Supporting information for:

Ligand retargeting by binding site analogy

Lars Wiedmer,[†] Claude Schärer,[‡] Dimitrios Spiliotopoulos,[†] Marianne Hürzeler,[‡] Paweł Śledź,[†] and Amedeo Caflisch[*,†]

Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland, and University of Applied Sciences and Arts Northwestern Switzerland, Hofackerstrasse 30, CH-4132 Muttenz, Switzerland

E-mail: caflisch@bioc.uzh.ch

[*]To whom correspondence should be addressed
[†]University of Zurich
[‡]University of Applied Sciences and Arts Northwestern Switzerland

Contents

1 Computational methods

1.1 Preparation of a library of aspartic protease inhibitors

A library of aspartic protease inhibitors with molecular weight larger than 200 Da and available crystal structure was compiled (in June 2015). The PDB entries of the 342 holo structures of aspartic proteases are listed in Figure S1. The 3D coordinates of the ligands were generated and minimized by open Babel[S2] (version 2.3.2) using the MMFF94 force field[S3–S6]. The library was subsequently docked as mentioned below.

1.2 Protein preparation

The docking of the 342 aspartic protease inhibitors was carried out using the human MTH1 protein in the complex with (R)-crizotinib (PDB code 4C9W). The docking of the nearly 5000 derivatives of fragment **7** was performed on the crystal structures 4C9X (complex with (S)-crizotinib) and 5ANS (complex with 1-[4-amino-2-(ethoxymethyl)-1H-imidazo[4,5-c]quinolin-1-yl]-2-methylpropan-2-ol). Note that the latter was not yet available when we screened the 342 aspartic protease inhibitors. For each structure of MTH1 used for docking, five templates were prepared with all possible protonation states of the Asp-Asp motif except for the unlikely double protonation.

1.3 Docking

Docking was carried out using the software package RDock[S7] which required about 7 seconds per molecule on a desktop computer with eight cores (Intel®Core^{TM} i7-4770 CPU @

3.40GHz). First, the molecular surface of the active site of MTH1 was generated by using a 1.0 Å probe radius within a preselected area defined by the (R)-crizotinib MTH1 inhibitor (PDB code: 4C9W). The molecules were then placed into the docking volume such that 20 poses per molecules were generated.

1.4 Scoring

A geometrical filter was applied before ranking the 6840 poses in MTH1 (20 poses for each of the 342 aspartic protease inhibitors). For each pose in MTH1, the atoms of the compound within a distance of 5 Å from the Asp-Asp motif were identified and compared with the atoms that are within 5 Å from the aspartic dyad in the crystal structure of the protease. Only poses with at least four non-hydrogen atoms within the corresponding pose in the aspartic protease were kept. For the nearly 5000 derivatives of fragment **7** only those poses were kept that had similar position and orientation of the fragment. The remaining poses were minimized in the rigid structure of MTH1 using CHARMM[S8] and a distance-dependent dielectric constant for the electrostatic energy. Parameters for the protein were generated using CHARMM36[S9] while the (candidate) ligands were parametrized by CGenFF[S10,S11]. The minimization consisted of 500 steps of steepest descent and 10,000 steps of conjugate gradient with a tolerance of the energy gradient of 0.01 kcal/(mol Å). Electrostatic desolvation penalties were evaluated in the continuum approximation using the finite-difference Poisson model[S12]. The CHARMM[S8] module PBEQ[S13] was used for the finite-difference Poisson calculations with dielectric constants of 4.0 and 78.5 for the solute and solvent, respectively. The total binding energy is the sum of intermolecular van der Waals and electrostatics with solvation. The final ranking was based on the total binding energy divided by the number of non-hydrogen atoms of the ligand.

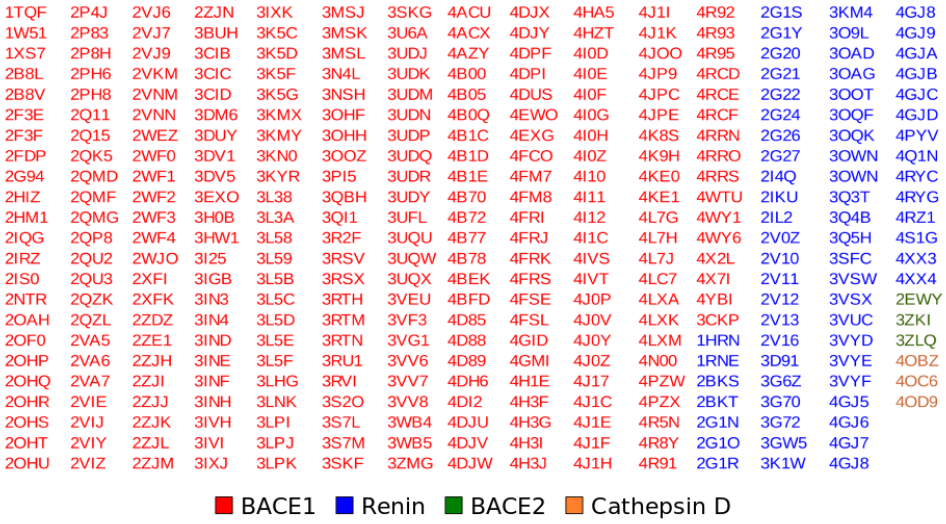


Figure S1: PDB codes of the 342 aspartic protease-ligand complexes used in this study.

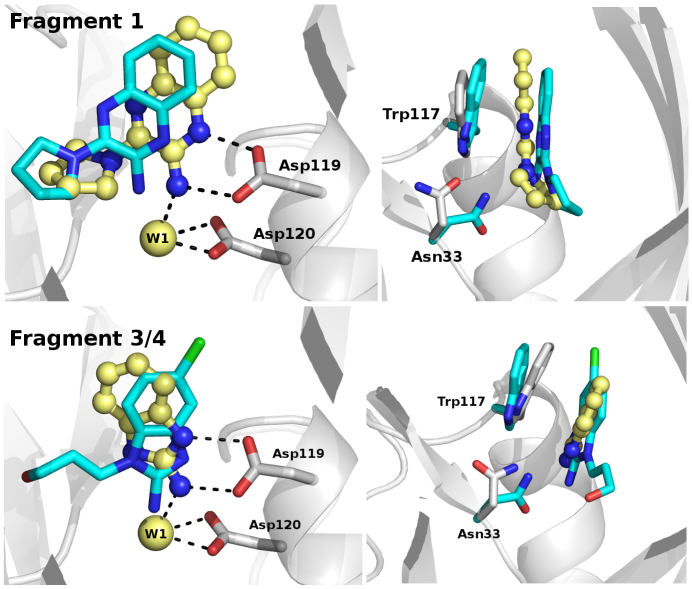
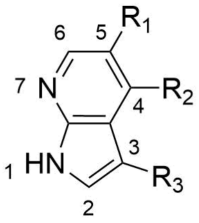
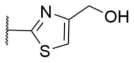
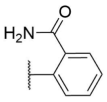
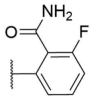
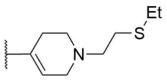
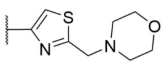
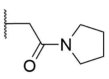


Figure S2: **Comparison of the pose obtained by rigid protein docking (carbon atoms in cyan) and the binding mode in the crystal structure (yellow).** (Top) Two orientations of the complex with fragment **1**. (Bottom) We were not able to solve the structure of MTH1 in the complex with fragment **3** so that the crystal structure with fragment **4** is used as reference. In all panels, the relevant side chains of MTH1 are shown for the crystal structure (gray sticks, PDB codes 6EQ6 and 6EQ5) and the structure used for docking (cyan, PDB code: 4C9W). Hydrogen bonds are shown only for the crystal structures (dashed lines).

Table S1: **Docking ranks and in vitro potency of the derivatives of the 7-azaindole 7.** Compounds **8-10** and **S1-S6** were identified by the docking campaign performed with analogues of fragment **7**. The single-dose measurement is the remaining MTH1 activity in the presence of 50 μM inhibitor with respect to DMSO. Thus, lower percentages indicate higher inhibition. The amount of conversion of dGTP to dGMP was measured by the colorimetric assay.



ID	R_1	R_2	R_3	Rank in 4C9X	Rank in 5ANS	single-dose at 50 μM (%)
8				24	46	1
9				61	5	0.4
10				26	54	13
S1				43	63	35
S2				51	77	51
S3				25	2326	61
S4				22	1614	33
S5	Br			48	799	89
S6				40	756	33

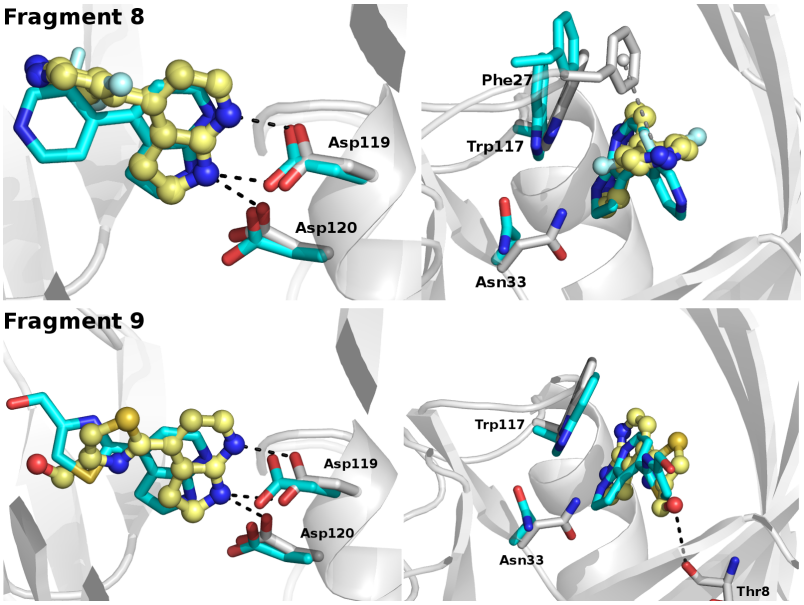


Figure S3: **Comparison of the pose obtained by rigid protein docking (carbon atoms in cyan) and the binding mode in the crystal structure (yellow).** Two orientations are shown for the complex with fragment **8** (top) and **9** (bottom). In all panels, the relevant side chains of MTH1 are shown for the crystal structure (gray sticks, PDB codes 6EQ4 and 6EQ3) and the structure used for docking (cyan, PDB code: 4C9X (top) and 5ANS (bottom)). Hydrogen bonds are shown only for the crystal structures (dashed lines).

