Supporting Information

Investigating mithramycin deoxysugar biosynthesis: enzymatic total synthesis of TDP-D-Olivose

Guojun Wang, Madan K. Kharel, Pallab Pahari, and Jürgen Rohr*

Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky,

Lexington, KY 40536, USA

jrohr2@email.uky.edu

Table of Contents

A. Experimental procedures ………………………………………………………………… S2

B. Primers used in the study ……………………………………………………………S6

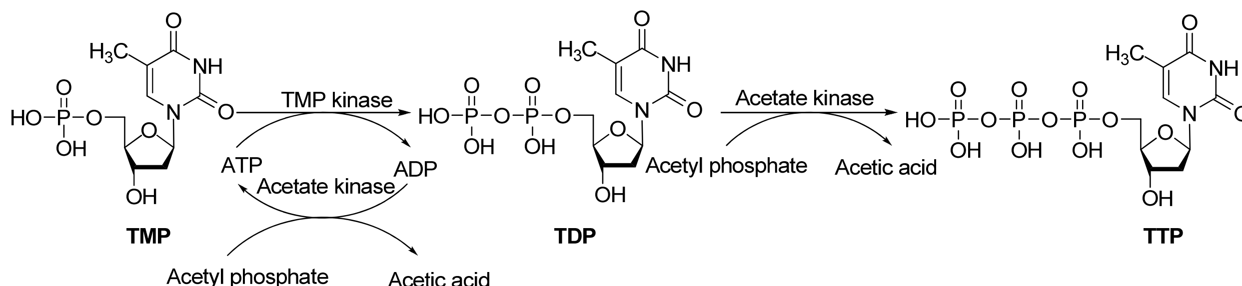
C. Spectral Data ……………………………………………………………………………S7

D. References …………………………………………………………………………………S17

A. Experimental procedures

General biological procedures: *Escherichia coli* strains XL1 Blue and BL21(DE3) were used for routine DNA cloning and protein over-expression, respectively. They were grown in Luria Broth (LB) using 100-ml flasks on a rotary shaker or on LB agar at 37 °C. Kanamycin sulfate, when necessary, was supplemented to a final concentration of 50 μg ml^{-1}. Vectors pET28a(+) and pET30 Xa/LIC (purchased from Novagen) were used for protein expression. Both vectors provide an *N*-terminal 6× His tag. Plasmid preparation and DNA fragment recovery from agrose gel were carried out using standard commercial products provided by Fermentas and Qiagen, respectively. Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs. *Pfu* DNA polymerase was obtained from Stratagene, and used for routine PCR using primers listed below with the following procedure: 96 °C, 3 min; followed by 25 cycles of 96 °C, 30 sec; 58 °C, 30 sec; 72 °C 2min; with final extension 72 °C for 5 min, in which annealing temperature and extension time might vary depending on Tm value of different primers and length of different target genes. Choromosomal DNA of the MTM producer *Streptomyces argillaceus* was used as template for *mtmV* and *mtmC*; while the sugar plasmid pLNR [1] was used as template for *oleV* , *oleW* and *urdR*, and the *zwf* gene for G6PD was amplified from *E. coli* genome. For those using pET28a(+) vector (*oleW*, *urdR* and *zwf*), gel purified PCR products were cloned into a blunt-end cloning vector TOPO-vector (Invitrogen^{TM}) and sequenced to ensure no PCR error, then, each insert was cut out using appropriate restriction enzymes, and ligated into pET28a(+). For, others using pET30Xa-LIC, PCR products were directly cloned into the vector following the manufacture's manual, correct constructs were then confirmed by sequencing. An earlier

established *in situ* ATP and TTP regeneration system using acetate kinase (ACK), TMP kinase (TMK) with acetyl phosphate as the phosphor donor was used herein (Scheme S1).[2]



Scheme S1. Enzymatic synthesis of thymidine-triphosphate (TTP).

In case of TDP-4-keto-6-deoxy-D-glucose-2, 3-dehydratase, in addition to *mtmV* and *oleV*, several other genes including *urdS* from urdamycin pathway, *cmmV* from chromomycin pathway and *jadO* from jadomycin pathway were cloned from each genome and tested for soluble expression in *E. coli*. However, only OleV can be successfully obtained from *E. coli*, all others were expressed as insoluble inclusion body.

Protein purification: All proteins used in this work contained a 6×His tag at *N*-terminal, and were purified using immobilized metal affinity chromatography (IMAC). For ACK, TMK, RfbA and RmlB[2], 1 % of overnight seed culture were inoculated into one liter of LB and grown at 37 °C until OD_{600} reached to 0.4~0.5, then the culture was induced by adding 200 μM IPTG for 12 h at 18 °C. For all other proteins, *E. coli* strains were cultured throughout 18 °C. After collecting *E. coli* cells by centrifugation (4000, 15min) at 4 °C, cells were washed twice with 20 ml of lysis buffer (50 mM KH_2PO_4, 300 mM KCl, 10 mM imidazole, pH 8.0), resuspended in 30 ml of lysis buffer, pressure-frenched, then centrifuged at 17, 000 × g for 30 min at 4 °C to remove cell debris. Supernatant was used for further protein purification. After IMAC, ACK, TMK, RfbA, RmlB, OleV, G6PD and MtmC were further desalted using Profinia protein purification system according to manufacture's protocol. All proteins were finally in 4 ml of desalting buffer (4.3 mM Na_2PO_4, 8.1 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH7.4), and used immediately. OleW was purified using gravity TALON affinity resin (Clontech Laboratories Inc.) with an elution buffer (50 mM KH_2PO_4, 300 mM KCl, 250 mM imidazole, pH 8.0). After elution,

protein was desalted using size exclusion filter units (Millipore) against 50 mM phosphate buffer, 150 mM NaCl, pH8.0, and then used immediately.

