

# Chemical discrimination between dC and <sup>5</sup>Me dC via their hydroxylamine adducts

Martin Münzel, Lukas Lercher, Markus Müller\* and Thomas Carell\*

Department of Chemistry, Center for Integrated Protein Science Munich, Ludwig-Maximilians University, Butenandtstr. 5-13, 81377 Munich, Germany

Received May 7, 2010; Revised June 28, 2010; Accepted July 30, 2010

## ABSTRACT

The presence of the methylated nucleobase <sup>5</sup>Me dC in CpG islands is a key factor that determines gene silencing. False methylation patterns are responsible for deteriorated cellular development and are a hallmark of many cancers. Today genes can be sequenced for the content of <sup>5</sup>Me dC only with the help of the bisulfite reagent, which is based exclusively on chemical reactivity differences established by the additional methyl group. Despite intensive optimization of the bisulfite protocol, the method still has specificity problems. Most importantly ~95% of the DNA analyte is degraded during the analysis procedure. We discovered that the reagent O-allylhydroxylamine is able to discriminate between dC and <sup>5</sup>Me dC. The reagent, in contrast to bisulfite, does not exploit reactivity differences but gives directly different reaction products. The reagent forms a stable mutagenic adduct with dC, which can exist in two states (E versus Z). In case of dC the allylhydroxylamine adduct switches into the E-isomeric form, which generates dC to dT transition mutations that can easily be detected by established methods. Significantly, the <sup>5</sup>Me dC-adduct adopts exclusively the Z-isomeric form, which causes the polymerase to stop. O-allylhydroxylamine does allow differentiation between dC and <sup>5</sup>Me dC with high accuracy, leading towards a novel and mild chemistry for methylation analysis.

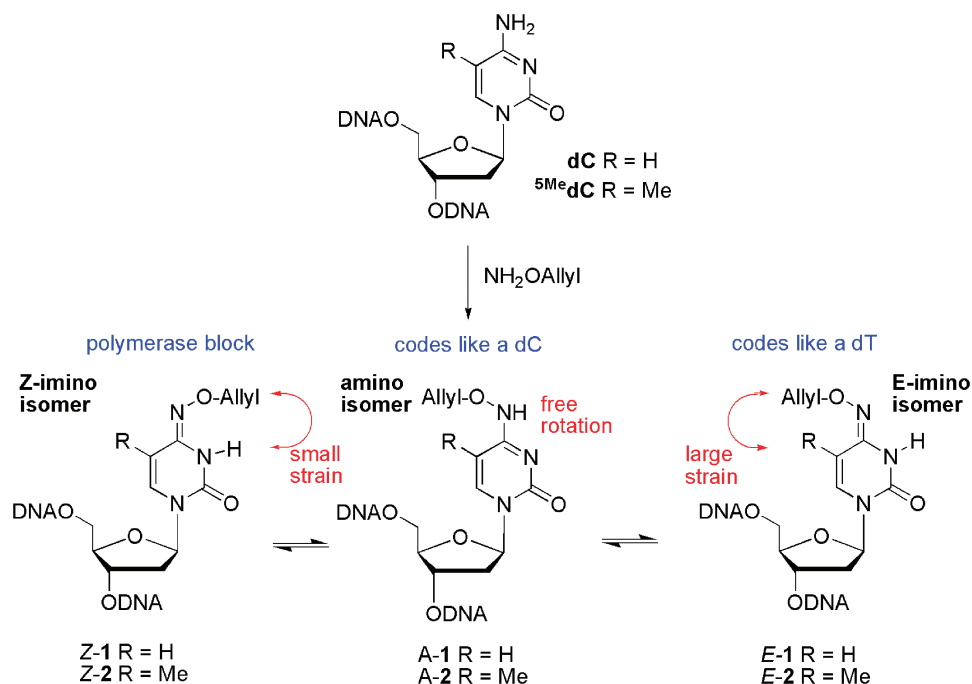
## INTRODUCTION

DNA methylation is an epigenetic mechanism for transcriptional regulation (1–2). The process controls cellular differentiation and is defective in many diseases including cancer (3–5). DNA methylation involves substitution of

the H atom at C5 of the cytosine base (dC) by a methyl group to give 5-methylcytosine (<sup>5</sup>Me dC) (6). These methylated <sup>5</sup>Me dC bases trigger processes that finally lead to gene silencing. Today, the detection of <sup>5</sup>Me dC bases in a given DNA sequence is possible with restriction enzymes, which cleave selectively unmethylated DNA (7), or with antibodies, which recognize methylated DNA (8–10). The current gold standard, however, is bisulfite sequencing. This reagent selectively deaminates dC to dU but does not affect <sup>5</sup>Me dC (11, 12). The position of the newly formed dU base is then detected with standard sequencing methods after PCR amplification of the treated DNA material (13). Bisulfite sequencing is currently in widespread use (14–18). However, during bisulfite analysis ~95% of the DNA material is degraded, which limits <sup>5</sup>Me dC detection if only a small amount of sample is available (19–21). This drawback fuels current research to find alternative <sup>5</sup>Me dC detection and sequencing methods (22). Research in this direction has furnished the result that Br<sup>+</sup> ions add selectively to <sup>5</sup>Me dC to give cleavable DNA sites (23). OsO<sub>4</sub> and reagents derived thereof were found to *cis*-hydroxylate dC bases with some selectivity (24, 25). Furthermore photochemical methods were reported (26, 27) that allow some discrimination and finally even compounds were reported that are added during bisulfite sequencing to trap intermediates of the bisulfite reaction (28). All reported approaches target the reactivity differences between dC and <sup>5</sup>Me dC imposed by the additional methyl group. However, as this difference is small, all new procedures are currently unable to distinguish between the dC and <sup>5</sup>Me dC efficiently or their general approach interferes with current sequencing techniques.

