# *Supporting Information*

# Structure-based *de novo* design and indentification of D816V mutant-selective c-KIT inhibitors

*Hwangseo Park,*[a] Soyoung Lee,[b] Suhyun Lee,[b] and Sungwoo Hong*[b]*

*[a]Department of Bioscience and Biotechnology, Sejong University, Seoul 143-747*

*[b]KoreaDepartment of Chemistry, Korea Advance Institute of Science and Technology (KAIST), Daejeon 305-701, Korea*

**I. Modeling.** To obtain the all-atom models for the two receptors, hydrogen atoms were added to each protein atom in the original X-ray crystal structures of the wild-type c-KIT and the homology-modeled structure of D816V mutant. A special attention was paid to assign the protonation states of the ionizable Asp, Glu, His, and Lys residues. The side chains of Asp and Glu residues were assumed to be neutral if one of their carboxylate oxygens pointed toward a hydrogen-bond accepting group including the backbone aminocarbonyl oxygen at a distance within 3.5 Å, a generally accepted distance limit for a hydrogen bond of moderate strength. Similarly, the lysine side chains were assumed to be protonated unless the NZ atom was in proximity to a hydrogen-bond donating group. The same procedure was also applied to determine the protonation states of ND and NE atoms in His residues.

**II. Rescoring of the generated 7-azaindole derivatives.** Although the effects of ligand solvation had been shown to be critically important in estimating the binding affinity of a protein-ligand complex, the scoring function of the LigBuilder program lacked a solvation term. Therefore, the derivatives of 7-azaindole generated with LigBuilder were further screened with a new binding free energy function constructed by combining an appropriate solvation free energy term to the original scoring function of the AutoDock program. This modified scoring function can be expressed as follows.

$$\Delta G_{bind}^{aq} = W_{vdW}\sum_{i=1}\sum_{j=1}\left(\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right)+W_{hbond}\sum_{i=1}\sum_{j=1}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}}-\frac{D_{ij}}{r_{ij}^{10}}\right)$$
$$+W_{elec}\sum_{i=1}\sum_{j=1}\frac{q_i q_j}{\varepsilon(r_{ij})r_{ij}}+W_{tor}N_{tor}+W_{sol}\sum_{i=1}S_i\left(O_i^{\max}-\sum_{j\neq i}V_j e^{-\frac{r_{ij}^2}{2\sigma^2}}\right) \quad (1)$$

Here, $W_{vdW}$, $W_{hbond}$, $W_{elec}$, $W_{tor}$, and $W_{sol}$ are the weighting factors of van der Waals, hydrogen bond, electrostatic interactions, torsional term, and solvation free energy of inhibitors, respectively. $r_{ij}$ represents the interatomic distance, and $A_{ij}$, $B_{ij}$, $C_{ij}$, and $D_{ij}$ are related to the depths of the potential energy well and the equilibrium separations between the protein and ligand atoms. The hydrogen bond term has an additional weighting factor, $E(t)$, representing the angle-dependent directionality. With respect to the distance-dependent dielectric constant ($\varepsilon(r_{ij})$), a sigmoidal function was used in computing the interatomic electrostatic interactions between c-KIT and 7-azaindole derivatives. Gasteiger-Marsili atomic charges were then used for both proteins and ligands to compute the electrostatic interaction energies. In the entropic term, $N_{tor}$ is the number of rotatable bonds in the ligand. In the

desolvation term, $S_i$ and $V_i$ are the solvation parameter and the fragmental volume of atom $i$, respectively, while $O_i^{max}$ stands for the maximum atomic occupancy.

In the calculation of molecular solvation free energy of 7-azaindole derivatives, we used the atomic parameters. The addition of this solvation free energy term is expected to increase the accuracy of the scoring function because the underestimation of ligand solvation often leads to the overestimation of the binding affinity of a ligand with many polar atoms. Indeed, the superiority of this modified scoring function to the previous one was well-appreciated in recent studies for virtual screening of kinase inhibitors.

**III. Molecular dynamics simulations.** MD simulations of wild-type c-KIT and the D816V mutant in complex with 7-azaindole derivatives were carried out using the AMBER program of version 7 with the force field. To derive the force-field parameters for the 7-azaindole derivatives unavailable in the standard force field database, we followed the procedure suggested by Fox and Kollman to be consistent with the standard AMBER force field. The starting coordinates of MD simulations were extracted from the structures of wild-type c-KIT and the D816V mutant in complex with 7-azaindole that were obtained with docking simulations. The all-atom models for the wild-type and D816V mutant in complex with the 7-azaindole derivatives were immersed in rectangular boxes containing about 12,000 TIP3P water molecules. After 200 cycles of energy minimization to remove the bad steric contacts, we equilibrated both systems beginning with 20 ps equilibration dynamics of the solvent molecules at 300 K. The next step involved the equilibration of the solute with a fixed configuration of the solvent molecules consecutively at 10, 50, 100, 150, 200, 250, and 300 K for 10 ps at each temperature. Then, the equilibration dynamics of the entire system was performed at 300 K for 100 ps. Following the equilibration procedure, 10.2 ns production simulations were carried out with periodic boundary conditions in the NPT ensemble. The temperature and pressure were kept at 300 K and 1 atm using Berendsen temperature coupling and isotropic molecule-based scaling, respectively. The SHAKE algorithm, with a tolerance of $10^{-6}$ Å, was applied to fix all bond lengths involving the hydrogen atom. We used a time step of 2.0 fs and a nonbond-interaction cutoff radius of 12 Å; the trajectory was sampled every 0.4 ps (200-step intervals) for analysis.

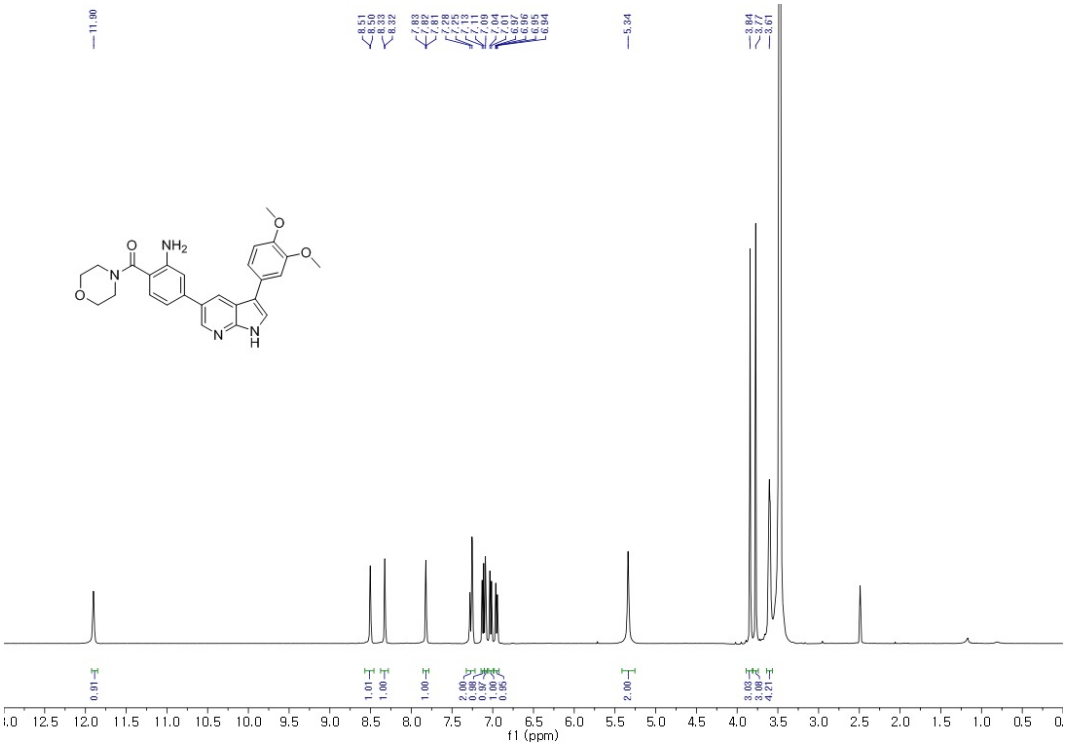
**IV. Kinase Assay.** The effects of compounds on the kinase activity of c-KIT and c-KIT$^{D816V}$ were analyzed the Reaction Biology Corp. (Malvern, PA, USA) using radiometric kinase assays ([$\gamma$-$^{33}$P]-ATP). Briefly, reactions contained Abltide (EAIYAAPFAKKK) in freshly prepared Base Reaction Buffer (20 mM HEPES (pH 7.5), 10

mM $MgCl_2$, 1 mM EGTA, 0.02% BRIJ-35, 0.02 mg/ml BSA, 0.1 mM $Na_3VO_4$, 2 mM DTT, 1% DMSO). c-KIT or c-$KIT^{D816V}$ were delivered into the substrate solution and gently mixed. The compounds in DMSO with indicated concentrations were then delivered to the reaction. Then, $^{33}$P-ATP was added to initiate the reaction, and the mixture was further incubated for 2 h at room temperature. Reactions are spotted onto P81 ion exchange paper (Whatman # 3698-915). The filters were washed extensively in 0.75% phosphoric acid. Compounds were tested in a 10-dose $IC_{50}$ mode with 3-fold serial dilutions starting at 10 μM.

