

## Supporting Information

### A Threonine Turnstile Defines a Dynamic Amphiphilic Binding Motif in an AAA ATPase p97 Allosteric Binding Site

James C. Burnett,<sup>§</sup> Chaemin Lim,<sup>‡</sup> Brian D. Peyser,<sup>&</sup> Lalith P. Samankumara,<sup>‡</sup> Marina Kovaliov,<sup>‡</sup> Raffaele Colombo,<sup>‡</sup> Stacie L. Bulfer,<sup>||</sup> Matthew G. LaPorte,<sup>‡</sup> Ann R. Hermone,<sup>§</sup> Connor F. McGrath,<sup>§</sup> Michelle R. Arkin,<sup>||</sup> Rick Gussio,<sup>\*,&</sup> Donna M. Huryn,<sup>\*,‡,†</sup> and Peter Wipf<sup>\*,‡,†</sup>

<sup>§</sup>Leidos Biomedical Research, Inc., P.O. Box B, Frederick, MD 21702, United States

<sup>‡</sup>University of Pittsburgh Chemical Diversity Center, University of Pittsburgh, Pittsburgh, PA 15260, United States

<sup>&</sup>Frederick National Laboratory for Cancer Research, Developmental Therapeutics Program, P.O. Box B, Frederick, MD 21702, United States

<sup>||</sup>Department of Pharmaceutical Chemistry, Small Molecule Discovery Center, University of California, San Francisco, CA 94158, United States

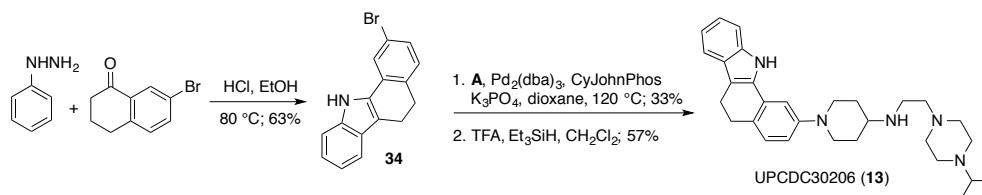
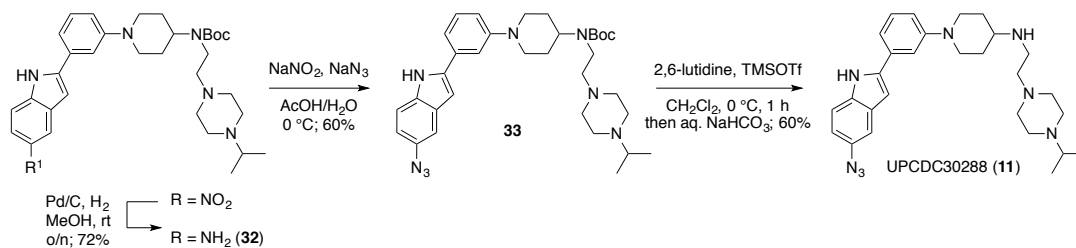
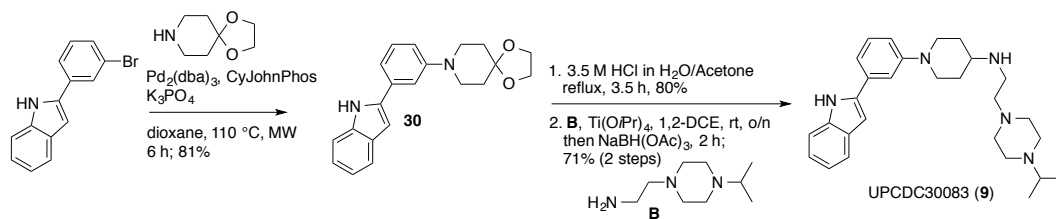
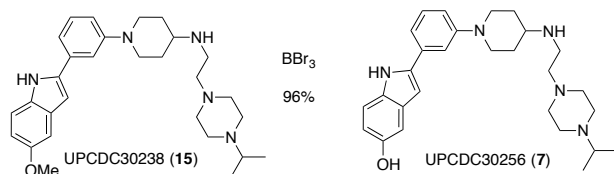
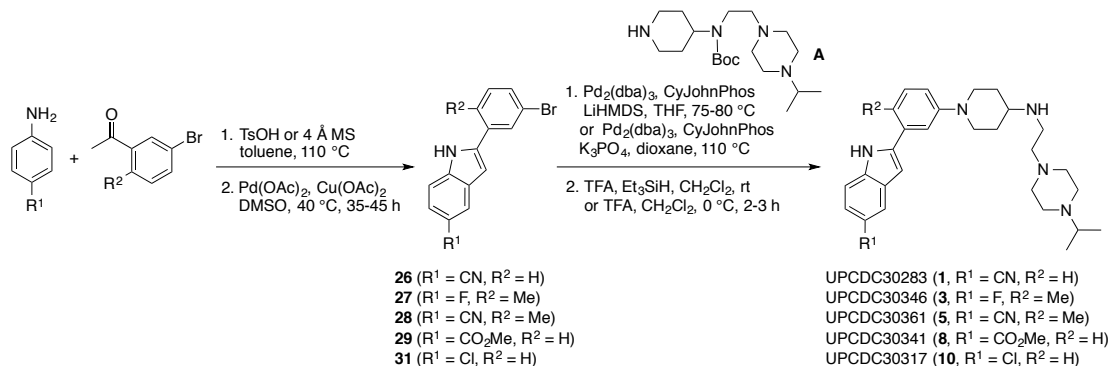
<sup>†</sup>Department of Pharmaceutical Sciences, University of Pittsburgh, Pittsburgh, PA 15261, United States

#### Contents

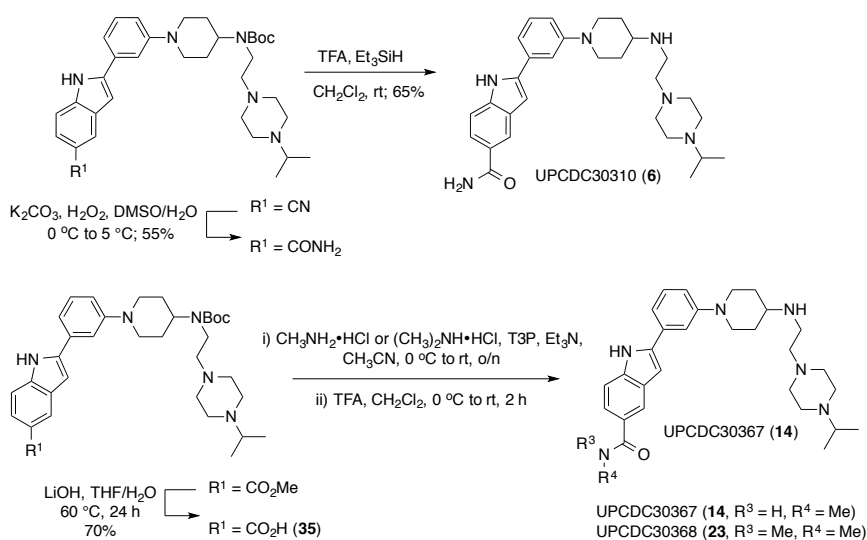
Synthetic Schemes .....	S2
Computational Methods.....	S5
Supplemental Figures. ....	S9
Supplemental Tables .....	S18
Description of the HINT Hydrophobic Scoring Function .....	S21
UPCDC30245 (4) Modeling and Scoring Method.....	S23
References.....	S25
Copies of <sup>1</sup> H NMR and <sup>13</sup> C NMR Spectra.....	S27

## Synthetic Schemes

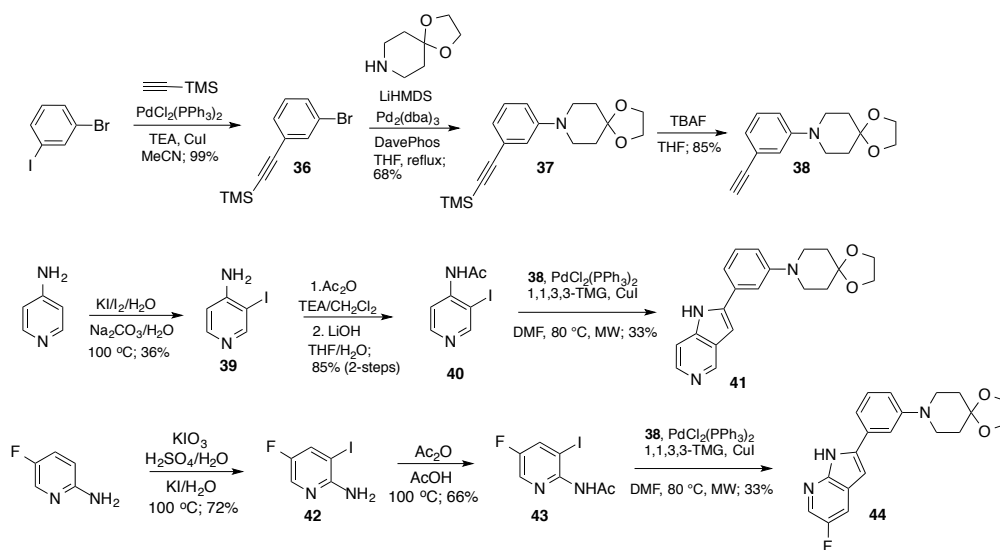
### Scheme 1. Syntheses of inhibitors **1**, **3**, **5**, **7**, **8**, **10**, and **13**.

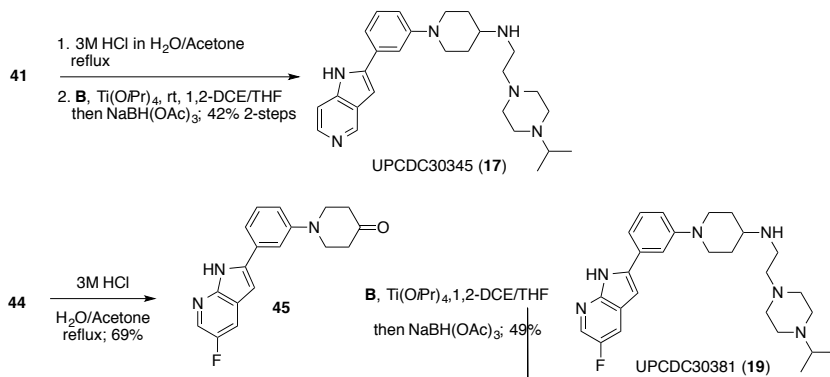


### Scheme 2. Syntheses of 5-substituted amide analogs **6**, **14**, and **23**.

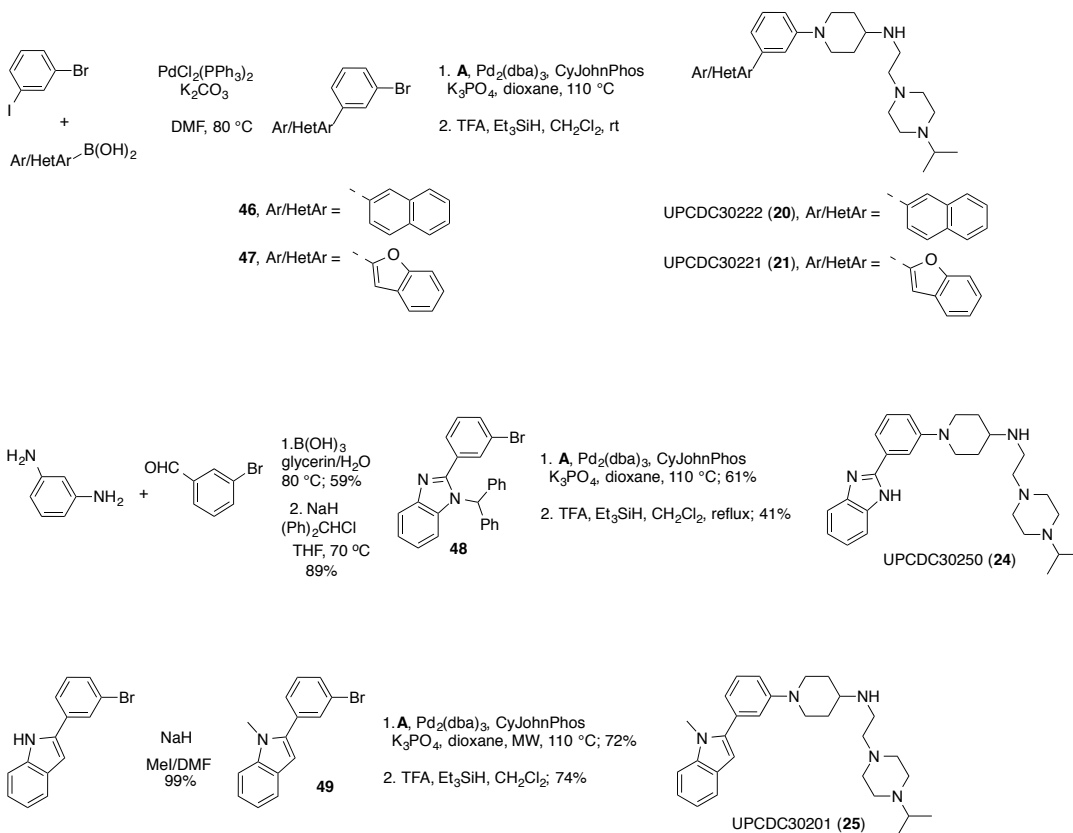


### Scheme 3. Syntheses of azaindole analogs **17** and **19**.





**Scheme 4.** Syntheses of other indole-replaced analogs (**20**, **21**, and **24**) and the *N*-methylated analog **25**.



## Computational Methods

*General.* All molecular modeling was performed on Linux workstations (Dell T7600). Insight II (2005) (Biovia, San Diego, CA), Discovery Studio 2016 (Biovia), and Maestro 2016 (Schrodinger, New York, NY) were used to build, energy refine, and inspect models, and for docking studies. During different stages of the calculations, various force fields were implemented, including (cff91, CHARMM, OPLS, MM2, and Amber). Materials Studio (Biovia), was used for quantum mechanics calculations to refine inhibitor structure geometries. For quantum mechanical optimizations and scanning, the mixed hybrid functional PW91 was employed during DFT calculations. The program HINT (eduSoft LC, Richmond, VA) was used to evaluate the quality of the models via the quantitation of intra- and intermolecular contacts (see Supplementary Information section 'Description of the HINT Hydrophobic Scoring Function' for program details). The HINT parameter settings used during the studies: steric term = Lennard-Jones 6-12 (for cff91 compatibility); lone pair vector focusing = 10; and distance dependence atom-atom interactions =  $\exp(1/r)$ .

*Initial protein refinement.* The cryo-EM structure with inhibitor UPCDC 30245 (**4**) bound to an allosteric site on p97 was used for model coordinates (PDB entry 5FTJ, resolution = 2.3 Å).<sup>1</sup> Two protomers (residues Asn 21 – Gln 763, chains A and B), encompassing the allosteric binding site of one inhibitor and the binding sites of four ADP molecules, were used for the docking studies. In the cryo-EM structure, the tetramine side chain of the inhibitor was oriented into the solvent. Therefore, torsional adjustments were made to orient this component of the inhibitor toward complimentary residue contacts. Subsequently, a tethered minimization strategy was used to energy refine the structure in preparation for further modeling studies. During the simulation, the distance-dependent dielectric was set to 1.0, and the entire system was hydrated with an 8.0 Å thick shell of water molecules. Following, hydrogen atoms only were optimized (100 steps steepest descents followed by conjugate gradients until the norm of the gradient was  $<0.01$  kcal/Å). Next, the side chains were relaxed using the same protocol, but with only backbone atoms and the

heavy atoms of the inhibitor phenyl-indole component and the four ADP molecules fixed in coordinate space. Finally, with only the heavy atoms of the ADP molecules fixed, a tethered minimization procedure was used to fully energy optimize the complex (protein, inhibitor, ADP molecules, and waters). For this stage of the optimization, 2000 kcal/mol per Å<sup>2</sup> of force was applied to the system, and then reduced by 100 kcal/mol decrements by minimizing with iterations of 100 steps of steepest descents followed by 200 steps of conjugate gradients until all external force was removed and the norm of the gradient was <0.001 kcal/Å.

