

SUPPLEMENTARY INFORMATION

Breaking Boundaries in Diabetic Nephropathy Treatment: Design and Synthesis of Novel Steroidal SGLT2 Inhibitors

Geetmani Singh Nongthombam,^{a, ‡} Semim Akhtar Ahmed,^{a,b, ‡} Kangkon Saikia,^a Sanjib Gogoi^c and Jagat Chandra Borah^{a,b,d, *}

^a Chemical Biology Laboratory, Life Sciences Division, Institute of Advanced Study in Science and Technology, Guwahati-781035, Assam, India.

^b Academy of Scientific and Innovative Research (AcSIR), Ghaziabad-201002, India.

^c Applied Organic Chemistry, Chemical Sciences & Technology Division, CSIR-North East Institute of Science and Technology, Jorhat 785006, India.

^d Department of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research, Guwahati 781101, Assam, India.

*E-mail: borahjc@gmail.com

Contents	Page No.
Materials and procedure	S4-S7
Table S1. Virtually screened steroidal compounds as SGLT2 inhibitors by high-precision molecular docking.	S8-S12
Spectral Data of Synthesised Steroidal pyrimidine compounds	S12-S15
Figure S1. ¹ H NMR spectrum of 5a in CDCl ₃ at 298 K.	S16
Figure S2. ¹³ C{ ¹ H} NMR spectrum of 5a in CDCl ₃ at 298 K.	S16
Figure S3. HRMS of 5a .	S17
Figure S4. ¹ H NMR spectrum of 5b in CDCl ₃ at 298 K.	S17
Figure S5. ¹³ C{ ¹ H} NMR spectrum of 5b in CDCl ₃ at 298 K.	S18
Figure S6. HRMS of 5b .	S18
Figure S7. ¹ H NMR spectrum of 5c in CDCl ₃ at 298 K.	S19
Figure S8. ¹³ C{ ¹ H} NMR spectrum of 5c in CDCl ₃ at 298 K.	S19
Figure S9. HRMS of 5c .	S20
Figure S10. ¹ H NMR spectrum of 5d in CDCl ₃ at 298 K.	S20
Figure S11. ¹³ C{ ¹ H} NMR spectrum of 5d in CDCl ₃ at 298 K.	S21
Figure S12. HRMS of 5d .	S21
Figure S13. ¹ H NMR spectrum of 5e in CDCl ₃ at 298 K.	S22
Figure S14. ¹³ C{ ¹ H} NMR spectrum of 5e in CDCl ₃ at 298 K.	S22
Figure S15. HRMS of 5e .	S23
Figure S16. ¹ H NMR spectrum of 6a in CDCl ₃ at 298 K.	S23
Figure S17. ¹³ C{ ¹ H} NMR spectrum of 6a in CDCl ₃ at 298 K.	S24
Figure S18. HRMS of 6a .	S24
Figure S19. ¹ H NMR spectrum of 6b in CDCl ₃ at 298 K.	S25
Figure S20. ¹³ C{ ¹ H} NMR spectrum of 6b in DMSO-D ₆ at 298 K.	S25
Figure S21. HRMS of 6b .	S26
Figure S22. ¹ H NMR spectrum of 6c in CDCl ₃ at 298 K.	S26
Figure S23. ¹³ C{ ¹ H} NMR spectrum of 6c in CDCl ₃ at 298 K.	S27

Figure S24. HRMS of 6c .	S27
Figure S25. ¹ H NMR spectrum of 6d in CDCl ₃ at 298 K.	S28
Figure S26. ¹³ C{ ¹ H} NMR spectrum of 6d in CDCl ₃ at 298 K.	S28
Figure S27. HRMS of 6d .	S29
Figure S28. ¹ H NMR spectrum of 8a in CDCl ₃ at 298 K.	S29
Figure S29. ¹³ C{ ¹ H} NMR spectrum of 8a in CDCl ₃ at 298 K.	S30
Figure S30. HRMS of 8a .	S30
Figure S31. ¹ H NMR spectrum of 8b in CDCl ₃ at 298 K.	S31
Figure S32. ¹³ C{ ¹ H} NMR spectrum of 8b in CDCl ₃ at 298 K.	S31
Figure S33. HRMS of 8b .	S32
Figure S34. ¹ H NMR spectrum of 8c in CDCl ₃ at 298 K.	S32
Figure S35. ¹³ C{ ¹ H} NMR spectrum of 8c in CDCl ₃ at 298 K.	S33
Figure S36. HRMS of 8c .	S33
Figure S37. ¹ H NMR spectrum of 8d in CDCl ₃ at 298 K.	S34
Figure S38. ¹³ C{ ¹ H} NMR spectrum of 8d in CDCl ₃ at 298 K.	S34
Figure S39. HRMS of 8d .	S35
Figure S40. ¹ H NMR spectrum of 9a in CDCl ₃ at 298 K.	S35
Figure S41. ¹³ C{ ¹ H} NMR spectrum of 9a in CDCl ₃ at 298 K.	S36
Figure S42. HRMS of 9a .	S36
Figure S43. ¹ H NMR spectrum of 9b in DMSO-D ₆ at 298 K.	S37
Figure S44. ¹³ C{ ¹ H} NMR spectrum of 9b in DMSO-D ₆ at 298 K.	S37
Figure S45. HRMS of 9b .	S38
Figure S46. ¹ H NMR spectrum of 9c in CDCl ₃ at 298 K.	S38
Figure S47. ¹³ C{ ¹ H} NMR spectrum of 9c in CDCl ₃ at 298 K.	S39
Figure S48. HRMS of 9c .	S39
Table S2. Calculated Physicochemical property of 9a	S40
Table S3. Calculated Medicinal Chemistry Parameters of 9a	S40-S42
Table S4. Calculated Absorption table of 9a	S42
Table S5. Calculated Distribution table of 9a	S43
Table S6. Calculated Metabolism table of 9a	S43-S44
Table S7. Calculated Excretion table of 9a	S44
Table S8. Calculated Toxicity table of 9a	S44-S46
Table S9. Calculated Environmental toxicity table of 9a	S46
Table S10. Calculated Tox21 pathway table of 9a	S46-S47
Table S11. Calculated Toxicophore Rules table of 9a	S47
Table S12. HPLC chromatogram of 9a with percentage purity	S48
Figure S49. Full-length blots of immunoblotting data	S49-S51
Figure S50. SGLT2 mediated glucose uptake analysis on HG-treated NRK-52E cells upon canagliflozin and 9a treatment cells.	S52

Materials and Procedure

General: Commercial reagents were purchased from Sigma-Aldrich, Merck and TCI and were used without further purification. The ^1H and $^{13}\text{C}\{^1\text{H}\}$ NMR spectra were recorded at ambient temperature on a 400 MHz (100 MHz for $^{13}\text{C}\{^1\text{H}\}$) NMR spectrometer with CDCl_3 or DMSO-d_6 as the solvent. Chemical shifts were referenced to the residual peaks of the solvent [CDCl_3 : $\delta = 7.26$ (^1H), 77.16 ppm (^{13}C); DMSO-d_6 : $\delta = 2.50$ (^1H), 39.52 ppm (^{13}C)]. Column chromatography was performed using EM Silica gel 60 (100-200 mesh). HRMS were obtained using a Xevo XS QT mass spectrometer, Waters ACQUITY UHPLC. Shimadzu Shimpack GIS C16 analytical column (5 μm , 250 \times 4.6 mm) was used for HPLC analysis with a mobile phase of water (pump A) and acetonitrile-methanol (50:50) (pump B). An isocratic method with 70 % Acetonitrile-Methanol (50:50) with a flow rate of 1 ml/min for 20 minutes was used to run a sample injection volume of 20 μl . The HPLC chromatogram trace and purity of our main target synthesized compound (**9a**) for which the detailed *in-vitro* study was conducted is included in the supporting information file (**Table S12**) (Purity > 96%) and the purity for all the synthesized compounds were also verified by NMR and HRMS to be > 95 %. NRK-52E cells were purchased from NCCS, Pune. The culture medium comprised of high glucose DMEM with 10 % FBS. Trypsin (Cat#25200072), antibiotic antimycotic (Cat#15240062), DPBS (Cat#21300025), and 6-NBDG (Cat# N23106) were purchased from Invitrogen. Sodium buffer (Na^+ buffer): 140 mM NaCl, 5 mM KCl, 2.5 mM CaCl_2 , 1 mM MgSO_4 , 1 mM KH_2PO_4 , 10 mM HEPES (pH 7.4). Sodium free buffer (Na^+ free buffer): 140 mM *N*-methyl-D-glucamine, 5 mM KCl, 2.5 mM CaCl_2 , 1 mM MgSO_4 , 1 mM KH_2PO_4 , 10 mM HEPES (pH 7.4). 33 mM high glucose solution, 6-NBDG solution, Triton X (0.1 %) pH 10 in 1XPBS, DCFDA solution, DMSO. PMSF (Roche, 10837091001), Halt protease and phosphatase inhibitor cocktail (Biorad 78441) were supplied by Thermo Scientific. Sodium dodecyl sulphate (SDS, Cat#L3771), Skim milk (Cat#GRM1254), Tris base (Cat#TC072) and Sodium Chloride (Cat#TC046) were purchased from HiMedia. SGLT2 and Anti-beta Actin (HRP) (Cat#49900) antibody were bought from Abcam. AMPK α (Cat#2532), phospho-AMPK α (Thr172) Rabbit mAb (Cat#2535), NF- κB (Cat#4882P), pNF- κB p65 (Ser536) (Cat#3031), Anti-rabbit IgG HRP linked secondary antibody (Cat#7074) were obtained from Cell Signaling Technology (Beverly, MA, USA). NOX-4 (Cat#MA5-32090) and IL-6 (Cat# M620) were purchased from Invitrogen. TGF beta 1 (Cat#A16640), Fibronectin (Cat#A12977) and Collagen-IV (Cat#A10710) were purchased from Abclonal.

Molecular docking and molecular dynamic simulation: The crystal structure of human SGLT2 complexed with empagliflozin was obtained from PDB (7VSI) and processed by removing waters, adding hydrogens, filling in missing side chains, generating alternate residue positions and restrain minimization in Modeller. The docking grid was generated near glucose entry site of SGLT2 based on the binding position of cocrystallized empagliflozin. The docking grid x, y and z coordinates were -2.68, 2.14 and -10.95 respectively. The docking method was validated by docking empagliflozin to the binding site and comparing its root mean square deviation (RMSD) to the cocrystallized empagliflozin, where RMSD value was less than 1 \AA . The ligand library was meticulously compiled using a diverse range of biologically potent *N*-heterocyclic scaffolds integrated with a steroid core. The structure of each of the steroidal heterocyclic compounds were generated using Marvin (Chemaxon, academic version 2024) followed by energy minimization and geometry optimization to generate the 3D conformation. The binding poses of each ligand with SGLT2 were derived by molecular docking using SeeSAR (BioSolveIt suite 2019). The compounds with highest docking score and binding energies were subjected to chemical synthesis and subsequent analysis.

The selected ligand bound SGLT2 complex with highest score was extracted in PDB format and subjected to molecular dynamic simulation. The MD simulation was carried out in Desmond academic version (Schrödinger suite 2024). The simulation system was generated within an orthorhombic periodic boundary with a minimum distance of 10 Å between the solute and edge of the boundary. The system comprised of TIP3P water, 0.15 M NaCl and docked SGLT2-**9a** complex in a POPC (300 K) membrane model. The membrane was placed within the transmembrane domains of SGLT2 as outlined in UniProt database (UniProt ID: P31639). The system was neutralized with required number of Na⁺/Cl⁻ ions and minimized with 10000 steps of steepest descent algorithm followed by equilibration with NVT and NPγT ensembles at 300 K and 1 bar for 100 ps. The reversible reference system propagator algorithms (r-RESPA) were used for trajectory integration with a 2 fs time step. The coulomb cut off and van der Waals radius were 1.5 Å and 10 Å. The system was thermostatically maintained in 300 K and 1 bar using Nosé-Hoover chain thermostat and Martyna Tobias Klein barostat using isotropic coupling and with a relaxation time of 2 ps for each. The production MD run was carried out for 100 ns using NPγT ensemble and trajectory was recorded within 10 ps interval. After completion of the MD simulation run the trajectory was analysed for root mean square deviations (RMSD) and interactions of ligand with SGLT2 residues throughout the simulation time.

Determination ADMET properties: The ADMET properties were determined using ADMETlab 3.0 server. The three dimensional structure **9a** interacting with SGLT2 was obtained in SDF format after the MD simulation. The ADMET attributes were then computed using prediction models trained by the Directed Message Passing Neural Network (DMPNN) framework in the ADMETlab 3.0 server.

Statistical Analysis: The results were presented as mean ± standard deviation (SD). Student's unpaired *t*-test was performed to analyse individual group statistical comparisons and multiple-group comparisons between diabetic control group and test groups were evaluated by performing one-way ANOVA and the Student-Newman-Keuls test in a proprietary scientific graphing and data analysis software. Values of *p* < 0.05, 0.01 and 0.001 were considered statistically significant.

Cell Culture and treatment with high glucose (HG) and synthetic compounds: NRK-52E cells were cultured in high glucose DMEM medium with 10 % FBS and 1 % antibiotic-antimycotic at 37 °C in a humidified atmosphere containing 5 % (v/v) CO₂. NRK-52E cells ranging in passages from 10 - 30 were used. Various concentrations of *D*-glucose (5.5 to 90 mM) and 24 h time were used for the investigation of concentration dependence. Further, the application of *D*-glucose (33 mM) for various times (0 - 24 h) was used for the investigation of time dependence. After obtaining the results, NRK-52E cells were exposed to high (33 mM *D*-glucose) glucose (HG) for 6 h, followed by with or without synthetic compounds (100 nM to 10 μM) for a 24 h treatment. Control cells were treated with media only. The cells were also treated with two standard drugs, canagliflozin (500 nM to 10 μM) and Resveratrol (500 nM to 10 μM). On termination of incubations, cells were lysed in radioimmunoprecipitation assay (RIPA) buffer (50 mM Tris pH 8, 150 mM NaCl, 1 % NP-40, 0.5 % deoxycholic acid, 0.1% SDS) supplemented with protease and phosphatase inhibitors (1 mM PMSF, 5 μg/mL leupeptin, 2 μg/mL aprotinin, 1 mM EDTA, 10 mM NaF, and 1 mM NaVO₄). Lysates were cleared by centrifugation and total protein concentrations were determined by bicinchoninic acid (BCA) assay (Pierce/Thermo Scientific, Rockford, IL). Each experiment was repeated three times, and the average of the three values was used.

Cell viability assay: Cell toxicity test of synthetic compounds **5a-9c** were carried out in NRK-52E murine proximal tubular cells using Alamar blue reduction bioassay as described

previously. When the cells reach 80-90 % confluency, the cells were treated with 33 mM of high glucose except the control group and after 6 h, it was treated the compounds **5a-9c** with doses ranging from 1 μ M - 50 μ M (0.1 % DMSO + 1XPBS as vehicle) and the control group were treated with vehicle for next 24 h. The original medium was removed after 24 h and incubated with Alamar blue dye for 4 h. After 4 h, OD values were assayed at two different wavelengths 570 nm and 600 nm in the microplate reader. Cell viability was expressed as the ratio of the absorbance to that of the control group.

Glucose uptake assay: Glucose uptake assay was determined using the 6-[*N*-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)amino]-2-deoxy-D-glucose (6-NBDG) as described previously with slight modifications. In brief, NRK-52E cells were evenly plated in a 96 well plate at a density of 1×10^4 cells/well. After 24 h of sub-confluence, cells were treated with HG (33 mM made in 1XPBS) for the next 6 h followed by treatment with **5a-9c** compounds and standard drug, canagliflozin with dose 1, 5 and 10 μ M (made in 0.1 % DMSO + 1XPBS as vehicle) and control cells were treated with the vehicle for next 24 h. Sodium buffer (Na^+ buffer), prepared for cell incubations in the presence of sodium conditions, contained 140 mM NaCl, 5 mM KCl, 2.5 mM CaCl_2 , 1 mM MgSO_4 , 1 mM $\text{KH}_2\text{-PO}_4$, and 10 mM HEPES (pH 7.4). Sodium free buffer (Na^+ -free buffer) contained 140 mM *N*-Methyl-*D*-glucamine instead of NaCl and was used for cell incubations to measure 6-NBDG uptake in the absence of sodium. For experiments, all culture medium was removed from each well and rinsed in Na^+ -free buffer for two times. NRK-52E cells were then incubated in 100 μ L Na^+ or Na^+ -free buffer in the presence of (6-NBDG; 20 μ M), fluorescent nonhydrolyzable glucose for 20 min in a humidified chamber with a 5 % CO_2 at 37 $^\circ\text{C}$. The 6-NBDG uptake reaction was stopped by removing the incubation medium and washing the cells two times with pre-cold Na^+ -free buffer. The cells were washed two times with 1XPBS after incubation and lysed with 70 μ L PBS containing 1 % Triton X (pH 10) and kept at dark for 10 min. After 10 min, 30 μ L of DMSO was added and the mixture was homogenized by pipetting up and down. Fluorescence was measured immediately with a microplate reader at emission and excitation wavelengths 470 nm and 540 nm respectively. Results were reported as fold change with respect to control. Further, dose dependent glucose uptake analysis of the compound **9a** was carried out at doses ranging from 100 nM to 10 μ M with the same protocol as described.

ROS scavenging assay: The ROS level in the NRK-52E cells post-HG treatment was determined using the fluorogenic probe 2',7'-dichlorofluorescein diacetate (H_2DCFDA) according to the reported procedure. Briefly, NRK-52E cells (1×10^6) at 80 % confluence were cultured in 24-well plates and kept overnight. Cells were treated with HG (33 mM made in 1XPBS) for 6 h followed by different concentrations of the compound **9a** (100 nM, 250 nM, 500 nM, 1 μ M, 5 μ M, 10 μ M) and control cells were treated with the vehicle, along with canagliflozin (500 nM, 5 and 10 μ M) and resveratrol (500 nM, 5 and 10 μ M) (0.1 % DMSO + 1XPBS as vehicle) for next 24 h. Media was aspirated out and cells were then washed with 1X PBS and incubated with 10 μ M H_2DCFDA in the dark for 30 min at 37 $^\circ\text{C}$. The cells were then washed twice with 200 μ L 1XPBS and incubated with 1 % Triton X-100 (170 μ L). The reaction was stopped by adding DMSO (130 μ L). Cells were scraped and transferred to the black plate. Absorbance was measured at excitation and emission wavelengths of 480 and 530 nm, respectively, using a microplate reader (in triplicates). The ROS production by HG-treated cells was considered 100 % (baseline), to which other activities were compared. In another set of experiments, intracellular ROS levels generated post-HG exposure in NRK-52E cells were measured by flow cytometry (BD-FACS-Melody, USA) analysis using H_2DCFDA , according to the reported procedure. Briefly, NRK-52E cells were treated with the compound **9a** after exposure with HG (33 mM) for 6 h at different concentrations (500 nM, 1 and 5 μ M), canagliflozin (5 μ M, as the positive control) and resveratrol (5 μ M, as the positive control) for

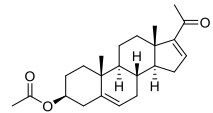
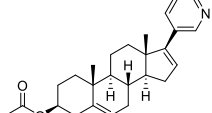
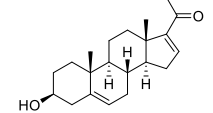
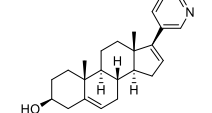
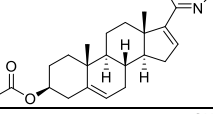
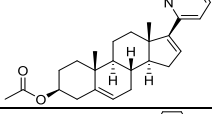
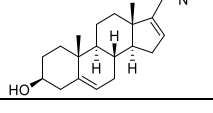
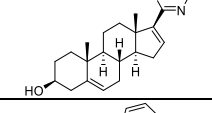
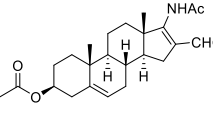
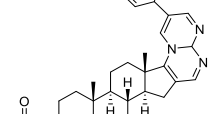
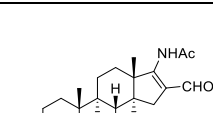
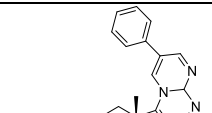
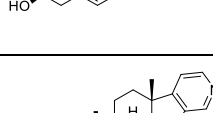
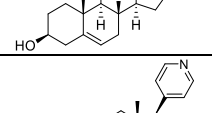
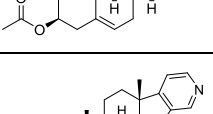
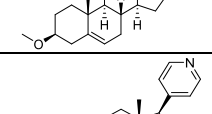
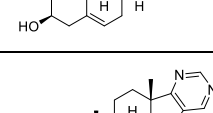
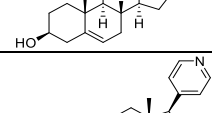
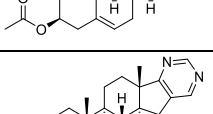
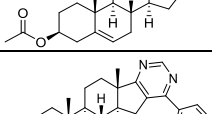
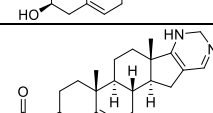
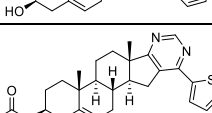
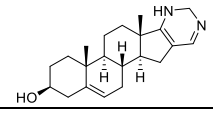
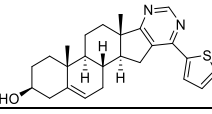
next 24 h. The adherent and non-adherent cells were collected and washed twice with 1XPBS (pH 7.4). The cells were then incubated with 10 μ M H₂DCFDA at 37 °C for 30 min in the dark, which was followed by washing twice with chilled 1XPBS. The fluorescence intensities of 2',7'-dichlorodihydrofluorescein (DCF) produced by intracellular ROS were analysed by flow cytometry with excitation and emission at 480 nm and 530 nm, respectively.

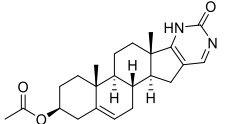
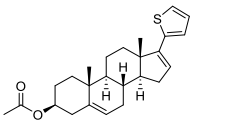
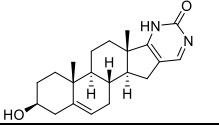
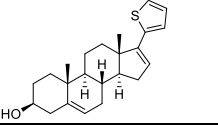
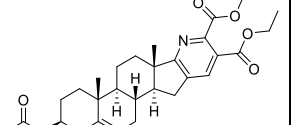
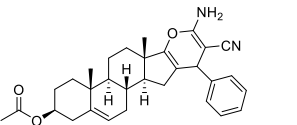
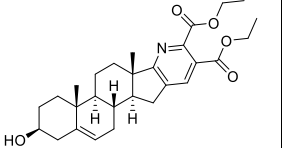
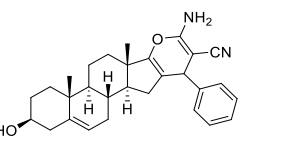
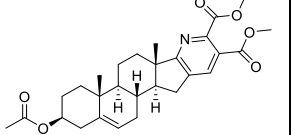
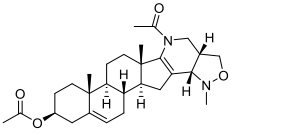
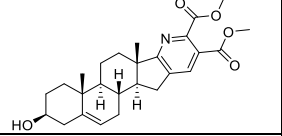
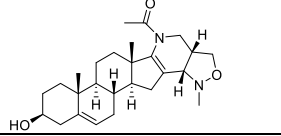
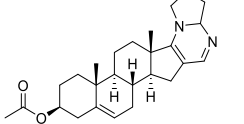
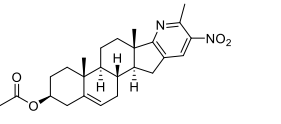
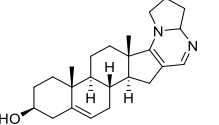
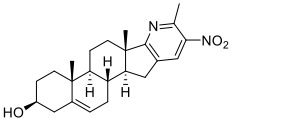
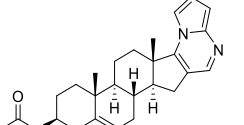
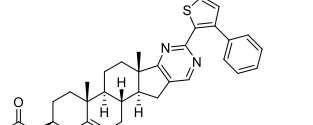
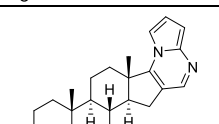
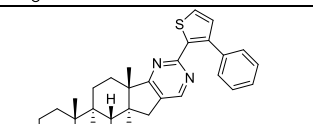
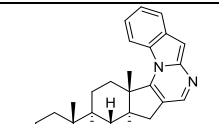
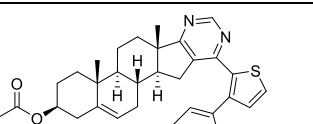
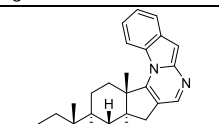
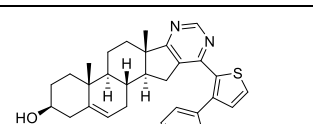
Measurement of mitochondrial membrane potential in HG-induced NRK-52E cells: Rhodamine 123 (Invitrogen, Life Technologies, Carlsbad, CA, USA) was used to measure MMP, as previously described. Briefly, 2×10^5 cells were plated in each well of a 6-well plate and allowed to attach for 22 - 24 h. After treatment for 24 h, the cells were treated with HG (33 mM) treatment for 6 h followed by the treatment of the best concentrations based on glucose uptake analysis and ROS analysis of the compound **9a** *i.e.* 500 nM, 1 μ M and 5 μ M for next 24 h. The standard drug, canagliflozin (5 μ M) and resveratrol (5 μ M) was used. After 24 h, the cells were washed with 1XPBS and incubated with 200 ng/mL of Rhodamine 123. After incubation for 30 min at 37 °C, the cells were washed thrice and resuspended in 1 mL of 1XPBS. Cytofluorimetric analysis was performed using flow cytometry (BD-FACS-Melody, USA) and FACS Chorus software (Beckman Coulter, Brea, CA, USA) using ~ 655 - 730 nm emission for PI.

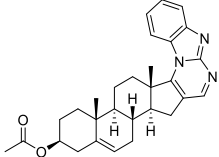
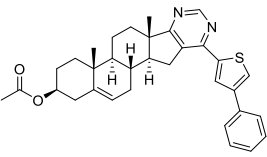
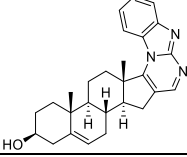
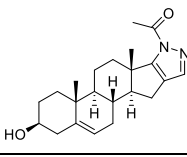
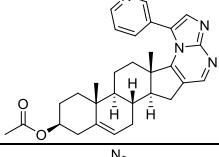
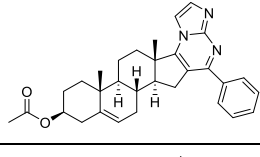
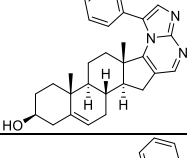
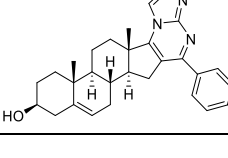
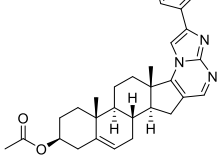
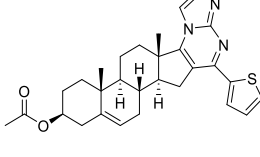
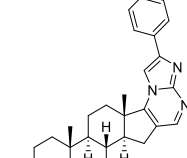
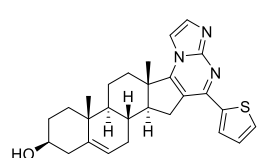
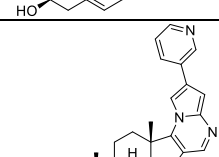
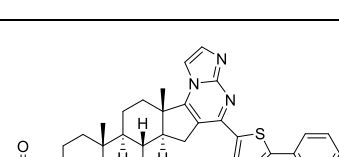
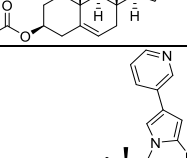
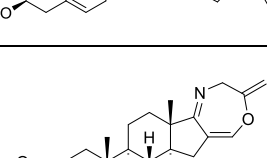
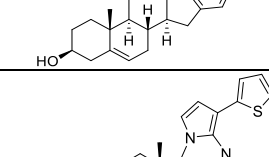
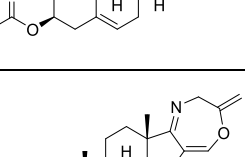
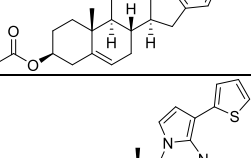
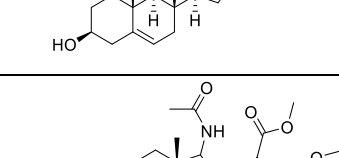
Western blot analysis: The total protein from NRK-52E cells were lysed in RIPA buffer as previously described. All samples consist of approximately same amount of protein (~ 20 - 30 μ g) were subjected to electrophoresis using 10 % sodium dodecyl polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to nitrocellulose membranes. Subsequently, the membranes were blocked in 1 % bovine serum albumin (BSA) for 40 min and incubated with the respective primary antibody overnight at 4 °C. Primary antibodies against SGLT2 (1:1000) were used. After overnight incubation, TBS-T (50 mmol/L Tris-HCl, pH 7.6, 150 mmol/L NaCl, 0.1 % Tween 20) were used to wash the membranes for 30 min and then incubated at room temperature with HRP-conjugated goat anti-rabbit secondary antibody (1:1000) for 2 h. The blotted membrane was detected by enhanced chemiluminescence method with ECL western blotting detection substrate and the intensity of each immunoblotting band was measured using the histogram tool of Adobe Photoshop CS5. HRP conjugated β -actin were used as an internal control.

Chemistry: Synthesis of Target Compounds **5**, **6**, **8**, **9**. General Procedure for synthesis: A 25 mL Schlenk tube equipped with a stir bar was charged with β -formyl enamide derivative of 16-DPA (1 mmol), aminopyrazole/ aminopyridine/ benzamidine/ aminobenzamidazole (1.2 mmol), Ni nanoparticles (5 mol %) and DMF (5 mL). The reaction mixture was stirred under 110 °C heating conditions in an oil bath. On completion, the tube was cooled to room temperature, the catalyst recovered using a magnetic rod and the solution poured into ice-cold water. The mixture was extracted with EtOAc (3 \times 10 mL) and washed with water. The organic layer was collected and concentrated in vacuum. The residue was purified by column chromatography on silica gel with hexane-EtOAc as the eluent to afford **5a-e**, **8a-d**. Some of the compounds were subjected to HPLC purification. The aforementioned compounds were hydrolysed using 10 % KOH to obtain the target molecules **6a-d**, **9a-c**.

Table S1. Virtually screened steroidal compounds as SGLT2 inhibitors using high-precision molecular docking.

