



HAL
open science

Identification of acylation products in SHAPE Chemistry

Chaoqi Lin, Salomé Poyer, Loussiné Zargarian, Jean-Yves Salpin, Philippe Fossé, Olivier Mauffret, Juan Xie

► **To cite this version:**

Chaoqi Lin, Salomé Poyer, Loussiné Zargarian, Jean-Yves Salpin, Philippe Fossé, et al.. Identification of acylation products in SHAPE Chemistry. *Bioorganic and Medicinal Chemistry Letters*, 2017, 27 (11), pp.2506-2509. 10.1016/j.bmcl.2017.03.096 . hal-01637313

HAL Id: hal-01637313

<https://hal.science/hal-01637313>

Submitted on 4 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

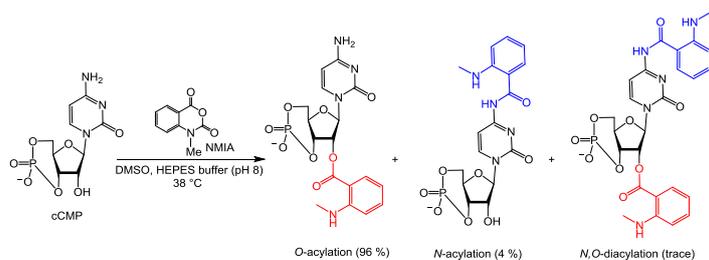
Graphical Abstract

To create your abstract, type over the instructions in the template box below.
Fonts or abstract dimensions should not be changed or altered.

Identification of acylation products in SHAPE Chemistry

Leave this area blank for abstract info.

Chaoqi Lin, Salomé Poyer, Loussiné Zargarian, Jean-Yves Salpin, Philippe Fossé, Olivier Mauffret, and Juan Xie



Identification of acylation products in SHAPE Chemistry

Chaoqi Lin^a, Salomé Poyer^c, Loussiné Zargarian^b, Jean-Yves Salpin^c, Philippe Fosse^b, Olivier Mauffret^b, and Juan Xie^{a*}

^a PPSM, CNRS, Institut d'Alembert, ENS Paris-Saclay, Université Paris-Saclay, 61 Avenue du P^r Wilson, F-94235 Cachan, France

^b LBPA, CNRS, Institut d'Alembert, ENS Paris-Saclay, Université Paris-Saclay, 61 Avenue du P^r Wilson, F-94235 Cachan, France

^c LAMBE, CNRS, Université d'Evry Val d'Essonne, CEA, Université Paris-Saclay, F-91025 Evry, France

ARTICLE INFO

Article history:

Received

Revised

Accepted

Available online

Keywords:

Acylation reaction

cNMP

LC-UV-MS²

NMR

RNA

SHAPE chemistry

ABSTRACT

SHAPE chemistry (selective 2'-hydroxyl acylation analyzed by primer extension) has been developed to specifically target flexible nucleotides (often unpaired nucleotides) independently to their purine or pyrimidine nature for RNA secondary structure determination. However, to the best of our knowledge, the structure of 2'-*O*-acylation products has never been confirmed by NMR or X-ray data. We have realized the acylation reactions between cNMP and NMIA under SHAPE chemistry conditions and identified the acylation products using standard NMR spectroscopy and LC-MS/MS experiments. For cAMP and cGMP, the major acylation product is the 2'-*O*-acylated compound (> 99 %). A trace amount of *N*-acylated cAMP has also been identified by LC-UV-MS². While for cCMP, the isolated acylation products are composed of 96 % of 2'-*O*-acylated, 4 % of *N,O*-diacylated, and trace amount of *N*-acylated compounds. In addition, the characterization of the major 2'-*O*-acylated compound by NMR showed slight differences in the conformation of the acylated sugar between the three cyclic nucleotides. This interesting result should be useful to explain some unexpected reactivity of the SHAPE chemistry.