*Optimization of the bis-Thr subsite.* A simulated annealing protocol was used to bring the side chain hydroxyl groups of loop residues Thr 509 and Thr 613 to within hydrogen bonding distance near the indole 5-F position of the inhibitor in the energy refined model of the p97 structure (indicated in the section above). Using closest distances observed between the β-carbons of the Thr 509 and Thr 613 residues of examined p97 PDB entries as guide posts, a distance constraint of  $4.0 \pm 1.0$  Å was imposed on the β-carbons of these two residues. A force constant of 100 kcal/per mol Å<sup>2</sup> was applied to the distance constraint, and the entire system was heated to 300 K for 1 ps, followed by 200 iterations of conjugate gradients minimization. This procedure iterated until the norm of the gradient was 0.01 kcal/Å. In the final step, the force constant was removed, and the structure was minimized using the procedure described in the section above. The final coordinates of the energy-refined structure were within the experimentally determined cryo-EM structure resolution; the final structure possessed an RMS deviation of 1.87 Å across all backbone atoms compared to the starting coordinates of the cryo-EM structure. The final distance between the β-carbons of Thr 509 and Thr 613 was 4.43 Å. In the refined structure, the side chain hydroxyl groups of Thr 509 and Thr 613 engaged in a hydrogen bond (2.2 Å distance), and coordinated a water molecule, thereby providing a polar subsite in the allosteric pocket. Rotation of the residues' χ<sub>1</sub> torsions by 150-180° resulted in the threonine side chain methyls inducing an apolar (or hydrophobic) subsite in the allosteric pocket.

*Inhibitor docking.* Inhibitors listed in Table 1 were manually docked in the p97 allosteric site using the Insight II graphical user interface, which allows for detailed small molecule positional and multi-torsional manipulation via dial box control. Initially, each inhibitor was placed in the allosteric site based on the coordinates of the energy refined binding mode of UPCDC 30245 (**4**). With the van der Waals bump set to 0.25 Å, translational, rotational, and torsional adjustments, as well as small torsional adjustments to protein residue side chains (within acceptable limits defined by rotamer libraries and empirical data) were used to eliminate unacceptable atom-atom overlaps and model optimal non-bonded interactions known to occur in high resolution X-ray structures. Inter- and intramolecular analyses of the docked inhibitors were performed using the HINT program, and iterative rounds of additional manual adjustments (translational, rotational, and torsional), tethered minimizations, and HINT scoring were used to direct precise, SAR-guided micro-environmental adjustments between protein side chain atoms and their corresponding inhibitor contacts. This method provided comprehensive, structure-based rationales for the *in vitro* binding data. In the final models, the total HINT scores for inhibitor-protein complexes correlated well with their respective empirical  $-\text{Log IC}_{50}$  values (Table S2 and Figure S5). Note, because the amphiphilic 5-F indole congener (UPCDC 30245 (**4**)) can bind in the allosteric site in both the polar and apolar bis-Thr conformations, its docking and scoring method is described in detail in below in section: *UPCDC30245 (4) Modeling and Scoring Method*.

*PDB analysis for bis-Thr pairs.* The PDB was searched using customized programs written in Python and the Schrodinger Maestro API.<sup>2</sup> **Search 1:** All PDB entries solved by X-ray crystallography with: 1) resolution  $\leq 2.5$  Å, 2)  $\geq 200$  residues in the longest chain, and 3) containing a ligand, were evaluated. The resulting collection of 19,881 entries was searched using customized programs written in Python and the Schrodinger Maestro API. The bis-Thr feature was defined as a pair of Thr residues with  $\beta$ -carbons  $\leq 5$  Å apart, with: 1) both Thr residues in secondary structural

features defined as loops, 2) both Thr residues  $\geq 5$  residues apart in the sequence, and 3) a ligand  $\leq 12$  Å away from either of the Thr  $\beta$ -carbons. Phosphate containing compounds (*e.g.s.*: ADP, ATP, NAD, FAD, etc.) were eliminated to bias the search toward ligand binding sites. The search resulted in the identification of 840 distinct PDB entries meeting the above criteria. All 840 entries were manually evaluated for potential interactions between bis-Thr pairs and ligands. **Search 2:** The method was the same as indicated for Search 1, but with the following modifications: 1) both Thr residues could occur in any secondary structure (*i.e.*, helices, sheets, and loops) and 2) the ligand was required to be  $\leq 5$  Å away from either of the Thr  $\beta$ -carbons. The search resulted in the identification of 263 distinct PDB entries meeting these criteria. All 263 entries were manually evaluated for potential interactions between bis-Thr pairs and ligands, with a main criterion being that at least one of the Thr residues must interact directly with the ligand. **Search 3:** All PDB entries generated by NMR with pairs of Thr residues with  $\beta$ -carbons  $\leq 5$  Å apart were evaluated. Following, using a customized program written in R, a subset of 8,687 non-redundant PDB entries (with each containing multiple conformer states), were analyzed and grouped by Uniprot identifiers,<sup>3</sup> and resulted in the identification of 271 entries, covering 243 distinct Uniprot identifiers, in which the  $\chi$  torsions of both threonine residues (*i.e.*, of the Thr pair) rotated by at least 50°. These hits were manually inspected for potential Thr-Thr turnstile interactions. Several examples of bis-Thr pairs engaging in cooperative, gear-like interactions in less solvent exposed protein sites, which would be more consistent with where a ligand might bind include PDB entries: 1YDU, 1YYC, 2FIN, 2G2K, 2LEN, and 2RLI.



**Supplementary Figures**

S1 – S3: CPK volume calculations of 5-substituted phenylindoles were calculated in Spartan 10 (Wave Function, Inc., Irvine, CA) with AM1 parametrization.

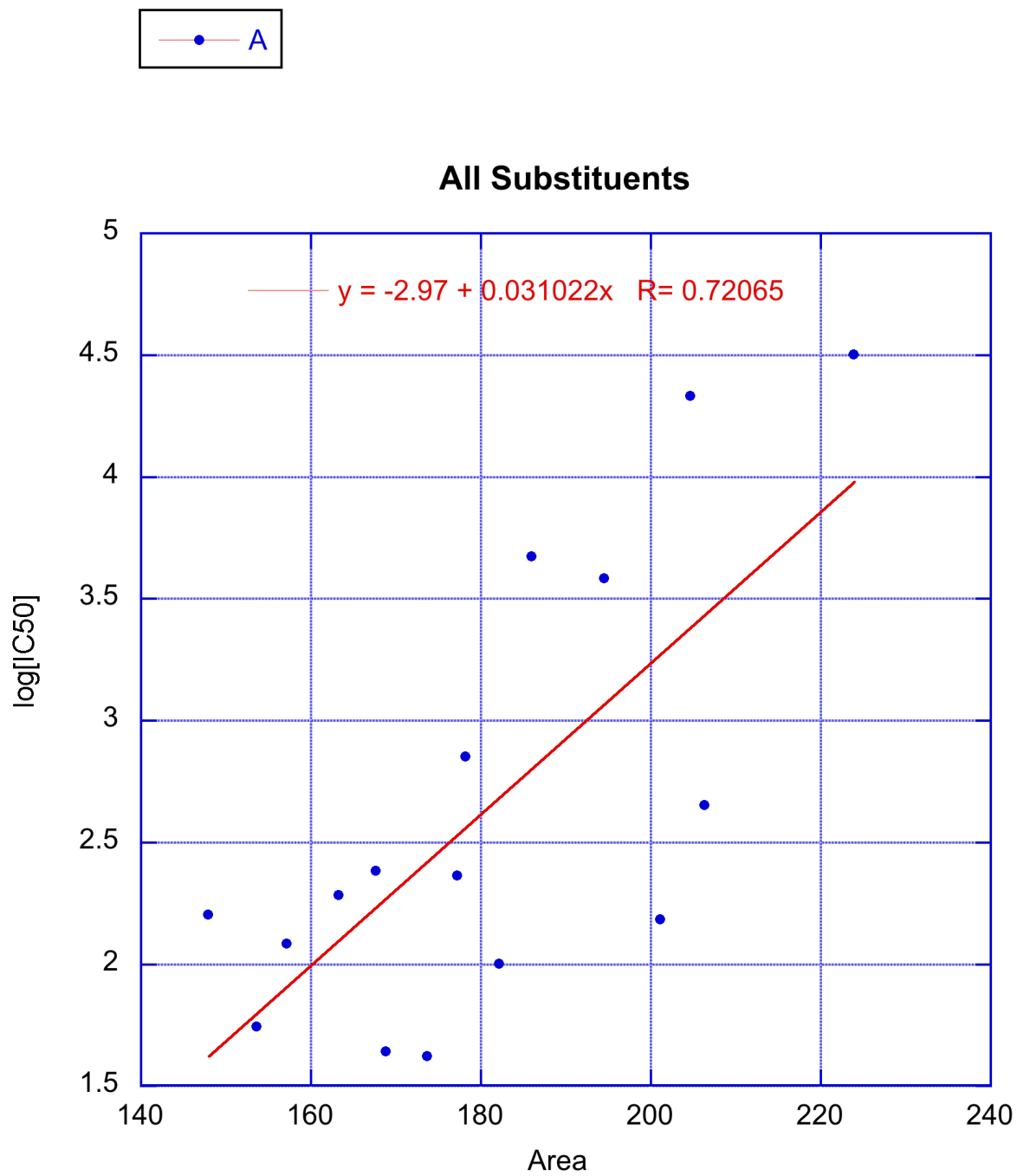
**Figure S1.**

Figure S2.

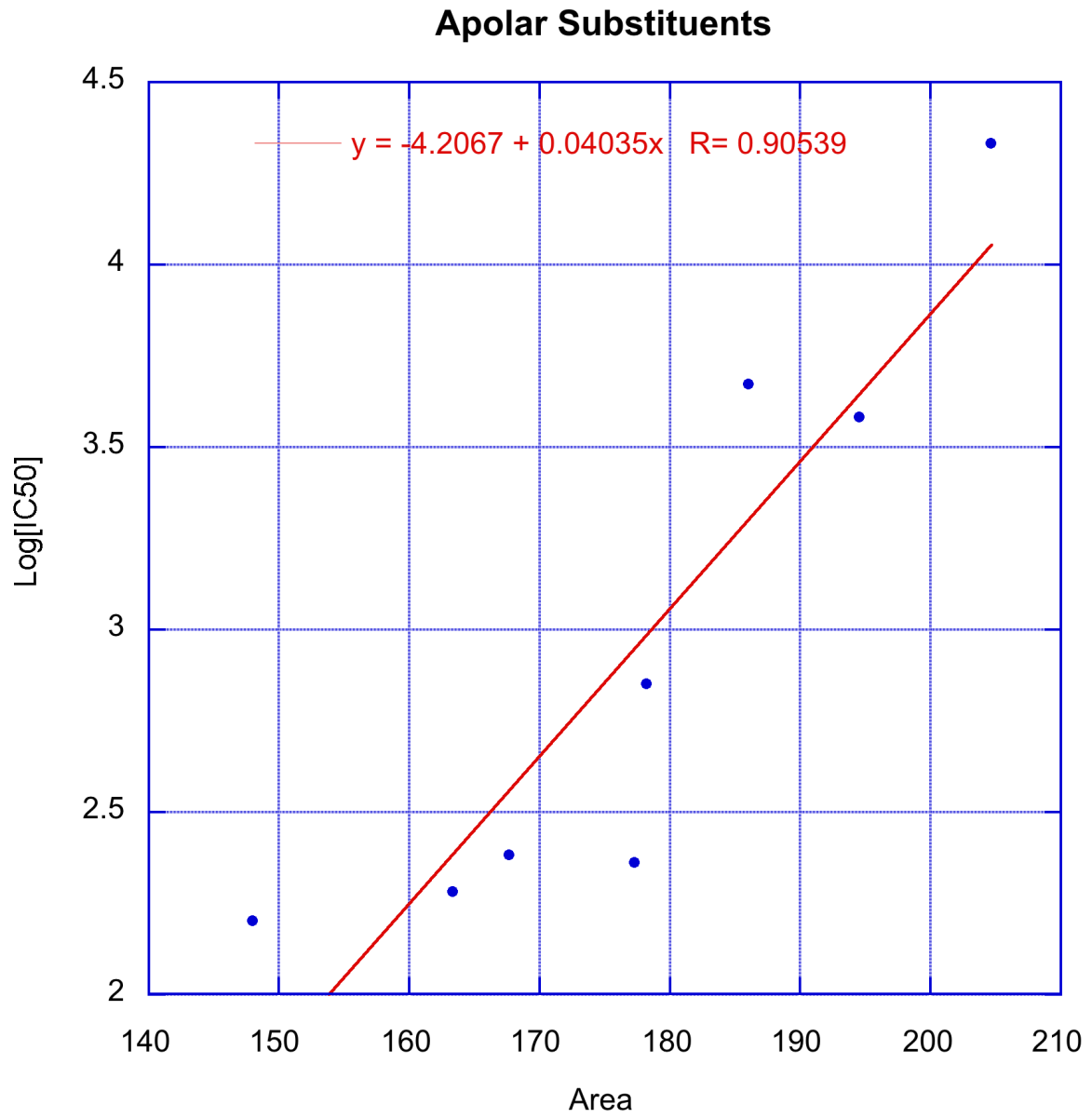
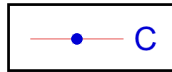
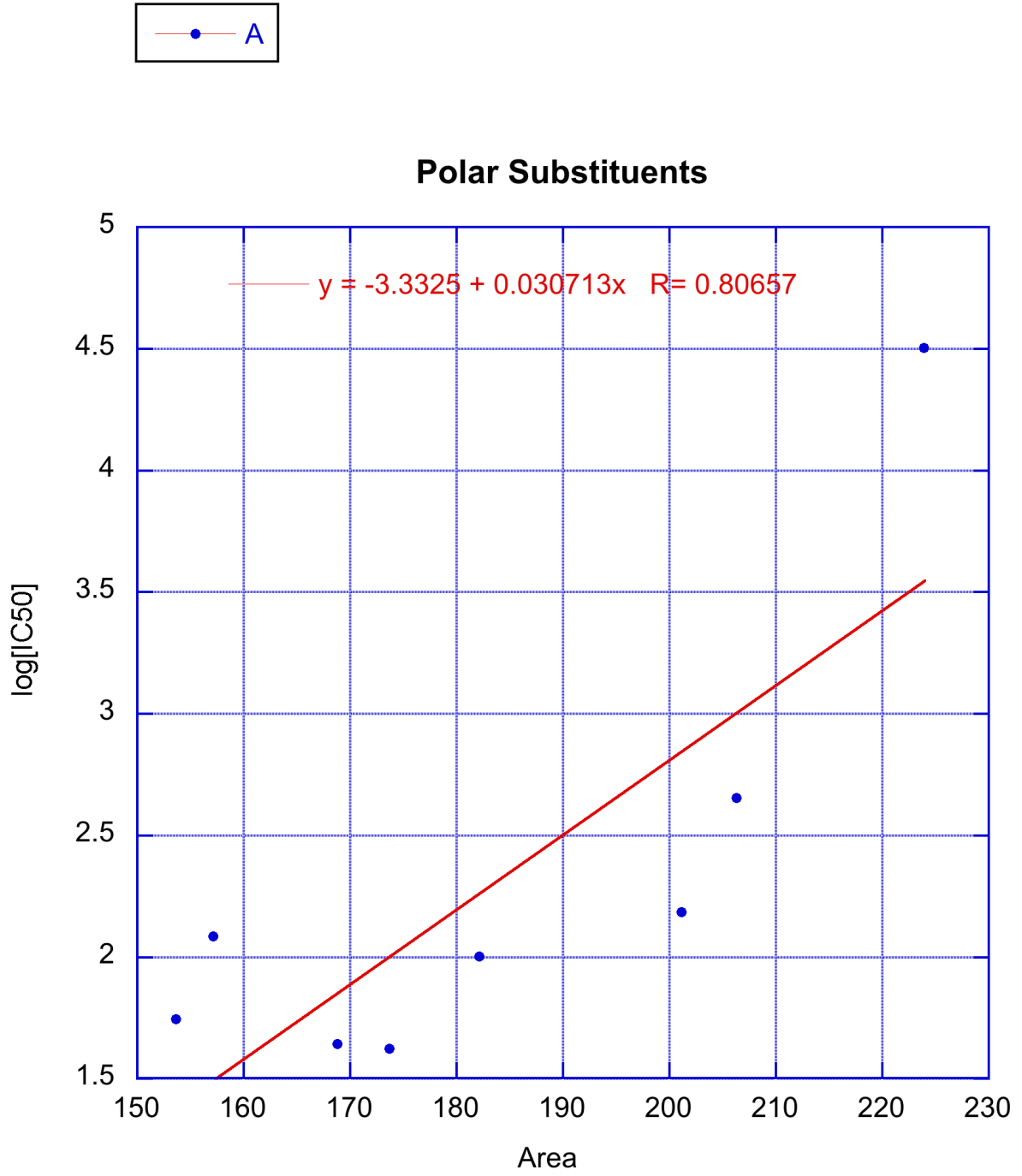
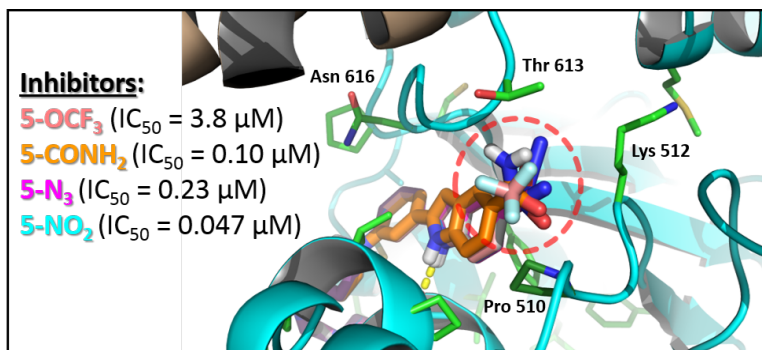
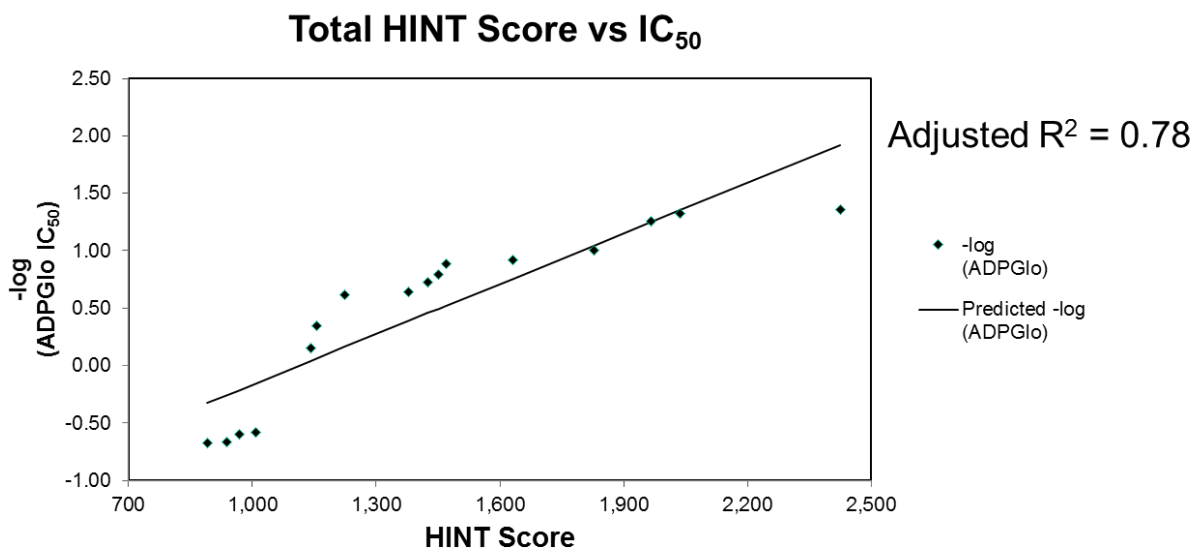