SI No.	Compound	Docking Score	SI No.	Compound	Docking Score
1		-9.0	51		-9.7
2		-7.9	52		-9.9
3		-9.3	53		-9.6
4		-9.6	54		-9.4
5		-9.5	55		-9.7
6		-8.4	56		-9.9
7		-9.7	57		-9.4
8		-9.6	58		-9.5
9		-9.5	59		-9.6
10		-9.4	60		-9.3
11		-9.5	61		-9.5
12		-8.9	62		-9.1

13		-9.7	63		-8.9
14		-9.4	64		-9.1
15		-8.2	65		-9.3
16		-8.0	66		-8.7
17		-8.5	67		-8.8
18		-8.3	68		-8.9
19		-9.3	69		-8.5
20		-9.1	70		-8.7
21		-9.6	71		-9.0
22		-9.5	72		-9.6
23		-9.9	73		-9.5
24		-9.9	74		-9.6

25		-10.2	75		-9.7
26		-10.2	76		-9.7
27		-9.7	77		-9.4
28		-9.3	78		-9.4
29		-9.9	79		-8.8
30		-9.6	80		-8.9
31		-9.1	81		-9.4
32		-9.9	82		-8.9
33		-10.0	83		-8.5
34		-9.5	84		-8.2

35		-9.5	85		-8.2
36		-9.6	86		-9.8
37		-9.5	87		-9.9
38		-9.4	88		-10.0
39		-9.5	89		-9.5
40		-9.4	90		-9.1
41		-10.1	91		-9.1
42		-10.5	92		-9.8
43		-9.9	93		-9.5
44		-9.9	94		-9.8
45		-10.2	95		-9.6

46		-10.3	96		-9.9
47		-9.5	97		-9.6
48		-9.4	98		-9.6
49		-9.4	99		-9.3
50		-9.3	100		-9.5
Canagli flozin		-12.3			

Spectral Data of Synthesised Steroidal pyrimidine compounds

*3β-Acetoxy-2'-phenyl-androst[16,17-*e*]pyrazolo(1',5'-*a*)pyrimidin-5-ene (5a)*: White solid (Yield: 85 %). ¹HNMR (400 MHz, CDCl₃) δ 8.34 (s, 1H), 7.96 (d, *J* = 6.9 Hz, 2H), 7.45 - 7.32 (m, 3H), 7.03 (s, 1H), 5.39 (d, *J* = 5.2 Hz, 1H), 2.99 (d, *J* = 15.4 Hz, 1H), 2.91 - 2.82 (m, 1H), 2.59 - 2.54 (m, 1H), 2.36 - 2.25 (m, 2H), 2.10 (d, *J* = 17.1 Hz, 1H), 1.99 (s, 3H), 1.88 - 1.69 (m, 8H), 1.61 - 1.53 (m, 2H), 1.22 (s, 3H), 1.18 - 1.10 (m, 2H), 1.08 (s, 3H) ppm. ¹³C{¹H}NMR (100 MHz, CDCl₃) δ 170.6, 156.9, 145.7, 140.3, 132.8, 129.2, 128.8, 126.7, 121.7, 120.2, 92.2, 73.7, 56.17, 50.1, 47.1, 38.1, 36.9, 36.8, 33.4, 31.3, 30.3, 29.7, 28.8, 27.7, 21.5, 20.3, 19.3, 14.4 ppm. HRMS (ESI) *m/z* calcd for C₃₁H₃₆N₃O₂ (M+H)⁺ 482.2808, found 482.2889.

*3β-Acetoxy-2'-(*p*-chlorophenyl)-androst[16,17-*e*]pyrazolo(1',5'-*a*)pyrimidin-5-ene (5b)*: White solid (Yield: 88 %). ¹HNMR (400 MHz, CDCl₃) δ 8.31 (s, 1H), 7.89 (d, *J* = 8.5 Hz, 2H), 7.36 (d, *J* = 8.5 Hz, 2H), 6.90 (s, 1H), 5.38 (d, *J* = 4.6 Hz, 1H), 4.60 - 4.52 (m, 1H), 2.96 (d, *J* = 12.8 Hz, 1H), 2.85 (dd, *J* = 14.5, 5.2 Hz, 1H), 2.56 - 2.50 (m, 1H), 2.36 - 2.25 (m, 2H), 2.09 (d, *J* = 14.8 Hz, 1H), 1.99 (s, 3H), 1.88 - 1.68 (m, 8H), 1.62 - 1.51 (m, 1H), 1.19 (s, 3H), 1.16 - 1.10 (m, 2H), 1.07 (s, 3H) ppm. ¹³C{¹H}NMR (100 MHz, CDCl₃) δ 170.6, 157.4, 155.1, 149.8, 146.7, 140.3, 134.7, 131.7, 128.9, 127.8, 121.7, 120.6, 92.50, 73.7, 56.2, 50.1, 46.9, 38.1, 36.9, 36.8, 33.5, 31.3, 30.2, 28.8, 27.7, 21.5, 20.3, 19.3, 14.4 ppm. HRMS (ESI) *m/z* calcd for C₃₁H₃₅ClN₃O₂ (M+H)⁺ 516.2418, found 516.2526.

*3β-Acetoxy-2'-(*p*-bromophenyl)-androst[16,17-*e*]pyrazolo(1',5'-*a*)pyrimidin-5-ene (5c)*: White solid, (Yield: 86 %). ¹HNMR (400 MHz, CDCl₃) δ 8.31 (s, 1H), 7.82 (d, *J* = 8.6 Hz, 2H), 7.51 (d, *J* = 8.6 Hz, 2H), 6.88 (s, 1H), 5.38 (d, *J* = 4.9 Hz, 1H), 4.59 - 4.52 (m, 1H), 2.95

(d, $J = 12.6$ Hz, 1H), 2.89 - 2.81 (m, 2H), 2.55 - 2.49 (m, 1H), 2.35 - 2.25 (m, 2H), 2.10 - 2.06 (m, 1H), 1.98 (s, 3H), 1.87 - 1.82 (m, 4H), 1.78 - 1.67 (m, 3H), 1.61 - 1.52 (m, 1H), 1.19 (s, 3H), 1.16 - 1.09 (m, 2H), 1.07 (s, 3H) ppm. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) δ 170.6, 157.2, 155.0, 150.0, 146.9, 140.3, 132.2, 131.9, 128.1, 122.9, 121.7, 120.7, 92.6, 73.7, 56.2, 50.1, 46.8, 38.1, 36.9, 36.8, 33.5, 31.3, 30.2, 28.8, 27.7, 21.5, 20.3, 19.3, 14.4 ppm. HRMS (ESI) m/z calcd for $\text{C}_{31}\text{H}_{35}\text{BrN}_3\text{O}_2$ ($\text{M}+2\text{H}$) $^{2+}$ 561.1991, found 561.2067.

*3 β -Acetoxy-3'-cyano-androst[16,17-*e*]pyrazolo(1',5'-*a*)pyrimidin-5-ene (5d)*: White solid (Yield: 77 %). ^1H NMR (400 MHz, CDCl_3) δ 8.56 (s, 1H), 8.30 (s, 1H), 5.37 (d, $J = 5.4$ Hz, 1H), 4.62 - 4.49 (m, 1H), 2.93 (dd, $J = 14.7, 5.7$ Hz, 1H), 2.79 - 2.76 (m, 1H), 2.62 - 2.56 (m, 1H), 2.35 - 2.23 (m, 1H), 2.11 - 2.07 (m, 2H), 1.98 (s, 3H), 1.88 - 1.67 (m, 7H), 1.60 - 1.50 (m, 2H), 1.17 (s, 3H), 1.15 - 1.08 (m, 2H), 1.06 (s, 3H) ppm. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) δ 170.6, 158.9, 150.4, 150.3, 147.6, 140.3, 124.1, 121.5, 113.3, 82.2, 73.6, 56.2, 49.9, 47.1, 38.1, 36.8, 36.8, 33.2, 31.2, 30.2, 28.8, 27.7, 21.5, 20.2, 19.3, 14.7 ppm. HRMS (ESI) m/z calcd for $\text{C}_{26}\text{H}_{31}\text{N}_4\text{O}_2$ ($\text{M}+\text{H}$) $^+$ 431.2447, found 431.2509.

*3 β -Acetoxy-5'-methoxy-phenyl-androst[16,17-*e*]indazolo(7',6'-*a*)pyrimidin-5-ene (5e)*: Yellow solid (Yield: 82%). ^1H NMR (400 MHz, CDCl_3) δ 8.57 (s, 1H), 7.48 - 7.44 (m, 1H), 7.37 (d, $J = 8.5$ Hz, 1H), 6.52 (d, $J = 7.5$ Hz, 1H), 5.38 (d, $J = 5.4$ Hz, 1H), 4.61 - 4.53 (m, 1H), 4.07 (s, 3H), 3.07 (d, $J = 12.8$ Hz, 1H), 2.97 (dd, $J = 14.7, 5.6$ Hz, 1H), 2.64 (dd, $J = 14.6, 3.2$ Hz, 1H), 2.35 - 2.25 (m, 2H), 2.10 (d, $J = 14.6$ Hz, 1H), 1.98 (s, 3H), 1.88 - 1.69 (m, 8H), 1.62 - 1.51 (m, 1H), 1.21 (s, 3H), 1.18 - 1.09 (m, 2H), 1.07 (s, 3H) ppm. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) δ 170.6, 155.7, 155.5, 153.2, 143.8, 143.6, 140.3, 130.4, 124.5, 121.7, 108.6, 104.5, 98.7, 73.7, 56.2, 56.0, 50.1, 47.1, 38.1, 36.9, 36.8, 33.5, 31.2, 30.3, 29.1, 27.7, 21.5, 20.3, 19.3, 14.0 ppm. HRMS (ESI) m/z calcd for $\text{C}_{30}\text{H}_{36}\text{N}_3\text{O}_3$ ($\text{M}+\text{H}$) $^+$ 486.2757, found 486.2895.

*3 β -Hydroxy-2'-phenyl-androst[16,17-*e*]pyrazolo(1',5'-*a*)pyrimidin-5-ene (6a)*: White solid (Yield: 92 %). ^1H NMR (400 MHz, CDCl_3) δ 8.32 (s, 1H), 7.95 (d, $J = 7.2$ Hz, 2H), 7.42 - 7.31 (m, 3H), 6.96 (s, 1H), 5.36 (d, $J = 5$ Hz, 1H), 3.54 - 3.46 (m, 1H), 3.01 (d, $J = 12.8$ Hz, 1H), 2.88 (dd, $J = 14.3, 5.4$ Hz, 1H), 2.58 - 2.52 (m, 1H), 2.31 - 2.22 (m, 2H), 2.12 - 2.07 (m, 1H), 1.89 - 1.68 (m, 7H), 1.57 - 1.43 (m, 2H), 1.21 (s, 3H), 1.16 - 1.08 (m, 2H), 1.06 (s, 3H) ppm. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) δ 156.5, 149.7, 146.2, 141.3, 132.9, 129.0, 128.8, 126.6, 120.8, 120.3, 92.4, 71.6, 56.3, 50.2, 46.9, 42.2, 37.1, 36.8, 33.5, 31.6, 31.3, 30.3, 29.7, 28.8, 20.4, 19.4, 14.4 ppm. HRMS (ESI) m/z calcd for $\text{C}_{29}\text{H}_{34}\text{N}_3\text{O}$ ($\text{M}+\text{H}$) $^+$ 440.2624, found 440.2628.

*3 β -Hydroxy-2'-(*p*-chlorophenyl)-androst[16,17-*e*]pyrazolo(1',5'-*a*)pyrimidin-5-ene (6b)*: White solid (Yield: 93 %). ^1H NMR (400 MHz, CDCl_3) δ 8.33 (s, 1H), 7.89 (d, $J = 8.3$ Hz, 2H), 7.37 (d, $J = 8.2$ Hz, 2H), 6.95 (s, 1H), 5.36 (d, $J = 5.0$ Hz, 1H), 3.54 - 3.46 (m, 1H), 2.97 (d, $J = 11.9$ Hz, 1H), 2.87 (dd, $J = 14.5, 5.5$ Hz, 1H), 2.59 - 2.53 (m, 1H), 2.32 - 2.19 (m, 2H), 2.12 - 2.07 (m, 1H), 1.86 - 1.71 (m, 8H), 1.54 - 1.43 (m, 1H), 1.21 (s, 3H), 1.18 - 1.08 (m, 2H), 1.06 (s, 3H) ppm. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, $\text{DMSO-}d_6$) δ 156.3, 154.1, 150.4, 148.3, 142.2, 133.9, 132.1, 129.4, 128.3, 121.7, 120.4, 93.1, 70.4, 55.7, 50.1, 46.6, 42.7, 37.2, 36.8, 33.5, 31.8, 31.1, 30.2, 28.7, 20.3, 19.5, 14.5 ppm. HRMS (ESI) m/z calcd for $\text{C}_{29}\text{H}_{33}\text{ClN}_3\text{O}$ ($\text{M}+\text{H}$) $^+$ 474.2312, found 474.2347.

*3 β -Hydroxy-2'-(*p*-bromophenyl)-androst[16,17-*e*]pyrazolo(1',5'-*a*)pyrimidin-5-ene (6c)*: White solid, (Yield: 95 %). ^1H NMR (400 MHz, CDCl_3) δ 8.32 (s, 1H), 7.83 (d, $J = 8.5$ Hz,

2H), 7.52 (d, $J = 8.5$ Hz, 2H), 6.91 (s, 1H), 5.35 (d, $J = 5.3$ Hz, 1H), 3.54 - 3.46 (m, 1H), 2.96 (d, $J = 12.7$ Hz, 1H), 2.85 (dd, $J = 14.4, 5.8$ Hz, 1H), 2.54 (dd, $J = 14.3, 11.0$ Hz, 1H), 2.31 - 2.18 (m, 2H), 2.12 - 2.07 (m, 1H), 1.89 - 1.67 (m, 8H), 1.54 - 1.43 (m, 1H), 1.19 (s, 3H), 1.15 - 1.08 (m, 2H), 1.06 (s, 3H) ppm. $^{13}\text{C}\{^1\text{H}\}\text{NMR}$ (100 MHz, CDCl_3) δ 157.6, 155.2, 149.8, 146.7, 141.3, 132.1, 131.9, 128.1, 123.0, 120.8, 120.7, 92.5, 71.6, 56.3, 50.2, 46.9, 42.2, 37.1, 36.78, 33.5, 31.6, 31.3, 30.3, 28.9, 20.4, 19.4, 14.4 ppm. HRMS (ESI) m/z calcd for $\text{C}_{29}\text{H}_{33}\text{BrN}_3\text{O}$ ($\text{M}+\text{H}$) $^+$ 518.1807, found 518.1883.

*3 β -Hydroxy-5'-methoxy-phenyl-androst[16,17-*e*]indazolo(7',6'-*a*)pyrimidin-5-ene (6d)*: Yellow solid (Yield: 92 %). $^1\text{H}\text{NMR}$ (400 MHz, CDCl_3) δ 8.56 (s, 1H), 7.49-7.41 (m, 1H), 7.37 (d, $J = 8.5$ Hz, 1H), 6.52 (d, $J = 7.5$ Hz, 1H), 5.34 (d, $J = 5.3$ Hz, 1H), 4.07 (s, 3H), 3.53 - 3.45 (m, 1H), 3.09 - 3.04 (m, 1H), 2.96 (dd, $J = 14.7, 5.9$ Hz, 1H), 2.64 (dd, $J = 14.6, 11.3$ Hz, 1H), 2.31 - 2.17 (m, 2H), 2.12 - 2.06 (m, 1H), 1.89 - 1.69 (m, 8H), 1.53-1.42 (m, 1H), 1.21 (s, 3H), 1.14 - 1.08 (m, 2H), 1.06 (s, 3H) ppm. $^{13}\text{C}\{^1\text{H}\}\text{NMR}$ (100 MHz, CDCl_3) δ 155.7, 155.5, 153.2, 143.8, 143.7, 141.4, 130.4, 124.5, 120.8, 108.6, 104.5, 98.7, 71.6, 56.3, 56.0, 50.2, 47.1, 42.2, 37.1, 36.8, 33.5, 31.6, 31.3, 30.3, 29.1, 20.3, 19.4, 14.0 ppm. HRMS (ESI) m/z calcd for $\text{C}_{28}\text{H}_{34}\text{N}_3\text{O}_2$ ($\text{M}+\text{H}$) $^+$ 444.2651, found 444.2646.

*3 β -Acetoxy-androst[16,17-*e*]benzimidazolo(1',2'-*a*)pyrimidin-5-en (8a)*: Yellow solid (Yield: 86 %). $^1\text{H}\text{NMR}$ (400 MHz, CDCl_3) δ 8.63 (s, 1H), 8.09 (d, $J = 8.5$ Hz, 1H), 7.97 (d, $J = 8.2$ Hz, 1H), 7.50 (t, $J = 7.7$ Hz, 1H), 7.32 (t, $J = 7.8$ Hz, 1H), 5.40 (d, $J = 5.0$ Hz, 1H), 4.60 - 4.52 (m, 1H), 2.98 (d, $J = 12.0$ Hz, 1H), 2.89 (dd, $J = 14.2, 5.5$ Hz, 1H), 2.63 - 2.57 (m, 1H), 2.37 - 2.25 (m, 2H), 2.10 (d, $J = 19.2$ Hz, 1H), 1.99 (s, 3H), 1.92 - 1.71 (m, 8H), 1.62 - 1.53 (m, 1H), 1.38 (s, 3H), 1.18 - 1.12 (m, 2H), 1.09 (s, 3H) ppm. $^{13}\text{C}\{^1\text{H}\}\text{NMR}$ (100 MHz, CDCl_3) δ 170.6, 159.7, 153.2, 152.1, 144.6, 140.1, 126.0, 126.0, 121.8, 121.4, 120.7, 120.5, 115.4, 73.6, 57.4, 49.8, 48.5, 38.0, 36.8, 36.8, 34.2, 30.9, 30.3, 28.6, 27.7, 21.5, 20.6, 19.3, 13.7 ppm. HRMS (ESI) m/z calcd for $\text{C}_{29}\text{H}_{34}\text{N}_3\text{O}_2$ ($\text{M}+\text{H}$) $^+$ 456.2651, found 456.2646.

*3 β -Acetoxy-6'-chloro-androst[16,17-*e*]benzimidazolo(1',2'-*a*)pyrimidin-5-ene (8b)*: Pale yellow solid (Yield: 35 %). $^1\text{H}\text{NMR}$ (400 MHz, CDCl_3) δ 8.63 (s, 1H), 7.99 (d, $J = 9$ Hz, 1H), 7.91 (s, 1H), 7.26 (dd, $J = 8.9, 2.0$ Hz, 1H), 5.39 (d, $J = 4.2$ Hz, 1H), 4.60 - 4.52 (m, 1H), 2.89 (d, $J = 13.3$ Hz, 2H), 2.60 (t, $J = 12.7$ Hz, 1H), 2.35 - 2.26 (m, 2H), 2.10 (d, $J = 17.6$ Hz, 1H), 1.99 (s, 3H), 1.89 - 1.70 (m, 8H), 1.62 - 1.53 (m, 1H), 1.36 (s, 3H), 1.18 - 1.13 (m, 2H), 1.09 (s, 3H) ppm. $^{13}\text{C}\{^1\text{H}\}\text{NMR}$ (100 MHz, CDCl_3) δ 170.6, 159.7, 153.5, 152.9, 145.8, 140.1, 131.7, 124.6, 121.8, 121.7, 121.1, 119.9, 116.2, 73.6, 57.4, 49.8, 48.5, 38.03, 36.8, 34.1, 30.9, 30.3, 28.6, 27.7, 21.5, 20.6, 19.3, 13.7 ppm. HRMS (ESI) m/z calcd for $\text{C}_{29}\text{H}_{33}\text{ClN}_3\text{O}_2$ ($\text{M}+\text{H}$) $^+$ 490.2261, found 490.2297.

*3 β -Acetoxy-5'-chloro-androst[16,17-*e*]benzimidazolo(1',2'-*a*)pyrimidin-5-ene (8c)*: Pale yellow solid (Yield: 53 %). $^1\text{H}\text{NMR}$ (400 MHz, CDCl_3) δ 8.63 (s, 1H), 8.06 (s, 1H), 7.88 (d, $J = 8.7$ Hz, 1H), 7.46 (dd, $J = 8.8, 1.8$ Hz, 1H), 5.40 (d, $J = 5.4$ Hz, 1H), 4.60 - 4.52 (m, 1H), 2.90 (dd, $J = 13.5, 6.4$ Hz, 2H), 2.60 (t, $J = 11.6$ Hz, 1H), 2.37 - 2.26 (m, 2H), 2.13 - 2.08 (m, 1H), 1.99 (s, 3H), 1.92 - 1.72 (m, 8H), 1.62 - 1.53 (m, 1H), 1.37 (s, 3H), 1.18 - 1.15 (m, 2H), 1.10 (s, 3H) ppm. $^{13}\text{C}\{^1\text{H}\}\text{NMR}$ (100 MHz, CDCl_3) δ 170.6, 159.6, 153.6, 152.6, 143.3, 140.2, 126.8, 126.7, 126.3, 121.7, 121.5, 121.1, 115.1, 73.6, 57.4, 49.8, 48.6, 38.0, 36.8, 36.8, 34.4, 30.9, 30.3, 28.6, 27.7, 21.5, 20.7, 19.3, 13.9 ppm. HRMS (ESI) m/z calcd for $\text{C}_{29}\text{H}_{33}\text{ClN}_3\text{O}_2$ ($\text{M}+\text{H}$) $^+$ 490.2261, found 490.2297.

*3β-Acetoxy-4'-chloro-androst[16,17-*e*]benzimidazolo(1',2'-*a*)pyrimidin-5-ene (8d)*: Yellow solid (Yield: 83 %). **¹H NMR (400 MHz, CDCl₃)** δ 8.63 (s, 1H), 7.98 (d, *J* = 8.4 Hz, 1H), 7.51 (d, *J* = 7.7 Hz, 1H), 7.24 - 7.22 (m, 1H), 5.39 (d, *J* = 5.0 Hz, 1H), 4.59 - 4.52 (m, 1H), 2.93 - 2.86 (m, 2H), 2.62 - 2.56 (m, 1H), 2.33 - 2.28 (m, 2H), 2.09 (d, *J* = 14.5 Hz, 1H), 1.98 (s, 3H), 1.86 - 1.69 (m, 8H), 1.61 - 1.51 (m, 1H), 1.36 (s, 3H), 1.15 - 1.11 (m, 2H), 1.08 (s, 3H) ppm. **¹³C{¹H} NMR (100 MHz, CDCl₃)** δ 170.6, 159.8, 153.8, 152.2, 142.2, 140.1, 126.9, 125.6, 125.2, 121.7, 121.4, 121.3, 113.9, 73.6, 57.3, 49.8, 48.5, 38.0, 36.8, 34.2, 30.9, 30.3, 29.7, 28.6, 27.7, 21.5, 20.6, 19.3, 13.7 ppm. HRMS (ESI) *m/z* calcd for C₂₉H₃₃ClN₃O₂ (M+H)⁺ 490.2261, found 490.2297.

*3β-Hydroxy-androst[16,17-*e*]benzimidazolo(1',2'-*a*)pyrimidin-5-en (9a)*: Pale Yellow solid (Yield: 91 %). **¹H NMR (400 MHz, CDCl₃)** δ 8.58 (s, 1H), 8.07 (d, *J* = 8.6 Hz, 1H), 7.93 (d, *J* = 8.3 Hz, 1H), 7.47 (t, *J* = 7.7 Hz, 1H), 7.30 (t, *J* = 7.8 Hz, 1H), 5.35 (d, *J* = 5.5 Hz, 1H), 3.56 - 3.48 (m, 1H), 3.42 (s, 1H), 2.95 (d, *J* = 12.4 Hz, 1H), 2.85 (dd, *J* = 14.1, 6.1 Hz, 1H), 2.60 - 2.54 (m, 2H), 2.34 - 2.24 (m, 2H), 2.07 (d, *J* = 12.4 Hz, 1H), 1.90 - 1.77 (m, 5H), 1.73 - 1.66 (m, 2H), 1.55 - 1.46 (m, 1H), 1.35 (s, 3H), 1.22 - 1.17 (m, 2H), 1.06 (s, 3H) ppm. **¹³C{¹H} NMR (100 MHz, CDCl₃)** δ 159.7, 153.1, 152.1, 144.8, 141.4, 126.0, 125.9, 121.4, 120.7, 120.5, 115.4, 71.4, 57.5, 49.9, 48.5, 42.2, 37.1, 36.7, 34.2, 31.5, 31.0, 30.3, 28.6, 20.6, 19.3, 13.6 ppm. HRMS (ESI) *m/z* calcd for C₂₇H₃₂N₃O (M+H)⁺ 414.2545, found 414.2543.

*3β-Hydroxy-5'-chloro-androst[16,17-*e*]benzimidazolo(1',2'-*a*)pyrimidin-5-ene (9b)*: Pale yellow solid (Yield: 90 %). **¹H NMR (400 MHz, DMSO-*d*₆)** δ 7.78 (s, 1H), 7.30 (s, 1H), 7.25 (d, *J* = 8.4 Hz, 1H), 6.97 (d, *J* = 8.3 Hz, 1H), 5.26 (d, *J* = 4.8 Hz, 1H), 4.60 (d, *J* = 4.5 Hz, 1H), 3.25 - 3.16 (m, 1H), 2.56 (dd, *J* = 8.8, 6.4 Hz, 1H), 2.15 - 1.96 (m, 3H), 1.72 - 1.53 (m, 7H), 1.41 - 1.15 (m, 4H), 0.92 (s, 3H), 0.90 - 0.79 (m, 2H), 0.76 (s, 3H) ppm. **¹³C{¹H} NMR (100 MHz, DMSO-*d*₆)** δ 151.3, 150.8, 142.1, 141.9, 134.5, 131.3, 125.4, 121.0, 120.5, 113.9, 70.4, 50.3, 49.9, 47.6, 42.7, 37.3, 36.8, 31.9, 31.8, 31.0, 30.9, 26.2, 20.5, 19.6, 14.7 ppm. HRMS (ESI) *m/z* calcd for C₂₇H₃₁ClN₃O (M+H)⁺ 448.2156, found 448.2155.

*3β-Hydroxy-4'-chloro-androst[16,17-*e*]benzimidazolo(1',2'-*a*)pyrimidin-5-ene (9c)*: Yellow solid (Yield: 92 %). **¹H NMR (400 MHz, CDCl₃)** δ 8.67 (s, 1H), 8.00 (d, *J* = 8.4 Hz, 1H), 7.53 (d, *J* = 7.7 Hz, 1H), 7.23 (t, *J* = 8.1 Hz, 1H), 5.37 (d, *J* = 5.3 Hz, 1H), 3.55 - 3.47 (m, 1H), 2.95 - 2.88 (m, 2H), 2.64 - 2.58 (m, 1H), 2.34 - 2.19 (m, 2H), 2.10 (dd, *J* = 16.9, 2.5 Hz, 1H), 1.92 - 1.71 (m, 7H), 1.59 - 1.48 (m, 2H), 1.37 (s, 3H), 1.14 - 1.11 (m, 2H), 1.08 (s, 3H) ppm. **¹³C{¹H} NMR (100 MHz, CDCl₃)** δ 159.9, 153.9, 152.2, 142.2, 141.2, 126.9, 125.7, 121.4, 120.8, 113.9, 71.5, 57.5, 49.9, 48.6, 42.1, 37.0, 36.7, 34.2, 31.9, 31.5, 30.3, 29.7, 28.6, 22.7, 20.6, 19.3, 13.7 ppm. HRMS (ESI) *m/z* calcd for C₂₇H₃₁ClN₃O (M+H)⁺ 448.2156, found 448.2150.

