A Marine Viral Halogenase that Iodinates Diverse Substrates

Authors: Danai S. Gkotsi,[1] Hannes Ludewig,[1] Sunil V. Sharma,[1] Jack A. Connolly,[1] Jagwinder Dhaliwal,[1] Yunpeng Wang,[1] William P. Unsworth,[2] Richard J. K. Taylor,[2] Matthew M.W. McLachlan,[3,4] Stephen Shanahan,[3] James H. Naismith,[5] and Rebecca J. M. Goss*[1]

Affiliations:

[1] School of Chemistry, University of St Andrews, North Haugh, St Andrews, Fife, KY16 9ST, UK. Biomedical Sciences Research Complex, University of St Andrews, North Haugh, St Andrews, Fife, KY16 9ST, UK. rjmg@st-andrews.ac.uk

[2] Department of Chemistry, University of York, Heslington, York YO10 5DD, UK.

[3] Syngenta, Jealott's Hill International Research Centre, Bracknell, Berkshire, RG42 6EY, UK.

[4] QEDDI, Staff House Road, The University of Queensland, Brisbane, QLD 4072, Australia.

[5] Division of Structural Biology, Wellcome Trust Centre of Human Genomics, Roosevelt Drive, Oxford, OX3 7BN; Research Complex at Harwell, Rutherford Laboratory, Didcot, OX11 0FA; The Rosalind Franklin Institute, Didcot OX11 0FA

Contents:

I. Materials and Methods

(A) Reagents, enzymes, media components, buffers, and solvents were obtained from Sigma Aldrich, Alfa Aesar, Fisher, Formedium, Promega or Melford Scientific. Microbial culturing was performed under sterile conditions maintained using a Faster BH-EN class II vertical laminar airflow cabinet or next to a Bunsen flame. *E. coli* strains were stored at -80 °C in 20% glycerol. Sterilisation was performed by autoclaving at 121 °C for 20 min unless otherwise stated, or by passage through a 0.2 μm syringe filter. Primers were synthesised by Sigma Aldrich at 0.1 mmol scale and purified by desalting by the manufacturer. Synthetic genes were purchased from Invitrogen. DNA sequencing was performed by GATC Biotech. Melting temperatures (T_m) of primers was calculated using Thermo Fisher Multiple Primer Analyzer under default settings, considering only the annealing region of the primer. Restriction enzymes, DNA modifying enzymes, DNA polymerases, and DNA purification kits were used according to the manufacturer's instructions unless otherwise described. Standard general cloning protocols were followed or adapted from Molecular Cloning: A Laboratory Manual.

Microbial cultures were incubated in New Brunswick Scientific I26 or I26R, Innova 4300, 44 or 42, or Stuart SI500 orbital incubator shakers or a Genlab static incubator. pH measurements were taken using a Fisherbrand Hydrus 300 pH meter. 18 MΩ water was generated using an ELGA Purelab Flex system fitted with 0.2 μm point-of-use filter. Autoclaving was performed using a Boxer Benchtop Denley autoclave. Pipetting was performed using LABNET Biopette or Eppendorf Xplorer pipettes, externally calibrated biennially by Starlab. Centrifugation was carried out using a Fisher Scientific accuSpin microcentrifuge, Thermo Scientific IEX CL30R centrifuge with T41 swinging bucket rotor, or Beckman JXN-26 centriguge with JS 5.3, JA 25.50, or JLA 8.100 rotors. PCR was performed using a Biorad T100 thermocycler. UV spectra were obtained using a BMG LABTECH FLUOstar OMEGA microplate reader using clear flat-bottomed 96-well plates or a quartz L-Vis plate. Sonication was performed using a Bandelin Sonopuls instrument with KE76 and MS73 flat titanium tips.

(B) Samples were analysed by UHPLC using a Waters Acquity UHPLC system equipped with a Waters Acquity BEH C_{18} 1.7 μm 2.1 × 50 mm column at 40 °C. An injection volume of 5 μl was always used for all samples. The analytes were eluted using an initial solvent composition of 90% solvent A (0.1% TFA in water) and 10% solvent B (acetonitrile) at a flow rate of 600 μl/min that was held for 0.5 min. This was followed by a linear gradient to 95% solvent B over 1.5 min. This solvent composition was held for 0.9 min before returning to initial conditions over 0.1 min. The initial conditions were held for 1 min before the next sample was injected. The elution was monitored by UV absorption or fluorescence response tuned to appropriate wavelengths.

(C) Samples were analysed by a LC-HRMS^2 using a Thermo Orbitrap Velos Pro system, equipped with a Waters Xbridge C_{18} μm 2.1 × 100 mm column at 40 °C. An injection volume of 5 μl was used for all samples. Analytes were eluted using an initial solvent composition of 90% water & 0.1% formic acid (solvent A) and 10% acetonitrile (solvent B) at a flow rate of 0.35 ml/min. This initial solvent composition was held for 1.5 min followed by a linear gradient to 95% solvent B over 8 min. This composition was held for another 2 min before returning to initial conditions over 0.5 min. The eluted analyte was passed through a PDA detector monitoring absorbance at 220-800 nm (2 nm resolution, 10 Hz) and a valve which diverted eluted analyte to the waste. After the first min, the valve was switched to pass the eluted analyte through to the inlet valve of the H-ESI source. The H-ESI source was set to positive ionisation mode using a 300 °C heater temperature, 350 °C capillary temperature, 50 U sheath gas flow, 20 U aux gas flow, 2 U sweep gas flow, 3.5 kV ionization voltage and 50% RF lens power. The scan cycle included one high-resolution survey scan and three data-dependent fragmentation scans. The survey scan was analysed in the orbitrap FTMS analyser at a resolution of 30,000 (at 400 m/z) over a range of 100-2000 m/z, based on a background ion corresponding to the $[M+H]^+$ charge state of n-butyl-benzenesulfonamide (exact mass 214.08963) as a lock- mass for internal scan-by-scan calibration. The top three peaks from that scan, were then identified, isolated and subsequently fragmented under the CID (collision-induced dissociation) or HCD (higher energy collisional dissociation) modes at 35% normalized energy before fragments were analysed in the standard resolution ITMS analyser (CID) or high-resolution FTMS analyser (HCD).

D) Proton NMR (^{1}H), and carbon NMR (^{13}C) were recorded on either a Bruker Ascend 500 (500 MHz) or Bruker 500 UltraShield (500 MHz) spectrometer. The NMR experiments were carried out in deuterated chloroform ($CDCl_3$), deuterated DMSO (d_6-DMSO) or deuterated methanol (d_4-MeOH). The chemical shifts (δ) are quoted in parts per million (ppm). Coupling constants are reported in Hertz (Hz). Using a DEPTQ sequence or an HSQC experiment with multiplicity editing, the ^{13}C NMR signals were identified to CH_3, CH_2, CH and C. Purification and flash chromatography was carried out on a Biotage Isolera Four using reverse-phase SNAP (C18, 12 g) or silica gel (10 g) column cartridges. Thin layer chromatography (TLC) was executed using aluminium sheets of silica gel 60 F254 and was visualised under a Mineralight model UVGL-58 lamp (254 nm). The plates were developed with alkaline potassium permanganate solutions.

II. General procedures

(A) Media, buffers and stock solutions. Solid and liquid cultures of *Escherichia coli* were always grown in Lysogeny Broth (LB) containing tryptone (10 g/L), NaCl (5 g/L), yeast extract (5 g/L) and glucose (1 g/L). The antibiotics used in this work, respective to the construct were either ampicillin (stock solution of 100 mg/ml in 1000x dilution) or kanamycin (stock solution of 50 mg/ml in 1000x dilution). General cell lysis buffer contained as a final concentration NaH_2PO_4 (50 mM) and NaCl (250 mM) at pH 7.8. The elution buffer used for NiNTA contained as a final concentration NaH_2PO_4 (50 mM) and NaCl (250 mM) and imidazole (250 mM) at pH 7.8. For storage of VirX1 after buffer exchange for the removal of imidazole from the final elution step there were three buffers that were used for this work depending on the experiment. Storage buffer a (SBa) contained as a final concentration Hepes (50 mM), glycerol (10% of total volume), halogen salt (50 mM NaX, where X is either Br or I) at pH 7.4. Storage buffer b (SBb) was the same as SBa however did not contain any halogen source. Storage buffer c (SBc), which was mainly used for the scaling up of the biotransformations contained KH_2PO_4 (50 mM), no halogen source and glycerol (10% of final volume) at pH 7.8.

(B) *In silico* identification of VirX1. The FxxPxxSxG variation of the *Fx.Px.Sx.G* [where x is any amino acid and each (.) represents independently the number of x between each conserved residue and any number for example 0, 1, 2 etc] Goss & Gkotsi motif (GB1803491.8) was used together with a list of previously experimentally verified FDHs (either deposited independently or in groups) as a probe to mine uncurated sequence data on NCBI. Pattern hit initiated (PHI)-blast program was used, with the default parameters of the algorithm changed to identify variants of low similarity (increased E value to 1000, increased hitlist size to 5000, introduced low complexity filter, used Phi pattern FxxPxxSxG) against the non-redundant protein sequences as a database. Several entries were excluded from the search including actinobacteria (taxid: 201174), fungi/metazoa group (taxid: 33154) as well as proteobacteria (taxid: 1224). Within those 5000 lower similarity hits to known FDHs, several deposits were coming up from cyanophage genomes. The parameters were adjusted and included all cyanophage genomes. One of the hits included the hypothetical protein CPUG_00131 from the cyanophage *Syn10*. The amino acid sequence deposited here:

MKIESVAIVGGGSSGWMTAAALSKLCPQLEIALIEDPNIKTVGVGESTLGHFNKFLHLLD
LKDEDWMPACNATYKNSIRFTNFREGKGEVFEYPFGPSLDVSFFSQTDGINTWGKLANK
YPEDFPPETFARFVNSNTYLAEHNRLTRNKDNKIPNFNFDWDTAYHIDAELFGQYLKEKI
ALPNGVKHIQGKVTGYQKESPNNHNFKYIILDQETAIFADLYIDCTGFKSLLLGEFMGEA
FSPFSKKLANDKAMATRIPYENREEEMHNVTDCHAMKNGWVWNIPLWNRIGTGYCYS
SRFVSKDDAEAEFREHLGERGKDAKIFHIDIGHGKRTRAWVNNCVGIGLSYGFIEPLEST
GLLTTHENIENLVYLINQRDGYVTQAERDGFNYTCDHQIDSFSDFVAMHYAYSMRTDTP
YWKWCTQMCNYMPESMGPHRQKQSTWQDLSTDTIGLNTWHINHNGISFIIAGHGLRPQ
SYDKLSEVLLKRNNESDYYYEDIRKDWLKHYESMVEYVKTLPTHYEFLRDEIYGSAE

(C) Comparative analysis of VirX1 against synechococcales genomes, other cyanobacterial genomes, non-redundant protein sequences (nr) and UniprotKB/Swiss-Prot databases. VirX1 was compared against all available genomes of synechococcales (taxid: 1890424) using protein-protein blast (blastp). That was repeated using PHI-blast and the FxxPxxSxG motif. The genome of the only organism that gave a hit above the expected threshold of 10 for the E-value (WP_110984764.1 from *Acaryochloris* sp. RCC1774 with 78% query coverage and 34% identity) as well as one that were found to be below the threshold and without the signature motif FxxPxxSxG (WP_036931167.1 from Prochlorococcus sp. MIT 0702 with 79% query coverage and 22% sequence identity) as well as the 177103 bp genome of Syn10 were submitted on antiSMASH to identify potential gene clusters where any potential FDHs could have evolved from. Similarly, VirX1 was also compared against non-redundant protein sequences (all hits below 36% sequence identity) and UniprotKB/Swiss-Prot databases (RebH 98% query coverage and 31% sequence identity, PrnA 99% sequence coverage and 30% sequence identity, CmdE 54% sequence coverage and 21% sequence similarity, PrnC 19% sequence coverage and 31% sequence similarity).

(D) Branching analysis and split decomposition of VirX1 against known FDHs and other flavoenzymes. The following amino acid sequences of VirX1 as well as those of PrnA, RebH, PrnC, PyrH, KtzR, KtzQ, Bmp5, Bmp2, KrmI, ThaI, PltA, HalB, HrmQ, CndH, Clohal, Rdc2, Asm12, ArmH1, MibH, VhaA, ThdH, SttH, ChlA, StaI, AclH, HalX, AoiQ, TiaM, Tcp2, BhaA, Th-Hal, AcOTAhal, HalY and BrvH and PltM including as a relevant outgroup several other flavoenzymes were used such as flavocytochrome c3 (FlavoC3), D-amino acid oxidase (DAAO), Vanillyl-alcohol oxidase (VAO), 4-Hydroxyphenylacetate 3-hydroxylase (HpaB), Tryptophan 2-monooxygenase (W2M), Styrene monooxygenase (StyA), Phenylacetone monooxygenase/Baeyer-Villiger monooxygenase (BVMO), Cyclohexanone monooxygenase (CHexM), Kynurenine 3-monooxynase (KMO), Dimethylaniline monooxygenase (DMAM) were selected. The amino acid sequences (fasta's included bellow) were submitted on CLC sequence viewer adjusting the parameters to very accurate alignment (with gap open cost 10.0 and gap

extension 1.0). The alignment was then used for branching analysis (Neighbour Joining used as tree constructing method, using both Jukes-Cantor and Kimura protein distance measurements) with 100 replicates of bootstrapping analysis. The same alignment was also submitted to splitstree software for split decomposition analysis.

>KMO_HS
MDSSVIQRKKVAVIGGGLVGSLQACFLAKRNFQIDVYEAREDTRVATFTRGRSINLALS
HRGRQALKAVGLEDQIVSQGIPMRARMIHSLSGKKSAIPYGTKSQYILSVSRENLNKDLL
TAAEKYPNVKMHFNHRLLKCNPEEGMITVLGSDKVPKDVTCDLIVGCDGAYSTVRSHL
MKKPRFDYSQQYIPHGYMELTIPPKNGDYAMEPNYLHIWPRNTFMMIALPNMNKSFTC
TLFMPFEEFEKLLTSNDVVDFFQKYFPDAIPLIGEKLLVQDFFLLPAQPMISVKCSSFHFK
SHCVLLGDAAHAIVPFFGQGMNAGFEDCLVFDELMDKFSNDLSLCLPVFSRLRIPDDHAI
SDLSMYNYIEMRAHVNSSWFIFQKNMERFLHAIMPSTFIPLYTMVTFSRIRYHEAVQRW
HWQKKVINKGLFFLGSLIAISSTYLLIHYMSPRSFLRLRRPWNWIAHFRNTTCFPAKAVD
SLEQISNLISR

>Clohal
MSENQEYDVIVIGGGPGGSMVSSLLADGGKKVLVLEVAKFPRYHIGESLLLGTVDLLDK
IGVREKLEAGDYIKKYGVEWVWGEQREPWKVDFRNAVSVTHDYTYQVERGPFDKMLL
DNAREHGVDVREQHRVTNFSIDDNGRSTVNYRRTDTGETGTATARWLVDASGQGGLV
TKRLHTQEWDPYLKNMAVWSYWKGVKRSEGEDAGNIFLPTFDDGWWWCIPLRDDITS
IGAVVDRESLNALKSTRVRQYYLDSIEKTPELASRTENAELVDDMHVQRDWSYIYDRFC
GDGYIAIGDAACFIDPLFSTGVHLAMLSGFLAATVVNTILDKPELDTARMLQFYEGAYR
KEFARLRDQVYFLYAGNKGSKESYFWHARSHFGVPSIEPEKAFVSLIAGAFAHRAWYN
HYAKHLDVSADLKKSVESMFNGAGIDLDTPLIASACGSVVNDFAVDGRYLREARSLVH
TNGASIAYTPAVRAAVEHADGKRTGREIIEIITDRGIDSEDRARSIVHEVISYGFVEPKA

>CndH
MSTRPEVFDLIVIGGGPGGSTLASFVAMRGHRVLLLEREAFPRHQIGESLLPATVHGICA
MLGLTDEMKRAGFPIKRGGTFRWGKEPEPWTFGFTRHPDDPYGFAYQVERARFDDMLL
RNSERKGVDVRERHEVIDVLFEGERAVGVRYRNTEGVELMAHARFIVDASGNRTRVSQ
AVGERVYSRFFQNVALYGYFENGKRLPAPRQGNILSAAFQDGWFWYIPLSDTLTSVGA
VVSREAAEAIKDGHEAALLRYIDRCPIIKEYLAPATRVTTGDYGEIRIRKDYSYCNTSFW
KNGMALVGDAACFVDPVFSSGVHLATYSALLVARAINTCLAGEMSEQRCFEEFERRYR
REYGNFYQFLVAFYDMNQDTDSYFWSARKIINTEERANEAFVRLIAGRSNLDEPVFQSV
AKDFFTEREGFGAWFGGLVTSMAKGDGGGLMVGEGATDATESTGFAPENFMQGFTREI
TELQHLAMFGEDRGPETPLWSGGLVPSRDGLAWAVESGEDAAG

>PrnC
MTQKSPANEHDSNHFDVIILGSGMSGTQMGAILAKQQFRVLIIEESSHPRFTIGESSIPETS
LMNRIIADRYGIPELDHITSFYSTQRYVASSTGIKRNFGFVFHKPGQEHDPKEFTQCVIPEL
PWGPESHYYRQDVDAYLLQAAIKYGCKVHQKTTVTEYHADKDGVAVTTAQGERFTGR
YMIDCGGPRAPLATKFKLREEPCRFKTHSRSLYTHMLGVKPFDDIFKVKGQRWRWHEG
TLHHMFEGGWLWVIPFNNHPRSTNNLVSVGLQLDPRVYPKTDISAQQEFDEFLARFPSIG
AQFRDAVPVRDWVKTDRLQFSSNACVGDRYCLMLHANGFIDPLFSRGLENTAVTIHAL
AARLIKALRDDDFSPERFEYIERLQQKLLDHNDDFVSCCYTAFSDFRLWDAFHRLWAVG

TILGQFRLVQAHARFRASRNEGDLDHLDNDPPYLGYLCADMEEYYQLFNDAKAEVEAV
SAGRKPADEAAARIHALIDERDFAKPMFGFGYCITGDKPQLNNSKYSLLPAMRLMYWT
QTRAPAEVKKYFDYNPMFALLKAYITTRIGLALKK

>PyrH
MERRKRERLGSLGRPTKKELRMIRSVVIVGGGTAGWMTASYLKAAFDDRIDVTLVESG
NVRRIGVGEATFSTVRHFFDYLGLDEREWLPRCAGGYKLGIRFENWSEPGEYFYHPFER
LRVVDGFNMAEWWLAVGDRRTSFSEACYLTHRLCEAKRAPRMLDGSLFASQVDESLG
RSTLAEQRAQFPYAYHFDADEVARYLSEYAIARGVRHVVDDVQHVGQDERGWISGVH
TKQHGEISGDLFVDCTGFRGLLINQTLGGRFQSFSDVLPNNRAVALRVPRENDEDMRPY
TTATAMSAGWMWTIPLFKRDGNGYVYSDEFISPEEAERELRSTVAPGRDDLEANHIQM
RIGRNERTWINNCVAVGLSAAFVEPLESTGIFFIQHAIEQLVKHFPGERWDPVLISAYNER
MAHMVDGVKEFLVLHYKGAQREDTPYWKAAKTRAMPDGLARKLELSASHLLDEQTIY
PYYHGFETYSWITMNLGLGIVPERPRPALLHMDPAPALAEFERLRREGDELIAALPSCYE
YLASIQ

>ThaI
MDNRIKTVVILGGGTAGWMTAAYLGKALQNTVKIVVLEAPTIPRIGVGEATVPNLQRAF
FDYLGIPEEEWMRECNASYKMAVKFINWRTPGEGSPDPRTLDDGHTDTFHHPFGLLPSA
DQIPLSHYWAAKRLQGETDENFDEACFADTAIMNAKKAPRFLDMRRATNYAWHFDAS
KVAAFLRNFAVTKQAVEHVEDEMTEVLTDERGFITALRTKSGRILQGDLFVDCSGFRGL
LINKAMEEPFIDMSDHLLCNSAVATAVPHDDEKNGVEPYTSSIAMEAGWTWKIPMLGR
FGSGHVYSDHFATQDEATLAFSKLWGLDPDNTEFNHVRFRVGRNRRAWVRNCVSVGL
ASCFVEPLESSGIYFIYAAIHMLAKHFPDKTFDKVLVDRFNREIEEMFDDTRDFLQAHYY
FSPRVDTPFWRANKELKLADSIKDKVETYRAGLPVNLPVTDEGTYYGNFEAEFRNFWT
NGSYYCIFAGLGLMPRNPLPALAYKPQSIAEAELLFADVKRKGDTLVESLPSTYDLLRQL
HGAS

>RebH
MSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVGEATIPNLQTAFF
DFLGIPEDEWMRECNASYKVAIKFINWRTAGEGTSEARELDGGPDHFYHSFGLLKYHEQ
IPLSHYWFDRSYRGKTVEPFDYACYKEPVILDANRSPRRLDGSKVTNYAWHFDAHLVA
DFLRRFATEKLGVRHVEDRVEHVQRDANGNIESVRTATGRVFDADLFVDCSGFRGLLIN
KAMEEPFLDMSDHLLNDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGRFG
TGYVYSSRFATEDEAVREFCEMWHLDPETQPLNRIRFRVGRNRRAWVGNCVSIGTSSCF
VEPLESTGIYFVYAALYQLVKHFPDKSLNPVLTARFNREIETMFDDTRDFIQAHFYFSPRT
DTPFWRANKELRLADGMQEKIDMYRAGMAINAPASDDAQLYYGNFEEEFRNFWNNSN
YYCVLAGLGLVPDAPSPRLAHMPQATESVDEVFGAVKDRQRNLLETLPSLHEFLRQQH
GR

>PltA
MSDHDYDVVIIGGGPAGSTMASYLAKAGVKCAVFEKELFEREHVGESLVPATTPVLLEI
GVMEKIEKANFPKKFGAAWTSADSGPEDKMGFQGLDHDFRSAEILFNERKQEGVDRDF
TFHVDRGKFDRILLEHAGSLGAKVFQGVEIADVEFLSPGNVIVNAKLGKRSVEIKAKMV
VDASGRNVLLGRRLGLREKDPVFNQFAIHSWFDNFDRKSATQSPDKVDYIFIHFLPMTN
TWVWQIPITETITSVGVVTQKQNYTNSDLTYEEFFWEAVKTRENLHDALKASEQVRPFK
KEADYSYGMKEVCGDSFVLIGDAARFVDPIFSSGVSVALNSARIASGDIIEAVKNNDFSK

SSFTHYEGMIRNGIKNWYEFITLYYRLNILFTAFVQDPRYRLDILQLLQGDVYSGKRLEV
LDKMREIIAAVESDPEHLWHKYLGDMQVPTAKPAF

>HalB
MSSAPAAGIDPAVSHARTFDVAILGSGIAGSMLGAILARNGARVLLVDASTHPKFAVGE
STIPYTLVALRTIAERYDVPEIKTLATFTNCTKVIGPQFGVKKHFGFLLHREGEPQDPREV
NQFDTPGLLHEASHMYRQATDAYLFHAAVKYGCVPRQNFRVVDVDFDDAGVTISGGD
GEQYRARYVVDASGFRSPLAEKFALRENPCRMKHHSRSLWNHMVNVPRTDALFDHSA
ADTPPVPWYEGTVHHMFDRGWFWVIGFDNHPASRNPLCSVGLTLDERKYPKDPSLTPQ
QEFESMAARYPDIARQFAGANPVREWVSTDRLQYSSKQTVGDRWCLLSHAAGFVDPLF
SRGLSNTAEAINALSWRLIAAIKDDDFSQERFEYVDRLQQGLFDYNDALVNASFIAFNH
YDLWTAVFRIWAWGSNAGTFFLQEALTKFLRDGRDEHFRAQEDVPHLGLYWPSHDGY
KKLFDEMIAQTDAYEQGLVTGKQAADALYDILVHADFVPKHFGFADRNRRFLHPTPNV
LRKTVRWAMREADPDVRRLMIGTGKEAVKRKLTGRRIF

>KrmI
MASGACTSHRSDRQVPHEHTAIKSICVVGGGTAGYLSALTLRRRMPHLKISLLESSRVPII
GVGEATTPSLPALLHGVLDMDIVDFFTQVKPTFKMGIRFEWGGTTPEHYFNYPFGAHGQ
LLEAMTYQGHIRDYSLGSMLMSADKMAIFRTGDGAYESRLHYGAISYAYHIDNQRFVS
WLQREAKRAGLHHIDANIQDVVTDDSGQNVVALLTDDGARHEYDFYVDCSGFRSLLLE
KALDSEWVPYDQTLFTDRAVTANVPHGGHIKPYTVAETMDHGWCWNIPMREDDHRG
YVFSSAFCDEETAIKEMQAKNPGMSDTKLVTFRSGRHRHFWKGNVVAIGNSYAFVEPL
ESTGVHMIAEEILAFIGNFPQSTHDDANRAALNRYMAAYWDELRWFLGLHFRFNTKLD
TPFWKACRADVDVSGFDDCLATYRKCAPLSYRNHYLTFNRLWGDHGRDVLLMGQQV
PANFLPPRESKQQWLRRVELARQAVEMAVDQAEAIRLLESNPDVLRQLVTDDGAWIHT
VGELFSSETALYSDVYAARLGHMKKPAGFGGAQNVILPSTL

>Bmp5
MRKKIAVIGAGLSGIAAIKQLTDGGHEVVCFEKAESFGGVFADKKIYEDLHLTISNYFMA
YSDYVPSQQKLKFWSKKEYVNYLGEYLARFELGQYIHYDHEVRKVEKQAGKWQVTTK
HGMAEQTDTFDMVAVCSGHFQKPKMPELAGLDMFEGEIEHSNDYRDKHKYAGKRVLC
VGLGESSADITSEISEVASKCILSLRRYPAVAPRYMAFQEDPYFTIDTSWLTSRIVNKLPH
RYHGGITKGIFTKYVNSRNDHVRIRGEWLKKSGPSHHQAVTKNERLFRPIADGKVVPNI
GGIVRFEKNAVVFQDGTREEIDAVVFCTGYQLSFPFLDVSISNMRDLYKQMFIPEMGDSL
SFIGFVRPQQGGIPVIAEMQCRYLSQLASGEKSLPPLSEMVDIIKYDTEHWQTEYKITPHV
ASLVNYCHYMDSMAKLVGCMPEIPSLFKDPLLRVKLLHNPQFAAQYRLDGPNKMTHT
ARSFLLGFPNISSWPRIIHFELALITQKLLSRLRLDGLRELSK

>Bmp2
MDQFKSYDVVIIGSGPAGSLCGIECRKKGLSVLCIEKDEFPRFHIGESLTGNAGQIIRDLG
LADEMNAAGFPDKPGVNVIGSLSKNEFFIPILAPTWQVRRSDFDNMLKRRALEHGVEYQ
QGLVKDVIKHEEKVVGAIYKADGVEHQVRSKVLVDASGQNTFLSRKGIAGKREIEFFSQ
QIASFAHYKNVERDLPPFSTNTTILYSKQYHWSWIIPISPDTDSLGIVIPKDLYYKECKNPD
DAIEWGMEHISPEIRRRFKNAERVGESQSMADFSYRIEPFVGDGWLCIGDAHRFLDPIFS
YGVSFAMKEGIKAADAIKRAIDGNDWKTPFYEYRDWSNGGQQIAADLIRYFWIYPIFFG
YQMQNPDLRDEVIRLLGGCCFDCEGWKAPTIFRNAIEEYDRKQMAG

>HrmQ
MSDFDYDIGIIGGGPAGSTMASYLAKAGISAIVLESAEFPRPHVGESLVTATTPVLQEIGA
LAKVDAAGFPKKYGAAWTSAAKETVPEMGFTGMTQGFRLASVEFQERDQPGVDQAYT
YHVDRGKYDQLLLQHAAELGARVRERTRVQDVEFTDDGVVLTAVGDEGTERLRVRM
VVDASGRQTMLGRKKKLKVPDPVFNQYAIHAWFDGLDRKALADDPEQEDFIFIHFLPVS
DTWVWQIPISETTTSVGVVTQKAKVKSSGQEREEFFWECLGTRPELRAALEKAERVREF
KTEGDYSYGMKEVVGDRWALIGDAARFVDPIFSSGVSVAMNSARLLAKDIVSAVERDD
FSKKSFGDYERVLRRGVSSWYEFISIYYRLNILFTAFVQDPRYRLDVLKMLQGDVYGEE
EPKALTRMREIVRRVEENPDHLWHPYLGTLRAPAAAPLF

>PrnA
MNKPIKNIVIVGGGTAGWMAASYLVRALQQQANITLIESAAIPRIGVGEATIPSLQKVFF
DFLGIPEREWMPQVNGAFKAAIKFVNWRKSPDPSRDDHFYHLFGNVPNCDGVPLTHYW
LRKREQGFQQPMEYACYPQPGALDGKLAPCLSDGTRQMSHAWHFDAHLVADFLKRW
AVERGVNRVVDEVVDVRLNNRGYISNLLTKEGRTLEADLFIDCSGMRGLLINQALKEPF
IDMSDYLLCDSAVASAVPNDDARDGVEPYTSSIAMNSGWTWKIPMLGRFGSGYVFSSH
FTSRDQATADFLKLWGLSDNQPLNQIKFRVGRNKRAWVNNCVSIGLSSCFLEPLESTGIY
FIYAALYQLVKHFPDTSFDPRLSDAFNAEIVHMFDDCRDFVQAHYFTTSRDDTPFWLAN
RHDLRLSDAIKEKVQRYKAGLPLTTTSFDDSTYYETFDYEFKNFWLNGNYYCIFAGLGM
LPDRSLPLLQHRPESIEKAEAMFASIRREAERLRTSLPTNYDYLRSLRDGDAGLSRGQRG
PKLAAQESL

>Rdc2
MSVPKSCTILVAGGGPAGSYAAAALAREGNDVVLLEADQHPRYHIGESMLPSLRPLLRF
IDLEDKFDAHGFQKKLGAAFKLTSKREGYTDFIAAHGPRGYSWNVVRSESDEILFNHAR
ESGAQAFQGIKINAIEFEPYEEEYPNGEKVANPGKPTSAKWSSKDGSSGDIAFKYLVDAT
GRIGIMSTKYLKNRHYNEGLKNLAIWGYYKNNIPWAQGTPRENQPFFEGMRDGAGWC
WTIPLHNGTVSVGAVMRKDLFFEKKKSLGENATNTQIMAECMKLCPTIGELLAPAELVS
DIKQATDYSYSATAYAGPNFRIVGDAGCFIDPFFSSGHHLAVAGALAAAVSINASIKGDC
TEYEASRWHAKKVDEGYTLFLLVVMAALKQIRMQENPVLSDVDEDGFDRAFQFLRPVI
QGSGSSEAVKKFTKEDVAQTIDFAVHALNNMAELDMDIPEHMINGDKEGENGVTNGNN
GAAKTAGLASNMEKLTNDEEKVLNGLRILGKAAPGGTLADFEGTAIDGLMPRLEHGKL
GLNKV

>SttH
MNTRNPDKVVIVGGGTAGWMTASYLKKAFGERVSVTLVESGTIGTVGVGEATFSDIRH
FFEFLDLREEEWMPACNATYKLAVRFQDWQRPGHHFYHPFEQMRSVDGFPLTDWWLQ
NGPTDRFDRDCFVMASLCDAGRSPRYLNGSLLQQEFDERAEEPAGLTMSEHQGKTQFP
YAYHFEAALLAEFLSGYSKDRGVKHVVDEVLEVKLDDRGWISHVVTKEHGDIGGDLFV
DCTGFRGVLLNQALGVPFVSYQDTLPNDSAVALQVPLDMEARGIPPYTRATAKEAGWI
WTIPLIGRIGTGYVYAKDYCSPEEAERTLREFVGPEAADVEANHIRMRIGRSEQSWKNN
CVAIGLSSGFVEPLESTGIFFIHHAIEQLVKHFPAGDWHPQLRAGYNSAVANVMDGVRE
FLVLHYLGAARNDTRYWKDTKTRAVPDALAERIERWKVQLPDSENVFPYYHGLPPYSY
MAILLGTGAIGLRPSPALALADPAAAEKEFTAIRDRARFLVDTLPSQYEYFAAMGQRV

>ThdH
MDNRIKTVVILGGGTAGWMTAAYLGKALQNTVKIVVLEAPTIPRIGVGEATVPNLQRAF
FDYLGIPEEEWMRECNASYKMAVKFINWRTPGEGSPDPRTLDDGHTDTFHHPFGLLPSA
DQIPLSHYWAAKRLQGETDENFDEACFADTAIMNAKKAPRFLDMRRATNYAWHFDAS
KVAAFLRNFAVTKQAVEHVEDEMTEVLTDERGFITALRTKSGRILQGDLFVDCSGFRGL
LINKAMEEPFIDMSDHLLCNSAVATAVPHDDEKNGVEPYTSSIAMEAGWTWKIPMLGR
FGSGHVYSDHFATQDEATLAFSKLWGLDPDNTEFNHVRFRVGRNRRAWVRNCVSVGL
ASCFVEPLESSGIYFIYAAIHMLAKHFPDKTFDKVLVDRFNREIEEMFDDTRDFLQAHYY
FSPRVDTPFWRANKELKLADSIKDKVETYRAGLPVNLPVTDEGTYYGNFEAEFRNFWT
NGSYYCIFAGLGLMPRNPLPALAYKPQSIAEAELLFADVKRKGDTLVESLPSTYDLLRQL
HGAS

>VhaH
MSVEDFDVVVAGGGPGGSTVATLVAMQGHRVLLLEKEVFPRYQIGESLLPATVHGVCR
MLGVADELAGAGFPIKRGGTFRWGARPEPWTFHFGISAKMAGSTSHAYQVERARFDEI
LLNNAKSKGVVVREGCSVHDVVEDGERVTGAKYTDPDGNEREVSARFVIDASGNKSRL
YSKVGGSRNYSEFFRSLALFGYFEGGKRLPEPVSGNILSVAFDNGWFWYIPLSDTLTSVG
AVVRREDADKIQGDREKALNALIAECPLISEYLSNATRVTTGKYGELRVRKDYSYQQET
YWRPGMVLVGDAACFVDPVFSSGVHLATYSALLAARSINSVLAGDLDEKTALNEFEAR
YRREYGVFYEFLVSFYQMNVNEESYFWQAKKVTQNQSTDIESFVELIGGVSSGETALTA
ADRIAARSAEFAAAVDEMAGGDGDNMVPMFKSTVVKQAMQEAGQVQMKALLGEDA
EPELPLFPGGLVTSPDGMKWLPHHPA

>MibH
MARSEESNTLARLFDVLGDDAAAAREWVTEPHRLIASNERLGTAPEAPADDDPEAIRTV
GVIGGGTAGYLTALALKAKRPWLDVALVESADIPIIGVGEATVSYMVMFLHHYLGIDPA
EFYQHVRPTWKLGIRFEWGSRPEGFVAPFDWGTGSVGLVGSLRETGNVNEATLQAML
MTEDRVPVYRGEGGHVSLMKYLPFAYHMDNARLVRYLTELATRRGVHHVDATVAEV
RLDGPDHVGDLITTDGRRLHYDFYVDCTGFRSLLLEKALGIPFESYASSLFTDAAITGTL
AHGGHLKPYTTATTMNAGWCWTIPTPESDHLGYVFSSAAIDPDDAAAEMARRFPGVTR
EALVRFRSGRHREAWRGNVIAVGNSYAFVEPLESSGLLMIATAVQILVSLLPSSRRDPLP
SNVANQALAHRWDAIRWFLSIHYRFNGRLDTPFWKEARAETDISGIEPLLRLFSAGAPLT
GRDSFARYLADGAAPLFYGLEGVDTLLLGQEVPARLLPPRESPEQWRARAAAARSLAS
RGLRQSEALDAYAADPCLNAELLSDSDSWAGERVAVRAGLR

>ArmH1
MEEQVPSSANILVIGGGPAGSYAATVLAREGFEVILLEKDVFPRYHIGESMLPSCRPFLKF
IDCEEKLKNHGFTMKPGAAVKLNQYKREGYTDFISTNPDNAARNVVRSEFDELLLRHAS
EMGVHVYEGVRVEEIHFAPDEPTRPISLTWSKEDKTRGTVSFNWLVDASGRNGLMSTR
YLKNRTFNKTLRNVAVWGYWTGTSCYAPGTTRENAPWFEALIDETGWAWFIPLHNGT
TSVGVVLVEEESKRKKSQYRSECKDKSPSEIQHDCYMADLQRAPGLIQLLGTATFEGKL
MSAGDYSYHATEYAGSNFRLAGDAGAFIDPFFSSGIHLALTGGLSAASTIAASIRGDCTE
SEACGFHNSKVGTAYTRFLLVVLGVYKQIRSQETAVLYDVDEDNFDKAFHFLRPVIQGC
ADADEGLTEAELQSTLDFCTNVFAPTQPETHEAAIQRLGSLADPDGPLLDTDTIDKLTGT
DMEAKHALWKINARKPLHSMYDCKRNFGTEIINGFYVKMEQGVLGLVRT

>StaI
MDGLSGDPLPSTIAKDSIEGWWEMSVETFDVVVAGGGPGGSTVAALVAMQGHRVLLL
EKEVFPRYQIGESLLPSTVHGVCRLLGVTDELAAAKFPVKRGGTFRWGARPEPWTFSFA
VSPRMTGPTSFAYQVERARFDEILLNNAKSKGVTVREGCSVTAVVEEGDRVTGLRYTDS
DGGEHEVGARFVIDASGNKSRLHASVGGTREYSEFFRSLALFGYFENGKRLPEPYSGNIL
SVAFDSGWFWYIPLSDTLTSVGVVVRREMAEKIQGDREKAYNALIAECPLIAEYLSDAR
RVTTGNYGELRVRKDYSYHQTKFWRPGMILVGDAACFVDPVFSSGVHLATYSALLAAR
SINSVLAGDLDEKTAFDEFEGRYRREYGVFYEFLVSFYQMNVNEESYFWHAKKVTNNE
HTEFESFVDLVGGVASGETALAASRLAERSAEFAGAVDQMAGGDDGDMVPLFKSHVV
KQVMQEGGQVQMKAVLGEDAAPEEPVFPGGLVSSPDGMKWLPYRG

>AcIH
MSIPNRCTVLVVGGGPAGSYAAAALAREGVDTVLLEADVFPRYHIGESMLPSIRHFLRFI
DLDSKFDSYGFVNKNGAAFKLNSKPEAYTDFIAAGGPGSHAWNVVRSEADHLMFKHA
GENGAQVFDGVKVNSIEFEQIDGLTVDPSLSELGRPVSATWSCKATGGKGSITFEYLIDA
TGRAGLVSTKYMKNRRYNQGLKNVASWGYWSNAGSYGVGTPREGDPYFEAIEDGSG
WVWLIPLHNGTTSIGVVMNQEAATAKKRETGATTKDLYLNTIKNTPGVWQLLDKAELQ
SDLKSASDWSYNASSYASPYLRIVGDAGCFIDPFFSSGVHLALASGLSAALTIRAAQRGD
CDEQAAVSWHSKKVAEGYTRFLLVVMSALKQISDREKPVLTDFDEDSFNRAFDFFRPII
QGTADVDKNLTQAEIAQTIEFCVQAFQTASTDEQDAVMTKVAAINSQNGTEGALRELH
ASLSADERRTLTTIQARQIIRSEDQMNIDNFTIDVIDGMVPRLERSSLGLARFVPKAQTSR
EDGLRATLGLPEKQKSIFSY

>HalX
SAQRADVVITGGGLAGLSLALQLRQRDPALAITVLERRAHPVREAAFKVGESTVEIGAH
YFADVLGLREHLEAEQIRKFGFRFFFSDKRDDIDRCTELGVSKILPTPSWQIDRGRFENFL
GERARAQGIVFLDSCSVKGVDLSDDDASDHRVRYERGGEAGTLDARWVVDASGRAGL
LKRKLGLTQDNAHDANAVWWRVEGLIDPNAWSQDSSWLQRCTPPDRWRSTNHMCGP
GYWFWLIPLSSGAHSLGIVCDAAMHPLETMNTHDKAMAWLRTHQPQVAATLDQTDYR
LQDFLFLRNFSYGCKQVFSTQRWALTGEAGVFLDPFYSPGSDFIAISNTYICELIGRDRAG
RSLSPYVELYEQLYFSFYENTLSLYQDQYALFGDAQVMPVKVIWDYTYYWSLLAPLFC
SGRIADLGLLARMKADFFYARDMNLAMQRVLHDWGQHNTALGTVTGDGRLLDQYLIG
WFNELNGALHDTLDDEAFAARIHANIARMAVLAREILEQARQRHPALPDHGLAALTAN
AVGDPILTAAWYADAA

>AoiQ
MTNVPEKCQVLVIGGGPAGSYSASALAREGIDVVLLEAEKFPRYHIGESMLPSMRHFLK
FIDAYDKWDAHGFNVKNGGAFRLNWSRPETYTDFIAAGGPGGYAWNVIRSEADELLFK
HAAECGVKTFDDTKVASIEFAPSEDANPLGRPVSATWTRKDGTSGTLAMDYIVDASGR
NGLISTKYLKNRTFNKGLRNVASWGYWKGGGVHGVGTHKEGAPYFEALKDASGWVW
FIPLHNGTHSVGVVQNQEMATEKKRKMAEPSSKGFYLESLEFVPGIKDLLSNAELISEVK
SASDWSYSASNYAFPGVRIAGDAGSFIDPFFSSGVHLALSGGLSAATTIAAAIRGDCDET
VAASWHDKKTSESYTRFLLVVSSALKQIRSQDEPVISDFDEQSFERAFDLFRPIIQGQADA

DAKGKLTQAEMSKTVEFCFRAFAHVSFEEKEALVQKLKSLGHDADANDEANRKALDE
LEKHLTPEEQAILKTLKGRRMVRPEDSLNIDNFTLDSIDGLAPRLERGNLGLSPAKKAEV
KYTTHDALSFLNGEARAAKKTLSNGESQTNGNHLCKDHDQTNGHTAINGHIENNSQGD
TNGHTEANGHSQTNGHSNGHAHTNGSHTTNAHDELNGHNNVEVRSVKSCMADLIAAE
KISSQTSIEEGTRHRLISSLHQSAEDLETPFDTVVRLVDAGRQTAMVCIGGDLGIFKSLVE
SKRPLSAEELAKATMADPLLVARIMRYMVASRLVGETGPDQYVASKKTYVFADPRIEH
PIRFFHAFSNPAFHALPEFLKETGYQNEPKGSAFQKALNTDLEPYPWLKQHPDVLKNFQ
AAMRLTRDANGVDMMPLDQSVSIGHDGAMFVDIGGNTGHQAAEVLSKYPELAGRVIV
QDRGEVIKCAPDIKGIQWMEHDFFQTQPVKGAKYYYLRAILHNWDDKNTVQILSNIVPA
MSADSLVAIDEVVVPEENAHVWPAGLDLQMYSLFSTTERTASQWDAILDKAGLRAVAV
KKYAPVMQSSVIFAAAK

>TiaM
MADAMPKVIVIGGGPGGSTAAGLLAKGGVEVKLLERAEFPRYHIGESITTSCRAIIDYVG
ALEKVNACGFTTKTGVLMRWGQEKDWVIDWTAQFGADTRSWQVDREIFDQVLLDNA
ADSGAEVVQQAHVKRVLFDGDRAVGVEWQPAGGELRTETADYVIDASGRAGLISAQH
YNNRRPHEIFRNVAIWGYWENAELLPDSPAGGINVISSPDGWYWLIPLRDNRFSVGFVT
HKTNFVERRREFDSQEAMLAAVLGESETMKQVLAQATQLPGVRIEQDFSYAADKFCGD
GYFLVGDSACFLDPLLSTGVHLAMYSGMLSAASILSLQRGEVNEPEAQAFFESLYRNAY
HRLFTLVISFYQQHAGKDRYFALADTLSRAAQGDTQSADLAFGEIIAGLTDLREAADRA
GTAAAPITETIDAAATGGAKPVQELLGAAEQARLRAEESSPNQRVTMAPMYIDANDLY
DAASGLYLVTEPRLGISRVTSEAAVAAG

>Tcp21
MTSLSSDDSIDGWWKMPVEEFDVVVAGGGPAGSTVATLVAMQGHRVLLLEKETFPRY
QIGESLLPSTVHGVCRMLGVSDELAAAGFPVKRGGTFRWGARPEPWTFSFAVSPLMNG
PTSFAYQVERARFDEILLRNAQRKGVDVREGCSVTGVIEGDERITGVRYTDPDGGEHEV
SARFVVDASGNKSRLYSQVGGVRNYSEFFRSLALFGYFENGKRLPEPYSGNILSVAFDSG
WFWYIPLSDTMTSVGAVVRREMAEKIQGDREKALQALIAECPLISDYLADARRITTGRY
GELRVRKDYSYHQTSYWRPGMILVGDAACFVDPVFSSGVHLATYSALLAARSINSVLA
EDLDEKTALSEFEARYRREYGVFYEFLVSFYQMNVNEESYFWQAKKVTNNKHTEIESF
VELIGGVSSGEAALATAGRIAERSAEFAAAVDQMAASADGNMVPMYKSEVVKQVMQE
GGQVQMKAVLGADAEPELPLFPGGLVVSPDGMKWLPYRPA

>BhaA
MSVEDFDVVVAGGGPGGSTVATLVAMQGHRVLLLEKEVFPRYQIGESLLPATVHGVCR
MLGISDELANAGFPIKRGGTFRWGARPEPWTFHFGISAKMAGSTSHAYQVERARFDEM
LLNNAKRKGVVVREGCAVTDVVEDGERVTGARYTDPDGTEREVSARFVIDASGNKSRL
YTKVGGSRNYSEFFRSLALFGYFEGGKRLPEPVSGNILSVAFDSGWFWYIPLSDTLTSVG
AVVRREDAEKIQGDREKALNTLIAECPLISEYLADATRVTTGRYGELRVRKDYSYQQET
YWRPGMILVGDAACFVDPVFSSGVHLATYSALLAARSINSVLAGDLDEKTALNEFELRY
RREYGVFYEFLVSFYQMNVNEESYFWQAKKVTQNQSTDVESFVELIGGVSSGETALTA
ADRIAARSAEFAAAVDEMAGGDGDNMVPMFKSTVVQQAMQEAGQVQMKALLGEDA
EPELPLFPGGLVTSPERMKWLPHHPA

>Th-Hal
LNNVVIVGGGTAGWMTASYLKAAFGDRIDITLVESGHIGAVGVGEATFSDIRHFFEFLGL
KEKDWMPACNATYKLAVRFENWREKGHYFYHPFEQMRSVNGFPLTDWWLKQGPTDR
FDKDCFVMASVIDAGLSPRHQDGTLIDQPFDEGADEMQGLTMSEHQGKTQFPYAYQFE
AALLAKYLTKYSVERGVKHIVDDVREVSLDDRGWITGVRTGEHGDLTGDLFIDCTGFR
GLLLNQALEEPFISYQDTLPNDSAVALQVPMDMERRGILPCTTATAQDAGWIWTIPLTG
RVGTGYVYAKDYLSPEEAERTLREFVGPAAADVEANHIRMRIGRSRNSWVKNCVAIGL
SSGFVEPLESTGIFFIHHAIEQLVKNFPAADWNSMHRDLYNSAVSHVMDGVREFLVLHY
VAAKRNDTQYWRDTKTRKIPDSLAERIEKWKVQLPDSETVYPYYHGLPPYSYMCILLG
MGGIELKPSPALALADGGAAQREFEQIRNKTQRLTEVLPKAYDYFTQ

>AcOTAhal
MAIPQKATVLVIGGGPGGSYSASALAREGIDTVVLEADVFPRYHIGESLVASIRPFLKFID
LDDTFVNYGFVRKNGAAFKLNNQKEAYTDFILEAGADTFAWNVVRSESDDLMFKHAA
KSGAQTFDGVRVTSIEFTDDDDDDDNNTNRPVSASWKAKDGRTGSIEFDYLVDASGRA
GITSTKYLKNRTFNNYLKNVASWGYWEGATPYGMGTPVEGQPFFEALQDGSGWVWFI
PLHNNTTSIGIVMNQELSTQKKKLSTTTSSRAFYLESLAGARGISRLLDPTTATLTSDIKH
ASDWSYNASAYGSPYLRIVGDAGAFIDPYFSSGVHLAVSGGLSAAVSIAASIRGDCPEEA
AWKWHSQGVANRYGRFLLVVLGATKQIRAGDRPVLNGVGDEGFDEAFMVIRPGEFPFF
FFSFSFFFLSSEGVGVVVWNCAN

>HalY
MVTDFDIGIIGGGPGGTTAASYLARAGLSVAVFEGELFPREHVGESLVPATTPVLREIGA
LAKLDEAGFPRKYGAAWTSPSGARASDTFGMADLAFAEDEDVLGGQSYTYHVDRGRF
DQLLLEHSSTLGAKVFEGARVRRVDFDGPHPRIVHASGGSESRTGVRMVIDASGRGTLL
GRQLGLKVTDPTFDQYAVHGWFTGLDREAIAGRDALADYLFVHFLPAEDSWVWQIPIT
DTITSVGVVSRKSRLRQAGGDREAFFWGTLEHRPELAAALRACTPVRPLKVDGDFSYA
MRQITGDGWAMVGDAARFVDPIFSSGVGIAMNGARLLSQDVLAAAEAGDFARQRFAE
YEKKQRRGLTNWYNFIAVYYRLNILFTVFVQDPRYRLDVLRMLQGNVYETDHPPVLTT
MQGVVETVESNPDHLWHPHLGSLRAPAAAPLF

>KtzR
MTAAYLKTAFGDRLSITVVESSRIGTIGVGEATFSDIQHFFQFLNLREQDWMPACNATY
KLGIRFENWRHVGHHFYQPFEQIRPVYGFPLTDWWLHDAPTDRFDTDCFVMPNLCEAG
RSPRHLDGTLADEDFVEEGDELANRTMSEHQGKSQFPYAYHFEAALLAKFLTGYAVDR
GVEHVVDDVLDVRLDQRGWIEHVVTAEHGEIHGDLFVDCTGFRGLLLNKALGVPFVSY
QDTLPNDSAVALQVPLDMQRRGIVPNTTATAREAGWIWTIPLFGRVGTGYVYAKDYLS
PEEAERTLREFVGPAAADVEANHIRMRIGRSQESWRNNCVAIGLSSGFVEPLESTGIFFIH
HAIEQLVKHFPAADWNPKSRDMYNSAVAHVMDGIREFLVIHYRGAARADNQYWRDTK
TRPLPDGLAERIECWQTQLPDTETIYPYYHGLPPYSYMCILMGGGAIRTPASAALALTDQ
GAAQKEFAAVRDRAAQLRDTLPSHYEYLARMRGLDV

>KtzQ
MDDNRIRSILVLGGGTAGWMSACYLSKALGPGVEVTVLEAPSISRIRVGEATIPNLHKVF
FDFLGIAEDEWMRECNASYKAAVRFVNWRTPGDGQATPRRRPDGRPDHFDHLFGQLPE
HENLPLSQYWAHRRLNGLTDEPFDRSCYVQPELLDRKLSPRLMDGTKLASYAWHFDAD
LVADFLCRFAVQKLNVTHVQDVFTHADLDQRGHITAVNTESGRTLAADLFIDCSGFRSV

LMGKVMQEPFLDMSKHLLNDRAVALMLPHDDEKVGIEPYTSSLAMRSGWSWKIPLLG
RFGSGYVYSSQFTSQDEAAEELCRMWDVDPAEQTFNNVRFRVGRSRRAWVRNCVAIG
VSAMFVEPLESTGLYFSYASLYQLVKHFPDKRFRPILADRFNREVATMYDDTRDFLQAH
FSLSPRDDSEFWRACKELPFADGFAEKVEMYRAGLPVELPVTIDDGHYYGNFEAEFRNF
WTNSNYYCIFAGLGFLPEHPLPVLEFRPEAVDRAEPVFAAVRRRTEELVATAPTMQAYL
RRLHQGT

>ChlA
MDPTDIQSDLKDKLFGARKNCSTPSGSTMEAVGNKHYDVVVIGAGIAGLSAVRHLLLKI
PELRANRVAVIDPRSSIRNNLLEDYKVGESTVEISGMFFARELELQDYLIENQPPKFSLQF
HWPKKVDKTDSMDDYHSTWAVKNPDIQAFQLNRSKMEKDLLQMVIDQGAVYYHGRV
KDVDIGMGDAVKTIDIQMLSEEVNFDDQKPLERINITTDYIADASGRNFIIGTRTDNVLK
DPKHLFGLQNASTWLRVKNTDKKYFDFNDQSVTSSWYYATNHFFGPGYWIWMIPIERA
THDFSIGVSYHRDKYDPSEFNSLDKFMAFLETNQKVLYNIIKSGEIVDFHRWPRLAHTSK
VFVSEDNWCVLGDAAAIFDPFYSTGMVMIAMEVECLTELTKYRLSGDKCATKTRQQAY
DKLIRAITQTNCHLIKKHSDHLGNASIMSWRIYLESLSYFGILLPAYIGKYHLCPIFGHEFA
DNNWRLIEFRDRLLNLMDYIGQQGQNVGFMDNHRGGQLIGSFCPTSSWDYEHCISMCM
YGHKRLNLHKYMIWTHFYQMAGVIKMLYRAFGARLLLNGDFQWMMTHTMKTMGFH
SVKSWVHYLSNISNPVTDDYYDRNQISFKEYQYTGEAVPWESSM

>BrvH
GAMDEIDDPRIRSVVIVGGGTAGWMTAAALVQHFRTAPLKITVVESSDIGTIGVGEATIP
TIRRFYGQLGLRDDDVMRATQATCKLGIRFLDWSGPGSDFIHPFGLYGQDVKGIGFHHY
WLKQRRAGDAAPLAAYSLGAALAAGGKFTLPSPHPPSQLSVFDWALHLDAGLFAQHLR
AYAEAGGCARIDARIRSVELRPEDGFVRALTLDDGREVEGDLFVDCSGFKGLVIGEALG
VGFEDWGRWLPCDAAYAVQSENRPGDAPAPFTRVTARSAGWQWGIPLRHRAGNGLVF
SSAHLSDDQALAELMPHLLGDPLTEPRRIPFRPGRRSQAWAKNCVAIGLSSGFLEPLEST
SIALIETGIERLKALFPDRRFAQPILDEFNDQTAREMERVRDFIILHYKLNRRTDTDFWRD
CREMPVPETLERKIALWTARGQFVRYRWEMFHPASWLAIYDGFGLYPDHHDPAVDAM
DPAYLARSLAEMRANIADLVARTPEHAQFLAGLDPAASAA

>FMO
MGKKVAIIGAGVSGLASIRSCLEEGLEPTCFEKSNDIGGLWKFSDHAEEGRASIYKSVFS
NSSKEMMCFPDFPFPDDFPNFMHNSKIQEYIIAFAKEKNLLKYIQFKTFVSSVNKHPDFA
TTGQWDVTTERDGKKESAVFDAVMVCSGHHVYPNLPKESFPGLNHFKGKCFHSRDYK
EPGVFNGKRVLVVGLGNSGCDIATELSRTAEQVMISSRSGSWVMSRVWDNGYPWDML
LVTRFGTFLKNNLPTAISDWLYVKQMNARFKHENYGLMPLNGVLRKEPVFNDELPASIL
CGIVSVKPNVKEFTETSAIFEDGTIFEGIDCVIFATGYSFAYPFLDESIIKSRNNEIILFKGVF
PPLLEKSTIAVIGFVQSLGAAIPTVDLQSRWAAQVIKGTCTLPSMEDMMNDINEKMEKK
RKWFGKSETIQTDYIVYMDELSSFIGAKPNIPWLFLTDPKLAMEVYFGPCSPYQFRLVGP
GQWPGARNAILTQWDRSLKPMQTRVVGRLQKPCFFFHWLKLFAIPILLIAVFLVLT

>MICAL
MASPTSTNPAHAHFESFLQAQLCQDVLSSFQELCGALGLEPGGGLPQYHKIKDQLNYWS
AKSLWTKLDKRAGQPVYQQGRACTSTKCLVVGAGPCGLRVAVELALLGARVVLVEKR
TKFSRHNVLHLWPFTIHDLRALGAKKFYGRFCTGTLDHISIRQLQLLLLKVALLLGVEIH
WGVTFTGLQPPPRKGSGWRAQLQPNPPAQLANYEFDVLISAAGGKFVPEGFKVREMRG

KLAIGITANFVNGRTVEETQVPEISGVARIYNQSFFQSLLKATGIDLENIVYYKDDTHYFV
MTAKKQCLLRLGVLRQDWPDTNRLLGSANVVPEALQRFTRAAADFATHGKLGKLEFA
QDAHGQPDVSAFDFTSMMRAESSARVQEKHGARLLLGLVGDCLVEPFWPLGTGVARG
FLAAFDAAWMVKRWAEGAESLEVLAERESLYQLLSQTSPENMHRNVAQYGLDPATRY
PNLNLRAVTPNQVRDLYDVLAKEPVQRNNDKTDTGMPATGSAGTQEELLRWCQEQTA
GYPGVHVSDLSSSWADGLALCALVYRLQPGLLEPSELQGLGALEATAWALKVAENELG
ITPVVSAQAVVAGSDPLGLIAYLSHFHSAFKSMAHSPGPVSQASPGTSSAVLFLSKLQRT
LQRSRAKENAEDAGGKKLRLEMEAETPSTEVPPDPEPGVPLTPPSQHQEAGAGDLCALC
GEHLYVLERLCVNGHFFHRSCFRCHTCEATLWPGGYEQHPGDGHFYCLQHLPQTDHKK
EGSDRGPESPELPTPSENSMPPGLSTPTASQEGAGPVPDPSQPTRRQIRLSSPERQRLSSLN
LTPDPEMEPPPKPPRSCSALARHALESSFVGWGLPVQSPQALVAMEKEEKESPFSSEEEE
EDVPLDSDVEQALQTFAKTSGTMNNYPTWRRTLLRRAKEEEMKRFCKAQTIQRRLNEIE
AALRELEAEGVKLELALRRQSSSPEQQKKLWVGQLLQLVDKKNSLVAEEAELMITVQE
LNLEEKQWQLDQELRGYMNREENLKTAADRQAEDQVLRKLVDLVNQRDALIRFQEER
RLSELALGTGAQG

>StyA
MKKRIGIVGAGTAGLHLGLFLRQHDVDVTVYTDRKPDEYSGLRLLNTVAHHAVTVQRE
VALDVNEWPSEEFGYFGHYYYVGGSQPMRFYGDLKAPSRAVDYRLYQPMLMRALEAR
GGKSCYDDVSTEDLEGLSEQYDLLVVCTGKNALGKVFEKQSENSPFEKPQRALCVGLF
KGIKEAPIRAVTMSFSPGHGELIEIPTLSFNGMSTALVLENHIGSDLEVLAHTKYDDDPRA
FLDLMLEKLRKHHPSVAERIDPAEFDLANSSLDILQGGVVPAFRDGHATLSNGKTIIGLG
DVQATVDPVLGQGANMASHAAWILGEEILAHSVYDLRFSEHLERRRQDRVLCATRWT
NFILSALSALPPEFLAFLQILSQSREMADEFTDNFNYPERQWDRFSSPERIGQWCNQYAPT
IAA

>W2M
MKNNRHPANGKKLITMFGPDFPFAFDDWIEHPKGLGSIPAEHHGAEVAIIGAGIAGLVA
AYELMKMGLKPVVYEASKMGGRLRSQEFEGAKGIVAELGGMRFPVSSTAFFHYVDKL
GLESRPFPNPLTAASGSTVIDLEGTTYYAQMLSDLPALFQEVADAWADALESGSQFGDI
QQAIRDRDVPRLKELWNKLVPLWDDRTFYDFVATSKAFAKLSFYHREVFGQVGFGTGG
WDSDFPNSMLEIFRVVMTNCDEHQHLIVGGVQQVPVGLWSHVPERCAHWPKGTSLSSL
HRGAPRPGVKRIARAEDGSFTVTDNWGDTRQYAAVLTTCQSWLLTTQIECEESLFSQK
MWMALDRTRYMQSSKTFVMVDRPFWKDKDPETGRDLMSMTLTDRLTRGTYLFDNGD
DKPGVICLSYSWMSDALKMLPQPIEKRVKLALDALKKIYPKVDIAARIIGDPITVSWEAD
PHFLGAFKGALPGHYRYNQRMYAHFMQQDMPSEQRGMFIAGDDVSWTPAWVEGAVQ
TSLNAVWGIMNHFGGKTHAENPGPGDVFHEIGPITLAD

>FlavoC3
ADNLAEFHVQNQECDSCHTPDGELSNDSLTYENTQCVSCHGTLEEVAETTKHEHYNAH
ASHFPGEVACTSCHSAHEKSMVYCDSCHSFDFNMPYAKKWQRDEPTIAELAKDKSERQ
AALASAPHDTVDVVVVGSGGAGFSAAISATDSGAKVILIEKEPVIGGNAKLAAGGMNA
AWTDQQKAKKITDSPELMFEDTMKGGQNINDPALVKVLSSHSKDSVDWMTAMGADLT
DVGMMGGASVNRAHRPTGGAGVGAHVVQVLYDNAVKRNIDLRMNTRGIEVLKDDKG
TVKGILVKGMYKGYYWVKADAVILATGGFAKNNERVAKLDPSLKGFISTNQPGAVGD
GLDVAENAGGALKDMQYIQAHPTLSVKGGVMVTEAVRGNGAILVNREGKRFVNEITTR
DKASAAILAQTGKSAYLIFDDSVRKSLSKIDKYIGLGVAPTADSLVKLGKMEGIDGKALT

ETVARYNSLVSSGKDTDFERPNLPRALNEGNYYAIEVTPGVHHTMGGVMIDTKAEVMN
AKKQVIPGLYGAGEVTGGVHGANRLGGNAISDIITFGRLAGEEAAKYSKKN

>DAAO
MRVVVIGAGVIGLSTALCIHERYHSVLQPLDVKVYADRFTPFTTTDVAAGLWQPYTSEP
SNPQEANWNQQTFNYLLSHIGSPNAANMGLTPVSGYNLFREAVPDPYWKDMVLGFRK
LTPRELDMFPDYRYGWFNTSLILEGRKYLQWLTERLTERGVKFFLRKVESFEEVARGGA
DVIINCTGVWAGVLQPDPLLQPGRGQIIKVDAPWLKNFIITHDLERGIYNSPYIIPGLQAV
TLGGTFQVGNWNEINNIQDHNTIWEGCCRLEPTLKDAKIVGEYTGFRPVRPQVRLEREQ
LRFGSSNTEVIHNYGHGGYGLTIHWGCALEVAKLFGKVLEERNLLTMPPSHL

>BVMO
MAGQTTVDSRRQPPEEVDVLVVGAGFSGLYALYRLRELGRSVHVIETAGDVGGVWYW
NRYPGARCDIESIEYCYSFSEEVLQEWNWTERYASQPEILRYINFVADKFDLRSGITFHTT
VTAAAFDEATNTWTVDTNHGDRIRARYLIMASGQLSVPQLPNFPGLKDFAGNLYHTGN
WPHEPVDFSGQRVGVIGTGSSGIQVSPQIAKQAAELFVFQRTPHFAVPARNAPLDPEFLA
DLKKRYAEFREESRNTPGGTHRYQGPKSALEVSDEELVETLERYWQEGGPDILAAYRDI
LRDRDANERVAEFIRNKIRNTVRDPEVAERLVPKGYPFGTKRLILEIDYYEMFNRDNVH
LVDTLSAPIETITPRGVRTSEREYELDSLVLATGFDALTGALFKIDIRGVGNVALKEKWA
AGPRTYLGLSTAGFPNLFFIAGPGSPSALSNMLVSIEQHVEWVTDHIAYMFKNGLTRSEA
VLEKEDEWVEHVNEIADETLYPMTASWYTGANVPGKPRVFMLYVGGFHRYRQICDEV
AAKGYEGFVLT

>CHexM
MTAQTTHTVDAVVIGAGFGGIYAVHKLHHELGLTTVGFDKADGPGGTWYWNRYPGAL
SDTESHLYRFSFDRDLLQESTWKTTYITQPEILEYLEDVVDRFDLRRHFKFGTEVTSALY
LDDENLWEVTTDHGEVYRAKYVVNAVGLLSAINFPNLPGLDTFEGETIHTAAWPEGKS
LAGRRVGVIGTGSTGQQVITSLAPEVEHLTVFVRTPQYSVPVGNRPVNPEQIAEIKADYD
RIWERAKNSAVAFGFEESTLPAMSVSEEERNRIFQEAWDHGGGFRFMFGTFGDIATDEA
ANEAAASFIRAKVAEIIEDPETARKLMPKGLFAKRPLCDSGYYEVYNRPNVEAVAIKENP
IREVTAKGVVTEDGVLHELDVLVFATGFDAVDGNYRRIEIRGRDGLHINDHWDGQPTSY
LGVSTANFPNWFMVLGPNGPFTNLPPSIETQVEWISDTIGYAERNGVRAIEPTPEAEAEW
TETCTEIANATLFTKGDSWIFGANIPGKKPSVLFYLGGLRNYRAVMAEVAADGYRGFEV
KSAEMVTV

>VAO
MSSSTVLPTPTLSEKHSGIPFRLYEKAQYAKSLILDIATKEQSERKRGPAIPAGVEKTTFL
KALDELSGQLGKENVEINDQPLKDGWYMEHPNTHDAMHVLDEEELVASAVVYPGSTE
EVQKIVLWANKYKIPIFPISIGRNLGYGGAAPRVRGSVVIDLGRRMNKILDINPVDHTCL
VEPGVTFYALYEEIQRRGYKHLWIDCPDLGGGSVLGNTLDRGIGYTVYGDHWACHSGL
EVVLPTGELIRTGMGAMANSSSWQIFPYGYGPMADGLFSQSNYGIVTKLGMTLMRNPG
GHESYLYTFPNEEDLAPLVDIIRPLRIGNILENVAQLRHVVQAIAYSGKPRSSYFQGEGQ
MTDEQVREIARKELKYGDFTWLYYGMSYGPKEIRQYKLDIIHKEFLKIPGARRIDPATLP
KTDYFWSRDRIATGIPDLEELQWVNWYPNGGHIAFSPVSPVRGPDATELWRMARSRAA
EFGHDIFPAFCVGLREMHLIVECVFNRDDPDSRKMALACMRAMIDEAASKGYGEYRTH
LVLMDQIAKTYNFNDHALMKFNERIKDTLDPNGILAPGKSGVWPARYRGRGWEMSGL
SERSEGDGVARDTATRL

>HpaB
MKPEDFRASTQRPFTGEEYLKSLQDGREIYIYGERVKDVTTHPAFRNAAASVAQLYDAL
HKPEMQDSLCWNTDTGSGGYTHKFFRVAKSADDLRQQRDAIAEWSRLSYGWMGRTPD
YKAAFGCALGANPGFYGQFEQNARNWYTRIQETGLYFNHAIVNPPIDRHLPTDKVKDV
YIKLEKETDAGIIVSGAKVVATNSALTHYNMIGFGSAQVMGENPDFALMFVAPMDADG
VKLISRASYEMVAGATGSPYDYPLSSRFDENDAILVMDNVLIPWENVLIYRDFDRCRRW
TMEGGFARMYPLQACVRLAVKLDFITALLKKSLECTGTLEFRGVQADLGEVVAWRNTF
WALSDSMCSEATPWVNGAYLPDHAALQTYRVLAPMAYAKIKNIIERNVTSGLIYLPSSA
RDLNNPQIDQYLAKYVRGSNGMDHVQRIKILKLMWDAIGSEFGGRHELYEINYSGSQDE
IRLQCLRQAQNSGNMDKMMAMVDRCLSEYDQDGWTVPHLHNNDDINMLDKLLK

>DMAM
MSKRVGIIGAGVSGLAAIWCCLEEGLEPTCFERSDDVGGLWKFSDHTEEGRASIYQSVFT
NSSKEMMCFPDFPYPDDYPNYIHHSKLQEYIKTYAQKKDLLRYIQFETLVSGIKKCPSFL
VTGQWVVVTEKDGKQESTIFDAVMICSGHHVYPNLPTDSFPGLDQFRGNYLHSRDYKN
PEAFKGKRVLVIGLGNSGSDIAVELSRLATQVIISTRSASWVMSRVWDDGYPWDMMYV
TRFASFLRNVLPSFISDWLYVQKMNTWFKHENYGLMPLNGSLRKEPVFNDELPSRILCG
TLSIKPSVKEFTETSAVFEDGTMFEAIDSVIFATGYDYSYPFLDETIMKSRNNEVTLFKGIF
PPLMEKPTLAVIGLVQSLGAAIPTADLQAWWAAKVFANSCTLPTTNEMMDDTDEKMG
KKLKCMFSSFFMFGQSQTLQTDYITYVDELGSFIGAKPNIPWLFLTDPRLALEVYFGPCS
PYQFRLMGPGKWDGARNAILTQWNRTVKPTRTRVVSEVQRPHPFYNLLKMLSFPLLLL
AVTLTFY

>PHBH
MKTQVAIIGAGPSGLLLGQLLHKAGIDNVILERQTPDYVLGRIRAGVLEQGMVDLLREA
GVDRRMARDGLVHEGVEIAFAGQRRRIDLKRLSGGKTVTVYGQTEVTRDLMEAREAC
GATTVYQAAEVRLHDLQGERPYVTFERDGERLRLDCDYIAGCDGFHGISRQSIPAERLK
VFERVYPFGWLGLLADTPPVSHELIYANHPRGFALCSQRSATRSRYYVQVPLTEKVEDW
SDERFWTELKARLPAEVAEKLVTGPSLEKSIAPLRSFVVEPMQHGRLFLAGDAAHIVPPT
GAKGLNLAASDVSTLYRLLLKAYREGRGELLERYSAICLRRIWKAERFSWWMTSVLHR
FPDTDAFSQRIQQTELEYYLGSEAGLATIAENYVGLPYEEIE

>KMO_Pf
MTATDNARQVTIIGAGLAGTLVARLLARNGWQVNLFERRPDPRIETGARGRSINLALAE
RGAHALRLAGLEREVLAEAVMMRGRMVHVPGTPPNLQPYGRDDSEVIWSINRDRLNRI
LLDGAEAAGASIHFNLGLDSVDFARQRLTLSNVSGERLEKRFHLLIGADGCNSAVRQAM
ASVVDLGEHLETQPHGYKELQITPEASAQFNLEPNALHIWPHGDYMCIALPNLDRSFTVT
LFLHHQSPAAQPASPCFAQLVDGHAARRFFQRQFPDLSPMLDSLEQDFEHHPTGKLATL
RLTTWHVGGQAVLLGDAAHPMVPFHGQGMNCALEDAVALAEHLQSAADNASALAAF
TAQRQPDALAIQAMALENYVEMSSKVASPTYLLERELGQIMAQRQPTRFIPRYSMVTFS
RLPYAQAMARGQIQEQLLKFAVANHSDLTSINLDAVEHEVTRCLPPLSHLC

>Antiport
MIVVGAGPAGSAAAYHLAKAGVDVLVLEKSEFPREKVCGDGLTPRAVHQLIRMGVDID
GPGWARSRGMRWVAGERQALIEWPKVGKYPDFGLSRSRYDFDDILAGHATGAGARLR
TGAKVTGPLTDRAGRVVGVTAELGPEKRPVEFRSRLVIAADGASARLALALGRQRDER

ALIATAARRYYRSPQRSQEEYLELWADLRGPDGDRLLPGYGWIFPMGDGRVNVGLGAL
RGRKSGQVDLRATMDRWLARTPEEWGLRPDADGHSENAEGRVQSAALPLGFNRHPQY
GRGLLLVGDSGGMVSPWNGEGIAQALEAGEVAAEAVAAALARPEGAGHEEALARYPA
EMARRWGRHYRLGVALSRHLFSRGGFQPLLNRWVMGSPTLLDTLARLLIGLTDDPPAD
TADRVLHTLLHLVPEQSPSV

>Digeran
MKSYKCDVLVVGGGPAGSNAARYAAKGADVILIEKKKEIGAPVQCAEGVSQGIFEKLE
MKQSSRYISSNIEFVKLISPNGTVIRLDGEKVKYWKSGMILDREVFDKAIAKEAARNGAD
FLMKTRFISAKRNSNGVTVTAKQMNEDIQIDAKIVIGADGPPSIVARSLGMDTTVPLRYL
ESGIQYLMMPMEVEPCIELYFGDCYAPGGYSWIFPKGDDQANVGLGVLSVKAKETAQY
YLDKFIERPRFKKAKIVEVNAGAIPVNGPLKEPYMDNLLLVGDAGRFVNPLTGGGIHTAI
ITGKHAGMLAAEAIDKGDYTSNFLKKYNDMWKPDIYDELDKCLKAQEAFMKLTEKDL
DSIADTLRDLKLDTISTISVLKAIVAKNPGLLFKLGKFM

(E) Designing of synthetic strings for *virX*1. Synthetic strings were designed, codon optimised for expression in *E. coli* and purchased from Invitrogen. The ORF of *vir*X1 was designed to be flanked by 30 bp at the 5' and 3' end, identical to pUC19 multiple cloning site (MCS). This way the synthetic string can be used as a primer for the amplification to the pUC19 blackbone on an initial PCR amplification. The *vir*X1 ORF was designed to start from nucleotide 45 and end at 1631, with protected sites 26-31 *Eco*RI, 36-41 *Nde*I, 1632-1637 *Sal*I and 1639-1644 *Hind*III. The areas that were protected from codon optimisation were nucleotides 1-41 and 1632-1671, which were the flanking regions identical to the pUC19 backbone. The motifs designed to be avoided from the areas to be codon optimised were *Eco*RI, *Nde*I, *Sal*I, *Hind*III, *Xma*I, *att*B1, *Nco*I, *Xho*I and attB2. The following sequences was added at the 5'-end and 3'-end of the synthetic string 5'-*acgacgttgtaaaacgacggccagtgaattcg*ccacat**atg** and gtcgac**t*aa***gcttggcgtaatcatggtcatagctgttcc-3'. The DNA sequence of the synthetic string is found bellow.