1.5 Substructure search

Substructure search was carried out by RDkit (www.rdkit.org). The tolerated elongation sites are directly integrated in the substructure search (Figure S4).

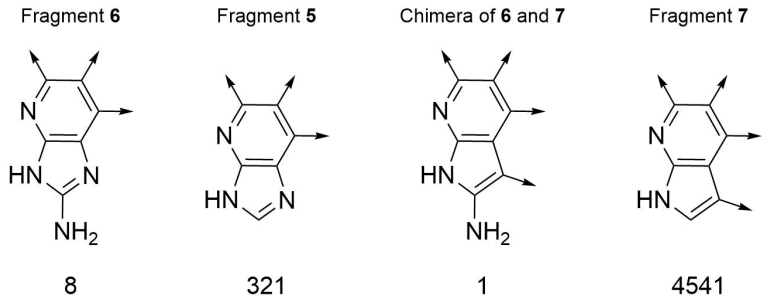


Figure S4: **Substructure search results.** Substructure searches were carried out within the ZINC database[S14] (all purchasable, 2015). The number of hits retrieved are listed in the bottom. The arrows indicate the tolerated elongation site which were directly integrated in the search.

1.6 Molecular dynamics simulations

The MD simulations were carried out with GROMACS[S15] using CHARMM PARAM36[S9] force field for the protein, CGenFF[S10,S11] for fragment **6** and TIP3P as water model. The ligand-protein complex was predicted by molecular docking and minimized as outlined in the docking section. The unit cell was filled with water and counter ions to neutralize the total charge by adding six Na^+ ions. Periodic boundary conditions were used in conjunction with Ewald summation for the long-range electrostatic interactions. The simulations were run at 300K and 1 atm by using the velocity rescaling thermostat and Parrinello Rahman barostat, respectively. The system was first minimized and equilibrated by constraining the heavy atoms of the ligand and the protein with a force constant of 1000 KJ/(mol nm). A total of five production runs of 200 ns each were carried out with different initial velocities. The calculations were performed on the Supercomputer Piz Daint at the Swiss national supercomputing center.

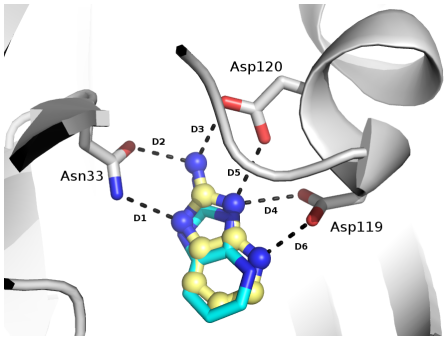


Figure S5: Comparison between docked pose (cyan) and crystal pose (yellow) of fragment **6**. The protein and the interacting residues are colored in gray.

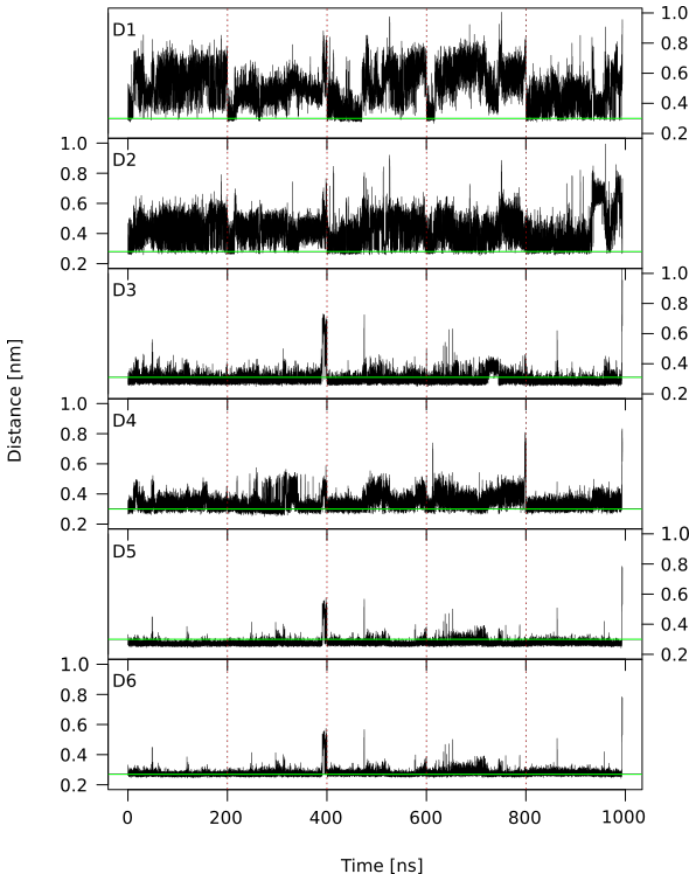


Figure S6: **Time series of hydrogen bond distances for fragment 6.** The distances in the crystal structure are shown for comparison (horizontal green lines). Distances D1-6 are shown in Figure S5. Five independent runs of 200 ns each are concatenated (vertical dashed lines).

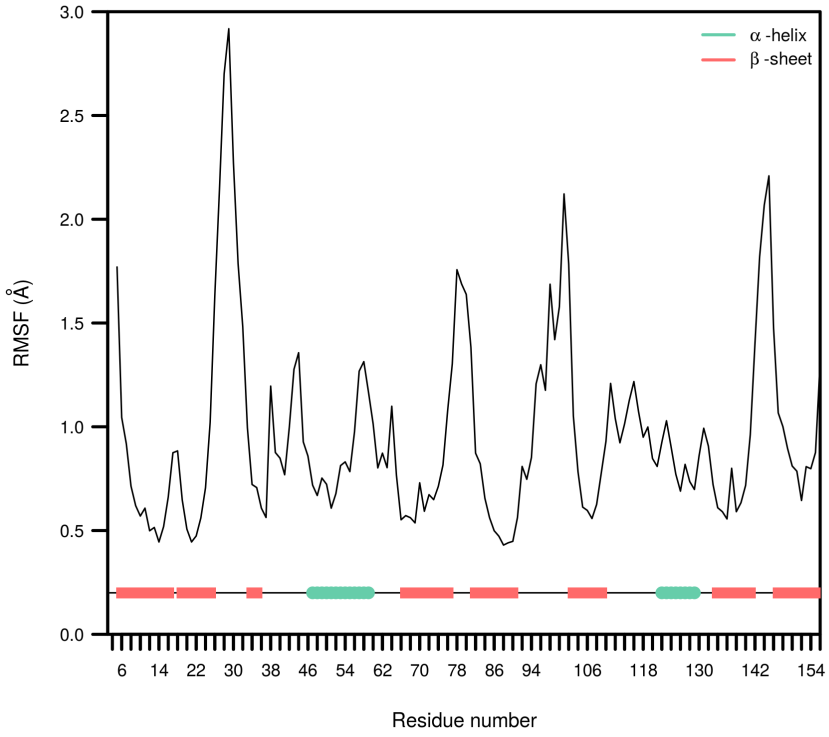


Figure S7: **RMSF analysis of MTH1 in complex with fragment 6.** The RMSF was calculated from the Cα atoms along all five MD simulation runs.

1.7 Aggregator advisor

Fragments **1** to **11** resulted negative at the Aggregator Advisor Web server[S16] except for fragments **3** (see main text) and **5** which is similar to a compound that has previously been reported as an aggregator[S17] (Figure S8).

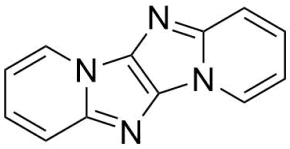


Figure S8: **Pyrido[2",1":2',3']imidazo[4',5':4,5]imidazo[1,2-a]pyridine.** The compound has been reported as aggregator[S17] and shows 74 % similarity to fragment **5**.

2 Compound purity

Compounds **1** and **3-10** were purchased from commercial vendors and the purity of all molecules was analyzed by HPLC-MS and is determined to be at least 95%.

3 Experimental methods

3.1 Expression and purification

The plasmid with MTH1 was kindly provided by Prof. Thomas Helleday (Karolinska Institutet) and, expressed and purified as previously reported[S18]. Unfortunately, we did not succeed to retrieve any crystals from this clone but used it in all enzymatic and binding assays. Another plasmid of MTH1 with an N-terminal His_6 tag and TEV (Tobacco Etch Virus) protease cleavage site, kindly provided by Nicola Burgess-Brown (Addgene plasmid # 74660), was used to retrieve all the crystals. MTH1 proteins from this plasmid were overexpressed in *Escherichia coli* BL21(DE3) cell upon induction with isopropyl thio-beta-D-galactoside (IPTG) for 20 h at 18 °C. The purification was then conducted as reported elsewhere[S19]. The protein was finally concentrated to 38 mg/ml, flash frozen and stored at -80 °C.

