Published in final edited form as:

Nanoscale. 2011 August ; 3(8): 3027–3039. doi:10.1039/c1nr10280j.

Labeling DNA for Single-Molecule Experiments: Methods of Labeling Internal Specific Sequences on Double-Stranded DNA

Hagar Zohar and Susan J. Muller*

Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA 94720, U.S.A.

Abstract

This review is a practical guide for experimentalists interested in specifically labeling internal sequences on double-stranded (ds) DNA molecules for single-molecule experiments. We describe six labeling approaches demonstrated in a single-molecule context and discuss the merits and drawbacks of each approach with particular attention to the amount of specialized training and reagents required. By evaluating each approach according to criteria relevant to single-molecule experiments, including labeling yield and compatibility with cofactors such as Mg^{2+} , we provide a simple reference for selecting a labeling method for given experimental constraints. Intended for non-specialists seeking accessible solutions to DNA labeling challenges, the approaches outlined emphasize simplicity, robustness, suitability for use by non-biologists, and utility in diverse single-molecule experiments.

1. Introduction

The burgeoning field of single-molecule biophysics has produced new techniques that characterize individual biological molecules, allowing for observations unattainable with conventional bulk methods.¹ A common challenge for single-molecule experiments is that each experimental setup must reconcile several constraints, including requirements of (i) the molecule(s) of interest (e.g. DNA, protein) and (ii) the platform of study (e.g. AFM, optical tweezers, fluorescence microscopy).² Furthermore, many such experiments require the labeling of internal specific sequences on long (kb) double-stranded (ds)DNA.³⁻⁸ Since single-molecule experiments are diverse in their requirements, there is a need for multiple labeling methods. Beyond the scope of single-molecule experiments, DNA labeling is of general interest to biotechnology as a strategy to map DNA⁹⁻¹¹ or achieve molecular scaffolding.¹²

This review is intended as a practical guide for experimentalists interested in specifically labeling dsDNA molecules for single-molecule experiments. We focus on six approaches demonstrated in a single-molecule context and describe relevant features of the molecular tools they utilize. The merits and drawbacks of each labeling method are considered with particular attention to the amount of specialized training and reagents required. By presenting a set of criteria relevant to single-molecule experiments (e.g. labeling yield, compatibility with cofactors such as Mg^{2+}) we provide a simple reference for selecting an approach for given experimental constraints. As DNA labeling is of great interest to many scientific disciplines, we note that several excellent reviews are devoted to this topic.¹³⁻¹⁵ This review, however, is intended for non-specialists interested in accessible solutions to

*corresponding author muller2@berkeley.edu .

DNA labeling challenges and the approaches outlined emphasize simplicity, robustness, suitability for use by non-biologists, and utility in diverse single-molecule experiments.

2. Features of the molecular “toolkit”

Sequence-specific DNA labeling methods utilize the properties of several small molecules and enzymes that, together, make up a molecular “toolkit.” In this context, labeling DNA refers to attaching small molecules or moieties to DNA that are not native to the DNA structure. Labels are often fluorophores^{4,16-20} or haptens^{3,21-24} that can create optical signals or attachment points, respectively. Fluorophores, such as the commonly used Cy dyes and Alexa Fluor dyes, allow for direct detection and localization via fluorescence microscopy. These and other fluorophores are commercially available (Table 1) and provide coverage over a wide range of excitation and emission spectra. Hapten labels can serve as specific attachment points for a larger protein-conjugated element, such as a microsphere, which would be challenging to attach directly to DNA.

Haptens are small molecules that form strong, non-covalent, bonds to a specific protein or antibody.²⁵ Commonly used haptens include biotin, digoxigenin, and fluorescein, which bind to avidin family, anti-digoxigenin, and anti-fluorescein proteins, respectively. Of the hapten-protein pairs, biotin-avidin is notable as the strongest, with nearly the strength of a covalent bond, and most widely used.²⁶ Since avidin proteins have four biotin binding sites, two or more biotinylated elements can be joined in the presence of a free avidin protein serving as a bridge. Hapten-conjugated nucleotides (e.g. biotin-dUTP, DIG-dUTP, fluorescein-dUTP) and protein-coated spheres (quantum dots (qdots), nano- and microspheres) are commercially available (Table 1).

There are several resources for detailed protocols that inform the use of the tools in the molecular “toolkit”. *Current Protocols in Molecular Biology* is a continually expanded lab manual with detailed methods and procedures. It includes fundamental techniques such as restriction enzyme digestions²⁷ and polymerase chain reaction (PCR).²⁸ *Cold Spring Harbor Protocols* is a journal of biological research protocols, including new and well-established techniques. SpringerProtocols (<http://www.springerprotocols.com>) is an electronic database that includes protocols from multiple journals and laboratory manuals. More advanced peer-reviewed procedures can be found in journals such as *Nature Protocols*. In the following section, we describe the relevant features of tools that are critical to the sequence-specific DNA labeling methods highlighted in this review.

2.1 Small molecules

Oligonucleotides (Oligos) and DNA fragments—An oligo is a short, typically less than 100 base, single-stranded sequence of DNA. Oligos of any desired sequence are commercially available (Table 1) and made to order via automated solid-phase synthesis.^{15,29} The synthesis can accommodate 5' and 3' modifications and alternative nucleotide bases. A DNA fragment is a portion of dsDNA that is hundreds or thousands of base pairs long. Fragments are typically created via restriction enzyme digest of a given dsDNA molecule and consist of the dsDNA between two cleavage sites or a cleavage site and the end of the DNA molecule. The length and sequence of the resulting fragments are constrained by the sequence of the complete DNA molecule and the locations of the restriction sites on the DNA molecule. Online tools like NEBcutter (<http://www.tools.neb.com>) can be used to identify appropriate enzyme-DNA combinations for generating fragments of a particular length from a known DNA sequence and commercially available enzymes (Table 1). Polymerase chain reaction (PCR) is a more flexible yet challenging method of creating fragments. A prescribed region of a known dsDNA sequence, up to several kb, can be selectively amplified and then isolated. A pair of

oligo primers, each roughly 20 bases, flanking the prescribed region must be designed according to well-known guidelines.³⁰ Tools like Invitrogen's OligoPerfect™ Designer and Vector NTI Software (<http://www.invitrogen.com>) aid oligo primer pair design.

Peptide Nucleic Acids (PNAs)—PNAs are synthetic molecules composed of a peptide backbone and nucleic acid bases, which confer upon them DNA sequence recognition.³¹⁻³⁴ They form strong and specific complexes with dsDNA, via four different binding modes, depending on their design.³⁵ PNAs bind optimally to target sequences roughly 8 to 10 bases in length via the strongest binding mode.^{3,31,35} Due to a neutrally-charged backbone, PNA binding to dsDNA deviates substantially from that of conventional oligos with respect to binding conditions and yield. The alternative backbone also makes PNAs nuclease and protease resistant.^{36,37} These differences are exploited in the many applications of PNAs, which are covered in a number of excellent general reviews.³⁸⁻⁴² PNAs are commercially available and made to order (Table 1). Design considerations for PNAs include the binding mode, the length of the targeted sequence, the incorporation of modifications, and the addition of positively charged features such as lysines to enhance binding rates. Vendors typically evaluate PNA design for issues affecting synthesis feasibility and yield.

2.2 Enzymes

Enzymes are proteins that catalyze chemical reactions. Some enzymes require cofactors such as Mg^{2+} and adenosine triphosphate (ATP) to function. Commercially available enzymes are provided with buffer solutions containing the requisite components for optimal performance. The enzymatic reactions of interest here include those that either create or break bonds along the backbone of DNA and those that modify bases on DNA. Below we focus on a very restricted set of enzymes and their properties that are useful for preparing constructs for single-molecule experiments. All of the enzymes described below are commercially available (Table 1) and are typically supplied with technical protocols for optimal performance.