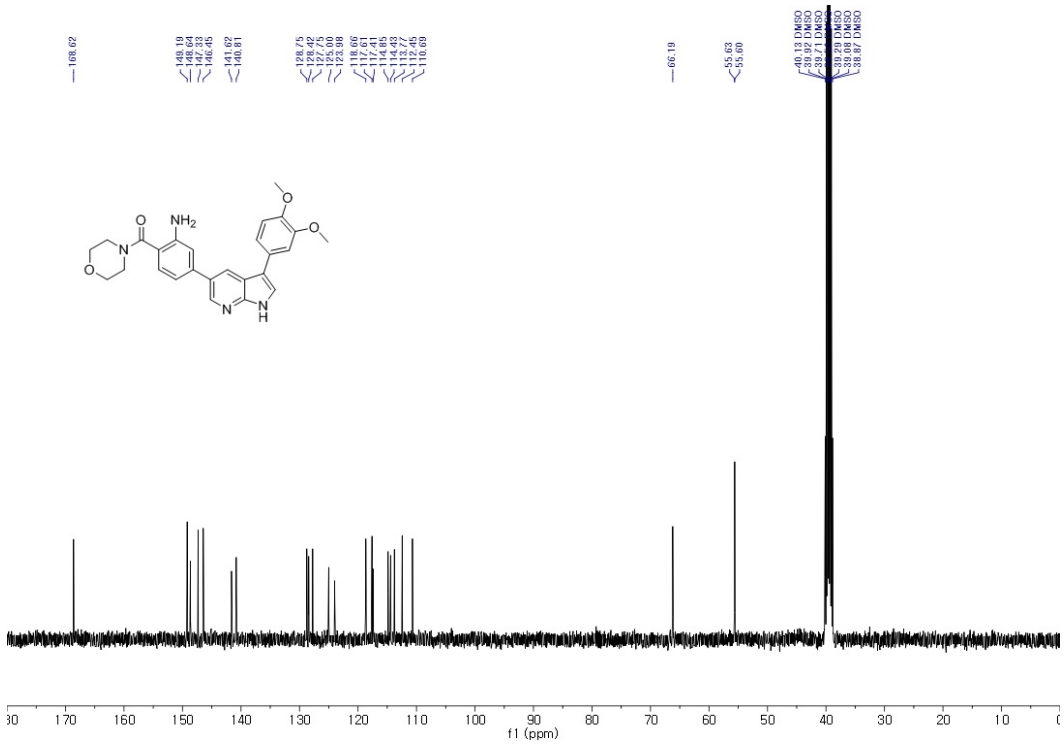
# *Appendix Ⅰ*

# Spectral Copies of $^{1}H$ and $^{13}C$ NMR Data Obtained in this Study

***(2-amino-4-(3-(3,4-dimethoxyphenyl)-1H-pyrrolo[2,3-b]pyridin-5-yl)phenyl)(morpholino)-methanone (5).***

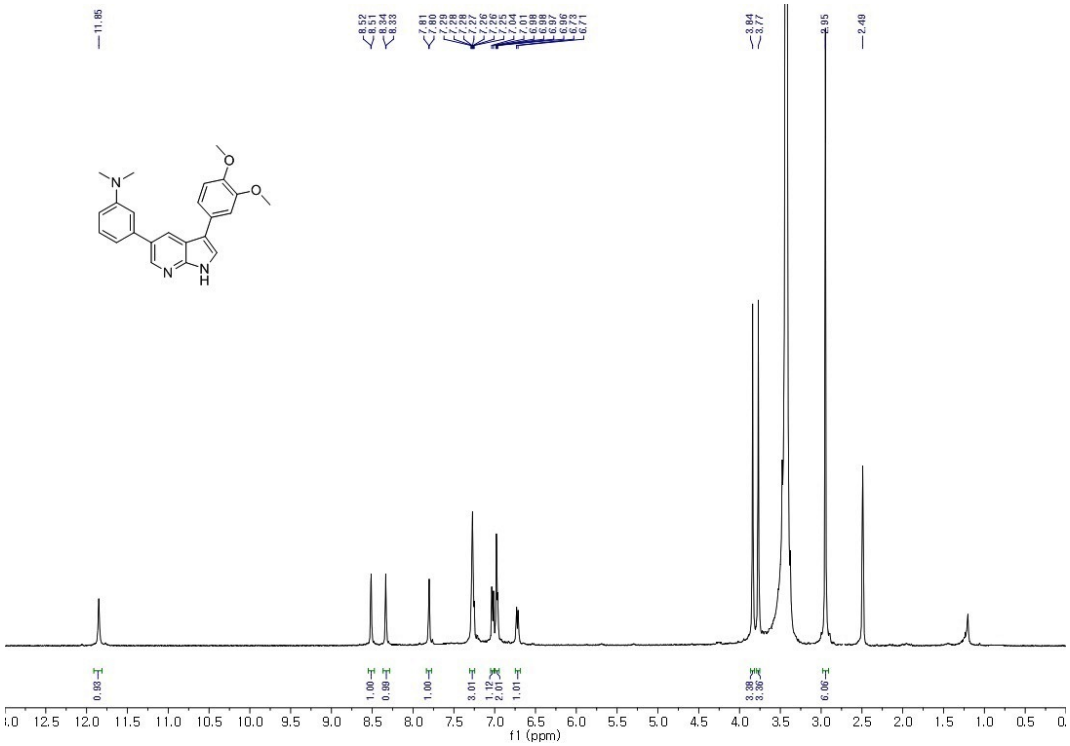



**400 MHz, $^{1}$H NMR in DMSO-$d_6$**

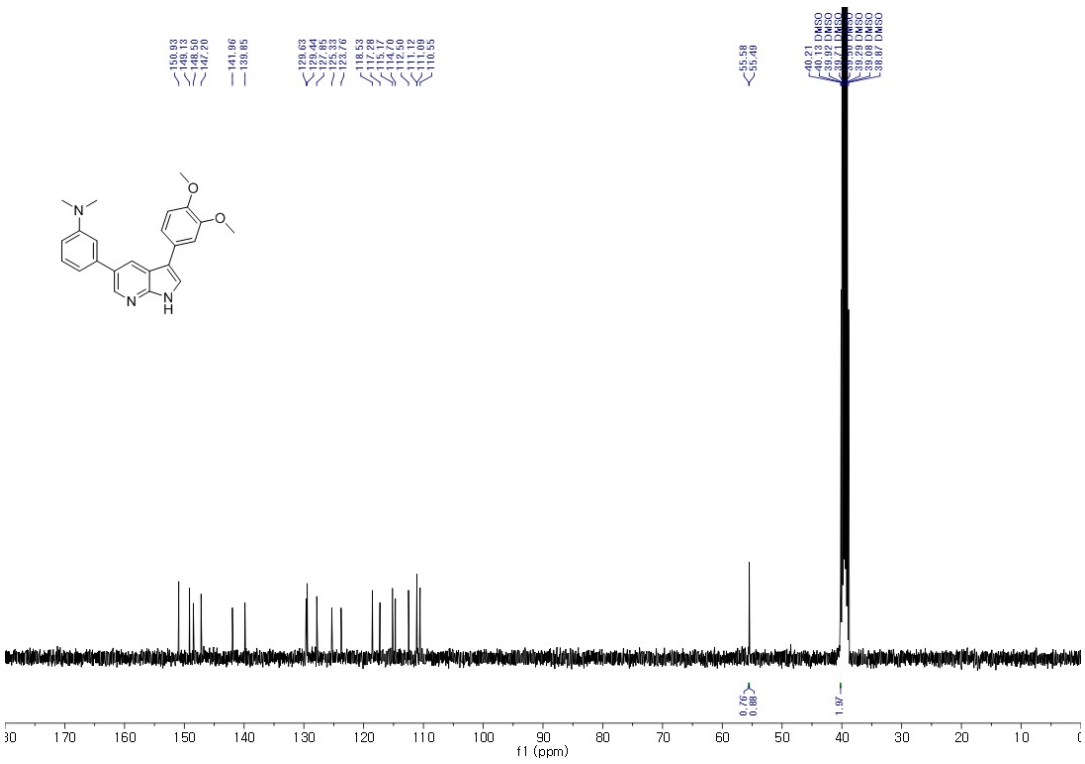



**100 MHz, $^{13}$C NMR in DMSO-$d_6$**

***3-(3-(3,4-dimethoxyphenyl)-1H-pyrrolo[2,3-b]pyridin-5-yl)-N,N-dimethylaniline (6).***

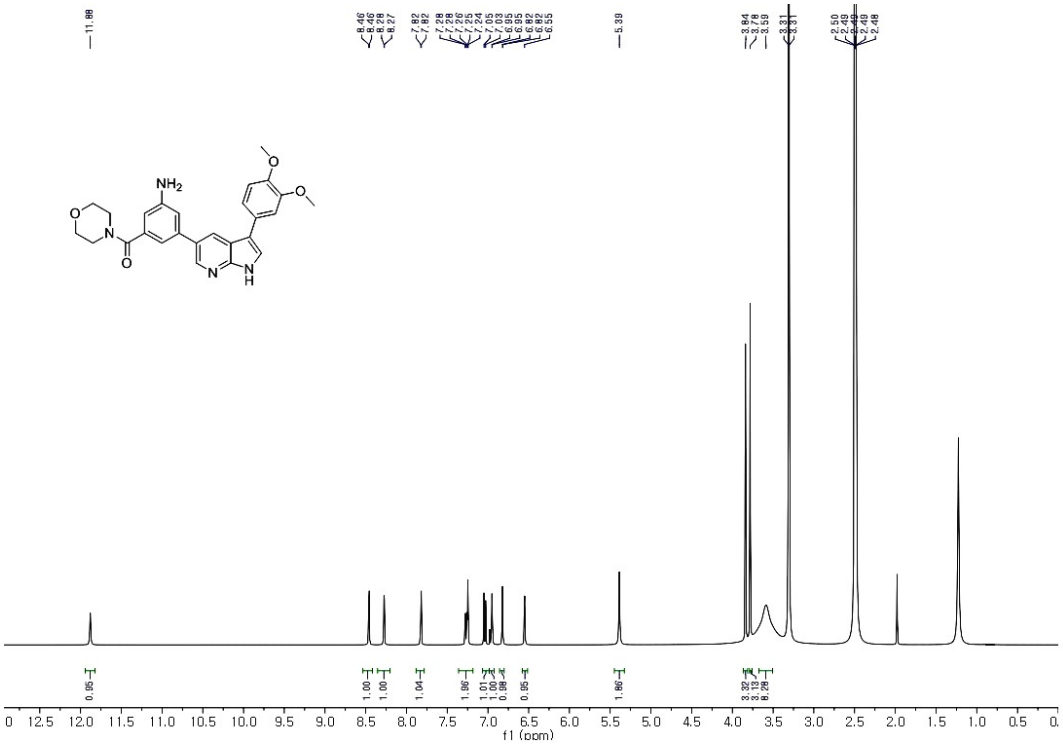



**400 MHz, $^{1}$H NMR in DMSO-$d_6$**

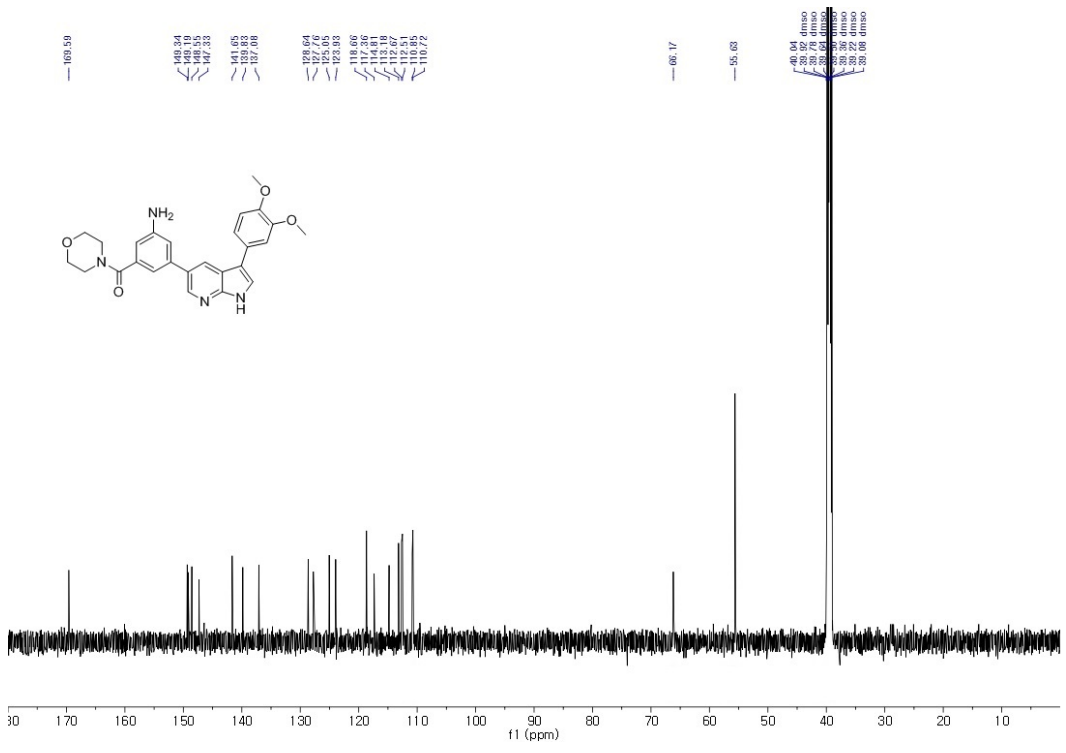



**100 MHz, $^{13}$C NMR in DMSO-$d_6$**

***(3-amino-5-(3-(3,4-dimethoxyphenyl)-1H-pyrrolo[2,3-b]pyridin-5-yl)phenyl)(morpholino)-methanone (7).***

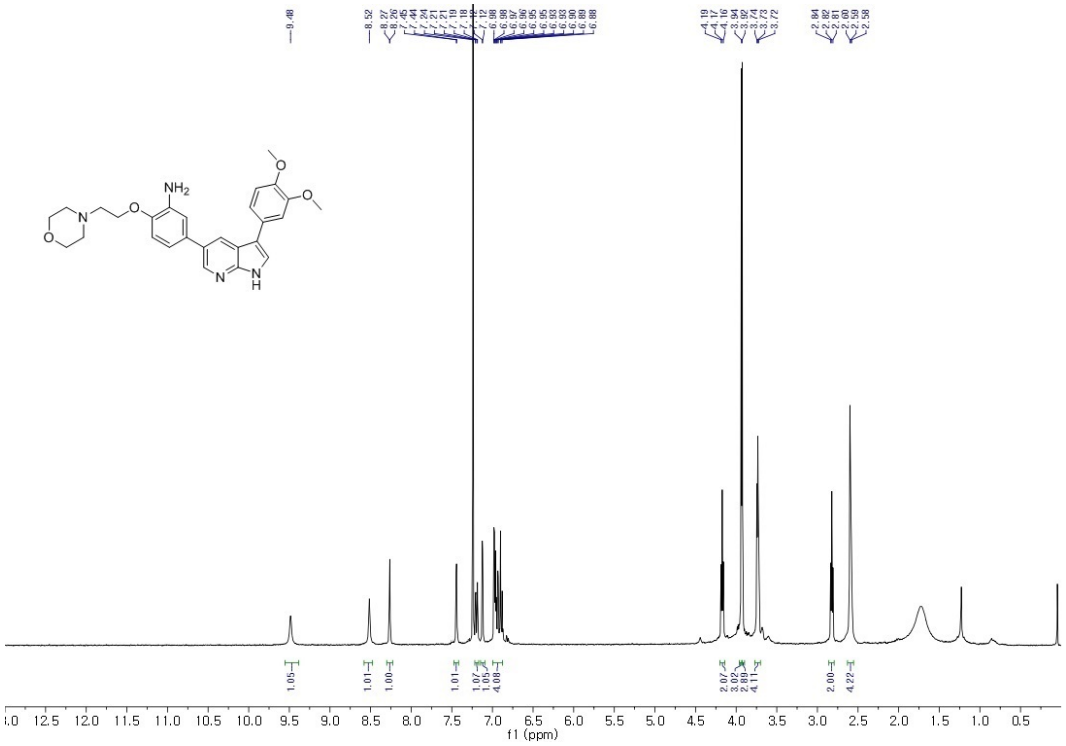