3.2 Protein crystallization

MTH1 was crystallized by vapor diffusion in hanging drops at 20 °C. Crystallization buffer contained 23-27 % (w/v) PEG3350, 200 mM lithium sulphate and 100 mM Na-acetate pH 4.5. The crystals were further improved by streak seeding. The crystals were smashed and seeded in a drop with the same condition as above but with 10 mg/ml of protein. Protein crystals were transferred to a solution containing 27% (w/v) PEG3350, 200 mM lithium sulphate, 100 mM sodium acetate pH 4.5, 20% (v/v) DMSO and 10 mM compounds. After overnight incubation, crystals were frozen in liquid nitrogen.

3.3 Data collection and structure solution

Diffraction data were collected at Swiss light source, Paul Scherrer Institue (Villigen, Switzerland) on the beamlines PXI and PXIII. Data were processed by XDS[S20] and structure was solved by molecular replacement with Phaser[S21] using coordinates of apo MTH1, previously solved in-house. Phenix[S22] iteratively refined the models which were build manually using COOT[S23].

Table S2: **Crystallization statistics.**

	Fragment **1**	Fragment **4**	Fragment **6**	Fragment **8**	Fragment **9**	Fragment **11**
	6EQ6	6EQ5	6EQ2	6EQ4	6EQ3	6EQ7
Space group	P 21 21 2	P 21 21 2	P 21 21 2	P 21 21 2	P 21 21 2	P 2 21 21
Unit cell						
a(Å)	60.47	60.80	60.41	60.67	60.75	36.02
b(Å)	66.95	66.47	66.31	65.94	65.97	60.59
c(Å)	36.16	36.24	36.14	35.89	36.20	66.34
alpha	90.00	90.00	90.00	90.00	90.00	90.00
beta	90.00	90.00	90.00	90.00	90.00	90.00
gamma	90.00	90.00	90.00	90.00	90.00	90.00
Resolution range (Å)	44.8-2.0	44.9-1.8	44.7-1.8	35.9-1.4	44.7-1.8	44.7-1.5
Unique reflections	17674(2798)	24073(4073)	24898(3917)	54076(8674)	24664(4036)	44725(7090)
$<I/\sigma(I)>$	12.6(4.6)	9.32(2.64)	21.26(10.53)	18.96(10.99)	23.41(9.71)	30.79(9.53)
R meas (%)	5.4(24.9)	7.3(46.7)	3.5(8.0)	5.9 (9.5)	3.6(11.3)	2.8(11.9)
Completeness (%)	96.9(91.3)	91.5(95.6)	96.0(94.2)	98.4(97.6)	94.2(95.2)	99.4(98.0)
Refinement						
Resolution range (Å)	44.876-2.002	44.864-1.801	44.656-1.802	35.887-1.399	44.689-1.798	44.738-1.499
R factor/R free	0.1979/0.2547	0.209/0.271	0.170/0.209	0.166/0.187	0.180/0.223	0.175/0.204
Mean B factors (A2)	30.0	30.0	23.0	16.0	17.0	18.0
RMS bonds (Å)	0.006	0.006	0.007	0.009	0.004	0.006
RMS angels(°)	0.778	0.772	0.919	1.032	0.644	0.874

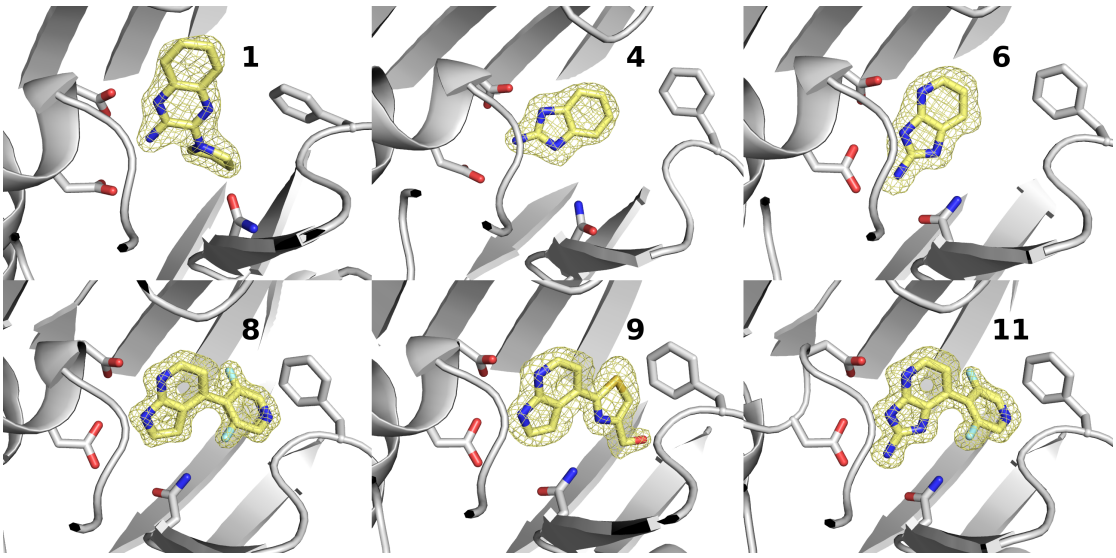


Figure S9: **Density map of crystallization poses.** The indices for the fragments are given next to the poses. The 2Fo-Fc maps are contoured at 1σ.

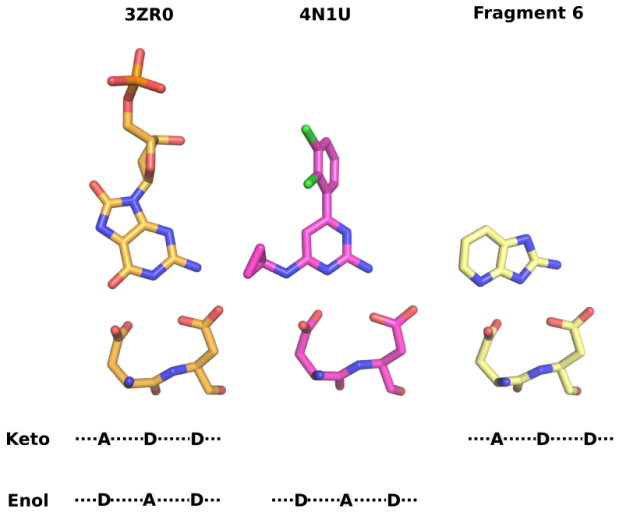
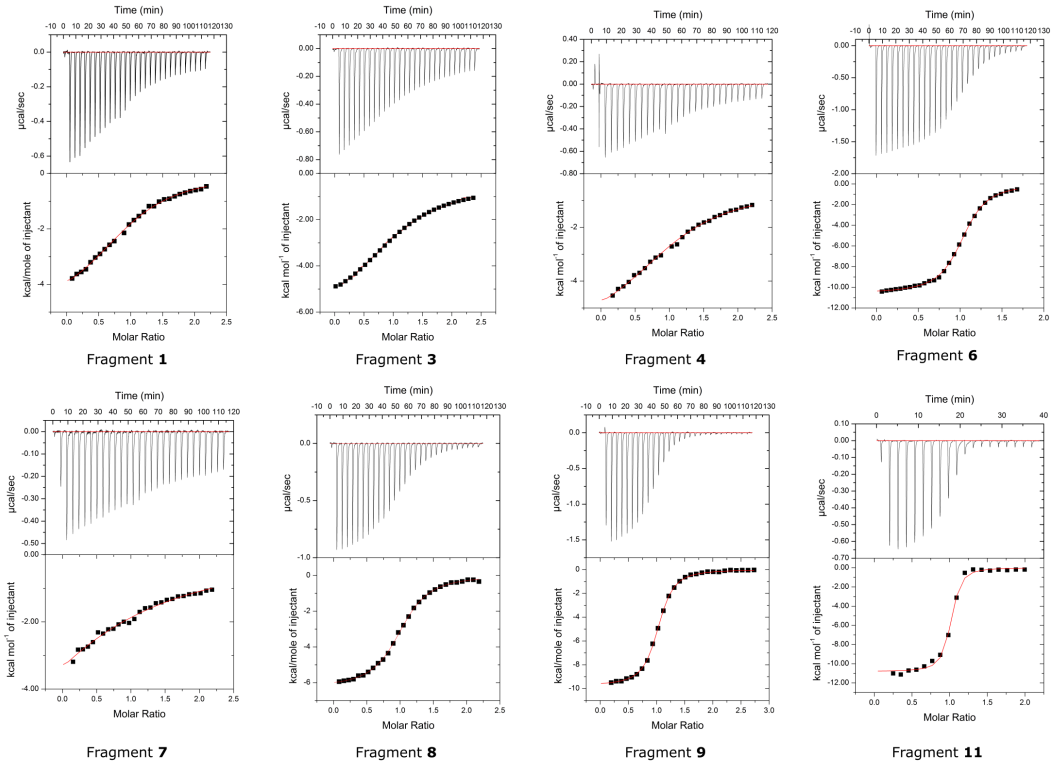


Figure S10: **Interaction pattern to the Asp-Asp motif.** From left to right is shown the product 8-oxo-dGMP (PDB:3ZR0), TH588 (PDB:4N1U) and fragment **6**. The pattern of hydrogen bonds to the Asp-Asp motif is annotated by 'D' (Donor) and 'A' (Acceptor). Inhibitors can mimic either the 6,8-diketo (Keto) or 6-enol-8-keto (Enol) form of 8-oxo-dGMP.