DNA Ligase—DNA ligase repairs nicks, which are discontinuities in one of the phosphodiester backbones of the DNA double helix, by forming a phosphodiester bond between the 5' phosphate and 3' hydroxyl termini of a DNA strand.⁴³ The nick-repairing action of ligase is critical to covalently joining together dsDNA fragments or an oligo to dsDNA. Fragment ends are either blunt ends lacking unpaired overhanging bases, or sticky ends consisting of single-stranded overhangs of one or more bases available to bind with other complementary overhangs. In the latter case, an annealing or hybridization step must first join together the fragments with complementary overhangs, creating a construct with nicks. Ligase then repairs these nicks and covalently joins the fragments. T4 DNA ligase is notable for also joining fragments with blunt ends. Since ligase forms the phosphodiester bond between the 5' phosphate and 3' hydroxyl termini of DNA, the 5' end must be phosphorylated. If the 5' end is a free hydroxyl group, it can be phosphorylated using polynucleotide kinase (PNK) to allow for subsequent ligation. If the 5' end is phosphorylated, the 5' phosphate can be removed with a phosphatase to prevent ligation.

Restriction Enzymes—The most commonly used restriction enzymes are type II, which typically bind to and cleave at target sites on dsDNA that are 4-8 nucleotides in length. In bacteria, they are employed to cleave and destroy foreign DNA. Binding of the enzyme to DNA, under optimized conditions, is highly specific to target sites called cognate sequences. Upon binding to target sites, most type II restriction enzymes require a cofactor such as Mg^{2+} for cleavage.⁴⁴ Restriction enzymes can cleave the two strands of DNA at staggered positions, forming short sticky-ends or cleave both DNA strands at the same location forming blunt ends. The fragments of DNA resulting from cleavage can be covalently

reattached by the action of DNA ligase. Type II restriction enzymes are composed of two subunits, each of which contains an active site where cleavage of one strand of DNA occurs. In a few instances, the subunits are not identical and one subunit can be mutated and rendered inactive such that the action of the dimer forms a single-stranded break, or nick, without cleaving the dsDNA molecule.⁴⁵ In the absence of the cofactor, some enzymes will bind to the target site and not cleave. Over 600 type II restriction enzymes, with 200 distinct recognition sequences, are commercially available (Table 1).^{46,47}

Nicking Endonucleases (Nickases)—Nickases cleave one strand of dsDNA, creating a single-stranded break or nick. Nickases, unlike restriction enzymes that cleave both DNA strands, are uncommon in nature but can be engineered.^{45,48} One approach to the engineering process, mentioned above, involves starting with restriction enzymes with disparate catalytic hydrolysis sites and inactivating only one of the sites by mutation.⁴⁵ The nickase DNase I produces randomly generated nicks while other nickases form nicks at specific sequences. Over 10 nicking enzymes with specific and known recognition sites are commercially available (Table 1) and over 200 have been characterized.^{46,47}

DNA Polymerases—DNA Polymerases use single-stranded (ss)DNA as a template to catalyze the synthesis of the complementary strand from deoxynucleotides (dNTPs).⁴³ The dNTPs include the four nucleoside triphosphates dATP, dCTP, dGTP, dTTP. DNA Polymerase adds a dNTP to the free 3'hydroxyl group at the end of the DNA strand complementary to the template strand. Consequently, DNA polymerase cannot begin synthesizing a new strand unless the beginning of the growing strand is present. An oligo can be used as a primer to initiate the synthesis of the growing strand. The synthesis of the growing strand can be terminated by the presence of 2',3'-dideoxynucleoside triphosphates, collectively referred to as ddNTPs, which are nucleoside triphosphates that lack the 3' hydroxyl group. When ddNTPs are incorporated into the chain by a polymerase, chain growth is terminated because addition of the subsequent nucleotide requires a free 3' hydroxyl group.⁴

Some DNA Polymerases, such as *E. Coli* DNA Polymerase I, perform other functions including 5' to 3' exonuclease activity that translates, or moves, a nick forward toward the 3' end of DNA, usually by at least tens of nucleotides. This is called nick translation, during which a polymerase excises a series of deoxynucleotides in the DNA molecule and replaces them with free dNTPs added to the reaction.^{43,49} Since some polymerases will incorporate modified dNTPs (e.g., biotin-dUTP), nick translation can be used to replace nucleotides in DNA with modified nucleotides introduced into solution.⁴ The extent of nick translation can be controlled by the addition of ddNTPs. A polymerase that lacks exonuclease activity, such as Vent (exo-) Polymerase, translates nicks similarly except that it produces a displaced strand during translation instead of excising nucleotides.¹⁶ Strand displacement activity varies among DNA polymerases and a DNA polymerase with strand displacement activity is required for some labeling methods, such as the strand displacement variation of nick translation, and is not desirable for others, such as padlock probes.

DNA Methyltransferases (MTases)—DNA MTases are a family of enzymes that sequence-specifically transfer a methyl group to the cytosines and adenines of dsDNA.⁵⁰ Because methylated cognate sites are protected from cleavage by some restriction enzymes, bacteria protect their own DNA from cleavage through methylation and employ restriction enzymes to attack foreign DNA, which is unmethylated at cognate sites. Some MTases perform other functions, but of interest here is the subset of MTases that only perform sequence-specific methylation of DNA using S-adenosyl-L-methionine (AdoMet) as a substrate for methylation.⁵⁰ These AdoMet-dependent MTases catalyze the transfer of methyl groups from the AdoMet substrate to target sequences on DNA. When used in

conjunction with modified AdoMet substrates, these MTases can be used to covalently bind chemical moieties other than methyl groups to specific sequences on dsDNA.⁵¹⁻⁵⁴ Over 35 DNA MTases, with over 20 distinct recognition sequences, are commercially available (Table 1).^{46,47}

3. Sequence-specific DNA labeling methods

Sequence-specific labeling of dsDNA molecules is commonly achieved by the attachment of a small modification such as a hapten or fluorophore to a target sequence. Fundamental considerations for sequence-specific labeling methods include the repertoire of target sequences and the range of available modifications. In the following section, we review the most common approaches to labeling arbitrary dsDNA for single-molecule experiments.

Nick translation

Site-specific labeling of dsDNA via nick translation fundamentally requires the consecutive action of two enzymes.⁴³ First, a nickase cuts one strand of dsDNA to produce a nick at the enzyme recognition site. Second, a polymerase then fills in that nick with nucleotide bases present in solution. If modified nucleotides are present in solution, some polymerases will incorporate them, along with modifications, starting at the nicking location of the nickase recognition sites.⁴ The polymerase promotes hydrolysis and synthesis simultaneously and results in the translation of the nick along the DNA duplex in the 5' to 3' direction. To ensure that nick translation does not continue unchecked, dideoxynucleotides (ddNTPs) can be added and their concentration controls the size of the modified region.⁴ To prevent the labeling of nicks existing prior to the action of the nicking enzyme, ligase is often used as a first step to fill in any non-specific nicks. Nick translation usually also concludes with ligation to repair the translated nicks (Figure 1). This method has been demonstrated to specifically label individual dsDNA molecules with fluorophores such as Alexa 647 and haptens such as biotin, for characterization by optical and force microscopy.^{4,16,20,55-57}

Several variations of the nick translation approach have been employed. One variation of this method uses the strand displacement activity of an (exo-) polymerase to create single-stranded extensions beyond the nicked dsDNA site. The single-stranded extensions, created by the nucleotide chain displaced as the nick was translated, can then be hybridized with modified oligos (Figure 2). Since strand displacement is critical to this method, the DNA polymerase used must have strand displacement activity. Das et al. used this approach in conjunction with nick translation to sequence-specifically modify DNA with the fluorophore Alexa 647 and optically map DNA.¹⁶ The single-stranded extension flaps, which were restricted to 50 bases in length, allowed for further selective labeling of each flap sequence with unique Cy3 fluorescently-labeled oligo probes. These labeled DNA molecules were then stretched by nanochannel confinement and imaged via fluorescence microscopy, thereby creating single-molecule optical maps of genomic-length DNA.