Here we report a novel concept that does not rely on reactivity differences but in contrast allows detection of <sup>5</sup>Me dC directly based on the reaction product. Incubation of dC and <sup>5</sup>Me dC with hydroxylamine derivatives gives *E*- and *Z*-configured oxime-type adducts (*Z/E*-1 and *Z/E*-2, see Scheme 1) (29,30). The *E*- and *Z*-configured products are in equilibrium via the amino isomeric forms A-1 and A-2. Each of these isomers

\*To whom correspondence should be addressed. Tel: +49 89 2180 77755; Fax: +49 89 2180 77756; Email: thomas.carell@cup.uni-muenchen.de  
Correspondence may also be addressed to Markus Müller. Tel: +49 89 2180 77752; Fax: +49 89 2180 77756;  
Email: markus.mueller@cup.uni-muenchen.de



**Scheme 1.** Reaction of dC and  ${}^{\text{5Me}}\text{dC}$  with hydroxylamine derivatives. Depiction of the tautomeric equilibrium and the interconversion process between the *Z*-isomers and the *E*-isomers. **1**: dC-*O*-allylhydroxylamine—adduct; **2**:  ${}^{\text{5Me}}\text{C}$ -*O*-allylhydroxylamine—adduct.

(*Z*, *A* and *E*) features different H-bonding characteristics. Whereas the amino tautomers (*A*) base pair like dC, the *E*-isomers are expected to code like dT. The *Z*-isomers likely cause stalling of polymerases during replication (Scheme 1) because they interfere with normal base pairing. NMR studies have shown that the pyrimidine oximes like **1** and **2** are present mainly in the imino tautomer and that the *Z*-isomers are the most stable forms (31). In the case of  ${}^{\text{5Me}}\text{dC}$ , we expected that the isomer *E*-2 would be disfavored because of the steric strain imposed between the methyl group and the *O*-allyl chain. The  ${}^{\text{5Me}}\text{dC}$ -adduct **2** should therefore prefer *Z*-configuration (*Z*-2) or exist in the amino tautomeric form (*A*-2). As such, we expected that this product would prefer base pairing with dG. In the case of the dC adduct **1** steric strain with the hydrogen atom at C5 is much lower and we hoped that this adduct would in contrast adopt at least to some extent the *E*-configuration, which would allow base pairing with dATP. The resulting dC to dT transition mutation we thought could be exploited as the desired signal that allows dC to be distinguished from  ${}^{\text{5Me}}\text{dC}$ . As our approach is based on mutagenicity, it is sequence independent and can also detect methylated cytosines that are not in a CpG context as found, for instance, in plant genomes. This is an advantage over restriction enzyme based methods that are limited by the specificity of the enzyme.

## MATERIALS AND METHODS

### Incubation of DNA strands

In a thermocycler (Eppendorf Mastercycler personal) with a heatable lid a single-stranded oligodeoxynucleotide ODN (10  $\mu\text{M}$ ) (Metabion, Martinsried, Germany) was

incubated with  $\text{NH}_2\text{OAllyl}$  (1 M, pH 5.2 with  $\text{HNEt}_2$ , Fluka, Buchs, Switzerland) for 4 h at 60°C. For incubation kinetics 200 pmol of the DNA were directly injected into a HPLC without prior desalting (for detailed incubation kinetics please refer to the Supplementary Data). If the sample was to be sequenced it was desalted using BioSpin 6 Tris columns (BioRad, Hercules, CA, USA) that were previously equilibrated three times with 500  $\mu\text{l}$  ddH<sub>2</sub>O. The filtrate was concentrated in a SpeedVac (Christ RVC-2-33 IR), re-dissolved in  $\text{NH}_3$  (28% in H<sub>2</sub>O, Fluka, Buchs, Switzerland) and heated to 60°C for 90 min in a thermocycler with a heatable lid. The solution was again desalted using BioSpin 6 Tris columns. The concentration of the solution was determined with a NanoDrop photometer (PecqLab ND-1000). The treatment with ammoniumhydroxide is optional but proved to be beneficial for the subsequent sequencing process because it converts the bis-hydroxylamineadduct (Supplementary Data) which is to a small extent formed into the desired compound **1** in base promoted elimination reaction.

### Hybridization of DNA

An amount of 1.2  $\mu\text{M}$  ODN, 1  $\mu\text{M}$  primer and a suitable buffer (NEBuffer 2, NEB, Frankfurt am Main, Germany) were mixed and heated to 95°C for 4 min in a thermocycler with a heatable lid. The solution was allowed to cool at a rate of 1°C per minute and stored at 4°C in the dark.

### Direct sequencing

An amount of 10 pmol dsDNA in 25  $\mu\text{l}$  Annealing Buffer (Qiagen, Hilden, Germany) were sequenced in a PyroMark Q24 Pyrosequencer. All bases except the lesion generated by incubation were sequenced using

standard conditions. Because the incorporation of nucleotides opposite the incubated dC or <sup>5</sup>Me dC is slower than opposite canonical bases, addition of the nucleotide was repeated 10 times. The data was analyzed by the software provided by the manufacturer. The software sums up the peak heights and determines the ratio of dATP $\alpha$ S/dGTP incorporation. The last methylation site cannot be detected, because the last position in a strand generally yields no results in pyrosequencing. Please note that due to the special pyrosequencing procedure dATP $\alpha$ S instead of dATP was used.