3.4 Isothermal titration calorimetry

Isothermal Titration Calorimetry (ITC) experiments were performed on a VP-ITC (MicroCal) or iTC_{200} instrument (GE Healthcare Life Sciences). The non-cleaved His-tagged protein was thoroughly dialyzed in the ITC buffer (100 mM Tris-acetate pH 7.5, 40 mM NaCl, 10 mM Mg-acetate, 0.0005% Tween-20 and 1 mM mercaptoethanol). Compound dissolved in ITC buffer with 1 % (v/v) DMSO was injected into the sample cell containing the protein with equal DMSO. The measurement was carried out at 20 °C while stirring at 300(VP-ITC)/800(iTC_{200}) rpm: after a control injection of 2/0.4 μl, 28/18 injections of 10/2 μl each (10/4 s duration, with a 4/2 min interval between) were performed. The raw data were integrated, normalized for concentration, and analyzed using a single binding site model, provided in the MicroCal Origin software package.



	Fragment **1**	Fragment **3**	Fragment **4**	Fragment **6**
N	1.00[a]	1.00[a]	1.00[a]	1.03
K_a (M^{-1})	6.742E4 ± 3.861E3	6.81E4 ± 2.62E3	5.58E4 ± 1.6E3	7.50E5 ± 2.12E4
ΔH (kcal/mol)	-5.0	-6.6	-6.6	-10.6
ΔS (cal/mol/deg)	5.0	-0.4	-0.8	-9.2

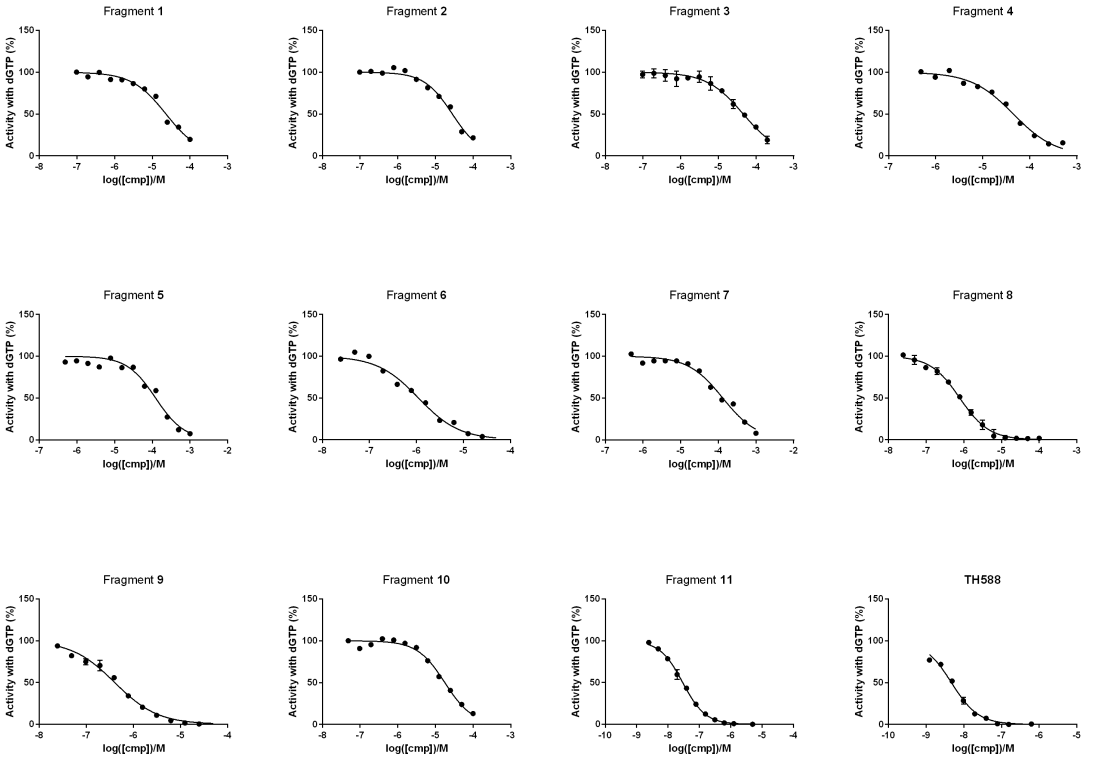
	Fragment **7**	Fragment **8**	Fragment **9**	Fragment **11**
N	1.00[a]	1.02	1.01	0.97
K_a (M^{-1})	1.05E4 ± 666	4.174E5 ± 1.328E4	1.311E6 ± 3.949E4	8.46E6 ± 1.55E6
ΔH (kcal/mol)	-9.5	-6.3	-9.7	-10.8
ΔS (cal/mol/deg)	-14.1	4.3	-5.3	-5.2

Figure S11: **ITC results.** The raw ITC data from the sequential injections and its apparent reaction heat derived from the integration and normalization are shown. The table below summarizes the calculated binding affinity(K_a), stoichiometry (N), enthalpy (ΔH) and entropy(ΔS). [a] Stoichiometry is fixed to 1 in case of weak binders[S24].

3.5 Pyrophosphatase-coupled colorimetric assay

The enzymatic assay was performed as already described elsewhere[S18]. Briefly, MTH1 converts the dGTP into dGMP and pyrophosphate in the assay buffer (100 mM Tris-acetate pH 7.5, 40 mM NaCl, 10 mM Mg acetate, 1 mM DTT and 0.005% Tween 20) at room tem-

perature (RT). The coupled inorganic pyrophosphatase catalyzed then the pyrophosphate to inorganic phosphate which can be quantified by malachite green reagent in a subsequent step. Inhibition of MTH1 leads to a decrease of inorganic phosphate, and therefore the absorbance at 650 nm measured by the GENios plate reader (Tecan) will drop. The enzymatic reaction was carried out in 100 μl assay buffer supplemented with 5 nM recombinant MTH1, 0.2 U/ml inorganic pyrophosphatase (Merck, I1643-100UN), 100 μM dGTP (Jena Bioscience, NU-1003L) and 1% (v/v) DMSO for 15 min. The reaction was subsequently stopped by 25 μl malachite green reagent[S25], followed by an additional incubation for 15 min at RT. The IC_{50} derived from a fitting of a dose-response curve to the data points using the non-linear regression analysis in GraphPad Prism version 8.00.



	Fragment **1**	Fragment **2**	Fragment **3**	Fragment **4**	Fragment **5**	Fragment **6**	Fragment **7**	Fragment **8**	Fragment **9**	Fragment **10**	Fragment **11**	**TH588**
IC_{50} (μM)	23	28	46	44	123	1.1	132	0.8	0.4	17	0.03	0.005
HillSlope	-0.96	-1.16	-0.9	-0.97	-1.14	-1.00	-0.95	-1.12	-0.95	-1.20	-1.20	-1.21
R^2	0.98	0.98	0.97	0.99	0.96	0.98	0.98	0.99	0.99	0.99	1.00	0.99

Figure S12: **IC_{50} determination.** Dose-response curves of all compounds mentioned in this study are shown with the best fitting (straight lines). The table below summarizes the calculated IC_{50} values and the various descriptors for the fitting model.

3.6 BACE1 assay

BACE1 inhibition was measured by the SensoLyte BACE1(β-secretase) assay kit (AnaSpec, AS-71144). In this kit, BACE1 cleaves the quenched QXL 520/ HiLyte Fluor 488 FRET substrate. The released HiLyte Fluor 488 was then monitored at excitation/emission (ex/em) = 490/520 nm. The signals from the compounds were normalized to the DMSO control.

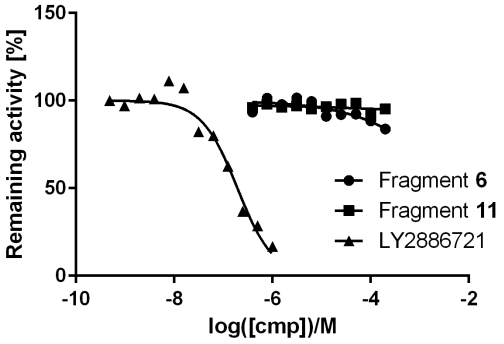


Figure S13: **Compounds 6 and 11 do not inhibit BACE1.** The BACE1 nanomolar inhibitor LY2886721, delivered in the kit, was used as positive control.

3.7 hERG liability

Compound **11** was outsourced to Reaction Biology Corp. (USA) for testing its hERG liability. In the assay, the compound has to compete against a fluorescently labeled tracer which binds to the membrane preparation containing hERG.

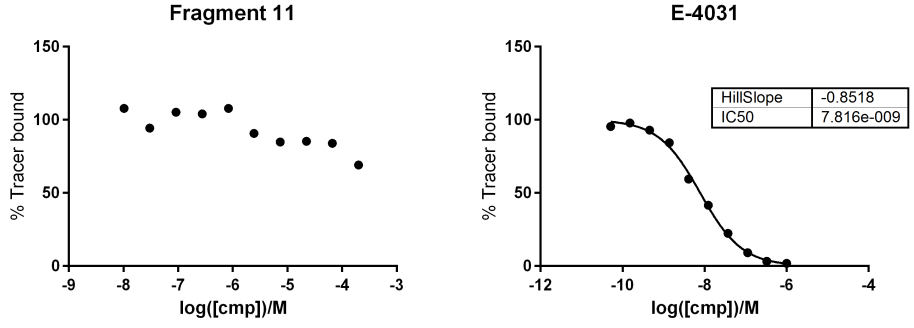


Figure S14: **Competition binding assay for hERG.** The binder E-4031 is used as control compound.