Protein concentrations were determined using Bradford reagent [3]. Purities were estimated by sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analyses. The measured sizes of purified proteins were in agreement with the calculated ones (Figure S1).

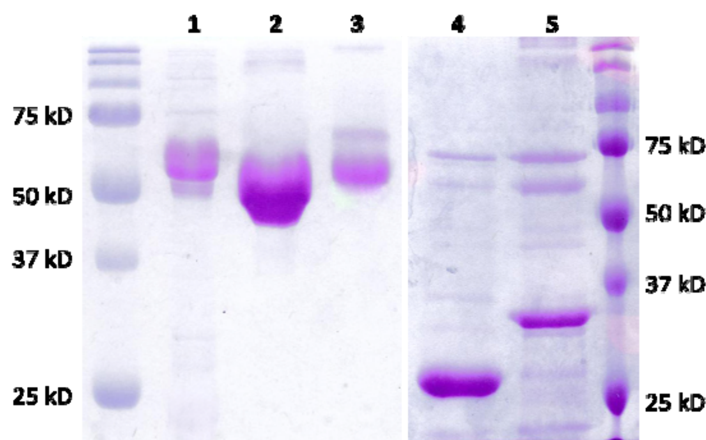


Figure S1. SDS-PAGE analysis of purified proteins used in this study. Lan1, OleV (53.06 kD); Lan2, MtmC (45.81 kD); Lan3, G6PD (55.7 kD); Lan2, UrdR (26.84 kD); Lan5, OleW (35.84 kD).

Enzymatic assays: 1) Large scale preparation of TDP-4-keto-2,6-dideoxy-D-glucose (**5**): A 25-ml reaction mixture contained 100 mM phosphate buffer (pH7.5), 10 mM $MgCl_2$, 12 mM **3**, 1 mM NADPH, 14 mM glucose-6-phosphate, 3.5 μM OleV, 11.6 μM OleW, and 10.8 μM G6PD. The reaction was kept at 28 °C, and monitored by HPLC. After complete **3**→**5** conversion, reaction mixture was lyophilized and loaded onto BioGel P2 column for purification for purification of **5**; on the other hand, this reaction mixture was used for following large scale preparation of **6** after filtration to remove all enzymes.

2) Enzymatic assays of UrdR activity was carried out using compound **3** by co-incubation with OleV and Olew, or compound **5**, using following conditions: a) 100 mM phosphate buffer (pH7.5), 10 mM $MgCl_2$, 160 μM **3**, 400 μM NADPH, 4 μM OleV, 10 μM OleW and 16.5 μM UrdR; or, b) 100 mM phosphate buffer (pH7.5), 10 mM $MgCl_2$, 160 μM **5**, 200 μM NADPH, and 16.5 μM UrdR.

3) Enzymatic assays of MtmC using pure **5** were typically carried out at 28 °C in a 100-µl aliquot containing 100 mM phosphate buffer (pH7.5), 10 mM $MgCl_2$, 160 µM **5**, and 10 µM MtmC. Different combinations of coenzymes were added to following final concentrations: NADPH and $NADP^+$ to 250 µM, and SAM to 320 µM. The reaction mixture was boiled for 5 min, centrifuged at 12, 000 ×g for 5 min, then the supernatant was loaded for HPLC analysis.

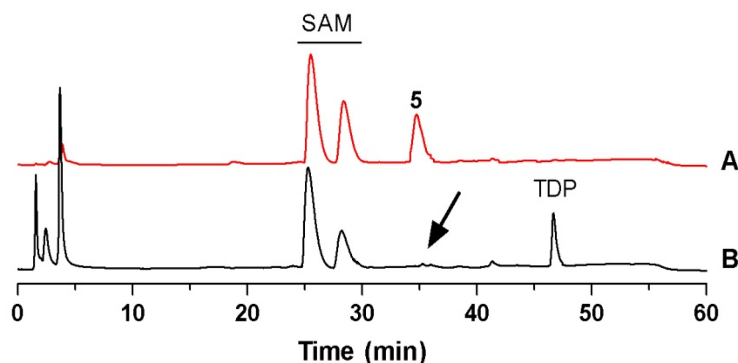


Fig. S2. HPLC profiles of MtmC incubated with pure **5** and SAM alone. Samples of zero time point (A) and overnight (B) were loaded for HPLC analysis. Arrow indicated the disappearance of TDP-sugars.

4) Large scale preparation of TDP-D-olivose (**6**): Above mentioned filtrated **3→5** reaction mixture was supplemented with an enzyme solution containing 100 mM phosphate buffer (pH7.5), 10 mM $MgCl_2$, 17.5 µM MtmC and 12.5 µM G6PD. The reaction was continued at 28 °C, and monitored by HPLC. After complete reaction of **5→6**, reaction mixture was lyophilized and loaded onto BioGel P2 column for purification. A typical reaction starting from 15 mg of substrate **3** yielded approximately 10.5 mg of TDP-D-olivose (**6**) after purification.

Deoxysugar purification: Deoxysugars were purified using either HPLC or BioGel P2 (BioRad) column as described previously [2, 4]. HPLC was carried out on a Waters Delta600 equipped with a Dionex Carbopac™ column (4×250 mm) using water as solvent A and 0.5 mM ammonium acetate as solvent B with following gradient of solvent B: 5~20% over 0~15min, 20~60% over 15~35 min, 60~100% over 35~37min, 100% over 37~40 min, decreased to 5% over 40~45 min, and 5 % over 45~60 min. Various size of BioGel P2 columns (2.5 cm × 100 cm, 1 cm × 100 cm, or 1 cm × 50 cm) were used to purify deoxysugars depending on the scale.

