Supplementary Information

The Development of a Broad-Spectrum Retaining β-*Exo*-Galactosidase Activity-Based Probe

Chi-Lin Kuo,[a] Qin Su,[a] Adrianus M. C. H. van den Nieuwendijk,[b] Thomas J. M. Beenakker,[a] Wendy A. Offen,[c] Lianne I. Willems,[a,c] Rolf. G. Boot,[a] Alexi J. Sarris,[b] André R. A. Marques,[a] Jeroen D. C. Codée,[b] Gijsbert A. van der Marel,[b] Bogdan I. Florea,[b] Gideon J. Davies,[c] Herman S. Overkleeft*[b] and Johannes M. F. G. Aerts*[a]

[a.] Department of Medical Biochemistry, Leiden Institute of Chemistry, Leiden University, P. O. Box 9502, 2300 RA Leiden, The Netherlands.
[b.] Bioorganic Synthesis group, Leiden Institute of Chemistry, Leiden University, P. O. Box 9502, 2300 RA Leiden, The Netherlands.
[c.] Department of Chemistry, University of York, Heslington, York, YO10 5DD York, United Kingdom.
* Corresponding author.

Table of Contents

Enzymatic assay and kinetic parameters measurement

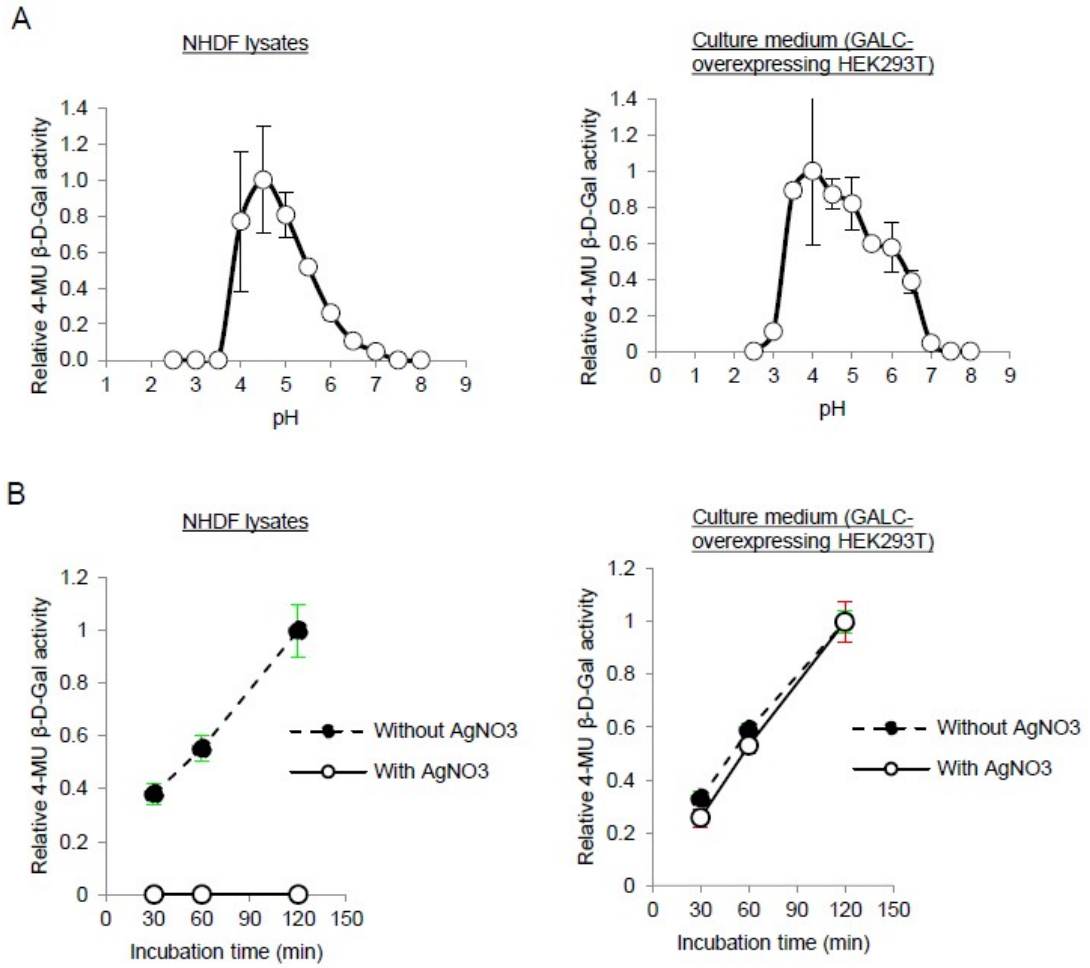


Figure S1 Enzymatic assay of discriminating activities from GLB1 and GALC in different samples. A) pH-dependent β-galactosidase activity (relative to the highest measured activity). B) Effect of $AgNO_3$ (11 µM). Error range = SD, *n* = 3 technical replicates

Table S1 Apparent IC_{50} values (nM) for compounds towards *Cj*GH35A, GBA1, and GBA2. Error range = SD, n = 2 biological replicates.

Compound	*Cj*GH35A	GBA1	GBA2
1	21.55 ± 2.14	1,560	3,290
3	27.16 ± 1.57	5,370	> 10,000
6	33.9 ± 2.15	85.8	752

Table S2 Kinetic data for compounds 1, 3 and 6 towards *Cj*GH35A. Error range = ± SD, *n* = 3 biological replicates.

	K_{inact} (min^{-1})	K_I (μM)	K_{inact} / K_I (min^{-1} $μM^{-1}$)
1	0.99 ± 0.11	1.90 ± 0.39	0.52 ± 0.06
3	2.19 ± 0.55	3.37 ± 1.20	0.65 ± 0.16
6	2.03 ± 0.63	0.11 ± 0.04	18.98 ± 5.88

