

# Influence of major-groove chemical modifications of DNA on transcription by bacterial RNA polymerases

Veronika Raindlová<sup>1,†</sup>, Martina Janoušková<sup>2,†</sup>, Michaela Slavíčková<sup>1</sup>, Pavla Perlíková<sup>1</sup>, Soňa Boháčová<sup>1</sup>, Nemanja Milisavljevič<sup>1</sup>, Hana Šanderová<sup>2</sup>, Martin Benda<sup>2</sup>, Ivan Barvík<sup>3</sup>, Libor Krásný<sup>2,\*</sup> and Michal Hocek<sup>1,4,\*</sup>

<sup>1</sup>Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Gilead & IOCB Research Center, Flemingovo nam. 2, CZ-16610 Prague 6, Czech Republic, <sup>2</sup>Department of Molecular Genetics of Bacteria, Institute of Microbiology, Academy of Sciences of the Czech Republic, CZ-14220 Prague 4, Czech Republic, <sup>3</sup>Division of Biomolecular Physics, Institute of Physics, Faculty of Mathematics and Physics, Charles University in Prague, Ke Karlovu 5, 121 16 Prague 2, Czech Republic and <sup>4</sup>Department of Organic Chemistry, Faculty of Science, Charles University in Prague, Hlavova 8, CZ-12843 Prague 2, Czech Republic

Received December 03, 2015; Revised March 03, 2016; Accepted March 04, 2016

## ABSTRACT

DNA templates containing a set of base modifications in the major groove (5-substituted pyrimidines or 7-substituted 7-deazapurines bearing H, methyl, vinyl, ethynyl or phenyl groups) were prepared by PCR using the corresponding base-modified 2'-deoxyribonucleoside triphosphates (dNTPs). The modified templates were used in an *in vitro* transcription assay using RNA polymerase from *Bacillus subtilis* and *Escherichia coli*. Some modified nucleobases bearing smaller modifications (H, Me in 7-deazapurines) were perfectly tolerated by both enzymes, whereas bulky modifications (Ph at any nucleobase) and, surprisingly, uracil blocked transcription. Some middle-sized modifications (vinyl or ethynyl) were partly tolerated mostly by the *E. coli* enzyme. In all cases where the transcription proceeded, full length RNA product with correct sequence was obtained indicating that the modifications of the template are not mutagenic and the inhibition is probably at the stage of initiation. The results are promising for the development of bioorthogonal reactions for artificial chemical switching of the transcription.

## INTRODUCTION

Gene expression is regulated on several levels. The primary level is transcription that itself is regulated at various steps. A critical step is the interaction of RNA polymerase (RNAP) and/or transcription factors (TF) with the DNA. This interaction depends on the affinity of the proteins for

specific DNA sequences and the ability of RNAP to form transcription-competent complex. This can be affected by post-replication modifications of nucleobases. Methylation of cytosine (5-methylcytosine, 5mC) in CpG islands in promoter or activator TF-binding sequences typically means switching off the particular genes due to inefficient binding of the protein (RNAP or TF) to the methylated sequence (1,2). More recently, other eukaryotic epigenetic modifications of cytosine, i.e. 5-hydroxymethyl-(5hmC), 5-formyl-(5fC) and 5-carboxycytosine (5caC), have been discovered (3–6) and studied extensively to prove that they are not only intermediates in oxidative demethylation (7–10) but, at least 5hmC is also a stable modification (11) and probably constitutes an independent epigenetic signal with direct impact on transcription. Another recent study showed even the presence of 5-hydroxymethyluracil (5hmU) in DNA in embryonic stem cells (12) but it is yet unclear what the biological function of this modification is. While the impact of DNA methylation on transcription silencing is well established (13,14), the influence of 5hmC, 5fC and 5caC has yet to be studied (15), although first reports on TF (16) or RNAP binding (17) and changes in DNA shape (18) have emerged very recently. Modifications of the DNA are present also in bacteria, namely 5mC as well as *N*<sup>6</sup>-methylation of A, and *N*<sup>4</sup>-methylation of C, which is specifically bacterial (19). Their main known function in bacteria is shielding the DNA against endogenous restriction endonucleases that protect the cell against invading DNA (20,21). Nevertheless, these epigenetic modifications can also affect chromosome replication, DNA-mismatch correction and transcription (22,23). While the number of these natural modifications is quite limited, it is possible to introduce new non-natural modifications, thereby expand-

\*To whom correspondence should be addressed. Tel: +420 220183324; Email: hocek@uochb.cas.cz

Correspondence may also be addressed to Libor Krásný. Tel: +420 241063208; Email: krasny@biomed.cas.cz

<sup>†</sup>These authors contributed equally to the paper as first authors.

ing the epigenetic landscape to offer new possibilities for regulation of gene expression. Though there were several reports (24,25) on transcription of DNA containing artificial base-pairs in the quest for the extension of the genetic alphabet and several reports on inhibition of transcription by DNA damage (26–32), to the best of our knowledge, no systematic study has ever been reported to focus on RNAP transcription from DNA containing non-natural modifications in the major groove (analogous to natural epigenetic modifications).

In contrast to the absence of relevant studies of transcription of non-natural base-modified DNA, recognition and cleavage of major-groove-modified DNA sequences by type II restriction endonucleases (REs) have been studied systematically in our lab. We found that virtually any non-natural modification at 5-position of cytosine (33,34) or at the 7-position of 7-deazaguanine (except for unsubstituted 7-deazaguanine,35) causes a complete block of cleavage, whereas some smaller modifications (ethynyl, vinyl) at 5-position of uracil (analogues of thymine) and at position 7 of 7-deazaadenine (analogues of adenine) are tolerated and the modified DNA is correctly recognized and cleaved at least by some REs (33,36), which suggests that perhaps the G:C pairs are more important for the sequence specific recognition in the major groove than A:T pairs. The fact that the bulky groups at A or U can block the recognition and cleavage whereas some small modifications are tolerated was applied in protection of genes for cloning to plasmid (37) and in transient protection and switching of the DNA cutting by REs using chemically (38) or photochemically (39) removable protecting groups. These promising results prompted us to systematically investigate how major groove modifications affect transcription by bacterial RNAP. Thus here, we report the synthesis of a set of modified DNA templates and the effects of these modifications on transcription by RNAP from two divergent bacterial species. This study thus establishes a foundation for understanding the effects of non-natural epigenetic DNA modifications on gene expression at the transcriptional level.

## MATERIALS AND METHODS

### General remarks – synthesis of nucleotides

NMR spectra were measured on a 500 MHz (499.8 or 500 MHz for  $^1\text{H}$ , 202.3 or 202.4 MHz for  $^{31}\text{P}$ , 125.7 MHz for  $^{13}\text{C}$ ) or a 600 MHz (600.1 MHz for  $^1\text{H}$ , 150.9 MHz for  $^{13}\text{C}$ ) spectrometers in  $\text{D}_2\text{O}$  (referenced to dioxane as internal standard,  $\delta_{\text{H}} = 3.75$  ppm,  $\delta_{\text{C}} = 69.30$  ppm); in  $\text{CD}_3\text{OD}$  (referenced to solvent signal,  $\delta_{\text{H}} = 3.31$  ppm,  $\delta_{\text{C}} = 49.00$  ppm); or  $\text{DMSO-d}_6$  (referenced to solvent signal,  $\delta_{\text{H}} = 2.50$  ppm,  $\delta_{\text{C}} = 39.70$  ppm). Chemical shifts are given in ppm ( $\delta$  scale), coupling constants ( $J$ ) in Hz.  $^{31}\text{P}$  chemical shifts were referenced to  $\text{H}_3\text{PO}_3$  as external reference; or to phosphate buffer signal 2.35 ppm in the case of measurement in phosphate buffer. Complete assignment of all NMR signals was achieved by using a combination of H,H-COSY, H,C-HSQC and H,C-HMBC experiments. Mass spectra and high resolution mass spectra were measured by ESI ionization technique. The MALDI-TOF spectra were measured on a MALDI-TOF/TOF mass spectrometer with 1 kHz smartbeam

II laser. The separation of nucleoside was performed by preparative flash chromatography on reverse phase (C18 RediSep column on a CombiFlash Teledyne ISCO system). Semi-preparative separation of nucleoside triphosphate was performed by HPLC on a column packed with  $10\ \mu\text{m}$  C18 reversed phase (Phenomenex, Luna C18 (2)). Synthesis and characterization data for 5-vinyl-2'-deoxyuridine 5'-*O*-triphosphate ( $\text{dU}^{\text{V}}\text{TP}$ ,34), 5-ethynyl-2'-deoxyuridine 5'-*O*-triphosphate ( $\text{dU}^{\text{E}}\text{TP}$ ,33), 5-phenyl-2'-deoxyuridine 5'-*O*-triphosphate ( $\text{dU}^{\text{Ph}}\text{TP}$ ,33), 5-vinyl-2'-deoxycytidine 5'-*O*-triphosphate ( $\text{dC}^{\text{V}}\text{TP}$ ,34), 5-ethynyl-2'-deoxycytidine 5'-*O*-triphosphate ( $\text{dC}^{\text{E}}\text{TP}$ ,33), 5-phenyl-2'-deoxycytidine 5'-*O*-triphosphate ( $\text{dC}^{\text{Ph}}\text{TP}$ ,33), 7-methyl-2'-deoxy-7-deazaguanosine 5'-*O*-triphosphate ( $\text{dG}^{\text{Me}}\text{TP}$ ,35), 7-vinyl-2'-deoxy-7-deazaguanosine 5'-*O*-triphosphate ( $\text{dG}^{\text{V}}\text{TP}$ ,35), 7-ethynyl-2'-deoxy-7-deazaguanosine 5'-*O*-triphosphate ( $\text{dG}^{\text{E}}\text{TP}$ ,35), 7-phenyl-2'-deoxy-7-deazaguanosine 5'-*O*-triphosphate ( $\text{dG}^{\text{Ph}}\text{TP}$ ,35), 7-vinyl-2'-deoxy-7-deazaadenosine 5'-*O*-triphosphate ( $\text{dA}^{\text{V}}\text{TP}$ ,34), 7-ethynyl-2'-deoxy-7-deazaadenosine 5'-*O*-triphosphate ( $\text{dA}^{\text{E}}\text{TP}$ ,36) and 7-phenyl-2'-deoxy-7-deazaadenosine 5'-*O*-triphosphate ( $\text{dA}^{\text{Ph}}\text{TP}$ ,36), were reported previously.