Lyophilized powder was dissolved in 5 ml of water, after centrifugation for 10 min at 12,000 × g, and loaded onto a BioGel P2 column which was pre-equilibrated with water. Separation was carried out at 4 °C with a flow rate of 0.2 ml min^{-1} of water. Aliquots of 4 ml were collected and checked for purity by HPLC at 254 mm. Pure fractions were combined, lyophilized, and stored at -80 °C for further use.

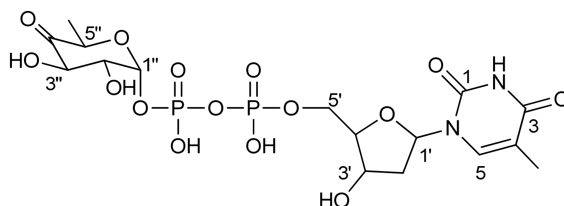
B. List of primers

Table S1. Primers used in this study. Restriction sites were underlined.

primer	sequence
mtmV_F30	GGTATTGAGGGTCGCATGACTCAAGCGATCATGAGCAG
mtmV_R30	AGAGGAGAGTTAGAGCCTCACCAGACGGACTGCAGACAGGCGA
mtmC_F30	GGTATTGAGGGTCGCATGGACATCTATGGGACCGCG
mtmC_R30	AGAGGAGAGTTAGAGCCTCACTCCCCGGGCGGGACGATCG
oleW_F_NdeI	AGCCATATGCCCTCCCCCCGTCTGCGCTTCG
oleW_R_XhoI	GTGCTCGAGTCAGCACCAGCGCACCCGCGCCAC
urdR-F-NdeI	AGCCATATGGACATTGTGGGAAATGGATTCC
urdR-R-XhoI	GTGCTCGAGTCAGATACGGACGGCGGAGGTG
oleV_F30	GGTATTGAGGGTCGCATGATATGGGGAATTCCCGCGATGAG
oleV_R30	AGAGGAGAGTTAGAGCCTCAGCTCAGGGCCTGGATGCAGGCG
cmmV_F_NdeI	TTGGATCCCATATGGCTCAGGCAGCGACACG
cmmV_R_XhoI	TTCTCGAGTCAGCCTCGGGCCCACAGCGGCTC
urdS_F_NdeI	TTGGATCCCATATGCTTTCTGAAAGCCTTCTCG
urdS_R_XhoI	TTCTCGAGTCATCGGCTGTTCACCGCCATGGCG
jadO_F_NdeI	TTGGATCCCATATGAGCACCGGCACCACCACG
jadO_R_XhoI	TTCTCGAGAGGTCGCACCGCTCCGGCGTG

C. Spectral data

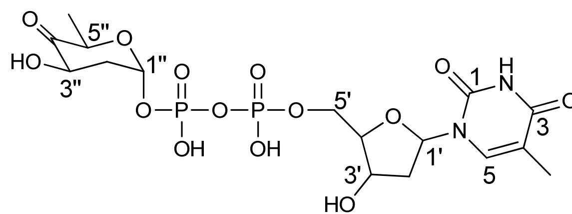
^{1}H NMR and ^{1}H-^{1}H COSY spectra of deoxysugars **3**, **5** and **6** were recorded on a Varian VNMR 500 MHz spectrometer in D_2O. MS analysis of **5** and **6** was carried out by MS facilities of University of Wisconsin, Madison, and University of Kentucky, respectively.



TDP-4-keto-6-deoxy-D-glucose (**3**)

Table S2. ^{1}H NMR (500 MHz, D_2O) of **3**:

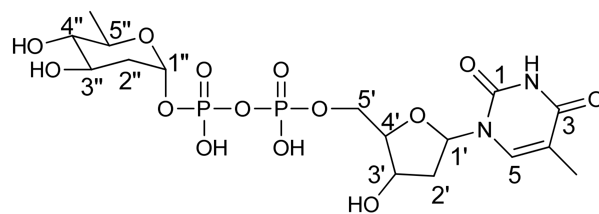
Position	δ (ppm)	Coupling constant (Hz)
4-Me	1.77 (s)	
5-H	7.57 (s)	
1'-H	6.19 (t)	7.0
2'-H	2.20-2.27 (m)	
3'-H	4.41-4.48 (m)	
4'-H	3.98-4.06 (m)	
5'-H	3.98-4.06 (m)	
1"-H	5.40 (dd)	6.5, 3.5
2"-H	3.50 (dt)	10.0, 3.5
3"-H	3.64 (d)	10.0
5"-H	3.93 (q)	6.5
5"-Me	1.08 (d)	6.5



TDP-4-keto-2, 6-dideoxy-D-glucose (**5**)

Table S3. ^{1}H NMR (500 MHz, D_2O) of **5**:

Position	δ (ppm)	Coupling constant (Hz)
4-Me	1.86 (s)	
5-H	7.68 (s)	
1’-H	6.28 (t)	6.5
2’-H	2.25-2.35 (m)	
3’-H	4.52-4.58 (m)	
4’-H	4.12-4.18 (m)	
5’-H	4.12-4.18 (m)	
1”-H	5.55-5.60 (m)	
2”-H*e*	2.08 (dd)	13.5, 5.0
2”-H*a*	1.79 (dd)	13.5, 13.5
3”-H	3.91-4.00 (m)	
5”-H	3.91-4.00 (m)	
5”-Me	1.14 (d)	6.0