Figure S3.

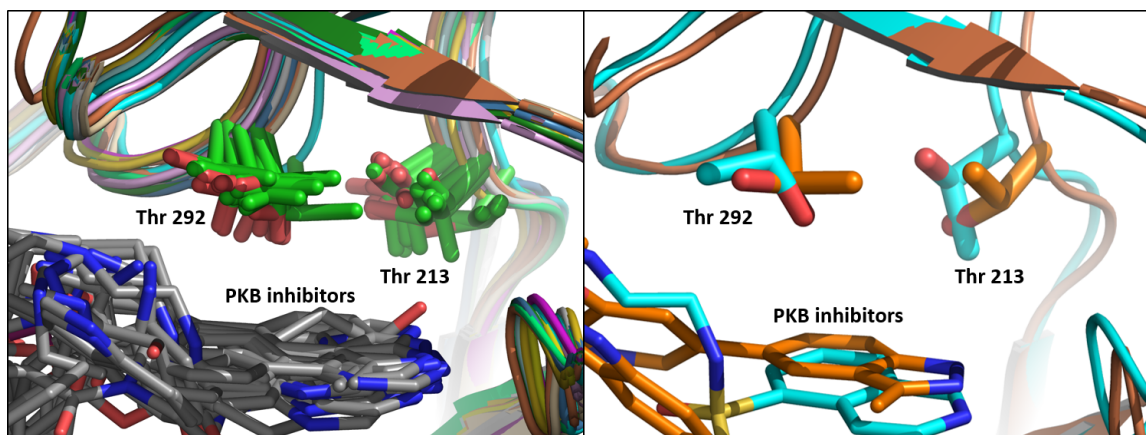




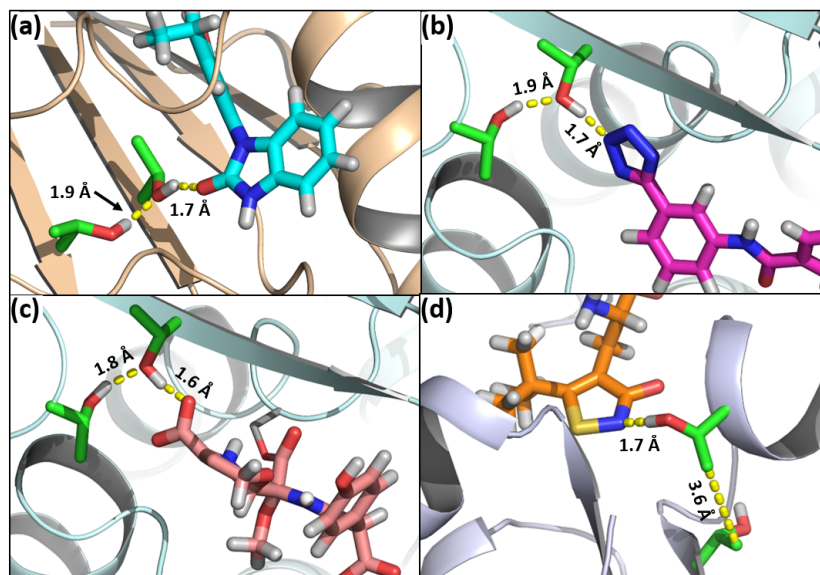
**Figure S4.** Docking inhibitors in unmodified co-cryo-EM structure 5FTJ<sup>1</sup> did not provide structure-based rationales to explain the SAR. For example, all of the inhibitors shown in the Figure are sterically accommodated by the binding site (red dashed circle), with no residue contacts, favorable or unfavorable, that distinguish relative inhibitory potencies. A hydrogen bond between the inhibitors and the backbone carbonyl oxygen of Val 493 is represented by a yellow dash.



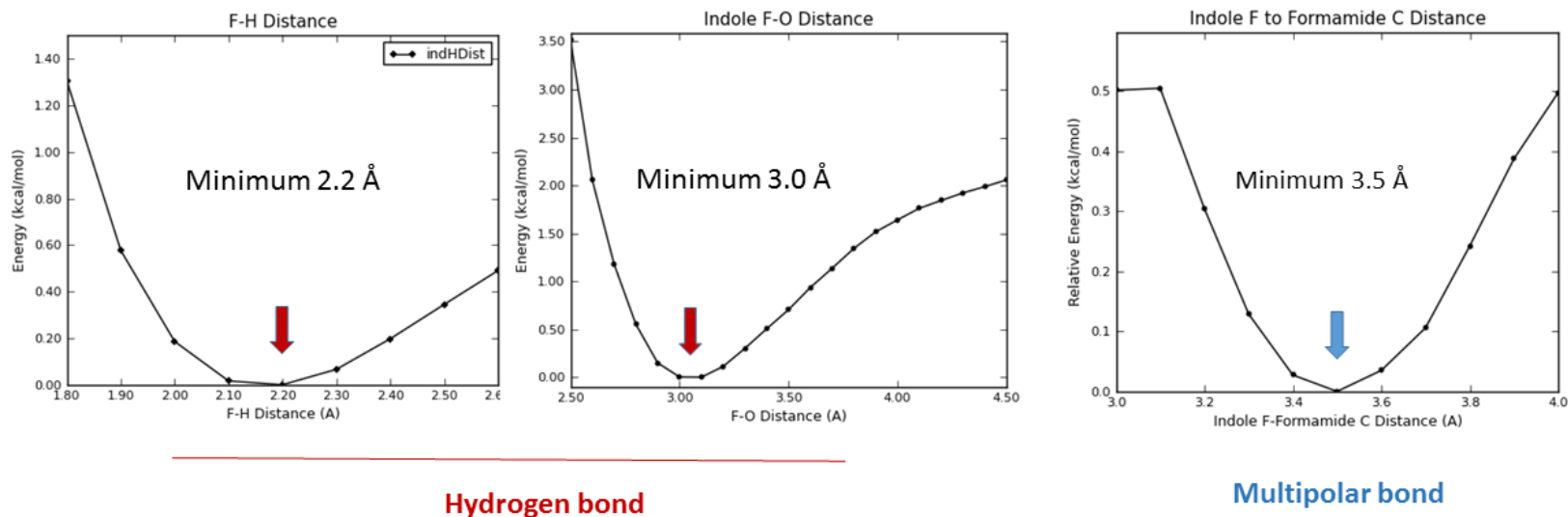
**Figure S5.** Correlation between IC<sub>50</sub> measurements and total HINT scores for inhibitors listed in Table S2.



**Figure S6.** (Left panel) PKB inhibitors bound to either PKB or a PKA-PKB chimera surrogate, which, via mutations,<sup>4</sup> is identical to PKB in the inhibitor binding site, and in overall secondary structure. A full range of side chain torsions are observed over several structures, including threonine pairs adopting the polar and apolar orientations described in the manuscript, as well as orientations in which the side chains' hydroxyl and methyl groups are oriented *trans* to one another, as depicted in Figure 5C of the manuscript. Inhibitors are shown in gray stick. The PDB entries shown include PKB-inhibitor co-crystals: 2JD0, 2JDR, 3QKK, 3QKL, 3MV5, 4EKK, 4EKL, and 4GV1, and PKA-PKB chimera-inhibitor co-crystals: 2JDT, 2JDV, 2UVX, 2UVY, 2UW3, 2UW4, 2UW8, and 4AXA. Note, for the PKA-PKB structures, PKB Thr 213 (as shown in the Figure) is numbered Thr 104 in the amino acid sequence provided in the PDB, and Thr 292 (as shown in the Figure) is numbered Thr 183 in the amino acid sequence provided in the PDB. (Right panel), PDB entries 2JDR (cyan carbons) and 2JDT (orange carbons) showing a view of the bis-Thr pair with both threonine  $\chi$  torsions rotated in opposite directions, indicating that neither Thr is locked in a single  $\chi$  rotamer in the PKB active site.

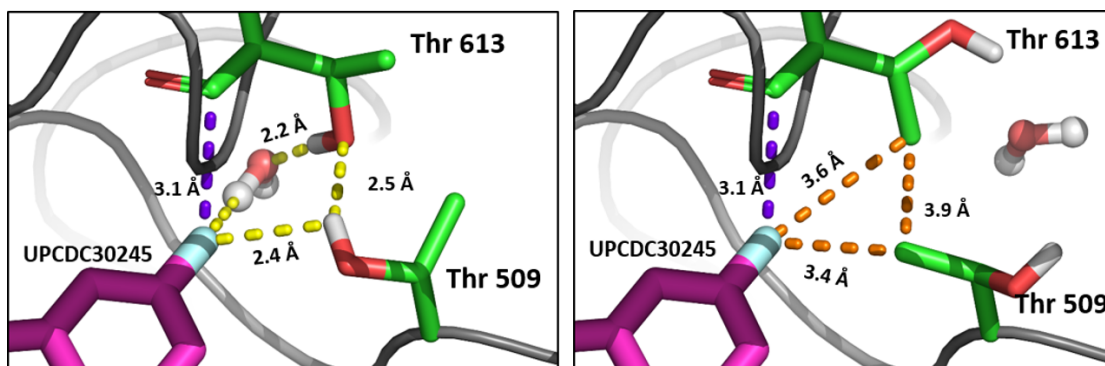


**Figure S7.** To broaden the scope of the PDB analysis for bis-Thr pairs occurring in ligand binding sites, the search criteria were relaxed to allow for the identification of bis-Thr pairs (within 7 Å of a ligand) occurring in any secondary structural feature (*i.e.*, helices, sheets, and loops). Additional X-ray co-crystal examples of ligands interacting with at least one threonine of a bis-Thr pair included: (a) heat shock protein 90 (HSP90) co-crystallized with inhibitors (example PDB entry 4YKQ shown). The HSP90 threonines both occur in  $\beta$  sheets; (b) class A  $\beta$ -lactamase CTX-M-9 co-crystallized with inhibitors (example PDB entry 4DE2 shown). The  $\beta$ -lactamase threonines occur on a  $\beta$  sheet and a loop; (c) class A  $\beta$ -lactamase BEL-1 co-crystallized with inhibitors (example PDB entry 5EUA shown). The  $\beta$ -lactamase threonines occur on a  $\beta$  sheet and a loop.; and (d) AMPA glutamate receptor 2 (GluR2) co-crystallized with agonists (example PDB entry 5FHM shown). The GluR2 threonines occur on a  $\beta$  sheet and an  $\alpha$  helix. In these examples, the bis-Thr pairs did not display the  $\chi$  rotations observed in PKB inhibitor co-crystal structures, as the interacting ligand moiety was consistently polar, and no SAR probing the potential for the Thr pairs to reorient to accommodate a hydrophobic moiety was evident from the analysis of additional PDB:ligand co-crystal structures of these examples, or from examining SAR in the literature.

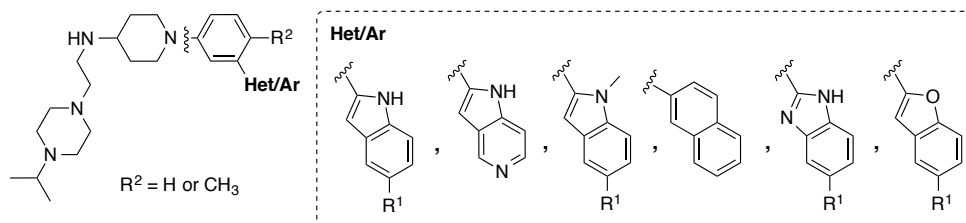


**Figure S8.** UPCDC30245 DFT determined F-H (water H) and F-O (water O) distances that comprise the hydrogen bond distance constraints used to the hydropathically refine the inhibitor-protein model by simulated annealing. Critical distances were computed from the relaxed scans and optimized geometries as stated in Supplementary Text under the section heading: UPCDC 30245 (4) Modeling and Scoring Method (below). The left and center panels (underscored with hydrogen bonds (red text)) display changes in bond energy over distance for fluorine hydrogen bonding. Similarly, the multipolar bond distance (right panel, underscored with blue text), was formed by applying constraints to enforce the F-C distance between the Thr 613 carbonyl backbone C (3.5 angstroms).





**Figure S9.** Binding models of the amphiphilic 5-F indole analog (UPCDC30245 (**4**)). Fluorine atoms are colored light blue. In the polar binding mode (left panel), the 5-F forms three enthalpic interactions, which include a water mediated hydrogen bond (yellow dash) with Thr 613, a direct hydrogen bond with Thr 509, and a multipolar bond (purple dash) using the distances obtained from DFT quantum mechanical scanning. In the hydrophobic binding mode to the bis-Thr subsite (right panel), the 5-F analog forms the same multipolar bond, and engages in favorable hydrophobic contacts (orange dash) with the bis-Thr methyl side chains.