### Synthesis and characterization of 7-Methyl-2'-deoxy-7-deazaadenosine

7-Iodo-2'-deoxy-7-deazaadenosine (200 mg, 0.53 mmol) and ammonium sulfate (0.053 mmol, 7 mg) were co-evaporated with hexamethyldisilazane at  $40^\circ\text{C}$  (5.3 mmol, 1.2 ml). Argon purged solution of  $\text{Pd}(\text{PPh}_3)_4$  (0.02 mmol, 30 mg) in 4 ml of anhydrous THF was rapidly added to the residue, followed by the addition of  $\text{AlMe}_3$  (2 M solution in hexanes, 6.3 mmol, 3.2 ml). The reaction mixture was refluxed at  $80^\circ\text{C}$  for 2 h. After cooling down to room temperature, the reaction mixture was slowly poured to saturated ice-cold solution of  $\text{NaH}_2\text{PO}_4$ . Resulting suspension was evaporated in vacuo, re-suspended in methanol, co-evaporated with silica gel and purified by high performance reverse phase flash chromatography (0→100% MeOH in water) using C18 RediSep column on a CombiFlash Teledyne ISCO system. Product was isolated as a white solid (75 mg, 55%). Spectral data are in accordance with those reported in literature (40).

### Synthesis and characterization of 7-Methyl-2'-deoxy-7-deazaadenosine-5'-*O*-triphosphate sodium salt

7-Methyl-2'-deoxy-7-deazaadenosine (30 mg, 0.11 mmol) was suspended in trimethyl phosphate (300  $\mu\text{l}$ ) in an argon purged vial and the suspension was cooled to  $0^\circ\text{C}$ . Then,  $\text{POCl}_3$  (13  $\mu\text{l}$ , 0.14 mmol) was added. The reaction mixture was stirred at  $0^\circ\text{C}$  for 1 h and then ice cold solution of  $(\text{NH}_4)_2\text{H}_2\text{P}_2\text{O}_7$  (311 mg, 0.567 mmol) and tributyl amine (110  $\mu\text{l}$ , 0.476 mmol) in anhydrous DMF (1 ml) was added. The reaction mixture was stirred at  $0^\circ\text{C}$  for another hour. Then, aqueous solution of TEAB (2 M, 2 ml, 4 mmol) was added and the mixture was evaporated under reduced pressure. The residue was co-evaporated several times with water. The product was purified by chromatography on DEAE Sephadex column (0→1.2 M aq. TEAB) and then by HPLC (C-18 column, 0.1 M TEAB in water

to 0.1 M TEAB in 50 % aq. MeOH), it was co-evaporated several times with water and converted to a sodium salt form (Dowex 50 in Na<sup>+</sup> cycle). Desired sodium salt of the triphosphate **dA<sup>Me</sup>TP** (5 mg, 8 %) was obtained as a white lyophilizate (water). <sup>1</sup>H NMR (500.0 MHz, D<sub>2</sub>O, ref<sub>dioxane</sub> = 3.75 ppm): 2.28 (s, 3H, CH<sub>3</sub>-5); 2.37 (ddd, 1H, *J*<sub>gem</sub> = 14.0, *J*<sub>2'b,1'</sub> = 6.1, *J*<sub>2'b,3'</sub> = 3.3, H-2'b); 2.55 (ddd, 1H, *J*<sub>gem</sub> = 14.0, *J*<sub>2'a,1'</sub> = 7.9, *J*<sub>2'a,3'</sub> = 5.9, H-2'a); 4.09-4.20 (m, 3H, H-4',5'); 4.68 (dt, 1H, *J*<sub>3',2'</sub> = 5.9, 3.3, *J*<sub>3',4'</sub> = 3.3, H-3'); 6.49 (dd, 1H, *J*<sub>1',2'</sub> = 7.9, 6.1, H-1'); 7.21 (s, 1H, H-6); 8.01 (s, 1H, H-2). <sup>13</sup>C NMR (125.7 MHz, D<sub>2</sub>O, ref<sub>dioxane</sub> = 69.30 ppm): 13.45 (CH<sub>3</sub>-5); 41.25 (CH<sub>2</sub>-2'); 68.16 (d, *J*<sub>C,P</sub> = 5.6, CH<sub>2</sub>-5'); 73.80 (CH-3'); 85.26 (CH-1'); 87.81 (d, *J*<sub>C,P</sub> = 9.0, CH-4'); 104.94 (C-4a); 115.62 (C-5); 122.66 (CH-6); 150.12 (CH-2); 150.94 (C-7a); 157.40 (C-4). <sup>31</sup>P{<sup>1</sup>H} NMR (202.3 MHz, D<sub>2</sub>O): -22.68 (t, *J* = 19.5, P<sub>γ</sub>); -11.15 (d, *J* = 19.5, P<sub>α</sub>); -9.22 (d, *J* = 19.5, P<sub>β</sub>). MS (ESI) *m/z* (%): 343 (45) [M-H<sub>2</sub>P<sub>2</sub>O<sub>6</sub>-H], 423 (100) [M-HPO<sub>3</sub>-H], 445 (45) [M-HPO<sub>3</sub>-2H+Na], 525 (5) [M-2H+Na], 547 (13) [M-3H+2Na], 569 (4) [M-4H+3Na]. HRMS-ESI [M-H]<sup>-</sup> calculated for C<sub>12</sub>H<sub>18</sub>O<sub>12</sub>N<sub>4</sub>P<sub>3</sub>: 503.01395, found: 503.01307.

### General remarks – biochemistry

Synthetic oligonucleotides (all primers, Temp<sup>Pveg50/10</sup>, Temp<sup>FVL-A</sup> and Temp<sup>L50nt</sup>) were purchased from Geneti Biotech. 339-mer template was prepared by PCR with forward and reverse primers from plasmid pRLG7558 (41) containing the Pveg promoter (SI, Section 1.3.). FAM-labeled DNA ladder, used in Gel Shift assay with promoter fragments, was prepared by mixing of **F100 nt** and **F50 nt**, the two short 5'-6-FAM labeled PCR products. Their amplification protocol is described in Supplementary Information Section 3. Vent(exo-) DNA polymerase as well as double-stranded 100 bp DNA Ladder (500 μg/ml) were purchased from New England Biolabs. Enzyme KOD XL DNA polymerase was purchased from Novagen and DyNAzymeII DNA polymerase from Finnzymes (part of Thermo Fisher Scientific). Natural nucleoside triphosphates (dATP, dTTP, dCTP, dGTP) as well as 2'-deoxyuridine 5'-*O*-triphosphate (dUTP) were obtained from Thermo Scientific. 5-Methyl-2'-deoxycytidine 5'-*O*-triphosphate (**dC<sup>Me</sup>TP**) was purchased from New England Biolabs; 2'-deoxy-7-deazapurine 5'-*O*-triphosphates (**dA<sup>7D</sup>TP** and **dG<sup>7D</sup>TP**) from Jena Biosciences.

The PCR products were purified with QIAquick PCR Purification Kit (Qiagen), spin columns in a microcentrifuge. In the case of DNA containing dUTP, E.Z.N.A. Gel Extraction Kit (Omega Bio-tek) was used for the final purification. All purification processes were done under RNase free conditions. All solutions were prepared in MilliQ water. Other chemicals were of analytical grade.

UV spectra (concentration of PCR products) were measured on NanoDrop1000 (Thermo Scientific).

### Agarose gel electrophoresis

PCR products containing 6X DNA loading dye (60 mM EDTA, 10 mM Tris-HCl (pH 7.6), 60 % glycerol, 0.03 % bromophenol blue, 0.03 % xylene cyanol FF, Thermo Scientific) were subjected to horizontal electrophoresis (Owl

EasyCastB, Thermo Scientific) and analyzed on either 1.3 % or 2 % agarose gels (containing 0.5x TBE buffer, pH 8). The gels were run at 118 V for ca. 90–120 min.

PCR products were visualized either with GelRed (Bio-otium, 10 000X in H<sub>2</sub>O) or by fluorescence imaging of 6-FAM-labeled (6-carboxyfluorescein at 5'-end) oligonucleotides using an electronic dual wave transilluminator equipped with GBox iChemi-XRQ Bio imaging system (Syngene, Life Technologies) or Typhoon<sup>TM</sup> FLA 9500 (GE Healthcare Life Sciences), respectively.

### PCR using base-modified dNTPs

Forty PCR cycles were run in PCR cycler, preheated to 80°C, under the following conditions: preheating for 3 min at 94°C, denaturation for 1 min at 94°C, annealing for 1 min at 70°C, extension either for 2 min at 72°C (in the case of all U<sup>X</sup>, A<sup>E</sup>, A<sup>Ph</sup> and C<sup>Me</sup>, C<sup>E</sup>) or 1 min at 72°C (in the case of all G<sup>X</sup>, and A<sup>H</sup>, A<sup>Me</sup>, A<sup>V</sup> and C<sup>V</sup>, C<sup>Ph</sup>), followed by final extension step of 5 min at 72°C. PCR products were analyzed on a 1.3 % agarose gel in 0.5x TBE buffer. The PCR reaction mixture (10 μl) was prepared by mixing of either KOD XL DNA Polymerase (2.5 U/μl, either 0.4 μl for all G<sup>X</sup>, U<sup>V</sup>, U<sup>E</sup>, A<sup>H</sup>, A<sup>Me</sup>, A<sup>E</sup>, A<sup>Ph</sup> and C<sup>Ph</sup>; or 0.5 μl for U<sup>Ph</sup>; or 0.8 μl for A<sup>V</sup> and C<sup>V</sup>); or Vent(exo-) DNA polymerase (2 U/μl, 0.8 μl for C<sup>Me</sup> and C<sup>E</sup>); or DyNAzyme II DNA polymerase (2 U/μl, 0.5 μl for U); natural dNTPs (4 mM, 0.5 μl), functionalized dNTPs (4 mM, 1 μl), primers (20 μM, 1 μl, Prim<sup>FOR</sup> and 20 μM, 1 μl, Prim<sup>REV</sup>) and template Temp<sup>Pveg</sup> (20 ng/μl, 1 μl) in appropriate enzyme reaction buffer (1 μl) supplied by the manufacturer. In the case of U, U<sup>E</sup>, A<sup>V</sup>, A<sup>E</sup>, C<sup>Me</sup> and C<sup>E</sup> also additives had to be inserted to the PCR reaction mixture (DMSO (100 %, 0.5 μl), formamide (5 %, 0.5 μl), betaine (0.75 M, 0.5 μl) and TMAC (50 mM, 0.5 μl)) to obtain full-length PCR products. For more details, see Supplementary Information.

### Sequencing of base-modified PCR products

The correct sequence of all prepared PCR products were confirmed by LIGHTRUN<sup>TM</sup> Sequencing service (GATC Biotech AG, Germany) using standard Sanger sequencing.

The PCR products, including positive controls, were prepared and purified as described above. Moreover, five modified DNAs (U<sup>V</sup>, A<sup>H</sup>, A<sup>Me</sup>, A<sup>Ph</sup>, C<sup>Ph</sup>) were additionally resolved on 2% agarose gel, cut out and purified from the gel with E.Z.N.A. Gel Extraction Kit (Omega Bio-tek). All PCR products were then diluted to the concentration of 40–50 ng/μl and sequenced using both primers Prim<sup>FOR</sup> and Prim<sup>REV</sup>.