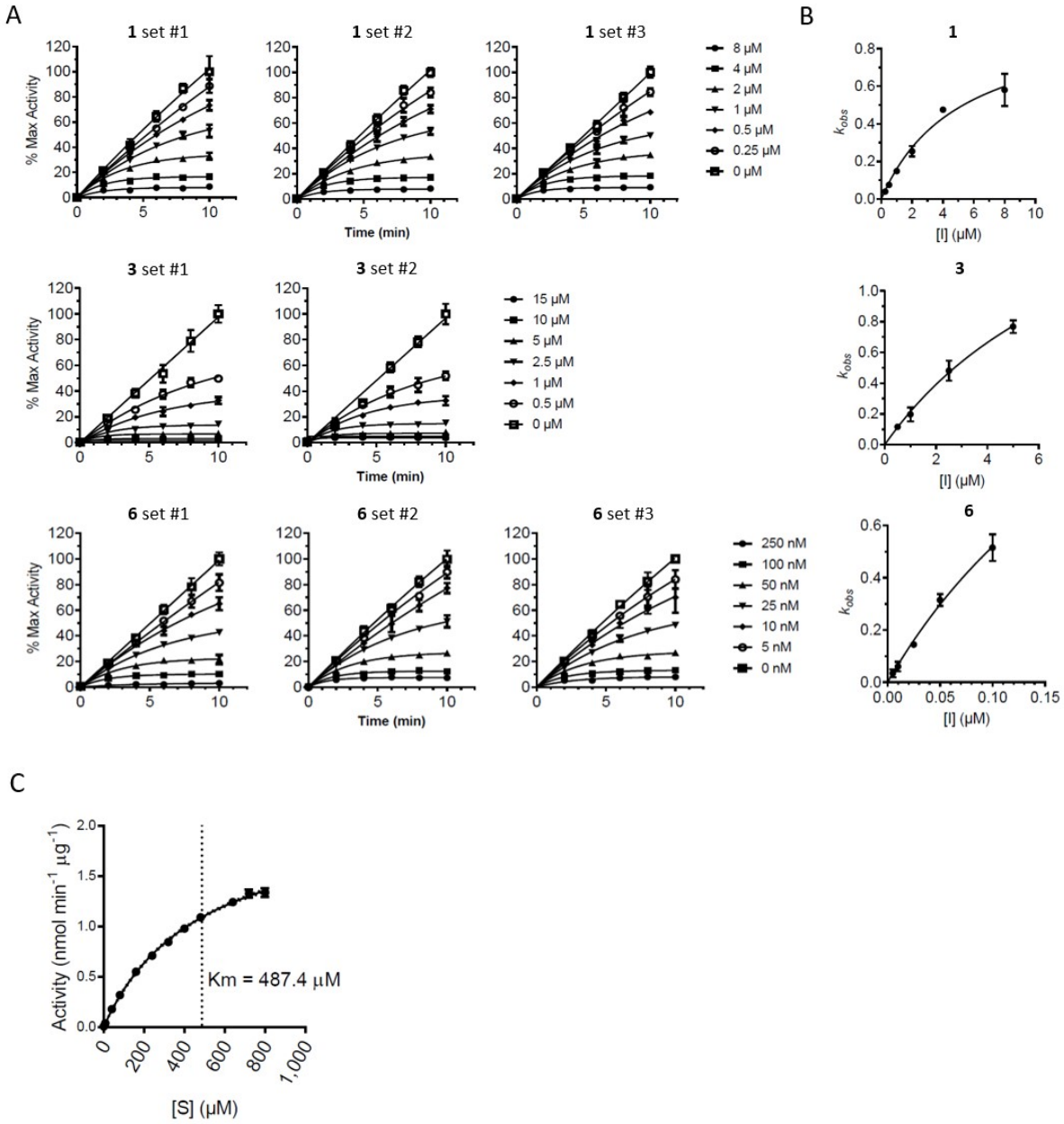


Figure S2 . Measurement of inhibition kinetic parameters by compounds **1**, **3** and **6** towards *Cj*GH35. A) Processing curves. Error range = SD, n = 3 technical replicates. B) kobs vs [I] plots, fitted with Michaelis- Menten equation. Error range = SD, n = 3 biological replicates. C) Michaelis-Menten plot of 4-MU-β-gal towards *Cj*GH35. Error range = SD, n = 3 technical replicates

Labeling of mouse tissue extracts with ABP 6

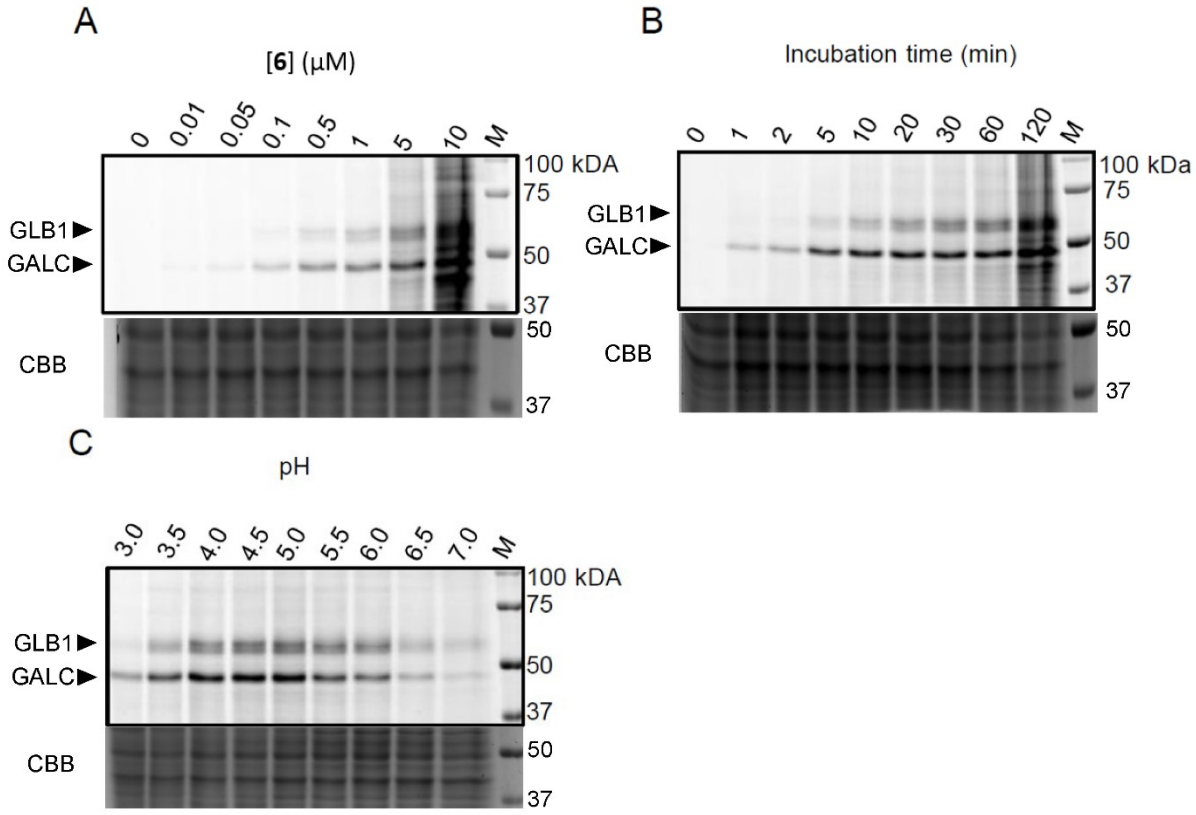


Figure S3-1 Coomassie Brilliant Blue (CBB) of labeling of mouse kidney extracts with Cy5-ABP **6**.

ABP 6 labels mouse tissue extracts

20 µg total protein from extracts of mouse kidney, brain, epididymis, testis, duodenum, and intestine were diluted in 10 µL McIlvaine buffer (150 mM, various pH from 3.0 to 7.0). Samples were then pre-incubated with 2.5 µL ABP **JJB70** [1] (diluted in McIlvaine buffer pH 3.0 – 7.0, 200 nM during incubation) for 30 min at 37 °C, and then 2.5 µL ABP **6** (diluted in McIlvaine buffer pH 3.0 – 7.0, 1 µM during incubation) for 30 min at 37 °C, before subjected to SDS-PAGE based fluorescent detection.

For deglycosylation analysis, ABP labeled samples (60 µg protein diluted consecutively in 20 µL McIlvaine buffer pH 6.0 (5 min on ice), 5 µL ABP **JJB70** (30 min, 37 °C) and 5 µL ABP **6** (30 min, 37 °C), at identical ABP concentrations as described) were firstly desalted using Pierce 7K polyacrylamide spin column (Thermo Fisher), and a 10 µL aliquot was treated with PNGase F according to the manufacturer's protocol (New England BioLabs). Non-treated samples (20 µg protein diluted in 10 µL McIlvaine buffer pH 6.0) were similarly labeled with ABPs. Both the non-treated and PNGase F-treated samples were subjected to SDS-PAGE and fluorescence detection.

