Deep learning enables rapid identification of potent DDR1 kinase inhibitors

Alex Zhavoronkov [1]*, Yan A. Ivanenkov[1], Alex Aliper[1], Mark S. Veselov[1], Vladimir A. Aladinskiy[1], Anastasiya V. Aladinskaya[1], Victor A. Terentiev[1], Daniil A. Polykovskiy[1], Maksim D. Kuznetsov[1], Arip Asadulaev[1], Yury Volkov[1], Artem Zholus[1], Rim R. Shayakhmetov[1], Alexander Zhebrak[1], Lidiya I. Minaeva[1], Bogdan A. Zagribelnyy[1], Lennart H. Lee [2], Richard Soll[2], David Madge[2], Li Xing[2], Tao Guo [2] and Alán Aspuru-Guzik[3,4,5,6]

[1]Insilico Medicine Hong Kong Ltd, Pak Shek Kok, New Territories, Hong Kong. [2]WuXi AppTec Co., Ltd, Shanghai, China. [3]Department of Chemistry, University of Toronto, Toronto, Ontario, Canada. [4]Department of Computer Science, University of Toronto, Toronto, Ontario, Canada. [5]Vector Institute for Artificial Intelligence, Toronto, Ontario, Canada. [6]Canadian Institute for Advanced Research, Toronto, Ontario, Canada. *e-mail: alex@insilico.com
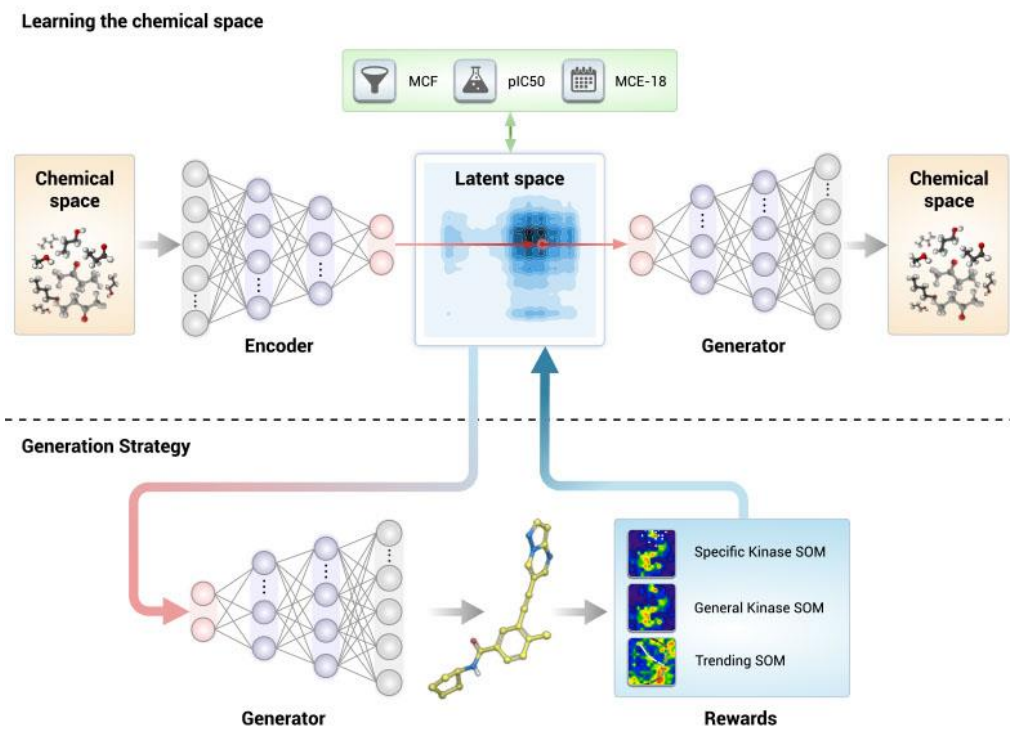
Supplementary Figure 1

Generative Tensorial Reinforcement Learning model.
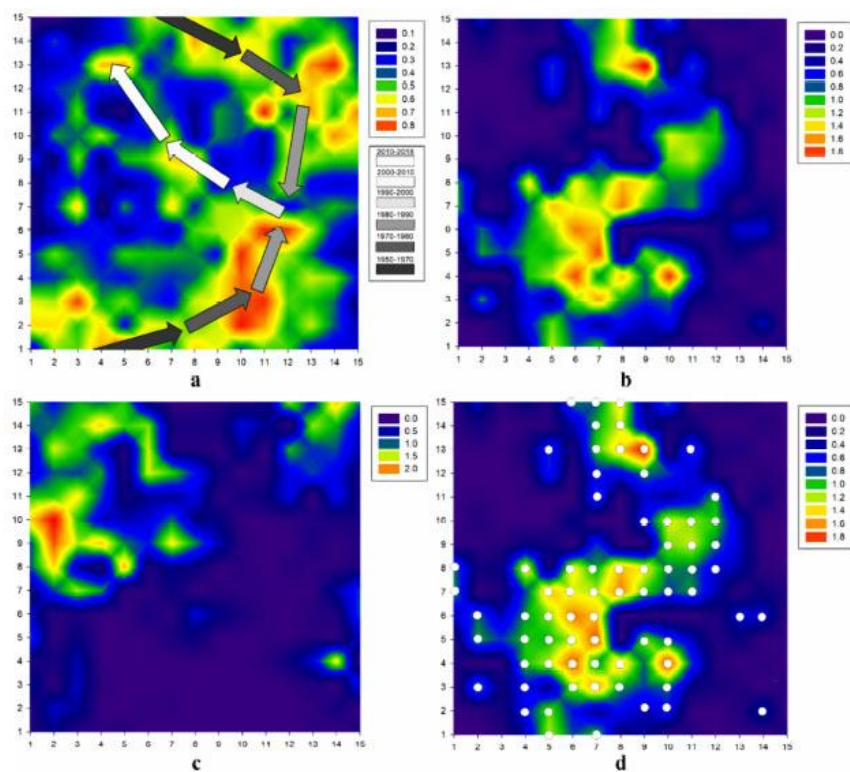
Supplementary Figure 2

Smoothed representation of the General Kinase and Trending SOMs.

(a) Representation of Trending SOM, a Kohonen-based reward function that discriminates "novel" compounds from "old" compounds considering the application priority date of lead compounds disclosed in patents by major pharmaceutical companies. (b) Representation of neurons populated with kinase inhibitors. (c) Representation of neurons populated by molecules with no experimental activity against kinases. (d) Neurons were selected based on PF (circles) and subsequently were used for reward. Within the Specific Kinase SOM (not depicted) we observed that DDR1 inhibitors were distributed in the ensemble of topographically proximal neurons. Finally, we selected those structures which were located in DDR1 associated neurons.
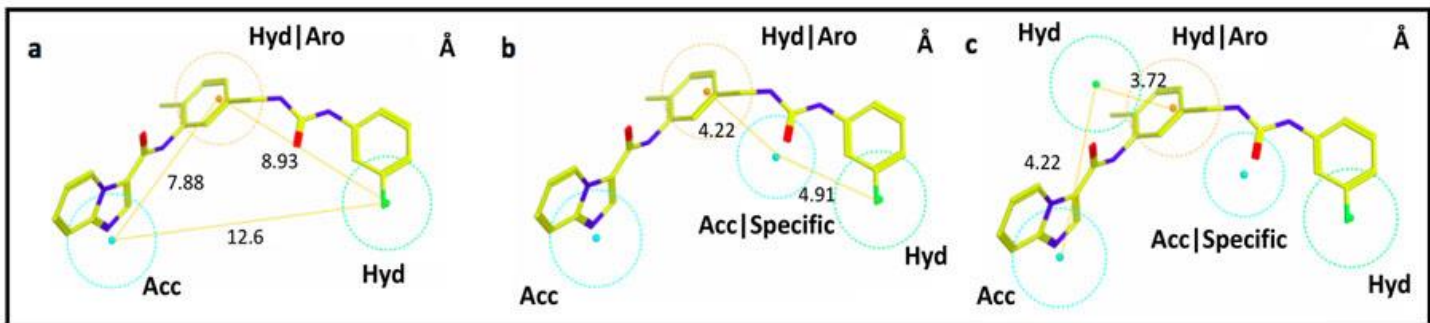
Supplementary Figure 3

Pharmacophore hypotheses.

(**a**) 3-Centered pharmacophore hypothesis: Acc - hydrogen bond acceptor (r = 2Å), Hyd|Aro - hydrophobic or aromatic center (r = 2Å), Hyd - hydrophobic center (r = 2Å). (**b**) 4-Centered pharmacophore hypothesis: Acc - hydrogen bond acceptor (r = 2Å), Hyd|Aro - hydrophobic or aromatic center (r = 2Å), Hyd - hydrophobic center (r = 2Å), Acc|Specific - hydrogen bond acceptor or a fragment with similar spatial geometry (e.g. double or triple bond, planar cycle) (r = 1.7Å). Non-depicted distances are the same as for 3-centered pharmacophore. (**c**) 5-Centered pharmacophore hypothesis containing the same points that are highlighted in **b** above with an additional hydrophobic feature. Non-depicted distances are the same as for 3-centered and 4-centered pharmacophores. Yellow: the reported small-molecule DDR1 inhibitor (PDB code: 5BVN).
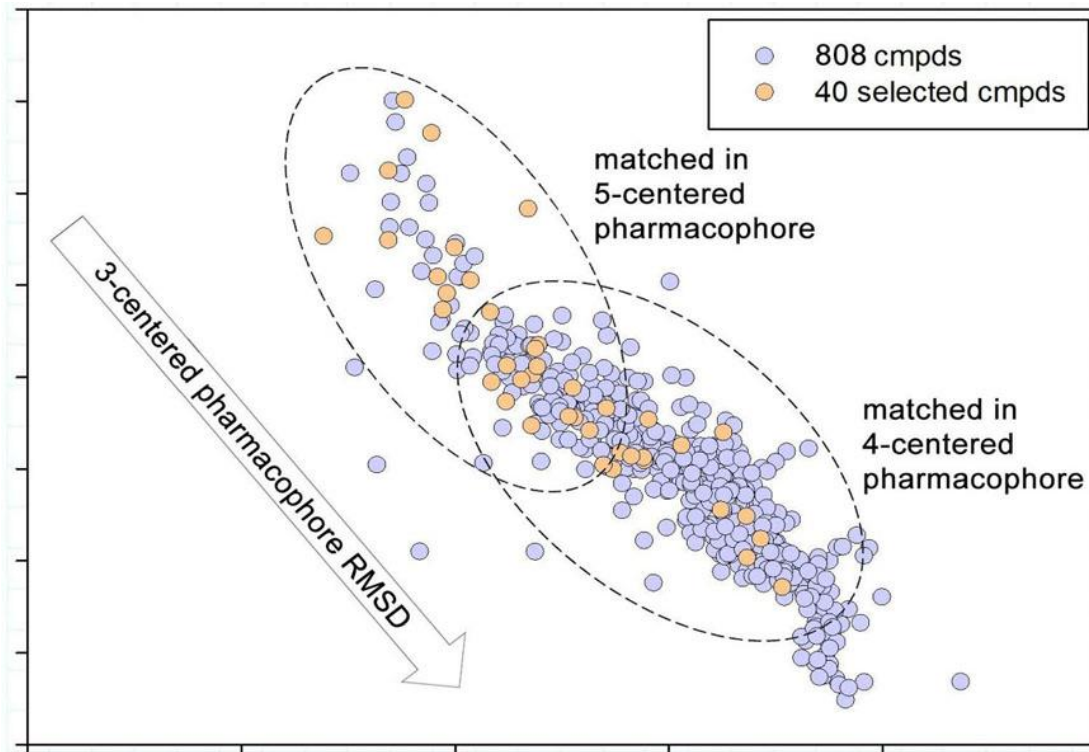
Supplementary Figure 4

Non-linear Sammon map.

The selected 40 molecules are marked by orange triangles. Areas of the best pharmacophore matching are highlighted by circles.
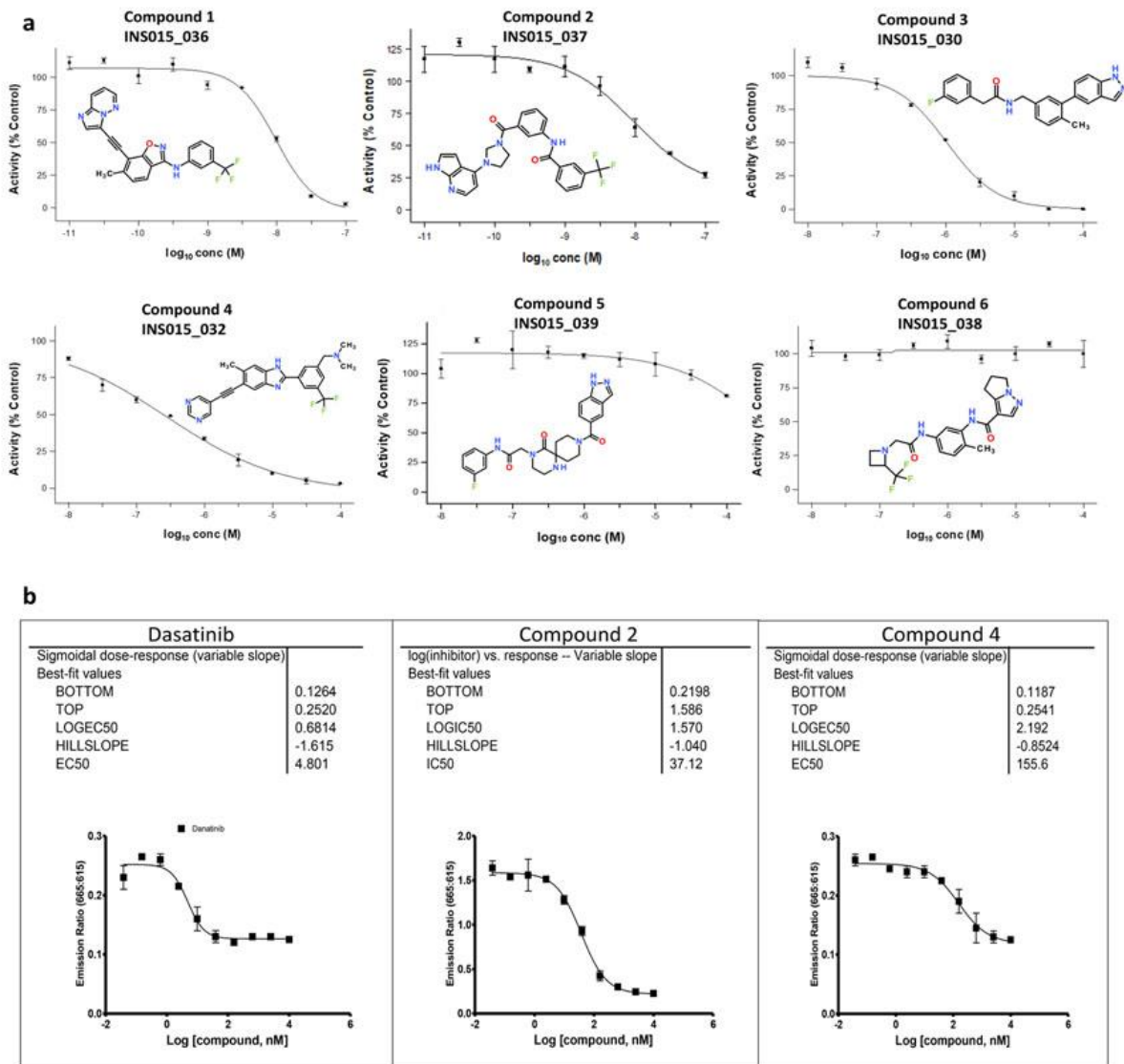
Supplementary Figure 5

The structures and dose-response curves for the generated molecules.