### Primer extension based sequencing

An amount of 10 pmol of the incubated sample were hybridized with a biotinylated primer. 0.1 U/ $\mu$ l Klenow(Exo-) polymerase, 500  $\mu$ M dNTPs and 1  $\mu$ M dsDNA in 20  $\mu$ l 1 $\times$  NEB Buffer 2 were incubated for 10–60 min at 60°C. To the solution were added 2  $\mu$ l Streptavidin Sepharose beads (GE Healthcare, Uppsala, Sweden), 40  $\mu$ l Binding Buffer (Qiagen) and 18  $\mu$ l ddH<sub>2</sub>O. After agitation at 1400 rpm for 15 min the beads were captured with a Vacuum Prep Tool (Qiagen), washed with 70% EtOH, 0.1 M NaOH and Washing Buffer (Qiagen). The beads were dissolved in 25  $\mu$ l Annealing Buffer (Qiagen) containing 10 pmol sequencing primer (Metabion). Pyrosequencing was performed on a PyroMark Q24 Pyrosequencer using standard conditions (Qiagen). The data was analyzed by the software provided by the manufacturer. The average dTTP incorporation of all experiments was calculated and the data plotted relative to this value. Negative deviations from average correspond to <sup>5</sup>Me dC, positive values to dC. The data displayed are the average of three measurements.

### Real-time primer extension based quantification

ODN1C and ODN1M were mixed at different ratios, incubated with NH<sub>2</sub>Oallyl and hybridized with ODN2 as described earlier. In a 96-well plate (NUNC, Roskilde, Denmark) were mixed per well: 10 pmol dsDNA (10  $\mu$ l), 0.15 U ATP sulfurylase (0.5  $\mu$ l), 1 U Klenow(exo-) polymerase (0.2  $\mu$ l), 500 pmol APS (1  $\mu$ l), 78.3  $\mu$ l ENLITEN luciferase, luciferin reagent (Promega, Madison, WI, USA). In a 96-well reader (TECAN Genios Pro) at 30°C 10  $\mu$ l dATP $\alpha$ S (20  $\mu$ M) were injected at 200  $\mu$ l/s. The plate was shaken for 5 s (amplitude 5 mm) and the luminescence measured every 250 ms for 400 cycles. The data was collected by the Microsoft Excel Makro provided by the manufacturer. The average of four blank measurements (enzymes without DNA) was subtracted from the average of four measurements of the sample. The data for dATP $\alpha$ S incorporation opposite undamaged dC was taken as baseline. In the first 15–20 s the curves are unstable due to mixing effects, so the data is only shown from second 20 onwards. The slope of the curves shows the speed of incorporation of dATP $\alpha$ S which corresponds to the fraction of dC in the sample.

### Protein purification

The polymerase subunit of DNA polymerase I from *Geobacillus stearothermophilus* (DSM No. 22) was

amplified via PCR and subcloned in pENTR-TEV-D-TOPO (Invitrogen, Karlsbad, USA) to introduce an N-terminal TEV-protease recognition site. The fragment was then transferred to the expression plasmid pDEST007 (32), yielding a cleavable N-terminal Strep-tag II. The protein was expressed in *Escherichia coli* BL21 at 37°C by induction at OD<sub>600</sub> = 1 with 0.2 mg/l anhydrotetracycline for 2 h. Cells were resuspended in 100 mM Tris-Cl pH 7.5, 150 mM NaCl, 10 mM  $\beta$ -ME and Complete protease inhibitor (Roche), lysed by French-press and heated to 50°C for 10 min to denature *E. coli* proteins. 0.1% NP-40 and Tween-20 were added to the lysate prior to heat-treatment and maintained during Strep-tag purification. The cleared lysate was loaded on a Streptactin column (IBA) and eluted with desthiobiotin. Subsequently, the tag was cleaved by incubation on ice with 10 U/mg AcTEV protease (Invitrogen) over night. Further purification via Heparin affinity chromatography and crystallization was performed as described in (33).

### Co-crystallization

For co-crystallization the template containing lesion **1** was annealed to the corresponding primer (for sequences see Supplementary Data) in the protein storage buffer (10 mM Na-cacodylate pH 7, 50 mM NaCl, 0.5 mM EDTA 10 mM MgSO<sub>4</sub>). Prior to crystallization protein and DNA were mixed in a 1 to 3 molar ratio. The final concentration of Pol I and dsDNA were 5 mg/ml and 0.5 mM, respectively. Crystals were grown by mixing an equal volume of protein-DNA complex with 47.5–51.0% (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 3.0–3.5% MPD and 100 mM MES pH 5.8, using the hanging-drop vapor diffusion method. The crystallization plates were incubated at 18°C and crystals appeared after 1–2 days. Crystals were frozen in 24% sucrose, 55% (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 3.0–3.5% MPD, 100 mM MES and stored in liquid nitrogen for data collection. Best crystals were obtained with 49.5% (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and 3% MPD.

### Collection and processing of X-ray diffraction data, phase determination and structure refinement

Data were collected at the beamline PXIII [Swiss Light Source (SLS), Villigen, Switzerland] The data were processed with the programs XDS (34) and SCALA (35, 36). Structure solution was carried out by molecular replacement with Phaser (37) using the coordinates of 1U45 (38). In order to reduce model bias, the temperature factors were reset to 20 for main chain and 40 for side chain and DNA atoms, respectively. Prior to model building in COOT (39) a simulated annealing omit map, removing the area around the lesion, was calculated with PHENIX (40). Restrained refinement was carried out on REFMAC5 (41). Data processing and refinement statistics are summarized in Supplementary Table S1.