ACGACGTTGTAAAACGACGGCCAGTGAATTCGCCACATATGGTCAAAATTGAAAGC
GTTGCAATTGTTGGTGGTGGTAGCAGCGGTTGGATGACCGCAGCAGCACTGAGCAA
ACTGTGTCCGCAGCTGGAAATTGCACTGATTGAAGATCCGAACATTAAAACCGTTGG
TGTTGGTGAAAGCACCCTGGGTCATTTTAACAAATTTCTGCATCTGCTGGATCTGAA
AGATGAAGATTGGATGCCTGCATGTAATGCCACCTATAAAAACAGCATTCGCTTTAC
CAATTTTCGCGAAGGTAAAGGCGAAGTGTTTGAATATCCGTTTGGTCCGAGCCTGGA
TGTTAGCTTTTTTAGCCAGACCGATGGCATTAATACCTGGGGTAAACTGGCAAACAA
ATACCCGGAAGATTTTCCGCCTGAAACCTTTGCACGTTTTGTTAATAGCAATACCTAT
CTGGCCGAACATAATCGTCTGACCCGTAACAAAGACAACAAAATCCCGAACTTTAA
CTTCGATTGGGATACCGCCTATCACATTGATGCAGAACTGTTTGGTCAGTACCTGAA

AGAAAAAATTGCCCTGCCGAATGGCGTTAAACACATTCAGGGTAAAGTTACCGGTT
ATCAGAAAGAAAGCCCGAACAACCACAACTTCAAATATATCATCCTGGATCAAGAA
ACCGCCATTTTTGCCGATCTGTATATTGATTGCACCGGCTTTAAAAGCCTGCTGCTGG
GTGAATTTATGGGTGAAGCATTTAGCCCGTTTAGCAAAAAACTGGCCAATGATAAA
GCAATGGCAACCCGTATTCCGTATGAAAATCGTGAAGAAGAAATGCACAACGTTAC
CGATTGTCATGCCATGAAAAATGGTTGGGTTTGGAACATTCCGCTGTGGAATCGTAT
TGGCACCGGTTATTGTTATAGCAGCCGTTTTGTGAGCAAAGATGATGCCGAAGCAGA
ATTTCGTGAACATCTGGGTGAACGTGGTAAAGATGCAAAAATCTTTCACATCGATAT
TGGTCATGGTAAACGTACCCGTGCATGGGTGAATAATTGTGTTGGTATTGGTCTGAG
CTATGGCTTTATTGAACCGCTGGAAAGCACCGGTCTGCTGACCACCCATGAAAATAT
TGAAAATCTGGTGTACCTGATTAACCAGCGTGATGGTTATGTTACCCAGGCAGAACG
TGATGGCTTTAACTATACCTGTGATCATCAGATCGATAGCTTTAGCGATTTTGTGGCA
ATGCATTATGCATATAGCATGCGTACCGATACCCCGTATTGGAAATGGTGTACACAG
ATGTGCAATTATATGCCGGAAAGCATGGGTCCGCATCGTCAGAAACAGAGCACCTG
GCAGGATCTGAGCACCGATACCATTGGCCTGAATACCTGGCATATTAATCATAACGG
CATCAGCTTTATTATCGCAGGTCATGGTCTGCGTCCGCAGAGCTATGATAAACTGAG
CGAAGTTCTGCTGAAACGCAATAATGAAAGCGATTATTACTATGAAGATATCCGCA
AAGACTGGCTGAAACACTATGAAAGTATGGTCGAATATGTTAAAACCCTGCCGACC
CATTATGAATTTCTGCGTGATGAAATTTATGGCAGCGCAGAAGTCGACTAAGCTTGG
CGTAATCATGGTCATAGCTGTTTCC

(F) Cloning of *virX1* into pUC19 and subcloning into pSG181 expression vector. All components were mixed on ice, in 50 µl reactions containing 37 µl sterile water, 5 µl 10X buffer E (Promega), 5 µl pUC19 (100 ng/µl), 1.5 µl of *Eco*RI, 1.5 µl of *Hind*III. The mixture was incubated at 37 °C over night or until complete digestion monitored by gel. FastAP alkaline phosphatase (1 µl) (Thermo) was added and the reaction incubated for a further 60 min. The mixture was purified with PCR clean up kit (Promega) and the product was analysed by agarose gel electrophoresis. Once the vectors had been prepared, PCR amplification was performed. The following were mixed on ice in 100 µl reactions that were later divided into 4 × 25 µl aliquots. The components included 67 µl of sterile water, 20 µl of 5X Phusion HF buffer, 2 µl of pUC19 (digested by *Eco*RI/*Hind*III, FastAP treated) as 5-20 ng/µl, 2 µl of synthetic gene as 50 ng/µl in sterile water, 8 µl of dNTPs (2.5 mM stock) and 1 µl of Phusion DNA polymerase. In parallel a negative control was prepared and analysed which did not include synthetic gene. The PCR conditions included 30

s of 98 °C as an initial denaturation step, followed by 30 cycles of denaturation for 15 s at 98 °C, gradient annealing temperature from 60-72 °C for 30 s and an extension for 3 min at 72 °C. After completion of the 30 cycles, PCR was extended for 10 min at 72 °C and the samples were held at 12 °C. After completion of PCR, 0.5 µl of *Dpn*I was added into each PCR reaction and the mixtures were again incubated at 37 °C overnight. 5 µl of each reaction and control were mixed together (control separately) and 5 µl of this mixture was used to transform 100 µl of chemically competent DH10B-T1 cells. After transformation, 4 colonies were picked to inoculate 10 ml of LB and the appropriate antibiotic. Plasmids were isolated using the Promega kit, and first verified by double restriction digest. Sequence was then validated by sequencing at ATCC. After confirmation by sequencing that the resulting pUC19 construct was carrying the *vir*X1 gene, *vir*X1 was subcloned into pSG181 (parent vector pET28) containing a T7lac IPTG-inducible promoter, kanamycin resistance and a TEV cleavable octa-histidine N-terminal tag (see section bellow for plasmid map and cloning flow chart).

(G) Site directed mutagenesis and cloning of VirX1 mutants

K79A, K79R, F353A, S359A and P356A were constructed using a one-step site-directed mutagenesis method described by Liu and Naismith[1] using the primers below. Each PCR reaction (final volume 50 µl) contained 40 ng of plasmid pDSG407 (encoding VirX1) as a template, 1 µM primer pair, 200 µM dNTPs and 0.5 units of Phusion DNA polymerase. The PCR cycles were initiated at 98 °C for 30 s to denature the template DNA, followed by 12 amplification cycles. Each amplification cycle consisted of 98 °C for 10 s, 57 °C for 30 s and 72 °C for 3.30 min (about 1 kb per 30 s for Phusion DNA polymerase). This was followed by 3 additional cycles. Each cycle consisted of 98 °C for 10 s, 45 °C for 30 s and 72°C for 3.30 min. The PCR cycles were finished with an extension step at 72°C for 10 min. The PCR products were treated with 10 U of DpnI at 37 °C for 4 h. 5 µl of each PCR reaction was analyzed by agarose gel electrophoresis. 5 µl of each PCR product was transformed respectively into *E. coli* DH10B chemically competent cells by heat shock. The transformed cells were spread on a Luria-Bertani (LB) plate containing kanamycin and incubated at 37 °C overnight. Five colonies from each plate were grown and the plasmid DNA was isolated the mutants were verified by DNA sequencing. Protein production, purification and assessment of VirX mutants identical to VirX1 sections.

Primers	Sequences

K79A F	CCACCTATGCGAACAGCATTCGCTTTACCAATTTTCGC
K79A R	CTGTTCGCATAGGTGGCATTACATGCAGGCATCCAATCTT
K79R F	CCACCTATCGCAACAGCATTCGCTTTACCAATTTTCGC
K79R R	CTGTTGCGATAGGTGGCATTACATGCAGGCATCCAATCTT
F353 F	TATGGCGCGATTGAACCGCTGGAAAGCACCGGTCTGC
F353 R	CGGTTCAATCGCGCCATAGCTCAGACCAATACCAACACAA
S359 F	CTGGAAGCGACCGGTCTGCTGACCACCCATGAAAAT
S359 R	GACCGGTCGCTTCCAGCGGTTCAATAAAGCCATAGCTCA
P356 F	TTTATTGAAGCGCTGGAAAGCACCGGTCTGCTG
P356 R	CCAGCGCTTCAATAAAGCCATAGCTCAGACCAATACCA

(H) Cloning and production of PrnA.

The gene *prnA* from *Pseudomonas fluorescence* was subcloned from pSG28 (pET28a(+) vector) into pSG181 containing a TEV-cleavable octahistidine tag at the N-terminus. This was done using *Nde*I and *Hind*III double digest and ligation with T4 promega ligase. Successful cloning was confirmed by DNA sequencing. pDSGpf was transformed into BL21 DE3 chemically competent cells by heat shocking. Protein production at various IPTG concentration (0, 0.1 mM, 0.5 mM and 1 mM) and temperatures (16, 21 and 25 °C) revealed that 0.5 mM and 21 degrees °C for 20 h gave more soluble protein between those conditions. The cultures were scaled up to 14 L using these conditions for protein isolation and purification, such as with VirX1 procedures, yielding 0.5 mg/L of pure his-free PrnA.

(I) Flavin reductase assays. The *prnF* gene from *Pseudomonas fluorescence* Pf-5 encoding for the flavin reductase of the pyrrolnitrin biosynthesis, had been previously cloned (and reported) within the Goss group into a pET-21a vector containing a C-terminal 6xHis-tag, between *Nde*I and *Xho*I sites.

MNAATETKVHDLLDAEGRDVRDARELRNVLGQFATGVTVITTRTADGRNVGVTANSFS
SLSLSPALVLWSLARTAPSLKVFCSASHFAINVLGAHQLHLSEQFARAAADKFAGVAHS
YGKAGAPVLDDVVAVLVCRNVTQYEGGDHLIFIGEIEQYRYSGAEPLVFHAGQYRGLG
SNRAESVLKHELEHHHHHH

The PrnF enzyme was produced and purified as per sections K and L and was stored in storage buffer containing 40 mM Hepes, 100 mM NaCl, 10% v/v glycerol at pH 7.8. The activity of PrnF was assessed by using 1 μM PrnF was added to a 100 μl assay containing 200 μM NADH, 30 μM FAD, 50 mM NaCl, 20 mM Tris-HCl at pH 7.5. PrnF was added half way through the time course and oxidation of NADH to NAD+ was monitored by the decreasing absorbance at 340 nm over 30

min. This was monitored using a plate reader and 96 well Greiner cellstar clear plates. An assay containing no FAD was used as a blank.

(J) Protein production screening of VirX1. The plasmid (pDSG407) harboring VirX1 was transformed into BL21 (DE3) chemically competent cells. A single colony was picked from the transformation plate and was used to inoculate a 10-ml culture in LB containing kanamycin, which was incubated overnight (37 °C, 200 rpm, 2.5 cm radius or rotor, 1 rcf). The starter culture was diluted 100-fold with fresh LB containing kanamycin and incubated as before until the cell density had reached an OD_{600} of 0.4-0.6. The culture was then split in 10 ml cultures into sterile plastic 50 ml falcon tubes. These tubes were induced at varied concentrations of IPTG including 0 mM (non-induced control), 1 mM, 0.5 mM and 0.1 mM and incubated at either 28 °C for 16 h, 16 °C for 20 h or 37 °C for 4 h. The pellets were then collected and lysed with lysozyme (2 mg ml^{-1} of lysis buffer 1: Tris-HCl 50 mM, NaCl 25 mM, 5% Glycerol, pH=7.4) for 2h in ice. Following centrifugation to separate the crude lysate from cell debris, the lysate was subjected to SDS-gel electrophoresis. Levels of protein production were assessed visually comparing the size of the band in SDS page relating to the VirX1. Relative quantification of each band could be done using Odyssey Imaging studio.

(K) Cell culturing for protein production and IMAC NiNTA purification of VirX1. As in previous section, a single colony was picked from a fresh transformation plate of pDSG407 plasmid in BL21 (DE3) cells and was used to inoculate a 100 ml culture overnight. 5 ml of the starter was used to inoculate 10 × 500 ml of fresh LB containing kanamycin in 2 L flasks, which were incubated (37 °C, 200 rpm, 2.5 cm radius or rotor, 1 rcf) until the cell density had reached an OD_{600} of 0.4-0.6. The temperature was then lowered to 16 °C and 0.5 mM IPTG was added to each culture for induction. The cultures were incubated for further 20 h (16 °C, 200 rpm, 2.5 cm radius or rotor, 1 rcf). The cells were harvested by centrifugation (using JLA 8.1000 rotor, 2000*g*, 20 min, 4 °C). Cell pellets were gently suspended in lysis buffer (5 ml/g of pellet) containing 50 mM NaH_2PO_4, 250 mM NaCl at pH 7.8. The portions of cell suspension were subjected to sonication on ice using the KE 76 titanium tapered tip for volumes of 50-100 ml per time, sonicating each sample for 2 cycles of 6 min at 40% power and 20% duty cycle. Samples were cooled on ice for at least 5 min between cycles. Cell lysate containing overexpressed 8Histag-VirX1 was harvested by centrifugation twice (using JA rotor 25.50, at 35,000 g, 45 min, 4 °C). The crude lysate was

then mixed with appropriate amount (as per manufacturer's instructions) of Thermo HisPurTM NiNTA resin (preequilibrated with same lysis buffer used above) at 4 °C on a rocker for 1 h. The lysate containing the nickel resin was then passed through a Bio-Rad Econo-Column fritted glass column which had been pre-washed with the lysis buffer. This was followed by sequential washes with lysis buffer containing 20 mM, 30 mM, 40 mM and 50 mM imidazole while monitoring using a mini-bradford assay (5 µl of eluate was added to 250 µl of 1X Bradford reagent until no blue colour was observed between washes) until no further protein was eluted between steps. The histidine-coordinated protein to the nickel beads was eluted using lysis buffer containing 250 mM imidazole and all fractions were collected in 1.5 ml eppendorf tubes and were monitored using a mini-bradford assay as above. The fractions containing protein were all combined. PD10 columns were used for exchanging the elution buffer containing 250 mM imidazole for the appropriate storage buffer (as described in section A).

(L) TEV cleavage of octa-histidine terminal tag. 8Histag-VirX1 and TEV were mixed in a 5:1 ratio (VirX1:TEV) and left at 4 °C in ice. After 12 h, a small sample of the VirX1-TEV mixture was removed and analysed by SDS-PAGE (12% resolving gel). Once full cleavage of the Histag had been achieved with the TEV protease, the protein sample was subjected to a second IMAC purification. The flow through containing the untagged VirX1 was collected.

(M) Size exclusion chromatography. The protein sample was concentrated to 5 ml using Merck Milipore Amicon® Ultra-15 10 kDa MWCO centrifugal filters and subjected to size exclusion chromatography using the GE AKTA pure FPLC system, typically equipped with the GE HiLoad 16/600 Superdex 200 pg column equilibrated and eluted with storage buffer containing halogen source (or no halogen source) of choice, while monitoring elution by UV absorbance at 280 nm. In a typical purification, the injection loop was washed with 5 volumes of GF buffer and the first 40 ml of eluate following injection were discarded. After this, the eluate was collected in fractions of 3 ml over an appropriate fractionation range for the protein of interest.

(N) Size exclusion chromatography-multiple angle laser light scattering (SEC-MALS). SEC-MALS experiments were performed using a Superdex 200 10/300 Increase column (GE Healthcare) and an AktaPure 25 System (GE Healthcare). The sample (100 µl, 2.5 mg/ml) was loaded onto the gel filtration column and eluted with one column volume (24 ml) of buffer, at a flow rate of 0.7 ml/min. The eluting protein was monitored using a DAWN HELEOS-II 18-angle light scattering detector (Wyatt Technologies) equipped with a WyattQELS dynamic light scattering module, a U9-M UV/Vis detector (GE Healthcare), and an Optilab T-rEX refractive index monitor (Wyatt Technologies). Data were analysed by using Astra (Wyatt Technologies) using a refractive increment value of 0.185 ml/g.

(O) Analytical ultra-centrifugation (AUC). Sedimentation velocity experiments were performed at 50000 rpm, using a Beckman Optima analytical ultracentrifuge with an An-60Ti rotor at 20 °C. Data were recorded using the absorbance (at 280 nm) and interference optical detection systems. The density and viscosity of the buffer was measured experimentally using a DMA 5000M densitometer equipped with a Lovis 2000ME viscometer module. The partial specific volume of the protein was calculated using SEDFIT from the amino acid sequence. Data were processed using SEDFIT, fitting to the c(s) model. Figures were made using GUSSI.

(P) Steady state kinetics. All assays were run in 200 µl reaction in 96 V-shaped plates. Initial rates were calculated based on substrate consumption at 0, 2, 4, and 6 min after addition of substrate to all other components. Substrate concentration was varied from 10 µM to 4 mM depending on the substrate and enzyme. The conditions for TEV-cleaved octahistidine tag-free VirX1 (1µM) or TEV-cleaved octahistidine tag-free VirX1 PrnA (1 µM) included an excess of the partner flavin reductase 10 µM PrnF, 2.5 mM NADH, 10 mM NaX (X is either Cl, Br, or I) 10 µM FAD and phosphate buffer pH 7.4 plus substrate resuspended in DMSO. Substrates were added at 30 second intervals and the reaction terminated with the addition of formic acid and equal amount of methanol. All assays were performed in triplicate. Analysis by UPLC against standard curves of starting material and calculation of initial rates followed by non-linear regression.

(Q) Biohalogenation assays analytical scale. 100 µl reactions were carried out using 10 µM of purified VirX1, 1 µl of each compound (10 mg in 1 ml of DMSO stock) from a 400-compound library, 10 mM NaX (where X was either NaBr or NaI or NaCl), 1 mM NADH, 10 µM

FAD and 40 mM Hepes at pH at pH 7.4. For a full 96 well plate a 10 ml mastermix was prepared, which contained everything but NADH and the substrate. Each substrate was added to a specific position in the well using a multichannel pipette. The mastermix was added in a sterile plastic container and was added in each well of the plate. NADH was added last in each well. After incubation for 90 min at 28 °C, the assays were placed on ice for 20 min, and the reactions were quenched by addition of the same amount of methanol or acetonitrile, using a multichannel electronic pipette, followed by centrifugation at 5000 g for 45 min and transfer into new and clean 96 well plates. Same exact conditions were used for mutant VirX1 activity screening. Assays with pure PrnA included 10 μM PrnA, 1 μM PrnF, 2.4 mM NADH, 10 μM FAD, 25 mM NaX, 10 mM KPhosphate buffer pH 7.2 in 200 μl reactions.

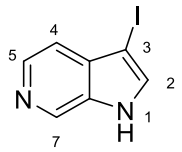
(R) UPLC analysis of biohalogenation reactions. Before commencing the assaying of the halogenases against these compounds, the compounds were first diluted 100-fold, 500-fold and 1000-fold with a 1:1 H_2O/MeOH solution. The profiles of each compound in the library were analysed by UHPLC using a Waters Acquity UHPLC system equipped with quaternary solvent manage, autosampler, PDA detector and FLR detector. The compounds were analysed using a C18 Waters Acquity BEH C18 1.7 μm 2.1 × 50 mm column at 45 °C monitoring with PDA detector at all wavelength ranges between 220-550 nm. For each compound, the following data were gathered in an excel spreadsheet: retention time, number of major peaks, λmax, peak shape, presence in 100-fold, 500-fold and 1000-fold dilution and finally the overall level of purity. Subsequently, the serial dilutions defined the limit of detection, so that the enzymatic assay could be planned accordingly. It was expected that the yields for non-native substrates would be low and as such it was important that the presence of halogenated products can be detected using this analysis. Therefore, the amount of substrate used for the halogenation assays was considered based on the limit of detection. As initially there was no positive control for the assays as the activity was unknown, each plate was divided so that one side would contain the same compound but without the enzyme. For example, if A1 contained compound X with the halogenase and cofactors, A7 would contain the same components, without the halogenase. All the samples were screened for the appearance of new peaks with a higher retention time than that of the known starting material.

(S) LC-HRMS analysis of biohalogenation reactions. When a novel product was identified by the UPLC analysis, it was be further verified by LCMS using a C18 RP-HPLC system and an ESI+ ion source coupled to a Thermo Orbitrap Velos Pro mass spectrometer capable of both high-resolution mass measurement and rapid MS^N fragmentation analysis. For the LCMS analysis, the first criterion was that the halogenated product should not be present in the negative control sample. The second criterion was that the halogenated compound had to agree with the calculated exact mass up to at least 3 decimal points. The third criterion required the presence of an isotope pattern in accordance with the natural abundance of the natural abundance of ^{79}Br: ^{81}Br was giving rise to the characteristic 1:1 ratio of n:n+2 peaks in a mass spectrum. It is important to note, that if one compound was halogenated once, the possibility of incorporation of more than one halogen was always assessed. Elemental composition analysis was used to calculate the best matching chemical formula for the selected mass. This formula was used to display the theoretical *m/z* values for the chemical formulas that the application can determine, which can be used to compare between the actual result. The fourth criterion was that the experimental and the theoretical values should agree. Finally, each n: n+2 isotope masses on the mass spectrum should be integrated to determine each of those masses would link with the same peak eluting at the same retention time on the TIC or EIC (extracted ion count). If not, they could be occurring by chance, and they are not isotope patterns of the same molecule. All positive hits were then repeated for iodination reactions. Bromide was used rather than iodide for initial screenings respective to the confidence over the n: n+2 of the isotope pattern of bromine.

(T) Isolation of halogenated products from scaled up biotransformations. For isolation of halogenated products, the biotransformations were scaled up to 50 ml reactions, with 0.5- 0.7 mM of substrate using the same conditions as with the 100 µl assays. After incubation for 12 h, the biotransformation mixture (~50 ml) was shaken with chloroform or ethyl acetate (30 ml) for 30 min and filtered over celite to remove precipitated proteins. The aqueous layer was further extracted with chloroform or ethyl acetate (3 × 20 ml). Combined organic extract was washed successively with water (2 × 30 ml), brine (30 ml) and dried using anhydrous sodium sulphate, filtered and solvent was evaporated on a rotary evaporator (bath temperature <35 °C). Crude mixture was purified by column chromatography as specified. Products were characterised and assigned based on 1D and 2D NMR.

Isolation of 3-Iodo-6-azaindole (1) from biotransformation

Purification by flash column chromatography (silica) using a 0-10% gradient of methanol in dichloromethane afforded 3.2 mg of the product.



3-iodo-1*H*-pyrrolo[2,3-*c*]pyridine

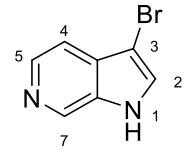
^{1}H NMR (500 MHz, CD_3OD) δ 8.69 (s, 1H, H7), 8.19 (d, *J* = 5.4 Hz, 1H, H5), 7.71 (s, 1H, H2), 7.41 (d, *J* = 5.4 Hz, 1H, H4).

^{13}C NMR (126 MHz, CD_3OD) δ 138.6 (C5), 136.9 (Cq), 135.8 (C2), 134.9 (C7), 116.3 (C4), 55.5 (C3-I) ppm. One of the quaternary carbons was not observed on 13C or HMBC NMR.

HRMS (FTMS-ESI$^+$) *m/z* $C_7H_6IN_2$ $[M+H]^+$ calculated 244.9570, found 244.9561.

Isolation of 3-Bromo-6-azaindole (1-Br) from biotransformation

Purification by flash column chromatography (silica) using a 0-10% gradient of methanol in ethyl acetate afforded 2.1 mg of the product.



3-bromo-1*H*-pyrrolo[2,3-*c*]pyridine

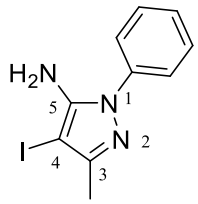
^{1}H NMR (500 MHz, CD_3OD) δ 8.72 (s, 1H, H7), 8.18 (d, *J* = 5.7 Hz, 1H, H5), 7.66 (s, 1H, H2), 7.53 (dd, *J* = 5.7, 0.9 Hz, 1H, H4) ppm.

^{13}C NMR (126 MHz, CD_3OD) δ 137.3 (C5), 133.9 (C2), 132.9 (Cq), 131.9 (Cq), 129.3 (C7), 113.1 (C4), 88.9 (C3-Br) ppm.

HRMS (FTMS-ESI$^+$) *m/z* $C_7H_6BrN_2$ $[M+H]^+$ calculated 196.9709, found 196.9708.

Isolation of 5-Amino-4-iodo-3-methyl-1-phenylpyrazole (2) from biotransformation

Purification by flash column chromatography (silica) using a 0-20% gradient of ethyl acetate in petroleum ether afforded 1.1 mg of the product. In NMR solution, the compound showed discoloration raising stability concerns. Hence, NMR data was obtained on the solution containing traces of petroleum hydrocarbons.



4-iodo-3-methyl-1-phenyl-1*H*-pyrazol-5-amine

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.54 (dd, *J* = 8.5, 1.3 Hz, 2H, ArH), 7.52 – 7.47 (m, 2H, ArH), 7.40 – 7.34 (m, 1H, ArH), 3.96 (s, 2H, NH_2), 2.26 (s, 3H, CH_3).

^{13}C NMR (126 MHz, $CDCl_3$) δ 150.1 (C3), 145.5 (C5), 138.7 (Ph-C), 129.6 (Ph-CH), 127.6 (Ph-CH), 123.65 (Ph-CH), 45.5 (C4-I), 14.2 (CH_3).

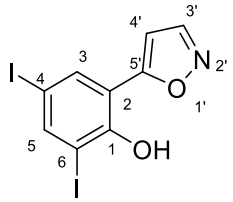
HRMS (FTMS-ESI$^+$) *m/z* $C_{10}H_{11}IN_3$ $[M+H]^+$ calculated 299.9992, found 299.9986

(U) General procedure for model reactions with HOBr or HOI. HOBr and HOI were prepared by adopting published procedures.[2] Sodium hypochlorite (from commercial bleach, Fisher) was diluted and standardised by measuring absorption at 292nm at pH 12 (ε_{292} 350 $M^{-1}cm^{-1}$) then diluted in phosphate buffer (50 mM, pH 7.8). HOBr was prepared by adding HOCl (100 mM, 200 µl) to a 1.5-fold molar excess of NaBr (100 mM, 300 µl). Its concentration was determined by measuring absorbance at pH 12 (ε_{329} 332 $M^{-1}cm^{-1}$). HOI was prepared in small volumes by mixing by adding HOCl (100 mM, 20 µl) to a 1.5-fold molar excess of sodium iodide (100 mM, 30 µl), and used within two min of mixing. The initial concentration of HOI was based on the assumption of 100% conversion of HOCl to HOI.[2] Model reactions were carried out by mixing HOBr or HOI (1 or 5 equiv.) with appropriate 6-azaindole or 4-(1-imidazolyl)-phenol (1 mM) in phosphate buffer (50 mM, pH 7.8). Reactions was stirred at room temperature in dark. Aliquots were diluted with 50% *v*/*v* aqueous methanol and analysed by LCMS.

(V) Monochlorodimenone (MCD) assay for hypoalous acid detection. In order to test for chemically reactive and enzymatically produced hypoalous acid in solution, the following solution was used, 10 µM VirX1 (or mutants of VirX1 depending on the reaction or PrnA), 1µM PrnF, 10 µM FAD, 2.5 mM NADH, 10 mM NaX, 50 µM monochlorodimedone (MCD) with and without presence of 6-azaindole. The reaction was analyzed by UPLC at 272 nm. As a control, HOI, HOBr and HOCl were generated as per section U, to demonstrate that MCD can be reduced in presence of hypoalous acids and the halogenated product does not absorb anymore at 272 nm.

(W) General procedure: Synthesis of iodinated standards. In a screw-cap vial, appropriate substrate (0.1 mmole) was suspended in acetonitrile or dichloromethane (1 ml) and stirred for 5 min. Hexafluoroisopropanol (0.1 ml) and *N*-iodosuccinimide (NIS, 23 mg, 0.1 mmole) were successively added. The reaction was stirred in dark, at room temperature for 18 h. Finally, reaction mixture was diluted with ethyl acetate (5 ml) and water (5 ml). Aqueous layer was extracted with ethyl acetate (2 × 5 ml). Combined organic extract was washed with diluted sodium thiosulfate solution, water and brine. Organic layer was dried with anhydrous sodium sulphate, filtered and solvent was removed by rotary evaporation. Crude mixture was purified by column chromatography as specified.

2-(5-Isoxazolyl)-4,6-diiodophenol (3)



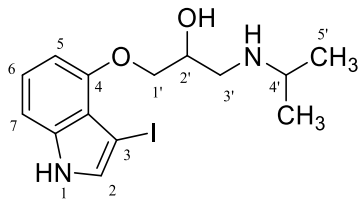
Purification was achieved by flash column chromatography (silica) using a 2-50% gradient of ethyl acetate in hexanes.

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.37 (d, *J* = 1.8 Hz, 1H, H4'), 8.21 (d, *J* = 2.0 Hz, 1H, H3), 8.04 (d, *J* = 2.0 Hz, 1H, H5), 6.87 (d, *J* = 1.8 Hz, 1H, H3'), 6.26 (s, 1H, OH).

^{13}C NMR (126 MHz, $CDCl_3$) δ 163.4 (C5'), 151.2 (C1), 151.0 (C4'), 146.7 (C5), 136.7 (C3), 116.1 (C2), 103.5 (C3'), 88.6 (C6-I), 83.2 (C4-I).

HRMS (FTMS-ESI$^+$) *m/z* $C_9H_6I_2NO_2$ $[M+H]^+$ calculated 413.8482, found 413.8488.

3-Iodo-pindolol (4)



Purification was achieved by reverse-phase chromatography (C-18 column) using a 5-95% gradient of acetonitrile in water.

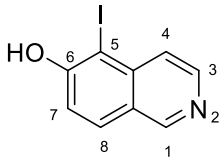
^{1}H NMR (500 MHz, CD_3OD) δ 7.21 (s, 1H, H2), 7.09 – 7.01 (m, 2H, H6-H7), 6.54 (dd, *J* = 7.0, 1.1 Hz, 1H, H5), 4.26 (dt, *J* = 7.5, 4.1 Hz, 1H, H2'), 4.14 (dd, *J* = 9.4, 4.7 Hz, 1H, H1'), 4.01 (dd,

J = 9.3, 6.9 Hz, 1H, H1'), 3.28 (dd, *J* = 12.1, 3.1 Hz, 1H, H3'), 2.96 (dt, *J* = 10.9, 5.5 Hz, 1H, H4'), 2.86 (dd, *J* = 12.0, 9.0 Hz, 1H, H3'), 1.16 (dd, *J* = 6.2, 3.5 Hz, 6H, 2×CH_3).

^{13}C NMR (126 MHz, CD_3OD) δ 152.0 (O-C4), 138.0 (C_q), 128.9 (C2), 122.6 (C6), 117.8 (C_q), 104.9 (C5), 100.2 (C7), 69.9 (C1'), 68.0 (C2'), 49.9 (C3'), 48.7 (C4'), 48.1 (C3-I)*, 21.0 (C5'), 20.9 (C5'). (*overlapped by CD_3OD, determined by HMBC correlation)

HRMS (FTMS-ESI$^+$) *m/z* $C_{14}H_{20}IN_2O_2$ $[M+H]^+$ calculated 375.0564, found 375.0554.

5-Iodo-isoquinolin-6-ol (6)



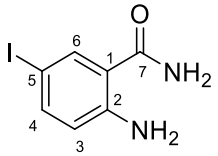
Product precipitated during aqueous work up, which was collected by filtration. The solid was washed successively with water, dichloromethane and dried under vacuum.

^{1}H NMR (500 MHz, DMSO-d_6) δ 9.03 (s, 1H, H1), 8.46 (d, *J* = 6.0 Hz, 1H, H3), 8.00 (d, *J* = 8.8 Hz, 1H, H8), 7.70 (d, *J* = 6.0 Hz, 1H, H4), 7.35 (d, *J* = 8.8 Hz, 1H, H7).

^{13}C NMR (126 MHz, DMSO-d_6) δ 159.8 (C6), 152.3 (C1), 145.1 (C3), 139.3 (Cq), 130.7 (C8), 124.8 (Cq), 122.7 (C4), 119.3 (C7), 82.2 (C5-I).

HRMS (FTMS-ESI$^+$) *m/z* C_9H_7INO $[M+H]^+$ calculated 271.9567, found 271.9559.

5-Iodo-anthranilamide (7)



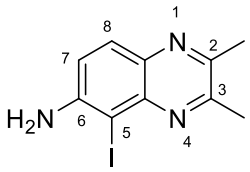
Purification was achieved by flash column chromatography (silica) using a 2-50% gradient of ethyl acetate in hexanes.

^{1}H NMR (500 MHz, CD_3OD) δ 7.82 (d, *J* = 2.1 Hz, 1H, H6), 7.43 (dd, *J* = 8.7, 2.1 Hz, 1H, H4), 6.58 (d, *J* = 8.7 Hz, 1H, H3).

^{13}C NMR (126 MHz, CD_3OD) δ 171.7 (C7), 149.4 (C2), 140.4 (C4), 136.6 (C6), 118.9 (C3), 116.4 (C1), 74.3 (C5-I).

HRMS (FTMS-ESI$^+$) *m/z* $C_7H_8IN_2O$ $[M+H]^+$ calculated 262.9676, found 262.9675.

5-Iodo-2,3-dimethylquinoxalin-6-amine (8)



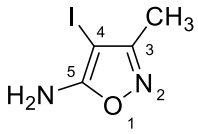
Purification was achieved by flash column chromatography (silica) using a 5-50% gradient of ethyl acetate in hexanes.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.73 (d, *J* = 8.9 Hz, 1H, H8), 7.16 (d, *J* = 8.9 Hz, 1H, H7), 4.72 (s, 2H, NH_2), 2.75 (s, 3H, CH_3), 2.68 (s, 3H, CH_3).

^{13}C NMR (126 MHz, $CDCl_3$) δ 154.4, 149.6 (C2, C3), 148.3, 142.1, 136.0, 129.1 (C8), 119.1 (C7), 83.6 (C5-I), 23.2, 22.1 (2×CH_3).

HRMS (FTMS-ESI$^+$) *m/z* $C_{10}H_{11}IN_3$ $[M+H]^+$ calculated 299.9992, found 299.9984.

5-Amino-4-iodo-3-methylisoxazole (9)



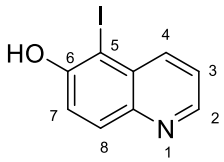
Purification was achieved by flash column chromatography (silica) using a 5-50% gradient of ethyl acetate in hexanes.

^{1}H NMR (500 MHz, $CDCl_3$) δ 4.62 (s, 2H, NH_2), 2.17 (s, 3H, CH_3).

^{13}C NMR (126 MHz, $CDCl_3$) δ 168.0 (C5), 162.2 (C3), 33.6 (C4-I), 12.4 (*C*H_3).

HRMS (FTMS-ESI$^+$) *m/z* $C_4H_6IN_2O$ $[M+H]^+$ calculated 224.9519, found 224.9511.

5-Iodoquinolin-6-ol (10)



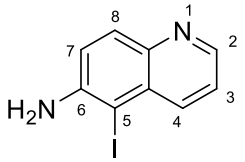
Purification was achieved by flash column chromatography (silica) using a 0-10% gradient of methanol in dichloromethane.

^{1}H NMR (500 MHz, CD_3OD) δ 8.63 (dd, *J* = 4.3, 1.4 Hz, 1H, H2), 8.50 – 8.44 (m, 1H, H4), 7.89 (d, *J* = 9.1 Hz, 1H, H7), 7.52 (dd, *J* = 8.6, 4.3 Hz, 1H, H3), 7.43 (d, *J* = 9.1 Hz, 1H, H8).

^{13}C NMR (126 MHz, CD_3OD) δ 156.2 (C6), 147.1 (C2), 143.2 (Cq), 139.3 (C4), 131.5 (Cq), 129.4 (C7), 122.5 (C3), 120.1 (C8), 81.6 (C5-I).

HRMS (FTMS-ESI$^+$) *m/z* C_9H_7INO $[M+H]^+$ calculated 271.9567, found 271.9555.

6-Amino-5-iodoquinoline (11)



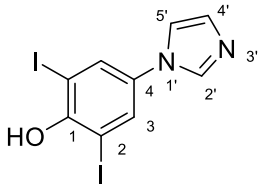
Purification was achieved by flash column chromatography (silica) using a 2-50% gradient of ethyl acetate in hexanes.

^{1}H NMR (500 MHz, CD_3OD) δ 8.46 (dd, *J* = 4.3, 1.4 Hz, 1H, H2), 8.28 (d, *J* = 8.6 Hz, 1H, H4), 7.74 (d, *J* = 9.1 Hz, 1H, H7), 7.43 (dd, *J* = 8.6, 4.3 Hz, 1H, H3), 7.34 (d, *J* = 9.0 Hz, 1H, H8).

^{13}C NMR (126 MHz, CD_3OD) δ 148.0 (C6), 145.3 (C2), 142.4 (Cq), 138.0 (C4), 131.5 (Cq), 128.8 (C7), 122.5 (C3), 120.6 (C8), 77.8 (C5-I).

HRMS (FTMS-ESI$^+$) *m/z* $C_9H_8IN_2$ $[M+H]^+$ calculated 270.9727, found 270.9718.

4-(1H-imidazol-1-yl)-2,6-diiodophenol (12)



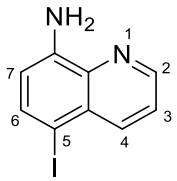
Di-iodo product precipitated during aqueous work up, which was collected by filtration. The solid was washed successively with water, ethyl acetate and dried under vacuum. Organic washings contained a mixture of unreacted starting material with small amounts of mono- and di-iodinated compounds which could not be separated.

^{1}H NMR (500 MHz, DMSO-d_6) δ 8.20 (bs, 1H, H5'), 8.03 (s, 2H, H3), 7.72 (t, *J* = 1.2 Hz, 1H, H4'), 7.05 (bs, 1H, H2').

^{13}C NMR (126 MHz, DMSO-d_6) δ 155.2 (C1), 136.1 (C5'), 132.2 (C4), 131.0 (C3), 130.0 (C2'), 118.7 (C4'), 88.0 (C2-I)

HRMS (FTMS-ESI$^+$) *m/z* $C_9H_7I_2N_2O$ $[M+H]^+$ calculated 412.8642, found 412.8629.

5-Iodo-8-aminoquinoline (13)



Purification was achieved by flash column chromatography (silica) using a 2-25% gradient of ethyl acetate in hexanes.

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.71 (dd, *J* = 4.1, 1.3 Hz, 1H, H2), 8.27 (dd, *J* = 8.5, 1.4 Hz, 1H, H3), 7.83 (d, *J* = 8.0 Hz, 1H, H6), 7.45 (dd, *J* = 8.5, 4.1 Hz, 1H, H3), 6.72 (d, *J* = 8.0 Hz, 1H, H7), 5.08 (s, 2H, NH_2).

^{13}C NMR (126 MHz, $CDCl_3$) δ 147.9 (C2), 144.9 (C_q), 140.1 (C4), 139.1 (C_q), 137.9 (C6), 130.1 (C_q), 123.0 (C3), 111.3 (C7), 81.0 (C5-I).

HRMS (FTMS-ESI$^+$) *m/z* $C_9H_8IN_2$ $[M+H]^+$ calculated 270.9727, found 270.9717.

(X) Structural analysis. Apo VirX crystals were obtained at 10.43 mg ml^{-1} in a hanging drop vapour diffusion crystallisation experiments with 2 µL drops (1:1 protein to reservoir ratio) with 500 µL reservoir solution containing 0.1 M Tris HCl pH 8.5, 0.2 M magnesium chloride hexahydrate, 30% w/v polyethylene glycol 4,000. Crystals were grown at room temperature. Crystals were harvested, cryoprotected with reservoir solution and 20 % (*v/v*) glycerol plus 20 % (*v/v*) polyethylene glycol 4,000 and subsequently flash frozen in liquid nitrogen. Data was collected at 100 K at beamline i03 (Diamond Light Source, UK). Data was processed with autoPROC.[3-6] Structure solution and refinement were performed using the CCP4 suite.[7] The apo structure was solved using molecular replacement employing MrBump.[8] The initial solution was completed through manual model building and refined using Coot and refmac5, respectively.[9,10] Model validation was carried out using MolProbity.[11] Structure figures were generated using ChimeraX[12] and substrate binding sites were mapped using LigPlot+ v.2.1.[13]

(Y) Docking studies. Docking experiments were performed using AutoDockTools-1.5.6 employing AutoGrid4 for grid pre-calculations and AutoDock4 for docking simulations[14] using VirX1 chain a as receptor with a grid size set to 126 × 126 × 126 points with 0.375 Å (volume: 105,488 $Å^3$) spacing centred on the receptor (x: 43.464 y: -47.549 z: -18.565). The Lamarckian genetic algorithm (LGA) with ten runs, population size of 150, maximum number of generations of 27,000 and a maximal number of energy evaluations of 25,000,000 was employed due its robustness and efficient performance.[15,16] Upon identification of a putative substrate binding site we refined the docking simulations by decreasing grid size (66 × 66 × 66; 0.375 Å spacing; volume: 15,160 $Å^3$) centred on the putative substrate binding site (x: 40.99 y: -45.115 z: -9.958) and allowing flexible side chains for residues: K79, E358, F99, G100, P101, Y97, H55, N463, N138 and L103 (total number of torsions: 21); using the rest of VirX1 chain A as the rigid part of the receptor, employing LGA again with the same parameters as above. Final binding modes were

judged by geometric restrains derived from insights into VirX1 regio-selectivity. For figure S14 we performed docking of compound 14 into chain b and chain c using the same docking parameters as above only changing the grid box origin to x: 0.214 y: -38.323 z: -24.017 and x: 23.0 y: -6.988 z: -9.94, respectively. In parallel rigid docking around the putative site (20 × 20 × 20 Å cubic search space, centred on the K79 side chain amine, x: 45.6, y: -50.3, z: -14.1) using LeDock[17,18] in order to bench mark docking results of flexible docking.

Number of rotatable bonds in docked compounds can be found in Supplementary Table 3.

(Z) Supplementary references

1. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol, *BMC Biotechnol.* **8**, 91 (2008).

2. Chapman, A. L. P. et al. Hypobromous acid and bromamine production by neutrophils and modulation by superoxide, *Biochem. J*. **417**, 773–781 (2009).

3. Vonrhein, C. et al. Data processing and analysis with the autoPROC toolbox, *Acta Cryst. D* **67**, 293-302 (2011).

4. Battye, T. G. G. et al. IMosflm: a new graphical interface for diffraction-image processing with MOSFLM, *Acta Cryst. D* **67**, 271-281 (2011).

5. Kabsch, W. XDS, *Acta Cryst. D* **66**, 125-132 (2010).

6. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Cryst. D* **69**, 1204–1214 (2013).

7. Winn, M. D. et al. Overview of the CCP4 suite and current developments, *Acta. Cryst. D* **67**, 235-242 (2011).

8. Keegan, R. M. & Winn, M. D. Automated search-model discovery and preparation for structure solution by molecular replacement, *Acta Cryst.* D **63**, 447 - 457 (2007).

9. Emsley, P. et al. Features and development of Coot, *Acta Cryst. D* **66**, 486-501 (2010).

10. Murshudov, G. N. et al. Refinement of macromolecular structures by the maximum-likelihood method, *Acta Cryst. D* **53**, 240-255 (1997).

11. Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids, *Nucleic Acids Res.* **35**, W375-W383 (2007).

12. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis, *Protein Sci.* **27**, 14-25 (2018).

13. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions, *Protein Eng*. **8**, 127-134 (1996).

14. Morris, G. M. et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, *J. Comput. Chem.* **30,** 2785-2791 (2009).

15. Morris, G. M. et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem.* **19**,1639-1662 (1998).

16. Brooijmans, N. & Kuntz, I. D. Molecular recognition and docking algorithms, *Annu. Rev. Bioph. Biom.* **32**, 335-373 (2003).

17. Wang, Z. et al. Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power, *Phys. Chem. Chem. Phys.* **18**, 12964-12975 (2016).

18. Zhao, H. & Caflisch, A. Discovery of ZAP70 inhibitors by high-throughput docking into a conformation of its kinase domain generated by molecular dynamics, *Bioorg. Med. Chem. Lett.* **23**, 5721-5726 (2013).

	VirX1
PDB ID	6QGM
Data collection	
Beamline	i03 (DLS)
Wavelength (Å)	0.9763
Space group	P 1 2_1 1
Cell dimensions	
a, b, c (Å)	101.13, 172.98, 110.43
α, β, γ (°)	90, 112.89, 90
Resolution	2.75 (2.75-2.80)
R_{meas}	0.119 (0.696)
I/δI	2.1
Completeness (%)	99.5
Multiplicity	3.5
Refinement	
Resolution (Å)	87.94-2.75
No. reflections	85,159
R_{work}/R_{free}	0.2158/0.2589
No. atoms	25,690
Protein	25,648
Ligand/ion	-
Water	63
Overall *B*-factor	39.291
R.m.s. deviations	
Bond length (Å)	0.007
Bond angles (°)	1.443

Supplementary Table 1. Data collection and refinement statistics (molecular replacement) for structural studies of apostructure of VirX1. *Values in parentheses for highest-resolution shell.*

Compound	Binding energy (kcal mol^{-1})	
	ADT, flexible	LeDock, rigid
1	-4.56	-3.25
2	-4.58	n.d.
3	-5.64	-3.39
4	-6.42	-4.77
5	-7.38	-4.29
6	-5.22	-3.27
7	-4.69	-3.69
8	-6.45	-3.55
9	n.d.	-2.8
10	-6.07	-3
11	-6.2	-3.09
12	n.d.	-3.81
13	n.d.	n.d.
14	-8.85	n.d.

Supplementary Table 2. Calculated binding energies derived from docking experiments.

Compound	Number of torsion angles
1	0
2	2
3	2
4	7
5	1
6	1
7	2
8	1
9	1
10	1
11	1
12	2
13	1
14	1

Supplementary Table 3. Number of rotational bonds for every docked compound.

	Peak 1	Peak 2	Peak 3
Mn (kDa)	162.6	389.9	662.9
Mw (kDa)	162.6	390.2	663.9
Polydispersity (Mw/Mn)	1.000	1.001	1.002

Supplementary Table 4. Molar mass of VirX1 species determined by SEC-MALS.

Monomer MW (kDa)	Detection method	Concentration (mg/ml)	Major species				f/f_0
			Peak 1		Peak 2		
			MW (kDa)	Sed. Co (S)	MW (kDa)	Sed. Co (S)	
61.5	Absorbance	1.20	Absorbance too high		-	-	-
		0.60	175.2	5.30	305.6	7.68	1.29
		0.12	171.1	5.29	328.7	8.18	1.27
	Interference	1.20*	193.1	5.28	-	-	1.38
		0.60*	186.2	5.27	-	-	1.35
		0.12*	174.8	5.14	-	-	1.31

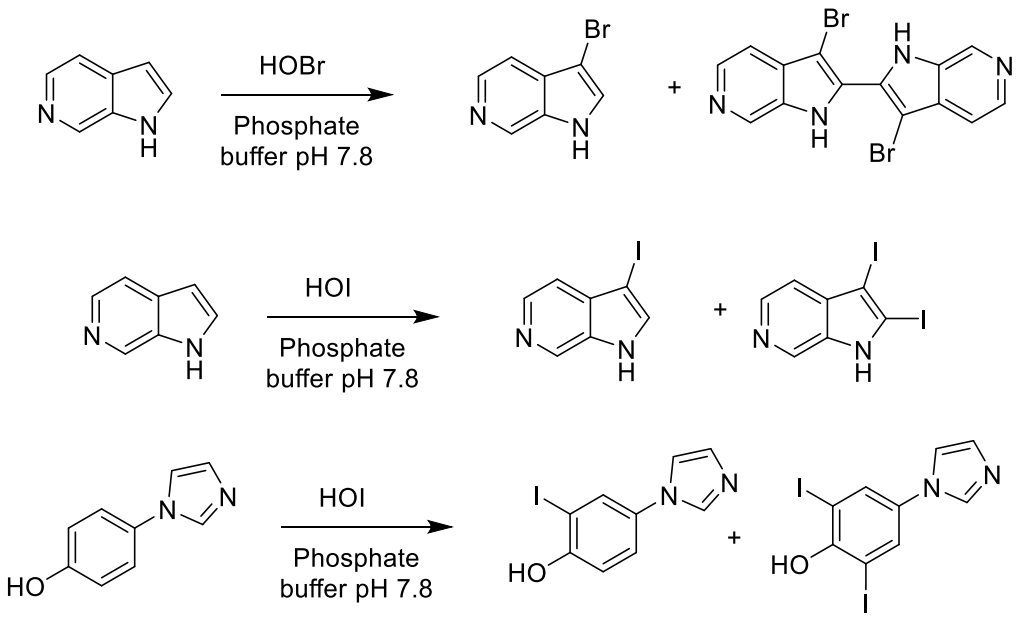
Supplementary Table 5. Species estimated molecular weights from the c(s) analysis.
For each sample concentration the signal-weighted sedimentation co-efficient and the estimated molecular weight of each species is shown, together with the best-fit frictional ratio for the distribution.

Compound	Iodination with PrnA %	Iodination with VirX1 %
1	20	95
2	10	30
3	ND	5
4	10	75
5	ND	30
6	ND	65
7	10	25
8	>1	20
9	ND	30
10	20	70
11	ND	20
12	ND	15
13	ND	10
14	ND	15
15	>1	20
16	15	35
17	>1	30
18	>1	10
19	25	50
20	>1	70
21	>1	>1
22	5	15
23	>1	35
24	ND	10
25	>1	20
26	10	15
27	>1	15
28	ND	>1
29	>1	>1
30	>1	>1
31	ND	20
32	>1	5

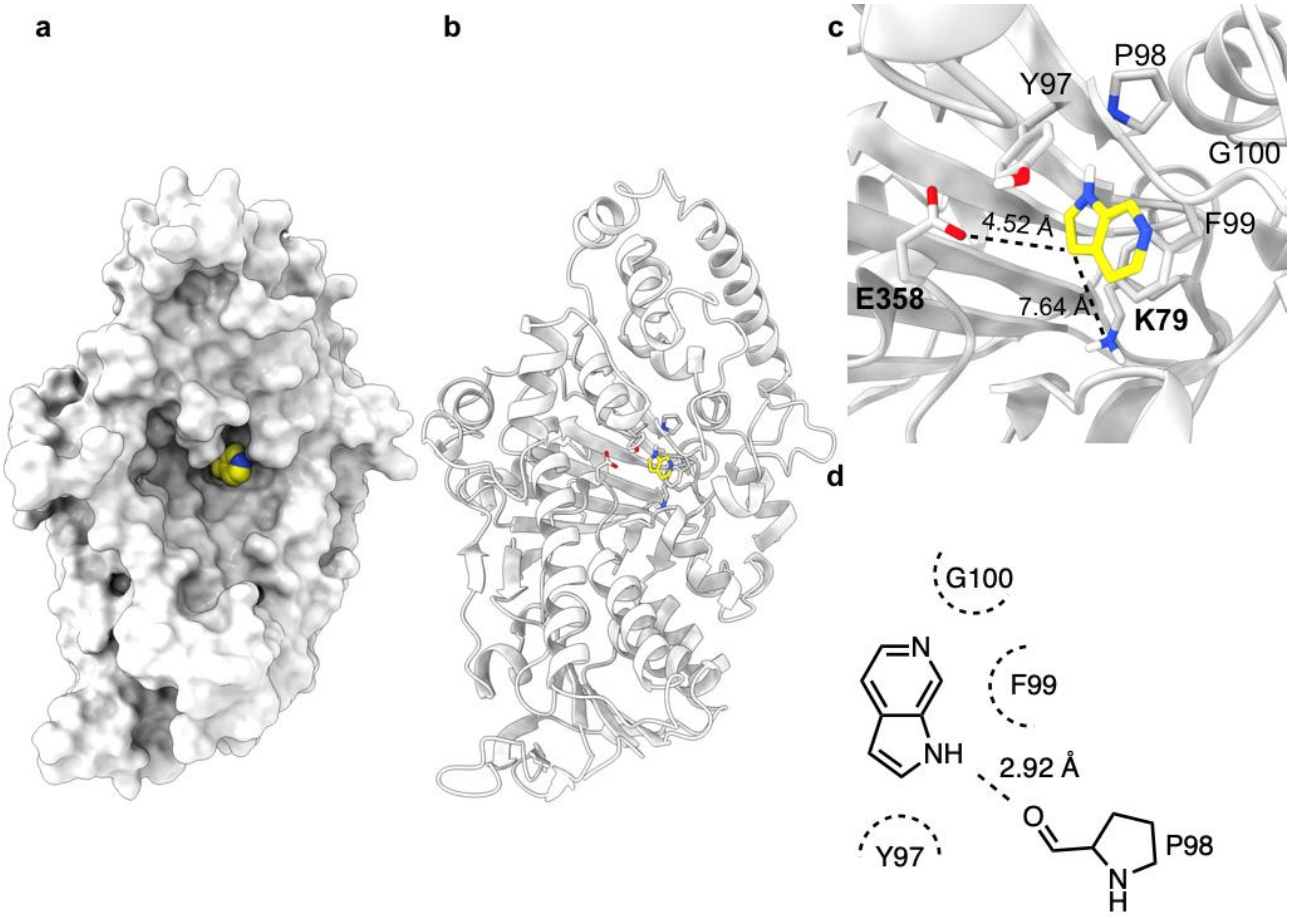
Supplementary Table 6. Iodination of compounds accepted by VirX1 vs PrnA.
*ND= product not detected.

Compound	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (min^{-1} μM^{-1})
7-Cl-L-tryptophan	3.6 ± 0.3	26 ± 1.5	0.138 ± 0.011
7-Br-L-tryptophan	2.2 ± 0.2	45 ± 3.1	0.048 ± 0.006

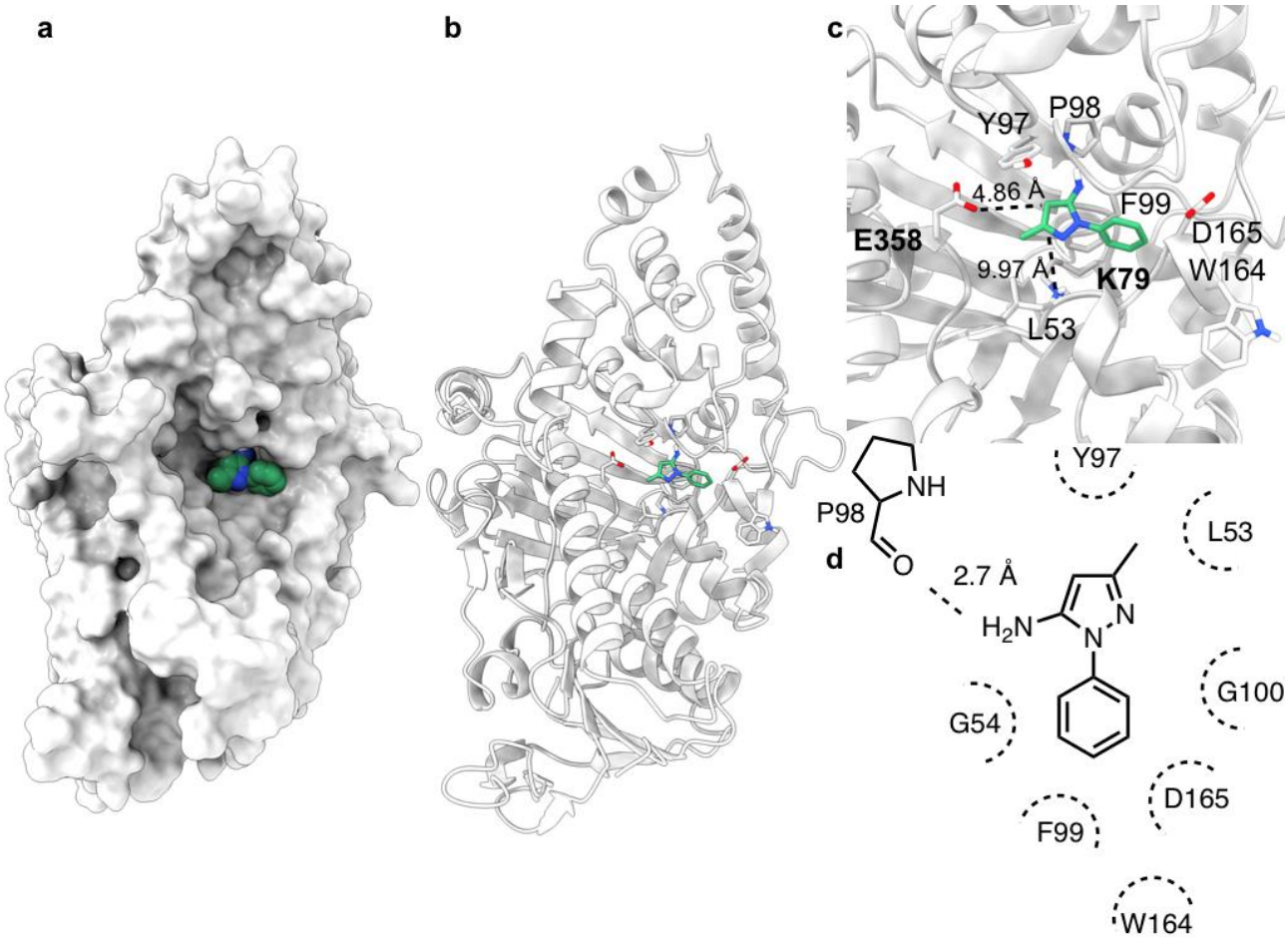
Supplementary Table 7. PrnA kinetic parameters obtained for L-tryptophan with both NaCl and NaBr



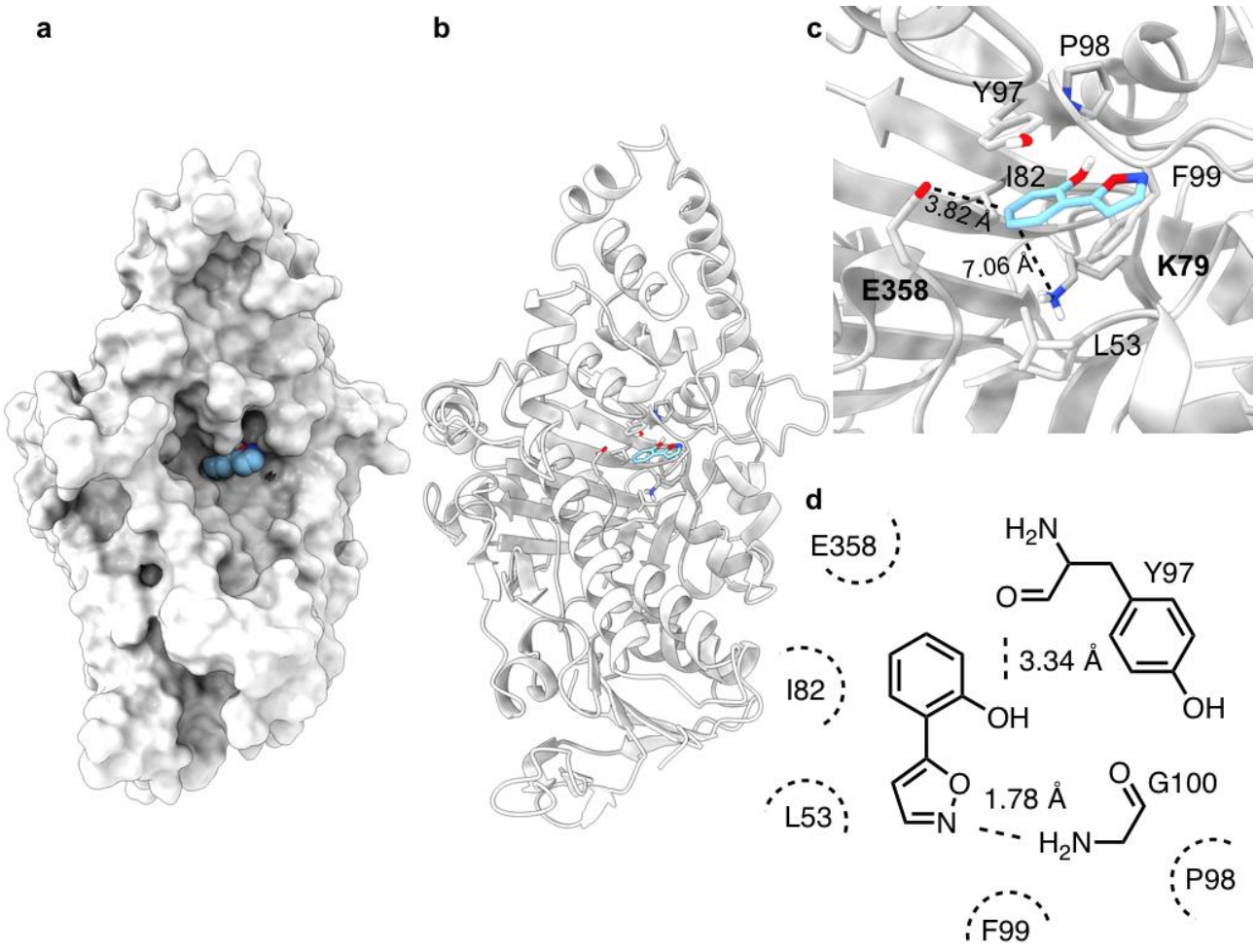
Supplementary Scheme 1. Proposed products of the reaction with HOBr or HOI identified by LCMS. These reactions proceed rapidly with HOX with formation of di-iodo products along with mono-iodo analogues at room temperature.



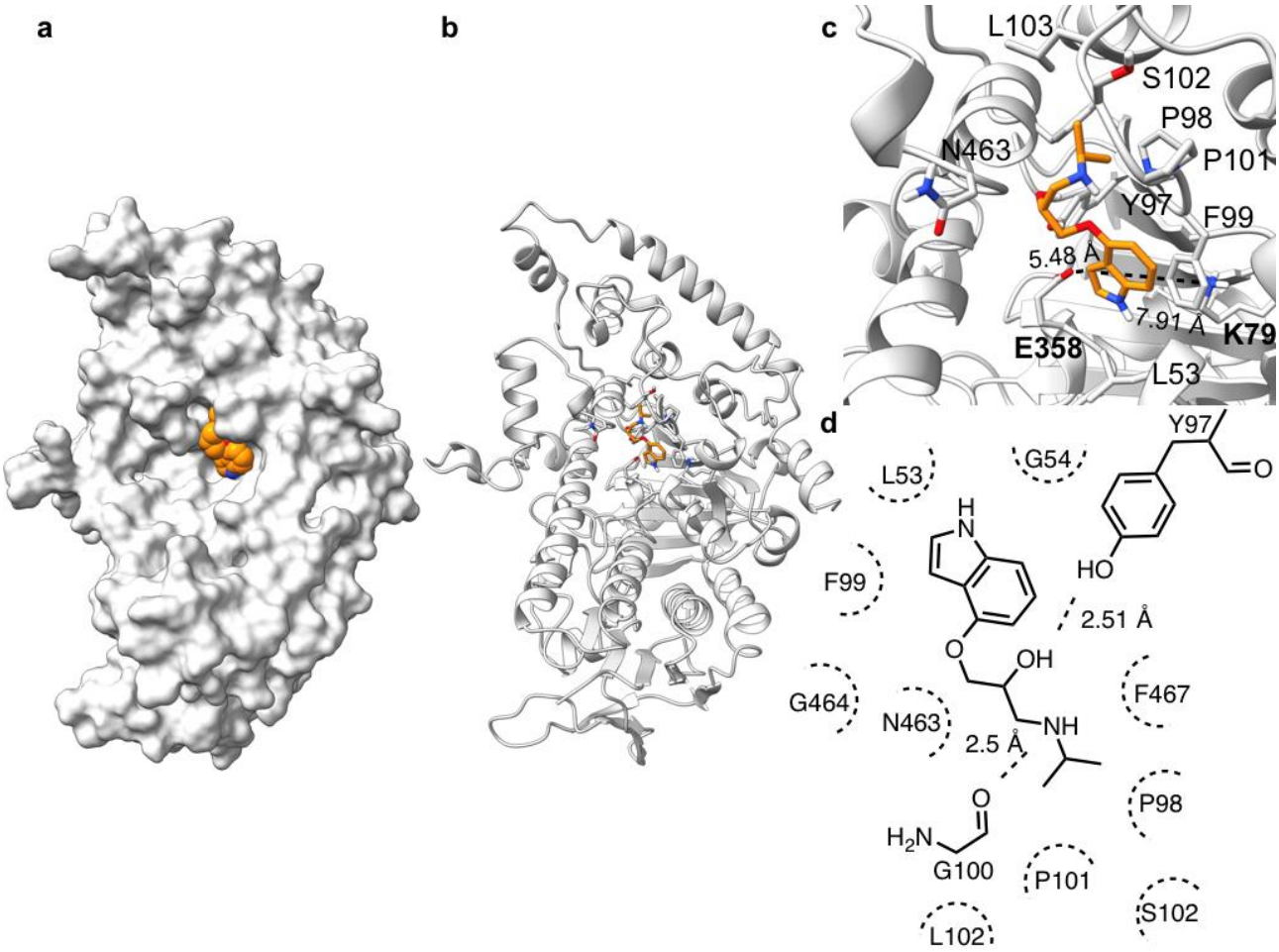
Supplementary Figure 1. Structural overview of VirX1 substrate binding site with bound compound 1 from flexible docking. a, Surface representation of VirX1 monomer with bound compound **1** (sphere style) in binding cleft. **b**, Cartoon representation of VirX1 monomer with bound compound1 (stick style) in binding cleft; binding site residues involved in binding compound1 as well as K79 and E358 are represented as sticks. **c**, Close up of VirX1 binding site, binding residues and compound **1** are depicted as sticks. For **a-c**, VirX1 surface style, cartoon style and carbons were coloured in light grey. Compound **1** is coloured in yellow (carbons). Nitrogen, oxygen and hydrogen atoms are coloured in blue, red and white, respectively. **d**, LigPlot+ v.2.1 derived representation of VirX1 substrate binding site. Dashed half circles represent hydrophobic contacts and dashed lines hydrogen bonds.



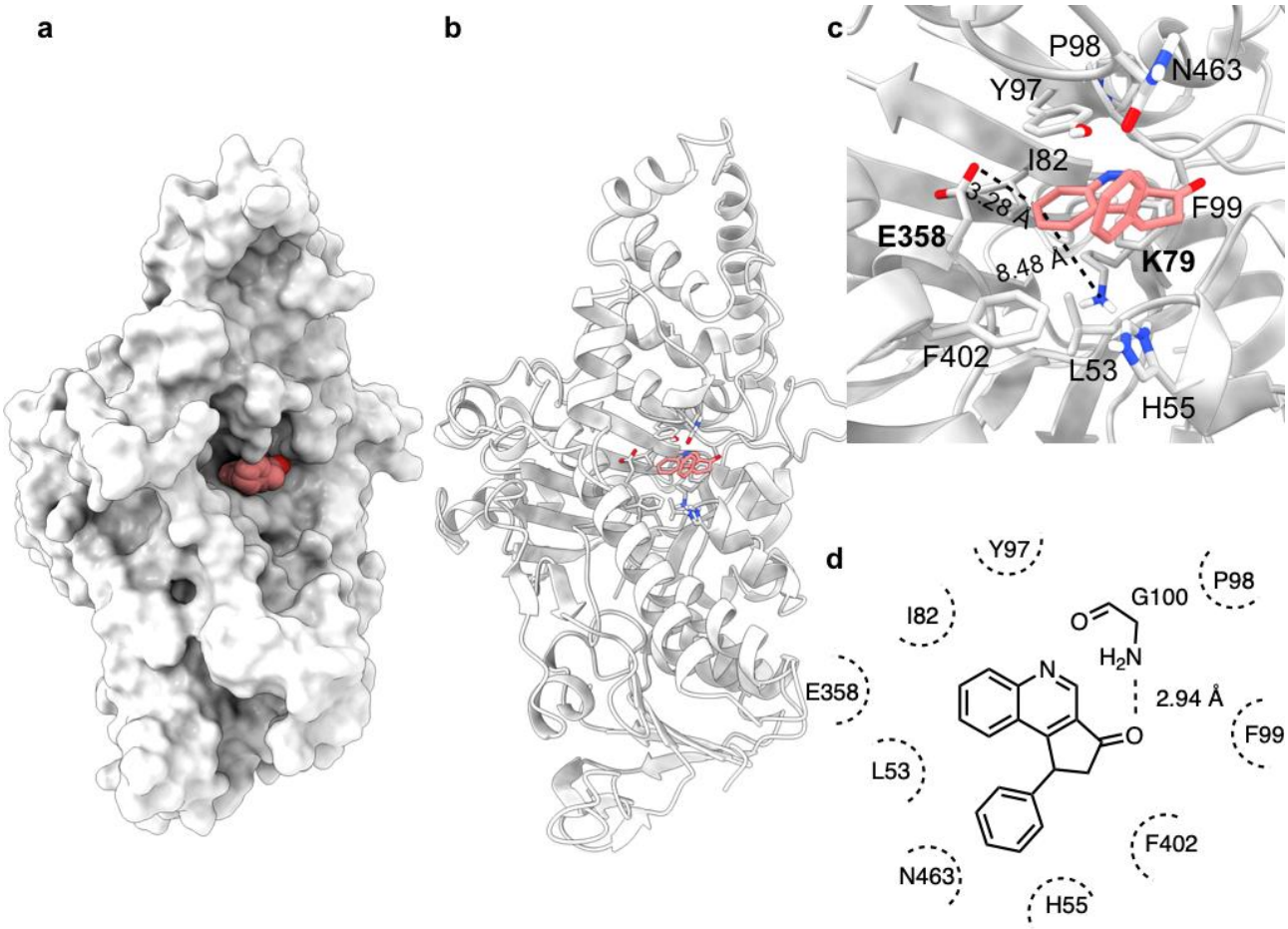
Supplementary Figure 2. Structural overview of VirX1 substrate binding site with bound compound 2 from flexible docking. a, Surface representation of VirX1 monomer with bound compound **2** (sphere style) in binding cleft. **b**, Cartoon representation of VirX1 monomer with bound compound **2** (stick style) in binding cleft; binding site residues involved in binding compound **2** as well as K79 and E358 are represented as sticks. **c**, Close up of VirX1 binding site, binding residues and compound **2** are depicted as sticks. For **a-c**, VirX1 surface style, cartoon style and carbons were coloured in light grey. Compound **2** is coloured in green (carbons). Nitrogen, oxygen and hydrogen atoms are coloured in blue, red and white, respectively. **d**, LigPlot+ v.2.1 derived representation of VirX1 substrate binding site. Dashed half circles represent hydrophobic contacts and dashed lines hydrogen bonds.