**Table S1. Structures and Activities of p97 Inhibitors.**

Entry	Compound ID	Indole Substitution	Structure	ADPGlo IC <sub>50</sub> [μM]	ADPGlo Std. Dev. (n)
1	UPCDC30283 (1)	5-CN		0.044	0.045 (15)
2	UPCDC30287 (2) <sup>5</sup>	5-NO <sub>2</sub>		0.047	0.040 (9)
3	UPCDC30346 (3)	5-F (o-tolyl)		0.050	0.044 (5)
4	UPCDC30245 (4)	5-F		0.055	0.087 (55)
5	UPCDC30361 (5)	5-CN (o-tolyl)		0.087	0.047 (5)
6	UPCDC30310 (6)	5-CONH <sub>2</sub>		0.10	0.032 (4)
7	UPCDC30256 (7)	5-OH		0.12	0.073 (4)
8	UPCDC30341 (8)	5-CO <sub>2</sub> Me		0.13	0.076 (4)
9	UPCDC30083 (9)	5-H		0.16	0.10 (17)
10	UPCDC30317 (10)	5-Cl		0.19	0.14 (7)
11	UPCDC30288 (11)	5-N <sub>3</sub>		0.23	0.18 (9)
12	UPCDC30318 (12)	5-CH <sub>3</sub>		0.24	0.11 (7)

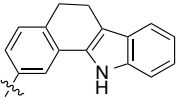
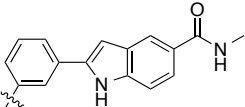
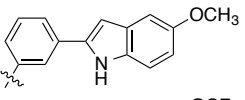
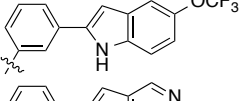
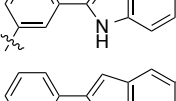
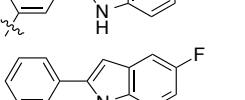
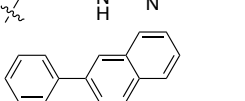
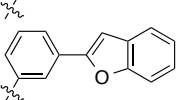
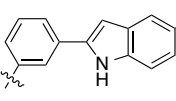
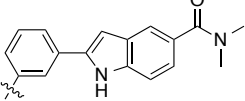
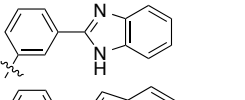
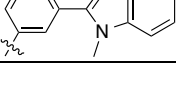

13	UPCDC30206 (13)	benzo[ $\alpha$ ]carbazole		0.39	0.069 (3)
14	UPCDC30367 (14)	5-CONHMe		0.45	0.12 (5)
15	UPCDC30238 (15)	5-OCH <sub>3</sub>		0.71	0.22 (4)
16	UPCDC30257 (16)	5-OCF <sub>3</sub>		3.8	0.8 (4)
17	UPCDC30345 (17)	5-azaindole		4.0	1.7 (5)
18	UPCDC30297 (18)	5-CF <sub>3</sub>		4.67	2.0 (4)
19	UPCDC30381 (19)	5-F, 7-azaindole		4.74	1.3 (5)
20	UPCDC30222 (20)	naphthalene		5.2	1.1 (4)
21	UPCDC30221 (21)	benzofuran		20.1	4.60 (4)
22	UPCDC30277 (22)	5-SF <sub>5</sub>		21.5	0.40 (4)
23	UPCDC30368 (23)	5-CONMe <sub>2</sub>		31.4	1.60 (3)
24	UPCDC30250 (24)	benzimidazole		38.5	6.60 (4)
25	UPCDC30201 (25)	N-Me indole		>50	-(1)

Table S2. Quantitated HINT Scores for Inhibitors with IC<sub>50</sub> < 5 μM.

Compound ID	Indole Substitution	ADPGlo IC <sub>50</sub> [μM]	ADPGlo Std. Dev. [μM]	Total HINT Score	Hydrogen Bond *	Acid/Base *	Hydrophobic *	Acid/Acid **	Base/Base **	Hydrophobic/Polar **
UPCDC30283 (1)	5-CN	0.044	0.045	<b>2.42E+03</b>	1.67E+03	7.33E+02	1.31E+03	-1.90E+02	-4.18E+02	-6.84E+02
UPCDC30287 (2)	5-NO <sub>2</sub>	0.047	0.04	<b>2.04E+03</b>	1.22E+03	8.15E+02	1.41E+03	-1.14E+02	-3.41E+02	-9.55E+02
UPCDC30245 (4)	5-F	0.055	0.087	<b>1.97E+03</b>	7.78E+02	8.08E+02	1.34E+03	-1.58E+02	-1.58E+02	-6.39E+02
UPCDC30310 (6)	5-CONH <sub>2</sub>	0.1	0.032	<b>1.83E+03</b>	1.69E+03	1.19E+03	1.39E+03	-6.37E+02	-5.78E+02	-1.23E+03
UPCDC30256 (7)	5-OH	0.12	0.073	<b>1.63E+03</b>	6.16E+02	8.29E+02	1.25E+03	-1.76E+02	-5.47E+01	-8.29E+02
UPCDC30341 (8)	5-CO <sub>2</sub> Me	0.13	0.076	<b>1.47E+03</b>	8.18E+02	9.16E+02	2.05E+03	-3.69E+02	-2.34E+02	-1.71E+03
UPCDC30083 (9)	5-H	0.16	0.1	<b>1.45E+03</b>	3.61E+02	7.36E+02	1.35E+03	-1.68E+02	-3.24E+01	-7.97E+02
UPCDC30317 (10)	5-Cl	0.19	0.14	<b>1.43E+03</b>	1.31E+02	8.92E+02	1.42E+03	-1.86E+02	-2.26E+01	-8.07E+02
UPCDC30288 (11)	5-N <sub>3</sub>	0.23	0.18	<b>1.38E+03</b>	1.72E+03	1.17E+03	2.14E+03	-1.60E+02	-2.33E+03	-1.15E+03
UPCDC30318 (12)	5-CH <sub>3</sub>	0.24	0.11	<b>1.22E+03</b>	9.05E+01	8.56E+02	1.51E+03	-1.86E+02	-1.78E+01	-1.03E+03
UPCDC30367 (14)	5-CONHMe	0.45	0.12	<b>1.16E+03</b>	8.97E+02	8.55E+02	1.93E+03	-1.89E+02	-2.59E+02	-2.08E+03
UPCDC30238 (15)	5-OCH <sub>3</sub>	0.71	0.22	<b>1.14E+03</b>	3.73E+02	6.21E+02	1.45E+03	-1.23E+02	-5.41E+01	-1.13E+03
UPCDC30257 (16)	5-OCF <sub>3</sub>	3.8	0.8	<b>1.01E+03</b>	5.47E+02	8.20E+02	1.38E+03	-1.63E+02	-8.52E+02	-7.24E+02
UPCDC30345 (17)	5-azaindole	4	1.7	<b>9.70E+02</b>	3.47E+02	8.27E+02	1.34E+03	-2.76E+02	-3.13E+01	-1.24E+03
UPCDC30297 (18)	5-CF <sub>3</sub>	4.67	2	<b>9.39E+02</b>	6.14E+02	4.53E+02	9.08E+02	-9.25E+01	-4.56E+02	-4.88E+02
UPCDC30381 (19)	5-F, 7-azaindole	4.74	1.3	<b>8.91E+02</b>	7.88E+02	7.74E+02	1.20E+03	-7.69E+02	-3.69E+01	-1.07E+03

\* Favorable interactions

\*\* Unfavorable interactions

## Description of the HINT Hydrophobic Scoring Function

The program HINT (for “Hydrophobic INTERactions”) scores a summation of the hydrophobic interactions between all atom pairs in a molecule at a given molecular geometry,<sup>5-7</sup> and was used to score the binding poses of the inhibitors from this study (Table S2). Hydrophobic profiles consist of a summation of the hydrophobic interactions between all atom pairs in a molecule at a given molecular geometry. Hydrophobic interactions are divided into 6 main classes. The “favorable” classes consist of hydrogen bonds, acid/base, and hydrophobic interactions, while acid/acid, base/base, and hydrophobic/polar interactions are the “unfavorable” categories. Each atom potential type has a corresponding hydrophobic constant  $a_i$  derived from the hydrophobic fragment constant approach of Hansch and Leo.<sup>6,8</sup> A summation of all of these constants for a given molecule results in the total molecular partition constant for that molecule. This methodology reduces the empirical information from bulk molecular solvent partitioning to discrete atom-atom interactions. Because of this approach, the hydrophobic constants derived include a linear free estimate of entropy, which is ignored in molecular mechanics models. Thus, a properly constructed set of comparisons using these hydrophobic parameters as its basis will yield a much richer set of information than a method based solely on parameters derived from mechanics.

The hydrophobic interaction value for an atom pair  $b_{ij}$  is a function of the hydrophobic constants for each atom, the distance between each of the atoms, and the solvent accessible surface area (SASA) of the two atoms.

$$b_{ij} = s_i a_i s_j a_j R_{ij}$$

The distance function in the form  $R_{ij} = e^{-r}$  has been reported to give a good fit to the published Leo polar proximity factors.<sup>1,8,9</sup> The SASA for atoms,  $s_i$ , is taken from literature values<sup>10</sup> for proteins. The interaction values  $b_{ij}$  for each atom pair are summed in one of the six aforementioned categories. Atoms pairs within an amino acid residue, atoms which were bonded to each other, and atoms which were

involved in 1-3 interactions were not included in the summation. A van der Waals component is also computed for each atom pair using published Lennard-Jones parameters.<sup>11,12</sup> This penalizes atom-atom interactions which are too close in a manner which is independent from the hydrophobic interaction.

## UPCDC 30245 (4) Modeling and Scoring Method

*Model of the 5-F-indole congener (UPCDC30245).* The amphiphilic 5-F substituent of UPCDC 30245 (4) can bind in both the polar and apolar conformations of the bis-Thr subsite. The modeling objective for this inhibitor was, therefore, to form geometrically optimal binding models in both conformations of the subsite, and to create a new aromatic fluorine hydrophobic constant for HINT quantitation – to provide final models that rank ordered in the SAR series.

Complicating the modeling of this inhibitor: 1) in the bis-Thr polar conformation, in addition to a hydrogen bond with Thr 509, the indole 5-F substituent can hydrogen bond with a coordinated water and engage in a multipolar bond with the backbone carbonyl carbon of Thr 613 and 2) in the bis-Thr apolar conformation, the indole 5-F substituent can bind in the subsite within acceptable distances to the bis-Thr side chain methyls (*i.e.*, 3.4 - 3.6 Å) and still engage in a multipolar bond with the backbone carbonyl carbon of Thr 613. However, the HINT program did not provide a hydrophobic constant for scoring these interactions. Therefore, to provide: 1) geometrically optimal binding modes for the 5-F analog in both conformations of the bis-Thr subsite and 2) a new HINT constant for the amphiphilic aromatic fluorine, definitive geometries for the intermolecular interactions of this substituent were calculated.

To obtain the geometries, we utilized: 1) DFT calculation models to place the coordinates of the 5-F-indole, 2) observed aromatic fluorine co-crystal data for non-bonded interactions, and 3) findings consistent with F-NMR data from the literature.<sup>13-17</sup>

Quantum mechanical DFT calculations were used to find the optimal distances between the F-water hydrogen bond (F-H and F-O distances) and the F-carbonyl carbon (multipolar bond) distance using formamide as a model for a backbone carbonyl carbon. For the relaxed coordinate scan, calculations were performed using the Jaguar2 quantum chemistry application in the Maestro software suite.<sup>18</sup> The distances of interest were held fixed while the rest of the

molecular coordinates were optimized using Hartree-Fock theory with the 6-31G\*\* basis set. Each point was optimized to an energy difference of .0005 Ha. Fluorine-hydrogen distances were held fixed at increments of 0.1 Å ranging from 1.8 to 2.6 Å, while fluorine-oxygen and fluorine-carbon distances were held fixed at increments of 0.1 Å ranging from 3.0 to 5.0 Å. The complexes were then optimized to a “fine” level using DFT with the PW911 functional.<sup>19,20</sup> The results are shown in Figure S8.

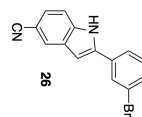
Figure S9 displays binding modes of the 5-F-indole derivative in both binding modes. The geometries of the final binding models were verified by searching the PDB for non-bonded binding interactions between protein side chains and aromatic fluorine geometries (*i.e.*, aromatic fluorines in ligands).<sup>13,16</sup> It was found that both binding modes were consistent with high resolution data. Moreover, probably the most studied non-bonded interaction work involving fluorinated ligands has been conducted by Vulpetti and colleagues.<sup>13-17</sup> Recently, Dalvit and Vulpetti<sup>21</sup> reported specific data with respect to 5-F-indole hydrogen bonding to water. It is clear from their investigations that the fluorine atom, when substituted on an aromatic group, presents a weak hydrogen bond acceptor that is nonetheless stable. They have further suggested that fluorine interactions in otherwise nonpolar binding subsites may in fact be favorable, and our results are in agreement with their “rule of shielding.”<sup>15</sup> We find this consistent with our DFT scanning studies, and numerous distances observed in the PDB. Given our geometries were consistent with the literature and X-ray data, we estimated a new hydrophobic HINT constant for an amphiphilic aromatic fluorine (details to be published elsewhere). Application of this constant in the calculation of the 5-F indole inhibitor binding modes resulted in a net increase of 288 favorable HINT interaction points. The total HINT score for the binding of UPCDC 30245 (**4**) was consistent with the SAR (Table S2, Figure S5).



## References

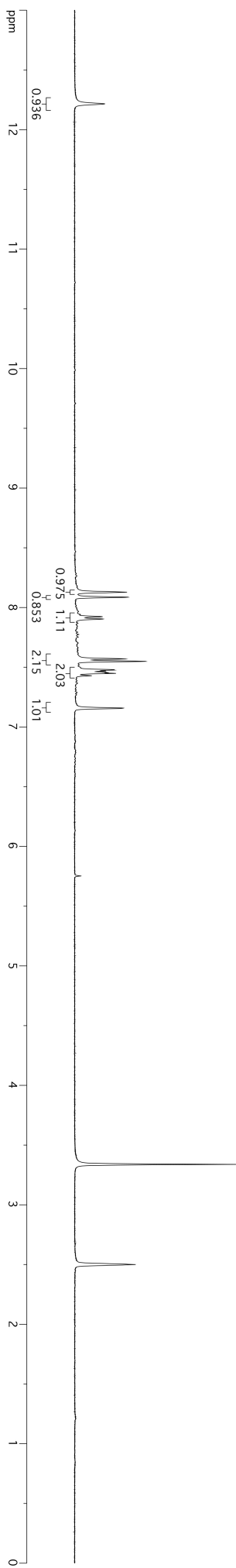
1. Banerjee, S. *et al.* 2.3 A resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. *Science* **351**, 871-875, (2016).
2. Schrodinger Maestro Python API, 2015-3; Schrodinger, LLC: 2015.
3. European Bioinformatics Institute FTP site  
[ftp://ftp.ebi.ac.uk/pub/databases/msd/sifts/flatfiles/csv/pdb\\_chain\\_uniprot.csv.gz](ftp://ftp.ebi.ac.uk/pub/databases/msd/sifts/flatfiles/csv/pdb_chain_uniprot.csv.gz) (accessed Sep 19, 2016).
4. Donald, A. *et al.* Rapid evolution of 6-phenylpurine inhibitors of protein kinase B through structure-based design. *J. Med. Chem.* **50**, 2289-2292, (2007).
5. Kellogg, G. E. & Abraham, D. J. KEY, LOCK, and LOCKSMITH: complementary hydrophobic map predictions of drug structure from a known receptor-receptor structure from known drugs. *J. Mol. Graph.* **10**, 212-217, 226 (1992).
6. Kellogg, G. E., Semus, S. F. & Abraham, D. J. HINT: a new method of empirical hydrophobic field calculation for CoMFA. *J. Comput. Aided Mol. Des.* **5**, 545-552 (1991).
7. Kellogg, G. E., HINT, 2.01; eduSoft: Ashland, VA, 1993.
8. Hansch, C. & Leo, A. Substituent Constants for Correlation Analysis in Chemistry. *John Wiley & Sons: New York* (1979).
9. Fauchere, J. L., Charton, M., Kier, L. B., Verloop, A. & Pliska, V. Amino acid side chain parameters for correlation studies in biology and pharmacology. *Int. J. Pept. Protein Res.* **32**, 269-278 (1988).
10. Shrake, A. & Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **79**, 351-371 (1973).
11. Levitt, M. Molecular dynamics of native protein. II. Analysis and nature of motion. *J. Mol. Biol.* **168**, 621-657 (1983).
12. Levitt, M. & Perutz, M. F. Aromatic rings act as hydrogen bond acceptors. *J. Mol. Biol.* **201**, 751-754 (1988).
13. Vulpetti, A. & Dalvit, C. Fluorine local environment: from screening to drug design. *Drug Discov. Today* **17**, 890-897, (2012).
14. Vulpetti, A., Hommel, U., Landrum, G., Lewis, R. & Dalvit, C. Design and NMR-based screening of LEF, a library of chemical fragments with different local environment of fluorine. *J. Am. Chem. Soc.* **131**, 12949-12959, (2009).
15. Dalvit, C. & Vulpetti, A. Fluorine-protein interactions and (1)(9)F NMR isotropic chemical shifts: An empirical correlation with implications for drug design. *ChemMedChem* **6**, 104-114, (2011).
16. Dalvit, C. & Vulpetti, A. Intermolecular and intramolecular hydrogen bonds involving fluorine atoms: implications for recognition, selectivity, and chemical properties. *ChemMedChem* **7**, 262-272, (2012).
17. Dalvit, C., Invernizzi, C. & Vulpetti, A. Fluorine as a hydrogen-bond acceptor: experimental evidence and computational calculations. *Chemistry* **20**, 11058-11068, (2014).
18. Maestro Version 19.4.017 Release 2016-1.