**400 MHz, $^{1}$H NMR in DMSO-$d_6$**

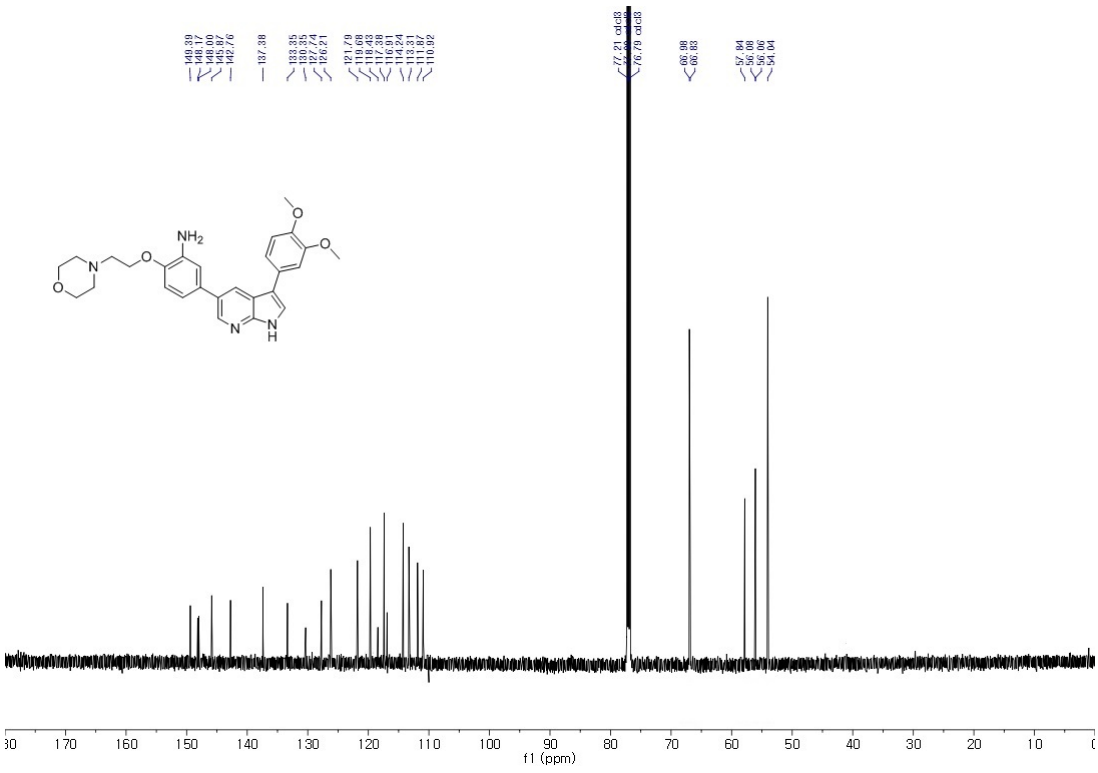


**150 MHz, $^{13}$C NMR in DMSO-$d_6$**

***5-(3-(3,4-dimethoxyphenyl)-1H-pyrrolo[2,3-b]pyridin-5-yl)-2-(2-morpholinoethoxy)aniline (8).***

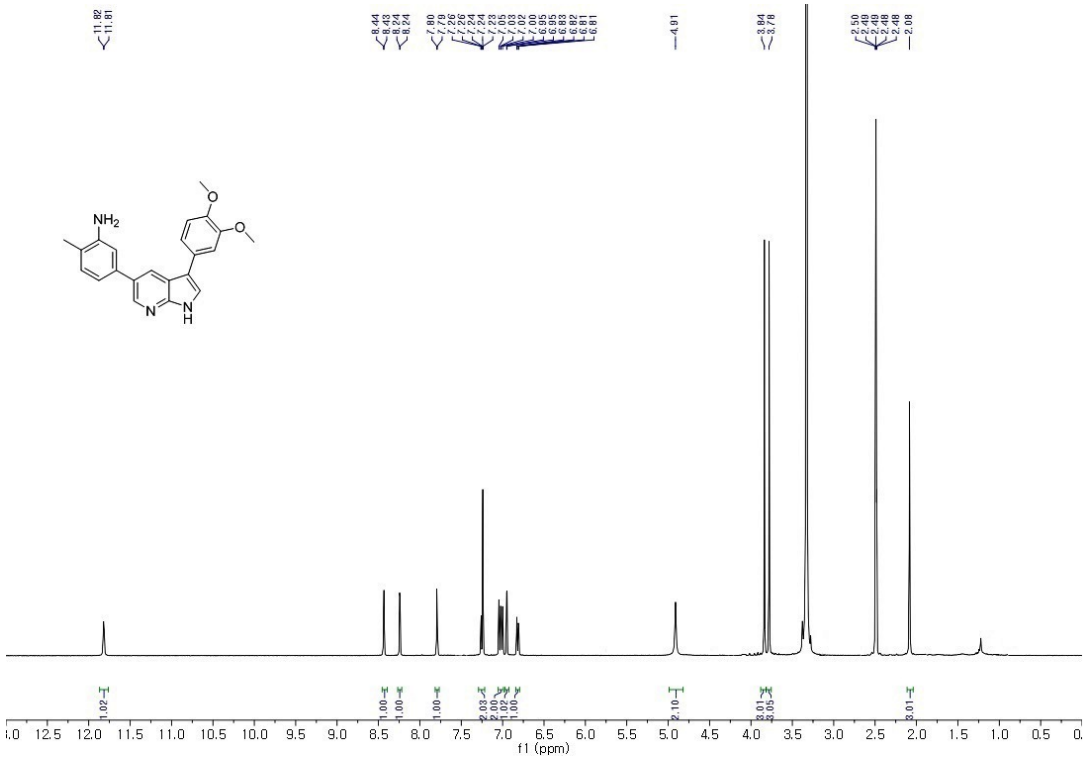



**400MHz, $^{1}$H NMR in Chloroform-*d***

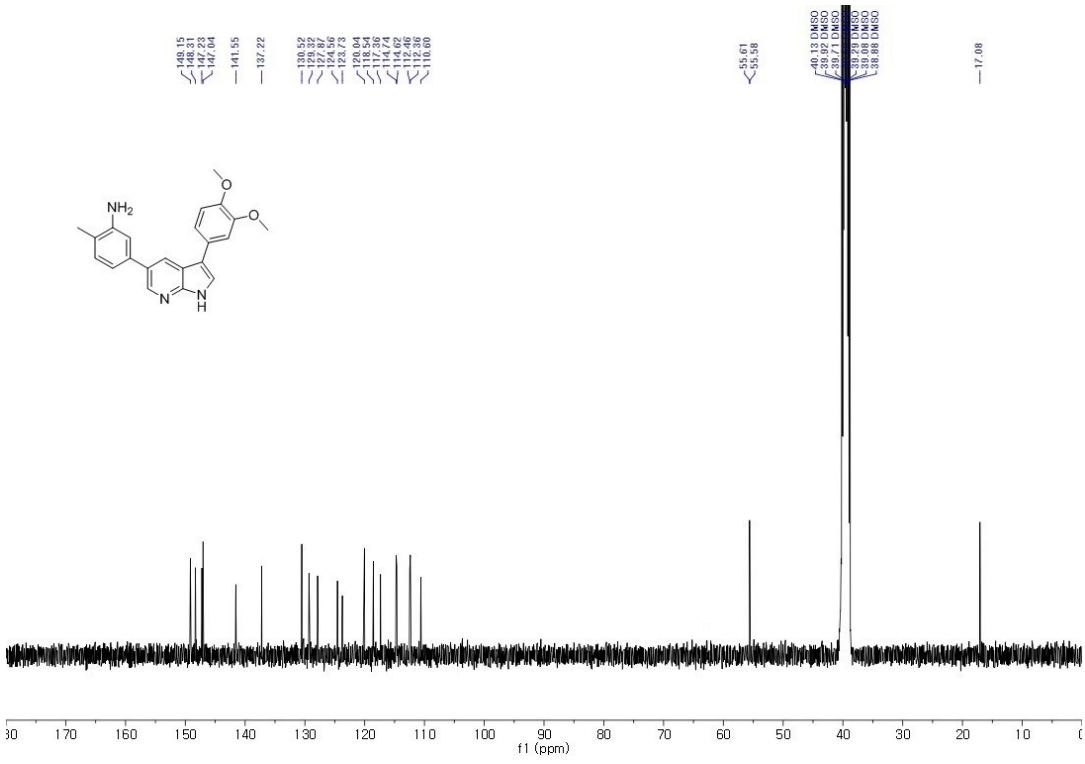



**150 MHz, $^{13}$C NMR in Chloroform-*d***

***5-(3-(3,4-dimethoxyphenyl)-1H-pyrrolo[2,3-b]pyridin-5-yl)-2-methylaniline (9)***

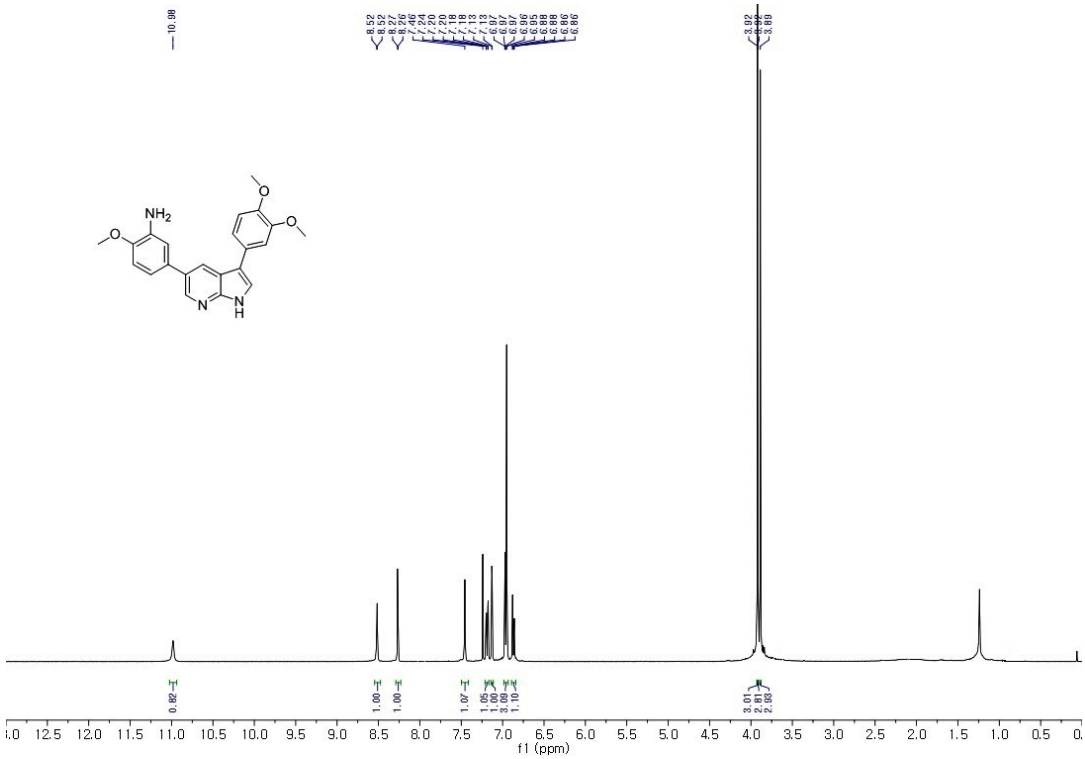


**400 MHz, $^{1}$H NMR in DMSO-$d_6$**

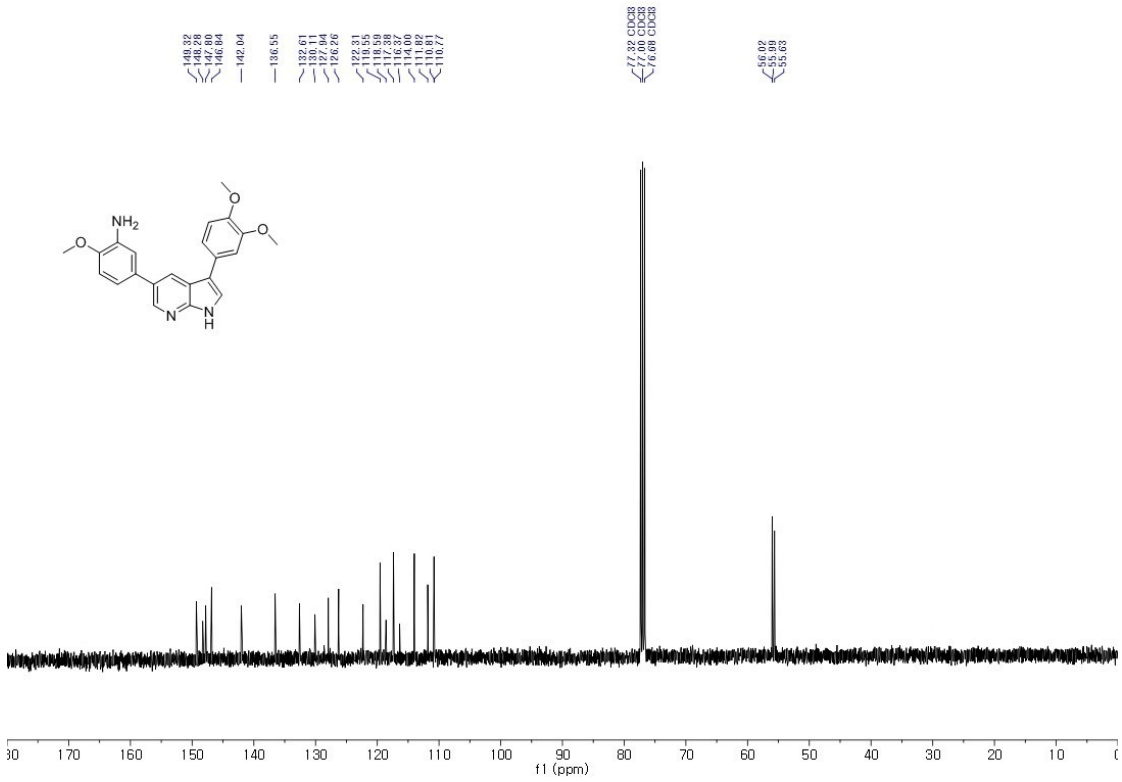


**100 MHz, $^{13}$C NMR in DMSO-$d_6$**

***5-(3-(3,4-dimethoxyphenyl)-1H-pyrrolo[2,3-b]pyridin-5-yl)-2-methoxyaniline (10).***

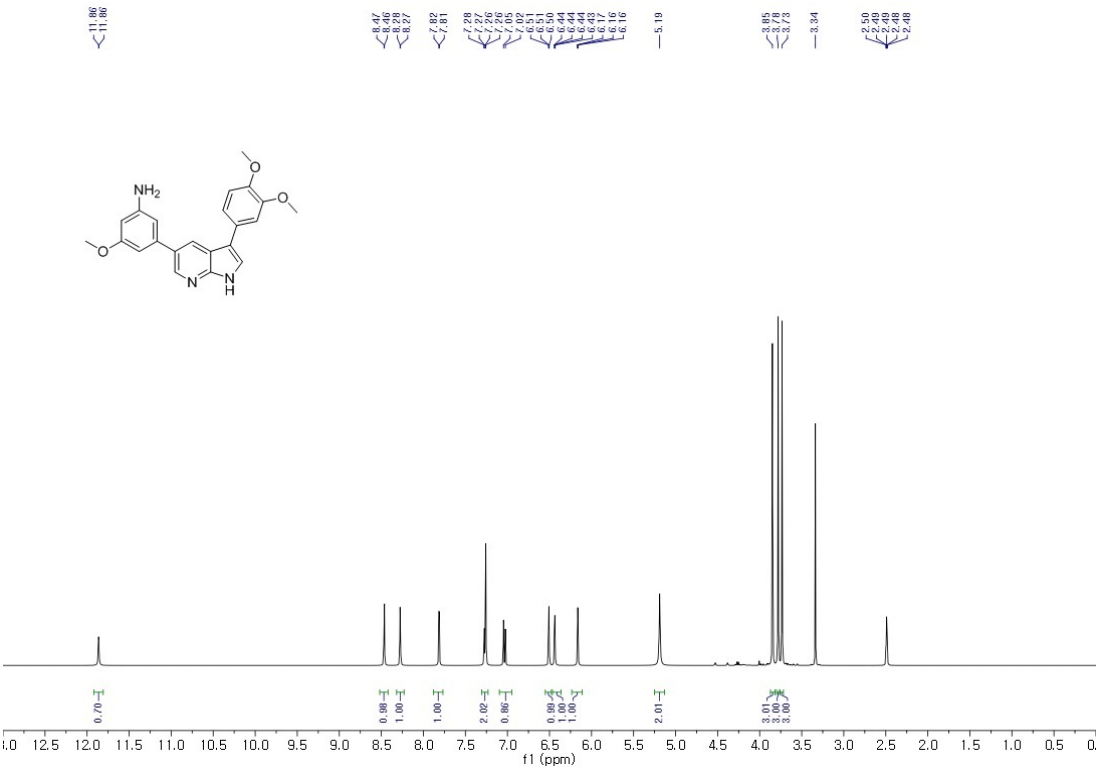



