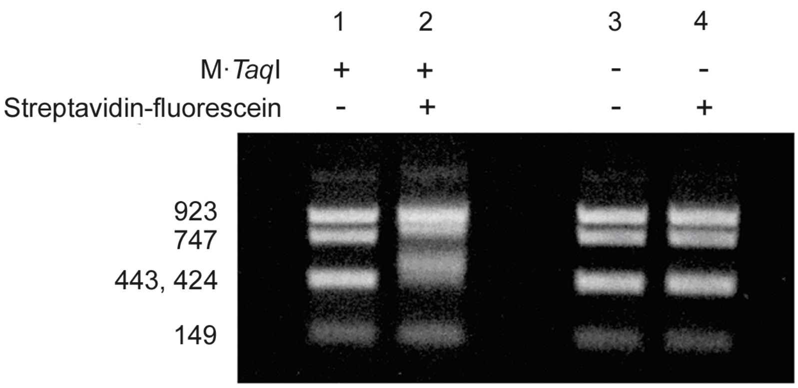


Fig. 3.4.4: Agarose gel of pUC19 restriction fragments after SMILing with BAZ and M·TaqI.

Lane one shows the fragments after the reaction. Four fragments are visible. Two fragments (fragments with the length 443 and 424 bp) cannot be separated on the agarose gel. When streptavidin-fluorescein is added to the reaction mixture the agarose gel shows a shift for the bands with a length of 443, 424 and 747 bp (lane 2). Almost all of those fragments are shifted upwards due to the higher molecular weight after streptavidin binding. The appropriate control experiment without M·*Taq*I in the reaction mixture is shown on the right half of the agarose gel (lane 3 and 4). The fragments in lane 3 behave as those in lane 1. However, only when streptavidin is added (lane 4) is no shift observed compared to lane 3. These results demonstrate that exactly the expected fragments are labeled with biotin and the appropriate control experiment reveals no labeling. This experiment also verifies the sequence-specificity of the SMILing DNA technique as only the fragments containing the M·*Taq*I recognition sequences are labeled and shifted. Furthermore, the experiment shows for the first time that the labeling reaction is almost quantitative with plasmid DNA. In summary, SMILing DNA is demonstrated for different DNA substrates, with different AZ cofactors and different DNA MTases. The reactions proceed slowly compared to the reactions with the natural cofactor but occur with the same specificity and quantity.

3.5 Crystal structure of the complex between M·*Taq*I and with BAZ biotinylated DNA

A very useful technique to visualize structures at atomic resolution is X-ray crystallography. Therefore, this technique was used to directly prove SMILing DNA with BAZ. For this purpose M·*Taq*I with cognate DNA and BAZ was crystallized. The structure obtained should not only visualize the covalently attached cofactor to the DNA but also give further insights into the mechanism of the action of the MTase.

3.5.1 Crystallisation conditions and data collection

Crystallization setups were performed using the hanging drop method. Crystals of the complex between M·*Taq*I and with BAZ biotinylated DNA could be obtained as described in the experimental section (5.5.2). Crystals grew only after micro seeding with crystals of the ternary complex between M·*Taq*I, DNA and the non-reactive cofactor analogue AETA yielding thin needles (maximal size 30 x 30 x 100 μm^3) and plates (maximal size of 10 x 30 x 100 μm^3) (Fig. 3.5.1a).

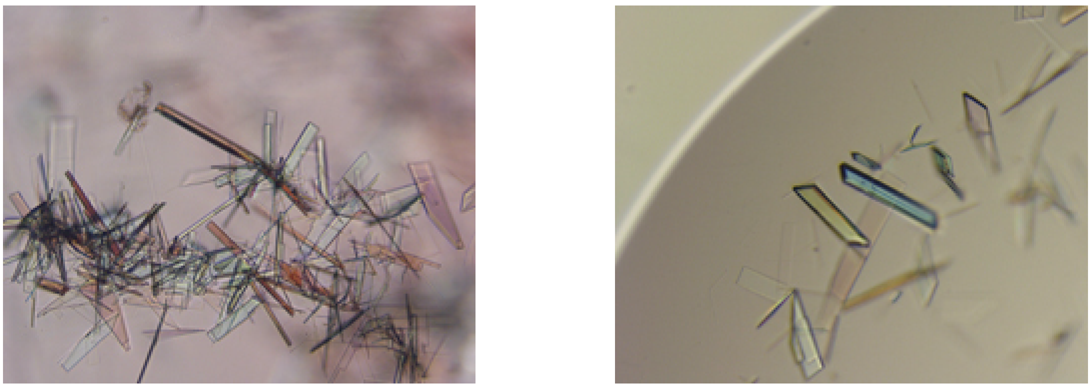


Fig. 3.5.1a: Crystals of the complex between M·TaqI and with BAZ biotinylated DNA. Crystals within the seed line (left). Single crystals suitable for data collection (right).

Crystals of the complex between M·*Taq*I and with BAZ biotinylated DNA, like in Fig. 3.5.1a (right), together with a cryoprotectant solution were frozen in liquid nitrogen. Data were collected at the European Synchrotron Radiation Facility (ESRF) in Grenoble (France), at the ID14 beamline for macromolecular crystallography. The measurement was performed at 100 K using a MAR CCD detector. An example of a rotation diffraction image is shown in Fig. 3.5.1b. The ring at 3.6 Å possibly resulted from ordered components in the cryoprotectant. Diffraction signals are visible up to 1.9 Å. Processing of data was performed using the programs XDS and XSCALE (Kabsch *et al.*, 1993). The crystals of the complex crystallize in the space group $P2_1$. The data statistics are given in Table 3.5.2 (top).

Fig. 3.5.1b: Rotation diffraction image of the co-crystals between M·TaqI and biotinylated DNA detected with a MAR CCD area detector (crystal to detector distance: 150 mm; Pixel size of 0.0793mm). Resolution circles are presented up to 1.9 Å. Intensities of reflexes up to 1.9 Å could be measured with the processing software.

3.5.2 Phase determination and structure refinement

The phases of the diffracted X-rays cannot be determined in the performed diffraction experiment. They were determined by molecular replacement using the ternary complex described by Goedecke and co-workers (2001) as starting model. Since the processing of data indicated that the obtained complex crystallizes in the same space group as the ternary complex between M·*Taq*I, DNA and AETA (Goedecke *et al.*, 2001) and since the cell parameters were similar it was not necessary to apply the translation and rotation search functions. The first electron density distribution was obtained by combination of the structure amplitudes from this experiment and the calculated phases from the structure of Goedecke *et al.* (2001). In order to avoid a bias by the structure of Goedecke *et al.* (2001) the flipped out adenine and the cofactor analogue were left out for the calculation. A first electron density map ($2F_{obs}$-F_{calc}) immediately shows the missing structural information of the flipped out base and the cofactor. Alternate rounds of refinement using the program CNS and modelling of the amino acid residues using the program O (Jones *et al.*, 1991) were performed until minimization of the residual factor (R-factor) between the calculated structure factor amplitudes (F_{calc}) and the observed structure factor amplitudes (F_{obs}) converged.

The final values for the crystallographic R-factor and the free R-factor are 0.2004 and 0.2359, respectively. For the calculation of the free R-factor 5 % of all reflexes were selected randomly (Brünger, 1992). A list of the refinement statistic is given in Table 3.5.2 (bottom).

Table 3.5.2: Crystallographic data and refinement results for the complex between M·*Taq*I and with BAZ biotinylated DNA.

Data statistics	
Beamline	ID14-1/ESRF
Temperature	100 K
Area detector	MAR CCD
Wavelength	0.934 Å
Space group	$P2_1$
Cell dimensions	a = 59.40 Å, b = 69.19 Å, c = 114.39 Å
	β = 92.16°
Number of recorded reflections	272,799
Average redundancy	4.1
R_{sym} [a]	7.0 %
Intensities (I/σ) [b]	15.6 (5.5)
Refinement statistics	
Resolution limit	1.9 Å
Number of unique reflections	72,765
Completeness of data (%) [c]	99.4 (92.1)
R_{cryst} [d] / R_{free} [e]	0.2004 / 0.2359
Rmsd bond lengths (Å)	0.005
Rmsd bond angle (deg.)	1.56
Mean/rms on B factors ($Å^2$)	
Backbone	18.68 / 7.54
Side chain	24.83 / 14.01
Covalently bound cofactor	23.67 / 6.43
DNA	17.07 / 6.83
Solvent (687 water molecules)	29.44 / 9.83

[a] $R_{sym} = 100 \times \sum I - \langle I \rangle / \sum I$, I is the observed intensity and $\langle I \rangle$ is the average intensity calculated from multiple observations of symmetry related reflections.

[b] The value in parentheses is calculated for the highest resolution shell collected (2.0-1.9 Å).

[c] The value in parentheses is calculated for the highest resolution shell used in refinement (2.0-1.9 Å).

[d] $R_{cryst} = \sum \left| F_{obs} - F_{calc} \right| / \sum_{hkl} \left| F_{obs} \right|$, where F_{obs} and F_{calc} are, respectively, observed and calculated structure factor amplitudes.

[e] R_{free} is an R_{cryst} calculated using 5 % of the processed data (3639 reflections), chosen randomly, kept constant and omitted from all structure refinement steps.

3.5.3 Three-dimensional structure of the complex between M·*Taq*I and with BAZ biotinylated DNA

The structure of M·*Taq*I and biotinylated DNA crystallizes with two complexes in the asymmetric unit. However, the differences between the two complexes are negligible. For this reason only complex A will be further described due to its better defined electron density map and smaller *B*-factors. The overall structure of the obtained complex is identical with the structure of M·*Taq*I, DNA and AETA presented by Goedecke and co-workers (2001). The root mean square deviation of the C*a* atoms (3.097 Å complex A, 0.199 Å complex B) between the two structures reveals that there is almost no change in the structure of the protein and also the shape of the DNA remains the same. The slightly higher root mean square deviation in complex A arises from a rotated loop which is probably due to crystal packing. A detailed description of the structure with all interactions between protein, DNA and cofactor analogue can be found in the thesis of Goedecke (Goedecke, 2000). Here, the attention is only given to the active site where the main differences are expected. The active site of the complex between M·*Taq*I, DNA and AETA is shown in Fig. 3.5.3a.

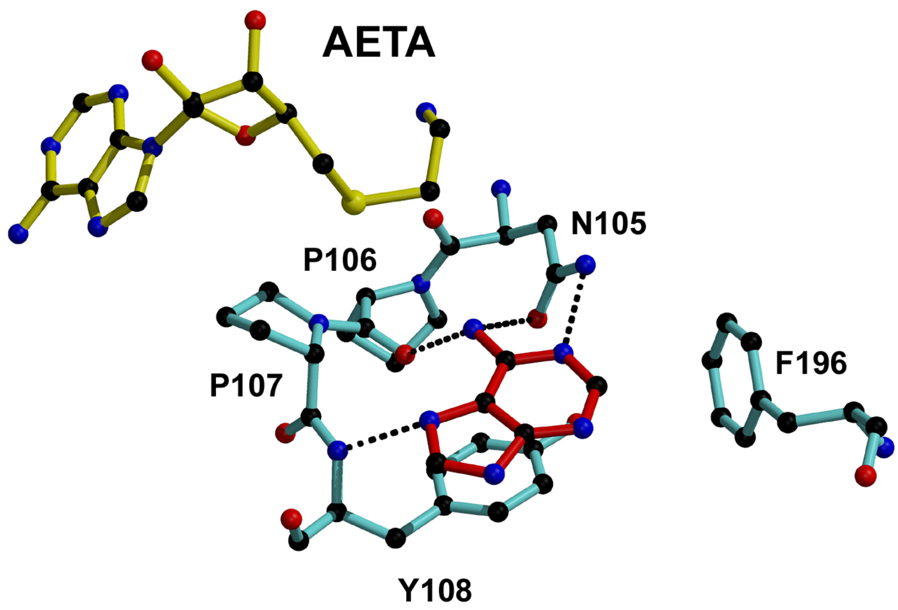


Fig. 3.5.3a: Image of the crystal structure of the active site of M·TaqI (the cofactor analogue AETA is yellow, the NPPY motif and F196 cyan and the flipped out base is red); Goedecke et al., 2001.

The amino acid residues from the so-called NPPY motif and F196 (cyan) stabilize the flipped out adenine (red) in its position. The cofactor analogue AETA is represented in yellow.

The crystallization of M·*Taq*I, DNA and BAZ can potentially result in two different structures. The crystal structure will either reveal an intact cofactor and DNA or a product in which the aziridine three-membered ring is opened by the N6 amino group of the flipped out adenine base. The first case would result in a structure of the active site similar to Fig. 3.5.3a. A model based on the structure M·*Taq*I, DNA and AETA containing the cofactor AZ instead of AETA was prepared. AZ was used instead of BAZ as the biotin moiety should not be important for the coupling reaction with the flipped out adenine. This modeled structure is shown in Fig. 3.5.3b. The adenosyl part of the AZ cofactor is expected to bind in the same way as observed for the cofactor analogue AETA. For the modeling of the aziridine ring one CH_2-unit of the aziridine moiety was chosen to point towards the N6-position of the flipped out base.

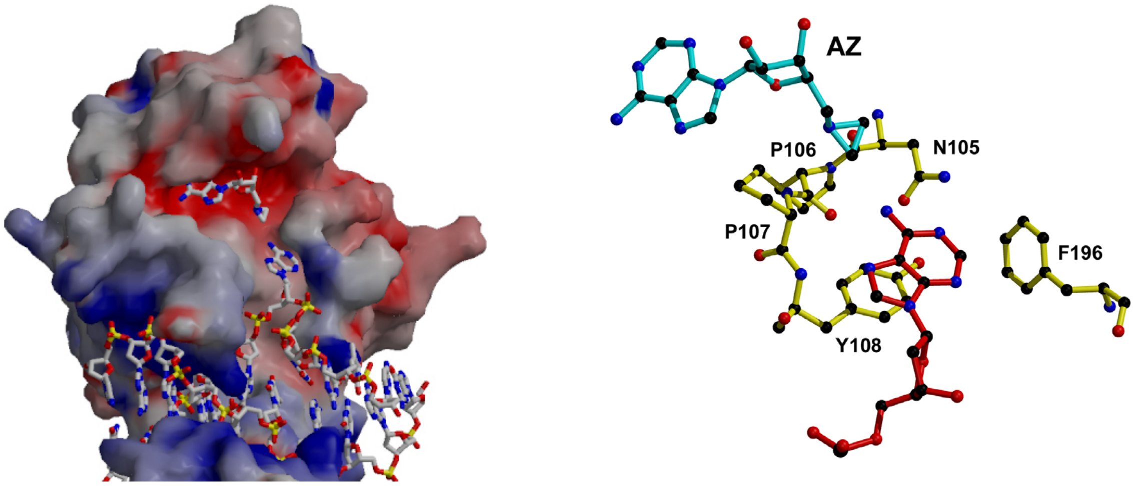


*Fig. 3.5.3b: Model of M·*Taq*I in complex with DNA and N-adenosylaziridine. The protein is represented as an electrostatic surface potential (red = negative, blue = positive and white = neutral) and DNA and modeled AZ are shown as ball and stick representation (left). Active site of M·*Taq*I with AZ represented in cyan, the NPPY motif and F196 in yellow and the flipped out nucleotide in red (right). The model is based on the crystal structure by Goedecke et al., 2001.*

Such a structure would be expected if the coupling reaction does not take place during crystallization. However, the crystallization time and the buffer conditions should allow for a reaction between DNA and BAZ on the enzyme. This would lead to a covalent link between the exocyclic amino group of the flipped out base and the aziridine moiety. The electron density of the obtained complex shows that this is indeed the case. The active site of M·*Taq*I with the electron density map ($2F_{obs}$-F_{calc}) for the flipped out nucleotide and the cofactor is shown in Fig. 3.5.3c (left). On the right side of Fig. 3.5.3c the previously shown modeled AZ and the extrahelical target nucleotide are modeled into the electron density map.

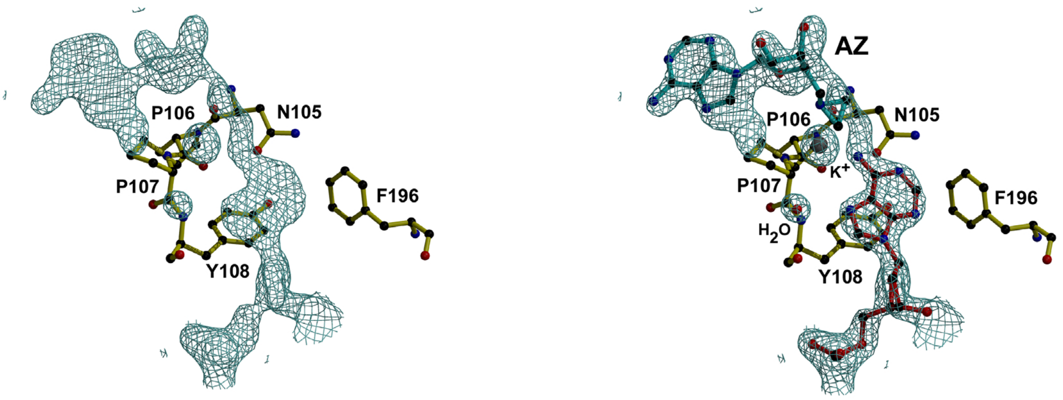


*Fig. 3.5.3c: Active site of M·*Taq*I superimposed with the electron density distribution ($2F_{obs}$-F_{calc}) contoured at 2.0 σ (left). For clarity reasons the electron density is only shown for the flipped out nucleotide and the cofactor. Active site of M·*Taq*I superimposed with the electron density distribution ($2F_{obs}$-F_{calc}) contoured at 2.0 σ with the modeled AZ, the extrahelical target nucleotide, a potassium ion and water (right).*

It can be clearly seen that AZ does not describe properly the determined electron density map. The adenosyl moiety is slightly out of the map and a continuous electron density map between the flipped out base and the cofactor analogue is visible. This clearly argues for a covalent bond between the cofactor and DNA. In Fig. 3.5.3d the structure of BAZ with an opened aziridine ring is fitted into the observed electron density.