TDP-D-olivose (**6**)

Table S4. ^{1}H NMR (500 MHz, D_2O) of **6**:

Position	δ (ppm)	Coupling constant (Hz)
4-Me	1.91 (s)	
5-H	7.74 (s)	
1'-H	6.33 (t)	7.5
2'-H	2.32-2.40 (m)	
3'-H	4.58-4.63 (m)	
4'-H	4.13-4.20 (m)	
5'-H	4.13-4.20 (m)	
1"-H	5.58-5.64 (m)	
2"-He	2.28 (dd)	12.5, 6.0
2"-Ha	1.70 (dd)	12.5, 12.5
3"-H	3.86-3.96 (m)	
4"-H	3.10 (t)	9.5
5"-H	3.86-3.96 (m)	
5"-Me	1.25 (d)	6.5

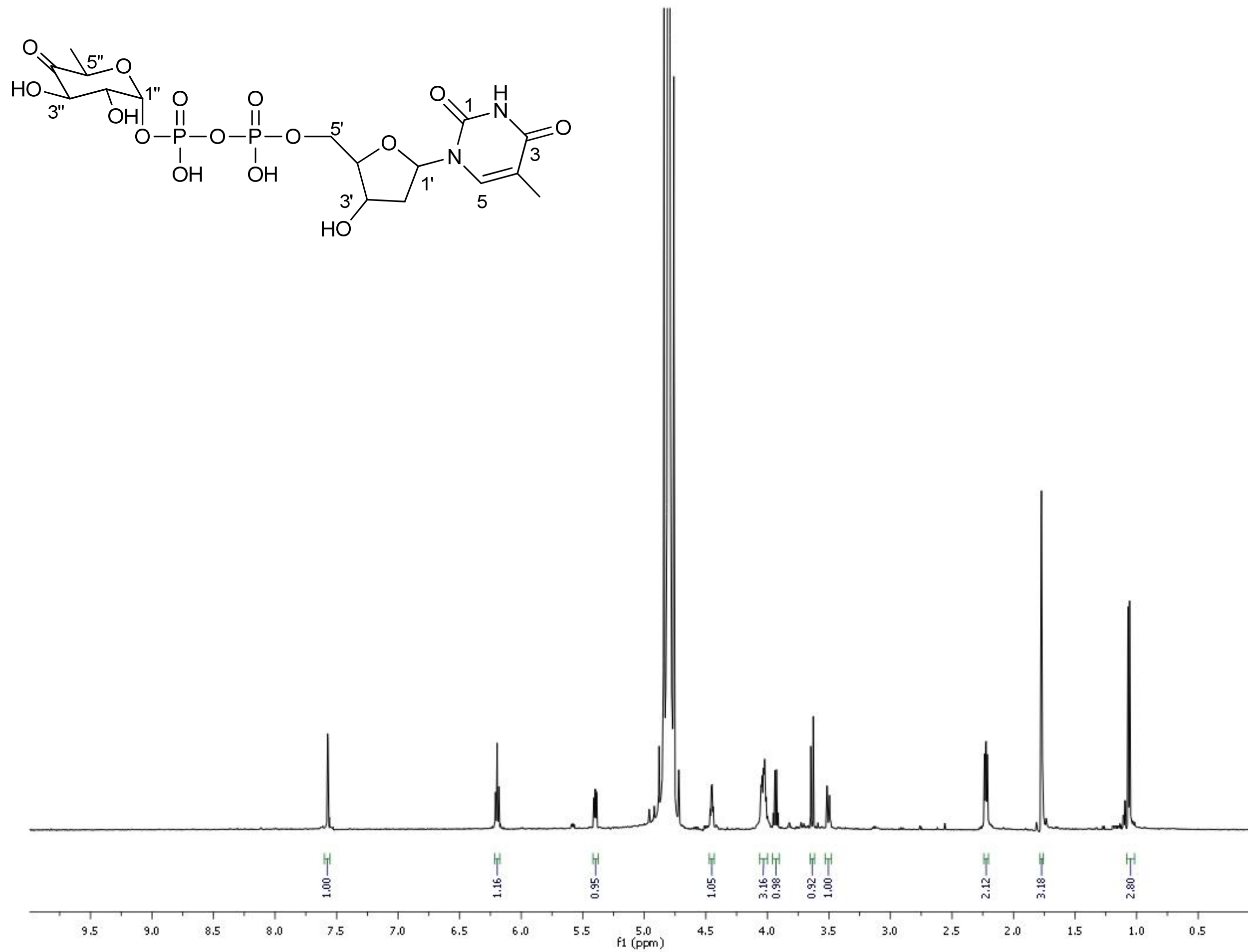


Figure S3. ^{1}H NMR spectrum of TDP-4-keto-6-deoxy-D-glucose (**3**).