## RESULTS

### What are the coding properties of the DNA lesions 1 and 2?

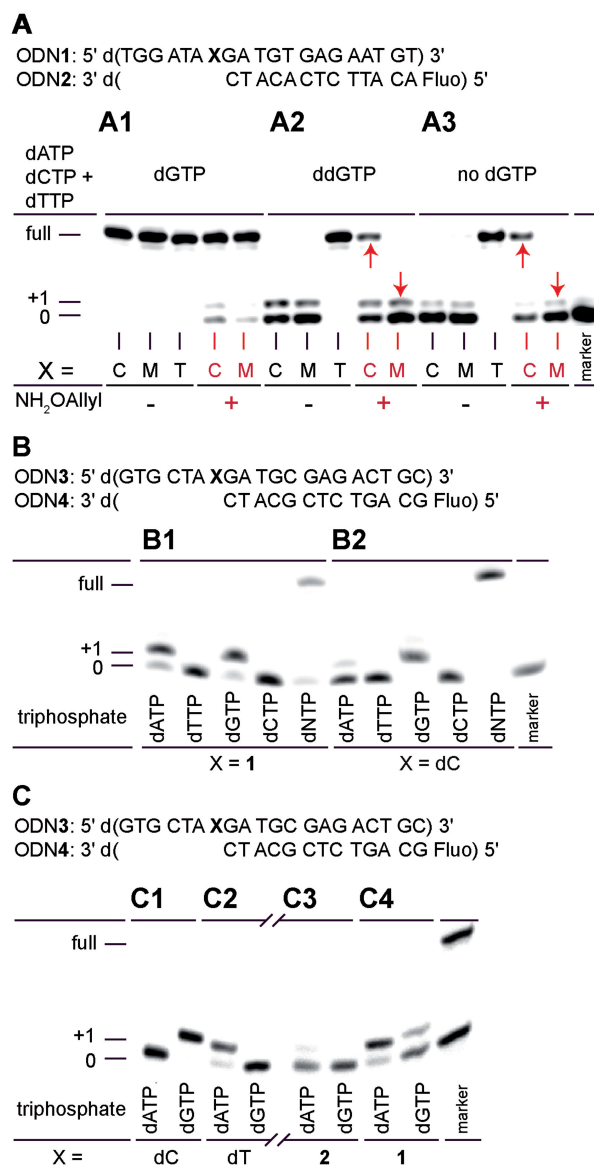
In order to investigate whether the dC-adduct **1** and <sup>5</sup>Me-dC-adduct **2** have different coding properties we reacted the oligodeoxynucleotide ODN1 (Figure 1A) containing either a dC or a <sup>5</sup>Me-dC (ODN1C and ODN1M) with various hydroxylamine derivatives (10 μM DNA, 1 M NH<sub>2</sub>OR, pH = 5.2, 4 h, 60°C). The best results were finally obtained with *O*-allylhydroxylamine, which furnished lesions **1** and **2** cleanly and in high yields (Supplementary Data). We observed a small reactivity difference between dC and <sup>5</sup>Me-dC, with the latter reacting three to five times slower (42). Most importantly, no DNA degradation or reaction with other DNA bases was observed.

To investigate the coding properties of the reaction products **1** and **2** we performed primer extension reactions with the primer ODN2, Klenow(exo-) polymerase and ODN1C or ODN1M (Figure 1A1–3). For both oligonucleotides we observed full extension before and after treatment with *O*-allylhydroxylamine. This proves that the polymerase can indeed read through the hydroxylamine adduct **1**. When dGTP was removed from the primer extension assay (Figure 1A3), we observed that after incubation with *O*-allylhydroxylamine only unmethylated ODN1C but not methylated ODN1M was copied. This shows that adduct **1** is replicatively bypassed and that it is base-paired with a triphosphate other than dGTP. Further studies, depicted in Figure 1B, showed that the polymerase is able to pair **1** with dATP. In order to investigate the relative incorporation efficiency of dATP and dGTP opposite **1** the primer extension studies were repeated with dGTP replaced by ddGTP (Figure 1A2). The data show that Klenow(exo-) pairs **1** with dGTP and dATP with similar efficiencies. Other polymerases tested including low fidelity polymerases such as Pol η and Pol κ showed the same behavior. In the case of ODN1M we observed full extension only when all four triphosphates are present (Figure 1A1). No elongation was detected in the absence of dGTP, showing that adduct **2** either only allows incorporation of dGTP or blocks the polymerase. Further studies, shown in Figure 1C, with ODN3 in which we synthetically inserted the <sup>5</sup>Me-dC adduct **2**, confirmed that bypass of adduct **2** is indeed difficult and not possible with dATP.

In summary, the results show that oxime **1** instructs a polymerase to introduce dG and dA into the primer. Adduct **2** in contrast hinders replication and instructs the polymerase to incorporate a dG. Both adducts **1** and **2** consequently differ dramatically regarding their coding characteristics, and only the adduct derived from dC gives rise to dC to dT transition mutations, providing the requested readout for epigenetic sequencing.