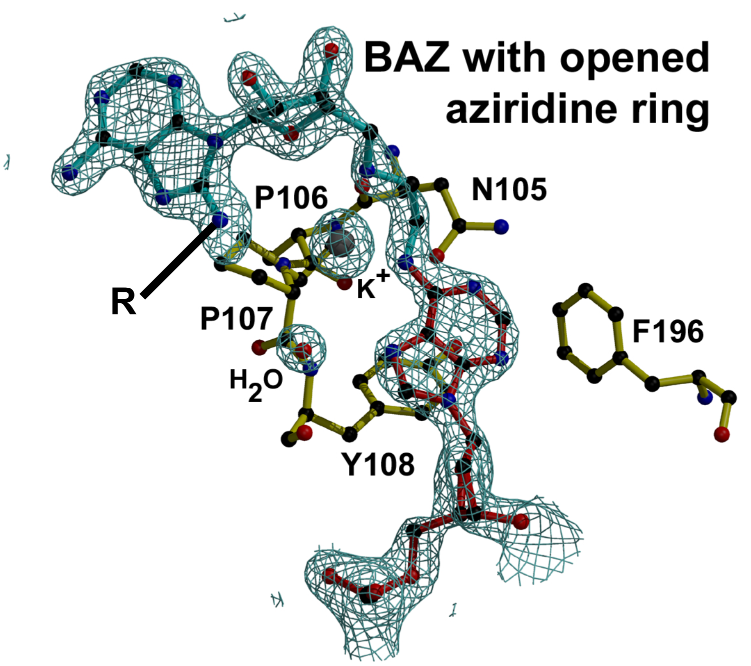


*Fig. 3.5.3d: Active site of M·*Taq*I superimposed with the electron density distribution ($2F_{obs}$-F_{calc}) contoured at 2.0 σ. Only the density for the cofactor and the flipped out base is shown. The cofactor BAZ with an opened aziridine ring, the extrahelical target nucleotide, a potassium ion and a water molecule are modeled into the electron density map. Only those parts of BAZ are modeled for which electron density is observed. R represents the non-modeled part of BAZ.*

Here the adenosyl moiety of BAZ fits nicely into the observed electron density map. Furthermore, the aziridine ring is undoubtedly opened forming a covalent bond with the exocyclic amino group of the flipped out adenine. This suggests that product formation took place in the crystals or a product complex was crystallized showing the already labeled DNA. For the tethered biotin no well defined electron density could be observed. The electron density describes the molecule nicely until the first linker nitrogen. This can be explained by a high flexibility of the linked biotin moiety. An additional feature visible in the co-crystal structure are two strong peaks of electron density which were assigned to potassium ions. One is located in the active site as illustrated in Fig. 3.5.3d and the other is found in the minor groove of the DNA. The potassium ion in the active site is coordinated by five atoms as shown in Table 3.5.3a. It is not clear if the potassium ion is involved in the reaction mechanism or just present due to the new coordination environment created by the N8-nitrogen of the 1,4-diaminobutane linker. These two potassium ions were not present in the structure of M·*Taq*I, DNA and AETA.

Table 3.5.3a: Coordination of the potassium ion in the active site.

Potassium coordinating atom	d / Å
Extrahelical adenine-N6	3.31
Extrahelical adenine-N7	3.86
Water molecule-O	3.44
BAZ-N8	3.78
BAZ-N10 (aziridine nitrogen)	3.07

Since the experiment resulted the product structure one can speculate about the motion of the substrates within the active site during the reaction by comparison of the model with AZ and the actual structure with the opened BAZ (Fig. 3.5.3e). The structures are superimposed based on the NPPY motif. One can imagine that the reaction occurs by the movement of the cofactor BAZ and the flipped out base towards each other. The flipped out adenine moves towards the cofactor and *vice versa*. F196 follows the flipped out nucleotide.

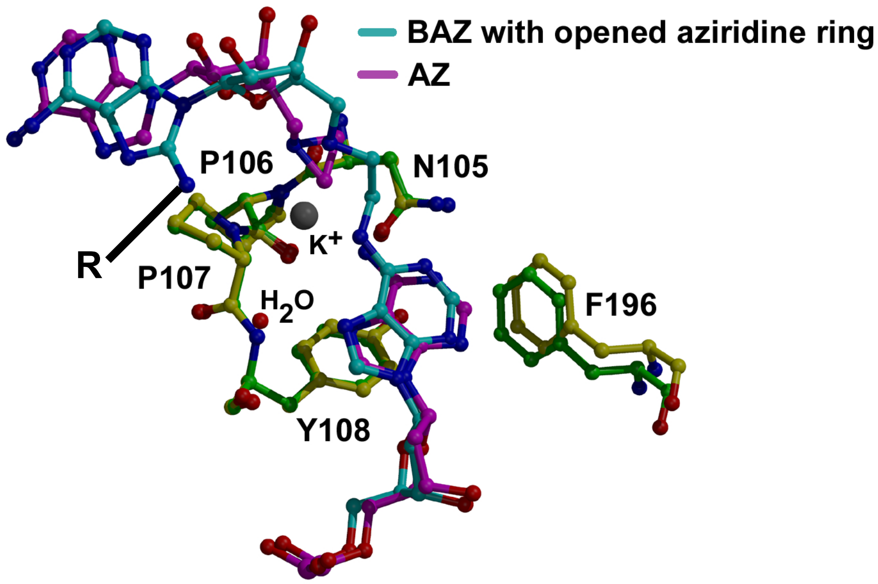


*Fig. 3.5.3e: Superposition of the co-crystal structure formed with M·*Taq*I, DNA and BAZ (NPPY motif and F196 are represented in green) and the model of the M·*Taq*I•DNA•AZ ternary complex (NPPY motif and F196 are represented in yellow) within the active site of M·*Taq*I. The opened BAZ cofactor and the flipped out nucleotide are cyan, the AZ model and the corresponding nucleotide are magenta. A potassium ion is represented with a grey sphere. R represents the non-modeled part of BAZ.*

This movement can also be followed by the comparison of the hydrogen bonding distances in the obtained structure and the structure of M·*Taq*I, DNA and AETA. The important distances between the exocyclic amino group of the flipped out base and its N1 position and the protein are illustrated with dotted lines in Fig. 3.5.3f.

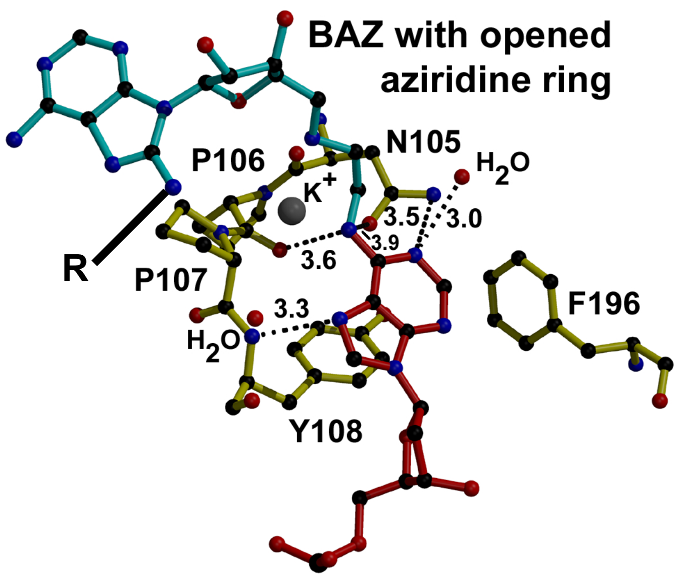


*Fig. 3.5.3f: Structure of the obtained complex within the active site of M·*Taq*I (compare Fig. 3.5.3d) showing the discussed hydrogen bonds as dotted lines. R represents the non-modeled part of BAZ.*

The illustrated hydrogen bonds are also listed in Table 3.5.3b, and compared with the corresponding hydrogen bonds in the ternary M·*Taq*I•DNA•AETA complex.

Table 3.5.3b: Hydrogen bond distances.

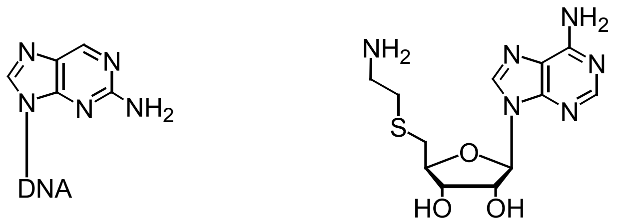
	d(structure M·*Taq*I•DNA•BAZ) / Å	d(Goedecke *et al.*, 2000) / Å
Water-O..N1	3.0	-
Y108-N..N7	3.3	3.1
N105-ND2..N1	3.5	3.0
N105-OD1..N6	3.9	3.2
P106-O..N6	3.6	2.9

It is striking that in the co-crystal structure formed from M·*Taq*I, DNA and BAZ the hydrogen bond distances between N1 and N6 of the flipped out adenine and the protein are longer than in the structure of M·*Taq*I, DNA and AETA by Goedecke and co-workers (2001). This illustrates that the flipped out adenine moves upwards (compare Fig. 3.5.3e) and towards the cofactor in the product complex. Another feature which is not present in the structure between M·*Taq*I, DNA and AETA is the water molecule in the active site which forms a hydrogen bond to N1 of the flipped out base.

The determined crystal structure of the complex formed from M·*Taq*I, DNA and BAZ reveals the existence of a covalent bond between the exocyclic amino group of the flipped out adenine and the opened aziridine moiety. The structure reveals that SMILing DNA occurs at the proposed position in DNA, thus demonstrating again the sequence-specificity. Further, the structure suggests that the enzyme-catalyzed labeling of DNA proceeds in analogy to the natural methylation reaction. It nicely demonstrates that the aziridine moiety of BAZ can replace the methyl sulfonium group of AdoMet as activated group in a MTase-catalyzed reaction.

3.6 Crystal structure of the ternary complex of M·*Taq*I, DNA containing 2-aminopurine (2AP) at the target position and 5'-[2-(amino)ethylthio]-5'-deoxy adenosine (AETA)

The development of different aziridine-based cofactors is essential for the extension of the SMILing technique. Attention in the development of new cofactor analogues should be given to cofactors which allow a faster labeling of DNA. In order to achieve this goal it would be helpful to obtain a better insight into the flexibility of the active site of M·*Taq*I. For this purpose structures with different cofactors and different target bases have to be determined to see the structural changes in the active site. Therefore, another very interesting crystal structure would be a ternary complex between M·*Taq*I, a cofactor analogue and a DNA in which the target adenine is replaced by the unnatural base 2-aminopurine (2AP; Fig. 3.6 left). This base analogue could even allow the crystallization of a ternary complex with the natural cofactor AdoMet because it does not contain a nucleophile at the 6-position. A second reason why 2AP is interesting is because it is a very popular probe for investigations of DNA base flipping enzymes. 2AP is fluorescent in solution. When incorporated into DNA it loses its fluorescence properties due to quenching by base stacking. When the base stacking interaction in DNA is broken by e.g. proteins the fluorescence increases again. Using this property of 2AP a base flipping mechanism was proposed for M·*Taq*I (Holz *et al.*, 1998). The ternary crystal structure of M·*Taq*I, DNA and AETA (Goedecke *et al.*, 2001) proved this proposal. However, at this time, there is no crystal structure known containing a flipped out 2AP which would prove that its increased fluorescence intensity indeed originates from an extrahelical position. Therefore, there is no direct evidence that 2AP behaves similar to the target adenine when bound to M·*Taq*I. Especially, regarding the structure of M·*Taq*I, DNA and AETA, a 2AP base cannot be placed the same way in the active site pocket as a target adenine. The amino group in the 2-position of 2AP would sterically overlap with the phenyl ring of F196 of the enzyme. Only a structure with a flipped out 2AP would remove these last uncertainties. Therefore, a complex between M·*Taq*I, DNA containing 2AP (2AP-DNA) and AdoMet was crystallized. The crystals diffracted well (up to 2 Å) but the electron density observed for AdoMet was not complete. The electron density suggested that the cofactor AdoMet has decayed to 5'-methylthio-5'-deoxyadenosine (MTA). This could be explained either by the intrinsic instability of AdoMet during crystallization or by radiation induced transformation. A second set of crystallization setups was performed with the complex between M·*Taq*I, DNA containing 2AP and the cofactor analogue AETA (Fig. 3.6 right). This allows a direct comparison with the reported structure of M·*Taq*I, DNA and AETA (Goedecke *et al.*, 2001).



2AP **AETA**

Fig. 3.6: Structures of 2-aminopurine (left) and AETA (right).

3.6.1 Crystallisation conditions and data collection

Crystallization setups were performed using the hanging drop method. Crystals of the ternary complex of M·*Taq*I, DNA containing 2AP and AETA could be obtained using the conditions described in 5.5.2. Crystals grew only after micro seeding with a ternary complex of M·*Taq*I, natural DNA and AETA. Thin needles (maximal size 30 x 30 x 100 μm^3) and plates (maximal size 10 x 30 x 100 μm^3) were obtained (Fig. 3.6.1a).

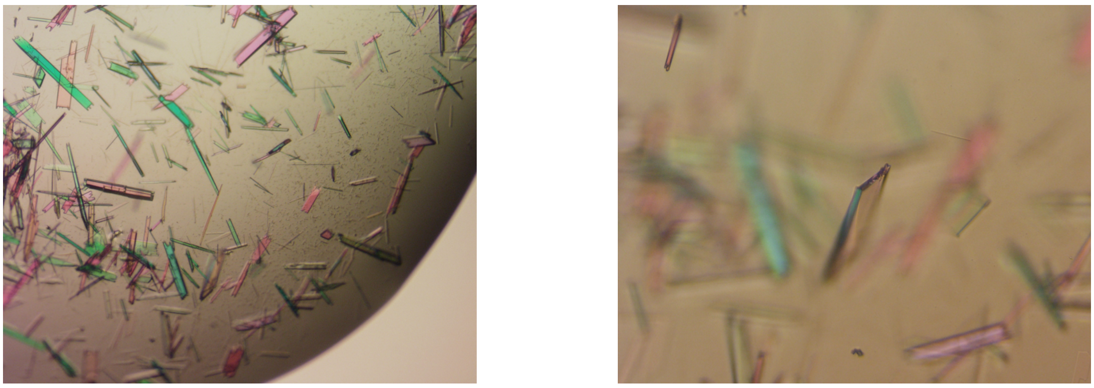


Fig. 3.6.1a: Co-crystals of the M·TaqI, AETA and DNA containing 2-aminopurine (left). Focus on a single crystal suitable for data collection (right).

Crystals of the ternary complex of M·*Taq*I, 2AP-DNA and AETA, like in Fig. 3.6.1a (right), together with a cryoprotectant solution were frozen in liquid nitrogen. As in the case of the crystals produced with M·*Taq*I, DNA and BAZ the data were collected at the European Synchrotron Radiation Facility (ESRF) in Grenoble (France), at the ID14 beamline for macromolecular crystallography. The measurment was also performed at 100 K using a MAR CCD detector. An example of a rotation diffraction image is shown in Fig. 3.6.1b. Diffraction signals are visible up to 2.4 Å. Processing of data was performed using the programs XDS and XSCALE (Kabsch *et al.*, 1993). The crystals of the ternary complex of M·*Taq*I, 2AP-DNA and AETA crystallize in the space group $P2_1$. The data statistics are given in Table 3.6.2 (top).

Fig. 3.6.1b: Rotation diffraction image of a M·TaqI, 2AP-DNA and AETA co-crystal recorded with a MAR CCD area detector. Resolution circles are presented up to 2.0 Å. Intensities of reflexes up to 2.4 Å could be measured with the data reduction software.

3.6.2 Phase determination and structure refinement

The phases were determined by molecular replacement using the ternary complex described by Goedecke and co-workers (2001) as starting model. In order to avoid a bias from the structure of M·*Taq*I, natural DNA and AETA the information about the cofactor and the flipped out base were ignored in the replacement. The missing structural information of the flipped out base and the cofactor was already visible in a first electron density map.

Subsequent refinement steps lead to a well interpretable electron density map. The final values for the crystallographic R-factor and the free R-factor are 0.2107 and 0.2596, respectively. A summary with the refinement statistic is given in Table 3.6.2 (bottom).

Table 3.6.2: Crystallographic data and refinement results for the complex between M·*Taq*I, 2AP-DNA and AETA.

Data statistics	
Beamline	ID14-1/ESRF
Temperature	100 K
Area detector	MAR CCD
Wavelength	0.934 Å
Space group	$P2_1$
Cell dimensions	a = 59.53 Å, b = 69.16 Å, c = 114.64 Å β = 92.19°
Asymetric unit	2 molecules
Number of recorded reflections	133,781
Average redundancy	3.7
R_{sym} [a]	11.5 %
Intensities (I/σ)[b]	9.3 (4.6)
Refinement statistics	
Resolution limit	2.4 Å
Number of unique reflections	36,299
Completeness of data (%)[c]	99.0 (97.5)
R_{cryst}[d] / R_{free}[e]	0.2107 / 0.2596
Rmsd bond lengths (Å)	0.006
Rmsd bond angle (deg.)	1.33
Mean/rms on B factors (Å^2)	
Backbone	18.16 / 8.56
Side chain	21.46 / 13.93
AETA	15.91 / 2.62
2AP-DNA	15.55 / 7.96
Solvent (229 water molecules)	25.28 / 7.21

[a] $R_{sym} = 100 \times \sum I - \langle I \rangle / \sum I$, I is the observed intensity and $\langle I \rangle$ is the average intensity calculated from multiple observations of symmetry related reflections.

[b] The value in parentheses is calculated for the highest resolution shell collected (2.5-2.4 Å).

[c] The value in parentheses is calculated for the highest resolution shell used in refinement (2.5-2.4 Å).

[d] $R_{cryst} = \sum |F_{obs} - F_{calc}| / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are, respectively, observed and calculated structure factor amplitudes.

[e] R_{free} is an R_{cryst} calculated using 5 % of the processed data (1815 reflections), chosen randomly, kept constant and omitted from all structure refinement steps.

3.6.3 Three-dimensional structure of the ternary complex of M·*Taq*I, 2AP-DNA and AETA

The ternary structure of M·*Taq*I, DNA containing 2AP at the target position and AETA crystallizes with two complexes in the asymmetric unit. Since the differences between the two complexes are negligible only complex A will be described due to its better defined electron density map and smaller *B*-factors. The overall crystal structure of M·*Taq*I, 2AP-DNA and AETA is identical with the structure of M·*Taq*I, natural DNA and AETA. A superposition of the two proteins from the ternary complex is shown in Fig. 3.6.3a.

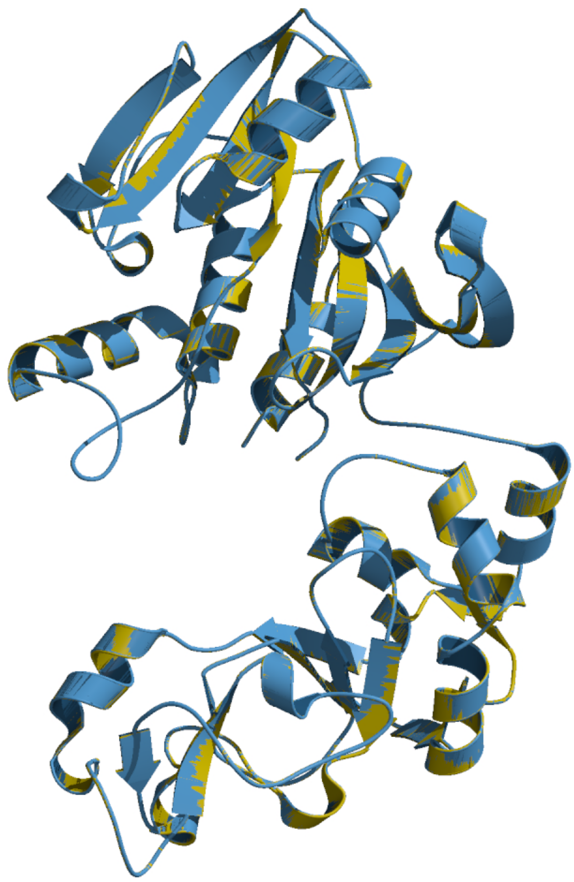


Fig. 3.6.3a: Superposition of the two proteins from the complex crystal structures between M·TaqI, DNA and AETA (blue) and M·TaqI, 2AP-DNA and AETA (gold).

The image reveals that no difference can be seen on the first sight. This is manifested in the root mean square deviation of (0.158 Å complex A, 0.186 Å complex B) of the Cα atoms of the amino acids. However, as already expected the structure differs slightly in the active site area. A superposition within the active site of the complex crystal structures between M·*Taq*I, DNA and AETA (all in blue) and M·*Taq*I (green), 2AP-DNA (red) and AETA (yellow) is shown in Fig. 3.6.3b.