(a) Six generated compounds were tested in a dose-dependent manner against DDR1 tyrosine kinase. Compounds 1 and 2 demonstrated the IC50 values in the low nanomolar range. **(b)** Compounds 2 and 4 were additionally rescreened towards DDR1 kinase using another biochemical assay (Thermo Fisher-PR6913A) and have demonstrated the IC50 values of 37.12 and 155.6 nM respectively (below). Measure of center is mean, error bars are s.d. (n=2 for each experiment).
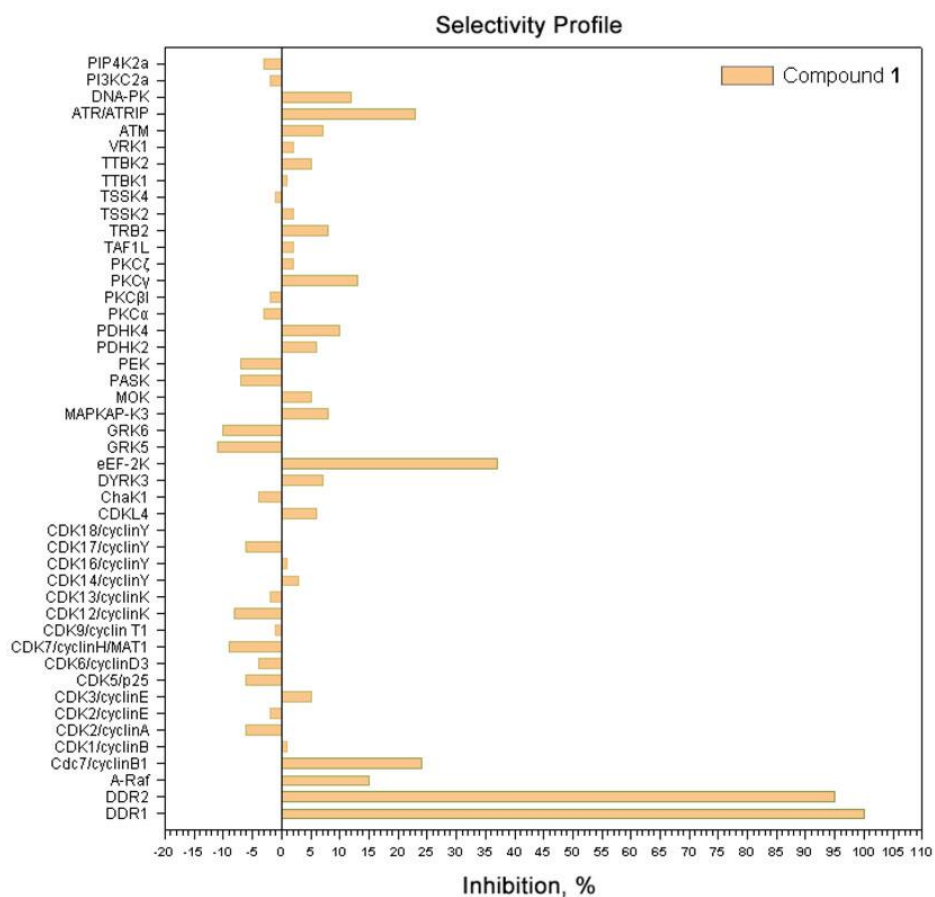
Supplementary Figure 6

Selectivity profile for compound 1 against 44 kinases panel.

The inhibition percent versus 44 non-target kinases was measured at 10μM concentration. The highest inhibition potency(%INH=37) within the panel was revealed against eEF-2K.
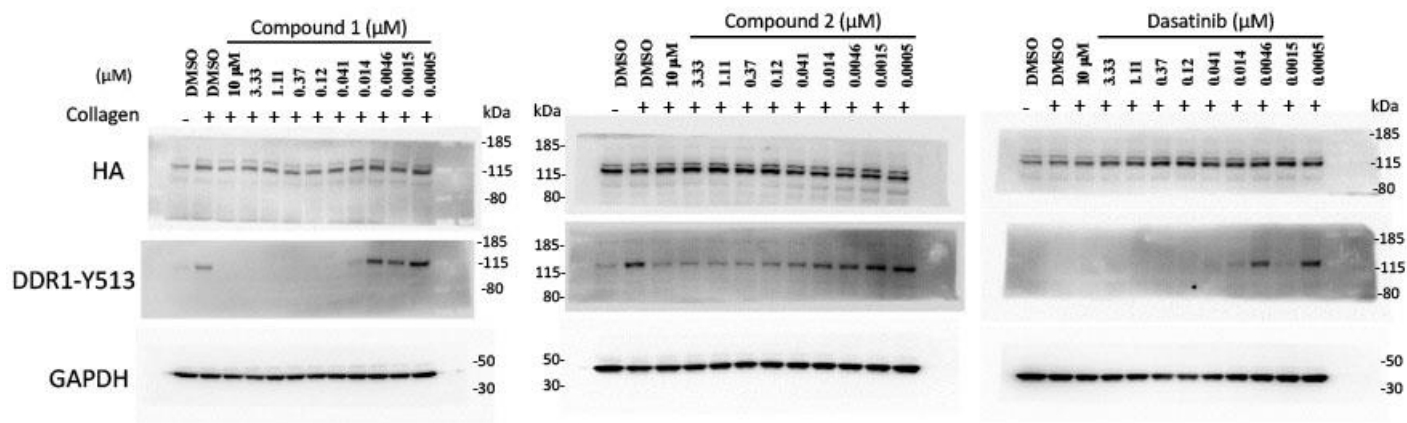
Supplementary Figure 7

Inhibition of DDR1 auto-phosphorylation in U2OS cells stimulated with collagen.

Representative blots of phosphorylated DDR1-Y513 in U2OS cells stimulated with collagen and treated with DDR1 inhibitors at different doses. Dasatinib was served as a positive control. Dasatinib, compounds **1** and **2** inhibited auto-phosphorylation in a dose-dependent manner. Experiments were repeated at least once and similar results were obtained.
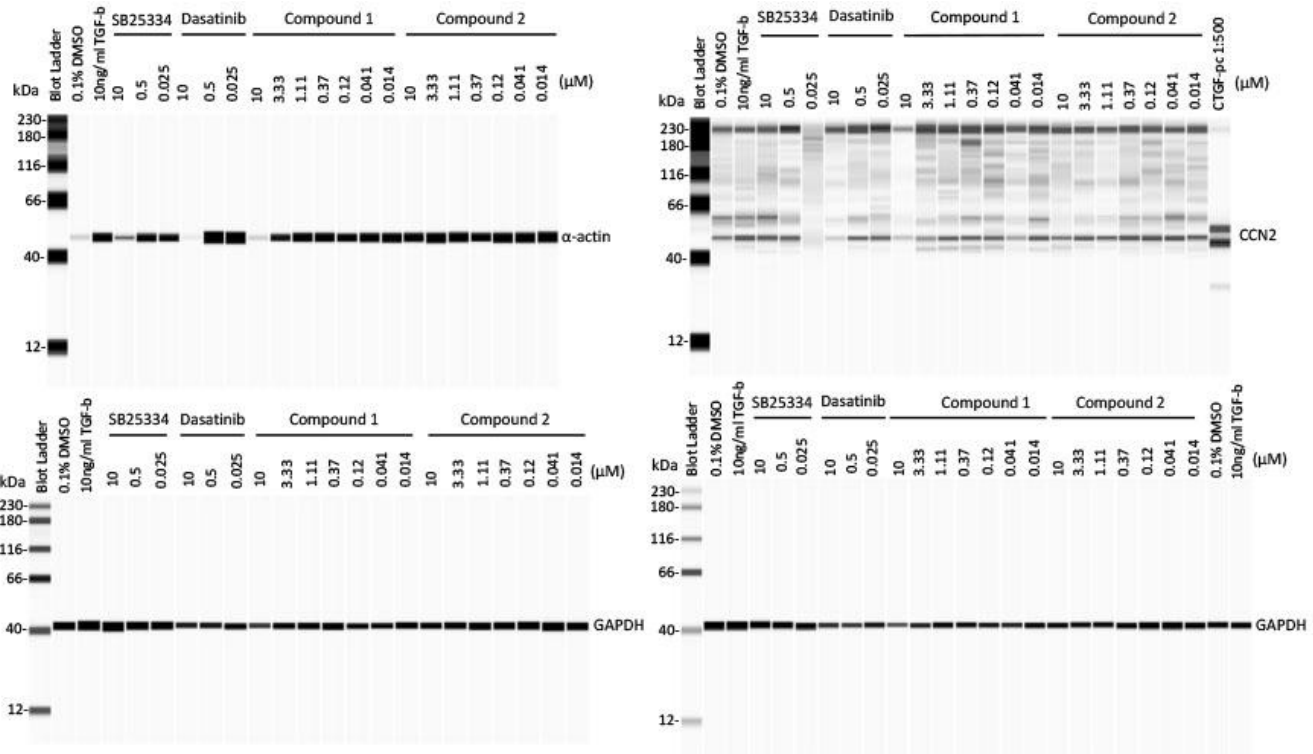
Supplementary Figure 8

Effects of compounds 1 and 2 on cellular fibrosis markers α-actin and CCN2 (normalized to GAPDH) in MRC-5 cells.

Representative blots of produced α-actin and CCN2 in MRC-5 cells treated with TGF-b in the presence of DDR1 inhibitors at different doses. SB25334 and dasatinib were served as a positive control. Dasatinib and compound **1** suppressed α-actin and CCN2 production at the concentration of 10 µM. SB25334 inhibited α-actin production at the dose of 10 µM. Experiments were repeated at least once and similar results were obtained.
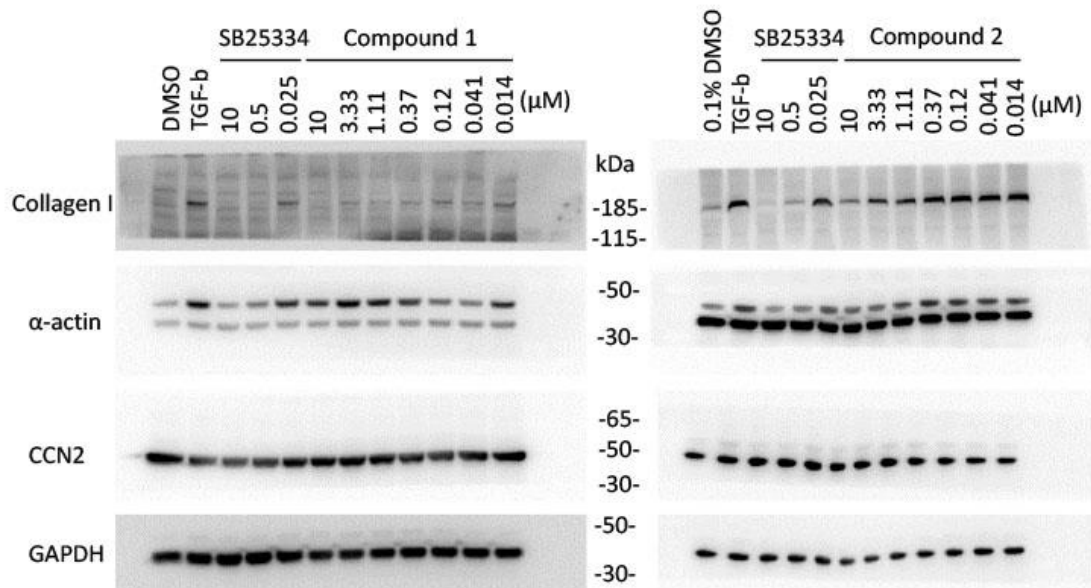
Supplementary Figure 9

Effects of compounds 1 and 2 on cellular fibrosis markers collagen I, α-actin and CCN2 (normalized to GAPDH) in LX-2 cells.

Representative blots of produced collagen I, α-actin and CCN2 in LX-2 cells treated with TGF-b in the presence of DDR1 inhibitors at different doses. SB25334 was served as a positive control. SB25334 and compound **1** suppressed collagen I production in a dose dependent manner. SB25334 inhibited α-actin production at the dose of 10 μM. Experiments were repeated at least once and similar results were obtained.
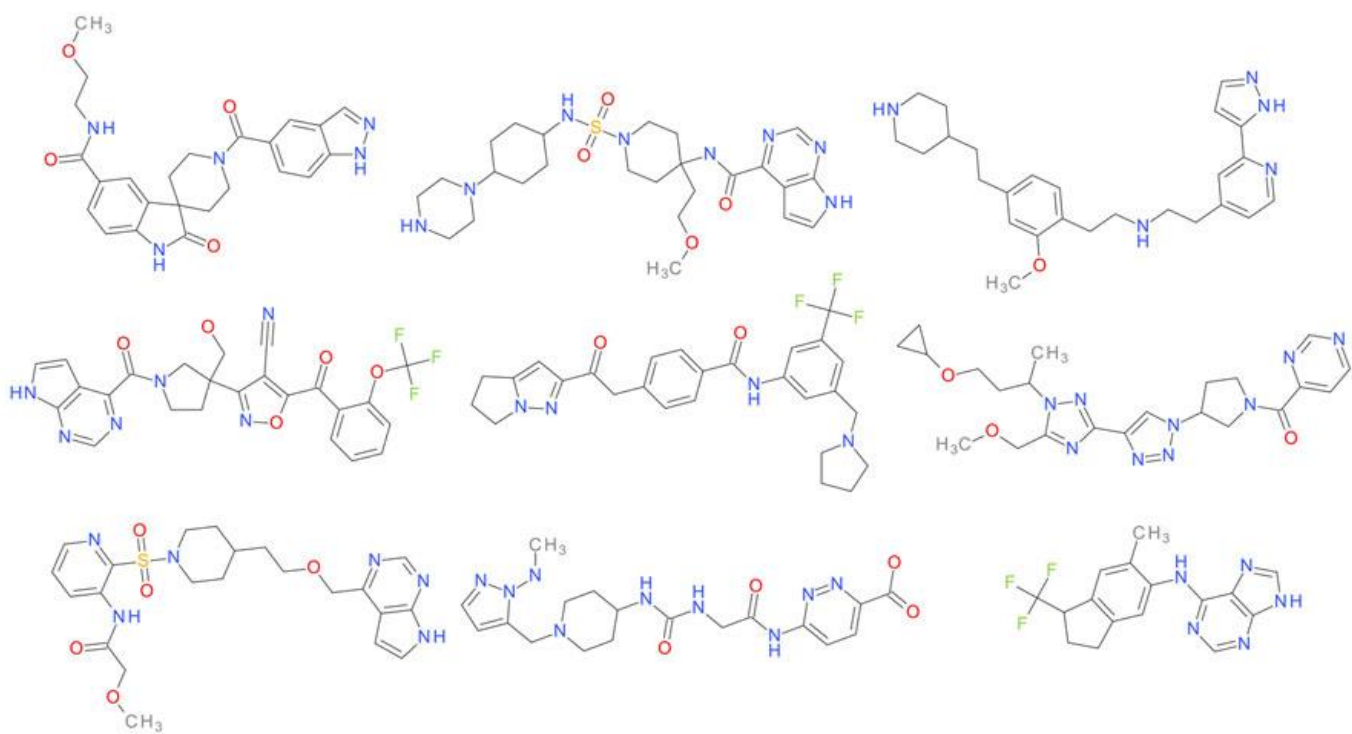
Supplementary Figure 10

Examples of molecules that were rejected during the prioritization step.

Supplementary Note

Chemical Synthesis and Analytical Data

Abbreviations

DCM   dichloromethane
DMF   dimethylformamide
DMSO   dimethylsulphoxide
EA   ethyl acetate
HPLC   high performance liquid chromatography
MeOH   methanol
PBS   phosphate buffered saline
THF   tetrahydrofuran
TFA   trifluoroacetic acid
TLC   thin layer chromatography
Py   pyridine
EDCI   1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
ACN   acetonitrile
T3P   1-Propanephosphonic anhydride solution
Pd(dppf)Cl$_2$ [1,1′-Bis(diphenylphosphino)ferrocene]dichloropalladium(II)
NIS   *N*-Iodosuccinimide
TEA   triethylamine
TFA   trifluoroacetic acid
HATU   1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium   3-oxid
            hexafluorophosphate
DIEA   diisopropylethylamine
NMP   N-Methyl-2-pyrrolidone
XPhos Pd G3   (2-Dicyclohexylphosphino-2′,4′,6′-triisopropyl-1,1′-biphenyl)[2-(2′-amino-
            1,1′-biphenyl)]palladium(II) methanesulfonate
MTBE   Methyl-*tret*-butyl ether
CDI   1,1'-Carbonyldiimidazole

Chemical Synthesis of INS015_030, INS015_32, INS015_036, INS015_037, INS015_038, INS015_039

NMR spectra were recorded on a Bruker 400 (400 MHz ^{1}H, 100 MHz ^{13}C, 400 MHz ^{19}F). Proton chemical shifts are reported in ppm (δ) referenced to the NMR solvent. Data are reported as follows: chemical shifts, multiplicity (br = broad, s = singlet, d = doublet, t = triplet, q = quartet, p = pentet, m = multiplet; coupling constant(s) in Hz; integration). NMR data were collected at 25°C. Analytical TLC was performed on 0.25 mm silica gel 60-F plates. Visualization was accomplished with UV light and I$_2$. Flash chromatography was performed using ISCO Combiflash. Reverse phase chromatography was performed using ISCO Combiflash (column: C18 (20-35μm)). Acidic condition: Mobile Phase A: 0.1% FA in water (v/v). Mobile Phase B: 0.1% FA in acetonitrile (v/v). Basic condition: Mobile Phase A: 0.1% NH$_3$·H$_2$O in water (v/v).