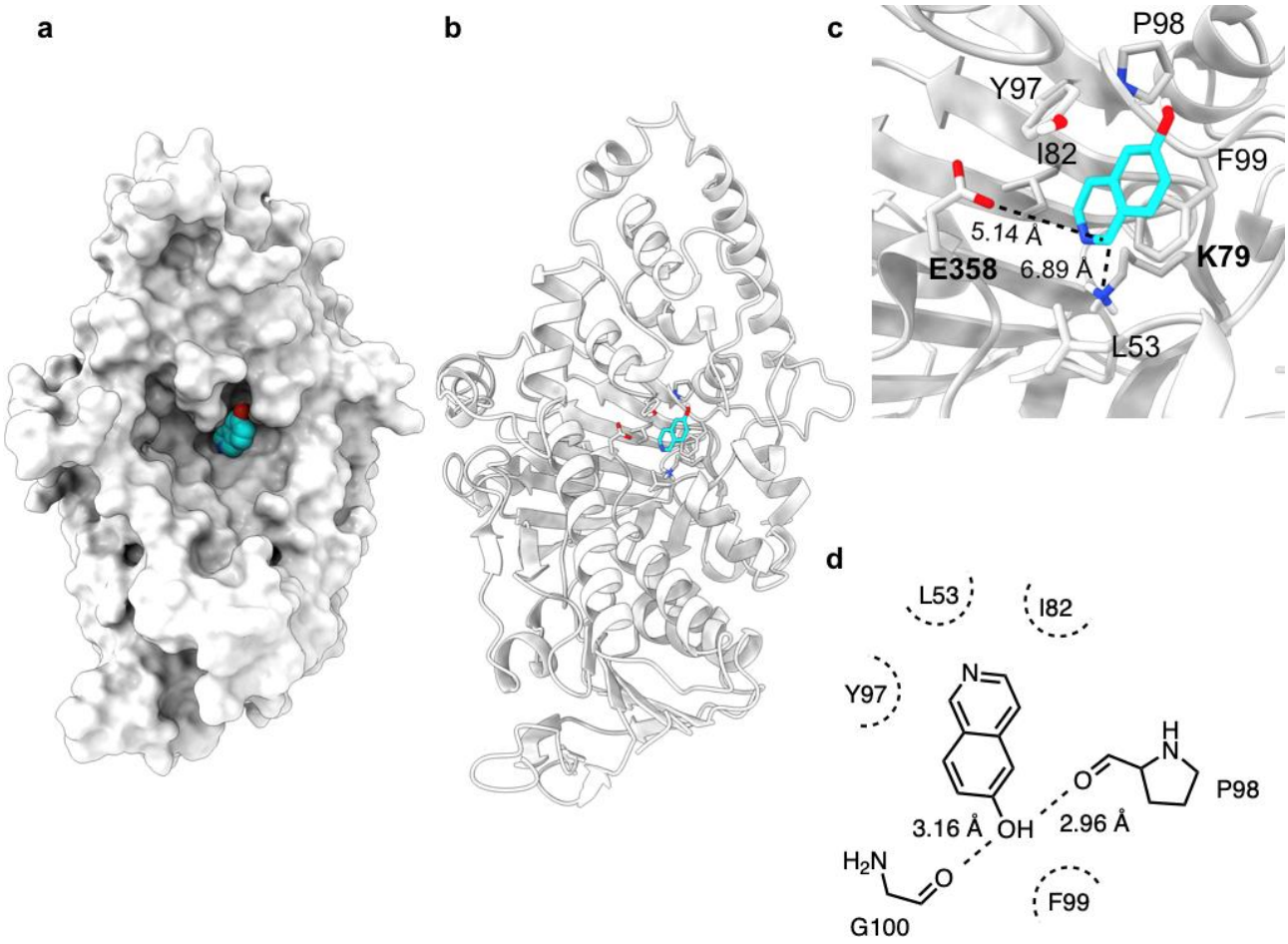
Supplementary Figure 3. Structural overview of VirX1 substrate binding site with bound compound 3 from flexible docking. a, Surface representation of VirX1 monomer with bound compound **3** (sphere style) in binding cleft. **b**, Cartoon representation of VirX1 monomer with bound compound **3** (stick style) in binding cleft; binding site residues involved in binding compound **3** as well as K79 are represented as sticks. **c**, Close up of VirX1 binding site, binding residues and compound **3** are depicted as sticks. For **a-c**, VirX1 surface style, cartoon style and carbons were coloured in light grey. Compound **3** is coloured in sky blue (carbons). Nitrogen, oxygen and hydrogen atoms are coloured in blue, red and white, respectively. **d**, LigPlot+ v.2.1 derived representation of VirX1 substrate binding site. Dashed half circles represent hydrophobic contacts and dashed lines hydrogen bonds.



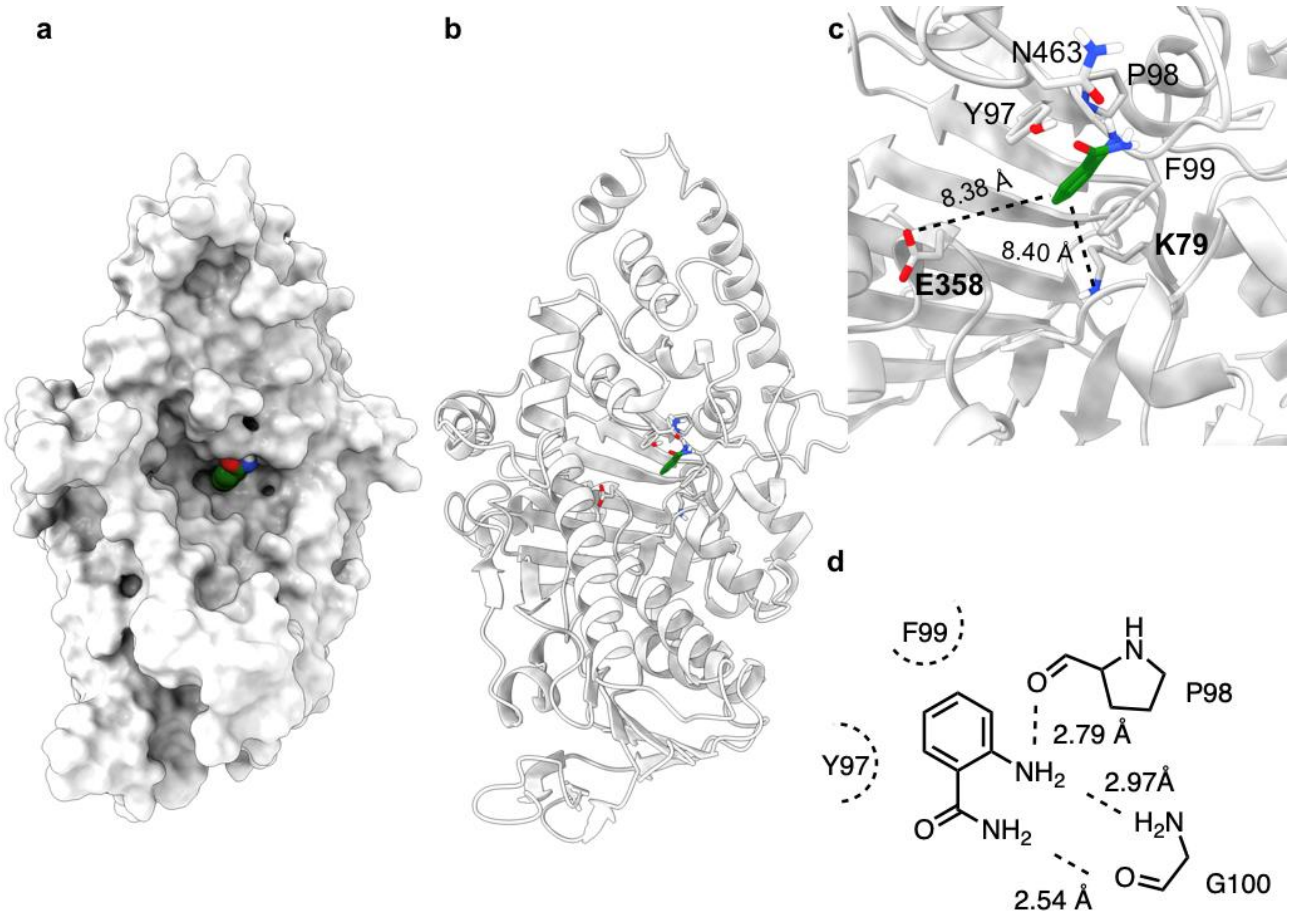
Supplementary Figure 4. Structural overview of VirX1 substrate binding site with bound compound 4 from flexible docking. a, Surface representation of VirX1 monomer with bound compound **4** (sphere style) in binding cleft. **b**, Cartoon representation of VirX1 monomer with bound compound **4** (stick style) in binding cleft; binding site residues involved in binding compound **4** as well as K79 and E358 are represented as sticks. **c**, Close up of VirX1 binding site, binding residues and compound **4** are depicted as sticks. For **a-c**, VirX1 surface style, cartoon style and carbons were coloured in light grey. Compound **4** is coloured in orange (carbons). Nitrogen, oxygen and hydrogen atoms are coloured in blue, red and white, respectively. **d**, LigPlot+ v.2.1 derived representation of VirX1 substrate binding site. Dashed half circles represent hydrophobic contacts and dashed lines hydrogen bonds.



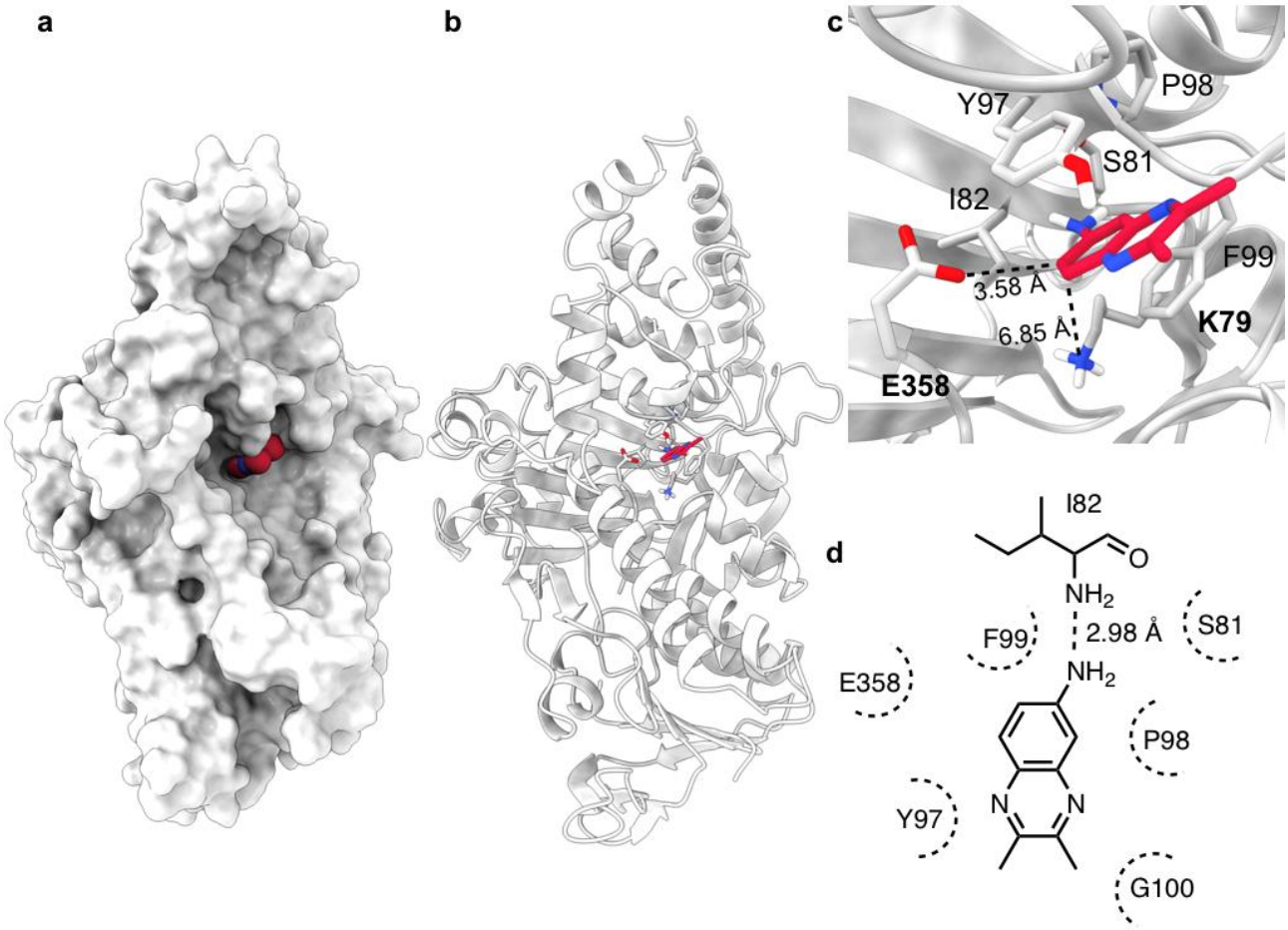
Supplementary Figure 5. Structural overview of VirX1 substrate binding site with bound compound 5 from flexible docking. a, Surface representation of VirX1 monomer with bound compound **5** (sphere style) in binding cleft. **b**, Cartoon representation of VirX1 monomer with bound compound **5** (stick style) in binding cleft; binding site residues involved in binding compound **5** as well as K79 and E358 are represented as sticks. **c**, Close up of VirX1 binding site, binding residues and compound **5** are depicted as sticks. For **a-c**, VirX1 surface style, cartoon style and carbons were coloured in light grey. Compound **5** is coloured in salmon (carbons). Nitrogen, oxygen and hydrogen atoms are coloured in blue, red and white, respectively. **d**, LigPlot+ v.2.1 derived representation of VirX1 substrate binding site. Dashed half circles represent hydrophobic contacts and dashed lines hydrogen bonds.



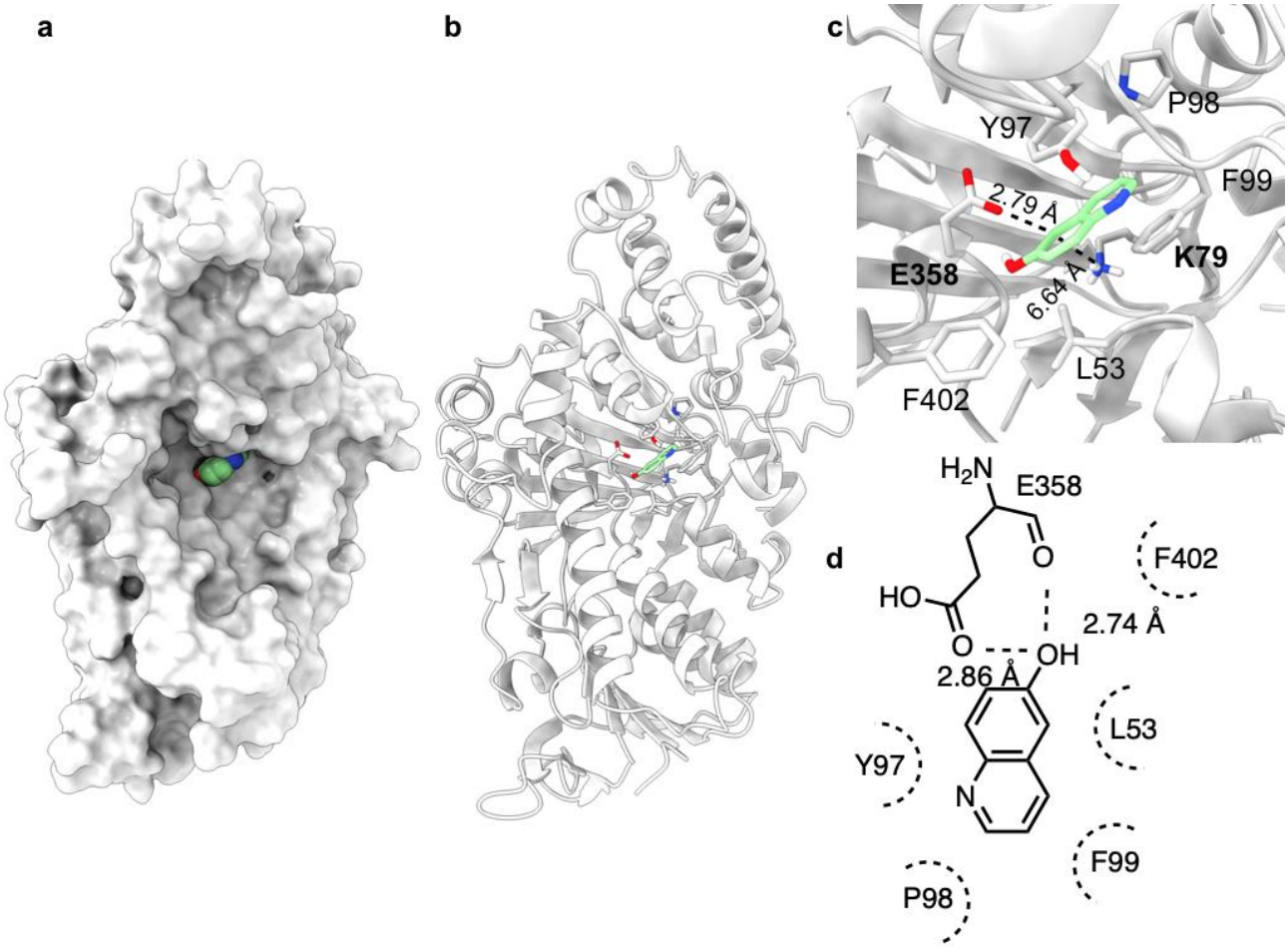
Supplementary Figure 6. Structural overview of VirX1 substrate binding site with bound compound 6 from flexible docking. a, Surface representation of VirX1 monomer with bound compound **6** (sphere style) in binding cleft. **b**, Cartoon representation of VirX1 monomer with bound compound **6** (stick style) in binding cleft; binding site residues involved in binding compound **6** as well as K79 and E358 are represented as sticks. **c**, Close up of VirX1 binding site, binding residues and compound **6** are depicted as sticks. For **a-c**, VirX1 surface style, cartoon style and carbons were coloured in light grey. Compound **6** is coloured in cyan (carbons). Nitrogen, oxygen and hydrogen atoms are coloured in blue, red and white, respectively. **d**, LigPlot+ v.2.1 derived representation of VirX1 substrate binding site. Dashed half circles represent hydrophobic contacts and dashed lines hydrogen bonds.



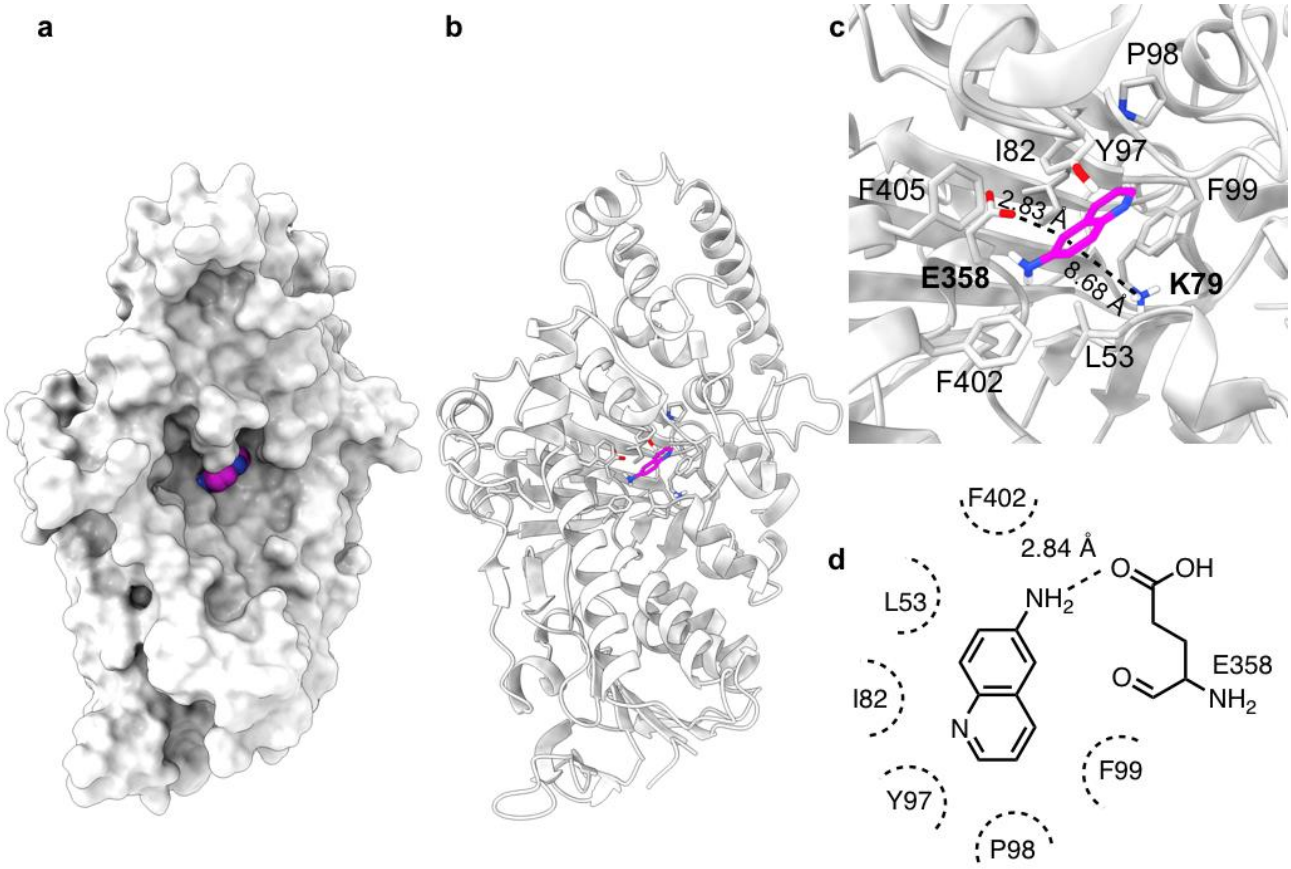
Supplementary Figure 7. Structural overview of VirX1 substrate binding site with bound compound 7 from flexible docking. a, Surface representation of VirX1 monomer with bound compound **7** (sphere style) in binding cleft. **b**, Cartoon representation of VirX1 monomer with bound compound **7** (stick style) in binding cleft; binding site residues involved in binding compound **7** as well as K79 and E358 are represented as sticks. **c**, Close up of VirX1 binding site, binding residues and compound **7** are depicted as sticks. For **a-c**, VirX1 surface style, cartoon style and carbons were coloured in light grey. Compound **7** is coloured in dark green (carbons). Nitrogen, oxygen and hydrogen atoms are coloured in blue, red and white, respectively. **d**, LigPlot+ v.2.1 derived representation of VirX1 substrate binding site. Dashed half circles represent hydrophobic contacts and dashed lines hydrogen bonds.



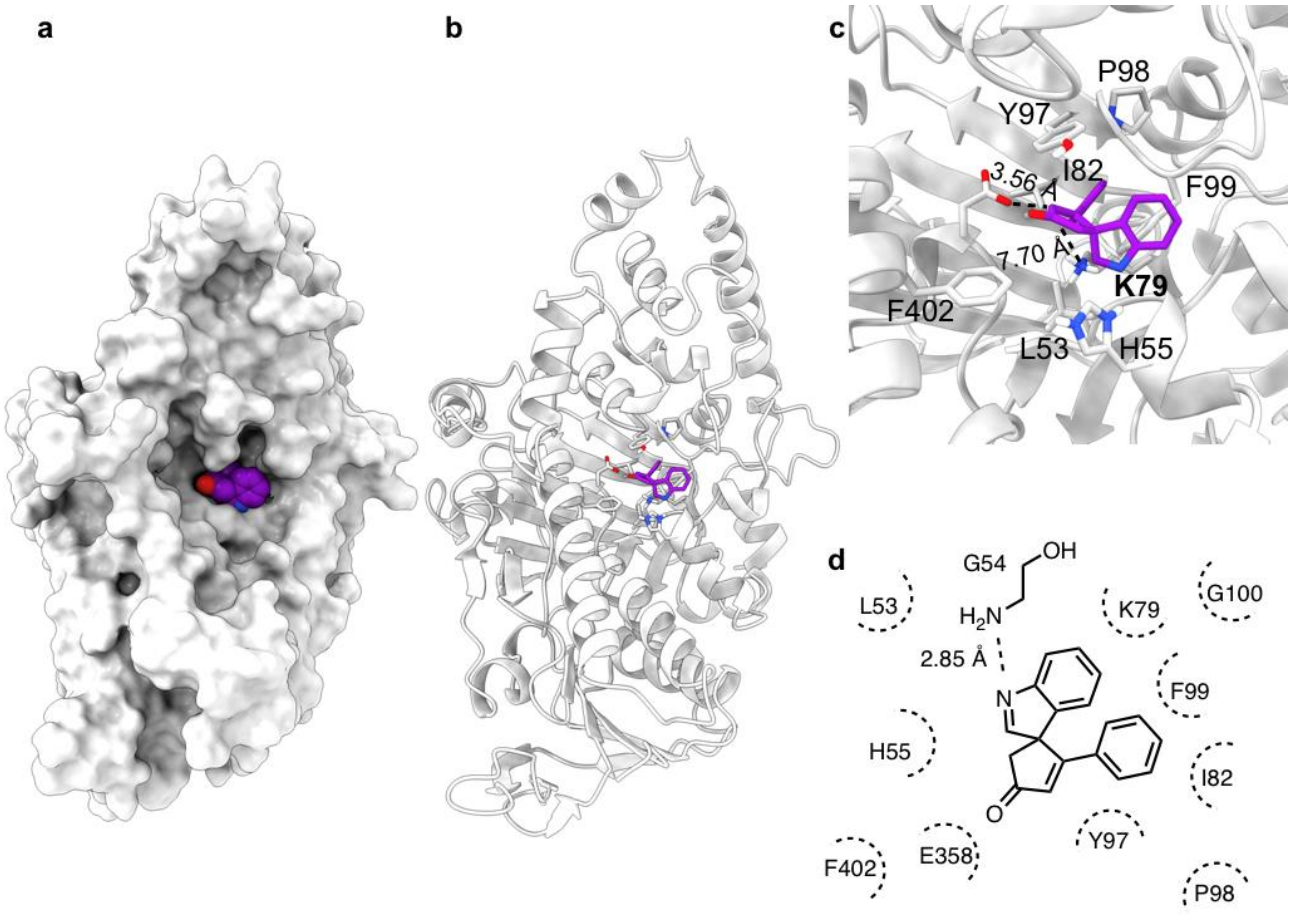
Supplementary Figure 8. Structural overview of VirX1 substrate binding site with bound compound 8 from flexible docking. a, Surface representation of VirX1 monomer with bound compound **8** (sphere style) in binding cleft. **b**, Cartoon representation of VirX1 monomer with bound compound **8** (stick style) in binding cleft; binding site residues involved in binding compound **8** as well as K79 and E358 are represented as sticks. **c**, Close up of VirX1 binding site, binding residues and compound **8** are depicted as sticks. For **a-c**, VirX1 surface style, cartoon style and carbons were coloured in light grey. Compound **8** is coloured in light red (carbons). Nitrogen, oxygen and hydrogen atoms are coloured in blue, red and white, respectively. **d**, LigPlot+ v.2.1 derived representation of VirX1 substrate binding site. Dashed half circles represent hydrophobic contacts and dashed lines hydrogen bonds.



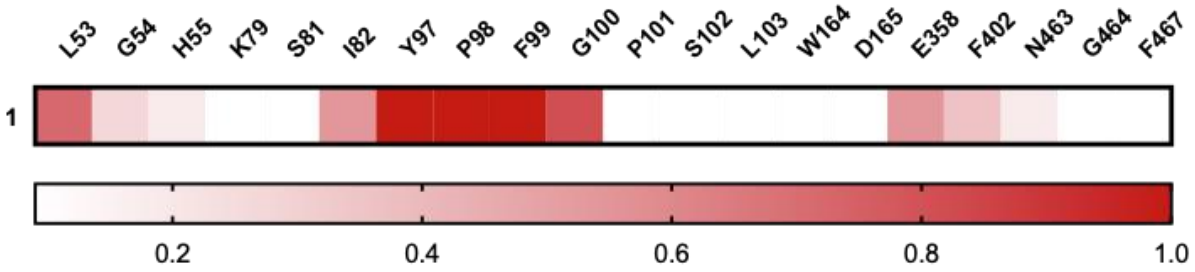
Supplementary Figure 9. Structural overview of VirX1 substrate binding site with bound compound 10 from flexible docking. a, Surface representation of VirX1 monomer with bound compound **10** (sphere style) in binding cleft. **b**, Cartoon representation of VirX1 monomer with bound compound **10** (stick style) in binding cleft; binding site residues involved in binding compound **10** as well as K79 and E358 are represented as sticks. **c**, Close up of VirX1 binding site, binding residues and compound **10** are depicted as sticks. For **a-c**, VirX1 surface style, cartoon style and carbons were coloured in light grey. Compound **10** is coloured in light green (carbons). Nitrogen, oxygen and hydrogen atoms are coloured in blue, red and white, respectively. **d**, LigPlot+ v.2.1 derived representation of VirX1 substrate binding site. Dashed half circles represent hydrophobic contacts and dashed lines hydrogen bonds.



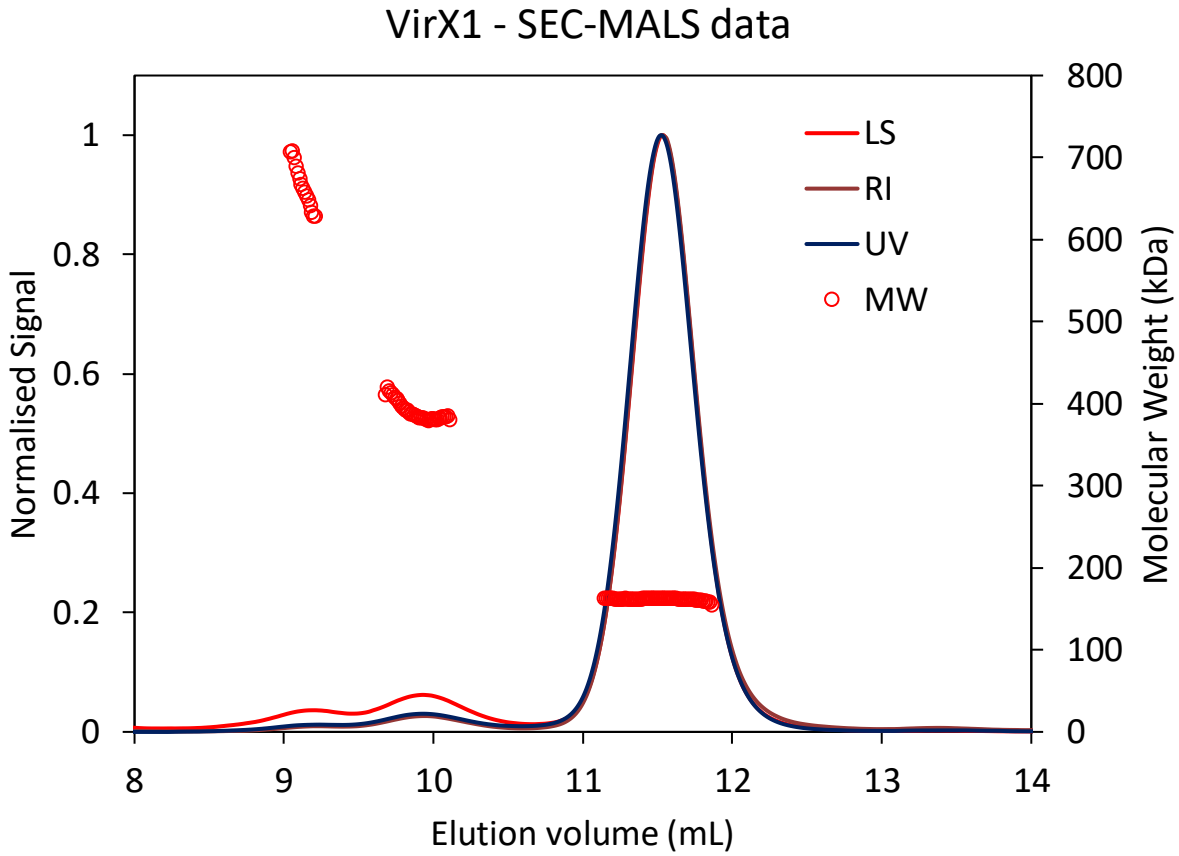
Supplementary Figure 10. Structural overview of VirX1 substrate binding site with bound compound 11 from flexible docking. a, Surface representation of VirX1 monomer with bound compound **11** (sphere style) in binding cleft. **b**, Cartoon representation of VirX1 monomer with bound compound **11** (stick style) in binding cleft; binding site residues involved in binding compound **11** as well as K79 and E358 are represented as sticks. **c**, Close up of VirX1 binding site, binding residues and compound **11** are depicted as sticks. For **a-c**, VirX1 surface style, cartoon style and carbons were coloured in light grey. Compound **11** is coloured in magenta (carbons). Nitrogen, oxygen and hydrogen atoms are coloured in blue, red and white, respectively. **d**, LigPlot+ v.2.1 derived representation of VirX1 substrate binding site. Dashed half circles represent hydrophobic contacts and dashed lines hydrogen bonds.



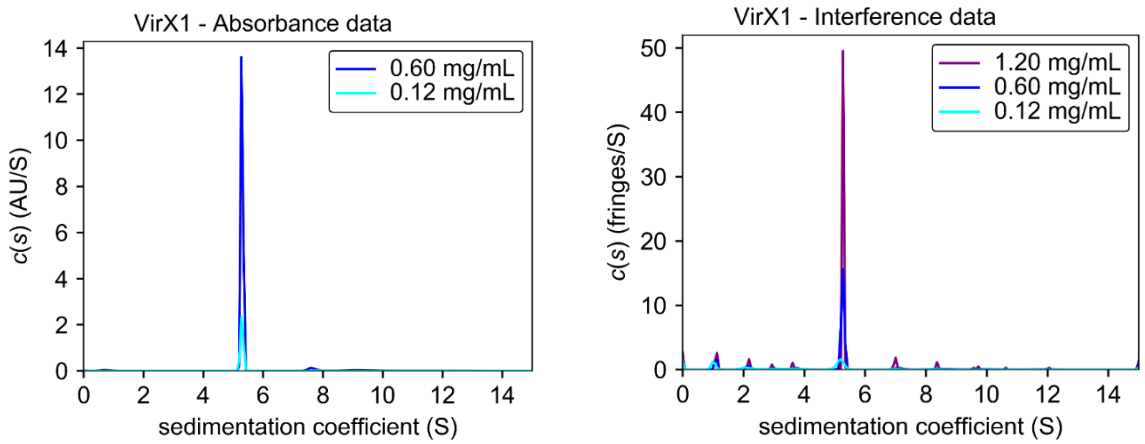
Supplementary Figure 11. Structural overview of VirX1 substrate binding site with bound compound 14 from flexible docking. a, Surface representation of VirX1 monomer with bound compound **14** (sphere style) in binding cleft. **b**, Cartoon representation of VirX1 monomer with bound compound **14** (stick style) in binding cleft; binding site residues involved in binding compound **14** as well as K79 and E358 are represented as sticks. **c**, Close up of VirX1 binding site, binding residues and compound **14** are depicted as sticks. For **a-c**, VirX1 surface style, cartoon style and carbons were coloured in light grey. Compound **14** is coloured in violet (carbons). Nitrogen, oxygen and hydrogen atoms are coloured in blue, red and white, respectively. **d**, LigPlot+ v.2.1 derived representation of VirX1 substrate binding site. Dashed half circles represent hydrophobic contacts and dashed lines hydrogen bonds.



Supplementary Figure 12. Occurrence frequency of residues involved in substrate binding. Sampled across all flexible docking solutions extracted from LigPlot+ v.2.1.

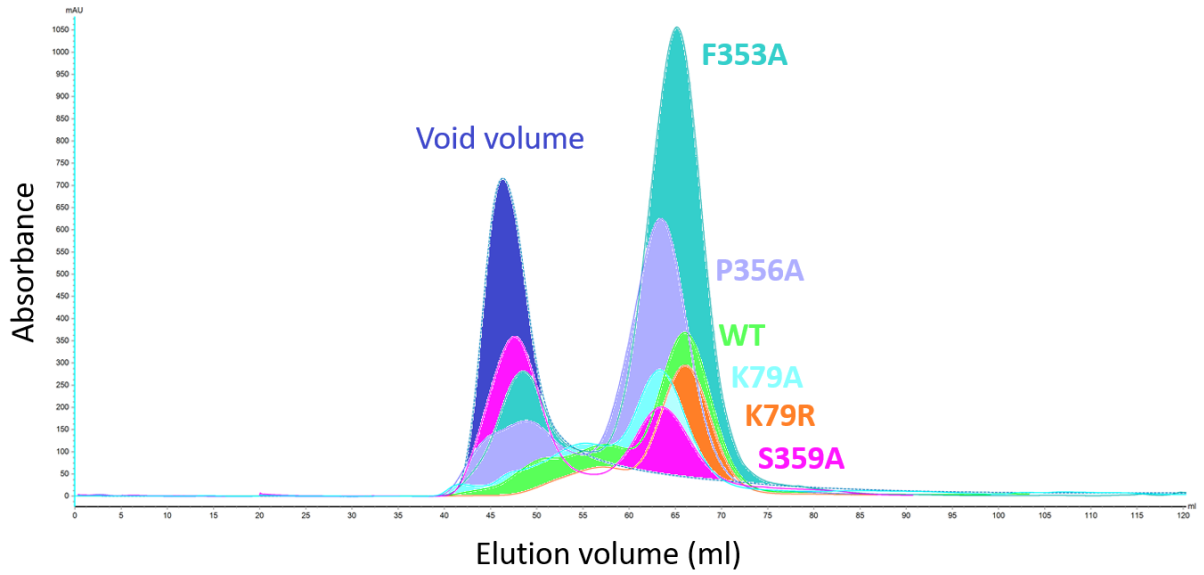


Supplementary Figure 13. SEC-MALS analysis of the oligomeric state of VirX1. LS – Light Scattering detection, RI – Refractive Index detection, UV – Absorbance detection at 280 nm, MW – Molecular Weight. Three species were detected in the SEC-MALS analysis. The main peak (peak 1) appeared with a molecular weight of 162.6 kDa, which is between the molecular weight expected for a dimer and a trimer. It is unlikely that the peak represents a population weighted average between two self-associating species as the peak appeared to be monodisperse, unlike peaks two and three where the leading edge of the peak has a higher molecular weight. The discrepancy may have arisen from either the use of the average dn/dc value, or from the presence of 10% glycerol in the buffer not being adequately accounted for.

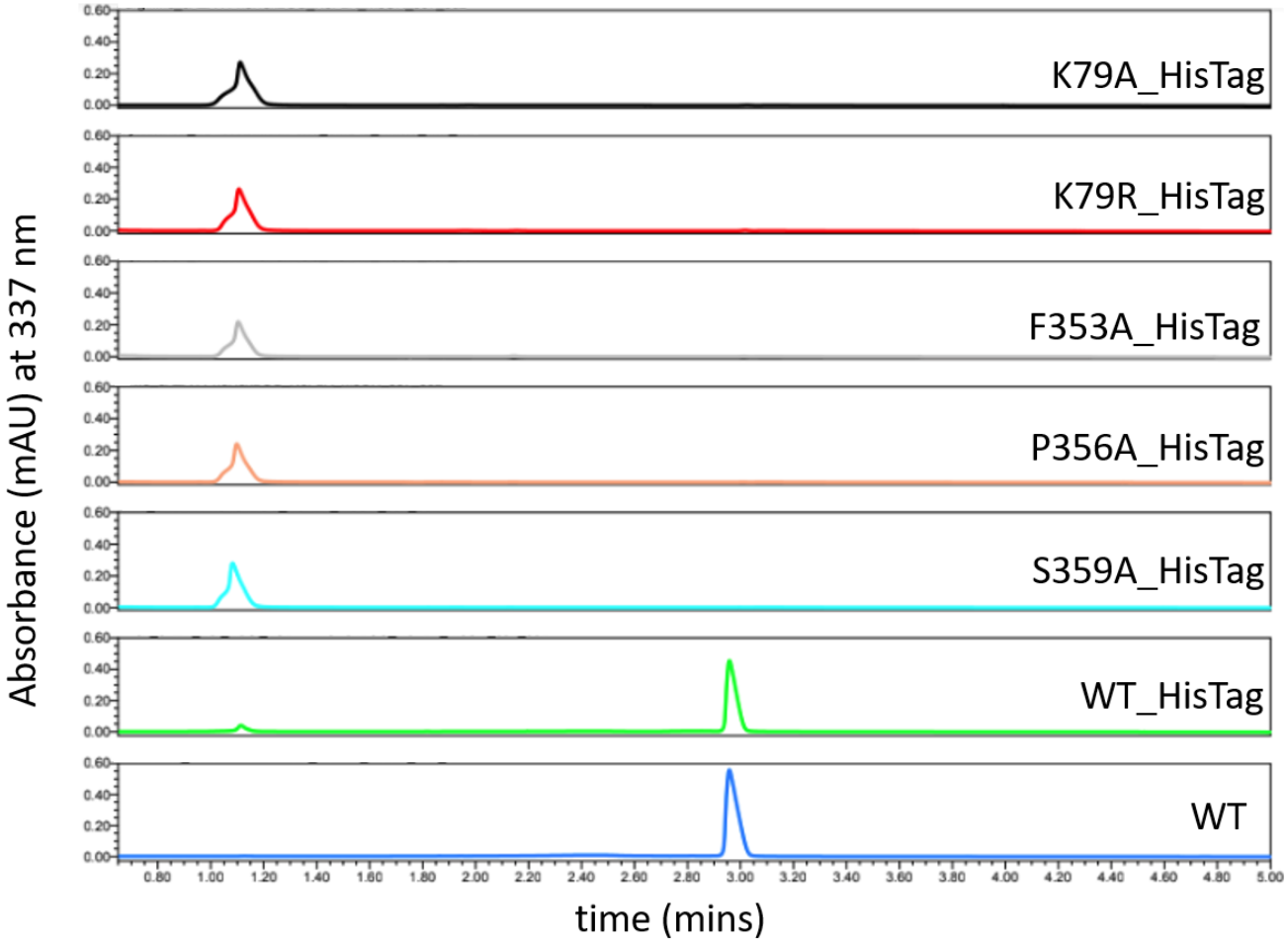


Supplementary Figure 14. Sedimentation coefficient distributions.

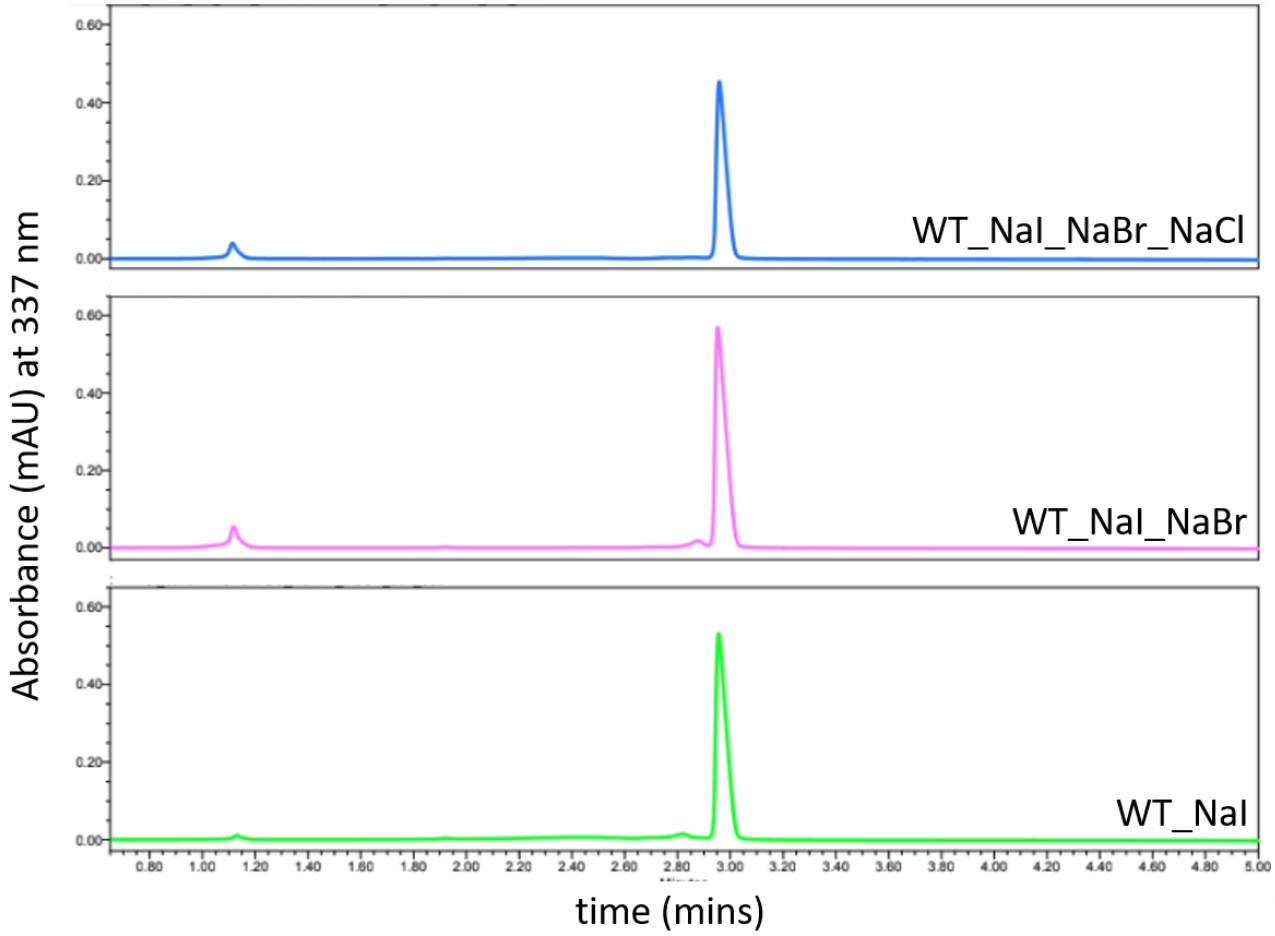
Analysis of the data collected using the absorbance optics found two species to be present, a main peak accounting for 95% of the material present and a second peak accounting for ~3% of the material. The main species present appeared with a molecular weight slightly less than that of a VirX1 trimer, but given the constraints of the c(s) analysis, this is indeed mostly likely a trimer. Analysis of the interference data also supports this, where the average molecular weight of this species was found to be 184.7 kDa, which is very close to the expected molecular weight of a trimer. Small amounts of other species were also seen in the interference data but reliably determining the molecular weights of these was not possible.



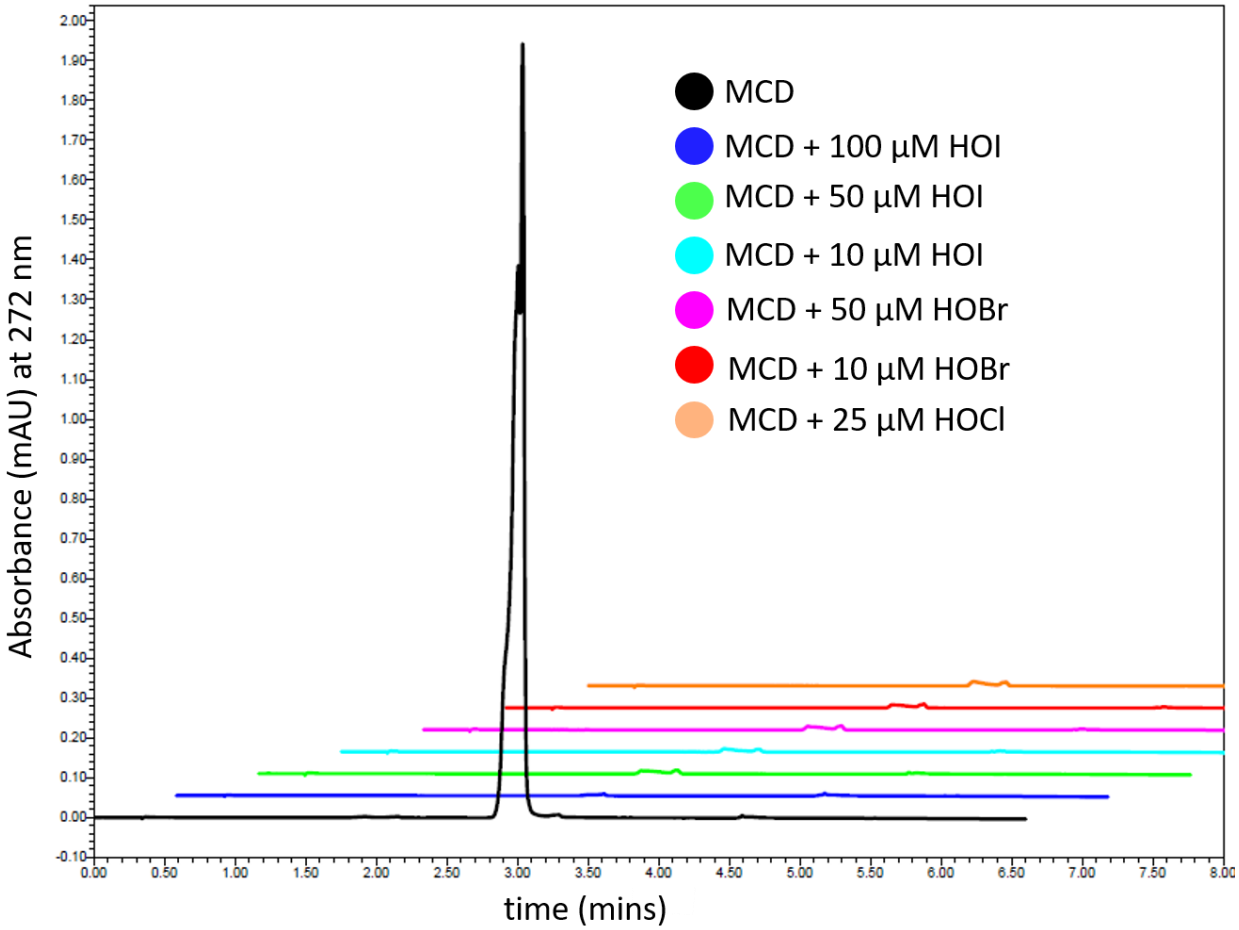
Supplementary Figure 15. Size exclusion chromatography profiles of wild type VirX1 and mutants F353A, P356A, K79A, K79R, S359A. Methods used for equilibration and elution of the respective protein samples were identical, in case of P356A, K79A and S359A the inlet for sample application did not retrieve the same volume of sample as in the case of the other samples compared, thus showing a slight shift in elution times by 2min. The left peak as defined by the elution of a blue dextran protein marker that defined the void volume corresponds to aggregates. The right peak corresponds to the elution volume of the trimeric VirX1 in solution indicating that all mutants are folded as WT VirX1, perhaps with different levels of stability, as indicated by the left peak.



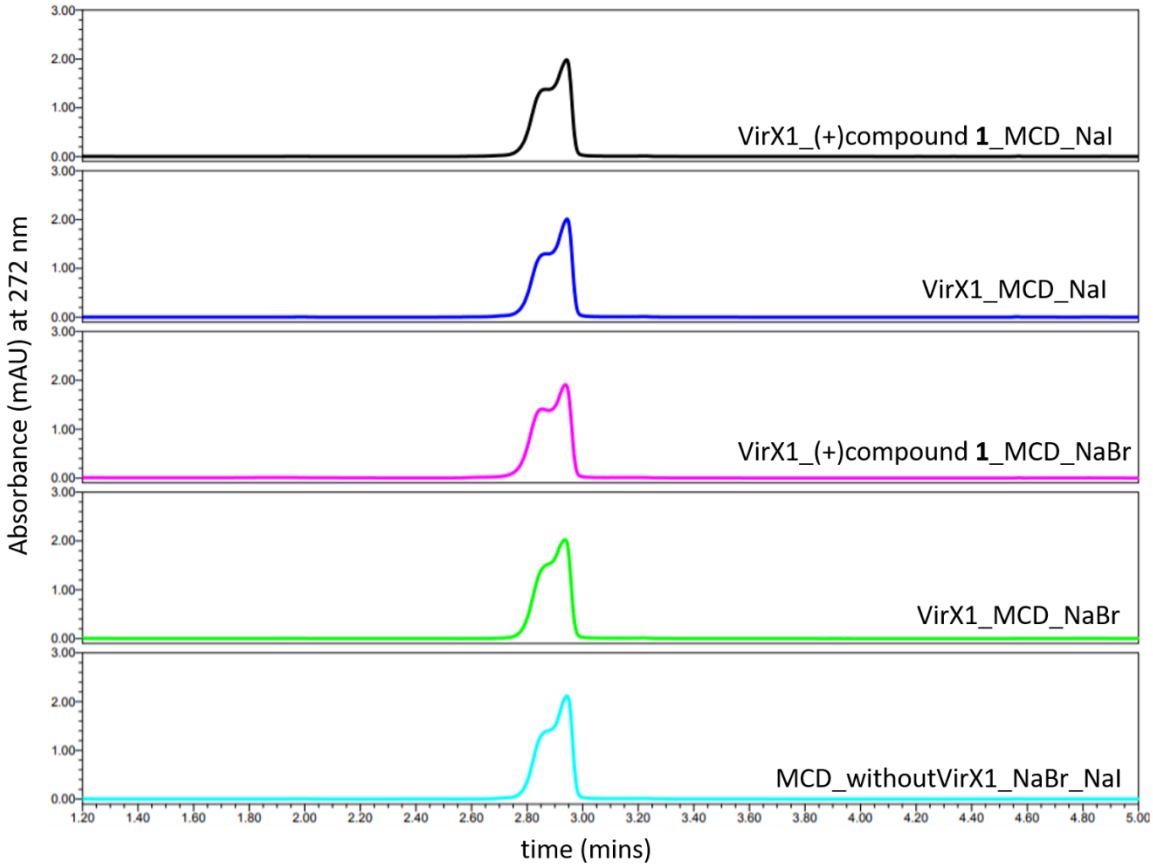
Supplementary Figure 16. UPLC chromatograms of activity assays with VirX1 mutants and WT enzyme and 6-azaindole (compound 1). Assays included 50 μM 6-azaindole (1), 10 μM of either VirX1 WT or VirX1 mutants, 1 μM of PrnF flavin reductase, 10 mM NaI 10 μM FAD, 2.5 mM NADH, in 40 mM HEPES pH 7.4. Iodinated product of VirX1 (peak eluting at 3 min) with 6-azaindole (peak eluting at 1.1 min) exhibits λmax at 337 nm, indicating that all selected mutations abolish activity of VirX.



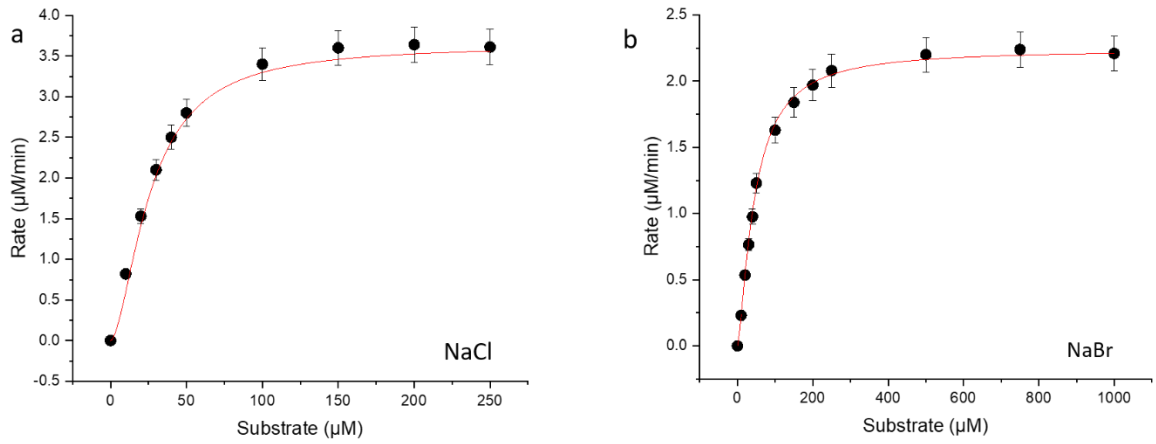
Supplementary Figure 17. UPLC chromatogram of competition experiments of VirX1 with NaI, NaBr, NaI and 6-azaindole. 3-iodo-6-azaindole elutes at 2.9 min. 3-bromo-6-azaiondole elutes at 2.6 min. Chlorinated product of 6-azaindole with VirX1 elutes at 2.2 min. 6-azaindole elutes at 1.1 min.



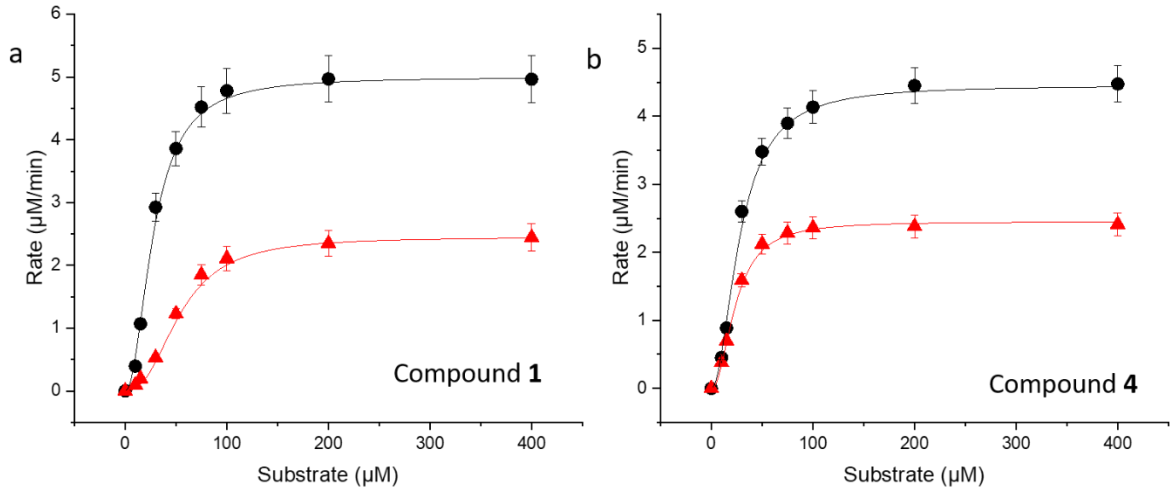
Supplementary Figure 18. UPLC chromatograms of 50 µM MCD with various concentrations of in situ generated hypoalous acids. MCD λmax at 272 nm.



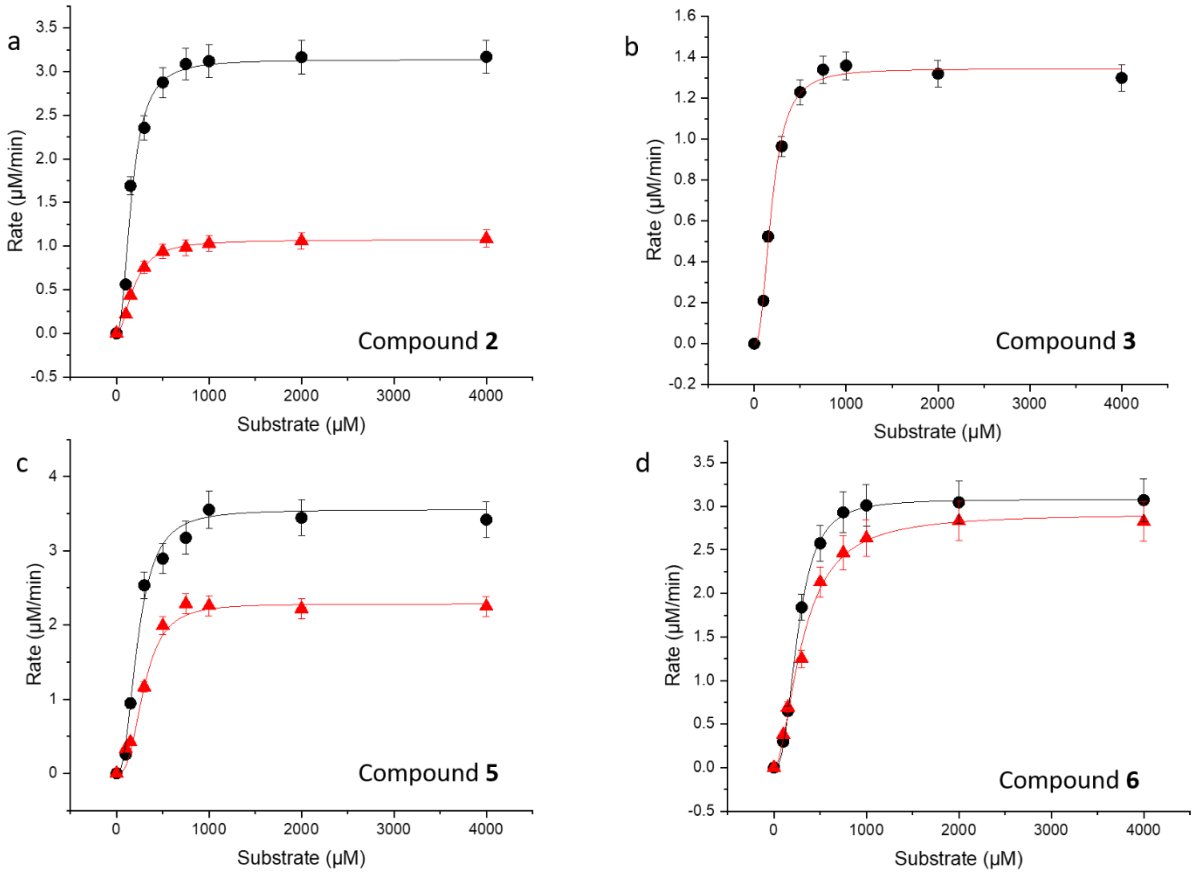
Supplementary Figure 19. UPLC chromatograms of MCD assays with VirX1 and NaI, NaBr reactions in presence and absence of one of the best substrates for VirX1, 6-azaindole (compound 1).



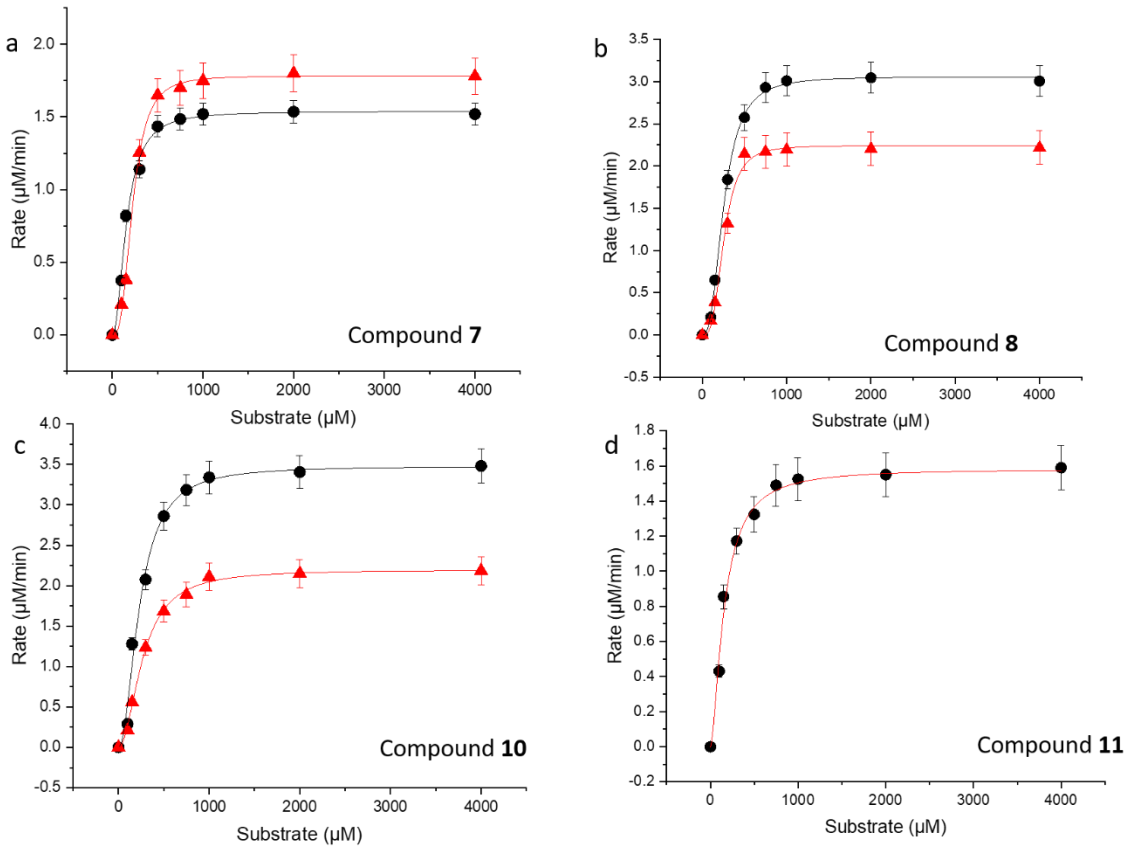
Supplementary Figure 20. Michaelis-Menten graph for PrnA and L-tryptophan with a) NaCl and b) NaBr



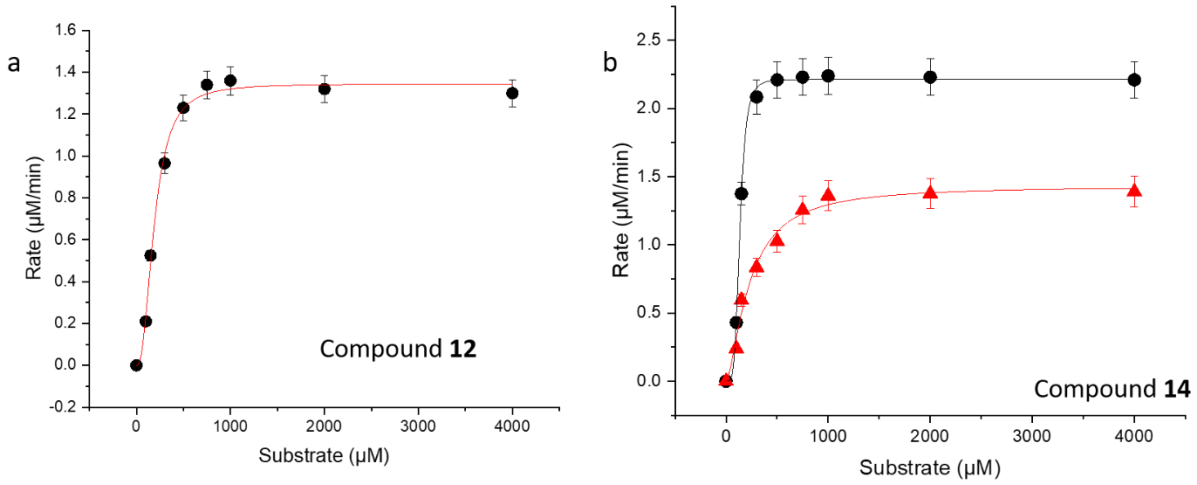
Supplementary Figure 21. Michaelis-Menten graph for VirX with a) compound 1 b) compound 4 with NaI and NaBr. Black curve is with NaI and red for NaBr.



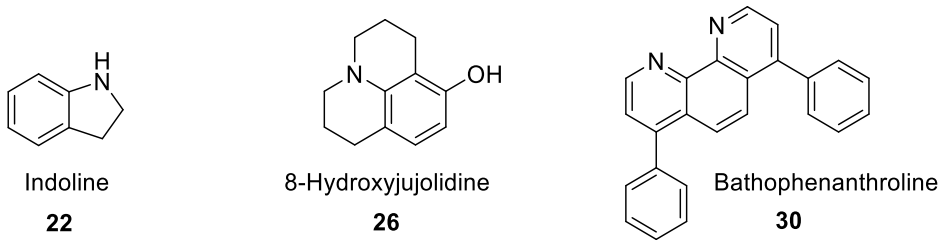
Supplementary Figure 22. Michaelis-Menten graph for VirX with a) compound 2 b) compound 3 c) compound 5 d) compound 6 with NaI and NaBr. For (a), (c) and (d) the black curve corresponds to reactions with NaI and red for NaBr. For compound **3** (b) only iodinated kinetic data were possible to be obtained.



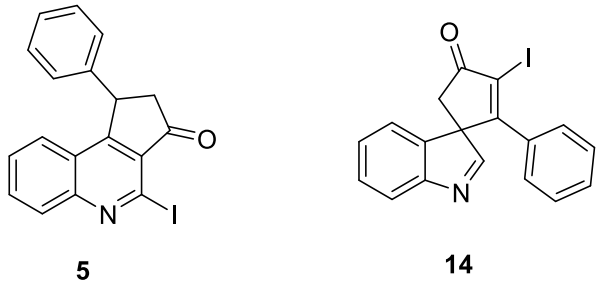
Supplementary Figure 23. Michaelis-Menten graph for VirX with a) compound 7 b) compound 8 c) compound 10 d) compound 11 with NaI and NaBr. For (a), (b) and (c) the black curve corresponds to reactions with NaI and red for NaBr. For compound **11** (d) only iodinated kinetic data were possible to be obtained.



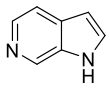
Supplementary Figure 24. Michaelis-Menten graph for VirX with a) compound 12 and b) compound 14. For (b) the black curve corresponds to reactions with NaI and red for NaBr. For compound **12** (a) only iodinated kinetic data were possible to be obtained.

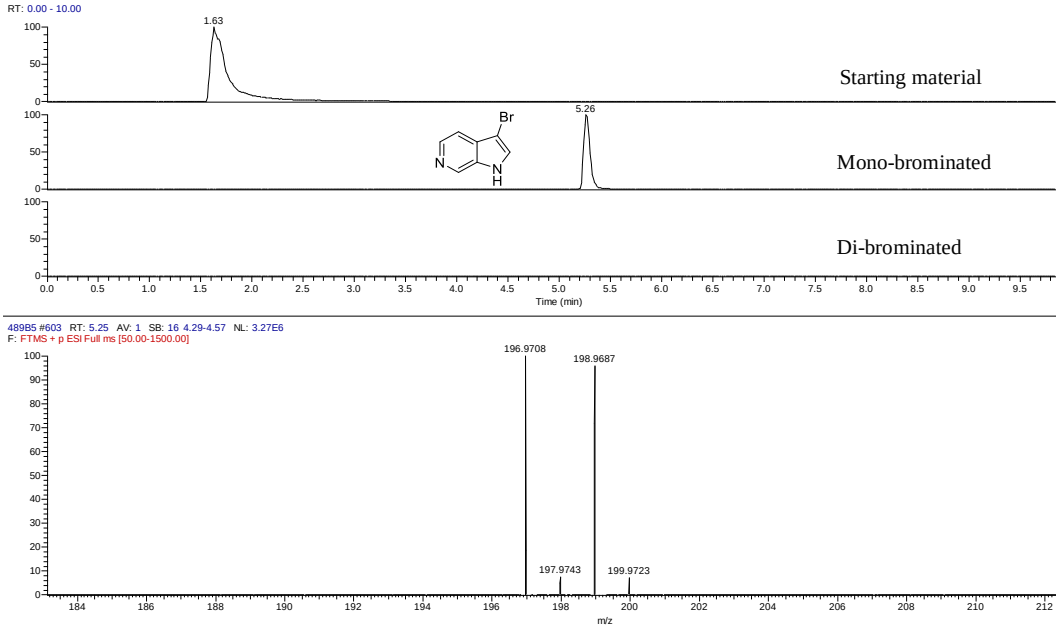


Supplementary Figure 25. Substrates that did not yield any isolable iodinated product upon reaction with equimolar amount of NIS using the general procedure given above.

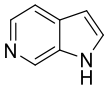


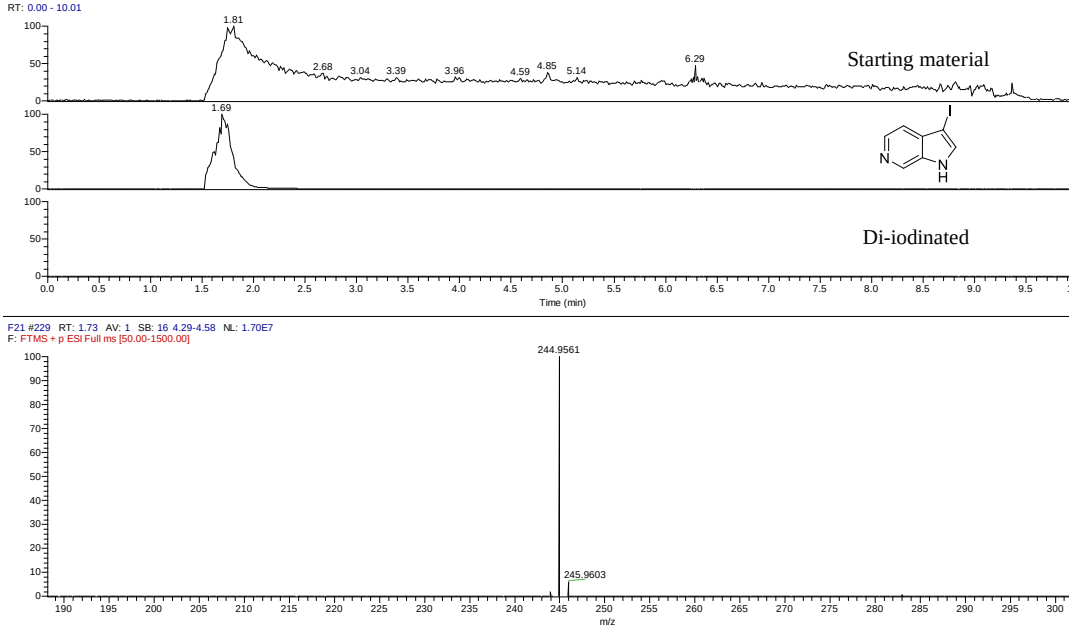
Supplementary Figure 26. Iodinated substrates synthesised via known literature methods (reference 17,18 of the manuscript).