Figure S1. ^1H NMR (400 MHz, CDCl_3) of 5a

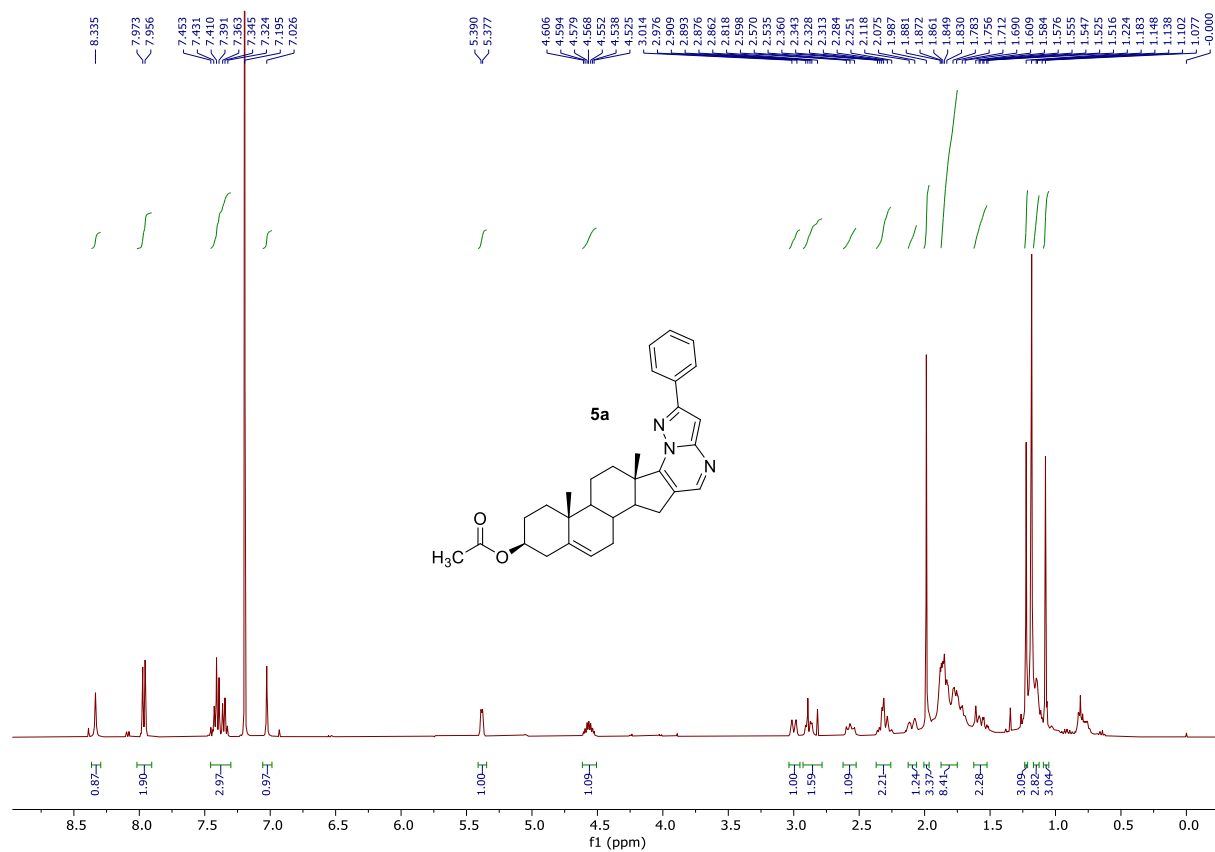


Figure S2. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) of 5a

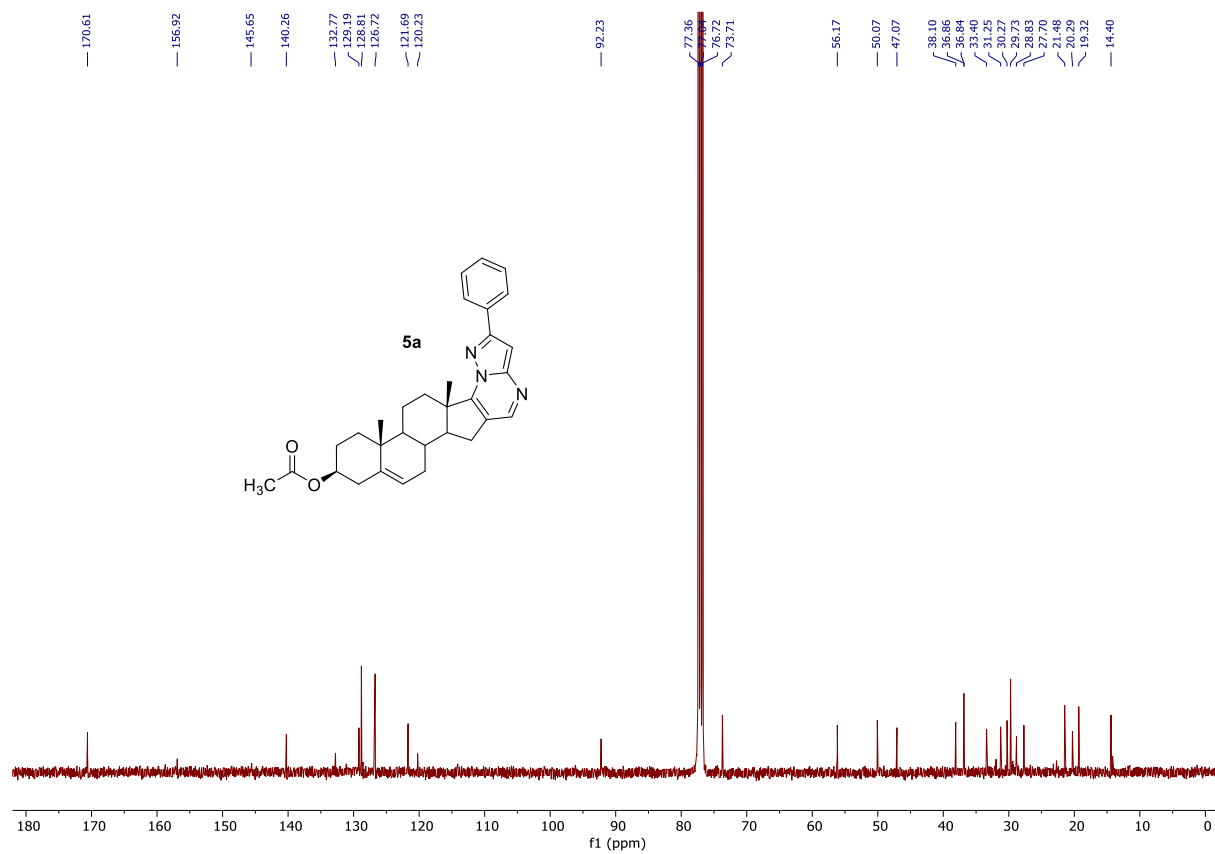


Figure S3. HRMS of 5a

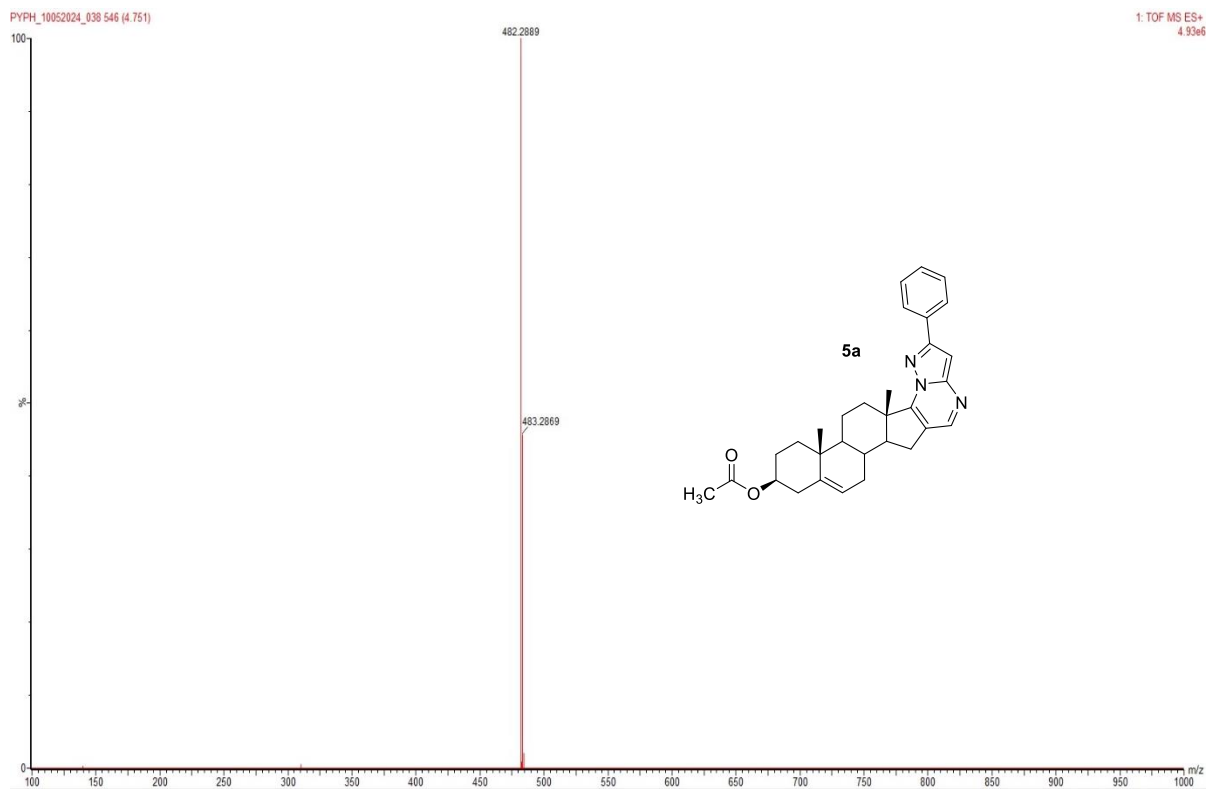


Figure S4. ¹HNMR (400 MHz, CDCl₃) of 5b

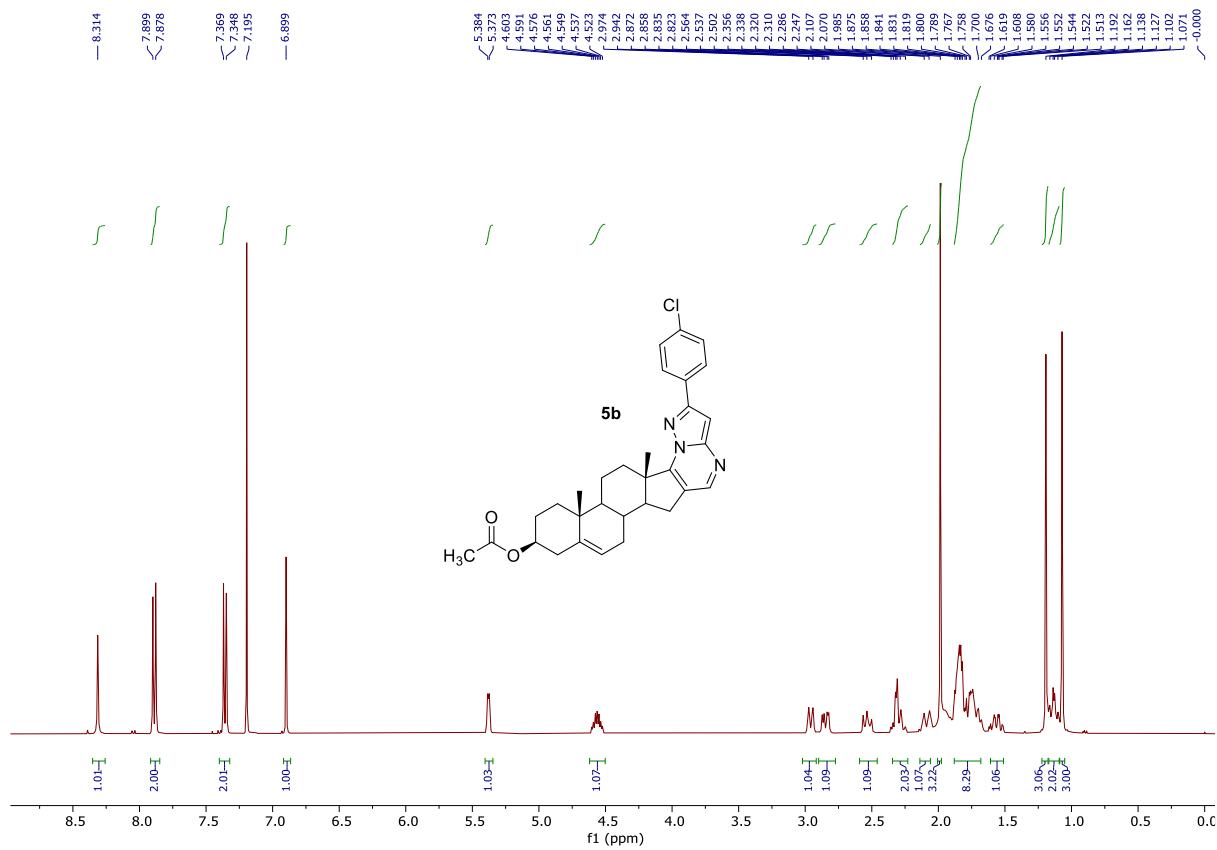


Figure S5. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) of **5b**

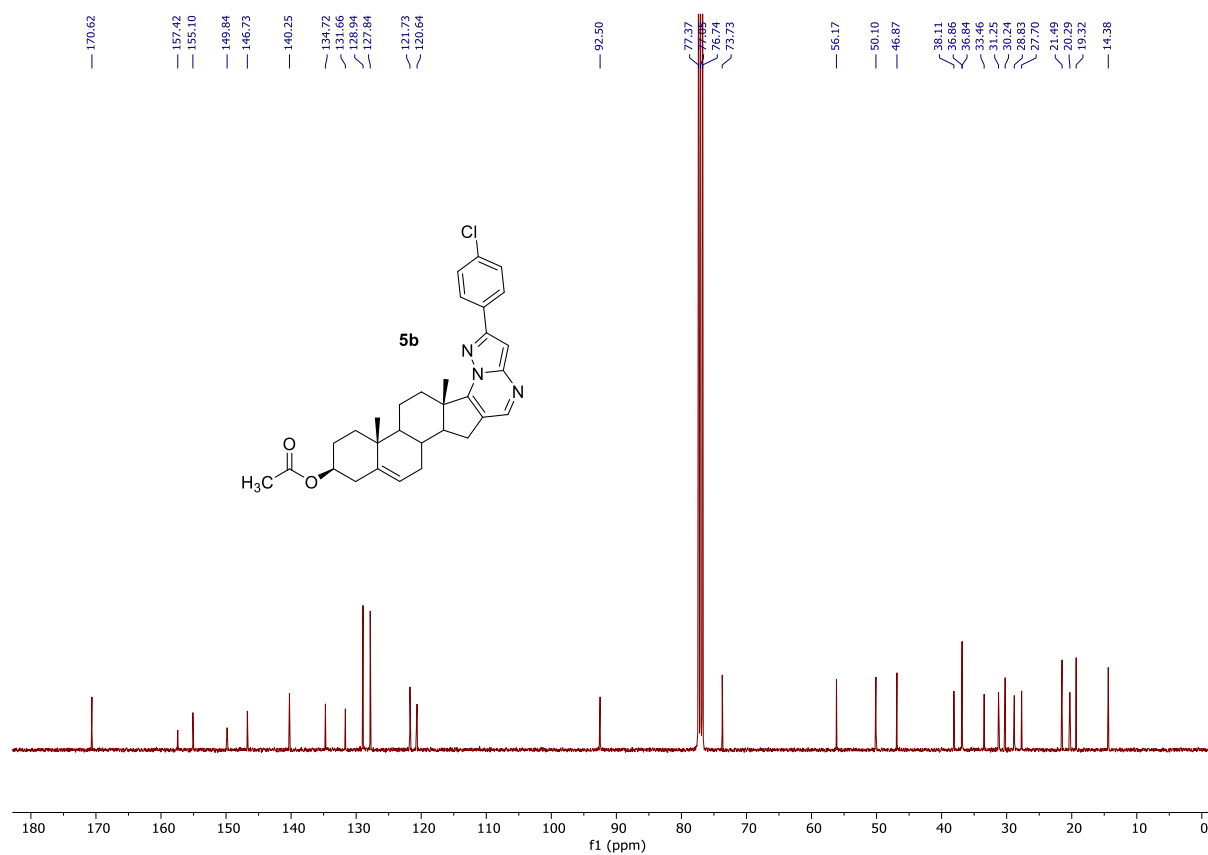


Figure S6. HRMS of **5b**

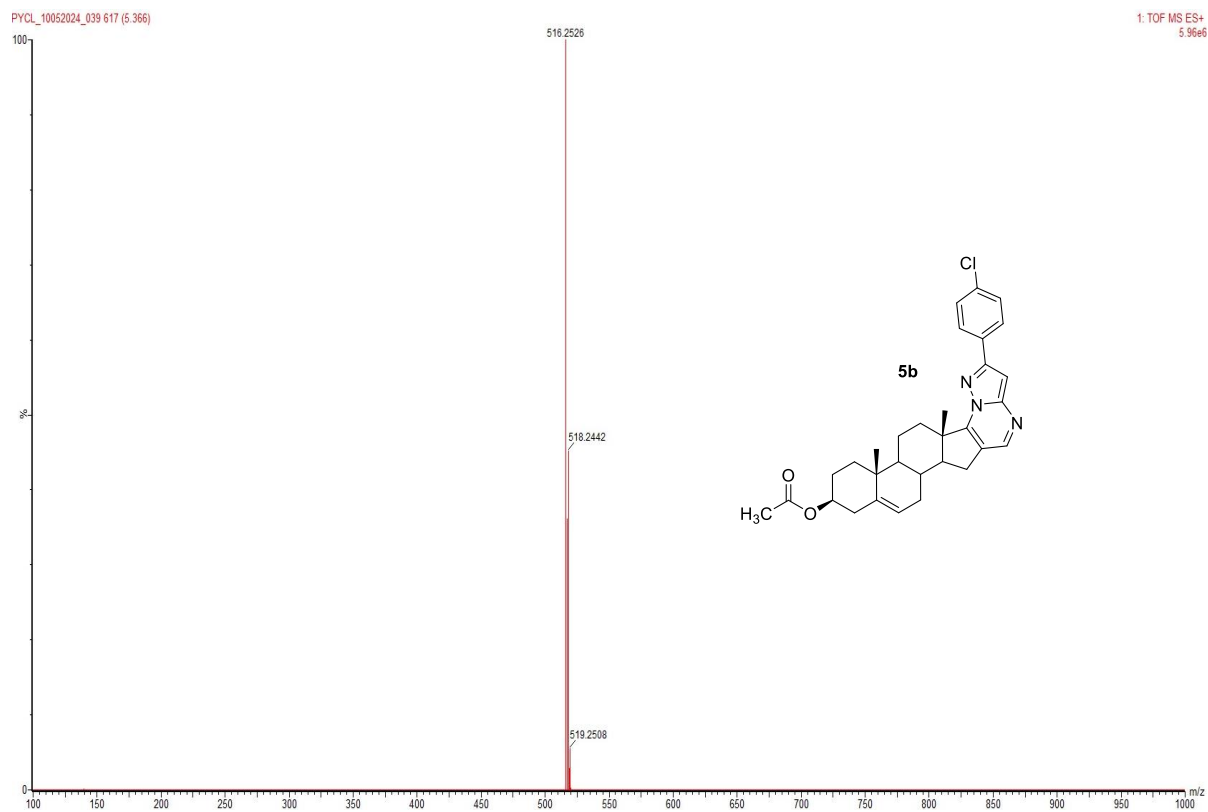


Figure S7. ^1H NMR (400 MHz, CDCl_3) of 5c

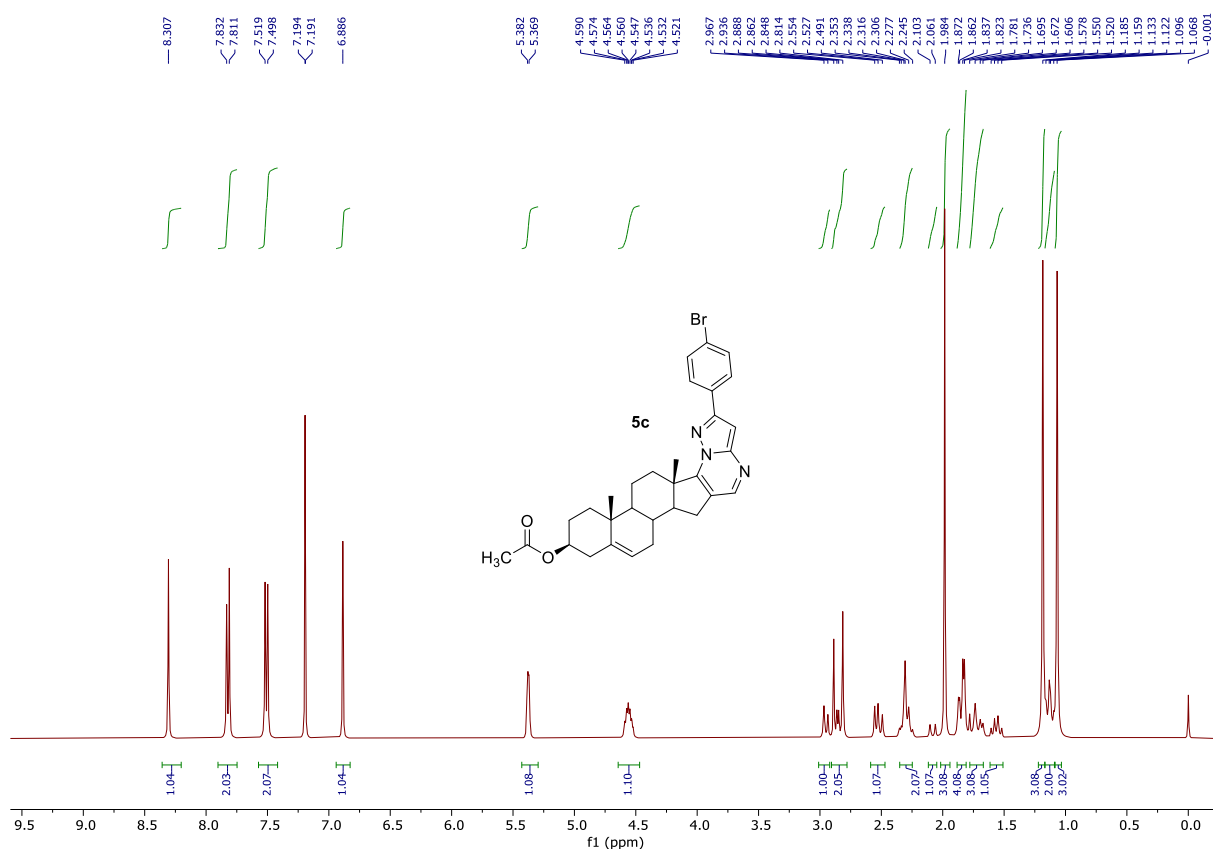


Figure S8. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) of 5c

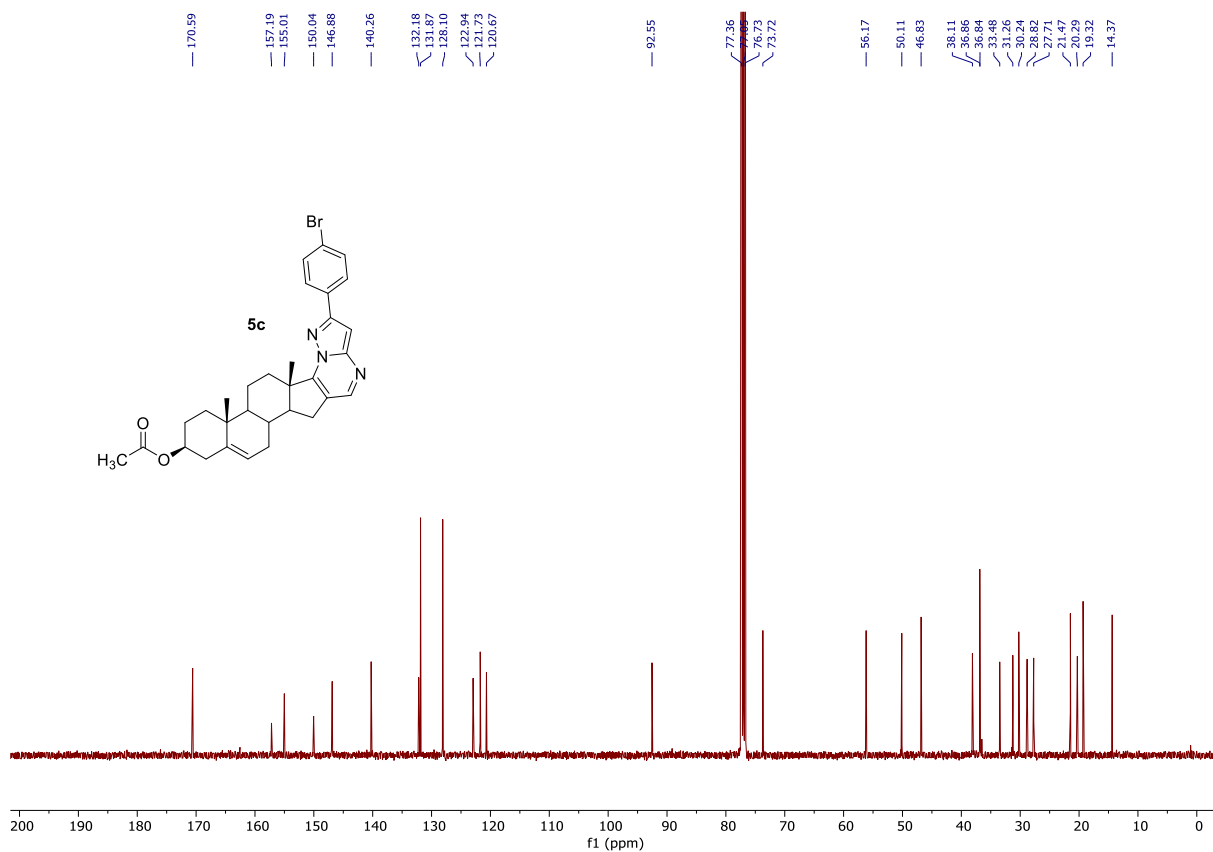


Figure S9. HRMS of 5c

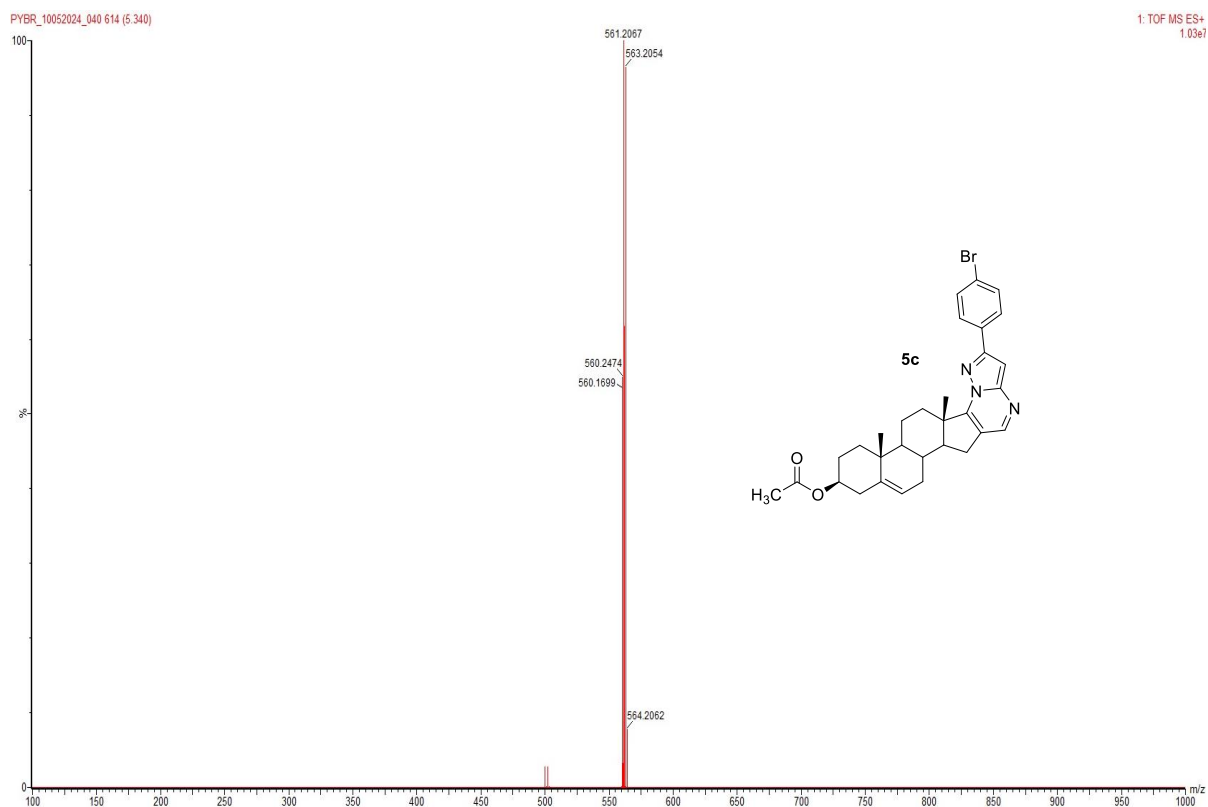


Figure S10. ¹HNMR (400 MHz, CDCl₃) of 5d

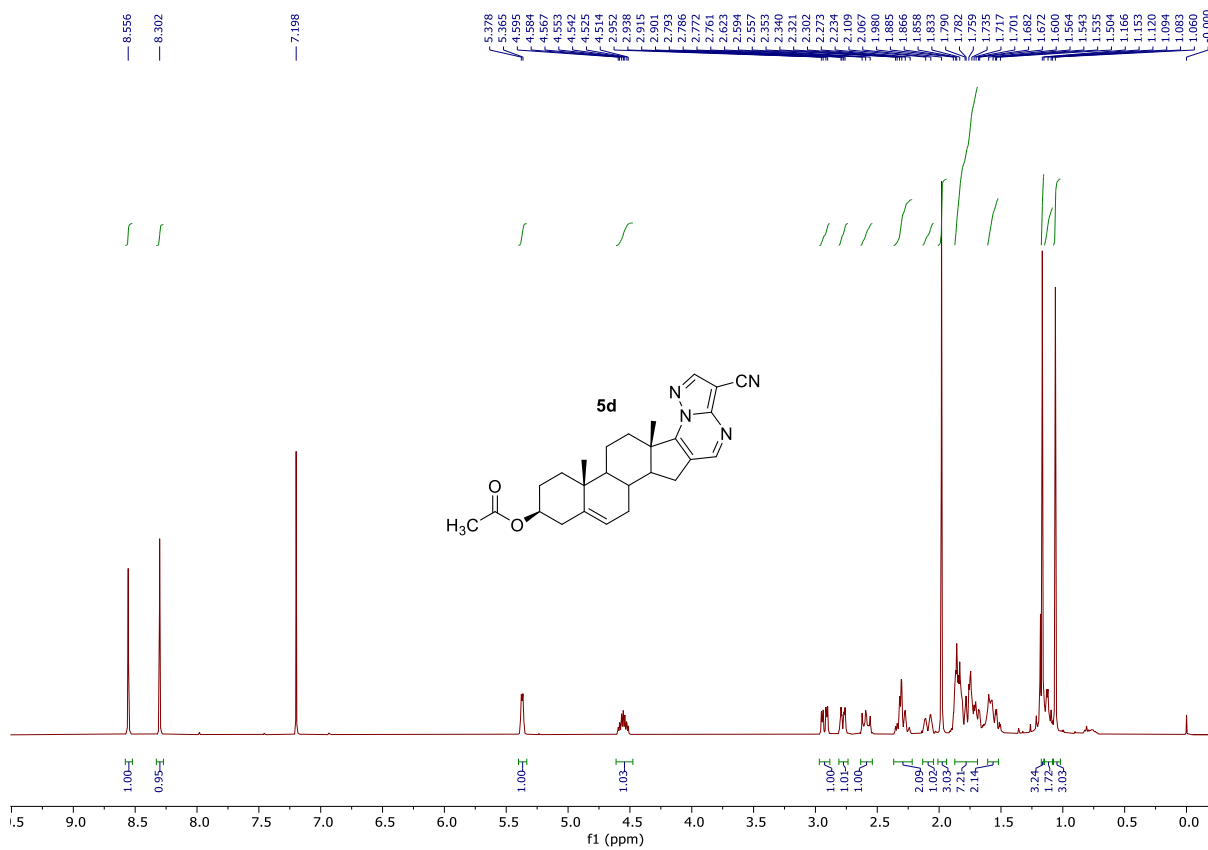