3.8 CETSA

Target engagement in cells was measured by the CETSA approach as described elsewhere[S19]. In brief, K562 cells were concentrated to $4*10^7$ cells/ml in Hank's Balanced Salt Solution (HBSS, gibco, 14025-050) and added to a compound dilution series in 1:1 ratio. The solutions were incubated for 60 min at 37 °C with another shaking every 15 min. Samples were then heated up to 58 °C for 5 min followed by cooling to room temperature. Before the centrifugation at 4 °C, cells were lysed by three freeze-thaw cycles. The amount of MTH1 in the supernatant was quantified by Western blot using MTH1 antibody (Santa Cruz Biotechnology, sc-67291)

3.9 Resazurin-based proliferation assay

HeLa cells were seeded at 10,000 cells per well in 96-well culture dishes and treated with compounds the day after. After 72 h incubation, the media was removed, and cells were washed with Phosphate-Buffered Saline (PBS, Merck, D8537) followed by resazurin staining. Fluorescence intensity was quantified after 4 h incubation using Infinite M1000 plate reader (Tecan) with 530/590 nm (ex/em). The GI_{50} derived from a fitting of a dose-response curve to the data points using the non-linear regression analysis in GraphPad Prism version 8.00.

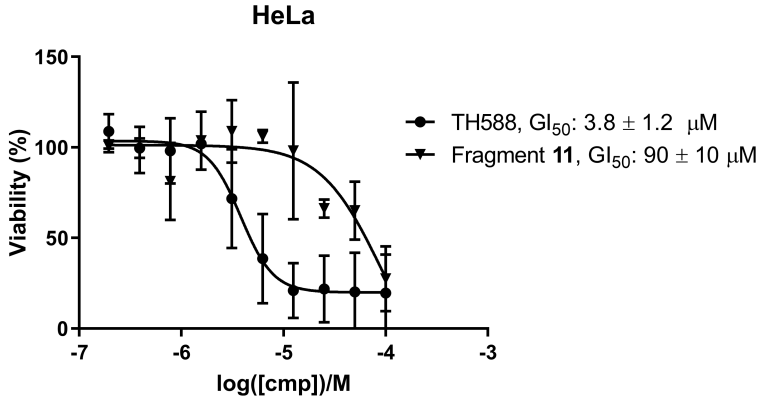


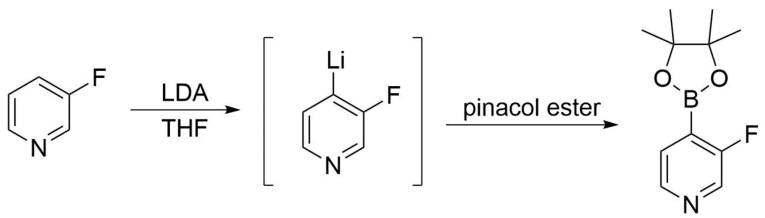
Figure S15: **Antiproliferative effects.** Fragment **11** does not show any relevant growth inhibitory activity on HeLa cells.

4 Synthetic methods

Unless otherwise stated, all reactions were performed under N_2. The reagents were commercially received and used without additional purification. The solvents were dried over activated molecular sieves of appropriate size. All reactions were monitored by HPLC, TLC, and NMR. Chromatography was carried out over silica gel. ^{1}H and ^{13}C NMR spectra were recorded on AV2 400 MHz Bruker spectrometer. The chemical shift is expressed in ppm and calibrated to the H and C signals of the solvents. The following abbreviations are used for the multiplicities: singlet(s), doublet (d) and triplet (t). Mass conformation for the synthetic products was accomplished with an Agilent 1290 Infinity LC system coupled to an Agilent 6540 quadrupole time-of-flight mass spectrometer. The jet stream electrospray source was operated in positive mode with following parameter settings: nebulizer pressure 35 psig, nozzle voltage 0 V, sheath gas flow 11 L/min, sheath gas temperature 375 °C, drying gas flow 8 L/min, drying gas temperature 250 °C, capillary voltage 3000 V and fragmentor voltage of 175 V. Accurate mass spectra were acquired in profile mode over an m/z range of 100 - 1000 by 1 spectrum per second. The Q-TOF instrument was operated in high-resolution mode with 1700 m/z instrument mass range at a resolving power of 33 000 (measured at m/z 322). The purity was acquired by HPLC on Agilent LC device using a NUCLEOSHELL RP18 column (50 x 4.6 mm, 2.7 μm) with MeCN and 0.01 M $(NH_4)_2HPO_4$ pH 6.6 as solvents.

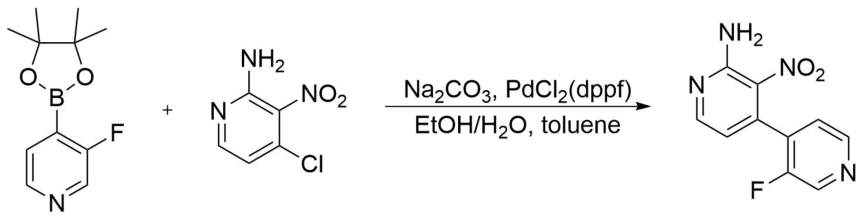
4.1 Synthesis of fragment 11

4.1.1 3-Fluoro-4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)pyridine



The reaction was performed as already described elsewhere with some modifications[S26,S27]. To a solution of diisopropylamine (1 ml, 7 mmol) in THF (10 ml) at -10 °C, was added dropwise a solution of 2.4 M nBuLi (2.9 ml, 7 mmol). The reaction mixture was cooled to -60 °C, then 3-fluoropyridine (500 mg, 5 mmol) was added. After 45 min at -60 °C, 2-isopropoxy-4,4,5,5-tetramethyl-1,3,2-dioxaborolane (2 ml, 10 mmol) was added. The reaction mixture was allowed to reach room temperature and was quenched with cold water (10 ml). The pH was adjusted to 6-7 with 1 M HCl. The product was extracted from the aqueous solution with dichloromethane, followed by filtering and evaporation. The white solid was subsequently washed several times with a mixture of toluene and pentane (1:9), whereby the precipitate gave the desired product as white solid. Yield: 25-32%. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.47 (s, 1H), 8.43 (dd, 1H, J=2.0, 4.6 Hz), 7.60 (t, 1H, J=4.8 Hz), 1.38 (s, 12H). ^{13}C NMR (100 MHz, $CDCl_3$) δ 162.9 (d, J_{CF}=260.1 Hz), 145.0 (d, J_{CF}=4.4 Hz), 138.2 (d, J_{CF}=26.7 Hz), 129.8 (d, J_{CF}=4.3 Hz), 84.7, 24.0.

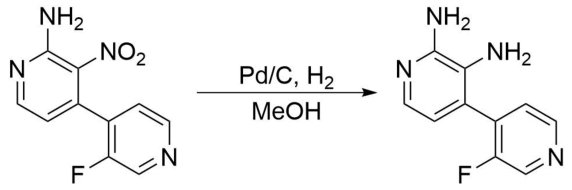
4.1.2 3'-Fluoro-3-nitro-[4,4'-bipyridin]-2-amine



The reaction was performed as already described for 4-(2-fluorophenyl)-3-nitropyridin-2-amine with small modifications[S28]. 4-Chloro-3-nitropyridin-2-amine (495 mg, 2.85 mmol), 3-fluoro-4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)pyridine (750 mg, 3.4 mmol), $PdCl_2$(dppf) (115 mg, 0.14 mmol) and Na_2CO_3 (600 mg, 5.7 mmol) were stirred in toluene (3 ml), H_2O (1.875 ml) and EtOH (0.625 ml) at 70 °C for 4 h. The reaction mixture was diluted in EtOAc and washed with brine. The organic layer was dried over Na_2SO_4, filtered and evaporated. The crude powder (1045 mg) was further purified chromatographically to obtain the desired product as yellow solid (450 mg, 67% yield): ^{1}H NMR (400 MHz ,DMSO-d6) δ 8.67 (d,

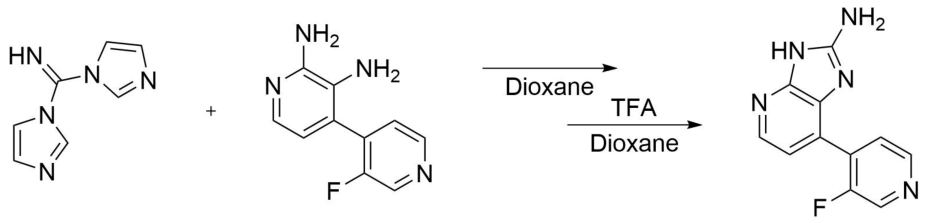
1H, J=1.7 Hz), 8.56 (dd, 1H, J=1.2, 4.8 Hz), 8.42 (d, 1H, J=4.7 Hz), 7.66 (bs, 2H), 7.56 (dd, 2H, J=4.9, 6.3 Hz), 6.72 (d, 1H, J=4.7 Hz). ^{13}C NMR (100 MHz , DMSO-d6) δ 155.6 (d, J_{CF}= 255.0 Hz), 154.7, 153.8, 147.0 (d, J_{CF}=5.1 Hz), 138.6, 138.1 (d, J_{CF}=23.5 Hz), 133.4 (d, J_{CF} =13.5 Hz), 127.4, 124.1, 114.9. HRMS calculated for $C_{10}H_7FN_4O_2$: 234.05530, found 234.05526.