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_7H_6BrN_2^+$ Exact Mass: 196.9709 m/z: 196.9709 (100.0%), 198.9688 (97.3%), 197.9742 (7.6%), 199.9722 (7.4%)	Chemical Formula: $C_7H_5Br_2N_2^+$ Exact Mass: 274.8814 m/z: 276.8794 (100.0%), 274.8814 (51.4%), 278.8773 (48.6%), 277.8827 (7.6%), 275.8848 (3.9%), 279.8807 (3.7%)
	Found 196.9708	Not found

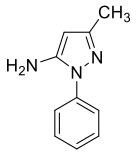


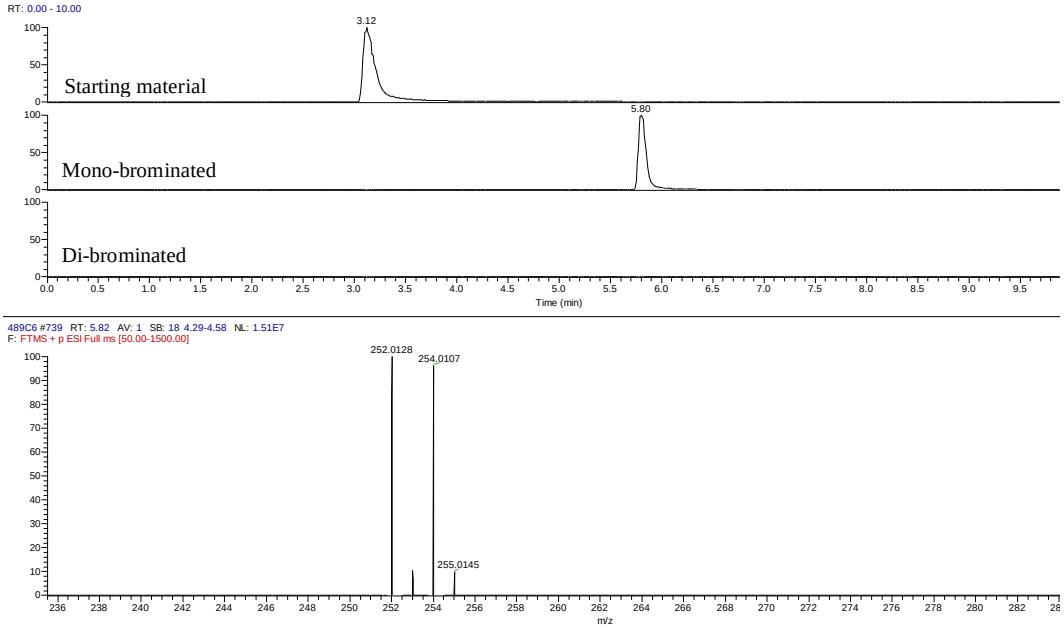
Supplementary Figure 27. LC-HRMS analysis of enzymatic bromination of compound 1 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. This product was isolated from the biotransformation and fully characterized. Regiochemistry was determined by NMR, the product was shown to be 3-Br-6-azaindole. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_7H_6IN_2^+$ Exact Mass: 244.9570 m/z: 244.9570 (100.0%), 245.9604 (7.6%)	Chemical Formula: $C_7H_5I_2N_2^+$ Exact Mass: 370.8537 m/z: 370.8537 (100.0%), 371.8570 (7.6%)
	Found 244.9561	Not found

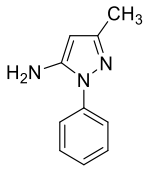


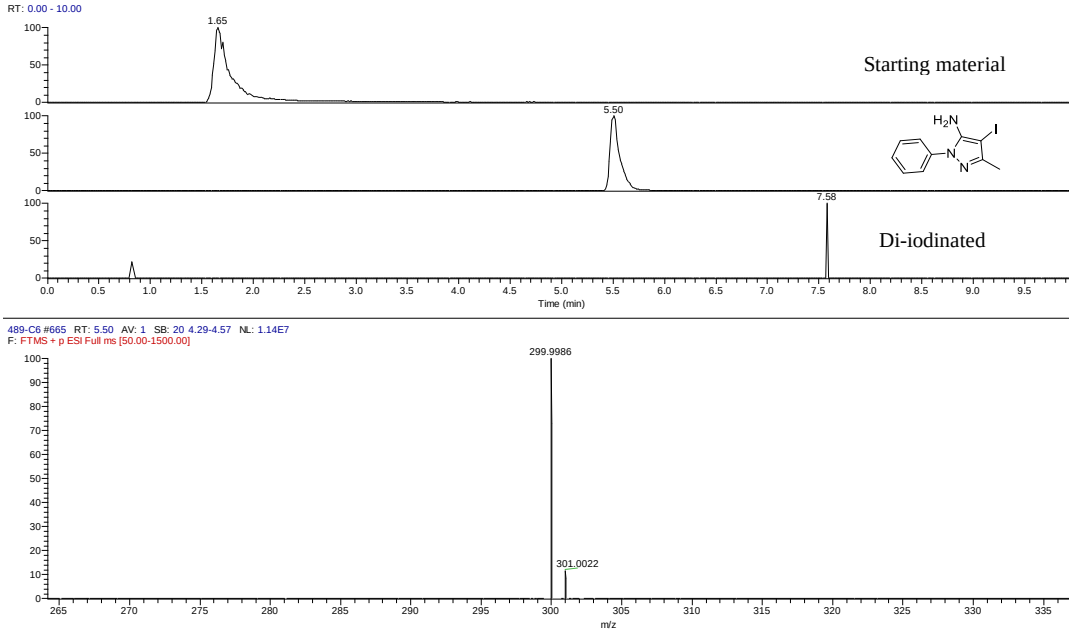
Supplementary Figure 28. LC-HRMS analysis of enzymatic iodination of compound 1 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. This product was isolated from the biotransformation and fully characterized. Regiochemistry was determined by NMR, the product was shown to be 3-iodo-6-azaindole. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_{10}H_{11}BrN_3^+$ Exact Mass: 252.0131 m/z: 252.0131 (100.0%), 254.0110 (97.3%), 253.0164 (10.8%), 255.0144 (10.5%), 253.0101 (1.1%), 255.0081 (1.1%)	Chemical Formula: $C_{10}H_{10}Br_2N_3^+$ Exact Mass: 329.9236 m/z: 331.9216 (100.0%), 329.9236 (51.4%), 333.9195 (48.6%), 332.9249 (9.7%), 334.9229 (5.3%), 330.9270 (4.4%)
	Found 252.0128	Not found

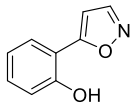


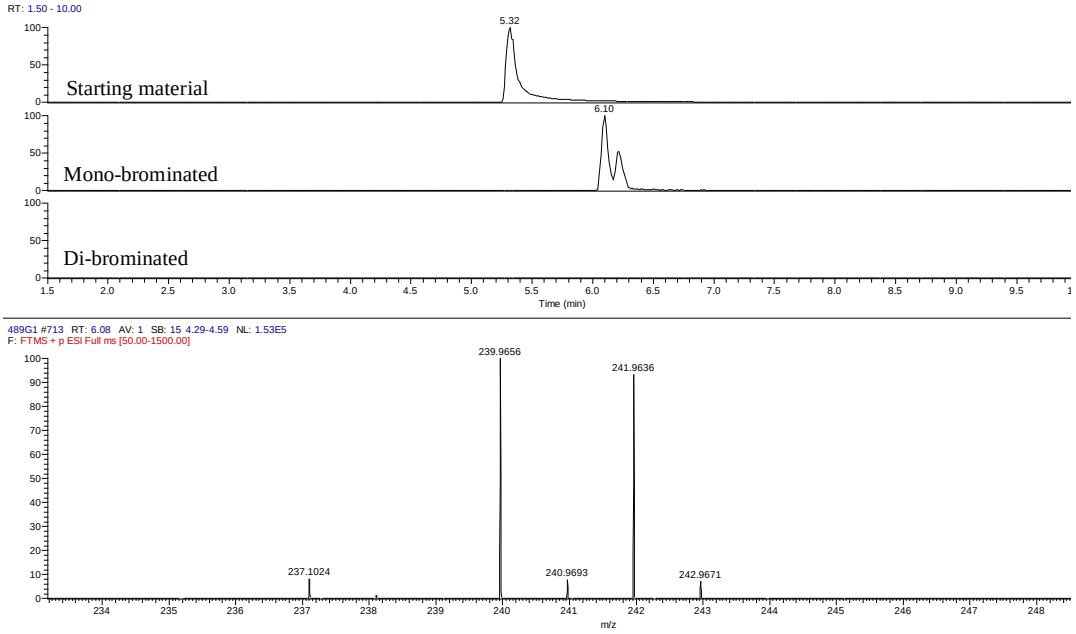
Supplementary Figure 29. LC-HRMS analysis of enzymatic bromination of compound 2 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_{10}H_{11}IN_3^+$ Exact Mass: 299.9992 m/z: 299.9992 (100.0%), 301.0026 (10.8%), 300.9963 (1.1%)	Chemical Formula: $C_{10}H_{10}I_2N_3^+$ Exact Mass: 425.8959 m/z: 425.8959 (100.0%), 426.8992 (10.8%), 426.8929 (1.1%)
	Found 299.9986	Not found

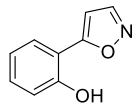


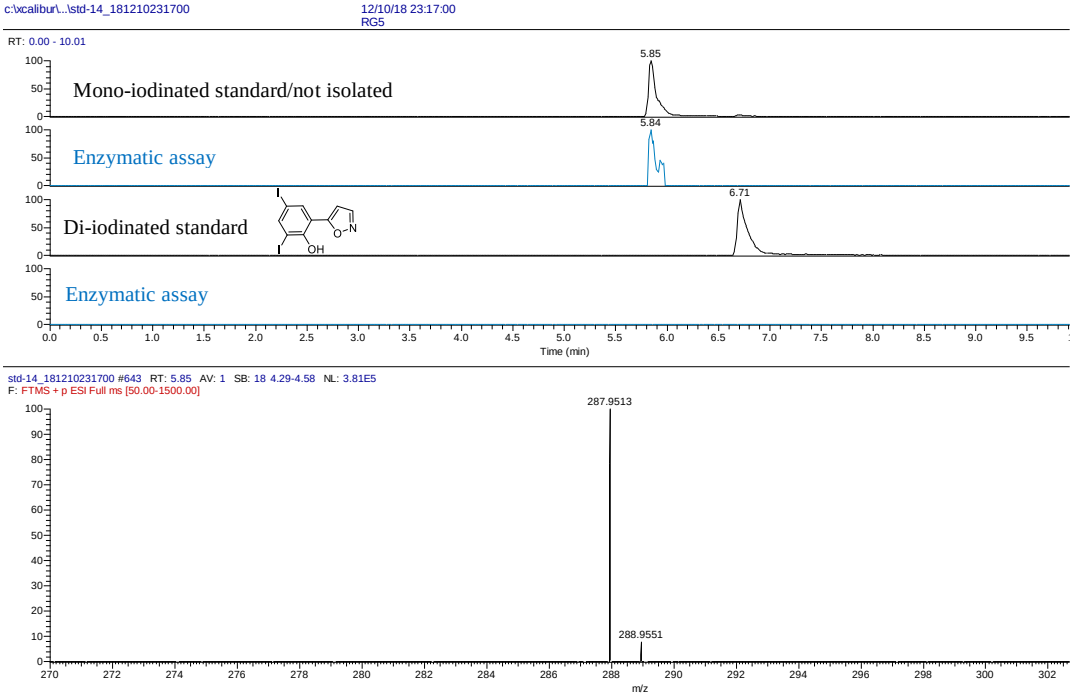
Supplementary Figure 30. LC-HRMS analysis of enzymatic iodination of compound 2 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product This product was isolated from the biotransformation and fully characterized. Regiochemistry was determined by NMR, the product was shown to be iodinated at the 4-position. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_9H_7BrNO_2^+$ Exact Mass: 239.9655 m/z: 239.9655 (100.0%), 241.9634 (97.3%), 240.9688 (9.7%), 242.9668 (9.5%)	Chemical Formula: $C_9H_6Br_2NO_2^+$ Exact Mass: 317.8760 m/z: 319.8739 (100.0%), 317.8760 (51.4%), 321.8719 (48.6%), 320.8773 (9.7%), 322.8752 (4.7%), 318.8793 (4.4%)
	Found 239.9656	Not found

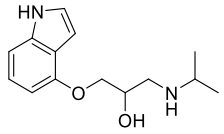


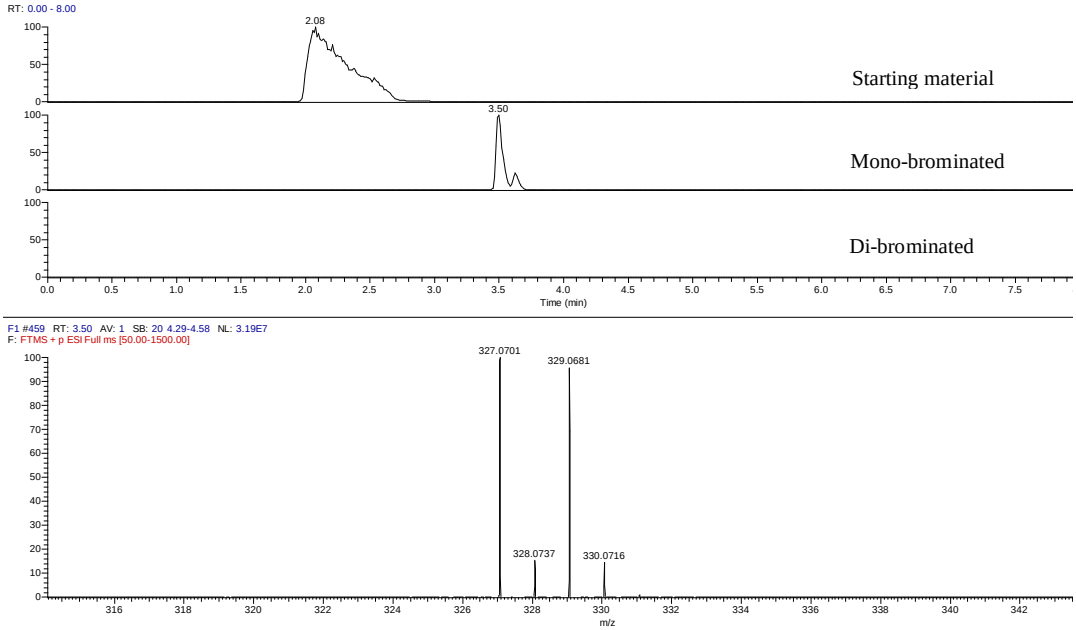
Supplementary Figure 31. LC-HRMS analysis of enzymatic bromination of compound 3 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. On EIC chromatogram (middle panel), both the major peak (6.1 min) and minor peak (6.3 min) correspond to the mono-brominated product indicating presence of potential regioisomers. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_9H_7INO_2^+$ Exact Mass: 287.9516 m/z: 287.9516 (100.0%), 288.9550 (9.7%)	Chemical Formula: $C_9H_5BrI_2NO_2^+$ Exact Mass: 491.7588 m/z: 491.7588 (100.0%), 493.7567 (97.3%), 492.7621 (9.7%), 494.7601 (9.5%)
	Found 287.9513	Not found

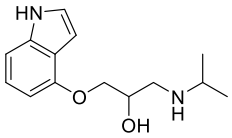


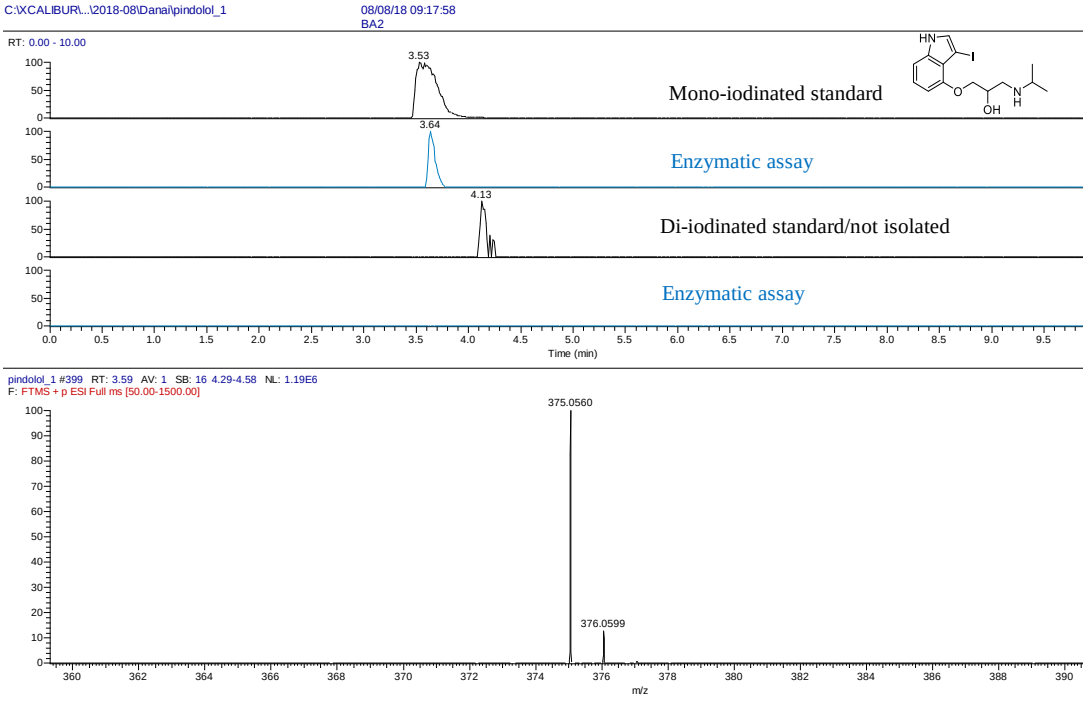
Supplementary Figure 32. LC-HRMS analysis of enzymatic iodination of compound 3 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product, and showing comparable retention time with synthetic mono-iodinated standard. On EIC chromatogram (middle panel), split peak corresponding to the mono-iodinated product was observed. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_{14}H_{20}BrN_2O_2^+$ Exact Mass: 327.0703 m/z: 327.0703 (100.0%), 329.0682 (97.3%), 328.0736 (15.1%), 330.0716 (14.7%), 329.0770 (1.1%), 331.0749 (1.0%)	Chemical Formula: $C_{14}H_{19}Br_2N_2O_2^+$ Exact Mass: 404.9808 m/z: 406.9787 (100.0%), 404.9808 (51.4%), 408.9767 (48.6%), 407.9821 (9.7%), 409.9800 (7.4%), 407.9821 (5.4%), 405.9841 (4.4%), 405.9841 (3.3%)
	Found 327.0701	Not found

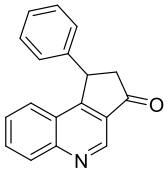


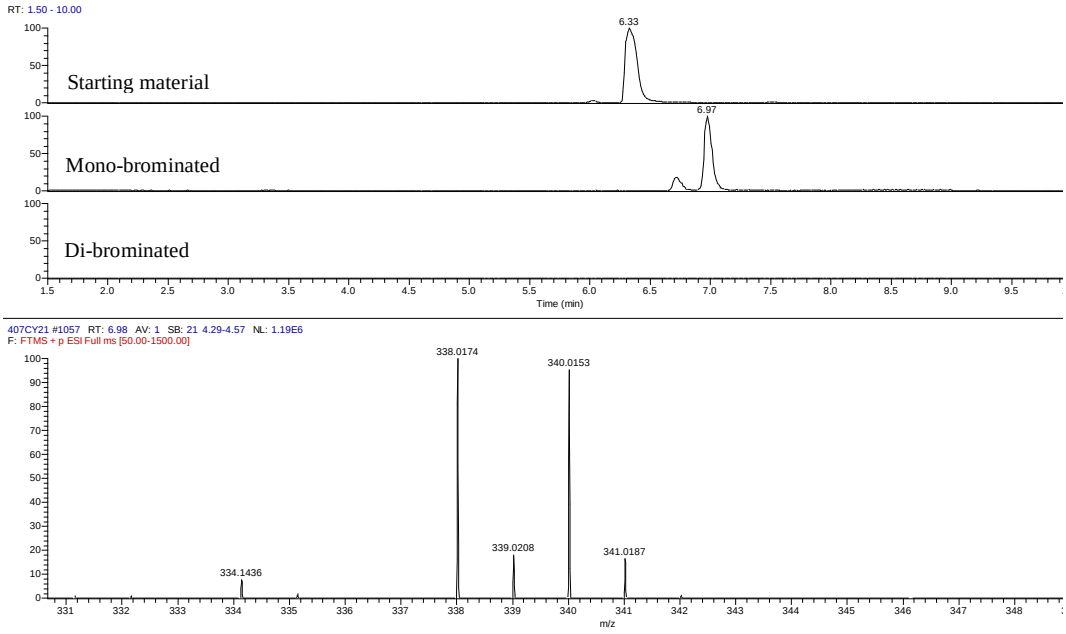
Supplementary Figure 33. LC-HRMS analysis of enzymatic bromination of compound 4 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Major peak (3.5 min) on EIC chromatogram (middle panel) corresponded to the mass of mono-brominated product, whereas minor peak (3.7 min) showed non-brominated impurity. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_{14}H_{20}IN_2O_2^+$ Exact Mass: 375.0564 m/z: 375.0564 (100.0%), 376.0598 (15.1%), 377.0631 (1.1%)	Chemical Formula: $C_{14}H_{19}I_2N_2O_2^+$ Exact Mass: 500.9530 m/z: 500.9530 (100.0%), 501.9564 (15.1%), 502.9597 (1.1%)
	Found 375.0560	Not found

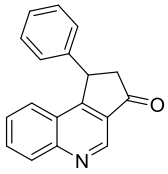


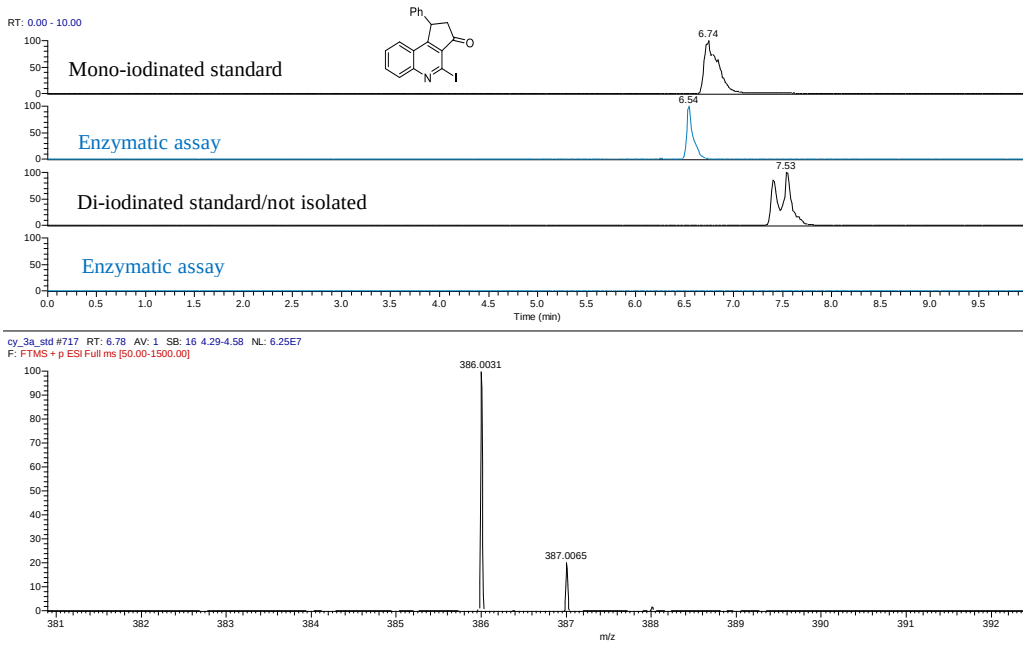
Supplementary Figure 34. LC-HRMS analysis of enzymatic iodination of compound 4 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product, and showing comparable retention time with synthetic mono-iodinated standard. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_{18}H_{13}BrNO^+$ Exact Mass: 338.0175 m/z: 338.0175 (100.0%), 340.0155 (97.3%), 341.0188 (18.9%), 339.0209 (16.2%)	Chemical Formula: $C_{18}H_{12}Br_2NO^+$ Exact Mass: 415.9280 m/z: 417.9260 (100.0%), 415.9280 (51.4%), 419.9239 (48.6%), 416.9314 (10.0%)
	Found 338.0174	Not found

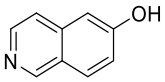


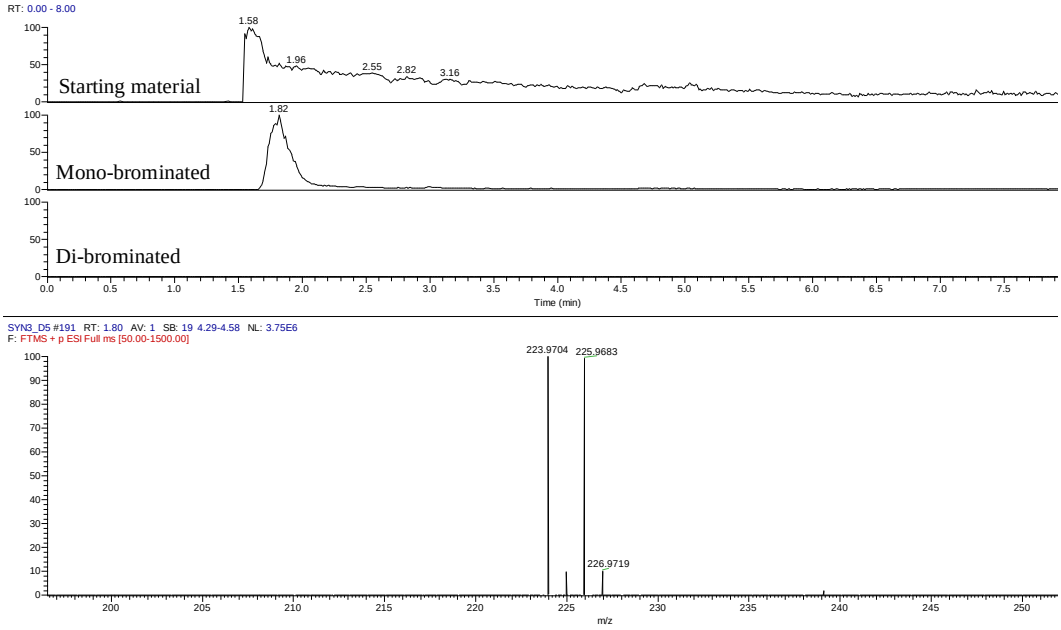
Supplementary Figure 35. LC-HRMS analysis of enzymatic bromination of compound 5 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. On EIC chromatogram (middle panel), both the major peak (6.9 min) and minor peak (6.7 min) correspond to the mono-brominated product indicating presence of potential regioisomers. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_{18}H_{13}INO^+$ Exact Mass: 386.0036 m/z: 386.0036 (100.0%), 387.0070 (19.5%), 388.0103 (1.8%)	Chemical Formula: $C_{18}H_{12}I_2NO^+$ Exact Mass: 511.9003 m/z: 511.9003 (100.0%), 512.9036 (19.5%), 513.9070 (1.8%)
	Found 386.0031	Not found

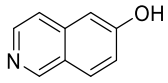


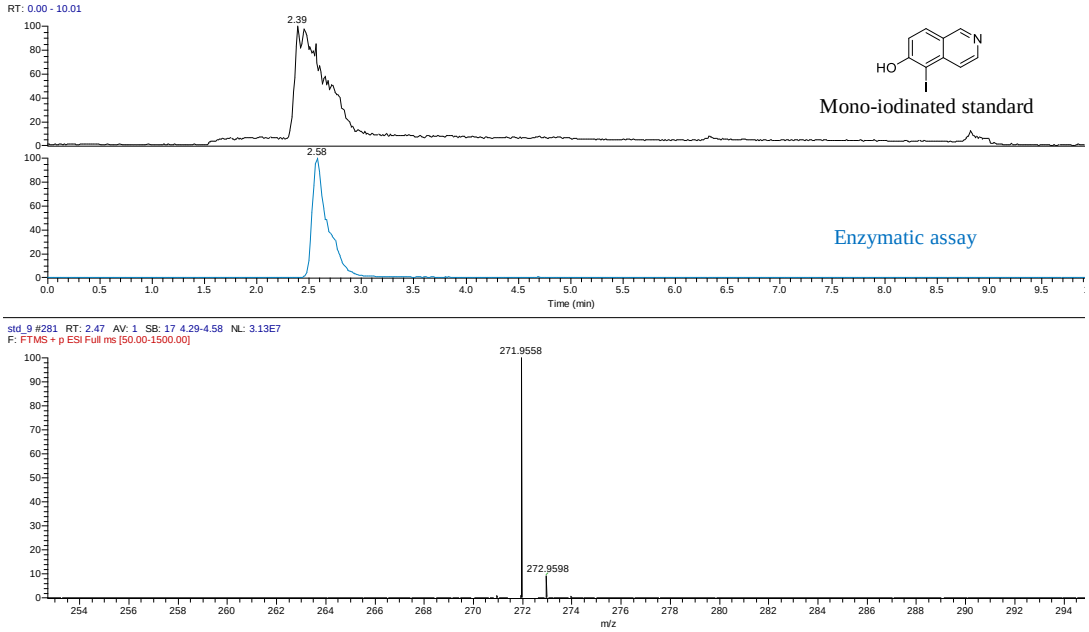
Supplementary Figure 36. LC-HRMS analysis of enzymatic iodination of compound 5 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product, and showing different retention time (6.5 min) in comparison with synthetic mono-iodinated standard (6.7 min). Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_9H_7BrNO^+$ Exact Mass: 223.9706 m/z: 223.9706 (100.0%), 225.9685 (97.3%), 224.9739 (9.7%), 226.9719 (9.5%)	Chemical Formula: $C_9H_6Br_2NO^+$ Exact Mass: 300.8790 m/z: 300.8790 (100.0%), 298.8811 (51.4%), 302.8770 (48.6%), 301.8824 (9.7%)
	Found 223.9704	Not found

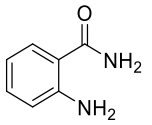


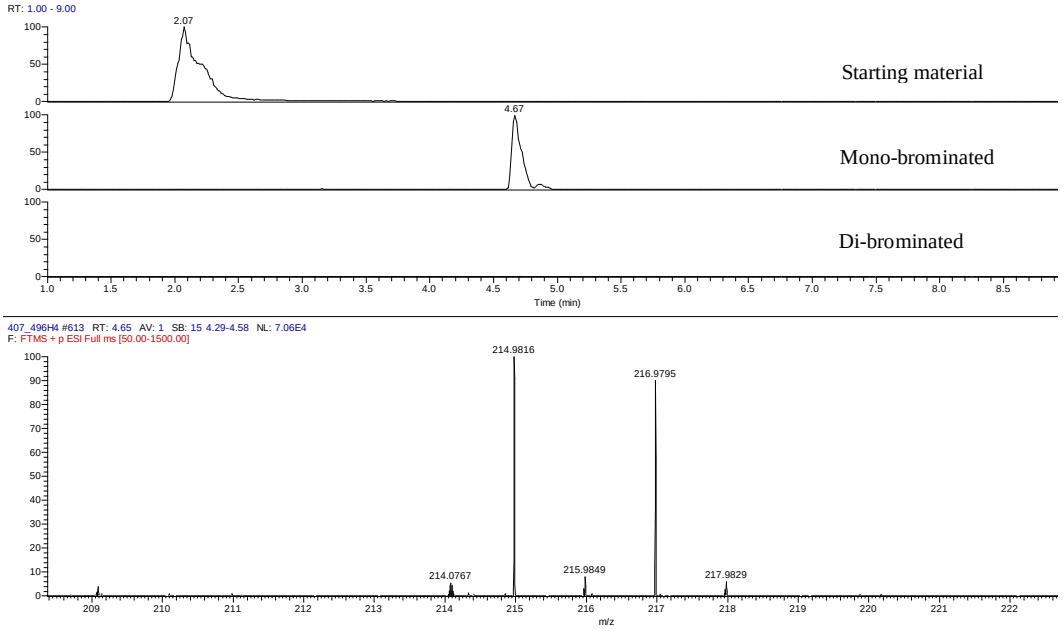
Supplementary Figure 37. LC-HRMS analysis of enzymatic bromination of compound 6 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_9H_7INO^+$ Exact Mass: 271.9567 m/z: 271.9567 (100.0%), 272.9600 (9.7%)	Chemical Formula: $C_9H_6I_2NO^+$ Exact Mass: 397.8533 m/z: 397.8533 (100.0%), 398.8567 (9.7%)
	Found 271.9558	Not found

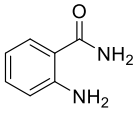


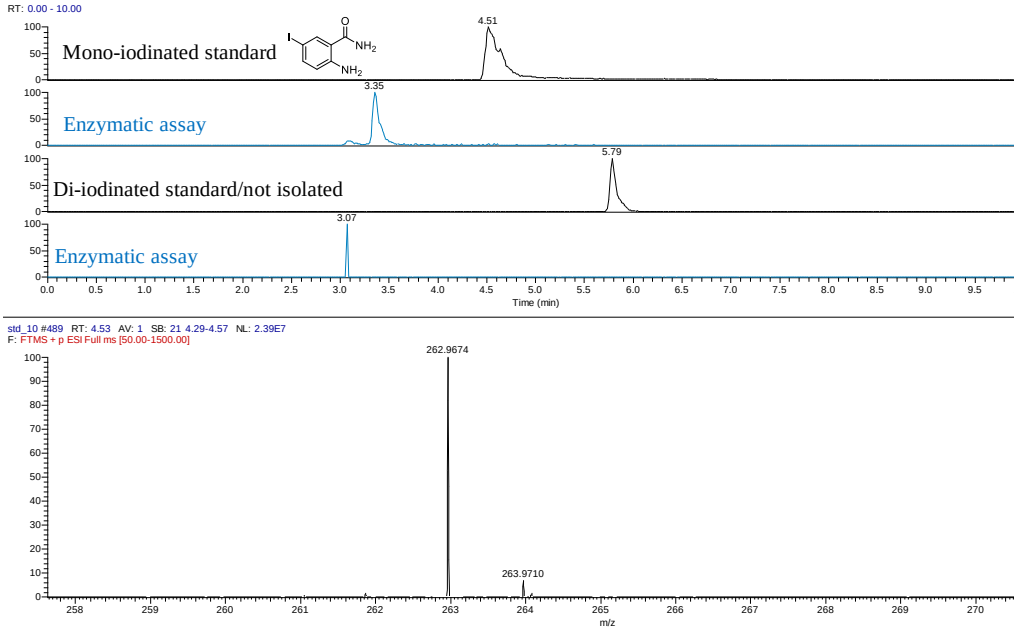
Supplementary Figure 38. LC-HRMS analysis of enzymatic iodination of compound 6 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product, and showing comparable retention time with synthetic mono-iodinated standard. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_7H_8BrN_2O^+$ Exact Mass: 214.9815 m/z: 214.9815 (100.0%), 216.9794 (97.3%), 215.9848 (7.6%), 217.9828 (7.4%)	Chemical Formula: $C_7H_7Br_2N_2O^+$ Exact Mass: 292.8920 m/z: 294.8899 (100.0%), 292.8920 (51.4%), 296.8879 (48.6%), 295.8933 (7.6%), 293.8953 (3.9%), 297.8912 (3.7%)
	Found 214.9816	Not found

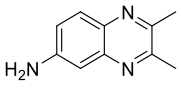


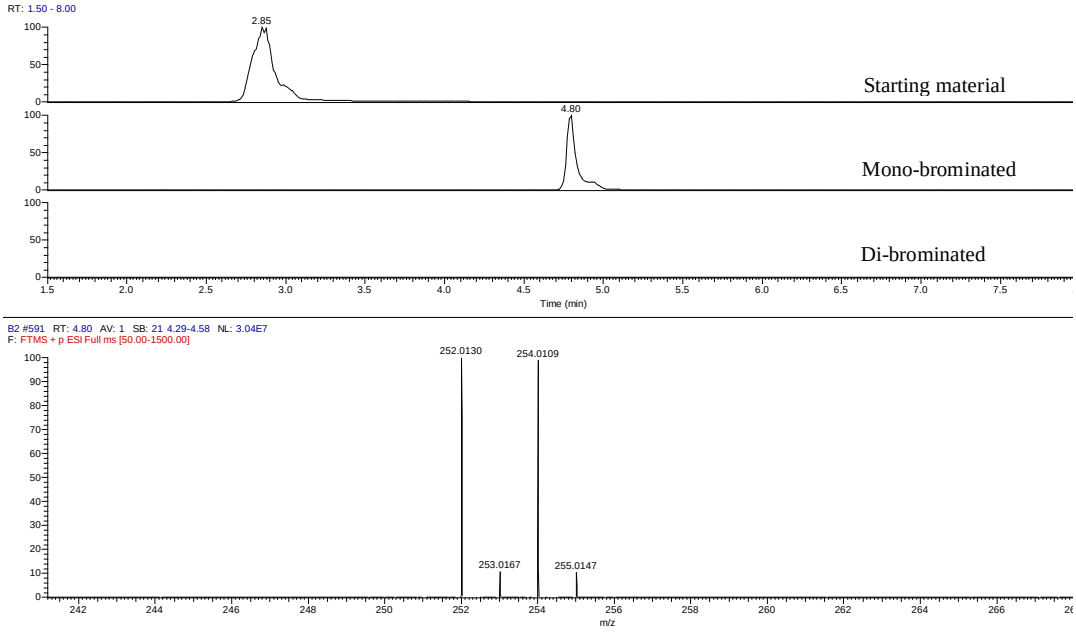
Supplementary Figure 39. LC-HRMS analysis of enzymatic bromination of compound 7 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_7H_8IN_2O^+$ Exact Mass: 262.9676 m/z: 262.9676 (100.0%), 263.9709 (7.6%)	Chemical Formula: $C_7H_7I_2N_2O^+$ Exact Mass: 388.8642 m/z: 388.8642 (100.0%), 389.8676 (7.6%)
	Found 262.9674	Not found

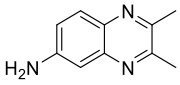


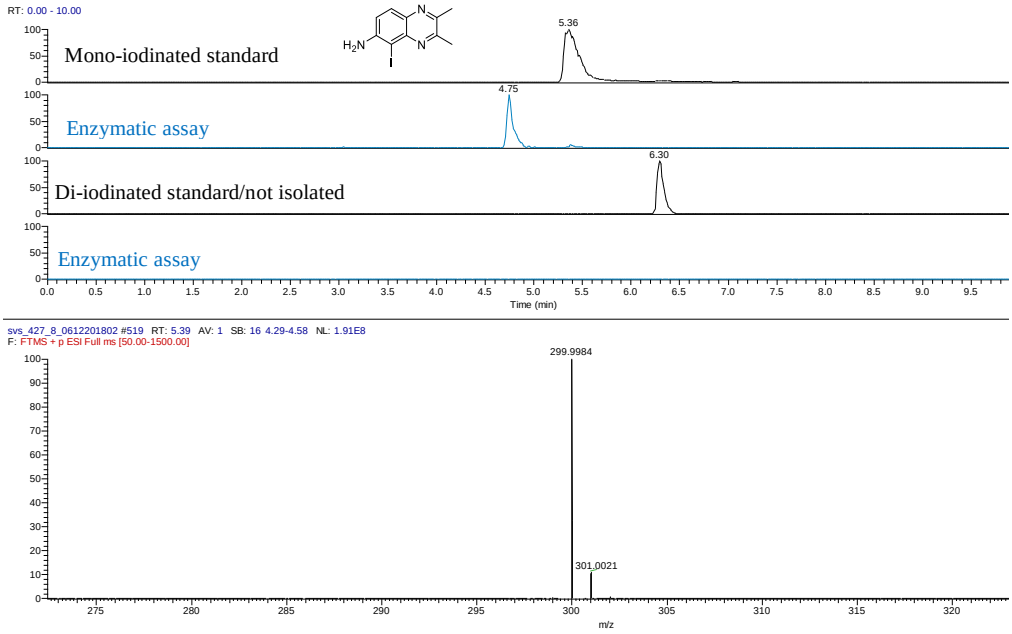
Supplementary Figure 40. LC-HRMS analysis of enzymatic iodination of compound 7 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product, and showing different retention time with synthetic mono-iodinated standard (indicating different regioisomer). Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_{10}H_{11}BrN_3^+$ Exact Mass: 252.0131 m/z: 252.0131 (100.0%), 254.0110 (97.3%), 253.0164 (10.8%), 255.0144 (10.5%)	Chemical Formula: $C_{10}H_{10}Br_2N_3^+$ Exact Mass: 329.9236 m/z: 331.9216 (100.0%), 329.9236 (51.4%), 333.9195 (48.6%), 332.9249 (9.7%)
	Found 252.0130	Not found

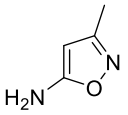


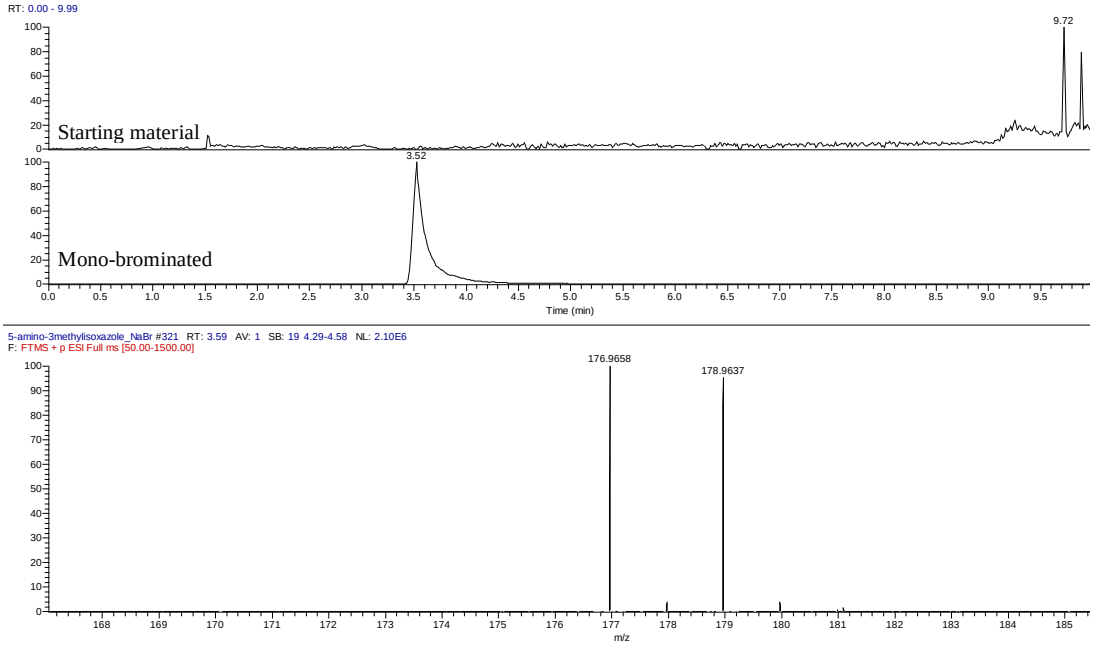
Supplementary Figure 41. LC-HRMS analysis of enzymatic bromination of compound 8 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_{10}H_{11}IN_3^+$ Exact Mass: 299.9992 m/z: 299.9992 (100.0%), 301.0026 (10.8%), 300.9963 (1.1%)	Chemical Formula: $C_{10}H_{10}I_2N_3^+$ Exact Mass: 425.8959 m/z: 425.8959 (100.0%), 426.8992 (10.8%), 426.8929 (1.1%)
	Found 299.9984	Not found

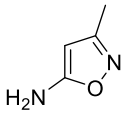


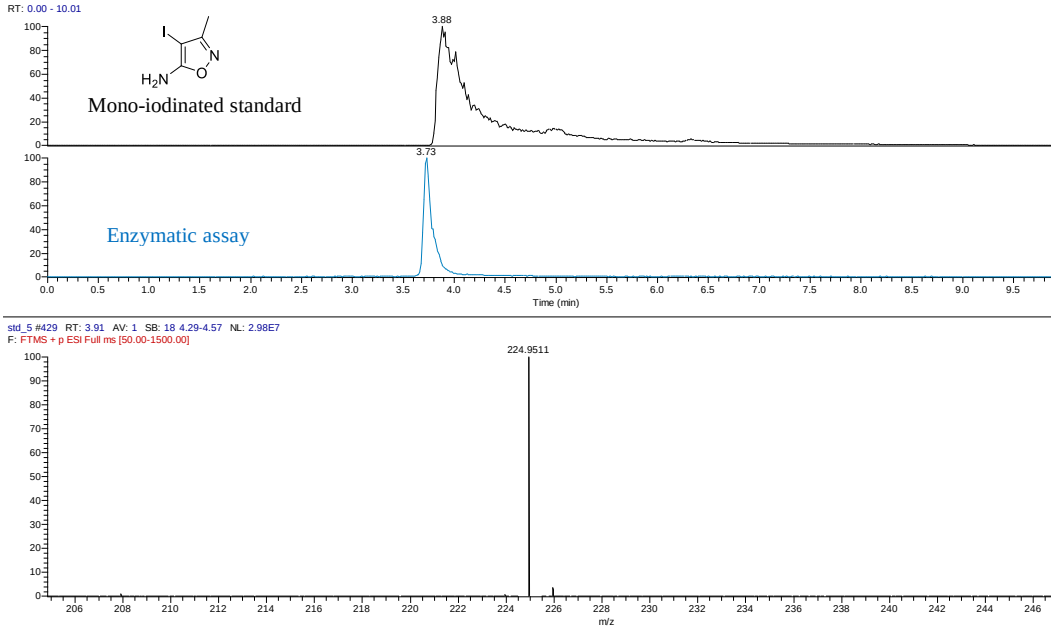
Supplementary Figure 42. LC-HRMS analysis of enzymatic iodination of compound 8 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product, and showing different retention time (4.7 min) in comparison with synthetic mono-iodinated standard (5.4 min) indicating different regioisomer. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated
	Chemical Formula: $C_4H_6BrN_2O^+$ Exact Mass: 176.9658 m/z: 176.9658 (100.0%), 178.9638 (97.3%), 177.9692 (4.3%), 179.9671 (4.2%)
	Found 176.9658

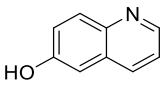


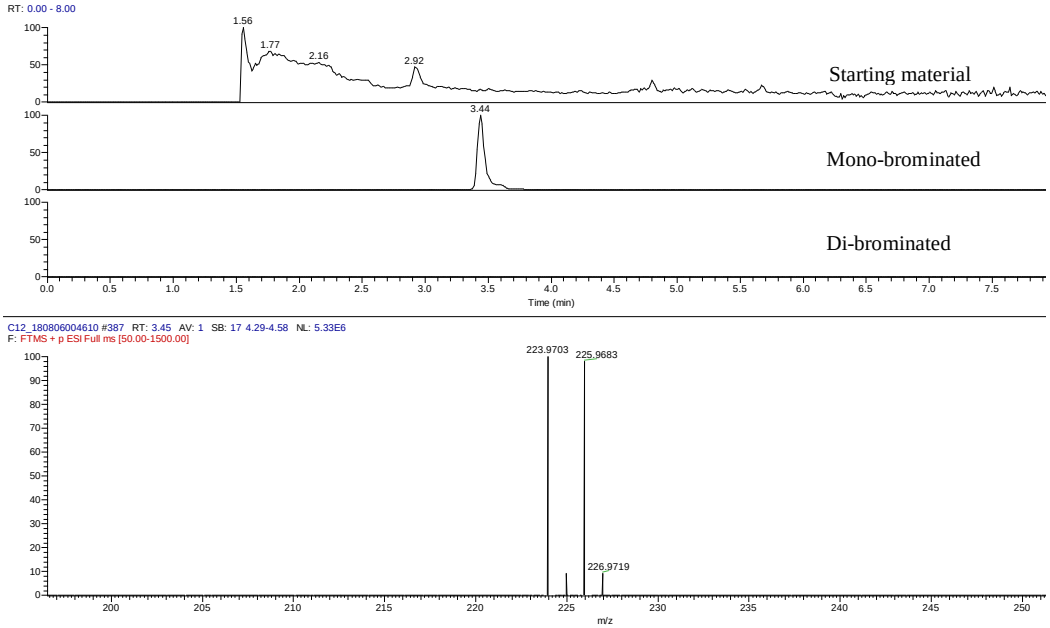
Supplementary Figure 43. LC-HRMS analysis of enzymatic bromination of compound 9 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated
	Chemical Formula: $C_4H_6IN_2O^+$ Exact Mass: 224.9519 m/z: 224.9519 (100.0%), 225.9553 (4.3%)
	Found 224.9511

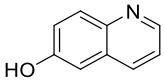


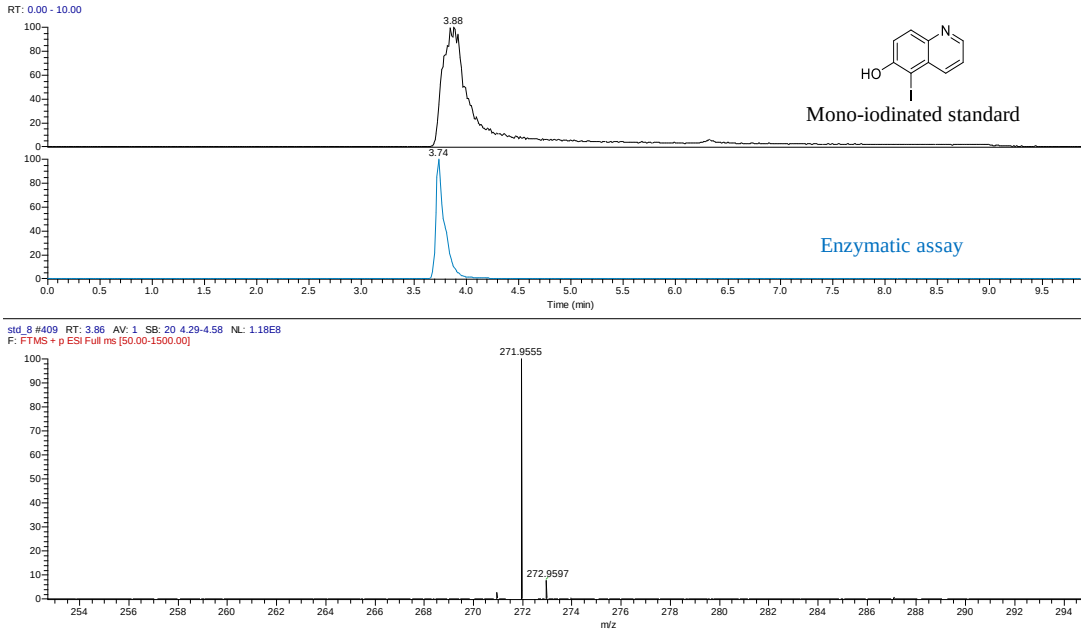
Supplementary Figure 44. LC-HRMS analysis of enzymatic iodination of compound 9 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product, and showing comparable retention time with synthetic mono-iodinated standard. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_9H_7BrNO^+$ Exact Mass: 223.9706 m/z: 223.9706 (100.0%), 225.9685 (97.3%), 224.9739 (9.7%), 226.9719 (9.5%)	Chemical Formula: $C_9H_6Br_2NO^+$ Exact Mass: 301.8811 m/z: 303.8790 (100.0%), 301.8811 (51.4%), 305.8770 (48.6%), 304.8824 (9.7%), 306.8803 (4.7%), 302.8844 (4.4%)
	Found 223.9703	**Not found**

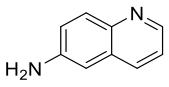


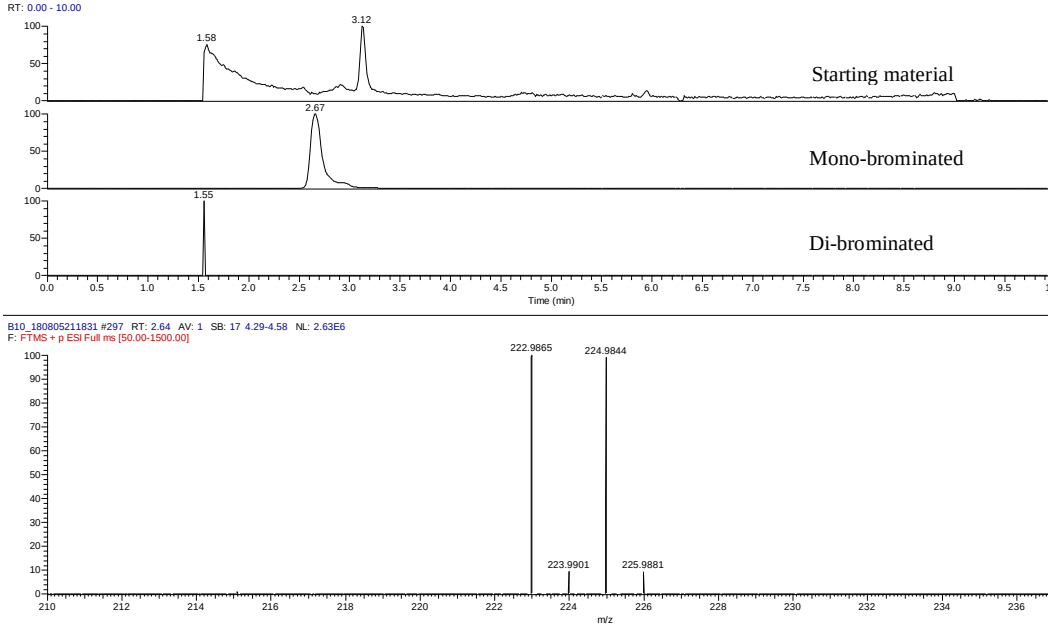
Supplementary Figure 45. LC-HRMS analysis of enzymatic bromination of compound 10 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_9H_7INO^+$ Exact Mass: 271.9567 m/z: 271.9567 (100.0%), 272.9600 (9.7%)	Chemical Formula: $C_9H_6I_2NO^+$ Exact Mass: 397.8533 m/z: 397.8533 (100.0%), 398.8567 (9.7%)
	Found 271.9555	Not found

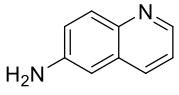


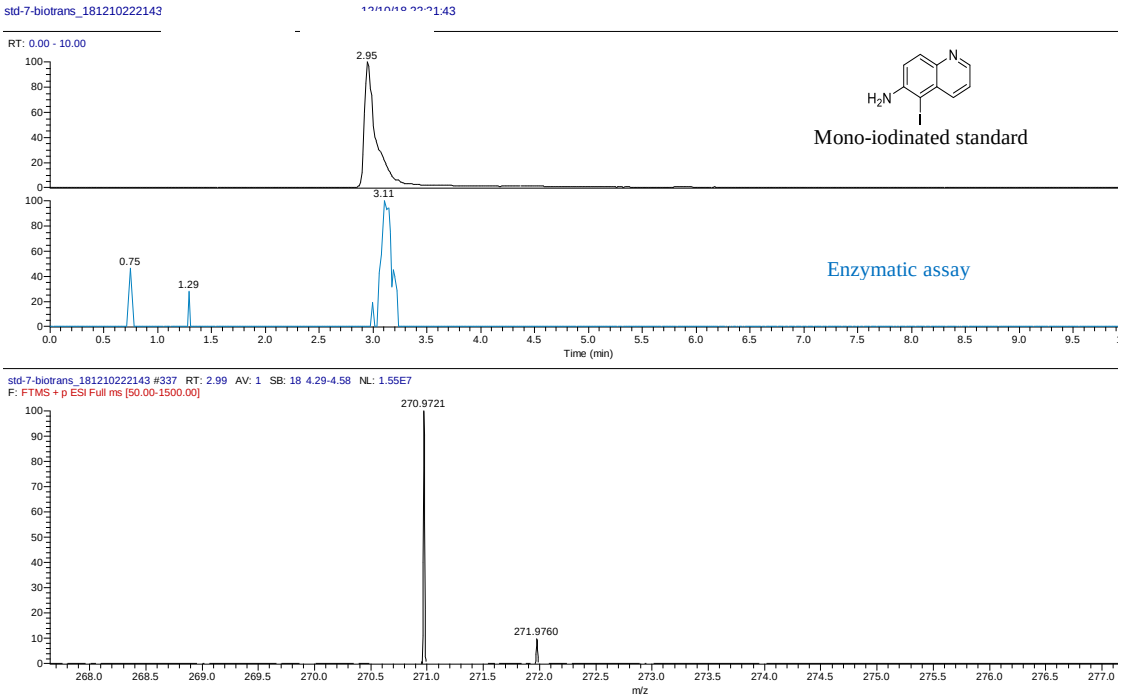
Supplementary Figure 46. LC-HRMS analysis of enzymatic iodination of compound 10 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product, and showing comparable retention time with synthetic mono-iodinated standard. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_9H_8BrN_2^+$ Exact Mass: 222.9865 m/z: 222.9865 (100.0%), 224.9845 (97.3%), 223.9899 (9.7%), 225.9878 (9.5%)	Chemical Formula: $C_9H_7Br_2N_2^+$ Exact Mass: 300.8970 m/z: 302.8950 (100.0%), 300.8970 (51.4%), 304.8930 (48.6%), 303.8984 (9.7%), 305.8963 (4.7%), 301.9004 (4.4%)
	Found 222.9865	Not found

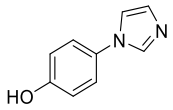


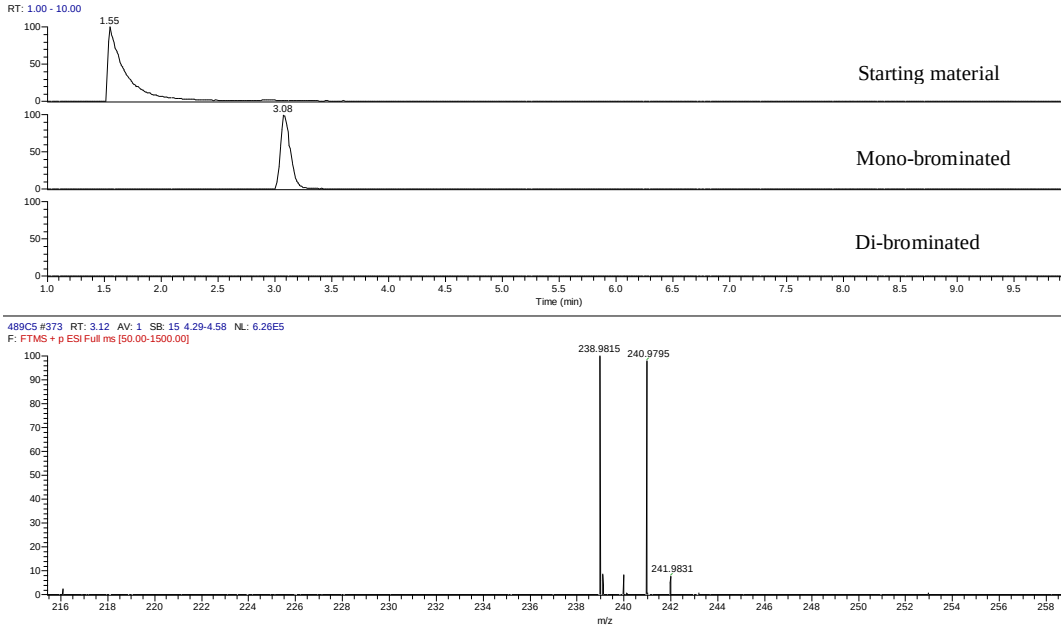
Supplementary Figure 47. LC-HRMS analysis of enzymatic bromination of compound 11 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_9H_8IN_2^+$ Exact Mass: 270.9727 m/z: 270.9727 (100.0%), 271.9760 (9.7%)	Chemical Formula: $C_9H_7I_2N_2^+$ Exact Mass: 396.8693 m/z: 396.8693 (100.0%), 397.8727 (9.7%)
	Found 270.9721	Not found

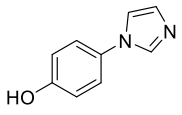


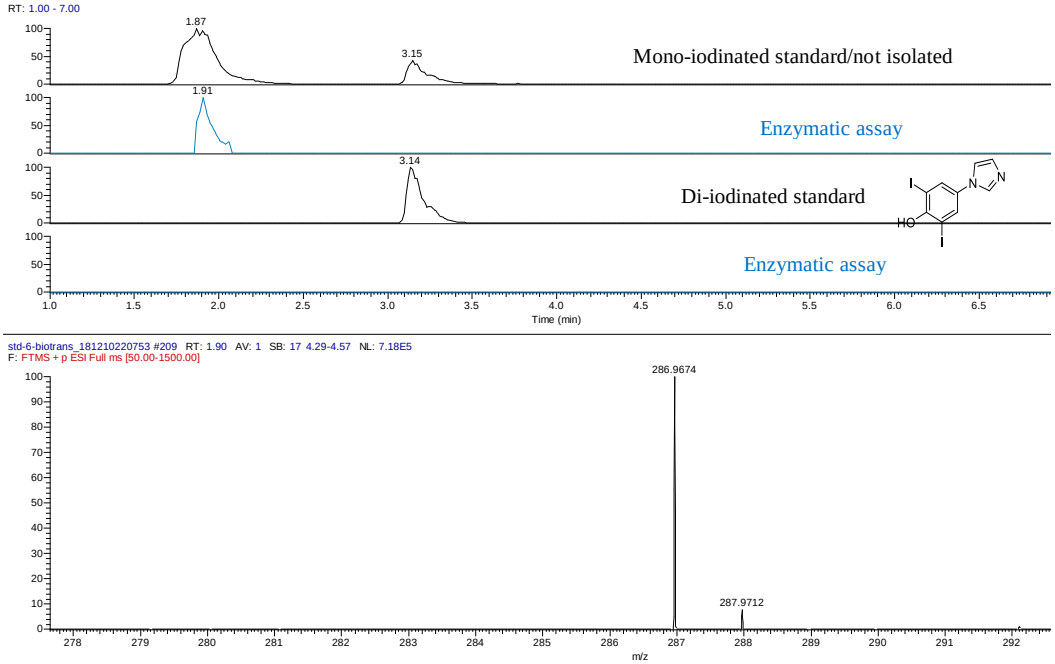
Supplementary Figure 48. LC-HRMS analysis of enzymatic iodination of compound 11 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product, and showing comparable retention time with synthetic mono-iodinated standard. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_9H_8BrN_2O^+$ Exact Mass: 238.9815 m/z: 238.9815 (100.0%), 240.9794 (97.3%), 239.9848 (9.7%), 241.9828 (9.5%)	Chemical Formula: $C_9H_7Br_2N_2O^+$ Exact Mass: 316.8920 m/z: 318.8899 (100.0%), 316.8920 (51.4%), 320.8879 (48.6%), 319.8933 (9.7%), 321.8912 (4.7%), 317.8953 (4.4%)
	Found 238.9815	Not found

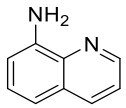


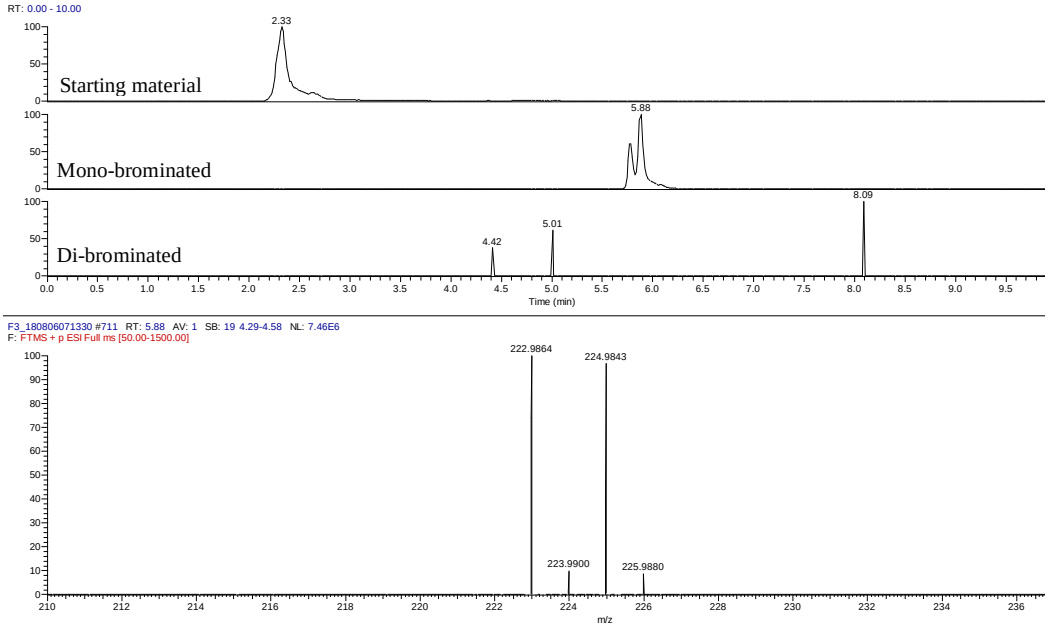
Supplementary Figure 49. LC-HRMS analysis of enzymatic bromination of compound 12 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_9H_8IN_2O^+$ Exact Mass: 286.9676 m/z: 286.9676 (100.0%), 287.9709 (9.7%)	Chemical Formula: $C_9H_7I_2N_2O^+$ Exact Mass: 412.8642 m/z: 412.8642 (100.0%), 413.8676 (9.7%)
	Found 286.9674	Not found

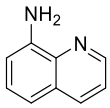


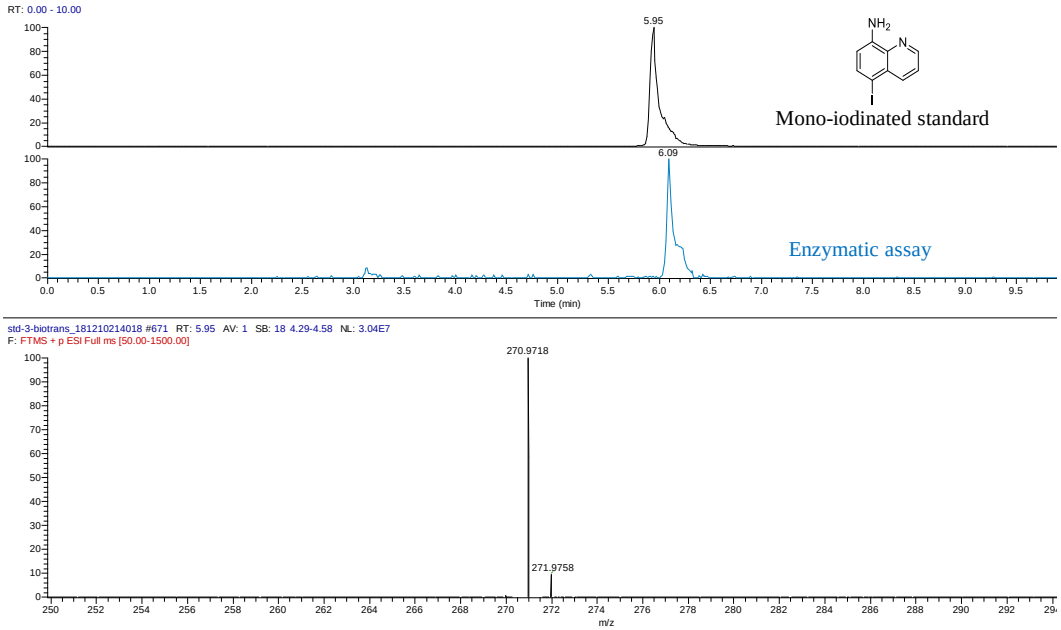
Supplementary Figure 50. LC-HRMS analysis of enzymatic iodination of compound 12 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product, and showing comparable retention time with synthetic mono-iodinated standard. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_9H_8BrN_2^+$ Exact Mass: 222.9865 m/z: 222.9865 (100.0%), 224.9845 (97.3%), 223.9899 (9.7%), 225.9878 (9.5%)	Chemical Formula: $C_9H_7Br_2N_2^+$ Exact Mass: 300.8970 m/z: 302.8950 (100.0%), 300.8970 (51.4%), 304.8930 (48.6%), 303.8984 (9.7%), 305.8963 (4.7%), 301.9004 (4.4%)
	Found 222.9864	Not found

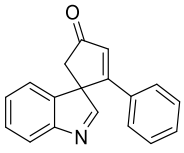


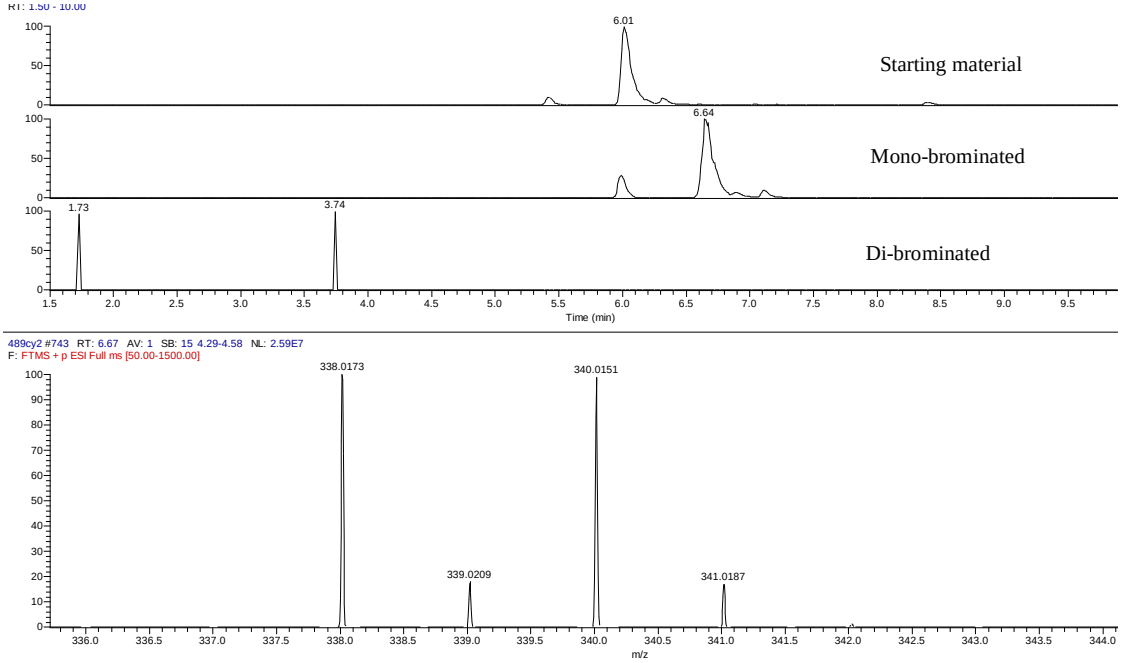
Supplementary Figure 51. LC-HRMS analysis of enzymatic bromination of compound 13 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. On EIC chromatogram (middle panel), both the major peak (5.9 min) and minor peak (5.6 min) correspond to the mono-brominated product indicating presence of potential regioisomers. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_9H_8IN_2^+$ Exact Mass: 270.9727 m/z: 270.9727 (100.0%), 271.9760 (9.7%)	Chemical Formula: $C_9H_7I_2N_2^+$ Exact Mass: 396.8693 m/z: 396.8693 (100.0%), 397.8727 (9.7%)
	Found 270.9718	Not found

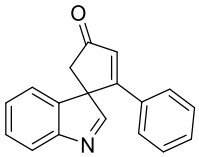


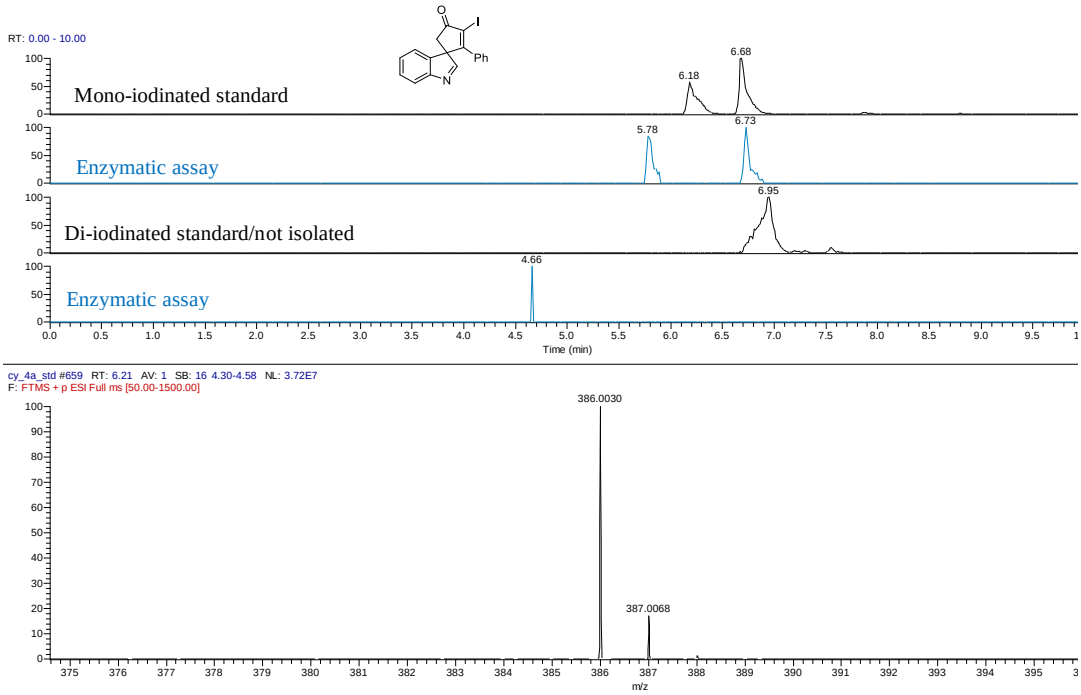
Supplementary Figure 52. LC-HRMS analysis of enzymatic iodination of compound 13 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product, and showing comparable retention time with synthetic mono-iodinated standard.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_{18}H_{13}BrNO^+$ Exact Mass: 338.0175 m/z: 338.0175 (100.0%), 340.0155 (97.3%), 341.0188 (18.9%), 339.0209 (16.2%), 339.0209 (3.2%), 342.0222 (1.2%), 340.0242 (1.1%)	Chemical Formula: $C_{18}H_{12}Br_2NO^+$ Exact Mass: 415.9280 m/z: 417.9260 (100.0%), 415.9280 (51.4%), 419.9239 (48.6%), 416.9314 (10.0%), 418.9293 (9.7%), 418.9293 (9.7%), 420.9273 (9.5%), 419.9327 (1.1%)
	Found 338.0175	Not found