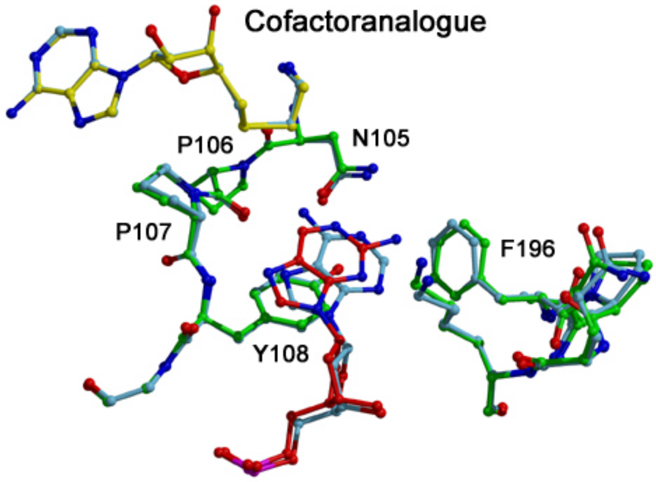


Fig. 3.6.3b: Superposition within the active site of the complex crystal structures between M·TaqI, DNA and AETA (all in blue; Goedecke et al. 2001) and M·TaqI (green), 2AP-DNA (red) and AETA (yellow).

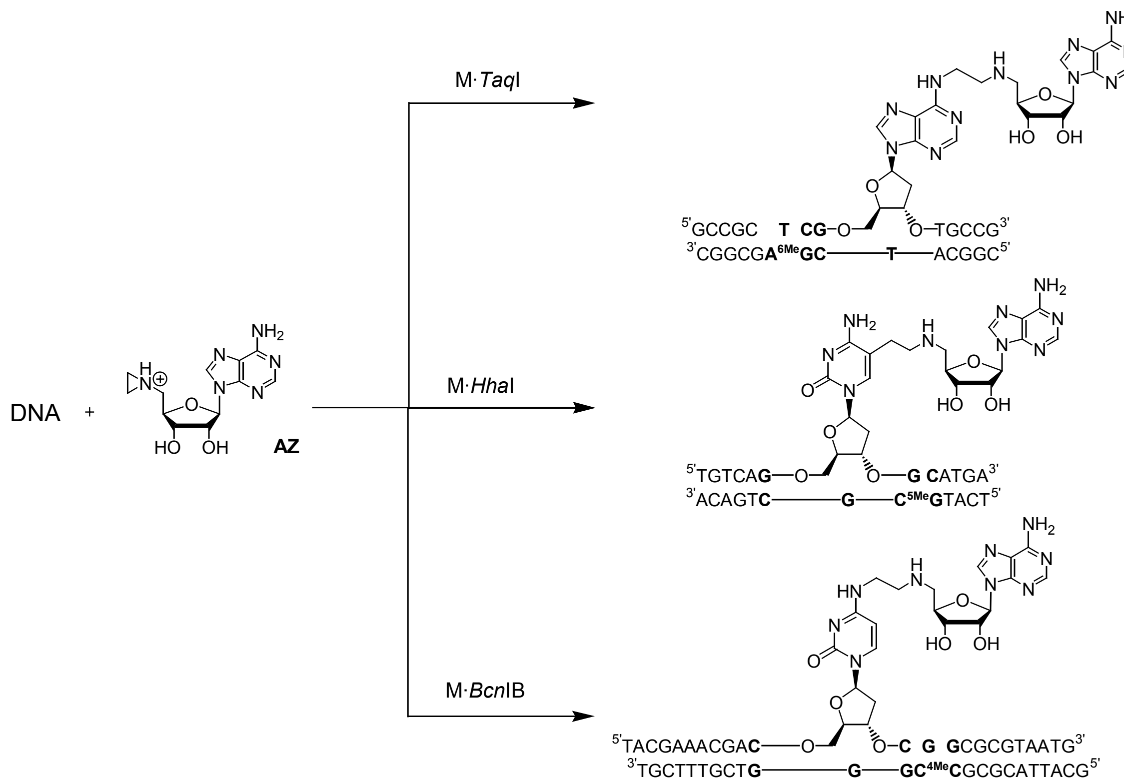
It is illustrated that only the target base 2AP and the amino acid F196 possess a slightly different position in the structure of M·*Taq*I, 2AP-DNA and AETA compared to the target base and F196 in the structure of M·*Taq*I, DNA and AETA. The phenyl ring of F196 is tilted slightly away from the flipped out 2AP base and 2AP is moved towards the proline residue of the NPPY motif. This difference can be explained by the 2-amino group of 2AP. The 2-amino group of 2AP placed in the same position as the adenine in the structure of Goedecke and co-workers (2001) would lead to a steric clash with F196. This problem is overcome in the M·*Taq*I•2AP-DNA•AETA complex by tilting F196 and a slightly shifting 2AP. Nevertheless, the overall structures show a striking similarity. The target base 2AP is extrahelical and occupies a similar position as the target base of a natural substrate of M·*Taq*I. It could be shown for the first time that the target base of a ternary complex between a base flipping enzyme (M·*Taq*I), DNA containing 2AP and a cofactor analogue (AETA) almost perfectly mimics the natural situation. Thus, the biochemical assay using 2AP as fluorescent probe is, at least for M·*Taq*I, reliable to detect and monitor DNA base flipping.

CHAPTER 4

Summary, Conclusions and Outlook

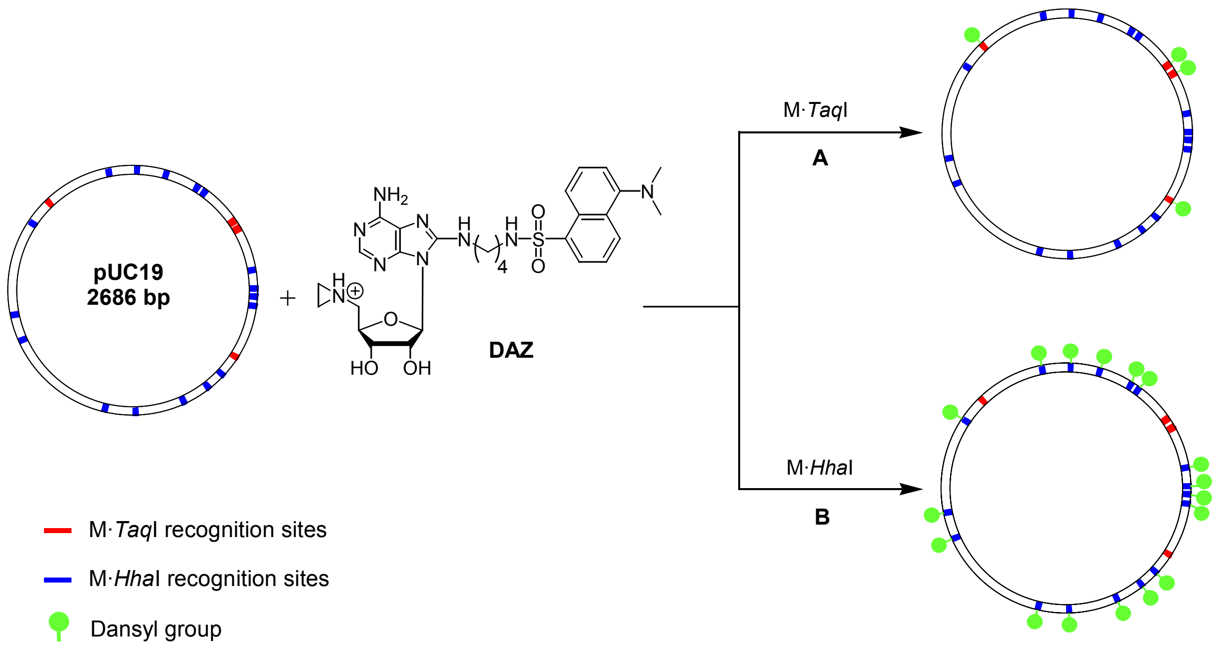
4 Summary, Conclusions and Outlook

Sequence-specific labeling of DNA is essential for a number of analytical and functional investigations of DNA. It represents a challenging task because defined DNA sequences have to be recognized and covalently modified. Different approaches such as triple helix formation with oligonucleotides or PNA, recognition of the minor groove with polyamides and targeting DNA with designed zinc finger proteins are reaching a state where a given sequence can be recognized. However, only in exceptional cases these systems were used for sequence-specific labeling of DNA. It was already demonstrated that the N6-adenine DNA MTase M·*Taq*I catalyzes sequence-specifically the labeling of DNA by the use of a novel aziridine cofactor. In this research project, **S**equence-specific **M**ethyltransferase-**I**nduced **L**abel**ing** of DNA (SMILing DNA) was demonstrated using all three classes of DNA MTases. In addition to M·*Taq*I also the C5-cytosine DNA MTase M·*Hha*I and the N4-cytosine DNA MTase M·*Bcn*IB were both able to catalyze the coupling of *N*-adenosylaziridine (AZ) to DNA. All these reactions occur sequence-specifically and reveal the general applicability of the SMILing technique (Scheme 4.1).



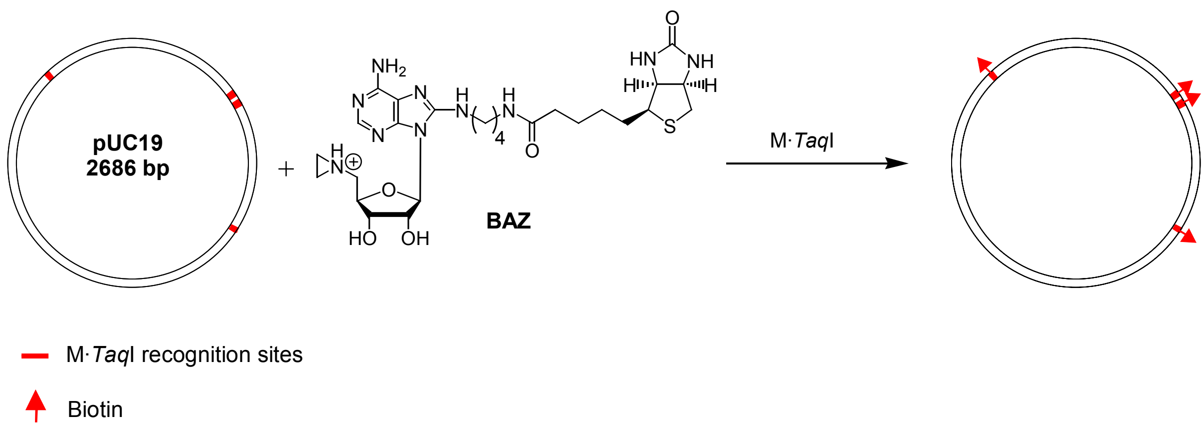
Scheme 4.1: General application of N-adenosylaziridine (AZ) as new cofactor for all three classes of DNA MTases using short duplex oligonucleotides.

By varying the pH value of the buffer solution, the M·*Taq*I-catalyzed coupling reaction of DNA with AZ was optimized with respect to the reaction rate. It was shown that the reaction rate increases with increasing acidity of the reaction solution. This can be explained by an activation of the aziridine moiety due to protonation. The modification of the cofactor at the 8-position appears not to influence the maximum velocity of the reaction. Furthermore, good nucleophiles, like DTT, were shown to have no significant influence on the reaction.
Additionally, labeling of plasmid DNA with the dansyl fluorophore could be demonstrated using the dansylated aziridine cofactor DAZ (Scheme 4.2, A). The reaction occurs in a sequence-specific manner. Furthermore, M·*Hha*I catalyzes the labeling of plasmid DNA with the dansylated cofactor DAZ (Scheme 4.2, B). This illustrates that the series of aziridine-based cofactors which are modified at the 8-position can also be used with the cytosine-specific DNA MTase M·*Hha*I for labeling of DNA.



Scheme 4.2: Sequence-specific fluorescence labeling of plasmid DNA with dansylated AZ cofactor DAZ. Labeling of plasmid using M·TaqI and M·HhaI. The M·TaqI-catalyzed coupling of DAZ with short duplex oligonucleotides was already analyzed in detail (Pljevaljčić, 1999).

All these experiments encouraged the development of a general synthesis for the novel cofactor family using a photolabile protection group. This general synthesis allows the attachment of different reporter groups at the end of the synthesis. Therefore, an easy route to a whole range of differently decorated cofactors paves the ground for SMILing of DNA with different reporter groups. The general route was used to synthesize two new cofactors, the biotinylated aziridine cofactor BAZ and the fluorescent cofactor Cy5AZ. Both were shown to bind strongly to M·*Taq*I. Using the BAZ cofactor it was demonstrated that short duplex DNA as well as plasmid DNA can be sequence-specifically labeled *via* M·*Taq*I-catalysis (Scheme 4.3).



Scheme 4.3: Sequence-specific biotinylation of plasmid DNA with the biotinylated AZ cofactor BAZ. Biotinylation was also achieved with short duplex oligonucleotides.

The reaction is almost quantitative. This cofactor analogue can further be used for indirect labeling using the biotin-streptavidin system in combination with the SMILing technique. A crystal structure of a complex formed from M·*Taq*I, DNA and BAZ proved that the DNA is labeled with biotin at the expected position, thus, revealing the sequence-specificity. In addition, a new application for the SMILing DNA technique was demonstrated. A new system for the selection of deoxyribozymes by internal functionalization could be obtained. In collaboration with the group of Dr. Andres Jäschke (FU Berlin, Germany) it was shown that a pool of DNA could be labeled with biotin. The biotinylated DNA is suitable as a template for PCR. The sequence of the biotinylated DNA can be amplified indicating no adverse effects of the biotin moiety on polymerase activity. To the best of my knowledge, this powerful internal functionalization of DNA is the easiest way to introduce specifically different groups into DNA.

Regarding the amount of known DNA MTases, SMILing DNA represents a very powerful technique to label a vast amount of different DNA sequences. Owing to the general design of the novel cofactor family and the conserved binding region of MTases in general it is conceivable that SMILing is applicable to most MTases, including the MTases for different substrates like RNA and proteins. However, the technique is not yet optimized to fulfill these requirements completely. The modification of the aziridine cofactor at the 8-position was originally designed for M·*Taq*I. Even if modification at this position turns out to be suitable for a lot of other MTases (e.g. M·*Hha*I), some other MTases might prefer a modification at another e.g. the 7-position (see table 3.1b). In this case, another synthetic route to 7-modified aziridine cofactors has to be developed. Further, the rate of SMILing DNA is still slow. As shown in the M·*Taq*I-catalyzed coupling reaction of DNA with AZ the highest reaction rate constant obtained was only 0.6 h^{-1}. This rate constant is about 2 orders of magnitute smaller than the rate constant of M·*Taq*I with the natural cofactor AdoMet. Therefore, further work has to be done to increase the reaction rate. A starting point could be the modification of the aziridine moiety. In the novel aziridine cofactors the methionine side chain of AdoMet was neglected because of synthetic ease. However, the methionine side chain is very important for the function of AdoMet which is revealed by the fact that 5'-deoxy-5'-dimethylthio adenosine is a very poor cofactor. The methionine side chain seems to fulfill the task of directing the methyl group in the direction of the flipped out base. The aziridine cofactors do not contain a side chain on the aziridine moiety which could direct the aziridine ring in the right orientation for the reaction. Regarding the determined crystal structure of M·*Taq*I in complex with biotinylated DNA, a new cofactor containing a hydroxymethyl group on the aziridine can be suggested to fulfill this task. The hydroxyl group would then be positioned to form a hydrogen bond with the protein and hence this side chain would fix the three-membered aziridine ring in a favourable position for the reaction. Furthermore, the electrophilicity of the aziridine ring could be increased by attachment of electron withdrawing groups to the ring. Therefore, a cyano group on the aziridine ring could lead to an increased reactivity of the cofactor. This modification however, could also lead to a possible uncatalyzed background reaction with DNA which was not observed with the aziridine cofactors used so far.

Another very important point is the investigation of the suitability of aziridine cofactors with other MTases. Not only different DNA MTases should be tested for labeling of DNA but other MTases like protein or RNA MTases could be used for the labeling of a variety of biopolymers. Thus, the SMILing technique could be extended to a general labeling technique of different biopolymers. This would open even broader application possibilities of the method.

Another question which was solved in this thesis, was about the applicability of the nucleobase analogue 2-aminopurine in base flipping assays. Till now, it remained unanswered whether the large fluorescence change observed upon binding to different DNA MTases really results from base flipping as observed in crystal structures with natural DNA. In the case of M·*Taq*I, a crystal structure of a complex between M·*Taq*I, DNA and an cofactor analogue suggested that a 2-aminopurine base would lead to steric interference with an amino acid side chain of M·*Taq*I. For this reason, a crystal structure of the complex between M·*Taq*I, DNA containing 2-aminopurine and the cofactor analogue AETA was determined. With the structure in hand it could be demonstrated that the nucleobase analogue 2-aminopurine is indeed in an extrahelical position but slightly shifted in the active site relative to the position of the natural adenine. Thus, the results could show for the first time that DNA with 2-aminopurine at the target position is a good probe for the detection and monitoring of base flipping by DNA modifying enzymes.

In summary, the developed SMILing DNA technique holds a great potential for applications in biotechnology, medical diagnosis (mutation detection) and nano biotechnology. Further experiments have to show whether the SMILing technique can also be expanded to the SMILing of other biopolymers.

CHAPTER 5

Materials and Methods

5 Materials and Methods

5.1 Materials

5.1.1 Chemicals and Proteins

Chemicals and proteins were purchased from the following companies with the highest purity available.

Table 5.1.1: Chemicals and Proteins.

Company	Product
AppliChem (Darmstadt, Germany)	Methanol.
Bayer AG (Leverkusen, Germany)	Baysilone paste (medium viscous)
Biorad (Munich, Germany)	Protein molecular weight standard, agarose.
Fluka (Neu-Ulm, Germany)	6-nitroveratryl chloroformate, Biotin succinimidylester, collidine, DMAP, DMF, DMSO, EDTA, formamide (99.5 %) glycerol, mesyl chloride, N-ethyldiisopropyl-amine, ninhydrin, PEG6000, pyridine, sodium cacodylate, THAP, triethanolamine, triethylamine.
Gerbu (Gaiberg, Germany)	Boric acid, DTT, HEPES, SDS, TEMED.
J.T. Baker (Deventer, Netherlands)	Acetic acid, acetone, acetonitrile, APS, calcium chloride, formic acid, chloroform, citric acid, dichloromethane, dipotassium hydrogenphosphate, hydrochloric acid, IPA, magnesium acetate, magnesium chloride, potassium acetate, potassium dihydrogen phosphate, potassium hydroxide, sodium azide, sulfuric acid.
Merck (Darmstadt, Germany)	Proteinase K.
Molecular Probes (Eugene, OR, USA)	Cy5-succinimidylate, streptavidin-fluorescein.
NEB (Frankfurt a.M., Germany)	R·*Taq*I, R·*Bsi*EI.
Pierce (Rockford, IL, USA)	BSA, Bradford reagent.
Riedel de Häen (Seelze, Germany)	Ethanol.
Roche Biochemicals (Mannheim, Germany)	Alkaline Phosphatase
Roth (Karlsruhe, Germany)	Acrylamide/Bis-acrylamide (29:1, 40 %), Tris, urea.
Serva (Heidelberg, Germany)	Acryl amide, bromphenol blue, ethidium bromide, Coomassie Brilliant Blue G250, Triton X-100.
Sigma Aldrich (Taufkirchen, Germany)	1,4-diaminobutane, 2-bromoethylamine hydrobromide, 2',3'-*O*-isopropylideneadenosine, bromine, sodium bisulfite, tricine.