Mobile Phase B: 0.1% in Acetonitrile(v/v). LC/MS spectra were obtained using Agilent 1200\G1956A or SHIMADZU LCMS-2020. Standard LC/MS conditions were as follows (running time 1.55 minutes): Acidic condition: Mobile Phase A: 0.0375% TFA in water (v/v). Mobile Phase B: 0.01875% TFA in acetonitrile (v/v); Column: Kinetex EVO C18 30*2.1mm, 5 µm. Basic condition: Mobile Phase A: 0.025% $NH_3 \cdot H_2O$ in water (v/v). Mobile Phase B: Acetonitrile; Column: Kinetex EVO C18 2.1X30mm, 5 µm. The gradient ran from 5% to 95% mobile phase B or 0 to 60% mobile phase B. HPLC spectra were obtained using SHIMADZU LC-20AB, Standard HPLC conditions were as follows (running time 4 minutes): Acidic condition: Mobile Phase A: 0.0375% TFA in water (v/v). Mobile Phase B: 0.01875% TFA in acetonitrile (v/v); Column: Kinetex EVO C18 50*4.6mm, 5 µm. Basic condition: Mobile Phase A: 0.025% $NH_3 \cdot H_2O$ in water (v/v). Mobile Phase B: Acetonitrile; Column: XBridge C18 2.1X50mm, 5 µm. The gradient ran from 5% to 95% mobile phase B or 0 to 60% mobile phase B. The final product was purified by Prep-HPLC using Gilson 281.

Chemical Synthesis of INS015_030
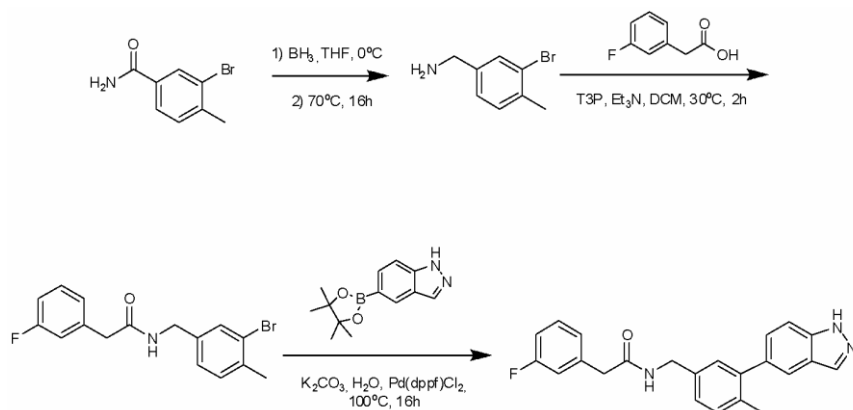
Scheme 2. Synthetic route to INS015_030

(3-bromo-4-methylphenyl)methanamine
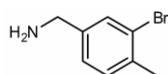
To a solution of 3-bromo-4-methylbenzamide (12 g, 56.06 mmol, 1 *eq*) in THF (100 mL) was added BH_3-Me_2S (10M, 14.01 mL, 2.5 *eq*) at 0°C, and then the mixture was heated up to 70°C and stirred for 16 h. Then the mixture was cooled to 25 °C and quenched with MeOH, and then adjusted the pH=2 with HCl/EA (4M). The mixture was heated to 80 °C and stirred for 2 h. TLC (PE/EA=1/1) showed starting material (Rf=0.50) was consumed completely and new point (Rf=0.0) was formed. The mixture was washed with 1N NaOH solution to adjust the pH=10 and extracted with EA (50 mL*3). The combined organic layers were washed with brine (50 mL*2), dried over Na_2SO_4, filtered and concentrated in vacuum to get the residue. The residue was used into next step without purification. (3-bromo-4-methylphenyl)methanamine (10 g, 49.98 mmol, 89.16% yield) was obtained as a white solid. LCMS: Retention time: 0.913 min, [M+H]$^+$ calcd. for $C_8H_{10}BrN$ 200.0; found 200.1.

N-(3-bromo-4-methylbenzyl)-2-(3-fluorophenyl)acetamide
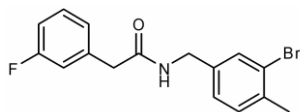
To a solution of 2-(3-fluorophenyl)acetic acid (2.3 g, 14.92 mmol, 1.0 *eq*) and (3-bromo-4-methylphenyl)methanamine (3.6 g, 17.99 mmol, 1.21 *eq*) in DCM (20 mL) was added Et₃N (4.53 g, 44.77 mmol, 6.23 mL, 3 *eq*) and T3P (7.12 g, 22.38 mmol, 6.66 mL, 1.5 *eq*) .The reaction mixture was stirred at 30°C for 2 h. LCMS showed desired MS. The mixture was washed with water (30 mL) and extracted with DCM (50 mL*3). The combined organic layers were washed with brine (30 mL*2), dried over Na₂SO₄, filtered and concentrated in vacuum to get the residue confirmed by ¹H-NMR. The crude product *N*-(3-bromo-4-methylbenzyl)-2-(3-fluorophenyl)acetamide (4 g, 11.90 mmol, 79.73% yield) as white solid was used into the next step without further purification.. ¹H-NMR (400MHz, METHANOL-d4) δ = 7.40 (d, *J*=1.2 Hz, 1H), 7.31 (dt, *J*=6.1, 7.9 Hz, 1H), 7.20 (d, *J*=7.7 Hz, 1H), 7.10 (d, *J*=7.7 Hz, 2H), 7.05 (dd, *J*=2.1, 9.9 Hz, 1H), 6.98 (dt, *J*=2.1, 8.6 Hz, 1H), 4.30 (s, 2H), 3.55 (s, 2H), 2.34 (s, 3H). LCMS: Retention time: 0.926 min, [M+H]⁺ calcd. for C₁₆H₁₅BrFNO 337.0; found 337.7.

N-(3-(1*H*-indazol-5-yl)-4-methylbenzyl)-2-(3-fluorophenyl)acetamide (INS015_030)
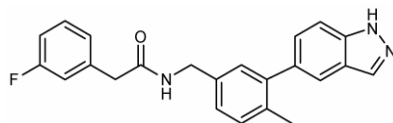
To a mixture of *N*-(3-bromo-4-methylbenzyl)-2-(3-fluorophenyl)acetamide (2 g, 5.95 mmol, 1 *eq*), 5-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1*H*-indazole (2.32 g, 9.52 mmol, 1.6 *eq*) and K₂CO₃ (2.47 g, 17.85 mmol, 3 *eq*) in H₂O (2 mL) and dioxane (20 mL) was added Pd(dppf)Cl₂ (217.64 mg, 297.44 μmol, 0.05 eq) Degassed and purged with N₂ for 3 times, and then the mixture was stirred at 100°C for 16 hr under N₂ atmosphere. TLC (PE/EA=1/2) showed starting material (Rf=0.60) was consumed completely and new main point (Rf=0.20) was formed. The mixture was washed with water (20 mL) and extracted with EA (30mL*3). The combined organic layers were washed with brine (20 mL*2), dried over Na₂SO₄, filtered and concentrated in vacuum to get the crude product confirmed by LCMS. The crude product (4.25 g, crude) was obtained without purification. 250mg of the crude product was purified by prep-HPLC (column: Waters Xbridge 150*25 5u; mobile phase: [water (0.05% ammonia hydroxide v/v)-ACN]; B%: 52%-82%,10min) and lyophilized to get the product. *N*-(3-(1*H*-indazol-5-yl)-4-methylbenzyl)-2-(3-fluorophenyl)acetamide (120.68 mg, 320.83 μmol, 5.39% yield, 99.275% purity) was obtained as a white solid. HPLC: Retention time: 2.271 min. ¹H-NMR (400MHz, METHANOL-d4) δ = 8.06 (s, 1H), 7.62 - 7.53 (m, 2H), 7.31 - 7.17 (m, 3H), 7.15 - 7.01 (m, 4H), 6.89 (dt, *J*=2.0, 8.6 Hz, 1H), 4.37 (s, 2H), 3.53 (s, 2H), 2.21 (s, 3H). ¹⁹F NMR (400MHz, METHANOL-d4) δ =115.38. ¹³C NMR (400MHz, METHANOL-d4) δ= 171.88, 164.02, 161.60, 142.23, 139.27, 138.30, 138.21, 135.91, 134.62, 134.22, 133.69, 130.07, 129.81, 129.73, 128.77, 128.36, 125.94, 124.57, 124.53, 122.99, 120.26, 115.53, 115.31, 113.28, 113.06,109.24, 42.50, 42.13, 42.11, 18.95. LCMS: Retention time: 0.973 min, [M+H]⁺ calcd. for C₂₃H₂₀FN₃O 374.2; found 374.3.

Chemical Synthesis of INS015_032

The first four synthesis steps of INS015_032 was performed by using synthetic methods from Britton *et al.*[1], Mavrova *et al.*[2], Morinaga *et al.*[3]
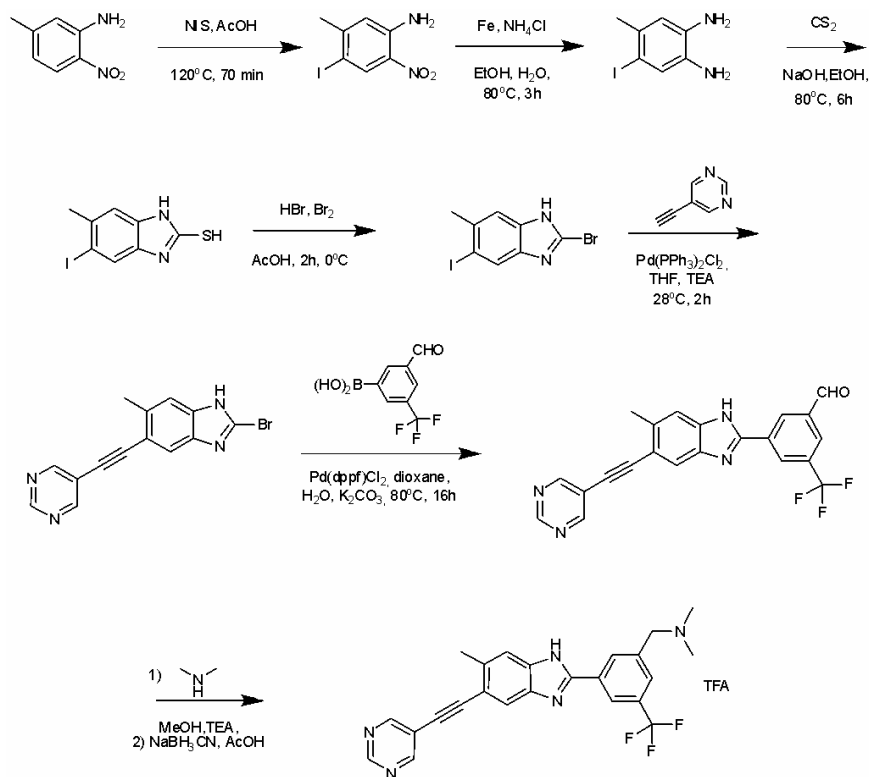
Scheme 3. Synthetic route to INS015_032

4-Iodo-5-methyl-2-nitroaniline
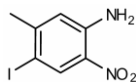
5-Methyl-2-nitroaniline (1 g, 6.57 mmol, 1 eq) and NIS (1.40 g, 6.24 mmol, 0.95 *eq*) in AcOH (60 mL) was refluxed at 120°C for 70 min. TLC (petroleum ether: ethyl acetate =10:1, twice, Rf= 0.55) showed 5-methyl-2-nitroaniline was consumed and a new spot was detected. The reaction mixture was cooled to room temperature and poured into ice-water (120 mL). The precipitate was collected by filtration. The filtered cake was washed with water (50 mL), petroleum ether (50 mL) and dried under the reduced pressure. 4-Iodo-5-methyl-2-nitroaniline (1.1 g, 3.96 mmol, 60.19% yield) was obtained as red solid. ^{1}H-NMR (400MHz, DMSO-d6) ppm= 8.28 (s, 1H), 7.44 (br s, 2H), 6.96 (s, 1H), 2.28 (s, 3H).

4-Iodo-5-methylbenzene-1,2-diamine
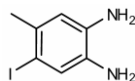
A suspension of 4-iodo-5-methyl-2-nitroaniline (1.1 g, 3.96 mmol, 1 *eq*), Fe (883.73 mg, 15.82 mmol, 4 *eq)* and NH$_4$Cl (2.12 g, 39.56 mmol, 10 *eq*) in EtOH (100 mL) and H$_2$O (20 mL) was stirred at 80°C for 3 h. The reaction mixture was cooled to room temperature and filtered through the celite. The filtrate was concentrated; the residue was dissolved in ethyl acetate (30 mL). The mixture was washed with water (10 mL*2), brine (10 mL) and dried over Na$_2$SO$_4$. After filtration and concentration, 4-iodo-5-methylbenzene-1,2-diamine (1.24 g, crude) was obtained as gray solid and was confirmed by ^{1}H-NMR. The crude product was used to the next step directly. ^{1}H-NMR (400MHz, METHANOL-d4) ppm = 6.90 (s, 1H), 6.47 (s, 1H), 4.63 (br s, 4H), 2.11 (s, 3H).

5-Iodo-6-methyl-1*H*-benzo[*d*]imidazole-2-thiol
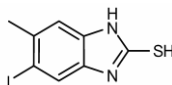
A solution of 4-iodo-5-methylbenzene-1,2-diamine (1.24 g, 5.00 mmol, 1 *eq*), CS$_2$ (3.81 g, 49.99 mmol, 3.02 mL, 10 *eq*) and NaOH (399.87 mg, 10.00 mmol, 2 *eq*) in EtOH (12 mL) and H$_2$O (3 mL) was stirred at 80°C for 6 h. LCMS showed 4-iodo-5-methylbenzene-1,2-diamine was consumed and the desired mass was detected. The reaction mixture was concentrated under the reduced pressure, the residue was dissolved in saturated NH$_4$Cl (20 mL). The precipitate was collected by the filtration. 5-Iodo-6-methyl-1*H*-benzo[*d*]imidazole-2-thiol (0.98 g, 3.38 mmol, 67.57% yield, 100% purity) was obtained as gray solid. ^{1}H-NMR (400 MHz, DMSO-d6) ppm = 7.51 (s, 1H), 7.13 (s, 1H), 2.39 (s, 3H). LCMS: Retention time: 0.813 min, [M+H]$^+$ calcd. for C$_8$H$_7$IN$_2$S 290,9; found 290.9.

2-Bromo-5-iodo-6-methyl-1*H*-benzo[*d*]imidazole

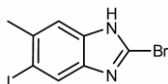A suspension of 5-iodo-6-methyl-1*H*-benzo[*d*]imidazole-2-thiol (0.5 g, 1.72 mmol, 1 *eq*) in HBr (29.80 g, 121.54 mmol, 20 mL, 33% in AcOH, 70.52 *eq*) was cooled to 0°C and Br$_2$ (1.10 g, 6.89 mmol, 355.37 μL, 4 *eq*) in AcOH (5 mL) was added slowly dropwise. The mixture was stirred for 2 h at 0°C. LCMS showed 5-iodo-6-methyl-1*H*-benzo[*d*]imidazole-2-thiol was consumed and a major peak with desired MS was detected. Water (100 mL) was added to the mixture slowly and the solid was precipitated. The solid was obtained by filtration. The solid was washed with water (20 mL) and EA (20 mL). 2-Bromo-5-iodo-6-methyl-1*H*-benzo[*d*]imidazole (380 mg, 1.13 mmol, 65.44% yield) was obtained as gray solid. ^{1}H-NMR (400 MHz, MeOD) ppm = 8.24 - 8.13 (m, 1H), 7.66 (d, *J* = 4.0 Hz, 1H), 2.61 (s, 3H). LCMS: Retention time: 0.886 min, [M+H]$^+$ calcd. for C$_8$H$_6$BrIN$_2$ 337.8; found 337.0

2-Bromo-6-methyl-5-(pyrimidin-5-ylethynyl)-1*H*-benzo[*d*]imidazole
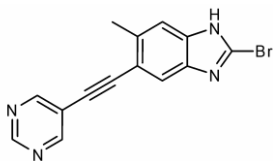
To the solution of 2-bromo-5-iodo-6-methyl-1*H*-benzo[*d*]imidazole (380 mg, 1.13 mmol, 1 *eq*) and 5-ethynylpyrimidine (234.82 mg, 2.26 mmol, 2 *eq*) in THF (6 mL) was added Pd(PPh₃)₂Cl₂ (79.16 mg, 112.77 μmol, 0.1 *eq*), CuI (21.48 mg, 112.77 μmol, 0.1 *eq*) and TEA (1.14 g, 11.28 mmol, 1.57 mL, 10 *eq*). The mixture was stirred at 28°C for 2 h. LCMS showed 2-bromo-5-iodo-6-methyl-1*H*-benzo[*d*]imidazole remained and desired MS was detected. The reaction mixture was quenched by addition water 10 mL, and then extracted with EA (10 mL*3). The combined organic layers were dried over anhydrous Na₂SO₄, filtered and concentrated under reduced pressure to give a residue. The residue was purified by column chromatography (SiO₂, Petroleum ether/Ethyl acetate=10/1 to 1:1) to give 2-bromo-6-methyl-5-(pyrimidin-5-ylethynyl)-1*H*-benzo[*d*]imidazole (150 mg, 411.39 μmol, 36.48% yield, 85.885% purity) as yellow solid. LCMS: Retention time: 0.797 min, [M+H]⁺ calcd. for C₁₄H₉BrN₄ 314.0; found 315.1.

