Sequence-specific covalent labelling of DNA

Anna Gottfried and Elmar Weinhold
Institute of Organic Chemistry, RWTH Aachen University, Landoltweg 1, 52056 Aachen, Germany

Abstract

Sequence-specific DNA modification is of significance for applications in bio- and nano-technology, medical diagnostics and fundamental life sciences research. Preferentially, labelling should be performed covalently, which avoids doubts about label dissociation from the DNA under various conditions. Several methods to label native DNA have been developed in the last two decades. Triple-helix-forming oligodeoxynucleotides and hairpin polyamides that bind DNA sequences specifically in the major and minor groove respectively were used as targeting devices for subsequent covalent labelling. In addition, enzyme-directed labelling approaches utilizing nicking endonucleases in combination with DNA polymerases or DNA methyltransferases have been employed. This review summarizes various techniques useful for functionalization of long native DNA.

Introduction

DNA labelling is of major importance for structural and functional investigations of DNA-based processes. Studies of DNA, DNA-binding proteins and DNA-modifying enzymes typically require making the DNA visible or manipulable by introducing labels. In medical diagnostics, DNA often needs to be conjugated with fluorescent dyes or affinity tags and constructing nanostructures with DNA can be facilitated by attaching unique functional groups.

A myriad of methods have been developed to covalently modify DNA either by chemical or enzymatic methods [1]. They involve de novo DNA synthesis or modification of natural DNA. Short labelled ODNs (oligodeoxynucleotides) are available by solid-phase DNA synthesis and the incorporation of modified phosphoramidite building blocks [2]. Extension of the labelled ODNs with DNA polymerases, e.g. by PCR, yields long DNA labelled at the 5′-ends. Alternatively, modified dNTPs can be incorporated by DNA polymerases or deoxynucleotidyltransferases for internal or 3′-end-labelling respectively. The modification may include the label or consist of a chemical group suitable for chemoselective labelling in a second step [3].

Many chemical modification reactions of native nucleic acids are known and include bisulfite-catalysed transamination of the amino group of cytosine residues, bromination of the C5 position of pyrimidines and the C8 position of purines followed by reaction with derivatized amines or photochemical coupling of azides such as photobiotin [1]. In addition, Heetebrij et al. [4] developed platinum(II)-based reagents to create stable adducts with the N7 position of guanine and, to a lesser extent, adenine bases. Although these methods have general applicability they cannot be directed to specific internal sequences. Thus sequence-specific covalent labelling of long native DNA represents a challenging task and suitable targeting approaches using synthetic molecules or enzymes will be discussed here.

Triple-helix formation by ODNs

Synthetic ODNs not only hybridize to complementary ssDNA (single-stranded DNA) according to the Watson–Crick base pairing rules but also bind to dsDNA (double-stranded DNA) forming triple-helix structures [5]. These TFO (triple-helix-forming ODN) bind preferentially to homopurine-homopyrimidine sequences, wind around the DNA in the major groove and form Hoogsteen-type hydrogen bonds with the Watson–Crick base pairs (Figure 1A). Escudé et al. [6] stabilized TFO binding by covalently linking the overhanging ends with T4 DNA ligase in the presence of a splint ODN forming catenanes of ssDNA and double-stranded plasmid DNA. These so-called ‘padlock’ ODNs offer a way to label native DNA with functional molecules (Figure 1B). For example, plasmid DNA was conjugated with a biotinylated padlock ODN for complex formation with streptavidin [7]. The same group also published a technique with stem–loop TFOs to attach fluorophores or a peptide to plasmid DNA. While specific binding to the target sequence occurs by triple-helix formation, the 5′- and 3′-ends of the TFO are designed to hybridize and produce a single-stranded overhang that is ligated to the sticky end of a labelled DNA duplex or a hairpin ODN [8,9].