### PCR of base-modified dNTPs-promoter fragment

Thirty PCR cycles were run in PCR cycler, preheated to 80°C, under the following conditions: preheating for 3 min at 94°C, denaturation for 1 min at 94°C, annealing for 1 min at 58°C, extension for 1 min at 72°C followed by final extension step of 5 min at 72°C. PCR products were analyzed on a 2% agarose gel in 0.5x TBE buffer. The PCR reaction mixture (10 μl) was prepared by mixing of either Vent(exo-) DNA Polymerase (for A<sup>H</sup>, A<sup>Me</sup>, A<sup>Ph</sup> and C<sup>Me</sup>, 2 U/μl, 0.5



$\mu\text{l}$ ) or DyNzyme II DNA Polymerase (for U, 2 U/ $\mu\text{l}$ , 0.5  $\mu\text{l}$ ), natural dNTPs (4 mM, 0.5  $\mu\text{l}$ ), functionalized dNTPs (4 mM, 1  $\mu\text{l}$ ), primers (10  $\mu\text{M}$ , 1  $\mu\text{l}$ , Prim<sup>FOR-GS</sup>-FAM and 10  $\mu\text{M}$ , 1  $\mu\text{l}$ , Prim<sup>REV-GS</sup>-FAM) and template Temp<sup>Pveg50/10</sup> (1  $\mu\text{M}$ , 0.5  $\mu\text{l}$ ) in appropriate enzyme reaction buffer (1  $\mu\text{l}$ ) supplied by the manufacturer. For more details, see Supplementary Information.

### *In vitro* transcription assays

**Enzymes.** *B. subtilis* RNAP (BsuRNAP) core with His<sup>10</sup>-tag on *rpoC* (RLG 7024 strain) was purified as described previously (42). The transcription factor  $\sigma^A$  was overproduced from the pCD2 plasmid and purified as described (43). *Escherichia coli* RNAP (EcoRNAP) holoenzyme (with  $\sigma^{70}$ ) was purchased from New England Biolabs. The BsuRNAP core was reconstituted with a saturating amount of  $\sigma^A$  for 10 min at 37°C.

**Multiple round transcriptions.** Multiple round transcription assays were performed essentially as described (44), unless stated otherwise. The BsuRNAP experiments were carried out in 10  $\mu\text{l}$  with 5 ng template, 100 mM Tris pH 8, 125 mM MgCl<sub>2</sub>, 50 mM DTT, 50 mM NaCl, 30 nM RNAP holoenzyme and NTPs (200  $\mu\text{M}$  ATP, 1000  $\mu\text{M}$  GTP, 200  $\mu\text{M}$  CTP, 10  $\mu\text{M}$  UTP plus 3.7 kBq [ $\alpha$ -<sup>32</sup>P]UTP). The EcoRNAP experiments were carried out under the same conditions except for the salt (90 mM KCl). RNAP was used diluted in 50 mM Tris-HCl pH 8.0, 0.1 M NaCl, 50 % glycerol. The samples were preheated for 10 min at 37°C. The reaction was allowed to proceed for 10 min at 37°C. Subsequently, the reaction was stopped by the addition of 10  $\mu\text{l}$  of formamide stop solution (95% formamide, 20 mM EDTA, pH 8.0). The samples were loaded onto 7% polyacrylamide gels. The gels were dried and exposed to Fuji MS phosphor storage screens and scanned with a Molecular Imager FX (BIORAD) and analyzed with Quantity One program (BIORAD).

**Single round transcriptions.** The experiments were performed essentially as described in Rabatnova *et al.* (45). Briefly, we selected K<sup>+</sup>, U, C<sup>Me</sup>, C<sup>Ph</sup>, A<sup>E</sup>, G<sup>V</sup> as the templates, used EcoRNAP and kept the enzyme, template, NTP concentrations and buffer composition the same as for multiple round transcriptions. Heparin was used as the competitor at a final concentration of 12.5  $\mu\text{g}/\text{ml}$ . In parallel, we performed competitor test experiments (reactions were initiated with RNAP after heparin had been added to the reaction), and in all cases heparin was able to abolish transcription. The experiments were stopped at 0.5, 1, 2, 4, 8 and 16 min, respectively. Polyacrylamide sequencing gel (9%) was pre-run for 45 min at 1500 V. Then, the samples were loaded onto the gel and resolved for 1 h at 1800 V. The gel was dried and analyzed as described in “Multiple round transcriptions”.

**Sequencing of *in vitro* transcription products.** The DNA templates (K<sup>+</sup>, U, U<sup>E</sup>, C<sup>V</sup>, A<sup>H</sup>, A<sup>Me</sup>, G<sup>H</sup>) were transcribed as described in “Multiple round transcriptions”. The volume of reactions was 5-fold scaled up compared to [ $\alpha$ -<sup>32</sup>P]UTP-labeled multiple round transcriptions. Concentrations

of the components stayed the same. BsuRNAP was used as the enzyme for all the templates; in addition, EcoRNAP was used also for K<sup>+</sup>, U and U<sup>E</sup>. RNA was purified and DNased by RNeasy<sup>®</sup> micro kit (Qiagen). The RNA was reverse transcribed (SuperScript<sup>®</sup> III, Invitrogen; random hexanucleotide primers from Metabion). Then, PCR was performed with Expand High Fidelity PCR System (Roche) and primers 1661 – 5' GCTTGGGTCCCACCTGACCC 3' and 1662 – 5' GAAAGGCCAGTCTTTTCGAC 3' (Metabion). Initial denaturation started at 94°C for 5 min, followed by 1x: [94°C – 15 s, 56°C – 30 s, 72°C – 90 s]. Then, 4x: [94°C – 15 s, 52°C – 30 s, 72°C – 90 s]. Finally, 23x: [94°C – 15 s, 48°C – 30 s, 72°C – 90 s]. Final elongation was carried out at 72°C for 7 min. RNA samples after DNasing but without reverse transcription were used as negative controls. The PCR products were resolved on 1.5% agarose gel. The DNA was cut out and purified from the gel with QIAquick<sup>®</sup> Gel extraction kit, Qiagen. The DNA was then sequenced with primers 1661 and 1662.

### Gel Shift assays

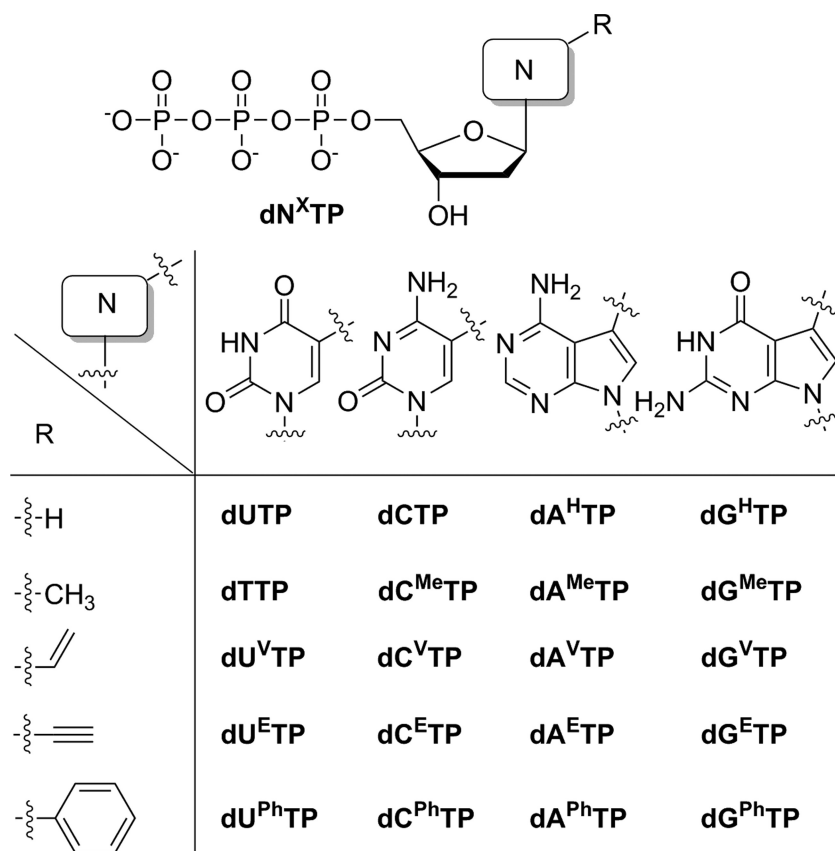
Reconstitution of EcoRNAP and respective fluorescently labeled templates was carried out with 0.25 pmol of the template and 0.023–5.625 pmol EcoRNAP (a dilution series at 3-fold increments) at 30°C, 15 min in 5  $\mu\text{l}$ . DNA template with no modification was used as positive control (K<sup>+</sup>). As a negative control (K<sup>-</sup>, Figure 3B), template with mutated –10 and –35 promoter sequences (Sigma Aldrich) was used. The samples were immediately loaded and resolved on native gradient gels 4–16% (Novex) with ice cold buffer at 150 V for 2 h. The gels were scanned with a Molecular Imager FX (BIORAD) and evaluated by Quantity One program (BIORAD).

## RESULTS

### Synthesis of modified dNTPs

In order to prepare modified DNA templates by PCR (for reviews on enzymatic synthesis of base-modified DNA, see refs. 46–48), we needed a complete set of modified dNTPs bearing substituents of increasing bulkiness (H, methyl, vinyl, ethynyl or phenyl group) at position 5 of pyrimidine or at position 7 of 7-deazapurine nucleotides (Scheme 1). The synthesis of most of them has been reported previously (33–36). The only new derivative was 7-methyl-7-deazadATP (dA<sup>Me</sup>TP) which was prepared by triphosphorylation (49,50) of known 7-methyl-7-deaza-2'-deoxyadenosine. At first we tested polymerase incorporation of the newly synthesised dA<sup>Me</sup>TP in a simple primer extension experiment (PEX) with different polymerases. KOD XL DNA polymerase was the best enzyme which readily and selectively incorporated 7-methyl-7-deazaadenine nucleotide (Supplementary Figure S10) giving the 31-mer full-length product. Apart from PAGE analysis with proper negative control experiment, the formation of methyl-modified oligonucleotide (ON) was also confirmed by MALDI analysis of 31-meric product containing 4 modified A<sup>Me</sup> bases (see Supplementary Information).





**Scheme 1.** Structures of modified  $dN^XTP$ s used in the study.

### Choice of the transcription system

To initiate transcription, RNAP first binds to the promoter DNA to form the closed complex in which the two DNA strands are not yet separated. Subsequently, RNAP isomerizes and forms the open complex, where the transcription bubble is established ( $-11$  to  $+2$ , where  $+1$  is the transcription start site). Upon binding the initiating nucleoside triphosphates (NTPs), the first covalent bond is formed and after the addition of  $\sim 10$ – $15$  nucleotides (nt) RNAP breaks its contacts with the promoter and proceeds to elongation. During the promoter escape stage, short abortive transcripts can be released without RNAP leaving the promoter (51).