3-(6-methyl-5-(pyrimidin-5-ylethynyl)-1*H*-benzo[*d*]imidazol-2-yl)-5-(trifluoromethyl)benzaldehyde
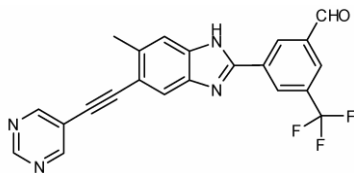
A mixture of 2-bromo-6-methyl-5-(pyrimidin-5-ylethynyl)-1*H*-benzo[*d*]imidazole (150 mg, 479.00 μmol, 1 *eq*), 3-formyl-5-(trifluoromethyl)phenylboronic acid (156.59 mg, 718.50 μmol, 1.5 *eq*), Pd(dppf)Cl₂ (35.05 mg, 47.90 μmol, 0.1 *eq*), and K₂CO₃ (132.40 mg, 958.00 μmol, 2 *eq*) in dioxane (2 mL) and H₂O (1 mL) was degassed and purged with N₂ for 3 times, and then the mixture was stirred at 80°C for 16 hr under N₂ atmosphere. LCMS showed 2-bromo-6-methyl-5-(pyrimidin-5-ylethynyl)-1*H*-benzo[*d*]imidazole was consumed and desired MS was detected. The reaction mixture was quenched by adding water 5 mL, and then extracted with EA (5 mL*3). The combined organic layers were dried over anhydrous Na₂SO₄, filtered and concentrated under reduced pressure to give a residue. The product was used for next step without purification. 3-(6-methyl-5-(pyrimidin-5-ylethynyl)-1*H*-benzo[*d*]imidazol-2-yl)-5-(trifluoromethyl)benzaldehyde (180 mg, crude) was obtained as black brown solid. LCMS: Retention time: 1.026 min, [M+H]⁺ calcd. for C₂₂H₁₃F₃N₄O 407.1; found 407.3.

N,*N*-Dimethyl-1-(3-(6-methyl-5-(pyrimidin-5-ylethynyl)-1*H*-benzo[*d*]imidazol-2-yl)-5-(trifluoromethyl)phenyl)methanamine INS015_032
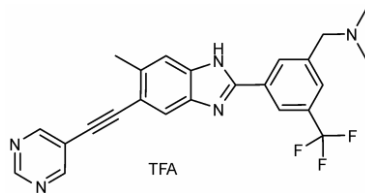
3-(6-methyl-5-(pyrimidin-5-ylethynyl)-1*H*-benzo[*d*]imidazol-2-yl)-5-(trifluoromethyl)benzaldehyde (180 mg, 442.96 μmol, 1 *eq*) and dimethyl amine (72.24 mg, 885.92 μmol, 81.17 μL, 2 *eq*, HCl) were dissolved in MeOH (2 mL). TEA (89.65 mg, 885.92 μmol, 123.31 μL, 2 *eq*) was added to the mixture. The mixture was stirred at 0°C for 30 minutes. NaBH₃CN (55.67 mg, 885.92 μmol, 2 *eq*) and AcOH (79.80 mg, 1.33 mmol, 76.00 μL, 3 *eq*) were added to the mixture. The mixture was stirred at 0°C for another 30 minutes. LCMS showed 3-(6-methyl-5-(pyrimidin-5-ylethynyl)-1*H*-benzo[*d*]imidazol-2-yl)-5-(trifluoromethyl)benzaldehyde was consumed and desired MS was detected. The mixture was quenched with by adding water (1 mL). The mixture was purified by Prep-HPLC (column: Luna C18 150*25 5u; mobile phase: [water (0.075%TFA)-ACN]; B%: 18%-48%, 9min). *N,N*-Dimethyl-1-(3-(6-methyl-5-(pyrimidin-5-ylethynyl)-1*H*-benzo[*d*]imidazol-2-yl)-5-(trifluoromethyl)phenyl)methanamine (34.14 mg, 61.21 μmol, 13.82% yield, 98.517% purity, TFA) was obtained as yellow solid. ¹H-NMR (400 MHz, METHANOL-d4) ppm= 9.15 - 9.02 (m, 1H), 9.00 - 8.85 (m, 2H), 8.58 - 8.50 (m, 2H), 8.11 (s, 1H), 7.93 -7.75 (m, 1H), 7.67 - 7.56 (m, 1H), 4.59 - 4.53 (m, 2H), 3.01 - 2.92 (m, 6H), 2.71 - 2.58 (m, 3H). LCMS: Retention time: 0.797 min, [M+H]⁺ calcd. for $C_{24}H_{20}F_3N_5$ 436.2; found 436.2.

Chemical Synthesis of INS015_036

The first four synthesis steps of INS015_036 was performed by adapting synthetic methods from Hirst *et al.*[4]
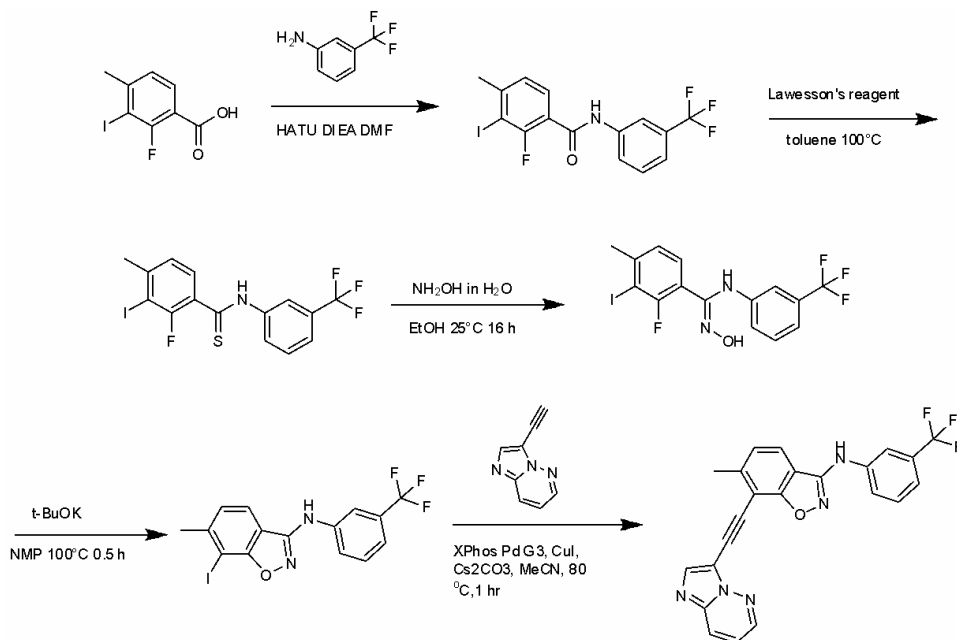
Scheme 4. Synthetic route to INS015_036

2-Fluoro-3-iodo-4-methyl-*N*-(3-(trifluoromethyl)phenyl)benzamide
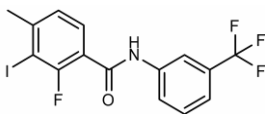
To a solution of 2-fluoro-3-iodo-4-methylbenzoic acid (500 mg, 1.79 mmol, 1 *eq*) in DMF (6 mL) was added HATU (814.68 mg, 2.14 mmol, 1.2 *eq*) and DIEA (692.29 mg, 5.36 mmol, 933.00 uL, 3 *eq*). The mixture was stirred at 25°C for 30 min. Then 3-(trifluoromethyl)aniline (316.46 mg, 1.96 mmol, 245.31 uL, 1.1 *eq*) was added to the mixture. The mixture was stirred at 25°C for 16 h. TLC (PE: EA=5:1, Rf = 0.8) and LCMS showed a major peak with desired mass was detected. To the mixture was added water (10 mL) and stirred for 5 min. The aqueous phase was extracted with ethyl acetate (10 mL*3). The combined organic phase was washed with brine (10 mL), dried with anhydrous Na_2SO_4, filtered and concentrated in vacuum. The residue was purified by flash silica gel chromatography (ISCO®; 80 g SepaFlash® Silica Flash Column, Eluent of 0~30% Ethyl acetate/Petroleum) to afford 2-fluoro-3-iodo-4-methyl-N-(3-(trifluoromethyl)phenyl)benzamide (500 mg, 1.18 mmol, 66.18% yield) as a white solid. LCMS: Retention time: 1.085 min, [M+H]$^+$ calcd. for $C_{15}H_{10}F_4INO$ 424.0; found 424.0.

2-Fluoro-3-iodo-4-methyl-*N*-(3-(trifluoromethyl)phenyl)benzothioamide
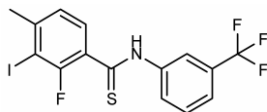
A solution of 2-fluoro-3-iodo-4-methyl-N-(3-(trifluoromethyl)phenyl)benzamide (500 mg, 1.18 mmol, 1 *eq*) in toluene (6 mL) was added LAWESSON'S REAGENT (477.93 mg, 1.18 mmol, 1 *eq*). The mixture was stirred at 100°C for 16 h. LCMS (the mixture was stirred at 100°C for 3 h) showed most of starting material was consumed and desired mass was detected. The reaction mixture was concentrated under reduced pressure to remove toluene. The residue was diluted with DCM 3 mL. The solution was purified by flash silica gel chromatography (ISCO®; 40 g SepaFlash® Silica Flash Column, Eluent of 0~40% Ethyl acetate/Petroleum ether gradient @ 40 mL/min) to afford 2-fluoro-3-iodo-4-methyl-*N*-(3-(trifluoromethyl)phenyl)benzothioamide (400 mg, 901.62 µmol, 76.30% yield, 99% purity) as a yellow solid. ^{1}H-NMR (400MHz, DMSO-d_6) ppm = 12.30 (s, 1H), 8.45 (s, 1H), 8.16 (d, *J*=8.0 Hz, 1H), 7.74 - 7.63 (m, 2H), 7.52 (t, *J*=7.6 Hz, 1H), 7.27 (d, *J*=7.6 Hz, 1H), 2.48 (s, 3H). LCMS: Retention time: 1.134 min, [M+H]$^+$ calcd. for $C_{15}H_{10}F_4INS$ 440.0; found 440.0.

2-Fluoro-*N*′-hydroxy-3-iodo-4-methyl-*N*-(3-(trifluoromethyl)phenyl)benzimidamide
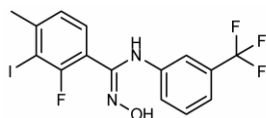
To a solution of 2-fluoro-3-iodo-4-methyl-*N*-(3-(trifluoromethyl)phenyl)benzothioamide (400

mg, 910.73 µmol, 1 *eq*) in EtOH (5 mL) was added NH$_2$OH·HCl (2.53 g, 18.21 mmol, 50% purity, 20 *eq*). The mixture was stirred at 25°C for 2 h. LCMS showed a major peak with desired mass was detected. The mixture was diluted with ACN (1 mL). The solution was purified by reversed-phase column (0.1% NH$_3$•H$_2$O) to give 2-fluoro-*N'*-hydroxy-3-iodo-4-methyl-*N*-(3-(trifluoromethyl)phenyl)benzimidamide (350 mg, 772.44 µmol, 84.82% yield, 96.7% purity) as a yellow solid. ^{1}H-NMR (400 MHz, DMSO-*d*$_6$) ppm = 10.75 (s, 1H), 8.90 (s, 1H), 7.38 (t, *J*=7.6 Hz, 1H), 7.30 - 7.20 (m, 2H), 7.11 (br d, *J*=8.0 Hz, 1H), 6.96 - 6.86 (m, 2H), 2.42 (s, 3H). LCMS: Retention time: 1.010 min, [M+H]$^+$ calcd. for C$_{15}$H$_{11}$F$_4$IN$_2$O 439.0; found 439.0.

7-Iodo-6-methyl-*N*-(3-(trifluoromethyl)phenyl)benzo[*d*]isoxazol-3-amine
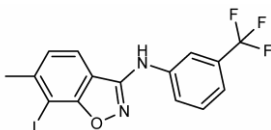
To a solution of 2-fluoro-*N'*-hydroxy-3-iodo-4-methyl-*N*-(3-(trifluoromethyl)phenyl)benzimidamide (160 mg, 365.17 µmol, 1 *eq*) in NMP (5 mL) was added *t*-BuOK (45.07 mg, 401.68 µmol, 1.1 *eq*). The mixture was stirred at 100°C for 0.5 h. LCMS showed a major peak with desired mass. The reaction mixture was poured into water (20 mL) and extracted with ethyl acetate (8 mL*3). The combined organic phase was concentrated in vacuum to give a residue. The residue was purified by flash silica gel chromatography (ISCO®; 40 g SepaFlash® Silica Flash Column, Eluent of 0~40% Ethyl acetate/Petroleum ether gradient @ 40 mL/min) to afford 7-iodo-6-methyl-*N*-(3-(trifluoromethyl)phenyl)benzo[*d*]isoxazol-3-amine (130 mg, 301.57 µmol, 82.58% yield, 97% purity) as a yellow solid.. ^{1}H-NMR (400MHz, DMSO-*d*$_6$) ppm = 9.96 (s, 1H), 8.09 (s, 1H), 7.98 (d, *J*=8.0 Hz, 1H), 7.90 (br d, *J*=8.4 Hz, 1H), 7.62 (t, *J*=7.6 Hz, 1H), 7.38 - 7.31 (m, 2H), 2.55 (s, 3H). LCMS: Retention time: 1.144 min, [M+H]$^+$ calcd. for C$_{15}$H$_{10}$F$_3$IN$_2$O 419.0; found 419.0

7-(Imidazo[1,2-*b*]pyridazin-3-ylethynyl)-6-methyl-*N*-(3-(trifluoromethyl)phenyl)benzo[*d*]isoxazol-3-amine (INS015_036)
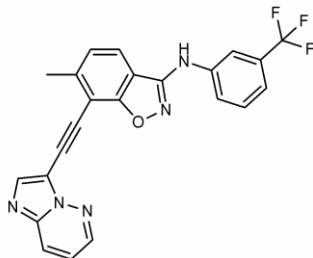
A mixture of 7-iodo-6-methyl-*N*-(3-(trifluoromethyl)phenyl)benzo[*d*]isoxazol-3-amine (101 mg, 241.54 µmol, 1 *eq*), XPhos Pd G3 (122.67 mg, 144.92 µmol, 0.6 *eq*), Cs$_2$CO$_3$ (204.62 mg, 628.00 µmol, 2.6 *eq*) and CuI (23.00 mg, 120.77 µmol, 0.5 *eq*) in anhydrous ACN (2.5 mL). 3-Ethynylimidazo[1,2-*b*]pyridazine (55.32 mg, 386.46 µmol, 1.6 *eq*) was then added and the

reaction mixture was stirred for 2 h at 80°C under nitrogen atmosphere. LCMS showed desired mass. Reaction mixture was poured into water (20 mL), extracted with ethyl acetate (8 mL*3). The combined organic phase was concentrated in vacuum to give a residue. The residue was purified by flash silica gel chromatography (ISCO®; 24 g SepaFlash® Silica Flash Column, Eluent of 0~100% Ethyl acetate/Petroleum ether gradient @ 40 mL/min). LCMS and HPLC showed the purity about 75%. After concentration, the residue was purified by prep-HPLC (neutral condition) and lyophilization to afford 7-(Imidazo[1,2-*b*]pyridazin-3-ylethynyl)-6-methyl-*N*-(3-(trifluoromethyl)phenyl)benzo[*d*]isoxazol-3-amine (14.11 mg, 32.56 μmol, 13.48% yield, 100% purity) as yellow solid. ^{1}H-NMR (400 MHz, DMSO-d_6) ppm= 10.04 (s, 1H), 8.79 - 8.71 (m, 1H), 8.31 - 8.26 (m, 2H), 8.12 (s, 1H), 8.07 (d, J=8.4 Hz, 1H), 7.92 (br d, J=8.4 Hz, 1H), 7.63 (t, J=8.0 Hz, 1H), 7.46 - 7.39 (m, 2H), 7.35 (br d, J=7.6 Hz, 1H), 2.70 (s, 3H). LCMS: Retention time: 1.051 min, [M+H]$^+$ calcd. for $C_{23}H_{14}F_3N_5O$ 434.1; found 434.2.