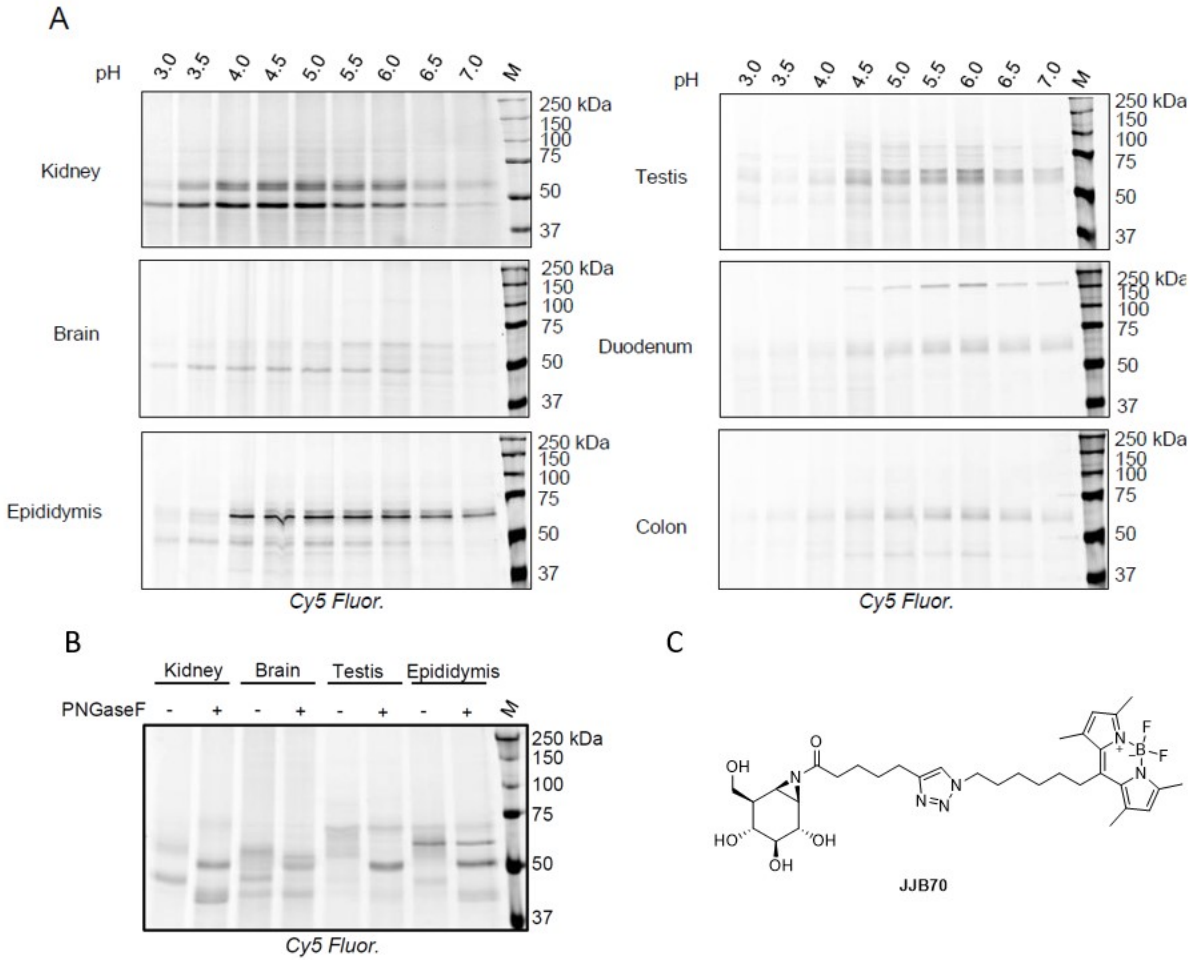


Figure S3-2 *In vitro* labeling with ABP **6** in mouse tissue extracts. A) pH-dependent labeling in mouse tissue extracts. B) ABP **6** Labeled mouse tissue extracts (pH 6.0) with or without deglycosylation by PNGaseF. C) Compounds (**JJB70**) used for the preincubation prior the incubation of ABP **6** with mouse kidney extracts

Crystallization

Methods for crystallization and complex formation of *Cj*GH35A with compound 7, data collection and structure refinement

*Cj*GH35A was expressed and purified as in Larsbrink *et al*. [2] Crystals of *Cj*GH35A were grown by the sitting drop vapour diffusion method using protein at 25 mg/ml (in 50 mM HEPES pH 7, 200 mM NaCl) mixed 1:1 with the well solution which was comprised of 2.7 M sodium acetate pH 7.4. The D-*galacto*-cyclophellitol ligand **7** was dissolved in water at 10 mM, and added to the drop containing crystals in a 1:1 volume ratio, giving a soak concentration of 5 mM. Crystals were fished after 2 days directly into liquid nitrogen without using cryoprotectant. Data were collected on beamline io3 at Diamond Light Source, processed using *DIALS* [3] and scaled with *AIMLESS.* [4] The structure was solved using PDB entry *4D1I* [2] and *REFMAC5* [5], and the ligand was built using *AceDRG*. [6] Manual model correction and ligand placement were performed using *Coot* [7] followed by cycles of REFMAC5. The programs were run in the *CCP4I2* interface. [8] In addition, the structure for *Cj*GH35A in complex with compound **1** was annotated in PDB (5JAW) and has been published in previous report. [9]

Table S3 . X-ray data collection and refinement statistics of *Cj*GH35A with **7** (PDB code: 8PEJ).

	***Cj*GH35-7**
Data collection	
Space group	*P1*
Cell dimensions	
a, *b*, *c* (Å)	99.3, 116.2, 116.4
α, β, γ (°)	90.2, 89.9, 90.1
Resolution (Å)	99.32-1.50 (1.53-1.50)
Total no. of reflections	2697222
No. unique reflections	802827
R_{sym} or R_{merge}	0.043(0.467)
R_{pim}	0.043(0.467)
$CC_{1/2}$	0.997 (0.719)
I / σI	10.4 (2.0)
Completeness (%)	96.3 (90.0)
Redundancy	3.4 (2.8)
Refinement	
No. reflections working set	762488
No. reflections test set	40267
R_{work} / R_{free}	0.16/0.18
No. atoms	
Protein	34404
Ions	35
Ligands	152
Water	4413
B-factors (Å^2)	
Protein	24.3
Ions	28.1
Ligands	23.3
Water	35.1
R.m.s. deviations	
Bond lengths (Å)	0.011
Bond angles (°)	1.709
Ramachandran plot residues	
In most favorable regions (%)	95.7
In allowed regions (%)	3.1
PDB code	8PEJ

Table S4 . X-ray data collection and refinement statistics of *Cj*GH35A with **1**

***Cj*GH35A in complex with 1**	
Data collection	
Space group	*P1*
Cell dimensions	
***a*, *b*, *c* (Å)**	98.9, 115.8, 116.0
α, β, γ (°)	90.2, 90.2, 90.4
Resolution (Å)	81.99-1.6 (1.63-1.60)
R_{merge}	0.080 (0.588)
R_{pim}	0.080 (0.588)
CC(1/2)	0.979 (0.482)
I /σI	6.0 (1.2)
Completeness (%)	95.7 (94.1)
Redundancy	1.8 (1.8)
Refinement	
Resolution (Å)	1.60
No. reflections	650665
R_{work} / R_{free}	0.20/0.21
No. atoms	
Protein	33482
Ligand/ion	125
Water	1637
***B*-factors (Å²)**	
Protein	25
Ligand/ion	20
Water	28
R.m.s. deviations	
Bond lengths (Å)	0.018
Bond angles (°)	1.927
PDB code	5JAW

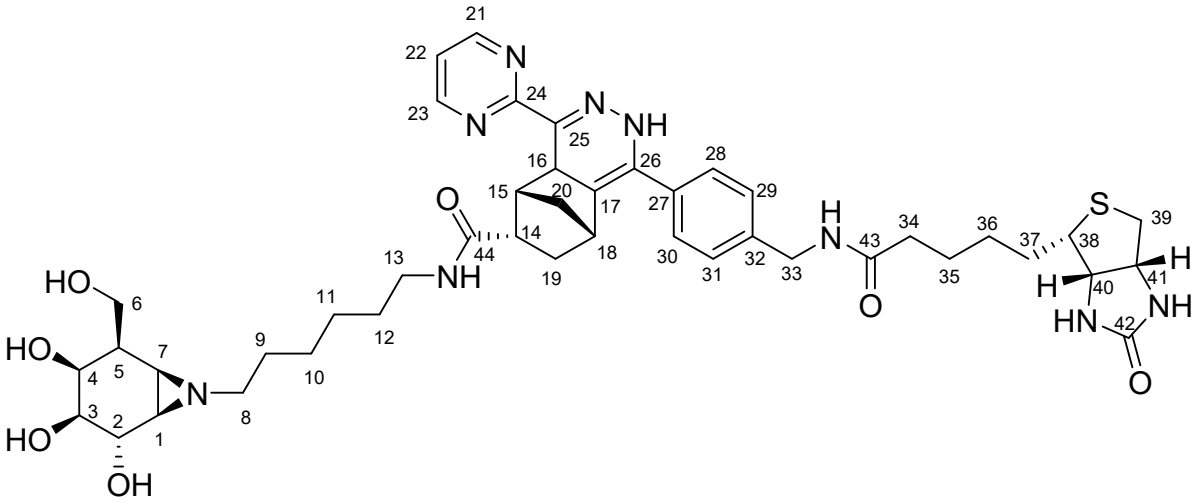
*Values in parentheses are for highest-resolution shell.