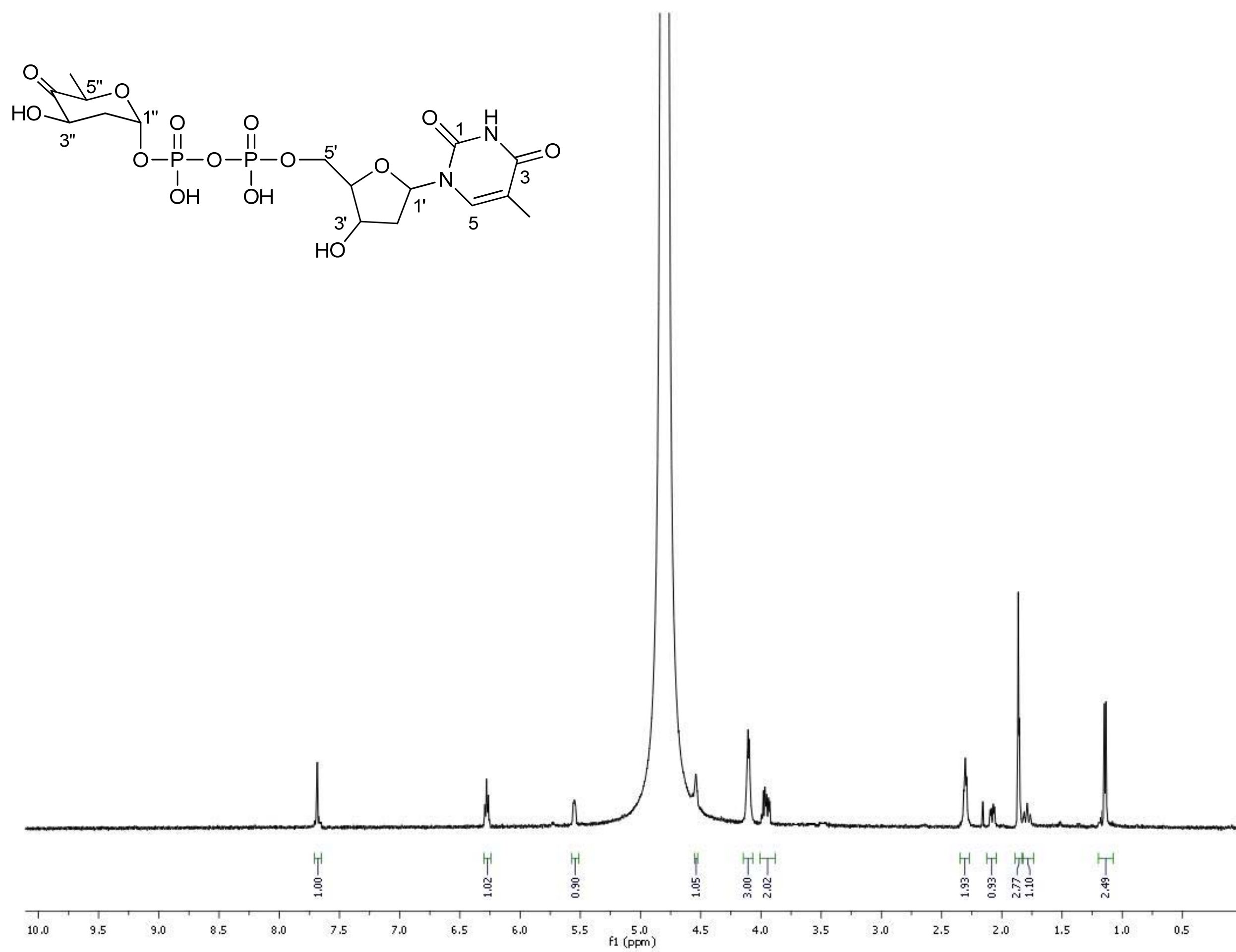


Figure S4. ^{1}H NMR spectrum of TDP-4-keto-2,6-dideoxy-D-glucose (**5**).

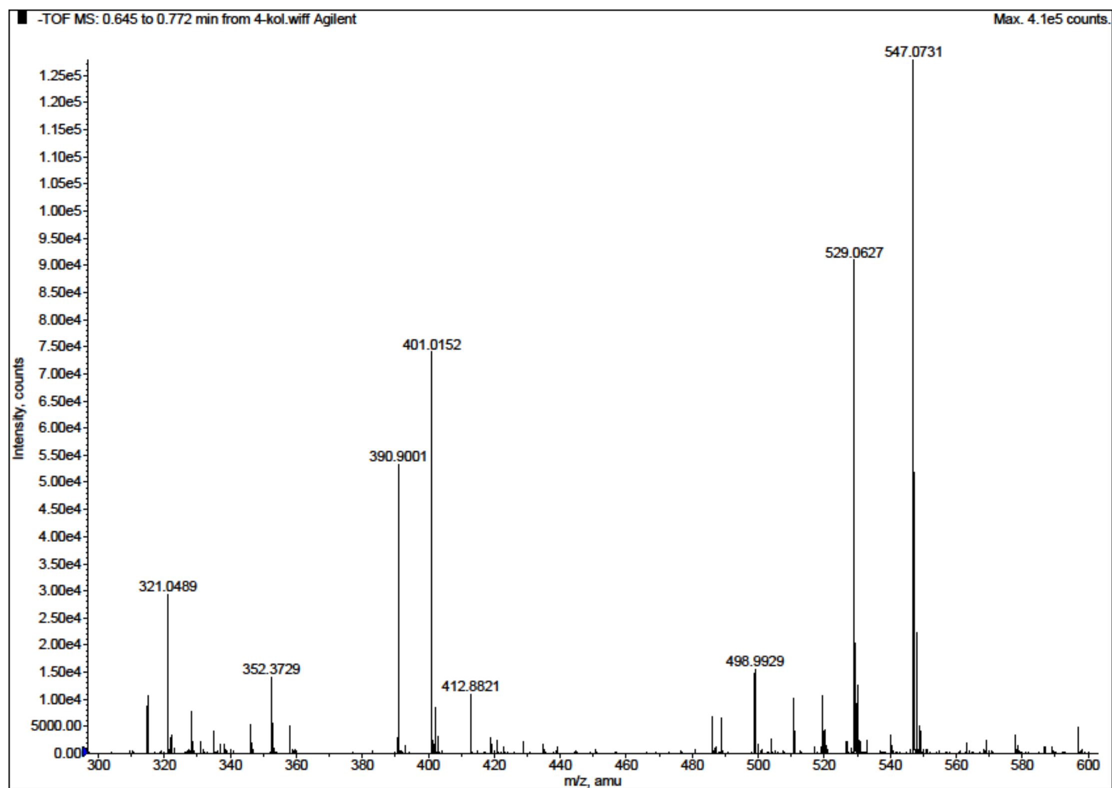
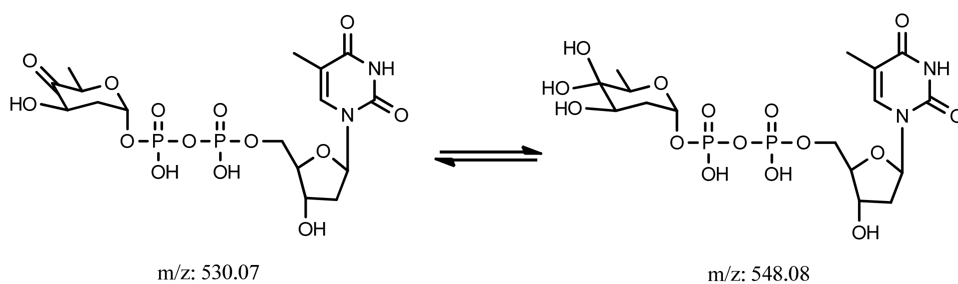