Chemical Synthesis of INS015_037

The first two synthesis steps of INS015_037 was performed by using synthetic methods from Ashweek *et al*.[5], Pan *et al*.[6]
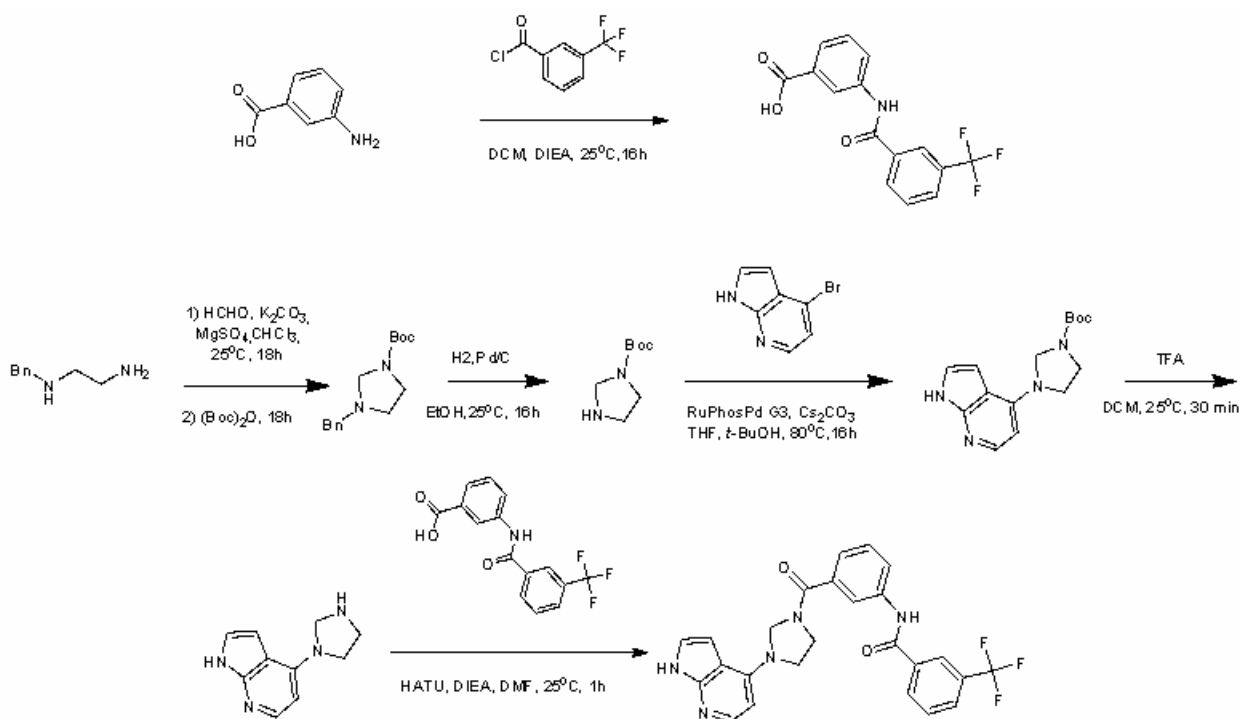
Scheme 5. Synthetic route to INS015_037

3-(3-(trifluoromethyl)benzamido)benzoic acid
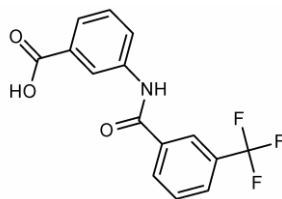
To a solution of 3-aminobenzoic acid (1.17 g, 5.61 m mol, 829.48 u L, 1 *eq*) in DCM (5 mL) 3-(trifluoromethyl)benzoyl

chloride (1 g, 7.29 mmol, 1.3 *eq*) was added at 0°C. And then was added DIEA (3.62 g, 28.04 mmol, 4.88 mL, 5 *eq*). The mixture was stirred at 25°C for 16 h. LCMS showed the 3-aminobenzoic acid was consumed and the desired MS (M+1, 310.1) was detected. The reaction mixture was concentrated under vacuum, and was poured into H_2O (30 mL) and extracted with MTBE (15 mL*3). And then was added citric acid to pH=3. The mixture was filtered and concentrated to give a white solid. The residue was used into the next step without purification. 3-(3-(trifluoromethyl)benzamido)benzoic acid (1.3 g, 4.20 mmol, 74.97% yield) as a white solid. ^{1}H-NMR (400MHz, DMSO-d_6) ppm = 10.72 - 10.45 (m, 1H), 8.51 - 8.12 (m, 4H), 8.07 (td, *J*=1.1, 7.1 Hz, 1H), 7.97 (br d, *J*=7.7Hz, 1H), 7.83 - 7.67 (m, 2H), 7.59 - 7.44 (m, 1H). LCMS: Retention time: 0.918 min, $[M+H]^+$ calcd. for $C_{15}H_{10}F_3NO_3$ 310.0; found 310.1.

Tert-butyl 3-benzylimidazolidine-1-carboxylate
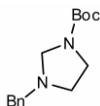
To a solution of N^1-benzylethane-1,2-diamine (2 g, 13.31 mmol, 2.00 mL, 1 *eq*) and $MgSO_4$ (6.41 g, 53.25 mmol, 4 *eq*) , K_2CO_3 (5.52 g, 39.94 mmol, 3 *eq*) and PARAFORMALDEHYDE (400 mg) in $CHCl_3$ (50 mL) was stirred at 25°C for 18 h. (Boc)$_2$O (2.91 g, 13.31 mmol, 3.06 mL, 1 *eq*) was added and the mixture was stirred at 25°C for a further 18 h. LCMS showed the starting material was consumed and the desired MS (M+1,263.3) was detected. The reaction mixture was poured into H_2O (40 mL) and extracted with EA (20 mL*3). The combined organic layer was washed with brine (10 mL), dried over Na_2SO_4, filtered and concentrated. The residue was purified by flash silica gel chromatography (ISCO®; 20 g SepaFlash® Silica Flash Column, Eluent of 0~30% Ethyl acetate/Petroleum ether gradient @ 35 m L/min) to give a spot (R f=0.4) as a white oil. *Tert*-butyl 3-benzylimidazolidine-1-carboxylate (2.8 g, 10.67 mmol, 80.16% yield) as a white oil was obtained. ^{1}H-NMR (400 MHz, CHLOROFORM-d) δ= 7.38 - 7.28 (m, 4H), 3.99 (br d, *J*=19.6 Hz, 2H), 3.64 (s, 2H), 3.43 (td, *J*=6.2, 18.6Hz, 2H), 2.83 (t, *J*=6.4 Hz, 2H), 1.45 (s, 9H). LCMS: Retention time: 1.025 min, $[M+H]^+$ calcd. for $C_{15}H_{22}N_2O_2$ 263.2; found 263.3

Tert-butyl imidazolidine-1-carboxylate

To a solution of *tert*-butyl 3-benzylimidazolidine-1-carboxylate (1 g, 3.81 mmol, 1 *eq*) in EtOH (5 mL) was added Pd/C (200 mg, 20% purity). The mixture was stirred at 25°C for 16 h under H_2 (15 psi). TLC (PE:EA=1:1) showed that most of starting material (R f=0.5) was consumed and a large new spot (R f=0.3) was formed. The reaction mixture was filtered and concentrated. The residue was used into the next step without purification. The crude product *tert*-butyl imidazolidine-1-carboxylate (600 mg, 3.48 m mol, 91.40% yield) as a white oil was obtained. ^{1}H-NMR (400 MHz, CHLOROFORM-d) ppm = 4.32 - 4.06 (m, 2H), 3.26 (br s, 2H), 3.20 - 3.05 (m, 2H), 2.05 - 2.00 (m, 1H), 1.49- 1.44 (m, 9H).

Tert-butyl 3-(1*H*-pyrrolo[2,3-*b*]pyridin-4-yl)imidazolidine-1-carboxylate
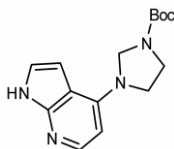
A mixture of *tert*-butyl imidazolidine-1-carboxylate (300 mg, 1.74 mmol, 1 *eq*), 4-bromo-1*H*-pyrrolo[2,3-*b*]pyridine (514.82 mg, 2.61 mmol, 1.5 *eq*) , Cs_2CO_3 (1.14 g, 3.48 mmol, 2 *eq*) , RuPhos Pd G3 (72.84 mg, 87.10 µmol, 0.05 *eq*) in THF (3 mL) and *t*-BuOH (3 mL) was degassed and purged with N_2 for 3 times, and then the mixture was stirred at 80°C for 16 h under N_2 atmosphere. LCMS showed the starting material was consumed and the desired MS (M+1,289.3) was detected. The reaction mixture was poured into H_2O (30 mL) and extracted with EA (20 mL*3). The combined organic layer was washed with brine (10 mL), dried over Na_2SO_4, filtered and concentrated. The residue was purified by flash silica gel chromatography (ISCO®; X g SepaFlash® Silica Flash Column, Eluent of 0~80% Ethyl acetate/Petroleum ether gradient @ 35 mL/min) to give a spot (R f=0.5) as a yellow solid. LCMS showed the yellow solid *tert*-butyl 3-(1*H*-pyrrolo[2,3-*b*]pyridin-4-yl)imidazolidine-1-carboxylate (120 mg, 416.17 µmol, 23.89% yield, N/A purity) the desired MS was detected. LCMS: Retention time: 0.961 min, [M+H]$^+$ calcd. for $C_{15}H_{20}N_4O_2$ 289.2; found 289.3.

4-(Imidazolidin-1-yl)-1*H*-pyrrolo[2,3-*b*]pyridine
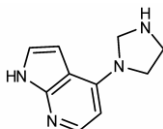
A solution of *tert*-butyl 3-(1*H*-pyrrolo[2,3-*b*]pyridin-4-yl)imidazolidine-1-carboxylate 346.81 µmol, 1 *eq*) in TFA (1 mL) and DCM (3 mL) was stirred at 25°C for 0.5 h. LCMS showed the starting material was consumed and the desired MS (M+1,189.3) was detected. The reaction mixture was concentrated under vacuum. The residue was used into the next step without purification. The crude product 4-(imidazolidin-1-yl)-1*H*-pyrrolo[2,3-*b*]pyridine (90 mg, 297.77 µmol, 85.86% yield, TFA) as a yellow oil was used into the next step without further purification. LCMS: Retention time: 0.704 min, [M+H]$^+$ calcd. for $C_{10}H_{12}N_4$ 189.1; found 189.3.

N-(3-(3-(1*H*-pyrrolo[2,3-*b*]pyridin-4-yl)imidazolidine-1-carbonyl)phenyl)-3-(trifluoromethyl)benzamide (INS015_037)
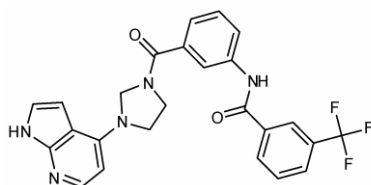
To a solution of 4-(Imidazolidin-1-yl)-1*H*-pyrrolo[2,3-*b*]pyridine (90 mg, 297.77 μmol, 1 *eq*, TFA) and 3-(3-(trifluoromethyl)benzamido)benzoic acid (73.66 mg, 238.21 μmol, 0.8 *eq*) in DMF (2 mL) was added HATU (169.83 mg, 446.65 μmol, 1.5 *eq*) and DIEA (115.45 mg, 893.30 μmol, 155.60 uL, 3 *eq*). The mixture was stirred at 25°C for 1h. LCMS showed the 4-(imidazolidin-1-yl)-1*H*-pyrrolo[2,3-*b*]pyridine was consumed and the desired MS (M+1,480.2) was detected. The reaction mixture was poured into H_2O (40 mL) and extracted with EA (10 mL*3). The combined organic layer was washed with brine (10 mL), dried over Na_2SO_4, filtered and concentrated. The residue was purified by prep-HPLC (column: Phenomenex Synergi C18 150*25*10um; mobile phase: [water (0.1%TFA)-ACN]; B%: 23%-43%, 10min). *N*-(3-(3-(1*H*-pyrrolo[2,3-*b*]pyridin-4-yl)imidazolidine-1-carbonyl)phenyl)-3-(trifluoromethyl)benzamide (19.58 mg, 40.66 μmol, 13.66% yield, 99.571% purity) as a yellow solid was obtained. ^{1}H-NMR (400 MHz, METHANOL-d_4) δ = 8.39 - 8.10 (m, 3H), 7.99 - 7.69 (m, 4H), 7.60 - 7.41 (m, 2H), 7.38 - 7.19 (m, 1H), 7.04 - 6.78 (m, 1H), 6.65 - 6.38 (m, 1H), 5.73 - 5.25 (m, 2H), 4.46 - 3.82 (m, 4H). LCMS: Retention time: 0.814 min, [M+H]$^+$ calcd. for $C_{25}H_{20}F_3N_5O_2$ 480.2; found 480.2

Chemical Synthesis of INS015_038
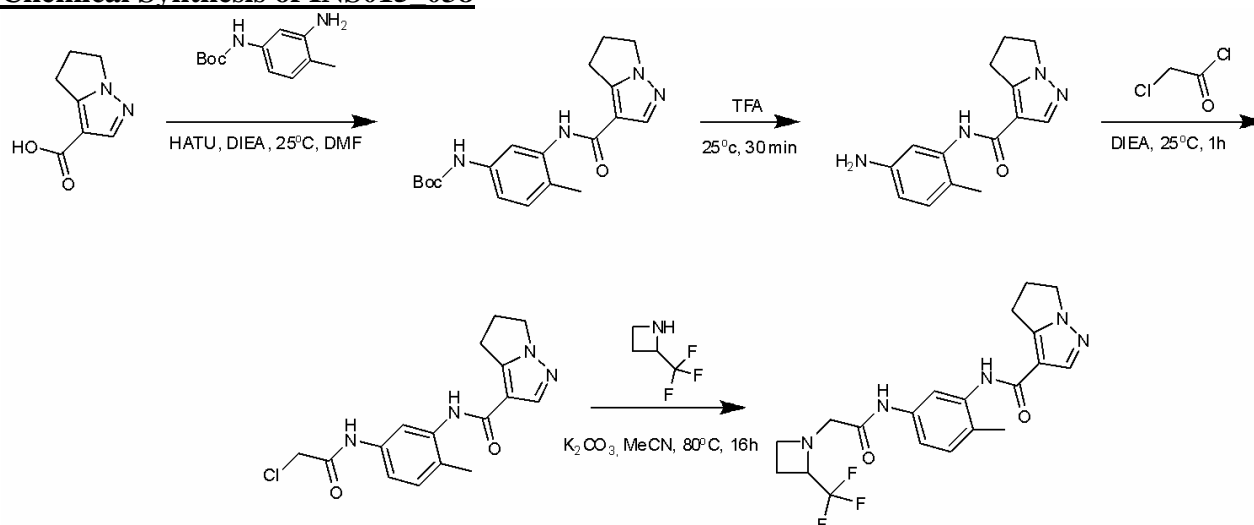
Scheme 6. Synthetic route to INS015_038

Tert-butyl 3-(5,6-dihydro-4*H*-pyrrolo[1,2-*b*]pyrazole-3-carboxamido)-4-

methylphenylcarbamate
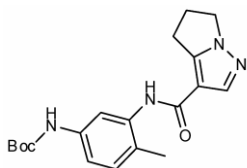
To a solution of 5,6-dihydro-4*H*-pyrrolo[1,2-*b*]pyrazole-3-carboxylic acid (164.28 mg, 1.08 mmol, 1.2 *eq*) in DMF (3 mL) was added HATU (513.17 mg, 1.35 mmol, 1.5 *eq*) stirred at 25°C for 10 min. And then was added DIEA (348.86 mg, 2.70 mmol, 470.16 μL, 3 *eq*) and *tert*-butyl 3-amino-4-methylphenylcarbamate (200 mg, 899.75 μmol, 1 *eq*). The mixture was stirred at 25°C for 1h. LCMS showed the starting material was consumed and the desired MS (M+1,357.2) was detected. Reaction mixture was added to the H_2O (30 mL) with stirred. And then was filtered and concentrated to give a white solid. The residue was used into the next step without purification. The crude product *tert*-butyl 3-(5,6-dihydro-4*H*-pyrrolo[1,2-*b*]pyrazole-3-carboxamido)-4-methylphenylcarbamate (300 mg, 841.71 μmol, 93.55% yield) as a white solid was obtained. ^{1}H-NMR (400 MHz, METHANOL-d_4) ppm = 7.79 - 7.70 (m, 1H), 7.27 - 7.09 (m, 2H), 6.59 - 6.51 (m, 1H), 4.24 - 4.16 (m, 2H), 3.00 - 2.95 (m, 2H), 2.72 - 2.60 (m, 2H), 2.26 - 2.23 (m, 3H), 1.53 - 1.50 (m, 9H). LCMS: Retention time: 0.993 min, [M+H]$^+$ calcd. for $C_{19}H_{24}N_4O_3$ 357.2; found 357.4.