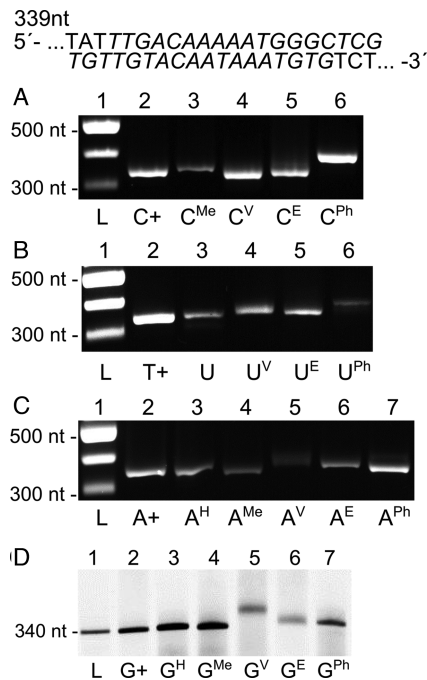
For transcription *in vitro*, RNAP, template DNA with a promoter sequence to initiate transcription, substrates (NTPs) and appropriate buffer milieu are required. As for our transcription system, we selected two bacterial RNAPs from model organisms *Escherichia coli* and *Bacillus subtilis*. These two species are phylogenetically distant (52), and they are extensively studied representatives of Gram-negative and -positive bacteria, respectively. RNAP in both species consists of the core subunits ( $\alpha_2\beta\beta'$   $\omega$ ) (53). In addition, *B. subtilis* RNAP contains two small subunits,  $\delta$  and  $\epsilon$  (54). The RNAP core is capable of transcription elongation but not initiation (55). To initiate, a  $\sigma$  factor is required, depending on the promoter. As our model promoter we selected *Pveg*, a strong constitutive promoter that is well characterized, recognized by both RNAPs and requires RNAP

in complex with the primary  $\sigma$  factor ( $\sigma^{70}$  in *E. coli* and  $\sigma^A$  in *B. subtilis*, respectively) to initiate transcription (56,57). The *veg* gene codes for a protein that was implicated in biofilm formation (58).

### Synthesis of base-modified templates

PCR was the obvious method of choice for the synthesis of the base-modified DNA templates for the transcription studies. We designed a 339-mer template (Temp<sup>Pveg</sup>) containing the promoter sequence that was cloned into the p770 plasmid (41). Transcription from the promoter would end at an intrinsic terminator, yielding  $\sim 145$  long transcripts.

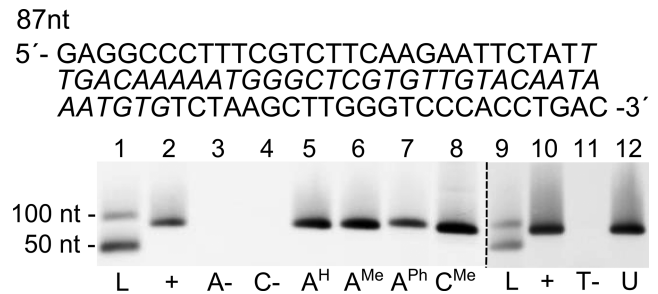
At first we tested the PCR method on the synthesis of a non-modified Temp<sup>Pveg</sup> which was prepared in 40 cycles using natural dNTPs and forward and reverse primers from plasmid containing *Pveg* constitutive promoter fragment using KOD XL DNA Polymerase (Supplementary Information). Base-modified DNA templates, with the use of selected 2'-deoxyribonucleoside triphosphates ( $dN^XTP$ s), *vide supra* Scheme 1, were then prepared analogously. From several tested commercially available DNA polymerases, the best results were obtained with KOD XL DNA polymerase which readily incorporated almost all modified  $dN^XTP$ s giving the corresponding full-length PCR products. In the case of **dC<sup>Me</sup>TP** and **dC<sup>E</sup>TP**, the KOD XL polymerase was not efficient and, therefore, Vent(exo-) DNA polymerase had to be used for the amplification. In the case of **dUTP**,



**Figure 1.** Agarose gel analysis of 339-mer PCR products, promoter region of the template in italics, with modified  $dN^XTPs$ : (A)  $dC^XTPs$ , (B)  $dU^XTPs$ , (C)  $dA^XTPs$  and (D)  $dG^XTPs$ ; Lines 1 (L): ladder (mix of dsDNA with specific number of base pairs); Lines 2 (N+): positive control-four natural dNTPs; Lines 3–7 ( $N^X$ ):  $dN^XTPs$ , X = H, Me, V, E, Ph, plus three natural dNTPs; 1.3% agarose gel stained with GelRed (A–C) or with the use of 6-FAM labeled primers (D).

the most efficient enzyme was DyNAzyme II DNA polymerase. In all cases, full-length PCR products were detected (Figure 1A–D) and the correct sequences were also confirmed by sequencing. Modified  $U^E$ ,  $A^{Me}$  and  $C^E$  DNAs were smoothly amplified with yields comparable to positive controls. Much better amplification proceeded in the case of  $A^H$  and  $C^{Me}$  where the yields were even higher than positive controls. Slightly worse yields were obtained for  $A^E$ ,  $A^{Ph}$ ,  $C^{Ph}$ ,  $U^{Ph}$ ,  $U^V$  and all  $G^Xs$  where also prolongation of time of extension from 1 to 2 min was helpful to form PCR products in sufficient yields. In the case of  $A^V$ ,  $C^V$  and  $U$ , the amplification was the most problematic and the volume of reactions had to be scaled up by  $\sim 1.5$ -fold, and also, for  $U$ , the prolongation of time of extension to 2 min was required to obtain the modified DNA template in sufficient amount for the transcription experiments.

All these PCR products were visualized by agarose gel electrophoreses with the use of GelRed as intercalating agent (Figure 1A–D), except PCR products containing modified  $dG^XTPs$ . Since 7-deazaguanosine moieties are known to quench fluorescence of intercalators (59), we used 5'-6-FAM labeled primers with fluorescence detection to analyze  $dG^XTPs$  PCR products (Figure 1D). The electrophoretic mobility of the PCR products containing  $C^{Ph}$ ,  $U^{Ph}$ ,  $A^V$  and  $G^V$  was somewhat slower compared to the positive control and other modified DNAs (Figure 1A and B lanes 6 and Figure 1C and D lanes 5) which may have been caused by increased size or some interactions of the modifications with agarose. However, also in these cases the se-



**Figure 2.** Agarose gel analysis of 87-mer PCR products used in Gel-Shift assay, promoter region of the template in italics, with modified  $dN^XTPs$ : N = A, C, U; Lines 1, 9 (L): ladder (mix of dsDNA with specific number of base pairs); Lines 2, 10 (+): positive control-four natural dNTPs; Lines 3, 4, 11 (N-): one dNTP missing; Lines 5–8, 12 ( $N^X$ ):  $dN^XTPs$ , X = H, Me, Ph, plus three natural dNTPs; 2% agarose gel with the use of 5'-6-FAM labeled primers.

quencing clearly showed the correct length and sequence of these amplification products. For more details of the PCR experiments and further gel electrophoresis images containing negative controls, see Supplementary Information.

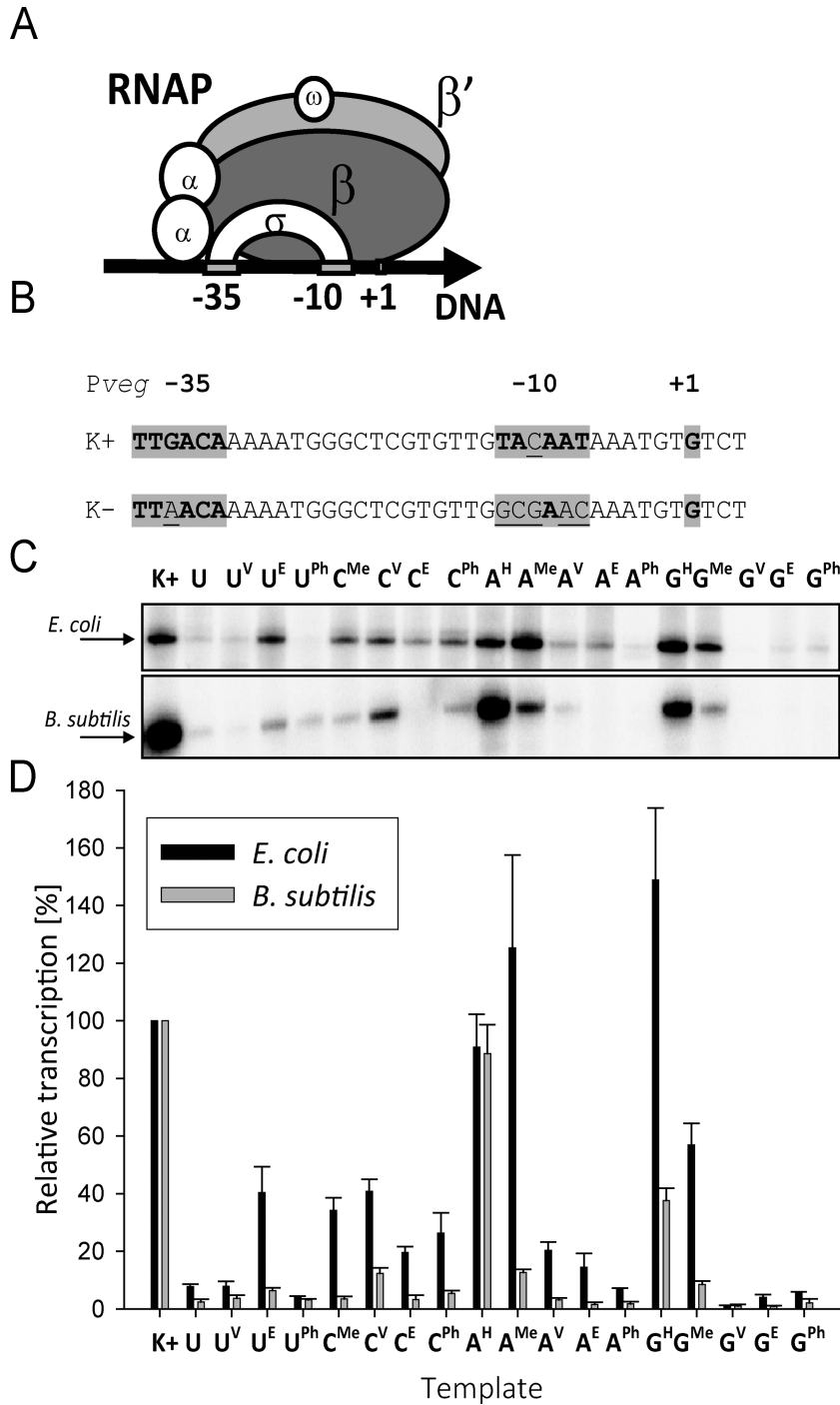
### Synthesis of short base-modified templates (EMSA)

In order to study the influence of the base-modification on the binding of the RNAP to the DNA template, we designed short a 87-mer template ( $Temp^{Pveg50/10}$ ) containing the promoter region. Five selected representative modified nucleoside triphosphates,  $dA^HTP$ ,  $dA^{Me}TP$ ,  $dA^{Ph}TP$ ,  $dC^{Me}TP$  and  $dUTP$ , were then used for the PCR synthesis of the modified  $Temp^{Pveg50/10}$  templates by PCR. Amplification of the 87-mer proceeded smoothly in 30 cycles for all five tested  $dN^XTPs$ . Vent(exo-) proved to be the best polymerase for  $dA^HTP$ ,  $dA^{Me}TP$ ,  $dA^{Ph}TP$  and  $dC^{Me}TP$  substrates, whereas  $dU$ -modified PCR product was efficiently amplified by DyNAzyme II DNA polymerase (Figure 2). Furthermore, to avoid working with radioactivity, 5'-6-FAM labeled primers were used for the agarose gel analysis of all base-modified 87-mer PCR products. For more details, see Supplementary Information.