5.1.2 General Instrumentation

Deionisation of water was performed with a deionisation system from Millipore (Eschborn, Germany). For the degasing of HPLC-buffers a ultrasonic bath Sonorex Super RK 103 from H. Bandelin (Berlin, Germany) was used.

Measurments of pH were performed using a pH-meter 761 from Calimatic Knick (Berlin, Germany).

Synthesized compounds were dried using the freeze dryer Alpha I-5 from Christ (Osterode, Germany).

For the hybridisation of DNA a heat block Dri-Block DB2A from Techne (Cambridge, UK) was used.

Centrifugation was carried out using a table centrifuge *Centrifuge* 5415 C from Eppendorf (Cologne, Germany).

SDS gel electrophoresis was performed with the Mini-Protean II system from Biorad (Munich, Germany).

For agarose gel electrophoresis an agarose mini gel system from MWG Biotech (Ebersberg, Germany) was used.

For polyacrylamide-urea gel electrophoresis a gel chamber from BioRad (Munich, Germany) was used.

Protein dialysis and concentration for crystallographic purposes was performed with a vacuum dialysis apparatus from Schleicher & Schuell (Dassel, Germany) using ultra thimbles (10 kD cutoff). For concentration the ILMVAC membrane pump (Ilmenau, Germany) was used.

Crystals were set up in Linbro cell culture plates from ICN (Meckenheim, Germany).

Data processing, refinement and model building were done using a graphics Workstation Silicon by Graphics (Mountain View, CA, USA).

5.1.3 Computer programs and packages

For three dimensional representations of crystal structures the programs MOLSCRIPT (Kraulis, 1991), GRASP (Nicholls *et al.*, 1991) and ADOBE PHOTOSHOP (Adobe Systems Inc., San Jose, CA, USA) were used.

Data fitting for the determination of binding constants was performed using GraFit 3.0 (Erithacus Software Ltd., Horley, UK) or Scientist 2.01 (MicroMath Inc., Salt Lake City, UT, USA).

5.1.4 Chromatography

For TLC silica gel 60 F_{254} (d = 0.25 mm) on glas plates from Merck (Darmstadt, Germany) and silica gel 60 F_{254} (d = 0.2 mm) on aluminium foil from Fluka (Neu-Ulm, Germany) were used. For TLC's of reactions with non-volatile solvents the samples were applied onto the silica gel and dried *in vacuo* (25 °C, 0.1 mbar) for 30 min before placing them into an elution chamber. Compounds were either detected by irradiation with UV light (254 nm and 366 nm) or by moisturizing with one of the following solutions and heating with a heat gun.

Solution for the detection of sugar derivatives: Ammoniummolybdate tetrahydrate $[(NH_4)_6Mo_7O_{24}\cdot 4H_2O]$ (15 g), cer(IV) sulfate tetrahydrate (400 mg), concentrated sulfuric acid (30 ml) diluted with water (1:10).

Solution for the detection of primary amines (ninhydrin solution): Ninhydrin (0.3 g), 2-propanol (95 ml), collidin (5 ml) and 5 ml acetic acid (96 %).

Column chromatographic separation of compounds was achieved by using 50-100 times the sample amount of silica gel 60 (230-400 mesh, particle size 0.040-0.063 mm) from Merck (Darmstadt, Germany).

High performance liquid chromatography (HPLC) was carried out with an apparatus from Beckman (System Gold). The spectrofluorometric detector RF-551 from Shimadzu (Kyoto, Japan) was connected to the HPLC system. The reversed phase C-18 PRONTOSIL AQ column (5 µm, 120 Å, 250 x 4.6 mm, flowrate 1 ml/min) was from Bischoff (Leonberg, Germany). The anion exchange column for plasmid analysis (Nucleogen 4000-7 DEAE, flow rate 1.5 ml/min) was from Macherey-Nagel (Düren, Germany) and the anion exchange column for the analysis of short oligonucleotides (Poros 10 HQ, flow rate 4 ml/min) was from Perseptive Biosystems (Weiterstadt, Germany).

Protein purification was carried out with a High-Load-System LKB GP-10 equipped with a P-50 pump, a LKB-UV-MII-detector, a conductivity monitor, a LKB-REC 102 writer and a fractions collector LKB-SuperFrac from Amersham-Pharmacia (Freiburg, Germany).

5.1.5 HPLC Buffers

The following buffer systems were used for the HPLC. In general, the pH values of solutions are adjusted at room temperature and given after all components which were present during pH adjustment. If further components follow, the pH value was not adjusted again.

Table 5.1.5: HPLC buffer systems.

	Buffer A	Buffer B
System 1	100 mM TEAAc, pH 7.0	100 mM TEAAc, (pH 7.0), 70 % CH_3CN
System 2	100 mM $TEAHCO_3$, pH 7.0	100 mM $TEAHCO_3$, (pH 7.0), 70 % CH_3CN
System 3	10 mM Tris/HCl, pH 7.0	1 M KCl + buffer A
System 4	30 mM KH_2PO_4, (pH 6.5) + 20 % CH_3CN	1.5 M KCl + buffer A

5.1.6 HPLC gradients

The gradients used for the HPLC are listed below.

Table 5.1.6: HPLC gradients.

Gradient 1	20 % buffer B (0-5 min), 20-50 % buffer B (5-10 min), 50-100 % buffer B (10-40 min), 100 % buffer B (40-45 min).
Gradient 2	20 % buffer B (0-5 min), 20-100 % buffer B (5-35 min), 100 % buffer B (35-40 min).
Gradient 3	7 % buffer B (0-10 min), 7-40 % buffer B (10-40 min), 40-100 % buffer B (40-50 min), 100 % buffer B (50-55 min).
Gradient 4	20 % buffer B (0-10 min), 20-60 % buffer B (10-50 min), 60-100 % buffer B (50-60 min).

5.1.7 Spectroscopy and spectrometry

All NMR spectra were recorded on a ADVANCE DRX_{500}-spectrometer from Bruker (Karlsruhe, Germany). The chemical shifts δ_H (^1H) and δ_C (^{13}C) are quoted in ppm and are referenced to TMS ($\delta = 0$). The signal of the used solvent was used for calibration (Hesse *et al.*, 1991). The following abbreviations were used for the signal multiplicities: s (singlet),

d (doublet), dd (doublet of doublets), t (triplet), q (quartet), m (multiplet), br. (boad); the coupling constants *J* are given in Hertz (Hz). The coupling protons are given in subscript.
Mass spectra were recorded with a LCQ mass spectrometer Finnigan MAT (San Jose, CA, USA) and a Voyager-DE Pro, BioSpectrometry Workstation from Perseptive Biosystems (Weiterstadt, Germany), respectively.
Fluorescence measurements in solution were performed using the Aminco-Bowman Series 2 luminescence spectrometer by SLM-AMINCO (Rochester, NY, USA).
UV-VIS spectra measurements were carried out using a Cary 3E spectrophotometer by Varian (Darmstadt, Germany).

5.1.8 Oligonucleotides

Unmodified oligonucleotides were purchased from MWG Biotech (Germany). The modified oligonucleotides were purchased from IBA-NAPS (Germany). The sequences of the oligonucleotides are listed from 5' to 3'. The extinction coefficients were calculated using the method by Cantor *et al.* (1970). The recognition sequences of the DNA MTases are highlighted with bold letters.

Table 5.1.8: Oligonucleotides.

Name	Sequence	Extinction coefficient/ (10^3·L·mol^{-1}·cm^{-1})
7 (KAz)	GCCGC**TCGA**TGCCG	122.7
8 (KAzMe)	CGGCA**TCGA^{6Me}**GCGGC	129.7
9 (UC)	TGTCA**GCGC**ATGA	126.4
10 (YM)	TCAT**GC^{5Me}GC**TGACA	120.2
UCSPEC	TGTCAGAGCATGA	133.0
YMSPEC	TCATGCTCTGACA	120.5
11 (B1)	TACGAAACGA**CCCGG**CGCGTAATG	234.8
12 (B2m)	GCATTACGCG**CC^{4Me}GGG**TCGTTTCGT	216.7
B1SPEC	TACGAAACGACACGGCGCGTAATG	240.0
B2SPEC	GCATTACGCGCCGTGTCGTTTCGT	215.3
O10	GT**TCGA**TGTC	93.6
O10Ap	GT**TCG2Ap**TGTC	82.0
U10	GACA**TCGA^{6Me}**AC	101.8

5.2 Chemical methods

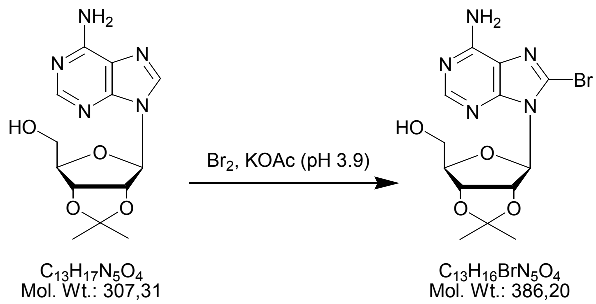
5.2.1 Deprotection of photolabile protecting group

The photolabile protective group NVOC was deprotected using two different exposure instruments. For analytical scale setups a cuvette was exposed to filtered light ($\lambda = 300$-370 nm) of a mercury lamp. For large scale setups a glass apparatus (max. 1 L) was illuminated with a mercury lamp.

5.2.2 Syntheses

All reactions were performed under argon atmosphere if not stated differently. Solvents were dried using standard drying techniques (Perrin & Armarego, 1988).

8-Bromo-2‘,3‘-*O*-isopropylidene adenosine (14)



13 **14**

To a solution of 2‘,3‘-*O*-isopropylidene adenosine (**13**) (2.3 g, 7.53 mmol) in KOAc buffer (300 ml, 1 M, pH 3.9) a bromine solution (0.62 ml, 12.1 mmol) in KOAc buffer (20 ml, 1 M, pH 3.9) was added within 15 min at 0 °C. Subsequently, the orange colored solution was stirred at room temperature for 15 h until no starting material could be detected by TLC. Excess of bromine was reduced using a saturated $NaHSO_3$ solution which lead to a white precipitate. The pH of the suspension was adjusted to 7 using NaOH solution (220 ml, 10 M). The white precipitate was collected by filtration, washed with water (350 ml) and freeze dried. Nucleoside **14** was obtained as a colourless solid (2.40 g, 83 %).

R_f = 0.72 (silica gel on aluminium foil, n-butanol/HOAc/H_2O 3:0.75:1.25).

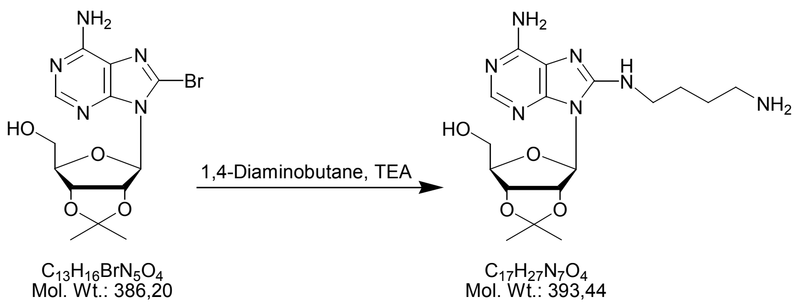
^{1}H-NMR (500 MHz, DMSO-d$_6$): δ = 1.32 (s, 3H, acetonide-CH_3), 1.54 (s, 3H, acetonide-CH_3), 3.40 - 3.52 (m, 2H, 5‘-H), 4.14 - 4.17 (m, 1H, 4‘-H), 5.02 (dd, $J_{3'\text{-H},4'\text{-H}}$ = 2.8 Hz, $J_{3'\text{-H},2'\text{-H}}$ = 6.0 Hz, 1H, 3‘-H), 5.10 (t, $J_{5'\text{-OH},5'\text{-H}}$ = 5.8 Hz, 1H, 5‘-OH), 5.65 (dd, $J_{2'\text{-H},1'\text{-H}}$ = 2.8

Hz, $J_{2‘\text{-H},3‘\text{-H}}$ = 6.0 Hz, 1H, 2‘-H), 6.01 (d, $J_{1‘\text{-H},2‘\text{-H}}$ = 2.8 Hz, 1H, 1‘-H), 7.53 (s, br., 2H, 6-NH_2), 8.14 (s, 1H, 2-H).

^{13}C-NMR (125.7 MHz, DMSO-d_6): δ = 25.45 (q, acetonide-CH_3), 27.34 (q, acetonide-CH_3), 61.71 (t, C-5‘), 81.82 (d, C-3‘), 82.14 (d, C-2‘), 87.34 (d, C-4‘), 91.22 (d, C-1‘), 113.48 (s, acetonide-<u>C</u>(CH_3)$_2$), 119.54 (s, C-5), 126.42 (s, C-8), 149.99 (s, C-4), 153.03 (d, C-2), 155.35 (s, C-6).

MS (ESI) m/z (%): 387.9 (100) $[M+H]^+$, 216.3 (75) $[\text{8-bromoadenine+H}]^+$.

8-Amino[1‘‘-(4‘‘-aminobutyl)]-2‘,3‘-*O*-isopropylidene adenosine (15)



14 **15**

To a solution of **14** (628 mg, 1.6 mmol) in dry DMSO (10 ml), dry TEA (2.26 ml, 16.3 mmol) and 1,4-diaminobutane (0.82 ml, 8.1 mmol) were added. The solution was stirred at 110 °C and the reaction progress monitored by TLC. After 4 h the solvent was removed under reduced pressure. The residue was dissolved in water (50 ml) and the pH was adjusted to 5.3 with HOAc (0.1 M). The crude product was purified by cation exchange chromatography (Dowex 50 x 4 in H^+-form, 100 g, elution with 600 ml water and subsequently with 1000 ml 1 M KOH). Fractions containing the product were collected, combined and extracted with chloroform. The organic layers were combined and the solvent was removed under reduced pressure. Yield: 639 mg (100 %).

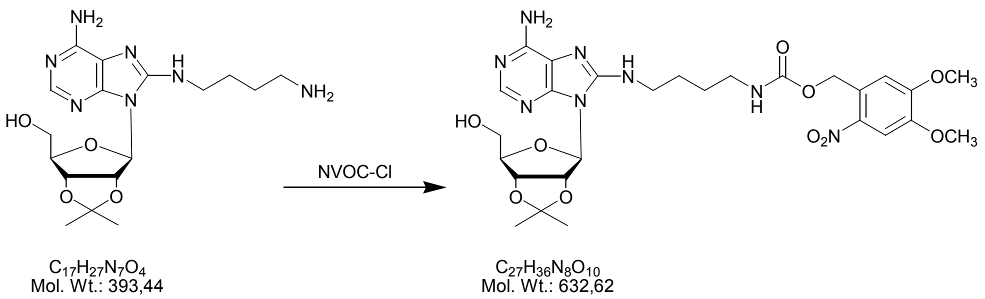
R_f = 0.44 (silica gel on glass plate, n-butanol/HOAc/H_2O 3:0.75:1.25).

^{1}H NMR (500 MHz, $CDCl_3$): δ = 1.33 (s, 3H; acetonide-H), 1.48-1.55 (m, 2H; linker-H), 1.61 (s, 3H; acetonide-H), 1.64-1.70 (m, 2H; linker-H), 2.66-2.73 (m, 2H; linker-H), 3.33-3.42 (m, 2H; linker-H), 3.77-3.91 (m, 2H; 5‘-H), 4.28-4.30 (m, 1H; 4‘-H), 4.99 (dd, $J_{3’\text{-H}, 4’\text{-H}}$ = 2.7, $J_{3’\text{-H}, 2’\text{-H}}$ = 6.3 Hz, 1H; 3‘-H), 5.08 (dd, $J_{2’\text{-H}, 1’\text{-H}}$ = 4.8, $J_{2’\text{-H}, 3’\text{-H}}$ = 6.3 Hz, 1H; 2‘-H), 5.39 (s, br., 2H; 6-NH_2), 6.15 (d, $J_{1'\text{-H},\,2'\text{-H}}$ = 4.5 Hz, 1H; 1‘-H), 6.55-6.60 (m, 1H; 8-NH), 8.10 (s, 1H; 2-H).

^{13}C-NMR (125.7 MHz, $CDCl_3$): δ = 25.30 (q; acetonide-CH_3), 25.73 (t; linker-C), 27.42 (q; acetonide-CH_3), 29.60 (t; linker-C), 40.46 (t; linker-C), 42.69 (t; linker-C), 61.17 (t; 5'-C), 80.59 (d; 3'-C), 82.19 (d; 2'-C), 84.48 (d; 4'-C), 89.21 (d; 1'-C), 114.50 (s; acetonide-$\underline{C}(CH_3)_2$), 117.68 (s; 5-C), 149.49 (d; 2-C), 149.95 (s; 8-C), 151.68 (s; 4-C), 151.72 (s; 6-C);

ESI-MS: m/z (%): 394.3 (25) $[M + H]^+$, 222.3 (100) [8-Amino[1’’-(4’’-aminobutyl)]-adenine + H]$^+$.

8-Amino[1‘‘-(*N*‘‘-6-nitroveratryl-oxocarbonyl)-4‘‘-aminobutyl]-2‘,3‘-*O*-isopropylidene adenosine (22)



15 **22**

To a solution of nucleoside **15** (224.7 mg, 0.571 mmol) in dry pyridine (10.0 ml), 6-nitroveratrylchloroformate (173.6 mg, 0.631 mmol) was added slowly, and the resulting solution was stirred at room temperature for 3 h. The progress of the reaction was monitored by TLC and ninhydrin test. After complete conversion the solution was treated with water (8 ml) at 0 °C. The solvent was removed under reduced pressure, and the crude product was purified by column chromatography (silica gel, 25 g, elution with CH_2Cl_2/CH_3OH 19:1). Yield: 168 mg (46 %).

R_f = 0.44 (silica gel on aluminium foil, CH_2Cl_2/CH_3OH 9:1).

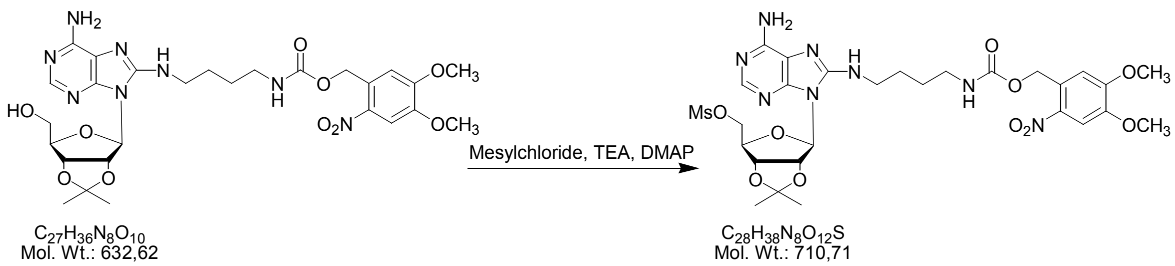
^{1}H-NMR (500MHz, DMSO-d_6): δ = 1.29 (s, 3H, isopropylidene-H); 1.47 - 1.49 (m, 2H, linker-H); 1.53 (s, 3H, isopropylidene-H); 1.58 - 1.61 (m, 2H, linker-H); 3.03 - 3.07 (m, 2H, linker-H); 3.29 - 3.30 (m, 2H, linker-H); 3.54 - 3.57 (m, 2H, 5‘-H); 3.85 (s, 3H, methoxy-H); 3.87 (s, 3H, methoxy-H); 4.14 - 4.15 (m, 1H, 4‘-H); 4.95 (dd, 1H, $J_{3'\text{-H},\,4'\text{-H}}$ = 2.9 Hz, $J_{3'\text{-H},\,2'\text{-H}}$ = 6.1 Hz, 3‘-H); 5.31 (s, 2H, NVOC-CH_2); 5.35 (dd, 1H, $J_{2'\text{-H},\,1'\text{-H}}$ = 3.7 Hz,

$J_{2'\text{-H},3'\text{-H}}$ = 6.5 Hz, 2'-H); 5.47 (t, 1H, 5'-OH); 6.03 (d, 1H, $J_{1'\text{-H},2'\text{-H}}$ = 3.7 Hz, 1'-H); 6.51 (s, 2H, 6-NH_2); 6.93 (t, 1H, Linker-NH); 7.16 (s, 1H, NVOC-arom. H); 7.46 (t, 1H, Linker-NH); 7.68 (s, 1H, NVOC-arom. H); 7.90 (s, 1H, 2-H).