N-(5-amino-2-methylphenyl)-5,6-dihydro-4H-pyrrolo[1,2-b]pyrazole-3-carboxamide
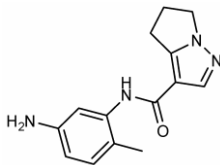
A solution of *tert*-butyl 3-(5,6-dihydro-4*H*-pyrrolo[1,2-*b*]pyrazole-3-carboxamido)-4-methylphenylcarbamate (200 mg, 561.14 μmol, 1 *eq*) in TFA (0.4 mL) and DCM (1.2 mL) was stirred at 25°C for 0.5 h. LCMS showed the starting material was consumed and the desired MS (M+1,257.1) was detected. The reaction mixture was concentrated under vacuum. The crude product N-(5-amino-2-methylphenyl)-5,6-dihydro-4*H*-pyrrolo[1,2-*b*]pyrazole-3-carboxamide (180 mg, 486.06 μmol, 86.62% yield, TFA) as a yellow oil was used into the next step without further purification. LCMS: Retention time: 0.819 min, [M+H]$^+$ calcd. for $C_{14}H_{16}N_4O$ 257.2; found 257.1

N-(5-(2-chloroacetamido)-2-methylphenyl)-5,6-dihydro-4H-pyrrolo[1,2-b]pyrazole-3-carboxamide
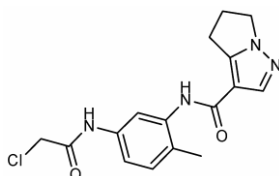
To a solution of *N*-(5-amino-2-methylphenyl)-5,6-dihydro-4*H*-pyrrolo[1,2-*b*]pyrazole-3-carboxamide (180 mg, 486.06 μmol, 1 *eq*, TFA) in DCM (2 mL) was added DIEA (188.46 mg, 1.46 mmol, 253.99 μL, 3 *eq*) and 2-chloroacetyl chloride (109.79 mg, 972.12 μmol, 77.32 μL, 2 *eq*) at 0°C. The mixture was stirred at 25°C for 1 h. LCMS showed the starting material was consumed and the desired MS (M+1,333.1) was detected. The reaction mixture was concentrated under vacuum. The residue was purified by flash silica gel chromatography (ISCO®; 12 g SepaFlash® Silica Flash Column, Eluent of 0~100% Ethyl acetate/Petroleum ether gradient @ 35 mL/min) to give a spot (R f=0.3) as a white solid. LCMS showed the white solid *N*-(5-(2-chloroacetamido)-2-methylphenyl)-5,6-dihydro-4*H*-pyrrolo[1,2-*b*]pyrazole-3-carboxamide (150 mg, 435.64 μmol, 89.63% yield, 96.650% purity) the desired MS was detected. LCMS: Retention time: 0.871 min, [M+H]$^+$ calcd. for $C_{16}H_{17}ClN_4O_2$ 333.1; found 333.1.

N-(2-methyl-5-(2-(2-(trifluoromethyl)azetidin-1-yl)acetamido)phenyl)-5,6-dihydro-4*H*-pyrrolo[1,2-*b*]pyrazole-3-carboxamide (INS015_038)
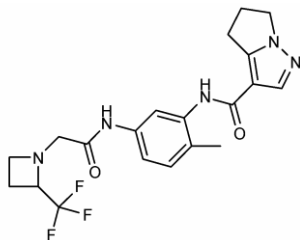
To a solution of *N*-(5-(2-chloroacetamido)-2-methylphenyl)-5,6-dihydro-4*H*-pyrrolo[1,2-*b*]pyrazole-3-carboxamide (100 mg, 300.50 μmol, 1 *eq*) and 2-(trifluoromethyl)azetidine (97.09 mg, 600.99 μmol, 2 *eq*, HCl) in MeCN (1 mL) was added K_2CO_3 (124.59 mg, 901.49 μmol, 3 *eq*) .The mixture was stirred at 80°C for 16 h. LCMS showed the starting material was consumed and desired MS (M+1,422.2) was detected. The reaction mixture was poured into H_2O (10 mL) and extracted with EA (10 mL*3). The combined organic layer was washed with brine (5 mL), dried over Na_2SO_4, filtered and concentrated to give a residue. The residue was purified by prep-HPLC (neutral condition, column: Waters Xbridge 150*25 5u; mobile phase: [Water-ACN]; B%: 35%-62%, 9min). Compound INS015_038 (31.17 mg, 73.41 μmol, 24.43% yield, 99.244% purity) as a yellow solid was obtained. ^{1}H-NMR (400 MHz, METHANOL-d_4) ppm = 7.92 (d, *J* = 2.0 Hz, 1H), 7.38 (dd, *J* = 2.2, 8.2 Hz, 1H), 7.20 (d, *J* = 8.3 Hz, 1H), 6.54(s, 1H), 4.17 (t, *J* = 7.3 Hz, 2H), 4.11 - 3.96 (m, 1H), 3.67 - 3.59 (m, 1H), 3.47 - 3.32 (m, 2H), 3.29 - 3.26 (m, 1H), 2.98 - 2.88(m, 2H), 2.72 - 2.55 (m, 2H), 2.38 - 2.23 (m, 5H). LCMS: Retention time: 0.955min, [M+H]$^+$ calcd. for $C_{20}H_{22}F_3N_5O_2$ 422.2; found 422.2.

Chemical Synthesis of INS015_039
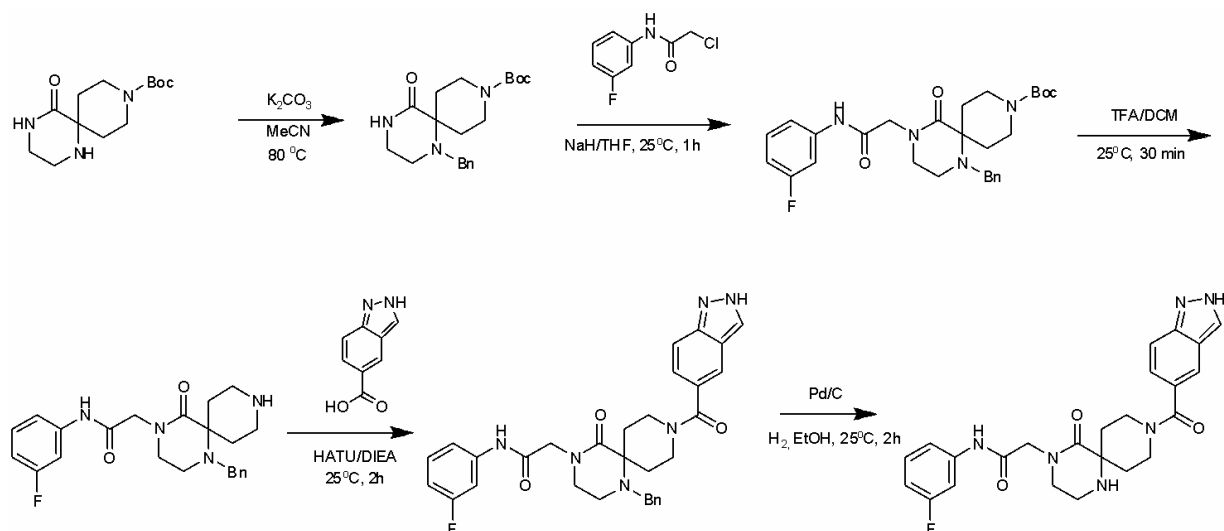
Scheme 7. Synthetic route to INS015_039

Tert-butyl 1-benzyl-5-oxo-1,4,9-triazaspiro[5.5]undecane-9-carboxylate
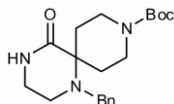
To a solution of *tert*-butyl 5-oxo-1,4,9-triazaspiro[5.5]undecane-9-carboxylate (250 mg, 928.20 μmol, 1 *eq.*) and BnBr (238.13 mg, 1.39 mmol, 165.37 μL, 1.5 *eq.*) in MeCN (5 mL) was added K$_2$CO$_3$ (256.57 mg, 1.86 mmol, 2 *eq.*), the mixture was stirred at 80°C for 1 h. LC-MS showed desired MS was detected. The mixture was filtered and concentrated under reduced pressure to give a residue and purified by column (SiO$_2$, PE:EA =5:1 to 1:2) to give *tert*-butyl 1-benzyl-5-oxo-1,4,9-triazaspiro[5.5]undecane-9-carboxylate (290 mg, 806.76 μmol, 86.92% yield). The white solid was confirmed by ^{1}H-NMR. ^{1}H-NMR (400 MHz, METHANOL-d_4) δ = 7.45 - 7.38 (m, 2H), 7.37 - 7.30 (m, 2H), 7.28 - 7.19 (m, 1H), 3.87 - 3.75 (m, 4H), 3.48 - 3.32 (m, 4H), 2.91 (t, *J* = 6.0 Hz, 2H), 2.03 - 1.91 (m, 2H), 1.67 - 1.60 (m, 2H), 1.47 (s, 9H).

Tert-butyl 1-benzyl-4-(2-(3-fluorophenylamino)-2-oxoethyl)-5-oxo-1,4,9-triazaspiro[5.5]undecane-9-carboxylate
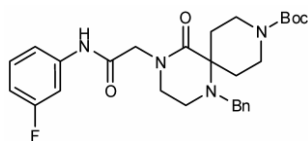
To a solution of 1-benzyl-5-oxo-1,4,9-triazaspiro[5.5]undecane-9-carboxylate (290 mg, 806.76 μmol, 1 *eq.*) in THF (10 mL) was added NaH (64.53 mg, 1.61 mmol, 60% purity, 2 *eq.*) at 0°C, the mixture was stirred at 25°C for 30 min, 2-chloro-*N*-(3-fluorophenyl)acetamide (227.02 mg, 1.21 mmol, 1.5 *eq.*) in THF (5 mL) was added dropwise to the mixture. The mixture was stirred at 25°C for 1 h. LC-MS showed desired MS was detected. The mixture was quenched by NH$_4$Cl

(saturation, 10 mL), diluted with EA (50 mL), washed with brine (10 mL), dried with Na_2SO_4, filtered and concentrated under reduced pressure to give a residue and purified by column (SiO_2, PE:EA = 5:1 to 1:1) to give *tert*-butyl 1-benzyl-4-(2-(3-fluorophenylamino)-2-oxoethyl)-5-oxo-1,4,9-triazaspiro[5.5]undecane-9-carboxylate (300 mg, 587.55 µmol, 72.83% yield) as a white solid. ^{1}H-NMR (400 MHz, METHANOL-d_4) δ = 8.88 - 8.62 (m, 1H), 7.54 - 7.45 (m, 1H), 7.41 - 7.28 (m, 5H), 7.26 - 7.18 (m, 1H), 7.17 - 7.03 (m, 1H), 6.88 - 6.72 (m, 1H), 4.12 - 4.07 (m, 2H), 3.96 - 3.68 (m, 4H), 3.60 - 3.20 (m, 4H), 3.06 - 2.85 (m, 2H), 2.15 - 2.05 (m, 4H), 1.51 - 1.39 (m, 9H).

2-(1-benzyl-5-oxo-1,4,9-triazaspiro[5.5]undecan-4-yl)-*N*-(3-fluorophenyl)acetamide
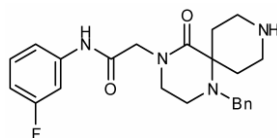
To a solution of *tert*-butyl 1-benzyl-4-(2-(3-fluorophenylamino)-2-oxoethyl)-5-oxo-1,4,9-triazaspiro[5.5]undecane-9-carboxylate (300 mg, 587.55 µmol, 1 *eq.*) in DCM (9 mL) was added TFA (4.62 g, 40.52 mmol, 3 mL, 68.96 *eq.*), the mixture was stirred at 25°C for 0.5 h. LC-MS showed desired MS was detected. The mixture was concentrated under reduced pressure to give 2-(1-benzyl-5-oxo-1,4,9-triazaspiro[5.5]undecan-4-yl)-*N*-(3-fluorophenyl)acetamide (300 mg, 571.97 µmol, 97.35% yield, TFA) as a colorless oil.

2-(1-benzyl-9-(2*H*-indazole-5-carbonyl)-5-oxo-1,4,9-triazaspiro[5.5]undecan-4-yl)-*N*-(3-fluorophenyl)acetamide
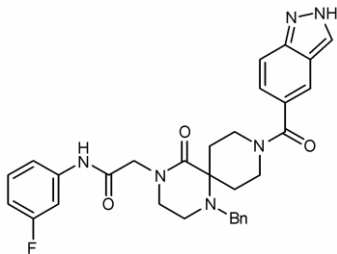
To a solution 2-(1-benzyl-5-oxo-1,4,9-triazaspiro[5.5]undecan-4-yl)-*N*-(3-fluorophenyl)acetamide (300 mg, 571.97 µmol, 1 *eq.*, TFA) and DIEA (221.77 mg, 1.72 mmol, 298.88 µL, 3 *eq.*) in DCM (8 mL) and DMF (2 mL) was added 2H-indazole-5-carboxylic acid (111.29 mg, 686.36 µmol, 1.2 *eq.*) and HATU (260.97 mg, 686.36 µmol, 1.2 *eq.*), the mixture was stirred at 25°C for 2 h. LC-MS showed desired MS was detected. The mixture was concentrated under reduced pressure to give a residue and purified by column (SiO_2, DCM:MeOH = 1:0 to 5:1) to give 2-(1-benzyl-9-(2*H*-indazole-5-carbonyl)-5-oxo-1,4,9-triazaspiro[5.5]undecan-4-yl)-*N*-(3-fluorophenyl)acetamide (250 mg, 450.76 µmol, 78.81% yield) as a colourless oil.

2-(9-(2*H*-indazole-5-carbonyl)-5-oxo-1,4,9-triazaspiro[5.5]undecan-4-yl)-*N*-(3-fluorophenyl)acetamide (INS015_039)
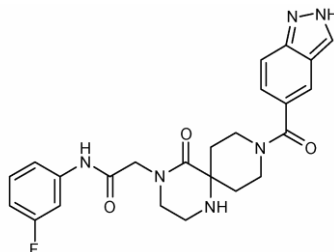
A mixture of 2-(1-benzyl-9-(2*H*-indazole-5-carbonyl)-5-oxo-1,4,9-triazaspiro[5.5]undecan-4-yl)-*N*-(3-fluorophenyl)acetamide (250 mg, 450.76 µmol, 1 *eq.*) and Pd/C (250 mg, 450.76 µmol, 10% purity, 1 *eq.*), in EtOH (10 mL) was degassed and purged with H_2 for 3 times, and then the mixture was stirred at 25°C for 2 h under H_2 (15 psi) atmosphere. LC-MS showed desired MS was detected. The mixture was filtered and concentrated under reduced pressure to give a residue and purified by pre-HPLC (water (10 mM NH_4HCO_3)-ACN) to give 2-(9-(2*H*-indazole-5-carbonyl)-5-oxo-1,4,9-triazaspiro[5.5]undecan-4-yl)-*N*-(3-fluorophenyl)acetamide (35.79 mg, 73.50 µmol, 16.31% yield, 95.387% purity) as a white solid. HPLC: Retention time: 1.675 min. ^{1}H-NMR (400 MHz, METHANOL-d_4) δ = 8.14 (s, 1H), 7.90 (s, 1H), 7.62 (d, *J* = 8.6 Hz, 1H), 7.55 - 7.44 (m, 2H), 7.35 - 7.20(m, 2H), 6.87 - 6.76 (m, 1H), 4.47 - 4.07 (m, 3H), 3.79 - 3.34 (m, 5H), 3.20 - 2.99 (m, 2H), 2.38 - 1.53 (m, 4H). ^{19}F-NMR (400MHz, METHANOL-d4) δ =114.135. ^{13}C-NMR (400MHz, METHANOL-d4) δ= 174.48, 171.53, 167.36, 164.11, 161.69, 140.48, 140.10, 139.99, 134.29, 129.89, 129.80, 128.13, 125.34, 122.37, 119.90, 114.82, 114.80, 110.18, 109.95, 106.63, 106.36, 56.81, 50.66, 50.34, 48.30, 48.09, 47.88, 47.66, 37.06. LCMS: Retention time: 0.800 min, [M+H]$^+$ calcd. for $C_{24}H_{25}FN_6O_3$ 465.2; found 465.2.

Hyperparameters for GENTRL model

1.   **Architecture**: The encoder $q_\phi(z \mid x)$ was a 2-layer recurrent neural network with gated recurrent units (GRU)[7] of hidden size 128. The decoder $p_\theta(x \mid z)$ was a 7-layer stacked dilated convolutions with 128 channels. The latent space z was 50 dimensional with 10 mixture components at each dimension. Core size of a tensor-train m was 30.

2.   **Autoencoder training**: We used multiple molecular properties y for learning the mapping of the chemical space onto a latent manifold. For all training molecules, we calculated MCE-18 and a binary flag MCF indicating if a molecule successfully passed medicinal chemistry filters. For molecules from a Kinase and "negative" dataset, we specified if a molecule was a kinase (for molecules outside these datasets the value was unknown). For known DDR1 inhibitors, we specified pIC_{50}. For each update, we constructed a batch containing 200 molecules: 60 active molecules from DDR1, 60 molecules from ZINC, 20 molecules from Kinase dataset, 20 from the negative dataset and 40 molecules from the patent records. We used an Adam[8] optimizer with a learning rate of 10^{-4} and ran the optimization procedure for 300,000 updates.