4.1.3 3'-Fluoro-[4,4'-bipyridin]-2,3-diamine



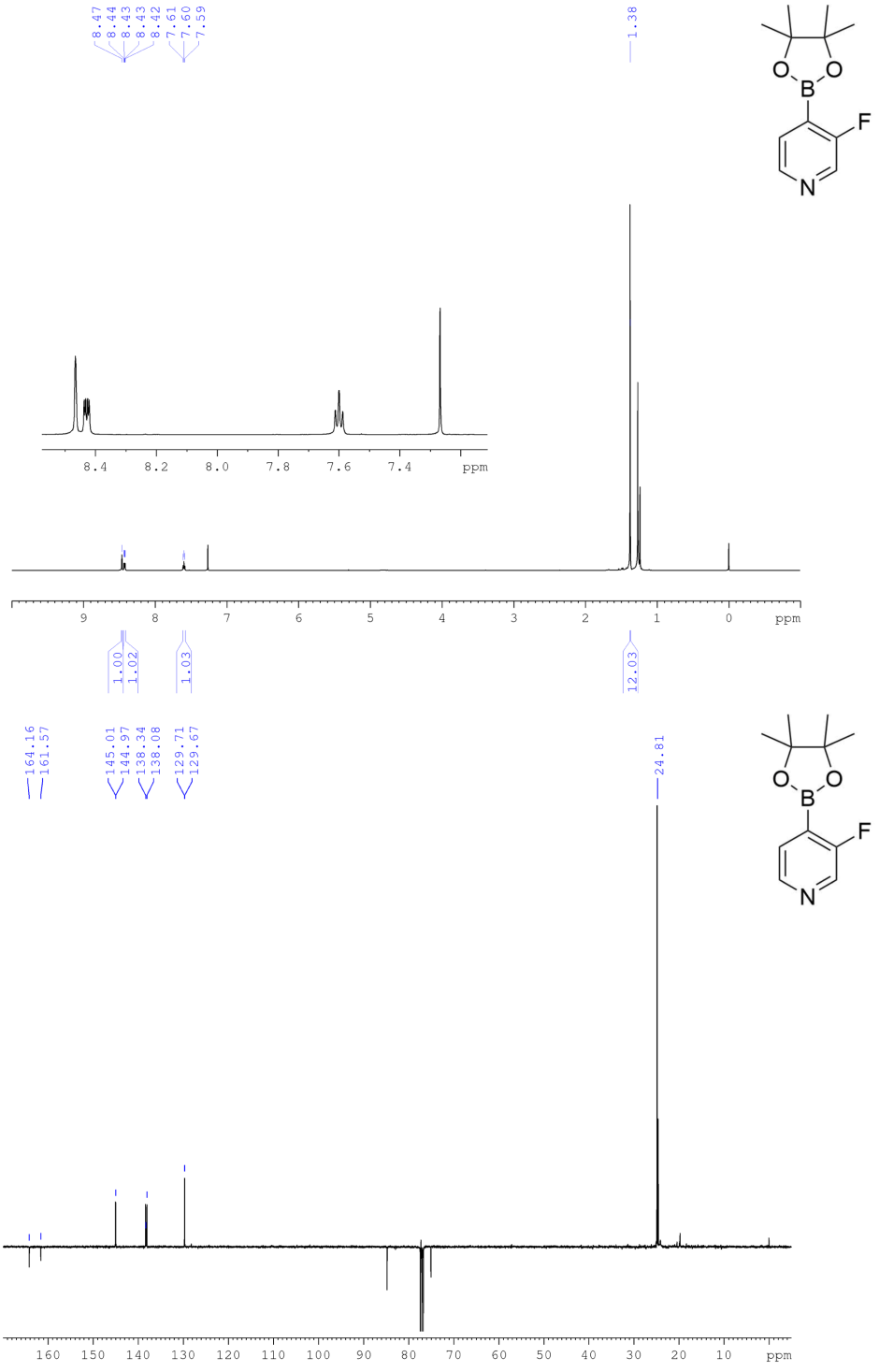
Reaction was performed as already described for 4-(2-fluorophenyl)pyridine-2,3-diamine with small modifications[S28]. A solution of 3-nitro[4,4-bipyridine]-2-amine (117 mg, 0.5 mmol) in MeOH (5 ml) and Pd/C(106 mg) was stirred under 5 bar of H_2 at room temperature for 2 h. The reaction mixture was filtered through Hyflo Super-Cel and the filtrate was concentrated *in vacuo* to give the product as brown solid (98 mg, 96% yield). ^{1}H NMR (400 MHz, DMSO-d6) δ 8.63 (bd, 1H, J=1.6 Hz), 8.48 (dd, 1H, J=0.9, 4.8 Hz), 7.43 (dd, 1H, J= 5.0, 6.27 Hz), 7.35 (d, 1H, J= 5.2 Hz), 6.31 (d, 1H, J=5.1 Hz), 5.70 (bs, 2H), 4.62 (bs, 2H). ^{13}C NMR (100 MHz, DMSO-d6) δ 156.8 (d, J_{CF}=255.6 Hz), 146.5 (d, J_{CF}=5.0 Hz), 139.0 (d, J_{CF}=24.4 Hz), 134.9, 133.8 (d, J_{CF}=13.8 Hz), 127.5, 126.1, 121.2, 114.1. HRMS calculated for $C_{10}H_9FN_4$: 204.08112, found 204.08096.

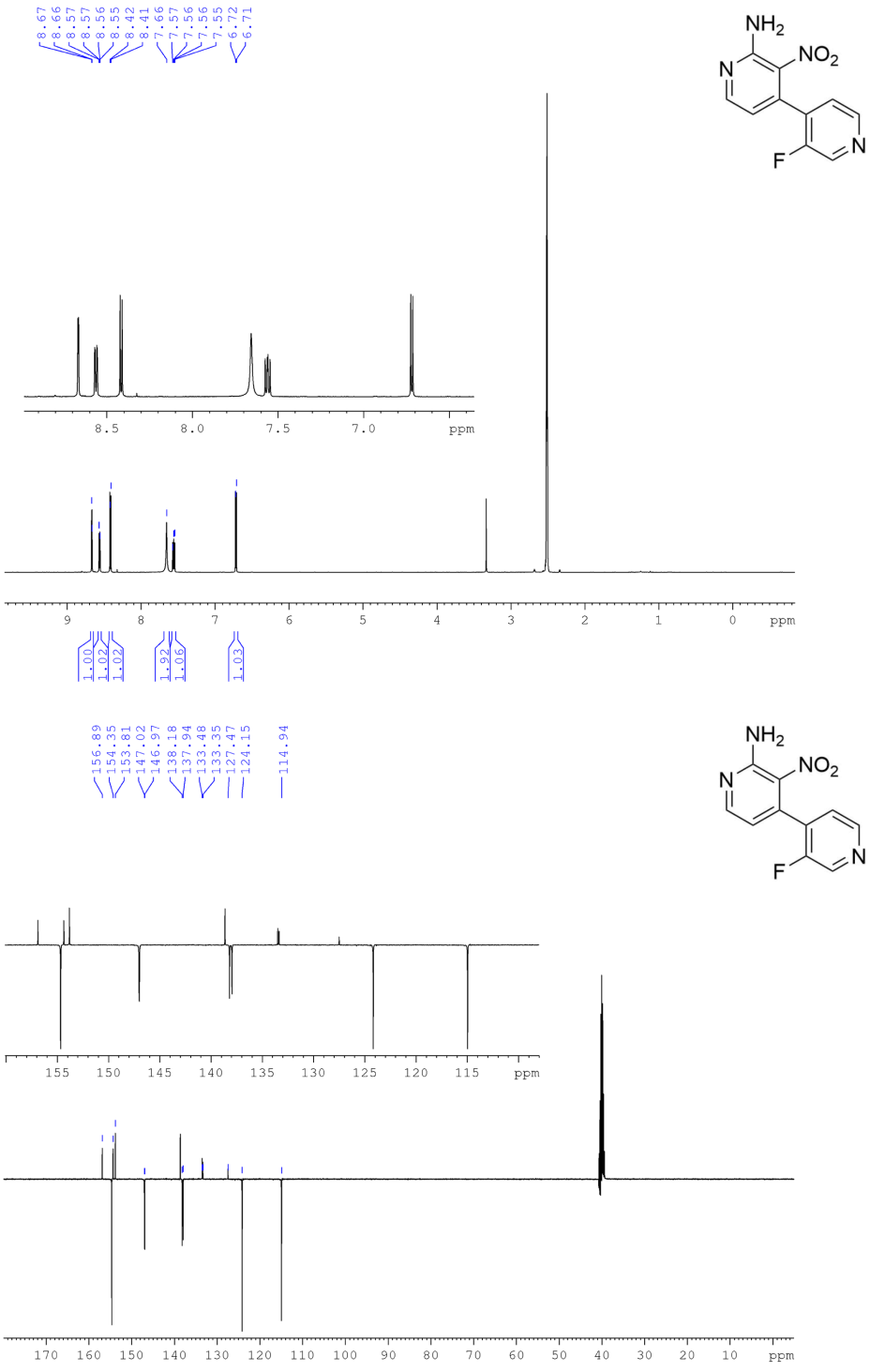
4.1.4 7-(3-Fluoropyridin-4-yl)-3H-imidazol[4,5-b]pyridin-2-amine[S1]