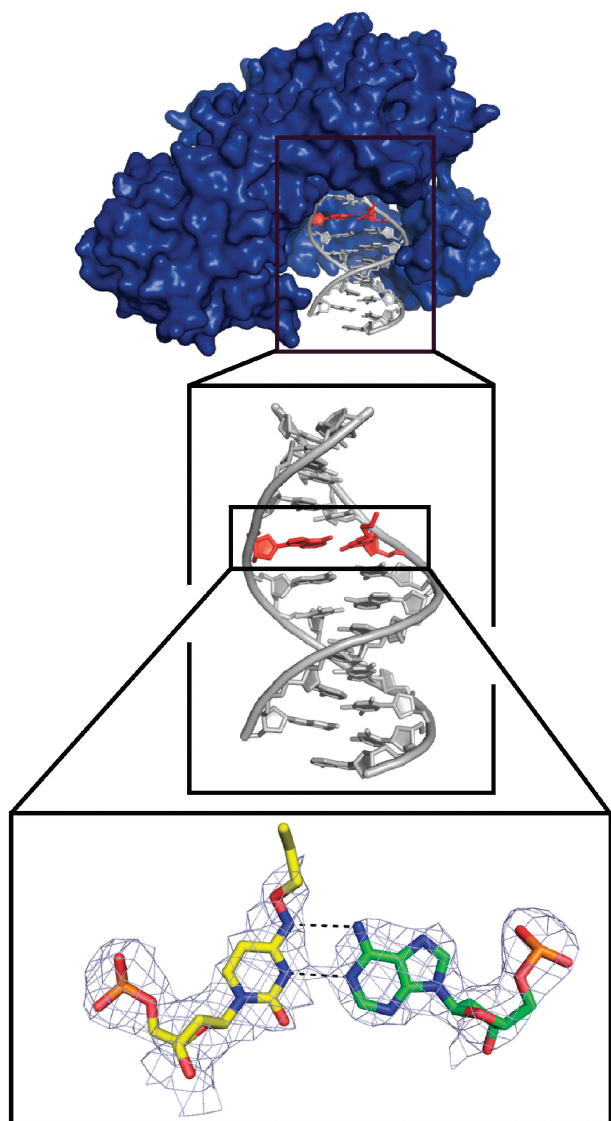
### What is the structural basis of the different coding properties?

In order to prove that the formation of a *E*-1:dA base pair is the reason for the mutational bypass we crystallized a



**Figure 1.** (A) Primer extension reactions of incubated (but not purified) ODN1 with primer ODN2 [1 U Klenow<sup>(exo-)</sup>, 5 μM dNTP, 10 min, 30°C]. ODN1C: X = dC, ODN1M: X = <sup>5</sup>Me-dC, ODN1T: X = dT. (A1) With all triphosphates present all strands are fully elongated before and after treatment. (A2) With ddGTP instead of dGTP present, all strands are elongated by one base. After incubation with *O*-allylhydroxylamine (red) only ODN1C but not ODN1M shows the dC→dT transition mutation and is fully elongated. (A3) Without dGTP again only ODN1C is elongated after *O*-allylhydroxylamine treatment. (B) Primer extension reactions of analytically pure ODN3 with Primer ODN4 [1 U Klenow<sup>(exo-)</sup>, 50 μM dNTP, 30°C, 10 min] of ODN3 with X = dC and X = 1. While dC can only base pair with dGTP, **1** can induce incorporation of dATP, thus generating a C→T transversion mutation. (C) Primer extension of analytically pure ODN3 with X = 1 and X = 2 and primer ODN4 using dATP and dGTP. [1 U Klenow<sup>(exo-)</sup>, 1.5 μM dNTP, 30°C, 10 min]. ODN3 with X = 1 can be elongated with dATP and dGTP, while elongation of the strand containing **2** is not possible.

duplex containing the base pair with the protein B.st Pol I (38). The full structure is depicted Figure 2. It shows the *E*-1:dA base pair determined at a resolution of 2.9 Å. The *O*-allyl-hydroxylamine-dC adduct **1** is indeed present in



**Figure 2.** Crystal structure of BstPolI in complex with a DNA double strand containing a 1:dATP base pair (PDB: 2xo7). 2Fo-Fc electron density for the base pair is depicted at the 1 $\sigma$  level.

the *E*-configuration and it base pairs with dA via two perfect Watson-Crick type H-bonds as expected.

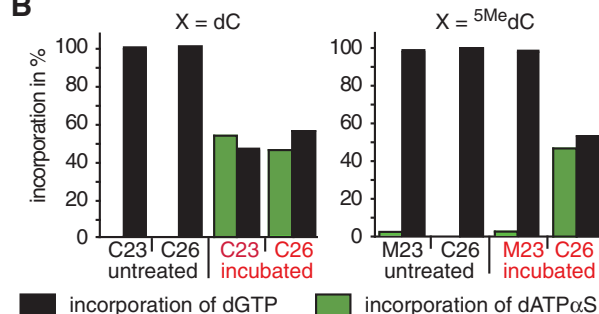
#### Can the coding properties be exploited for a direct readout of the methylation status?

In order to study if the coding difference of the dC and <sup>5</sup>Me dC oxime adducts **1** and **2** can be exploited for epigenetic sequencing, we analyzed the promoter sequence of the cyclin-dependent kinase inhibitor gene *p15*, known to be aberrantly methylated in acute myeloid leukemia patients (43–45). For the study we used ODN5 (Figure 3A) containing two either methylated or unmethylated CpG sites and one dC site that does not neighbor a dG. After incubation of ODN5 with *O*-allylhydroxylamine followed by hybridization with a sequencing primer ODN6 the three DNA constructs were directly subjected to pyrosequencing, which visualizes elongation of the

#### A

ODN5:  
3' d(TGT GAG AAG GGA AGA AAG GGT G<sup>23</sup>XG A<sup>26</sup>CG AGG X) 5'  
5' d(ACA CTC TTC CCT TCT TTC CCA C ) 3'  
ODN6:

#### B



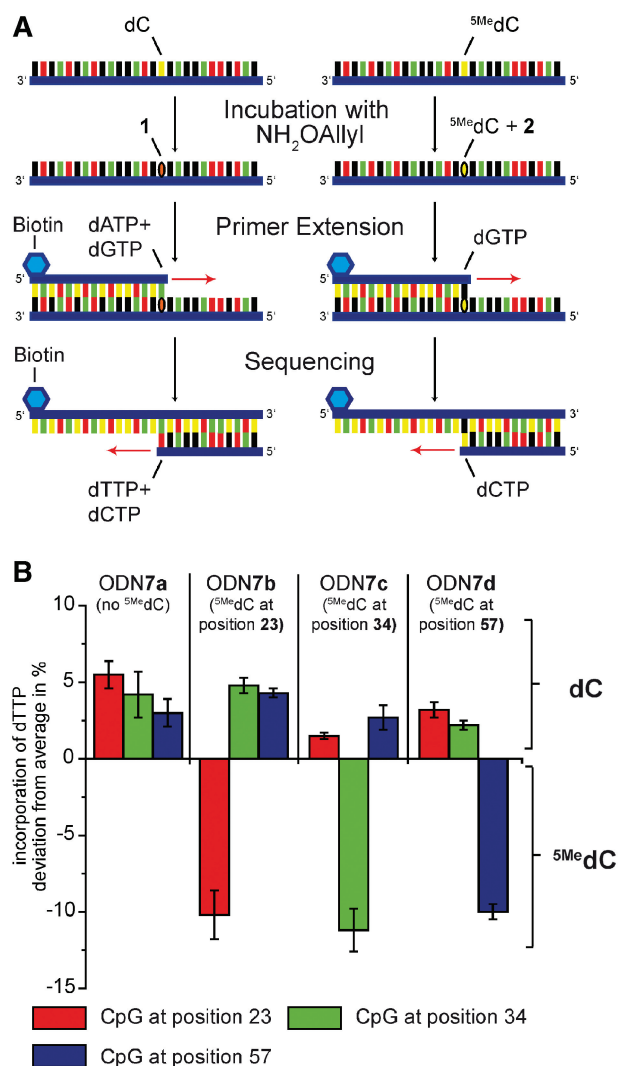
**Figure 3.** (A) Fragment of the *p15* promoter used for direct sequencing. (B) Direct sequencing of ODN5 with primer ODN6; Workflow: Incubation with NH<sub>2</sub>Oallyl converts dC to **1**, <sup>5</sup>Me dC is to a small extent converted to **2**. **1** mispairs with dATP $\alpha$ S, while <sup>5</sup>Me dC pairs with dGTP and **2** blocks. Note that position 31 cannot be sequenced because pyrosequencing does not allow sequencing of the last position.

primer strand (ODN6) by a luminescence signal (46). The results for position 23 and 26 are depicted in Figure 3B. As expected, both dC and <sup>5</sup>Me dC in untreated oligonucleotides direct the addition of dGTP to the primer. The treated DNA strands, in contrast, showed significant incorporation of dATP opposite the dC. Opposite <sup>5</sup>Me dC this misincorporation was in all cases not detected (<5% which is the error range of the instrument, Figure 3B). Important is the finding that the *O*-allylhydroxylamine modified dC base **1** instructs incorporation of ~50% dATP $\alpha$ S regardless of the position of the dC base, showing that no sequence effects occur.

#### Does the novel method allow sequencing of DNA fragments usually employed in pyrosequencing?

We extended the new epigenetic sequencing possibility to the analysis of a longer promoter sequence. For this experiment we choose an 84-bp fragment with three either methylated or unmethylated CpG sites (ODN7, CpG at positions 23, 34 and 57, for sequence see Supplementary Data) of the *p15* promoter. Such a promoter sequence is a typical target of bisulfite sequencing. For the experiment (Figure 4A) we first incubated the different 84-mer oligonucleotides with *O*-allylhydroxylamine and subsequently hybridized the treated strands with the biotinylated primer ODN8, (for sequence see Supplementary Data). We next performed the primer extension reaction with the Klenow(exo-) polymerase and isolated the product DNA strand via streptavidin-coated sepharose beads. Following hybridization with the sequencing primer (ODN9, for sequence see Supplementary Data) the obtained oligonucleotides were finally analyzed via pyrosequencing. The results are depicted in Figure 4B.

The sequencing data clearly furnished a dT signal at all original dC positions in the promoter. The signal was in

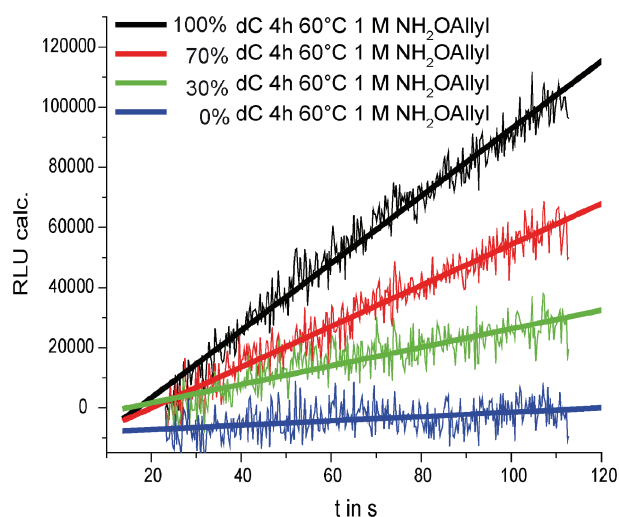


**Figure 4.** Primer extension based sequencing of ODN7 in four different methylation states. (A) Workflow: Incubation with  $\text{NH}_2\text{Oallyl}$  converts dC to **1**,  $^5\text{Me dC}$  is to a small extent converted to **2**. **1** mispairs with dATP, while  $^5\text{Me dC}$  pairs with dGTP and **2** blocks. After primer extension original dC is thus detected by a dT signal, while original  $^5\text{Me dC}$  still only shows a dC signal. (B) Results of the measurements, the deviation of dTTP incorporation from average is shown. Positive values show unmethylated, negative values show methylated cytosines. For sequences of ODN 7–9 see Supplementary Data.

all cases at least 15–20% above background. Again, the same signal intensities are obtained at all dC positions showing again that sequence effects are negligible. The *O*-allylhydroxylamine based epigenetic sequencing method in the reported form is consequently only limited by the pyrosequencing method, which currently limits sequencing to <100- to 150-base long oligonucleotides. As the result provides the same final readout as bisulfite sequencing it is fully compatible with current epigenetic sequencing equipment.