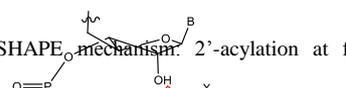
RNAs play essential roles in many fundamental cellular processes such as gene regulation and host defense. Long non-coding RNAs are key actors in development, epigenetics, cancer, brain function and hereditary diseases,¹⁻³ while small RNAs and the untranslated regions of mRNAs regulate cellular processes such as transcription and translation.⁴ The diverse functions of RNAs are closely linked to their structures. The secondary structure of an RNA molecule (the pattern of base pairing) dictates the tertiary folding. A RNA function may depend on sequence, secondary and tertiary structures, or a combination of these features. Secondary structures are usually considered as the key point for RNA functions.

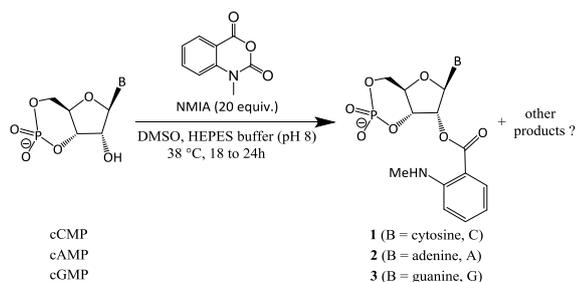
A wide range of chemicals can be used to determine RNA secondary structure. Recently, methods for genome-wide RNA secondary structure determination by dimethyl sulfate or an acylating reagent have been developed.⁵⁻⁷ Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) is used to identify the flexible nucleotides (often unpaired nucleotides) in RNAs. SHAPE relies on the use of anhydride, imidazolide or acyl cyanide acylating reagents such as NMIA (*N*-methylisatoic anhydride), BzCN (benzoyl cyanide), 1M7 (1-methyl-7-nitroisatoic anhydride) and FAI (2-methyl-3-furoic acid imidazolide) and NAI (2-methylnicotinic acid imidazolide).⁸⁻¹² The identity of nucleotide has a low influence on ribose 2'-hydroxyl reactivity.¹³ SHAPE-directed RNA modeling is a powerful approach but it is not always unambiguous.¹⁴ In the SHAPE chemistry, the acylating reagent is supposed to acylate only the 2'-hydroxyl function (Figure 1),^{5,15} based on the reactivity of cAMP and cGMP with NMIA under basic conditions (with 2N NaOH to adjust the pH to 9.6).¹⁶ To the best

of our knowledge, no structural data based on NMR or X-ray methods establish that the acylation of the 2'-OH group is the unique product of the reaction. Besides, use of sodium hydroxide solution could increase the nucleophilic property of hydroxyl group by deprotonation. Actually, nucleobase¹⁷⁻¹⁹ or cyclic phosphate²⁰ could also be acylated or alkylated. These reactions, even occurred partially, might explain the unexpected reactivity of some nucleotides towards acylating reagents. We then decided to identify acylation sites in ribonucleotides with NMR and mass spectrometry (MS) methods,^{21,22} by using cyclic ribonucleotides (cAMP, cGMP and cCMP) as model substrates. Cyclic (3',5') nucleosides monophosphate (cNMPs) were chosen because these cyclic forms were shown to be "hyper-reactive" toward SHAPE agents.¹³

Figure 1. Proposed SHAPE mechanism of 2'-acylation at flexible RNA nucleotides.

We then performed the acylation reaction with NMIA under SHAPE conditions at pH 8 and 38 °C. The reaction at pH 8 is much slower than that at pH 9.6.^{16,23,24} Total consumption of the starting cNMP needed 20 equivalents of NMIA after one-night reaction (Scheme 1).^{25,26} 78 to 84% of fluorescent nucleotides **1-3** have been isolated.²⁷ These compounds showed a maximum absorption at 360 nm and maximum fluorescence emission at 424 nm in acetonitrile ($\lambda_{\text{ex}} = 360 \text{ nm}$) (Figure S1).





Scheme 1. Acylation reaction in a mixture of 1:1 DMSO-HEPES buffer (pH8, 50 mM HEPES-KOH, 200 mM KOAc, 5 mM MgCl₂).