^{13}C-NMR (125.7 MHz, DMSO-d$_6$): δ = 27.15 (isopropylidene-C); 28.02 (linker-C); 29.07 (isopropylidene-C); 41.90 (linker-C); 43.92 (linker-C); 58.02 (methoxy-C); 58.10 (methoxy-C); 63.18 (5'-C); 64.15 (NVOC-CH_2); 82.90 (3'-C); 83.26 (2'-C); 87.04 (4'-C); 89.64 (1'-C); 110.04 (NVOC-arom. C); 112.55 (2 -C); 115.15; 119.02; 129.83; 141.26; 149.63; 150.68 (NVOC-arom. C); 151.07; 153.01; 154.35; 155.32; 157.61.

MS (ESI) m/z (%): 633.2 (100) $[M+H]^+$.

8-Amino[1''-(*N*''-6-nitroveratryl-oxocarbonyl)-4''-aminobutyl]-2',3'-*O*-isopropylidene-5'-*O*-mesyl adenosine (17)



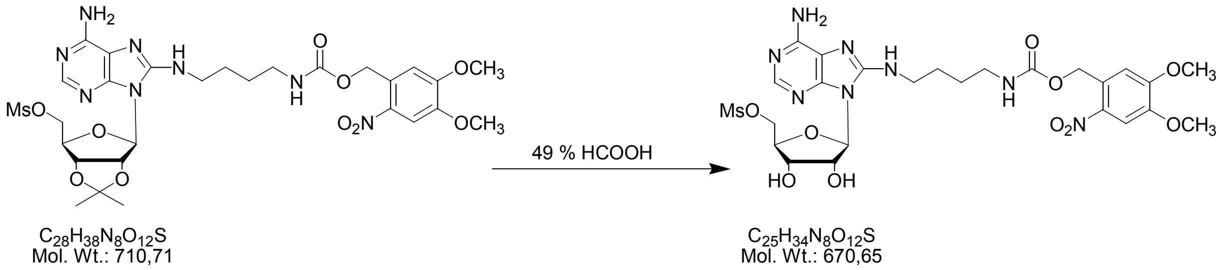
22 **17 (R = NVOC)**

To a solution of **22** (202.5 mg, 0.320 mmol) and DMAP (39.6 mg, 0.324 mmol) in dry CH_2Cl_2 (15 ml), dry TEA (1.12 ml, 8.04 mmol) was added and the resulting solution was cooled to 0 °C. Mesyl chloride (200 µl, 2.6 mmol) was added and the solution was stirred for 90 min. The reaction was quenched with a cold, saturated $NaHCO_3$ solution (3 ml). The solution was extracted three times with cold $CHCl_3$ (5 ml). The organic phases were combined and the solvent removed under reduced pressure. The crude product was purified by column chromatography (silica gel, 11 g, elution with CH_2Cl_2/CH_3OH 97:3). Yield: 126 mg (55 %).

R$_f$ = 0.45 (silica gel on aluminium foil, CH_2Cl_2/CH_3OH 9:1).

MS (ESI) m/z (%): 711.3 (71) $[M+H]^+$, 615.5 (100) $[\text{cyclonucleoside}]^+$.

8-Amino[1''-(*N*''-6-nitroveratryl-oxocarbonyl)-4''-aminobutyl]-5'-*O*-mesyl adenosine (18)



17 (R = NVOC) **18 (R = NVOC)**

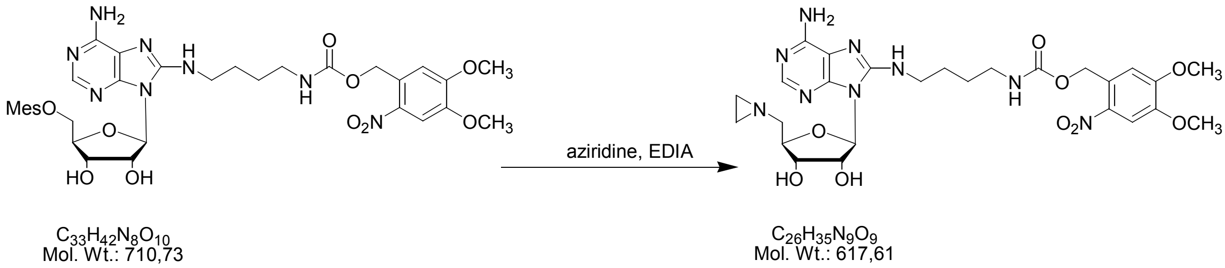
Nucleoside **17** (143.6 mg, 202 µmol) was dissolved in aqueous formic acid (49 %, 20 ml), and the resulting solution was stirred at room temperature for 5 d. After complete conversion the solvent was removed under reduced pressure and remaining solvent was co-evaporated three times with a mixture of water and methanol (1:1, 5 ml). The crude product was purified by column chromatography (silica gel, 15 g, elution with CH_2Cl_2/CH_3OH 9:1). After removing the solvent under reduced pressure a yellowish product was obtained (132.2 mg, 98 %).
$\mathbf{R_f}$ = 0.23 (silica gel on aluminium foil, CH_2Cl_2/CH_3OH 9:1).
MS (ESI) m/z (%): 671.2 (45) $[M+H]^+$, 575.2 (100) $[cyclonucleoside]^+$.

Aziridine (29)

28 **29**

The synthesis was performed based on Gabriel (1888). KOH (13.3 g, 237 mmol) was dissolved in water (25.0 ml) at 0 °C. 2-Bromoethylamin hydrobromide (12.0 g, 58.6 mmol) was added at 10 °C to the KOH solution and stirred for 1 h before the solution was distilled on KOH plates. A second distillation was performed prior to use. Aziridine (1.3 ml, 43 %) was obtained as a colorless liquid (head temperature 55 °C).
^{1}H-NMR (500MHz, CDCl$_3$): δ = 1.24 (s, 4H).

8-Amino[1''-(*N*''-6-nitroveratryl-oxocarbonyl)-4''-aminobutyl]-5'-(1-aziridinyl)-5'-deoxy adenosine (19)



18 (R = NVOC) **19**

Nucleoside **18** (44.8 mg, 66.8 µmol) was dissolved in dry aziridine (1.1 ml, 21.2 mmol) and *N*-ethyldiisopropylamine (0.40 ml, 2.34 mmol) and stirred at room temperature for 3 d. The reaction was monitored by analytical reversed phase HPLC. The solvent was removed under reduced pressure after completeness of the reaction. The crude yellow product was purified by column chromatography (silica gel, 2.8 g, elution with CH_2Cl_2/CH_3OH 9:1). Yield: 8 mg (20 %).

R_f = 0.23 (silica gel on aluminium foil, dichloromethane/methanol 9:1).

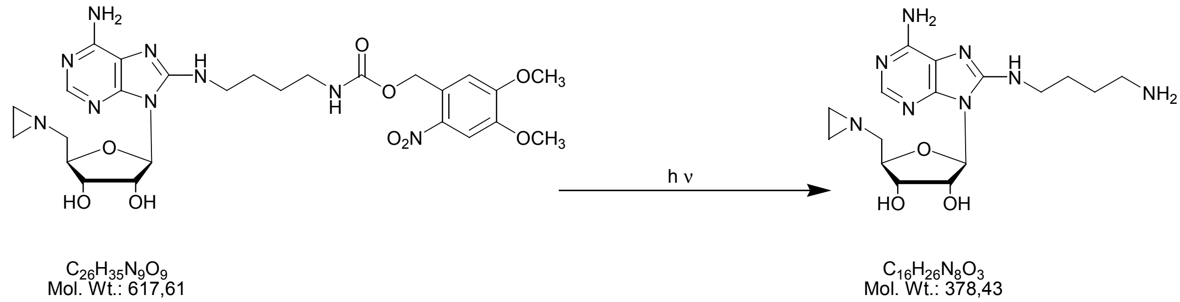
R_t = 48.9 min (C-18 reversed phase column, system 1, gradient 4).

^{1}H-NMR (500MHz, DMSO-d$_6$): δ = 1.25 - 1.27 (m, 2H, linker-H); 1.37 - 1.39 (m, 2H, linker-H); 1.43 - 1.45 (m, 2H, aziridine-H); 1.46 - 1.50 (m, 2H, aziridine-H); 1.67 - 1.69 (m, 2H, linker-H); 1.98 (dd, 1H, $J_{5'\text{-Ha, }5'\text{-Hb}}$= 13.0 Hz, $J_{5'\text{-Ha, }4'\text{-H}}$= 3.1 Hz, 5'-Ha); 2.97 (dd, 1H, $J_{5'\text{-Hb, }5'\text{-Ha}}$= 13.0 Hz, $J_{5'\text{-Hb, }4'\text{-H}}$= 3.4 Hz, 5'-Hb); 3.00 - 3.04 (m, 2H, linker-H); 3.85 (s, 3H, methoxy-H); 3.86 (s, 3H, methoxy-H); 3.97 (m, 1H, 4'-H); 4.23 (m, 1H, 3'-H); 4.71 (m, 1H, 2'-H); 5.18 - 5.19 (m, 1H, OH), 5.31 (s, 2H, NVOC-CH_2); 5.92 (d, 1H, *J* = 7.2 Hz, 1'-H); 6.41 (s, 2H, 6-NH_2); 7.16 (s, 1H, arom. H); 7.49 (t, 1H, *J* = 5.7 Hz, linker-NH); 7.61 (t, 1H, *J* = 5.7 Hz, linker-NH); 7.68 (s, 1H, arom. H); 7.88 (s, 1H, 2-H).

^{13}C-NMR (125.7 MHz, DMSO-d$_6$): δ = 25.63 (linker-C); 26.28 (aziridine-C); 26.39 (linker-C); 26.87 (aziridine-C); 28.56 (linker-C); 41.94 (linker-C); 56.21 (methoxy-C); 59.69 (methoxy-C); 61.71 (5'-C); 62.19 (NVOC-CH_2); 69.95 (2'-C); 70.95 (3'-C); 83.92 (4'-C); 86.01 (1'-C); 108.19 (NVOC-arom. C); 110.73 (2-C); 116.96; 127.82; 139.39; 148.51 (NVOC-arom. C); 150.16; 151.29; 152.09; 153 28; 155.67; 161.14.

MS (ESI) m/z (%): 618.3 (100) $[M+H]^+$.

8-Amino(4''aminobutyl)-5'-aziridinyl-5'-deoxy adenosine (20)



19 **20**

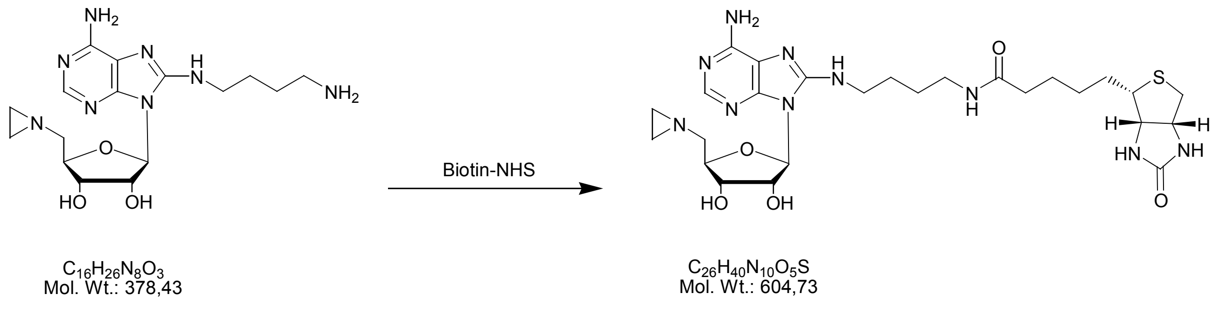
A yellowish solution of nucleoside **19** (0.38 mg, 615 nmol) in DMSO (200 µl) was added to triethanolamine buffer (750 µl, 100 mM, pH 8.0 adjusted with 1 M HCl) and illuminated for 40 min at room temperature using a mercury lamp.

Product formation was analyzed by HPLC. Since complete turnover occurred, the solution was utilized directly in the next step.

R$_t$ = 22 min (C-18 reversed phase column, system 1, gradient 3).

MS (ESI) m/z (%): 379.3 (100) $[M+H]^+$.

8-Amino[1''-(*N*''-biotinyl)-4''aminobutyl]-5'-aziridinyl-5'-deoxy adenosine (26)



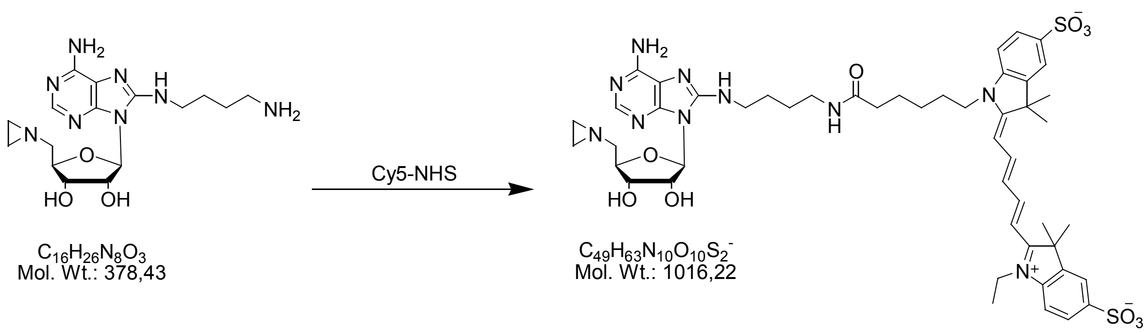
20 **26**

A solution of biotinsuccinimidyl ester (0.66 mg, 1.93 µmol) in DMSO (500 µl) was added to the yellow solution of **20** in DMSO (250 µl) and triethanolamine buffer (750 µl, 100 mM, pH 8.0) and stirred for 40 min at room temperature. The crude product was purified using HPLC chromatography. The amount of product was determined using UV spectroscopy (0.17 mg, 45 %). After freeze drying **26** was obtained as a white solid.

R$_t$ = 32.5 min (C18 reversed phase column, system 2, gradient 3).

^{1}H-NMR (500MHz, DMSO-d$_6$): δ = 1.28 - 1.30 (m, 2H, aliphat. H); 1.31 - 1.32 (m, 2H, aliphat. H); 1.36 - 1.48 (m, 8H, 4x aliphat. H, 2x aziridine-H); 1.55 - 1.59 (m, 2H, aliphat. H); 1.69 - 1.71 (m, 2H, aliphat. H); 1.98 (dd, 1H, $J_{5'\text{-Hb}, 4'\text{-H}}$ = 3.0 Hz, $J_{5'\text{-Hb}, 5'\text{-Ha}}$ = 13.7 Hz, 5'-Hb); 2.02 (t, 2H, aliphat. H); 2.56 (d, 1H, J = 12.8 Hz, S-CH$_2$, Ha); 2.80 (dd, 1H J = 4.9 Hz, J = 12.8 Hz, S-CH$_2$, Hb); 2.98 (dd, 1H, $J_{5'\text{-Ha}, 4'\text{-H}}$ = 3.1 Hz, $J_{5'\text{-Ha}, 5'\text{-Hb}}$ = 13.1 Hz, 5'-Ha); 3.01 - 3.03 (m, 1H, Biotin S-CH); 3.15 - 3.26 (m, 2H, aliphat. H); 3.97 (dt, 1H, $J_{4'\text{-H}, 3'\text{-H}}$ = 2.1 Hz, $J_{4'\text{-H}, 5'\text{-H}}$ = 3.7 Hz, 4'-H); 4.10 (dd, 1H, J = 4.5 Hz, J = 7.9 Hz, biotin-bridgehead S-CHRC$\underline{\text{H}}$); 4.23 (dd, 1H, $J_{3'\text{-H}, 4'\text{-H}}$= 2.1 Hz, $J_{3'\text{-H}, 2'\text{-H}}$= 5.3 Hz, 3'-H); 4.28 (dd, 1H J = 4.7 Hz, J = 7.7 Hz, biotin-bridgehead S-CH$_2$-C$\underline{\text{H}}$); 4.69 (dd, 1H, $J_{2'\text{-H}, 3'\text{-H}}$= 5.1 Hz, $J_{2'\text{-H}, 1'\text{-H}}$= 7.3 Hz, 2'-H); 5.92 - 5.93 (d, 1H, $J_{1'\text{-H}, 2'\text{-H}}$ = 7.3 Hz, 1'H); 6.34 (s, br, 1H, biotin-NH); 6.39 (s, 2H, 6-NH$_2$); 6.40 (s, 1H, biotin-NH) 7.57 (t, 1H, linker-NH); 7.78 (t, 1H, linker-NH); 7.88 (s, 1H, 2-H).
^{13}C-NMR (125.7 MHz, DMSO-d$_6$): δ = 25.31, 25.65, 26.83, 27.02 (2x linker-C, 2x aziridine-C); 28.35 (linker-C); 28.55 (linker-C); 28.97 (linker-C); 35.55 (linker-C); 38.53 (biotin S-$\underline{\text{C}}$H-R); 42.32 (linker-C); 55.73; 59.51 (biotin-bridgehead S-CH$_2$-$\underline{\text{C}}$H); 61.36 (biotin-bridgehead S-CHR-$\underline{\text{C}}$H); 62.06 (5'-C); 70.28 (2'-C); 71.30 (3'-C); 74.83 (4'-C); 84.30 (1'-C); 86.30; 117.27; 148.83 (2-C); 150.49; 151.59; 152.41; 163.02; 172.23.
MS (ESI) m/z (%): 605.3 (100) $[M+H]^+$.

8-Amino[1''-(*N*''-Cy5)-4''-aminobutyl]-5'-aziridinyl-5'-deoxy adenosine (27)



20 **27**

A yellowish solution of nucleoside **20** (6 mg, 9.7 µmol) in triethanolamine buffer (1 ml, 100 mM, pH 8.0) was treated with Cy5-NHS ester (6 mg, 7.8 µmol) in triethanolamine buffer (1 ml, 100 mM, pH 8.0). The solution was incubated for 30 min at room temperature. The crude product was purified by reversed phase HPLC yielding **27**. The amount of product was determined by UV-spectroscopy (1.7 mg, 17 %) and the solution freeze dried.
R$_t$ = 42 min (C-18 reversed phase column, system 2, gradient 3).

MS (ESI) m/z (%): 1017.4 (100) $[M+H]^+$.