With UV-reactive TFOs the nucleic acid target can be modified irreversibly. Pfannschmidt et al. [10] conjugated a TFO with a psoralen moiety that leads to photo-cross-linking with nearby thymine residues. This covalent adduct is stable.
Figure 1  | Schematic representation of triple-helix complexes

(A) A TFO (black) binds to a target sequence in the major groove of dsDNA. (B) Different strategies to stabilize triple-helix complexes by covalent bond formation. The overhanging ends of a TFO can be joined by T4 DNA ligase in the presence of a complementary splint ODN, leading to stable attachment of a label (indicated as star). Alternatively, complementary overhanging ends can form a stem-loop for ligation to the sticky end of a labelled hairpin ODN. Furthermore, the TFO may contain a photo-reactive group (black oval) leading to covalent bond formation with the DNA after UV irradiation.

under triple-helix-dissociating conditions. Perrouault et al. [11] introduced sequence-specific breaks after forming the triple helix with a TFO containing a photo-reactive ellipticine derivative. Photo-induced cleavage of the duplex occurred, but photo-cross-linking of the TFO and target DNA was also observed to some extent. Additional examples of cross-linking and cleaving DNA can be found in [5].

**Hairpin polyamides**

Several natural products (e.g. netropsin and distamycin) and synthetic ligands (Hoechst 33258) can bind specifically in the minor groove of A/T-rich DNA sequences. Inspired by these minor groove binders, the Dervan group developed hairpin polyamides (Figure 2) which sequence-specifically recognize DNA in the minor groove [12]. These synthetic cell-permeable polyamides typically contain two segments of aromatic heterocycles which are connected head-to-tail by a $\gamma$-aminobutyric acid linker [13]. This arrangement allows folding back of the polyamides and pairing of two aromatic heterocycles in an antiparallel fashion to discriminate one Watson–Crick base pair from the other three base pair combinations. A pair of two $N$-methylpyrroles (Py/Py) recognizes A/T over C/G base pairs, a pair between $N$-methylimidazole and $N$-methylpyrrole (Im/Py) distinguishes G/C from C/G, and a pair of 3-hydroxy-1-methylpyrrole and $N$-methylpyrrole (Hp/Py) discriminates T/A over A/T [14,15]. Binding of hairpin polyamides to DNA leads to widening of the minor groove along with a bend in the helix towards the major groove [16]. This allosteric perturbation is expected to be the reason for inhibition of transcription factor binding.

Hairpin polyamides typically recognize 3–5 bp in DNA with dissociation constants in the nanomolar range and high sequence specificity [14]. Based on the recognition rules mentioned above they represent programmable targeting devices and were used for sequence-specific fluorescence detection of DNA. Conjugation of TO (Thiazole Orange) to the C-terminal end (Figure 2A) or attachment of TMR (tetramethylrhodamine) to an internal pyrrole position caused strongly increased fluorescence signals in the presence of matched DNA [17,18]. In the case of TO, this effect was attributed to directed intercalation of the fluorophore, while for TMR, a strong quenching was observed for the free conjugate that is abrogated upon DNA binding. Binding specificity can be enhanced by linking two hairpin polyamides. These tandem hairpin polyamides were tagged with a fluorophore at the C-terminus and used to visualize insect and vertebrate telomeric repeats in fixed cells. From the fluorescence intensity of the spots it was also possible to estimate the relative telomere lengths [19].

In addition to non-covalent labelling, polyamides can be used to direct covalent bond formation. DNA interstrand cross-linking agents are of considerable interest in molecular biology and human medicine because they effectively block both DNA replication and gene expression. Sugiyama and
**Figure 2** | Hairpin polyamides

(A) Structure of the heterocyclic units N-methylpyrrole (Py), N-methylimidazole (Im) and 3-hydroxy-1-methylpyrrole (Hp), which are connected by peptide bonds to form synthetic hairpin polyamides of different sequence and length. A fluorophore, peptide or another chemical moiety (illustrated as shadowed circle) can be attached to the C-terminal end (shown) or to a heterocyclus of the molecule. (B) Model of hairpin polyamide binding to the minor groove of DNA. The aromatic heterocycles are indicated as black ovals.