Another variation uses nicking endonucleases to create vicinal nicks, nicks within 15-25 base pairs on the same strand of DNA. The single-stranded segment between the nicks can then be substituted in a strand replacement reaction with an oligo probe. If the oligo probe is modified, the modification is covalently incorporated into the DNA upon ligation, thereby labeling DNA at a prescribed sequence (Figure 3). In a study by Kuhn et al., the oligo probe instead contained an excess of bases at the 3' terminus creating an overhang or flap of single-stranded DNA 20 bases long.⁵⁵ The flap was used downstream as a substrate for a fluorescent reporting system, linear rolling-circle amplification (RCA), that demonstrated highly sequence-specific targeting of viral DNA. While this particular study did not image individual DNA molecules, this approach could easily be used to modify DNA for single-molecule experiments. Furthermore, different flap sequences can be designed to bind to

corresponding modified oligo probes, thereby differentially labeling each flap. Luzzi et al. developed another variation of nick translation specifically for attaching rotor-beads to internal DNA sequences for magnetic tweezer experiments.⁵⁷

There are multiple schemes to utilize nick translation for site-specific labeling with high yield (some greater than 90%).^{4,16,20,55-57} While non-specific nicks are common on DNA and could result in non-specific labeling, treating DNA with ligase first can dramatically reduce this non-specific contribution.⁴ However, the repertoire of target sequences is limited to those with an associated nicking enzyme.

Stem-Loop Triplex Forming Oligonucleotides (TFOs)

TFOs are short single-stranded segments of DNA, typically binding to targets 10-30 bases long, that recognize and bind to the major groove of dsDNA by winding around the target sequence and forming a triple helix.⁵⁸ A special class of TFOs, stem-loop TFOs, are designed such that the center of the TFO binds to the target sequence, roughly 30 bases, and the ends of the TFO hybridize to one another and leave a single-stranded overhang, or flap.¹⁷ In the first of two steps, a looped structure of roughly 70 bases, with a stem acting as an available flap, hybridizes around the target sequence. Lastly, a modified dsDNA fragment with an overhang complementary to the flap is hybridized to the flap and covalently attached to the flap using ligase (Figure 4). The attachment of fragments as long as 500 bp has been demonstrated by Escude et al. and Geron-Landre et al.^{17,59} Binding of the dsDNA fragment to the TFO overhang is followed by ligation to prevent unbinding and excess TFOs can be removed later in the experimental protocol. Interestingly, Cherney et al. demonstrated in an early single-molecule electron-microscopy experiment that even TFOs 12 to 17 bases in length modified with biotin, that do not form stem-loop structures and are not ligated, can bind with initial yields exceeding 70%.⁶⁰

Geron-Landre et al. demonstrated 59 base stem-loop TFOs binding to a 15 bp target with a labeling yield exceeding 60% and noted that the maximum yield was dependent on the TFO design.⁶¹ The 500 bp fragment that was ligated to the stem flap was modified with AlexaFluor 546 and detected with single-molecule fluorescence microscopy. Escude et al. demonstrated this method with unmodified fragments of 200 and 500 bp, that were directly visualized as branched structures off of the DNA backbone using electron microscopy.⁵⁹ They reported a similar yield to Geron-Landre.

One limitation of stem-loop TFOs is that target sequences for this binding motif are restricted to homopurine sequences, which have a homopyrimidine complementary strand in dsDNA. The TFOs bind via different motifs but always to the homopurine strand of the target.⁶² The pyrimidine binding motif is more stable at acidic pH when the cytosine is protonated and can facilitate Hoogsteen binding.⁶³ A modified cytosine base, that is protonated at neutral pH, can also be employed and there are strategies that permit the recognition of mixed homopurine and homopyrimidine sequence dsDNA at physiological pH.^{17,64} An advantage to this method is that the DNA fragment ligated to the flap can contain multiple modifications (i.e. multiple fluorophores and/or haptens) and combinations of modifications. In this way, the method is similar to a variation of nick translation, where single-stranded overhangs are designed to bind to modified oligo probes. Variations on this method include two types of padlock probes, described below.

Padlock Probes

Padlock probes are linear oligos that are similar in design to stem-loop TFOs and also utilize the triplex-forming capability of oligos. While stem-loop TFOs are designed such that the center of the TFO binds to the target sequence and the ends of the TFO hybridize to one

another, padlock probes are designed in the opposite manner such that the ends of the oligo hybridize to the target sequence and the center of the oligo connects the ends to form a ring (Figure 5 and Figure 6). The hybridizing regions are typically 20 bases each and are connected by a roughly 50 base central segment. To make a linear oligo into a closed ring, or padlock, the gaps and nicks present along the oligo backbone upon hybridization to dsDNA are sealed via polymerase and ligase, respectively.

There are two approaches to attach a modification using padlock probes, the labeled probe method and the gap-fill method.⁶⁵ In both approaches, the linear oligo goes on to form a complete modified ring. In the labeled probe method, the central segment of the oligo contains modifications and the padlock probe is designed such that there is a nick between where the two ends hybridize to the target sequence. First, the ends of the probe oligo hybridize to the target sequence creating a target-oligo triplex with a nick in the oligo strand. Second, the nick is sealed via ligase, forming a covalently-bonded modified oligo ring around the target sequence (Figure 5). In the gap-fill method, the padlock probe is designed such that there is a gap of at least one base between where the two ends hybridize to the target sequence. First, the ends of the probe oligo hybridize to the target sequence leaving a gap. Second, this gap is filled in with modified dNTPs using polymerase and the remaining nick is sealed via ligation (Figure 6). The gap-filled ring is similar to that produced with the labeled-probe approach, with the modification placed at the target-binding segment of the oligo instead of in the central segment. Both approaches benefit from the sensitivity of ligase to correct base pairing, as mismatched oligos are not ligated efficiently.⁶⁶

Nilsson et al. used the labeled probe approach with 90 base padlock probes and a biotin-modified linker to detect repeated sequences in human chromosome 12.⁶⁷ The locations of the biotin-modified padlock probes were visualized using fluorescence microscopy by incubating them sequentially with fluorescein-labeled avidin, biotinylated antibodies against avidin, and a second layer of fluoresceinated avidin. While these experiments did not image individual DNA molecules, this technique can be applied to single-molecule experiments. Xiao et al. used the gap-fill approach with 80 to 100 base padlock probes to achieve single-molecule haplotyping.¹⁹ The gap-filling nucleotides were modified with Cy3 and Cy5 fluorophores such that the closed padlocks attached fluorophores to the target sites. Individual labeled DNA molecules were deposited onto a surface and the locations of the padlocks were determined via total internal reflection fluorescence (TIRF) microscopy. The labeling efficiency for a single site was roughly 25% and the labeling efficiency is limited by competition between hybridization of the padlock probe and the native dsDNA renaturation. Variations of this approach use an additional oligo to close the padlock and achieve a multi-stranded stable structure similar to a padlock probe.^{68,69}