Purification and Characterization of ABP 5 and NMR data

HPLC purification was performed on an Agilent 1260 Infinity II HPLC using a reversed phase Gemini C18 column (10 x 250 mm size, particle size 5 µm) employing a 22 – 25% buffer B gradient (10 min) at a 5 mL / min flow. (Buffer A = aqueous 50 mM NH_4HCO_3 ; Buffer B = acetonitrile).

HRMS was measured on a Waters Synapt G2-Si apparatus equipped with an electrospray ion source in positive mode (ESI-TOF), injection of 2 µl of a 2 µM solution via a Waters NanoEquity system, using LeuEnk (m/z=556.2771) as "lock mass". Source voltage 3,5 kV, at 275 °C, mass range m/z = 160-2000. Eluent used: acetonitrile : water = 1:1 (v/v), supplemented with 0.1% formic acid.

Atom numbering in β-Gal-aziridine ABP **5** used for NMR analysis:

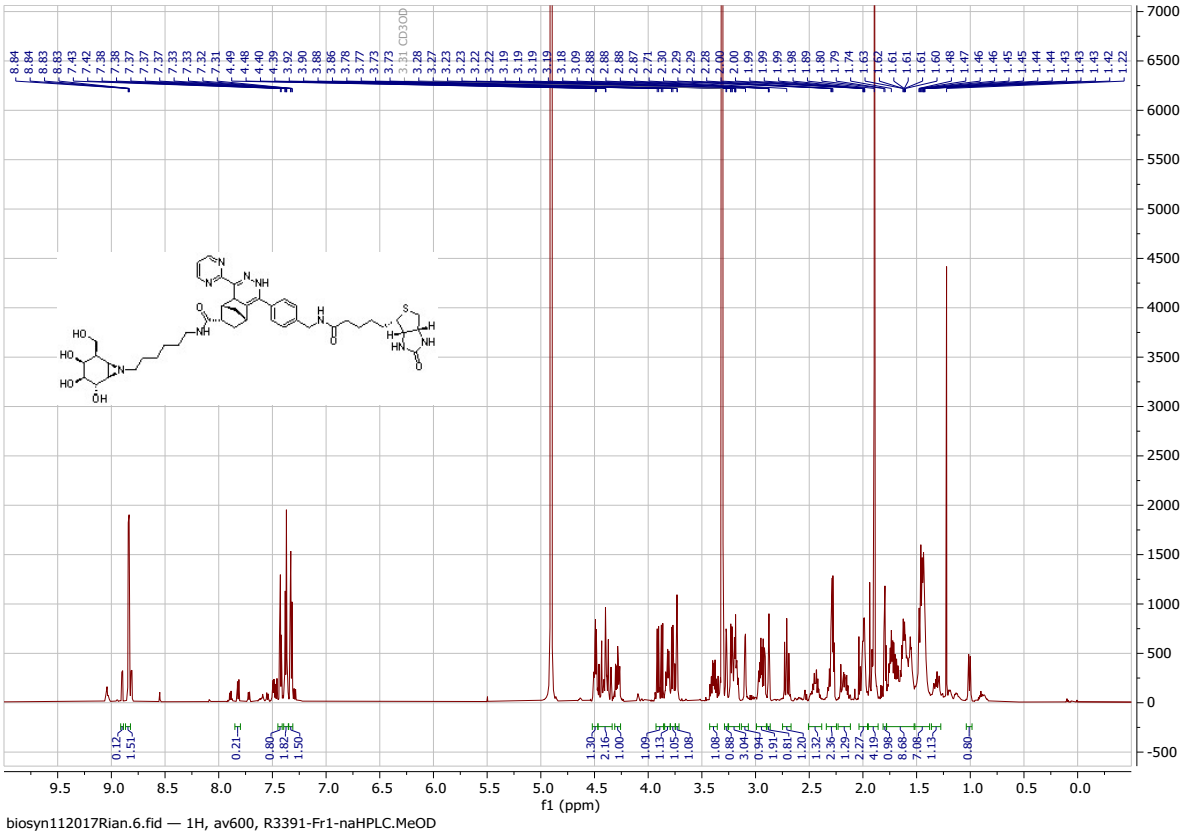


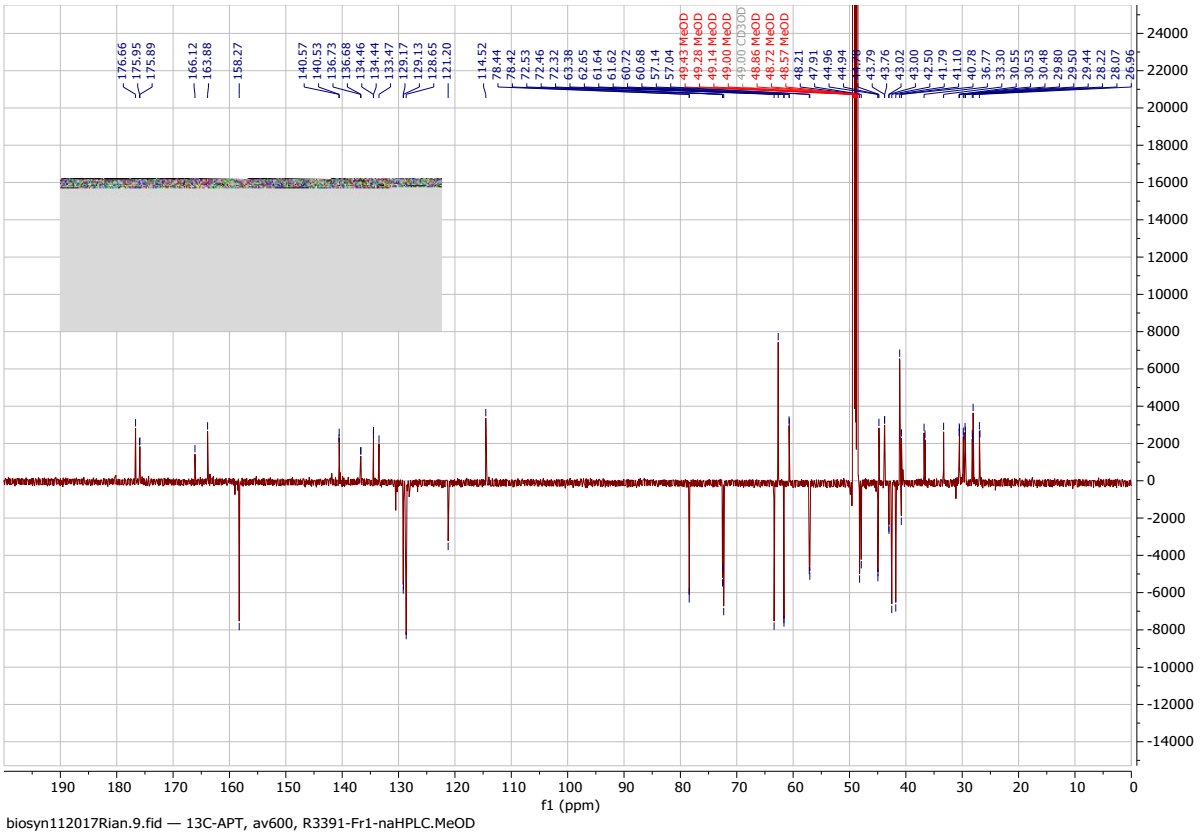
Fraction 1 (2.77 mg after lyophilization, major isomers reported (1:1 ratio)): ^{1}H NMR (600 MHz, MeOD) δ 8.84 (dd, *J* = 4.9, 1.0 Hz, 2H, Ar H-21,H-23), 7.43 (t, *J* = 4.9 Hz, 1H, Ar, H-22), 7.40 – 7.35 (m, 2H, Ar H-28, H-30), 7.32 (dd, *J* = 8.4, 1.8 Hz, 2H, Ar H-29, H-31), 4.52 – 4.47 (m, 1H, H-41), 4.46 – 4.34 (m, 2H, CH_2-33), 4.28 (ddd, *J* = 12.0, 7.9, 4.5 Hz, 1H, H-40), 3.91 (d, *J* = 7.9 Hz, 0.5H, H-2), 3.87 (d, *J* = 7.9 Hz, 0.5H, H-2), 3.82 (ddd, *J* = 10.4, 7.0, 1.7 Hz, 1H, H-6), 3.79 – 3.75 (m, 1H, H-6'), 3.73 (t, *J* = 2.3 Hz, 1H, H-4), 3.53 – 3.36 (m, 1H, H-13), 3.27 (d, *J* = 3.6 Hz, 1H, H-18), 3.25 – 3.15 (m, 3H, H-3, H-13', H-38), 3.10 (dt, *J* = 4.3, 1.4 Hz, 1H, H-16), 2.99 – 2.90 (m, 2H, H-14, H-39), 2.88 (d, *J* = 2.4 Hz, 1H, H-15), 2.71 (t, *J* = 12.0 Hz, 1H, H-39'), 2.50 – 2.39 (m, 1H, H-8), 2.28 (td, *J* = 7.3, 3.9 Hz, 2H, CH_2-34), 2.22 – 2.13 (m, 1H, H-8'), 2.02 – 1.96 (m, 2H, H-1, H-5), 1.92 – 1.86 (m, 2H, CH_2-19 (obscured by solvent)), 1.79 (t, *J* = 5.7 Hz, 1H, H-7), 1.77 – 1.53 (m, 8H, 4 x CH_2), 1.50 – 1.39 (m, 7H, H-20, 3 x CH_2), 1.01 (d, *J* = 8.5 Hz, 1H, H-20'). ^{13}C NMR (151 MHz, MeOD) δ 176.66, 175.95 & 175.89 (C-44, C-43), 166.12 (C-42), 163.88 (C-24), 158.27 (C-21,23), 140.57 & 140.53 (C-25), 136.73 & 136.68 (C-26), 134.46 & 134.44 (C-32), 133.47 (C-27), 129.17 (C-28), 129.13 (C-30), 128.65 (C29,31), 121.20 (C-22), 114.52 (C-17), 78.44 & 78.42 (C-3), 72.53 & 72.46 (C-2), 72.32 (C-4), 63.38 (C-40), 62.65 (C-6), 61.64 & 61.62 (C-41), 60.72 & 60.68 (C-8), 57.14 & 57.04 (C-38), 48.21 (C-14), 47.91 (C-16), 44.96 & 