Figure S11. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) of **5d**

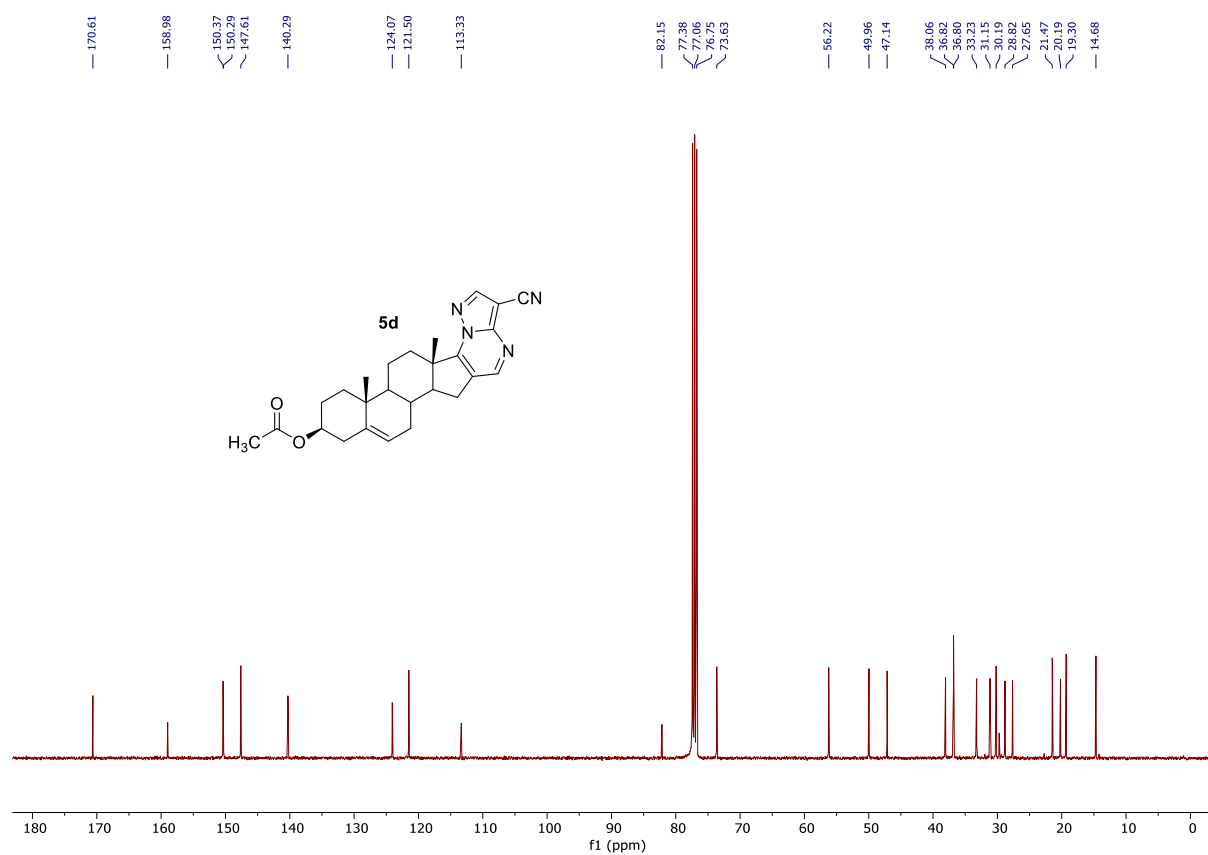


Figure S12. HRMS of **5d**

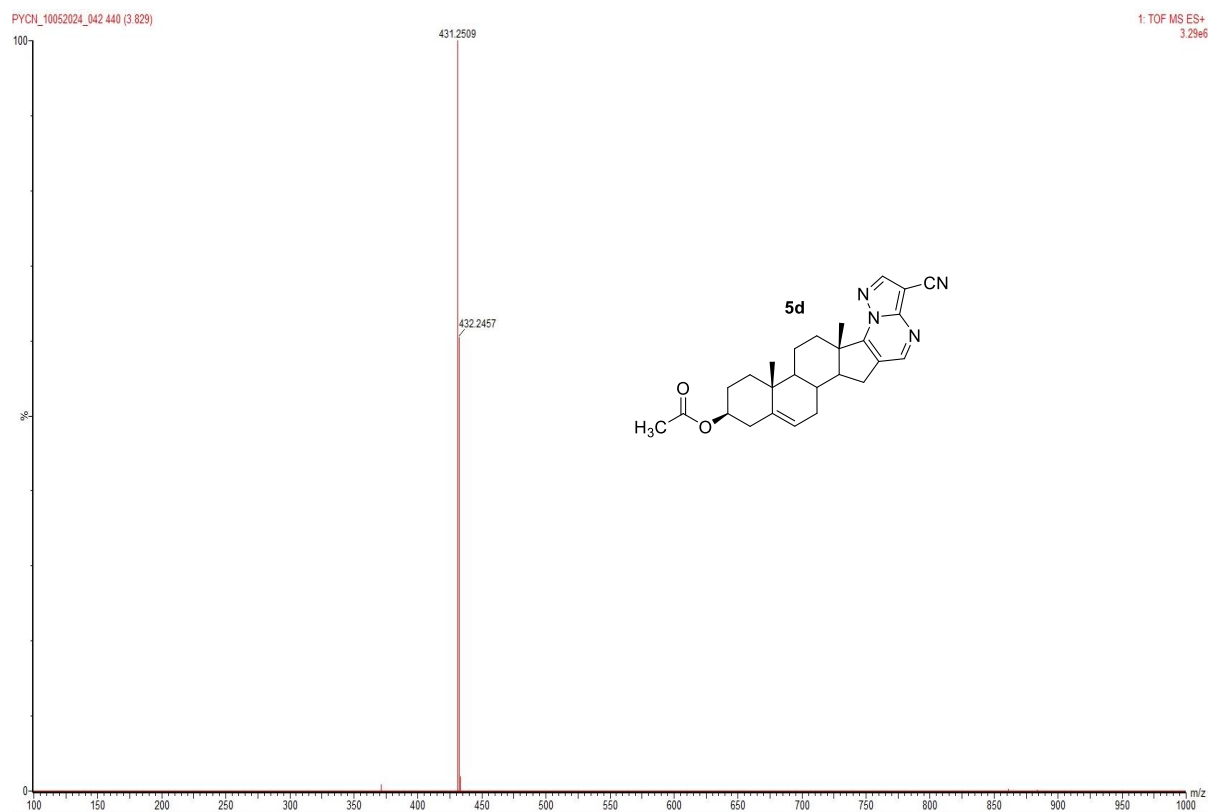


Figure S13. ^1H NMR (400 MHz, CDCl_3) of **5e**

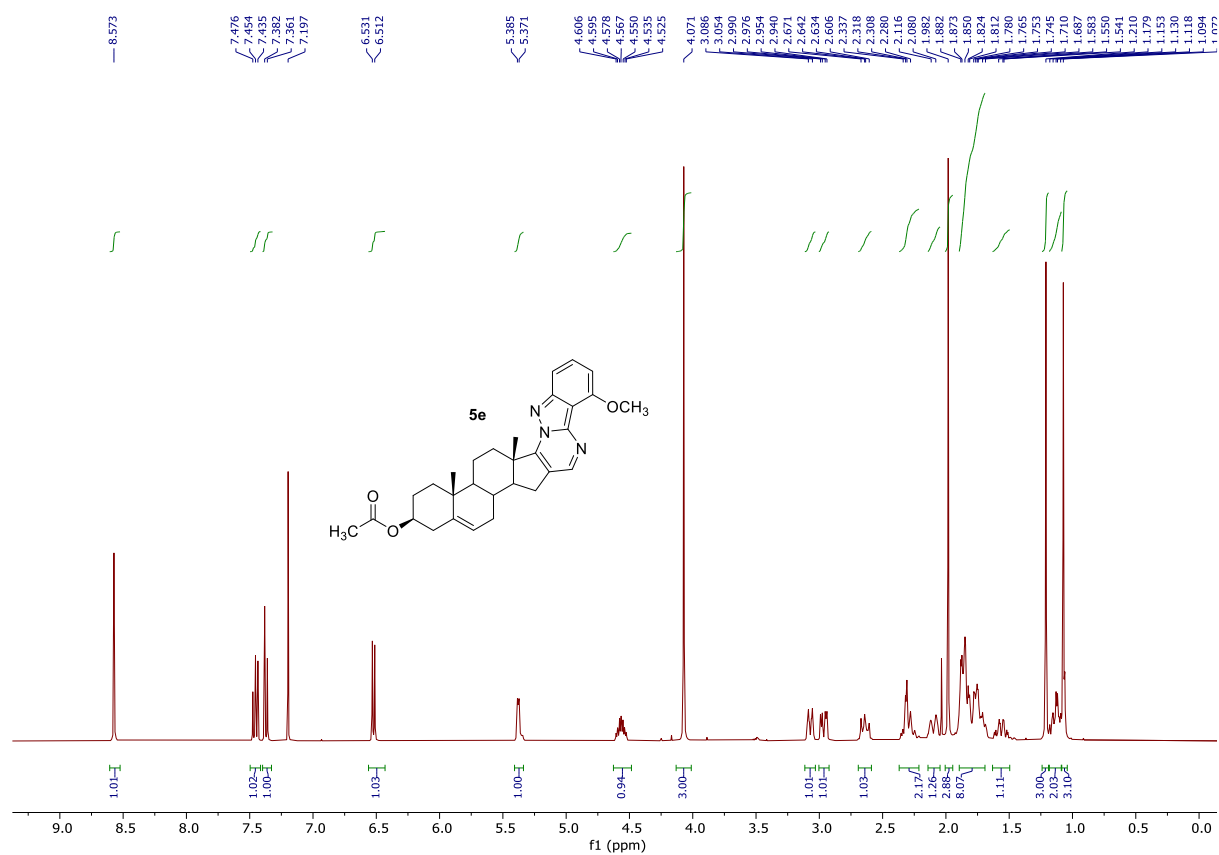


Figure S14. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) of **5e**

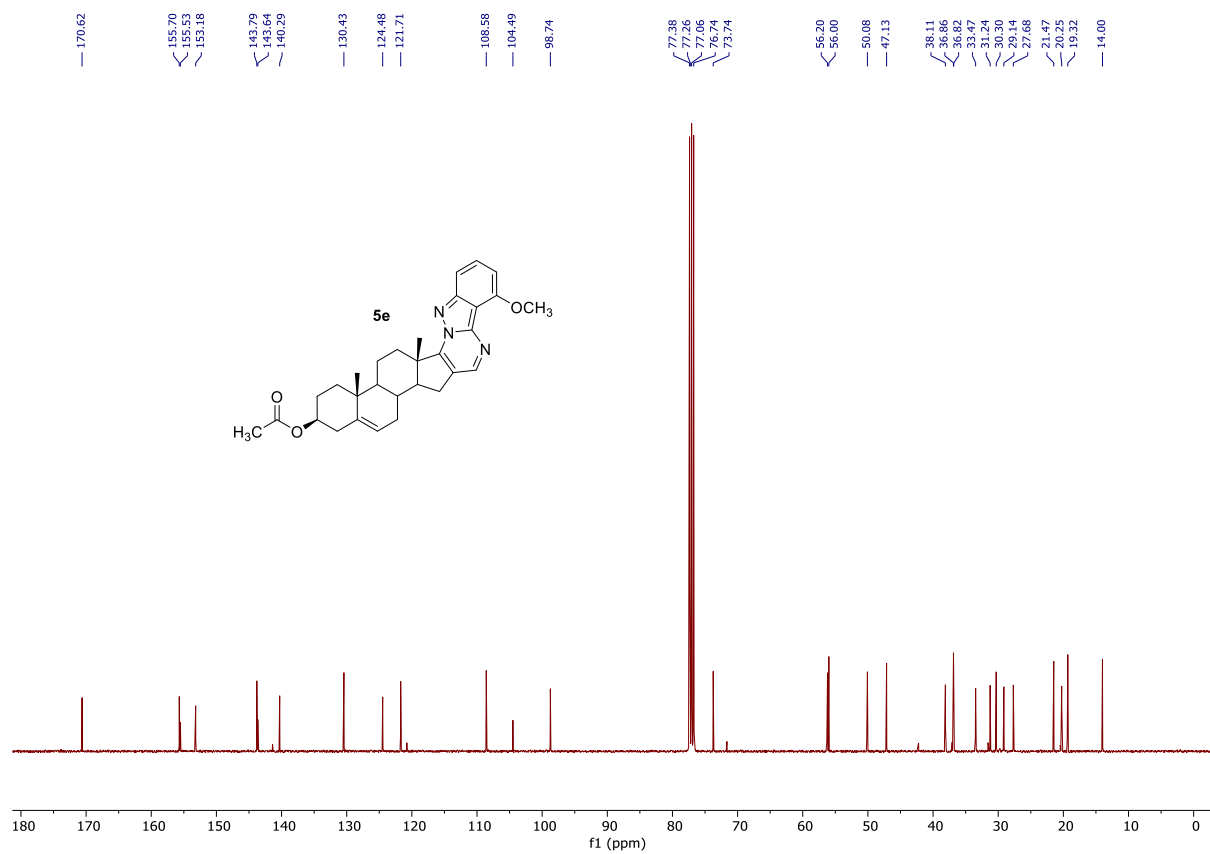


Figure S15. HRMS of 5e

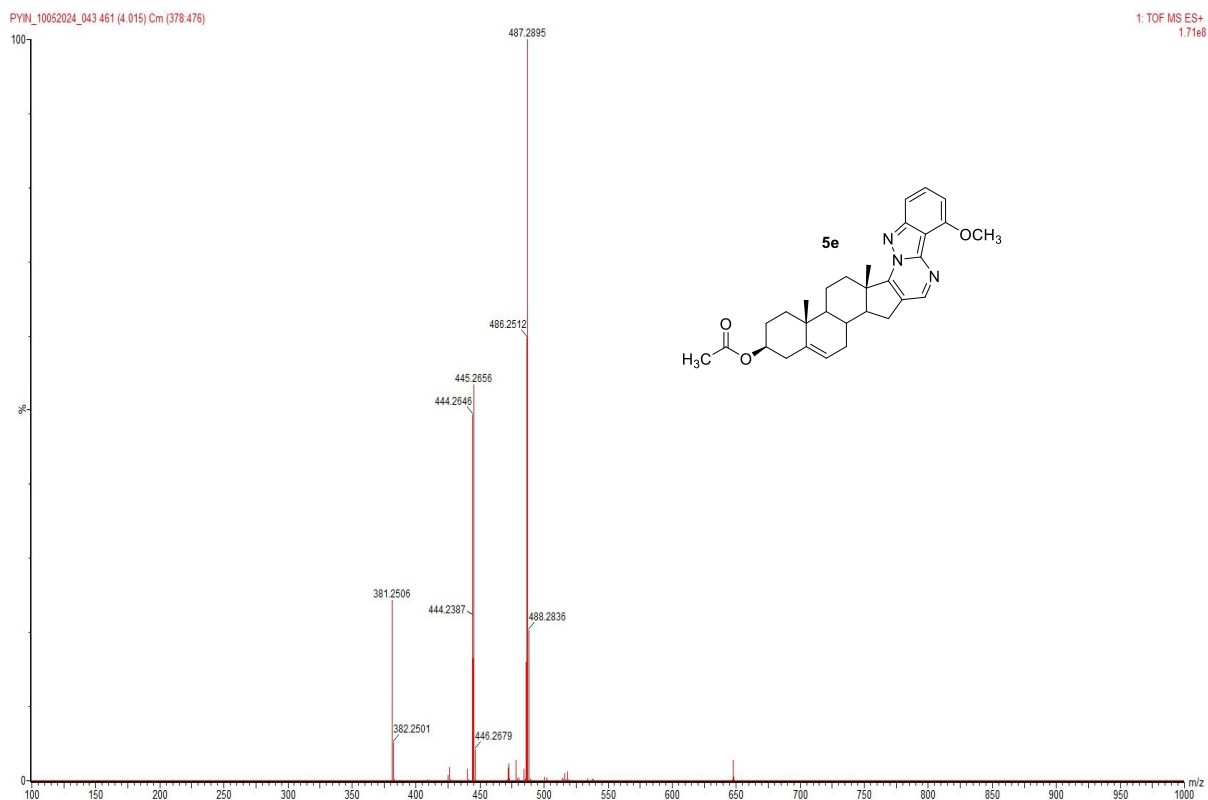


Figure S16. ¹HNMR (400 MHz, CDCl₃) of 6a

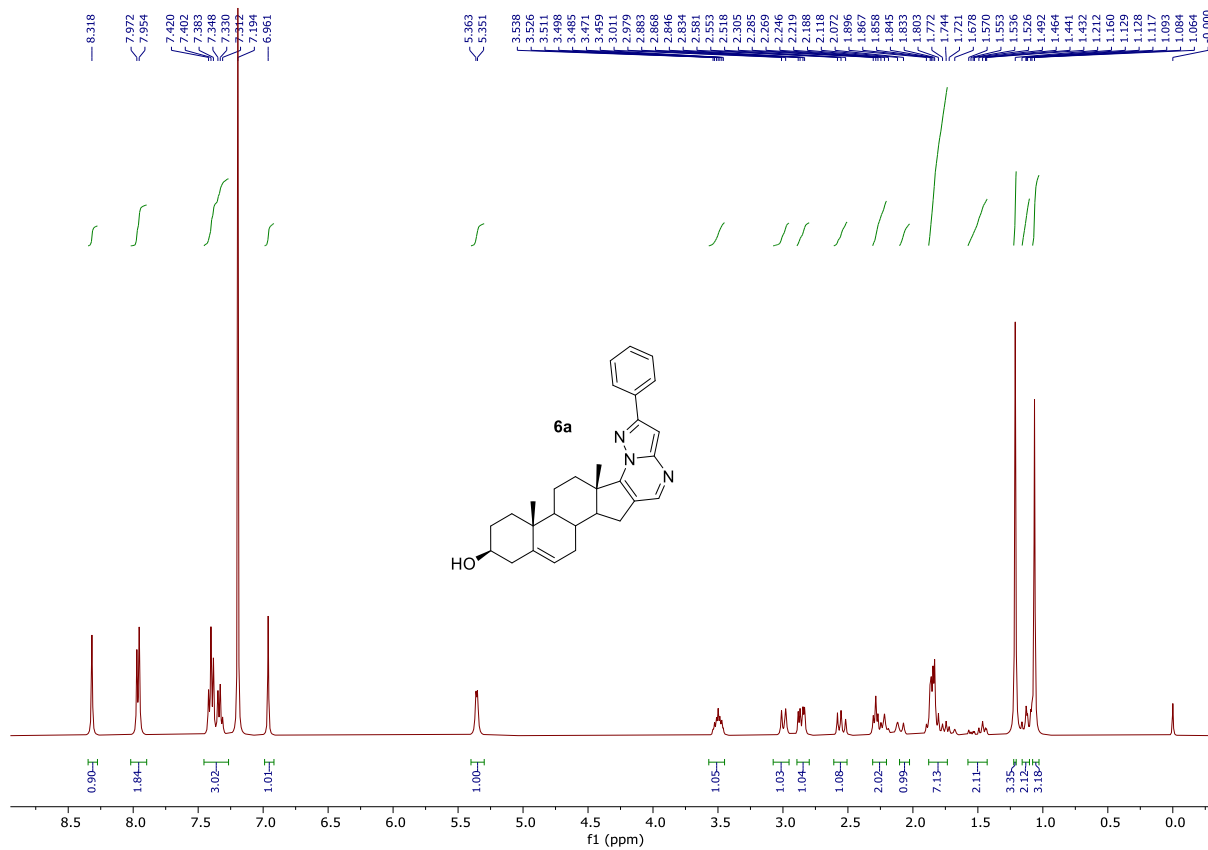


Figure S17. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) of 6a

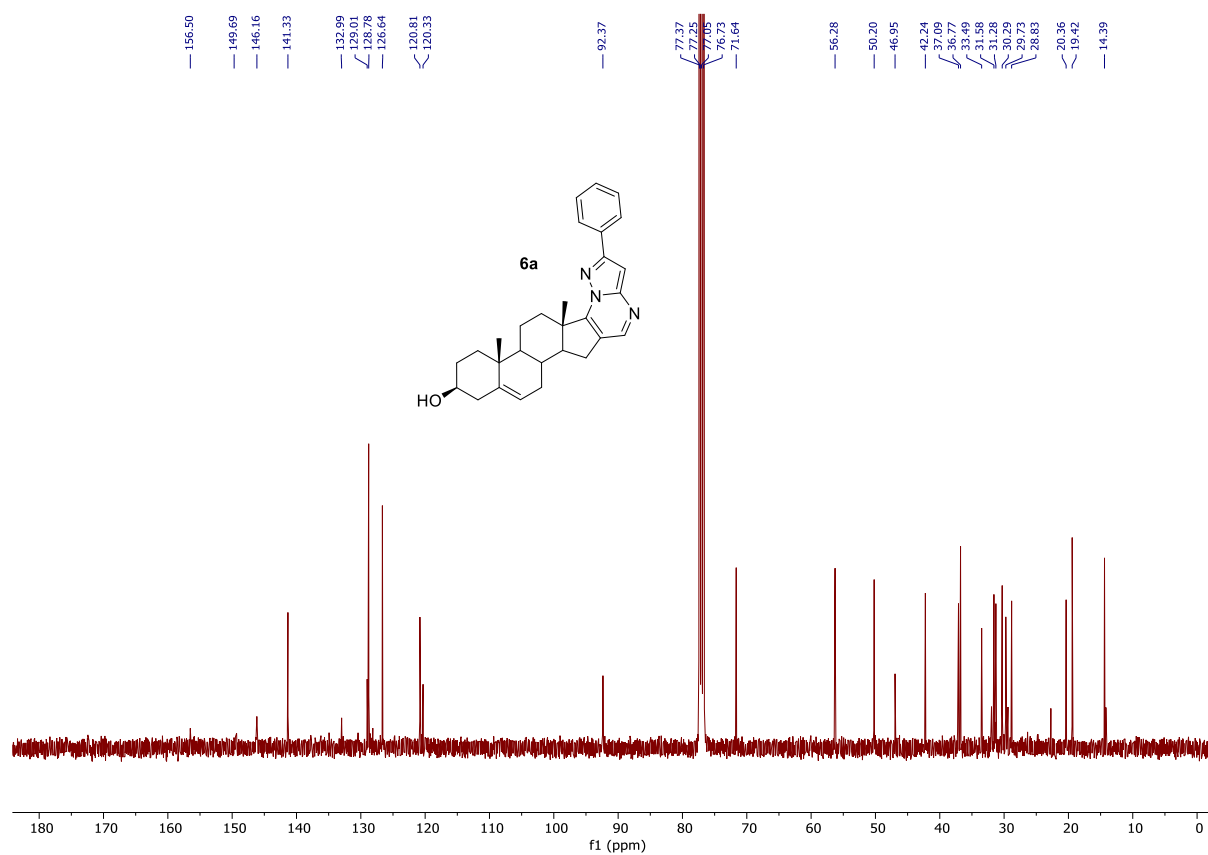


Figure S18. HRMS of 6a

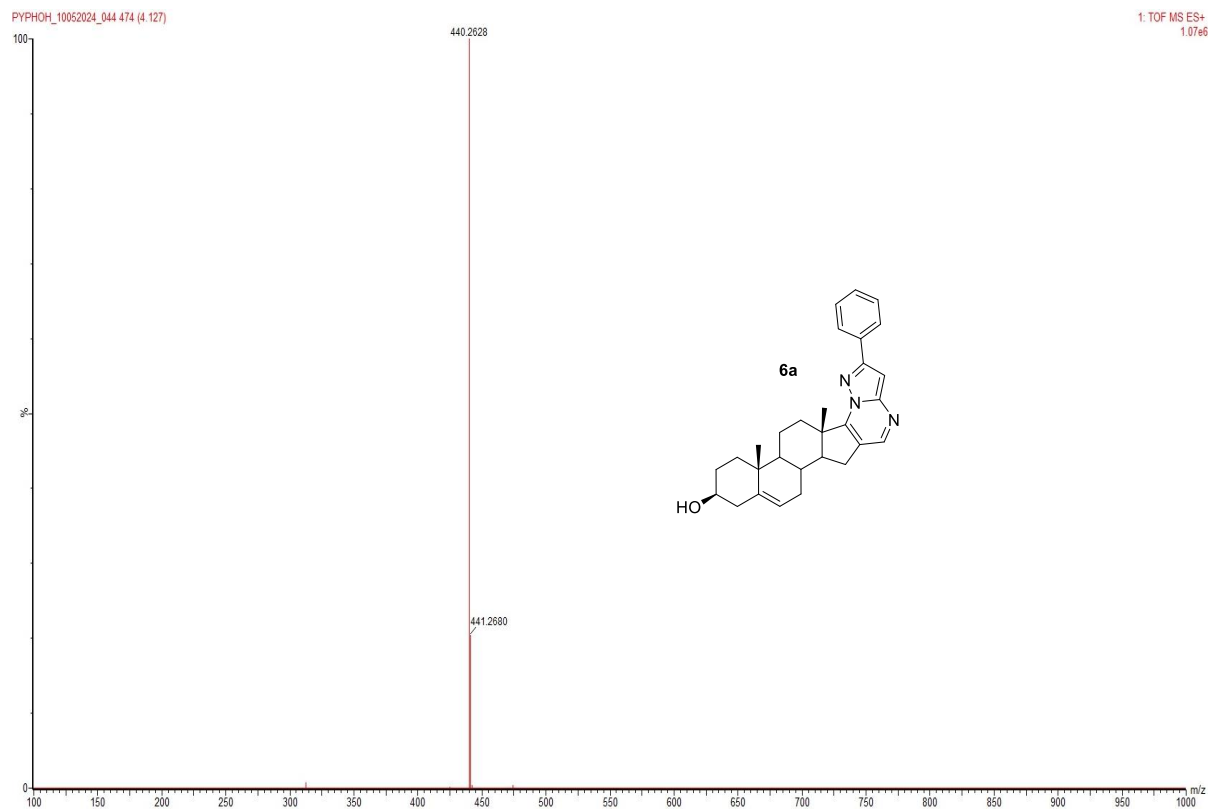


Figure S19. ^1H NMR (400 MHz, CDCl_3) of **6b**

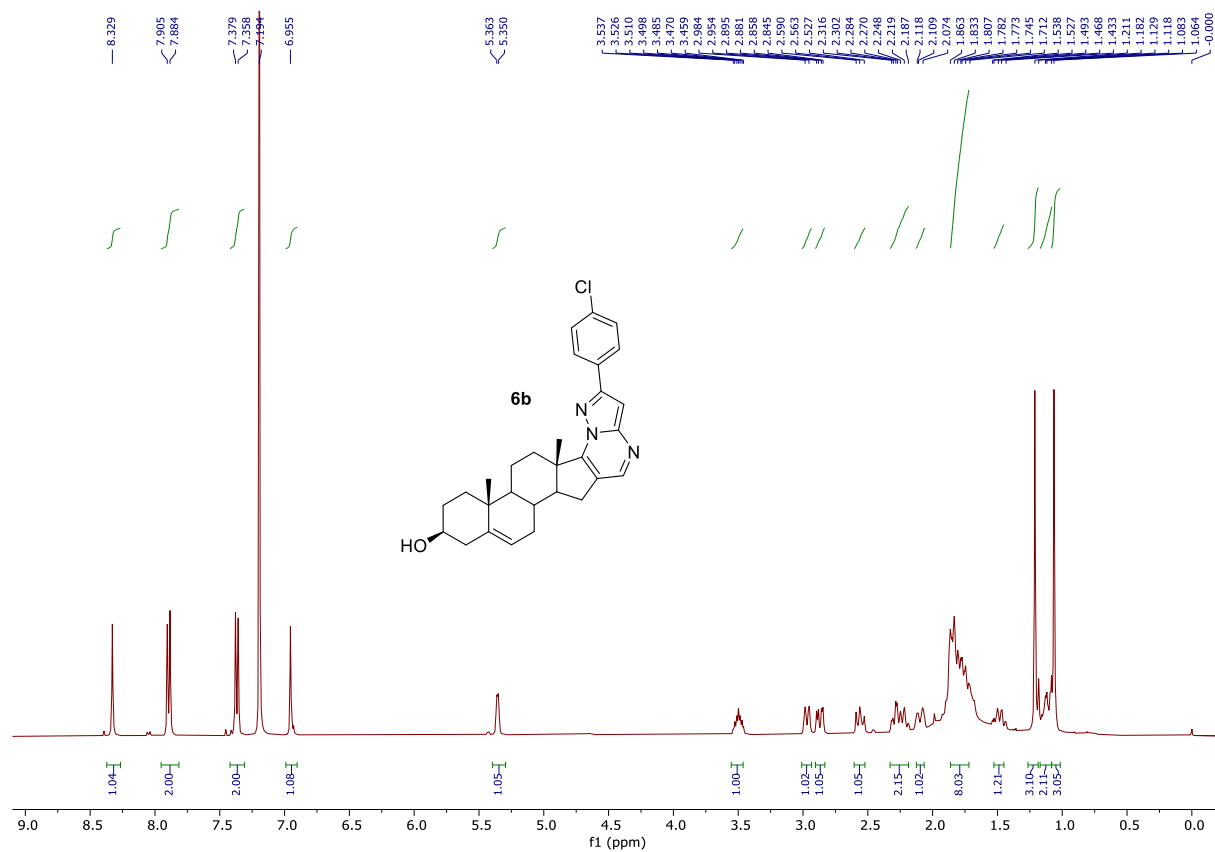


Figure S20 $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, $\text{DMSO}-d_6$) of **6b**