Modified Proteins

Using modified proteins to sequence-specifically label DNA takes advantage of both the known sequence-specific binding properties of some proteins and the well-developed protocols and commercially available reagent kits for labeling proteins with fluorophores and haptens (Table 1). In particular, type II restriction enzymes such as EcoRI and EcoRV that, in the absence of Mg^{2+} , can bind to cognate sites on dsDNA without cleaving them are well-suited for labeling DNA (Figure 7).^{21,22,70,71} They can be labeled with reagents that react with the side chains of their cysteine or lysine residues. Care must be taken so that the labeling of those residues does not hinder the binding site of the enzyme and impair binding to the DNA target sequence. Additionally, cysteine or lysine residues should be solvent-accessible to facilitate labeling. One strategy involves labeling a mutant enzyme that is engineered to have only solvent-accessible cysteines or lysines far from the binding pocket.²¹ Another strategy involves labeling a wild type enzyme while protecting those cysteines or lysines at the binding site from labeling.²² Other proteins, such as T7 RNA

Polymerase (RNAP), that would not naturally be immobilized at cognate sites, can also be labeled and covalently cross-linked to dsDNA at these sites.⁷²

There have been several single-molecule demonstrations of fluorescent labeling of dsDNA using the EcoRI restriction enzyme, which has a 6 base recognition sequence, since its binding, cleavage, and structure have been extremely well-studied.⁷³⁻⁷⁵ Oana et al. biotinylated EcoRI using the Sulfo-NHS-LC-Biotin reagent that reacts with the amino groups of one or more of the 22 lysine residues in EcoRI.²² To preserve binding activity, lysines in the binding site were protected from biotinylation by binding EcoRI to heparin prior to biotinylation. The biotinylated EcoRI was incubated with rhodamine-avidin complexes, forming fluorescent EcoRI. Taylor et al. conjugated EcoRI to carboxylate functionalized fluorescent nanoparticles using the cross-linking agent 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDAC).⁷¹ The carboxylate groups on the nanoparticle surface react with the EDAC, forming an active intermediate that is attacked by the amine side chains of the lysine residues in EcoRI. The nanoparticle-bound EcoRI demonstrates site-specific cleavage activity and, when bound to DNA in the absence of Mg^{2+} , the EcoRI nanoparticles label dsDNA at the EcoRI binding sites. Dylla-Spears et al. achieved a similar result to Taylor et al. by biotinylating mutant EcoRI, binding it to dsDNA in the absence of Mg^{2+} , and then incubating the dsDNA with Neutravidin-coated fluorescent nanospheres.²¹ Here control of the biotin position with respect to the enzyme's binding pocket was achieved by biotinylating the enzyme at the mutation site, a nonessential, solvent accessible lysine that was mutated into a cysteine. Biotinylation was performed using the EZ-link Maleimide-PEO2-biotin reagent which forms a thioether bond with the cysteine. By biotinylating a single nonessential site, Dylla-Spears et al. limited the reduction in binding efficiency that can accompany modification of multiple residues.

EcoRV, another well-studied enzyme with a 6 base recognition sequence, has been labeled with qdots and the Cy3 fluorophore. While single-molecule studies of EcoRV have focused on its translocation behavior on dsDNA, it can be used to label DNA in a manner similar to EcoRI. Interestingly, EcoRV, unlike EcoRI, requires Ca^{2+} (and the absence of Mg^{2+}) to bind sequence-specifically to dsDNA without cleaving.⁷⁶ Bonnet et al. labeled an EcoRV variant that contained only a single cysteine located far from the binding site.⁷⁷ A Cy3B-Maleimide Mono-reactive pack reagent was used to directly label the cysteine with the Cy3 fluorophore. Biebricher et al. similarly biotinylated the cysteine on the same EcoRV variant using the Maleimide-PEO2-biotin reagent.⁷⁰ Streptavidin-coated qdots were then used to fluorescently label the EcoRV and the qdot-conjugated EcoRV was shown to retain its sequence specific binding activity in the presence of Ca^{2+} and the absence of Mg^{2+} .

T7 RNAP is a protein that binds to promoter sequences on the T7 bacteriophage genome and then initiates the transcription of RNA from the DNA template. Ebenstein et al. mapped the locations of promoter sequences on individual dsDNA molecules by cross-linking T7 RNAP to DNA at cognate sites upon initiation of transcription.⁷² Formaldehyde was used to form covalent bonds between proximal amino or imino groups such as a lysine residue of the T7 RNAP in contact with a cytosine base of DNA. Biotinylated T7 RNAP was created using a protein expression protocol and then fluorescently labeled with streptavidin-qdots. For dynamic single-molecule studies of T7 RNAP, Kim and Larson used a different, more accessible, approach to fluorescently label T7 RNAP.⁷⁸ T7 RNAP was labeled with the rhodamine fluorophore by serial incubation with two commercially available reagents (Table 1), T7 RNAP mouse monoclonal antibody followed by rhodamine conjugated to anti-mouse IgG.

Unless engineered enzymes, which may not be accessible to the non-specialist, are used the set of target sites accessible with modified binding proteins is limited to the cognate sites of

commercially available proteins.^{46,47} Since labeling methods based on type II restriction enzymes usually require modification of cytosine or lysine residues, preserving the activity of modified enzymes is non-trivial and may require a mutant enzyme, which may also not be available to the non-specialist. Furthermore, some single-molecule experiments, such as studies of motor proteins, require the presence of Mg^{2+} . Modified type II restriction enzymes would not be useful labels for such studies, since in the presence of Mg^{2+} the enzymes would cleave dsDNA instead of serving as sequence-specific labels. However, the Modrich group has developed an EcoRI mutant, EcoRI(Gln-111), that binds with high specificity and does not cleave in the presence of Mg^{2+} .^{79,80} Furthermore, the large number of commercially available type II restriction enzymes and the multitude of commercially available reagent kits for biotinylating or otherwise functionalizing proteins make modified proteins a useful DNA labeling strategy for many applications.

Methyltransferases (MTases)

There are two DNA labeling strategies that exploit the targeted functionalization activity of MTases, both requiring synthetically prepared modified AdoMet analog cofactors.⁵⁰ In the first strategy, Sequence-specific Methyltransferase-Induced Labeling of DNA (SMILing DNA) developed by Pljevaljcic, Schmidt, and Weinhold, a modified AdoMet analogue is designed to be transferred in its entirety by the MTase to DNA.⁸¹ A synthetically prepared modified cofactor is composed of a central moiety for cofactor binding, a reactive aziridine group, and a modification attached through a short linker. Since the final step of the AdoMet analogue synthesis is the coupling of a primary amine with various activated esters, a variety of modifications can be used and both biotin and a fluorophore have been demonstrated. The reaction creates product inhibitors, however, so the stoichiometry must be precisely controlled.⁵⁰ In the second strategy, MTase-Directed Transfer of Activated Groups (mTAG) developed by Lukinavicius et al., the methyl group of AdoMet is replaced with a different functional group of interest such as an amine group.⁸² The functional group is transferred by the MTase to its recognition sites and can be used to attach fluorophores and haptens in a subsequent coupling step (Figure 8). The reaction rate decreases as the size of the transferable group increases but there are methods to increase transfer efficiency.⁵⁰

Pljevaljcic et al. biotinylated DNA using the SMILing DNA approach with M.TaqI, which has a 4 base long cognate site.⁸³ A biotinylated synthetic AdoMet analogue was transferred in its entirety through the opening of its aziridine ring to the M.TaqI cognate sites. Upon treating with streptavidin, biotinylated sites on individual DNA molecules were visualized using electron microscopy. An analogous cofactor, with a Cy5 fluorophore modification instead of a biotin, was also demonstrated to work effectively with M.TaqI. Similarly, Braun et al. biotinylated DNA using M.TaqI and M.BseCI, which has a 6 base long cognate site, and a modified cofactor.¹² A biotinylated synthetic AdoMet analogue was transferred in its entirety to the cognate sites and biotinylated gold nanoparticles were attached to biotinylated DNA sites using free streptavidin as a bridge. Labeled DNA was deposited on mica and individual molecules were directly visualized via AFM.