44.94 (C-7), 44.78 (C-20), 43.79 & 43.76 (C-33), 43.02 & 43.00 (C-15), 42.50 (C-1), 41.79 (C-5), 41.10 (C-39), 40.81 (C-18), 40.78 (C-13), 36.77 & 36.60 (C-34), 33.30 (C-19), 30.55 & 30.53, 30.48, 29.80 & 29.71, 29.50 & 29.44, 28.22 & 28.18, 28.07, 26.96 & 26.85 (7 x CH_2, C-9 – C-12 and C-35 – C-37). HRMS calculated for $[C_{44}H_{59}N_9O_7S+H]^+$ 858.4331; Found 858.4331.

Fraction 2 (4.09 mg after lyophilization, major isomers reported (1:1 ratio)): ^{1}H NMR (600 MHz, MeOD) δ 8.91 (dd, *J* = 4.8, 1.2 Hz, 2H, Ar H-21,H-23), 7.49 (d, *J* = 8.2 Hz, 2H, Ar H-28, H-30), 7.45 (tdd, *J* = 4.8, 2.0, 0.7 Hz, 1H, Ar, H-22), 7.36 (d, *J* = 8.2 Hz, 2H, Ar H-29, H-31), 4.49 (ddd, *J* = 7.9, 5.0, 0.9 Hz, 1H, H-41), 4.46 – 4.36 (m, 2H, H-33), 4.28 (dt, *J* = 7.9, 4.0 Hz, 1H, H-40), 4.17 (d, *J* = 3.1 Hz, 1H, H-15), 3.89 (dd, *J* = 9.0, 7.9 Hz, 1H, H-2), 3.84 – 3.78 (m, 1H, H-6), 3.78 – 3.74 (m, 1H, H-6'), 3.73 – 3.70 (m, 1H, H-4), 3.25 – 3.14 (m, 4H, H-3, H-13, H-14, H-38), 3.01 (m, 1H, H-13'), 2.93 (ddd, *J* = 12.8, 5.0, 2.2 Hz, 1H, H-39), 2.79 (d, *J* = 4.4 Hz, 1H, H-18), 2.71 (dd, *J* = 12.6, 2.2 Hz, 1H, H-39'), 2.63 (br s, 1H, H-16), 2.37 – 2.30 (m, 1H, H-8), 2.28 (t, *J* = 7.3 Hz, 2H, CH_2-34), 2.18 (td, *J* = 12.5, 4.7 Hz, 1H, H-19), 2.13 – 2.03 (m, 1H, H-8'), 2.00 – 1.94

(m, 2H, H-1, H-5), 1.86 (dt, *J* = 12.6, 3.8 Hz, 1H, H-19'), 1.77 (dd, *J* = 5.7, 4.5 Hz, 1H, H-7), 1.75 – 1.58 (m, 4H, CH_2-12, CH_2-35), 1.57 (dd, *J* = 9.7, 1.7 Hz, 1H, H-20), 1.44 (p, *J* = 7.7 Hz, 2H, CH_2-36), 1.41 – 1.27 (m, 4H, CH_2-9, CH_2-37), 1.26 – 1.06 (m, 4H, CH_2-10, CH_2-11), 1.04 (dt, *J* = 9.7, 2.2 Hz, 1H, H-20'). ^{13}C NMR (151 MHz, MeOD) δ 176.78 (C-43), 175.98 (C-44), 166.13 (C-42), 160.39 & 160.38 (C-24), 158.63 (C-21,23), 143.25 & 143.23 (C-25), 141.08 (C-26), 135.72 & 135.71 (C-32), 131.85 (C-27), 129.95 (C-28,30), 128.41 (C-29,31), 122.72 (C-17), 121.10 (C-22), 78.44 & 78.43 (C-3), 72.52 & 72.51 (C-2), 72.26 (C-4), 63.37 & 63.35 (C-40), 62.61 & 62.59 (C-6), 61.62 (C-41), 60.64 (C-8), 57.06 & 57.03 (C-38), 48.20 (C-14), 45.88 (C-16), 44.92 & 44.90 (C-7), 44.41 & 44.38 (C-15), 44.19 (C-20), 43.82 (C-33), 42.46 (C-1), 41.78 & 41.76 (C-5), 41.08 (C-39), 40.33 (C-13), 39.87 (C-18), 36.75 & 36.74 (C-34), 33.92 (C-19), 30.41 (C-9), 30.29 (C-37), 29.79 & 29.77 (C-36), 29.51 & 29.48 (C-12), 28.03 & 27.99 (C-10), 27.76 & 27.73 (C-11), 26.91 & 26.89 (C-35). HRMS calculated for $[C_{44}H_{59}N_9O_7S+H]^+$ 858.4331; Found 858.4336.