### Transcription on modified DNA

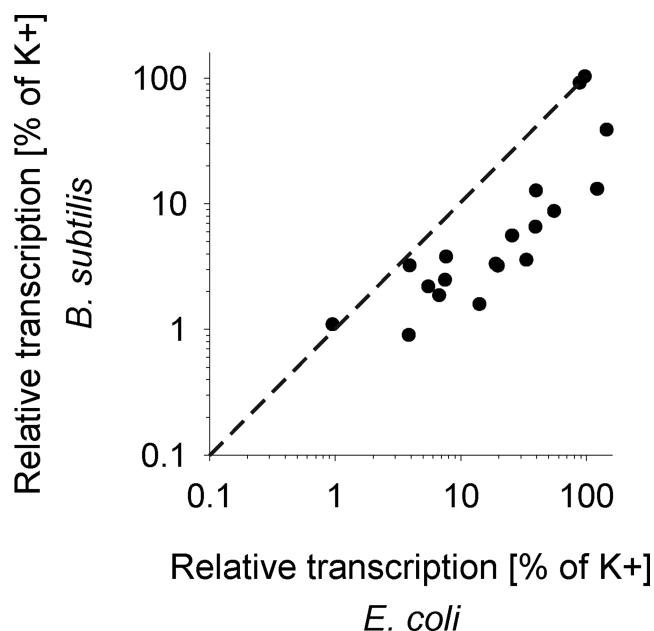
Having all the base-modified DNA templates in hand, we performed multiple round transcriptions with RNAPs from *E. coli* (EcoRNAP) and *B. subtilis* (BsuRNAP), respectively (Figure 3). The two enzymes (EcoRNAP and BsuRNAP) showed similar trends with respect to their ability to transcribe the various templates. While the trends were similar, transcription with EcoRNAP resulted in higher yields of transcripts than transcription with BsuRNAP (Figure 4).

The presence of the bulkiest phenyl modification resulted in virtually complete inhibition of transcription with the exception of  $C^{Ph}$ . Ethynyl and vinyl groups showed similar inhibition with EcoRNAP and BsuRNAP, but the degree of inhibition depended on the identity of the base. The most pronounced inhibition was achieved when these modifications were positioned on 7-deazaguanine; less pronounced effects were observed for A and C. Interestingly,  $U^E$  dis-



**Figure 3.** Effect of DNA modifications on transcription. (A) A cartoon of RNAP and its interactions with DNA. -35 and -10 regions, and the transcription start site (+1) are indicated. The arrow shows the direction of transcription. (B) Promoter sequences used in the study. K+, the transcriptionally active Pveg sequence with +1 G (wt Pveg has an A at +1; nevertheless, the activity/regulation of the promoter are not affected by this difference [60]). K-, a mutated Pveg promoter sequence - negative control for experiments addressing the binding of RNAP to DNA. -35, -10, +1 are highlighted in grey. Bases identical with consensus are in bold. Differences from consensus are underlined. (C) Representative results from *in vitro* multiple round transcriptions with *E. coli* and *B. subtilis* RNAPs. The transcripts (145 nt in length) are indicated with arrows. The modifications are indicated. (D) Quantitation of transcription results. The bars show the mean, the error bars +SD (-SD was the same, and it was omitted for clarity). Experiments with *B. subtilis* RNAP were repeated five times, experiments with *E. coli* RNAP three times. The modifications are indicated.





**Figure 4.** Correlation between transcription with BsuRNAP and EcoRNAP. Each axis shows relative transcription expressed as % of **K+** (set as 100 for each RNAP). Each dot represents transcription from one type of template with the two RNAPs. The dashed diagonal line shows where the dots should align if the transcription of each template would be identically efficient with both BsuRNAP and EcoRNAP enzymes.

played a higher activity than **U<sup>V</sup>** and **U**. Most interestingly, **U** almost abolished transcription with both RNAPs.

In the gels, we did not observe any prominent shorter transcripts than the full length RNA (145 nt) which indicated that the transcription of the studied modified templates was not blocked during elongation. Nevertheless, single round transcriptions performed as a function of time could be more sensitive to detect paused or blocked RNAPs (26). Therefore, we performed these experiments with the control (**K+**) and selected modified templates (**U**, **U<sup>E</sup>**, **C<sup>Me</sup>**, **C<sup>Ph</sup>**, **A<sup>E</sup>**, **G<sup>V</sup>**). Transcription on the wt (**K+**) template was the strongest, correlating with the multiple round results (Figure 5A and B). Importantly, transcription on the modified templates including the one with the bulkiest modification (**C<sup>Ph</sup>**) did not lead to transiently paused RNAPs. Likewise, the elongation rate did not seem to be negatively affected. This suggested that (i) the modifications affected transcription mainly at the initiation stage, and, therefore, (ii) if positioned only within the promoter, some of the *in vitro* tested modifications may not be targeted by the transcription-coupled nucleotide repair system *in vivo* that acts on stalled RNAPs during transcription as reported previously (27, 31). To assess whether transcription from the modified templates could have perhaps caused misincorporations of NTPs into the RNA, we sequenced reverse transcribed RNA generated from **K+**, **U**, **U<sup>E</sup>**, **C<sup>V</sup>**, **A<sup>H</sup>**, **A<sup>Me</sup>**, **G<sup>H</sup>**. In neither case we observed any deviation from the wt (**K+**) sequence, including **U** and **U<sup>E</sup>** as these substitute the **Ts** and, therefore, do not change the coding.

**Table 1.**  $K_d$  values for selected templates and EcoRNAP

Template	$K_d$ [nM]*	$\pm$ SD
<b>K+</b>	30.6	3.0
<b>A<sup>H</sup></b>	35.5	2.9
<b>A<sup>Me</sup></b>	54.3	4.4
<b>A<sup>Ph</sup></b>	35.8	5.0
<b>C<sup>Me</sup></b>	102.8	13.5
<b>U</b>	617.5 <sup>†</sup>	57.5 <sup>†</sup>
<b>K-</b>	>1125.0 <sup>†</sup>	n.d.

\* $K_d$  represents the mean value calculated from at least three independent experiments.

<sup>†</sup>These  $K_d$ s are the means from two experiments and the error shows the range. n. d., not determined.

### Binding of RNAP to modified DNA templates

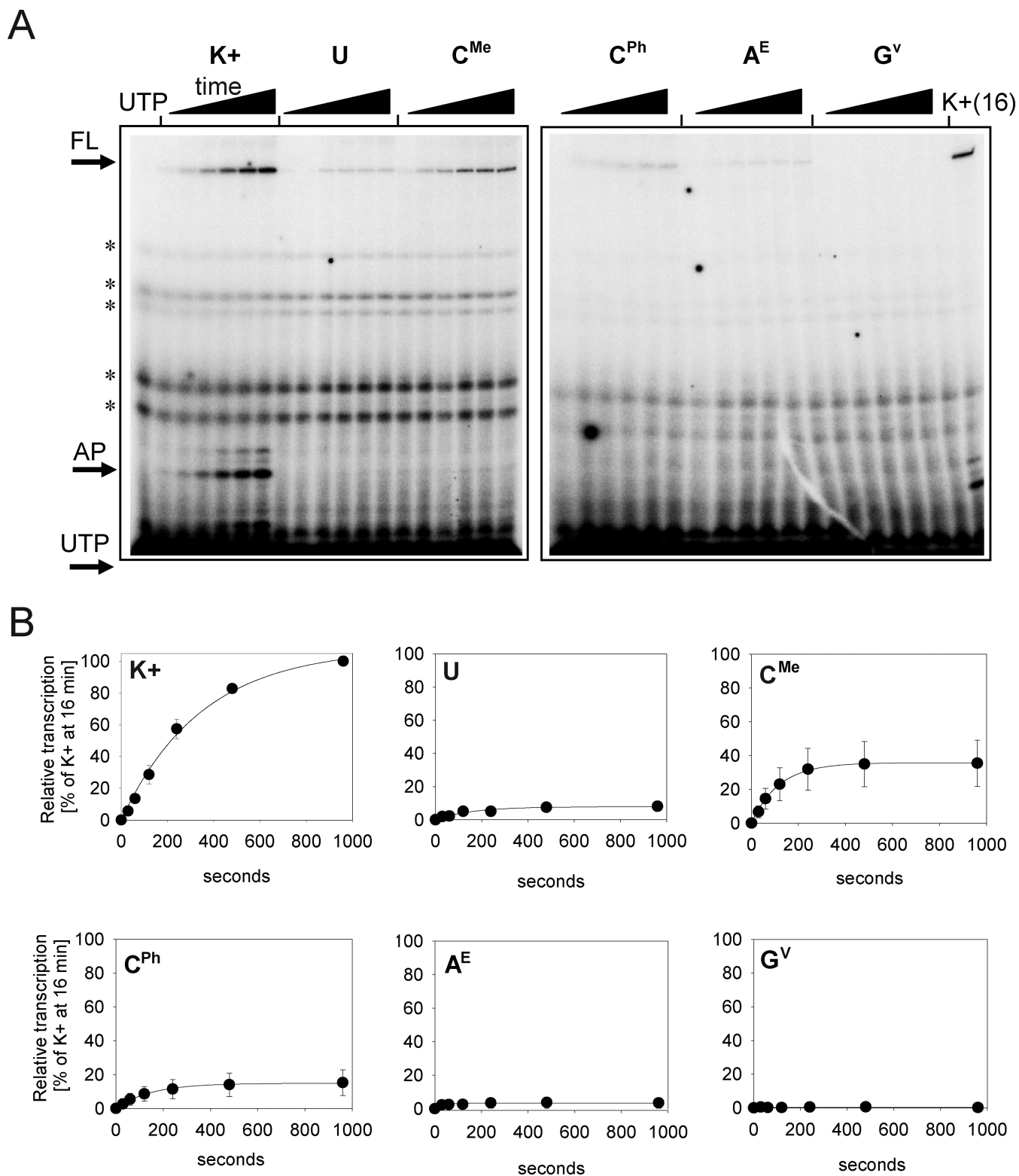
Based on the results of transcription, we asked how the modifications affected formation of complexes between RNAP and selected modified templates. We performed electrophoretic mobility shift assays (EMSA) with selected templates (**K+**, **K-** [mutated *Pveg* without base modifications; Figure 3B], **A<sup>H</sup>**, **A<sup>Me</sup>**, **A<sup>Ph</sup>**, **C<sup>Me</sup>** and **U**) and EcoRNAP to determine the respective  $K_d$ s (Table 1). The non-modified template (**K+**) and **A<sup>H</sup>** displayed similar  $K_d$ s, and the  $K_d$  of **A<sup>Me</sup>** was only slightly higher, which is in accord with the ability of RNAP to transcribe these templates efficiently. Unexpectedly, **A<sup>Ph</sup>** (which completely inhibits transcription) showed similarly low value of  $K_d$  as **A<sup>H</sup>** indicating that it also binds to the template with high affinity (see Discussion *vide infra*).