Neely et al. used the mTAG approach with an engineered version of the MTase M.HhaI, which recognizes a 4 base long sequence, to densely label DNA with fluorophores and construct optical DNA maps via fluorescence microscopy from individual DNA molecules deposited on a surface. First, DNA was incubated with engineered M.HhaI and a synthetically prepared modified AdoMet analog cofactor, Ado-11-amino, that contained an appended transferable amine group. With the modified cofactor as a substrate, M.HhaI covalently transferred the amine group to its recognition sequence on DNA. Second, the amine-modified DNA was incubated with an amine-reactive fluorophore Atto-647N. This process produced DNA molecules with 50-60% of the M.HhaI target sites fluorescently labeled. The efficiency of the initial amine modification by the MTase is thought to be near

complete and this reduced yield is attributed to the lower efficiency of the amine-reactive fluorophore coupling in the second step.

Since the specificity of a DNA MTase is, to a first approximation, equivalent to that of its companion restriction enzyme, the functionalization activity of MTases is considered highly sequence specific.⁸⁴ The lengths of the recognition sequences in the above examples are typical as over 70% of the commercially available MTases recognize sequences with lengths between 4 to 6 bases.^{46,47} Furthermore, the short length of some MTase recognition sequences allows for highly dense labeling. At least 25% of the commercially available MTases have recognition sequences with lengths of 4 bases or less and shorter recognition sequences are likely to occur more often than longer sequences on a random DNA molecule. Ultimately, however, the utility of MTase-based labeling depends on the availability or ease of synthesis of modified AdoMet analogs, which to our knowledge are not commercially available, and the library of recognition sequences of available MTases.

Peptide Nucleic Acids (PNAs)

Upon simple incubation, PNAs bind to dsDNA via one of four binding modes (Figure 9), determined by the PNA design.³⁵ The binding mode dictates the base content of the DNA sequences that can be targeted. The first binding mode, triplex invasion, is the most stable and requires two homopyrimidine PNA molecules to simultaneously bind to the homopurine strand of the DNA recognition sequence. A bisPNA, which is a construct of the two triplex-forming PNA molecules covalently attached by a central linker, is often used to enhance binding rates for this binding mode. The second mode, double duplex invasion, requires two sequence complementary PNA molecules to bind to the two complementary strands of the target sequence on dsDNA. The PNAs contain sterically hindered analogs of the adenine and thymine bases so that the two sequence complementary PNA molecules will prefer to bind to a complementary DNA strand rather than each other. For the double duplex invasion binding mode, the base composition of the PNA sequence needs to have at least 50% combined cytosine and guanine base content. The third mode, triplex, which creates a triple helix by binding to the major groove of dsDNA, seems to require high cytosine content. The fourth and final mode is duplex invasion, where a single PNA molecule binds to a DNA strand by displacing the complementary DNA strand, and requires a guanine-rich homopurine PNA sequence. Even given the sequence restrictions imposed by each binding mode, PNAs can be employed to bind to an extensive repertoire of target binding sequences. Furthermore, they are commercially available (Table 1) and made to order with a choice of conjugated modifications including haptens and fluorophores.

Demidov et al. demonstrated physical mapping of DNA molecules using biotinylated bisPNAs with a 10 base recognition sequence.⁸⁵ Individual biotinylated PNA-DNA complexes were conjugated to streptavidin protein, which could be directly visualized via electron microscopy. The streptavidin protein served as the electron microscopy markers and appeared as 'beads' located at the target sites. Kim et al. and Qu et al. demonstrated optical mapping of DNA using PNAs with 15 and 7 base recognition sequences, respectively, modified with Alexa fluorophores.^{18,86} Kim et al. determined the position of their Alexa 532 labeled PNA along the DNA molecule by concurrent scanning near-field optical and atomic force microscopy (SNOM/AFM), while Qu et al. measured the location of their Alexa 488 fluorophore-conjugated PNA on the DNA by nanometer-localized multiple single molecule (NALMS) fluorescence microscopy.

Chan et al. and Phillips et al. demonstrated mapping of DNA using bisPNAs conjugated to Alexa 546 or tetramethylrhodamine fluorophores with 7 or 8 base recognition sequences.^{3,23} The PNA-DNA complexes, created with roughly 90% yield, were introduced into a continuous flow device for direct linear analysis (DLA), a scheme that uses flow to linearize

the DNA and provides the spatial locations of the PNAs along the extended DNA molecule. The fluorescently labeled DNA was analyzed as it flowed past multicolor confocal fluorescence detectors with single-fluorophore sensitivity. Zohar et al. demonstrated mapping of DNA using biotinylated bisPNAs with 8 base recognition sequences that were incubated with Neutravidin-coated fluorescent nanospheres.²⁴ Labeled DNA was introduced into a microfluidic cross-slot where molecules were trapped and stretched in a stagnation point flow. The locations of the bisPNAs along the backbone of the DNA were determined via fluorescence microscopy. Zohar et al. further demonstrated that these same biotinylated PNAs can be used as robust and benign tethers in conjunction with avidin-coated microspheres for optical tweezer experiments and that they bind to DNA with an average strength of roughly 60 pN.

While creating PNA-DNA complexes is relatively simple, the incubation conditions must be carefully optimized for yield and specificity for a given PNA molecule.^{3,23,24,87} Incubation conditions affecting yield or specificity include incubation time and temperature,^{3,88} pH,^{89,90} salt concentration,^{39,88,91,92} and the ratio of PNA to DNA.^{3,39} Compared to other labeling approaches, however, PNAs are advantageous because they are commercially available with modifications, compatible with cofactors such as Mg^{2+} , and non-destructive to dsDNA. Furthermore, they do not require ligation and, under optimized conditions, bind with high specificity and yield.^{3,23} Though PNAs are expensive relative to other commercially available reagents, a typical minimum order of roughly 40 nmol may exceed the amount needed for bulk optimization and a full course of single-molecule experiments.²⁴ PNAs also offer much more flexibility in the accessible target sequence than the enzyme-based labeling approaches using currently commercially available nickases, modified proteins, and MTases.

4. Conclusions and outlook

Single-molecule DNA experiments allow for unprecedented insights yet are challenging and diverse in their requirements. Frequently, they require the internal sequence-specific labeling of long and arbitrary dsDNA molecules. As experimentalists of different backgrounds are increasingly engaging in single-molecule DNA experiments and each new experimental system brings new constraints, there is a need for multiple labeling options that are accessible to non-specialists. Ideally, a sequence-specific dsDNA labeling approach should be straightforward, require only commercially available reagents, and offer flexibility in target sequences and modifications for labeling (Table 2). Furthermore, depending on the experimental demands, the labeling method may need to attach modifications covalently, not require ligase, be robust to cofactors such as Mg^{2+} , or exhibit high labeling yield (Table 3). It is important to note that the labeling yield refers to the initial modification made to DNA and that subsequent labeling steps, such as the attachment of a nanosphere to a modification, will result in lower overall yield.

When considering the six approaches described, each has advantages and disadvantages and, ultimately, the approach of choice will depend on individual experimental constraints. The approaches that use an enzyme to introduce a modification - nick translation, modified proteins, and MTases - share a limitation that the repertoire of target sequences is limited to the cognate sites of the commercially available enzymes. While nick translation and MTases produce covalently-bound modifications, modified proteins are not covalently-bound. Furthermore, modified restriction enzymes are not compatible with Mg^{2+} as it will result in enzymatic cleavage, though specially-engineered mutant enzymes may overcome this restriction. The enzymatic approaches share the advantage that the binding conditions for high yield sequence-specific activity are known for commercially available enzymes and require little to no optimization. Nick translation is particularly straightforward, MTases are

easy to use with the main challenge being to secure a modified cofactor, and modifying proteins is possible with commercially available reagents but requires specialized techniques. MTases, which can have short cognate sequences, are particularly useful when high-density labeling is required. Furthermore, nick translation and MTase-based approaches produce modified DNA constructs without leaving physical obstacles on the DNA molecule.