---

Co-workers used polyamides to target DNA cross-linking to specific sites [20]. They synthesized conjugates of polyamides and CPI (cyclopropapyrroloindole), the reactive segment of the antitumour antibiotic duocarmycin A that alkylates DNA at A/T base pairs. Two ImPy-CPI conjugates were fused via their N-terminal ends (tail-to-tail) and in the presence of the triamide ImImPy the double-headed alkylation agent (ImPy-CPI)$_2$ produced interstrand cross-links. In contrast with the hairpin polyamides described above that form 1:1 complexes, DNA cross-linking involves a 1:2:1 complex between (ImPy-CPI)$_2$, ImImPy and the symmetrical DNA sequence. Using a 400 bp DNA fragment an interstrand cross-linking yield of approx. 40% was achieved [21].

**NEases (nicking endonucleases)**

In addition to synthetic targeting devices, DNA-modifying enzymes can be used to recognize and covalently manipulate DNA. A two-step strategy for covalent labelling of DNA exploiting NEases was developed [22,23]. These enzymes recognize specific DNA sequences and introduce strand breaks (nicks) by cleaving only one phosphodiester backbone of dsDNA. The NEase Nb.BbvCI, for example, recognizes the 7 bp 5’-GCTGAGG-3’ sequence and cleaves after C in the upper strand. Afterwards, DNA polymerases are used for template-directed incorporation of a fluorescent nucleotide terminator or for nick translation with fluorophore-labelled deoxyribonucleotides. Pre-existing nicks that would lead to non-specific labelling can be repaired or blocked by using T4 DNA ligase and DNA polymerase-mediated incorporation of deoxyribonucleotides before labelling. DNA genotyping can be performed by reading the fluorescence barcode after linear stretching of the dye-labelled DNA and fluorescence microscopy of single DNA molecules.

Another approach employs two NEases cleaving on the same strand and producing two nicks at a distance of 15–25 nt [24]. A probe ODN is inserted by strand displacement hybridization and covalently attached by enzymatic ligation. The ODN was not fluorescence labelled but contained a 3’ target-independent overhang that served as a template for detection by rolling circle amplification.

**DNA MTases (DNA methyltransferases)**

We developed methods for sequence-specific covalent labelling of DNA using DNA MTases [25]. This class of enzymes naturally transfers the activated methyl group from the ubiquitous cofactor AdoMet (S-adenosylmethionine)
to the exocyclic amino groups of adenine and cytosine (DNA adenine-N6 and DNA cytosine-N4 MTases) or carbon at the 5-position of cytosine (DNA cytosine-C5 MTases) within specific recognition sequences [26]. More than 8000 DNA MTases with hundreds of distinct recognition sequences ranging from 2 to 8 bp are currently listed in REBASE, a database of restriction endonucleases and DNA MTases [27]. By replacing the methionine side chain of AdoMet with an aziridine ring, we obtained a new cofactor for DNA MTases (Figure 3, left). The enzymes are tricked into catalysing a nucleophilic aziridine ring opening reaction instead of a methyl group transfer, and couple the whole aziridine cofactor to their target nucleobase in DNA [28]. By attaching chemical groups to different positions of the adenine ring the aziridine cofactors work in combination with DNA MTases as delivery systems for reporter groups such as fluorophores [29] or biotin [30,31]. Thus this one-step method was termed SMILing DNA (sequence-specific methyltransferase-induced labelling of DNA) [32] and used to study cell transfection with fluorescently labelled DNA [33] or to nanostructure DNA and position gold nanoparticles on DNA with biotinylated DNA [34,35].