**400MHz, $^{1}$H NMR in Chloroform-*d***

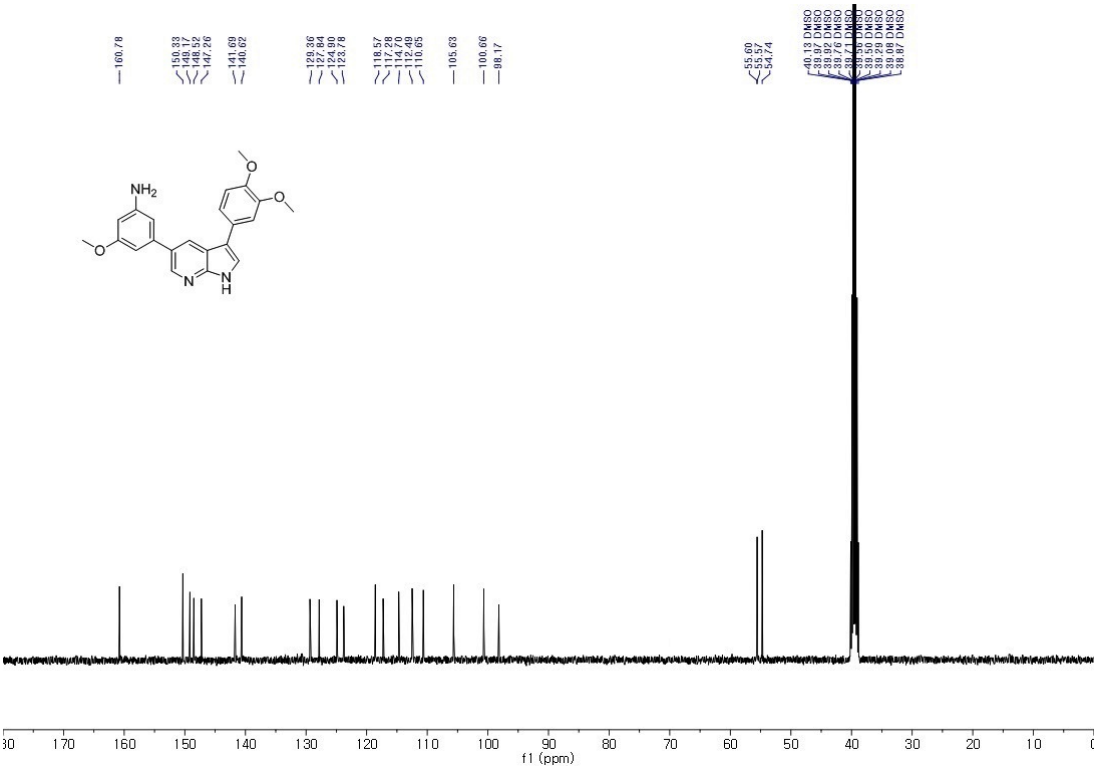



**100 MHz, $^{13}$C NMR in Chloroform-*d***

***3-(3-(3,4-dimethoxyphenyl)-1H-pyrrolo[2,3-b]pyridin-5-yl)-5-methoxyaniline (11).***

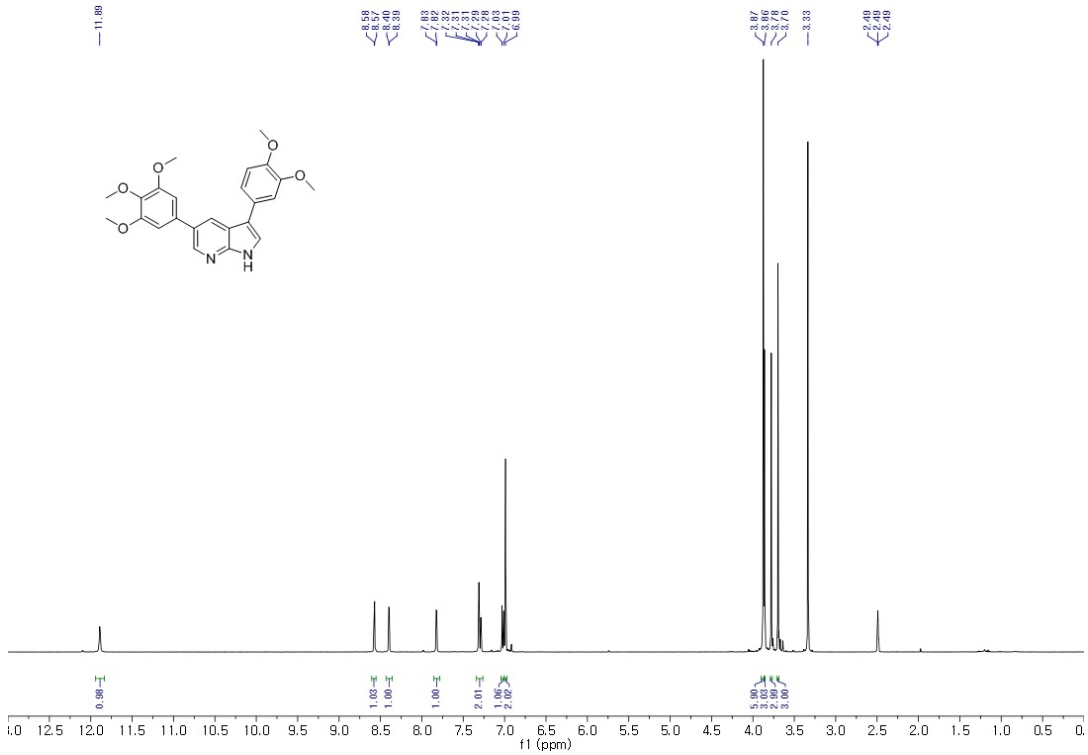



**400 MHz, $^{1}$H NMR in DMSO-$d_6$**

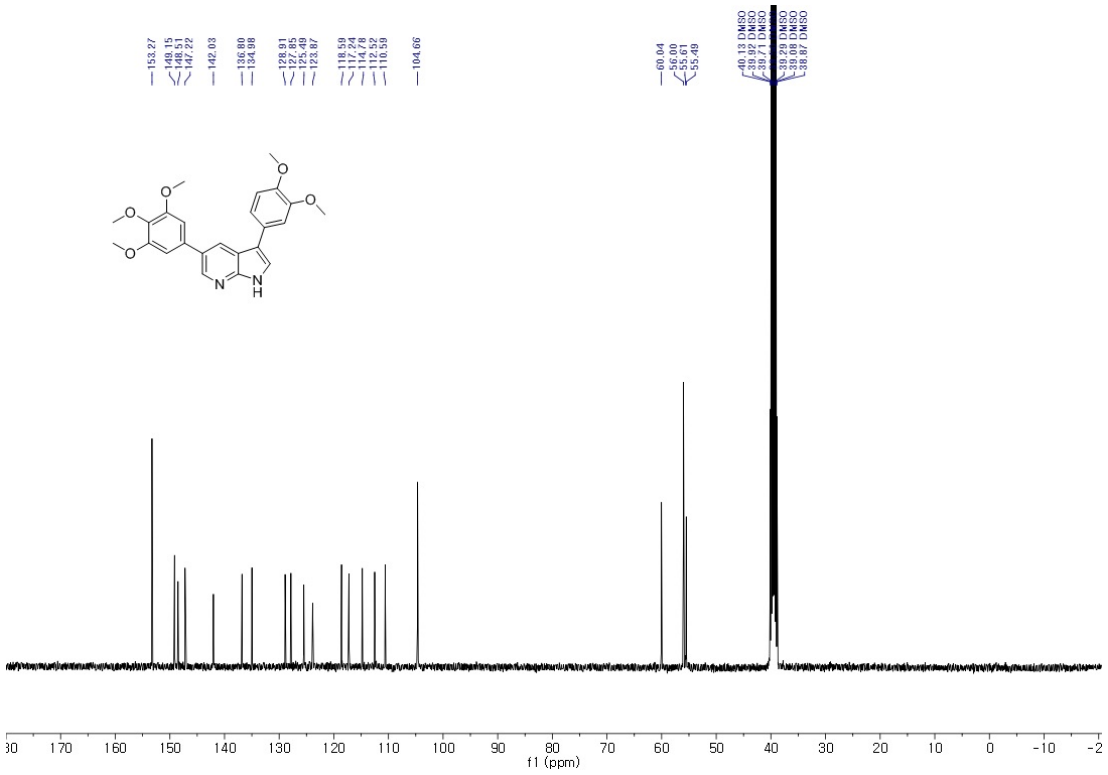



**100 MHz, $^{13}$C NMR in DMSO-$d_6$**

***3-(3,4-dimethoxyphenyl)-5-(3,4,5-trimethoxyphenyl)-1H-pyrrolo[2,3-b]pyridine (12).***

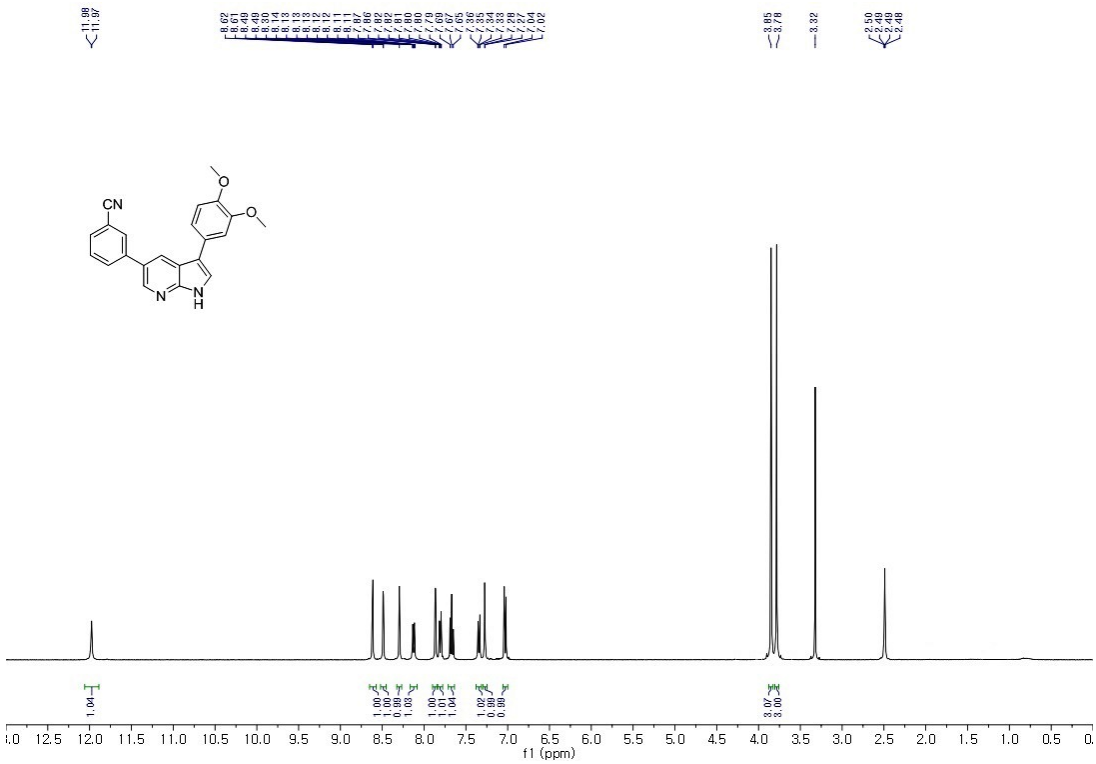



**400 MHz, $^{1}$H NMR in DMSO-$d_6$**

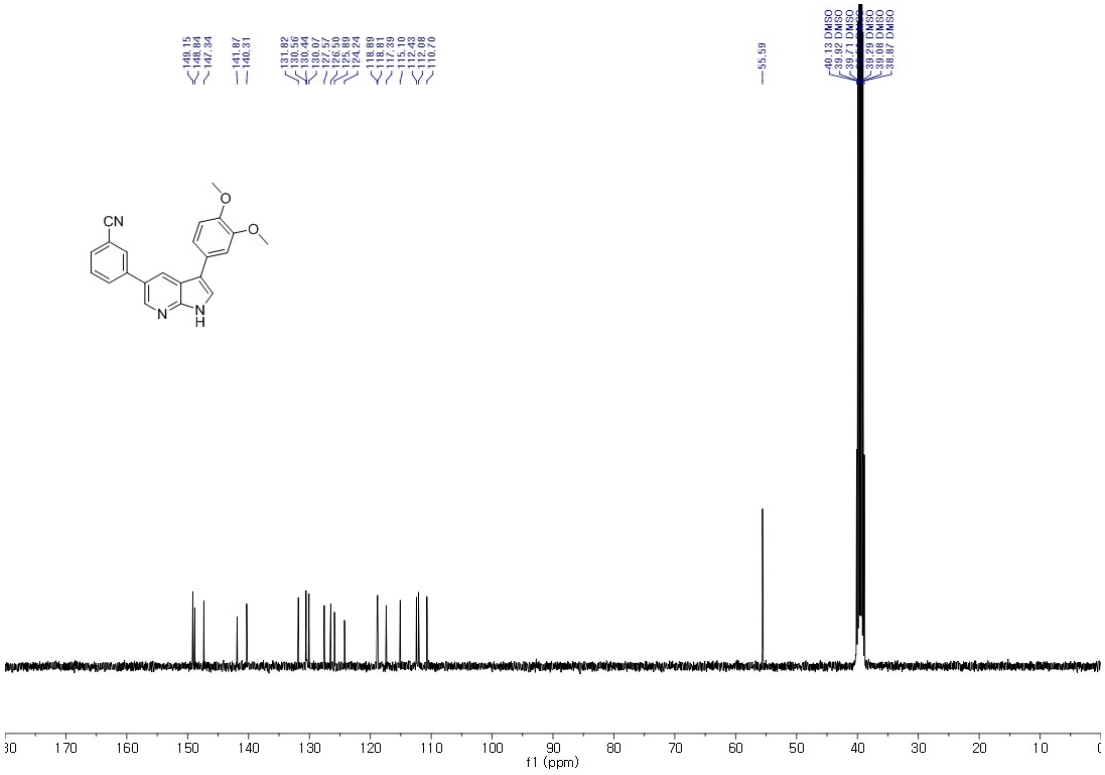



**100 MHz, $^{13}$C NMR in DMSO-$d_6$**

***3-(3-(3,4-dimethoxyphenyl)-1H-pyrrolo[2,3-b]pyridin-5-yl)benzonitrile (13).***

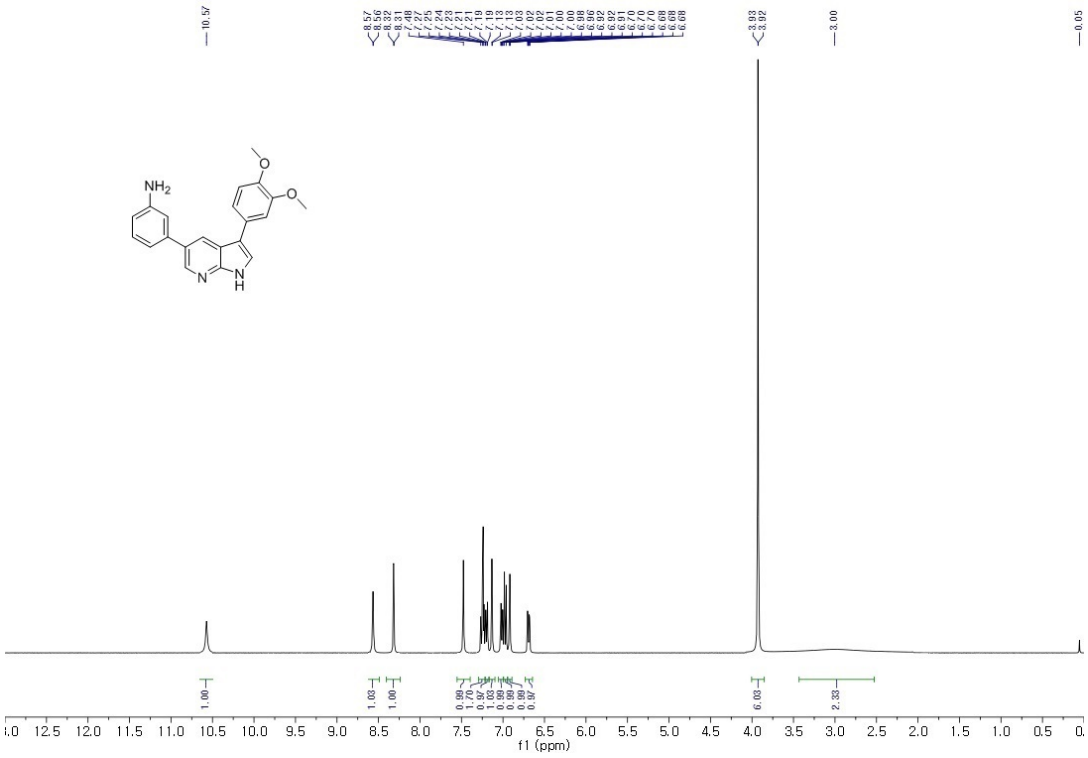