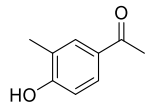


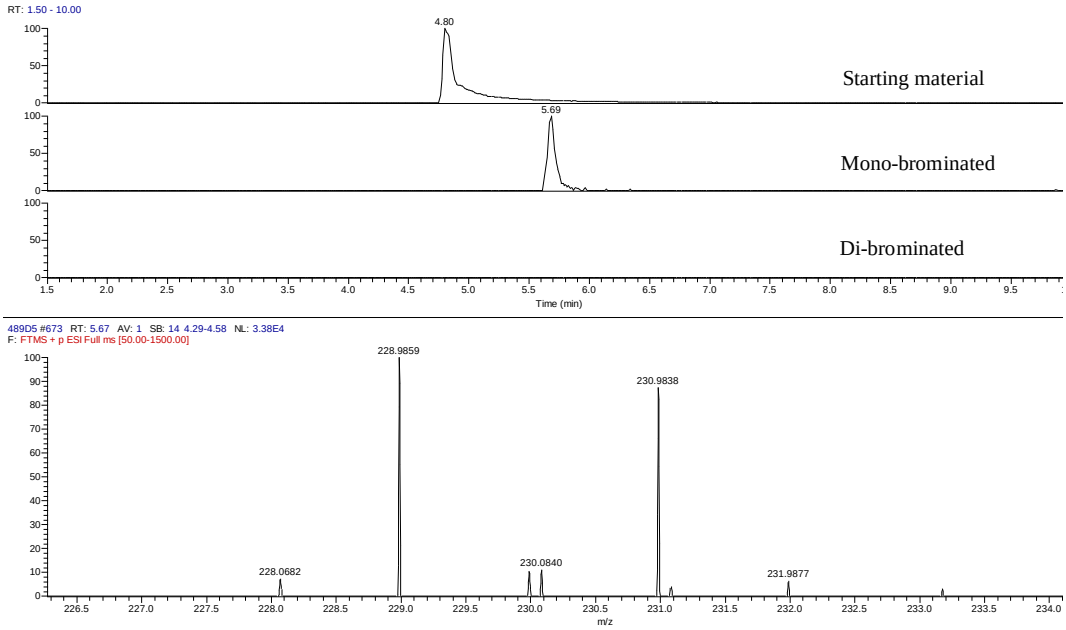
Supplementary Figure 53. LC-HRMS analysis of enzymatic bromination of compound 14 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Major peak (retention 6.6 min) on EIC (middle panel) corresponds to the mono-brominated product, whereas minor peak (retention 6.0 min) revealed mass of the starting material. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_{18}H_{13}INO^+$ Exact Mass: 386.0036 m/z: 386.0036 (100.0%), 387.0070 (19.5%), 388.0103 (1.8%)	Chemical Formula: $C_{18}H_{12}I_2NO^+$ Exact Mass: 511.9003 m/z: 511.9003 (100.0%), 512.9036 (19.5%), 513.9070 (1.8%)
	Found 386.0030	Not found

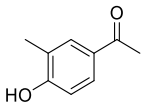


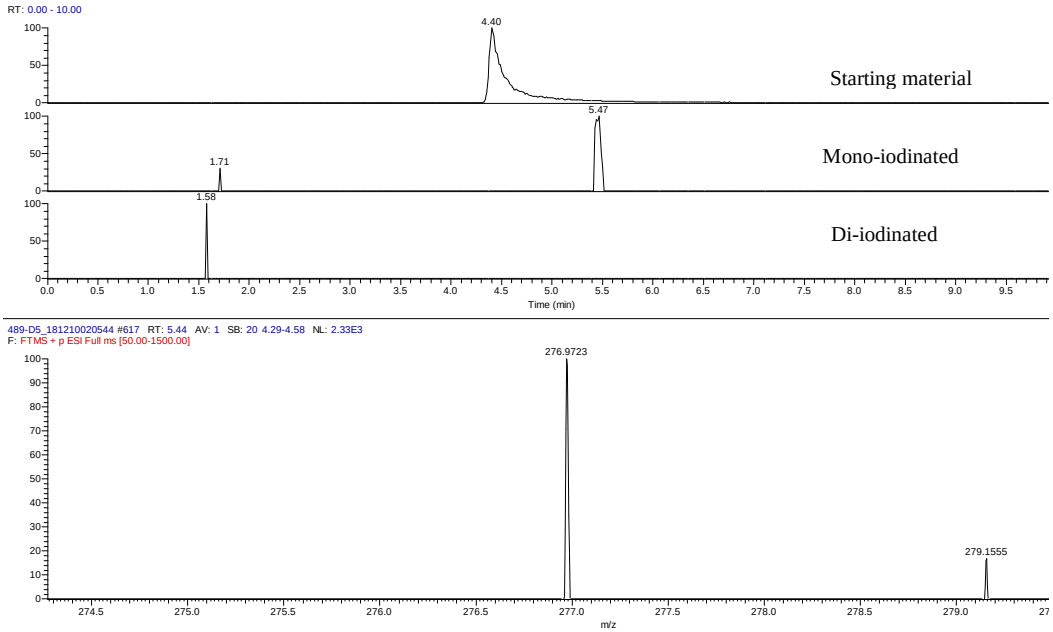
Supplementary Figure 54. LC-HRMS analysis of enzymatic iodination of compound 14 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product, and showing comparison with synthetic mono-iodinated standard. Each of the two peaks observed on EIC of the synthetic standard (6.1 and 6.7 min) as well as enzymatic reaction (5.7 and 6.7 min) corresponded to the mass of mono-iodinated product (potentially either regioisomers or diasteromers). Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_9H_{10}BrO_2^+$ Exact Mass: 228.9859 m/z: 228.9859 (100.0%), 230.9838 (97.3%), 229.9892 (9.7%), 231.9872 (9.5%)	Chemical Formula: $C_9H_9Br_2O_2^+$ Exact Mass: 306.8964 m/z: 308.8943 (100.0%), 306.8964 (51.4%), 310.8923 (48.6%), 309.8977 (9.7%), 311.8956 (4.7%), 307.8997 (4.4%)
	Found 228.9859	Not found

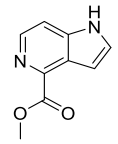


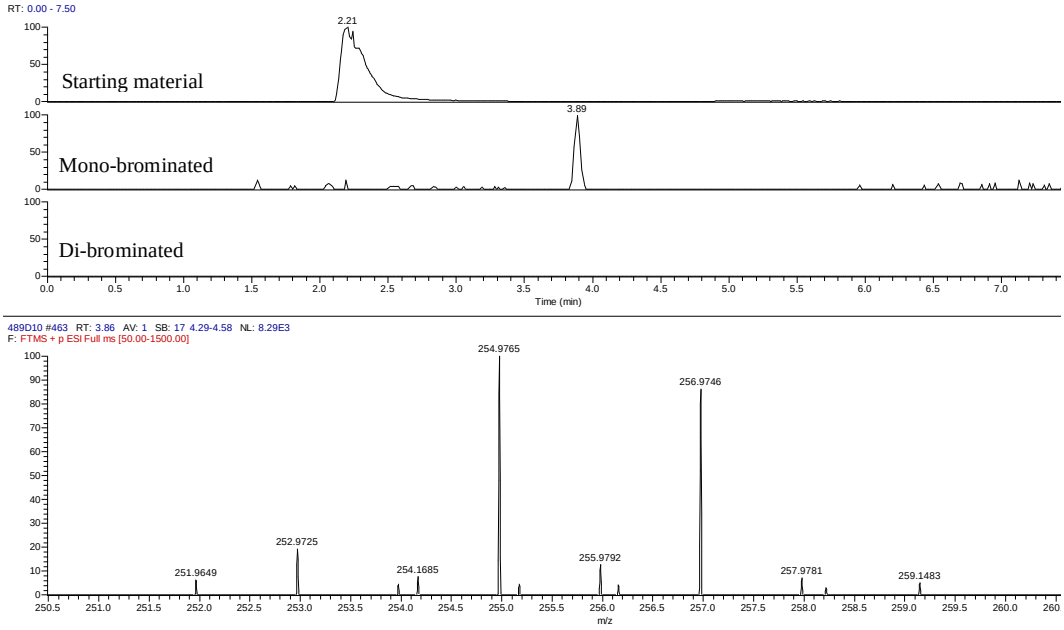
Supplementary Figure 55. LC-HRMS analysis of enzymatic bromination of compound 15 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_9H_{10}IO_2^+$ Exact Mass: 276.9720 m/z: 276.9720 (100.0%), 277.9754 (9.7%)	Chemical Formula: $C_9H_9I_2O_2^+$ Exact Mass: 402.8686 m/z: 402.8686 (100.0%), 403.8720 (9.7%)
	Found 276.9723	Not found

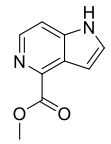


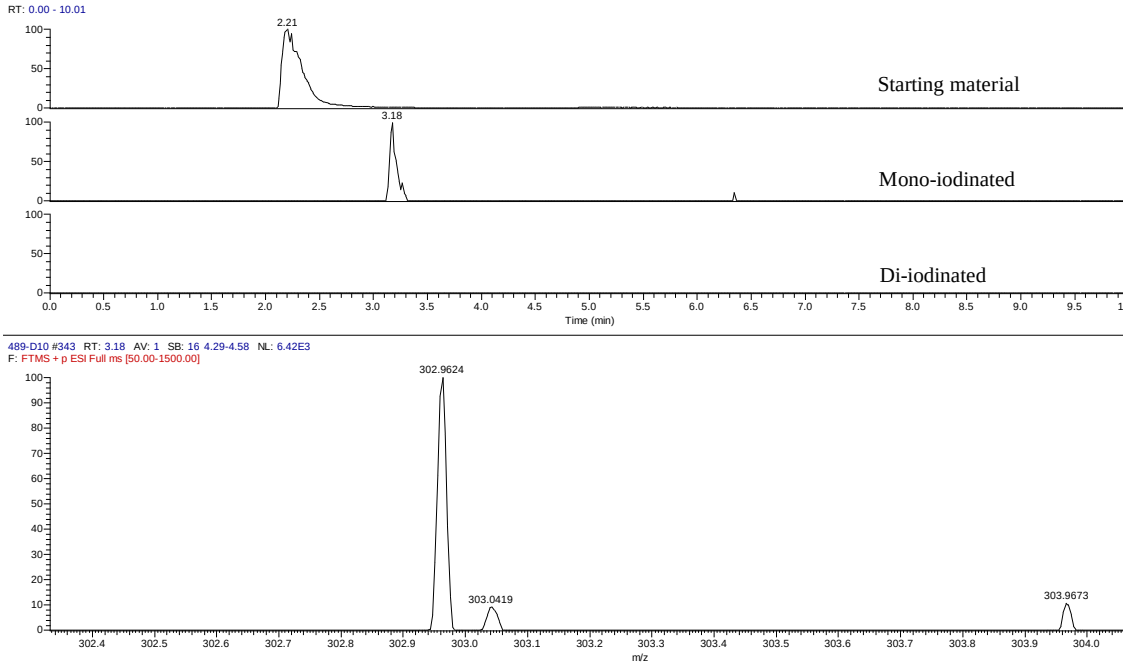
Supplementary Figure 56. LC-HRMS analysis of enzymatic iodination of compound 15 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_9H_8BrN_2O_2^+$ Exact Mass: 254.9764 m/z: 254.9764 (100.0%), 256.9743 (97.3%), 255.9797 (9.7%), 257.9777 (9.5%)	Chemical Formula: $C_9H_9Br_2N_2O_2^+$ Exact Mass: 334.9025 m/z: 336.9005 (100.0%), 334.9025 (51.4%), 338.8984 (48.6%), 337.9038 (9.7%), 339.9018 (4.7%), 335.9059 (4.4%)
	Found 254.9765	Not found

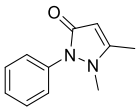


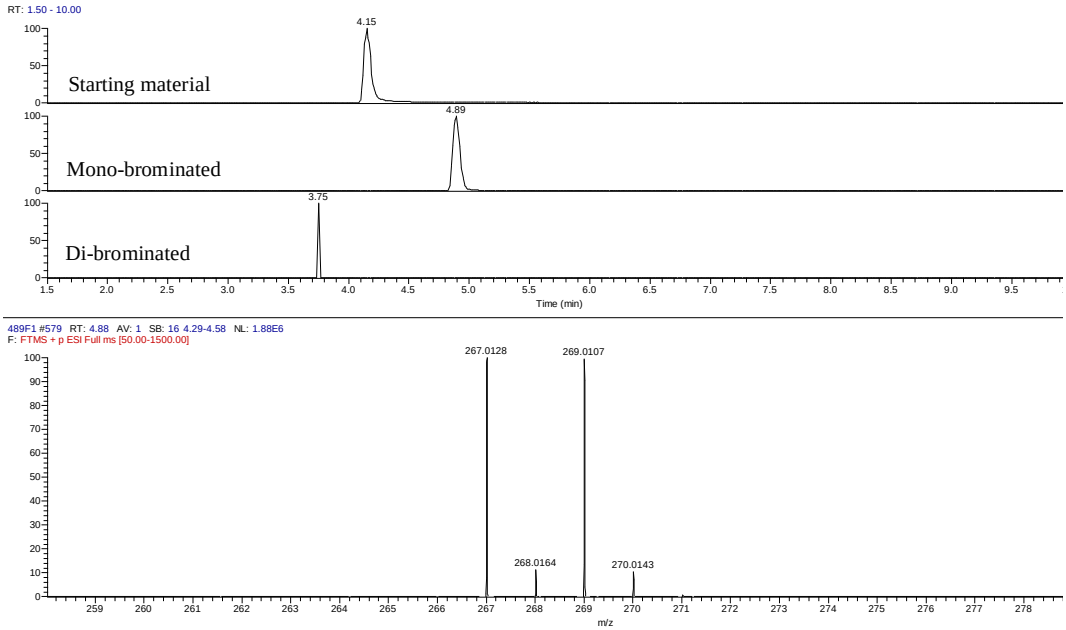
Supplementary Figure 57. LC-HRMS analysis of enzymatic bromination of compound 16 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_9H_8IN_2O_2^+$ Exact Mass: 302.9625 m/z: 302.9625 (100.0%), 303.9659 (9.7%)	Chemical Formula: $C_9H_7I_2N_2O_2^+$ Exact Mass: 428.8591 m/z: 428.8591 (100.0%), 429.8625 (9.7%)
	Found 302.9624	Not found

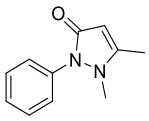


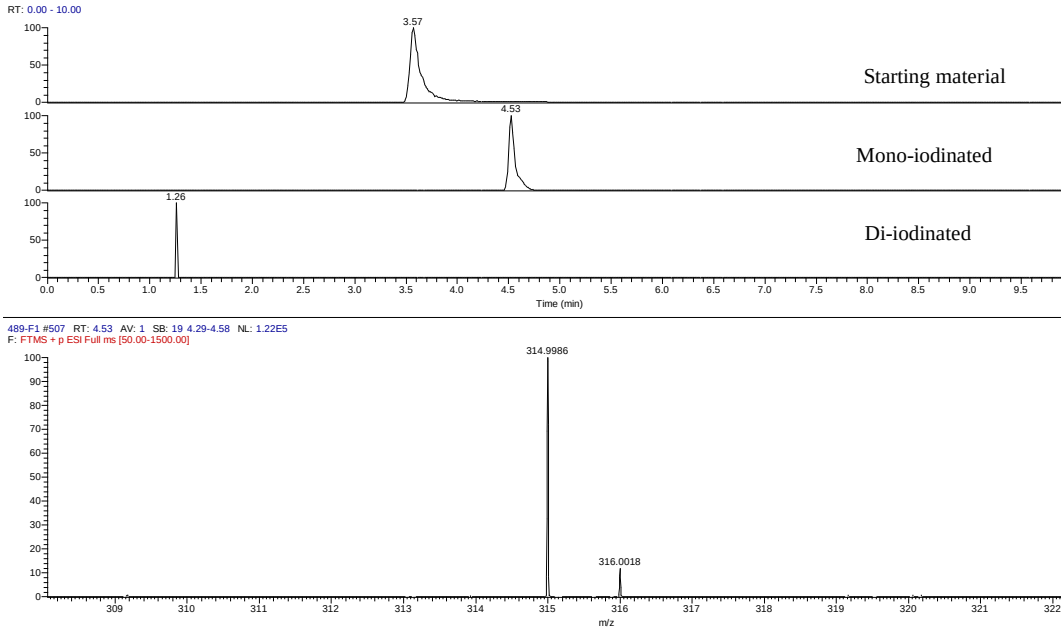
Supplementary Figure 58. LC-HRMS analysis of enzymatic iodination of compound 16 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_{11}H_{12}BrN_2O^+$ Exact Mass: 267.0128 m/z: 267.0128 (100.0%), 269.0107 (97.3%), 268.0161 (11.9%), 270.0141 (11.6%)	Chemical Formula: $C_{11}H_{11}Br_2N_2O^+$ Exact Mass: 344.9233 m/z: 346.9212 (100.0%), 344.9233 (51.4%), 348.9192 (48.6%), 347.9246 (11.9%), 349.9225 (5.8%), 345.9266 (4.4%), 345.9266 (1.7%)
	Found 267.0128	Not found

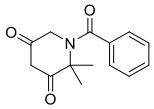


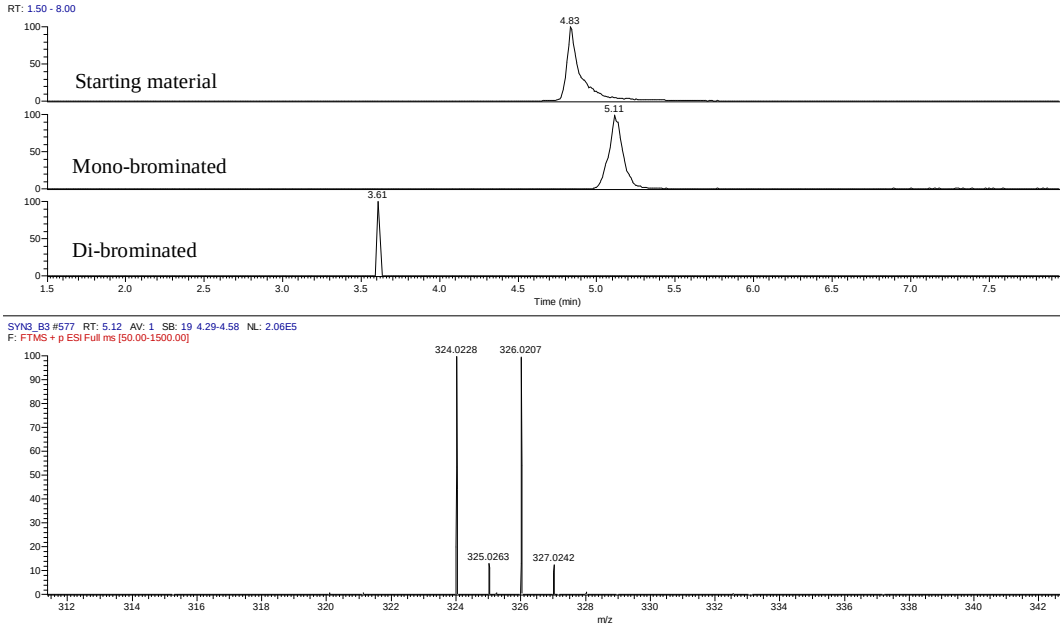
Supplementary Figure 59. LC-HRMS analysis of enzymatic bromination of compound 17 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_{11}H_{12}IN_2O^+$ Exact Mass: 314.9989 m/z: 314.9989 (100.0%), 316.0022 (11.9%)	Chemical Formula: $C_{11}H_{11}I_2N_2O^+$ Exact Mass: 444.8955 m/z: 444.8955 (100.0%), 445.8989 (11.9%)
	Found 314.9986	Not found

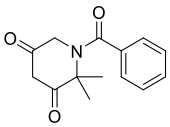


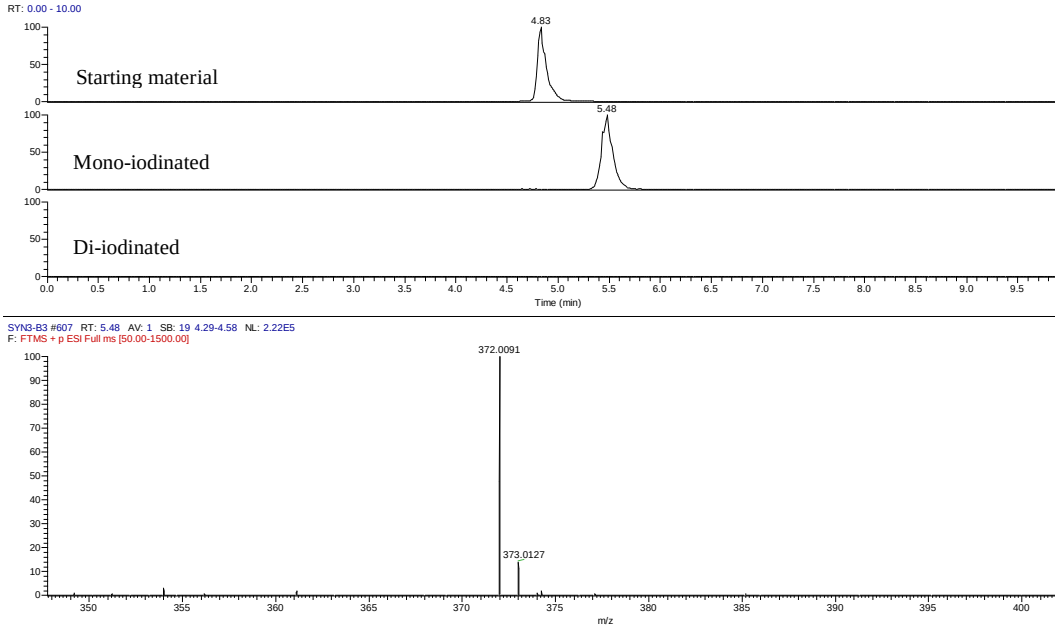
Supplementary Figure 60. LC-HRMS analysis of enzymatic iodination of compound 17 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_{14}H_{15}BrNO_3^+$ Exact Mass: 324.0230 m/z: 324.0230 (100.0%), 326.0209 (97.3%), 325.0263 (15.1%), 327.0243 (14.7%), 326.0297 (1.1%), 328.0276 (1.0%)	Chemical Formula: $C_{14}H_{14}Br_2NO_3^+$ Exact Mass: 401.9335 m/z: 403.9314 (100.0%), 401.9335 (51.4%), 405.9294 (48.6%), 404.9348 (15.1%), 406.9328 (7.4%), 402.9368 (4.4%), 402.9368 (3.3%)
	Found 324.0228	Not found

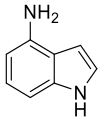


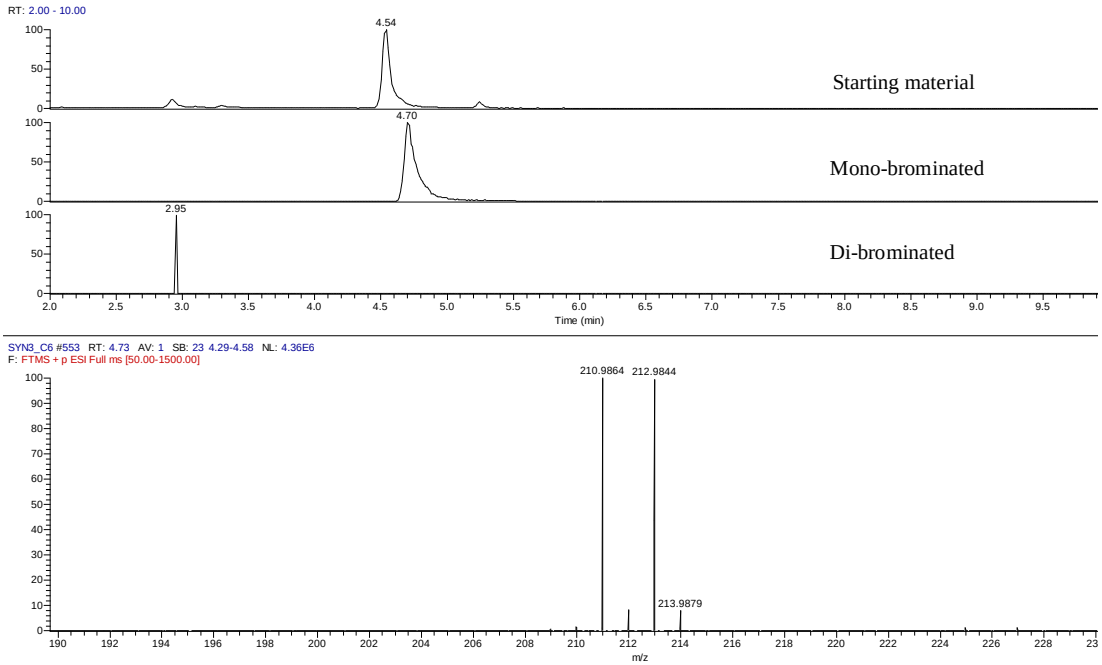
Supplementary Figure 61. LC-HRMS analysis of enzymatic bromination of compound 18 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_{14}H_{15}INO_3^+$ Exact Mass: 372.0091 m/z: 372.0091 (100.0%), 373.0125 (15.1%), 374.0158 (1.1%)	Chemical Formula: $C_{14}H_{14}I_2NO_3^+$ Exact Mass: 497.9058 m/z: 497.9058 (100.0%), 498.9091 (15.1%), 499.9125 (1.1%)
	Found 372.0091	Not found

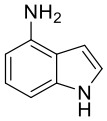


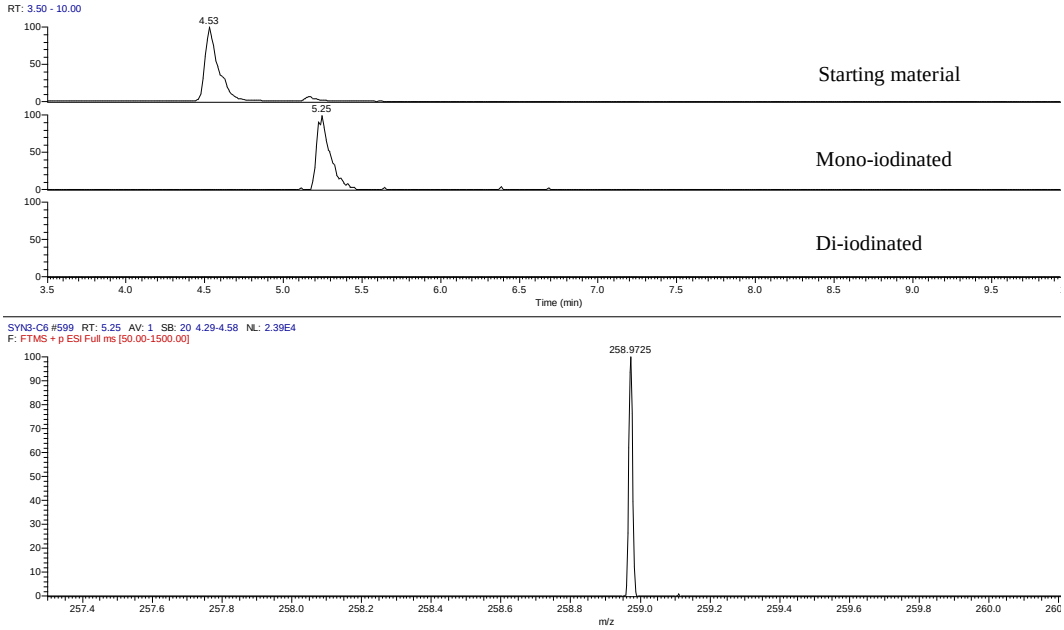
Supplementary Figure 62. LC-HRMS analysis of enzymatic iodination of compound 18 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_8H_8BrN_2^+$ Exact Mass: 210.9865 m/z: 210.9865 (100.0%), 212.9845 (97.3%), 211.9899 (8.7%), 213.9878 (8.4%)	Chemical Formula: $C_8H_7Br_2N_2^+$ Exact Mass: 282.8770 m/z: 282.8770 (100.0%), 280.8970 (51.4%), 284.8930 (48.6%), 293.8984 (8.7%),
	Found 210.9864	Not found

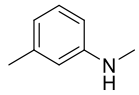


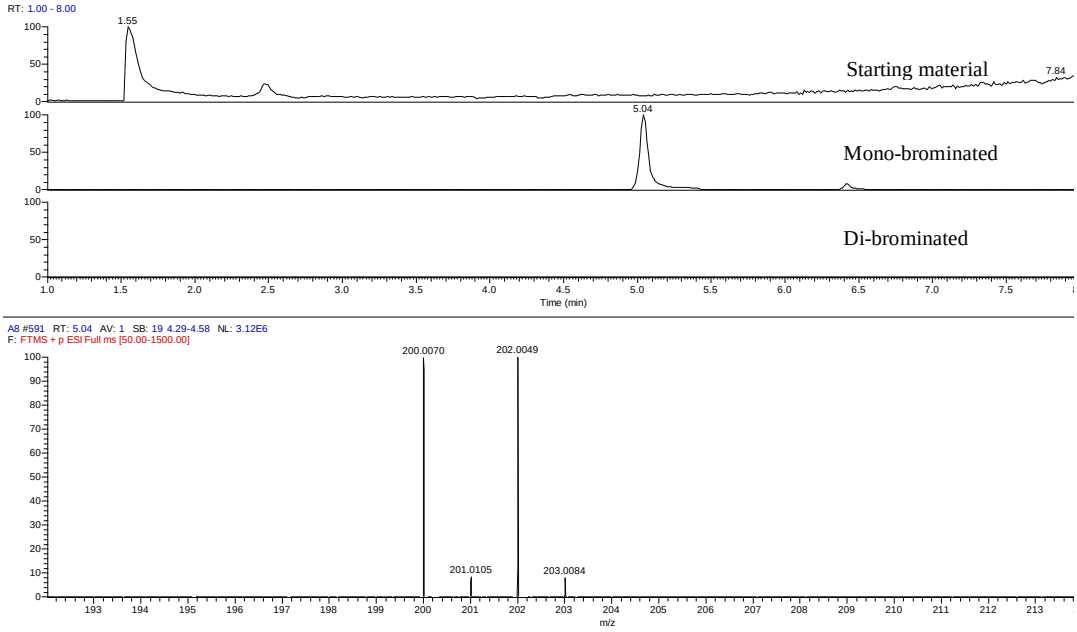
Supplementary Figure 63. LC-HRMS analysis of enzymatic bromination of compound 19 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_8H_8IN_2^+$ Exact Mass: 258.9727 m/z: 258.9727 (100.0%), 259.9760 (8.7%)	Chemical Formula: $C_8H_7I_2N_2^+$ Exact Mass: 384.8693 m/z: 384.8693 (100.0%), 385.8727 (8.7%)
	Found 258.9725	Not found

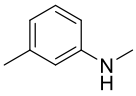


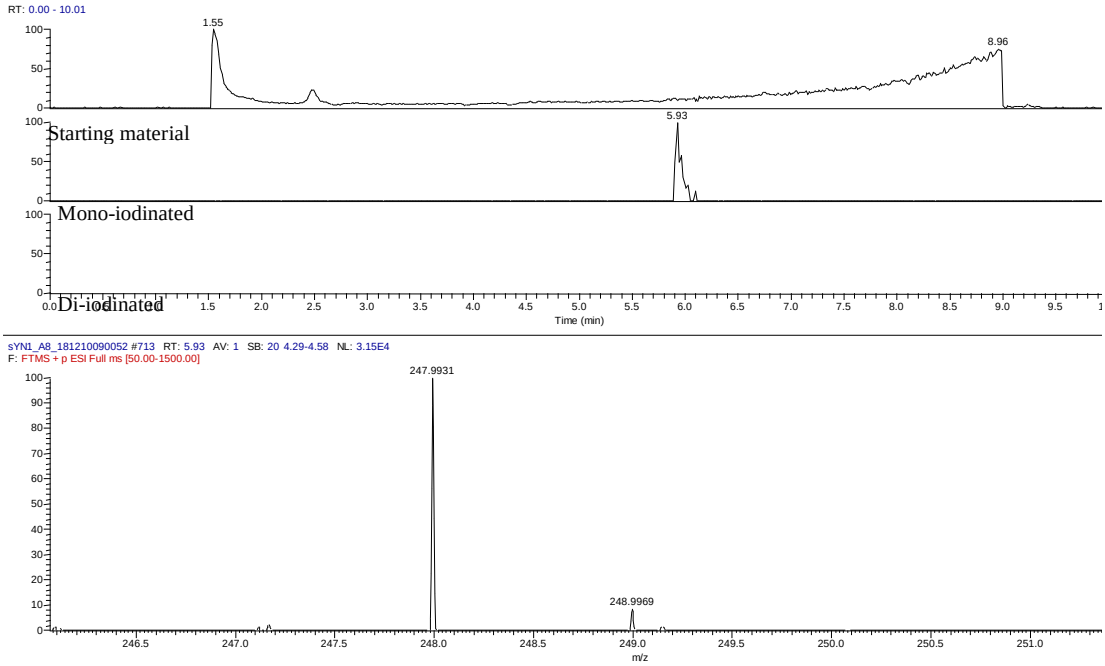
Supplementary Figure 64. LC-HRMS analysis of enzymatic iodination of compound 19 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_8H_{11}BrN^+$ Exact Mass: 200.0069 m/z: 200.0069 (100.0%), 202.0049 (97.3%), 201.0103 (8.7%), 203.0082 (8.4%)	Chemical Formula: $C_8H_{10}Br_2N^+$ Exact Mass: 277.9175 m/z: 279.9154 (100.0%), 277.9175 (51.4%), 281.9134 (48.6%), 280.9188 (8.7%), 278.9208 (4.4%), 282.9167 (4.2%)
	Found 200.0070	Not found

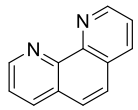


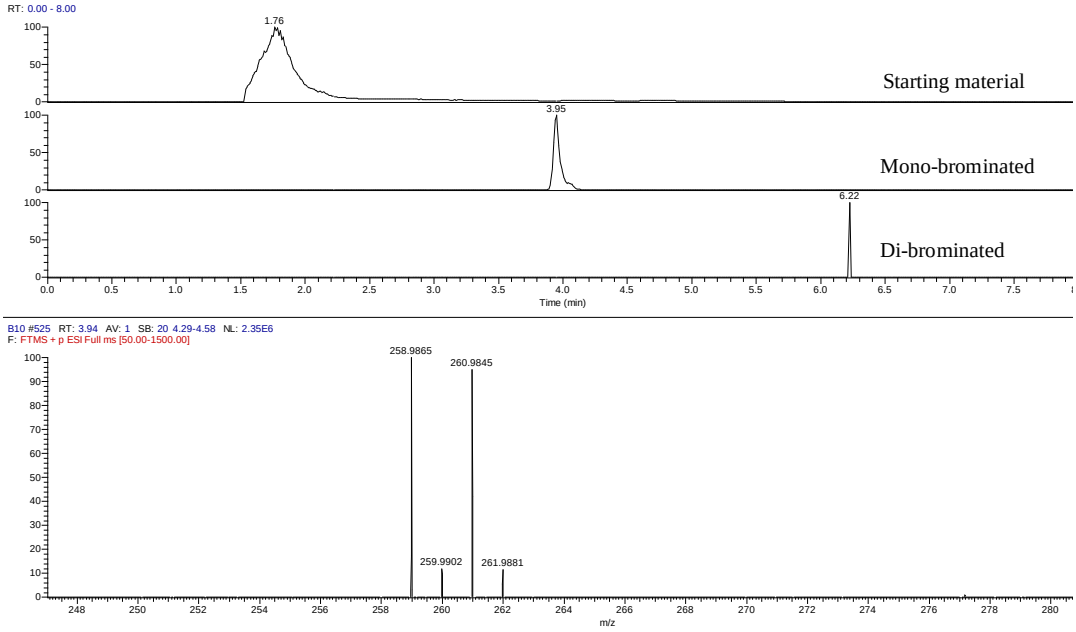
Supplementary Figure 65. LC-HRMS analysis of enzymatic bromination of compound 20 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_8H_{11}IN^+$ Exact Mass: 247.9931 m/z: 247.9931 (100.0%), 248.9964 (8.7%)	Chemical Formula: $C_8H_{10}I_2N^+$ Exact Mass: 373.8897 m/z: 373.8897 (100.0%), 374.8931 (8.7%)
	Found 247.9933	Not found

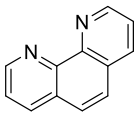


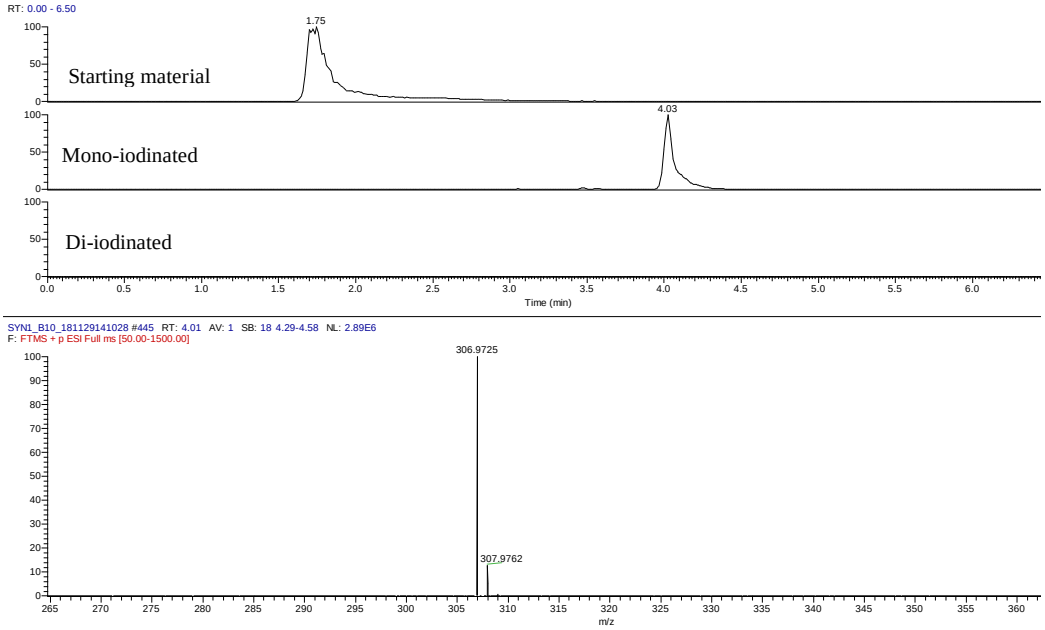
Supplementary Figure 66. LC-HRMS analysis of enzymatic iodination of compound 20 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_{12}H_8BrN_2^+$ Exact Mass: 258.9865 m/z: 258.9865 (100.0%), 260.9845 (97.3%), 259.9899 (13.0%), 261.9878 (12.6%)	Chemical Formula: $C_{12}H_7Br_2N_2^+$ Exact Mass: 336.8970 m/z: 338.8950 (100.0%), 336.8970 (51.4%), 340.8930 (48.6%), 339.8984 (9.7%), 337.9004 (6.7%), 341.8963 (6.3%), 339.8984 (3.2%)
	Found 258.9865	Not found

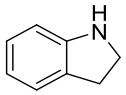


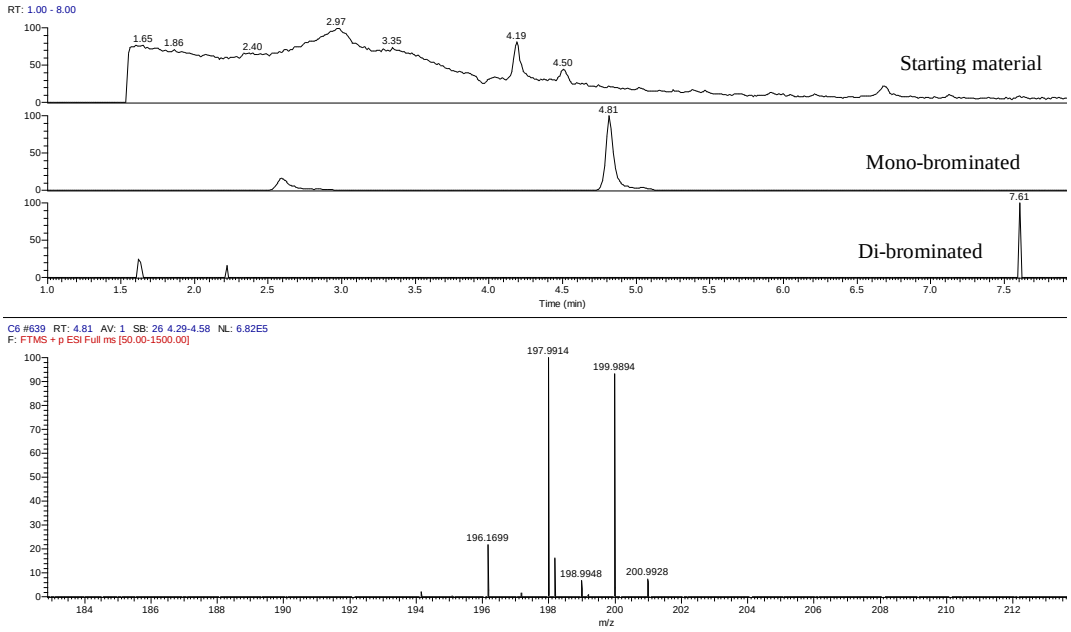
Supplementary Figure 67. LC-HRMS analysis of enzymatic bromination of compound 21 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_{12}H_8IN_2^+$ Exact Mass: 306.9727 m/z: 306.9727 (100.0%), 307.9760 (13.0%)	Chemical Formula: $C_{12}H_7I_2N_2^+$ Exact Mass: 432.8693 m/z: 432.8693 (100.0%), 433.8727 (13.0%)
	Found 306.9725	Not found

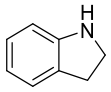


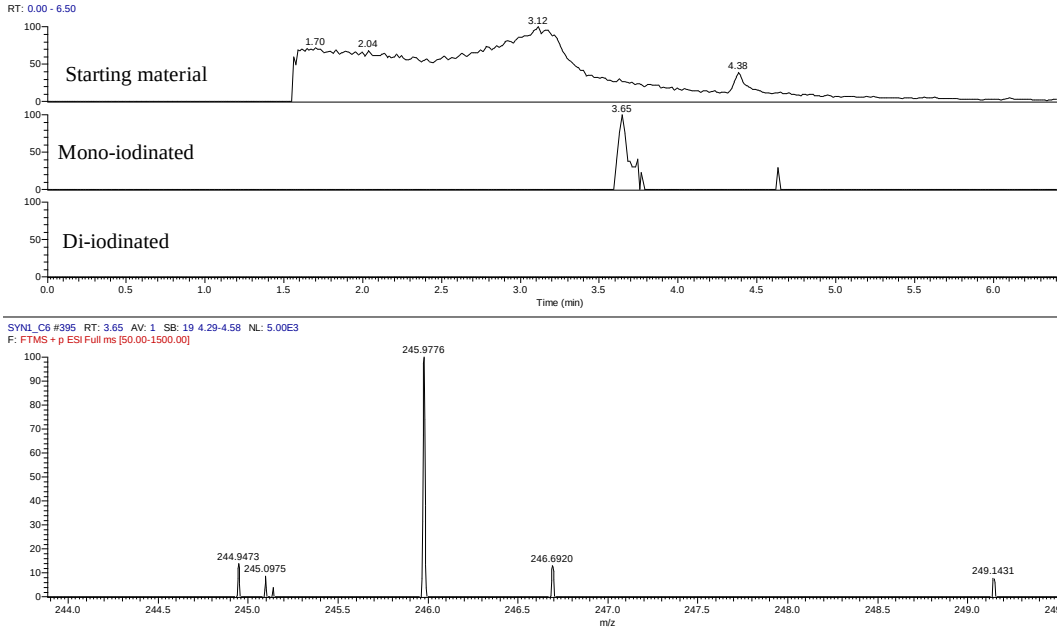
Supplementary Figure 68. LC-HRMS analysis of enzymatic iodination of compound 21 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_8H_9BrN^+$ Exact Mass: 197.9913 m/z: 197.9913 (100.0%), 199.9892 (97.3%), 198.9946 (8.7%), 200.9926 (8.4%)	Chemical Formula: $C_8H_8Br_2N^+$ Exact Mass: 275.9018 m/z: 277.8998 (100.0%), 275.9018 (51.4%), 279.8977 (48.6%), 278.9031 (8.7%)
	Found 197.9914	Not found

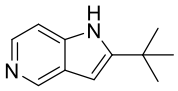


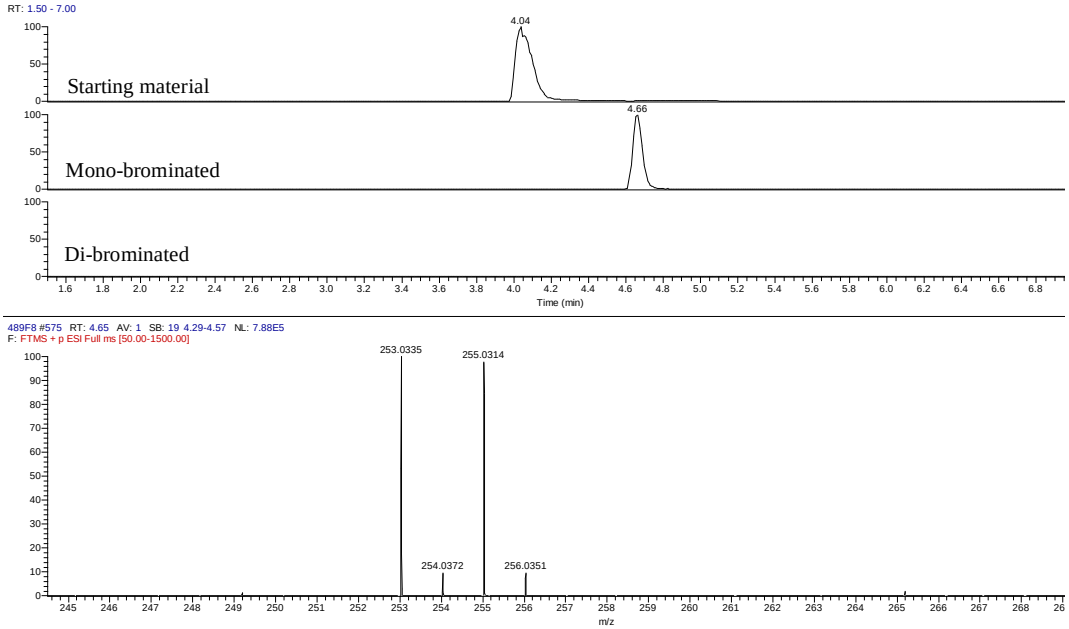
Supplementary Figure 69. LC-HRMS analysis of enzymatic bromination of compound 22 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_8H_9IN^+$ Exact Mass: 245.9774 m/z: 245.9774 (100.0%), 246.9808 (8.7%)	Chemical Formula: $C_8H_8I_2N^+$ Exact Mass: 371.8741 m/z: 371.8741 (100.0%), 372.8774 (8.7%)
	Found 245.9776	Not found

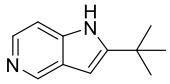


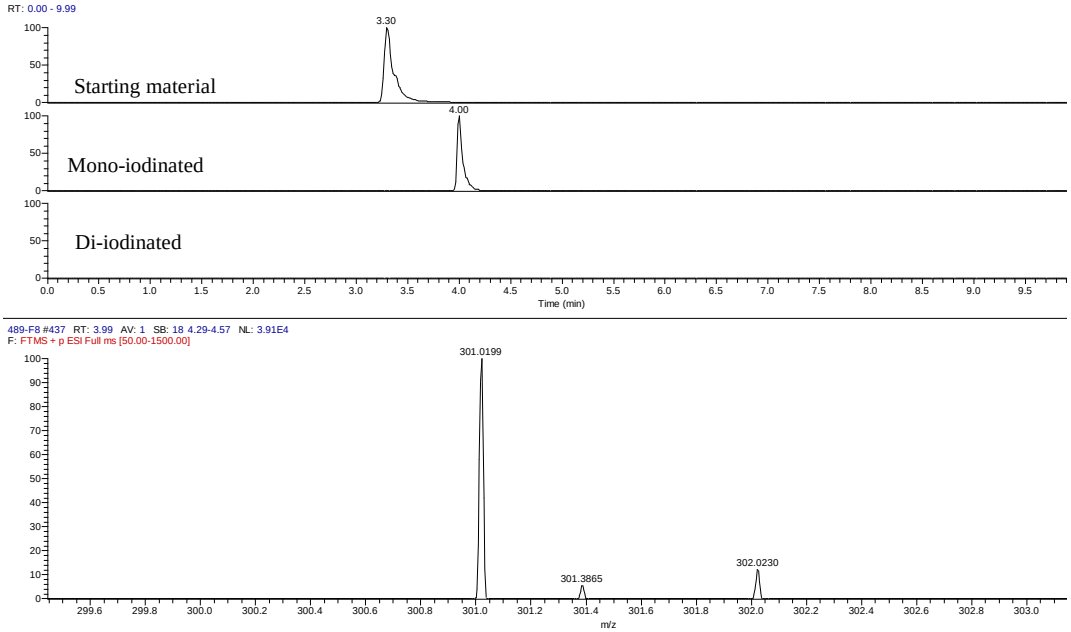
Supplementary Figure 70. LC-HRMS analysis of enzymatic iodination of compound 22 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_{11}H_{14}BrN_2^+$ Exact Mass: 253.0335 m/z: 253.0335 (100.0%), 255.0314 (97.3%), 254.0368 (11.9%), 256.0348 (11.6%)	Chemical Formula: $C_{11}H_{13}Br_2N_2^+$ Exact Mass: 330.9440 m/z: 332.9420 (100.0%), 330.9440 (51.4%), 334.9399 (48.6%), 333.9453 (9.7%), 335.9433 (5.8%), 331.9474 (4.4%), 333.9453 (2.2%), 331.9474 (1.7%)
	Found 253.0335	Not found

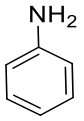


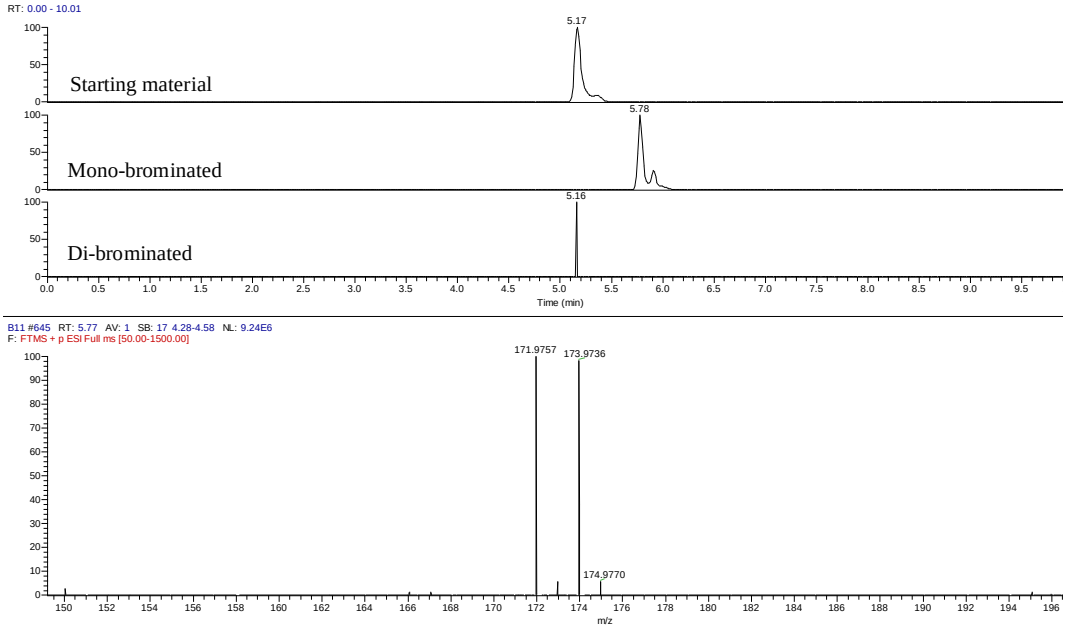
Supplementary Figure 71. LC-HRMS analysis of enzymatic bromination of compound 23 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_{11}H_{14}IN_2^+$ Exact Mass: 301.0196 m/z: 301.0196 (100.0%), 302.0230 (11.9%)	Chemical Formula: $C_{11}H_{13}I_2N_2^+$ Exact Mass: 426.9163 m/z: 426.9163 (100.0%), 427.9196 (11.9%)
	Found 301.0199	Not found

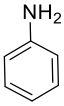


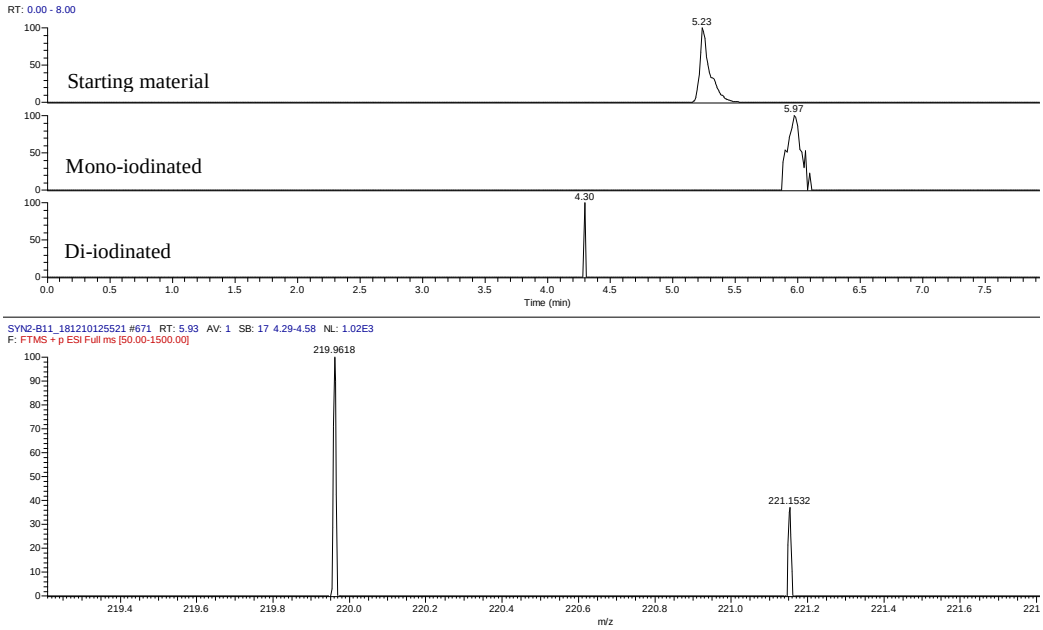
Supplementary Figure 72. LC-HRMS analysis of enzymatic iodination of compound 23 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_6H_7BrN^+$ Exact Mass: 171.9756 m/z: 171.9756 (100.0%), 173.9736 (97.3%), 172.9790 (6.5%), 174.9769 (6.3%)	Chemical Formula: $C_6H_6Br_2N^+$ Exact Mass: 249.8862 m/z: 251.8841 (100.0%), 249.8862 (51.4%), 253.8821 (48.6%), 252.8875 (6.5%), 250.8895 (3.3%), 254.8854 (3.2%)
	Found 171.9757	Not found

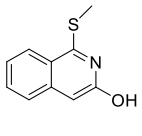


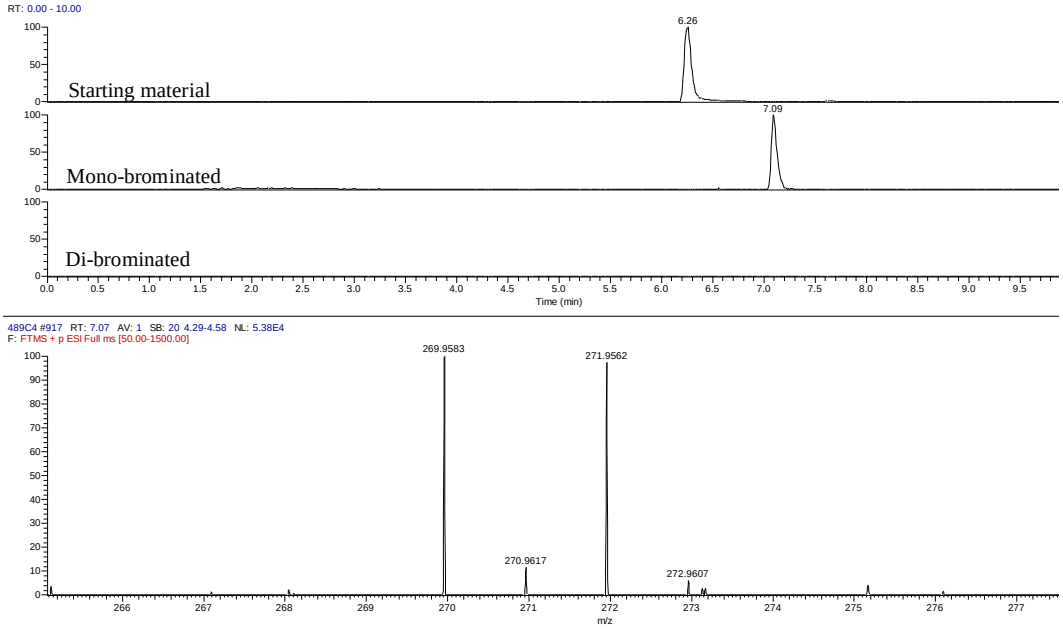
Supplementary Figure 73. LC-HRMS analysis of enzymatic bromination of compound 24 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. On EIC chromatogram (middle panel), both the major peak (5.7 min) and minor peak (5.9 min) correspond to the mono-brominated product indicating presence of potential regioisomers. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_6H_7IN^+$ Exact Mass: 219.9618 m/z: 219.9618 (100.0%), 220.9651 (6.5%)	Chemical Formula: $C_6H_6I_2N^+$ Exact Mass: 345.8584 m/z: 345.8584 (100.0%), 346.8618 (6.5%)
	Found 219.9618	Not found

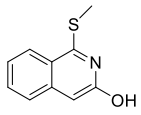


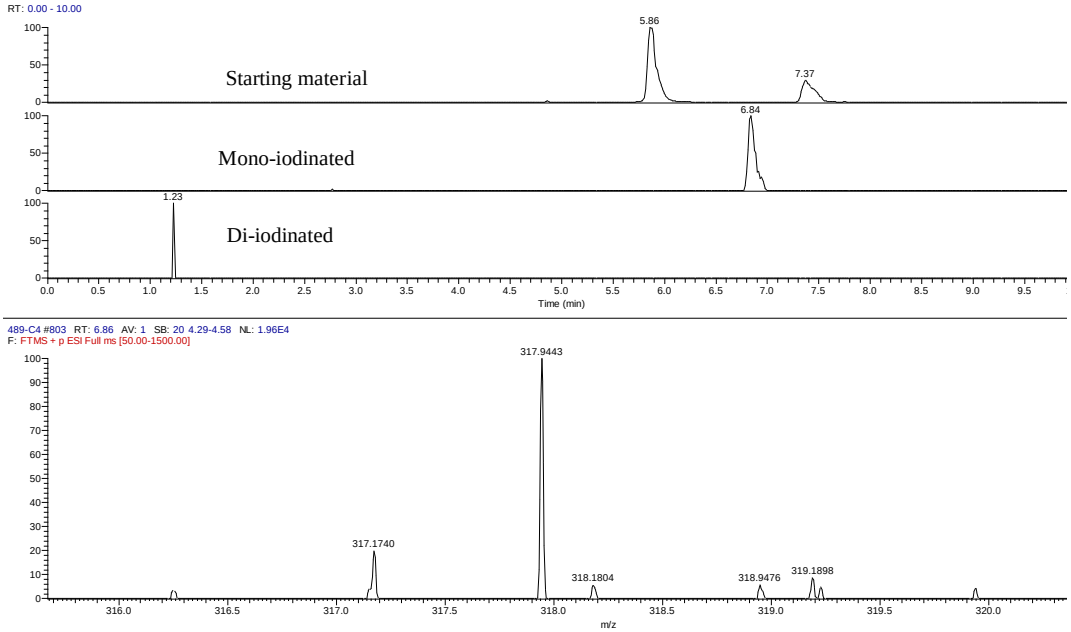
Supplementary Figure 74. LC-HRMS analysis of enzymatic iodination of compound 24 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_{10}H_9BrNOS^+$ Exact Mass: 269.9583 m/z: 269.9583 (100.0%), 271.9562 (97.3%), 270.9616 (10.8%), 272.9596 (10.5%), 271.9541 (4.5%), 273.9520 (4.4%)	Chemical Formula: $C_{10}H_8Br_2NOS^+$ Exact Mass: 347.8688 m/z: 349.8667 (100.0%), 347.8688 (51.4%), 351.8647 (48.6%), 350.8701 (9.7%), 352.8680 (5.3%), 351.8625 (4.5%), 348.8721 (4.4%)
	Found 269.9583	Not found

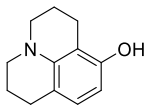


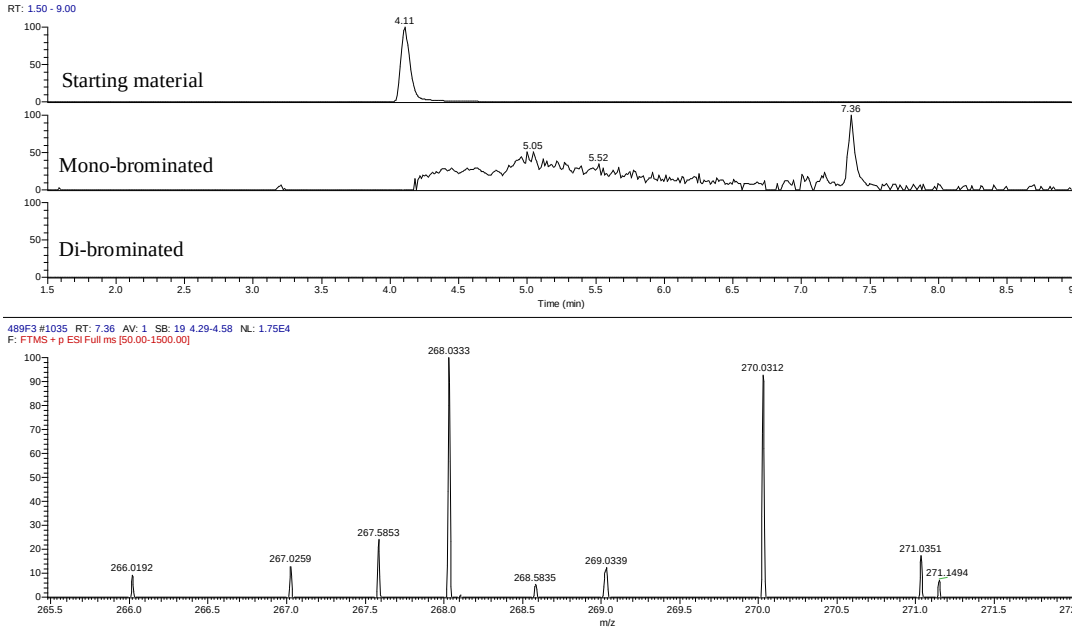
Supplementary Figure 75. LC-HRMS analysis of enzymatic bromination of compound 25 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_{10}H_9INOS^+$ Exact Mass: 317.9444 m/z: 317.9444 (100.0%), 318.9478 (10.8%), 319.9402 (4.5%)	Chemical Formula: $C_{10}H_8I_2NOS^+$ Exact Mass: 443.8410 m/z: 443.8410 (100.0%), 444.8444 (10.8%), 445.8368 (4.5%)
	Found 317.9443	Not found

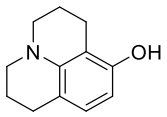


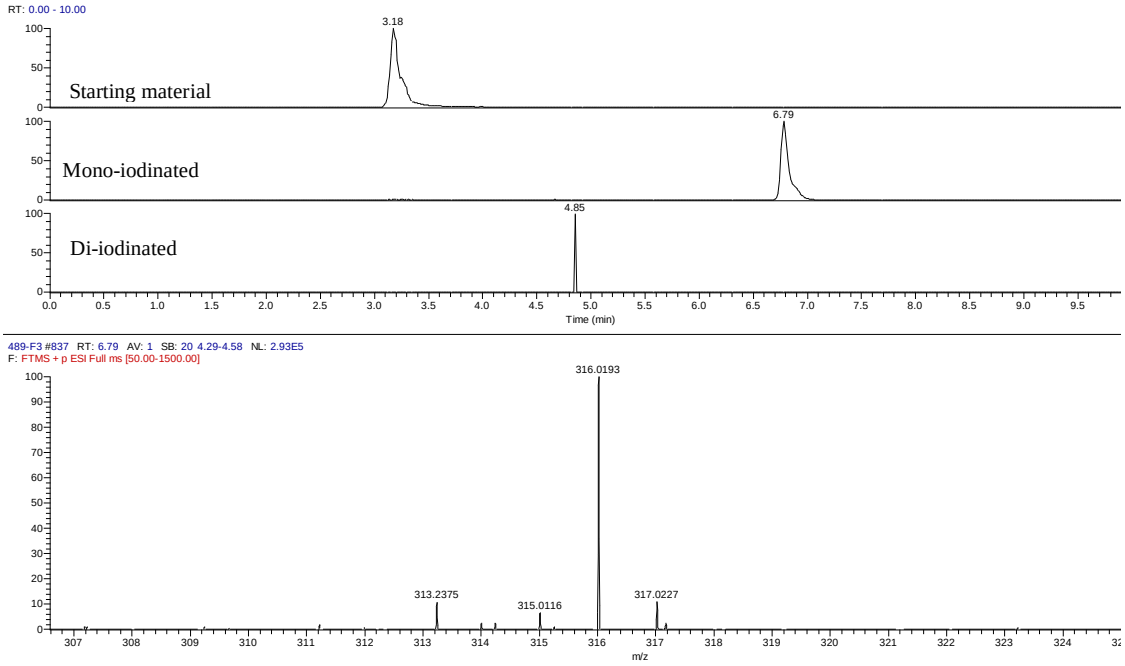
Supplementary Figure 76. LC-HRMS analysis of enzymatic iodination of compound 25 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_{12}H_{15}BrNO^+$ Exact Mass: 268.0332 m/z: 268.0332 (100.0%), 270.0311 (97.3%), 269.0365 (13.0%), 271.0345 (12.6%)	Chemical Formula: $C_{12}H_{14}Br_2NO^+$ Exact Mass: 345.9437 m/z: 347.9416 (100.0%), 345.9437 (51.4%), 349.9396 (48.6%), 348.9450 (9.7%), 350.9429 (6.3%), 346.9470 (4.4%), 348.9450 (3.2%), 346.9470 (2.2%)
	Found 268.0333	Not found

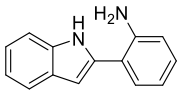


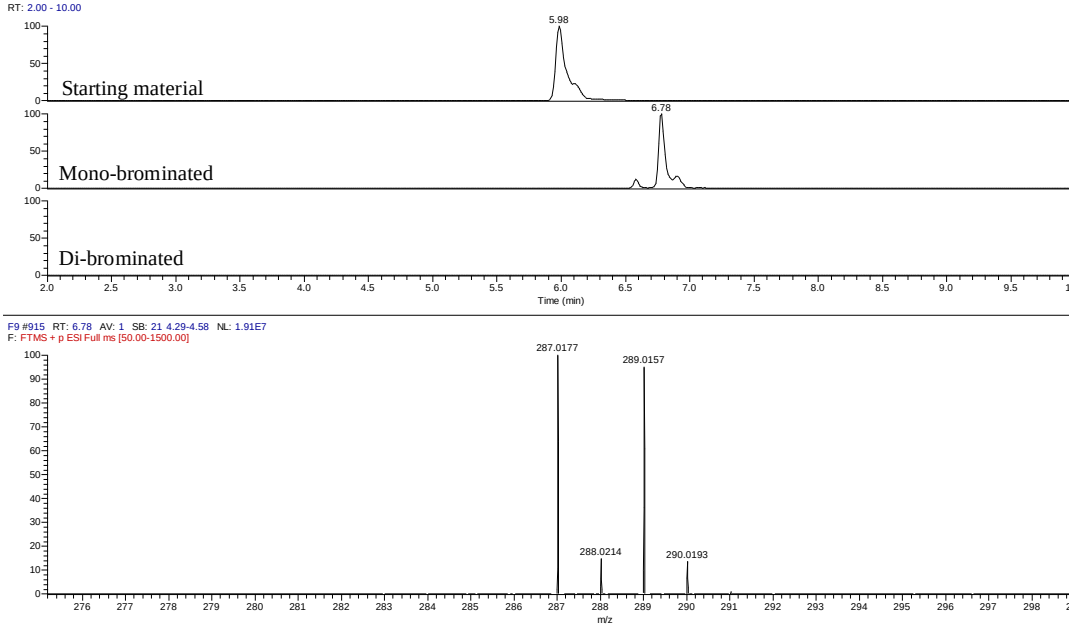
Supplementary Figure 77. LC-HRMS analysis of enzymatic bromination of compound 26 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_{12}H_{15}INO^+$ Exact Mass: 316.0193 m/z: 316.0193 (100.0%), 317.0226 (13.0%)	Chemical Formula: $C_{12}H_{14}I_2NO^+$ Exact Mass: 441.9159 m/z: 441.9159 (100.0%), 442.9193 (13.0%)
	Found 316.0193	Not found

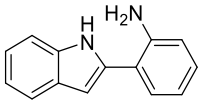


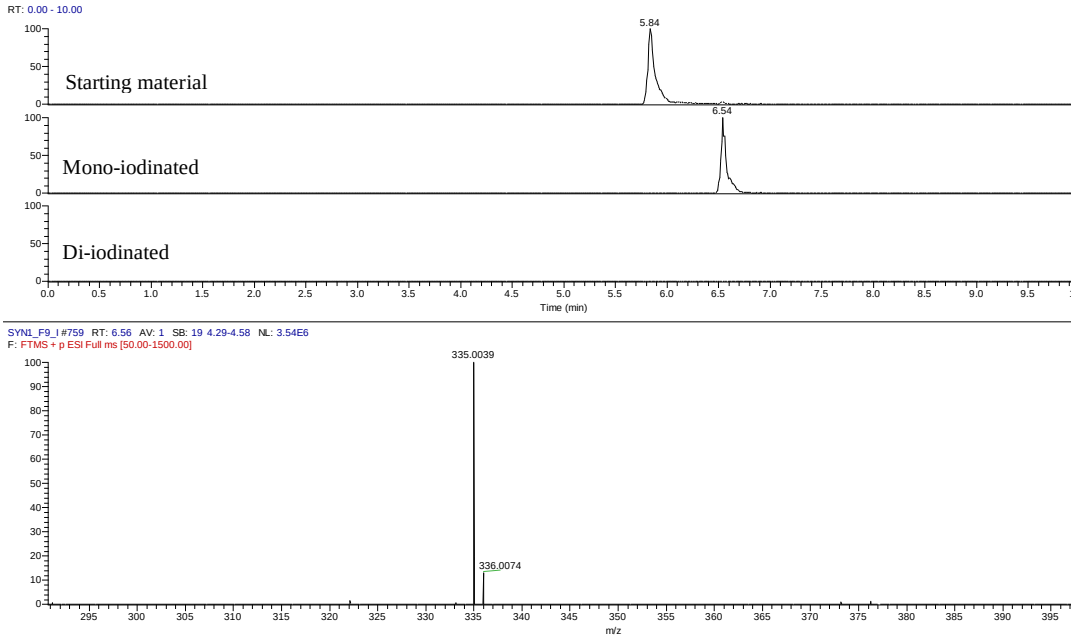
Supplementary Figure 78. LC-HRMS analysis of enzymatic iodination of compound 26 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_{14}H_{12}BrN_2^+$ Exact Mass: 287.0178 m/z: 287.0178 (100.0%), 289.0158 (97.3%), 288.0212 (15.1%), 290.0191 (14.7%), 289.0245 (1.1%), 291.0225 (1.0%)	Chemical Formula: $C_{14}H_{11}Br_2N_2^+$ Exact Mass: 346.9283 m/z: 346.9263 (100.0%), 344.9283 (51.4%), 348.9243 (48.6%), 347.9297 (9.7%), 345.9317 (7.8%), 349.9276 (7.4%), 347.9297 (5.4%)
	Found 287.0177	Not found