Both **C<sup>Me</sup>** and **U** displayed much lower affinities for RNAP, consistent with their adverse effect on transcription. The *Pveg* promoter fragment with mutations in the consensus hexamers and bearing no artificial base modifications (**K-**) displayed the lowest affinity for RNAP – in fact, the binding was so weak that the  $K_d$  was out of range of the tested RNAP concentrations (Table 1).

### DISCUSSION

We have synthesized and evaluated a complete set of templates containing non-natural nucleobase derivatives modified in the major groove by substituents of increasing bulkiness (H, Me, vinyl, ethynyl, phenyl) at the position 5 of pyrimidines or at the position 7 of 7-deazapurines. The non-natural modifications were designed for testing the structural requirements of RNAP in complex with the main sigma factor to recognize the promoter sequence and to read through the modified template during transcription. The goal was to find some (presumably smaller) modifications which would be fully tolerated by the RNAP and allow transcription, and some (presumably bulkier) modifications which would completely block transcription. Apart from deeper understanding of the sequence-specific recognition of DNA by these proteins, these findings might pave the way to the development of biorthogonal reactions for switching (ON and/or OFF) the transcription.

The corresponding base-modified dNTPs were good substrates for DNA polymerases and were efficiently used for PCR synthesis of the modified templates. In these templates, both strands were fully modified with the exception of the primer parts which were non-modified. For example, when



**Figure 5.** Single round *in vitro* transcription assays on K<sup>+</sup>, U, C<sup>Me</sup>, C<sup>Ph</sup>, A<sup>E</sup> and G<sup>V</sup> modified templates. (A) Representative results from *in vitro* single round transcriptions with *E. coli* RNAP as a function of time. Respective symbols for DNA modifications are shown above the time arrows. The time course was 0.5, 1, 2, 4, 8 and 16 min, and it is indicated with the triangle. UTP, [ $\alpha$ -<sup>32</sup>P] UTP only was run in this lane. Non-specific signals originated from [ $\alpha$ -<sup>32</sup>P] UTP are indicated with asterisks. K<sup>+</sup>(16) is single round transcription with K<sup>+</sup> from the 16 min time point. The full length transcripts (FL), abortive products (AP) and unincorporated UTP are indicated with arrows. (B) Quantitation of full length transcription signals normalized to K<sup>+</sup> at 16 min (set as 100%). The DNA modifications are indicated. The averages were calculated from three independent experiments and the error bars show  $\pm$ SD.

using modified **dA<sup>X</sup>TP**, all adenines in both strands (except for primers) were replaced by the **A<sup>X</sup>** nucleobase. This inherently means, that if the full length transcript is generated, RNAP has recognized the sequence and the RNAP was able to transcribe through the modified template.

The systematic transcription study of all the modified templates revealed several important findings. As expected, some modifications were tolerated and the modified templates gave efficient transcription, whereas some modified templates did not give any transcription. The absence of prominent bands shorter than the full length transcript suggested that the modifications acted already at the transcription initiation stage. Interestingly, several shorter transcripts generated from **K+** appeared near the bottom of the gel (Figure 5A). These were likely products of abortive transcription (AP). This was consistent with the relatively long time it took RNAP on **K+** to reach the plateau: RNAP struggled to sever its contacts with the promoter. These bands disappeared in transcriptions with modified DNA, suggesting that RNAP was not limited in promoter escape on these templates. Despite using two phylogenetically distant polymerases (EcoRNAP and BsuRNAP), both displayed similar trends with respect to their abilities to transcribe the modified templates (Figure 4). This is consistent with the structural-functional conservation of RNAPs across prokaryotes. The architecture of the active site is conserved also in eukaryotes (61,62). Nevertheless, EcoRNAP was generally more efficient, reflecting its ability to form more stable complexes with the DNA than BsuRNAP (63). Importantly, the modifications appeared to have no mutagenic effect with respect to misincorporation of NTPs as proved by sequencing of several transcription products.

### The effect of U and C modifications

It was previously reported that dihydrouracil was efficiently bypassed by *E. coli* RNAP when positioned within the transcribed region (26). However, transcription of the **dU**-containing template was severely hindered. A likely reason is the importance of Ts within the  $-10$  hexamer (consensus 5'-T<sub>-12</sub>A<sub>-11</sub>T<sub>-10</sub>A<sub>-9</sub>A<sub>-8</sub>T<sub>-7</sub>-3') of the non-transcribed strand of DNA that interacts with sigma region 2. Crystal structures show a lock-and-key fit, where the A<sub>-11</sub> and T<sub>-7</sub> are flipped out of the base stack, with their bases buried in pockets on the protein surface (64). Only an A base can fit in the A<sub>-11</sub> pocket, and the T base is specifically recognized in the T<sub>-7</sub> pocket. The intervening bases of T<sub>-10</sub>A<sub>-9</sub>A<sub>-8</sub> remain stacked, directed away from the protein surface. The methyl group of T<sub>-7</sub> is oriented toward the A<sub>-9</sub>A<sub>-8</sub> phosphate group (65–67) and sterically blocks thermal movements of the phosphate group toward the  $-7$  base. This space remains unoccupied in the **dU**-containing DNA (see the modeled structure in Figure 6). This likely causes disturbance of the recognition between the modified DNA and RNAP and contributes to the decreased transcription from the template.

Compared to the **dU**-template, the addition of ethynyl at **dU** was able to partially restore the transcription activity of RNAP. The presence of ethynyl group at **dU<sup>E</sup>** probably mimics the methyl group of the natural T base. On the other hand, the presence of a vinyl group or bulky phenyl

group at **dU** completely blocked the transcription. Interestingly, the presence of **C<sup>Me</sup>** (the only naturally occurring base from the studied modifications) was partially tolerated by EcoRNAP, whereas with BsuRNAP the transcription was blocked. Surprisingly, vinyl derivative **C<sup>V</sup>** was tolerated by both enzymes, whereas the ethynyl and phenyl derivatives **C<sup>E</sup>** and **C<sup>Ph</sup>** inhibited the *E. coli* transcription and **C<sup>E</sup>** fully blocked BsuRNAP. Previously, it was shown that 5-propynyl cytosine (**C<sup>P</sup>**) blocked transcription when positioned within the transcribed region (28). In contrast, **C<sup>Ph</sup>**, while significantly decreasing the full-length transcription, did not result in a ladder of shorter transcripts indicating blocked RNAP at each modification (Figure 5A). A possible explanation could be the 20-fold higher competitor concentration used in the **C<sup>P</sup>** experiment: heparin destabilizes RNAP–DNA complexes and may contribute to premature termination. In our case, we did not use a higher heparin concentration as it interfered with RNAP–promoter DNA complex formation and resulted in no transcription. Nevertheless, some other modifications, such as 8-oxoguanine, were reported to cause stalling of mammalian RNA polymerase II during elongation (29) and affected promoter clearance with EcoRNAP when positioned within the early transcribed region (28). Yet another modification, 5-iodocytosine, negatively affected transcription from the *E. coli trp* promoter (30).

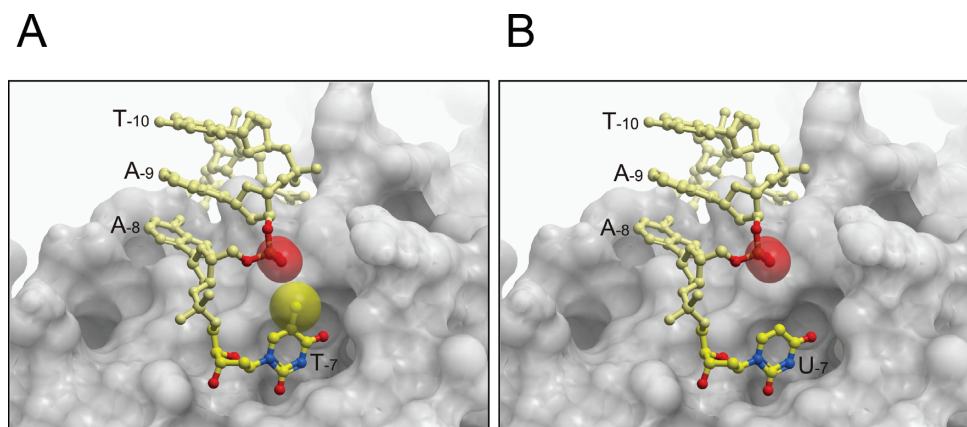
### The effect of A and G modifications

Modifications of A resulted generally in less pronounced effects than modifications of Gs. Both 7-deazaA (**A<sup>H</sup>**) and 7-deazaG (**G<sup>H</sup>**) were fully tolerated indicating that the N7 nitrogen is not crucial for the sequence recognition by RNAP. Also 7-methyl-7-deazapurines (**A<sup>Me</sup>** and **G<sup>Me</sup>**) gave significant transcription, especially with the EcoRNAP. **A<sup>V</sup>**, **A<sup>E</sup>** and **A<sup>Ph</sup>** gave some residual transcription with the *E. coli* (but not *B. subtilis*) enzyme, whereas the bulkier G-modifications blocked both enzymes.