Rajski and co-workers also synthesized aziridine cofactors or nitrogen mustards, which are presumably converted into aziridines in situ [36,37]. These cofactor analogues were equipped with functional groups for chemoselective labelling after DNA MTase-mediated coupling. They demonstrated subsequent conversion of an azide-modified duplex ODN by Staudinger ligation or an alkyne-modified duplex ODN by CuAAC (Cu-catalysed azide-alkyne 1,3-dipolar cycloaddition) click chemistry [36,37]. An azide-modified duplex ODN was appended with Cu-phenanthroline and used to induce DNA strand scissions mainly proximal to the DNA MTase recognition sequence [38].

However, an inherent feature of the aziridine cofactors is that covalent coupling of the two substrates leads to potent product inhibitors preventing further turnovers. Thus the DNA MTases have to be used in stoichiometric amounts with respect to target sites on DNA. Although prokaryotic DNA MTases can be easily obtained in milligram quantities, and microgram amounts of labelled DNA are generally sufficient for various applications, this feature prompted the development of more efficient AdoMet analogues.

Replacement of the methyl group in AdoMet by larger saturated carbon chains such as ethyl or propyl leads to a drastic decline of transfer rates by MTases [39]. We synthesized AdoMet analogues with extended side chains carrying an unsaturated bond in β-position to the sulfonium centre and tested them with various DNA MTases in collaboration with the Klimešauskas group [40,41]. They turned out to be much more reactive in DNA MTase-catalysed transalkylations than the analogues with extended saturated side chains, indicating that steric constraints in the active site of the enzymes are not of major importance. In fact,
the double or triple bonds in the allylic or propargylic side chains are likely to lower the transition state energy of the nucleophile substitution reaction at the neighbouring carbon atom by conjugative stabilization. Since the extended side chains are directly transferred to DNA, no problem with product release is observed with these cofactor analogues (Figure 3, right).

Further extension of the allylic or propargylic AdoMet analogues with suitable functional groups gives the possibility to sequence-specifically functionalize native DNA which can be further modified in chemoselective reactions. Using this approach, called mTAG (methyltransferase-directed transfer of activated groups), fluorescent reporter groups or biotin were attached to amino-functionalized plasmid DNA [42,43]. Such AdoMet analogues in combination with a large variety of DNA MTases are envisioned as versatile tools to modify DNA sequence specifically with labels of wide diversity. They are also substrates for small-molecule MTases [44,45] and recently we demonstrated that an AdoMet analogue with a terminal alkyne group is not only a cofactor for DNA MTases but also for RNA and protein MTases [46,47]. This opened the way to label RNA and proteins sequence-specifically by CuAAC click chemistry.

Conclusions and outlook
Various methods are available to gain access to modified synthetic ODN and DNA but sequence-specific labelling of native DNA represents a more challenging task. The few convenient methods to covalently functionalize natural DNA, such as genomic nucleic acid and plasmids, at defined sequences are discussed in this review. The main problem is to direct the labelling process to specific targets in a sea of putative binding sites. Small synthetic molecules that bind sequence specifically in the major or minor groove of DNA, including TFO and hairpin polyamides as well as DNA-modifying enzymes such as NEases or DNA MTases, can be used as targeting devices. DNA MTases offer the advantage that a large number of enzymes with a wide variety of recognition sequences ranging from 2 to 8 bp are found in Nature. In addition, it is possible to direct DNA MTases to a subset of their natural recognition sequence by fusing the enzymes with zinc-finger proteins [48,49] or a TFO [50], giving the opportunity to label DNA at very few sites. The modified AdoMet cofactors in combination with DNA MTases serve as a new labelling technology with potential applications in DNA genotyping, DNA methylation detection and (nano)biotechnology.

Acknowledgements
We thank our collaboration partners Professor Saulius Klimašauskas, Dr Marie-Paule Teulade-Fichou, Professor Norbert Reich, Professor Mark Helm and Professor Bernhard Lüscher for very fruitful cooperation regarding the modified AdoMet cofactors.

Funding
This work was supported by a grant from the German Research Foundation [grant number WE 1453/4-1 and 4-2].

References

©The Authors Journal compilation ©2011 Biochemical Society