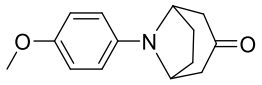


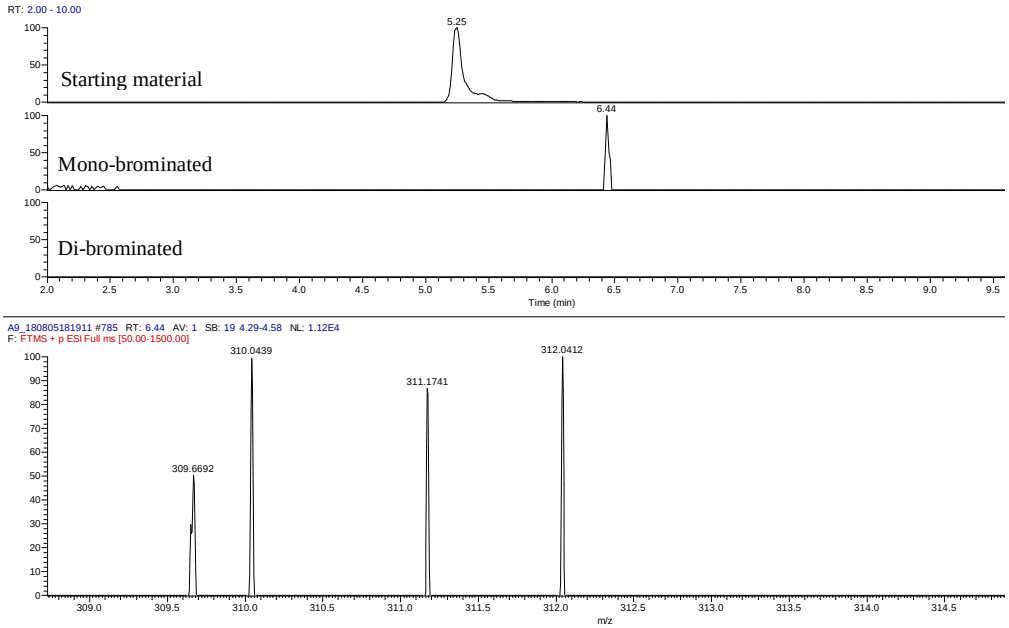
Supplementary Figure 79. LC-HRMS analysis of enzymatic bromination of compound 27 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Major peak (6.7 min) on EIC chromatogram (middle panel) corresponded to the mass of mono-brominated product, whereas minor peak (6.6 min) showed presence of starting material and minor peak (6.9 min) indicated non-brominated impurity. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_{14}H_{12}IN_2^+$ Exact Mass: 335.0040 m/z: 335.0040 (100.0%), 336.0073 (15.1%), 337.0107 (1.1%)	Chemical Formula: $C_{14}H_{11}I_2N_2^+$ Exact Mass: 460.9006 m/z: 460.9006 (100.0%), 461.9040 (15.1%), 462.9073 (1.1%)
	Found 335.0039	Not found

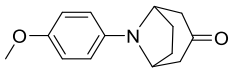


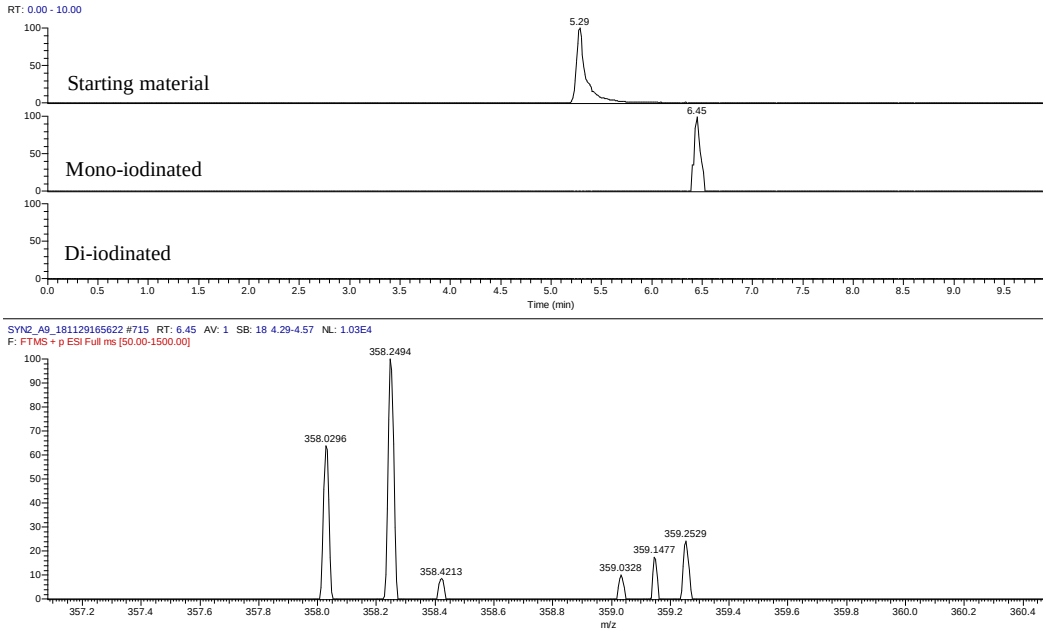
Supplementary Figure 80. LC-HRMS analysis of enzymatic iodination of compound 27 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_{14}H_{17}BrNO_2^+$ Exact Mass: 310.0437 m/z: 310.0437 (100.0%), 312.0417 (97.3%), 311.0471 (15.1%), 313.0450 (14.7%)	Chemical Formula: $C_{14}H_{16}Br_2NO_2^+$ Exact Mass: 387.9542 m/z: 389.9522 (100.0%), 387.9542 (51.4%), 391.9501 (48.6%), 390.9555 (9.7%)
	Found 310.0439	Not found

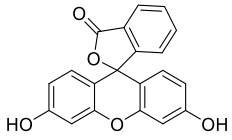


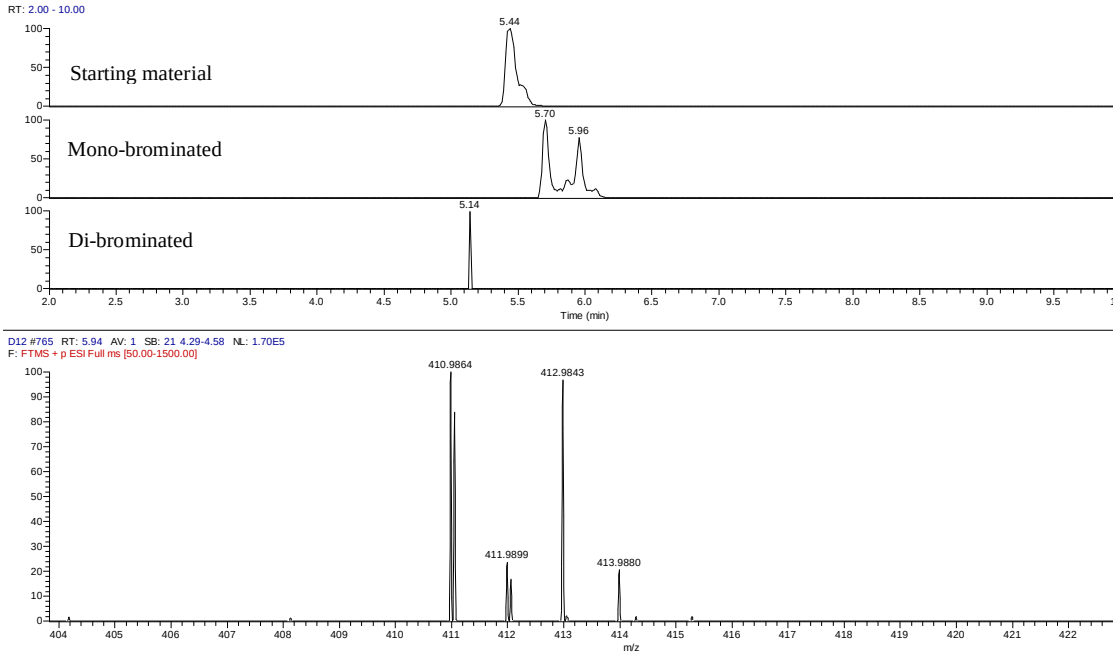
Supplementary Figure 81. LC-HRMS analysis of enzymatic bromination of compound 28 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_{14}H_{17}INO_2^+$ Exact Mass: 358.0298 m/z: 358.0298 (100.0%), 359.0332 (15.1%), 360.0366 (1.1%)	Chemical Formula: $C_{14}H_{16}I_2NO_2^+$ Exact Mass: 483.9265 m/z: 483.9265 (100.0%), 484.9298 (15.1%), 485.9332 (1.1%)
	Found 358.0296	Not found

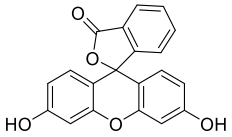


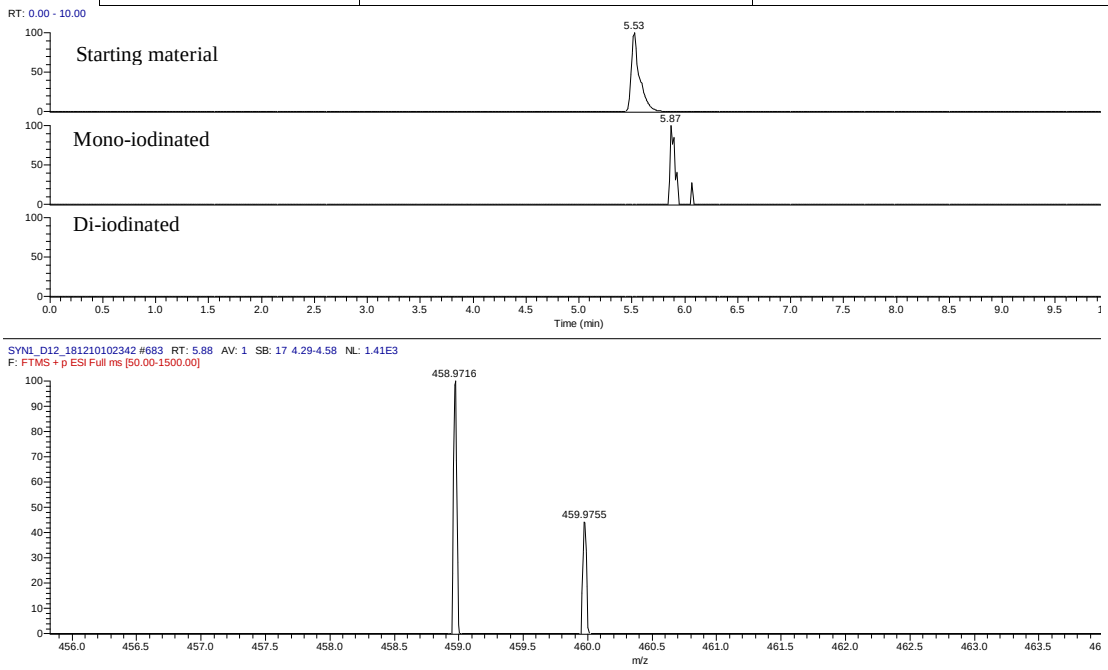
Supplementary Figure 82. LC-HRMS analysis of enzymatic iodination of compound 28 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_{20}H_{12}BrO_5^+$ Exact Mass: 410.9863 m/z: 410.9863 (100.0%), 412.9842 (97.3%), 413.9876 (21.0%), 411.9896 (16.2%)	Chemical Formula: $C_{20}H_{11}Br_2O_5^+$ Exact Mass: 488.8968 m/z: 490.8947 (100.0%), 488.8968 (51.4%), 492.8927 (48.6%), 491.8981 (11.9%)
	Found 410.9864	Not found

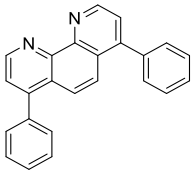


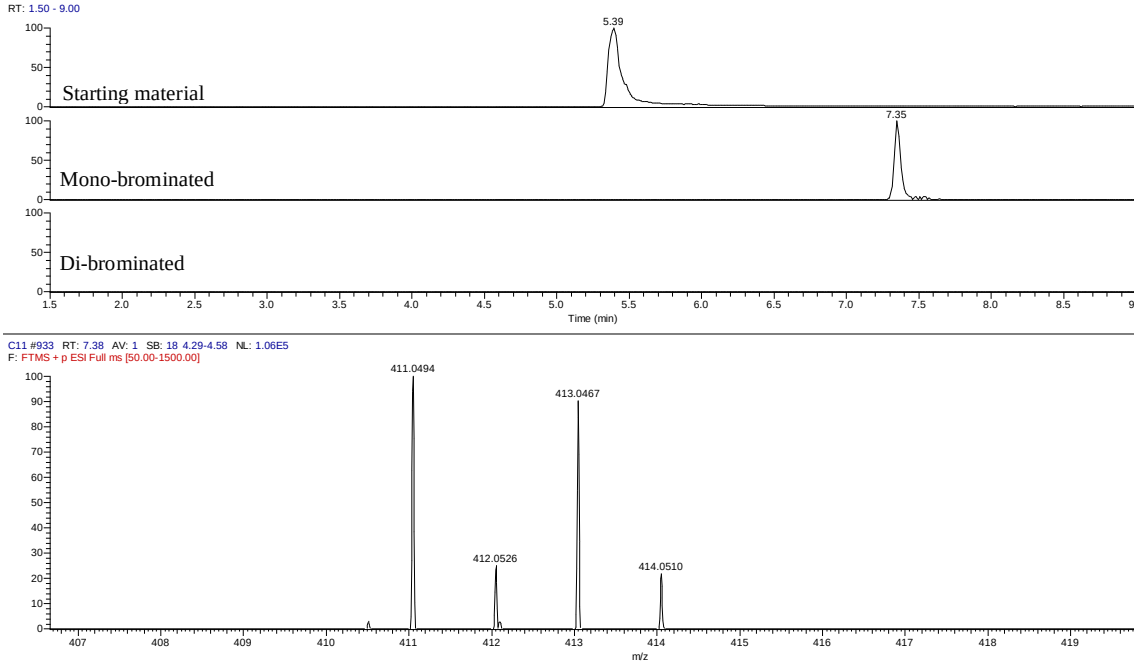
Supplementary Figure 83. LC-HRMS analysis of enzymatic bromination of compound 29 using VirX1. Only mono-brominated product was observed on EIC chromatogram (middle panel), both peaks (5.7 and 5.9 min) correspond to the mass of mono-brominated product, potentially indicating regioisomers. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono- iodinated Calculated	Di- iodinated Calculated
	Chemical Formula: $C_{20}H_{12}IO_5^+$ Exact Mass: 458.9724 m/z: 458.9724 (100.0%), 459.9757 (21.6%)	Chemical Formula: $C_{20}H_{11}I_2O_5^+$ Exact Mass: 584.8690 m/z: 584.8690 (100.0%), 585.8724 (21.6%)
	Found 458.9710	Not found

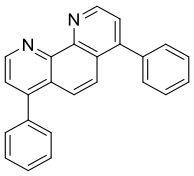


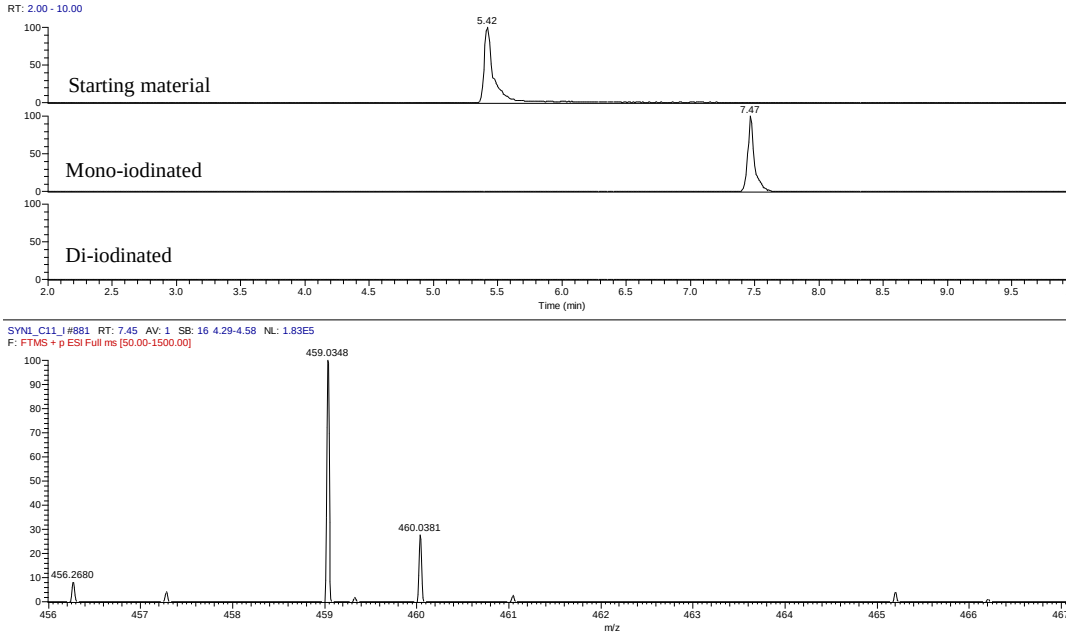
Supplementary Figure 84. LC-HRMS analysis of enzymatic iodination of compound 29 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Major peak (5.9 min) on EIC chromatogram (middle panel) corresponded to the mass of mono-iodinated product, whereas minor peak (6.1 min) showed non-iodinated impurity. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated calculated	Di-brominated calculated
	Chemical Formula: $C_{24}H_{16}BrN_2^+$ Exact Mass: 411.0491 m/z: 411.0491 (100.0%), 413.0471 (97.3%), 414.0504 (25.3%), 412.0525 (16.2%)	Chemical Formula: $C_{24}H_{15}Br_2N_2^+$ Exact Mass: 488.9597 m/z: 490.9576 (100.0%), 488.9597 (51.4%), 492.9556 (48.6%), 491.9610 (16.2%)
	Found 411.0494	Not found

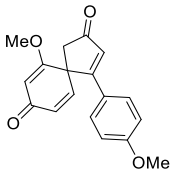


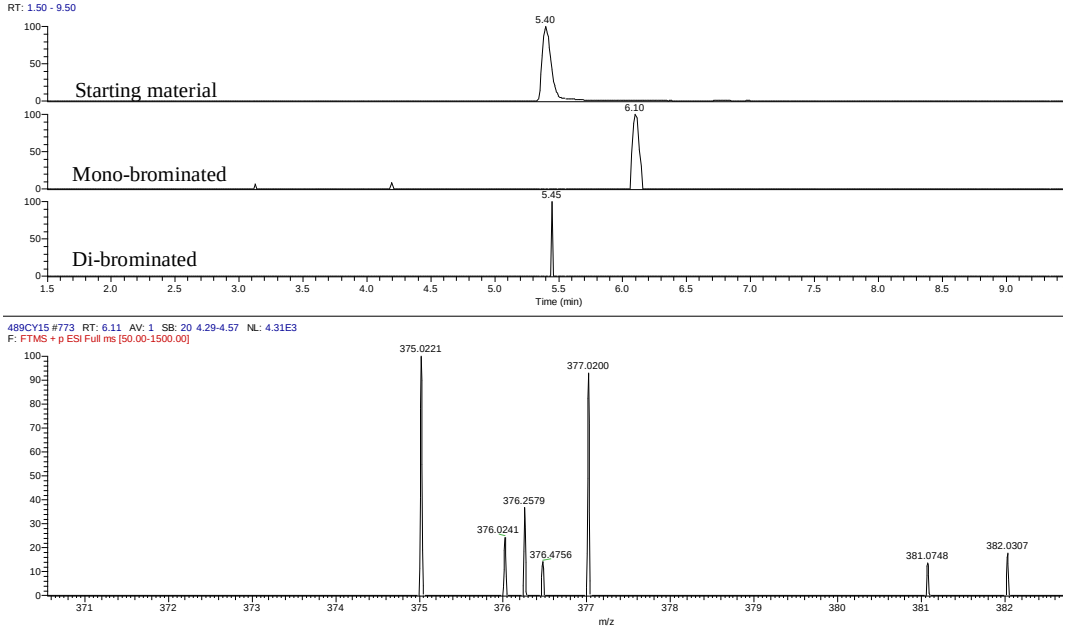
Supplementary Figure 85. LC-HRMS analysis of enzymatic bromination of compound 30 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated calculated	Di-iodinated calculated
	Chemical Formula: $C_{24}H_{16}IN_2^+$ Exact Mass: 459.0353 m/z: 459.0353 (100.0%), 460.0386 (26.0%), 461.0420 (2.7%)	Chemical Formula: $C_{24}H_{15}I_2N_2^+$ Exact Mass: 584.9319 m/z: 584.9319 (100.0%), 585.9353 (26.0%), 586.9386 (3.2%)
	Found 459.0348	Not found

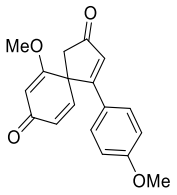


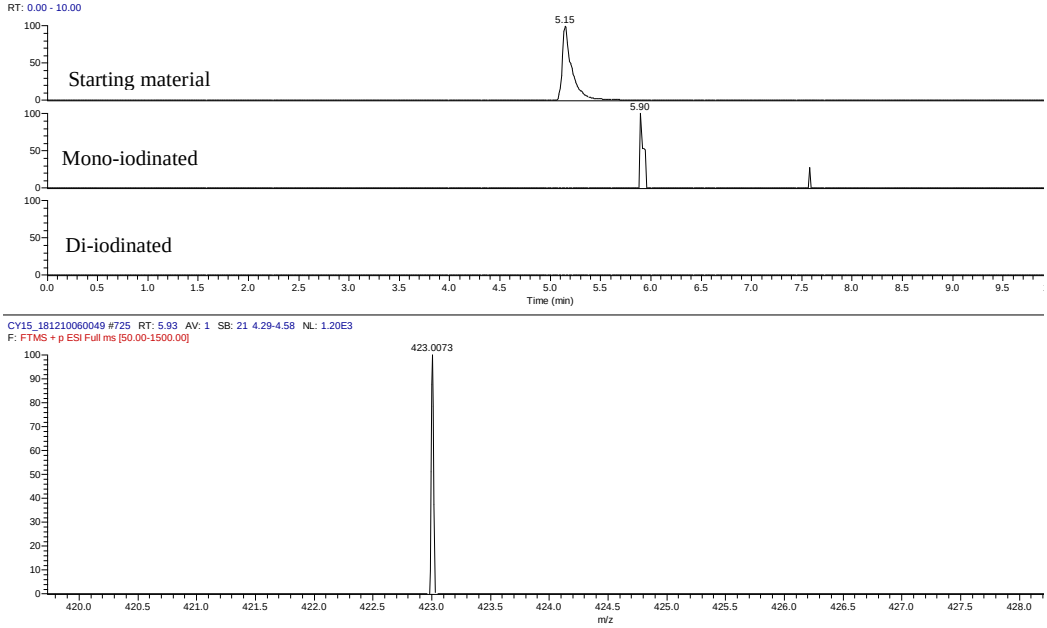
Supplementary Figure 86. LC-HRMS analysis of enzymatic iodination of compound 30 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_{18}H_{16}BrO_4^+$ Exact Mass: 375.0226 m/z: 375.0226 (100.0%), 377.0206 (97.3%), 378.0240 (18.9%), 376.0260 (16.2%)	Chemical Formula: $C_{18}H_{15}Br_2O_4^+$ Exact Mass: 452.9332 m/z: 454.9311 (100.0%), 452.9332 (51.4%), 456.9291 (48.6%), 455.9345 (9.7%)
	Found 375.0221	Not found

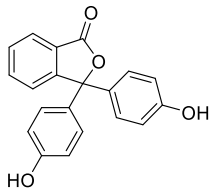


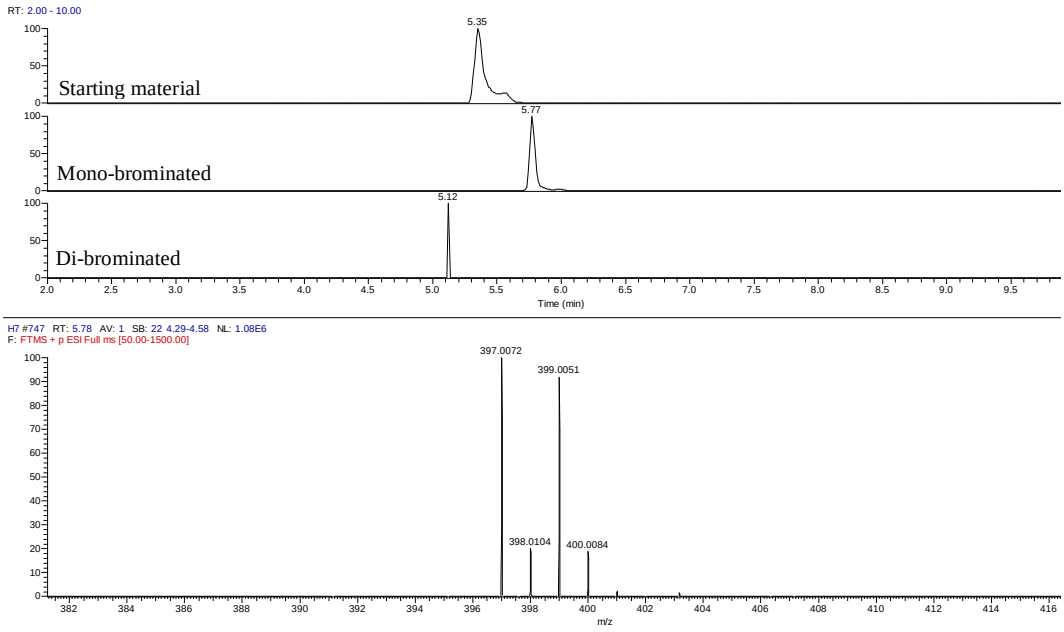
Supplementary Figure 87. LC-HRMS analysis of enzymatic bromination of compound 31 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_{18}H_{16}IO_4^+$ Exact Mass: 423.0088 m/z: 423.0088 (100.0%), 424.0121 (19.5%), 425.0155 (1.8%)	Chemical Formula: $C_{18}H_{15}I_2O_4^+$ Exact Mass: 548.9054 m/z: 548.9054 (100.0%), 549.9088 (19.5%), 550.9121 (1.8%)
	Found 423.0073	Not found

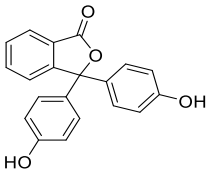


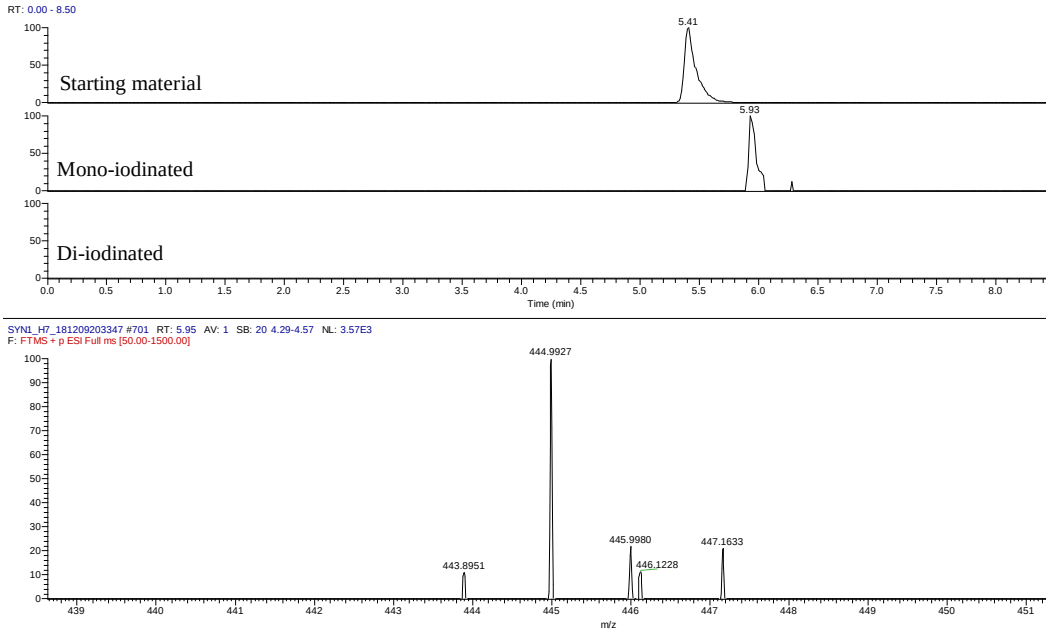
Supplementary Figure 88. LC-HRMS analysis of enzymatic iodination of compound 31 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-iodinated product.

	Mono-brominated Calculated	Di-brominated Calculated
	Chemical Formula: $C_{20}H_{14}BrO_4^+$ Exact Mass: 397.0070 m/z: 397.0070 (100.0%), 399.0050 (97.3%), 400.0083 (21.0%), 398.0104 (16.2%)	Chemical Formula: $C_{20}H_{13}Br_2O_4^+$ Exact Mass: 474.9175 m/z: 476.9155 (100.0%), 474.9175 (51.4%), 478.9134 (48.6%), 477.9188 (11.9%)
	Found 397.0072	Not found

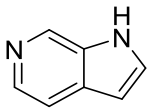
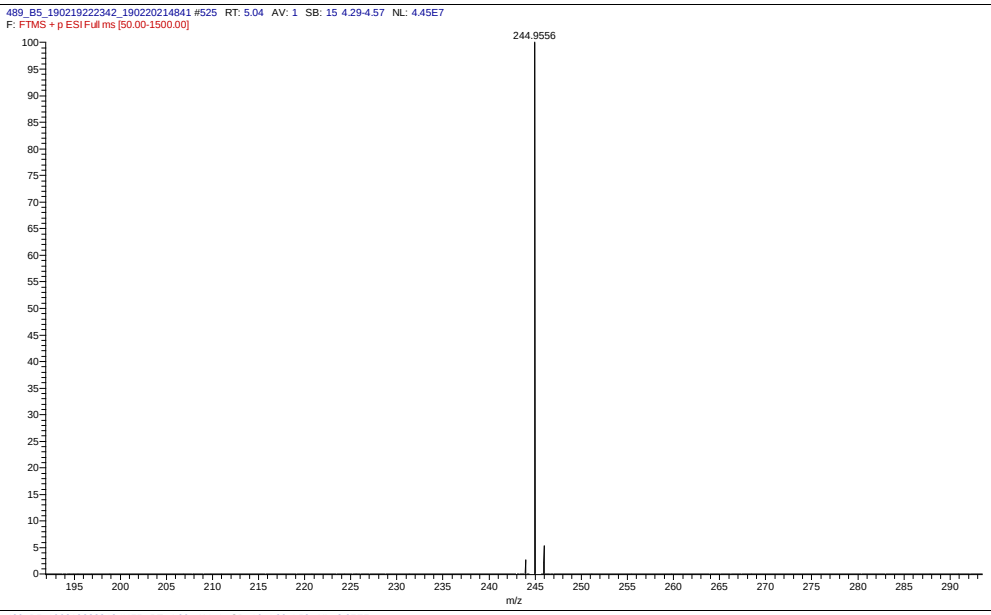
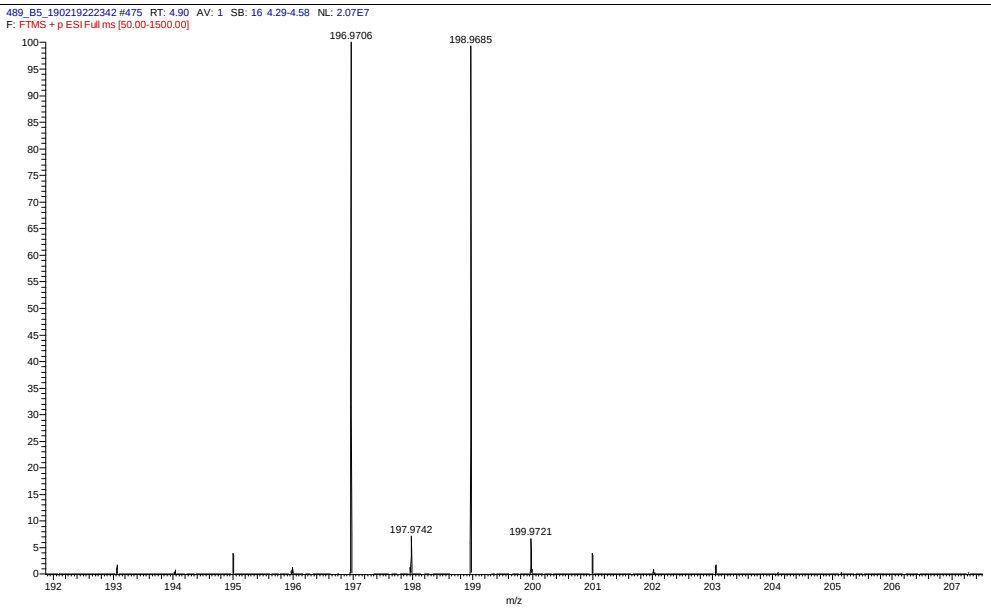


Supplementary Figure 89. LC-HRMS analysis of enzymatic bromination of compound 32 using VirX1. Only mono-brominated product was observed on EIC chromatogram. HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-brominated product.

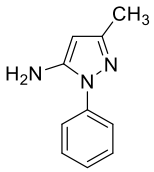
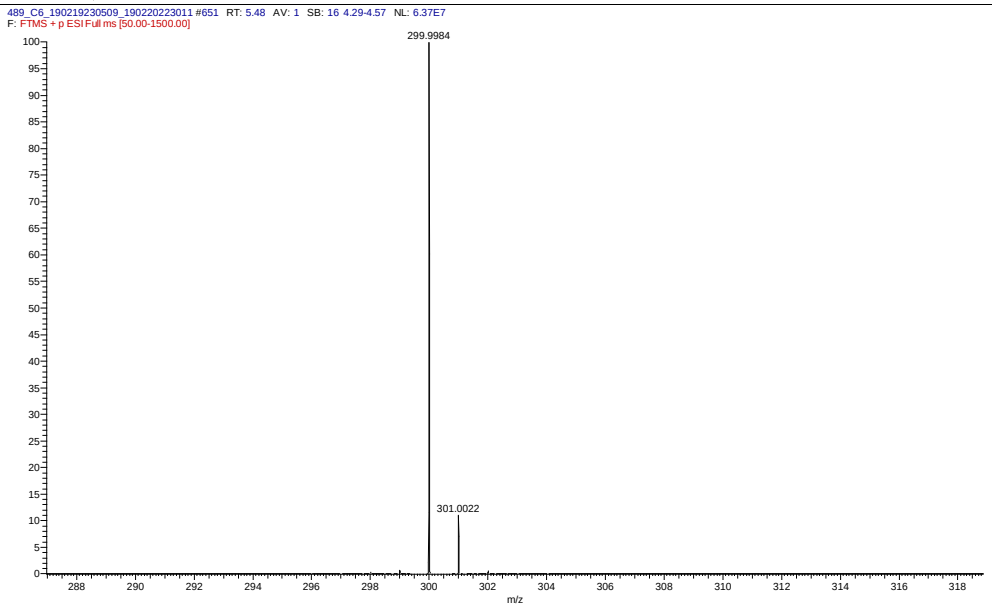
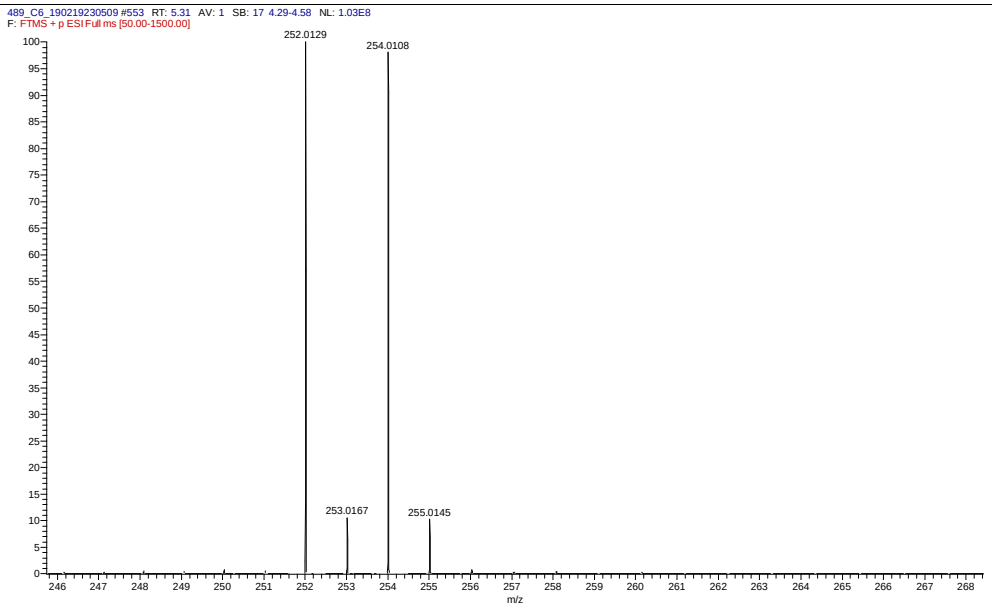
	Mono-iodinated Calculated	Di-iodinated Calculated
	Chemical Formula: $C_{20}H_{14}IO_4^+$ Exact Mass: 444.9931 m/z: 444.9931 (100.0%), 445.9965 (21.6%), 446.9998 (2.2%)	Chemical Formula: $C_{20}H_{13}I_2O_4^+$ Exact Mass: 570.8898 m/z: 570.8898 (100.0%), 571.8931 (21.6%), 572.8965 (2.2%)
	Found 444.9927	Not found



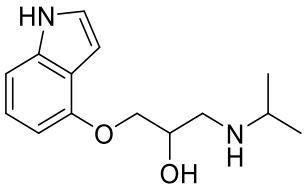
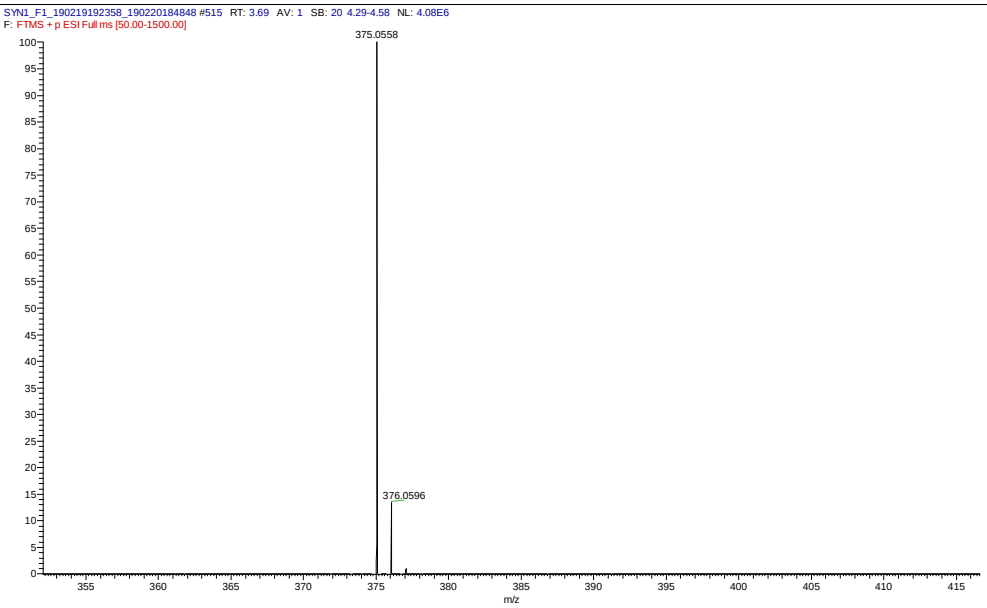
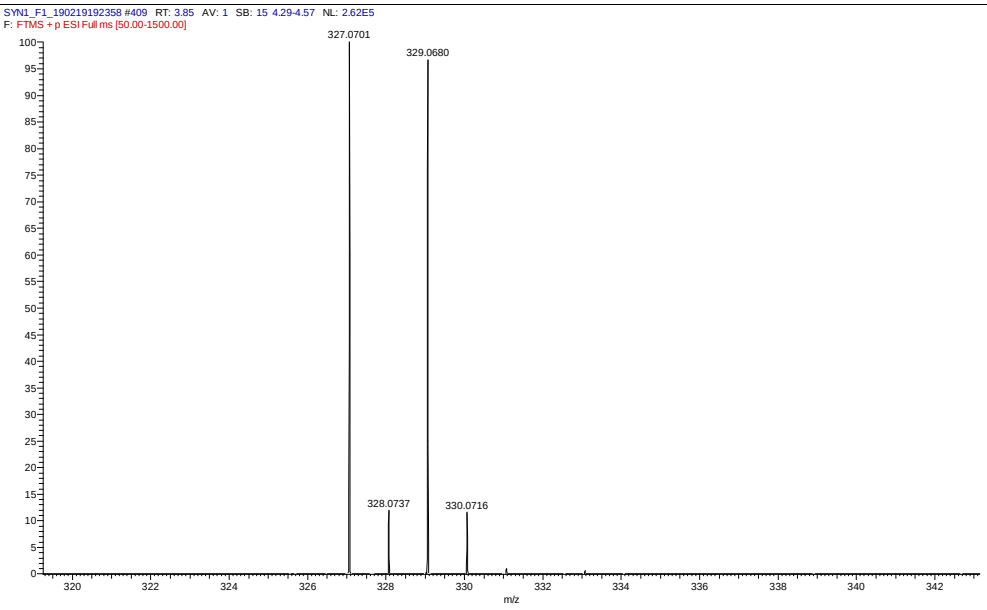
Supplementary Figure 90. LC-HRMS analysis of enzymatic iodination of compound 32 using VirX1. Only mono-iodinated product was observed on EIC chromatogram, while HRMS confirmed formation of the depicted product. Mass spectrum (bottom panel) for mono-iodinated product.

Compound	
NaI	
NaBr	

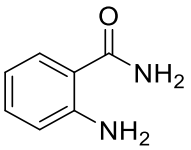
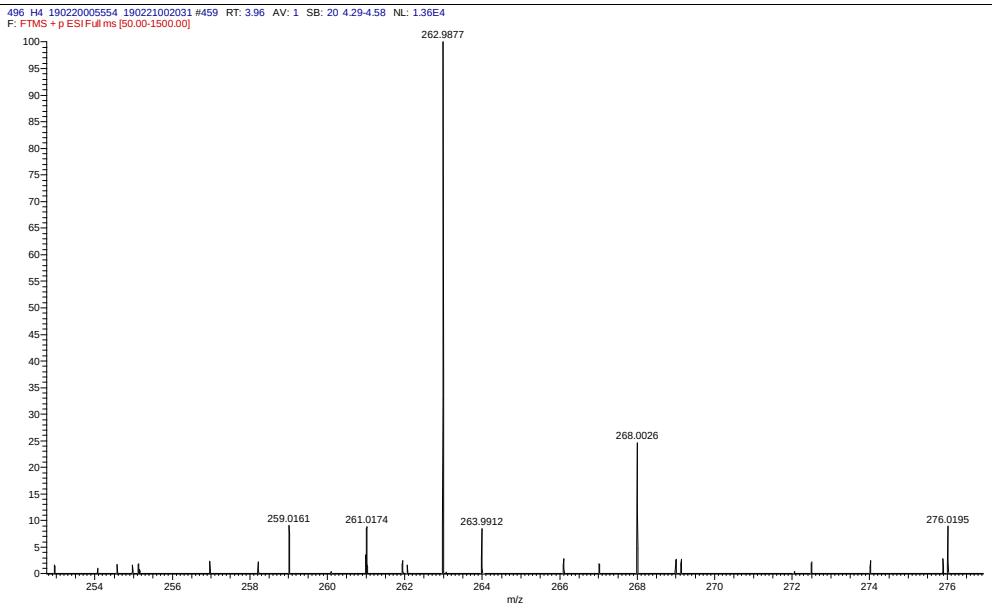
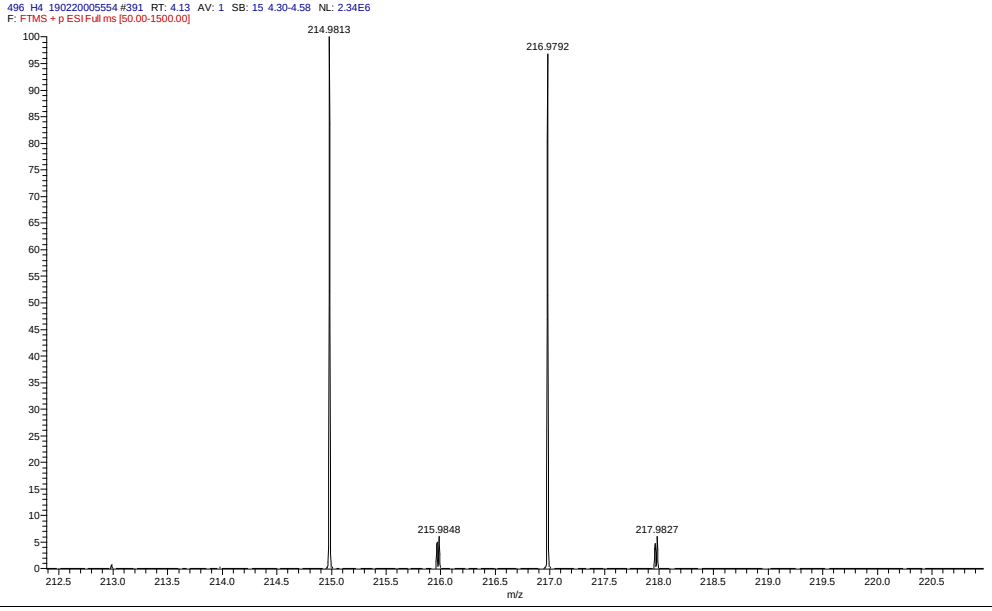
Supplementary Figure 91. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 1.

Compound	
NaI	
NaBr	

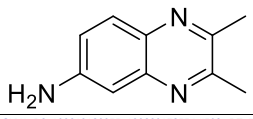
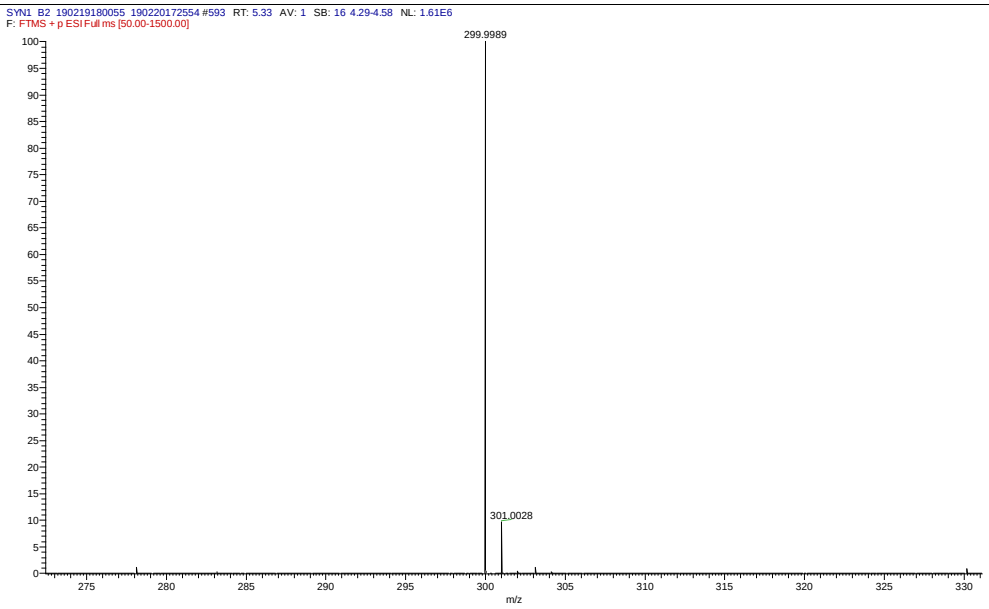
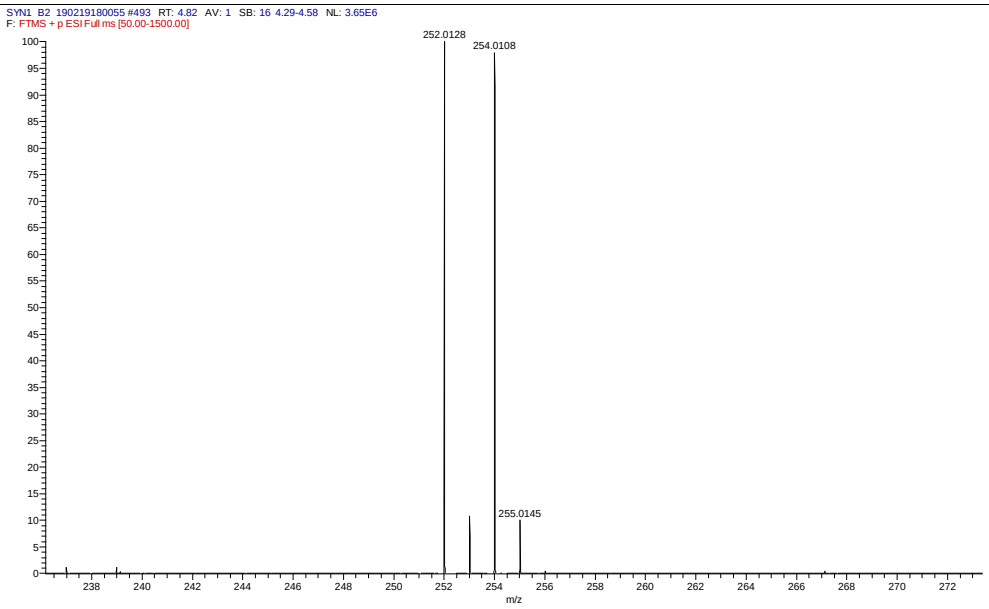
Supplementary Figure 92. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 2.

Compound	
NaI	
NaBr	

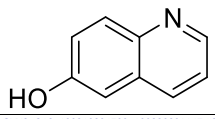
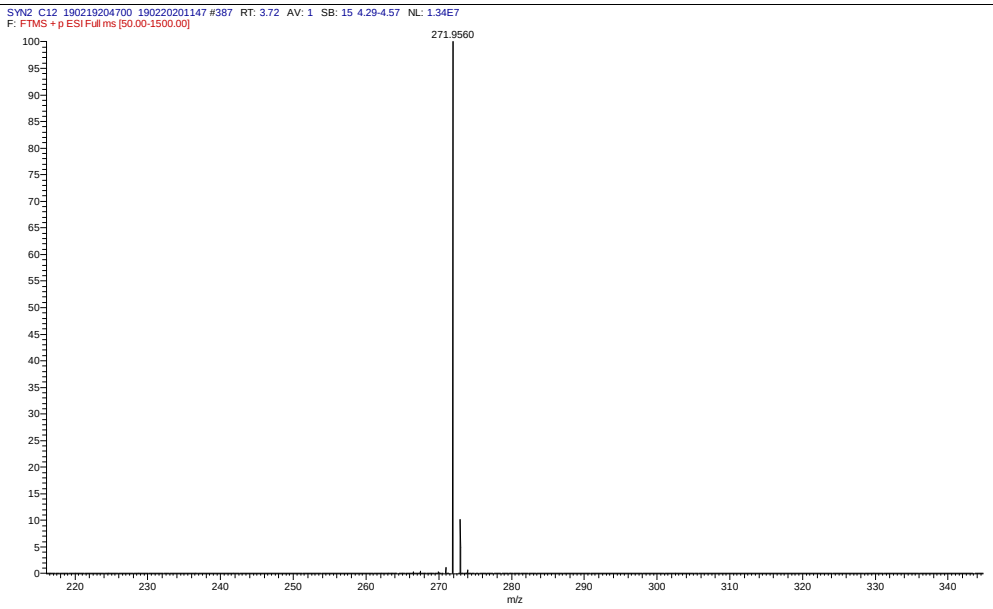
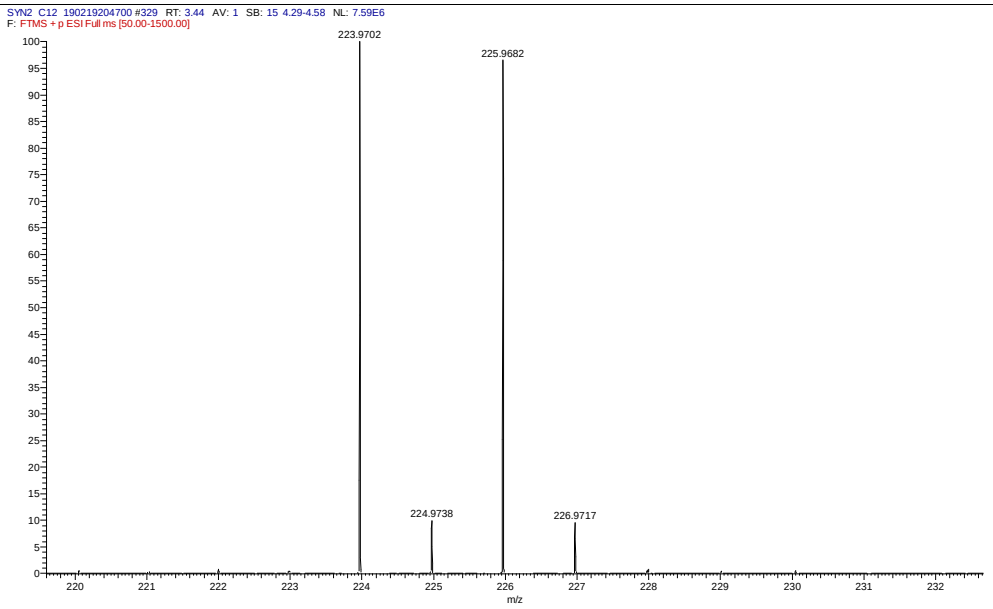
Supplementary Figure 93. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 4.

Compound	
NaI	
NaBr	

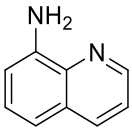
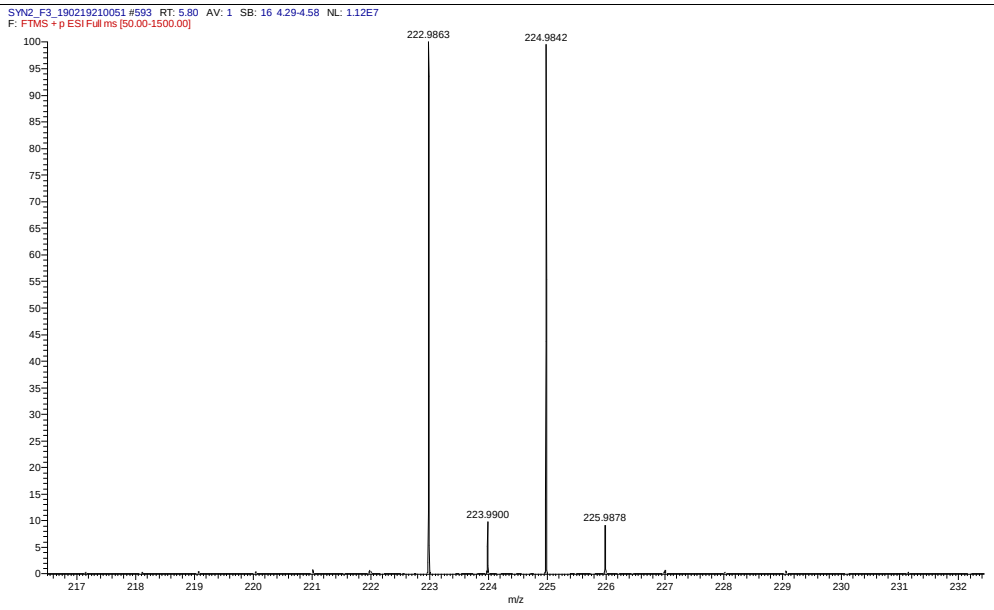
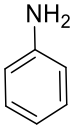
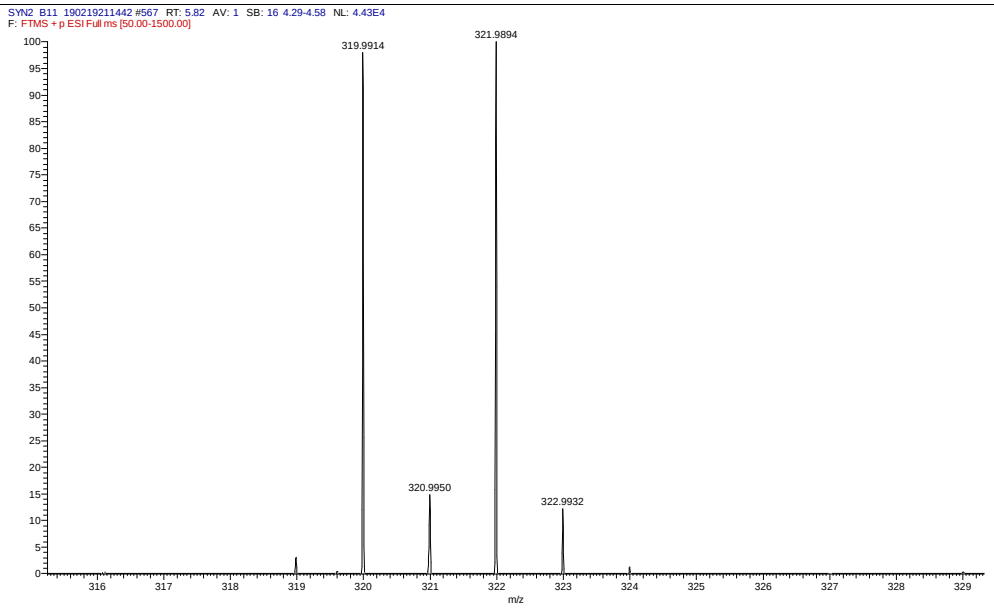
Supplementary Figure 94. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 7.

Compound	
NaI	
NaBr	

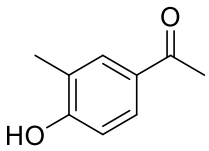
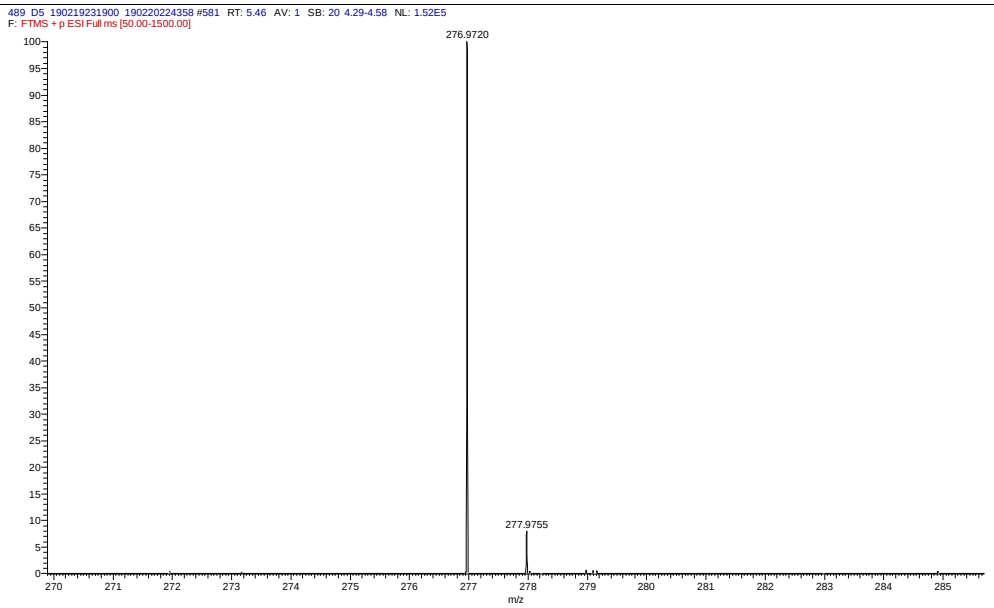
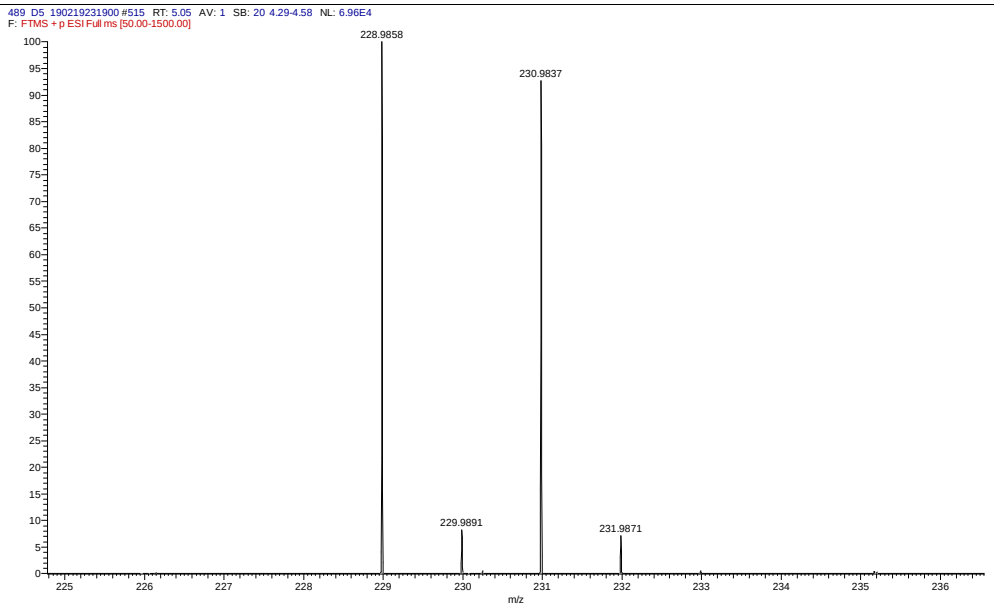
Supplementary Figure 95. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 8.

Compound	
NaI	
NaBr	

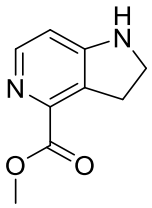
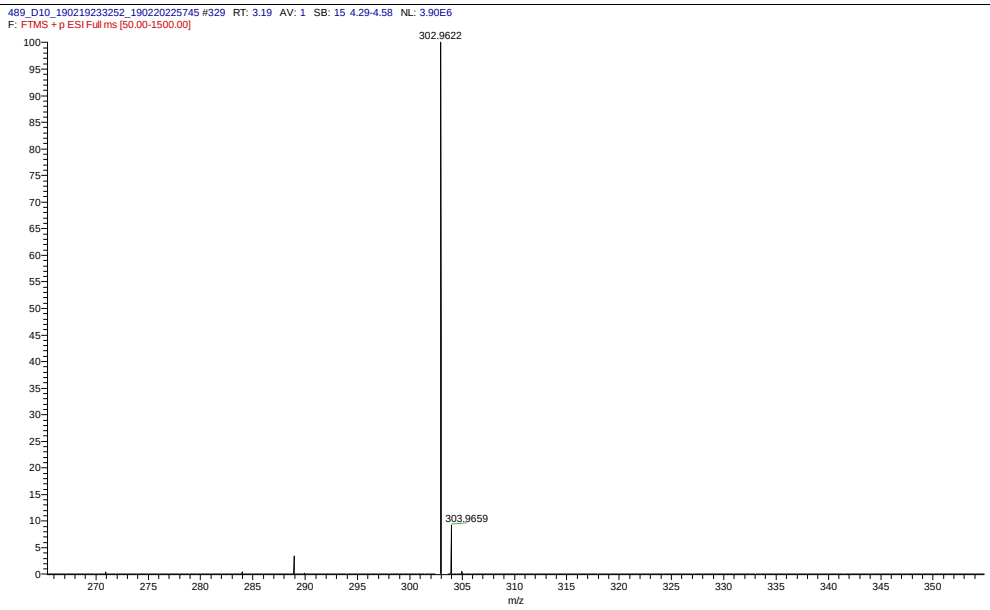
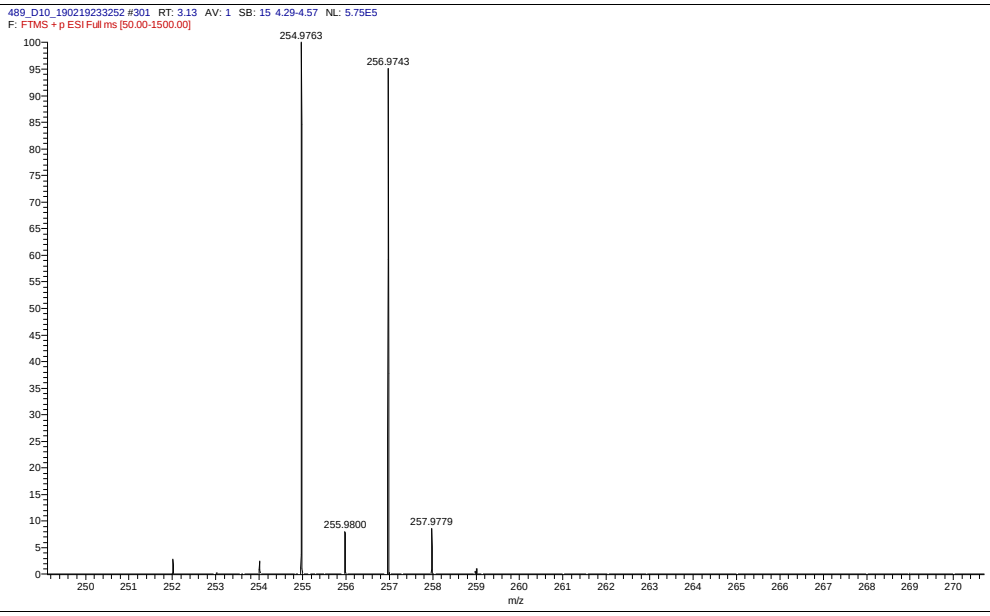
Supplementary Figure 96. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 11.

Compound	
NaBr	
Compound	
NaBr	

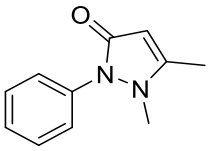
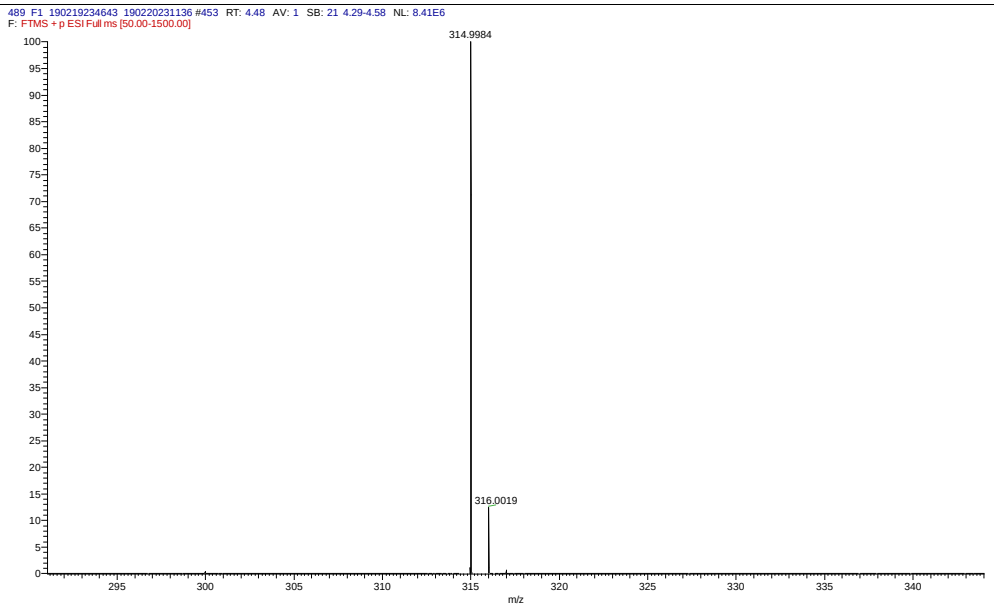
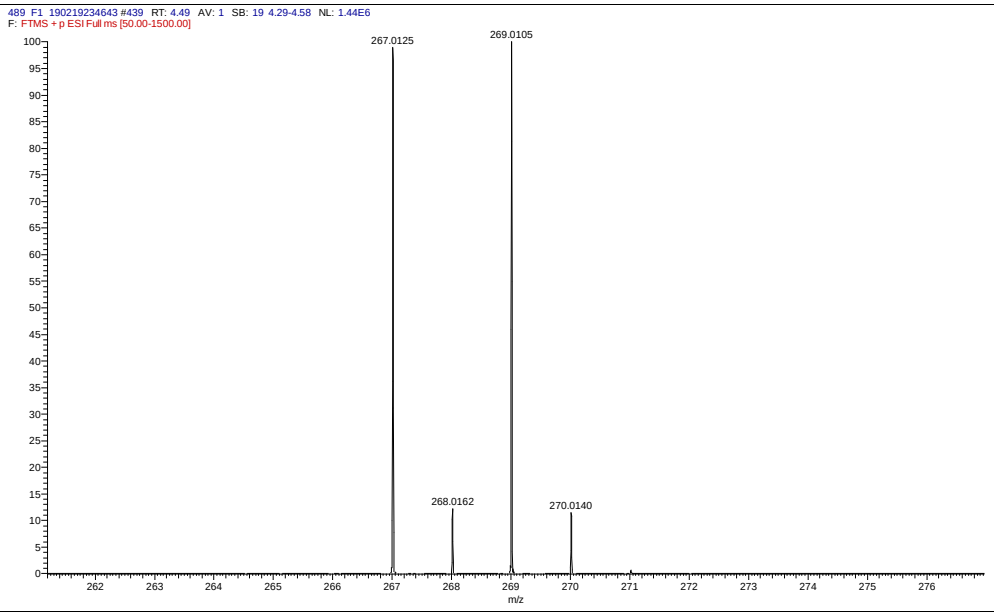
Supplementary Figure 97. LC-HRMS spectra for brominated products of PrnA with compound 13 (top) and compound 24 (bottom). Only brominated products were observed.

Compound	
NaI	
NaBr	

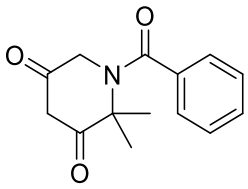
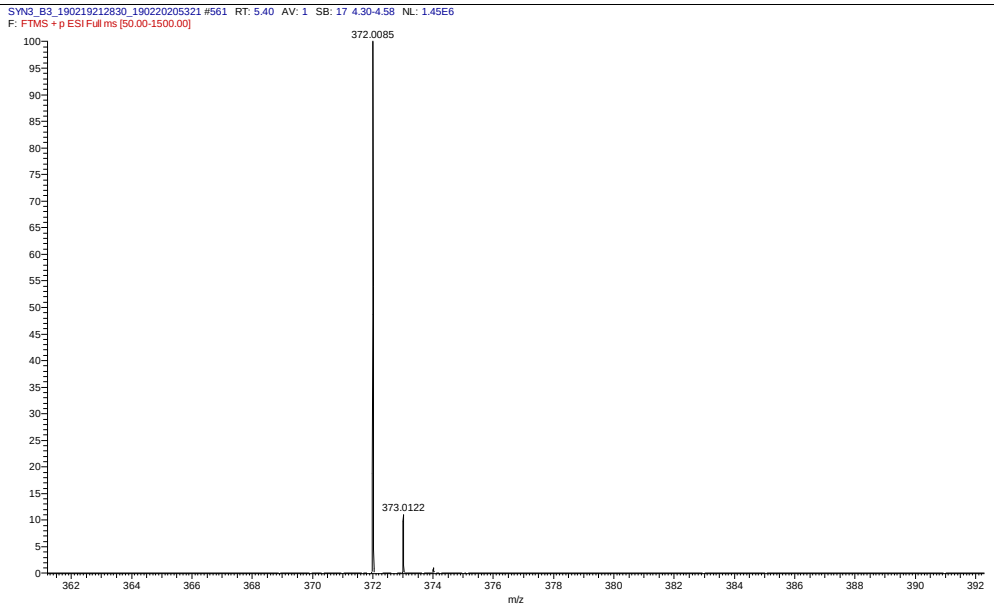
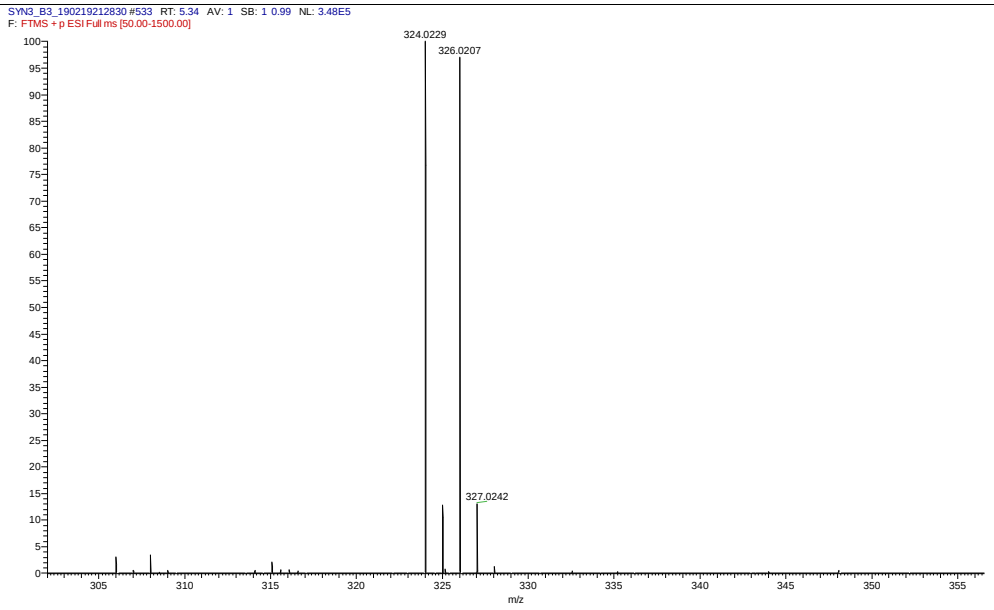
Supplementary Figure 98. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 15.