The approaches that use a triplex-former to introduce a modification – stem-loop TFOs, padlock probes, and PNAs – require that each new probe be custom-designed and share a limitation that the binding conditions must be optimized for each designed probe. Triplex-formers, however, can access a variety of target sequences and are not limited to commercially available enzymatic cognate sites. Stem-loop TFOs and padlock probes are not directly covalently bound to the target sequence but form ligated ring-like structures that are topologically constrained to the target sequence location. PNAs are also not covalently bound to the target sequence yet have been shown to be robust and suitable as optical tweezer handles. Labeling approaches based on triplex-formers locally alter DNA structure and leave a physical obstacle on the DNA molecule. This property can be undesirable or it can be exploited, for example when obstacles for translocases are desired.

Labeling approaches that result in a covalently-bound modification (as indicated in Table 3) produce the most stable labeling of DNA. The remaining non-covalent labeling approaches, modified proteins and PNAs, produce labeled DNA that is inherently less stable. The off rates of modified proteins will vary for different proteins and the off rates of PNAs will depend on characteristics of the molecule, including the base sequence and charge. Our experiences with a limited number of modified proteins and PNAs suggest that there is no substantial reduction in labeling yield over the course of a day, for modified proteins, and several days, for PNAs.

In an effort to make single-molecule DNA experiments more accessible to non-specialists, we have reviewed six demonstrated approaches to sequence-specific labeling of DNA. There are other approaches, including those based upon zinc fingers and locked nucleic acids (LNAs) and other triplex-formers that could potentially be used to label dsDNA for single-molecule experiments.^{15,62,93-98} Our choice of examples here has been illustrative and selective, and related experiments, omitted in the interest of conciseness, may be found in the literature. Moreover, new techniques, specialized reagents, and demonstrations of variations of existing approaches are appearing with increasing speed in this burgeoning field. We hope that this review serves as an introduction to those new to the field and provides a framework for evaluating new options as they develop.

Acknowledgments

This work was supported by NIH award R21HG4342. H.Z. gratefully acknowledges an NSF Graduate Research Fellowship. SJM gratefully acknowledges support from the Radcliffe Institute for Advanced Study during the preparation of this manuscript. The authors thank Craig L. Hetherington for helpful and insightful discussions.

References

1. Walter NG, Huang C, Manzo AJ, Sobhy MA. Nat. Methods. 2008; 5:475–489. [PubMed: 18511916]
2. Ritort F. J. Phys.: Condens. Matter. 2006; 18:R531–R583. [PubMed: 21690856]
3. Chan EY, Goncalves NM, Haeusler RA, Hatch AJ, Larson JW, Maletta A, Yantz G, Carstea E, Fuchs M, Wong G, Gullans S, Gilmanishin R. Genome Res. 2004; 14:1137–1146. [PubMed: 15173119]

4. Jo K, Dhingra DM, Odijk T, de Pablo JJ, Graham MD, Runnheim R, Forrest D, Schwartz DC. *Proc. Natl. Acad. Sci. U. S. A.* 2007; 104:2673–2678. [PubMed: 17296933]
5. Neely RK, Dedecker P, Hotta J, Urbanavičiūtė G, Klimasauskas S, Hofkens J. *Chemical Science.* 2010; 1:453–460.
6. Bryant Z, Stone MD, Gore J, Smith SB, Cozzarelli NR, Bustamante C. *Nature.* 2003; 424:338–341. [PubMed: 12867987]
7. Gore J, Bryant Z, Stone MD, Nollmann MN, Cozzarelli NR, Bustamante C. *Nature.* 2006; 439:100–104. [PubMed: 16397501]
8. Hugel T, Michaelis J, Hetherington CL, Jardine PJ, Grimes S, Walter JM, Falk W, Anderson DL, Bustamante C. *PLoS Biol.* 2007; 5:e59. [PubMed: 17311473]
9. Olson MV. *Science.* 1995; 270:394–394. [PubMed: 7569992]
10. Olson M, Hood L, Cantor C, Botstein D. *Science.* 1989; 245:1434–1434. [PubMed: 2781285]
11. Roberts L. *Science.* 1989; 245:1438. [PubMed: 2781288]
12. Braun G, Diechtierow M, Wilkinson S, Schmidt F, Hüben M, Weinhold E, Reich NO. *Bioconjugate Chem.* 2008; 19:476–479.
13. Ghosh I, Stains CI, Ooi AT, Segal DJ. *Mol. BioSyst.* 2006; 2:551–560. [PubMed: 17216036]
14. Kwok PY, Xiao M. *Human Mutation.* 2004; 23:442–446. [PubMed: 15108275]
15. Weisbrod SH, Marx A. *Chem. Commun.* 2008:5675–5685.
16. Das SK, Austin MD, Akana MC, Deshpande P, Cao H, Xiao M. *Nucleic Acids Res.* 2010; 38:e177. [PubMed: 20699272]
17. Geron-Landre B, Roulon T, Escude C. *FEBS J.* 2005; 272:5343–5352. [PubMed: 16218964]
18. Qu XH, Wu D, Mets L, Scherer NF. *Proc. Natl. Acad. Sci. U. S. A.* 2004; 101:11298–11303. [PubMed: 15277661]
19. Xiao M, Gordon MP, Phong A, Ha C, Chan T, Cai D, Selvin PR, Kwok P. *Human Mutation.* 2007; 28:913–921. [PubMed: 17443670]
20. Xiao M, Phong A, Ha C, Chan T, Cai D, Leung L, Wan E, Kistler AL, DeRisi JL, Selvin PR, Kwok P. *Nucleic Acids Res.* 2007; 35:e16. [PubMed: 17175538]
21. Dylla-Spears R, Townsend JE, Sohn LL, Jen-Jacobson L, Muller SJ. *Anal. Chem.* 2009; 81:10049–10054. [PubMed: 19908852]
22. Oana H, Ueda M, Washizu M. *Biochem. Biophys. Res. Commun.* 1999; 265:140–143. [PubMed: 10548504]
23. Phillips KM, Larson JW, Yantz GR, D'antoni CM, Gallo MV, Gillis KA, Goncalves NM, Neely LA, Gullans SR, Gilmanshin R. *Nucleic Acids Res.* 2005; 33:5829–5837. [PubMed: 16243782]
24. Zohar H, Hetherington C, Bustamante C, Muller SJ. *Nano Lett.* 2010; 10:4697–4701. [PubMed: 20923183]
25. Landsteiner, K. *The Specificity of Serological Reactions.* Dover Publications; Mineola: 1962.
26. Diamandis EP, Christopoulos TK. *Clin. Chem.* 1991; 37:625. [PubMed: 2032315]
27. Bloch K, Grossmann B. *Current Protocols in Molecular Biology.* 1995; (Supplement 31)
28. Kramer M, Coen D. *Current Protocols in Molecular Biology.* 2001; (Supplement 56)
29. Kool ET. *Chem. Rev.* 1997; 97:1473–1487. [PubMed: 11851456]
30. Dieffenbach, CW.; Dveksler, GS. *PCR Primer: A Laboratory Manual.* Cold Spring Harbor Laboratory Press; Cold Spring Harbor: 2003.
31. Nielsen PE. *Curr. Opin. Biotechnol.* 1999; 10:71–75. [PubMed: 10047504]
32. Nielsen PE. *Curr. Opin. Biotechnol.* 2001; 12:16–20. [PubMed: 11167067]
33. Nielsen, PE. *PNA Technology.* Nielsen, PE., editor. Vol. vol. 208. Humana Press; Totowa: 2002. p. 3–26.
34. Nielsen PE, Haaima G. *Chem. Soc. Rev.* 1997; 26:73–78.
35. Nielsen PE. *Methods in Enzymology.* 2001; 340:329–340. [PubMed: 11494857]
36. Pellestor F, Paulasova P. *European Journal of Human Genetics.* 2004; 12:694–700. [PubMed: 15213706]
37. Pellestor F, Paulasova P. *Chromosoma.* 2004; 112:375–380. [PubMed: 15156326]