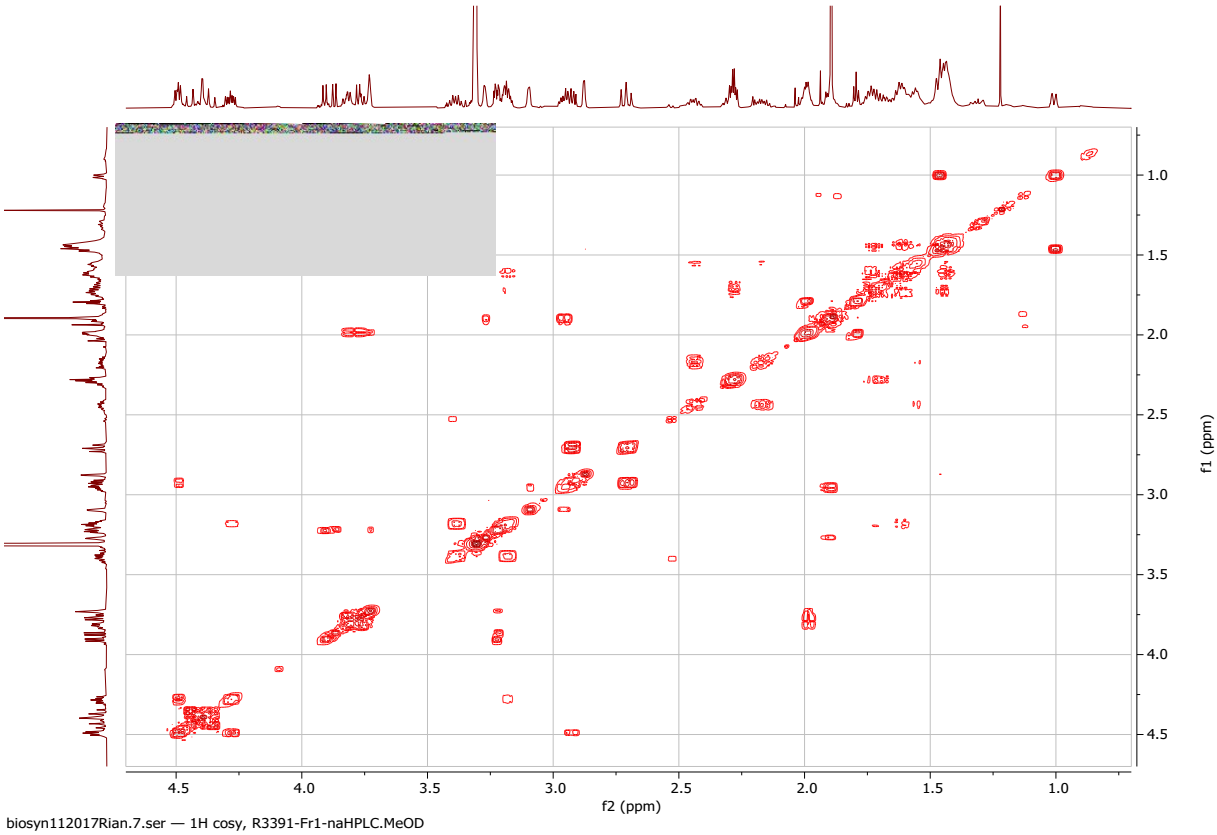
^{1}H- and ^{13}C-NMR for β-Gal-aziridine ABP **5** Fraction-1 in CD_3OD.



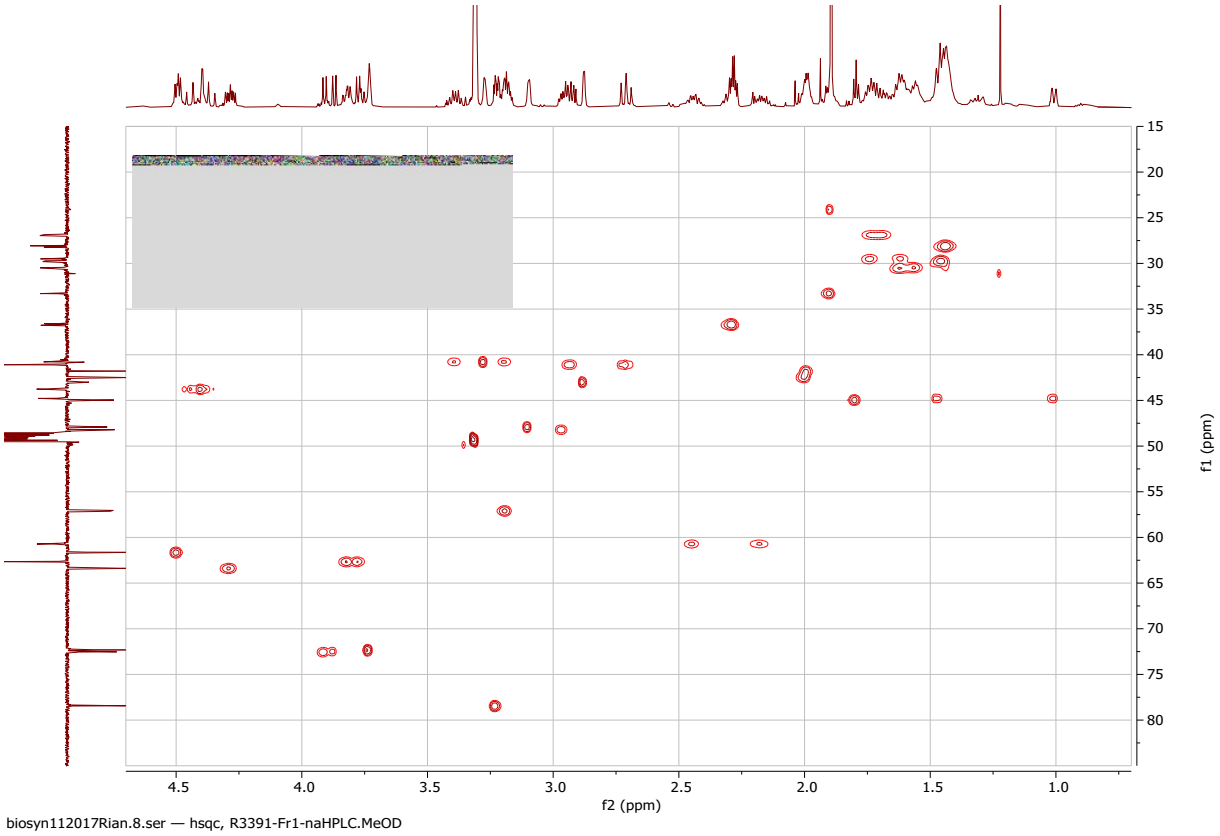


biosyn112017Rian.9.fid — 13C-APT, av600, R3391-Fr1-naHPLC.MeOD

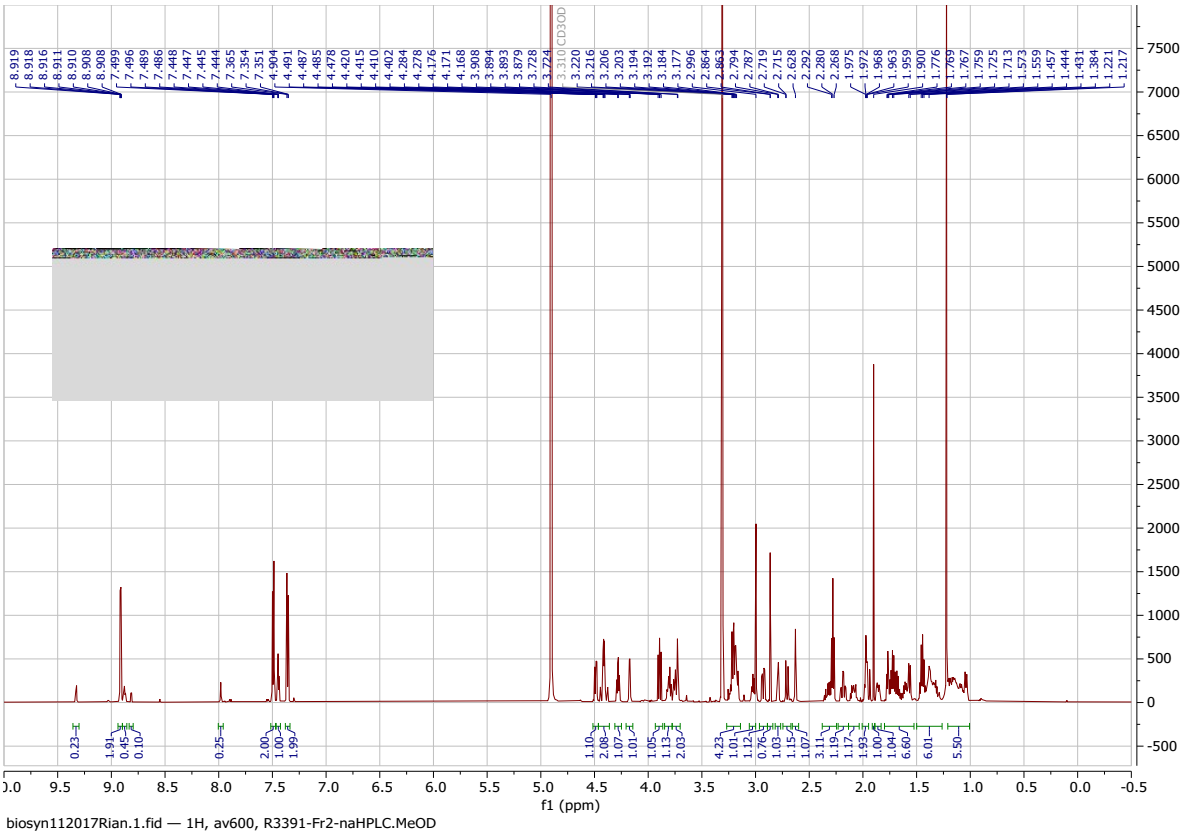
^{1}H-COSY and HSQC 2D-NMR spectra for β-Gal-aziridine ABP **5** Fraction-1 in CD_3OD.