**400 MHz, $^{1}$H NMR in DMSO-$d_6$**

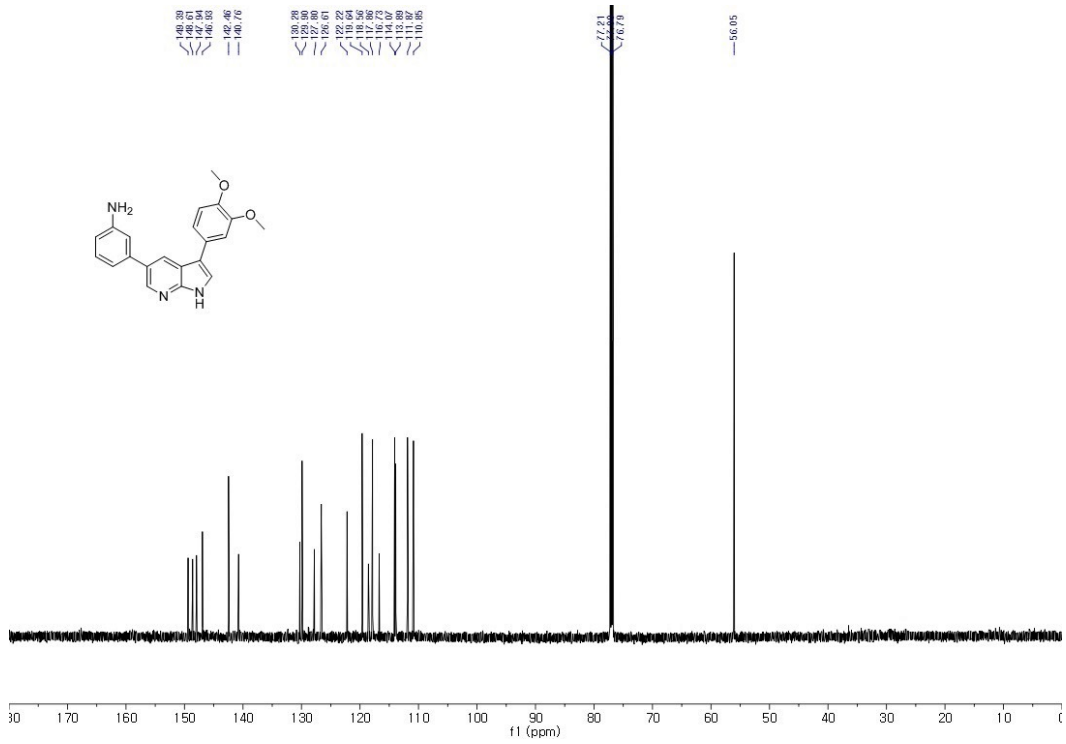



**100 MHz, $^{13}$C NMR in DMSO-$d_6$**

***3-(3-(3,4-dimethoxyphenyl)-1H-pyrrolo[2,3-b]pyridin-5-yl)aniline (14).***

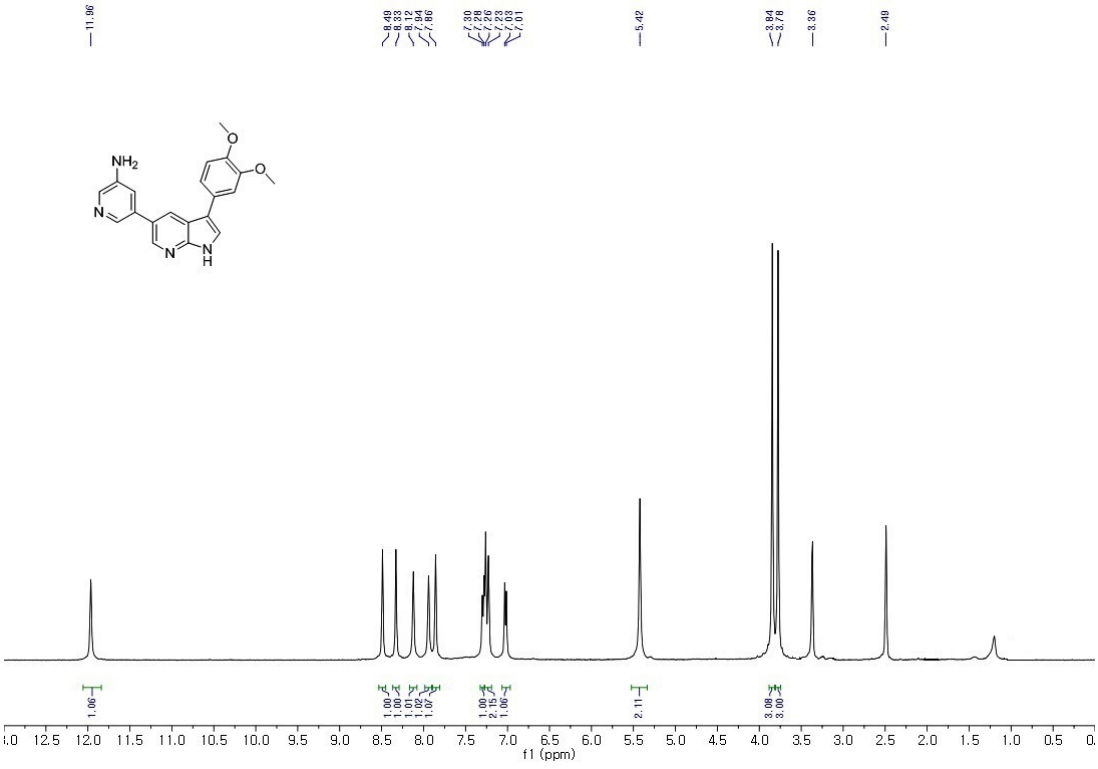



**400MHz, $^{1}$H NMR in Chloroform-*d***

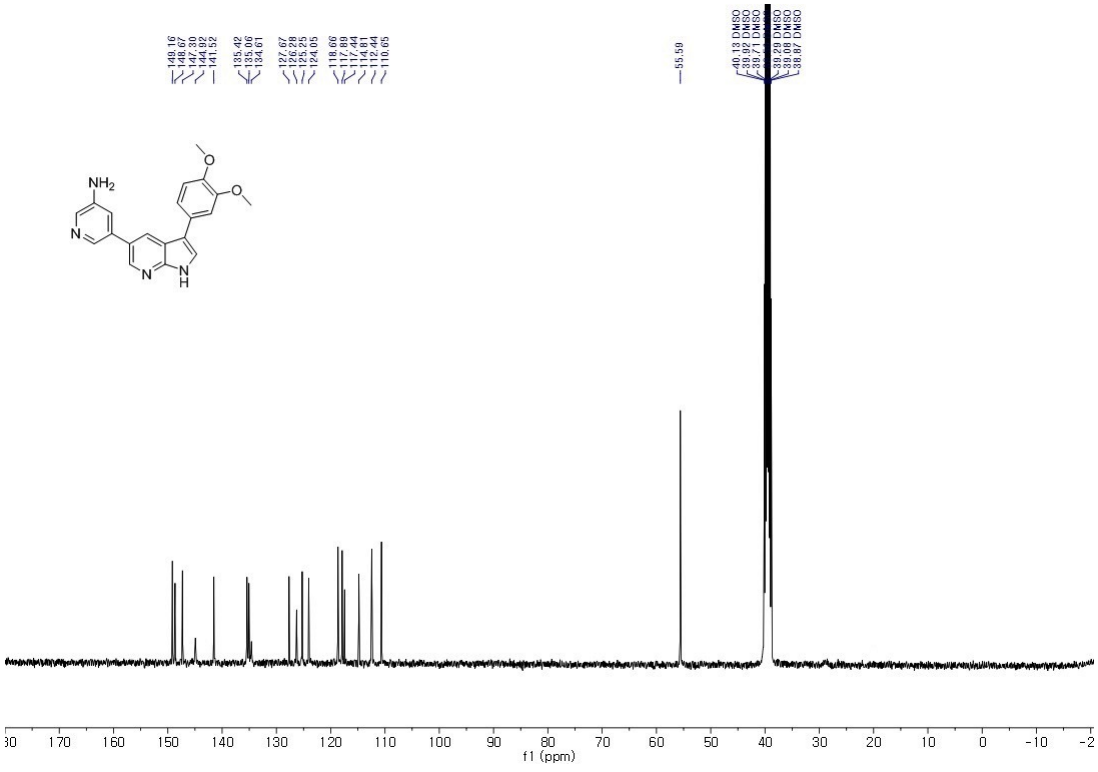



**150 MHz, $^{13}$C NMR in Chloroform-*d***

***5-(3-(3,4-dimethoxyphenyl)-1H-pyrrolo[2,3-b]pyridin-5-yl)pyridin-3-amine (15).***

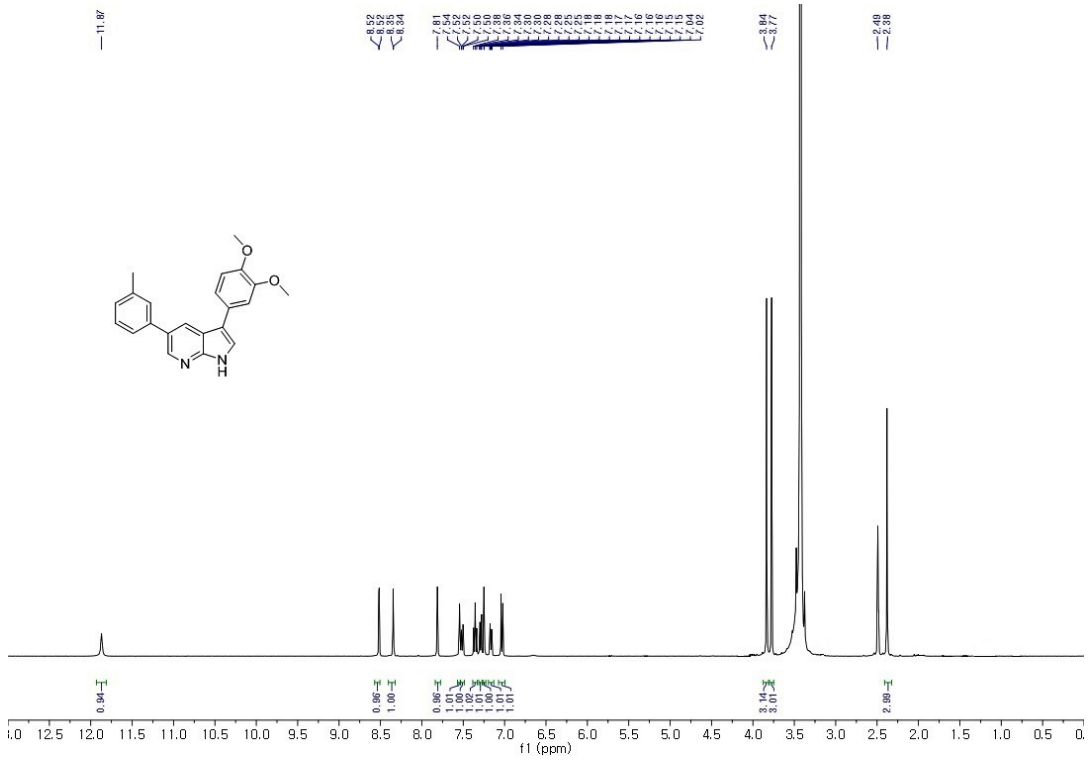



**400 MHz, $^{1}$H NMR in DMSO-$d_6$**

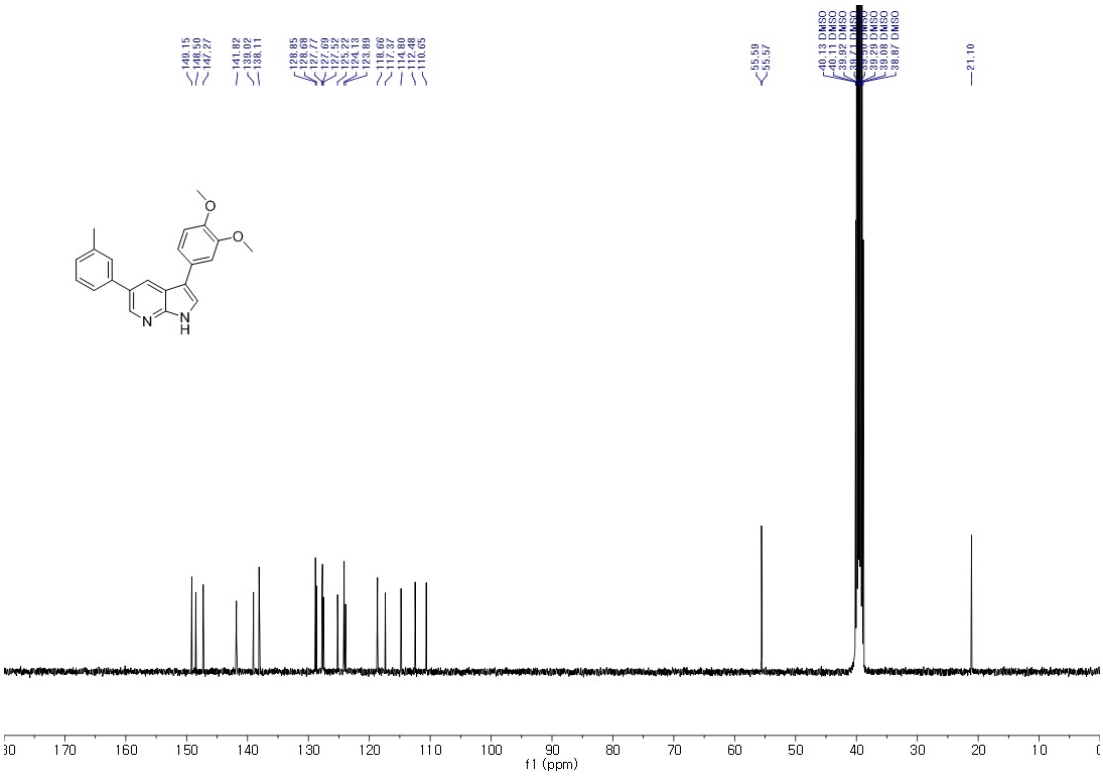



**100 MHz, $^{13}$C NMR in DMSO-$d_6$**

***3-(3,4-dimethoxyphenyl)-5-(m-tolyl)-1H-pyrrolo[2,3-b]pyridine (16).***

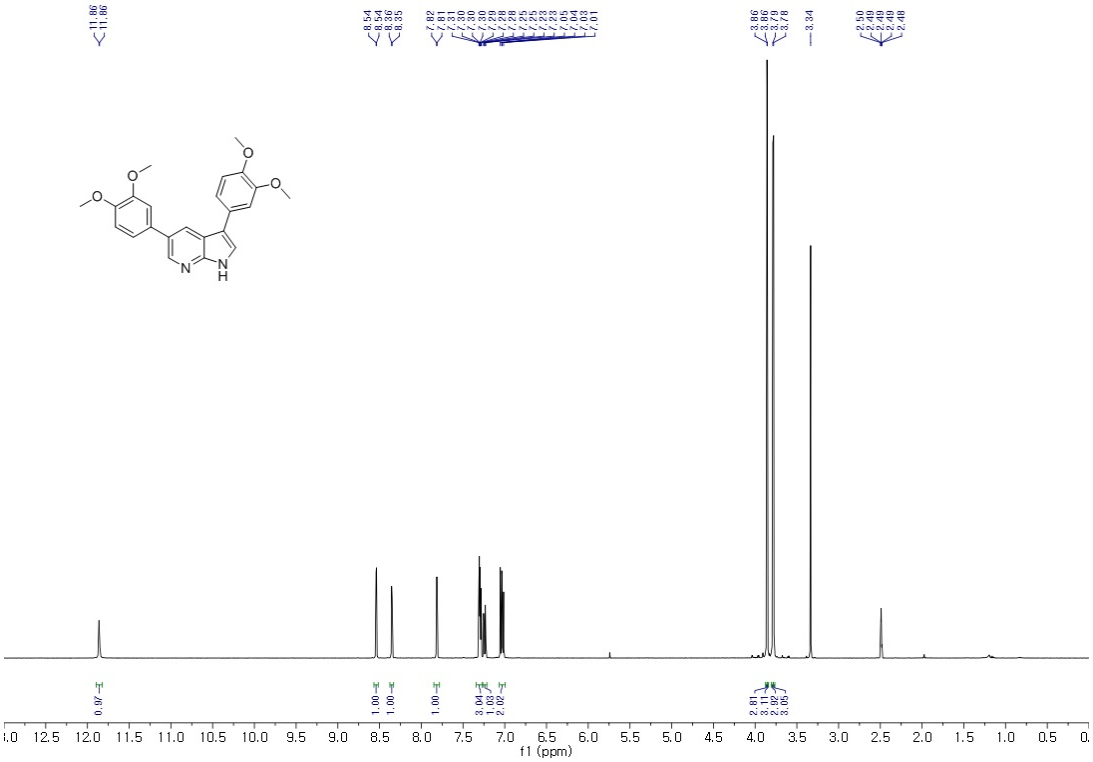



**400 MHz, $^{1}$H NMR in DMSO-$d_6$**

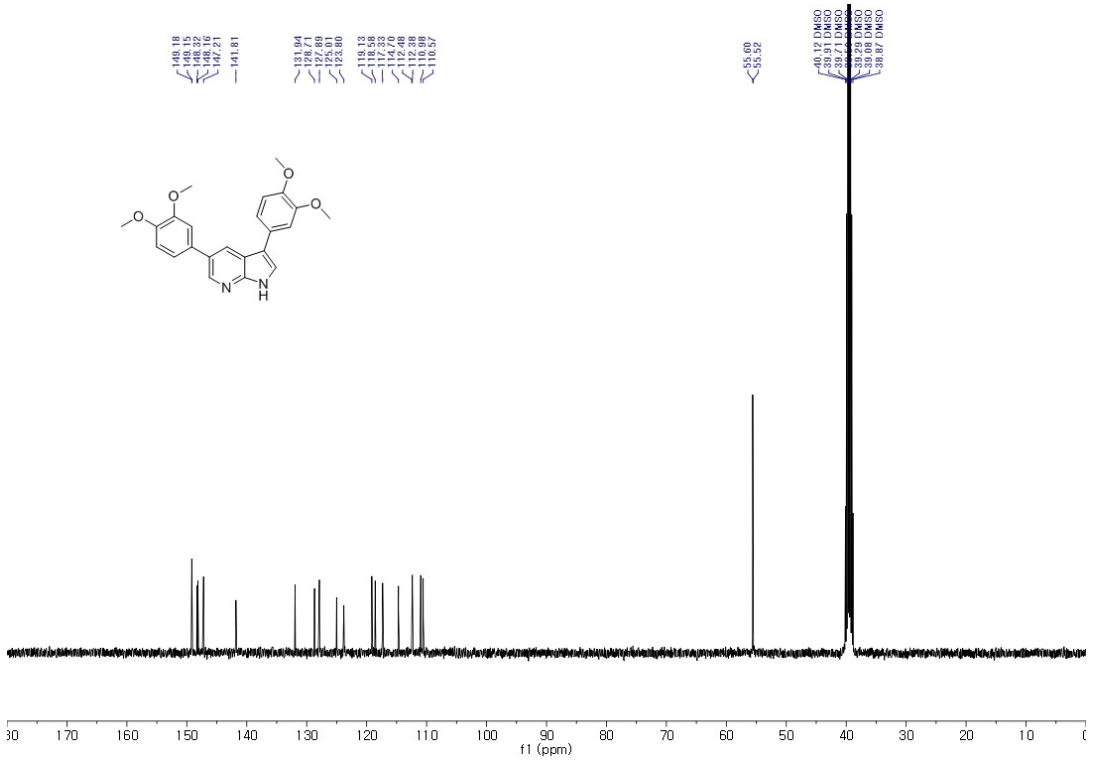