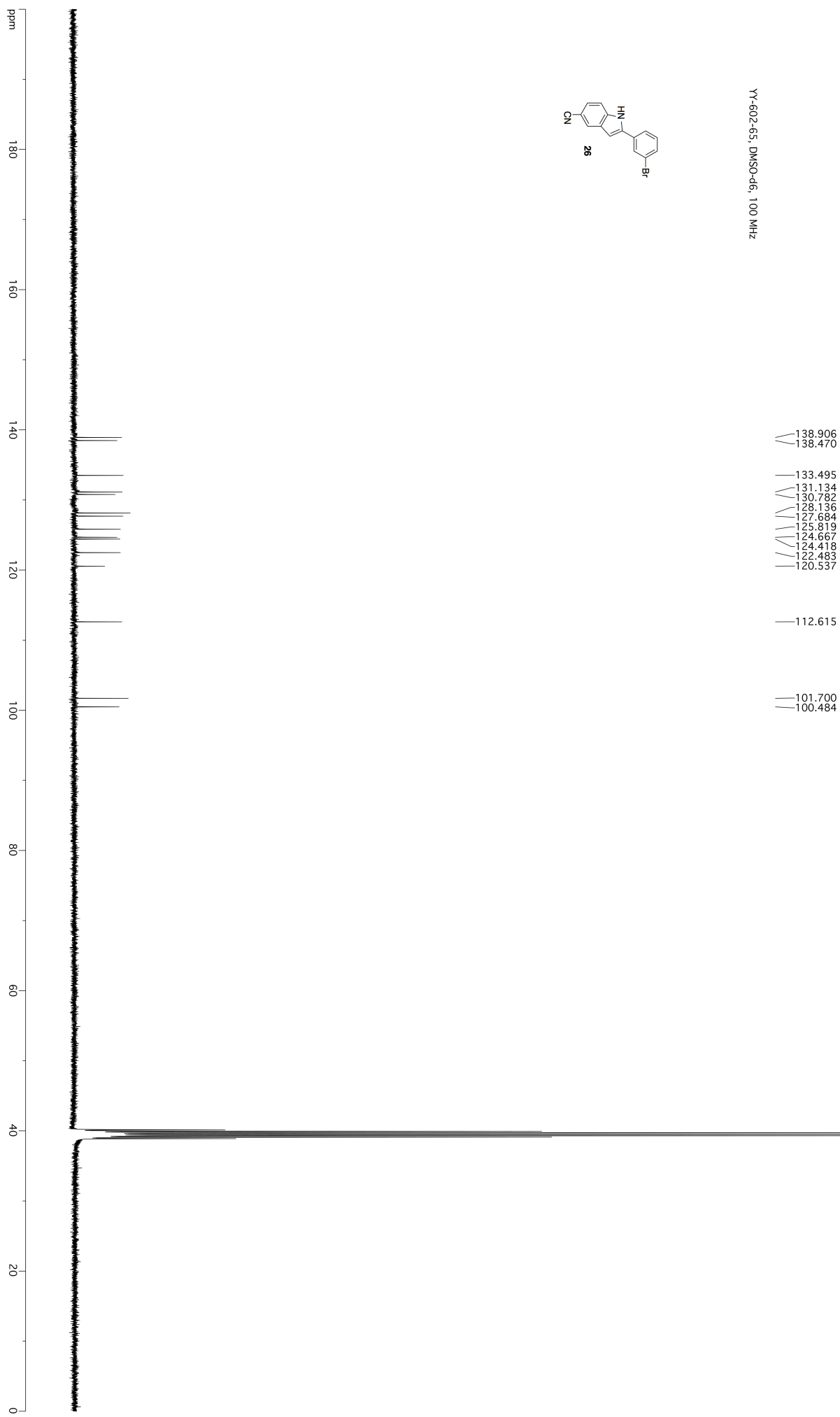
19. Perdew, J. P. & Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B Condens. Matter* **45**, 13244-13249 (1992).
20. Bochevarov, A. D. *et al.* Jaguar: a high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem.* **113**, 2110–2142 (2013).
21. Dalvit, C. & Vulpetti, A. Weak Intermolecular Hydrogen Bonds with Fluorine: Detection and Implications for Enzymatic/Chemical Reactions, Chemical Properties, and Ligand/Protein Fluorine NMR Screening. *Chemistry* **22**, 7592-7601, (2016).

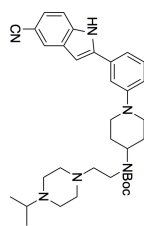
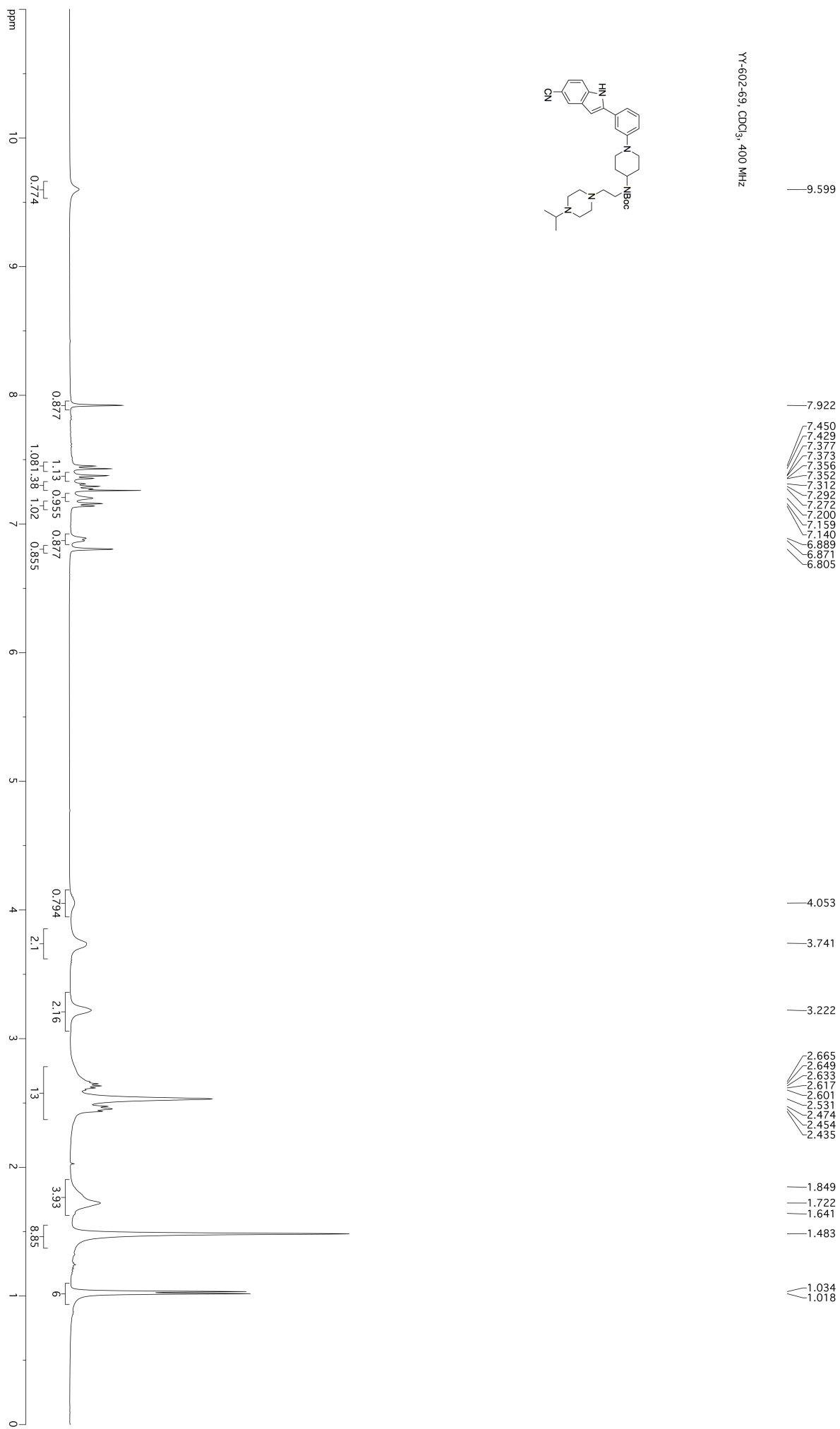


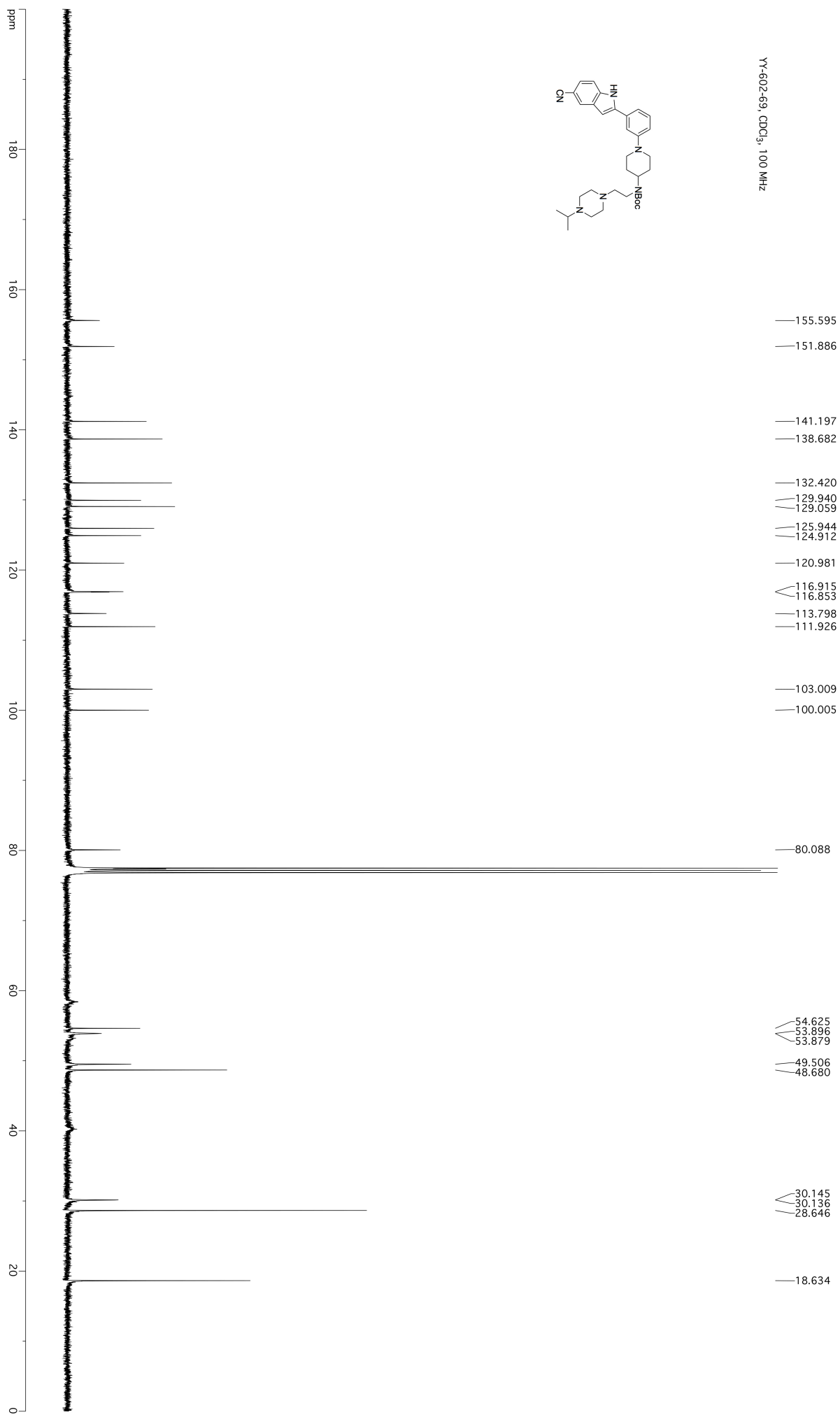
—12.216  
YY-602-65, DMSO-d6, 400 MHz

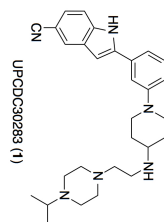
8.128  
8.086  
7.923  
7.904  
7.570  
7.549  
7.480  
7.476  
7.468  
7.459  
7.455  
7.448  
7.429  
7.158