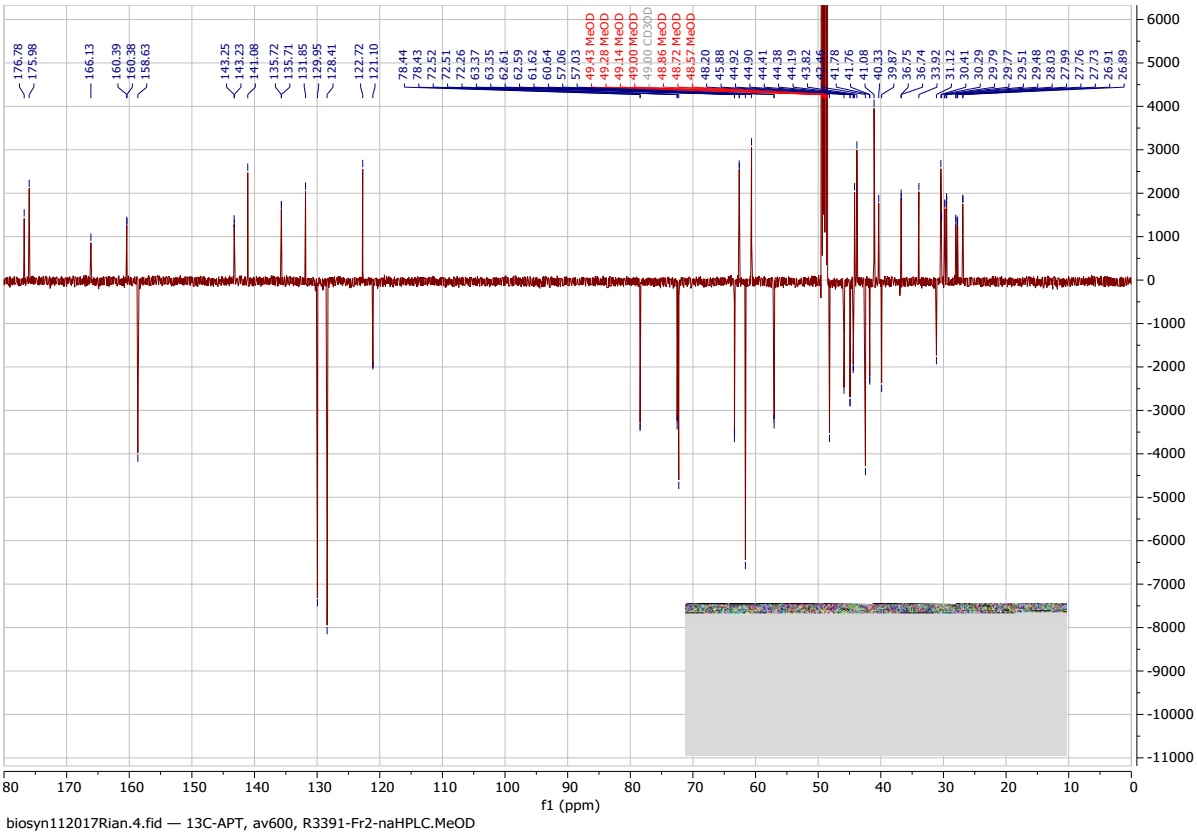
biosyn112017Rian.7.ser — 1H cosy, R3391-Fr1-naHPLC.MeOD



^{1}H- and ^{13}C-NMR for β-Gal-aziridine ABP **5** Fraction-2 in CD_3OD.