Structural characterization of the isolated acylated products have been realized by standard 1D and 2D homo- and heteronuclear NMR spectroscopy and LC-UV-MS² experiments. The NMR analysis of the different reaction products allowed identification of a major reaction product for the three NMIA reactants cAMP, cGMP and cCMP. Interestingly, while for cAMP a unique compound was detected, the spectra of the reaction products for cCMP and cGMP showed that, besides a major compound, a minor product (~2%) was observed. For the major compound, we integrated separately the resonances related to the nucleoside mono-phosphate part (cNMP; N stands for any nucleotide A, C, or G) and those related to NMIA part, this permitted to establish the formation of a mono-acylated product. In a second step, we analyzed in detail the connectivities detected for the major product by using COSY, HSQC and HMBC spectra for the three studied cNMPs. This led to the unambiguous assignment of the ¹H and ¹³C resonances of three monoacylated cNMPs. The analyses allowed determining that the major acylation site is at the O2' position. We identified in the HMBC spectra a direct ³J scalar coupling between the H2' proton of ribose sugar and the carbonyl C1'' formed consequently to the acylation reaction (Figures 2, S2 and S3). This latter carbonyl showed also ³J and ⁴J couplings to the NMIA protons. These observations were made for the major reaction products with cAMP and cGMP and constitute a direct proof of the acylation at the O2' position. The perturbations of proton chemical shifts consecutive to the acylation are in good agreement with the information obtained from the scalar couplings. Indeed, a downfield shift of 1.3-1.5 ppm (according to the cNMP considered) was observed for the H2' proton while smaller shifts were observed for the H1' (0.2-0.3 ppm) and H3' (0.1-0.3 ppm) protons (Figure S4). Besides, the absence of 2'-OH on the acylated product has been confirmed by exchange experiment with D₂O (Figure S5). In contrast, other protons of the acylated cNMPs showed very weak variations. The very similar pattern of chemical shifts modifications due to acylation with the three cyclic mononucleotides strongly suggest that the acylation site is the same, that is at the O2' position. In addition, the chemical shifts of the NMIA part are extremely close suggesting a similar environment with the three cNMPs. Interestingly, small differences were observed in the scalar couplings ³J ¹H-¹H (*J*_{2',3'} between 5.3 and 5.9 Hz) and the intensities of HMBC H2'-C1'' peaks vary significantly between cAMP/cGMP and cCMP (Figures 2, S2 and S3) indicating that the corresponding ³J H2'-C1'' could be different in the three final reaction products. These data indicate the existence of differences in the sugar conformation between the cAMP and cGMP versus the cCMP products. Indeed, it was not expected that the acylated products of cyclic and therefore constrained mononucleotides should exhibit significant differences in the ribose sugar conformations. This interesting observation suggests that the ribose pucker is an important factor for the acylation reaction. For all the cNMPs,

the small proportion (< 3%) of the minor isomer hampered its 2D analysis.

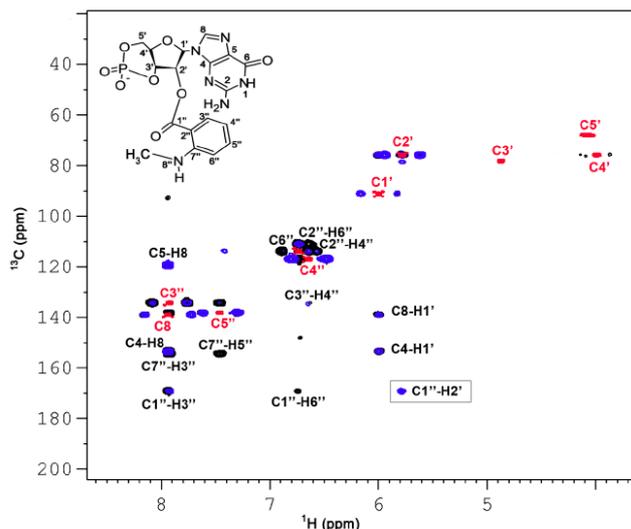


Figure 2. ¹H-¹³C spectra of the cGMP acylation product **3** with assignments of CH and CH₂ carbons as well as those of some cross-peaks. HSQC (red) and two HMBC spectra with two different values of delays were chosen for the visualization of long-range couplings *J* = 4 Hz (blue) and *J* = 6 Hz (black) (See also Figures S6-8 for more details). The name of CH and CH₂ carbons are indicated in red at the position of the HSQC cross-peaks. The ³J coupling C1''-H2' is boxed.