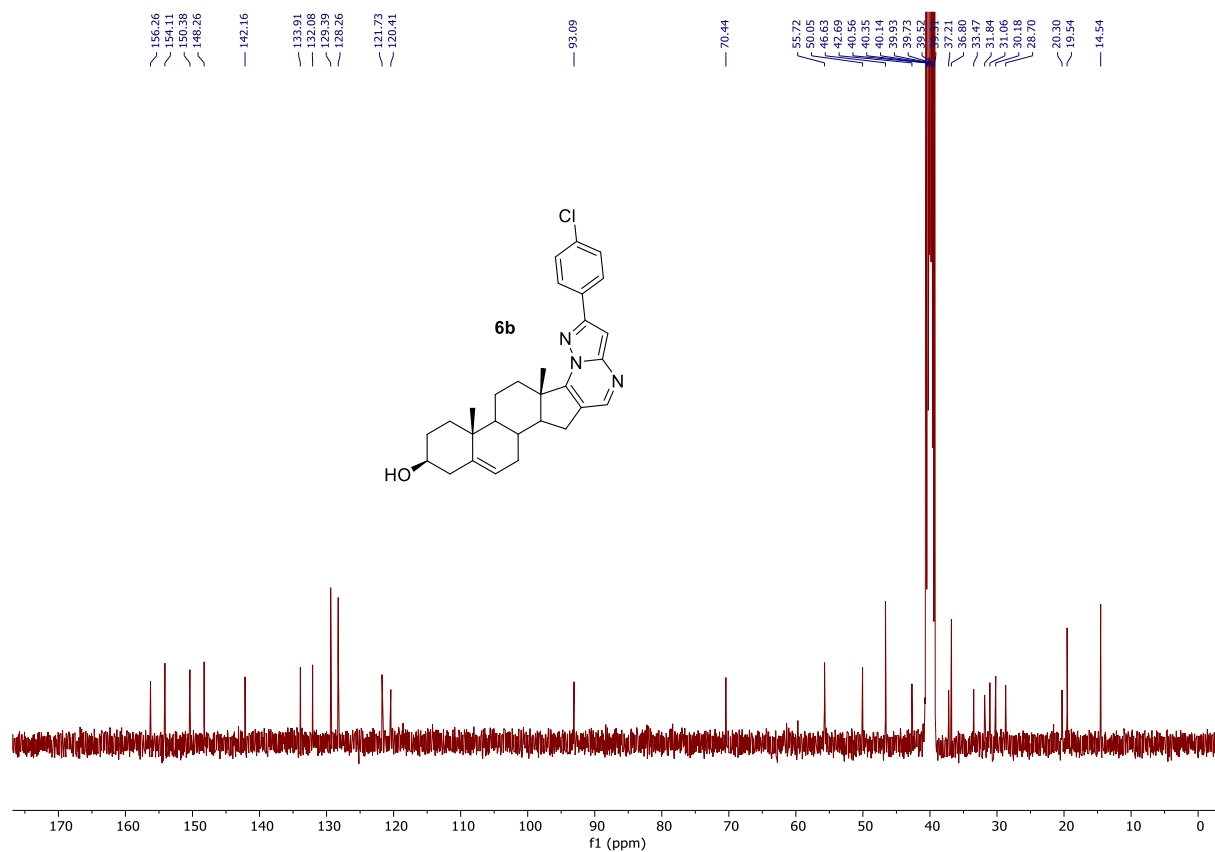


Figure S21. HRMS of 6b

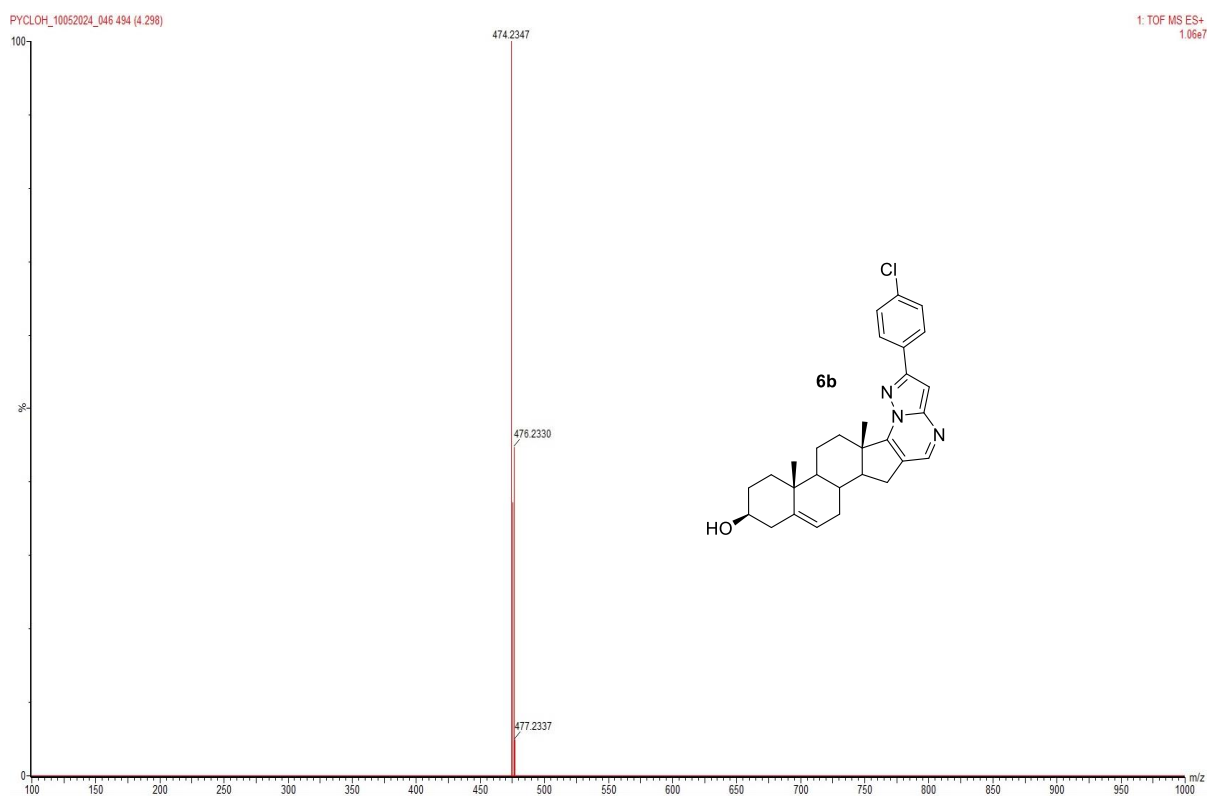


Figure S22. ¹HNMR (400 MHz, CDCl₃) of 6c

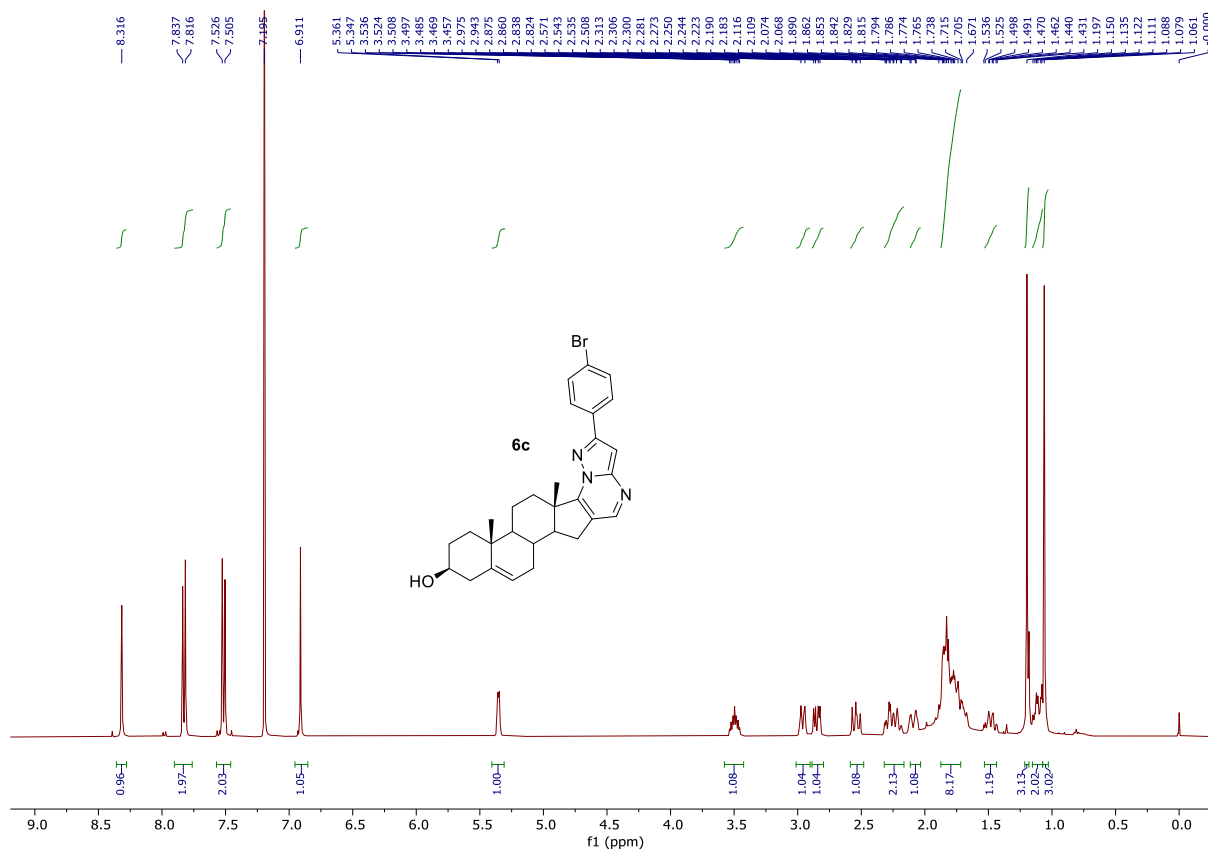


Figure S23. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) of **6c**

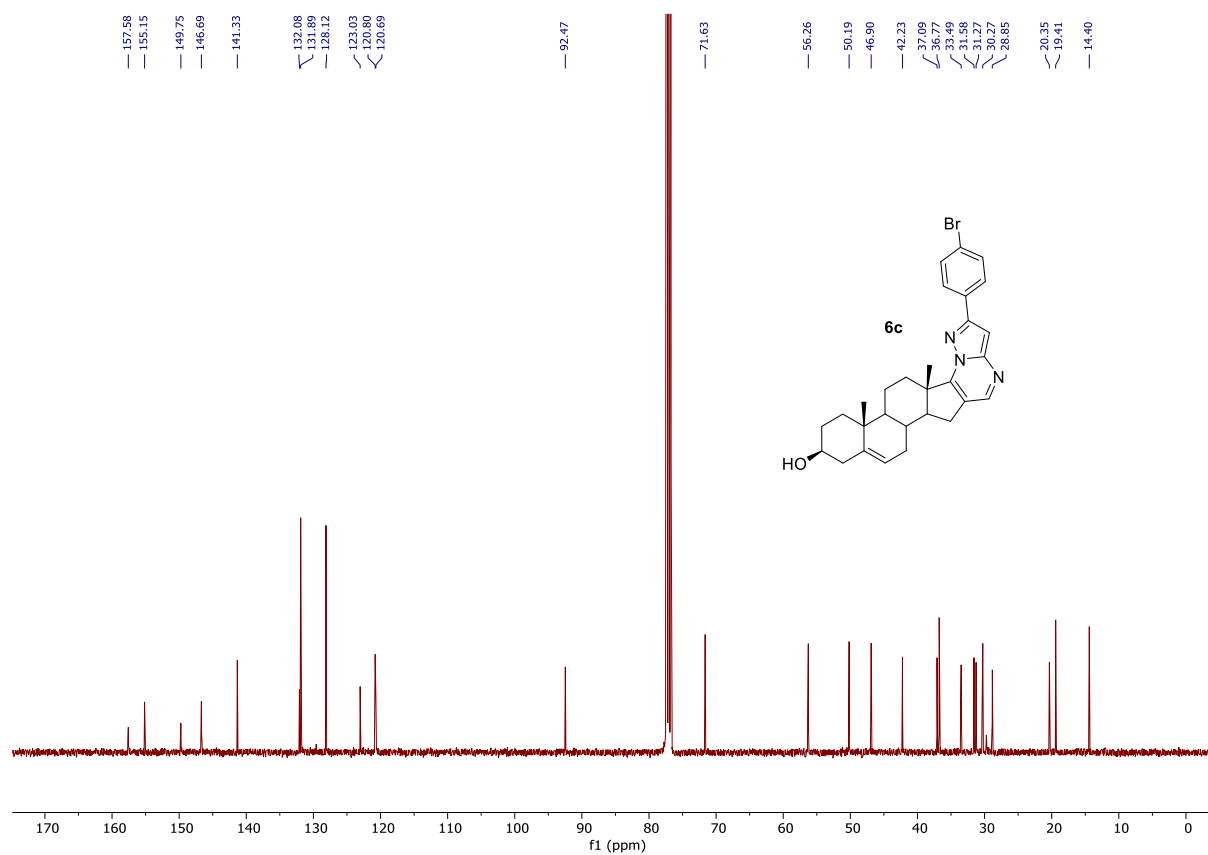


Figure S24. HRMS of **6c**

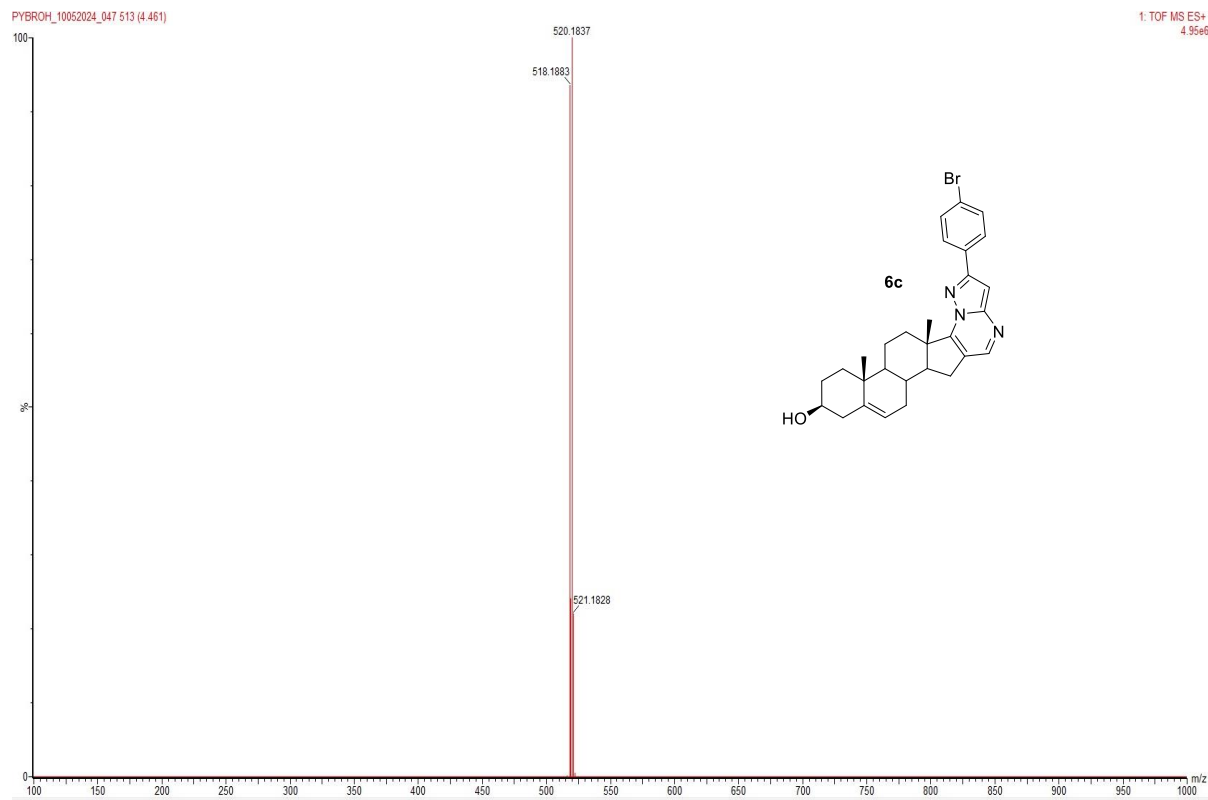


Figure S25. ^1H NMR (400 MHz, CDCl_3) of 6d

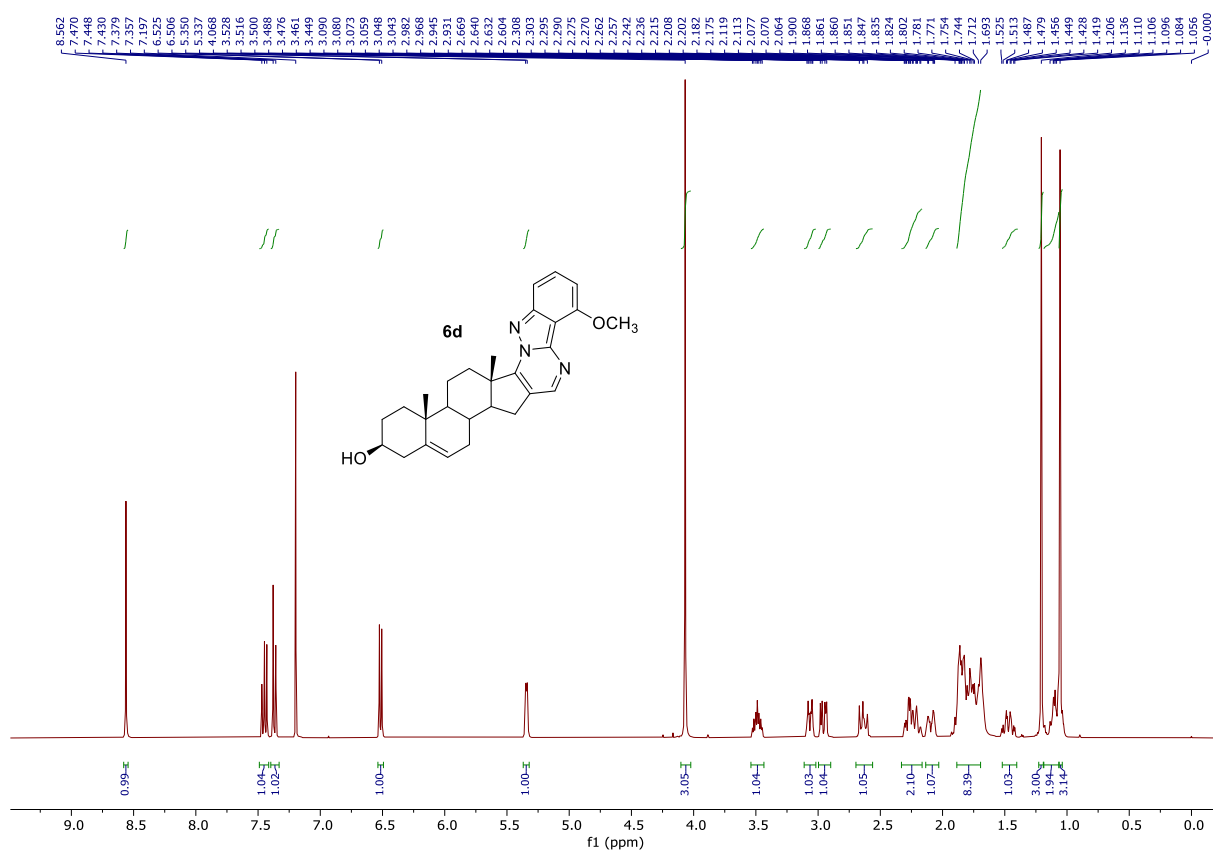


Figure S26. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) of 6d

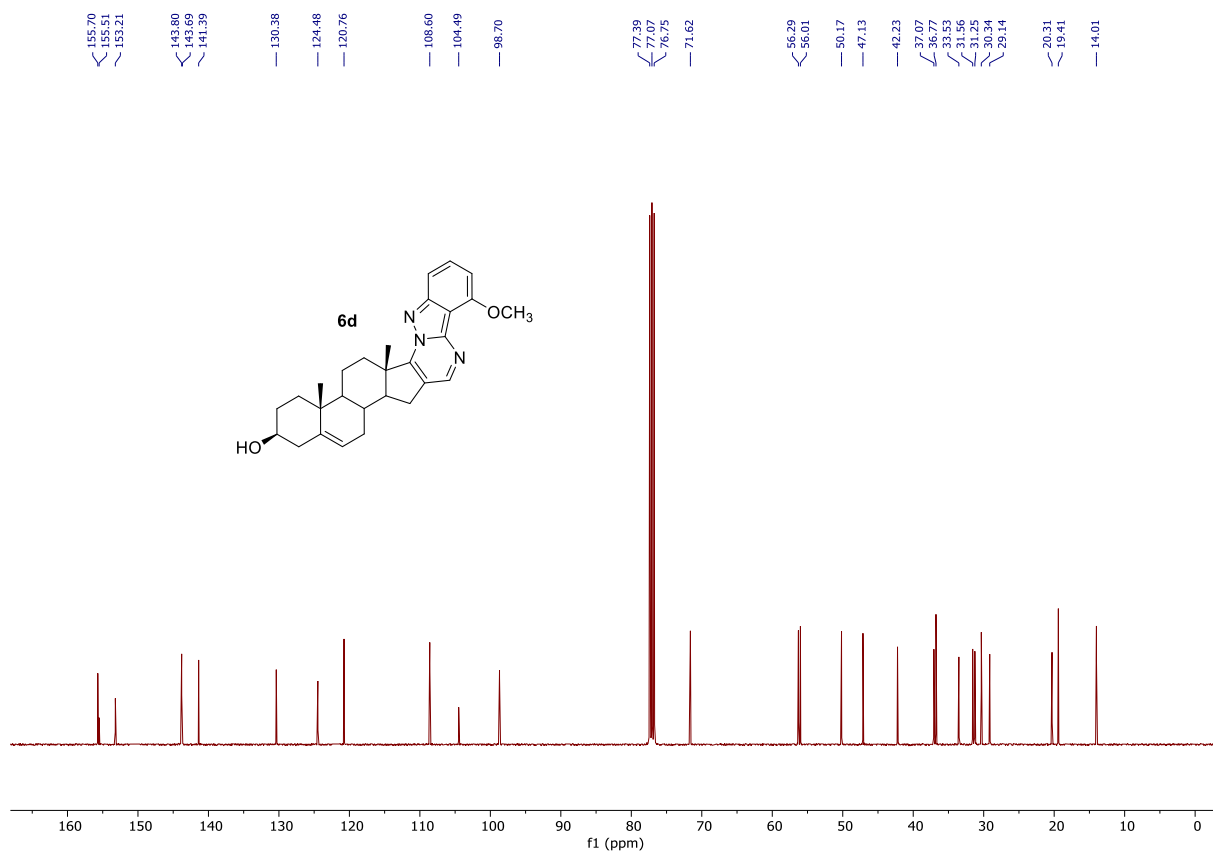


Figure S27. HRMS of 6d

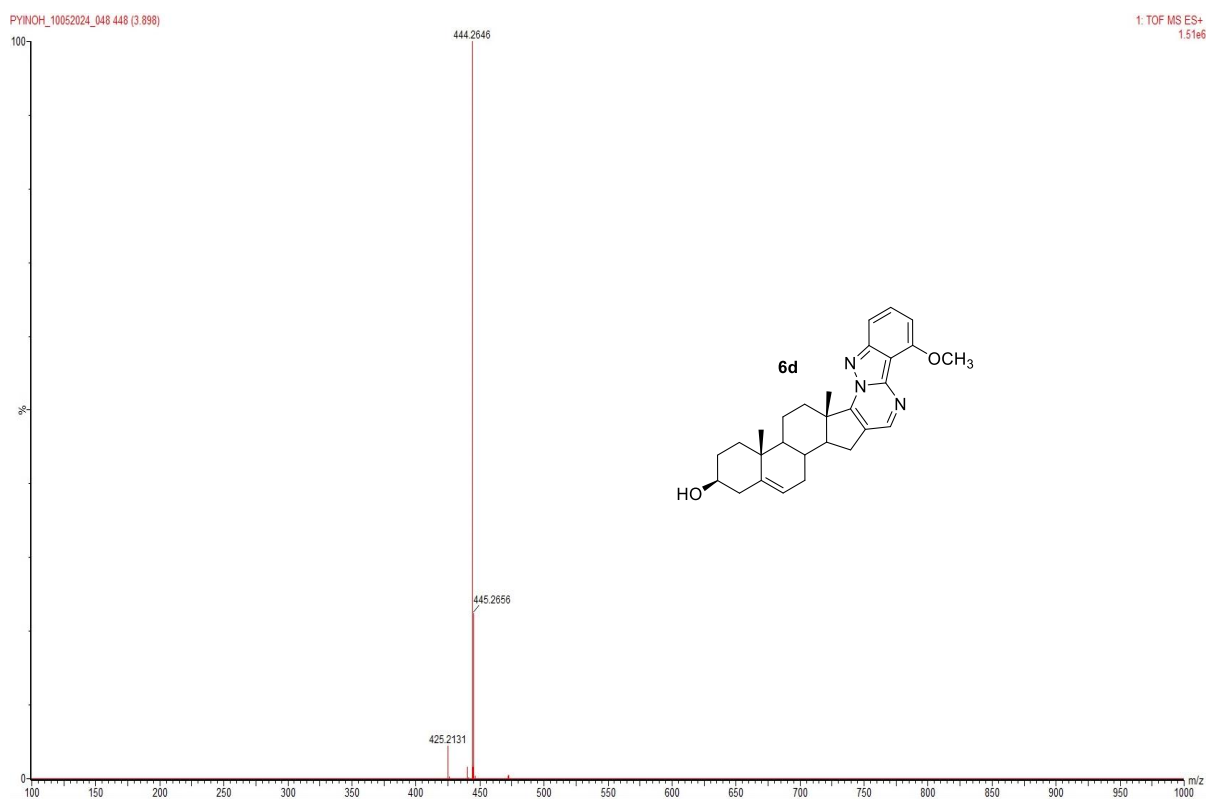


Figure S28. ¹HNMR (400 MHz, CDCl₃) of 8a

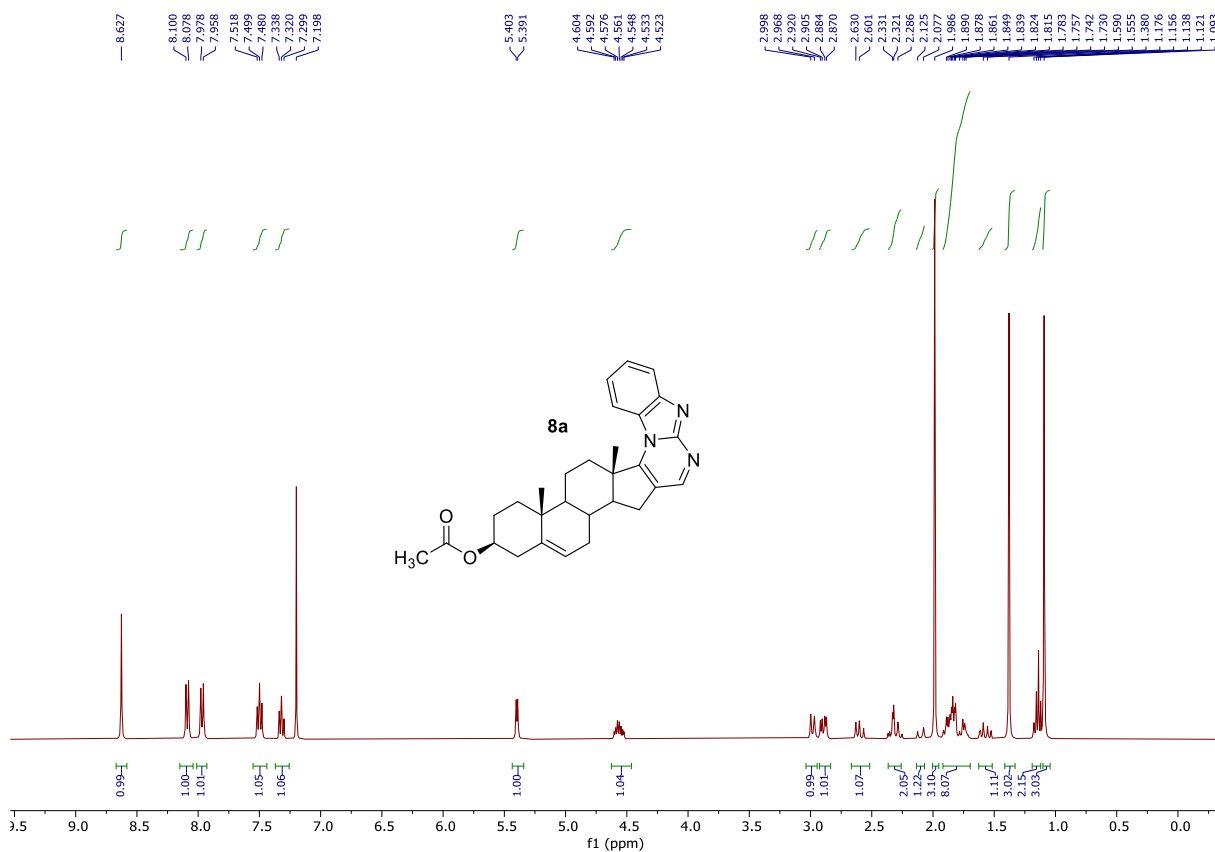


Figure S29. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) of 8a

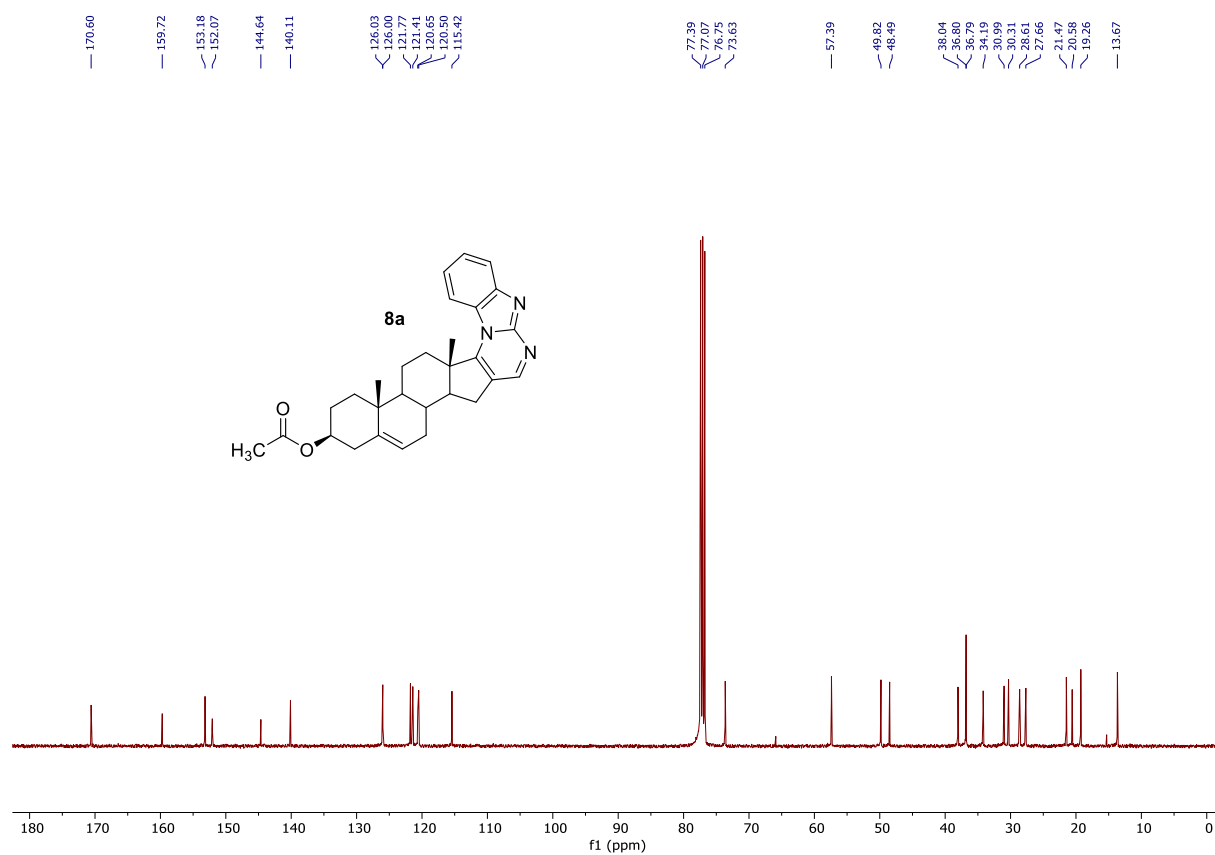


Figure S30. HRMS of 8a

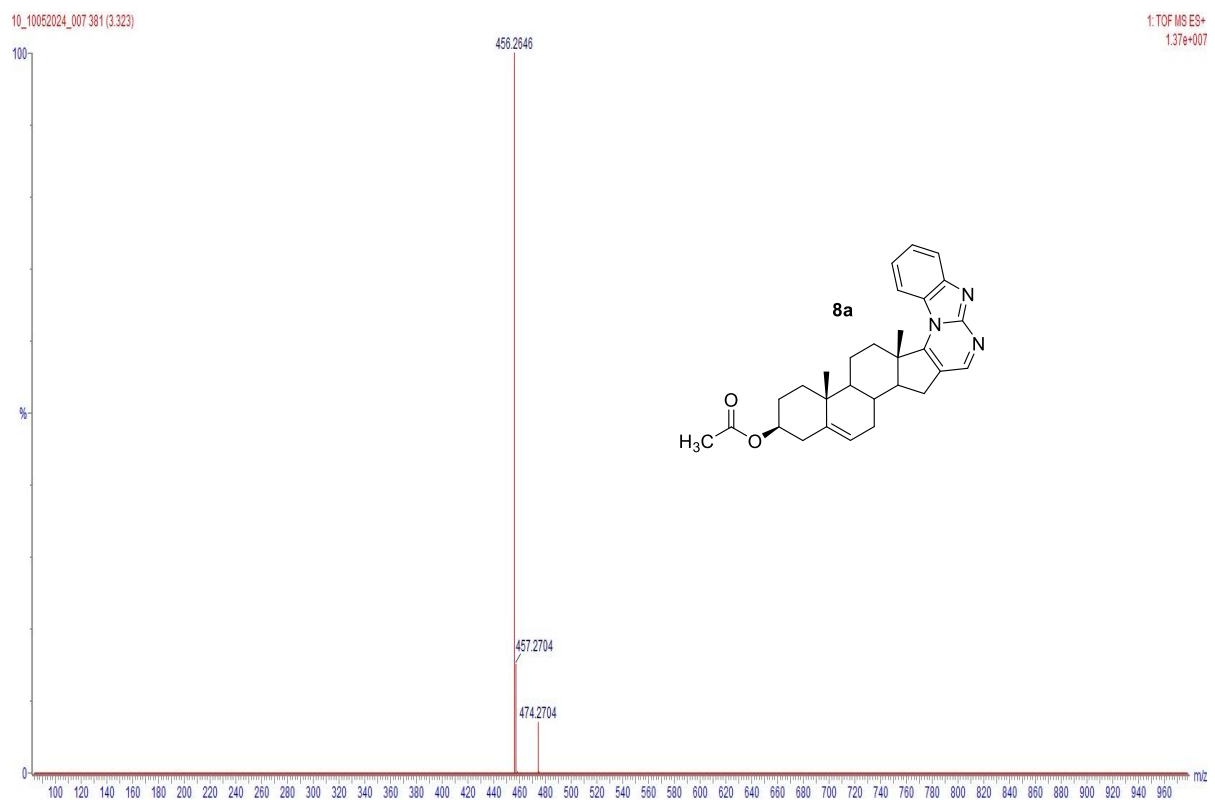


Figure S31. ^1H NMR (400 MHz, CDCl_3) of **8b**

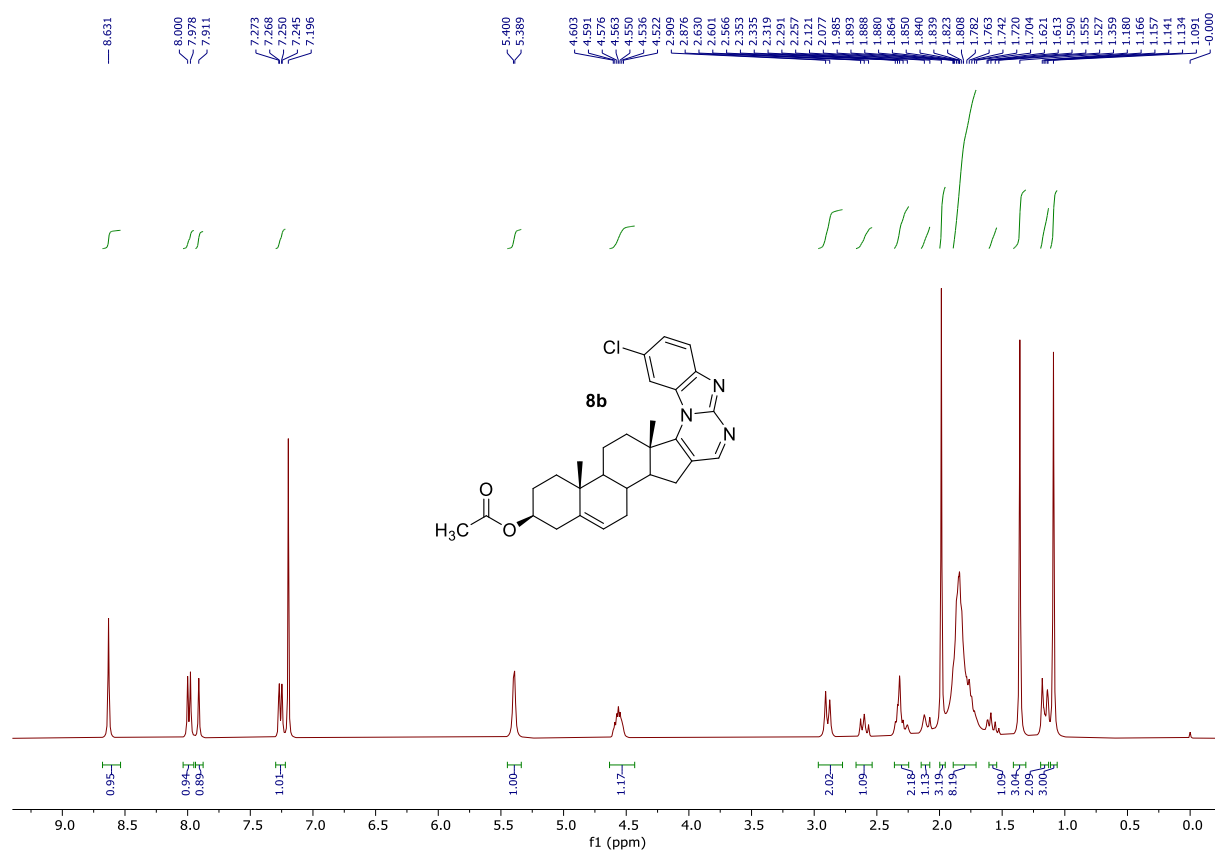


Figure S32. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) of **8b**

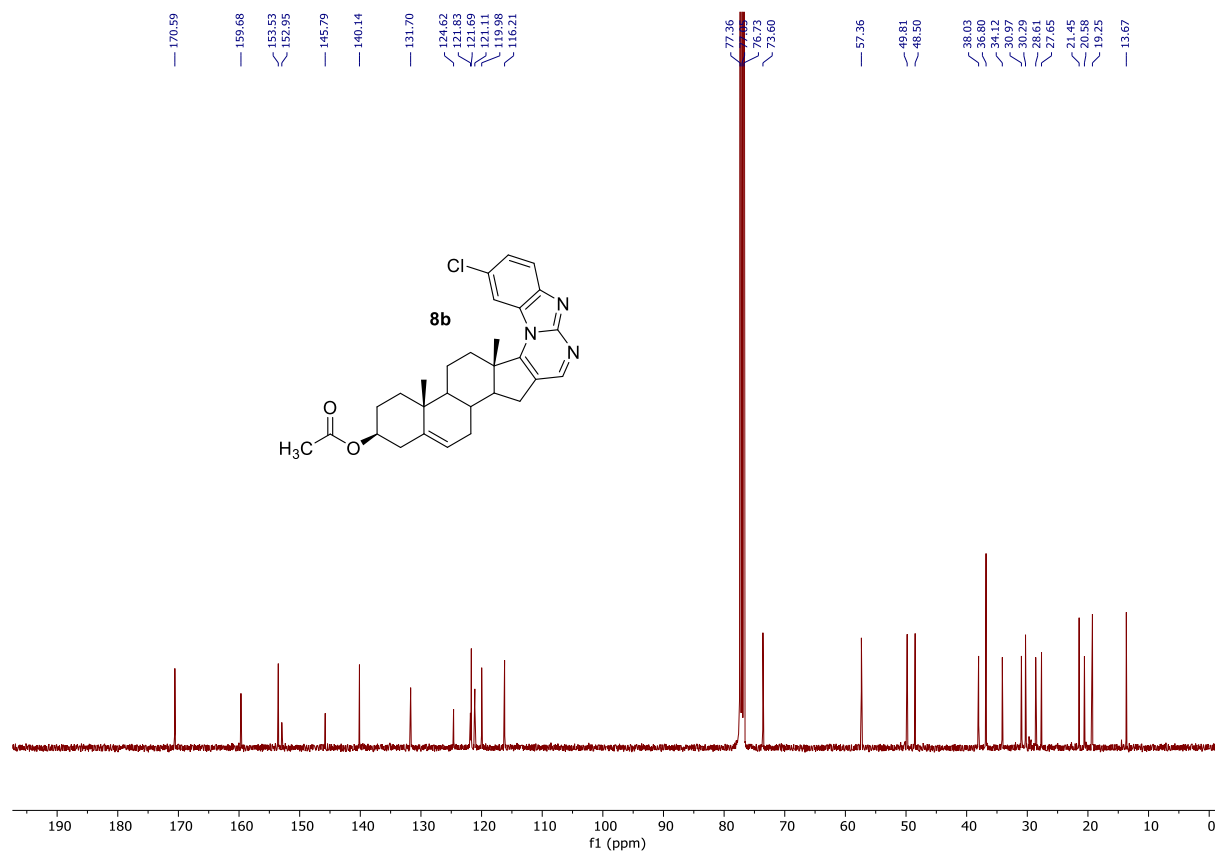


Figure S33. HRMS of 8b

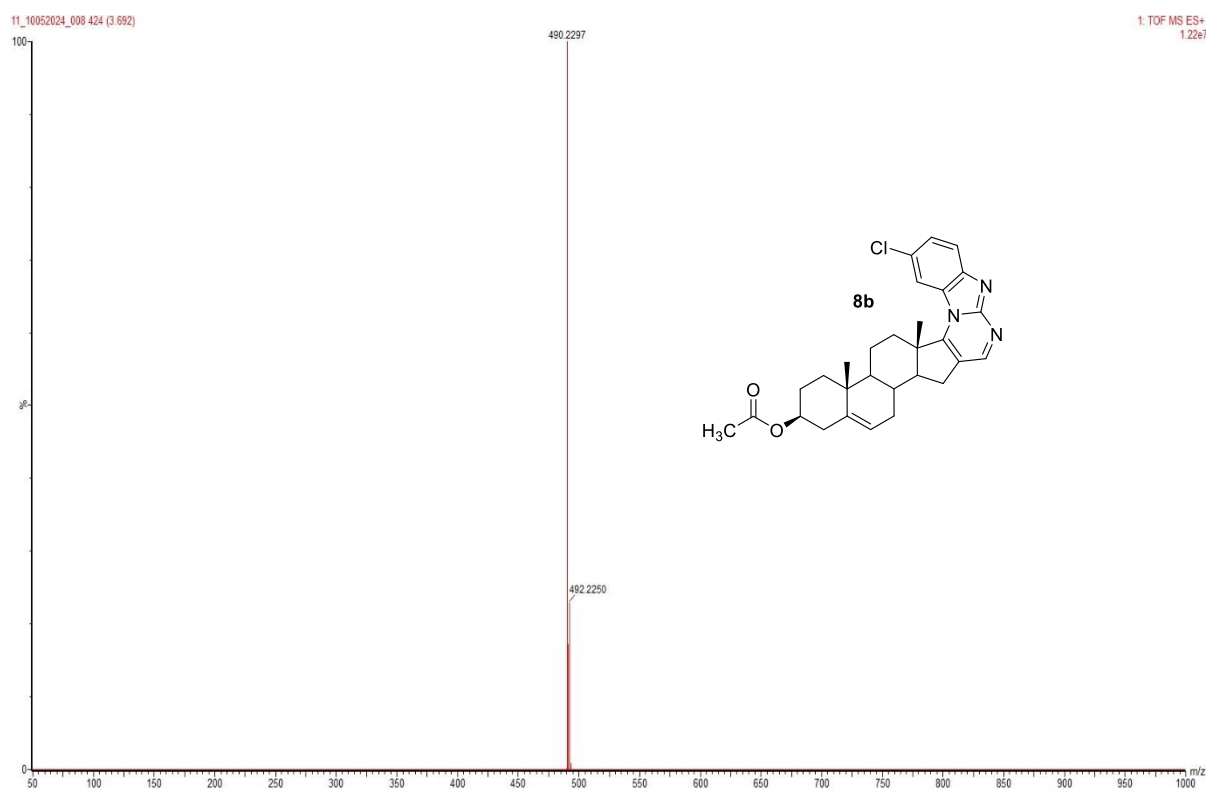


Figure S34. ¹HNMR (400 MHz, CDCl₃) of 8c

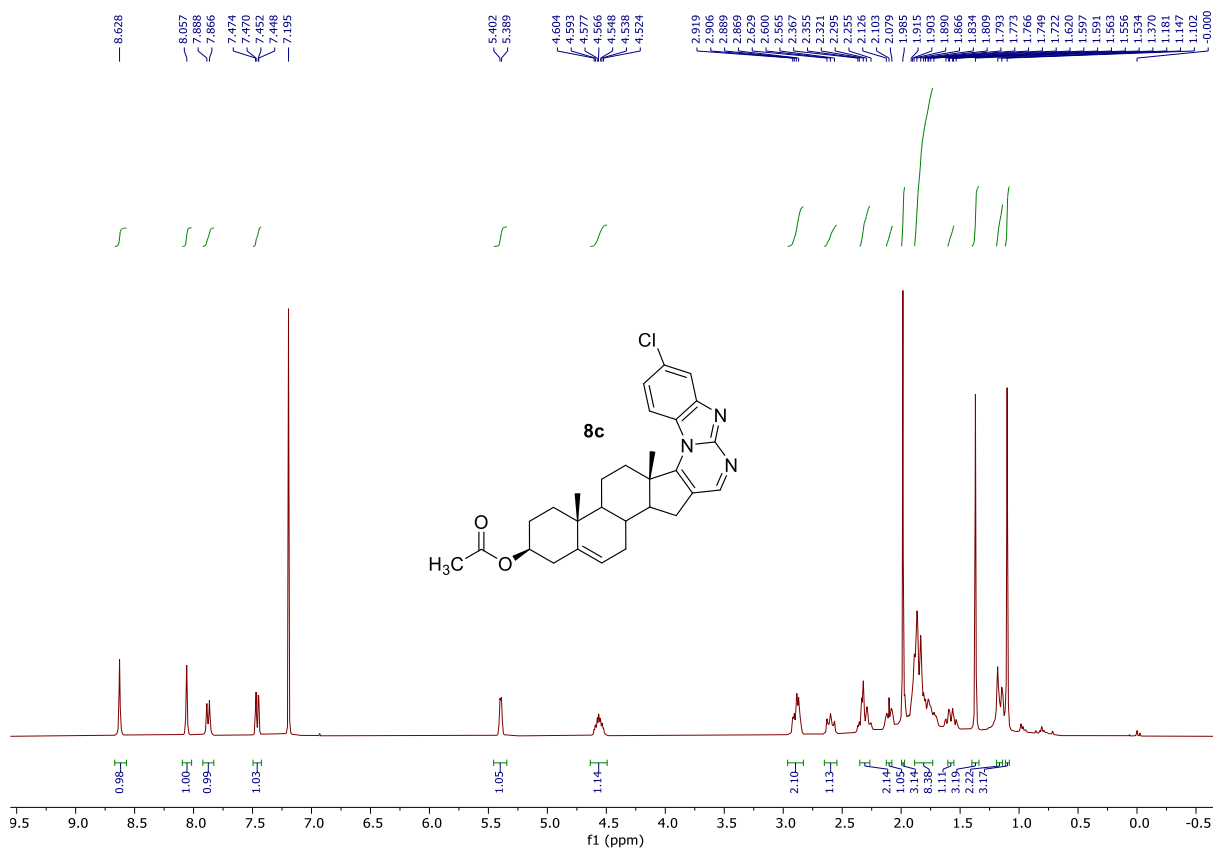


Figure S35. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) of **8c**

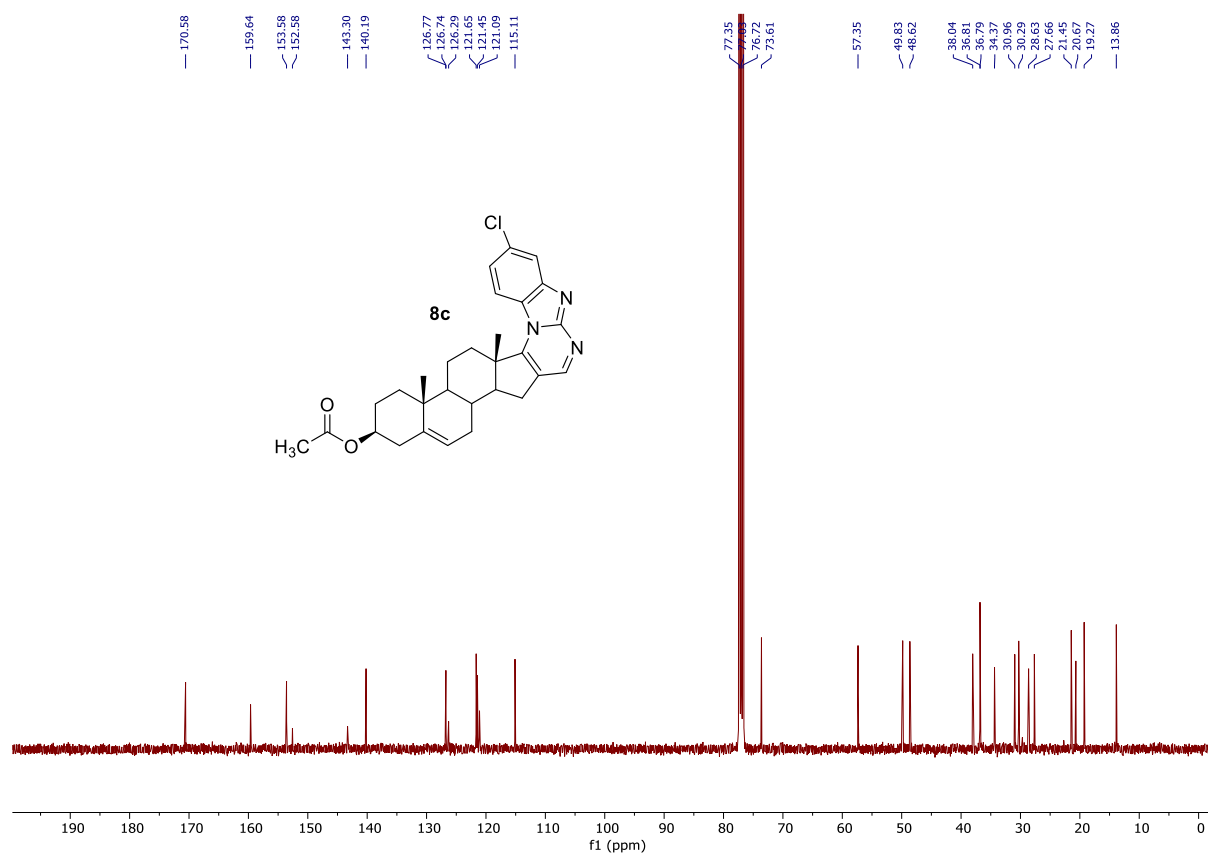


Figure S36. HRMS of **8c**

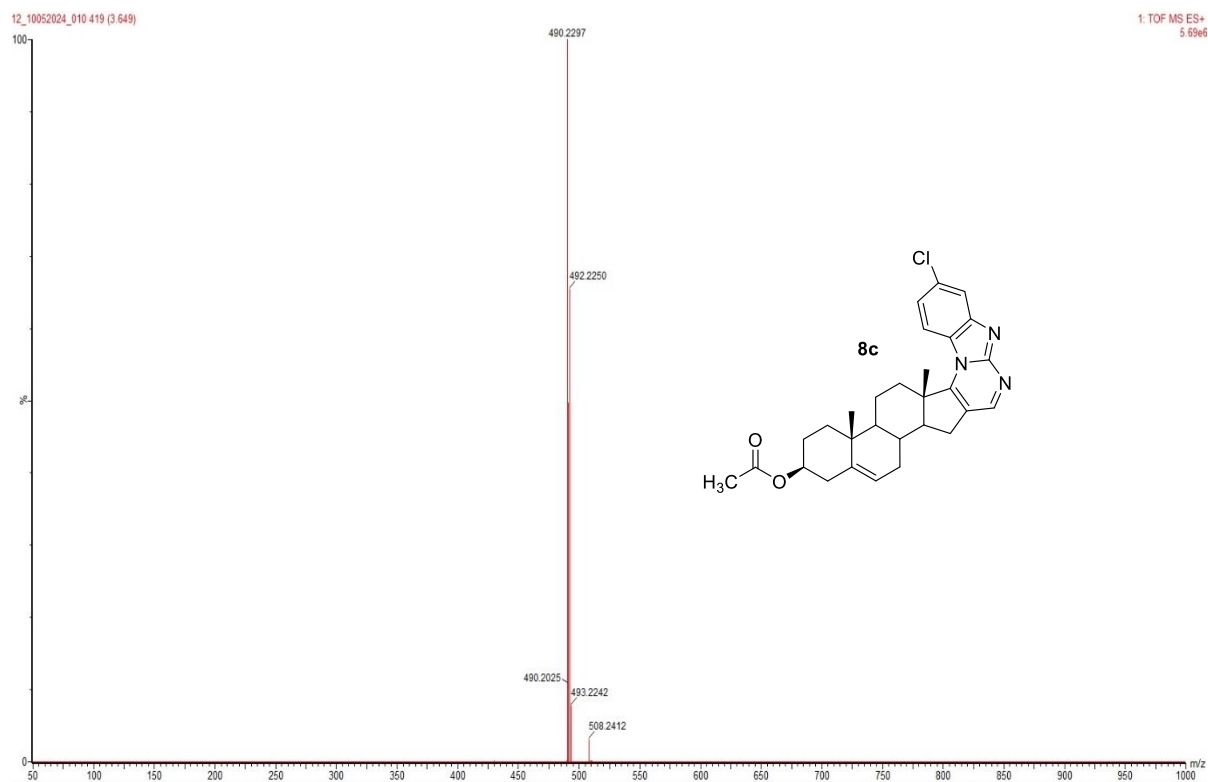


Figure S37. ^1H NMR (400 MHz, CDCl_3) of **8d**

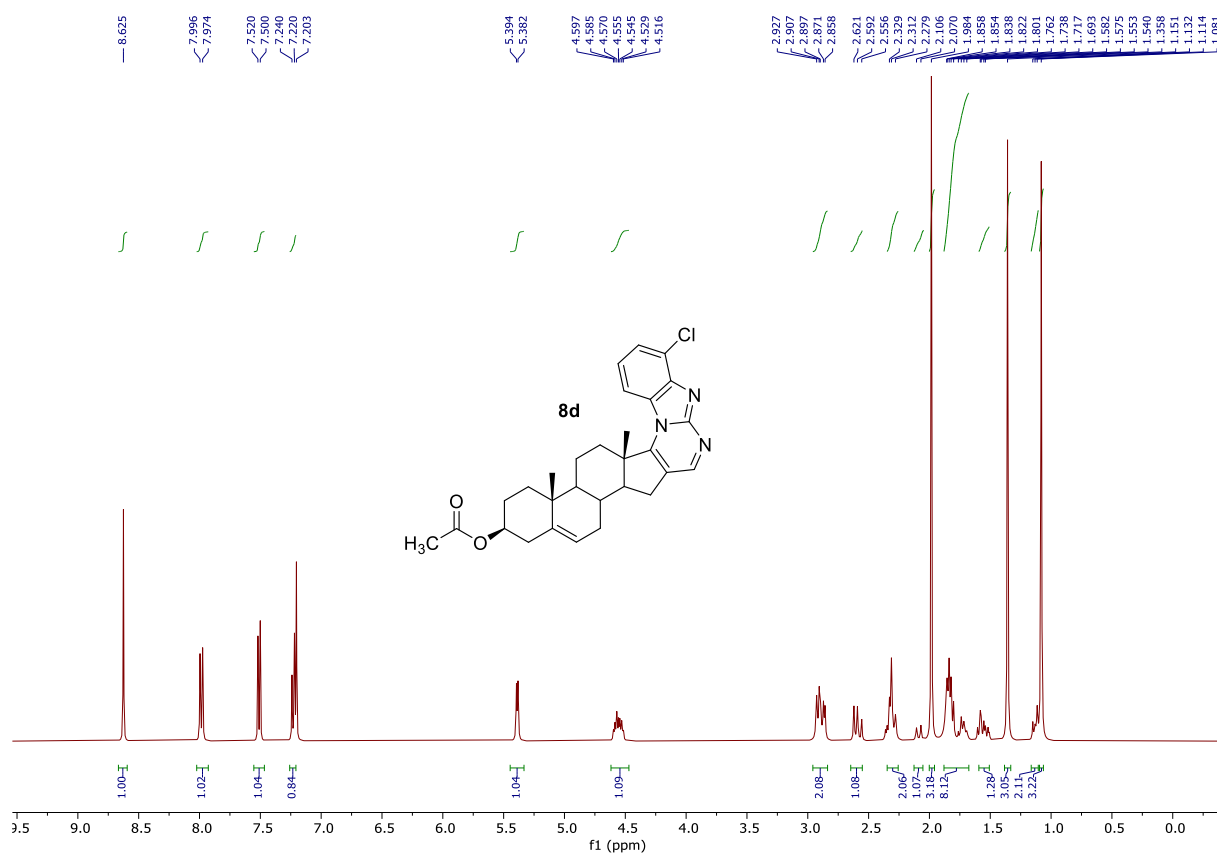


Figure S38. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) of **8d**

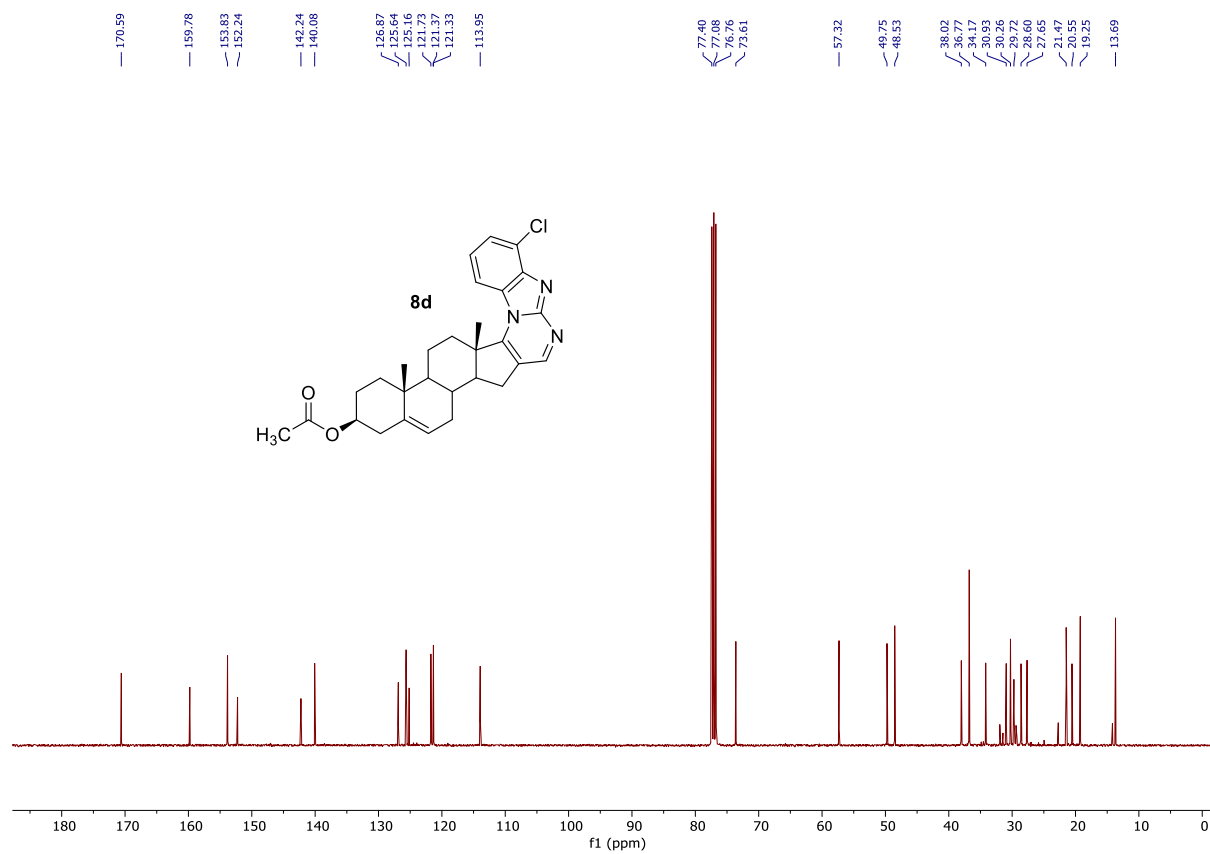


Figure S39. HRMS of 8d

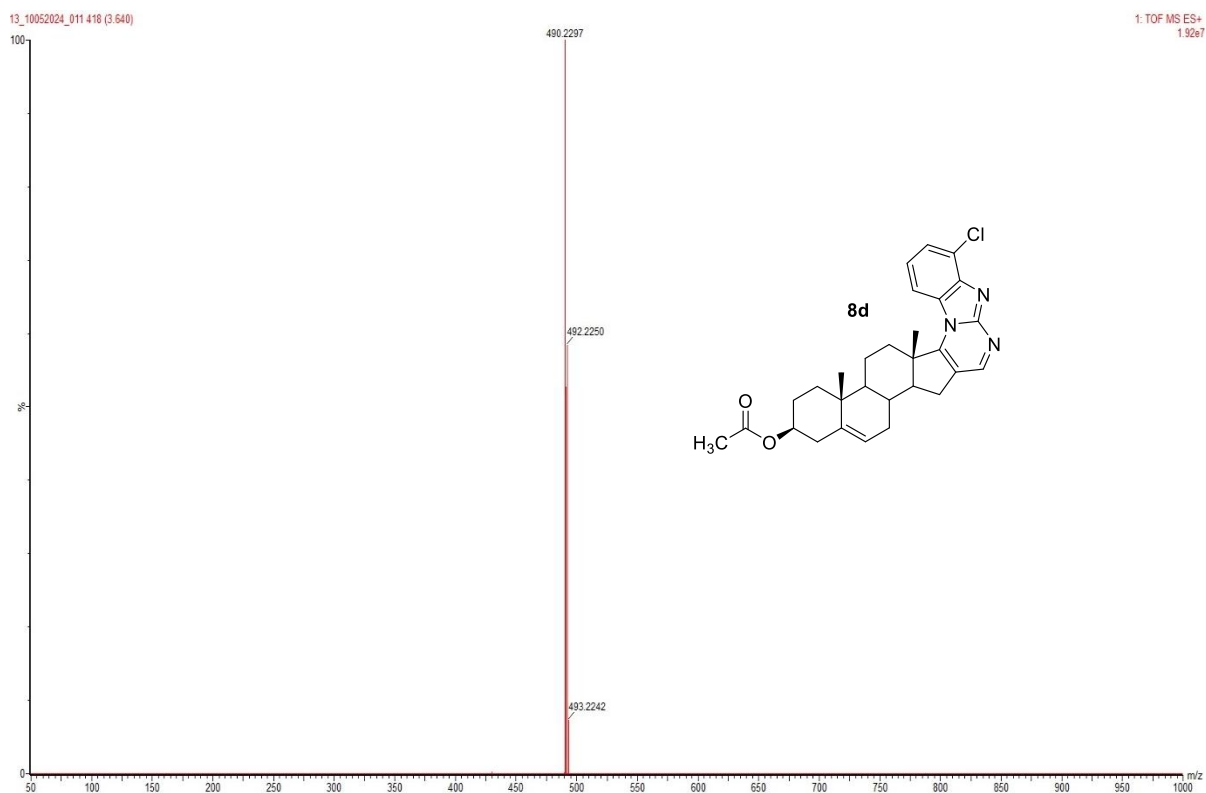


Figure S40. ¹HNMR (400 MHz, CDCl₃) of 9a

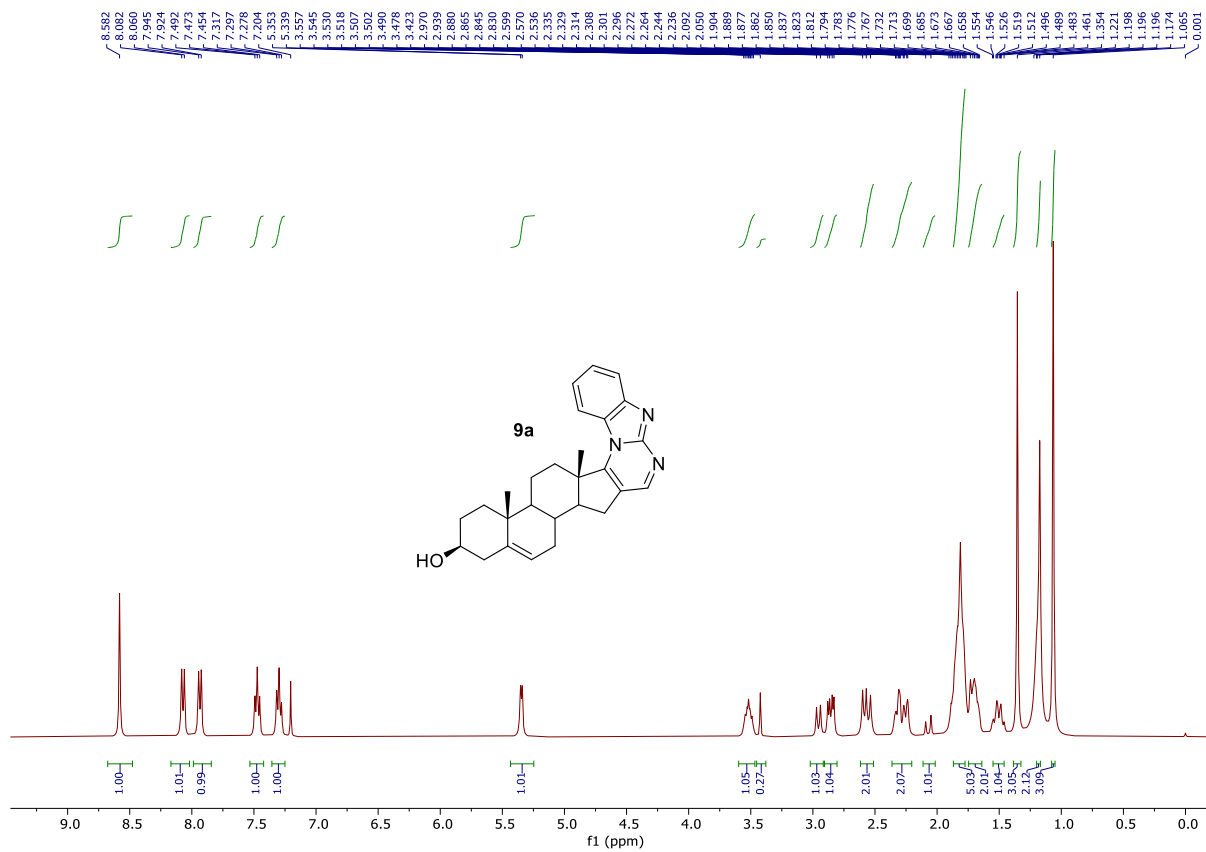


Figure S41. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) of 9a

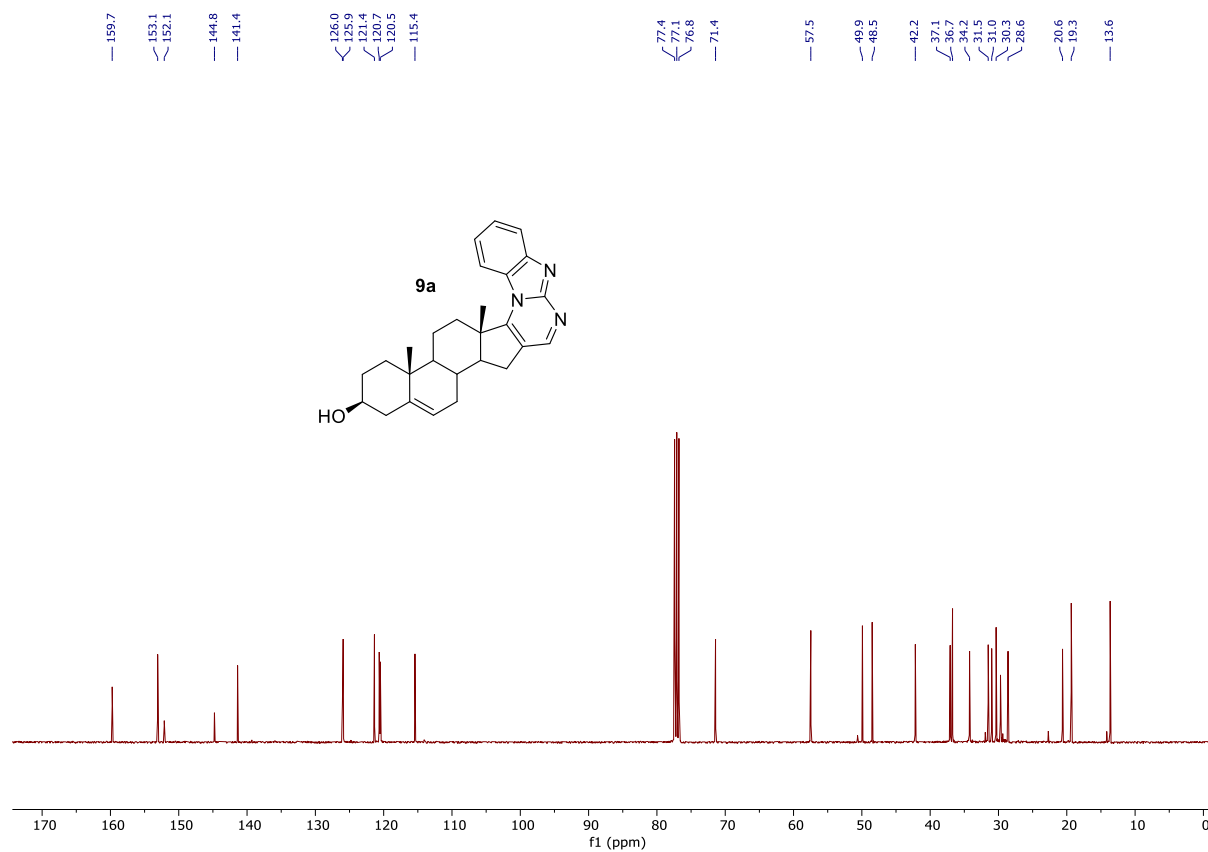


Figure S42. HRMS of 9a

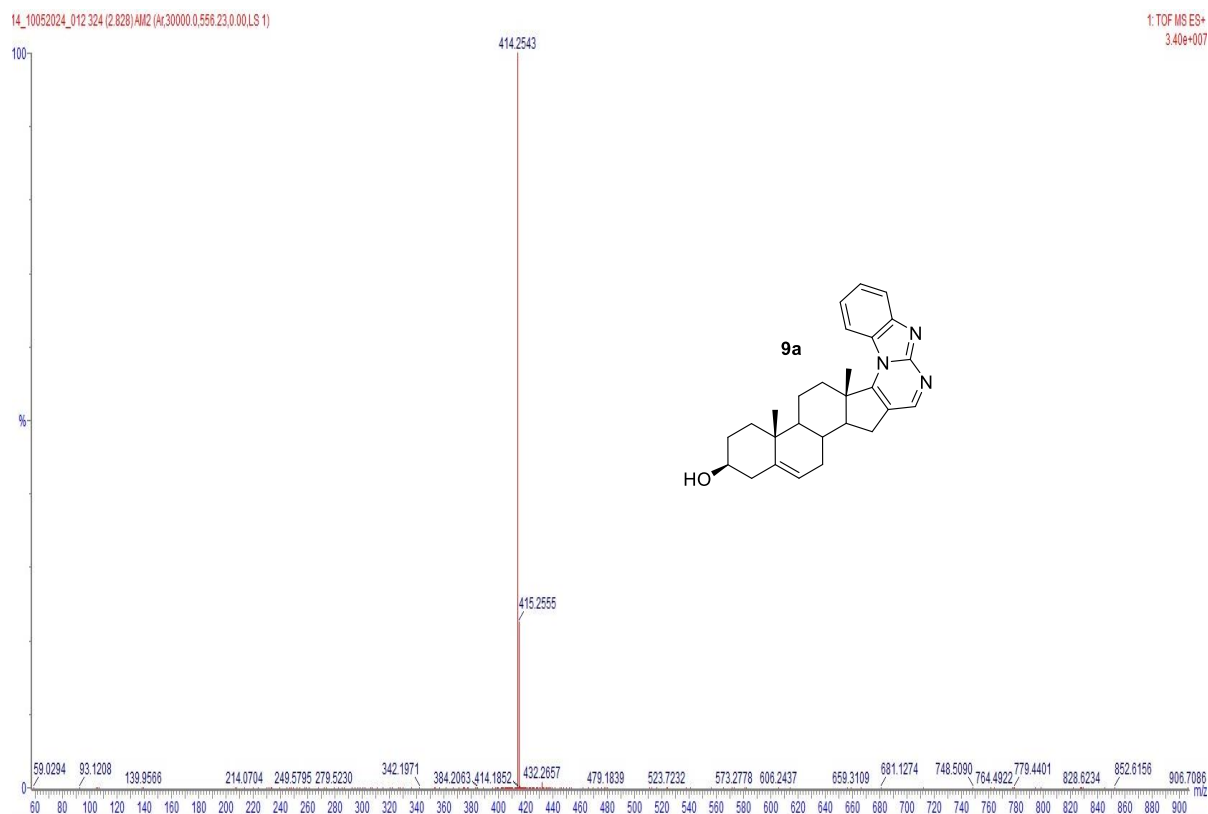


Figure S43. ^1H NMR (400 MHz, $\text{DMSO-}d_6$) of 9b

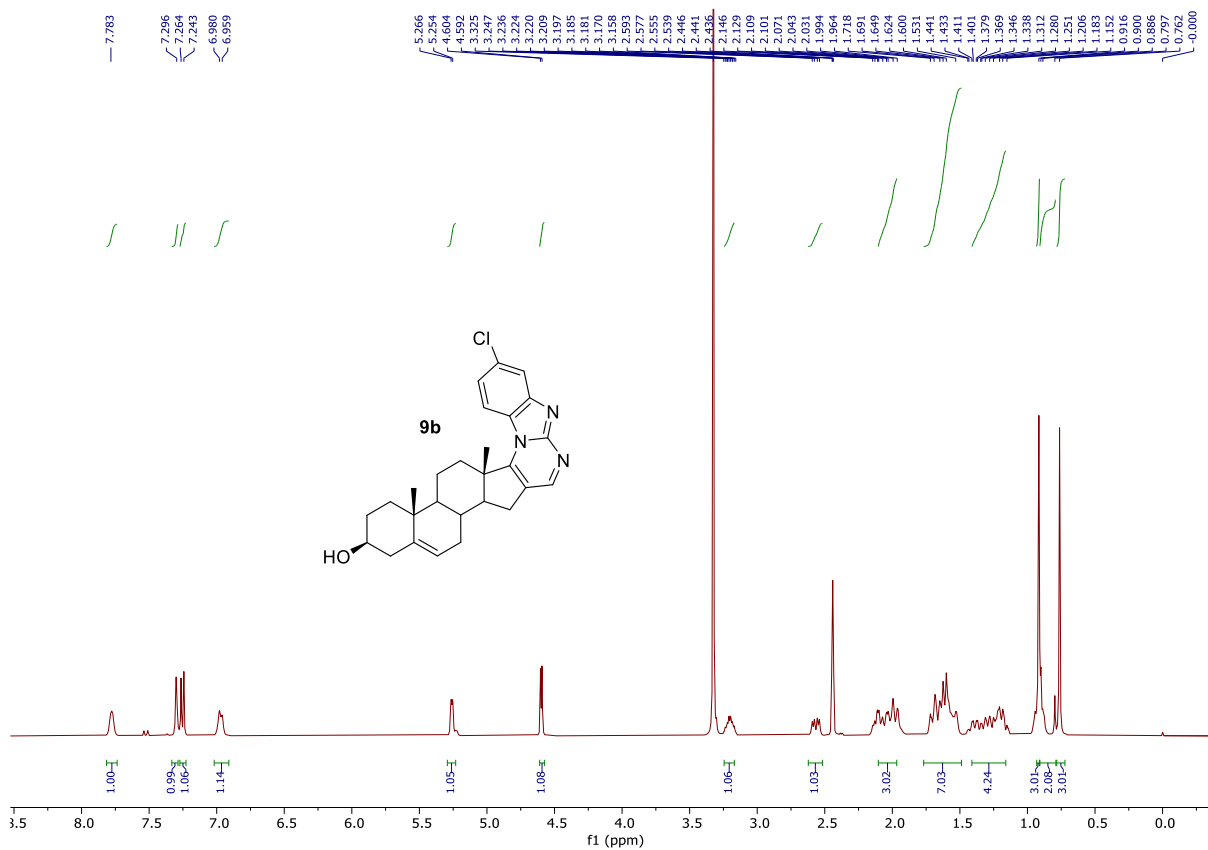


Figure S44. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, $\text{DMSO-}d_6$) of 9b

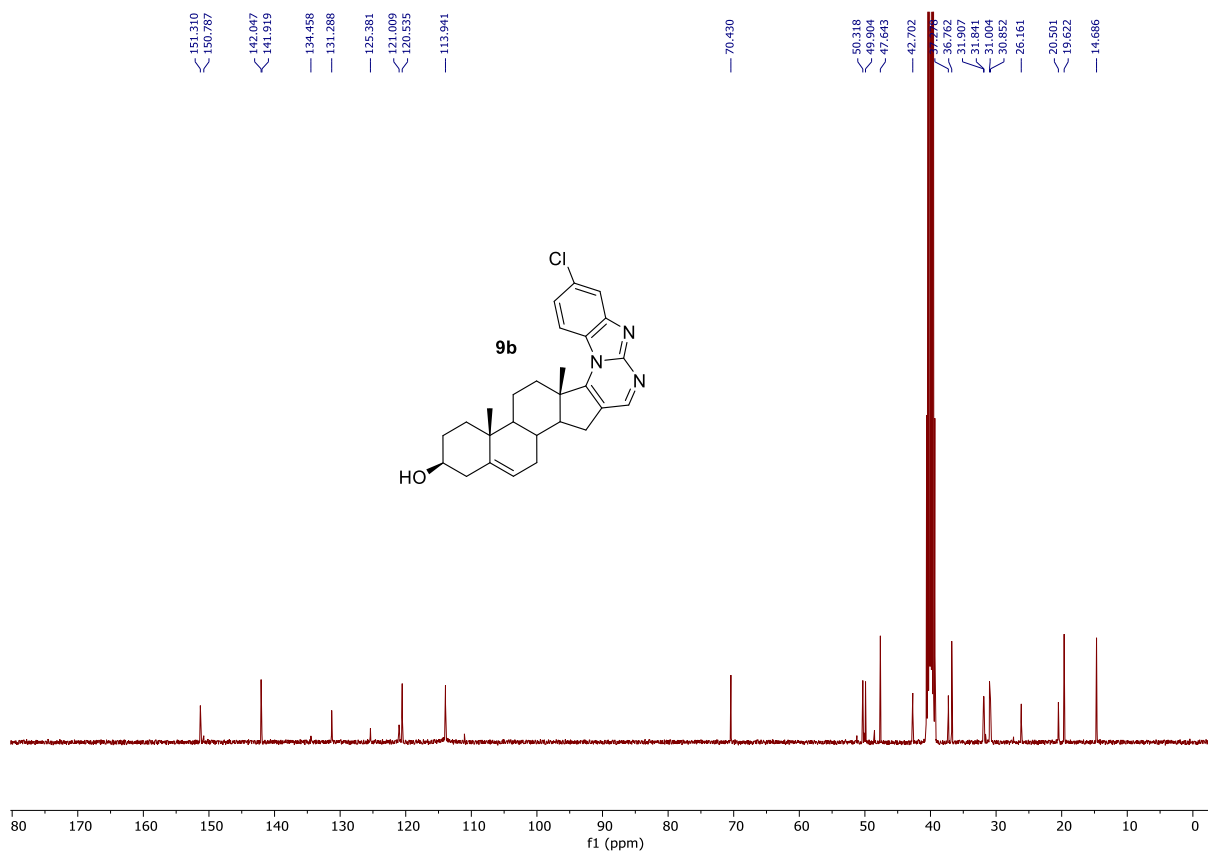


Figure S45. HRMS of 9b

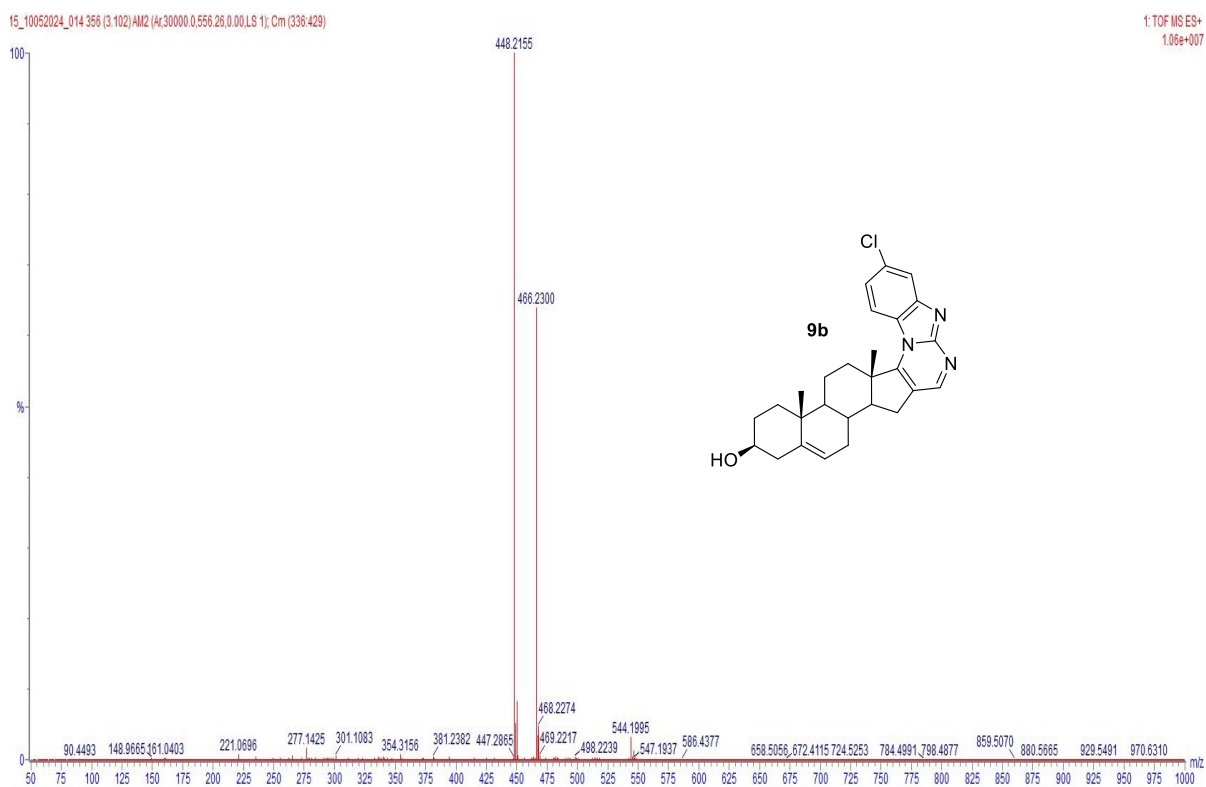


Figure S46. ¹HNMR (400 MHz, CDCl₃) of 9c

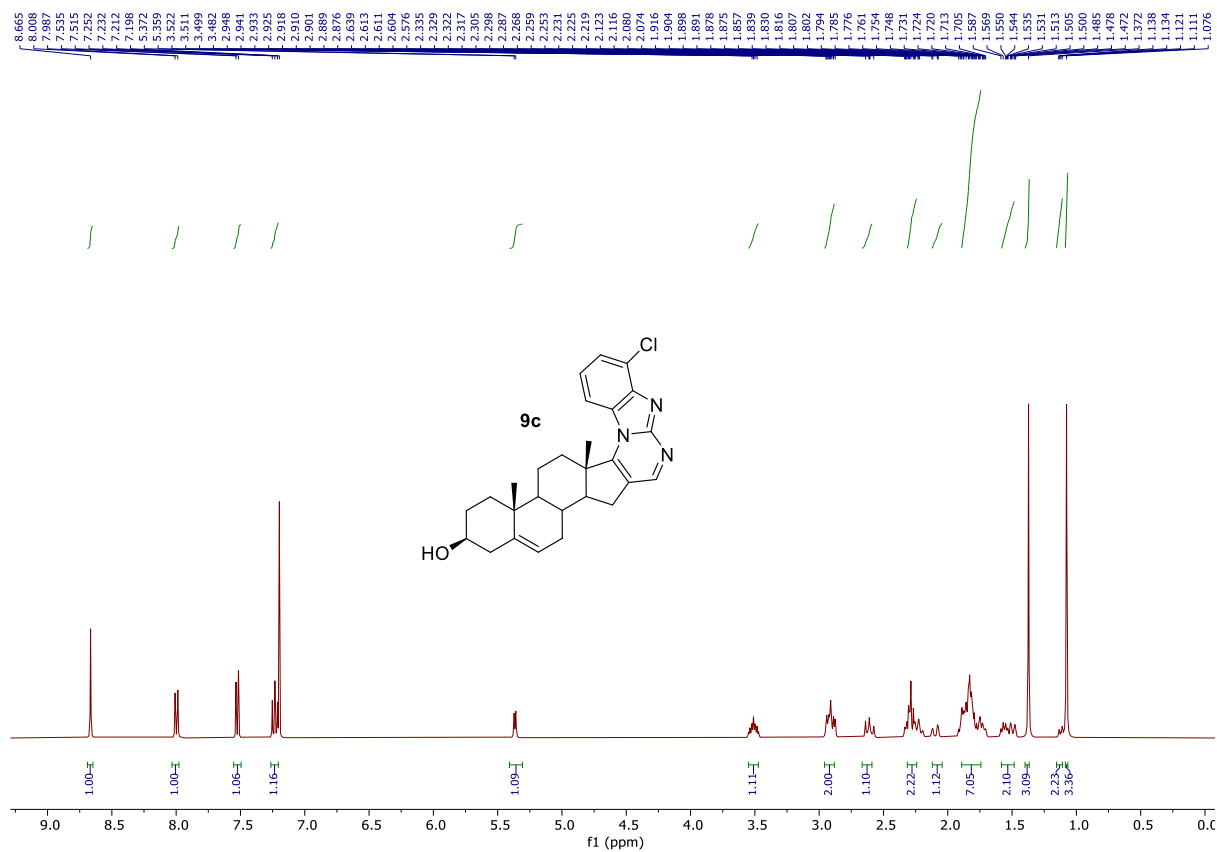


Figure S47. $^{13}\text{C}\{^1\text{H}\}$ NMR (100 MHz, CDCl_3) of **9c**