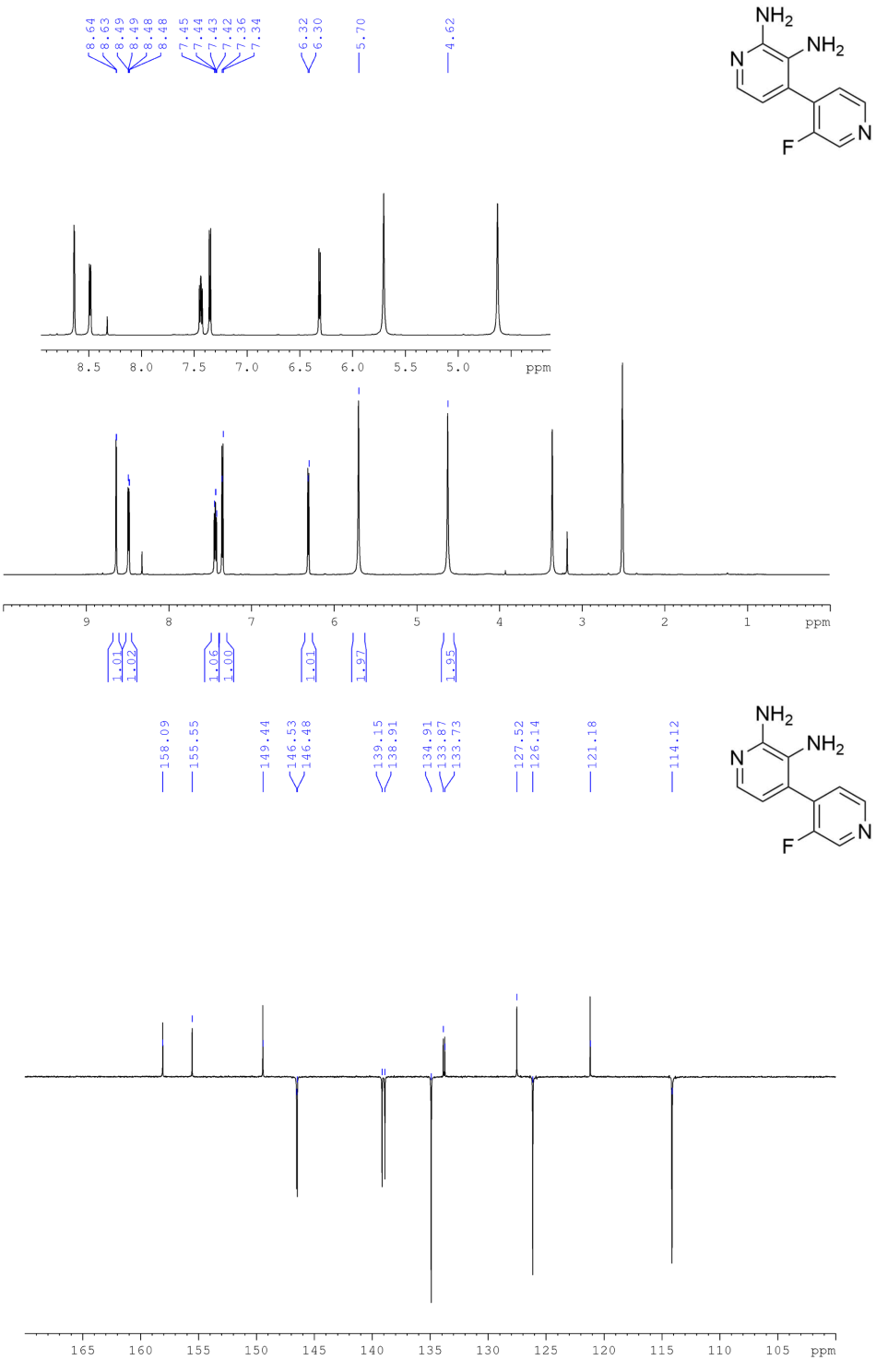


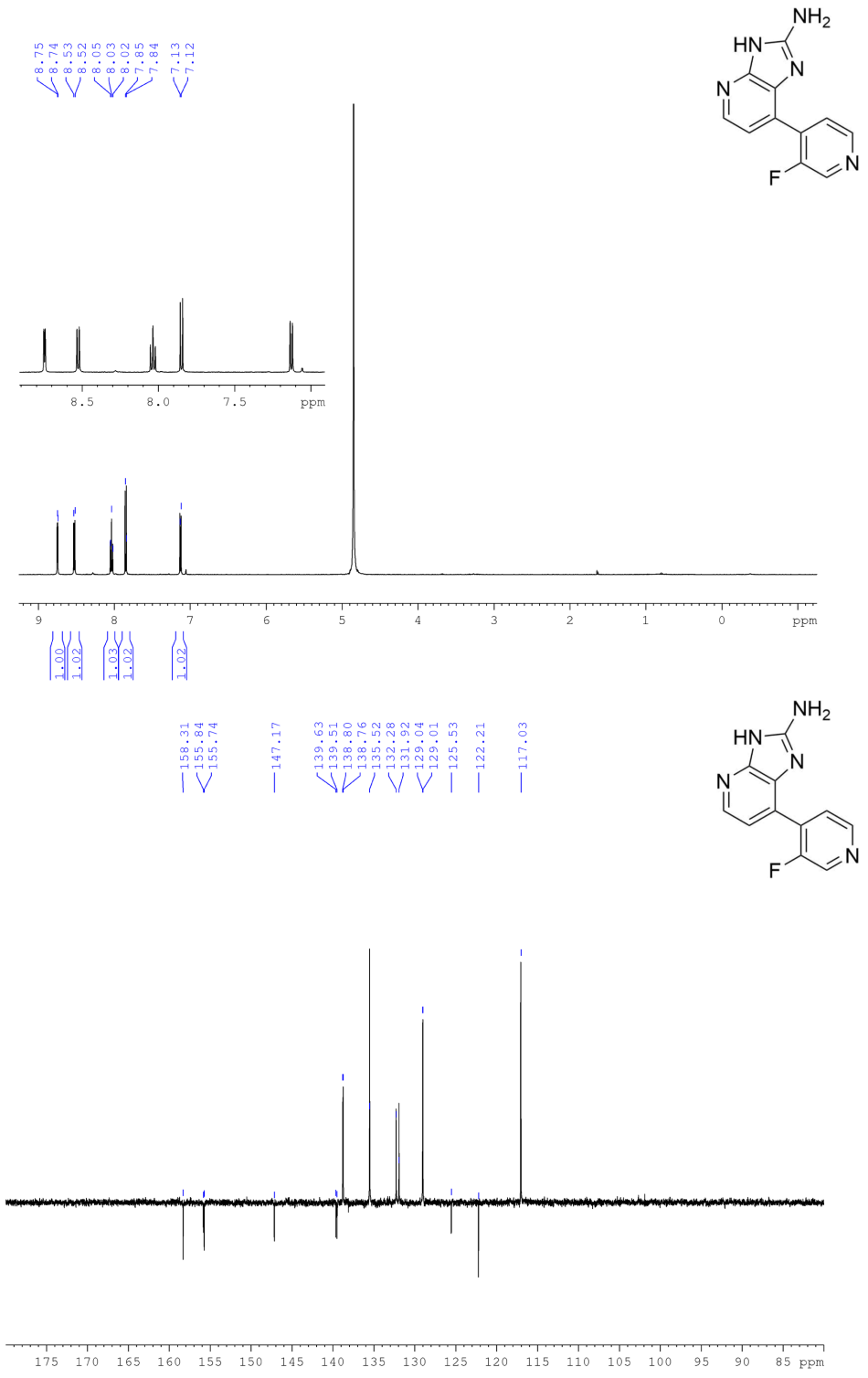
A solution of di(imidazole-1-yl)methanimine (64 mg, 0.55 mmol) in 1,4-dioxane (2.4 ml) and [4,4-bipyridine]-2,3-diamine (112 mg, 0.5 mmol) was stirred at 90 °C for 2 days. The reaction mixture was cooled to room temperature. The precipitation was removed by centrifugation, while the supernatant was further purified chromatographically. The intermediate (140 mg, 0.47 mmol) was further solved in dioxane (3 ml) and treated with TFA (56 μl) at 80 °C for 3 h. The reaction mixture was cooled to room temperature. The precipitation was collected by centrifugation and washed three times with dioxane (3 ml). The solid was dissolved in EtOAc (50 ml) and washed with saturated $NaHCO_3$ solution. The organic layer was dried over Na_2SO_4 and evaporated. The residual product was subsequently dissolved into 1 M HCl (400 μl) and precipitated again by basifying with 2 M NaOH. The precipitate was dried *in vacuo* to give the desired product as a brown solid (25 mg, 22% yield). ^{1}H NMR (400 MHz, D_2O-DCl) δ 8.75 (d, 1H, J=3.5 Hz), 8.52 (d, 1H, J=6.0 Hz), 8.03 (t, 1H, J=6.3 Hz), 7.85 (d, 1H, J=6.1 Hz), 7.13 (d, 1H, J=6.1 Hz). ^{13}C NMR (100 MHz, D_2O-DCl) δ 157.9 (d, J_{CF}=258.9 Hz), 155.8, 147.2, 139.6 (d, J_{CF}=12.7 Hz), 138.8 (d, J_{CF}=4.9 Hz), 135.5, 132.1 (d, J_{CF}=36.2 Hz), 129.0 (d, J_{CF}=3.1 Hz), 125.5, 122.2, 117.0. HRMS calculated for $C_{11}H_8FN_5$: 229.07637, found 229.07622.