The LC-ESI-MS/MS analysis, thanks to its high sensitivity, permitted the analysis of the different products of the reaction, even in the case of a very low amount of products. LC-UV-MS² in negative ionization mode was performed to characterize each compound of the mixture. Specificity of fragment ions for *O*- and *N*-acylation can be determined from fragmentation pathways. In the case of cCMP, three different peaks in 4/96/trace ratio were quantified from UV spectra recorded at 232 nm and identified by MS (Figure 3). The dominant species detected at *m/z* 437 exhibited fragment ions at *m/z* 326 and *m/z* 150 (Table S1, Figure S9) specific of the *O*-acylation (**1**). In fact, the detection of fragment ions at *m/z* 150 corresponding to C₈H₈NO₂ (2-NHMe-Ph-CO₂) which can only be formed from *O*-acylation product. In addition, the direct loss of 111 Daltons (*m/z* 326), which corresponds to the cytosine nucleobase, can only be lost from *O*-acylated compounds. A mechanism has been proposed in Scheme 2.

The species detected at *m/z* 570 exhibited two fragment ions detected at *m/z* 243 and *m/z* 376, characteristic of the *N*-acylation (Table S2, Figure S10). In fact, *N*-acylated compounds show acylated cytosine fragment ion (*m/z* 243), allowing the unambiguous determination of *N*-acylation. In addition, a fragment ion at *m/z* 326 specific of *O*-acylation was also observed, clearly suggesting that the species detected at *m/z* 570 was *N,O*-diacylated (**1b**) (proposed mechanism in Scheme 3). By contrast, the third compound detected at *m/z* 437 and present at trace amount, showed specific fragment ions of the *N*-acylation at *m/z* 243 and *m/z* 376 (**1c**) (Scheme S1). Moreover, the successive losses of 43 (*m/z* 394) and 18 (*m/z* 376) mass units corresponding to CONH and H₂O, respectively, indicate the presence of a free hydroxyl group meaning that *O*-acylation did not occur. ESI-HRMS² analysis of these compounds was used to confirm molecular formulae of the specific fragment ions described (Table S1).

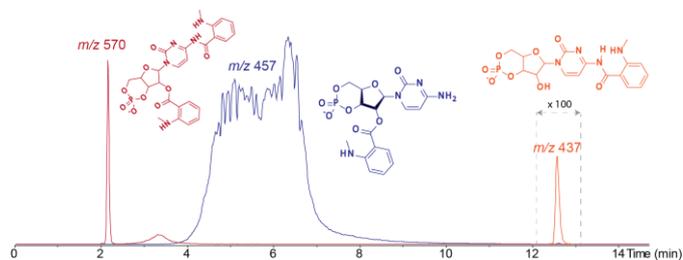
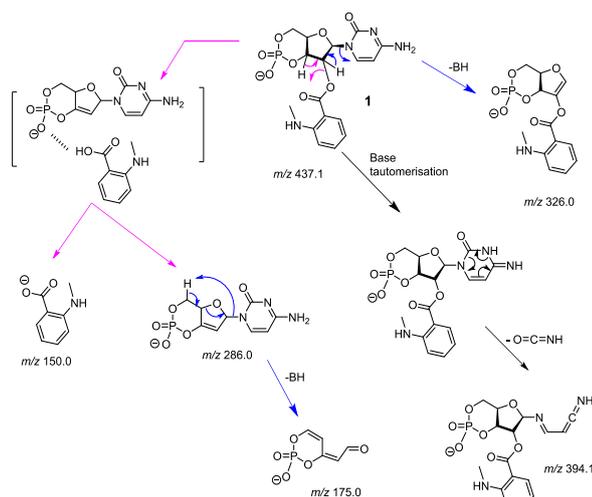
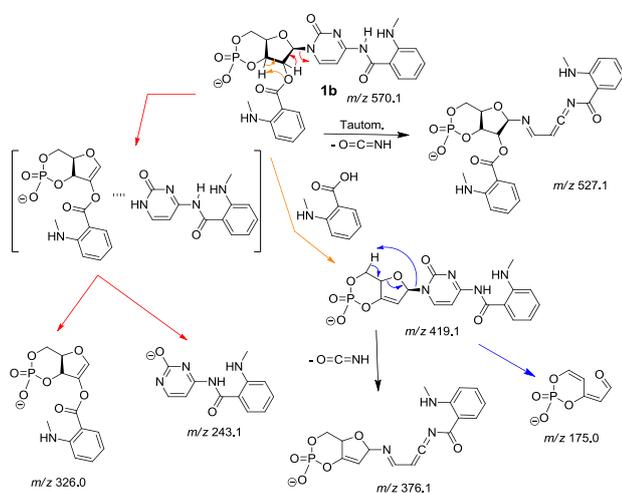