**100 MHz, $^{13}$C NMR in DMSO-$d_6$**

***3,5-bis(3,4-dimethoxyphenyl)-1H-pyrrolo[2,3-b]pyridine (17).***

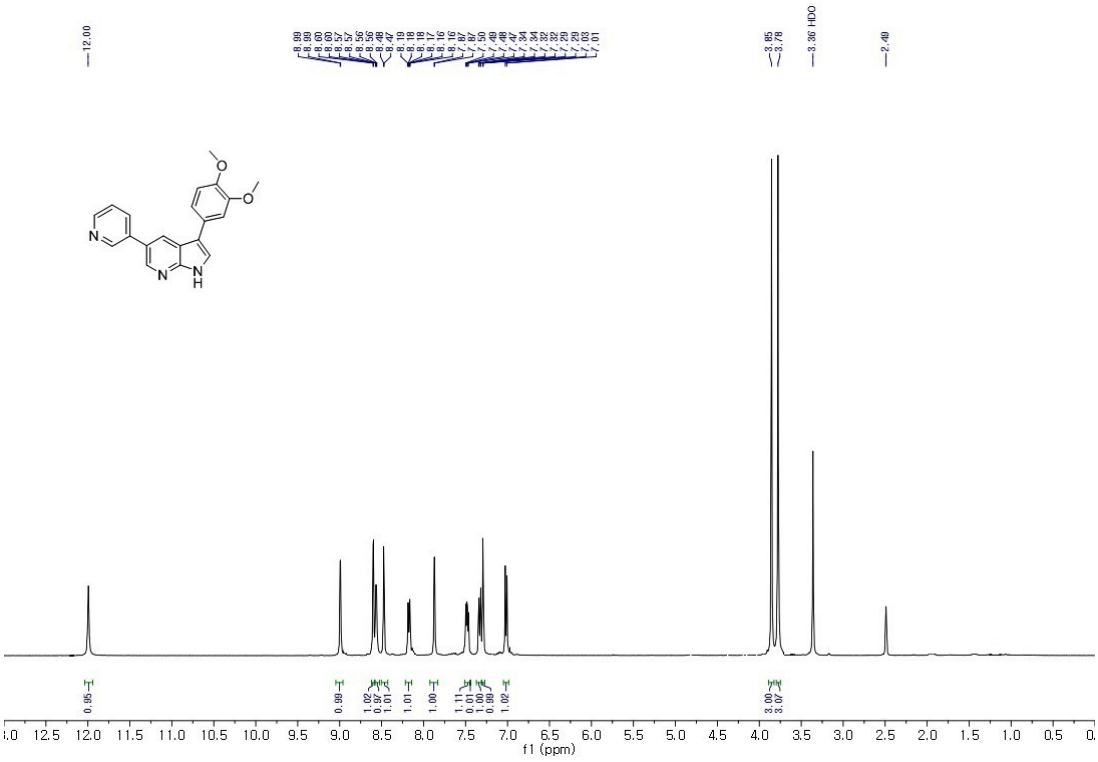



**400 MHz, $^{1}$H NMR in DMSO-$d_6$**

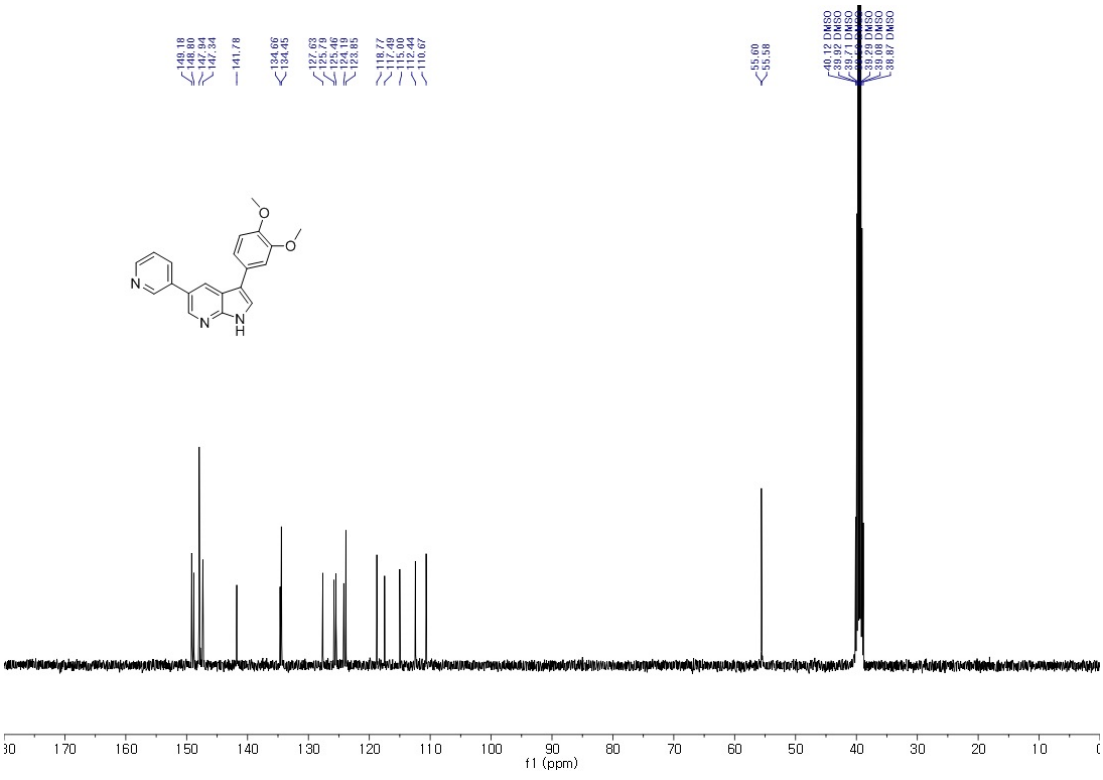



**100 MHz, $^{13}$C NMR in DMSO-$d_6$**

***3-(3,4-dimethoxyphenyl)-5-(pyridin-3-yl)-1H-pyrrolo[2,3-b]pyridine (18).***

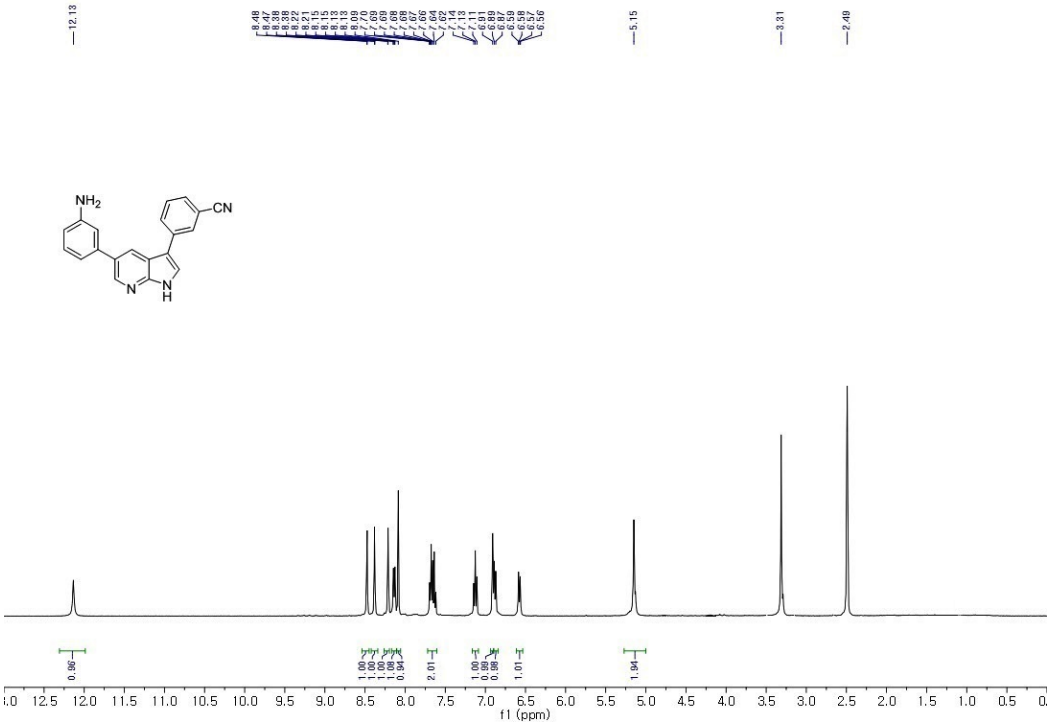



**400 MHz, $^{1}$H NMR in DMSO-$d_6$**

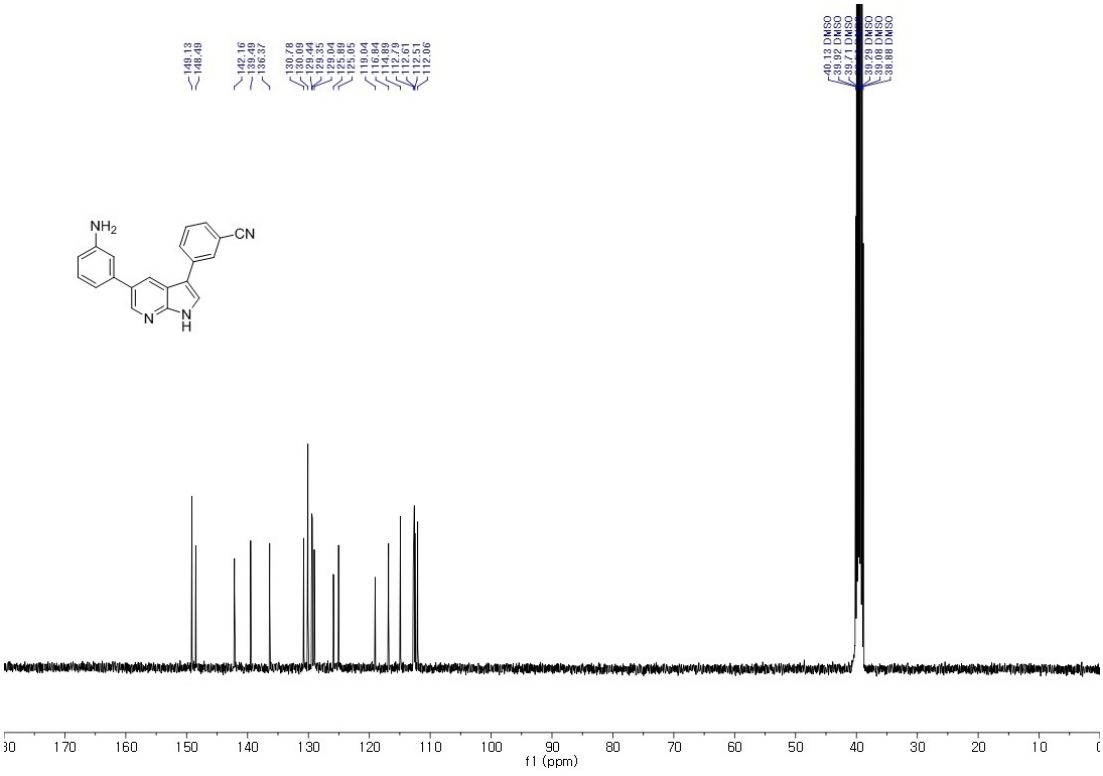



**100 MHz, $^{13}$C NMR in DMSO-$d_6$**

***3-(5-(3-aminophenyl)-1H-pyrrolo[2,3-b]pyridin-3-yl)benzonitrile (19).***

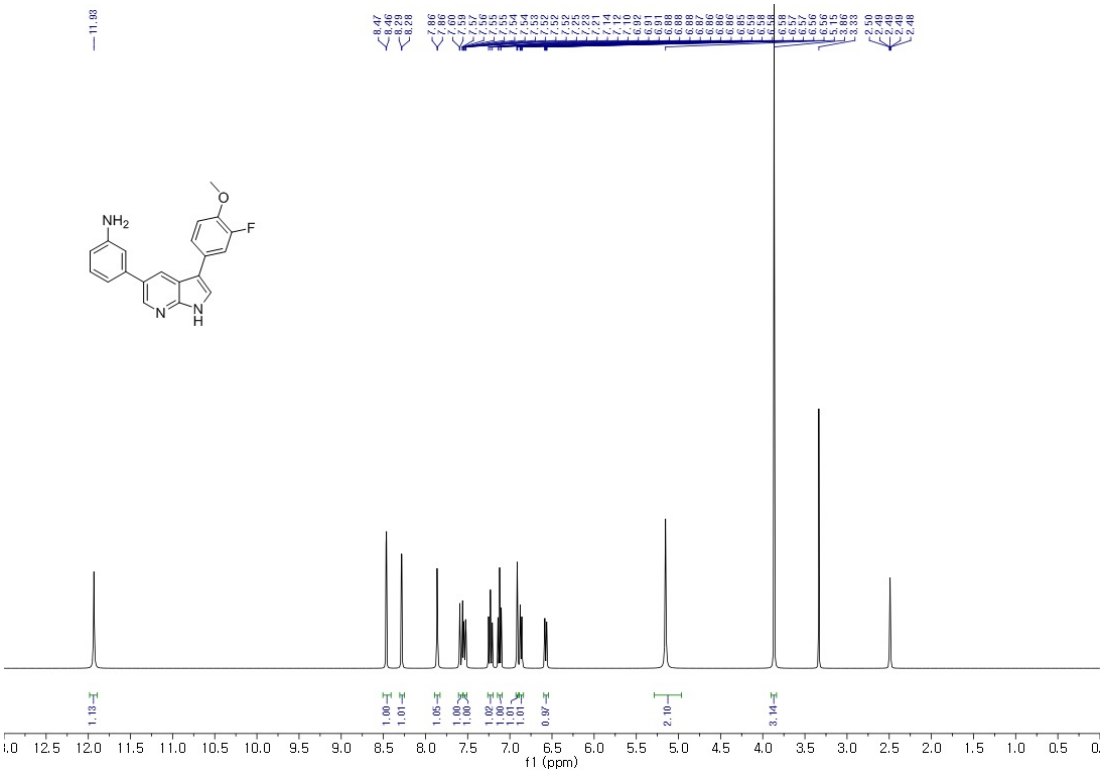



**400 MHz, $^{1}$H NMR in DMSO-$d_6$**

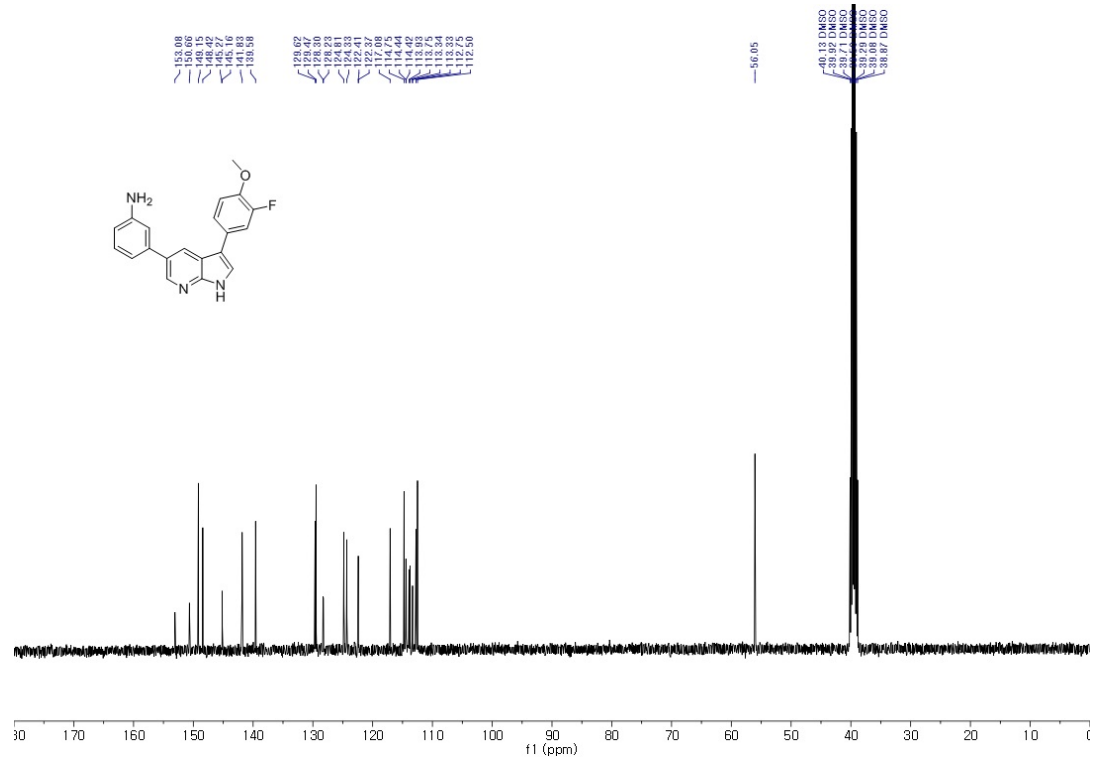