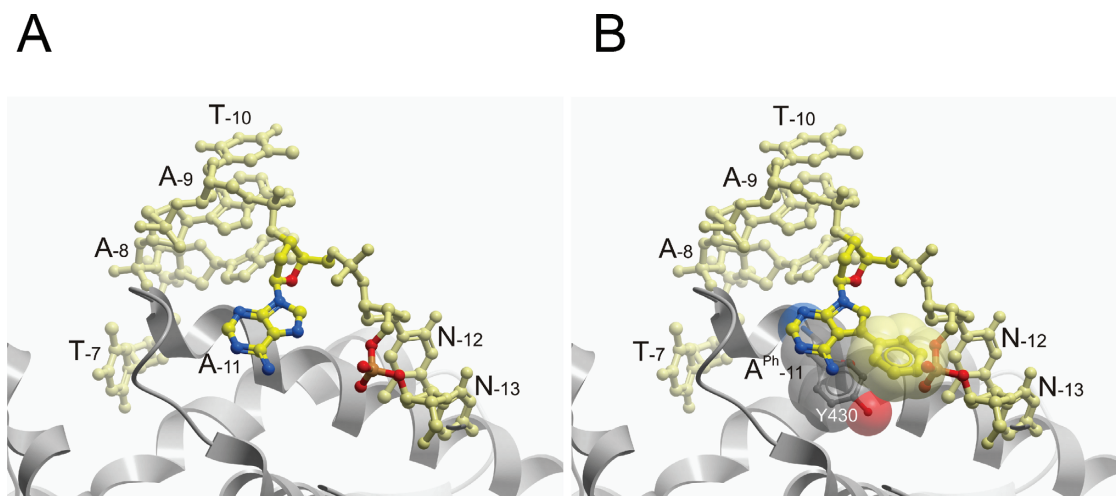
K<sub>d</sub>s determined for selected A-modified templates correlated with transcription activity except for **A<sup>Ph</sup>**, which, surprisingly, formed complexes with EcoRNAP with low K<sub>d</sub>. A<sub>-11</sub>Ph from the  $-10$  hexamer in the non-template DNA should be able to bind into a pocket on the protein surface. Even stabilizing stacking interactions between the phenyl moiety and Y430 could be envisioned. Nevertheless, the phenyl group would protrude into the space that is normally occupied by the phosphodiester linkage connecting important nucleotides N<sub>-12</sub> and N<sub>-13</sub> of the non-template strand (see the modeled structure in Figure 7). These nucleotides are the first ones that base-pair at the upstream end of the transcription bubble. This suggests that while **A<sup>Ph</sup>** does not decrease the ability of the template to interact with RNAP, steps such as melting the DNA and/or formation of the transcription competent complex are likely affected.

Guanine is important within the  $-35$  element (consensus T<sub>-35</sub>T<sub>-34</sub>G<sub>-33</sub>A<sub>-32</sub>C<sub>-31</sub>A<sub>-30</sub>), and this was reflected in almost complete inhibition of transcription on templates containing bulkier G-modifications (**G<sup>V</sup>**, **G<sup>E</sup>**, **G<sup>Ph</sup>**). On the other hand, the transcription of templates containing 7-deazaguanine (**G<sup>H</sup>**) was very efficient and the *E. coli* RNAP also efficiently transcribed template containing **G<sup>Me</sup>**. Crys-





**Figure 6.** Interaction of the non-template strand with the sigma subunit of RNAP: T<sub>-7</sub>. (A) Segment of the crystal structure (PDB 4yln, 66). The methyl group of T<sub>-7</sub> is indicated with the yellow van der Waals sphere. The red spheres indicate the oxygens of the A<sub>-8</sub>A<sub>-9</sub> phosphate group. The DNA is shown in light yellow. The sigma subunit is shown in grey. (B) Modeled modified DNA containing U<sub>-7</sub> instead of T<sub>-7</sub>.



**Figure 7.** Interaction of the non-template strand with the sigma subunit of RNAP: A<sub>-11</sub>. (A) Segment of the crystal structure (PDB 4yln, 66). The A<sub>-11</sub> base is labeled and shown in bright colors. The phosphodiester linkage between N<sub>-12</sub> and N<sub>-13</sub> is also highlighted. The DNA is in light yellow. The sigma subunit is shown in grey. (B) Modeled modified DNA containing A<sup>Ph</sup><sub>-11</sub> instead of A<sub>-11</sub> clashes with the N<sub>-12</sub> and N<sub>-13</sub> phosphodiester linkage. The phenyl group at A<sub>-11</sub> is shown including the van der Waals spheres of individual atoms. Tyrosine 430 that can stack with the phenyl group is shown in dark grey, the van der Waals sphere of the Tyr phenyl oxygen is shown in red. A<sub>-8</sub> and A<sub>-9</sub> are facing away from the protein and so the phenyl groups were not modeled there.

tal structure (66) shows that *E. coli* Arg584 in sigma region 4.2 donates two hydrogen bonds to the O6 and N7 acceptors of -31'G. Our results with efficient transcription of G<sup>H</sup>-containing templates indicate that hydrogen bonding to O6 may be sufficient for recognition whereas the interaction with N7 is not crucial. A substitution of the corresponding position in *Thermus aquaticus* sigma (Arg409Cys) decreased expression from the wild-type *lac* promoter. Accordingly, N7 methylation at this position by dimethyl sulphate was reported (68) to have a negative effect on promoter binding by RNAP and transcriptional activity. We observed a similar inhibition of transcription of template containing 7-methyl-7-deazaguanine (G<sup>Me</sup>) by *B. subtilis* RNAP (but not by *E. coli* RNAP).

## CONCLUSIONS

We synthesized a complete set of base-modified dNTPs bearing modifications of diverse bulkiness in the major

groove and used them as templates for enzymatic construction of the corresponding base-modified templates by PCR. The non-natural major-groove modifications were designed to study the structural requirements of the transcription factor and RNAP in recognition of the DNA template and the ability of the RNAP to read through the modified template during transcription. This study has brought several important findings. The major-groove modifications are either tolerated or inhibit transcription but they do not cause significant formation of shorter transcripts or mutagenesis, which means that they primarily influence transcription initiation rather than subsequent steps. The replacement of the N7 in purines by a carbon is always well tolerated indicating that this nitrogen is not crucial as H-bond acceptor in the interactions with RNAP. The presence of U (instead of T) in the template almost completely blocks transcription due to the decreased binding of RNAP and suboptimal interactions of U during initiation. Both

in pyrimidines and in deazapurines, some smaller modifications (Me, ethynyl, vinyl) are at least partly tolerated. This gives us a good chance for the design and development of biorthogonal chemical transformations that may convert some more bulky substituents into these small (tolerated) ones and vice versa, and ultimately develop artificial chemical switches for transcription. Studies in this direction are under way.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

Academy of Sciences of the Czech Republic [RVO: 61388963 and Premium Academiae for M.H.]; Czech Science Foundation [14-04289S to M. J., M. S., S. B., H. Š., L. K. and M. H.]; Gilead Sciences, Inc. (Foster City, CA, U.S.A.). Part of the project was also supported by the C4Sys infrastructure. Funding for open access charge: Research grant of the Czech Science Foundation [14-04289S].

*Conflict of interest statement.* None declared.

## REFERENCES

- Jeltsch, A. (2002) Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *ChemBioChem*, **3**, 274–293.
- Fu, Y. and He, C. (2012) Nucleic acid modifications with epigenetic significance. *Curr. Opin. Chem. Biol.*, **16**, 516–524.
- 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.
- Kriaucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. *Science*, **324**, 929–930.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C. and Zhang, Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300–1303.
- Münzel, M., Globisch, D. and Carell, T. (2011) 5-Hydroxymethylcytosine, the Sixth Base of the Genome. *Angew. Chem. Int. Ed.*, **50**, 6460–6468.
- Song, C.-X. and He, C. (2013) Potential functional roles of DNA demethylation intermediates. *Trends Biochem. Sci.*, **38**, 480–484.
- Lu, X., Han, D., Zhao, B.S., Song, C.-X., Zhang, L.-S., Doré, L.C. and He, C. (2015) Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics. *Cell Res.*, **25**, 386–389.
- Liutkevičiūtė, Z., Kriukienė, E., Ličytė, J., Rudytė, M., Urbanavičiūtė, G. and Klimišauskas, S. (2014) Direct decarboxylation of 5-Carboxylcytosine by DNA C5- Methyltransferases. *J. Am. Chem. Soc.*, **136**, 5884–5887.
- Schiesser, S., Pfaffeneder, T., Sadeghian, K., Hackner, B., Steigenberger, B., Schröder, A.S., Steinbacher, J., Kashiwazaki, G., Höfner, G., Wanner, K.T. *et al.* (2013) Deamination, oxidation, and C-C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine. *J. Am. Chem. Soc.*, **135**, 14593–14599.
- Bachman, M., Uribe-Lewis, S., Yang, X., Burgess, H.E., Iurlaro, M., Reik, W., Murrell, A. and Balasubramanian, S. (2015) 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.*, **11**, 555–557.
- Pfaffeneder, T., Spada, F., Wagner, M., Brandmayr, C., Laube, S.K., Eisen, D., Truss, M., Steinbacher, J., Hackner, B., Kotljarova, O. *et al.* (2014) Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.*, **10**, 574–581.
- Kass, S.U., Pruss, D. and Wolffe, A.P. (1997) How does DNA methylation repress transcription? *Trends Genet.*, **13**, 444–449.
- Law, J.A. and Jacobsen, S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.*, **11**, 204–220.
- Schröder, A., Steinbacher, J., Steinberger, B., Gnerlich, F.A., Schiesser, S., Pfaffeneder, T. and Carell, T. (2014) Synthesis of a DNA promoter segment containing all four epigenetic nucleosides: 5-Methyl-, 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxy-2'-deoxycytidine. *Angew. Chem. Int. Ed.*, **53**, 315–318.
- Lercher, L., McDonough, M.A., El-Sagheer, A.H., Thalhammer, A., Kriaucionis, S., Brown, T. and Schofield, C.J. (2014) Structural insights into how 5-hydroxymethylation influences transcription factor binding. *Chem. Commun.*, **50**, 1794–1796.
- Wang, L., Zhou, Y., Xu, L., Xiao, R., Lu, X., Chen, L., Chong, J., Li, H., He, C., Fu, X.-D. *et al.* (2015) Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature*, **523**, 621–625.
- Raiber, E.-A., Murat, P., Chirgadze, D.Y., Beraldi, D., Luisi, B.F. and Balasubramanian, S. (2015) 5-Formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.*, **22**, 44–49.
- Jeltsch, A. (2002) Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA Methyltransferases. *ChemBioChem*, **3**, 274–293.
- Loenen, W.A.M., Dryden, D.T.F., Raleigh, E.A., Wilson, G.G. and Murray, N.E. (2014) *Nucleic Acids Res.*, **42**, 2–19.
- Bickle, T.A. and Kruger, D.H. (1993) Biology of DNA Restriction. *Microbiol. Rev.*, **57**, 434–450.
- Sanchez-Romero, M.A., Cota, I. and Csadesús, J. (2015) DNA methylation in bacteria: from the methyl group of the methylome. *Curr. Opin. Microbiol.*, **25**, 9–16.
- Reisenauer, A. and Shapiro, L. (2002) DNA methylation affects the cell cycle transcription of the CtrA global regulator in *Caulobacter*. *EMBO J.*, **21**, 4969–4977.
- Seo, Y.J., Matsuda, S. and Romesberg, F.E. (2009) Transcription of an expanded genetic alphabet. *J. Am. Chem. Soc.*, **131**, 5046–5047.
- Ishizuka, T., Kimoto, M., Sato, A. and Hirao, I. (2012) Site-specific functionalization of RNA molecules by an unnatural base pair transcription system via click chemistry. *Chem. Commun.*, **48**, 10835–10837.
- Liu, J. and Doetsch, P. W. (1998) *Escherichia coli* RNA and DNA polymerase bypass of dihydrouracil: mutagenic potential via transcription and replication. *Nucleic Acids Res.*, **26**, 1707–1712.
- You, C., Wang, J., Dai, X. and Wang, Y. (2015) Transcriptional inhibition and mutagenesis induced by N-nitroso compound-derived carboxymethylated thymidine adducts in DNA. *Nucleic Acids Res.*, **43**, 1012–1018.
- Viswanathan, A. and Doetsch, P. W. (1998) Effects of nonbulky DNA base damages on *Escherichia coli* RNA polymerase-mediated elongation and promoter clearance. *J. Biol. Chem.*, **273**, 21276–21281.
- Kuraoka, I., Endou, M., Yamaguchi, Y., Wada, T., Handa, H. and Tanaka, K. (2003) Effects of endogenous DNA base lesions on transcription elongation by mammalian RNA polymerase II. Implications for transcription-coupled DNA repair and transcriptional mutagenesis. *J. Biol. Chem.*, **278**, 7294–7299.
- Farnham, P.J. and Platt, T. (1982) Effects of DNA base analogs on transcription termination at the tryptophan operon attenuator of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 998–1002.
- Kitsera, N., Stathis, D., Lühnsdorf, B., Müller, H., Carell, T., Epe, B. and Khobta, A. (2011) 8-Oxo-7, 8-dihydroguanine in DNA does not constitute a barrier to transcription, but is converted into transcription-blocking damage by OGG1. *Nucleic Acids Res.*, **39**, 5926–5934.
- You, C. and Wang, Y. (2015) Quantitative measurement of transcriptional inhibition and mutagenesis induced by site-specifically incorporated DNA lesions *in vitro* and *in vivo*. *Nat. Protoc.*, **10**, 1389–1406.
- Macíková-Cahová, H., Pohl, R. and Hocek, M. (2011) Cleavage of functionalized DNA containing 5-modified pyrimidines by Type II restriction endonucleases. *ChemBioChem*, **12**, 431–438.
- Mačková, M., Pohl, R. and Hocek, M. (2014) Polymerase synthesis of DNA bearing vinyl groups in major groove and their cleavage by restriction endonucleases. *ChemBioChem*, **15**, 2306–2312.