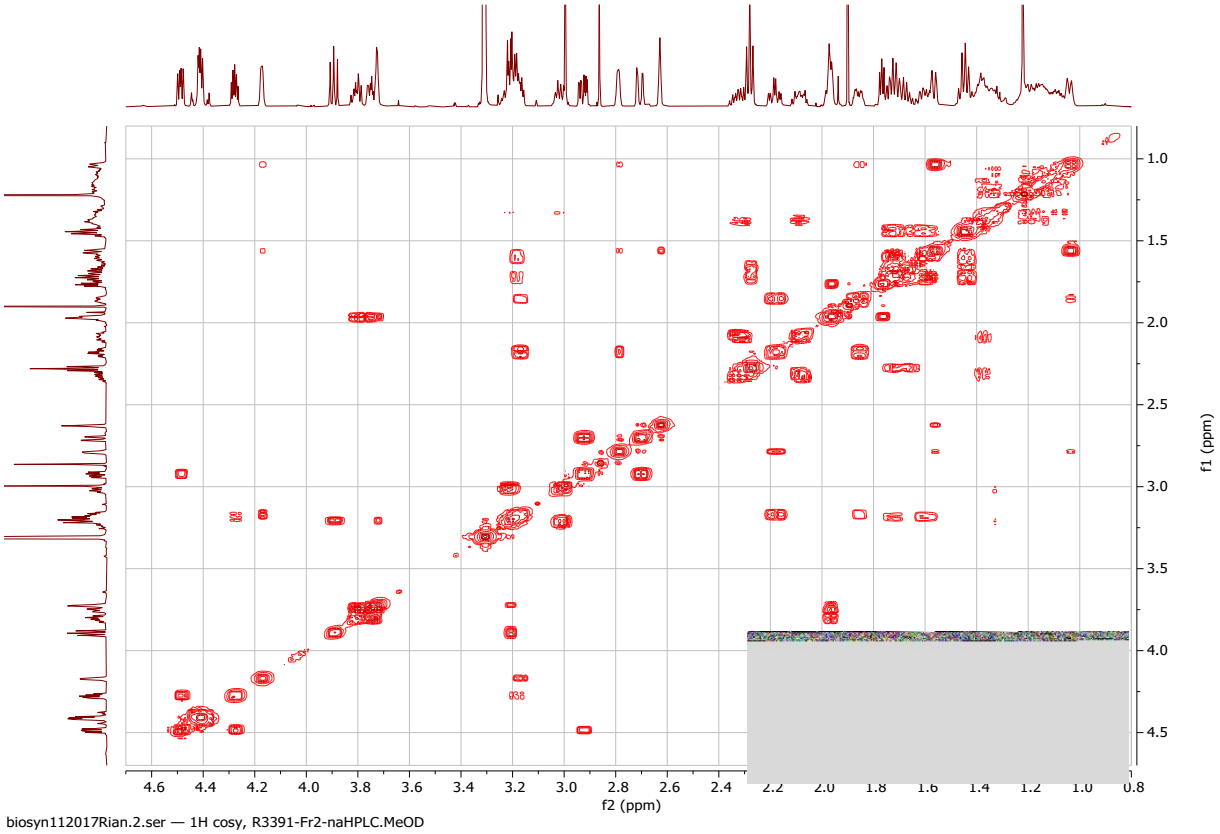


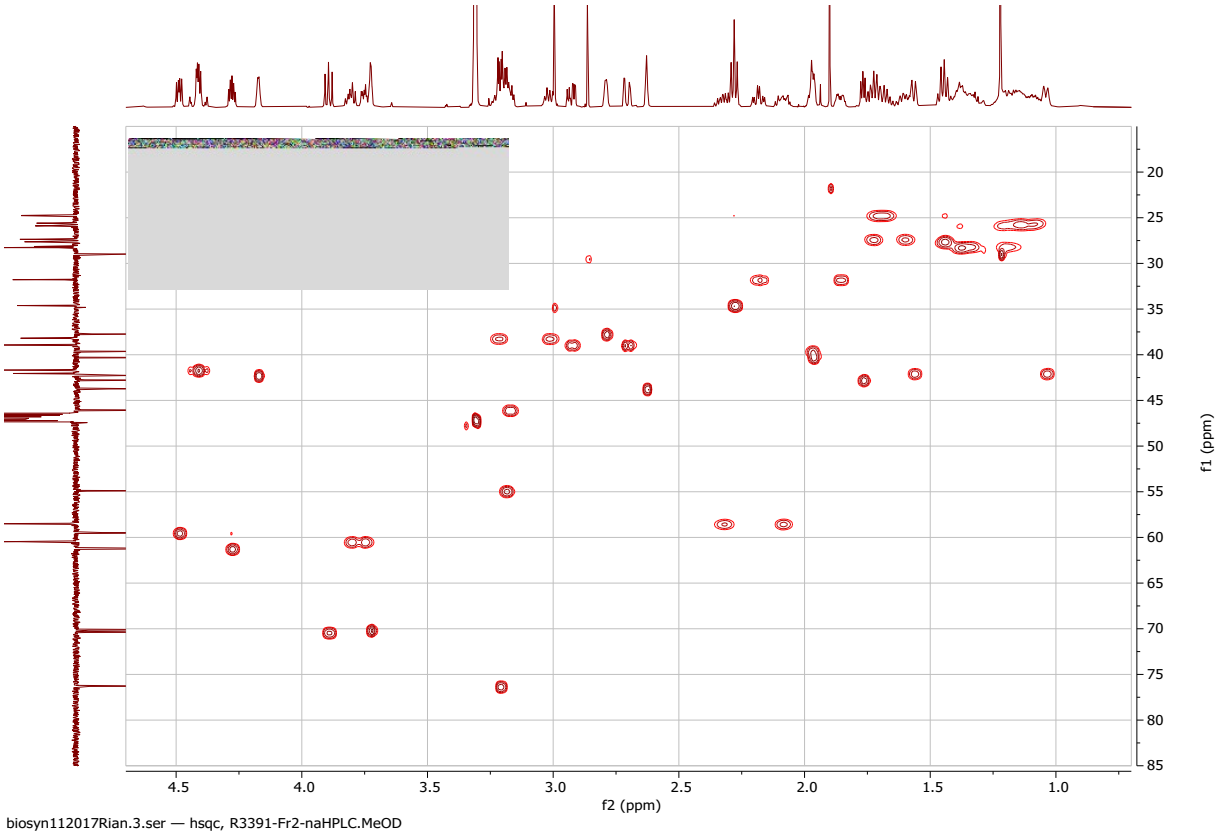
biosyn112017Rian.1.fid — 1H, av600, R3391-Fr2-naHPLC.MeOD



biosyn112017Rian.4.fid — 13C-APT, av600, R3391-Fr2-naHPLC.MeOD

^{1}H-COSY and HSQC 2D-NMR spectra for β-Gal-aziridine ABP **5** Fraction-2 in CD_3OD.





References

1. W. W. Kallemeijn, K. Y. Li, M. D. Witte, A. R. Marques, J. Aten, S. Scheij, J. Jiang, L. I. Willems, T. M. Voorn-Brouwer, C. P. van Roomen, R. Ottenhoff, R. G. Boot, H. van den Elst, M. T. Walvoort, B. I. Florea, J. D. Codee, G. A. van der Marel, J. M. Aerts and H. S. Overkleeft, *Angew Chem Int Ed Engl*, 2012, **51**, 12529-12533.
2. J. Larsbrink, A. J. Thompson, M. Lundqvist, J. G. Gardner, G. J. Davies and H. Brumer, *Mol Microbiol*, 2014, **94**, 418-433.
3. G. Winter, D. G. Waterman, J. M. Parkhurst, A. S. Brewster, R. J. Gildea, M. Gerstel, L. Fuentes-Montero, M. Vollmar, T. Michels-Clark, I. D. Young, N. K. Sauter and G. Evans, *Acta Crystallogr D Struct Biol*, 2018, **74**, 85-97.
4. P. R. Evans and G. N. Murshudov, *Acta Crystallogr D Biol Crystallogr*, 2013, **69**, 1204-1214.
5. G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin, *Acta Crystallogr D Biol Crystallogr*, 2011, **67**, 355-367.
6. F. Long, R. A. Nicholls, P. Emsley, S. Graaeulis, A. Merkys, A. Vaitkus and G. N. Murshudov, *Acta Crystallogr D Struct Biol*, 2017, **73**, 112-122.
7. P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr D Biol Crystallogr*, 2010, **66**, 486-501.
8. J. Agirre, M. Atanasova, H. Bagdonas, C. B. Ballard, A. Basle, J. Beilsten-Edmands, R. J. Borges, D. G. Brown, J. J. Burgos-Marmol, J. M. Berrisford, P. S. Bond, I. Caballero, L. Catapano, G. Chojnowski, A. G. Cook, K. D. Cowtan, T. I. Croll, J. E. Debreczeni, N. E. Devenish, E. J. Dodson, T. R. Drevon, P. Emsley, G. Evans, P. R. Evans, M. Fando, J. Foadi, L. Fuentes-Montero, E. F. Garman, M. Gerstel, R. J. Gildea, K. Hatti, M. L. Hekkelman, P. Heuser, S. W. Hoh, M. A. Hough, H. T. Jenkins, E. Jimenez, R. P. Joosten, R. M. Keegan, N. Keep, E. B. Krissinel, P. Kolenko, O. Kovalevskiy, V. S. Lamzin, D. M. Lawson, A. A. Lebedev, A. G. W. Leslie, B. Lohkamp, F. Long, M. Maly, A. J. McCoy, S. J. McNicholas, A. Medina, C. Millan, J. W. Murray,

G. N. Murshudov, R. A. Nicholls, M. E. M. Noble, R. Oeffner, N. S. Pannu, J. M. Parkhurst, N. Pearce, J. Pereira, A. Perrakis, H. R. Powell, R. J. Read, D. J. Rigden, W. Rochira, M. Sammito, F. Sanchez Rodriguez, G. M. Sheldrick, K. L. Shelley, F. Simkovic, A. J. Simpkin, P. Skubak, E. Sobolev, R. A. Steiner, K. Stevenson, I. Tews, J. M. H. Thomas, A. Thorn, J. T. Valls, V. Uski, I. Uson, A. Vagin, S. Velankar, M. Vollmar, H. Walden, D. Waterman, K. S. Wilson, M. D. Winn, G. Winter, M. Wojdyr and K. Yamashita, *Acta Crystallogr D Struct Biol*, 2023, **79**, 449-461.

9. S. P. Schroder, J. W. van de Sande, W. W. Kallemeijn, C. L. Kuo, M. Artola, E. J. van Rooden, J. Jiang, T. J. M. Beenakker, B. I. Florea, W. A. Offen, G. J. Davies, A. J. Minnaard, J. Aerts, J. D. C. Codee, G. A. van der Marel and H. S. Overkleeft, *Chem Commun (Camb)*, 2017, **53**, 12528-12531.