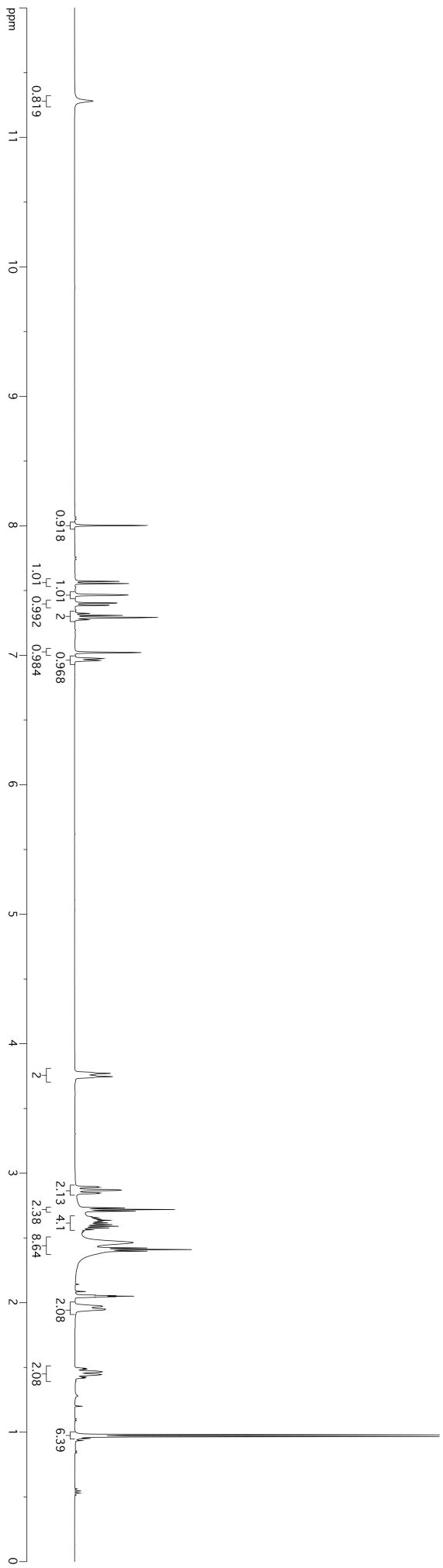
YY-602-69, CDCl<sub>3</sub>, 400 MHz

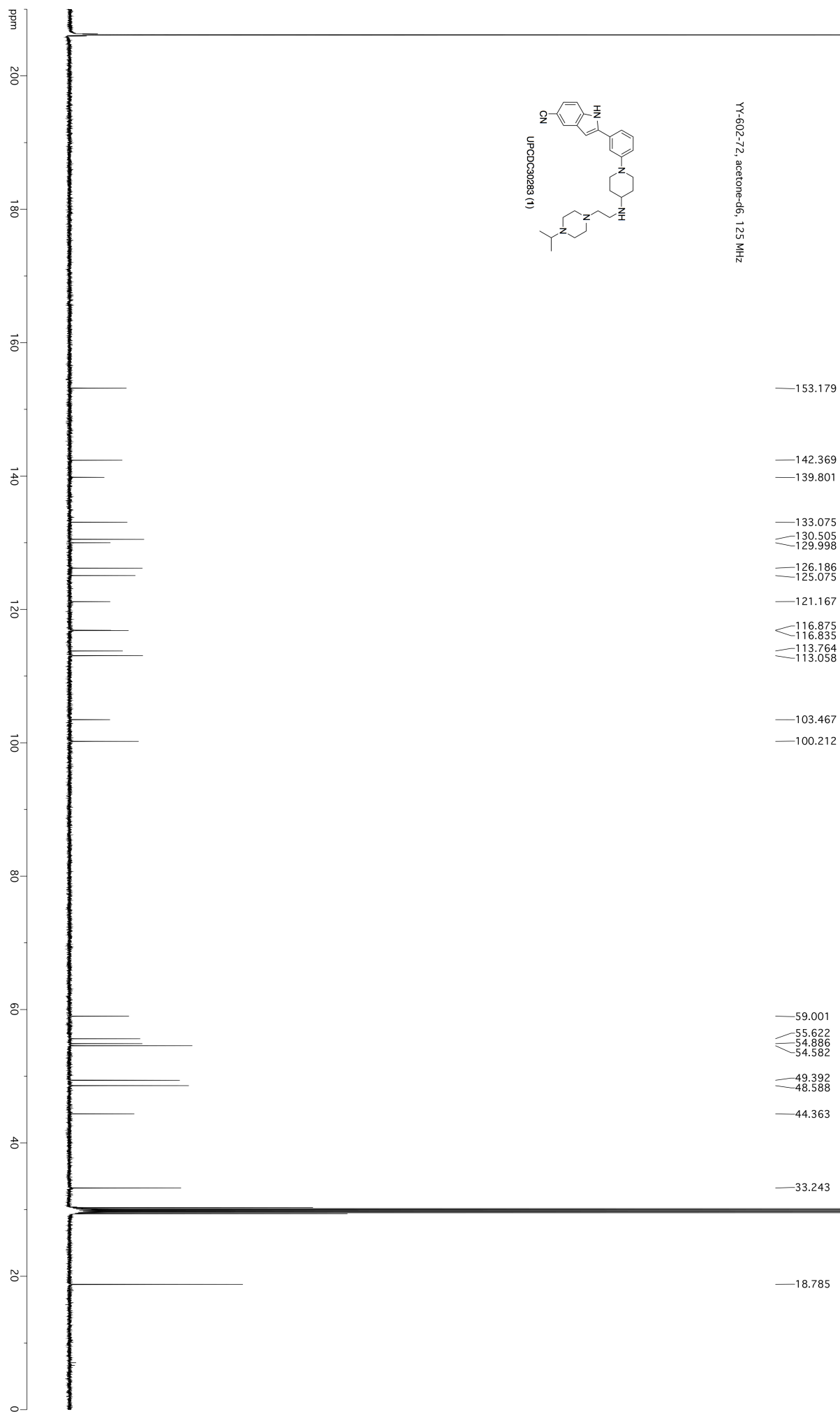




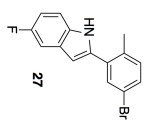
YY-602-72, acetone-d6, 500 MHz

- 11.281
- 8.002
- 7.571
- 7.554
- 7.466
- 7.405
- 7.402
- 7.388
- 7.386
- 7.323
- 7.308
- 7.293
- 7.275
- 7.021
- 6.975
- 6.960
- 3.783
- 3.771
- 3.746
- 3.736
- 2.895
- 2.891
- 2.869
- 2.848
- 2.844
- 2.732
- 2.719
- 2.707
- 2.664
- 2.656
- 2.644
- 2.636
- 2.629
- 2.616
- 2.603
- 2.590
- 2.577
- 2.564
- 2.551
- 2.464
- 2.423
- 2.410
- 2.398
- 1.975
- 1.970
- 1.950
- 1.945
- 1.493
- 1.486
- 1.466
- 1.445
- 1.425
- 1.418
- 0.980
- 0.966







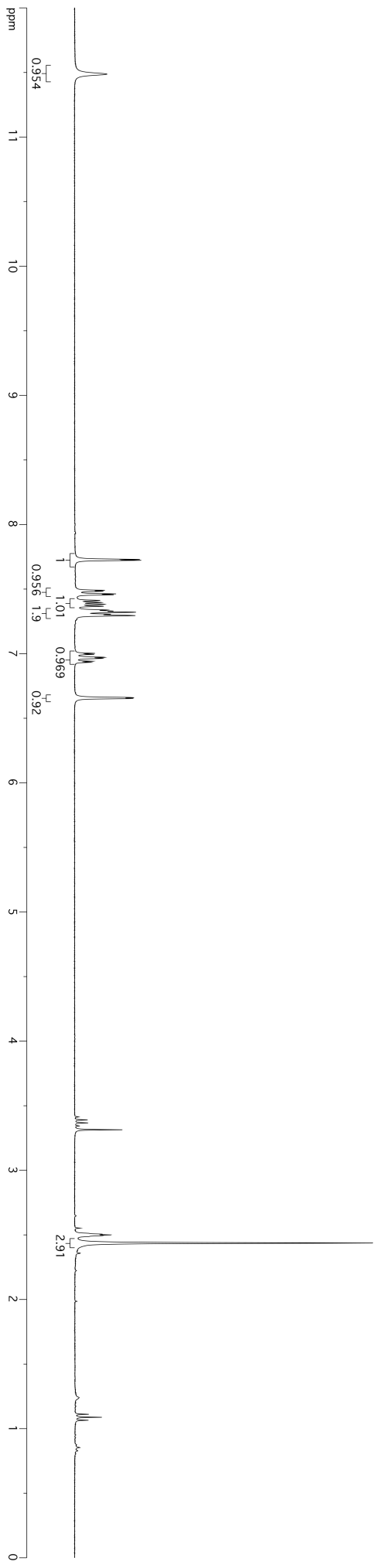


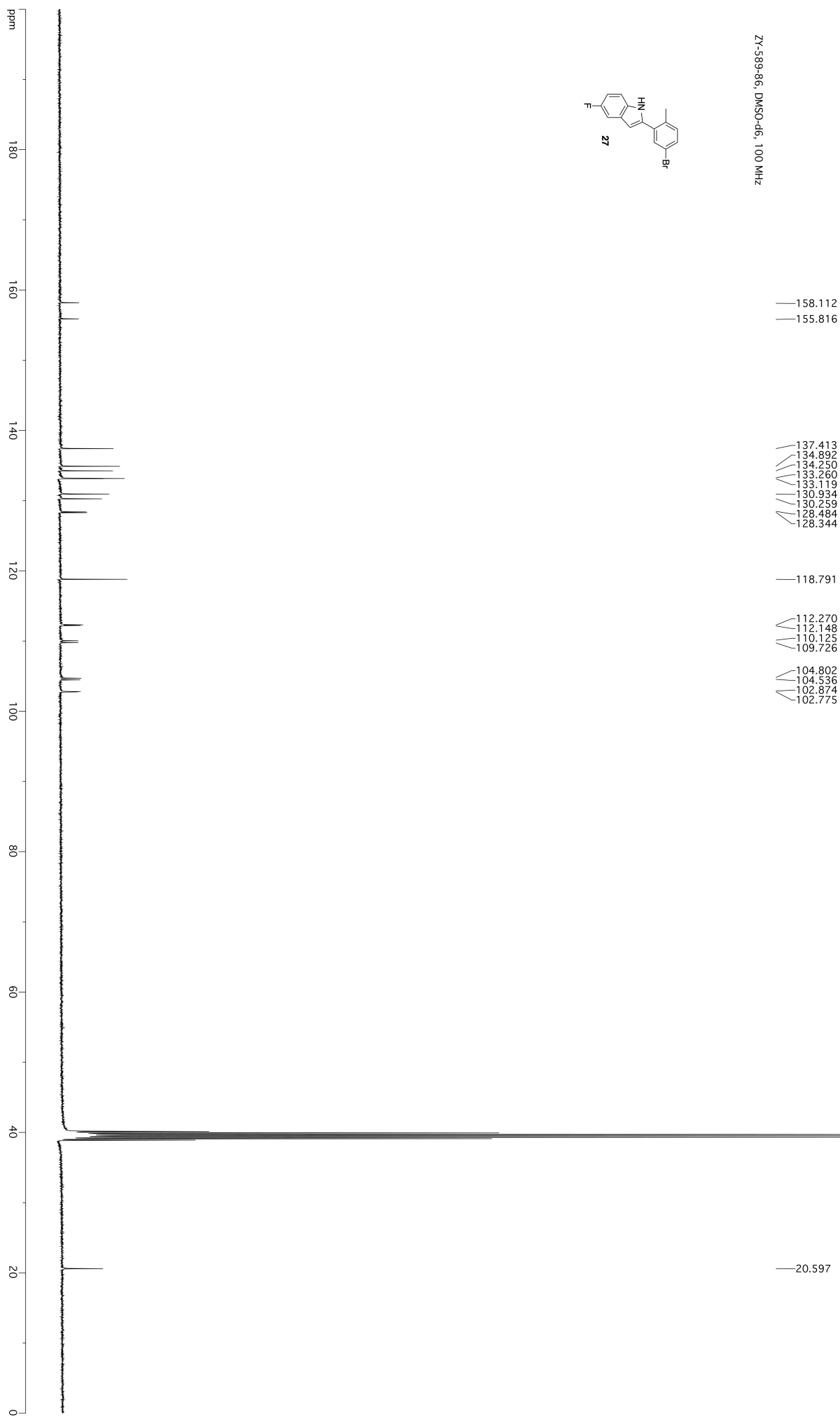
ZY-589-86, DMSO-d6, 300 MHz

— 11.487

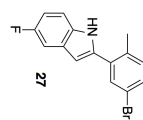
- 7.730
- 7.723
- 7.490
- 7.483
- 7.463
- 7.456
- 7.412
- 7.397
- 7.383
- 7.367
- 7.338
- 7.330
- 7.321
- 7.305
- 7.295
- 7.003
- 6.994
- 6.971
- 6.964
- 6.941
- 6.933
- 6.659
- 6.654

— 2.437





ZY-583-86, DMSO-d6



ppm

-20

-40

-60

-80

-100

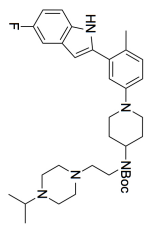
-120

-140

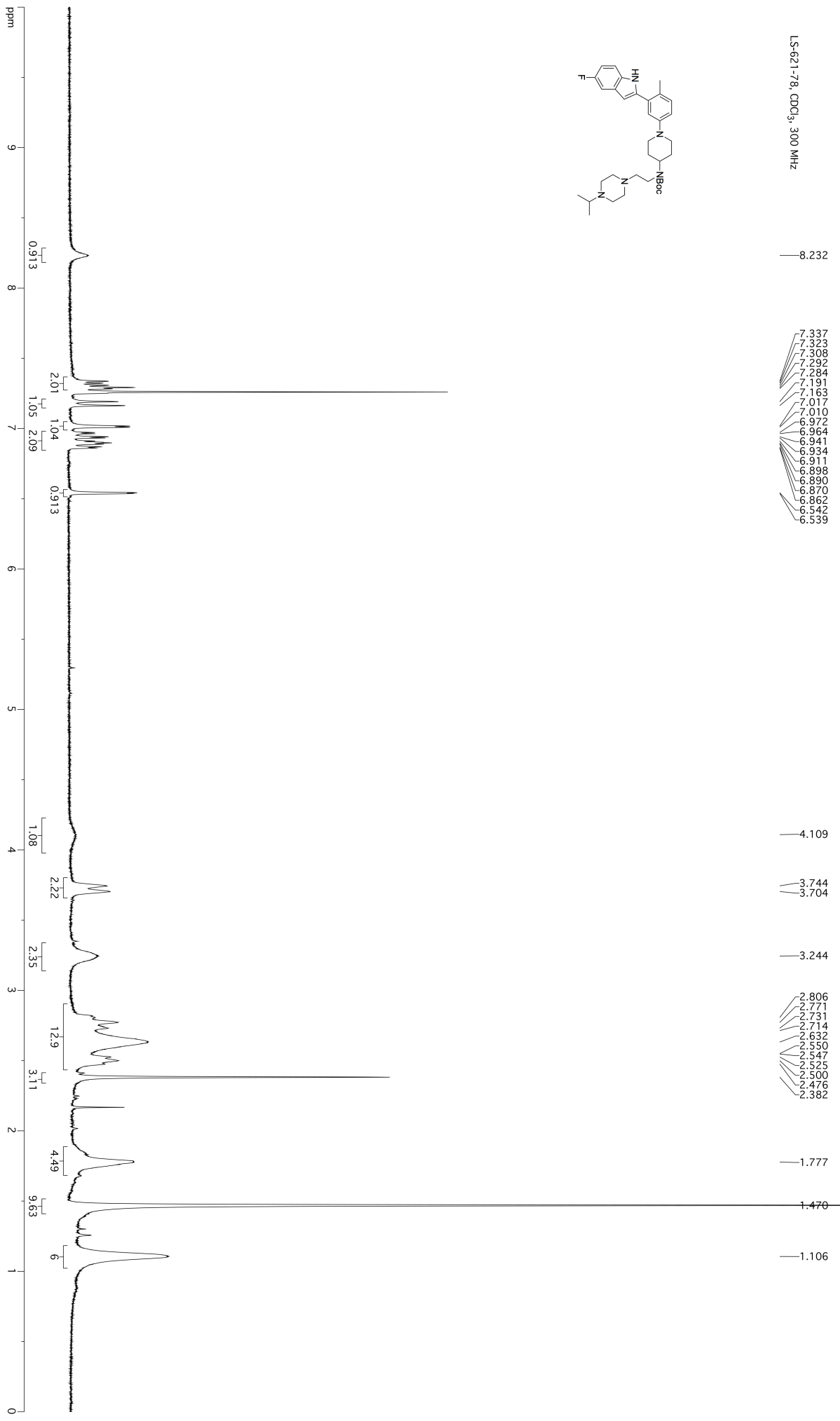
-160

-180

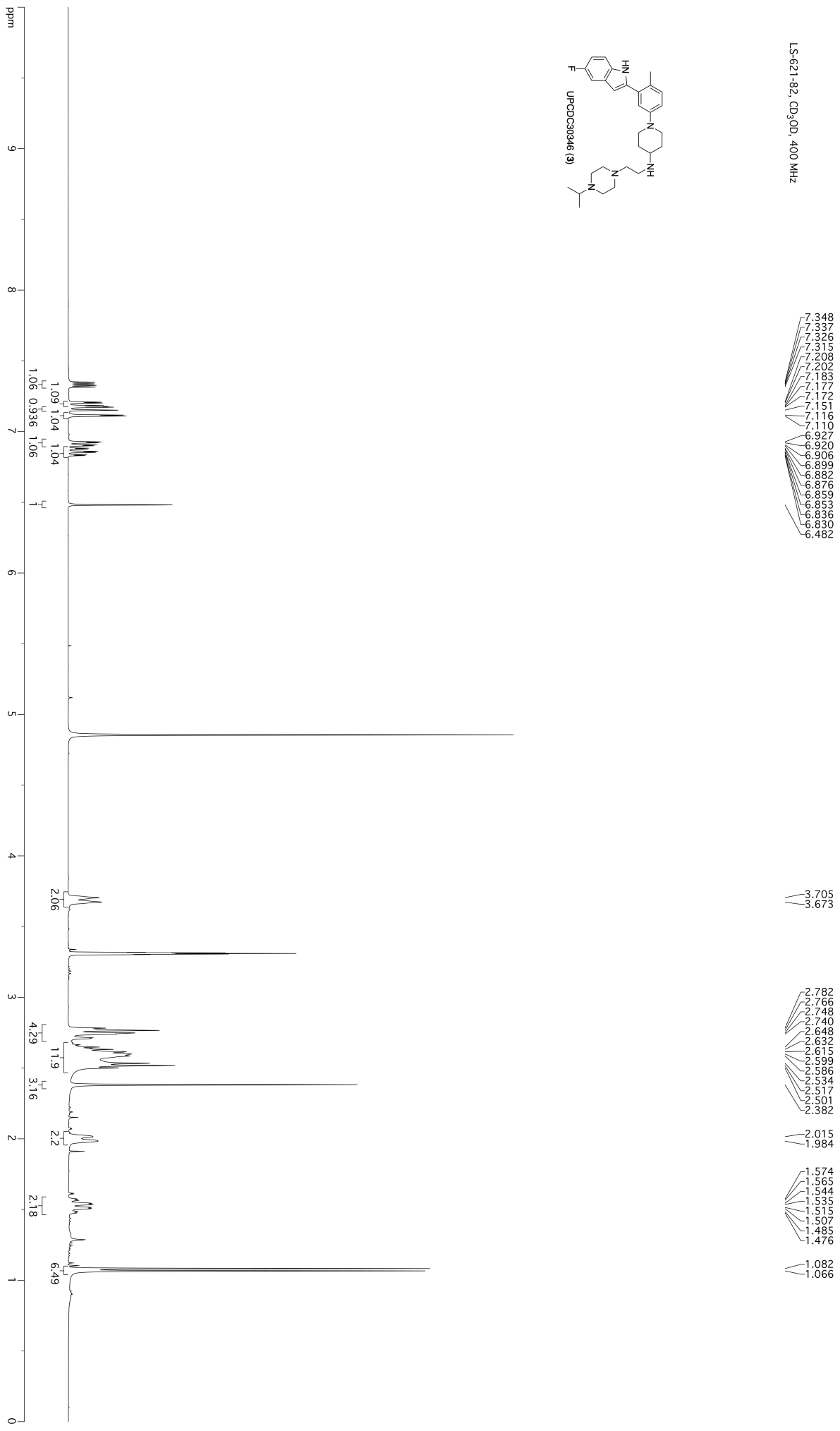
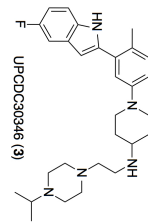
—124.605