Compound	
NaI	
NaBr	

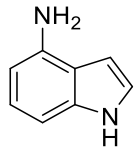
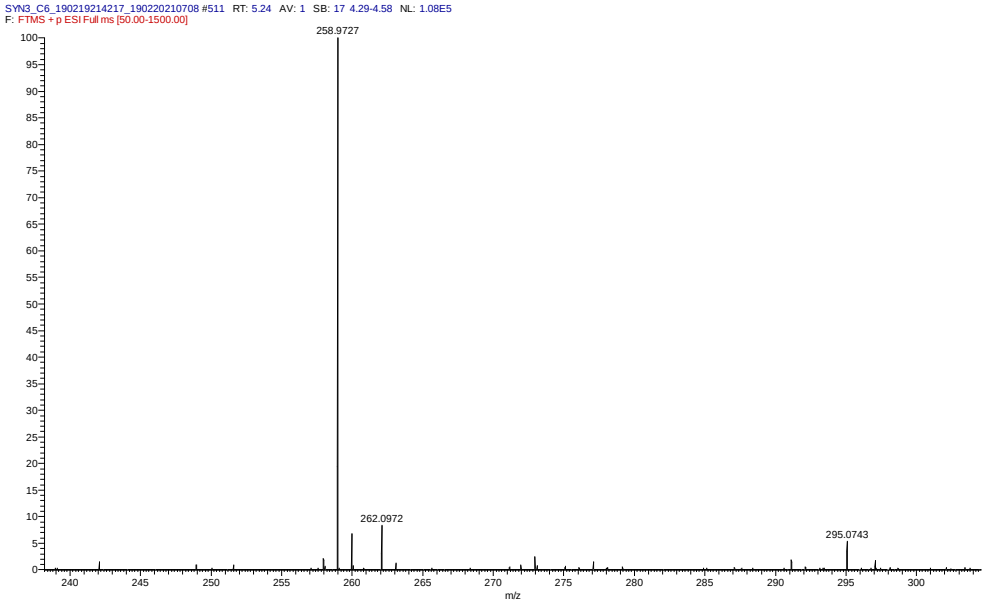
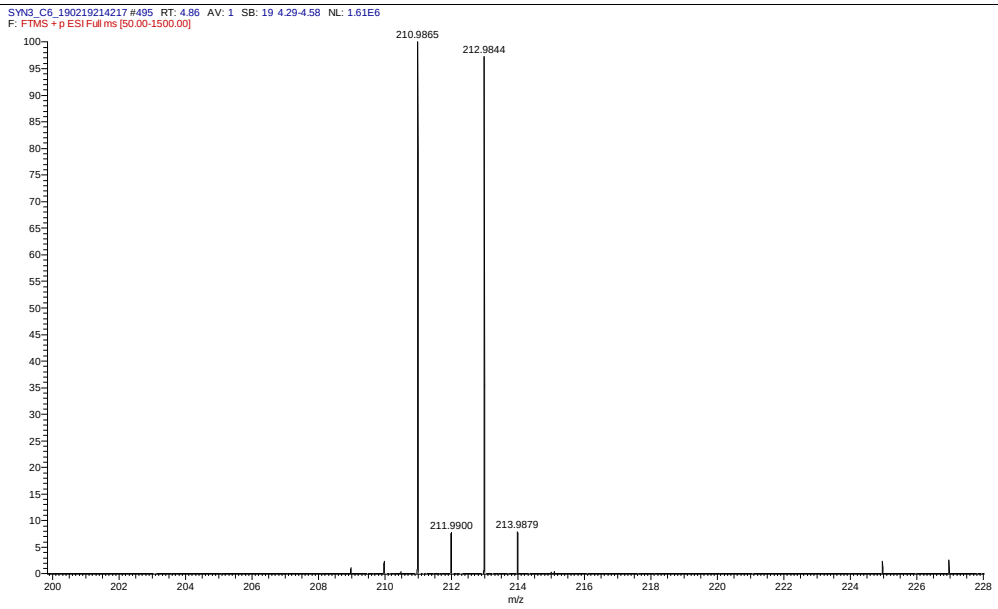
Supplementary Figure 99. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 16.

Compound	
NaI	
NaBr	

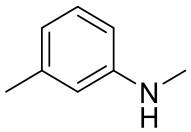
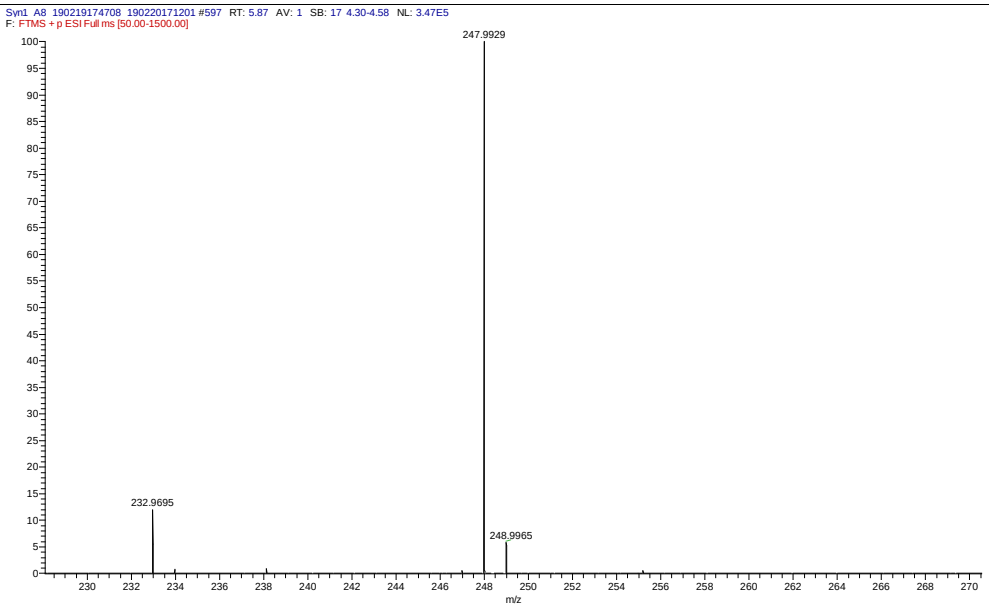
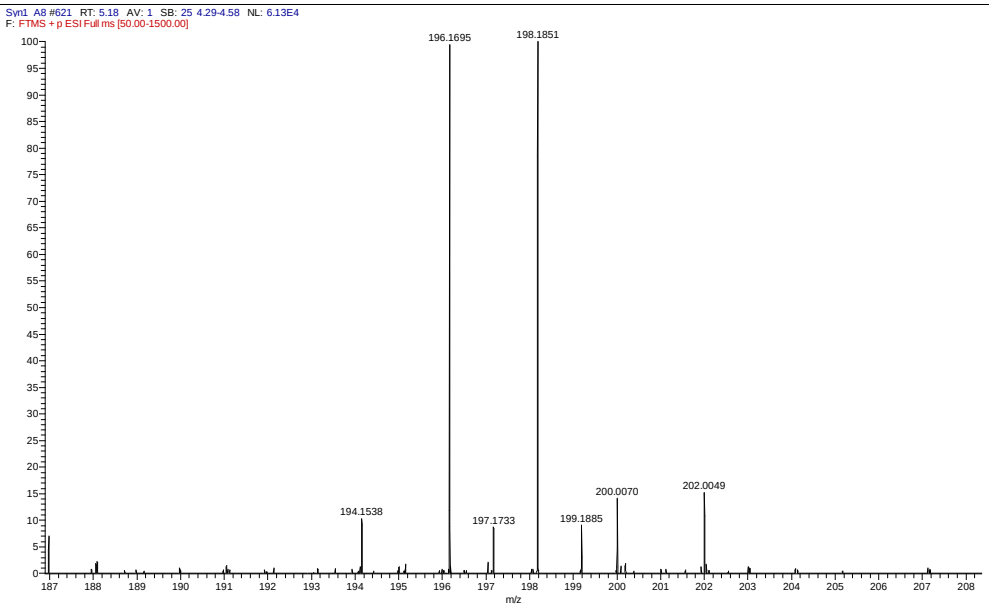
Supplementary Figure 100. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 17.

Compound	
NaI	
NaBr	

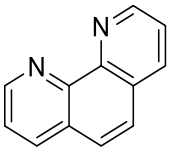
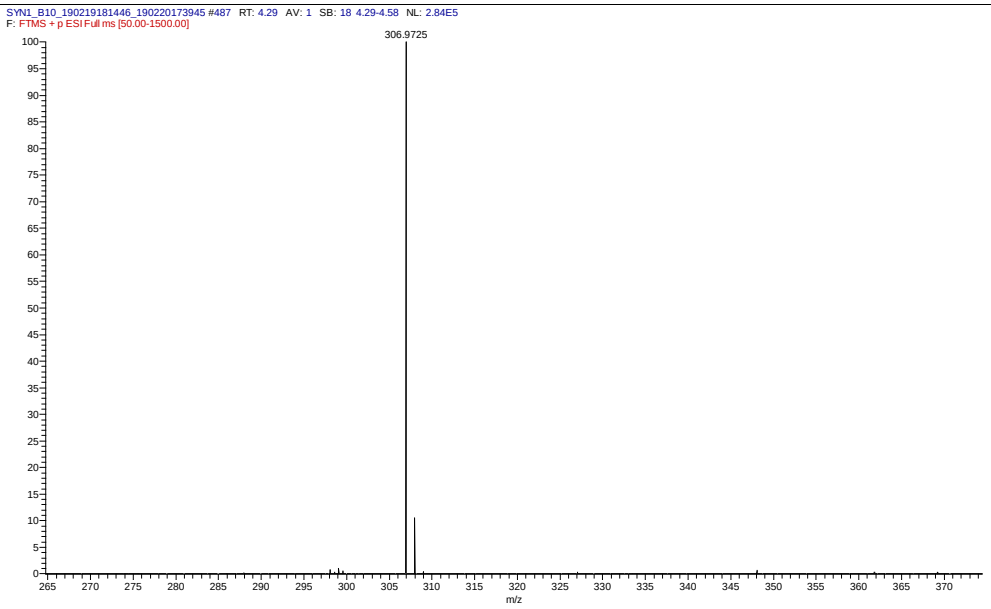
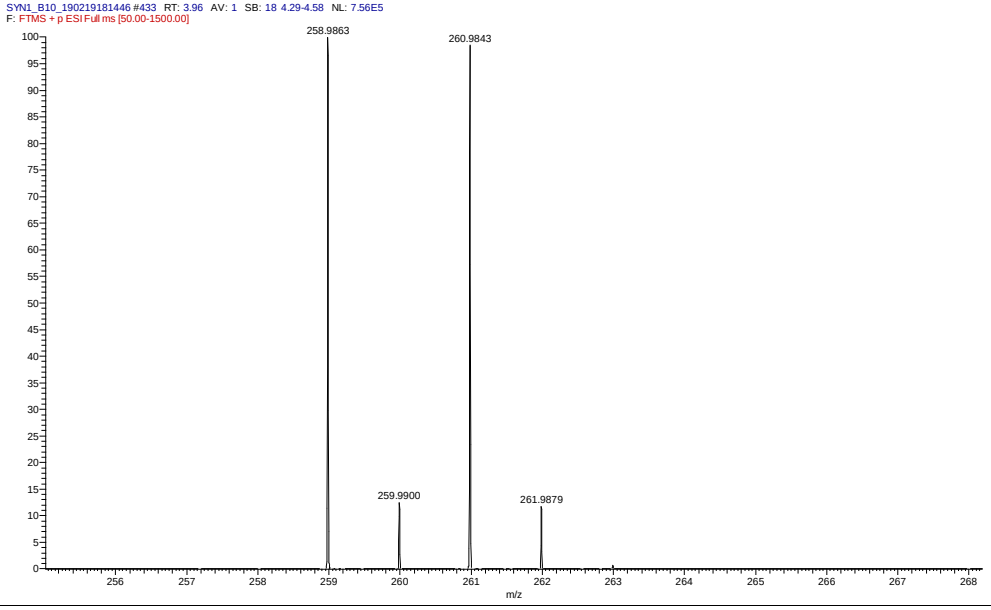
Supplementary Figure 101. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 18.

Compound	
NaI	
NaBr	

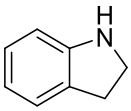
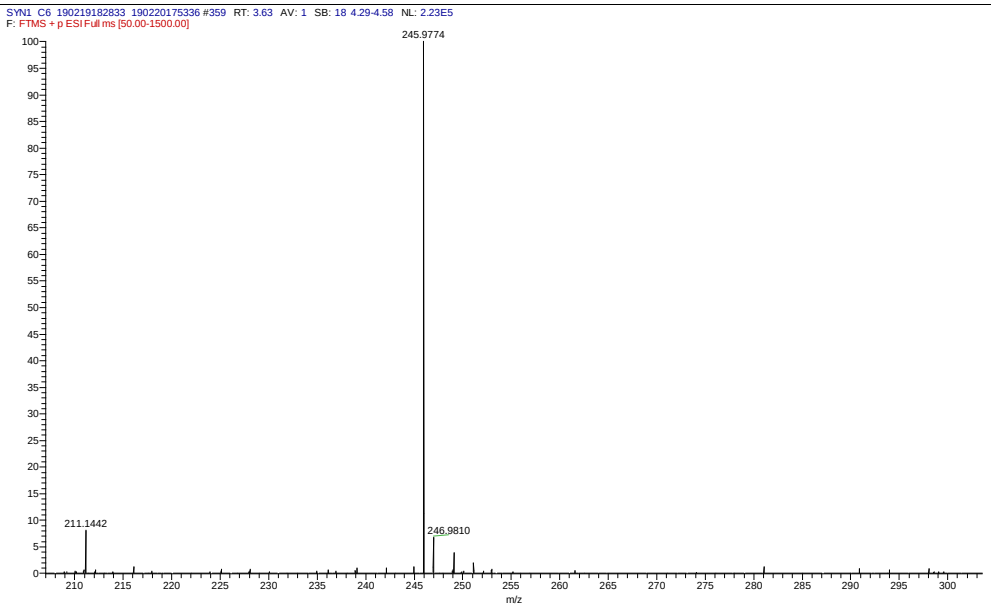
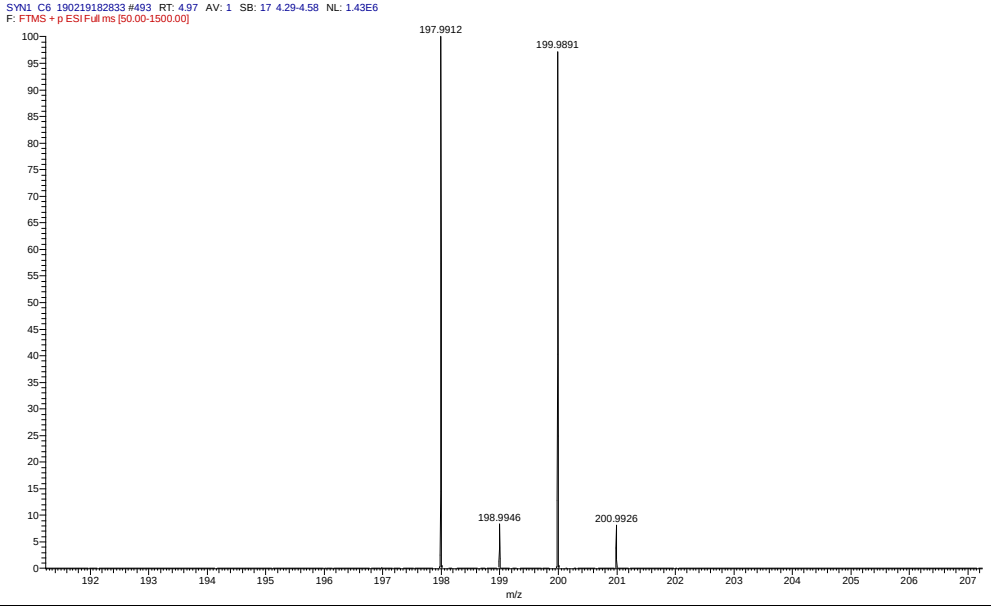
Supplementary Figure 102. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 19.

Compound	
NaI	
NaBr	

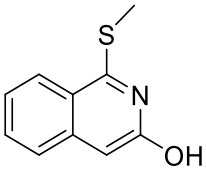
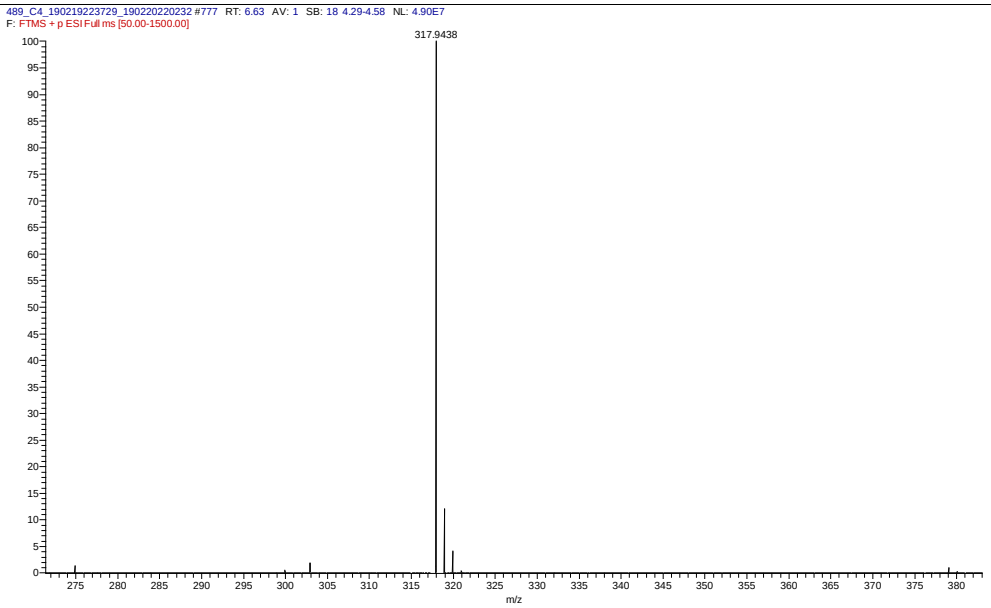
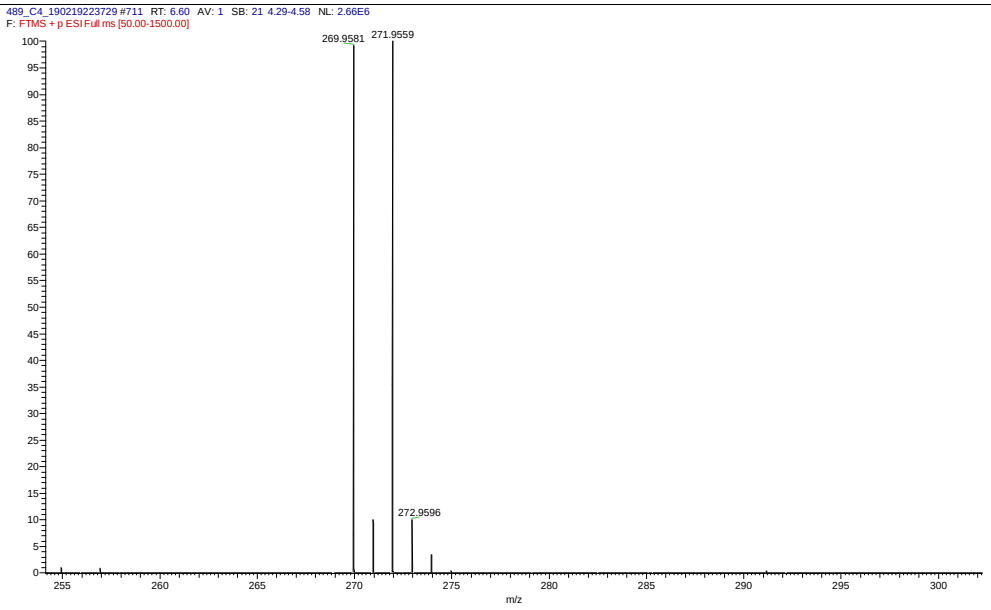
Supplementary Figure 103. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 20.

Compound	
NaI	
NaBr	

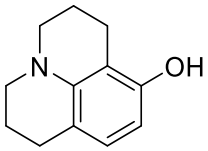
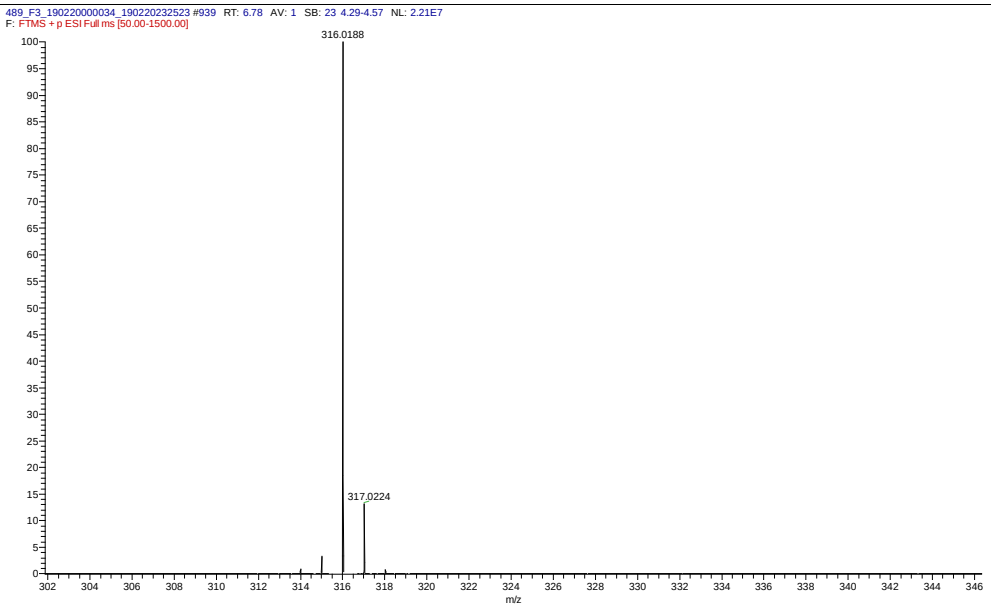
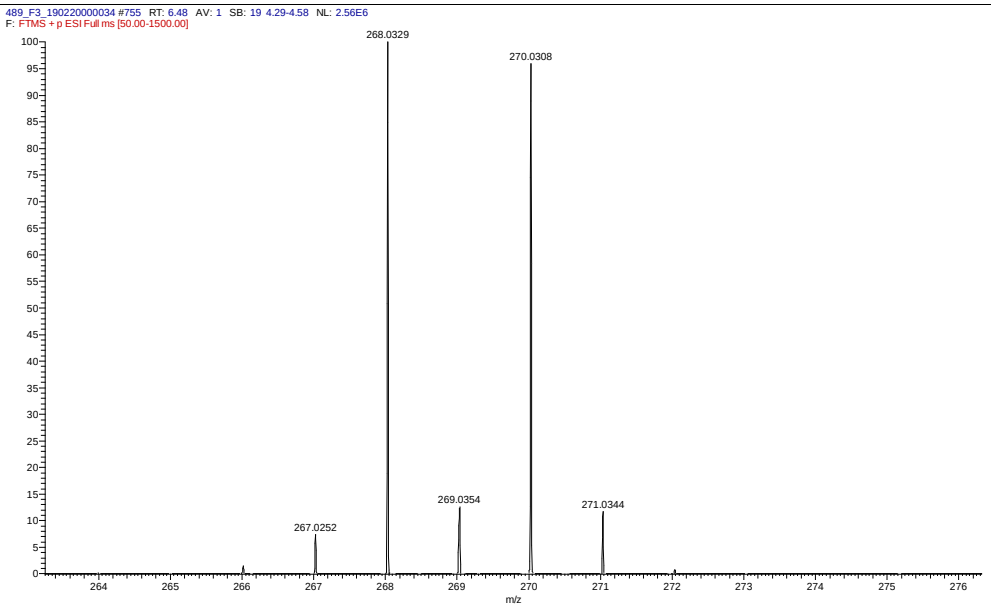
Supplementary Figure 104. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 21.

Compound	
NaI	
NaBr	

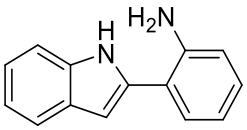
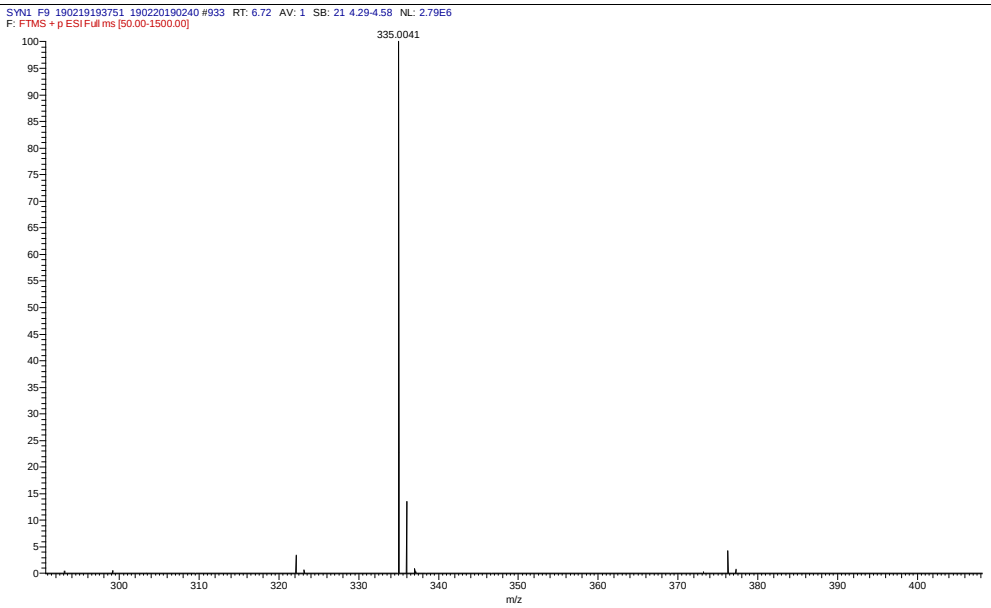
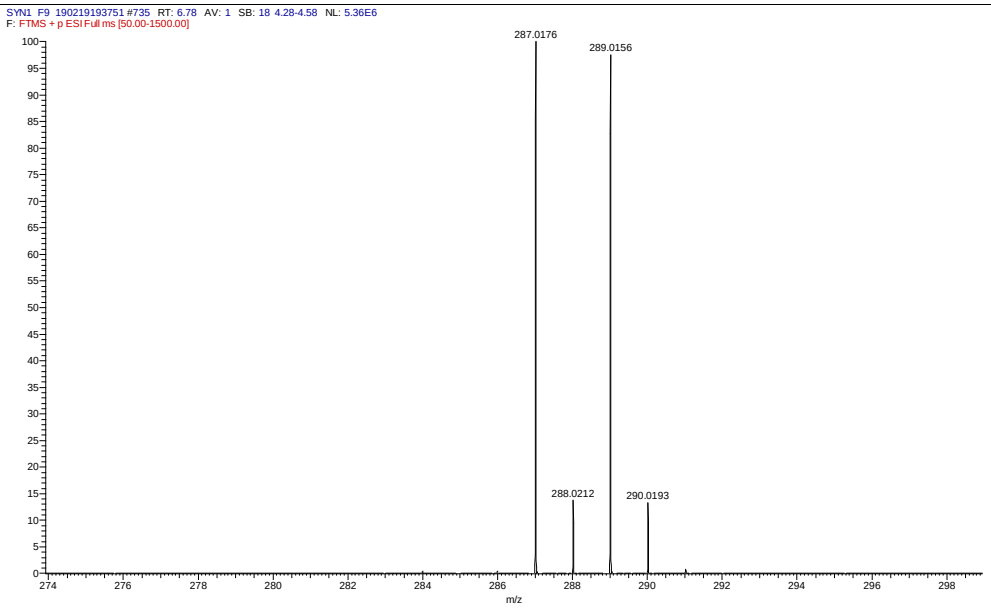
Supplementary Figure 105. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 22.

Compound	
NaI	
NaBr	

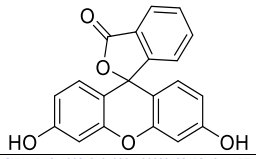
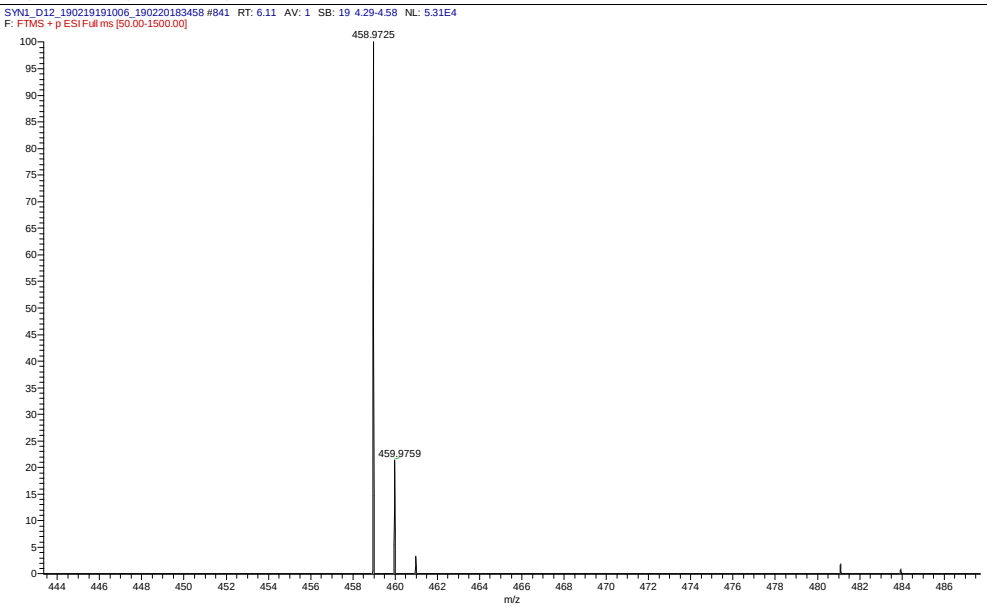
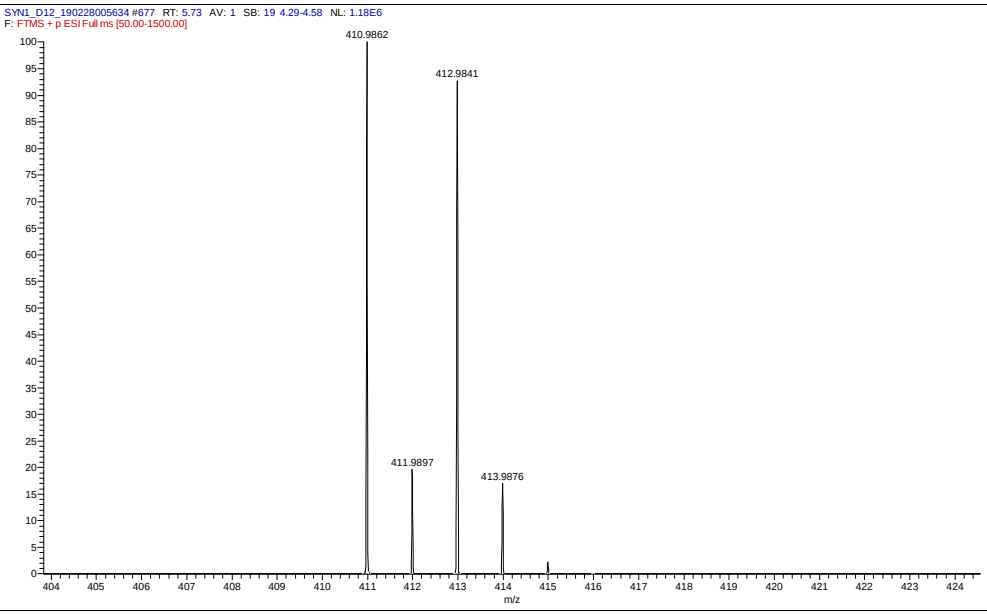
Supplementary Figure 106. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 25.

Compound	
NaI	
NaBr	

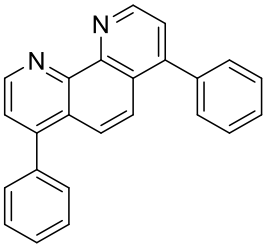
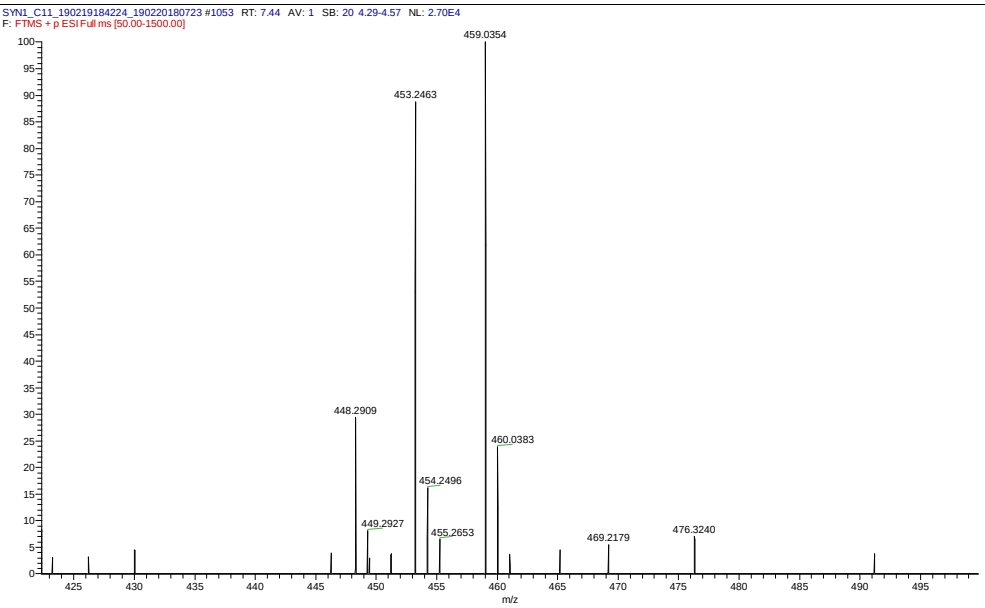
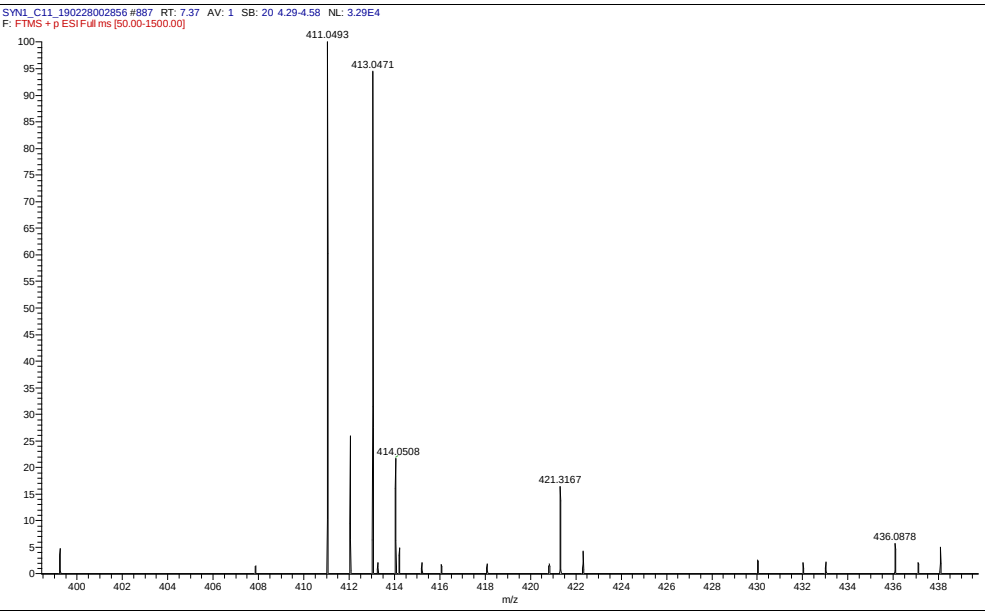
Supplementary Figure 107. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 26.

Compound	
NaI	
NaBr	

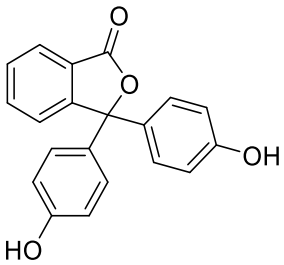
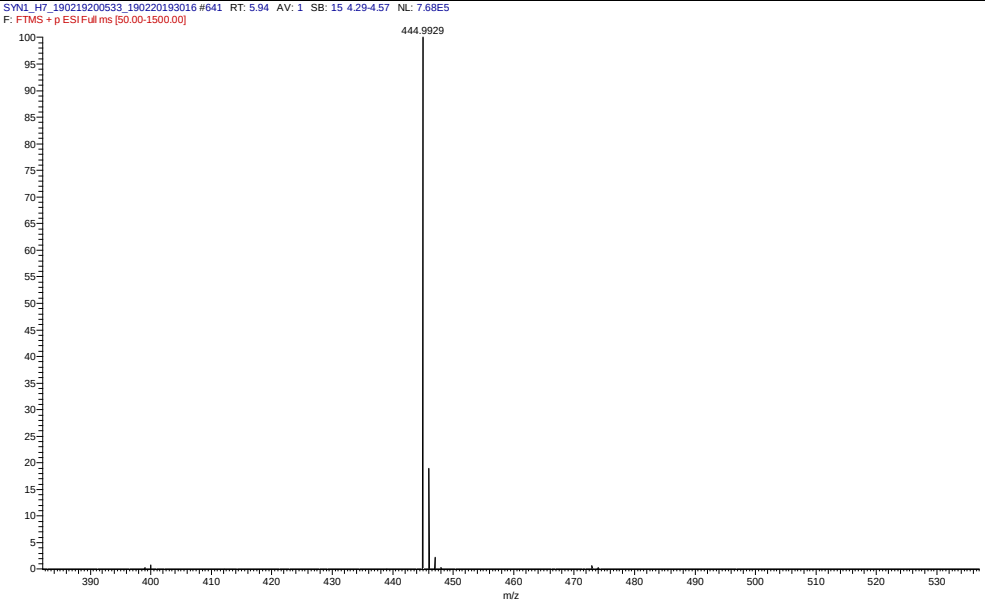
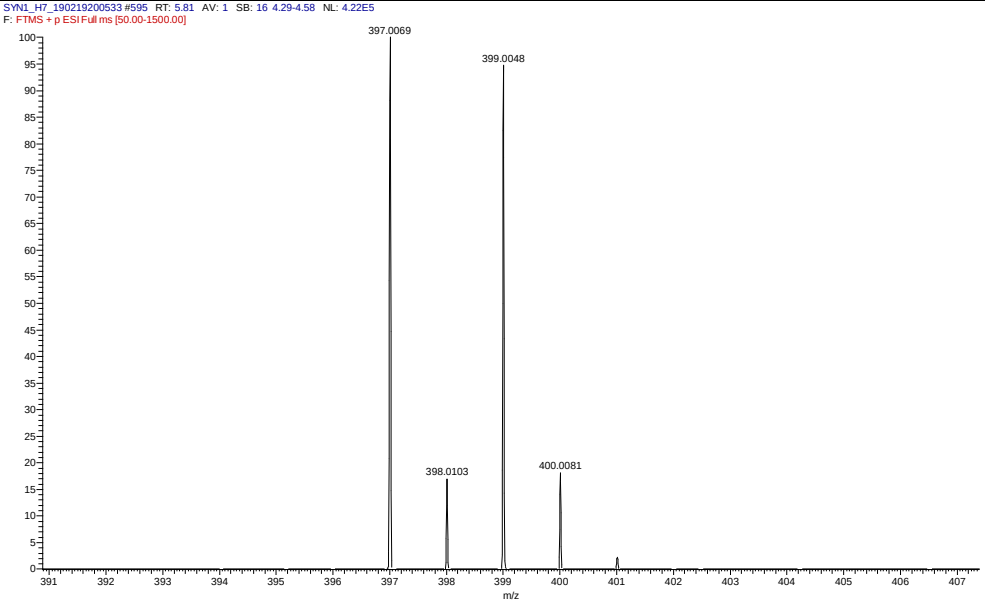
Supplementary Figure 108. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 27.

Compound	
NaI	
NaBr	

Supplementary Figure 109. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 29.

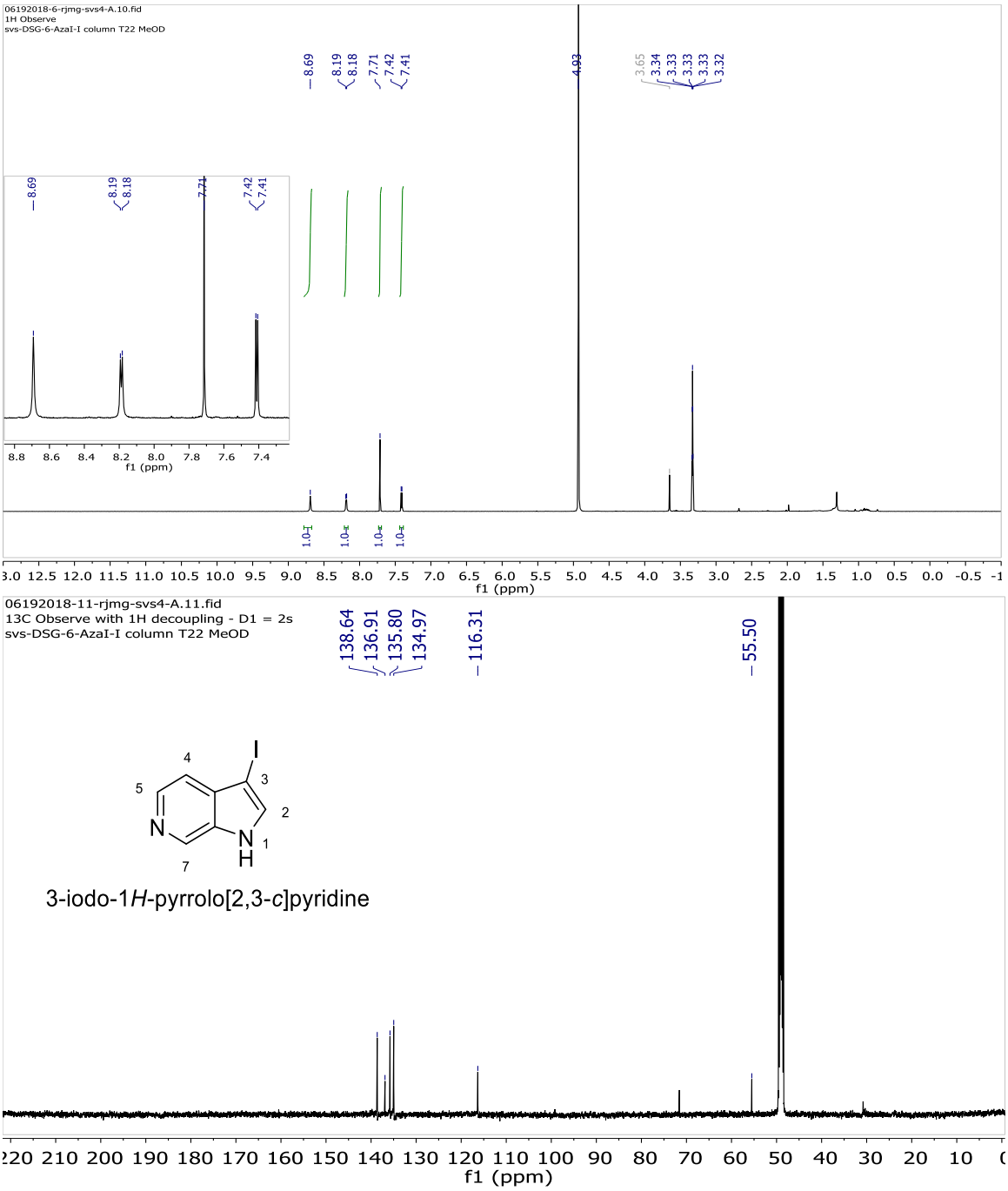
Compound	
NaI	
NaBr	

Supplementary Figure 110. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 30.

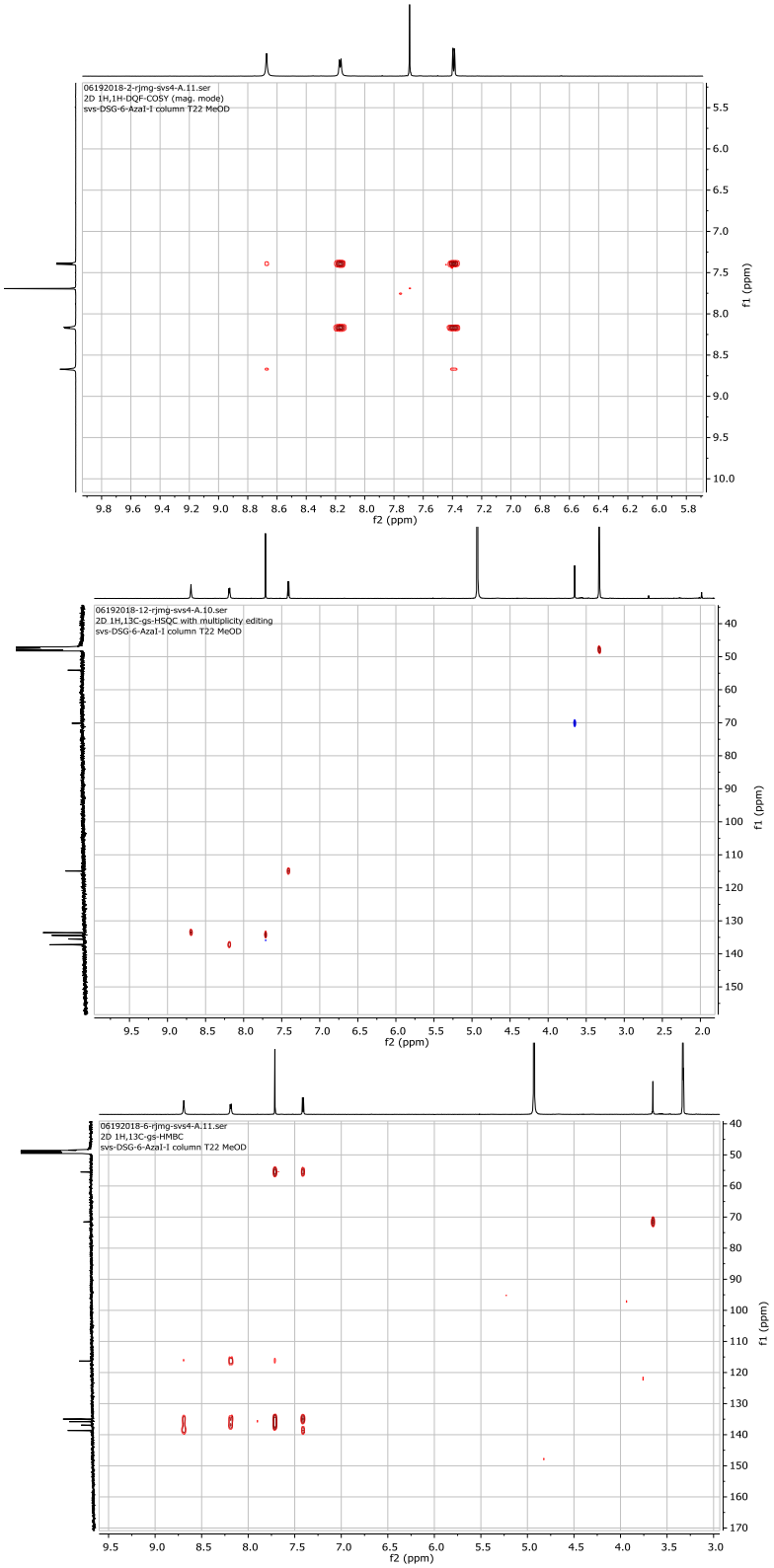
Compound	
NaI	
NaBr	

Supplementary Figure 111. LC-HRMS spectra for brominated and iodinated product of PrnA with compound 32.

3-Iodo-6-azaindole (1)

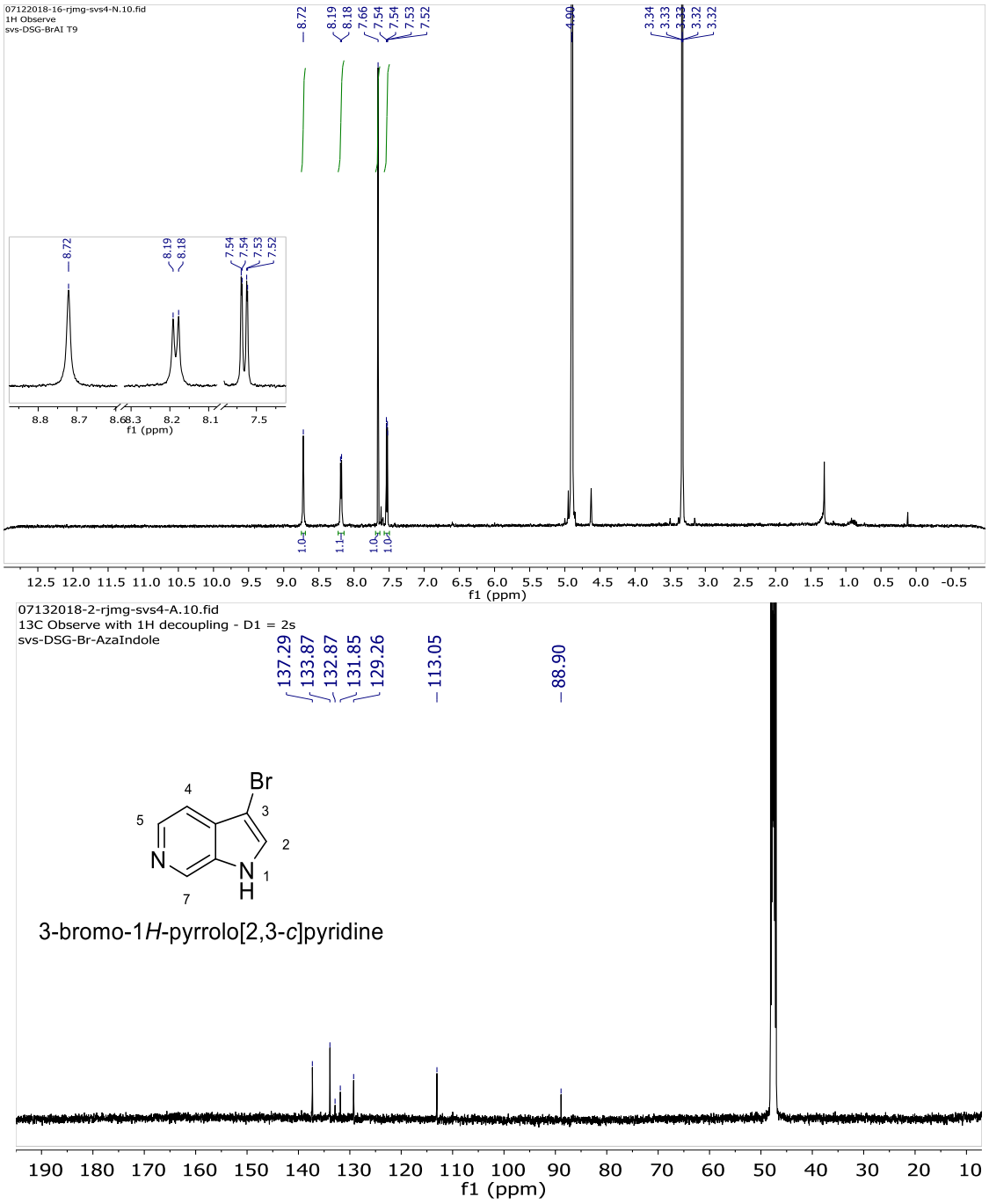


Supplementary Figure 112. ^{1}H NMR (500 MHz, CD_3OD) and ^{13}C NMR (126 MHz, CD_3OD) of 3-iodo-6-azaindole

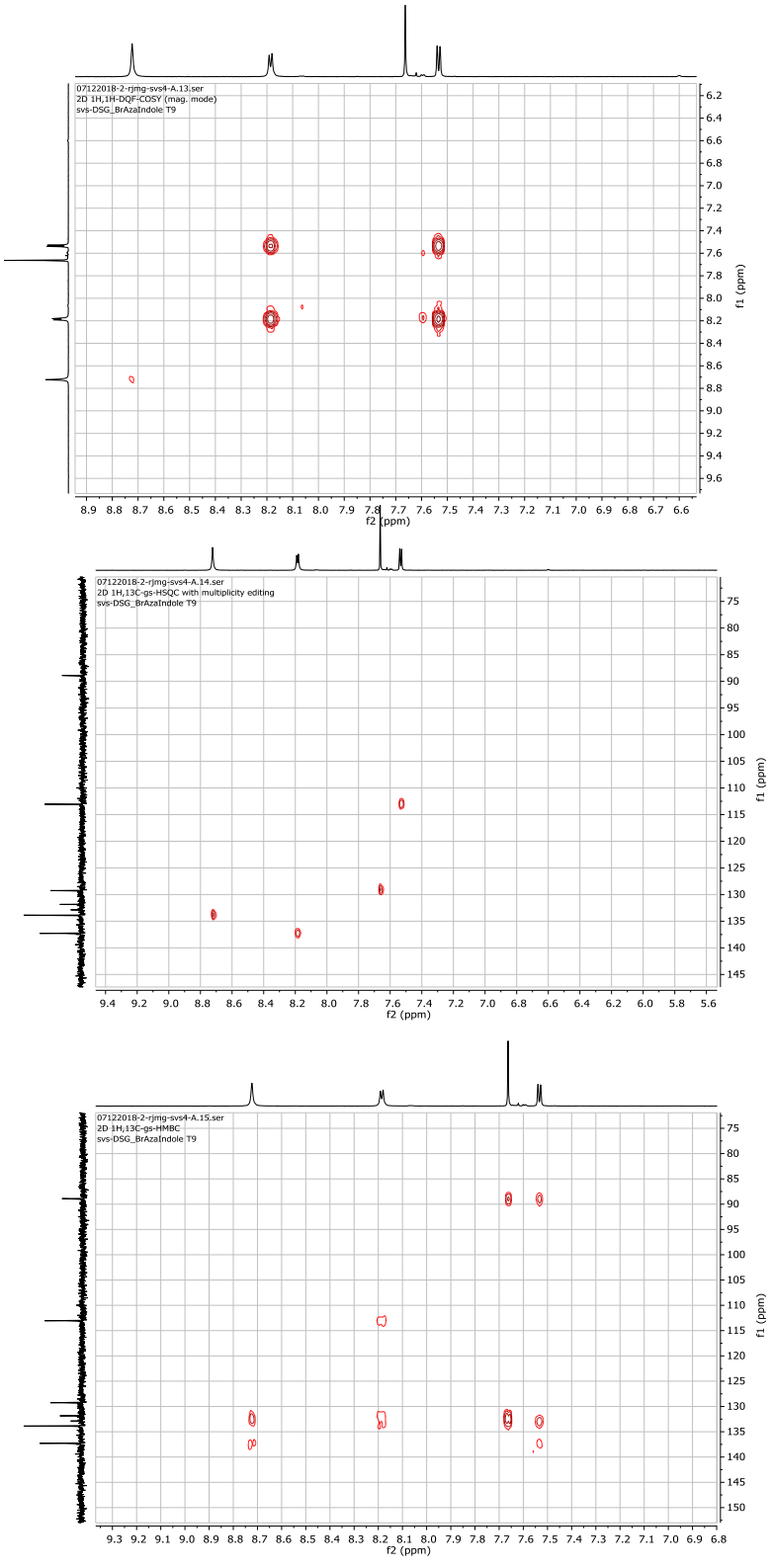


Supplementary Figure 113. 2-D NMR (500 MHz, CD_3OD), COSY, HS QC and HMBC spectra of 3-iodo-6-azaindole

3-Bromo-6-azaindole

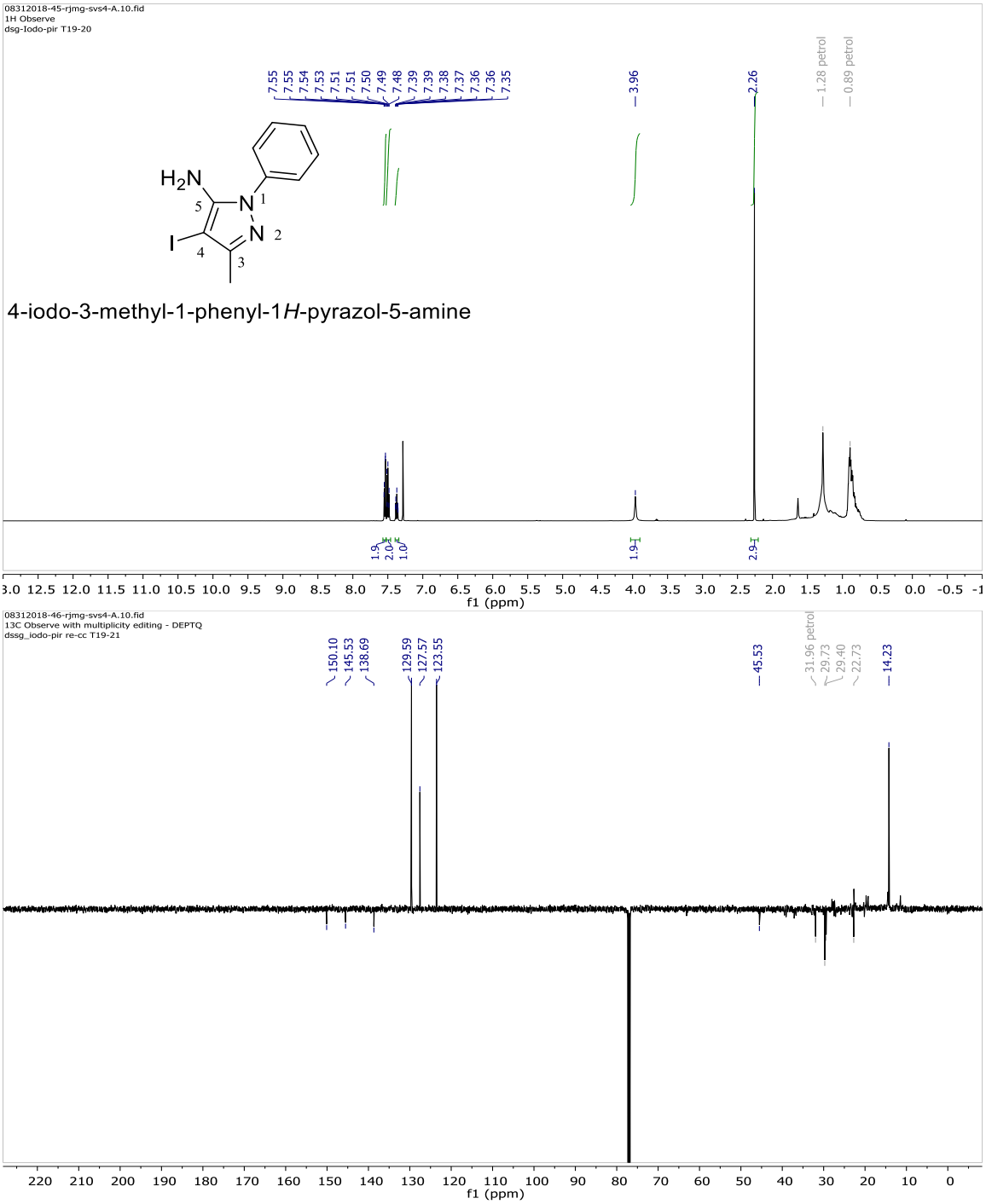


Supplementary Figure 114. ^{1}H NMR (500 MHz, CD_3OD) and ^{13}C NMR (126 MHz, CD_3OD) of 3-bromo-6-azaindole

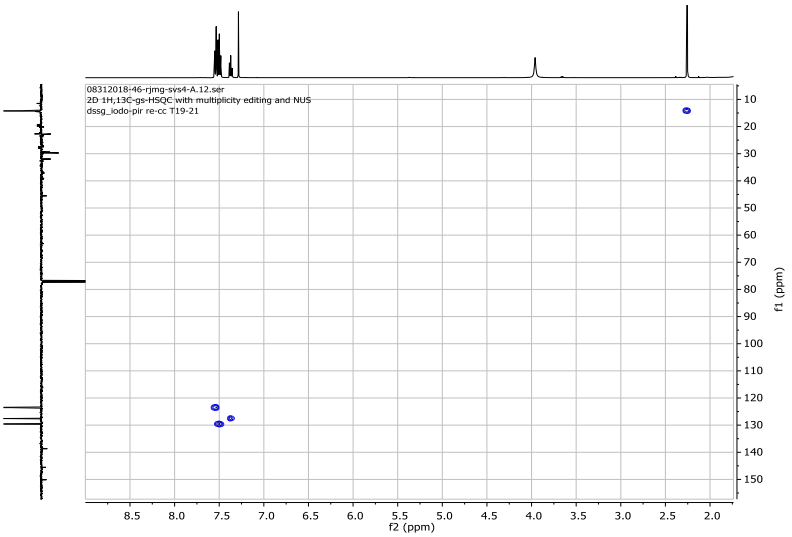


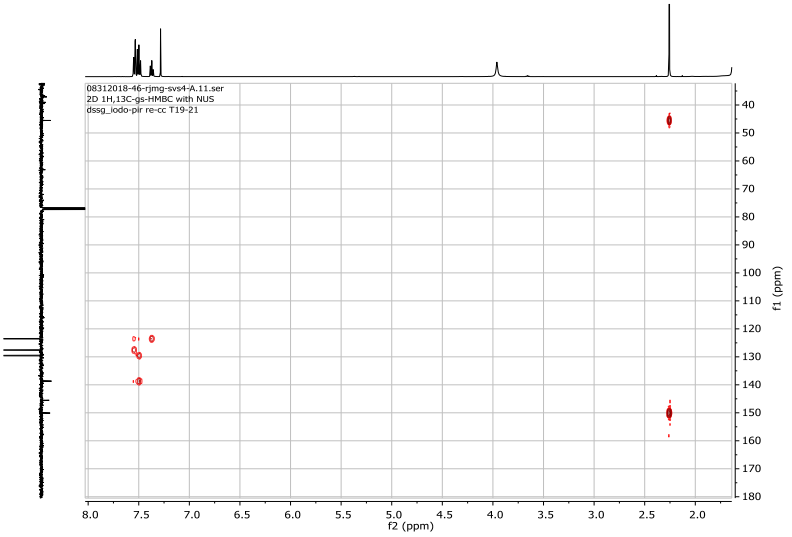
Supplementary Figure 115. 2-D NMR (500 MHz, CD_3OD), COSY, HS QC and HMBC spectra of 3-bromo-6-azaindole

5-Amino-4-iodo-3-methyl-1-phenylpyrazole (2)



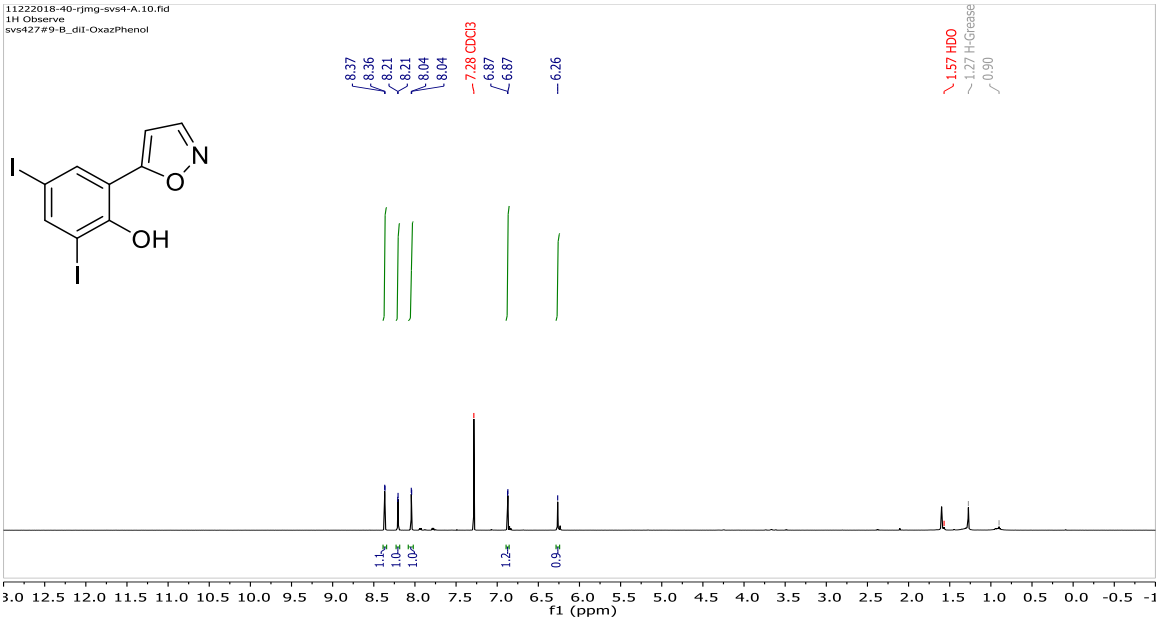
Supplementary Figure 116. ^{1}H NMR (500 MHz, $CDCl_3$) and ^{13}C NMR (126 MHz, $CDCl_3$) of 5-amino-4-iodo-3-methyl-1-phenylpyrazole

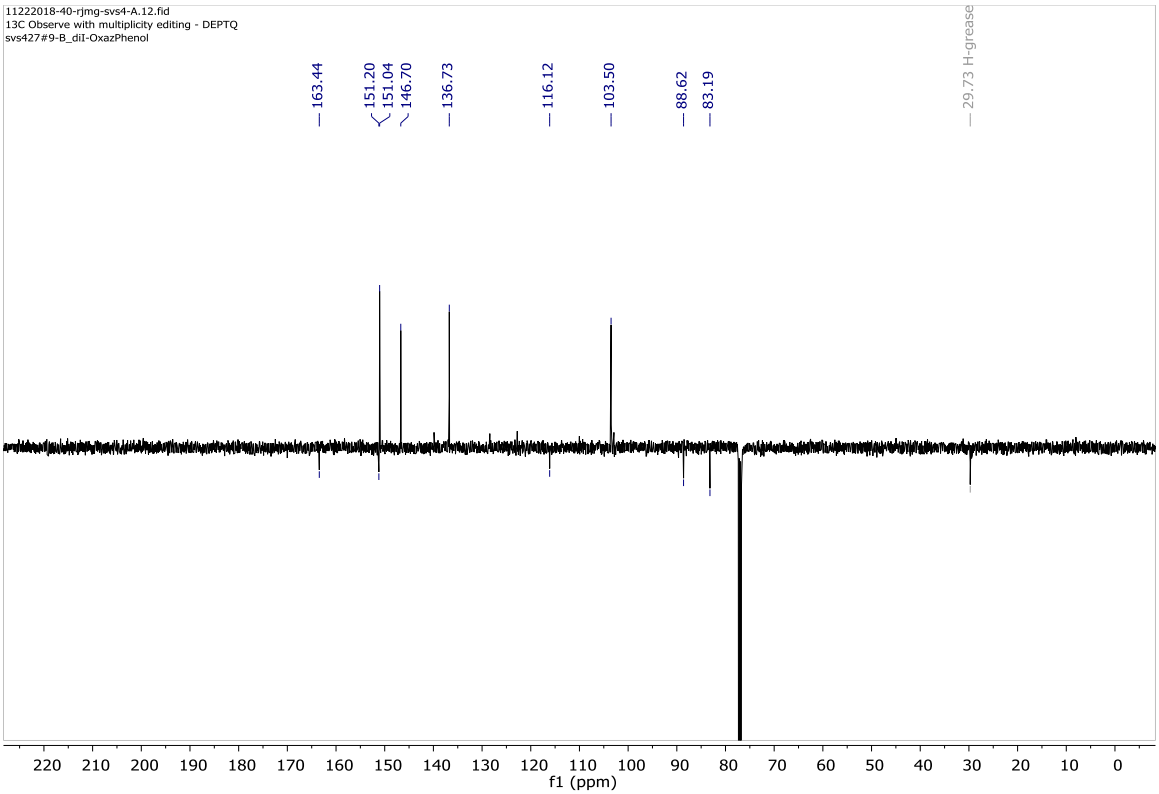




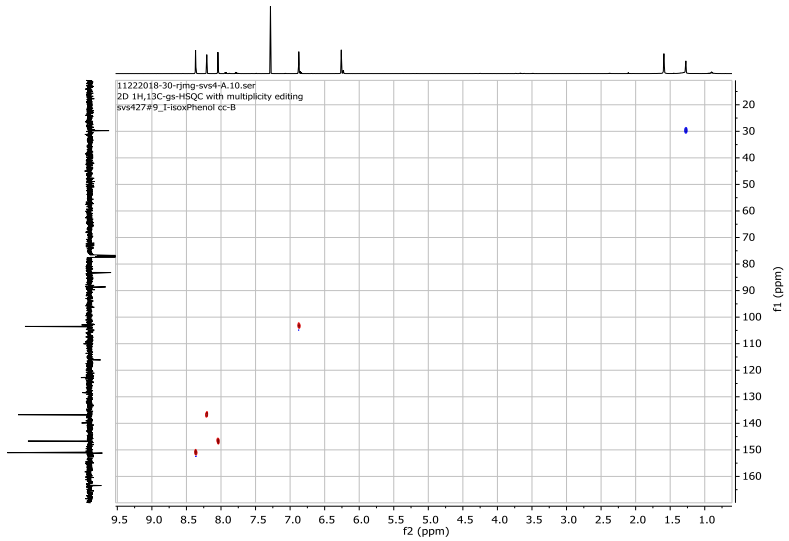
Supplementary Figure 117. 2-D NMR (500 MHz, $CDCl_3$), HSQC and HMBC spectra of 5-amino-4-iodo-3-methyl-1-phenylpyrazole

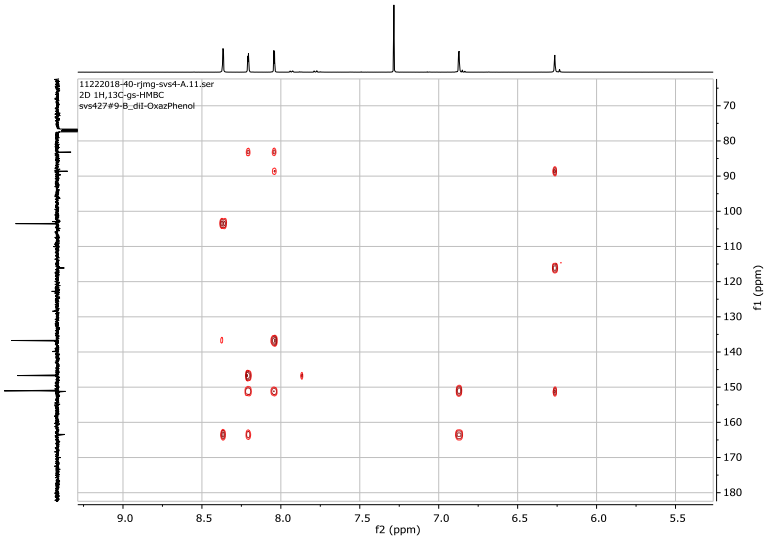
2-(5-Isoxazolyl)-4,6-diiodophenol (3)





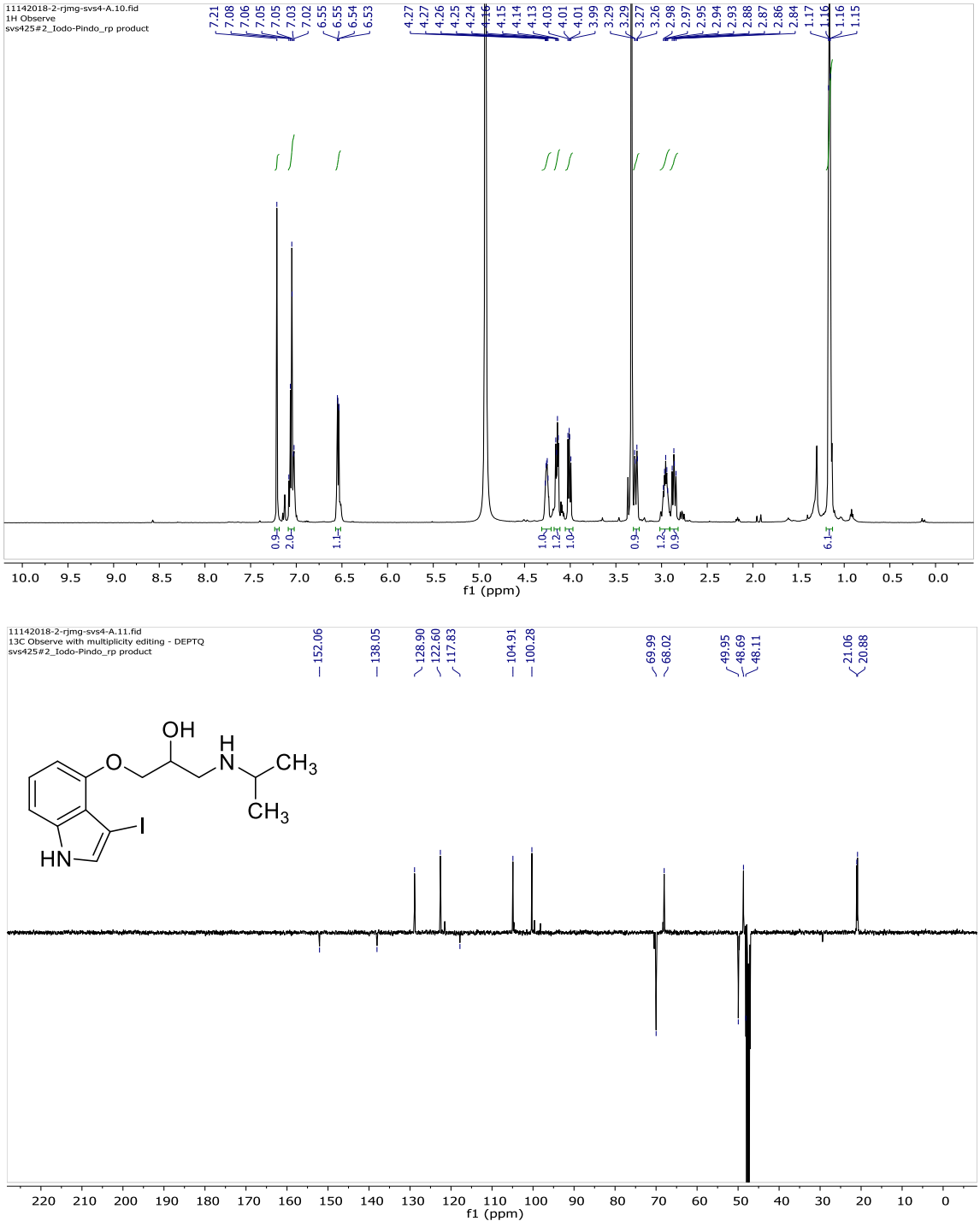
Supplementary Figure 118. ^{1}H NMR (500 MHz, $CDCl_3$) and ^{13}C NMR (126 MHz, $CDCl_3$) of 2-(5-isoxazolyl)-4,6-diiodophenol