3.   **Reinforcement learning**: We trained a model with the REINFORCE algorithm for 2,000 updates with Adam optimizer, learning rate $2 \cdot 10^{-5}$ and a batch size 200. We sampled exploratory batches $z^{explore}$ with probability 0.1 and standard batches $z \sim p_\psi(z)$.

Estimated from 50,000 randomly sampled molecules, 97.8% were valid and 73% were unique.

Comparison of GENTRL with RANC, ATNC, ORGAN

In this section, we compare the proposed GENTRL model with previous methods: reinforced adversarial neural computer (RANC)[9], adversarial threshold neural computer (ATNC)[10] and objective-reinforced generative adversarial network (ORGAN)[11,12]. We compare on two toy reward functions derived from the commonly used penalized logP[13] and penalized quantitative estimation of drug-likeness (QED)[14]

$$\text{plogP}(x) = [\text{logP}(x) - \text{SA}(x) - \text{large\_rings}(x)] \cdot \text{MCF}$$

$$\text{pQED}(x) = [5 \cdot \text{QED}(x) - \text{SA}(x)] \cdot \text{MCF},$$

where logP is a water-octanol partition coefficient[15], SA is a synthetic accessibility score[16], QED[17] is a quantitative estimation of drug-likeness, and large_rings is a number of rings with more than 6 atoms. All properties were calculated with RDKit[18]. MCF is a binary flag indicating if a molecule passes medicinal chemistry filters taken from MOSES[19] platform. The MCF penalty does not allow the model to collapse to long carbon chains, providing more meaningful molecules for comparison.

We trained all models on the training set of the MOSES dataset and generated 30,000 molecules with each model. We recorded the fraction of valid molecules, 10-th percentile and median of the rewards. We optimized the hyperparametrs for all models independently with a random search. We also found it profitable to train a decoder jointly with a latent space on the RL phase of training on these tasks.

Quality of generated molecules for plogP and pQED rewards. The proposed model finds molecules with higher reward than other baselines.

Model	plogP			pQED		
	Valid	Top10%	Median	Valid	Top10%	Median
ORGAN	**0.85**	4.11	2.88	0.9	2.8	2.16
RANC	0.65	1.80	0.42	0.63	2.65	1.83
ATNC	0.81	1.42	-0.08	0.81	2.34	1.57
GENTRL	**0.85**	**7.57**	**6.34**	**0.97**	**3.14**	**2.94**

Comparing results on both metrics, we found that the generated molecules have higher reward values than all the baseline models.
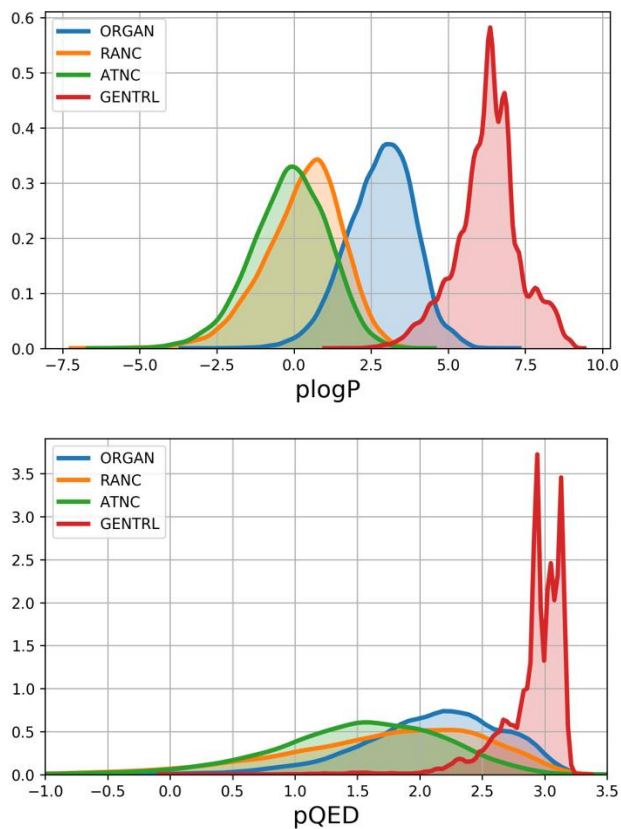
Distribution of rewards (plogP and pQED) for different models. Note that since our model learns a multimodal prior, the distribution curve is also highly multimodal.

Supplementary Tables

Supplementary Table 1. Datasets used for AI-driven DRR1 inhibitors generation and pharmacophore modelling.

№	Title	Reference	Type	Quantity
	1. Sources used to collect data on DDR1 inhibitors			
1	ZINC	[20]	Database	904,801
	2. Sources used to collect data on DDR1 inhibitors			
1	Wang Z, et al. (2017)	[21]	Article	7
2	Wang Z, et al. (2016)	[22]	Article	4
3	Gao M, et al. (2013)	[23]	Article	30
4	Kim H, et al. (2013)	[24]	Article	3
5	Elkamhawy A, et al. (2016)	[25]	Article	1
6	Liu L, et al. (2017)	[26]	Article	19
7	Canning P, et al. (2014)	[27]	Article	6
8	Li Y, et al. (2015)	[28]	Article	1
9	Fraser C, et al. (2016)	[29]	Article	3
10	Wang Q, et al. (2016)	[30]	Article	1
11	Murray C, et al. (2015)	[31]	Article	2
12	WO2013161851	[32]	Pat.App.	394
13	WO2013161853(A1)	[33]	Pat.App.	334
14	WO2017005583(A1)	[34]	Pat.App.	169
15	WO2017137334(A1)	[35]	Pat.App.	168
16	WO2017038873(A1)	[36]	Pat.App.	75
17	WO2017038871(A1)	[37]	Pat.App.	47
18	WO2016064970(A1)	[38]	Pat.App.	28
19	WO2015004481(A1)	[39]	Pat.App.	10
20	WO2013100632(A1)	[40]	Pat.App.	5
21	Integrity	[41]	Database	47
22	ChemBL	[42]	Database	16
	3. Sources used to collect data on common kinase inhibitors (positive set)			
1	Integrity	[41]	Database	16,417
2	ChemBL	[42]	Database	6,961
	4. Sources used to collect data on molecules acting on non-kinase targets (negative set)			
1	Integrity	[41]	Database	9,890
2	ChemBL	[42]	Database	6,802
	5. Sources used to collect data on patent data for biologically active molecules that have been claimed by pharmaceutical companies			
1	Integrity	[41]	Database	17,000
	6. Sources used to collect data on 3D structures for DDR1 inhibitors			
1	PDB	[43]	Database	18

Supplementary Table 2. Prioritization process (for more details see Methods)

№	Step	Key comments	Compounds
1	Phys-Chem Filters	Compounds meeting pre-defined ranges of molecular descriptors.	12147
2	MCFs	Compounds containing alert substructures which are usually undesirable in medicinal chemistry.	7912
3	Clustering/Diversity	Tanimoto-based clustering followed by diversity maximization per each cluster, removing (≤5 cmpds in cluster) of similar compounds and chemical space normalization.	5542
4	Similarity	Tanimoto-based similarity towards compounds available within vendors` (MolPort, ZINC) stocks (threshold ≤0.5).	4642
5	General Kinase SOM	Structures which were classified as kinase inhibitors vs. non-kinase chemistry.	2570
6	Specific Kinase SOM	Structures were selected from the neurons containing at least one ref. DDR1 inhibitor to overcome the bias.	1951
7	Pharmacophore searching	Structures which successfully passed the pharmacophore modeling	848
8	Sammon mapping	Compounds were subjected to Sammon learning procedure to randomly select the final set of structures.	40

Supplementary Table 3. RMSD values (Å) for 40 molecules matching the pharmacophore hypothesis

ID	3c_rmsd*	4c_rmsd	5c_rmsd	ID	3c_rmsd	4c_rmsd	5c_rmsd
INS015_001	0.21	1.18	NA	INS015_021	0.15	0.97	1.30
INS015_002	0.18	0.68	1.25	INS015_022	0.63	1.12	NA
INS015_003	0.31	0.80	0.96	INS015_023	1.07	1.00	NA
INS015_004	0.44	1.01	1.37	INS015_024	0.89	1.05	NA
INS015_005	0.18	0.79	1.23	INS015_025	0.49	0.65	1.52
INS015_006	0.50	1.07	NA	INS015_026	0.06	0.99	0.91
INS015_007	0.39	0.82	1.34	INS015_027	0.55	0.92	NA
INS015_008	0.32	0.67	0.96	INS015_028	0.75	0.98	NA
INS015_009	0.24	0.66	0.83	INS015_029	0.41	0.52	1.41
INS015_010	0.34	0.87	1.08	INS015_030	0.60	1.32	NA
INS015_011	0.09	0.43	NA	INS015_031	0.15	0.35	0.72
INS015_012	0.47	1.05	1.19	INS015_032	0.40	0.77	0.99
INS015_013	0.07	0.13	NA	INS015_033	0.17	0.41	0.79
INS015_014	0.35	0.77	1.35	INS015_034	0.19	0.50	0.66
INS015_015	0.14	0.84	0.94	INS015_035	0.18	0.84	0.82
INS015_016	0.21	0.75	0.83	INS015_036	0.25	0.55	0.79
INS015_017	0.48	0.76	1.08	INS015_037	0.94	1.38	NA
INS015_018	0.45	0.70	0.99	INS015_038	0.43	0.85	1.28
INS015_019	0.18	0.38	NA	INS015_039	0.31	0.57	1.54
INS015_020	0.38	0.92	1.01	INS015_040	0.72	0.65	1.29

3c_rmsd – RMSD value for a compound matched 3-centered pharmacophore hypothesis
4c_rmsd – RMSD value for a compound matched 4-centered pharmacophore hypothesis
5c_rmsd – RMSD value for a compound matched 5-centered pharmacophore hypothesis
NA – a compound did not match pharmacophore hypothesis

Supplementary Table 4. IP status of generated molecules

ID	Results of Similarity Searching, %*	Markush***	Example***	MolPort***	Zinc****
INS015_001	<70	0	0	0.27	0.27
INS015_002	70 - 74 (4)	0	0	0.24	0.24
INS015_003	<70	0	0	0.21	0.29
INS015_004	85-89 (2); 80-84 (10); 75-79 (29); 70-74 (57)	0	0	0.39	0.39
INS015_005	75-79 (9); 70-74 (20)	0	0	0.27	0.27
INS015_006	80-84 (9); 75-79 (22); 70-74 (62)	0	0	0.39	0.39
INS015_007	95-98 (1); 90-94 (10); 85-89 (8); 80-84 (17); 75-79 (61); 70-74 (336)	0	0	0.36	0.36
INS015_008	<70	0	0	0.21	0.19
INS015_009	75-70 (16)	0	0	0.38	0.38
INS015_010	71-70 (4)	0	0	0.38	0.38
INS015_011	<70	0	0	0.22	0.22
INS015_012	75-79 (2); 70-74 (11)	0	0	0.33	0.33
INS015_013	<70	0	0	0.21	0.21
INS015_014	80-84 (1); 75-79 (16); 70-74 (149)	0	0	0.25	0.28
INS015_015	<70	0	0	0.38	0.4
INS015_016	<70	0	0	0.22	0.27
INS015_017	<70	0	0	0.25	0.24
INS015_018	75-79 (5); 70-74 (85)	0	0	0.31	0.31
INS015_019	<70	0	0	0.23	0.25
INS015_020	70-71 (6)	0	0	0.22	0.22
INS015_021	<70	0	0	0.28	0.28
INS015_022	<70	0	0	0.33	0.34
INS015_023	80-84 (1); 75-79 (10); 70-74 (46)	0	0	0.36	0.37
INS015_024	70-74 (6)	0	0	0.36	0.39
INS015_025	70-74 (9)	0	0	0.41	0.41
INS015_026	70-74 (3)	0	0	0.31	0.31
INS015_027	75-79 (1); 70-74 (25)	0	0	0.33	0.33
INS015_028	70-74 (27)	0	0	0.48	0.46
INS015_029	<70	0	0	0.31	0.31
INS015_030	75-79 (21); 70-74 (222)	0	0	0.41	0.43
INS015_031	70-74 (2)	0	0	0.25	0.25
INS015_032	75-79 (1); 70-74 (45)	0	0	0.23	0.23
INS015_033	70-74 (10)	0	0	0.2	0.21
INS015_034	<70	0	0	0.19	0.19
INS015_035	70-74 (1)	1	0	0.36	0.36
INS015_036	70-74 (1)	0	0	0.29	0.23
INS015_037	75-79 (37); 70-74 (426)	0	0	0.48	0.48
INS015_038	80-84 (1); 75-79 (3); 70-74 (37)	0	0	0.43	0.43
INS015_039	75-79 (6); 70-74 (106)	0	0	0.33	0.33
INS015_040	70-74 (10)	0	0	0.38	0.38

* Similarity searching in SciFinder database was carried out. The ranges of structural similarity are presented. The number of similar compounds in the certain range is enclosed in parenthesis. In the absence of molecules with similarity >70% it was displayed as <70%.
** The number of patent Markush structures a compound match.

*** The presence of a compound among the examples in patents
**** The maximal similarity between generated compound and molecules from MolPort and ZINC databases

Supplementary Table 5. Microsomal stability results summary for compounds 1 (n=6) and 2 (n=6).

Testosterone (n=6 in HLM and DLM, n=2 in RLM, n=4 in MLM), Diclofenac (n=6) and Propafenone (n=6 in HLM and DLM, n=3 in RLM, n=4 in MLM) were served as controls.

Sample Name	HLM 0.5					
	R^2	$T_{1/2}$ (min)	$CL_{int(mic)}$ (μL/min/mg)	$CL_{int(liver)}$ (mL/min/kg)	Remaining (T=60min)	Remaining (*NCF=60min)
Compound 1	0.1518	>145	<9.6	<8.6	96.80%	105.20%
Compound 2	0.8843	12.8	108.2	97.3	2.80%	88.70%
Testosterone	0.9998	15.6	88.6	79.7	6.90%	82.40%
Diclofenac	0.9964	10.7	129.9	116.9	1.90%	87.90%
Propafenone	0.9868	8.3	166.3	149.7	0.70%	103.30%
Sample Name	RLM 0.5					
	R^2	$T_{1/2}$ (min)	$CL_{int(mic)}$ (μL/min/mg)	$CL_{int(liver)}$ (mL/min/kg)	Remaining (T=60min)	Remaining (*NCF=60min)
Compound 1	0.8652	>145	<9.6	<17.3	74.60%	88.80%
Compound 2	0.9272	17.2	80.6	145	7.10%	88.30%
Testosterone	1	1.1	1309.5	2357.2	0.00%	86.80%
Diclofenac	0.9934	23.6	58.7	105.7	17.50%	89.80%
Propafenone	0.9897	1.5	895.7	1612.3	0.40%	87.10%
Sample Name	DLM 0.5					
	R^2	$T_{1/2}$ (min)	$CL_{int(mic)}$ (μL/min/mg)	$CL_{int(liver)}$ (mL/min/kg)	Remaining (T=60min)	Remaining (*NCF=60min)
Compound 1	0.9698	65.122	21.283	30.64752	51.00%	92.40%
Compound 2	0.9614	8.5	162.6	234.1	0.60%	99.90%
Testosterone	0.9988	17.5	79.4	114.3	9.50%	88.60%
Diclofenac	0.6896	>145	<9.6	<13.8	85.70%	96.90%
Propafenone	0.9458	8.2	169.8	244.5	0.60%	101.60%
Sample Name	MLM 0.5					
	R^2	$T_{1/2}$ (min)	$CL_{int(mic)}$ (μL/min/mg)	$CL_{int(liver)}$ (mL/min/kg)	Remaining (T=60min)	Remaining (*NCF=60min)
Compound 1	0.8361	>145	<9.6	<38.0	78.70%	99.70%
Compound 2	0.9226	15.1	91.8	363.6	5.30%	86.00%
Testosterone	0.9975	3.4	409.4	1621.1	0.00%	75.40%
Diclofenac	0.9347	58.5	23.7	93.8	50.20%	91.80%
Propafenone	0.9654	2.9	472.2	1869.7	0.40%	88.00%

*NCF: the abbreviation of no co-factor. No NADPH regenerating system is added into NCF sample (replaced by buffer) during the 60 min-incubation, if the NCF remaining is less than 60%, then Non-NADPH dependent occurs
R^2 is the correlation coefficient of the linear regression for the determination of kinetic constant.