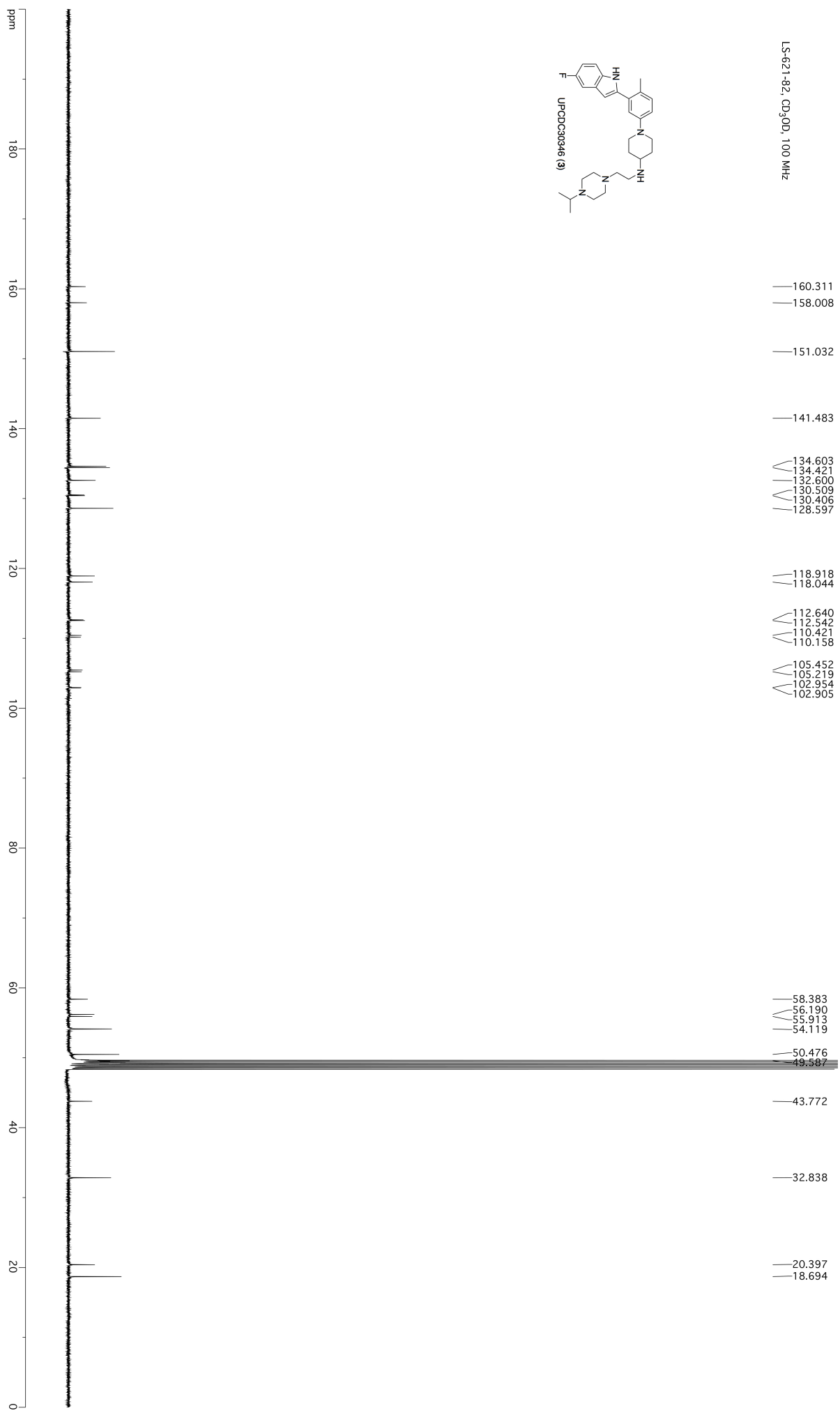


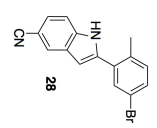
LS-621-78, CDCl<sub>3</sub>, 300 MHz



LS-621-82, CD<sub>3</sub>OD, 400 MHz

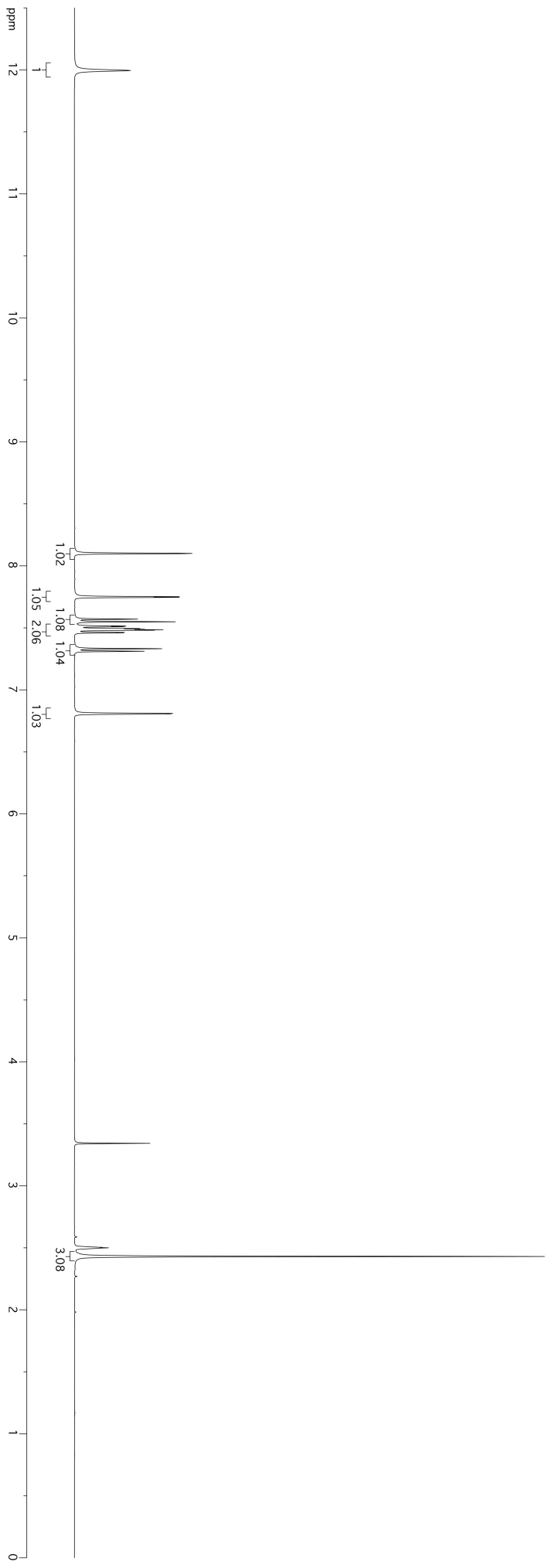




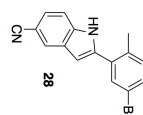
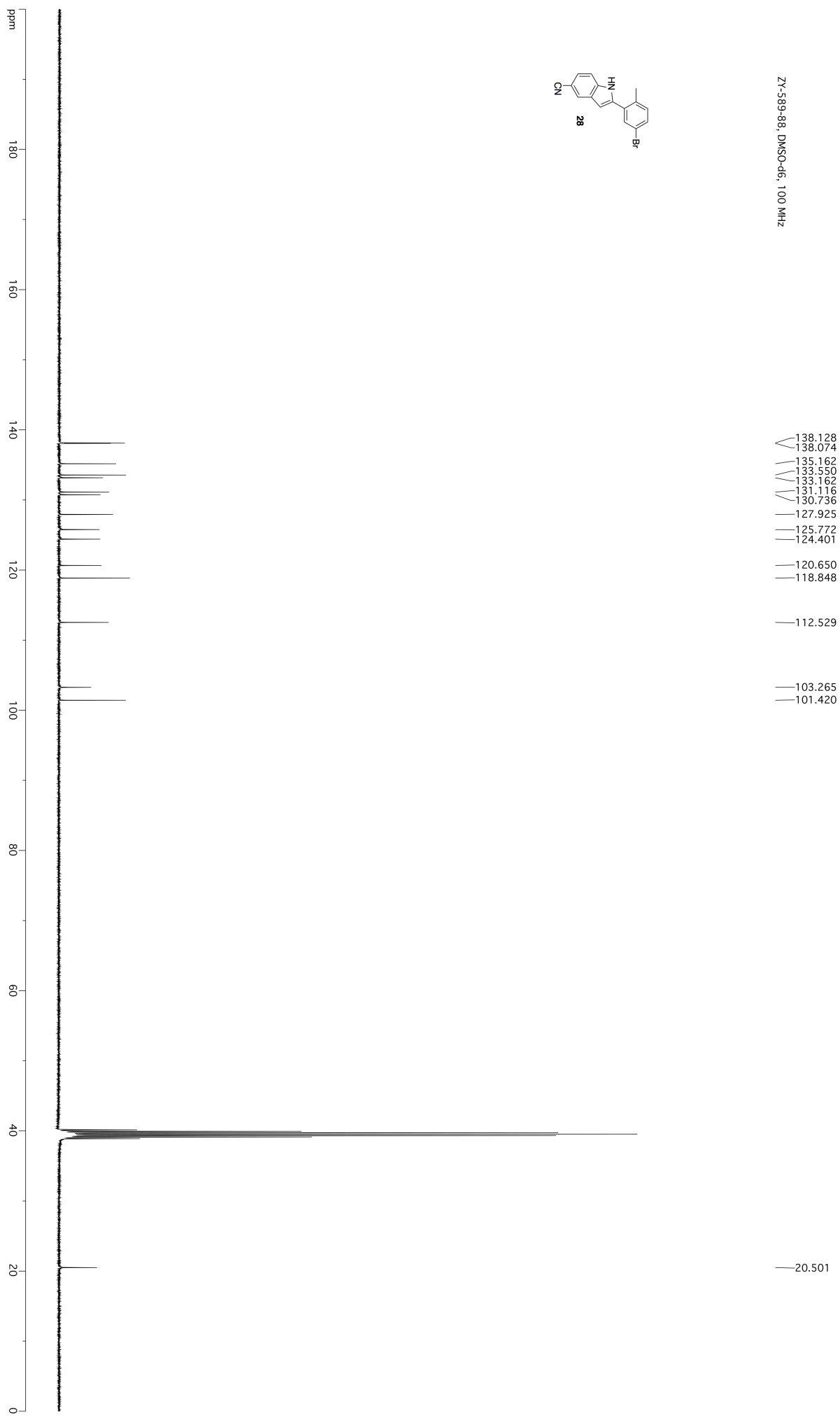


11.995  
ZY-589-88, DMSO-d6, 400 MHz

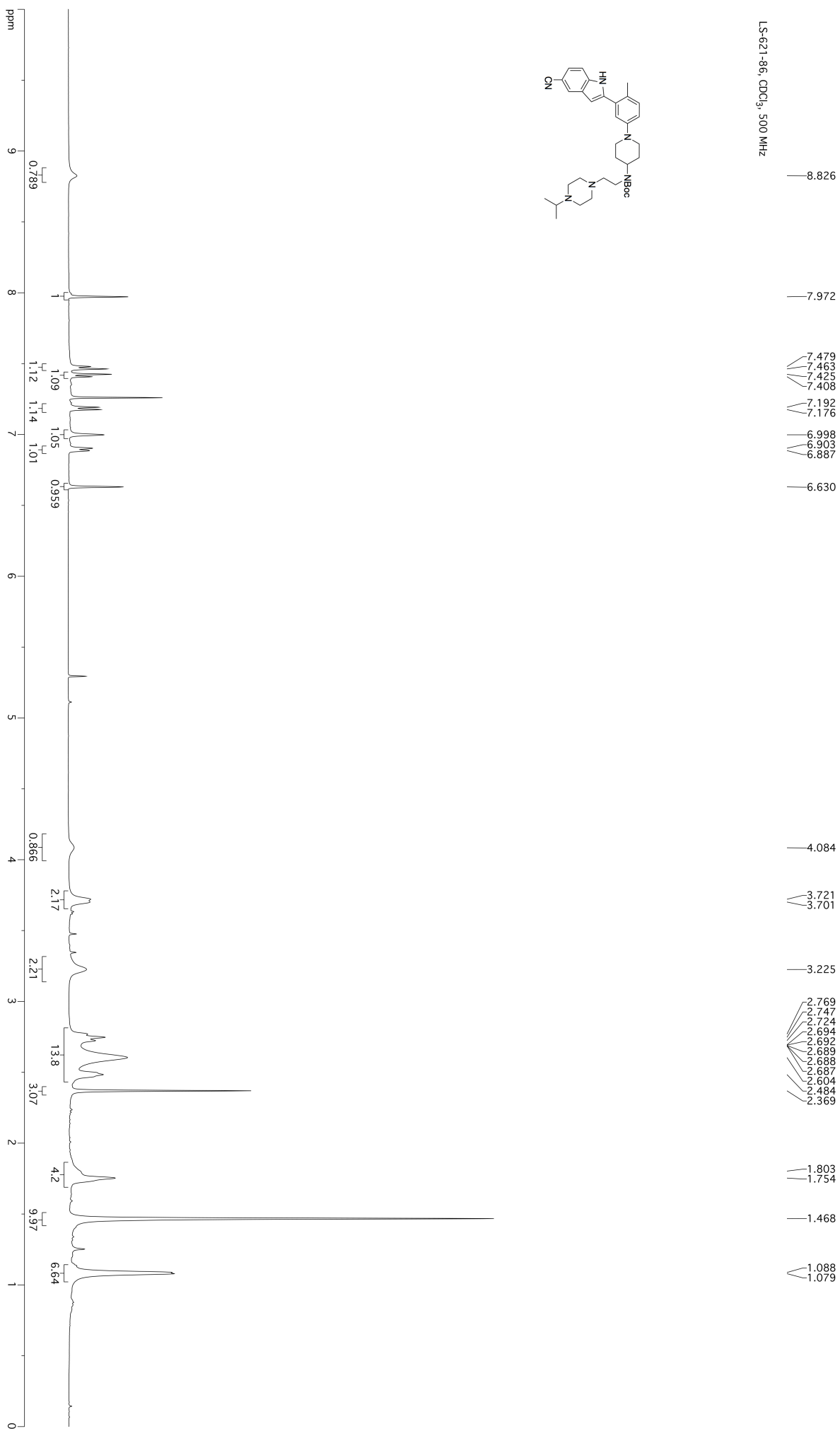
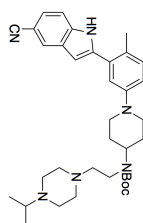
- 8.100
- 7.751
- 7.746
- 7.570
- 7.549
- 7.515
- 7.510
- 7.494
- 7.489
- 7.484
- 7.481
- 7.463
- 7.459
- 7.332
- 7.311
- 6.809
- 6.806