Figure 3. Extracted ion chromatogram from LC-UV-MS experiments of a solution mixture of **1** diluted at 100 μ M in acetonitrile. The concentration of the solution mixture was voluntary too important in order to observe *N*-acylation compound at 12.6 min but generated peak duplication.



Scheme 2. Proposed mechanism associated with the main fragment ions of the m/z 437 ion corresponding to *O*-acylated cCMP **1**.



Scheme 3. Proposed mechanism associated with the main fragment ions of the m/z 570 ion corresponding to *N,O*-diacylated cCMP **1b**.

For cAMP, two compounds at m/z 461 were detected from LC-UV-MS² experiments ($t_R = 3.5$ and 11.8 min), corresponding to acylated cAMP (Figure S11, Table S3). CID analysis of the major compound ($t_R = 3.5$ min, > 99 % from UV quantitation) exhibited fragment ions at m/z 326 (loss of adenine) and 150 (C₈H₈NO₂) specific of the *O*-acylation (*vide supra*) (Scheme S2). The compound at m/z 461 presented at trace level ($t_R = 11.8$ min) showed fragment ion at m/z 267 (*N*-acylated adenine), specific of the *N*-acylation (Scheme S3). Fragment ions composition of the *O*-acylated compound was confirmed by HRMS² (Table S3). Concerning cGMP, one major product (> 99 % at m/z 477) and four minor compounds (including the starting material cGMP at

m/z 344) have been detected (Figure S12). Acylated cGMP was observed at m/z 477 from LC-UV-MS² spectra. CID spectra of this ion showed specific fragment ions at m/z 326 and m/z 150, characteristics of the *O*-acylation, as confirmed by ESI-HRMS² (Table S4, Scheme S4).

In conclusion, we have identified for the first time the acylation products of cNMP with NMIA carried out under the SHAPE chemistry conditions, by NMR and mass spectrometry. The LC-ESI-MS/MS analysis clearly demonstrate that beside the major 2'-*O*-acylated product, there are also *N*- and *N,O*-diacylated products with cNMPs, especially in the case of cCMP. This interesting result should be useful to explain some unexpected reactivity of the SHAPE chemistry.

Acknowledgement

C.L. gratefully acknowledges China Scholarship Council (CSC) for a doctoral scholarship. This work has been supported by CHARMMAT ANR - 11-LABX-0039 for a post-doctoral fellowship to S.P. and