Figure S5. HRMS spectrum (-ve mode) of **5**. Calculated MW of **5** is 530.31, with a MW of 548.33 for the hydrate form.



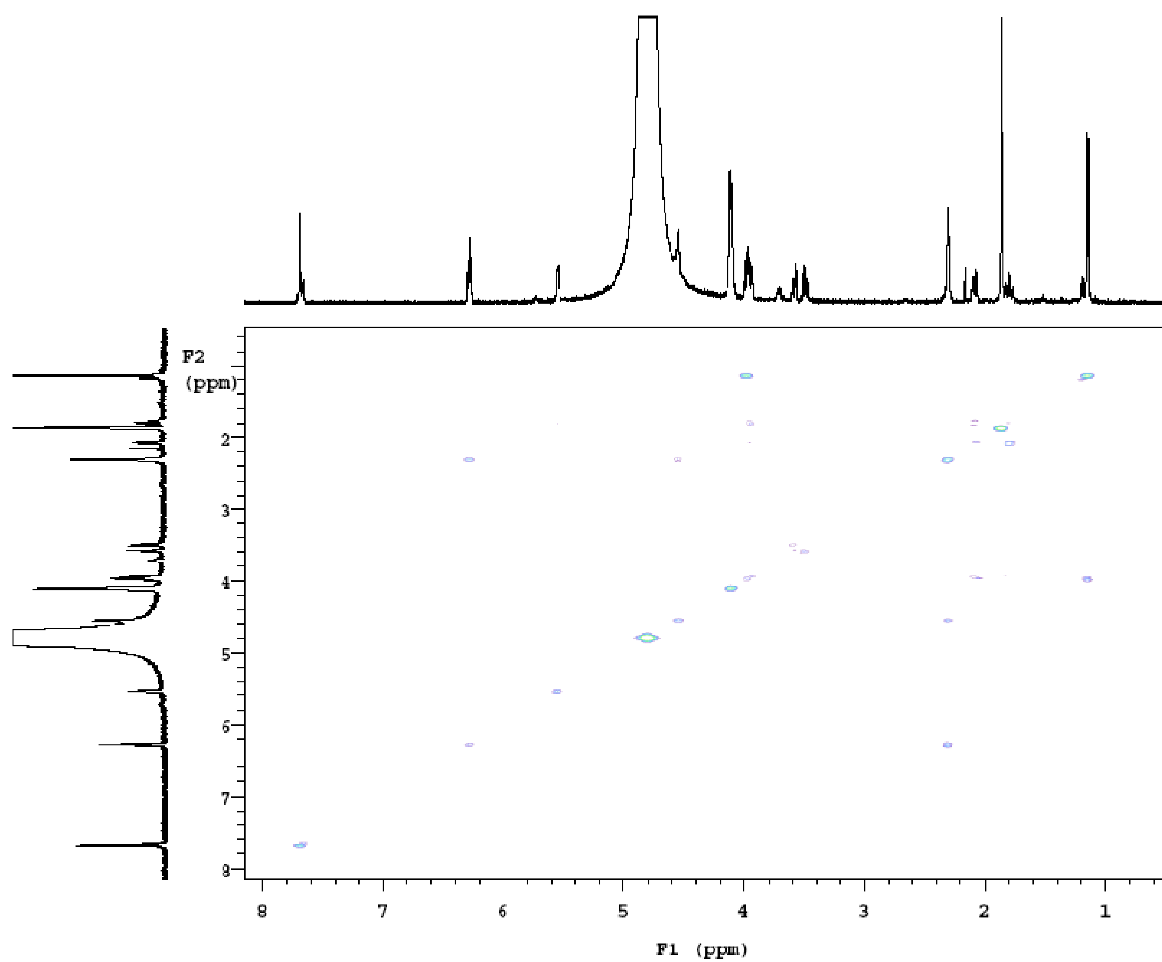


Figure S6. ^{1}H-^{1}H COSY spectrum of **5**.