4.2 NMR traces of intermediate compounds and fragment 11

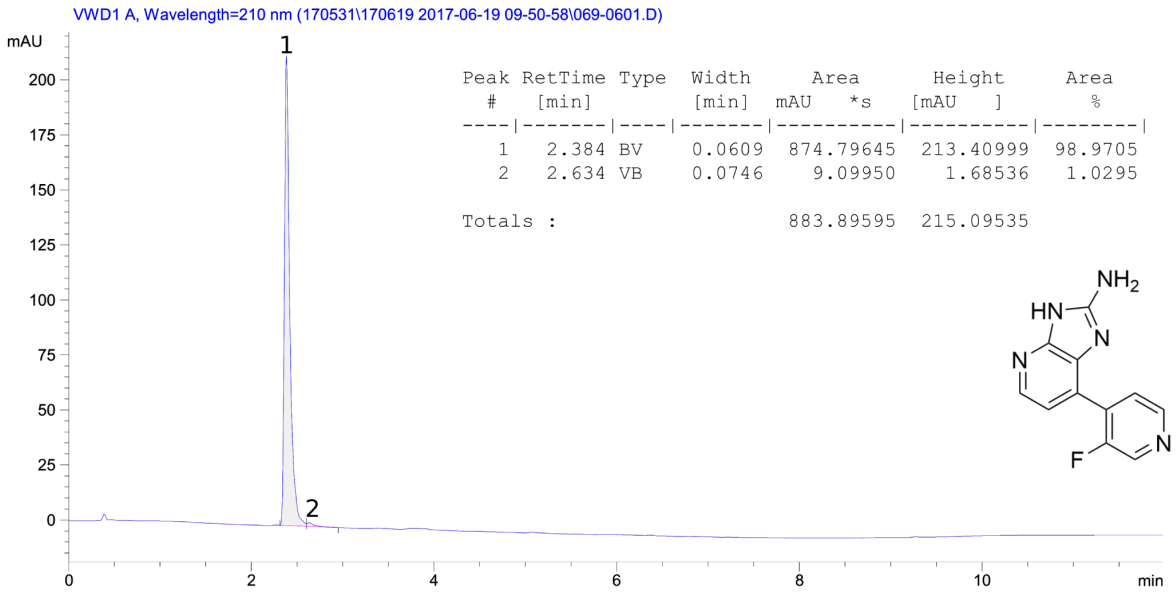



8.67
8.66
8.57
8.57
8.56
8.55
8.42
8.41
7.66
7.57
7.56
7.55
6.72
6.71
NH2
NO2
N
F
N
8.5
8.0
7.5
7.0
ppm
9
8
7
6
5
4
3
2
1
0
ppm
1.00
1.02
1.02
1.92
1.06
1.03
156.89
154.35
153.81
147.02
146.97
138.18
137.94
133.48
133.35
127.47
124.15
114.94
NH2
NO2
N
F
N
155
150
145
140
135
130
125
120
115
ppm
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
ppm


8.64
8.63
8.49
8.49
8.48
8.48
7.45
7.44
7.43
7.42
7.36
7.34
6.32
6.30
5.70
4.62
NH2
NH2
N
F
N
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
ppm
9
8
7
6
5
4
3
2
1
ppm
1.01
1.02
1.06
1.00
1.01
1.97
1.95
158.09
155.55
149.44
146.53
146.48
139.15
138.91
134.91
133.87
133.73
127.52
126.14
121.18
114.12
NH2
NH2
N
F
N
165
160
155
150
145
140
135
130
125
120
115
110
105
ppm


8.75
8.74
8.53
8.52
8.05
8.03
8.02
7.85
7.84
7.13
7.12
NH2
HN
N
N
F
N
8.5
8.0
7.5
ppm
9
8
7
6
5
4
3
2
1
0
ppm
1.00
1.02
1.02
1.03
1.02
1.02
158.31
155.84
155.74
147.17
139.63
139.51
138.80
138.76
135.52
132.28
131.92
129.04
129.01
125.53
122.21
117.03
175
170
165
160
155
150
145
140
135
130
125
120
115
110
105
100
95
90
85 ppm

4.3 HPLC trace (for purity) of fragment 11



References

(S1) Wu, Y.-Q.; Limburg, D. C.; Wilkinson, D. E.; Hamilton, G. S. Formation of nitrogen-containing heterocycles using di(imidazole-1-yl)methanimine. *J. Heterocycl. Chem.* **2009**, *40*, 191–193.

(S2) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.

(S3) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.

(S4) Halgren, T. A. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17*, 520–552.

(S5) Halgren, T. A.; Nachbar, R. B. Merck molecular force field. IV. conformational energies and geometries for MMFF94. *J. Comput. Chem.* **1996**, *17*, 587–615.

(S6) Halgren, T. A. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem.* **1996**, *17*, 616–641.

(S7) Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, A. B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R. E.; Morley, S. D. rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput. Biol.* **2014**, *10*, e1003571.

(S8) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(S9) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; Mackerell, A. D. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.

(S10) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144–3154.

(S11) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155–3168.

(S12) Marchand, J.-R.; Dalle Vedove, A.; Lolli, G.; Caflisch, A. Discovery of Inhibitors of Four Bromodomains by Fragment-Anchored Ligand Docking. *J. Chem. Inf. Model.* **2017**, *57*, 2584–2597.

(S13) Im, W.; Beglov, D.; Roux, B. Continuum solvation model: Computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput. Phys. Commun.* **1998**, *111*, 59 – 75.

(S14) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: a free tool to discover chemistry for biology. *J. Chem. Inf. Model.* **2012**, *52*, 1757–1768.

(S15) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(S16) Irwin, J. J.; Duan, D.; Torosyan, H.; Doak, A. K.; Ziebart, K. T.; Sterling, T.; Tumanian, G.; Shoichet, B. K. An Aggregation Advisor for Ligand Discovery. *J. Med. Chem.* **2015**, *58*, 7076–7087.

(S17) Ferreira, R. S.; Simeonov, A.; Jadhav, A.; Eidam, O.; Mott, B. T.; Keiser, M. J.; McKerrow, J. H.; Maloney, D. J.; Irwin, J. J.; Shoichet, B. K. Complementarity between a docking

and a high-throughput screen in discovering new cruzain inhibitors. *J. Med. Chem.* **2010**, *53*, 4891–4905.

(S18) Gad, H.; Koolmeister, T.; Jemth, A.-S.; Eshtad, S.; Jacques, S. A.; Ström, C. E.; Svensson, L. M.; Schultz, N.; Lundbäck, T.; Einarsdottir, B. O.; Saleh, A.; Göktrk, C.; Baranczewski, P.; Svensson, R.; Berntsson, R. P.-A.; Gustafsson, R.; Strömberg, K.; Sanjiv, K.; Jacques-Cordonnier, M.-C.; Desroses, M.; Gustavsson, A.-L.; Olofsson, R.; Johansson, F.; Homan, E. J.; Loseva, O.; Bräutigam, L.; Johansson, L.; Höglund, A.; Hagenkort, A.; Pham, T.; Altun, M.; Gaugaz, F. Z.; Vikingsson, S.; Evers, B.; Henriksson, M.; Vallin, K. S. A.; Wallner, O. A.; Hammarstrm, L. G. J.; Wiita, E.; Almlöf, I.; Kaldern, C.; Axelsson, H.; Djureinovic, T.; Puigvert, J. C.; Häggblad, M.; Jeppsson, F.; Martens, U.; Lundin, C.; Lundgren, B.; Granelli, I.; Jensen, A. J.; Artursson, P.; Nilsson, J. A.; Stenmark, P.; Scobie, M.; Berglund, U. W.; Helleday, T. MTH1 inhibition eradicates cancer by preventing sanitation of the dNTP pool. *Nature* **2014**, *508*, 215–221.

(S19) Kettle, J. G.; Alwan, H.; Bista, M.; Breed, J.; Davies, N. L.; Eckersley, K.; Fillery, S.; Foote, K. M.; Goodwin, L.; Jones, D. R.; Kck, H.; Lau, A.; Nissink, J. W. M.; Read, J.; Scott, J. S.; Taylor, B.; Walker, G.; Wissler, L.; Wylot, M. Potent and Selective Inhibitors of MTH1 Probe Its Role in Cancer Cell Survival. *J. Med. Chem.* **2016**, *59*, 2346–2361.

(S20) Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125–132.

(S21) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658–674.

(S22) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 213–221.

(S23) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126–2132.

(S24) Turnbull, W. B.; Daranas, A. H. On the value of c: can low affinity systems be studied by isothermal titration calorimetry? *J. Am. Chem. Soc.* **2003**, *125*, 14859–14866.

(S25) Baykov, A. A.; Evtushenko, O. A.; Avaeva, S. M. A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. *Anal. Biochem.* **1988**, *171*, 266–270.

(S26) Alessi, M.; Larkin, A. L.; Ogilvie, K. A.; Green, L. A.; Lai, S.; Lopez, S.; Snieckus, V. Directed ortho metalation-boronation and Suzuki-Miyaura cross coupling of pyridine derivatives: a one-pot protocol to substituted azabiaryls. *J. Org. Chem.* **2007**, *72*, 1588–1594.

(S27) Bouillon, A.; Lancelot, J. C.; Collot, V.; Bovy, P. R.; Rault, S. Synthesis of novel halopyridinylboronic acids and esters. Part 3: 2, or 3-Halopyridin-4-yl-boronic acids and esters. *Tetrahedron* **2002**, *58*, 4369 – 4373.

(S28) Kumar, S.; Wallace, D. M.; Cao, J.; Chiruta, C.; Hood, J. 2-(1H-indazol-3-yl)-2H-imidazo[4,5-B]pyridine and therapeutic uses thereof. patent US 2016/0090380 A1, 2016.