**100 MHz, $^{13}$C NMR in DMSO-$d_6$**

***3-(3-(3-fluoro-4-methoxyphenyl)-1H-pyrrolo[2,3-b]pyridin-5-yl)aniline (20).***

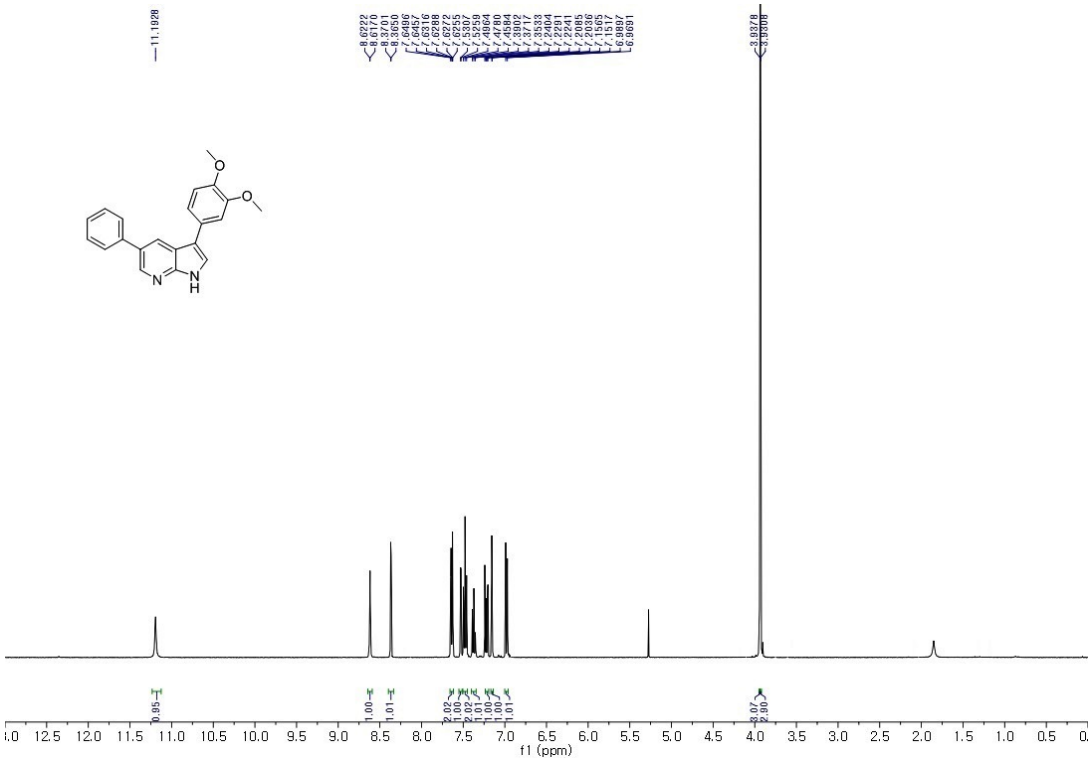



**400 MHz, $^{1}$H NMR in DMSO-$d_6$**

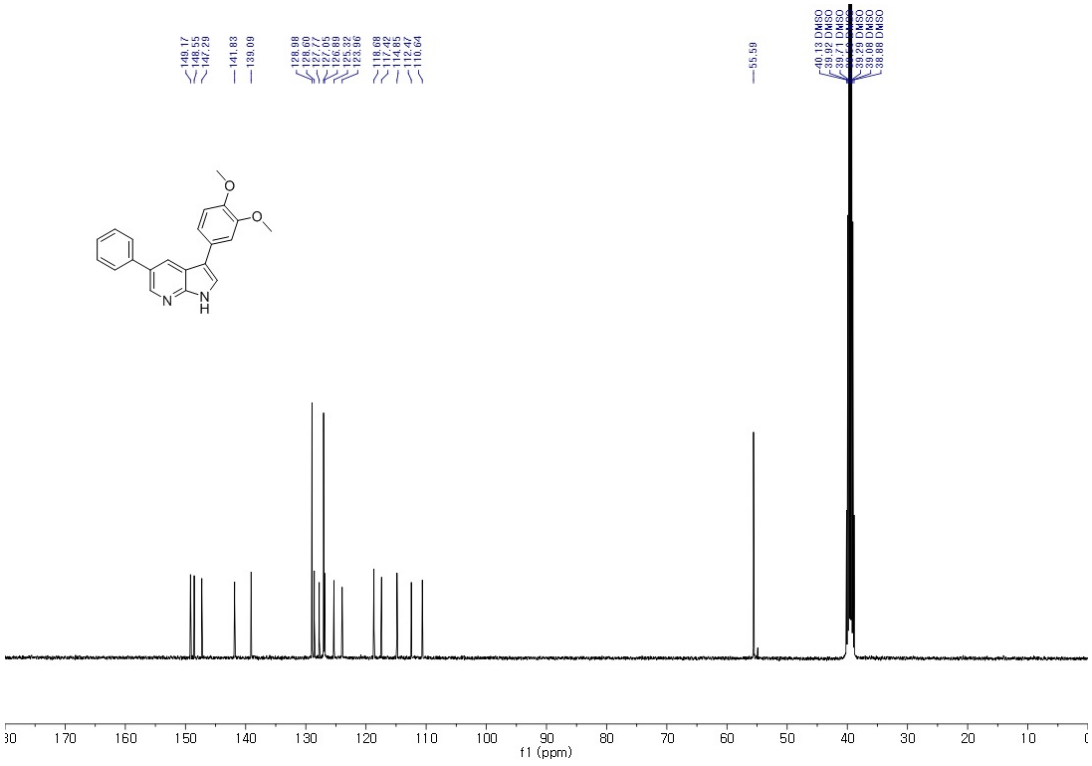



**100 MHz, $^{13}$C NMR in DMSO-$d_6$**

***3-(3,4-dimethoxyphenyl)-5-phenyl-1H-pyrrolo[2,3-b]pyridine (21).***

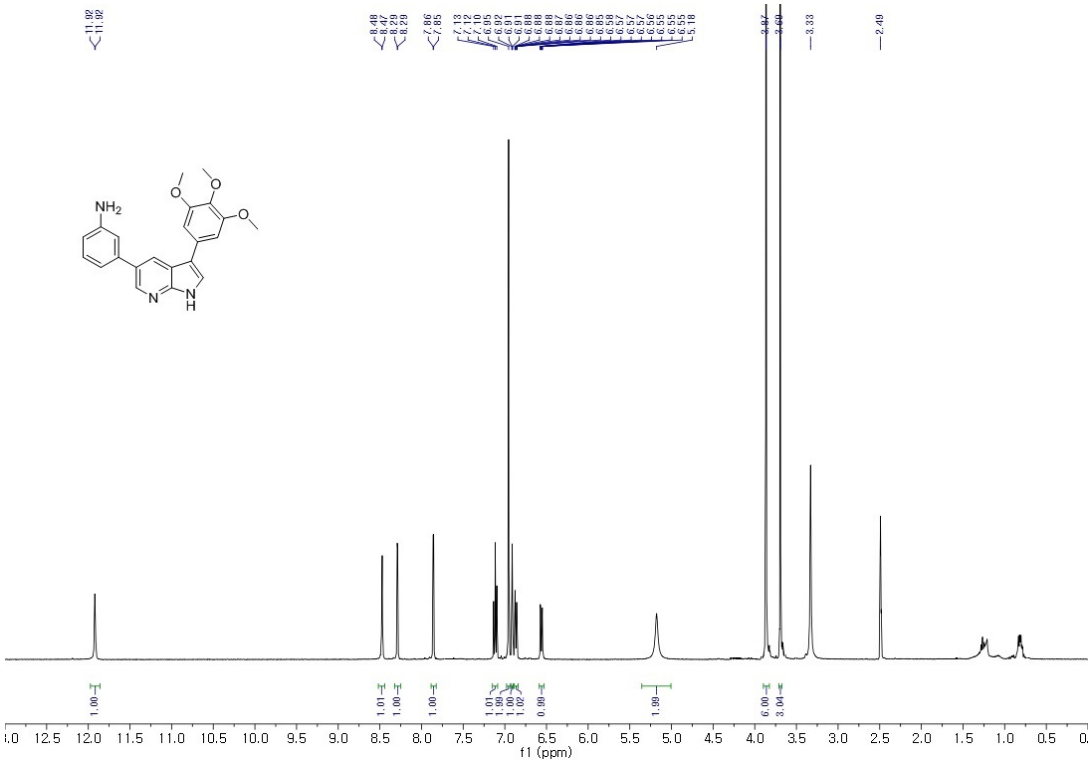



**400MHz, $^{1}$H NMR in Chloroform-*d***

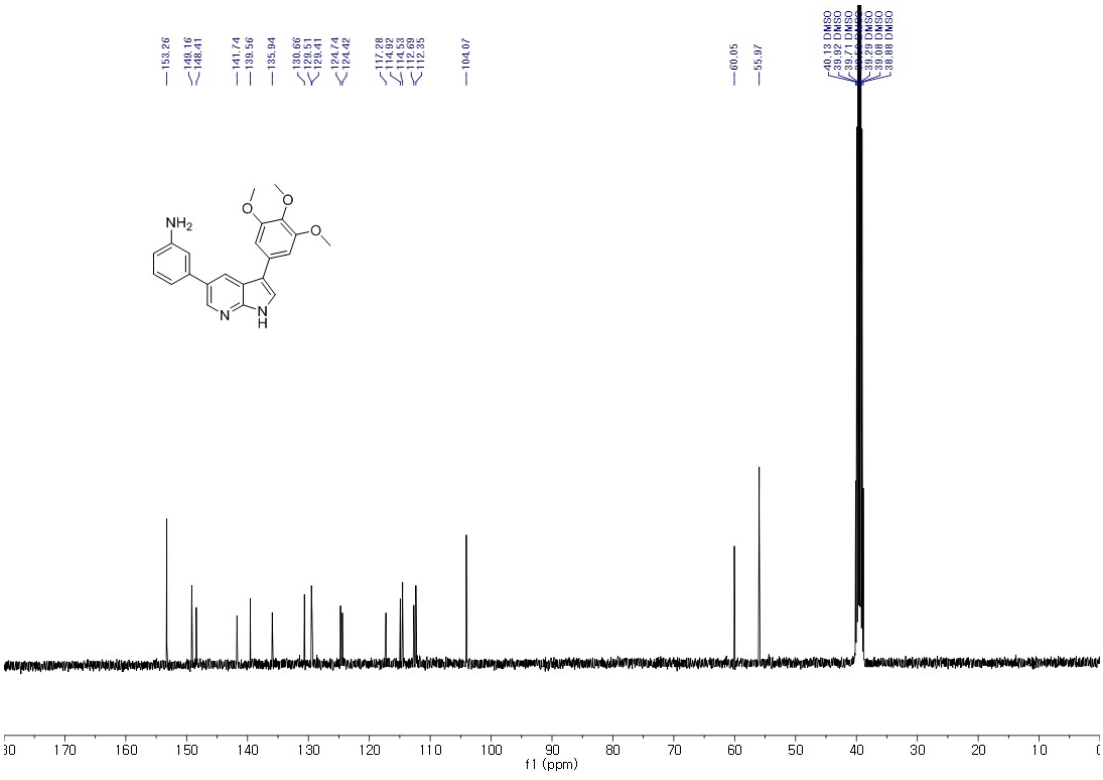



**100 MHz, $^{13}$C NMR in DMSO-$d_6$**

***3-(3-(3,4,5-trimethoxyphenyl)-1H-pyrrolo[2,3-b]pyridin-5-yl)aniline (22).***

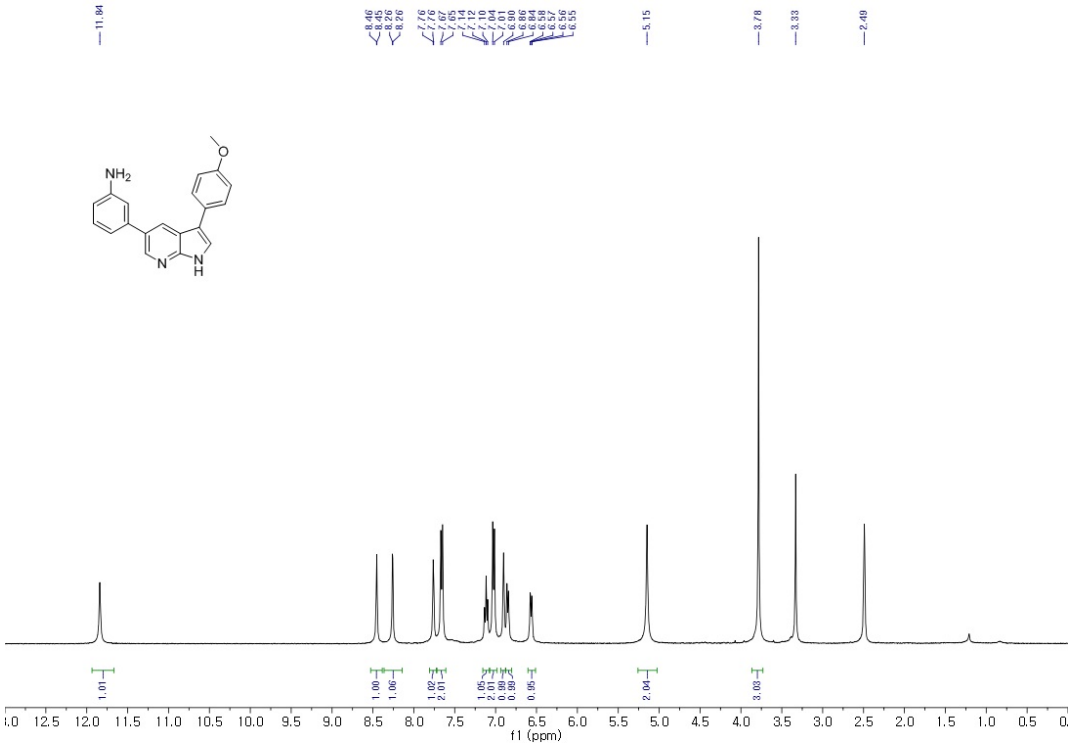



**400 MHz, $^{1}$H NMR in DMSO-$d_6$**

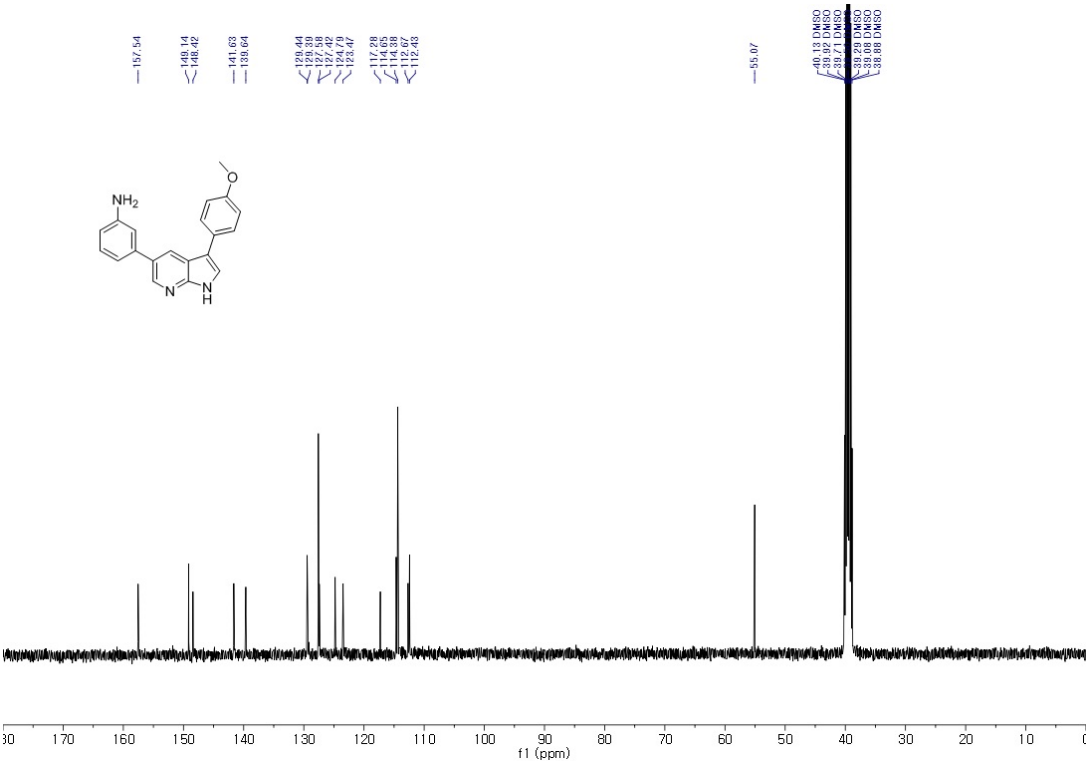



**100 MHz, $^{13}$C NMR in DMSO-$d_6$**

***3-(3-(4-methoxyphenyl)-1H-pyrrolo[2,3-b]pyridin-5-yl)aniline (23).***

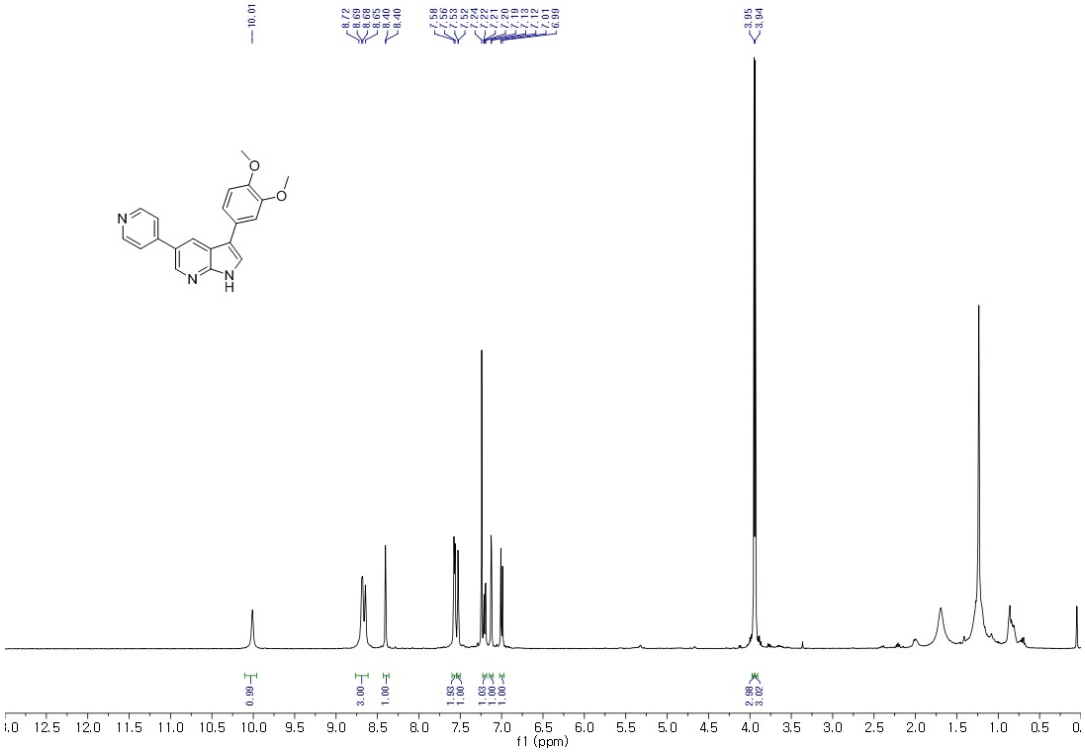