#### Can the ratio of dC to $^5\text{Me dC}$ be determined in a sample?

Generally, the extent of the methylation can be determined by comparing the relative heights of the dT



**Figure 5.** Real time primer extension based quantification of the extent of the methylation in ODN1. The slope of the curves shows the speed of incorporation of dATP $\alpha$ S which corresponds to the fraction of dC in the sample.

and dC signals as done in bisulfite sequencing. Additionally, we aimed to develop an alternative approach that is based on a real time primer extension reaction. By coupling a polymerase to a luciferase, elongation of a DNA strand can be monitored by light. In the case of  $^5\text{Me dC}$  no light upon addition of dATP $\alpha$ S can be detected. If a mixture of dC and  $^5\text{Me dC}$  is present the amount of luminescence represents the fraction of dC in the sample. As shown in Figure 5 the extent of methylation of a specific CpG site can be readily determined after calibration.

#### Can the method distinguish between $^5\text{Me dC}$ and $^5\text{HOMe dC}$ ?

Recently, 5-hydroxymethylcytosine ( $^5\text{HOMe dC}$ ) has been found as a second post-replicatively formed DNA base (47, 48). Newly developed methods allow for precise quantification of  $^5\text{HOMe dC}$  levels in tissues, but do not yield sequence-specific information (49). It is highly desired to have a method that can distinguish  $^5\text{HOMe dC}$  and  $^5\text{Me dC}$  at a single base resolution. Unfortunately, as the detection principle presented in this work is based on sterics and the new base  $^5\text{HOMe dC}$  even imposes a larger steric strain it is not possible to distinguish  $^5\text{Me dC}$  and  $^5\text{HOMe dC}$  after incubation with *O*-allylhydroxylamine.

## CONCLUSION

In conclusion, we show that the *E/Z* equilibrium of *O*-allylhydroxylamine dC/ $^5\text{Me dC}$  adducts is influenced by the presence of a methyl group at C5. The lesion generated from dC after *O*-allylhydroxylamine treatment is bypassed by base pairing with dA and dG. In contrast, after incubation of  $^5\text{Me dC}$  only incorporation of dG can be observed. This allows for the first time detection of the methylation status of a cytosine not only based on reaction rates but based on different reaction products. We used this observation to establish the first bisulfite

independent  $^{5\text{Me}}\text{dC}$  sequencing principle. The selectivity of the method is controlled by two chemical filters. First, dC reacts faster with  $\text{NH}_2\text{Oallyl}$  and secondly, only the dC adduct **1** but not the  $^{5\text{Me}}\text{dC}$  adduct **2** can induce dC to dT transition mutations, which can be detected by standard sequencing methods. As this is the same readout as for the bisulfite method our system is fully compatible with current sequencing equipment. The next focus in our research will be to couple our novel detection principle to modern sequencing-by-synthesis methods that could allow our method to be routinely applied for the methylation analysis of small amounts of DNA. Alternatively one could imagine PCR based approaches that are employed in bisulfite sequencing or to exploit the mutagenic potential of hydroxylamines by subcloning in bacteria.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online. The crystal structure can be accessed under pdb code 2xo7.

## ACKNOWLEDGEMENTS

The authors thank Claudia Szeibert and Andrea Kneutinger for their help during the project, Sabine Schneider for her help solving the structure and the beamline staff at the SLS for technical support. M. Münzel likes to thank the Verband der chemischen Industrie (VCI) for a Kekulé Fellowship.

## FUNDING

SFB 646 and the Center for Integrated Protein Science, Munich (CIPSM); Verband der chemischen Industrie (VCI) for a Kekulé Fellowship. Funding for open access charge: Center for Integrated Protein Science Munich (CIPSM).

*Conflict of interest statement.* None declared.