$T_{1/2}$ is half life and $CL_{int(mic)}$ is the intrinsic clearance

$CL_{int(mic)} = 0.693/\text{half life/mg microsome protein per mL}$

$CL_{int(liver)} = CL_{int(mic)} * \text{mg microsomal protein/g liver weight} * \text{g liver weight/kg body weight}$

mg microsomal protein / g liver weight: 45 mg/g for 5 species

Liver weight: 88 g/kg, 40g/kg, 32 g/kg, 30 g/kg and 20 g/kg for mouse, rat, dog, monkey and human.

Supplementary Table 6. Buffer stability results for compound 2.

Data reported as mean, n=2 independent samples were used.

Buffer	Compound ID	Final Concentration	Incubation Time (min)	Analyte Peak Area	IS1(Labetalol) Peak Area	IS2(tolbutamide) Peak Area	PAR	Mean (n=2)	% Remaining	% CV
50 mM phosphate buffer, pH 7.4	Compund 2	10 µM	0	1929905	6033253	2370965	0.81	0.82	100.00	
				1897033	5767532	2312289	0.82			
			120	2014536	5997906	2443293	0.82	0.82	100.65	0.35
				1993993	6362157	2430358	0.82			
			240	1959004	6227270	2506330	0.78	0.79	97.05	1.98
				1927245	6082958	2395458	0.80			
			360	2143657	6248040	2497383	0.86	0.82	100.07	7.03
				1900890	6111501	2446086	0.78			
			1440	2030579	5699870	2293050	0.89	0.88	107.23	1.61
				2002559	5723871	2309804	0.87			
8 mM MOPS pH 7.0, 0.2 mM EDTA, pH7.0	Compound 2	10 µM	0	1562845	6211850	2369241	0.66	0.64	100.00	
				1603841	6233248	2576393	0.62			
			120	1573942	6034502	2394434	0.66	0.65	101.41	1.60
				1627098	6275956	2530982	0.64			
			240	1649971	6239927	2452320	0.67	0.67	103.81	1.62
				1542025	6088004	2343021	0.66			
			360	1694387	6053386	2438806	0.69	0.68	106.18	3.11

				163193 2	6188803	2448220	0.67			
			1440	170903 6	6175205	2435887	0.70	0.68	105.37	5.7 6
				162530 3	6005431	2502723	0.65			

Supplementary Table 7. CYP inhibition results for compound 1 and 2.

Compound ID	IC_{50} (μM)				
	CYP1A2	CYP2C9	CYP2C19	CYP2D6	CYP3A4-M
Compound 1	7.36	>50	>50	>50	>50
Compound 2	10.6	2.70	6.56	6.97	7.36

Supplementary Table 8. Summary of physiochemical properties of compound 1 and 2.

ID	MW	logP*	TPSA	HBD	HBA	IC$_{50}$ (nM)	N	LE**
1	433.39	4.94	68.25	1	4	10	32	0.35
2	479.45	3.88	81.33	2	3	21	35	0.31

*Predicted
**LE = 1.4*(pIC50)/N, N – number of heavy atoms

Supplementary Table 9. Mouse PK study results (IV administration).

Full mouse PK study results for 10mg/kg compound **1** IV administration. Formulation: 5 mg/mL in NMP/PEG400/H20=1 : 7 : 2, clear solution.

IV							
IV Time (h)	**M01**	**M02**	**M03**	**Mean IV**		**SD**	**CV (%)**
0.0830	2140	2810	2120	**2357**	±	**393**	**16.7**
0.250	2000	2090	1760	**1950**	±	**171**	**8.75**
0.500	1630	2070	1430	**1710**	±	**327**	**19.1**
1.00	1400	1490	1170	**1353**	±	**165**	**12.2**
2.00	893	886	862	**880**	±	**16.3**	**1.85**
4.00	417	456	528	**467**	±	**56.3**	**12.1**
8.00	244	261	261	**255**	±	**9.81**	**3.84**
24.0	9.37	12.5	11.3	**11.1**	±	**1.58**	**14.3**
PK Parameters	**M01**	**M02**	**M03**	**Mean IV**		**SD**	**CV (%)**
Rsq_adj	0.993	0.996	0.999	--	±	--	--
No. points used for $T_{1/2}$	3.00	3.00	3.00	**3.00**	±	--	--
C_0 (ng/mL)	2213	3255	2325	**2598**	±	**572**	**22.0**
$T_{1/2}$ (h)	3.58	3.79	3.58	**3.65**	±	**0.123**	**3.36**
Vd_{ss} (L/kg)	5.24	4.85	5.55	**5.21**	±	**0.349**	**6.69**
Cl (mL/min/kg)	19.8	18.0	19.3	**19.1**	±	**0.970**	**5.09**
T_{last} (h)	24.0	24.0	24.0	**24.0**	±	--	--
AUC_{0-last} (ng.h/mL)	6555	7223	6712	**6830**	±	**349**	**5.11**
AUC_{0-inf} (ng.h/mL)	6603	7291	6770	**6888**	±	**359**	**5.21**
MRT_{0-last} (h)	4.22	4.26	4.56	**4.35**	±	**0.187**	**4.29**
MRT_{0-inf} (h)	4.41	4.50	4.78	**4.56**	±	**0.193**	**4.23**
AUC_{Extra} (%)	0.732	0.938	0.863	**0.845**	±	**0.104**	**12.3**
$AUMC_{Extra}$ (%)	4.85	6.15	5.27	**5.42**	±	**0.663**	**12.2**

ND = Not determined (Parameters not determined due to inadequately defined terminal elimination phase)
BQL = Below the lower limit of quantitation (LLOQ)
If the adjusted rsq (linear regression coefficient of the concentration value on the terminal phase) is less than 0.9, T1/2 might not be accurately estimated.
If the % AUC_{Extra} > 20%, AUC_{0-inf}, Cl, MRT_{0-inf} and Vd_{ss} might not be accurately estimated.
If the % $AUMC_{Extra}$ > 20%, MRT_{0-inf} and Vd_{ss} might not be accurately estimated.
The adjusted linear regression coefficient of the concentration value on the terminal phase is less than 0.9, $T_{1/2}$ might not be accurately estimated.
a: Bioavailability (%) was calculated using AUC_{0-inf} (% AUC_{Extra} < 20%) or AUC_{0-last} (% AUC_{Extra} > 20%) with administered dose

Supplementary Table 10. Mouse PK study results (PO administration).

Full mouse PK study results for 15mg/kg compound **1** PO administration. Formulation: 3 mg/mL in NMP/PEG400/H20=1 : 7 : 2, clear solution

PO							
PO Time (h)	M04	M05	M06	Mean PO		SD	CV (%)
0.250	65.4	203	47.1	**105**	±	85.2	81.0
0.500	179	254	99.7	**178**	±	77.2	43.5
1.00	291	322	185	**266**	±	71.8	27.0
2.00	317	259	175	**250**	±	71.4	28.5
4.00	226	200	189	**205**	±	19.0	9.27
8.00	69.9	71.2	164	**102**	±	54.0	53.1
24.0	2.18	3.85	3.33	**3.12**	±	0.855	27.4
PK Parameters	M04	M05	M06	Mean PO		SD	CV (%)
Rsq_adj	0.994	0.995	0.950	--	±	--	--
No. points used for $T_{1/2}$	3.00	4.00	3.00	**ND**	±	--	--
C_{max} (ng/mL)	317	322	189	**276**	±	75.4	27.3
T_{max} (h)	2.00	1.00	4.00	**2.33**	±	1.53	65.5
$T_{1/2}$ (h)	3.04	3.59	3.24	**3.29**	±	0.278	8.45
T_{last} (h)	24.0	24.0	24.0	**24.0**	±	--	--
AUC_{0-last} (ng·h/mL)	1843	1841	2004	**1896**	±	93.7	4.94
AUC_{0-inf} (ng.h/mL)	1852	1860	2019	**1911**	±	94.3	4.93
MRT_{0-last} (h)	4.84	5.10	6.69	**5.54**	±	1.000	18.0
MRT_{0-inf} (h)	4.97	5.36	6.86	**5.73**	±	0.999	17.4
AUC_{Extra} (%)	0.517	1.07	0.771	**0.787**	±	0.278	35.4
$AUMC_{Extra}$ (%)	2.96	5.84	3.22	**4.01**	±	1.60	39.8
Bioavailability (%)[a]	--	--	--	**17.8**	±	--	--

ND = Not determined (Parameters not determined due to inadequately defined terminal elimination phase)
BQL = Below the lower limit of quantitation (LLOQ)
If the adjusted rsq (linear regression coefficient of the concentration value on the terminal phase) is less than 0.9, T1/2 might not be accurately estimated.
If the % AUC_{Extra} > 20%, AUC_{0-inf}, Cl, MRT_{0-inf} and Vd_{ss} might not be accurately estimated.
If the % $AUMC_{Extra}$ > 20%, MRT_{0-inf} and Vd_{ss} might not be accurately estimated.
The adjusted linear regression coefficient of the concentration value on the terminal phase is less than 0.9, $T_{1/2}$ might not be accurately estimated.
a: Bioavailability (%) was calculated using AUC_{0-inf} (% AUC_{Extra} < 20%) or AUC_{0-last} (% AUC_{Extra} > 20%) with administered dose

References

1. Britton, D., Noland, W.E., Pinnow, M.J. & Young, V.G.Jr. Crystal packing: an examination of the packing of molecules approximately isosteric with 4,5-dichlorophthalic anhydride. *Helv. Chim. Acta* **86 (4)**, 1175–1192 (2003).

2. Mavrova, A.Ts. *et al.* Synthesis and antitrichinellosis activity of some 2-substituted-[1,3]thiazolo[3,2-a]benzimidazol-3(2H)-ones. *Bioorg. Med. Chem.* **13 (19)**, 5550–5559 (2005).

3. Morinaga, A., Nagao, K., Ohmiya, H. & Sawamura, M. Synthesis of 1,1-diborylalkenes through a bronsted base catalyzed reaction between terminal alkynes and bis(pinacolato)diboron. *Angew. Chem. (International Edition).* **54 (52)**, 15859–15862 (2015).

4. Hirst, G. *et al.* Pyrazolopyrimidines as therapeutic agents. US2002156081 (A1).

5. Ashweek, N.J., Coldham, I., Haxell, T.F.N. & Howard, S. Preparation of diamines by lithiation–substitution of imidazolidines and pyrimidines. *Org. Biomol. Chem.* **1(9)**, 1532–1544 (2003).

6. Pan, Zh. & Lin, X. Kinase inhibitor and method for treatment of related diseases. US2014256759 (A1).

7. Kyunghyun, C. *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1724–1734 (2014).

8. Kingma D.P. & Ba J.A.: A method for stochastic optimization, 3rd International Conference for Learning Representations, San Diego, CA, US, 7-9 May 2015.

9. Putin, E. *et al.* Reinforced adversarial neural computer for de novo molecular design. *J. Chem. Inf. Model.* **58,** 1194–1204 (2018).

10. Putin, E. *et al.* Adversarial threshold neural computer for molecular de novo design. *Mol. Pharm.* **15,** 4386–4397 (2018).

11. Sanchez-Lengeling, B., Outeiral, C., Guimaraes, G.L. & Aspuru-Guzik, A. Optimizing distributions over molecular space. An objective-reinforced generative adversarial network for inverse-design chemistry (ORGANIC). Preprint at https://chemrxiv.org/articles/ORGANIC_1_pdf/5309668 (2017).

12. Guimaraes, G.L., Sanchez-Lengeling, B., Farias, P.L.C. & Aspuru-Guzik, A. Objective- reinforced generative adversarial networks (ORGAN) for sequence generation models. Preprint at https://arxiv.org/abs/1705.10843 (2017).

13. Kusner, M.J., Paige, B. & Hernández-Lobato, J.M. Grammar variational autoencoder. In D. Precup and Y. W. Teh, editors, International Conference on Machine Learning, volume 70 of Proceedings of Machine Learning Research, Sydney, Australia, 06–11 Aug 2017.

14. Gómez-Bombarelli, R. *et al.* Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Sci.* **4**, 268–276 (2018).

15. Bickerton, G.R., Paolini, G.V., Besnard, J., Muresan, S. & Hopkins, A.L. Quantifying the chemical beauty of drugs. *Nat. Chem.* **4**, 90–98 (2012).

16. Ertl, P. & Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminform.* **1**, 8 (2009).

17. Wildman, S.A. & Crippen, G.M. Prediction of physicochemical parameters by atomic contributions. *J. Chem. Inf. Comput. Sci.* **39,** 868–873 (1999).

18. Landrum, G. RDKit: Open-source cheminformatics. Available at: http://www.rdkit.org/ (Accessed: 23rd August 2018).

19. Polykovskiy, D *et al.* Molecular Sets (MOSES): a benchmarking platform for molecular generation models. Preprint at https://arxiv.org/abs/1811.12823 (2018).

20. Irwin, J.J. *et al.* ZINC: a free tool to discover chemistry for biology *J. Chem. Inf. Model.* **52**, 1757–1768 (2012).

21. Wang, Z. *et al.* Tetrahydroisoquinoline-7-carboxamide derivatives as new selective discoidin domain receptor 1 (DDR1) inhibitors. *ACS Med. Chem. Lett.* **8**, 327–332 (2017).

22. Wang, Z. *et al.* Structure-based design of tetrahydroisoquinoline-7-carboxamides as selective discoidin domain receptor 1 (DDR1) inhibitors. *J. Med. Chem.* **59**, 5911–5916 (2016).

23. Gao, M. *et al.* Discovery and optimization of 3-(2-(Pyrazolo[1,5-a]pyrimidin-6-yl)ethynyl)benzamides as novel selective and orally bioavailable discoidin domain receptor 1 (DDR1) inhibitors. *J. Med Chem.* **56**, 3281–3295 (2013).

24. Kim, H.G. *et al.* Discovery of a potent and selective DDR1 receptor tyrosine kinase inhibitor. *ACS Chem. Biol.* **8**, 2145–2150 (2013).

25. Elkamhawy, A. *et al.* Discovery of a broad spectrum antiproliferative agent with selectivity for DDR1 kinase: cell line-based assay, kinase panel, molecular docking, and toxicity studies. *J. Enzyme Inhib. Med. Chem.* **31**, 158–166 (2016).

26. Liu, L. *et al.* Synthesis and biological evaluation of novel dasatinib analogues as potent DDR1 and DDR2 kinase inhibitors. *Chem. Biol. Drug Des.* **89**, 420–427 (2017).

27. Canning, P. *et al.* Structural mechanisms determining inhibition of the collagen receptor DDR1 by selective and multi-targeted type II kinase inhibitors. *J. Mol. Biol.* **426**, 2457–2470 (2014).

28. Li, Y. *et al.* Small molecule discoidin domain receptor kinase inhibitors and potential medical applications. *J. Med. Chem.* **58**, 3287–3301 (2015).

29. Fraser, C., Carragher, N. O. & Unciti-Broceta, A. eCF309: a potent, selective and cell-permeable mTOR inhibitor. *Med. Chem. Commun.* **7**, 471–477 (2016).

30. Wang, Q. *et al.* Discovery of N-(3-((1-Isonicotinoylpiperidin-4-yl)oxy)-4-methylphenyl)-3-(trifluoromethyl)benzamide (CHMFL-KIT-110) as a selective, potent, and orally available type II c-KIT kinase inhibitor for gastrointestinal stromal tumors (GISTs). *J. Med. Chem.* **59**, 3964–3979 (2016).

31. Murray, C.W. *et al.* Fragment-based discovery of potent and selective DDR1/2 inhibitors. *ACS Med. Chem. Lett.* **6**, 798–803 (2015).

32. Murata, T. *et al.* Benzamide derivative. WO2013161851 (A1).

33. Murata, T. *et al.* Quinazolinedione derivative. WO2013161853 (A1).

34. Buettelmann, B. *et al.* Triaza-spirodecanones as DDR1 inhibitors. WO2017005583 (A1).

35. Buettelmann, B. *et al.* Spiroindolinones as DDR1 inhibitors. WO2017137334 (A1).

36. Nishio, Y. *et al.* Urea derivative and use threrefor. WO2017038873 (A1).

37. Nishio, Y. *et al.* Urea derivative and use threrefor. WO2017038871 (A1).

38. Brekken, R.A. *et al.* Small-molecule inhibitors targeting discoidin domain receptor 1 and uses thereof. WO2016064970 (A1).

39. Saxty, G. et al. Imidazo-condensed bicycles as inhibitors of discoidin domain receptors (DDRs). WO2015004481 (A1).

40. Bae, I.H. Thieno[3,2-d]pyrimidine derivatives having inhibitory activity for protein kinases. WO2013100632 (A1).

41. Clarivate Analytics Integrity. Available at: https://integrity.thomson-pharma.com/integrity/xmlxsl/. (Accessed: 23rd August 2018).

42. EBI Web Team. ChEMBL. Available at: https://www.ebi.ac.uk/chembl/. (Accessed: 23rd August 2018).

43. PDB. RCSB Protein Data Bank. Available at: https://www.rcsb.org (Accessed: 19th July 2018).