5.3 Protein synthesis and protein analytical methods

5.3.1 Protein expression and purification

5.3.1.1 M·*Taq*I expression and purification

The expression and purification of M·*Taq*I was carried out as described in Holz *et al.* (1998). The protein had to be free of the natural cofactor in order to measure cofactor analogue binding correctly. The natural cofactor AdoMet was removed as described in (Goedecke *et al.*, 2001). M·*Taq*I was stored in TrisOAc (20 mM), KCl (300 mM), KOAc (50 mM), $Mg(OAc)_2$ (10 mM), pH 7.9, DTT (1 mM) and glycerol (55 %) at -20 °C.

5.3.1.2 M·*Hha*I and M·*Bcn*IB expression and purification

The C5-cytosine DNA MTase M·*Hha*I and the N4-cytosine DNA MTase M·*Bcn*IB were kindly provided by Saulius Klimasauskas (Vilnius, Lithuania).

M·*Hha*I was stored KH_2PO_4 (10 mM), EDTA (5 mM), NaCl (100 mM), pH 7.4, 2-mercaptoethanol (10 mM) and glycerol (50 %) at -20 °C.

M·*Bcn*IB was stored in KH_2PO_4 (10 mM), EDTA (0.1 mM), KCl (100 mM), pH 7.5, 2-mercaptoethanol (7 mM) and glycerol (50 %) at -20 °C.

5.3.2 Determination of protein concentration

The protein concentration was determined using the method described by Bradford (1976). The method is based on the emission maximum shift from 465 nm to 595 nm of Coomassie Blue G-250 after binding to proteins. Protein solutions were diluted with water (up to 100 µl) and Bradford reagent was added to give a final volume of 1000 µl. After 10 min the absorption of the solution at 595 nm was measured and the protein concentration calculated using a bovine serum albumin standard.

5.3.3 Sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE)

The SDS-PAGE gel electrophoresis was carried out using the buffer system described by Schägger and von Jagow (1987). Proteins were taken up in probe buffer (2 % SDS, 10 % glycerol, 62 mM Tris/HCl, 5 % 2-mercapto ethanol, 0.001 % bromphenol blue, pH 6.8) and denatured for 5 min at 95 °C. As a molecular weight standard for proteins a mixture of phosphorylase β (94 kD), bovine serum albumine (67 kD), ovalbumine (43 kD), carbonicanhydrase (30 kD), trypsin inhibitor (20,1 kD) und lysozyme (14,4 kD) was used.

After gel electrophoretic separation of the proteins at 40 mA per gel, the gels were stained with Coomassie Brilliant Blue G250.

For the gel preparation the following solutions were used:

Tab. 5.3.3: Gel composition.

	Stacking gel solution for 10 gels, 4 % acryl amide	Resolving gel solution for 10 gels, 12 % acrylamide
Acryl amide (49.5 %)/ *N*,*N*'-bisacrylamide (3 %)	1.8 ml	9 ml
Gel buffer	6 ml	12.5 ml
H_2O	14.5 ml	16 ml
APS (10 %)	125 µl	190 µl
TEMED	20 µl	20 µl

Gel buffer:	363.5 g Tris, 3 g SDS, 0.1 g sodium azide, pH 8.45 (per liter).
Anode buffer:	241.5 g Tris/HCl, pH 8.9 (per liter).
Cathode buffer:	121.2 g Tris, 180 g tricine, 10 g SDS, 0.1 g sodium azide, pH 8.25 (per liter).
Staining solution:	15 % Coomassie Blue in destaining solution I.
Destaining solution I:	2 L H_2O, 2 L ethanol and 1 L acetic acid.
Destaining solution II:	10 % acetic acid in water.

5.3.4 Activity assay for M·*Taq*I

For the determination of the activity of M·*Taq*I a protection activity assay was carried out. Solutions (20 µl) of λ-DNA (50 ng/ml, 2.1 nM, 254 nM M·*Taq*I recognition sequences), AdoMet (80 µM), BSA (0.1 mg/ml) and 1.0, 2.0, 3.9, 7.8, 15.6, 31.3, 62.5, 125, 250 or 500 ng/ml M·*Taq*I in TrisOAc (20 mM, pH 7.9), $Mg(OAc)_2$ (10 mM), KOAc (50 mM), DTT (1 mM) and Triton X-100 (0.01 %) were incubated for 60 min at 65 °C.

Subsequently, a solution (30 µl) of R·*Taq*I (40 U) in Tris/HCl (10 mM, pH 8.4), NaCl (100 mM), $MgCl_2$ (10 mM) and BSA (0.1 mg/ml) was added to each of the solutions and the resulting solutions were incubated for 15 min at 65 °C. The solutions were treated with the probe buffer (10 µl: 0.25 % bromphenol blue and 30 % glycerol) before aliquots (12 µl) of the solutions were loaded onto an agarose gel (1 %). The DNA was completely protected against restriction at a concentration of 7.8 ng/ml M·*Taq*I.

5.4 Methods related to DNA

5.4.1 Hybridisation of Oligodeoxynucleotides

Equimolar amounts of complementary oligonucleotides in appropriate buffer were treated 2 min at 95 °C in a heat block. The hybridization of DNA occurred during the cooling period (2-3 h) in the heat block.

5.4.2 Agarose gel electrophoresis

Large DNA fragments were analyzed using horizontal agarose gel electrophoresis. Gels of 0.7-1 % agarose in 1 x TBE-Buffer (pH 8.9, composition see 5.4.3) containing ethidium bromide (10 µg/ml) were used. Stained DNA bands were visualized using UV light at 302 nm and documented using a Polaroid camera.

5.4.3 Analytical polyacrylamide-urea gel electrophoresis

Analytical polyacrylamide-urea gel electrophoresis is used for the high resolution separation of DNA and RNA fragments, respectively. A 0.4 mm thick gel of 40 cm height was prepared in a gel chamber from BioRad (Munich, Germany). Depending on the length of the nucleic acid fragments which had to be separated either 10 % or 20 % acrylamide solution with urea (7 M) in 1 x TBE buffer were prepared. In a pre-run the gels were heated to 56 °C before the samples were loaded. The samples were mixed with formamide buffer and denatured for 3 min at 95 °C. The gel electrophoresis was carried out at 2000-2500V in order to keep the gel temperature at 50-56 °C. The running buffer was 1 x TBE buffer. After electrophoresis the gels were fixed in acetic acid (10 %) for 5 min and dried in a gel drier on Whatman GB 002 blotting paper for 2 h at 80 °C. The analysis of the gels was performed with a Fuji gel scanner.

Tab. 5.4.3: Gel composition.

	10 % gel	20 % gel
Acrylamide solution (40 %)	25 ml	50 ml
10 x TBE buffer	10 ml	10 ml
Urea	42 g	42 g
H_2O	65 ml	40 ml
APS (10 %)	250 µl	250 µl
TEMED	100 µl	100 µl

TBE buffer (10 x): 108 g Tris, 55 g boric acid, 3.35 g EDTA (per liter, pH not adjusted).

Formamide buffer: 80 % Formamide, 1 x TBE.

5.4.4 Restriction of DNA

Analytical amounts of plasmid DNA (200 ng - 1 µg) were treated with 10 U per µg plasmid of the appropriate restriction enzyme for 1 h at the designated temperatures using the recommended buffers supplied by the manufacturers.

5.5 X-Ray crystallography

5.5.1 Setups for crystallization

M·*Taq*I was concentrated and transferred into the crystallization buffer using vacuum dialysis. The protein solution (400 µl) was mixed with 400 µl crystallization buffer (10 mM Tris/HCl, 300 mM NaCl, pH 7.3) and dialyzed overnight at 4 °C against the crystallization buffer using ultra thimbles (10 kD cutoff). Subsequently, the protein solution was concentrated (approx. 1 h) until the desired concentration was obtained. A solution (30 µl) of the concentrated protein and cofactor analogue in Tris/HCl (10 mM) and NaCl (300 mM) was mixed with a solution (30 µl) of previously hybridized DNA (100 µM) in Tris/HCl (10 mM) and NaCl (300 mM). The resulting mixture was centrifuged for 10 min prior to use for crystallization setups.

5.5.2 Crystallization and mounting crystals

Crystallization setups were performed in Linbro cell culture plates using the hanging drop method. The edges of the 24 wells were first greased using Baysilone paste, medium viscous (Bayer AG, Germany). Previously prepared reservoir solutions (1 ml) were added to the appropriate wells. The reservoir solution (3 µl) and the protein solution (1 µl) were mixed on siliconized glass cover slides to form a drop. The obtained solution was micro seeded using crystal solutions described by (Goedecke *et al.*, 2001). The cover slides were turned upside down and attached carefully to the wells. Crystal growth was achieved at 20 °C.
Crystallization conditions were chosen by varying the conditions described by Goedecke *et al.* (2001), explicitly by varying the concentration of the precipitant. Crystals of all three complexes grew after 2-3 days. Optimal crystallization conditions for each complex are listed below.

Tab. 5.5.2: Crystallization conditions.

	Ternary complex formed from M·*Taq*I, DNA, BAZ	Ternary complex M·*Taq*I, DNA (2AP), AETA	Ternary complex M·*Taq*I, DNA (2AP), AdoMet
Protein Solution	M·*Taq*I (95 µM) O10/U10 (100 µM) NaCl (300 mM) Tris/HCl (10 mM, pH 7.3) BAZ (250 µM)	M·*Taq*I (95 µM) O10Ap/U10 (100 µM) NaCl (300 mM) Tris/HCl (10 mM, pH 7.3) AETA (1 mM)	M·*Taq*I (95 µM) O10Ap/U10 (100 µM) NaCl (300 mM) Tris/HCl (10 mM, pH 7.3) AdoMet (1 mM)
Reservoir	KCl (100 mM) $MgCl_2$ (25 mM) IPA (21 %) Sodium cacodylate (50 mM, pH 6)	KCl (100 mM) $MgCl_2$ (25 mM) IPA (6 %) Sodium cacodylate (50 mM, pH 6)	KCl (100 mM) $MgCl_2$ (25 mM) IPA (6 %) Sodium cacodylate (50 mM, pH 6)

The obtained crystals were measured under cryo conditions. Therefore, crystals with a dimension from 10 x 10 x 100 μm^3 to 10 x 30 x 100 μm^3 were harvested with a nylon fibre loop out of the drop, transferred quickly from the mother liquor into the cryo protectant (12 % PEG6000, 200 mM NaCl, 100 mM Tris/HCl, pH 7.3, 17.1 % glycerol) and frozen with liquid nitrogen. The loop containing the crystal was placed onto a goniometer head and measured under a nitrogen stream at 100 K.

5.5.3 Data collection and structure determination

Due to the small size of crystals, X-ray diffraction data were collected with a synchrotron beam source. The measurements were performed in Grenoble at the ESRF beamline ID14 by Dr. Axel Scheidig, Max-Planck Institut Dortmund, Germany. Data statistics are given in the result section.

Processing of the measured reflexes was performed using the programs XDS/XSCALE (Kabsch, 1993). Phase determination of the structure amplitudes was achieved using molecular replacement. As a model the ternary structure from Goedecke *et al.* (2001) was used omitting the cofactor, the extrahelical base and water molecules in order to avoid model bias. Refinement was performed using the program CNS. The interpretation of the electron density was carried out using the program O. The computational calculations were carried out by Dr. Axel Scheidig, Max-Planck Institut Dortmund, Germany.

5.6 Coupling of aziridine cofactors to different DNA substrates

5.6.1 Coupling of AZ with 7·8 using M·*Taq*I

A solution of double-stranded oligonucleotide **7·8** (10 µM), M·*Taq*I (10 µM) and the aziridine cofactor AZ (1 mM) in TrisOAc (20 mM), Triton X-100 (0.01 %), $Mg(OAc)_2$ (10 mM) and KOAc (50 mM) at different pH values (5.5-7.9) was incubated for 24 h at 37 °C. The reaction was analyzed using an anion exchange column (Poros 10 HQ, system 3, gradient 1). The following retention times were observed: Complex of DNA/protein/cofactor 8.8 min, modified DNA $\mathbf{7^{AZ}\cdot 8}$ 22.8 min and double-stranded DNA **7·8** 24.1 min.

5.6.2 Coupling of AZ with 9·10 using M·*Hha*I

A solution of double-stranded oligonucleotide **9·10** (10 µM), M·*Hha*I (10 µM) and aziridine cofactor AZ (1 mM) in Tris/HCl (10 mM), EDTA (0.5 mM) and NaOAc (50 mM) at different pH values (5.8-7.4) was incubated for 24 h at 37 °C. The reaction was analyzed using an anion exchange column (Poros 10 HQ; system 3, gradient 1). The following retention times were observed: Complex of DNA/protein/cofactor 11.2 min, modified DNA $\mathbf{9^{AZ}\cdot 10}$ 21.4 min and double-stranded DNA **9·10** 22.4 min.

5.6.3 Coupling of AZ with 11·12 using M·*Bcn*IB

A solution of double-stranded oligonucleotide **11·12** (10 µM), M·*Bcn*IB (10 µM) and aziridine cofactor AZ (1 mM) in TrisOAc (20 mM), $Mg(OAc)_2$ (10 mM) and KOAc (50 mM), pH 6.9, was incubated for 24 h at 37 °C. The reaction was analyzed using an anion exchange column (Poros 10 HQ; system 3, gradient 1). The following retention times were observed: modified DNA **11^{AZ}·12** 29.3 min, double-stranded DNA **11·12** 30.2 min.

5.6.4 Labeling of pUC19 with DAZ using M·*Taq*I

Labeling of pUC19 (4 recognition sites for M·*Taq*I) with the dansyl fluorophore was achieved using M·*Taq*I and the dansylated aziridine cofactor DAZ. A solution of pUC19 (38.9 nM), M·*Taq*I (186.7 nM) and DAZ (20 µM) in TrisOAc (20 mM), Triton X-100 (0.01 %), $Mg(OAc)_2$ (10 mM) and KOAc (50 mM), pH 6, was incubated for 8 h at 65 °C. The reaction was analyzed by anion exchange chromatography (Nucleogen 4000-7 DEAE; system 4, gradient 2). The plasmid elutes with a retention time of 21.6-21.9 min.

5.6.5 Labeling of pUC19 with DAZ using M·*Hha*I

Labeling of pUC19 (13 recognition sites for M·*Hha*I) with the dansylated fluorophore was achieved using M·*Hha*I and the dansylated aziridine cofactor DAZ. A solution of pUC19 (38.9 nM), M·*Hha*I (730 nM) and DAZ (20 µM) in Tris/HCl (10 mM), EDTA (0.5 mM) and NaOAc (50 mM), pH 6.85, was incubated for 20 h at 37 °C. The reaction was analyzed by anion exchange chromatography (Nucleogen 4000-7 DEAE; system 4, gradient 2). The plasmid elutes with a retention time of 21.6-21.9 min.

5.6.6 Labeling of 7·8 with biotin using M·*Taq*I

Labeling of **7·8** with biotin was achieved using M·*Taq*I and the biotinylated aziridine cofactor BAZ. A solution of double-stranded oligonucleotide **7·8** (10 µM), M·*Taq*I (10 µM) and BAZ (80 µM) in TrisOAc (20 mM), Triton X-100 (0.01 %), $Mg(OAc)_2$ (10 mM) and KOAc (50 mM) at different pH values (5.6-7.9) was incubated for 24 h at 37 °C. The reaction was analyzed by anion exchange chromatography (Poros 10 HQ; system 3, gradient 1). The following retention times were observed: Complex of DNA/protein/cofactor 8.2 min, modified DNA **7^{BAZ}·8** 24.6 min and double-stranded DNA **7·8** 25.6 min.

5.6.7 Labeling of pUC19 with biotin using M·*Taq*I

Labeling of pUC19 (4 recognition sites for M·*Taq*I) with biotin was achieved using M·*Taq*I and the biotinylated aziridine cofactor BAZ. A solution (100 µl) of pUC19 (27.8 nM, 5 µg), M·*Taq*I (133.3 nM) and BAZ (80 µM) in TrisOAc (20 mM), Triton X-100 (0.01 %), $Mg(OAc)_2$ (10 mM) and KOAc (50 mM), pH 6, was incubated for 4 h at 65 °C. The product was treated with proteinase K (see 5.6.8) and purified with the QiagenPCR purification Kit according to the instructions given by the manufacturers. The purified plasmid (3 µg, 60 % yield) was treated with the restriction enzyme R·*Bsi*EI (30 U) in Tris/HCl (10 mM), $MgCl_2$ (10 mM), NaCl (50 mM), pH 7.9, and DTT (1 mM) for 1 h at 60 °C. The resulting DNA fragments were analyzed by agarose gel electrophoresis. To an aliquot of the solution (2 µl containing 200 ng DNA) streptavidin-fluorescein (1 µg) was added and the solution was incubated 15 min at room temperature. The solutions with and without streptavidin-fluorescein were loaded onto an agarose gel (0.7 %) and biotinylated DNA bands were visible by a gel mobility shift.

5.6.8 Release of modified DNA from MTase-DNA complexes by Proteinase K

The labeling solutions described above were all treated with proteinase K in order to release the modified DNA from tightly bound MTase. The pH of the solutions was adjusted to 8 and proteinase K (4 mg/ml) was added (10 µg per µg DNA). Further the DNA was purified using the Qiagen PCR purification Kit.

5.6.9 Mass spectra of duplex deoxyoligonucleotides

The MALDI mass spectra of duplex oligonucleotides were obtained using a Voyager system. The desalted sample of the DNA was pipetted (0.5 µl) together with the matrix (0.5 µl; 10 mg/ml THAP, 50 mg/ml citric acid and 50 % acetonitrile) onto a MALDI plate. The drop was air dried before citric acid was used to remove remaining salt. The drop was air dried again before loading the plate into the mass spectrometer. The measurements were performed using the instrument settings listed in Table 5.6.9.

Table 5.6.9: Instrument settings for MALDI mass spectrometry.

Mode of operation	Reflector and linear
Extraction mode	Delayed
Polarity	Negative
Acquisition control	Manual
Accelerating voltage	20000 V
Grid voltage	76 %
Mirror voltage ratio	1.12
Guide wire 0	0.03 %
Extraction delay time	450 nsec
Laser intensity	2969
Laser Repetition Rate	2.8 Hz
Timed ion selector	Off

5.7 Biophysical methods

5.7.1 Fluorescence spectroscopy for the determination of K_D-values: Binding to M·*Taq*I

Fluorescence spectroscopy was applied to determine the affinity of the synthesized cofactor analogues to M·*Taq*I. All measurements were performed in TrisOAc (20 mM), Triton X-100 (0.01 %), $Mg(OAc)_2$ (10 mM) and KOAc (50 mM) at pH 7.9 and 25 °C. In the case of an attached fluorophore to the cofactor a direct titration and in the case of a biotinylated cofactor a competitive titration was carried out using MANT-AdoHcy as competitor.