## REFERENCES

- Bird, A. (2007) Perceptions of epigenetics. *Nature*, **447**, 396–398.
- Suzuki, M.M. and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.*, **9**, 465–476.
- Egger, G., Liang, G., Aparicio, A. and Jones, P.A. (2004) Epigenetics in human disease and prospects for epigenetic therapy. *Nature*, **429**, 457–463.
- Esteller, M. (2008) Epigenetics in cancer. *N. Engl. J. Med.*, **358**, 1148–1159.
- Jones, P.A. and Baylin, S.B. (2002) The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.*, **3**, 415–428.
- Goll, M.G. and Bestor, T.H. (2005) Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.*, **74**, 481–514.
- Bird, A.P. and Southern, E.M. (1978) Use of restriction enzymes to study eukaryotic DNA methylation. *J. Mol. Biol.*, **118**, 27–47.
- Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P. (1994) Purification of CpG islands using a methylated DNA binding column. *Nat. Genet.*, **6**, 236–244.
- Irizarry, R.A., Ladd-Acosta, C., Carvalho, B., Wu, H., Brandenburg, S.A., Jeddloh, J.A., Wen, B. and Feinberg, A.P. (2008) Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res.*, **18**, 780–790.
- Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L. and Schübeler, D. (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.*, **37**, 853–862.
- Hayatsu, H., Wataya, Y., Kai, K. and Iida, S. (1970) Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry*, **9**, 2858–2865.
- Shapiro, R., Servis, R. and Welcher, M. (1970) Reactions of uracil and cytosine derivatives with sodium bisulfite. *J. Am. Chem. Soc.*, **92**, 422–424.
- Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L. and Paul, C.L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl Acad. Sci. USA*, **89**, 1827–1831.
- Ball, M.P., Li, J.B., Gao, Y., Lee, J.H., LeProust, E.M., Park, I.H., Xie, B., Daley, G.Q. and Church, G.M. (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.*, **27**, 361–368.
- Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M. and Jacobsen, S.E. (2008) Shotgun bisulfite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, **452**, 215–219.
- Deng, J., Shoemaker, R., Xie, B., Gore, A., LeProust, E.M., Antosiewicz-Bourget, J., Egli, D., Maherali, N., Park, I.-H., Yu, J. et al. (2009) Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat. Biotechnol.*, **27**, 353–360.
- Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M. et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315–322.
- Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H. and Ecker, J.R. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*, **133**, 523–536.
- Grunau, C., Clark, S.J. and Rosenthal, A. (2001) Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res.*, **29**, e65.
- Wang, R.Y., Gehrke, C.W. and Ehrlich, M. (1980) Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. *Nucleic Acids Res.*, **8**, 4777–4790.
- Warnecke, P.M., Stirzaker, C., Song, J., Grunau, C., Melki, J.R. and Clark, S.J. (2002) Identification and resolution of artifacts in bisulfite sequencing. *Methods*, **27**, 101–107.
- Okamoto, A. (2009) Chemical approach toward efficient DNA methylation analysis. *Org. Biomol. Chem.*, **7**, 21–26.
- Bareyt, S. and Carell, T. (2008) Selective detection of 5-methylcytosine sites in DNA. *Angew. Chem. Int. Ed.*, **47**, 181–184.
- Tanaka, K., Tainaka, K., Kamei, T. and Okamoto, A. (2007) Direct labeling of 5-methylcytosine and its applications. *J. Am. Chem. Soc.*, **129**, 5612–5620.
- Tanaka, K., Tainaka, K., Umemoto, T., Nomura, A. and Okamoto, A. (2007) An osmium-DNA interstrand complex: application to facile DNA methylation analysis. *J. Am. Chem. Soc.*, **129**, 14511–14517.
- Ogino, M., Taya, Y. and Fujimoto, K. (2008) Highly selective detection of 5-methylcytosine using photochemical ligation. *Chem. Commun.*, **45**, 5996–5998.
- Tanabe, K., Yamada, H. and Nishimoto, S.-i. (2007) One-electron photooxidation and site-selective strand cleavage at 5-methylcytosine in DNA by sensitization with 2-methyl-1,4-naphthoquinone-tethered oligonucleotides. *J. Am. Chem. Soc.*, **129**, 8034–8040.
- Oka, Y., Peng, T., Takei, F. and Nakatani, K. (2009) Synthesis and reaction of DNA oligomers containing modified cytosines related to bisulfite sequencing. *Org. Lett.*, **11**, 1377–1379.
- Brown, D.M. and Schell, P. (1965) The reaction of hydroxylamine with cytosine and related compounds. *J. Chem. Soc.*, **47**, 208–215.

30. Freese, E., Bautz, E. and Freese, E.B. (1961) The chemical and mutagenic specificity of hydroxylamine. *Proc. Natl Acad. Sci. USA*, **47**, 845–855.
31. Niedzwiecka-Kornas, A., Kierdaszuk, B., Stolarski, R. and Shugar, D. (1998) Tautomerism, acid-base properties and conformation of methylated analogues of the promutagenic N4-hydroxycytosine. *Biophys. Chem.*, **71**, 87–98.
32. Ober, M., Müller, H., Pieck, C., Gierlich, J. and Carell, T. (2005) Base pairing and replicative processing of the formamidopyrimidine-dG DNA lesion. *J. Am. Chem. Soc.*, **127**, 18143–18149.
33. Kiefer, J.R., Mao, C., Hansen, C.J., Basehore, S.L., Hogrefe, H.H., Braman, J.C. and Beese, L.S. (1997) Crystal structure of a thermostable Bacillus DNA polymerase I large fragment at 2.1 Å resolution. *Structure*, **5**, 95–108.
34. Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.*, **26**, 795–800.
35. CCP4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 760–763.
36. Evans, P. (1997) SCALA. *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography*, **33**, 22–24.
37. McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. and Read, R.J. (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr. D Biol. Crystallogr.*, **61**, 458–464.
38. Hsu, G.W., Ober, M., Carell, T. and Beese, L.S. (2004) Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. *Nature*, **431**, 217–221.
39. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
40. Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.-W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K. and Terwilliger, T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 1948–1954.
41. Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.
42. Janion, C. and Shugar, D. (1965) Mutagenicity of hydroxylamine: reaction with analogues of cytosine, 5(6)-substituted cytosines and some 2-keto-4-ethoxypyrimidines. *Acta Biochim. Pol.*, **12**, 337–355.
43. Dodge, J.E., List, A.F. and Futscher, B.W. (1998) Selective variegated methylation of the p15 CpG island in acute myeloid leukemia. *Int. J. Cancer*, **78**, 561–567.
44. Herman, J.G., Civin, C.I., Issa, J.P., Collector, M.I., Sharkis, S.J. and Baylin, S.B. (1997) Distinct patterns of inactivation of p15INK4B and p16INK4A characterize the major types of hematological malignancies. *Cancer Res.*, **57**, 837–841.
45. Melki, J.R., Vincent, P.C. and Clark, S.J. (1999) Concurrent DNA hypermethylation of multiple genes in acute myeloid leukemia. *Cancer Res.*, **59**, 3730–3740.
46. Ronaghi, M., Uhlen, M. and Nyren, P. (1998) A sequencing method based on real-time pyrophosphate. *Science*, **281**, 363–365.
47. Kriaucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
48. Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
49. Münzel, M., Globisch, D., Brückl, T., Wagner, M., Welzmler, V., Michalakis, S., Müller, M., Biel, M. and Carell, T. (2010) Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed.*, **49**, 5375–5377.