38. Larsen HJ, Bentin T, Nielsen PE. *Biochim. Biophys. Acta*. 1999; 1489:159–166. [PubMed: 10807005]
39. Lundin, KE.; Good, L.; Stromberg, R.; Graslund, A.; Smith, CIE. *Advances in Genetics*. Hall, JC.; Dunlap, JC.; Friedman, T.; van Heyningen, V., editors. Vol. vol. 56. Elsevier; 2006. p. 1-51.
40. Paulasova P, Pellestor F. *Annales de Genetique*. 2004; 47:349–358. [PubMed: 15581832]
41. Ray A, Norden B. *FASEB J*. 2000; 14:1041–1060. [PubMed: 10834926]
42. Shakeel S, Karim S, Ali A. *J. Chem. Technol. Biotechnol*. 2006; 81:892–899.
43. Sambrook, J.; Russell, DW. *Molecular Cloning*. Cold Spring Harbor Laboratory Press; New York: 2001.
44. Pingoud A, Jeltsch A. *Nucleic Acids Res*. 2001; 29:3705–3727. [PubMed: 11557805]
45. Heiter DF, Lunnen KD, Wilson GG. *J. Mol. Biol*. 2005; 348:631–640. [PubMed: 15826660]
46. Roberts RJ, Vincze T, Posfai J, Macelis D. *Nucleic Acids Res*. 2005; 33:D230–D232. [PubMed: 15608184]
47. Roberts RJ, Vincze T, Posfai J, Macelis D. *Nucleic Acids Res*. 2010; 38:D234–D236. [PubMed: 19846593]
48. Zhang P, Too PH, Samuelson JC, Chan S, Vincze T, Doucette S, Bäckström S, Potamouisis KD, Schramm TM, Forrest D, Schwartz DC, Xu S. *Protein Expression Purif*. 2010; 69:226–234.
49. Kelly R, Cozzarelli N, Deutscher M, Lehman I, Kornberg A. *J. Biol. Chem*. 1970; 245:39–45. [PubMed: 4904090]
50. Klimasauskas S, Weinhold E. *Trends in Biotechnol*. 2007; 25:99–104.
51. Liutkeviciute Z, Lukinavicius G, Masevicius V, Daujotyte D, Klimasauskas S. *Nat. Chem. Biol*. 2009; 5:400–402. [PubMed: 19430486]
52. Pignot M, Pljevaljcic G, Weinhold E. *Eur. J. Org. Chem*. 2000:549–555.
53. Pljevaljcic G, Pignot M, Weinhold E. *J. Am. Chem. Soc*. 2003; 125:3486–3492. [PubMed: 12643710]
54. Pljevaljcic, G.; Schmidt, F.; Peschlow, A.; Weinhold, E.; Niemeyer, CM. *Sequence-specific DNA labeling using methyltransferases*. Vol. vol. 283. Humana Press Inc; Totowa: 2004. p. 145-161.
55. Kuhn H, Frank-Kamenetskii MD. *Nucleic Acids Res*. 2008; 36:e40. [PubMed: 18344522]
56. Pfannschmidt C, Langowski J. *J. Mol. Biol*. 1998; 275:601–611. [PubMed: 9466934]
57. Luzzietti N, Brutzer H, Klaue D, Schwarz FW, Staroske W, Clausing S, Seidel R. *Nucleic Acids Res*. 2011; 39:e15. [PubMed: 21071409]
58. Knauert MP, Glazer PM. *Hum. Mol. Genet*. 2001; 10:2243. [PubMed: 11673407]
59. Escude C, Roulon T, Lyonnais S, Le Cam E. *Anal. Biochem*. 2007; 362:55–62. [PubMed: 17250797]
60. Cherney DI, Malkov VA, Volodin AA, Frank-Kamenetskii MD. *J. Mol. Biol*. 1993; 230:379–383. [PubMed: 8464052]
61. Géron-Landre B, Roulon T, Desbiolles P, Escudé C. *Nucleic Acids Res*. 2003; 31:e125. [PubMed: 14530458]
62. Duca M, Vekhoff P, Oussedik K, Halby L, Arimondo PB. *Nucleic Acids Res*. 2008; 36:5123–5138. [PubMed: 18676453]
63. Thuong NT, Helene C. *Angew. Chem., Int. Ed*. 1993; 22:666–690.
64. Rusling DA, Powers EC, Ranasinghe RT, Wang Y, Osborne SD, Brown T, Fox KR. *Nucleic Acids Res*. 2005; 33:3025–3032. [PubMed: 15911633]
65. Gordon MP. 2006:1–86.
66. Landegren U, Kaiser R, Sanders J, Hood L. *Science*. 1988; 241:1077–1080. [PubMed: 3413476]
67. Nilsson M, Helena M, Samiotaki M, Kwiatkowski M, Chowdhary BP. *Science*. 1994; 265:2085–2088. [PubMed: 7522346]
68. Shigemori Y, Haruta H, Okada T, Oishi M. *Genome Res*. 2004; 14:2478–2485. [PubMed: 15574826]
69. Roulon T, Coulaud D, Delain E, Le Cam E, Hélène C, Escudé C. *Nucleic Acids Res*. 2002; 30:E12. [PubMed: 11809900]