References and notes

- Xie, C.; Yuan, J.; Li, H.; Li, M.; Zhao, G.; Bu, D.; Zhu, W.; Wu, W.; Chen, R.; Zhao, Y. *Nucleic Acids Res.* **2014**, *42*, D98-103.
- Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Martin, D.; Merkel, A.; Knowles, D. G.; Lagarde, J.; Veeravalli, L.; Ruan, X.; Ruan, Y.; Lassmann, T.; Carninci, P.; Brown, J. B.; Lipovich, L.; Gonzalez, J. M.; Thomas, M.; Davis, C. A.; Shiekhattar, R.; Gingeras, T. R.; Hubbard, T. J.; Notredame, C.; Harrow, J.; Guigó, R. *Genome Res.* **2012**, *9*, 1775.
- Novikova, I. V.; Hennelly, S. P.; Tung, C. S.; Sanbonmatsu, K. Y. *J. Mol. Biol.* **2013**, *425*, 3731.
- Cech, T. R.; Steitz, J. A. *Cell* **2014**, *157*, 77.
- Ding, Y.; Tang, Y.; Kwok, C. K.; Zhang, Y.; Bevilacqua, P. C.; Assmann, S. M. *Nature* **2014**, *505*, 696.
- Rouskin, S.; Zubradt, M.; Washietl, S.; Kellis, M.; Weissman, J. S. *Nature* **2014**, *505*, 701.
- Spitale, R. C.; Flynn, R. A.; Zhang, Q. C.; Crisalli, P.; Lee, B.; Jung, J. W.; Kuchelmeister, H. Y.; Batista, P. J.; Torre, E. A.; Kool, E. T.; Chang, H. Y. *Nature* **2015**, *519*, 486.
- Merino, E. J.; Wilkinson, K. A.; Coughlan, J. L.; Weeks, K. M. *J. Am. Chem. Soc.* **2005**, *127*, 4223.
- Mortimer, S. A.; Weeks, K. M. *J. Am. Chem. Soc.* **2007**, *129*, 4144.
- Mortimer, S. A.; Weeks, K. M. *Nat. Protoc.* **2009**, *4*, 1413.
- Spitale, R. C.; Crisalli, P.; Flynn, R. A.; Torre, E. A.; Kool, E. T.; Chang, H. Y. *Nat. Chem. Biol.* **2013**, *9*, 18.
- Nodin, L.; Noël, O.; Chaminade, F.; Maskri, O.; Barbier, V.; David, O.; Fossé, P.; Xie, J. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 566.
- Wilkinson, K. A.; Vasa, S. M.; Deigan, K. E.; Mortimer, S. A.; Giddings, M. C.; Weeks, K. M. *RNA* **2009**, *15*, 1314.
- Kladwang, W.; VanLang, C. C.; Cordero, P.; Das, R. *Biochemistry* **2011**, *50*, 804.
- McGinnis, J. L.; Dunkle, J. A.; Cate, J. H.; Weeks, K. M. *J. Am. Chem. Soc.* **2012**, *134*, 6617.
- Hiratsuka, H. *J. Biol. Chem.* **1982**, *257*, 13354.
- Matsuda, A.; Yasuoka, J.; Sasaki, T.; Ueda, T. *J. Med. Chem.* **1991**, *34*, 999.
- Sekine, M. *J. Org. Chem.* **1989**, *54*, 2321.
- Kataoka, S.; Imai, J.; Yamaji, N.; Kato, M.; Kawada, T.; Imai S. *Chem. Pharm. Bull.* **1991**, *38*, 1596.
- Kataoka, S.; Isono, J.; Yamaji, N.; Kato, M. *Chem. Lett.* **1986**, 1221.
- Neubauer, S.; Rugova, A.; Chu, D. B.; Drexler, H.; Ganner, A.; Sauer, M.; Mattanovich, D.; Hann, S.; Koellensperger, G. *Anal. Bioanal. Chem.*, **2012**, *404*, 799.
- Zhao, H.-Q.; Wang, X.; Li, H.-M.; Yang, B.; Yang, H.-J.; Huang, L. *Molecules* **2013**, *18*, 9755.
- Hiratsuka, T. *Biochim. Biophys. Acta* **1983**, *742*, 496.
- Reinecke, D.; Schwede, F.; Genieser, H.-G.; Seifert, R. *Plosone* **2013**, *8*, e54158.
- Because of the availability and cost, cUMP was not tested.