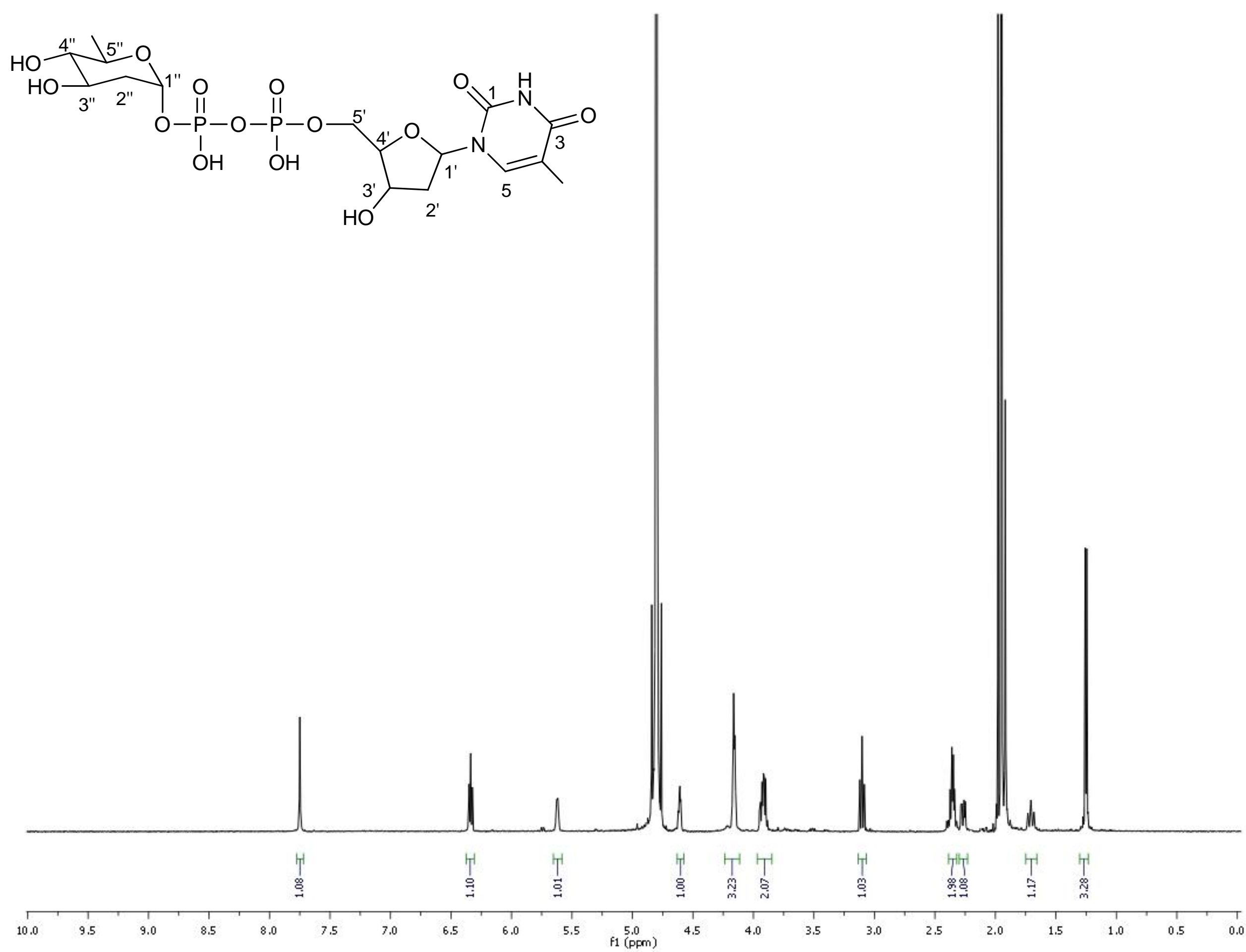


Figure S7. ^{1}H NMR spectrum of TDP-D-olivose (**6**).