Direct titration

To a solution (300 µl) of cofactor analogue (1 µM) a solution (400 µl) of M·*Taq*I (75 µM) and cofactor analogue (1 µM) was stepwise added until no fluorescence change could be observed. The fluorescence was measured after every single addition of protein. The resulting data were fitted using the solution of the quadratic binding equation which describes the complex concentration in a bimolecular equilibrium of 1:1 stochiometry. The quadratic equation is derived from the reaction equation of a reversible equilibrium of enzyme (E) and ligand (L) as described below.

$$E + L \rightleftharpoons EL$$

The equilibrium dissociation constant is derived from the mass action law (1).

$$K_D = \frac{\left[E_{eq}\right]\left[L_{eq}\right]}{\left[EL_{eq}\right]} \tag{1}$$

The equilibrium concentrations $[E_{eq}]$,$[L_{eq}]$ and $[EL_{eq}]$ are related to starting concentration $[E_0]$ and $[L_0]$ using mass conservation law (2).

$$\left[E_{eq}\right] = \left[E_0\right] - \left[EL_{eq}\right] \text{ and } \left[L_{eq}\right] = \left[L_0\right] - \left[EL_{eq}\right] \tag{2}$$

By inserting equation (2) in (1) and reorganizing the quadratic binding equation results.

$$0 = \left[EL_{eq}\right]^2 - \left[EL_{eq}\right]\left(K_D + [L_0] + [E_0]\right) + \left([L_0][E_0]\right) \tag{3}$$

Its reasonable solution is

$$\left[EL_{eq}\right] = \frac{\left(K_D + [L_0] + [E_0] - \sqrt{\left(K_D + [L_0] + [E_0]\right)^2 - 4[L_0][E_0]}\right)}{2} \tag{4}$$

The partial increase in fluorescence emission (F-F_{min}) of the fluorophore per titration step equals the binding rate ($[EL_{eq}]/[L_0]$) times maximal fluorescence (F_{max}-F_{min}). The total fluorescence (F) equals the sum of the starting fluorescence (F_{min}) and the partial fluorescence increase of a titration step:

$$F = F_{\min} + \left(F_{\max} - F_{\min}\right) \cdot \frac{\left[EL_{eq}\right]}{\left[L_0\right]} \tag{5}$$

Inserting equation (4) into (5) leads to

$$F = F_{\min} + \left(F_{\max} - F_{\min}\right)\frac{\left(K_D + [L_0] + [E_0] - \sqrt{\left(K_D + [L_0] + [E_0]\right)^2 - 4[L_0][E_0]}\right)}{2[L_0]} \tag{6}$$

The measured titration curves represent F as a function of $[E_0]$. Fitting of the data using equation (6) gives the K_D value of the investigated equilibrium.

Competitive titration

The determination of K_D values of cofactor analogues without an intrinsic fluorescent property to M·*Taq*I was carried out by competitive titration using MANT-AdoHcy as fluorescent competitor. The prerequisite for such a competitive titration is that the competitor and the cofactor analogue bind at the same location of the enzyme and that the K_D value of the competitor is known.

To a solution (250 µl) of MANT-AdoHcy (1 µM) and M·*Taq*I (2 µM) a solution (400 µl) of BAZ (200 µM), MANT-AdoHcy (1 µM) and M·*Taq*I (2 µM) was added stepwise leading to a fluorescence decrease due to the replacement of MANT-AdoHcy.

As the dissociation constant of MANT-AdoHcy and M·*Taq*I was already determined in a direct titration (Pignot, 1999) the program Scientist could be used to determine iteratively the K_D value of the cofactor analogue BAZ. The following model was used for the iterative description of the measured data (Table 5.7.1a).

Table 5.7.1a: Model for the competitive titration with slope.

// Comp. Titration	Definitions	
IndVars: C	A:	[MANT-AdoHcy]
DepVars: A, B, AB, BC, Cf, F	B:	[M·*Taq*I]
Params: K1, K2, ATOT, BTOT, Ya, Yab, Ysl	C:	[cofactor analogue BAZ]$_0$
AB=A*B/K1	Cf:	[cofactor analogue BAZ]
BC=B*Cf/K2	F:	fluorescence
ATOT=A+AB	ATOT:	[MANT-AdoHcy]$_0$
BTOT=B+AB+BC	BTOT:	[M·*Taq*I]$_0$
C=Cf+BC	Ya:	F of free A
0<A<ATOT	Yab:	F of AB
0<B<BTOT	Ysl:	F due to inner filter effect
0<Cf<C	K1:	K_D of complex AB
F=A*Ya+AB*Yab-Ysl*Cf	K2:	K_D of complex BC

For the calculation of K2, Ya and Yab the independently determined values of K1 (14.9 µM) and Ysl (9.1 μM^{-1}) were hold constant.
Table 5.6.1b shows the instrument settings and the type of titration for each cofactor.

Table 5.7.1b: Instrument settings and titration type.

Cofactor	Titration method	λ_{Ex} / nm (slit width)	λ_{Em} / nm (slit width)
BAZ	Competitive	295 (1)	430 (16)
Cy5AZ	Direct	649 (1)	665 (16)

5.7.2 Concentration determination of cofactor analogues

The concentration of the cofactor analogues was determined by UV-spectroscopy. The extinction coefficients for the determination were either of adenosine or the fluorescent group attached. For the N8-modified adenosine analogues with no fluorescent group (BAZ) the extinction coefficient ε^{278} = 20450 L cm^{-1} $mol^{-1}$of 8-Amino[1''-(6''-aminohexyl)]-adenosine-5'-phosphate (Barker *et al.*, 1974) was used.

CHAPTER 6

References

6 References

Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M. & Harrison, S.C. (1988). Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science* **242**, 899-907.

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U. & Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl-CpG-binding protein 2. *Nature Genet.* **23**, 185-188.

Ban, N., Nissen, P., Hansen, J., Moore, P.B. & Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 angstrom resolution. *Science* **289**, 905-920.

Bailly, C., Braña, M. & Waring, M.J. (1996). Sequence-selective intercalation of antitumour bis-naphthalimides into DNA - Evidence for an approach via the major groove. *Eur. J. Biochem.* **240**, 195-208.

Barany, F., Slatko, B., Danzitz, M., Cowburn, D., Schildkraut, I. & Wilson, G.G. (1992). The corrected nucleotide sequences of the TaqI restriction and modification enzymes reveal a thirteen-codon overlap. *Gene* **112**, 91-95.

Barker, R., Trayer, I.P. & Hill, R.L. (1974). Nucleoside phosphates attached to Agarose. *Methods Enzymol.* **34**, 479-491.

Beal, P.A. & Dervan, P.B. (1991). Second structural Motif for Recognition of DNA by Oligonucleotide-Directed Triple-Helix Formation. *Science* **251**, 1360-1363.

Bestor, T.H. & Tycko, B. (1996). Creation of genomic methylation patterns. *Nature Genet.* **12**, 363-367.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analyt. Biochem.* **72**, 248-254.

Braña, M.F., Cacho, M., Gradillas, A., de Pascual-Teresa, B. & Ramos, A. (2001). Intercalators as Anticancer Drugs. *Curr. Pharm. Design* **7**, 1745-1780.

Brünger, A.T. (1992). The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-474.

Buist, G.J. & Lucas, H.J. (1957). Basicity Constants and Rates of Hydration of Some Imines. *J. Am. Chem. Soc.* **79**, 6157-6160.

Bussiere, D.E., Muchmore, S.W., Dealwis, C.G., Schluckebier, G., Nienaber, V.L., Edalji, R.P., Walter, K.A., Ladror, U.S., Holzman, T.F. & Abad-Zapatero, C. (1998). Crystal Structure of ErmC', an rRNA Methyltransferase Which Mediates Antibiotic Resistance in Bacteria. *Biochemistry* **37**, 7102-7112.

Cantor, C.R., Warshaw, M.M. & Shapiro, H. (1970). Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers* **9**, 1059-1077.

Caruthers, M.H., Barone, A.D., Beaucage, S.L., Dodds, D.R., Fisher, E.F., McBride, L.J., Matteucci, M., Stabinsky, Z. & Tang, J.Y. (1987). Chemical synthesis of deoxyoligonucleotides. *Methods Enzymol.* **154**, 287-313.

Caserta, M., Zacharias, W., Nwankwo, D., Wilson, G.G. & Wells, R.D. (1987). Cloning, sequencing, in vivo promoter mapping, and expression in Escherichia coli of the gene for the HhaI methyltransferase. *J. Biol. Chem.* **262**, 4770-4777.

Chen, L., MacMillan, A.M. & Verdine, G.L. (1993). Mutational Separation of DNA Binding from Catalysis in a DNA Cytosine Methyltransferase. *J. Am. Chem. Soc.* **115**, 5318-5319.

Chen, X., Ramakrishnan, B., Rao, S.T. & Sundaralingam, M. (1994). Binding of two Distamycin A molecules in the minor groove of an alternating B-DNA duplex. *Nat. Struct. Biol.* **1**, 169-175.

Chen, R.Z., Petterson, U., Beard, C., Jackson-Grusby, L. & Jaenisch, R. (1998). DNA hypomethylation leads to elevated mutation rates. *Nature* **395**, 89-93.

Clemons, W.M. Jr., May, J.L.C., Wimberly, B.T., McCutcheon, J.P., Capel, M.S. & Ramakrishnan, V. (1999). Structure of a bacterial 30S ribosomal subunit at 5.5Å resolution. *Nature* **400**, 833-840.

Coll, M., Frederick, C.A., Wang, A.H. & Rich, A. (1987). A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8385-8389.

Cox, R. & Irving, C.C. (1977). Inhibition of DNA Methylation by S-Adenosylethionine with the Production of Methyl-deficient DNA in Regenerating Rat Liver. *Cancer Research* **37**, 222-225.

Dervan, P.B. (1986). Design of sequence-specific DNA-binding molecules. *Science* **232**, 464-471.

Dervan, P.B. & Bürli, R.W. (1999). Sequence-specific DNA recognition by polyamides. *Curr. Opinion Struct. Biol.* **3**, 688-693.

Dervan, P.B. (2001). Molecular Recognition of DNA by Small Molecules. *Bioorg. Med. Chem.* **9**, 2215-2235.

Dey, S. & Sheppard, T.L. (2001). Ketone-DNA: A Versatile Postsynthetic DNA Decoration Platform. *Org. Lett.* **3**, 3983-3986.

Djordjevic, S. & Stock, A.M. (1997). Crystal structure of the chemotaxis receptor methyltransferase CheR suggests a conserved structural motif for binding S-adenosylmethionine. *Structure* **5**, 545-558.

Ellenberger, T.E., Brandl, C.J., Struhl, K. & Harrison, S.C. (1992). The GCN4 Basic Region Leucine Zipper Binds DNA as a Dimer of Uninterrupted α Helices: Crystal Structure of the Protein-DNA Complex. *Cell* **71**, 1223-1237.

Escudé, C., Sun., J.-S., Rougeé, M., Garestier, T. & Hélène, C. (1992). Stable triple helices are formed upon binding of RNA oligonucleotides and their 2'-O-methyl derivatives to double-helical DNA. *C.R. Acad. Sci. Paris*, Ser. III **315**, 521-525.

Escudé, C., Nguyen, C.H., Kukreti, S., Janin, Y., Sun, J.-S., Bisagni, E., Garestier, T. & Hélène, C. (1998). Rational design of a triple helix-specific intercalating ligand. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3591-3596.

Felsenfeld, G., Davies, D.R. & Rich, A. (1957). Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.* **79**, 2023-2024.

Fersht, A. (1999). Structure and Mechanism in Protein Science: A guide to Enzyme Catalysis and Protein Folding. *W.H. Freeman*, New York, 163.

Finer, J.T., Simmons, R.M. & Spudich, J.A. (1994). Single myosin molecule mechanics: piconewton force and nanometer steps. *Nature* **368**, 113-119.

Friederich, A. (2001). Untersuchungen zur DNA-Bindung der DNA-Methyltransferase aus *Thermus aquaticus*. *Dissertation*.

Gabriel, S. (1888). Ueber Vinylamin und Bromäthylamin. *Chem. Ber.* **21**, 2665.

Goedecke, K. (2000). Kristallstruktur der N6-Adenin-DNA-Methyltransferase aus *Thermus aquaticus* im Komplex mit DNA und einem Cofaktoranalogon - Der katalytische Mechanismus und der durch DNA-Kompression induzierte Nukleotidausklappmechanismus. *Dissertation*.

Goedecke, K., Pignot, M., Goody, R.S., Scheidig, A.J. & Weinhold, E. (2001). Structure of the N6-adenine DNA methyltransferase M·*Taq*I in complex with DNA and a cofactor analog. *Nat. Struct. Biol.* **8**, 121-125.

Gong, W., O'Gara, M., Blumenthal, R.M. & Cheng, X. (1997). Structure of PvuII DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment. *Nucleic Acids Res.* **25**, 2702-2715.

Guelev, V., Harting, M., Lokey, R.S. & Iverson, B.L. (2000). Altered sequence specificity identified from a library of DNA-binding small molecules. *Chem. Biol.* **7**, 1-8.

Guelev, V., Lee, J., Ward, J., Sorey, S., Hoffman, D.W. & Iverson, B.L. (2001). Peptide *bis*-intercalator binds DNA via threading mode with sequence specific contacts in the major groove. *Chem. Biol.* **8**, 415-425.

Haines, J.A., Reese, C.B. & Todd, L. (1964). The methylation of nucleosides and mononucleotides with diazamethane. *J. Chem. Soc.*, 1406-1412.

Hansen, R.S., Wijmenga, C., Luo, P., Stanek, A.M., Canfield, T.K., Weemaes, C.M.R. & Gartler, S.M. (1999). The *DNMT3B* DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc. Natl. Acad. Sci. USA* **96**, 14412-14417.

Hélène, C. & Lancelot, G. (1982). Interactions between functional-groups in protein-nucleic acid associations. *Prog. Biophys. Mol. Biol.* **39**, 1-68.

Hélène, C. & Toulmè, J.J. (1990). Specific regulation of gene-expression by antisense, sense and antigene nucleic-acids. *Biochim. Biophy. Acta* **1049**, 99-125.

Hélène, C. (1991). The anti-gene strategy – Control of gene-expression by triplex-forming-oligonucleotides. *Anti-Cancer Drug Design* **6**, 569-584.

Hesse, M., Meier, H. & Zeeh, B. (1991). Spektroskopische Methoden in der organischen Chemie. *G. Thieme Verlag*, Stuttgart.

Hodel, A.E., Gershon, P.D., Shi, X. & Quiocho, F.A. (1996). The 1.85 Å Structure of Vaccinia Protein VP39: A Bifunctional Enzyme That Participates in the Modification of Both mRNA Ends. *Cell* **85**, 247-256.

Holz, B., Klimasauskas, S., Serva, S. & Weinhold, E. (1998). 2-Aminopurine as a fluorescent probe for DNA base flipping by methyltransferases. *Nucleic Acids Res.* **26**, 1076-1083.

Holz, B., Dank, N., Eickhoff, J.E., Lipps, G., Krauss, G. & Weinhold E. (1999). Identification of the binding site for the extrahelical target base in N-6-adenine DNA methyltransferases by photo-cross-linking with duplex oligodeoxyribonucleotides containing 5-iodouracil at the target position. *J. Biol. Chem.* **274**, 15066-15072.

Janulaitis, A.A., Povilionis, P. & Sasnauskas, K. (1982). Cloning of the modification methylase gene of Bacillus centrosporus in Escherichia coli. *Gene* **20**, 197-204.

Janulaitis, A.A., Klimasauskas, S., Petrusyte, M. & Butkus, V. (1983). Cytosine modification in DNA by BcnI methylase yields N4-methylcytosine. *FEBS Lett.* **161**, 131-134.

Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in these Models. *Acta Cryst.* **A47**, 110-119.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795-800.

Kessler, C. (1992). Nonradioactive Labeling and Detection of Biomolecules. *Springer Verlag*, Berlin, 13-14.

Klimasauskas, S., Kumar, S., Roberts, R.J. & Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357-369.

Klug, A. & Rhodes, D. (1987). Zinc fingers: a novel protein fold for nucleic acid recognition. *Cold Spring Harbor Symposia on Quantitative Biology* **52**, 473-482.

Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.

Kremer, E.J., Pritchard, M. Lynch, M. Yu, S., Holman, K., Baker, E., Warren, S.T., Schhlessinger, D., Sutherland, G.R. & Richards, R.I. (1991). Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science* **252**, 1711-1714.

Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R.J. & Wilson, G.G. (1994). The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.* **22**, 1-10.

Labahn, J., Granzin, J., Schluckebier, G., Robinson, D.P., Jack, W.E., Schildkraut, I. & Saenger, W. (1994). Three-dimensional structure of the adenine-specific DNA methyltransferase M·*Taq*I in complex with the cofactor S-adenosylmethionine. *Proc. Natl. Acad. Sci. USA* **91**, 10957-10961.

Lauster, R., Trautner, T.A. & Noyer-Weidner, M. (1989). Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognition domains. *J. Mol. Biol.* **206**, 305-312.

Le Doan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J.-L., Thuong, N.-T., Lhomme, J. & Hélène, C. (1987). Sequence-Specific Recognition, Photo-Cross-Linking and cleavage of the DNA double helix by an oligo-[alpha]-thymidylate covalently linked to an azidoproflavine derivative. *Nucleic Acids Res.* **15**, 7749-7760.

Lee, J.S., Woodsworth, M.L., Latimer, L.J.P. & Morgan, A.R. (1984). Poly(pyrimidine).Poly(purine) synthetic DNAs containing 5-Methylcytosine form stable triplexes at neutral pH. *Nucleic Acids Res.* **12**, 6603-6614.

Lerman, L.S. (1961). Structural considerations in interaction of DNA and acridines. *J. Mol. Biol.* **3**, 18-&.

Li, E., Bestor, T.H. & Jaenisch, R. (1992). Target Mutation of the DNA methyltransferase Gene Results in Embryonic Lethality. *Cell* **69**, 915-920.

Li, T., Stark, M.R., Johnson, A.D. & Wolberger, C. (1995). Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA. *Science* **270**, 262-269.

Malone, T., Blumenthal, R.M., & Cheng, X. (1995). Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* **253**, 618-632.

Merkiene, E., Vilkaitis, G. & Klimasauskas, S. (1998). A Pair of Single-Strand and Double-Strand Cytosine-N4 Methyltransferase from *Bacillus centrosporus*. *Biol. Chem.* **379**, 569-571.

Moser, H. & Dervan, P.B. (1987). Sequence-specific Cleavage of Double Helical DNA by Triple Helix Formation. *Science* **238**, 645-650.

Mrksich, M., Parks, M.E. & Dervan, P.B. (1994). Hairpin peptide motif – a new class of oligopeptides for sequence-specific recognition in the minor-groove of double-helical DNA. *J. Am. Chem. Soc.* **116**, 7983-7988.

Nicholls, A., Sharp, K. & Honig, B. (1991). Protein Folding and Association: Insights from the Interfacial and Thermodynamic Properties of Hydrocarbons. *Proteins* **11**, 281-296.

Nielsen, P.E. & Egholm, M. (1999). An Introduction to Peptide Nucleic Acid. *Current Issues Mol. Biol.* **1**, 89-104.

Noji, H., Yasuda, R., Yoshida, M. & Kinosita, K. (1997). Direct observation of the rotation of F1 -ATPase. *Nature* **386**, 299-302.

Oberle, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M.F. & Mandel, J.L. (1991). Instability of a 550-base pair segment and abnormal methylation in Fragile X Syndrome. *Science* **252**, 1097-1102.

Osterman, D.G., Depillis, G.D., Wu, J.C., Matsuda, A. & Santi, D.V. (1988). 5-Fluorocytosine in DNA is a mechanism-based inhibitor of Hhai methylase. *Biochemistry* **27**, 5204-5210.

Okano, M., Bell, D.W., Haber, D.A. & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257.

Parel, S.P. & Leumann, C.J. (2001). Triple-helix formation in the antiparallel binding motif of oligodeoxynucleotides containing N^9- and N^7-2-aminopurine deoxynucleosides. *Nucleic Acids Res.* **29**, 2260-2267.

Patschornik, A., Amit, B. & Woodward, R.B. (1970). Photosensitive protecting groups. *J. Am. Chem. Soc.* **92**, 6333-6335.