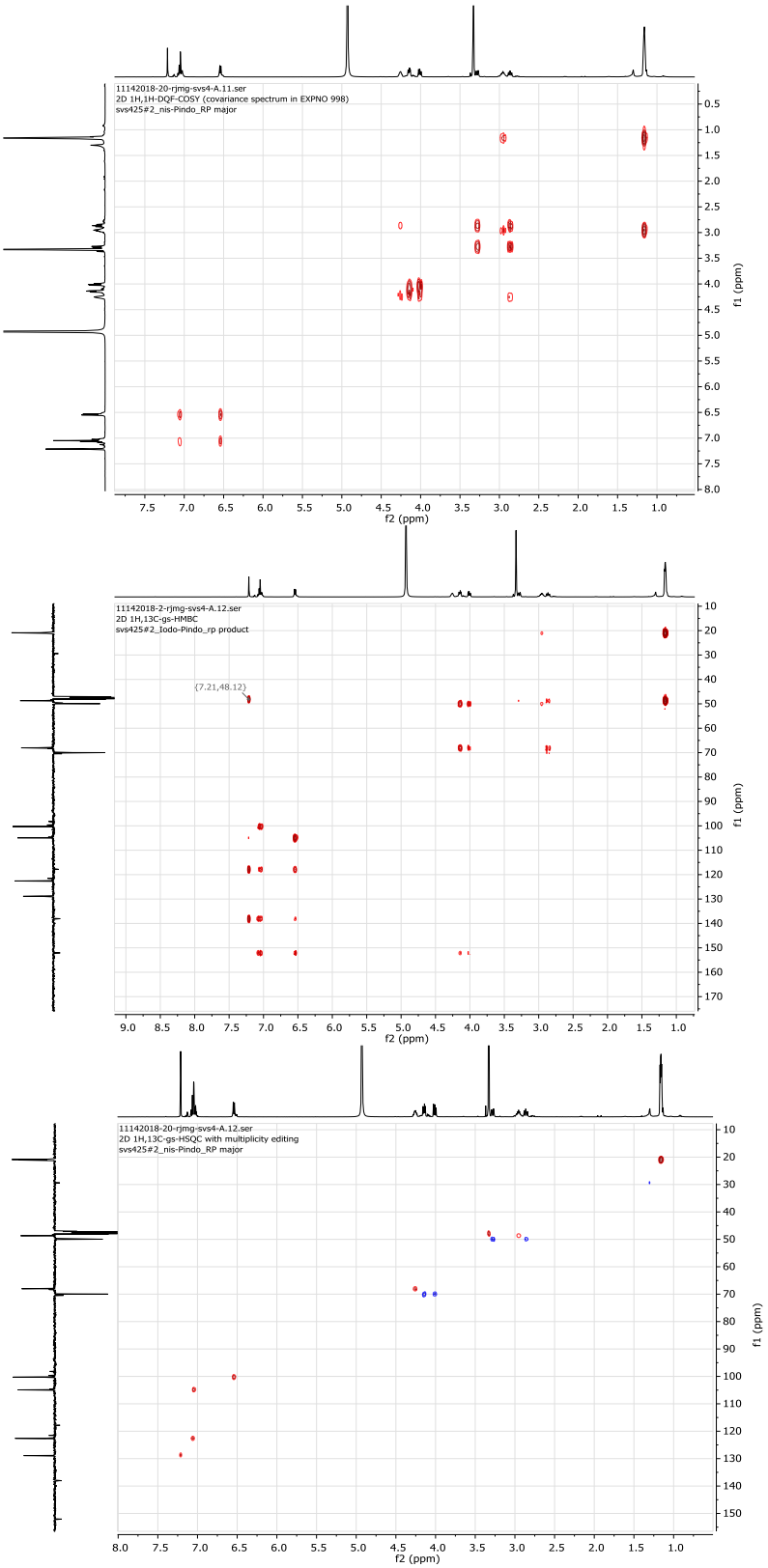


Supplementary Figure 119. 2-D NMR (500 MHz, $CDCl_3$), HSQC and HMBC spectra of of 2-(5-isoxazolyl)-4,6-diiodophenol

3-Iodo pindolol (4)

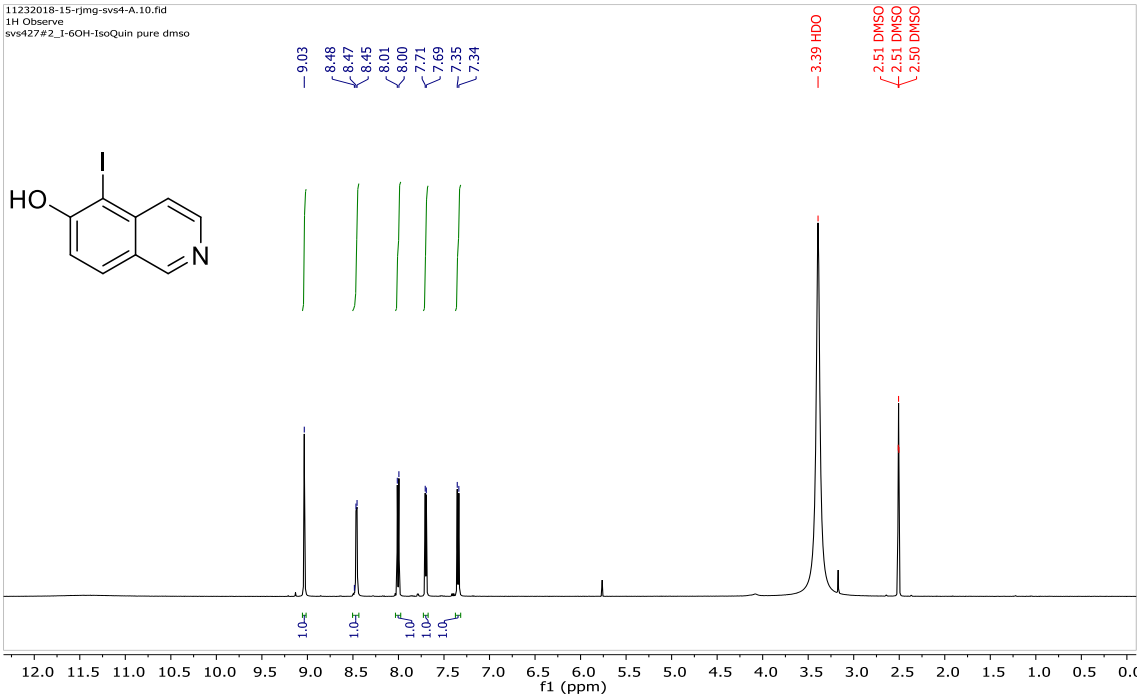


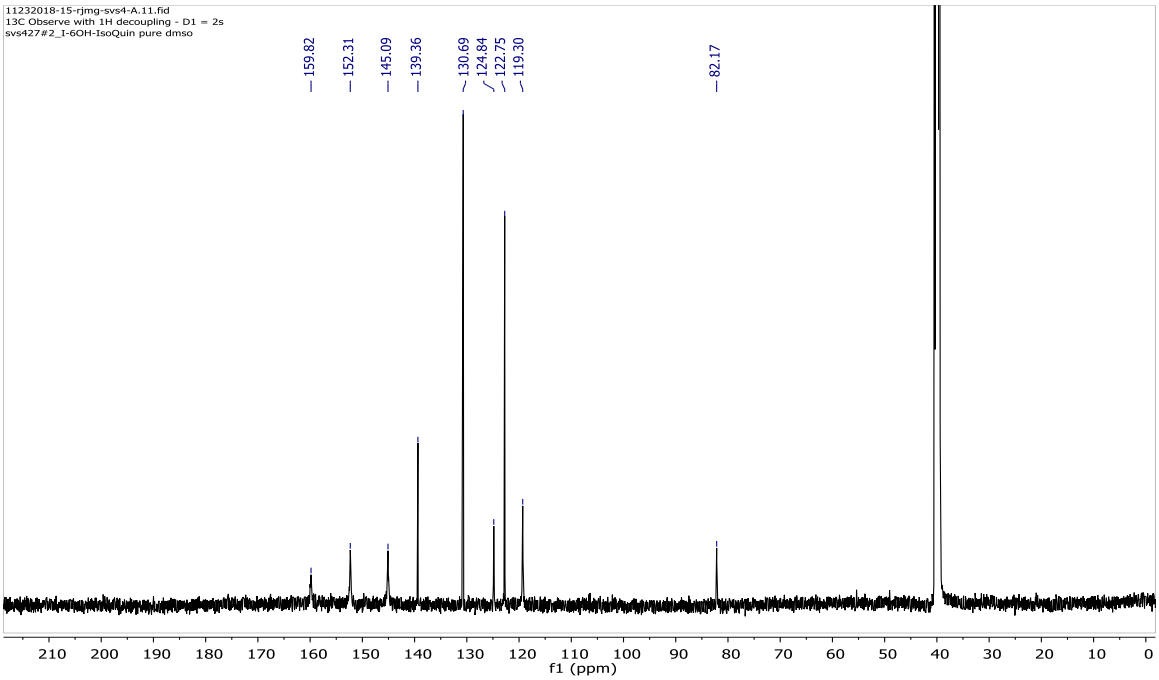
Supplementary Figure 120. ^{1}H NMR (500 MHz, CD_3OD) and ^{13}C NMR (126 MHz, CD_3OD) spectra of 3-iodo pindolol



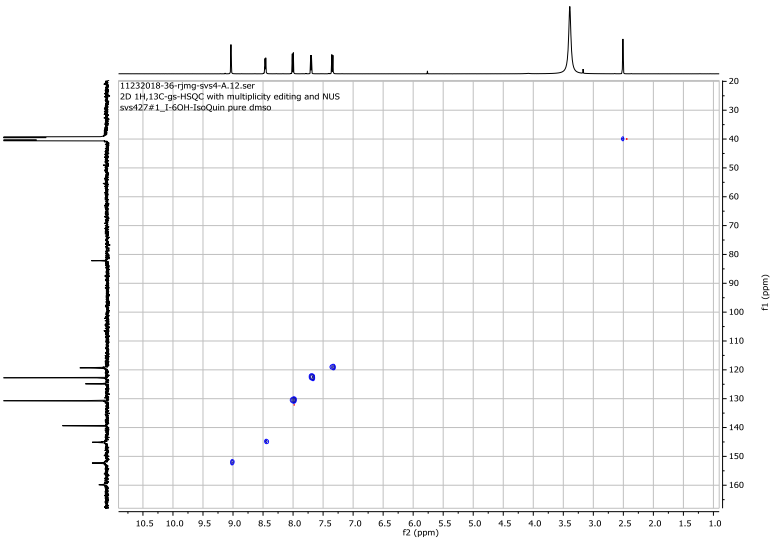
Supplementary Figure 121. 2-D NMR (500 MHz, CD_3OD), COSY, HS QC and HMBC spectra of 3-iodo pindolol

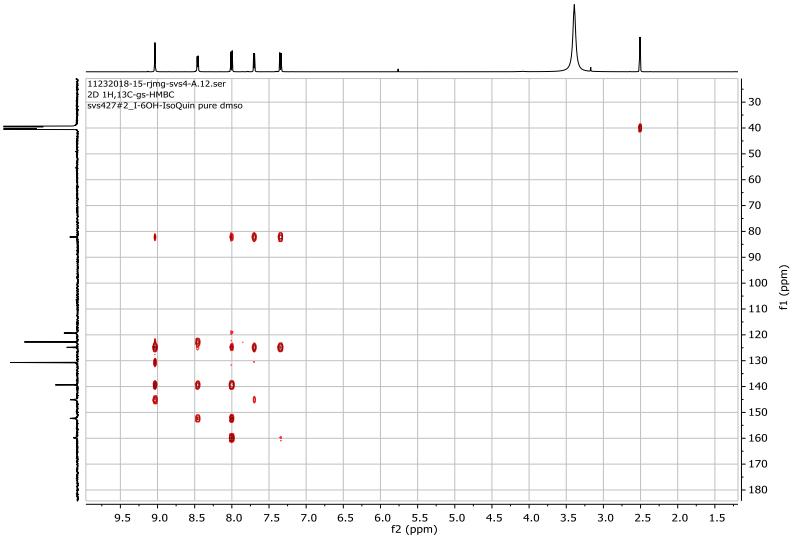
5-Iodo-isoquinolin-6-ol (6)





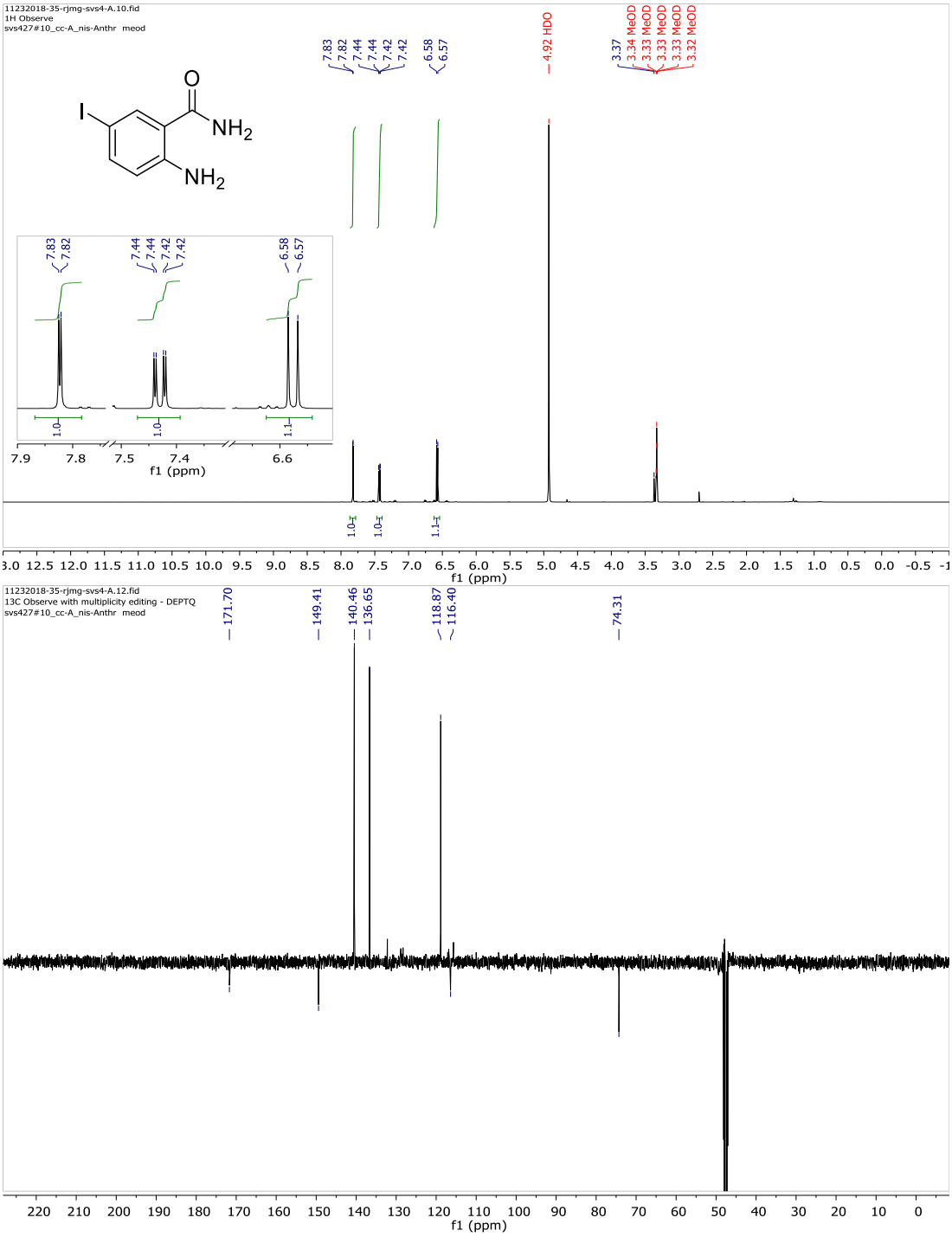
Supplementary Figure 122. ^{1}H NMR (500 MHz, DMSO-d_6) and ^{13}C NMR (126 MHz, DMSO-d_6) spectra of 5-iodo-isoquinolin-6-ol



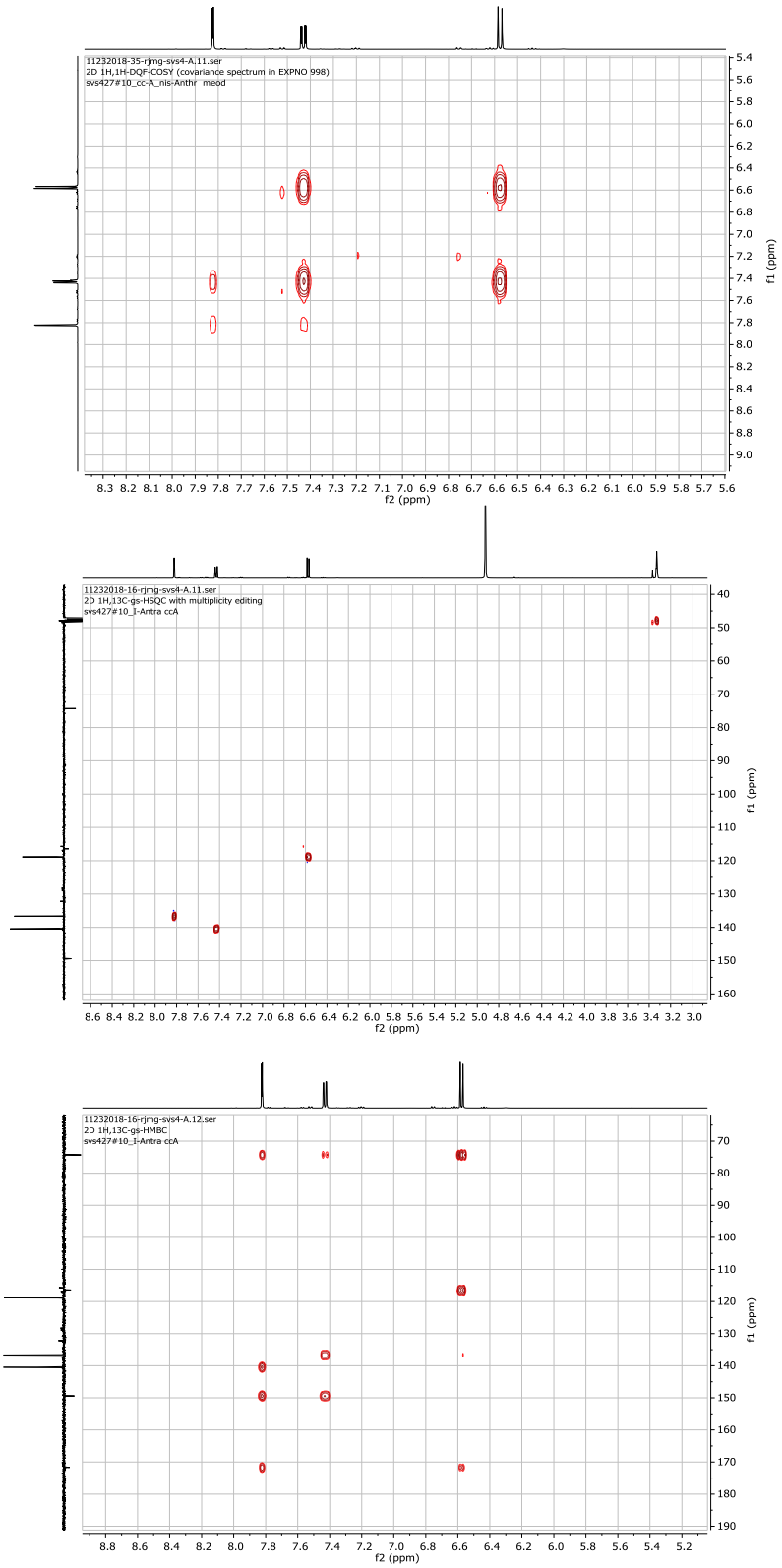


Supplementary Figure 123. 2-D NMR (500 MHz, DMSO-*d*$_6$), HSQC and HMBC spectra of 5-iodo-isoquinolin-6-ol

5-Iodo-anthranilamide (7)

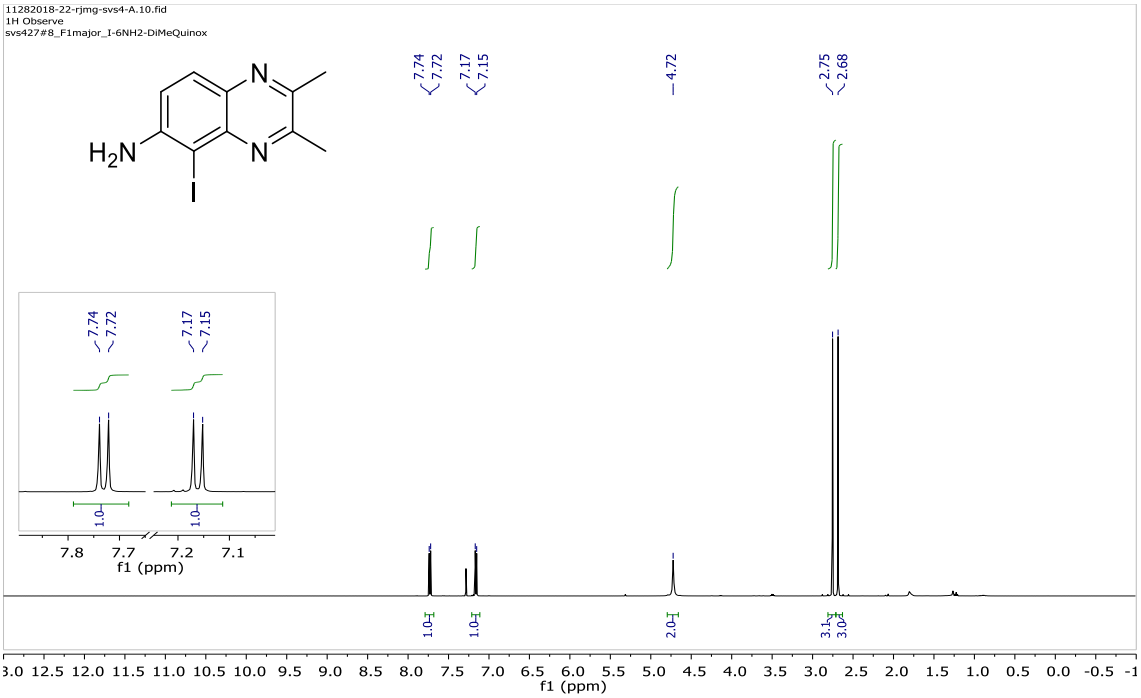


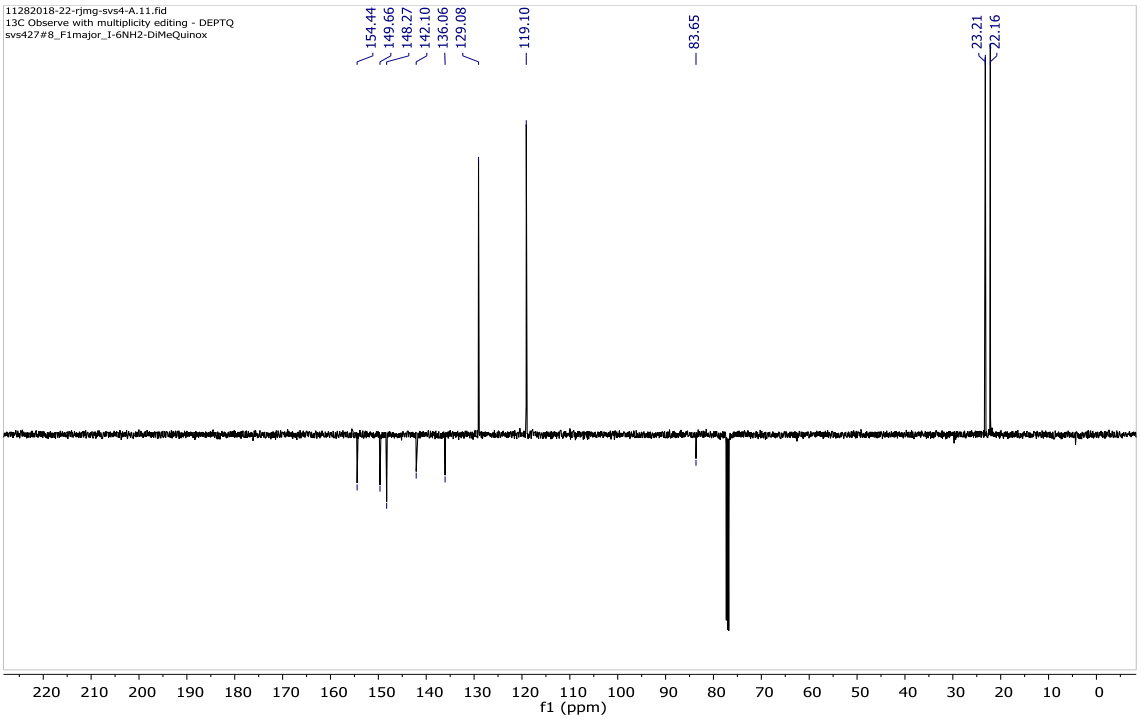
Supplementary Figure 124. ^{1}H NMR (500 MHz, CD_3OD) and ^{13}C NMR (126 MHz, CD_3OD) spectra of 5-iodo-anthranilamide



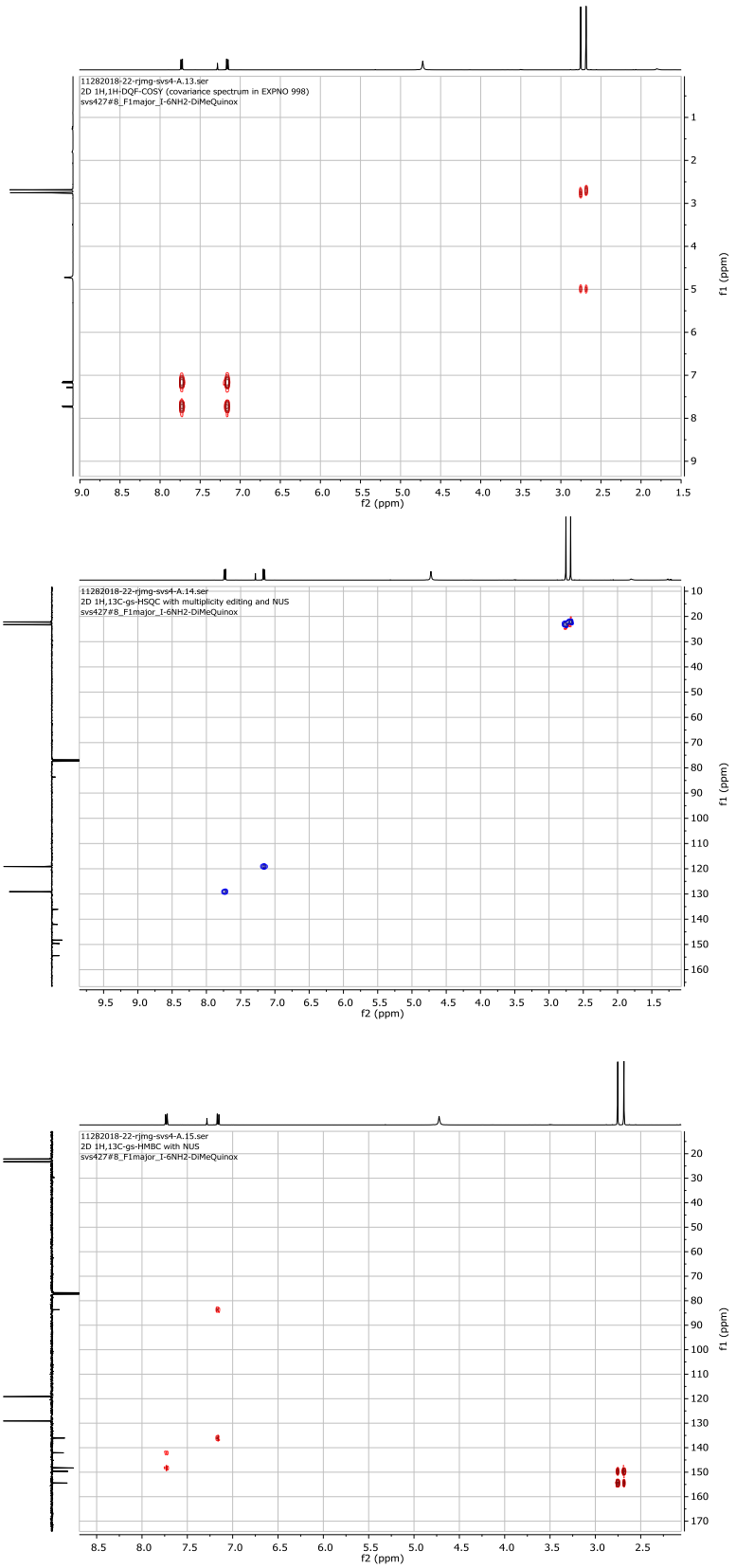
Supplementary Figure 125. 2-D NMR (500 MHz, CD_3OD), COSY, HSQC and HMBC spectra of 5-iodo-anthranilamide

5-Iodo-2,3-dimethylquinoxalin-6-amine (8)



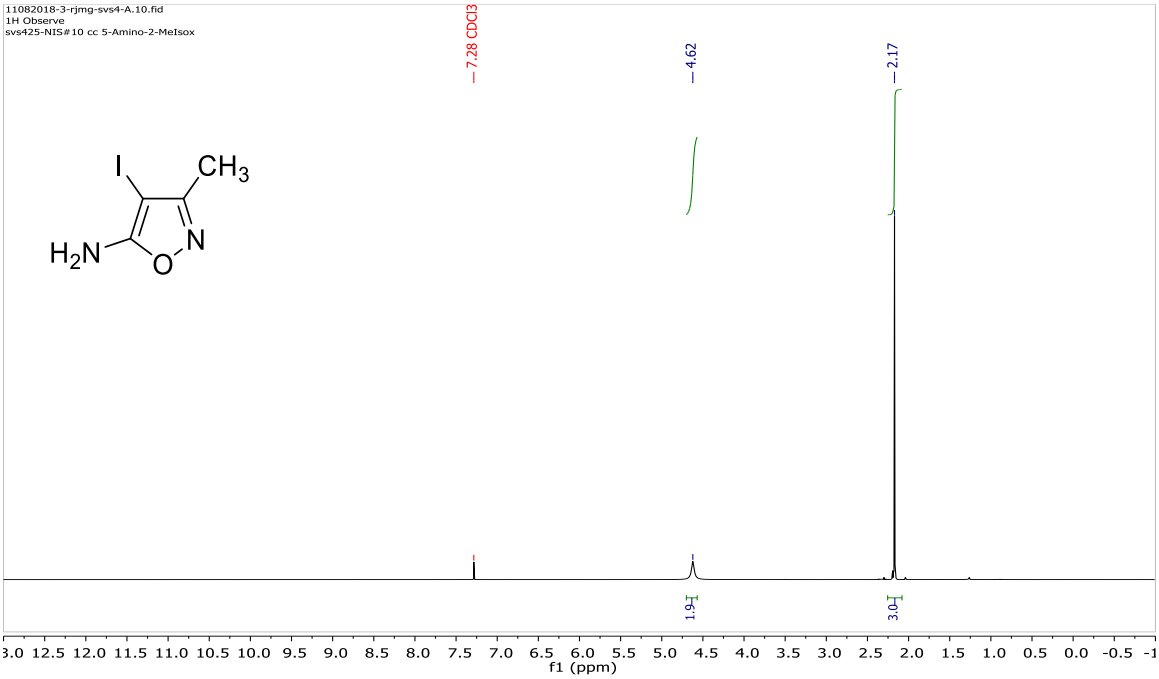


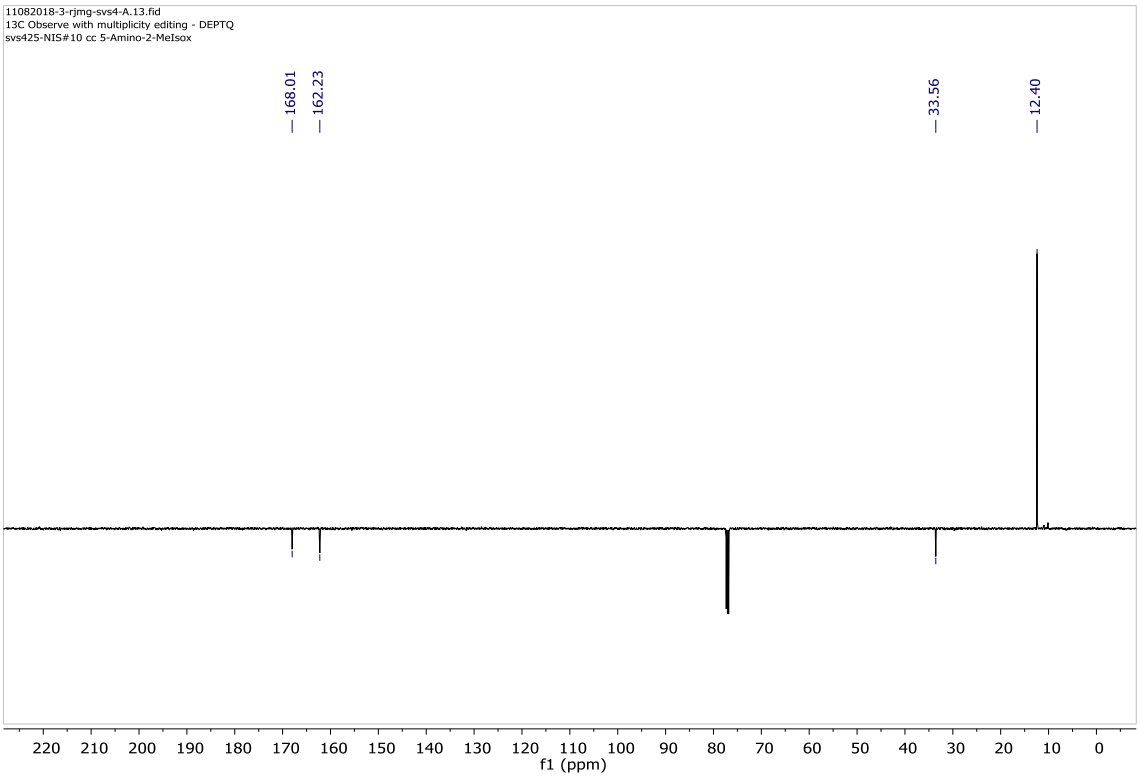
Supplementary Figure 126. ^{1}H NMR (500 MHz, $CDCl_3$) and ^{13}C NMR (126 MHz, CD_3OD) of 5-iodo-2,3-dimethylquinoxalin-6-amine



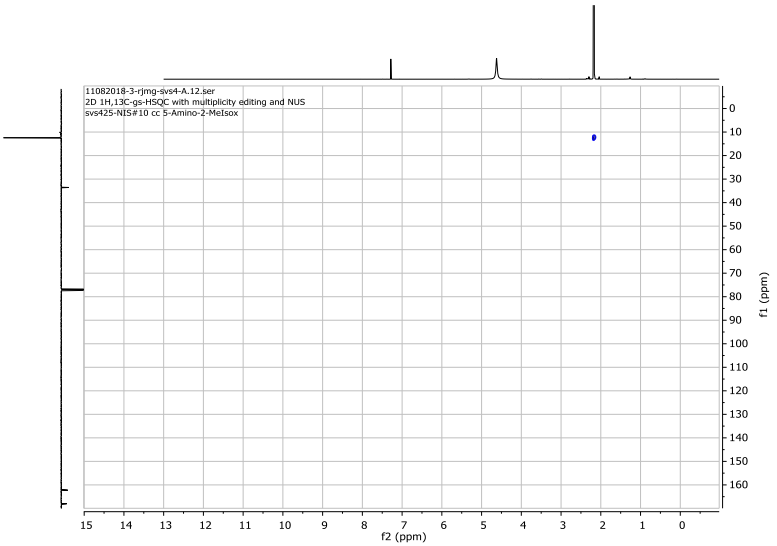
Supplementary Figure 127. 2-D NMR (500 MHz, CD_3OD), COSY, HSQC and HMBC spectra of 5-iodo-2,3-dimethylquinoxalin-6-amine

5-Amino-4-iodo-3-methylisoxazole (9)



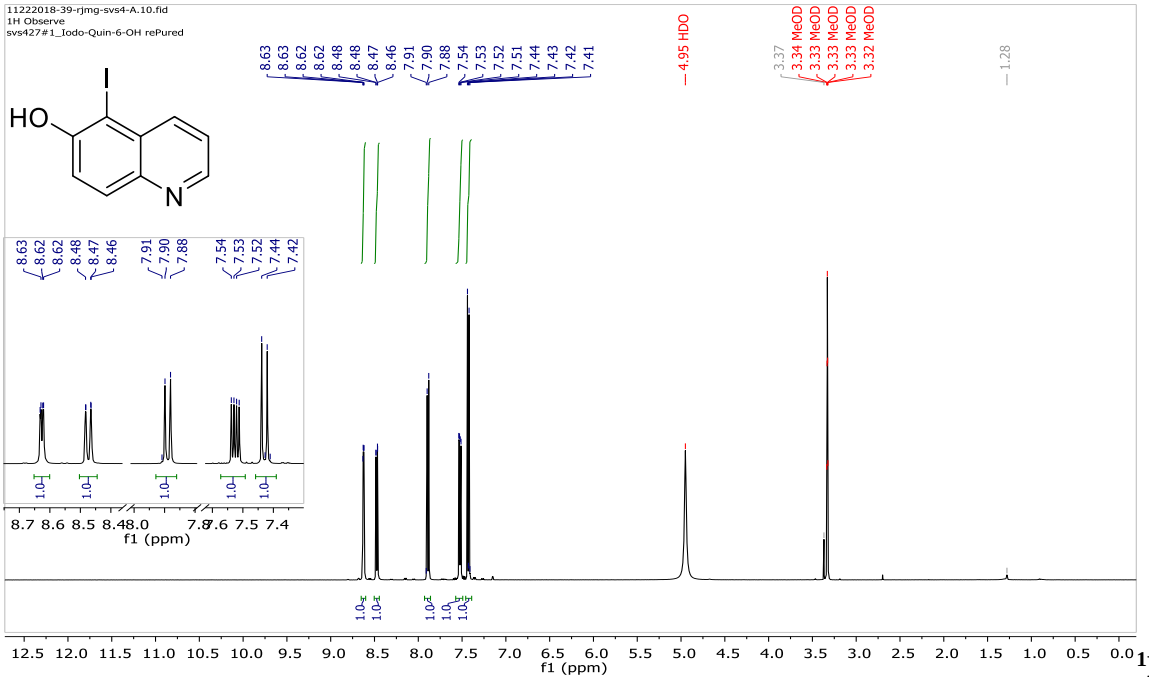


Supplementary Figure 128. ^{1}H NMR (500 MHz, $CDCl_3$) and ^{13}C NMR (126 MHz, $CDCl_3$) of 5-amino-4-iodo-3-methylisoxazole

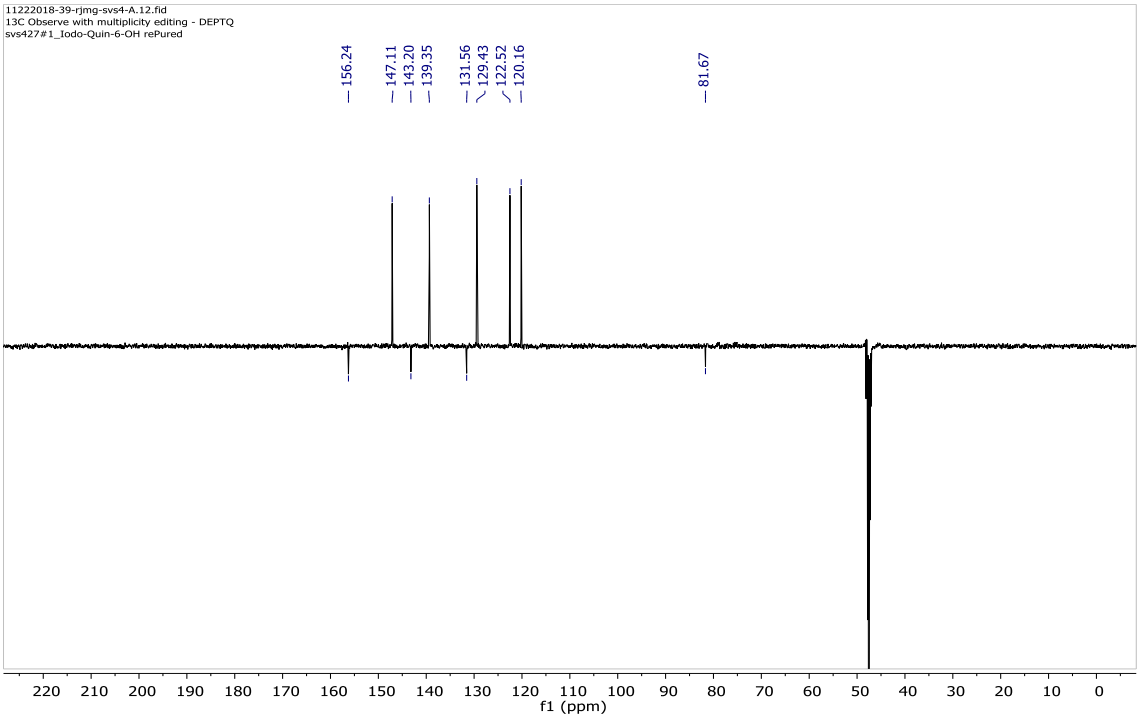


Supplementary Figure 129. HSQC NMR ($CDCl_3$) of 5-amino-4-iodo-3-methylisoxazole

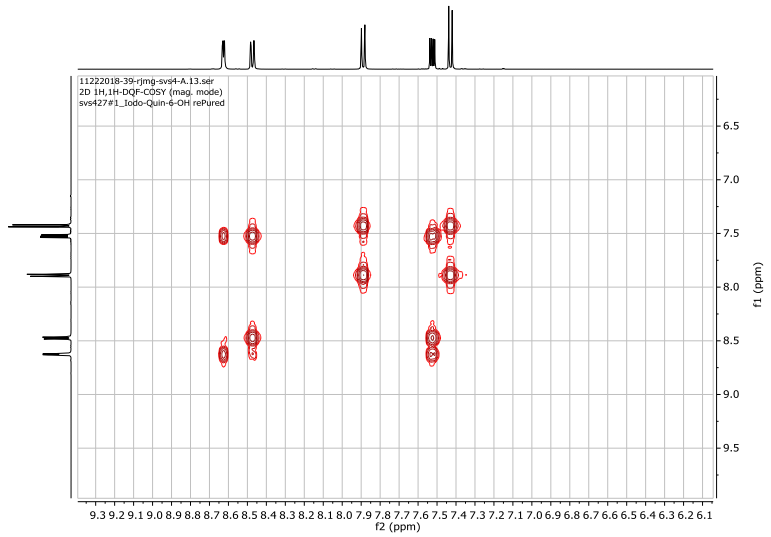
5-Iodoquinolin-6-ol (10)

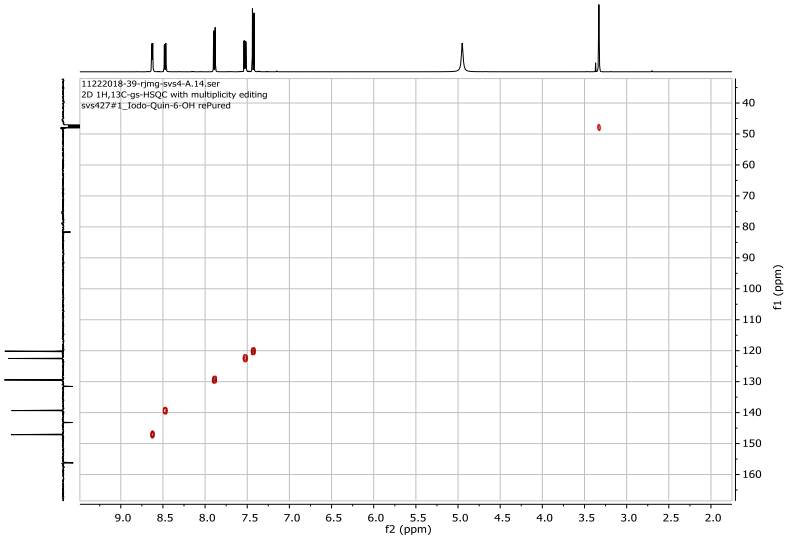


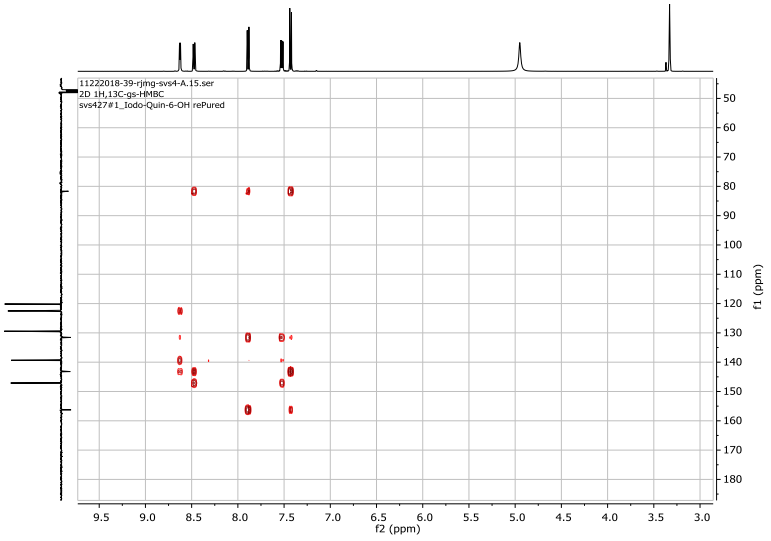
^{1}H



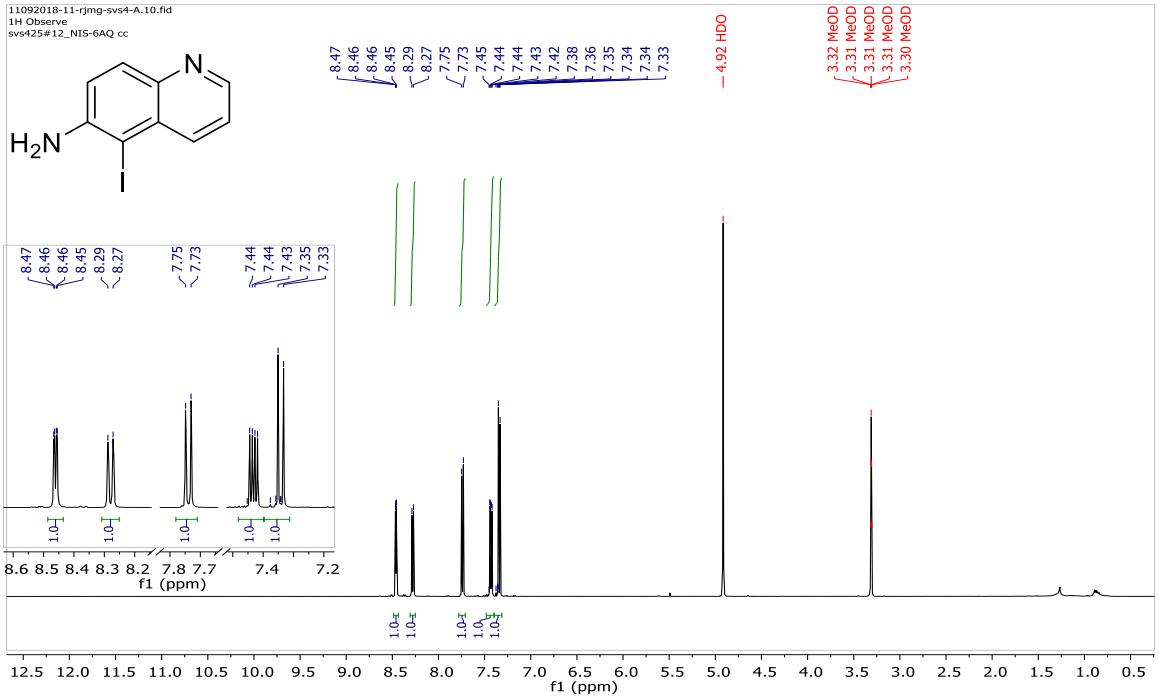
Supplementary Figure 130. ^{1}H NMR (500 MHz, CD_3OD) and ^{13}C NMR (126 MHz, CD_3OD) spectra of 5-iodoquinolin-6-ol

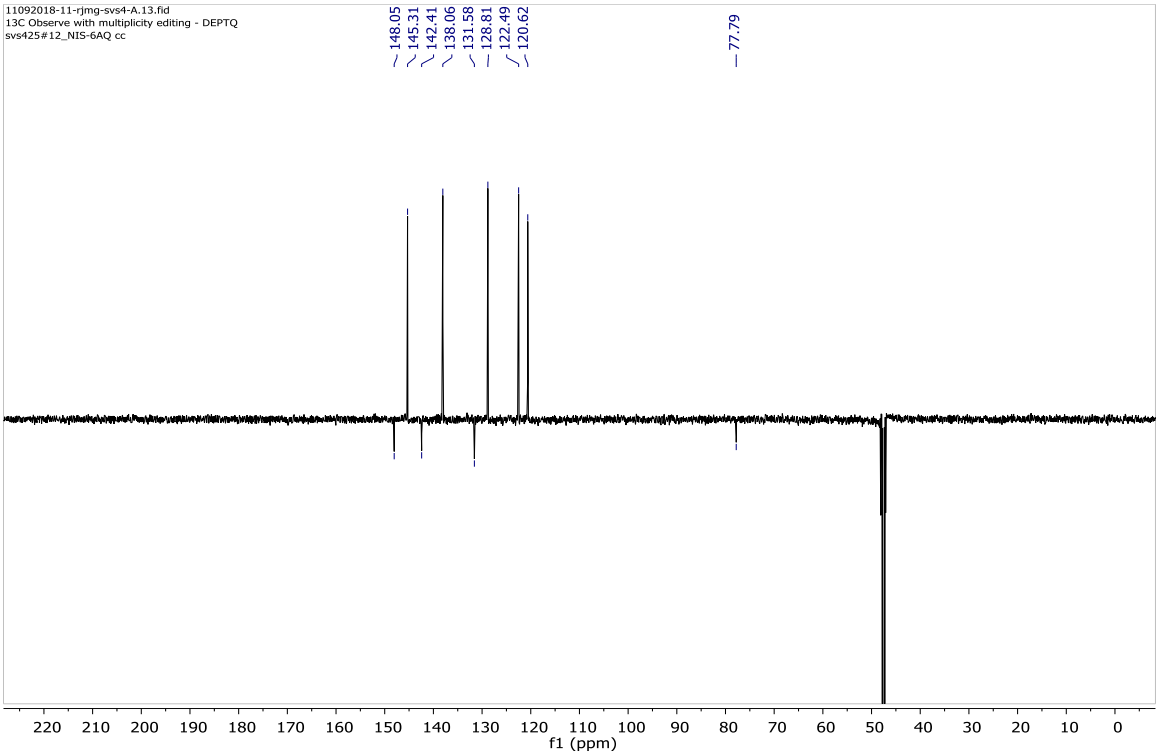

11222018-39-rjmg-svs4-A.13.ser
2D 1H,1H-DQF-COSY (mag. mode)
svs427#1_Iodo-Quin-6-OH rePured
f1 (ppm)
f2 (ppm)


11222018-39-rjmg-svs4-A.14.ser
2D 1H,13C-gs-HSQC with multiplicity editing
svs427#1_Iodo-Quin-6-OH rePured
f1 (ppm)
f2 (ppm)

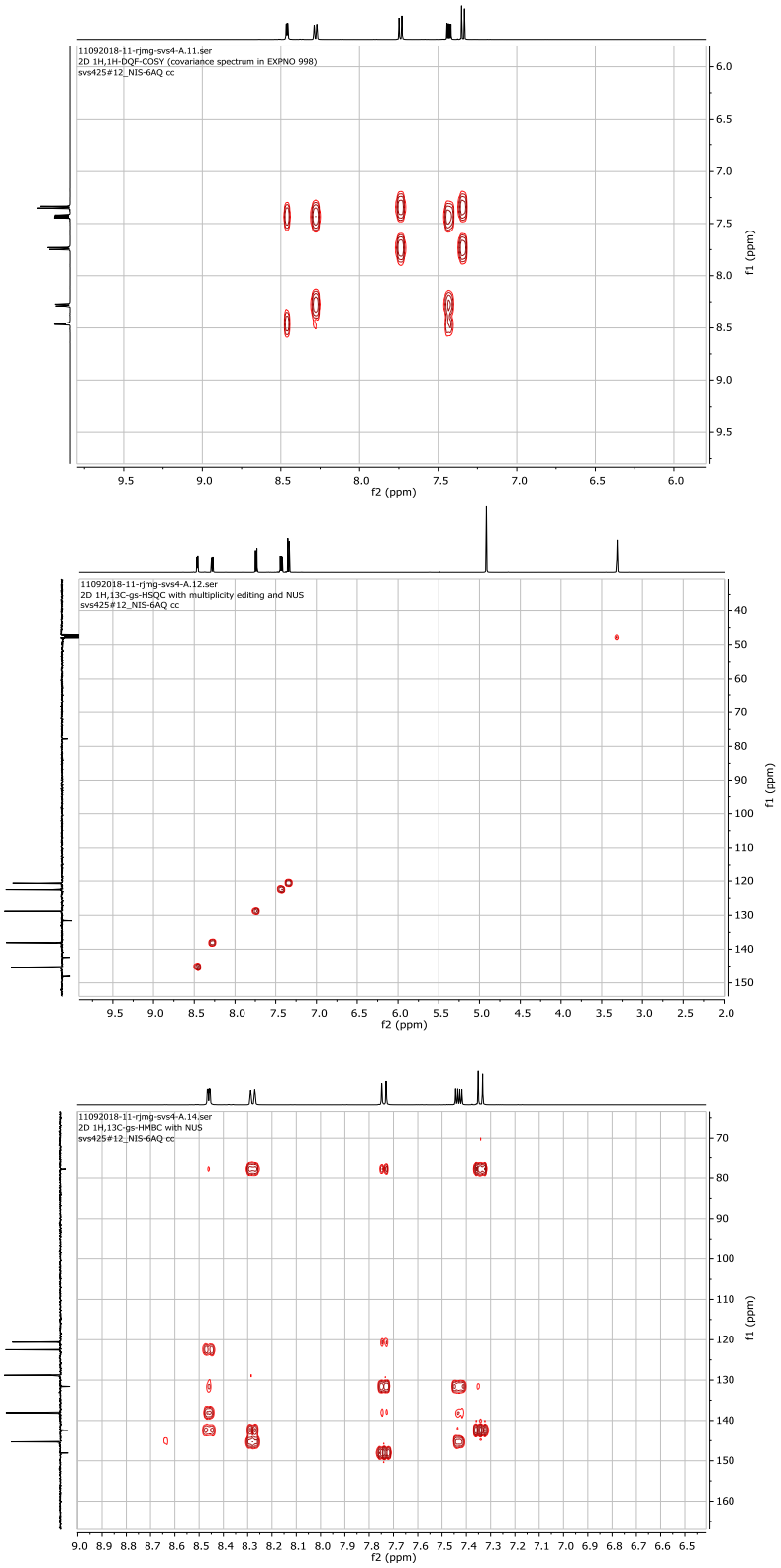

11222018-39-rjmg-svs4-A.15.ser
2D 1H,13C-gs-HMBC
svs427#1_Iodo-Quin-6-OH rePured
f1 (ppm)
f2 (ppm)

Supplementary Figure 131. 2-D NMR (500 MHz, CD_3OD), COSY, HSQC and HMBC spectra of 5-iodoquinolin-6-ol
6-Amino-5-iodoquinoline (11)



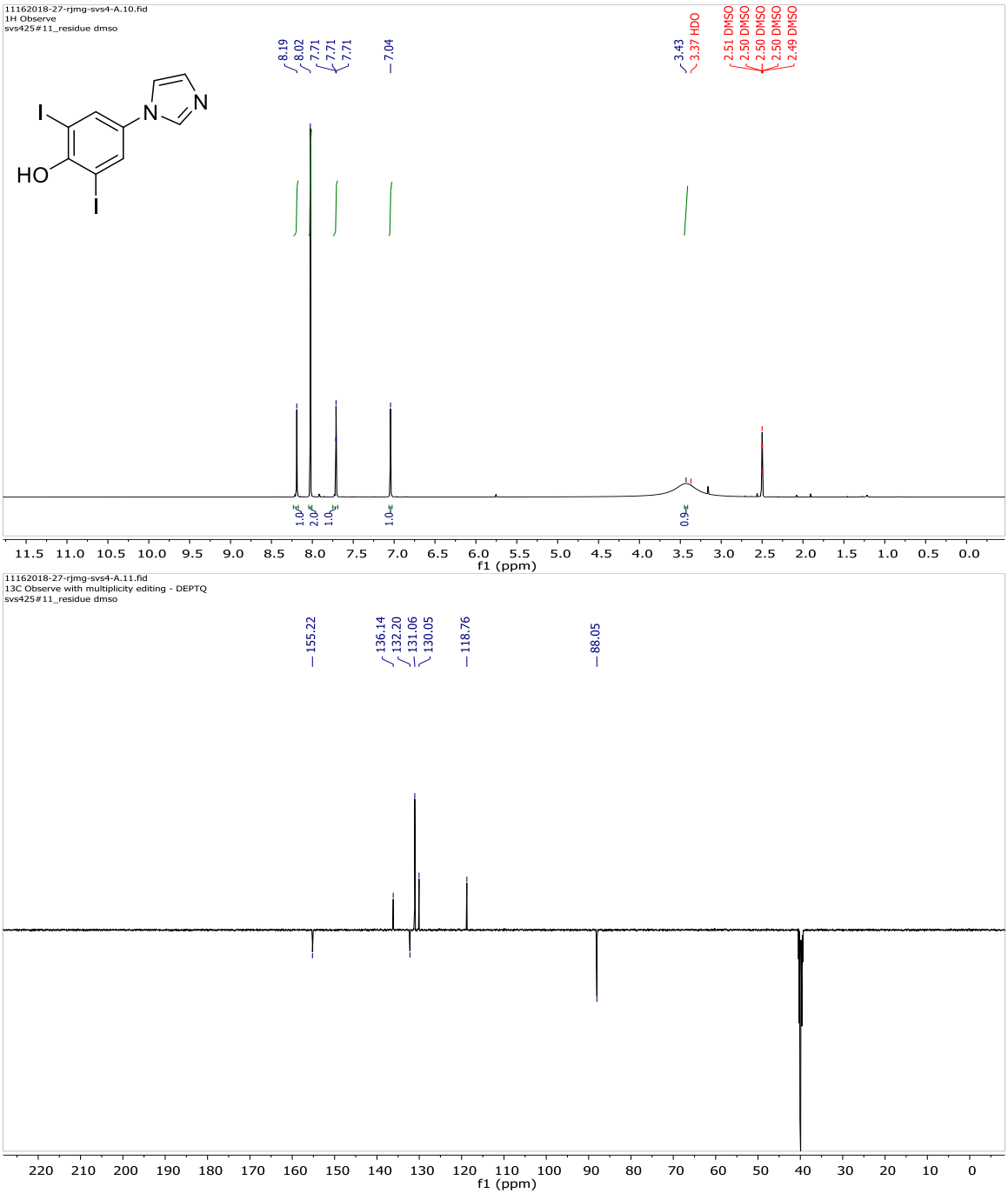


Supplementary Figure 132. ^{1}H NMR (500 MHz, CD_3OD) and ^{13}C NMR (126 MHz, CD_3OD) spectra of 6-amino-5-iodoquinoline

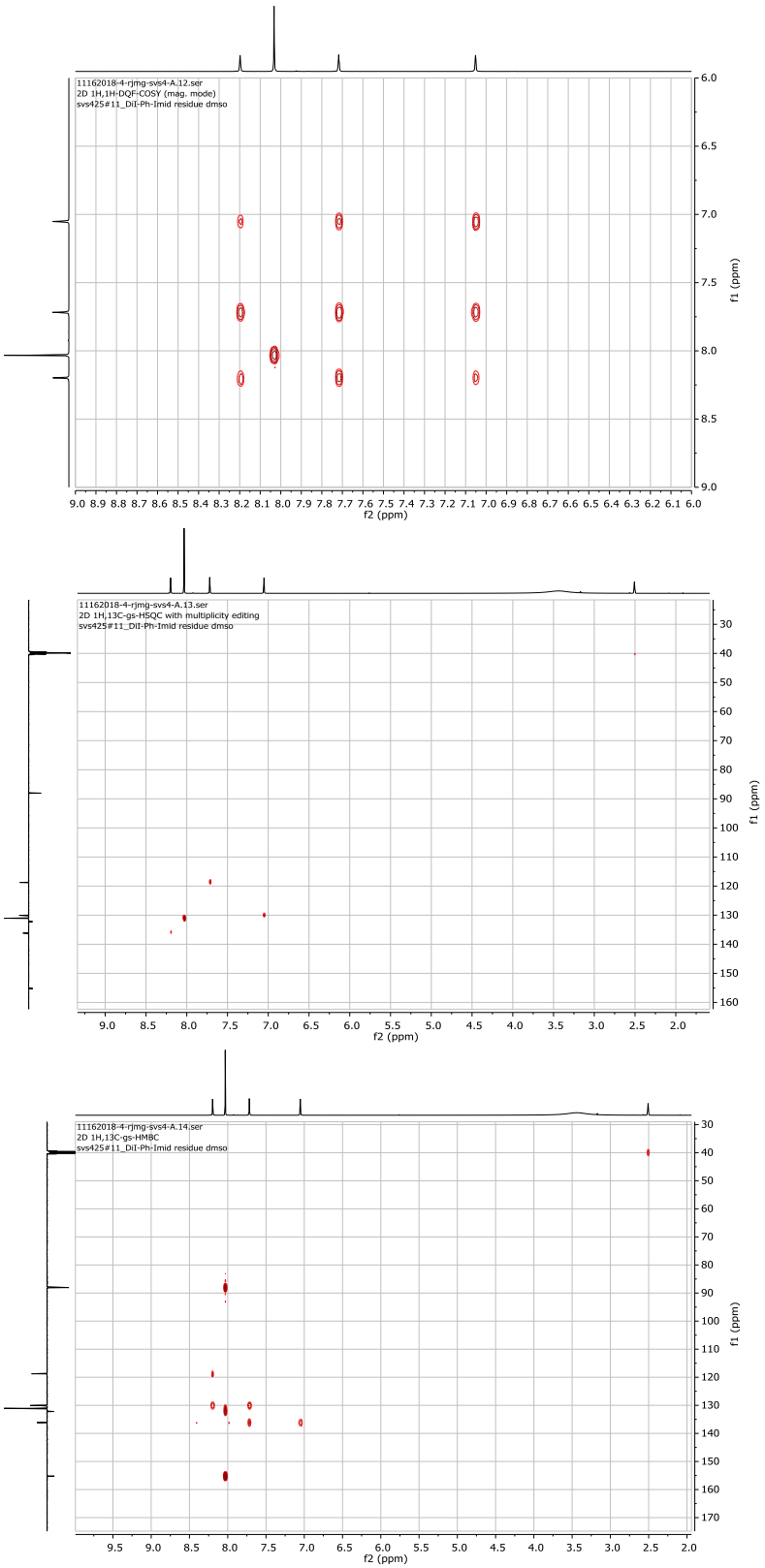


Supplementary Figure 133. 2-D NMR (500 MHz, CD_3OD), COSY, HSQC and HMBC spectra of 6-amino-5-iodoquinoline

4-(1H-imidazol-1-yl)-2,6-diiodophenol (12)

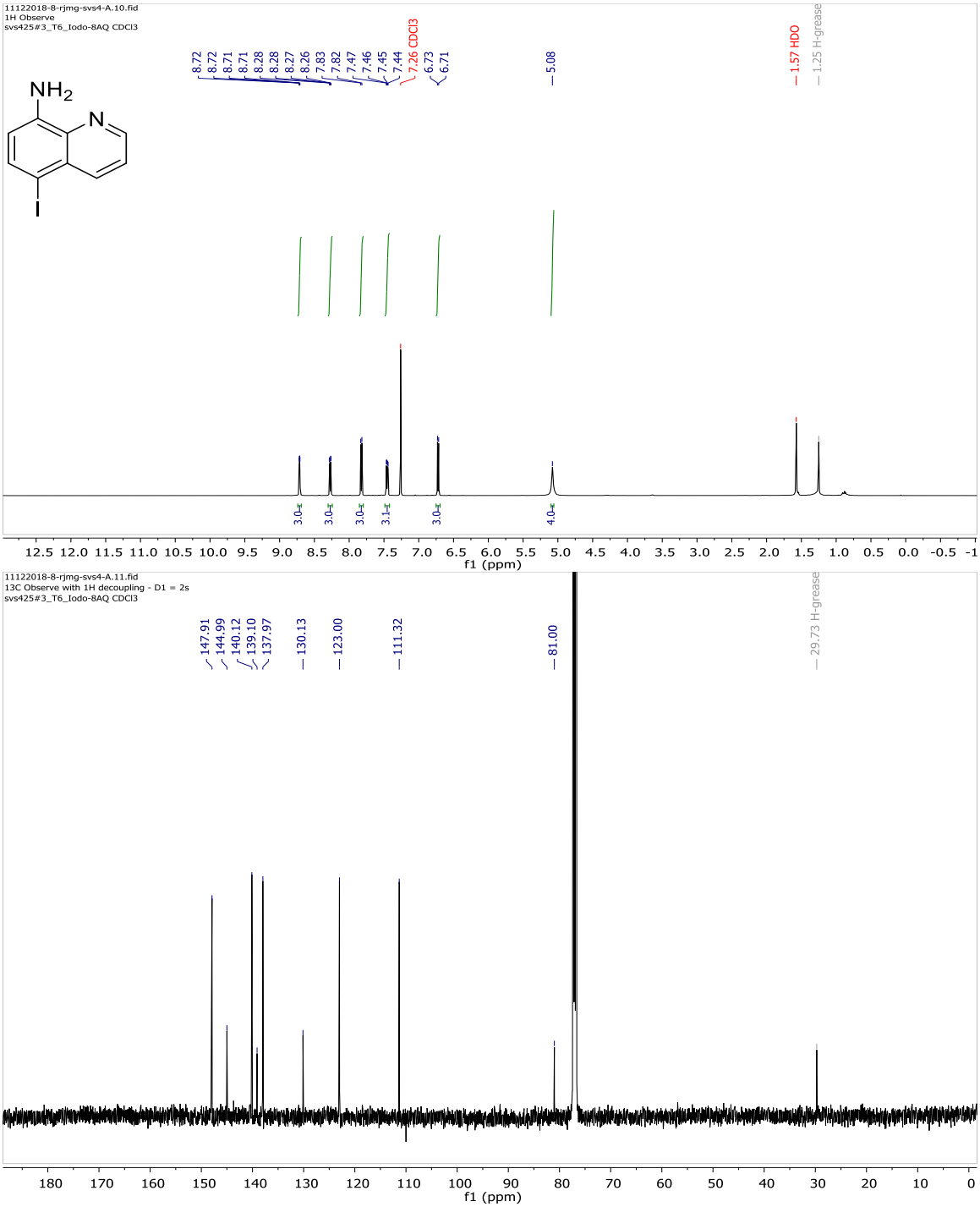


Supplementary Figure 134. ^{1}H NMR (500 MHz, DMSO-*d*$_6$) and ^{13}C NMR (126 MHz, DMSO-*d*$_6$) spectra of 4-(1H-imidazol-1-yl)-2,6-diiodophenol

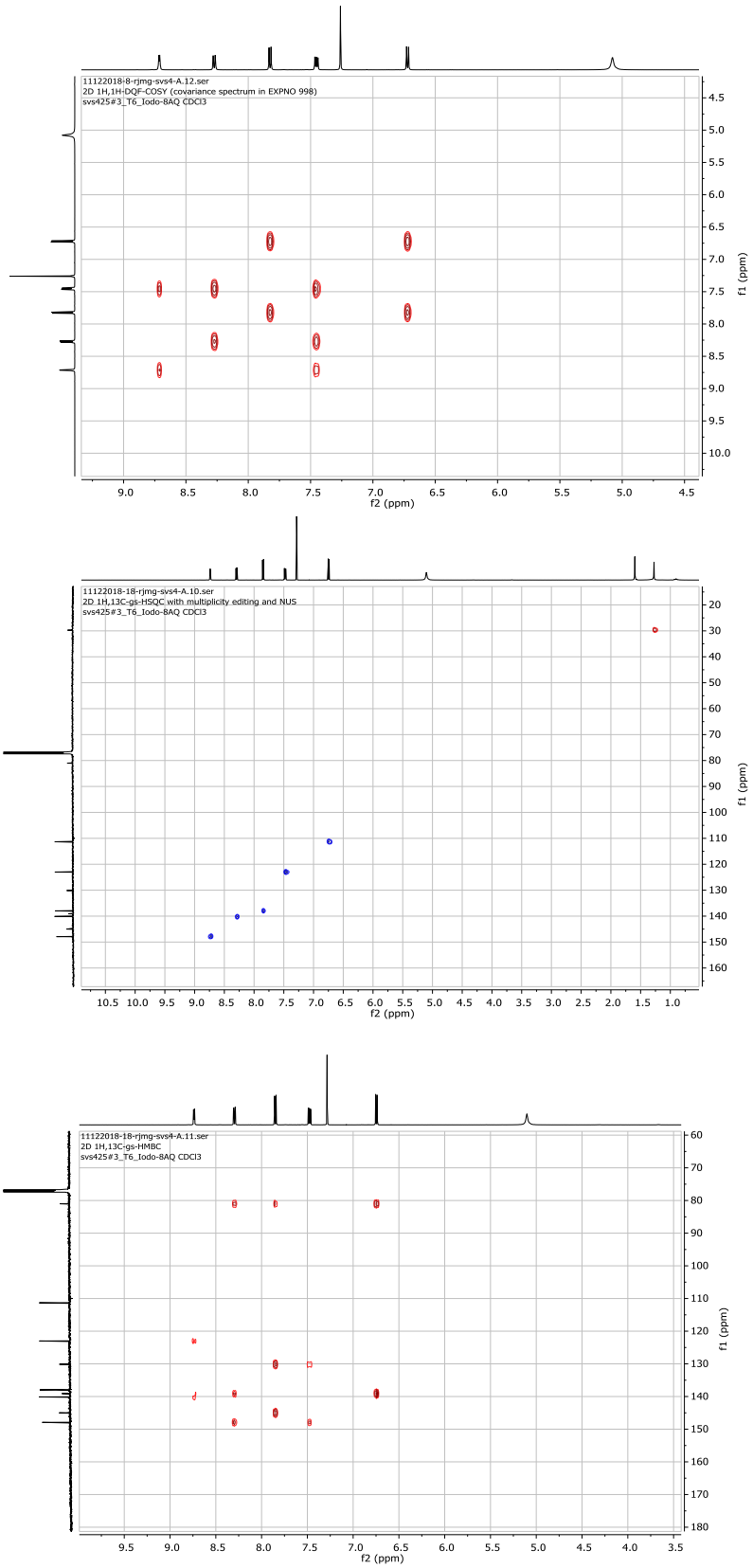


Supplementary Figure 135. 2-D NMR (500 MHz, DMSO-*d*$_6$), COSY, HSQC and HMBC spectra of 4-(1H-imidazol-1-yl)-2,6-diiodophenol

5-Iodo-8-aminoquinoline (13)



Supplementary Figure 136. ^{1}H NMR (500 MHz, $CDCl_3$) and ^{13}C NMR (126 MHz, $CDCl_3$) spectra of 5-iodo-8-aminoquinoline



Supplementary Figure 137. 2-D NMR (500 MHz $CDCl_3$), COSY, HSQC and HMBC spectra 5-iodo-8-aminoquinoline