26. Reaction of cNMP with BzCN under SHAPE's condition was much more difficult to lead to completion, probably due to the instability of BzCN in aqueous medium. We have failed to isolate pure and sufficient amount of acylation products for analysis, despite many attempts.
27. General procedures for the acylation: To a mixture of cNMP (1 equiv.) in HEPES buffer (pH 8)/DMSO system, was added NMIA (20 equiv.). The resulting mixture was stirred at 38 °C overnight. Then the solvent was removed under reduced pressure and the residue purified by column chromatography (CH₂Cl₂/MeOH, 2/1) to give the acylated compound. Compound **1**: white solid, 84%, **R_f**: 0.08 (CH₂Cl₂/MeOH, 5/1); ¹H NMR (DMSO) δ 7.89 (d, 1H, *J* = 8 Hz, H-3''), 7.61 (d, 1H, *J* = 7.4 Hz, H-6), 7.44 (t, 1H, *J* = 7.5 Hz, H-5''), 6.72 (d, 1H, *J* = 8.7 Hz, H-6''), 6.62 (t, 1H, *J* = 7.6 Hz, H-4''), 5.78 (d, 1H, *J* = 7.3 Hz, H-5), 5.76 (s, 1H, H-1'), 5.49 (d, 1H, *J* = 5.3 Hz, H-2'), 4.61-4.67 (m, 1H, H-3'), 4.09-4.16 (m, 2H, H-5'a, H-5'b), 3.84-3.90 (m, 1H, H-4'), 2.84 (d, 3H, *J* = 5.0 Hz, *N*-Me); ¹³C NMR (DMSO) δ 169.0 (C1''), 168.5 (C4), 157.0 (C2), 154.1 (C7''), 145.8 (C6), 137.8 (C5''), 134.1 (C3''), 116.8 (C4''), 113.6 (C6''), 111.0 (C2''), 97.2 (C1'), 95.8 (C5), 77.8 (C3'), 75.7 (C2'), 75.3 (C4'), 67.9 (C5'), 31.9 (Me). Compound **2**: white solid, 80%, **R_f**: 0.13 (CH₂Cl₂/MeOH, 5/1); ¹H NMR (DMSO) δ 8.31 (s, 1H, H-8), 8.20 (s, 1H, H-2), 7.93 (d, 1H, *J* = 8.0 Hz, H-3''), 7.45 (t, 1H, *J* = 7.8 Hz, H-5''), 6.72 (d, 1H, *J* = 8.5 Hz, H-6''), 6.64 (t, 1H, *J* = 7.5 Hz, H-4''), 6.23 (s, 1H, H-1'), 5.89 (d, 1H, *J* = 5.9 Hz, H-2'), 5.04-5.09 (m, 1H, H-3'), 4.05-4.14 (m, 2H, H-5'a, H-5'b), 3.96-4.02 (m, 1H, H-4'), 2.83 (d, 3H, *J* = 5.1 Hz, *N*-Me); ¹³C NMR (DMSO) δ 169.0 (C1''), 155.6 (C2), 154.1 (C7''), 143.0 (C8), 151.6 (C4), 137.9 (C5''), 134.1 (C3''), 121.5 (C5), 116.8 (C4''), 113.7 (C6''), 111.0 (C2''), 91.8 (C1'), 77.8 (C3'), 75.7 (C2', C4'), 67.8 (C5'), 158.9 (C6), 31.9 (Me). Compound **3**: white solid, 78%, **R_f**: 0.08 (CH₂Cl₂/MeOH, 5/1); ¹H NMR (DMSO) δ 7.93 (s, 1H, H-8); 7.91 (d, 1H, *J* = 8.0 Hz, H-3''), 7.45 (t, 1H, *J* = 7.8 Hz, H-5''), 6.72 (d, 1H, *J* = 8.6 Hz, H-6''), 6.63 (t, 1H, *J* = 7.4 Hz, H-4''), 5.98 (s, 1H, H-1'); 5.77 (d, 1H, *J* = 5.9 Hz, H-2'); 4.79-4.87 (m, 1H, H-3'); 4.02-4.14 (m, 2H, H-5'a, H-5'b), 3.94-4.01 (m, 1H, H-4'), 2.84 (d, 3H, *J* = 5.0 Hz, *N*-Me); ¹³C NMR (DMSO) δ 169.0 (C1''), 154.3 (C7''), 153.4 (C4), 139.5 (C8), 138.0 (C5''), 134.1 (C3''), 119.6 (C5), 116.8 (C4''), 113.7 (C6''), 111.0 (C2''), 91.0 (C1'), 78.1 (C3'), 75.8 (C2'), 75.6 (C4'), 68.0 (C5'), 156.5 (C2), 159.7 (C6), 32.0 (Me).