Pavletich, N.P. & Pabo, C.O. (1991). Zinc Finger-DNA Recognition: Crystal structure of a Zif268-DNA Complex at 2.1 Å. *Science* **252**, 809-817.

Pelton, J.G. & Wemmer, D.E. (1989). Structural characterization of a 2-1 distamycin A-d(CGCAAATTTGGC) complex by two-dimensional NMR. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5723-5727.

Perrin, D.D. & Armarego, L.F. (1988). Purification of Laboratory Chemicals. *Pergamon Press Ltd.*, Oxford.

Pignot, M., Siethoff, C., Linscheid, M. & Weinhold, E. (1998). Coupling of a nucleoside with DNA by a methyltransferase. *Angew. Chem.* **110**, 3050-3053; *Angew. Chem. Intl. Ed. Engl.* **37**, 2888-2891.

Pignot, M. (1999). Cofaktoranaloga-Synthesen des S-Adenosyl-L-methionins und deren Einsatz für Struktur-Reaktivitäts-Untersuchungen an DNA-Methyltransferasen. *Dissertation*.

Pignot, M., Pljevaljcic, G. & Weinhold, E. (2000). Efficient Synthesis of S-Adenosyl-L-Homocysteine Natural Product Analogues and Their Use to Elucidate the Structural Determinant for Cofactor Binding of the DNA Methyltransferase M·HhaI. *Eur. J. Org. Chem.*, 549-555.

Pljevaljčić, G. (1999). Synthese eines fluoreszierenden neuen Cofactors für Methyltransferasen und seine enzymkatalysierte, sequenzspezifische Kupplung mit DNA zur Fluoreszenzmarkierung. *Diploma thesis*.

Posfai, J., Bhagwat, A.S., Posfai, G. & Roberts, R.J. (1989). Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.* **17**, 2421-2435.

Povilionis, P.I., Lubys, A.A. & Janulaitis, A.A. (1988). Cloning of the restriction-modification genes of Bacillus centrosporus in Escherichia coli. *Genetika* **24**, 210-215.

Povsic, T.J. & Dervan, P.B. (1990). Sequence-specific alkylation of double-helical DNA by oligonucleotide-directed triple-helix formation. *J. Am. Chem. Soc.* **112**, 9428-9430.

Povsic, T.J., Strobel, S.A. & Dervan P.B. (1992). Sequence-specific double-strand alkylation and cleavage of DNA mediated by triple-helix formation. *J. Am. Chem. Soc.* **114**, 5934-5941.

Reinisch, K.M., Chen, L., Verdine, G.L. & Lipscomb, W.N. (1995). The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell* **82**, 143-153.

Roberts, R.J., Myers, P.A., Morrison, A. & Murray, K. (1976). A specific endonuclease from Haemophilus haemolyticus. *J. Mol. Biol.* **103**, 199-208.

Scavetta, R.D., Thomas, C.B., Walsh, M.A., Szegedi, S., Joachimiak, A., Gumport, R.I. & Churchill, M.E.A. (2000). Structure of *Rsr*I methyltransferase, a member of the *N*6-adenine β class of DNA methyltransferases. *Nucleic Acids Res.* **28**, 3950-3961.

Schägger, H. & von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal.Biochem.* **166**, 368-379.

Schlichting, I. & Goody, R.S. (1997). Triggering methods in crystallographic enzyme kinetics. Macromolecular Crystallography, PT B, *Methods Enzymol.* **277**, 467-490.

Schluckebier, G., O’Gara, M., Saenger, W. & Cheng, X. (1995). Universal Catalytic Domain Structure of AdoMet-dependent Methyltransferases. *J. Mol. Biol.* **247**, 16-20.

Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F. & Yonath, A. (2000). Structure of functionally activated small ribosomal subunit at 3.3 angstrom resolution. *Cell* **102**, 615-623.

Seelig, B. & Jäschke, A. (1998). A small catalytic RNA motif with Diels-Alderase activity. *Chem. Biol.* **6**, 167-176.

Serva, S., Weinhold, E., Roberts, R.J. & Klimasauskas, S. (1998). Chemical display of thymine residues flipped out by DNA methyltransferases. *Nucleic Acids Res.* **26**, 3473-3479.

Shimizu, M., Konishi, A., Shimada, Y., Inoue, H. & Ohtsuka, E. (1992). Oligo(2'-O-methyl)ribonucleotides – effective probes for duplex DNA. *FEBS Lett.* **302**, 155-158.

Slatko, B.E., Benner, J.S., Jager-Quinton, T., Moran, L.S., Simcox, T.G., Van Cott, E.M. & Wilson, G.G. (1987). Cloning, sequencing and expression of the TaqI restriction-modification system. *Nucleic Acids Res.* **15**, 9781-9796.

Smit, W.A., Krimer, M.Z. & Vorob'eva, E.A. (1975). Generation and chemical reactions of episulfonium ions. *Tetrahedron Lett.* **29**, 2451-2454.

Szewczyk, J.W., Baird, E.E. & Dervan, P.B. (1996a). Sequence-specific recognition of DNA by a major and minor groove binding ligand. *Angew. Chem. Intl. Ed. Engl.* **35**, 1487-1489.

Szewczyk, J.W., Baird, E.E. & Dervan, P.B. (1996b). Cooperative triple-helix formation via a minor groove dimerization domain. *J. Am. Chem. Soc.* **118**, 6778-6779.

Tanner, M.E. & Miao, S. (1994). The Synthesis and Stability of Aziridino-Glutamate, an Irreversible Inhibitor of Glutamate Racemase. *Tetrahedron Lett.* **35**, 4073-4076.

Tao, Z.-F., Fujiwara, T., Saito, I. & Sugiyama, H. (1999). Rational design of sequence-specific DNA alkylating agents based on duocarmycin A and pyrrole-imidazole hairpin polyamides. *J. Am. Chem. Soc.* **121**, 4961-4967.

Tomasz, M. (1995). Mitomycin C: small, fast and deadly (but very selective). *Chem. Biol.* **2**, 575-579.

Townsend, L.B. (1988). Chemistry of Nucleosides and Nucleotides. *Plenum Press*, New York, 249-250.

Tran, P.H., Korszun, R.Z., Cerritelli, S., Springhorn, S.S. & Lacks, S.A. (1998). Crystal structure of the DpnM DNA adenine methyltransferase from the *Dpn*II restriction system of *Streptococcus pneumoniae* bound to S-adenosylmethionine. *Current Biol.* **6**, 1563-1575.

Trauger, J.W., Baird, E.E., Mrksich, M. & Dervan, P.B. (1996). Extension of sequence-specific recognition in the minor groove of DNA by pyrrole-imidazole polyamides to 9-13 base pairs. *J. Am. Chem. Soc.* **118**, 6160-6166.

Tycko, B. (1997). DNA methylation in genomic imprinting. *Mutation Research* **386**, 131-140.

Vassylyev, D.G., Kashiwagi, T., Mikami, Y., Ariyoshi, M., Iwai, S., Ohtsuka, E. & Morikawa, K. (1995). Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: Structural basis for damaged DNA recognition. *Cell* **83**, 773-782.

Vidgren, J., Svensson, L.A. & Liljas, A. (1994). Crystal structure of catechol O-methyltransferase. *Nature* **368**, 354-358.

Wan, M., Lee, S.J.L., Zhang, X., Houwink-Manville, I., Song, H.-R., Amir, R.E., Budden, S., Naidu, S., Pereira, J.L.P., Lo, I.F.M., Zoghbi, H.Y., Schanen, N.C. & Francke, U. (1999). Rett syndrome and beyond: recurrent spontaneous familial MECP2 mutations at CpG hotspots. *Am. J. Hum. Genet.* **65**, 1520-1529.

Wang, C.C.C., Ellervik, U. & Dervan, P.B. (2001). Expanding the Recognition of the Minor Groove of DNA by Incorporation of β-Alanine in Hairpin Polyamides. *Bioorg. Med. Chem.* **9**, 653-657.

Watson, J.D. & Crick, F.H.C. (1953). Molecular structure of nucleic acids. *Nature* **171**, 737-745.

Weiss, M.A., Ellenberger, T., Wobbe, C.R., Lee, J.P., Harrison, S.C. & Struhl, K. (1990). Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature* **347**, 575-578.

White, S., Szewczyk, J.W., Turner, J.M., Baird, E.E. & Dervan, P.B. (1998). Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* **391**, 468-471.

Wilson, D.S., Guenther, B., Desplan, C. & Kuriyan, J. (1995). High resolution crystal structure of paired (Pax) class cooperative homeodomain dimmer on DNA. *Cell* **82**, 709-719.

Wölcke, J. (1998). Der kinetische Mechanismus der DNA-Methyltransferase aus Thermus aquaticus und Selektion eines DNA-bindenden Peptids mittels Phage Display. *Dissertation*.

Wu, J.C. & Santi, D.V. (1987). Kinetic and katalytic mechanism of *Hha*I methyltransferase. *J. Biol. Chem.* **263**, 319-326.

Wu, J.C. & Santi, D.V. (1988). High level expression and purification of HhaI methyltransferase. *Nucleic Acids Res.* **16**, 703-717.

Wurtz, N.R. & Dervan, P.B. (2000). Sequence specific alkylation of DNA by hairpin pyrrole-imidazole polyamide conjugates. *Chem. Biol.* **7**, 153-161.

Xu, G.-L., Bestor, T.H., Bourc'his, D., Hsieh, C.-L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J. & Viegas-Péquignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**, 187-191.

APPENDIX

Appendix A

DNA Methyltransferase from *Thermus aquaticus* (M·*Taq*I)

Type	N6-adenine
Recognition sequence	5’-TCG**A**-3’
Length	421 amino acids
Molecular weight	47,848 Da
Theoretical pI	8.98

Sequence

1	**11**	**21**	**31**	**41**
MGLPPLLSLP	SNSAPRSLGR	VETPPEVVDF	MVSLAEAPRG	GRVLEPACAH
51	**61**	**71**	**81**	**91**
GPFLRAFREA	HGTGYRFVGV	EIDPKALDLP	PWAEGILADF	LLWEPGEAFD
101	**111**	**121**	**131**	**141**
LILGNPPYGI	VGEASKYPIH	VFKAVKDLYK	KAFSTWKGKY	NLYGAFLEKA
151	**161**	**171**	**181**	**191**
VRLLKPGGVL	VFVVPATWLV	LEDFALLREF	LAREGKTSVY	YLGEVFPQKK
201	**211**	**221**	**231**	**241**
VSAVVIRFQK	SGKGLSLWDT	QESESGFTPI	LWAEYPHWEG	EIIRFETEET
251	**261**	**271**	**281**	**291**
RKLEISGMPL	GDLFHIRFAA	RSPEFKKHPA	VRKEPGPGLV	PVLTGRNLKP
301	**311**	**321**	**331**	**341**
GWVDYEKNHS	GLWMPKERAK	ELRDFYATPH	LVVAHTKGTR	VVAAWDERAY
351	**361**	**371**	**381**	**391**
PWREEFHLLP	KEGVRLDPSS	LVQWLNSEAM	QKHVRTLYRD	FVPHLTLRML
401	**411**	**421**		
ERLPVRREYG	FHTSPESARN	F		

Appendix B

DNA Methyltransferase from *Haemophilus haemolyticus* (M·*Hha*I)

Type	C5-cytosine
Recognition sequence	5'-G**C**GC-3'
Length	327 amino acids
Molecular weight	36,996 Da
Theoretical pI	7.58

Sequence

1	**11**	**21**	**31**	**41**
MIEIKDKQLT	GLRFIDLFAG	LGGFRLALES	CGAECVYSNE	WDKYAQEVYE
51	**61**	**71**	**81**	**91**
MNFGEKPEGD	ITQVNEKTIP	DHDILCAGFP	CQAFSISGKQ	KGFEDSRGTL
101	**111**	**121**	**131**	**141**
FFDIARIVRE	KKPKVVFMEN	VKNFASHDNG	NTLEVVKNTM	NELDYSFHAK
151	**161**	**171**	**181**	**191**
VLNALDYGIP	QKRERIYMIC	FRNDLNIQNF	QFPKPFELNT	FVKDLLLPDS
201	**211**	**221**	**231**	**241**
EVEHLVIDRK	DLVMTNQEIE	QTTPKTVRLG	IVGKGGQGER	IYSTRGIAIT
251	**261**	**271**	**281**	**291**
LSAYGGGIFA	KTGGYLVNGK	TRKLHPRECA	RVMGYPDSYK	VHPSTSQAYK
301	**311**	**321**		
QFGNSVVINV	LQYIAYNIGS	SLNFKPY		

Appendix C

DNA Methyltransferase from *Bacillus centrosporus* (M·*Bcn*IB)

Type	N4-cytosine
Recognition sequence	5'-C**C**(C/G)GG-3'
Length	321 amino acids
Molecular weight	37,222 Da
Theoretical pI	7.8

Sequence

1	**11**	**21**	**31**	**41**
MINLNKQLEQ	VIQQDIVENT	DCREILRKIP	KESIDLLVTS	PPYWAKRVYN
51	**61**	**71**	**81**	**91**
EDEEGELGSE	PTPEEFVKVL	ADYFDIFRPY	IKKEGNLFVN	IGDTFFGSGA
101	**111**	**121**	**131**	**141**
GAWNKYLDEE	GNTTSFQKER	KEKYFTLKPL	QPKLKQDGKL	YQNKQLLLIP
151	**161**	**171**	**181**	**191**
SRFAIEMQSR	GWILRDDIIW	RKPNRIPASV	KDRFNNTYEH	VFHFVQAKKY
201	**211**	**221**	**231**	**241**
YFDLDSVKIM	GANGKLKNPG	DVWDINTQPL	RGSHTATFPE	ALVDICIKCG
251	**261**	**271**	**281**	**291**
SPKNGLVFDP	FMGTGTSWIV	ARRLNRRFIG	SEINPEFYQF	SIKRFLESLS
301	**311**	**321**		
ERNKEHGETI	PSNTQTHLDP	G		

CURRICULUM VITAE

PERSONAL DETAILS

Name:	**Goran Pljevaljčić**
Birth date:	**April 2, 1974**
Place of birth:	**Dortmund, Germany**
Citizenship:	**Germany**
Work address:	**Otto-Hahn-Str. 11**
	D-44227 Dortmund
	Department of Physical Biochemistry
	Max-Planck-Institute of Molecular Physiology
E-mail:	**goran.plj@mpi-dortmund.mpg.de**
Telephone:	**+49-231-133 2306 (work); +49-231-133 2398 (fax)**

EDUCATION

1999 - 2002:	PhD at the Max-Planck-Institute of Molecular Physiology, Department of Prof. Dr. R. S. Goody, PhD
1997 - 1999:	University of Dortmund, Germany, Diplom (January 1999) Max-Planck-Institute of Molecular Physiology, Department of Prof. Dr. R. S. Goody
1996 - 1997:	King's College London, University of London. Supervisor: Prof. Dr. Keith Jones
1993 - 1996:	University of Dortmund, Germany, Vordiplom (1995).
1988 - 1993:	Helmoltz-Gymnasium, Dortmund, Germany. Finished with Abitur with the main subjects Mathematics, Chemistry, History and German.
1984 - 1988:	Gertrud-Baeumler-Realschule, Dortmund, Germany.
1980 - 1984:	Elizabeth Grundschule, Dortmund, Germany.

AWARDS AND GRANTS

- Award for Excellent Poster Presentation, Gordon Research Conference for Bioorganic Chemistry, July 2001.
- Travel grant, SmithKline Beecham Foundation, April 2001.
- EU-Grant (ERASMUS scheme) for abroad studies, 1996/97.
- Book prize, 3rd year organic chemistry, University of Dortmund, July 1996.
- Book prize, Fonds der Chemischen Industrie, May 1993.

PUBLICATIONS

Invited Minireview/Concept article from ChemBioChem on the topic of "SMILing: Sequence-Specific Methyltransferase-Induced Labeling of DNA".

Publications on Sequence-Specific Methyltransferase-Induced Labeling of DNA (SMILing DNA), in preparation.

Stephen T. Hilton, Tim C. T. Ho, Goran Pljevaljcic, Marcus Schulte and Keith Jones (2001). A tandem radical approach to the ABCE-rings of the *Aspidosperma* and *Strychnos* alkaloids. *Chem. Comm.*, 209-210.

M. Pignot, G. Pljevaljcic, E. Weinhold (2000). Efficient synthesis of S-adenosyl-L-homocysteine natural product analogues and their use to elucidate the structural determinant for cofactor binding of the DNA methyltransferase M.HhaI. *Eur. J. Org. Chem*, 549-555.

Stephen T. Hilton, Tim C. T. Ho, Goran Pljevaljcic, and Keith Jones (2000). A New Route to Spirooxindoles. *Org. Lett.*, **2**, 17, 2639-2641.

E. Weinhold, M. Pignot, G. Pljevaljcic (1999). Sequence-specific Labeling of DNA Using New Cofactors for DNA Methyltransferases. *Biol. Chem.*, **380**, S207. (ABSTRACT)

CONTRIBUTIONS TO CONFERENCES AND MEETINGS

- Poster presentation at the Gordon Conference of Bioorganic chemistry (2001), Andover NH, USA.
- Invited oral presentation at Kingston University (2001), Kingston, UK.
- Poster presentation at the german organic chemistry conference, Orchem (2000), Bad Nauheim, Germany.
- Oral presentation at the EU-network meeting on methyltransferases and restriction endonucleases (2000), Siena, Italy.
- Oral presentation at the EU-network meeting on methyltransferases and restriction endonucleases (1999), Iraklion (Crete), Greece.
- Poster presentation at the conference of the german chemical society GdCH on life science (1999), Berlin, Germany.

Acknowledgements

Without the encouragement and participation of many people, this project would never have been accomplished. My very deep appreciation goes to each of the persons mentioned below.

I would like to thank Prof. Dr. Elmar Weinhold for his supervision, advice and continous support. Thanks are also due to my PhD supervisor Prof. Dr. Roger S. Goody for the possibility to work in his department, for his advice and support. I thank Prof. Dr. Wolfgang Kreiser for the backup supervision, his support and continous interest in my work.

I am also very grateful to Dr. Axel Scheidig who introduced me to the secrets of protein crystallography and for enormous help with the crystallographic part of the thesis.
Thanks are due to Christine Beuck who was involved in the search for suitable protecting groups in the synthesis of the novel aziridine cofactors and Alexander Peschlow who worked on the optimization of the deprotection step in the synthesis.
I would also like to thank Dr. Andres Jäschke and Jörg Schlatterer for a fruitful collaboration on the development of a selection system for the isolation of deoxyribozymes.
I am indebted to Prof. Dr. Keith Jones for proofreading the manuscript and for the experience I gained in his lab during my undergraduate research project.

I also have to thank Nathalie Bleimling for many preparations of cofactor-free M·*Taq*I and Egle Merkiene for providing M·*Hha*I and M·*Bcn*IB. Bernhard Griewel for recording the NMR spectra. Dr. Petra Janning for useful hints with MALDI spectrometry.
I thank all in department 3 and the scientific service facilities of the institute for a lot of help and support.

My collegues and friends Christian Becker, Gereon Franzen, Andrea Beste, Almut Friederich, Karsten Goedecke, Lilianna Wilitzek and Nathalie Bleimling are responsible for the brilliant atmosphere in the lab and the joy I had during my PhD project. Lilianna Wilitzek, Silke Huber, Tresfore Dambe, Paul Rothwell and Janina Cramer were responsible for the brilliant atmosphere in the office.

Special thanks go to my family and Carmen for their great support.