**400 MHz, $^{1}$H NMR in DMSO-$d_6$**

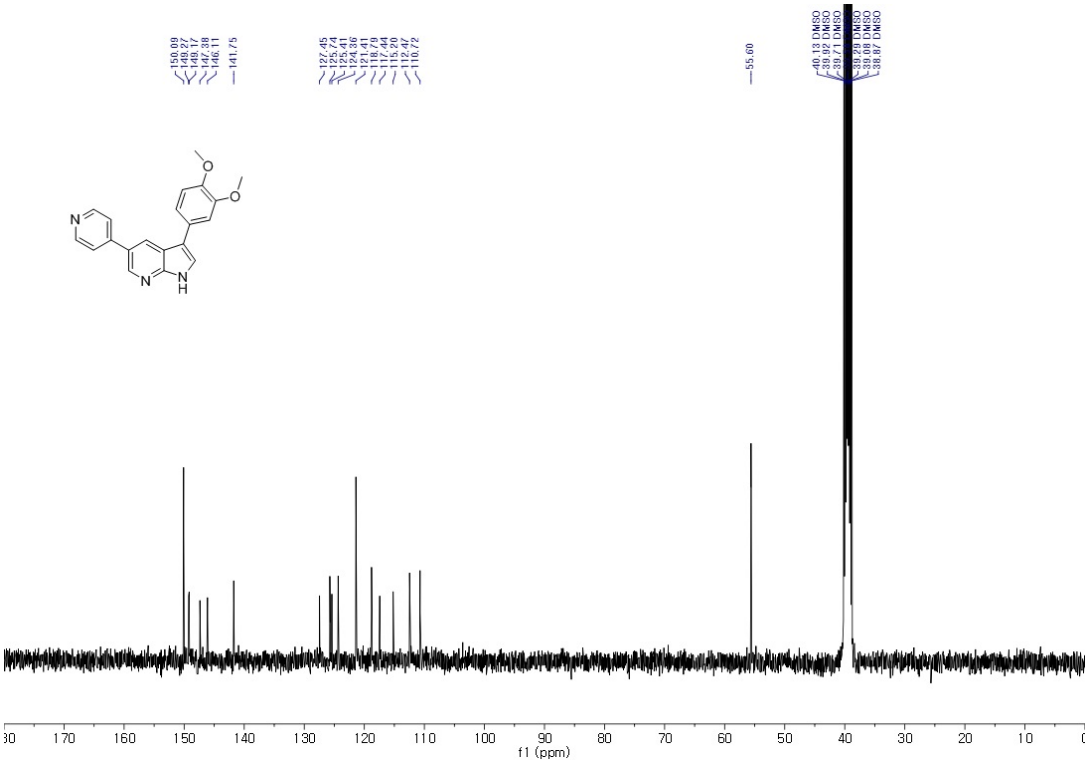



**100 MHz, $^{13}$C NMR in DMSO-$d_6$**

***3-(3,4-dimethoxyphenyl)-5-(pyridin-4-yl)-1H-pyrrolo[2,3-b]pyridine (24).***

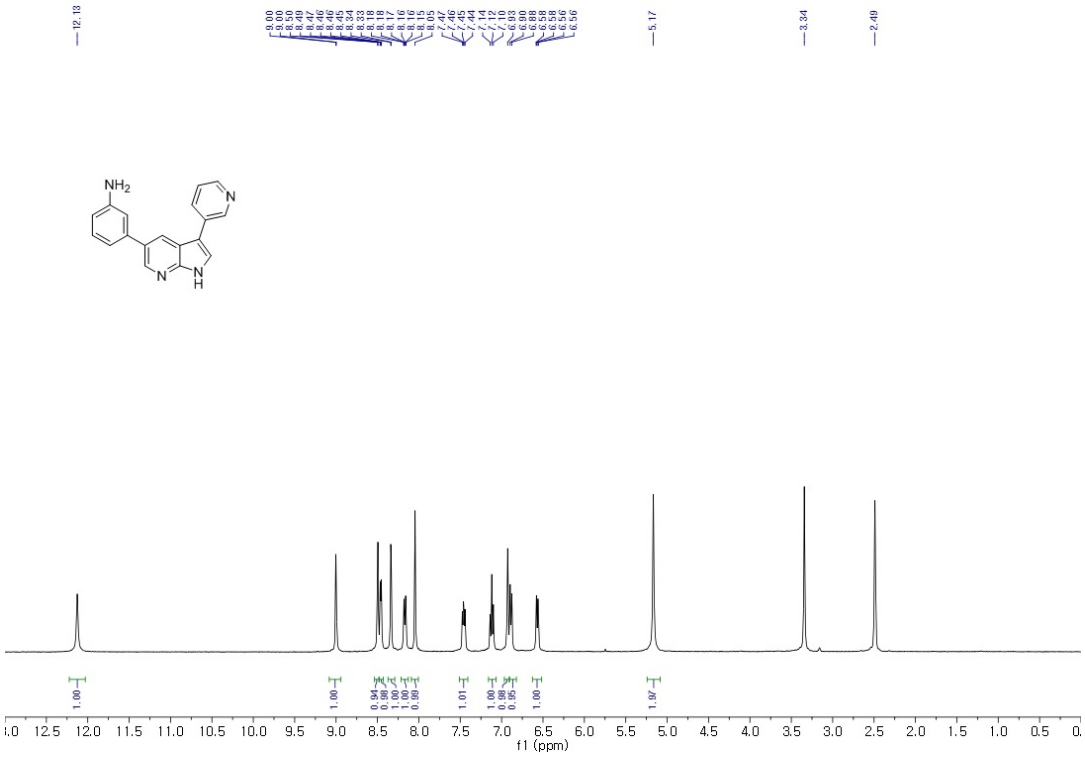



**400MHz, $^{1}$H NMR in Chloroform-*d***

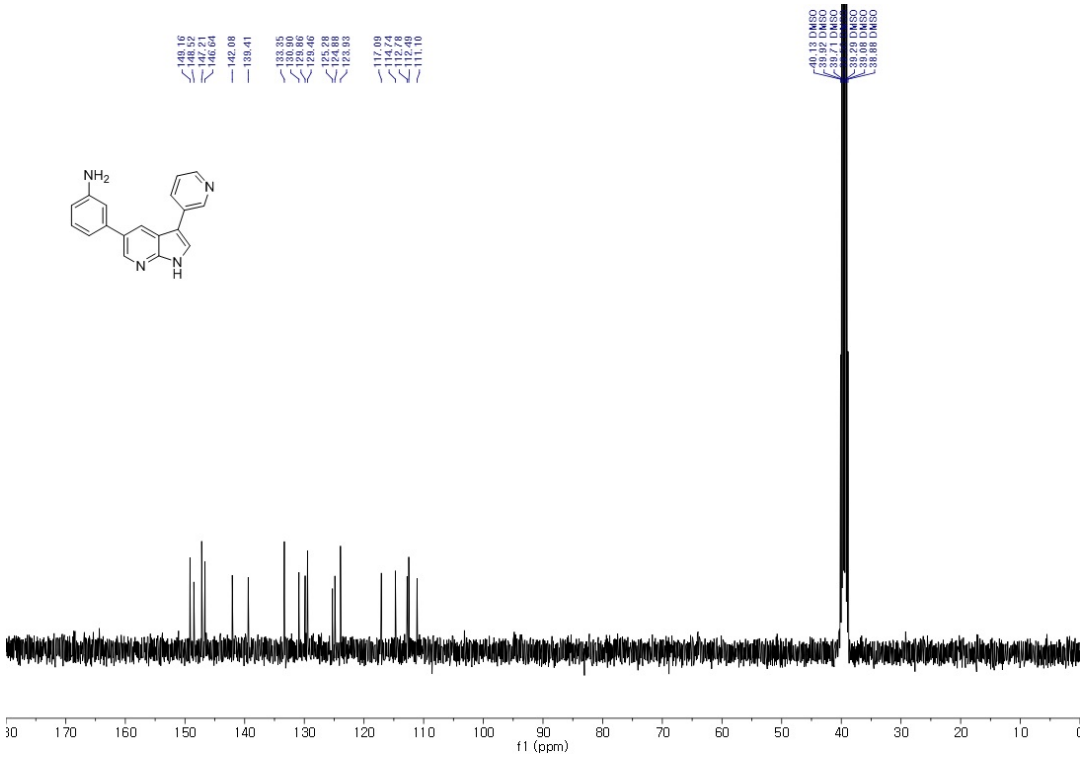



**100 MHz, $^{13}$C NMR in DMSO-$d_6$**

***3-(3-(pyridin-3-yl)-1H-pyrrolo[2,3-b]pyridin-5-yl)aniline (25).***

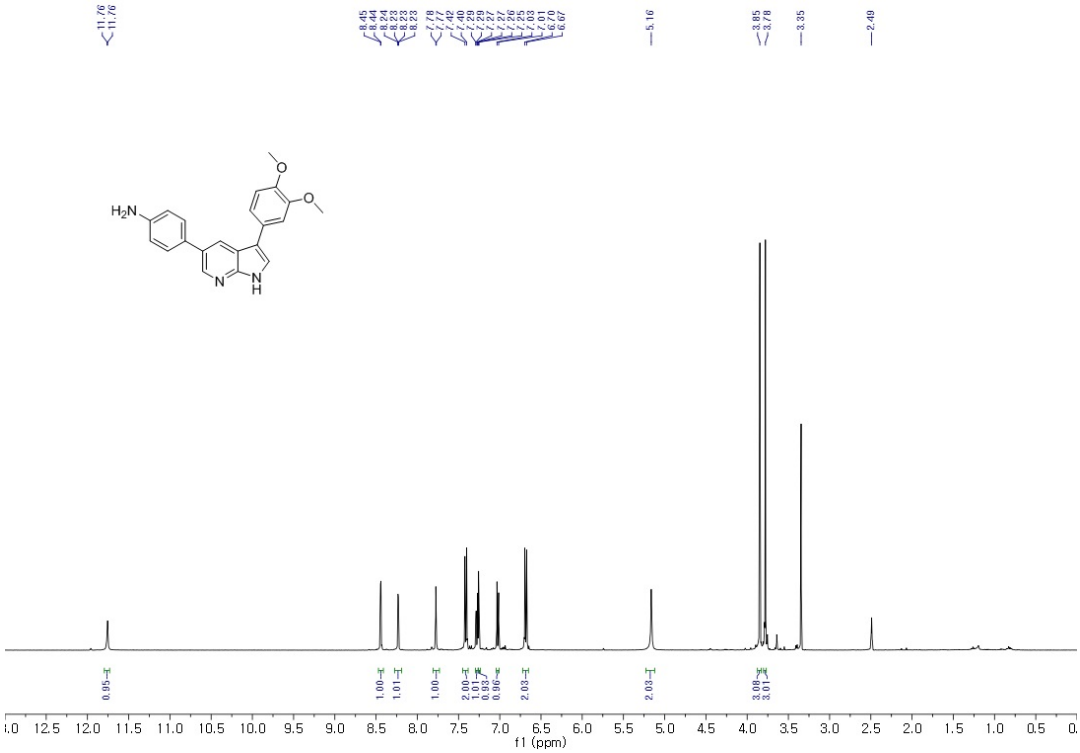


**400 MHz, $^{1}$H NMR in DMSO-$d_6$**

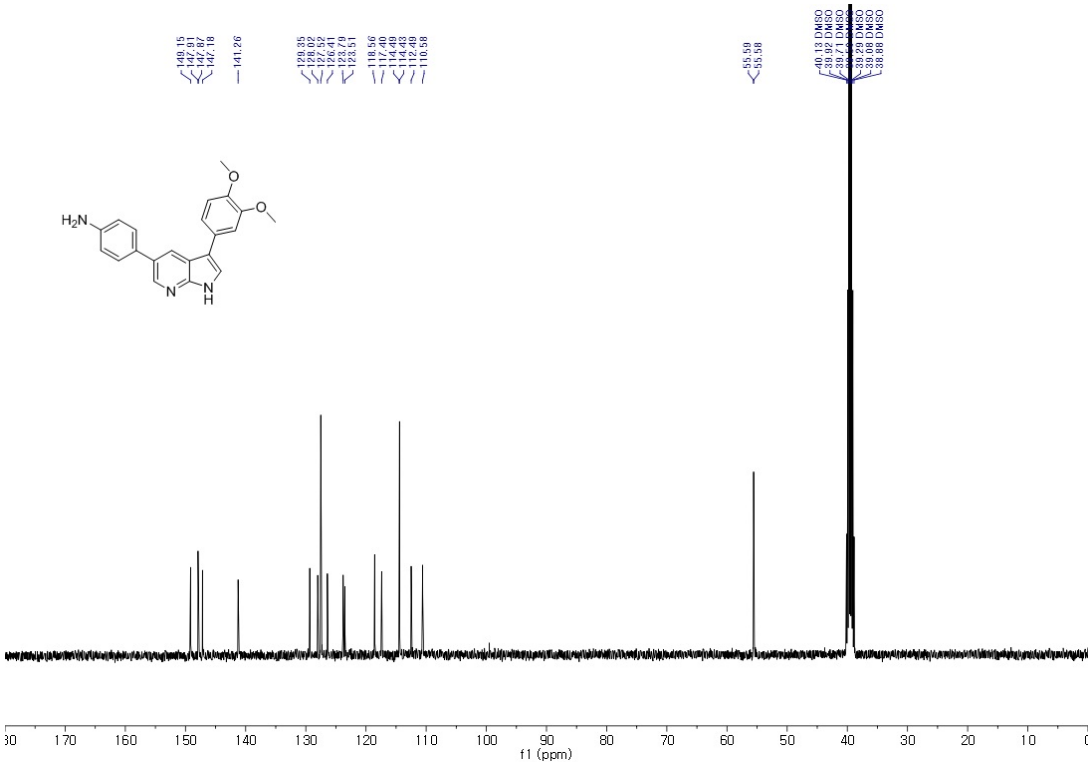


**100 MHz, $^{13}$C NMR in DMSO-$d_6$**

***4-(3-(3,4-dimethoxyphenyl)-1H-pyrrolo[2,3-b]pyridin-5-yl)aniline (26).***




**400 MHz, $^{1}$H NMR in DMSO-$d_6$**




**100 MHz, $^{13}$C NMR in DMSO-$d_6$**