35. Mačková, M., Boháčová, S., Perlíková, P., Poštová Slavětínská, L. and Hocek, M. (2015) Polymerase synthesis and restriction enzyme cleavage of DNA containing 7-substituted 7-deazaguanines. *ChemBioChem*, **16**, 2225–2236.
36. Macíčková-Cahová, H. and Hocek, M. (2009) Cleavage of adenine-modified functionalized DNA by type II restriction endonucleases. *Nucleic Acids Res.*, **37**, 7612–7622.
37. Kielkowski, P., Brock, N.L., Dickschat, J.S. and Hocek, M. (2013) Nucleobase protection strategy for gene cloning and expression. *ChemBioChem*, **14**, 801–804.
38. Kielkowski, P., Macíčková-Cahová, H., Pohl, R. and Hocek, M. (2011) Transient and switchable (triethylsilyl)ethynyl protection of DNA against cleavage by restriction endonucleases. *Angew. Chem. Int. Ed.*, **50**, 8727–8730.
39. Vaníková, Z. and Hocek, M. (2014) Polymerase synthesis of photocaged DNA resistant against cleavage by restriction endonucleases. *Angew. Chem. Int. Ed.*, **53**, 6734–6737.
40. Seela, F. and Thomas, H. (1994) Synthesis of certain 5-Substituted 2'-deoxytubercidin derivatives. *Helv. Chim. Acta*, **77**, 897–903.
41. Krásný, L. and Gourse, R.L. (2004) An alternative strategy for bacterial ribosome synthesis: *Bacillus subtilis* rRNA transcription regulation. *EMBO J.*, **23**, 4473–4483.
42. Anthony, L.C., Artsimovitch, I., Svetlov, V., Landick, R. and Burgess, R.R. (2000) Rapid purification of His(6)-tagged *Bacillus subtilis* core RNA polymerase. *Protein Expr. Purif.*, **19**, 350–354.
43. Chang, B.Y. and Doi, R.H. (1990) Overproduction, purification, and characterization of *Bacillus subtilis* RNA polymerase sigma A factor. *J. Bacteriol.*, **172**, 3257–3263.
44. Wiedermannová, J., Sudzinová, P., Kovař, T., Rabatinová, A., Šanderová, H., Ramaniuk, O., Rittich, Š., Dohnálek, J., Zhihui, F., Halada, P. et al. (2014) Characterization of HeID, an interacting partner of RNA polymerase from *Bacillus subtilis*. *Nucleic Acids Res.*, **42**, 5151–5163.
45. Rabatinová, A., Šanderová, H., Jiráť Matějčková, J., Korelusová, J., Sojka, L., Barvík, I., Papoušková, V., Sklenář, V., Židek, L. and Krásný, L. (2013) The  $\delta$  subunit of RNA polymerase is required for rapid changes in gene expression and competitive fitness of the cell. *J. Bacteriol.*, **195**, 2603–2611.
46. Hocek, M. and Fojta, M. (2008) Cross-coupling reactions of nucleoside triphosphates followed by polymerase incorporation. Construction and applications of base-functionalized nucleic acids. *Org. Biomol. Chem.*, **6**, 2233–2241.
47. Hollenstein, M. (2012) Nucleoside triphosphates - building blocks for the modification of nucleic acids. *Molecules*, **17**, 13569–13591.
48. Hocek, M. (2014) Synthesis of base-modified 2'-deoxyribonucleoside triphosphates and their use in enzymatic synthesis of modified DNA for applications in bioanalysis and chemical biology. *J. Org. Chem.*, **79**, 9914–9921.
49. Ludwig, J. (1981) A new route to nucleoside 5'-triphosphates. *Acta Biochim. Biophys. Acad. Sci. Hung.*, **16**, 131–133.
50. Kovacs, T. and Otvos, L. (1988) Simple synthesis of 5-vinyl and 5-ethynyl-2'-deoxyuridine-5'-triphosphates. *Tetrahedron Lett.*, **29**, 4525–4528.
51. Ruff, E.F., Drennan, A.C., Capp, M.W., Poulos, M.A., Artsimovitch, I. and Record, M.T. Jr (2015) *E. coli* RNA polymerase determinants of open complex lifetime and structure. *J. Mol. Biol.*, **427**, 2435–2450.
52. Bralley, P., Chang, S.A. and Jones, G.H. (2005) A phylogeny of bacterial RNA nucleotidyltransferases: *Bacillus halodurans* contains two tRNA nucleotidyltransferases. *J. Bacteriol.*, **187**, 5927–5936.
53. Murakami, K.S. (2015) Structural biology of bacterial RNA polymerase. *Biomolecules*, **5**, 848–862.
54. Weiss, A. and Shaw, L.N. (2015) Small things considered: the small accessory subunits of RNA polymerase in Gram-positive bacteria. *FEMS Microbiol. Rev.*, **39**, 541–554.
55. Ruff, E.F., Record, M.T. Jr and Artsimovitch, I. (2015) Initial events in bacterial transcription initiation. *Biomolecules*, **5**, 1035–1062.
56. Sojka, L., Kouba, T., Barvík, I., Šanderová, H., Maderová, Z., Jonák, J. and Krásný, L. (2011) Rapid changes in gene expression: DNA determinants of promoter regulation by the concentration of the transcription initiating NTP in *Bacillus subtilis*. *Nucleic Acids Res.*, **39**, 4598–4611.
57. Fukushima, T., Ishikawa, S., Yamamoto, H., Ogasawara, N. and Sekiguchi, J. (2003) Transcriptional, functional and cytochemical analyses of the veg gene in *Bacillus subtilis*. *J. Biochem.*, **133**, 475–483.
58. Lei, Y., Oshima, T., Ogasawara, N. and Ishikawa, S. (2013) Functional analysis of the protein Veg, which stimulates biofilm formation in *Bacillus subtilis*. *J. Bacteriol.*, **195**, 1697–1705.
59. Měnová, P., Dziuba, D., Güixens-Gallardo, P., Jurkiewicz, P., Hof, M. and Hocek, M. (2015) Fluorescence quenching in oligonucleotides containing 7-substituted 7-deazaguanine bases prepared by the Nicking Enzyme Amplification Reaction. *Bioconjugate Chem.*, **26**, 361–366.
60. Krásný, L., Tišerová, H., Jonák, J., Rejman, D. and Šanderová, H. (2008) The identity of the transcription +1 position is crucial for changes in gene expression in response to amino acid starvation in *Bacillus subtilis*. *Mol. Microbiol.*, **69**, 42–54.
61. Lane, W.J. and Darst, S.A. (2010) Molecular evolution of multisubunit RNA polymerases: sequence analysis. *J. Mol. Biol.*, **395**, 671–685.
62. Lane, W.J. and Darst, S.A. (2010) Molecular evolution of multisubunit RNA polymerases: structural analysis. *J. Mol. Biol.*, **395**, 686–704.
63. Whipple, F.W. and Sonenshein, A.L. (1992) Mechanism of initiation of transcription by *Bacillus subtilis* RNA polymerase at several promoters. *J. Mol. Biol.*, **223**, 399–414.
64. Liu, X., Bushnell, D.A. and Kornberg, R.D. (2011) Lock and key to transcription: sigma-DNA Interaction. *Cell*, **147**, 1218–1219.
65. Feklistov, A. and Darst, S.A. (2011) Structural basis for promoter -10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell*, **147**, 1257–1269.
66. Zuo, Y. and Steitz, T.A. (2015) Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. *Mol. Cell*, **58**, 534–540.
67. Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R. and Darst, S.A. (2015) Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *eLife*, **4**, e08504.
68. Campbell, E.A., Muzzin, O., Chlenov, M., Sun, J.L., Olson, C.A., Weinman, O., Trester-Zedlitz, M.L. and Darst, S.A. (2002) Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol. Cell*, **9**, 527–539.