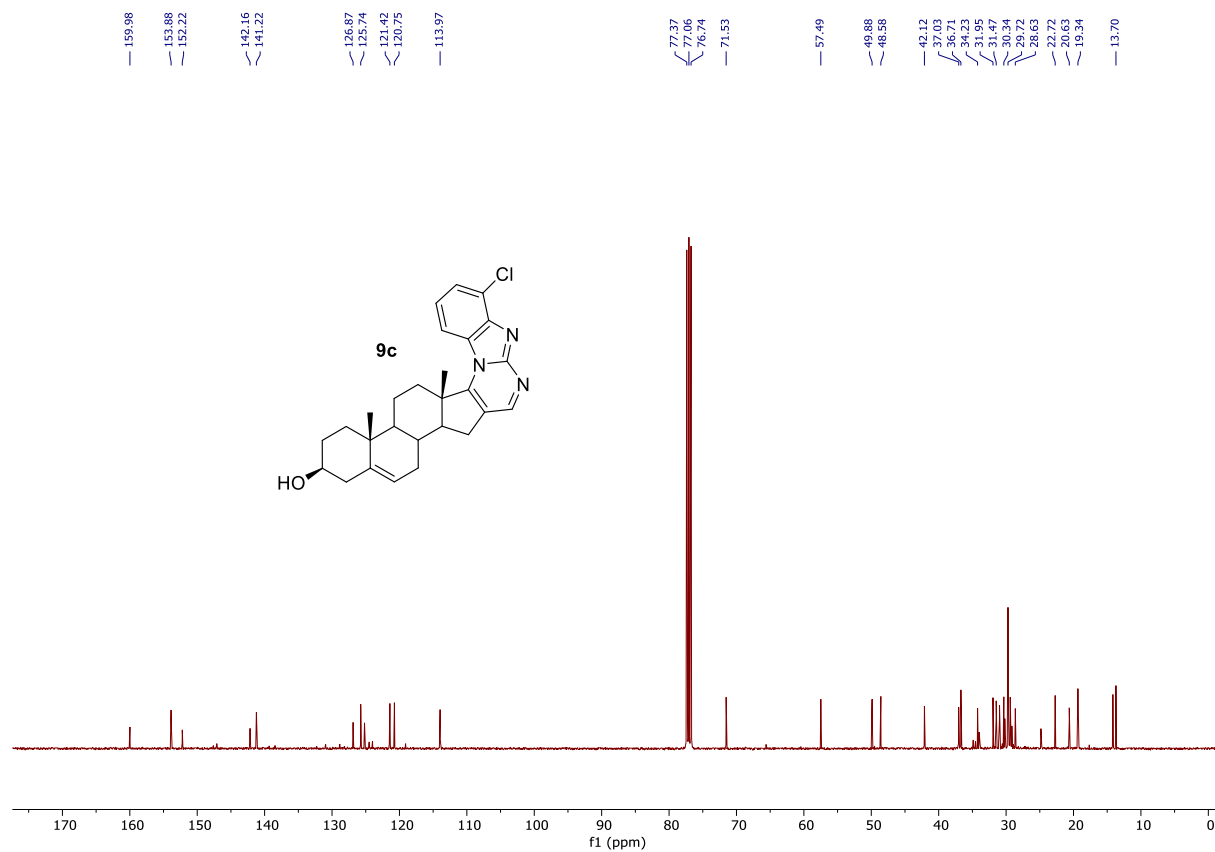


Figure S48. HRMS of **9c**

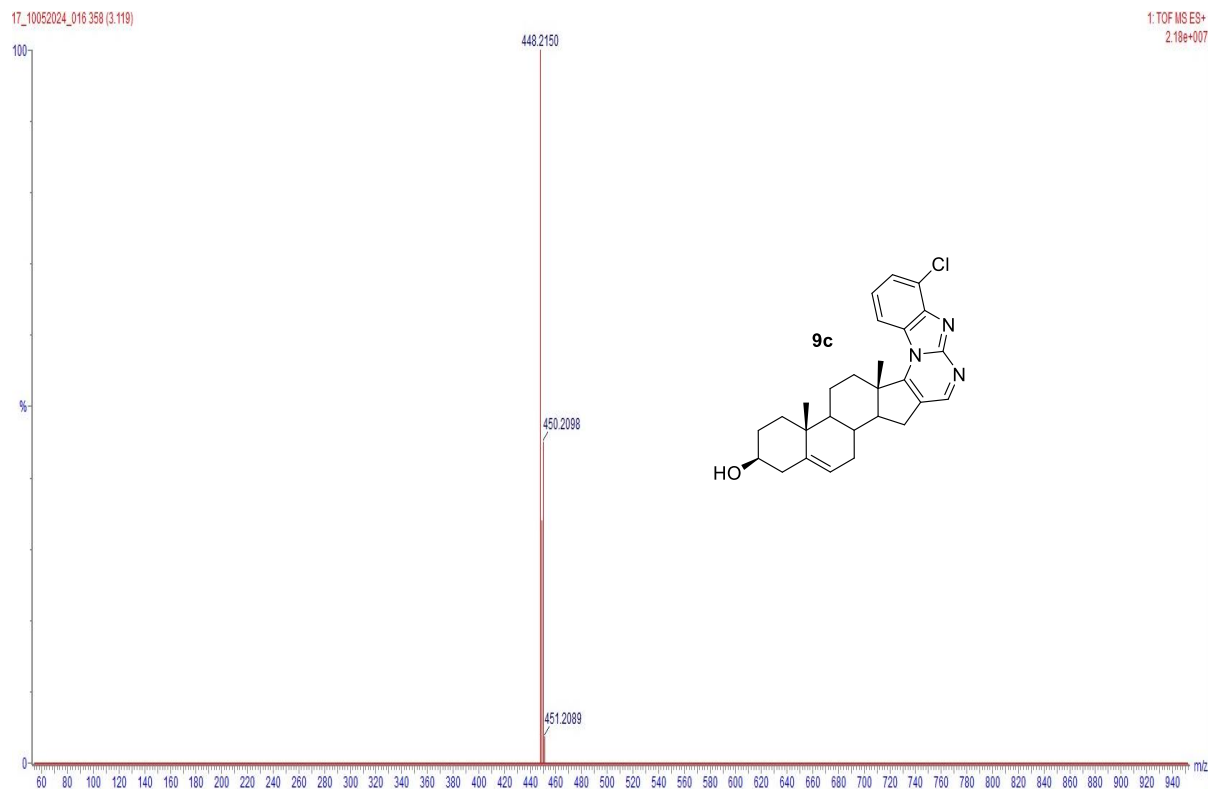


Table S2. Calculated Physicochemical property of **9a**

Property	Value	Comment
Molecular Weight	427.36	Contain hydrogen atoms. Optimal:100~600
Volume	457.433	van der Waals volume
Density	0.934	Density = MW / Volume
nHA	4.0	Number of hydrogen bond acceptors. Optimal:0~12
nHD	3.0	Number of hydrogen bond donors. Optimal:0~7
nRot	0.0	Number of rotatable bonds. Optimal:0~11
nRing	1.0	Number of rings. Optimal:0~6
MaxRing	28.0	Number of atoms in the biggest ring. Optimal:0~18
nHet	4.0	Number of heteroatoms. Optimal:1~15
fChar	0.0	Formal charge. Optimal:-4 ~4
nRig	34.0	Number of rigid bonds. Optimal:0~30
Flexibility	0.0	Flexibility = nRot /nRig
Stereo Centers	12.0	Stereo Centers. Optimal: ≤ 2
TPSA	47.53	Topological Polar Surface Area. Optimal:0~140
logS	-4.69	The logarithm of aqueous solubility value.
logP	3.856	The logarithm of the n-octanol/water distribution coefficients at pH = 7.4.
logD	3.75	The logarithm of the n-octanol/water distribution coefficient.
pKa (Acid)	10.817	Acid-base dissociation constant (pKa) value represents the strength of a drug molecule's acidity or basicity.
pKa (Base)	8.801	Acid-base dissociation constant (pKa) value represents the strength of a drug molecule's acidity or basicity.
Melting point	258.232	The predicted melting point of a compound is expressed in degrees Celsius (°C). Melting points below 25 °C are classified as liquids, while melting points above 25 °C are classified as solids.
Boiling point	335.774	The predicted melting point of a compound is expressed in degrees Celsius (°C). A normal boiling point below 25 °C is categorized as a gas.

Table S3. Calculated Medicinal Chemistry Parameters of **9a**

Property	Value	Decision	Comment
QED	0.545	●	<ul style="list-style-type: none"> ■ A measure of drug-likeness based on the concept of desirability; ■ Attractive: > 0.67; ■ unattractive: 0.49~0.67; ■ too complex: < 0.34
GASA	1.0	●	<ul style="list-style-type: none"> ■ ES: Easy to synthesize; HS: Hard to synthesize; ■ The output value represents the probability of being difficult to synthesize, ranging from 0 to 1.

Synth	5.0	●	<ul style="list-style-type: none"> ■ Synthetic accessibility score is designed to estimate ease of synthesis of drug-like molecules. ■ SAscore ≥ 6, difficult to synthesize; SAscore < 6, easy to synthesize