70. Biebricher A, Wende W, Escudé C, Pingoud A, Desbiolles P. *Biophys. J.* 2009; 96:L50–L52. [PubMed: 19383444]
71. Taylor JR, Fang MM, Nie S. *Anal. Chem.* 2000; 72:1979–1986. [PubMed: 10815954]
72. Ebenstein Y, Gassman N, Kim S, Antelman J, Kim Y, Ho S, Samuel R, Michalet X, Weiss S. *Nano Lett.* 2009; 9:1598–1603. [PubMed: 19290670]
73. Halford S, Johnson N, Grinsted J. *Biochem. J.* 1980; 191:581. [PubMed: 6263249]
74. Kim Y, Grable JC, Love R, Greene PJ, Rosenberg JM. *Science.* 1990; 249:1307–1309. [PubMed: 2399465]
75. Lesser DR, Kurpiewski MR, Jen-Jacobson L. *Science.* 1990; 250:776–786. [PubMed: 2237428]
76. Vipond IB, Halford SE. *Biochemistry.* 1995; 34:1113–1119. [PubMed: 7827059]
77. Bonnet I, Biebricher A, Porte P, Loverdo C, Benichou O, Voituriez R, Escude C, Wende W, Pingoud A, Desbiolles P. *Nucleic Acids Res.* 2008; 36:4118–4127. [PubMed: 18544605]
78. Kim JH, Larson RG. *Nucleic Acids Res.* 2007; 35:3848–3858. [PubMed: 17526520]
79. Allison DP, Kerper PS, Doktycz MJ, Spain JA, Modrich P, Larimer FW, Thundat T, Warmack RJ. *Proc. Natl. Acad. Sci. U. S. A.* 1996; 93:8826–8829. [PubMed: 8799111]
80. Wright DJ, King K, Modrich P. *J. Biol. Chem.* 1989; 264:11816–11821. [PubMed: 2745418]
81. Pljevaljcic G, Schmidt F, Weinhold E. *ChemBioChem.* 2004; 5:265–269. [PubMed: 14997517]
82. Lukinavicius G, Lapiene V, Stasevskij Z, Dalhoff C, Weinhold E, Klimasauskas S. *J. Am. Chem. Soc.* 2007; 129:2758–2759. [PubMed: 17309265]
83. Pljevaljcic G, Schmidt F, Scheidig AJ, Lurz R, Weinhold E. *ChemBioChem.* 2007; 8:1516–1519. [PubMed: 17654629]
84. Lopez OJ, Quintanar A, Padhye NV, Nelson M. *J. Immunoassay Immunochem.* 2003; 24:11–28. [PubMed: 12680604]
85. Demidov VV, Cherny D, Kurakin A, Yavnilovich M, Malkov V, Frank-Kamenetskii M, Sonnichsen S, Nielsen PE. *Nucleic Acids Res.* 1994; 22:5218–5222. [PubMed: 7816609]
86. Kim J, Hirose T, Sugiyama S, Ohtani T, Muramatsu H. *Nano Lett.* 2004; 4:2091–2097.
87. Singer A, Wanunu M, Morrison W, Kuhn H, Frank-Kamenetskii M, Meller A. *Nano Lett.* 2010; 10:738–742. [PubMed: 20088590]
88. Kosaganov YN, Stetsenko DA, Lubyako EN, Kvitko NP, Lazurkin YS, Nielsen PE. *Biochemistry.* 2000; 39:11742–11747. [PubMed: 10995242]
89. Hansen ME, Bentin T, Nielsen PE. *Nucleic Acids Res.* 2009; 37:4498–4507. [PubMed: 19474349]
90. Kuhn H, Demidov VV, Nielsen PE, Frank-Kamenetskii MD. *J. Mol. Biol.* 1999; 286:1337–1345. [PubMed: 10064701]
91. Abibi A, Protozanova E, Demidov VV, Frank-Kamenetskii MD. *Biophys. J.* 2004; 86:3070–3078. [PubMed: 15111421]
92. Kuhn H, Demidov VV, Frank-Kamenetskii MD, Nielsen PE. *Nucleic Acids Res.* 1998; 26:582–587. [PubMed: 9421519]
93. Demidov VV. *Trends in Biotechnol.* 2003; 21:4–7.
94. Kauppinen S, Vester B, Wengel J. *Drug Discovery Today: Technologies.* 2005; 2:287–290.
95. Petersen M, Wengel J. *Trends in Biotechnol.* 2003
96. Wahlestedt C, Salmi P, Good L, Kela J, Johnsson T, Hökfelt T, Broberger C, Porreca F, Lai J, Ren K, Ossipov M, Koshkin A, Jakobsen N, Skouv J, Oerum H, Jacobsen MH, Wengel J. *Proc. Natl. Acad. Sci. U. S. A.* 2000; 97:5633–5638. [PubMed: 10805816]
97. Xu G, Bestor TH. *Nat. Genet.* 1997; 17:376–378. [PubMed: 9398832]
98. Kiss A, Weinhold E. *ChemBioChem.* 2008; 9:351–353. [PubMed: 18189249]

Nick Translation

Figure 1. Scheme for covalently modifying DNA via nick translation. Pentagons represent any modifications available conjugated to nucleotides.

Strand Displacement and Nick Translation

Figure 2.

Scheme for covalently modifying DNA using a variation of nick translation. Pentagons represent any modifications available conjugated to nucleotides. The square represents any modification that can be conjugated to an oligo. This approach allows for differential dual labeling of the nicking site and flap.

Vicinal Nick Translation

Figure 3. Scheme for covalently modifying DNA using a variation of nick translation. Pentagons represent any modification that can be conjugated to oligos.

Stem-loop TFO

Figure 4. Scheme for modifying DNA using stem-loop TFOs. Pentagons represent any modification that can be conjugated to oligos.

Padlock Probe: labeled probe

TFO forms a loop around target with a nick at the target sequence

Ligase seals the nick creating a closed modified padlock

Figure 5. Scheme for modifying DNA using a labeled probe padlock. Pentagons represent any modification that can be conjugated to oligos.

Padlock Probe: gap-fill

Figure 6. Scheme for modifying DNA using a gap-filled probe padlock. Pentagons represent any modifications available conjugated to nucleotides.

Modified Protein

Type II restriction enzyme is modified far from the binding site to preserve activity

Modified type II restriction enzyme binds to cognate sequence and does not cleave (in absence of Mg^{2+})

Figure 7. Scheme for modifying DNA using a modified protein. Pentagons represent any modification that can be conjugated to a protein.

Methyltransferase and Modified Cofactor

Methyltransferase has a modified
cofactor as a substrate

Methyltransferase transfers modification
from cofactor to cognate sequence

DNA is covalently modified,
methyltransferase and reacted substrate
are released

Figure 8. Scheme for covalently modifying DNA using Methyltransferase-directed Transfer of Activated Groups (mTAG). Pentagons represent a transferable reactive group appended to an AdoMet analog cofactor. Circles represent the remainder of the AdoMet analog cofactor.

Peptide Nucleic Acid Binding Modes

Triplex Invasion
(bisPNA)

Double Duplex Invasion

Triplex

Duplex Invasion

Figure 9.

Schematics of PNA-DNA binding modes. Pentagons represent any modification that can be conjugated to a PNA. The PNA-DNA complexes are created by simple incubation and the binding mode is determined by the PNA design.

Table 1

Representative suppliers for commercially available reagents used in the experiments described.

Reagent	Suppliers
DNA	NEB, Fermentas
Proteins (restriction enzymes, methyltransferases, ligase, antibodies, etc.)	NEB, Fermentas, Novagen, Invitrogen
Custom DNA Oligos	Integrated DNA Technologies (IDT), Invitrogen
Modified Nucleotides	Invitrogen
Peptide Nucleic Acids	Biosynthesis Inc., Cambridge Research Biochemicals
Protein functionalization kits	Pierce, Amersham
Hapten-conjugated fluorophores	Invitrogen
Functionalized quantum dots and microspheres	Invitrogen, Spherotech

Table 2

Accessibility and versatility of each labeling approach.

Approach	Commercially available modified reagents	Ease of use for non-specialist	Target Sequence Repertoire	Modification Repertoire
Nick Translation	yes	+++	Limited to nicking endonuclease cognate sites	Available modified nucleotides
Stem-loop TFOs	yes	++ requires binding optimization	Limited to homopurine or homopyrimidine	Available custom oligo modifications
Padlock Probes	yes	++ requires binding optimization	Limited to homopurine or homopyrimidine	Available modified nucleotides (gap-fill) or custom oligo (linker probe) modifications
Modified Proteins	No, may require mutant enzymes. For wild enzymes, the proteins and modifying kits are available	+ requires binding optimization and specialized enzyme knowledge	Limited to protein cognate sites	Available modifying kits that react with lysine or cysteine residues
Methyltransferases and modified cofactors	No, modified cofactors are not available	+++	Limited to methyltransferase cognate sites	Available modified cofactors
Peptide Nucleic Acids	yes	++ requires binding optimization	Depends on binding mode: strongest limited to homopurine	Available custom PNA modifications

Table 3

Characteristics of binding and labeled DNA for each labeling approach.

Approach	Covalently-bound modification	Requires ligase	Compatible with Mg ²⁺ cofactor	Typical reported modification yield
Nick Translation	✓	✓	✓	> 90% ⁵⁵
Stem-loop TFOs	✓ ^a	✓	✓	> 60 % ^{59,61}
Padlock Probes	✓ ^a	✓	✓	25-50% ^{19,65}
Modified Proteins				> 50% ²¹
Methyltransferases and modified cofactors	✓		✓	Quantitative, can be near complete ^{5,12,51,82}
Peptide Nucleic Acids			✓	80-90% ³

^a Modifications are on topologically constrained ligated ring-like structures