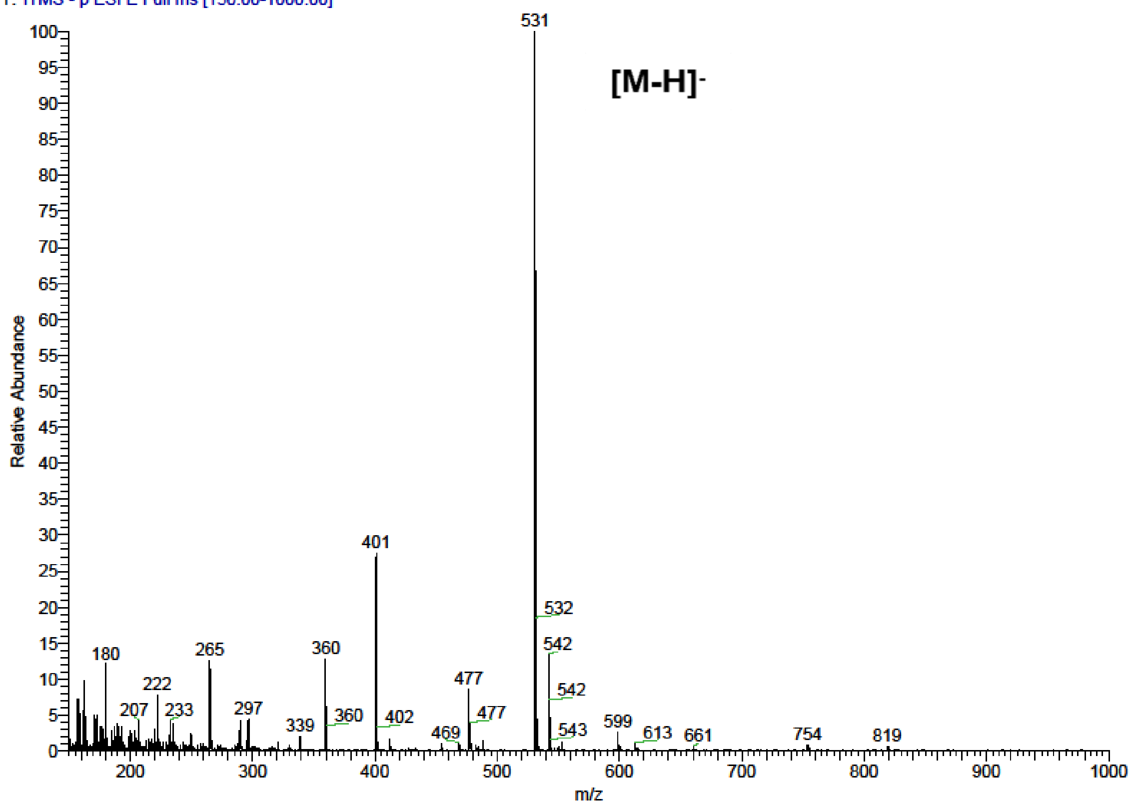


Figure S8. Low resolution ESI-MS spectra (-ve mode) of **6**. Calculated MW of **6** is 532.33.

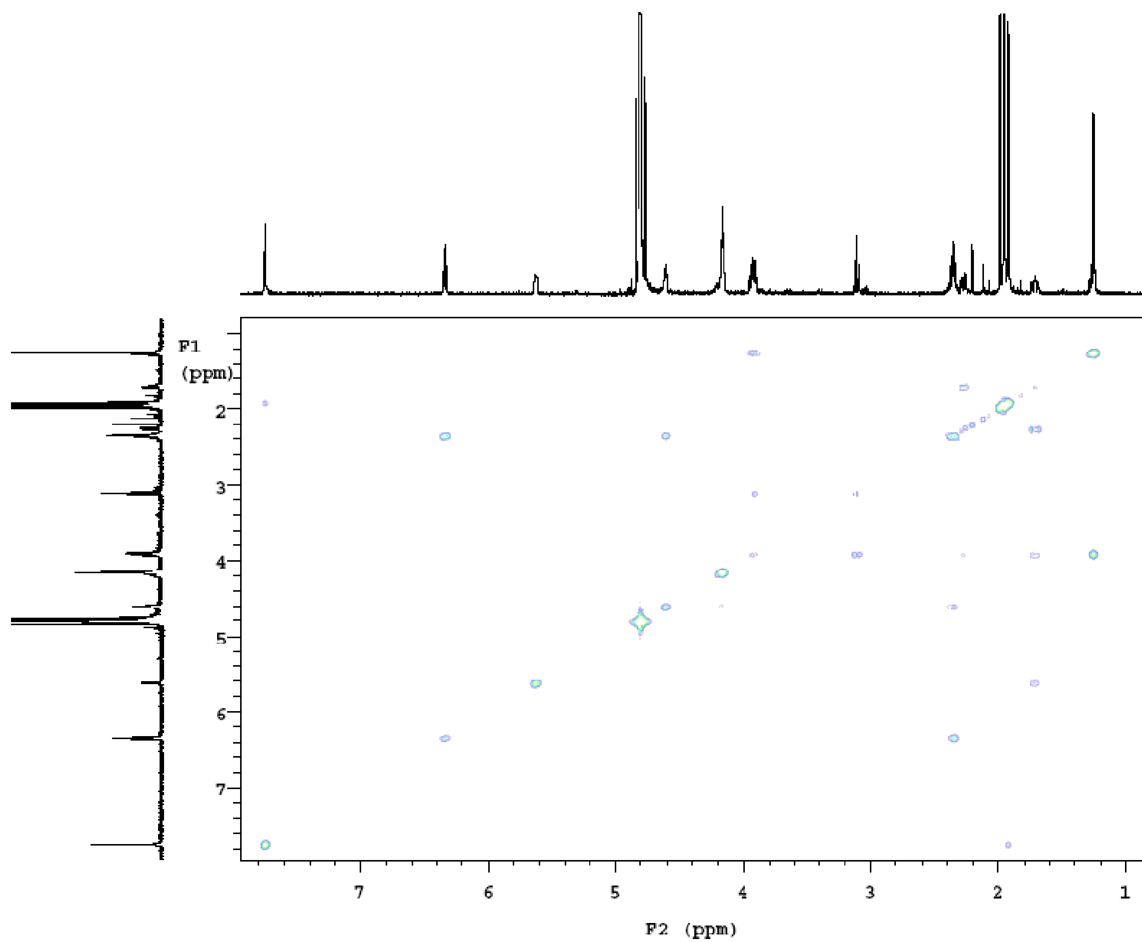


Figure S9. ^{1}H-^{1}H COSY spectrum of **6**.

References

[1] L. Rodriguez, I. Aguirrezabalaga, N. Allende, A. F. Brana, C. Mendez, J. A. Salas, *Chem. Biol.* **2002**, *9*, 721-729.

[2] M. K. Kharel, H. Lian, J. Rohr, *Org. Biomol. Chem.* **2011**, *9*, 1799-1808.

[3] M. M. Bradford, *Anal Biochem* **1976**, *72*, 248-254.

[4] J. White-Phillip, C. J. Thibodeaux, H. W. Liu, *Methods Enzymol.* **2009**, *459*, 521-544.