Fsp3	1.0	●	<ul style="list-style-type: none"> ■ The number of sp³ hybridized carbons / total carbon count, correlating with melting point and solubility. ■ Fsp³ ≥ 0.42 is considered a suitable value.
MCE-18	111.185	●	<ul style="list-style-type: none"> ■ MCE-18 stands for medicinal chemistry evolution. ■ MCE-18 ≥ 45 is considered a suitable value.
NPscore	1.844	-	<ul style="list-style-type: none"> ■ Natural product-likeness score. ■ This score is typically in the range from -5 to 5. ■ The higher the score is, the higher the probability is that the molecule is a NP.
Lipinski Rule	0.0	●	<ul style="list-style-type: none"> ■ MW ≤ 500; logP ≤ 5; Hacc ≤ 10; Hdon ≤ 5 ■ If two properties are out of range, a poor absorption or permeability is possible, one is acceptable.
Pfizer Rule	1.0	●	<ul style="list-style-type: none"> ■ logP > 3; TPSA < 75 ■ Compounds with a high log P (> 3) and low TPSA (< 75) are likely to be toxic.
GSK Rule	1.0	●	<ul style="list-style-type: none"> ■ MW ≤ 400; logP ≤ 4 ■ Compounds satisfying the GSK rule may have a more favorable ADMET profile
Golden Triangle	0.0	●	<ul style="list-style-type: none"> ■ $200 \leq MW \leq 500$; $-2 \leq \log D \leq 5$ ■ Compounds satisfying the Golden Triangle rule may have a more favorable ADMET profile.
PAINS	0 alerts	-	frequent hitters, Alpha-screen artifacts and reactive compound 480 substructures (J Med Chem 201053: 2719-40)
ALARM NMR	0 alerts	-	Thiol reactive compounds.
BMS	0 alerts	-	undesirable, reactive compounds 176 substructures (J Chem Inf Model 200646:1060-8)
Chelator Rule	0 alerts	-	Chelating compounds.
Colloidal aggregators	0.707	-	<ul style="list-style-type: none"> ■ Category 0: non-colloidal aggregators; ■ Category 1: colloidal aggregators. ■ The output value is the probability of being colloidal aggregators, within the range of 0 to 1.
FLuc inhibitors	0.0	●	<ul style="list-style-type: none"> ■ Category 0: non-fLuc inhibitors; ■ Category 1: fLuc inhibitors. ■ The output value is the probability of being fLuc inhibitors, within the range of 0 to 1.
Blue fluorescence	0.04	●	<ul style="list-style-type: none"> ■ Category 0: non-blue fluorescence; ■ Category 1: blue fluorescence. ■ The output value is the probability of being blue fluorescence, within the range of 0 to 1.
Green fluorescence	0.0	●	<ul style="list-style-type: none"> ■ Category 0: non-green fluorescence; ■ Category 1: green fluorescence. ■ The output value is the probability of being green fluorescence, within the range of 0 to 1.

Reactive compounds	0.002	●	<ul style="list-style-type: none"> ■ Category 0: non-reactive compound; ■ Category 1: reactive compound. ■ The output value is the probability of being reactive compound, within the range of 0 to 1.
Promiscuous compounds	0.162	●	<ul style="list-style-type: none"> ■ Category 0: non-promiscuous compound; ■ Category 1: promiscuous compound. ■ The output value is the probability of being promiscuous compound, within the range of 0 to 1.

Table S4. Calculated Absorption table of **9a**

Property	Value	Decision	Comment
Caco-2 Permeability	-5.337	●	Optimal: higher than -5.15 Log unit
MDCK Permeability	-5.225	●	<ul style="list-style-type: none"> ■ low permeability: $< 2 \times 10^{-6}$ cm/s ■ medium permeability: $2-20 \times 10^{-6}$ cm/s ■ high passive permeability: $> 20 \times 10^{-6}$ cm/s
PAMPA	0.005	●	<ul style="list-style-type: none"> ■ The experimental data for Peff was logarithmically transformed (logPeff). ■ Molecules with log Peff values below 2.0 were classified as low-permeability (Category 0), while those with log Peff values exceeding 2.5 were classified as high-permeability (Category 1).
Pgp-inhibitor	0.057	●	<ul style="list-style-type: none"> ■ Category 1: Inhibitor; ■ Category 0: Non-inhibitor; ■ The output value is the probability of being Pgp-inhibitor
Pgp-substrate	0.976	●	<ul style="list-style-type: none"> ■ Category 1: substrate; ■ Category 0: Non-substrate; ■ The output value is the probability of being Pgp-substrate
HIA	0.0	●	<ul style="list-style-type: none"> ■ Human Intestinal Absorption ■ Category 1: HIA+ (HIA < 30 %); ■ Category 0: HIA- (HIA \geq 30 %); ■ The output value is the probability of being HIA+
F _{20%}	0.845	●	<ul style="list-style-type: none"> ■ 20 % Bioavailability ■ Category 1: F 20 % + (bioavailability < 20 %); ■ Category 0: F 20 % - (bioavailability \geq 20 %); ■ The output value is the probability of being F 20 % +
F _{30%}	0.978	●	<ul style="list-style-type: none"> ■ 30 % Bioavailability ■ Category 1: F 30% + (bioavailability < 30 %); ■ Category 0: F 30% - (bioavailability \geq 30 %); ■ The output value is the probability of being F 30 % +
F _{50%}	0.966	●	<ul style="list-style-type: none"> ■ 50 % Bioavailability ■ Category 1: F 50 % + (bioavailability < 50 %); ■ Category 0: F 50 % - (bioavailability \geq 50 %); ■ The output value is the probability of being F 50 % +

Table S5. Calculated Distribution table of **9a**

Property	Value	Decision	Comment
PPB	54.74	●	<ul style="list-style-type: none"> ■ Plasma Protein Binding Optimal: < 90 %. ■ Drugs with high protein-bound may have a low therapeutic index.
VDss	0.091	●	<ul style="list-style-type: none"> ■ Volume Distribution ■ Optimal: 0.04-20 L/kg
BBB	0.032	●	<ul style="list-style-type: none"> ■ Blood-Brain Barrier Penetration ■ Category 1: BBB+; Category 0: BBB-; ■ The output value is the probability of being BBB+
Fu	45.384	●	<ul style="list-style-type: none"> ■ The fraction unbound in plasmas ■ Low: < 5 %; Middle: 5~20 %; High: > 20 %
OATP1B1 inhibitor	0.217	●	<ul style="list-style-type: none"> ■ Category 0: Non-inhibitor; Category 1: inhibitor. ■ The output value is the probability of being inhibitor, within the range of 0 to 1.
OATP1B3 inhibitor	0.001	●	<ul style="list-style-type: none"> ■ Category 0: Non-inhibitor; Category 1: inhibitor. ■ The output value is the probability of being inhibitor, within the range of 0 to 1.
BCRP inhibitor	0.024	●	<ul style="list-style-type: none"> ■ Category 0: Non-inhibitor; Category 1: inhibitor. ■ The output value is the probability of being inhibitor, within the range of 0 to 1.
MRP1 inhibitor	0.993	●	<ul style="list-style-type: none"> ■ Category 0: Non-inhibitor; Category 1: inhibitor. ■ The output value is the probability of being inhibitor, within the range of 0 to 1.

Table S6. Calculated Metabolism table of **9a**

Property	Value	Decision	Comment
CYP1A2 inhibitor	0.0	●	<ul style="list-style-type: none"> ■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
CYP1A2 substrate	0.0	●	<ul style="list-style-type: none"> ■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP2C19 inhibitor	0.0	●	<ul style="list-style-type: none"> ■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
CYP2C19 substrate	0.084	●	<ul style="list-style-type: none"> ■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP2C9 inhibitor	0.0	●	<ul style="list-style-type: none"> ■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
CYP2C9 substrate	0.0	●	<ul style="list-style-type: none"> ■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP2D6 inhibitor	0.0	●	<ul style="list-style-type: none"> ■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
CYP2D6 substrate	0.001	●	<ul style="list-style-type: none"> ■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP3A4 inhibitor	0.0	●	<ul style="list-style-type: none"> ■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.

CYP3A4 substrate	0.001	●	<ul style="list-style-type: none"> ■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP2B6 inhibitor	0.428	●	<ul style="list-style-type: none"> ■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
CYP2B6 substrate	0.0	●	<ul style="list-style-type: none"> ■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP2C8 inhibitor	0.005	●	<ul style="list-style-type: none"> ■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
HLM Stability	0.053	●	<ul style="list-style-type: none"> ■ human liver microsomal (HLM) stability ■ Category 0: stable+ (HLM > 30 min); Category 1: unstable- (HLM ≤ 30 min). The output value is the probability of human liver microsomal instability, where a value closer to 1 indicates a higher likelihood of instability. The range is between 0 and 1.

Table S7. Calculated Excretion table of **9a**

Property	Value	Decision	Comment
CL _{plasma}	10.84	●	<ul style="list-style-type: none"> ■ The unit of predicted CL_{plasma} penetration is mL/min/kg. > 15 mL/min/kg: high clearance; 5-15 mL/min/kg: moderate clearance; < 5 mL/min/kg: low clearance.
T _{1/2}	1.411	●	<ul style="list-style-type: none"> ■ The unit of predicted T_{1/2} is hours. ■ ultra-short half-life drugs: 1/2 < 1 hour; short half-life drugs: T_{1/2} between 1-4 hours; intermediate short half-life drugs: T_{1/2} between 4-8 hours; long half-life drugs: T_{1/2} > 8 hours.

Table S8. Calculated Toxicity table of **9a**

Property	Value	Decision	Comment
hERG Blockers	0.174	●	<ul style="list-style-type: none"> ■ Molecules with IC₅₀ ≤ 10 μM or ≥ 50 % inhibition at 10 μM were classified as hERG+ (Category 1), ■ while molecules with IC₅₀ > 10 μM or < 50 % inhibition at 10 μM were classified as hERG - (Category 0). ■ The output value is the probability of being hERG+, within the range of 0 to 1.
hERG Blockers (10 μM)	0.353	●	<ul style="list-style-type: none"> ■ Molecules with IC₅₀ ≤ 10 μM are classified as hERG+ (Category 1), ■ and molecules with IC₅₀ > 10 μM are classified as hERG- (Category 0). ■ The output value is the probability of being hERG+, within the range of 0 to 1.
DILI	0.078	●	<ul style="list-style-type: none"> ■ Drug Induced Liver Injury. ■ Category 1: drugs with a high risk of DILI; ■ Category 0: drugs with no risk of DILI. ■ The output value is the probability of being toxic.
AMES Mutagenicity	0.307	●	<ul style="list-style-type: none"> ■ AMES Toxicity ■ Category 1: Ames positive (+); ■ Category 0: Ames negative (-); ■ The output value is the probability of being toxic.

Rat Oral Acute Toxicity	0.445	●	<ul style="list-style-type: none"> ■ Rat Oral Acute Toxicity. ■ Category 0: low-toxicity, > 500 mg/kg; ■ Category 1: high-toxicity; < 500 mg/kg. ■ The output value is the probability of being toxic, within the range of 0 to 1.
FDAMDD	0.796	●	<ul style="list-style-type: none"> ■ FDA Maximum (Recommended) Daily Dose. ■ Category 1: FDAMDD (+); ■ Category 0: FDAMDD (-); ■ The output value is the probability of being positive.
Skin Sensitization	0.911	●	<ul style="list-style-type: none"> ■ Category 1: Sensitizer; ■ Category 0: Non-sensitizer. ■ The output value is the probability of being toxic, within the range of 0 to 1.
Carcinogenicity	0.495	●	<ul style="list-style-type: none"> ■ Category 1: carcinogens; ■ Category 0: non-carcinogens; ■ The output value is the probability of being toxic.
Eye Corrosion	0.168	●	<ul style="list-style-type: none"> ■ Eye Corrosion ■ Category 1: corrosives; ■ Category 0: noncorrosives; ■ The output value is the probability of being corrosives.
Eye Irritation	0.53	●	<ul style="list-style-type: none"> ■ Eye Irritation ■ Category 1: irritants; ■ Category 0: nonirritants; ■ The output value is the probability of being irritants.
Respiratory	0.863	●	<ul style="list-style-type: none"> ■ Category 1: respiratory toxicants; ■ Category 0: non-respiratory toxicants. ■ The output value is the probability of being toxic, within the range of 0 to 1.
Human Hepatotoxicity	0.525	●	<ul style="list-style-type: none"> ■ Human Hepatotoxicity ■ Category 1: H-HT positive (+); ■ Category 0: H-HT negative (-); ■ The output value is the probability of being toxic.
Drug-induced Nephrotoxicity	0.644	●	<ul style="list-style-type: none"> ■ Category 0: non-nephrotoxic (-); ■ Category 1: nephrotoxic (+). ■ The output value is the probability of being nephrotoxic (+), within the range of 0 to 1.
Ototoxicity	0.472	●	<ul style="list-style-type: none"> ■ Category 0: non-ototoxicity (-); ■ Category 1: ototoxicity (+). ■ The output value is the probability of being ototoxicity (+), within the range of 0 to 1.
Hematotoxicity	0.346	●	<ul style="list-style-type: none"> ■ Category 0: non-hematotoxicity (-); ■ Category 1: hematotoxicity (+). ■ The output value is the probability of being hematotoxicity (+), within the range of 0 to 1.
Genotoxicity	0.082	●	<ul style="list-style-type: none"> ■ Category 0: non-Genotoxicity (-); ■ Category 1: Genotoxicity (+). ■ The output value is the probability of being ototoxicity (+), within the range of 0 to 1.
RPMI-8226 Immunitoxicity	0.058	●	<ul style="list-style-type: none"> ■ Category 0: non-cytotoxicity (-); ■ Category 1: cytotoxicity (+). ■ The output value is the probability of being ototoxicity (+), within the range of 0 to 1.

A549 Cytotoxicity	0.456	●	<ul style="list-style-type: none"> ■ Category 0: non-cytotoxicity (-); ■ Category 1: cytotoxicity (+). ■ The output value is the probability of being cytotoxicity (+), within the range of 0 to 1.
Hek293 Cytotoxicity	0.531	●	<ul style="list-style-type: none"> ■ Category 0: non-cytotoxicity (-); ■ Category 1: cytotoxicity (+). ■ The output value is the probability of being cytotoxicity (+), within the range of 0 to 1.
Drug-induced Neurotoxicity	0.158	●	<ul style="list-style-type: none"> ■ Category 0: non-neurotoxic (-); ■ Category 1: neurotoxic (+). ■ The output value is the probability of being neurotoxic (+), within the range of 0 to 1.

Table S9. Calculated Environmental toxicity table of **9a**

Property	Value	Comment
Bioconcentration Factors	1.946	<ul style="list-style-type: none"> ■ Bioconcentration factors are used for considering secondary poisoning potential and assessing risks to human health via the food chain. ■ The unit is $-\log_{10} [(mg/L) / (1000 * MW)]$
IGC ₅₀	3.466	<ul style="list-style-type: none"> ■ Tetrahymena pyriformis 50 percent growth inhibition concentration. ■ The unit is $-\log_{10} [(mg/L) / (1000 * MW)]$
LC ₅₀ FM	4.108	<ul style="list-style-type: none"> ■ 96-hour fathead minnow 50 percent lethal concentration. ■ The unit is $-\log_{10} [(mg/L) / (1000 * MW)]$
LC ₅₀ DM	4.364	<ul style="list-style-type: none"> ■ 48-hour daphnia magna 50 percent lethal concentration. ■ The unit is $-\log_{10} [(mg/L) / (1000 * MW)]$

Table S10. Calculated Tox21 pathway table of **9a**

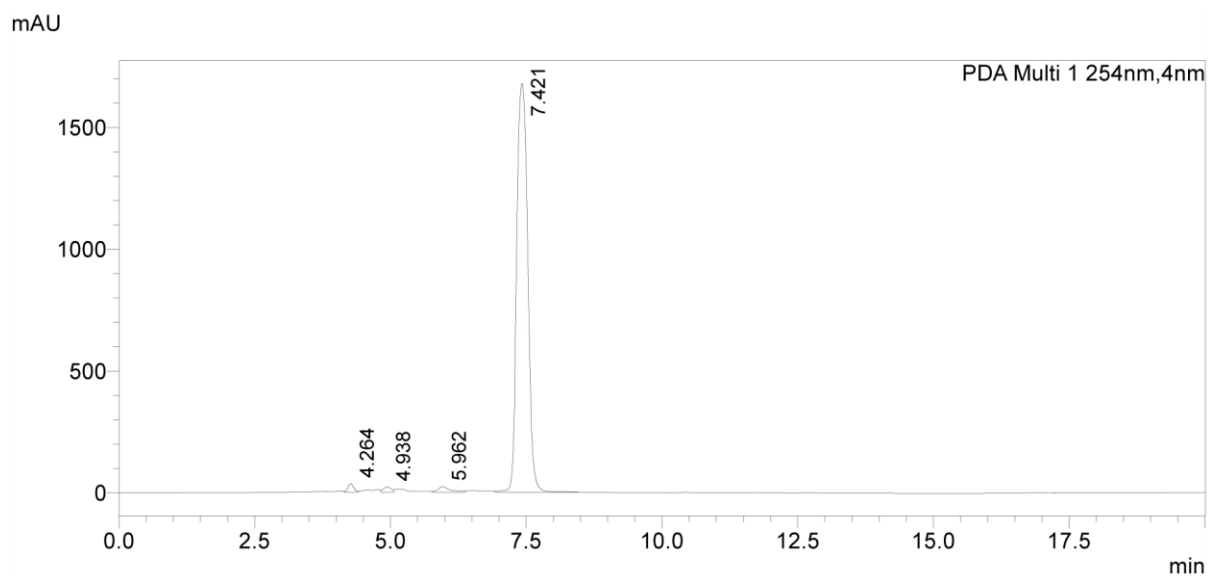
Property	Value	Decision	Comment
NR-AhR	0.0	●	<ul style="list-style-type: none"> ■ Aryl hydrocarbon receptor ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
NR-AR	0.063	●	<ul style="list-style-type: none"> ■ Androgen receptor ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
NR-AR-LBD	0.0	●	<ul style="list-style-type: none"> ■ Androgen receptor ligand-binding domain ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
NR-Aromatase	0.012	●	<ul style="list-style-type: none"> ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
NR-ER	0.165	●	<ul style="list-style-type: none"> ■ Estrogen receptor ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.

NR-ER-LBD	0.004	●	<ul style="list-style-type: none"> ■ Estrogen receptor ligand-binding domain ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
NR-PPAR-gamma	0.0	●	<ul style="list-style-type: none"> ■ Peroxisome proliferator-activated receptor gamma ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
SR-ARE	0.007	●	<ul style="list-style-type: none"> ■ Antioxidant response element ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
SR-ATAD5	0.006	●	<ul style="list-style-type: none"> ■ ATPase family AAA domain-containing protein 5 ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
SR-HSE	0.001	●	<ul style="list-style-type: none"> ■ Heat shock factor response element ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
SR-MMP	0.001	●	<ul style="list-style-type: none"> ■ Mitochondrial membrane potential ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
SR-p53	0.03	●	<ul style="list-style-type: none"> ■ p53, a tumor suppressor protein ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.

Table S11. Calculated Toxicophore Rules table of **9a**

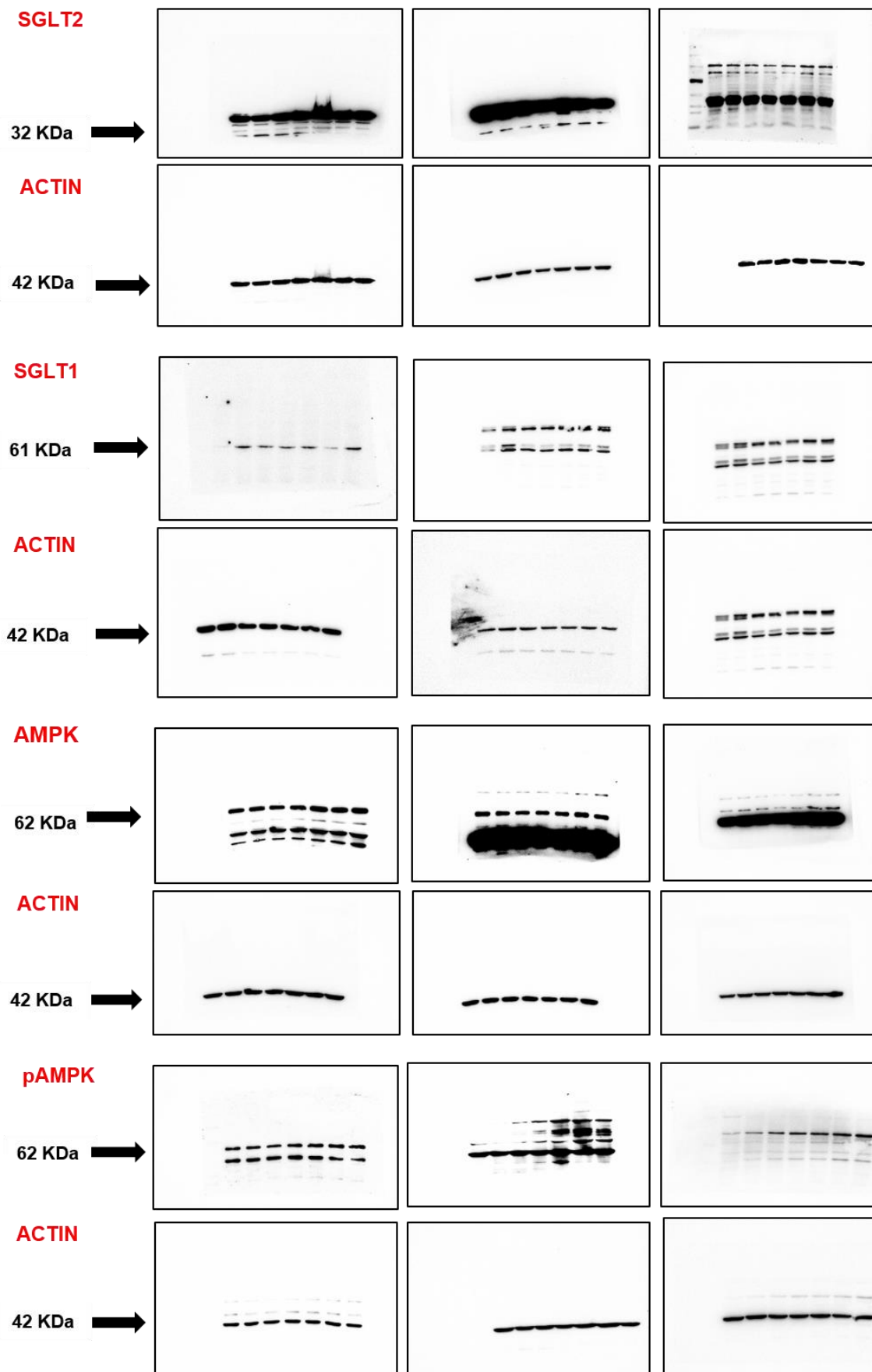
Property	Value	Comment
Acute Toxicity Rule	0	<ul style="list-style-type: none"> ■ 20 substructures; ■ acute toxicity during oral administration
Genotoxic Carcinogenicity Rule	0	<ul style="list-style-type: none"> ■ 117 substructures; ■ carcinogenicity or mutagenicity
Non Genotoxic Carcinogenicity Rule	0	<ul style="list-style-type: none"> ■ 23 substructures; ■ carcinogenicity through nongenotoxic mechanisms
Skin Sensitization Rule	1 alerts	<ul style="list-style-type: none"> ■ 155 substructures; ■ skin irritation
Aquatic Toxicity Rule	0	<ul style="list-style-type: none"> ■ 99 substructures; ■ toxicity to liquid (water)
Non-Biodegradable Rule	0	<ul style="list-style-type: none"> ■ 19 substructures; ■ non-biodegradable
SureChEMBL Rule	0	<ul style="list-style-type: none"> ■ 164 substructures; ■ MedChem unfriendly status
Toxicophores Rule	0	<ul style="list-style-type: none"> ■ 154 toxic substructures from FAF-Drug4

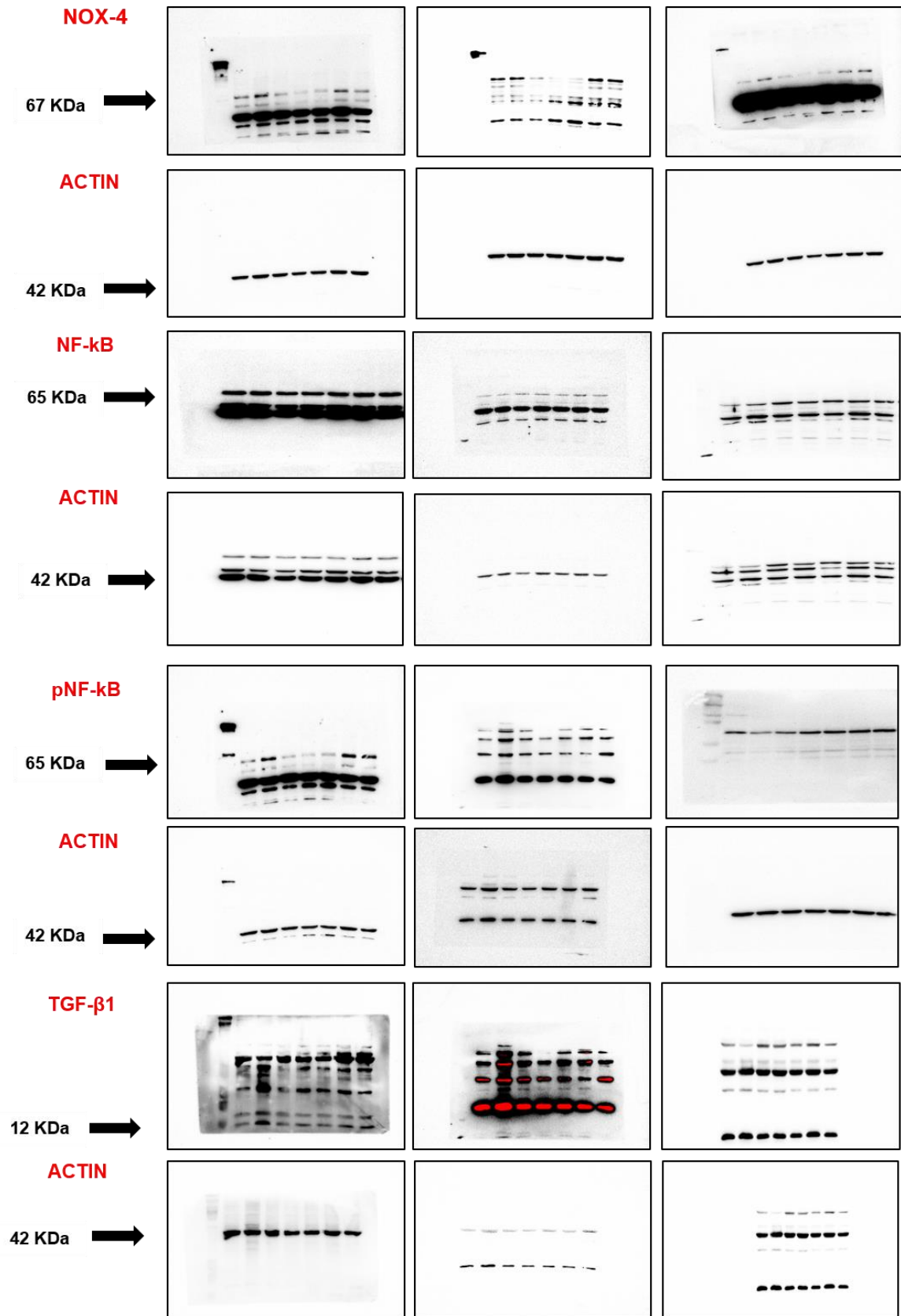
Table S12. HPLC chromatogram of **9a** with percentage purity.

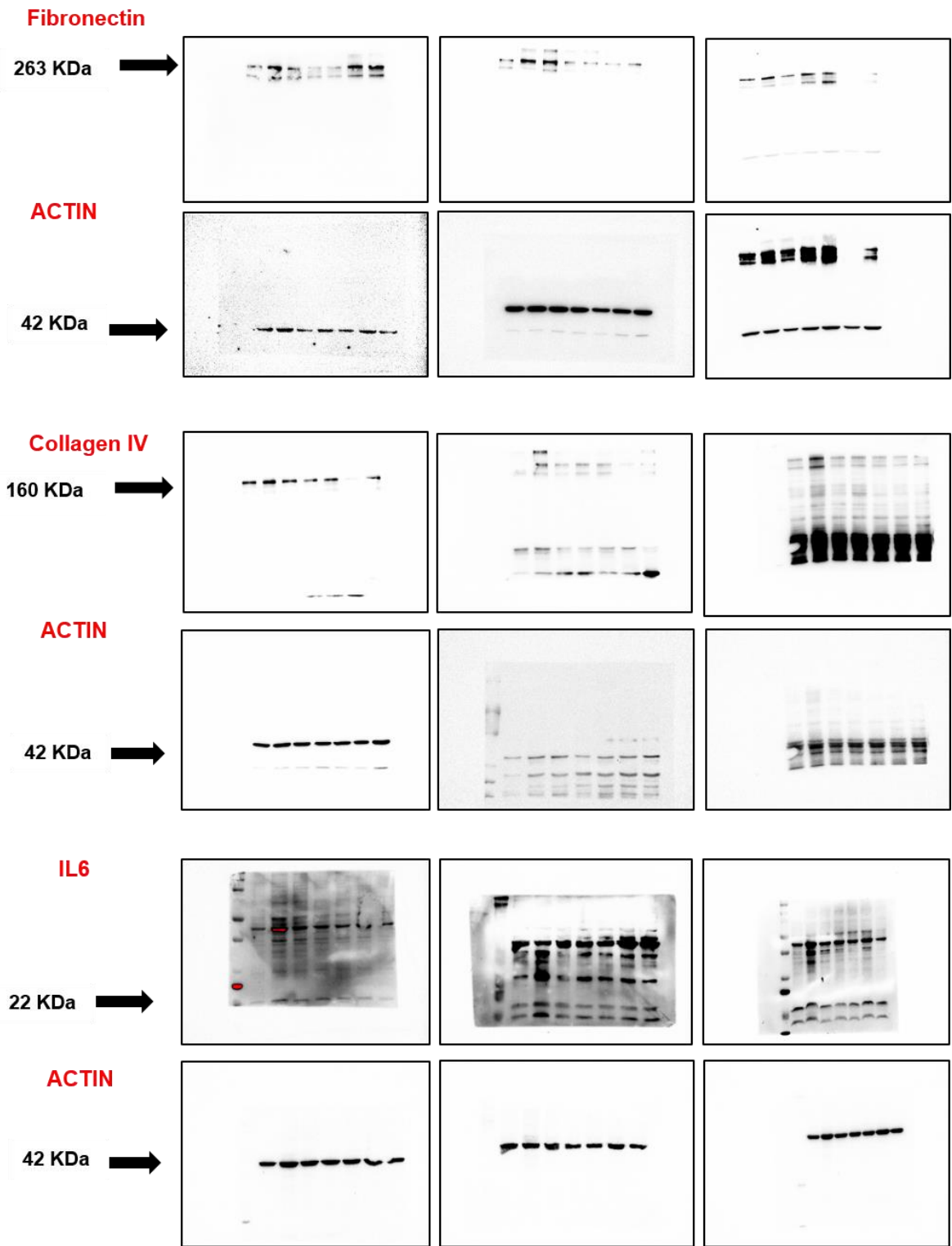


Peak#	Retention Time	Area	Height	Percentage (%)
1	4.264	264836	35258	1.073102
2	4.938	229031	22220	0.928022
3	5.962	397587	23135	1.611002
4	7.421	23788029	1678033	96.38787
Total		24679483	1758646	100

Figure S49. Full-length blots of immunoblotting data







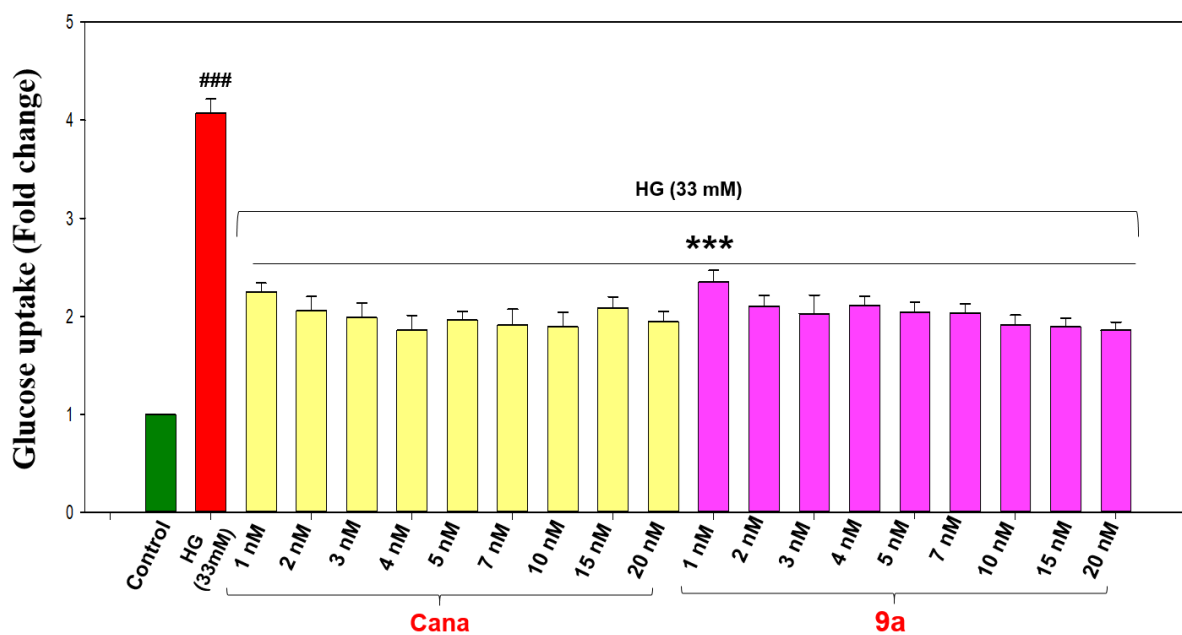


Figure S50. SGLT2 mediated glucose uptake analysis on HG-treated NRK-52E cells upon canagliflozin and **9a** treatment cells. Results are mean \pm SEM. (###) $p < 0.001$ vs control group; (***) $p < 0.001$ and (**) $p < 0.01$ vs HG group. $n = 3$ independent experiments for each group.