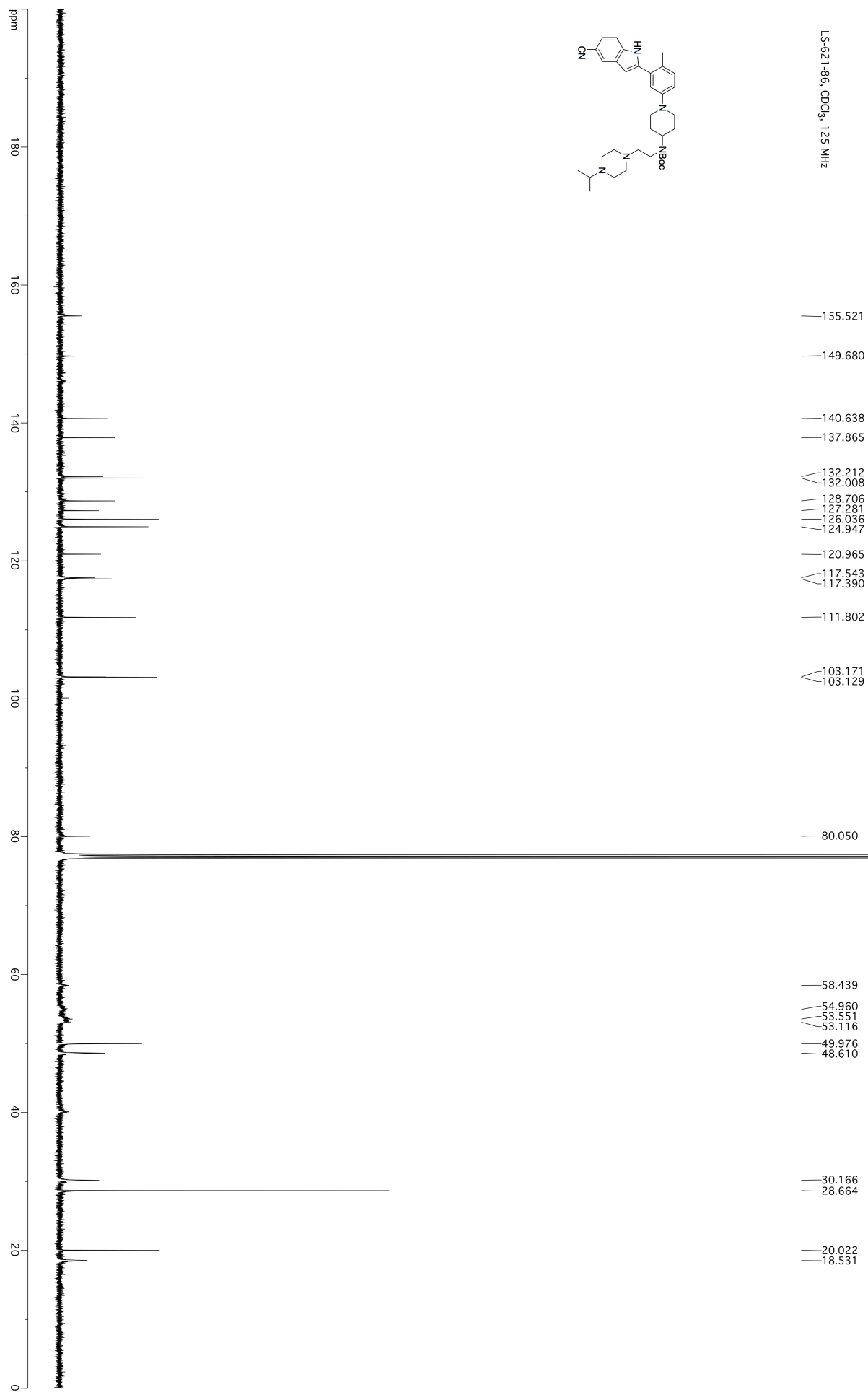


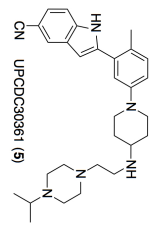
2.430

ZV-589-88, DMSO-d<sub>6</sub>, 100 MHz

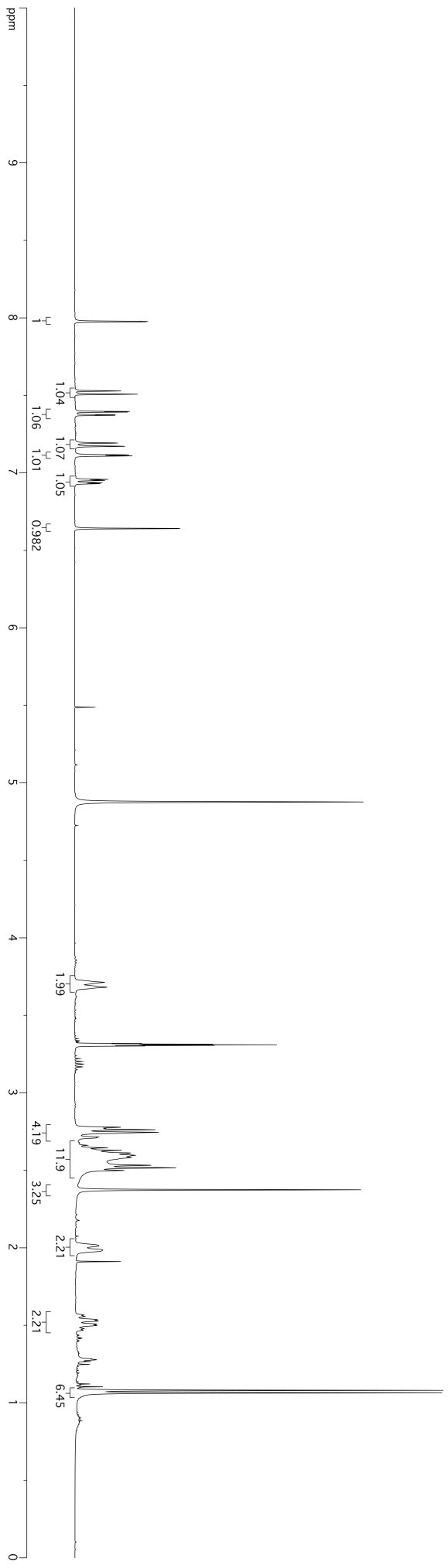


LS-621-86, CDCl<sub>3</sub>, 500 MHz

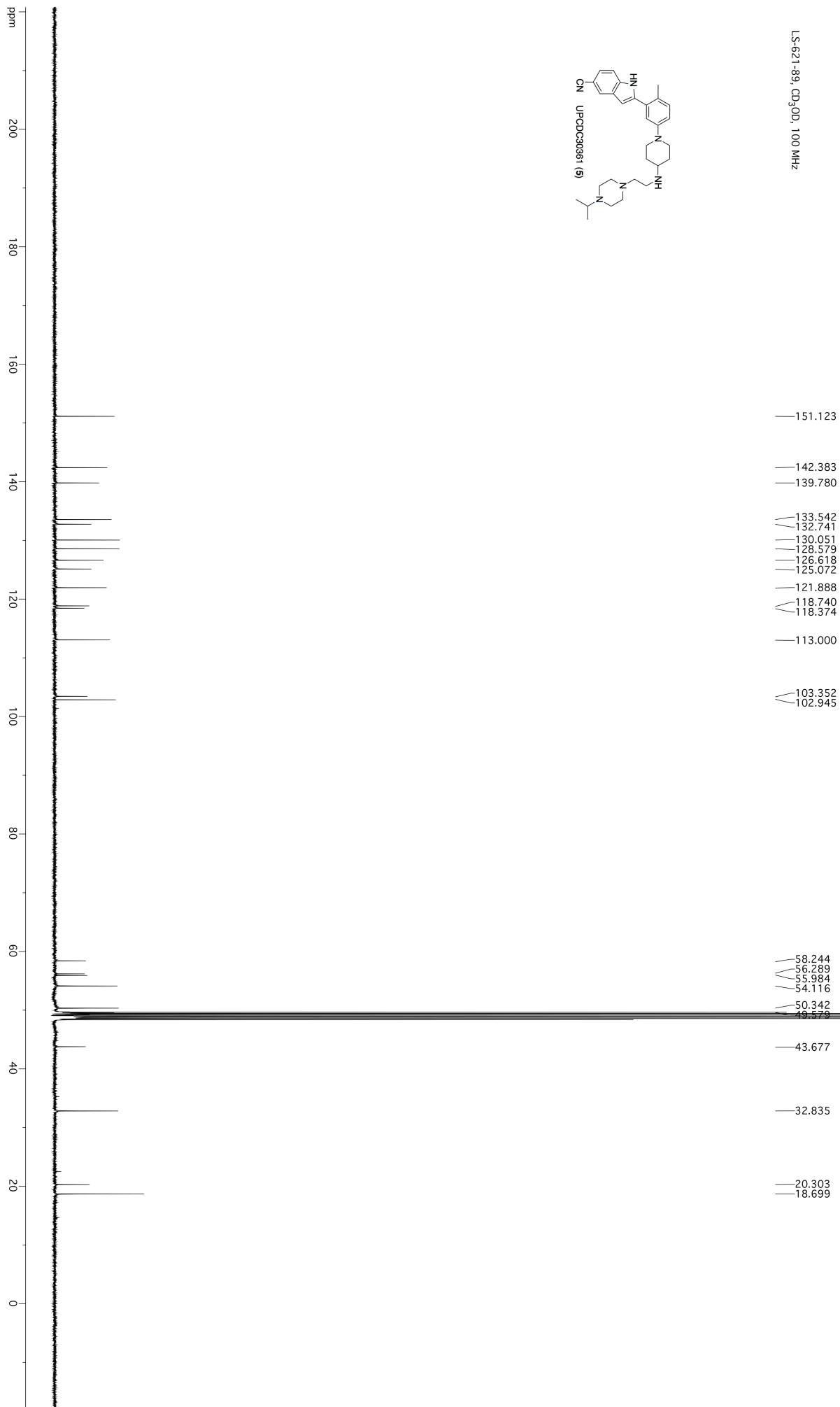


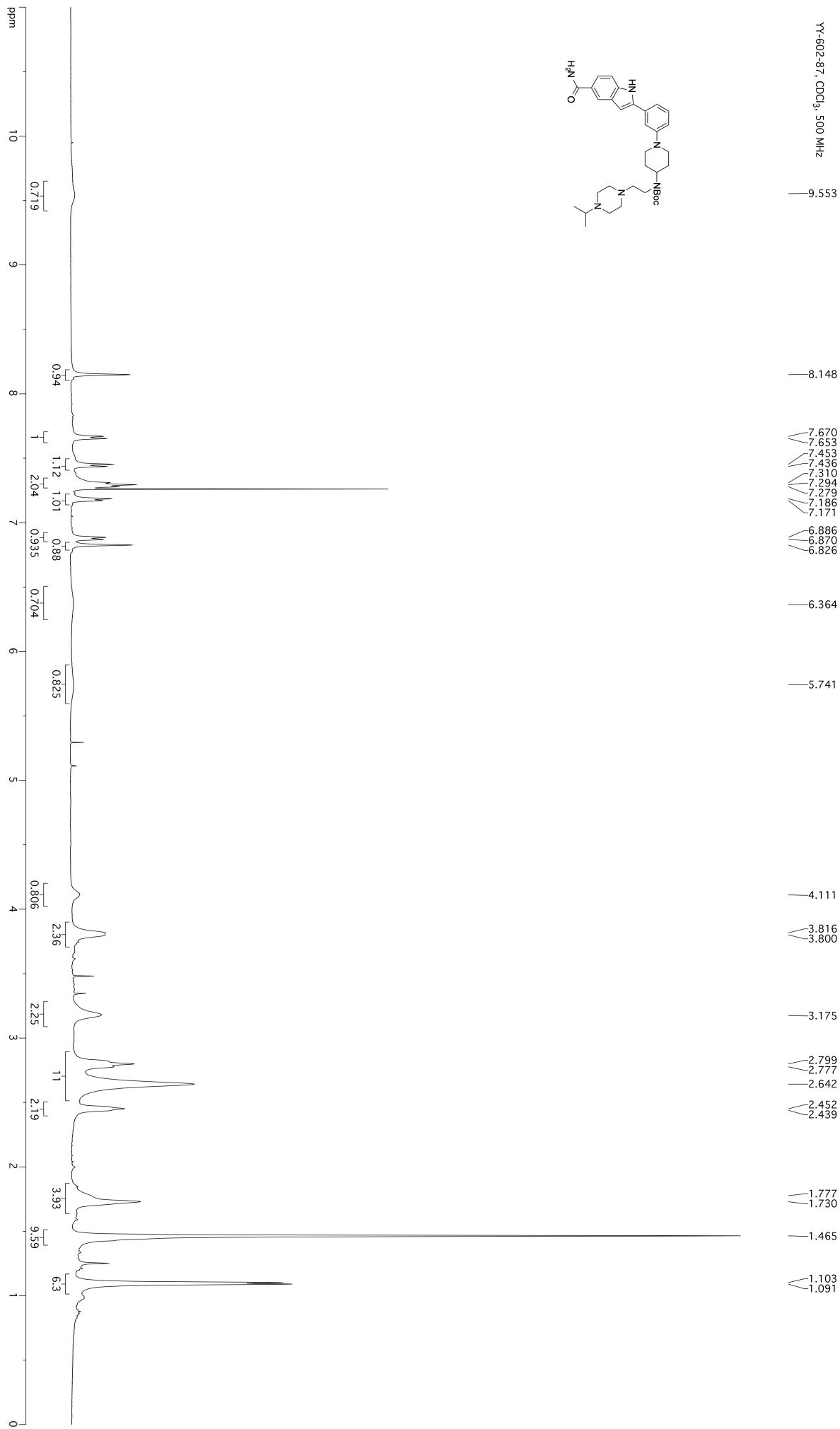


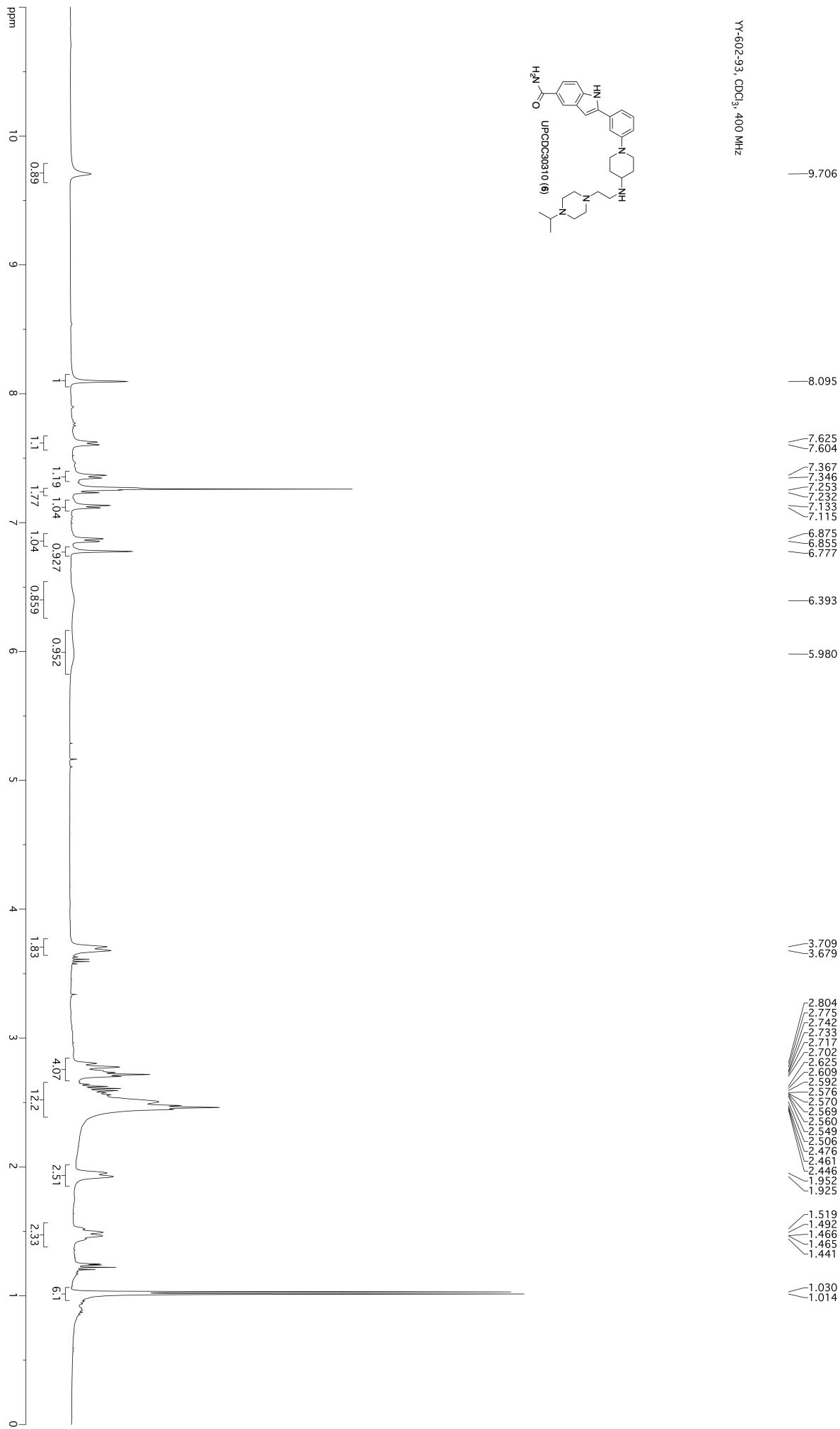
LS-621-89, CD<sub>3</sub>OD, 400 MHz

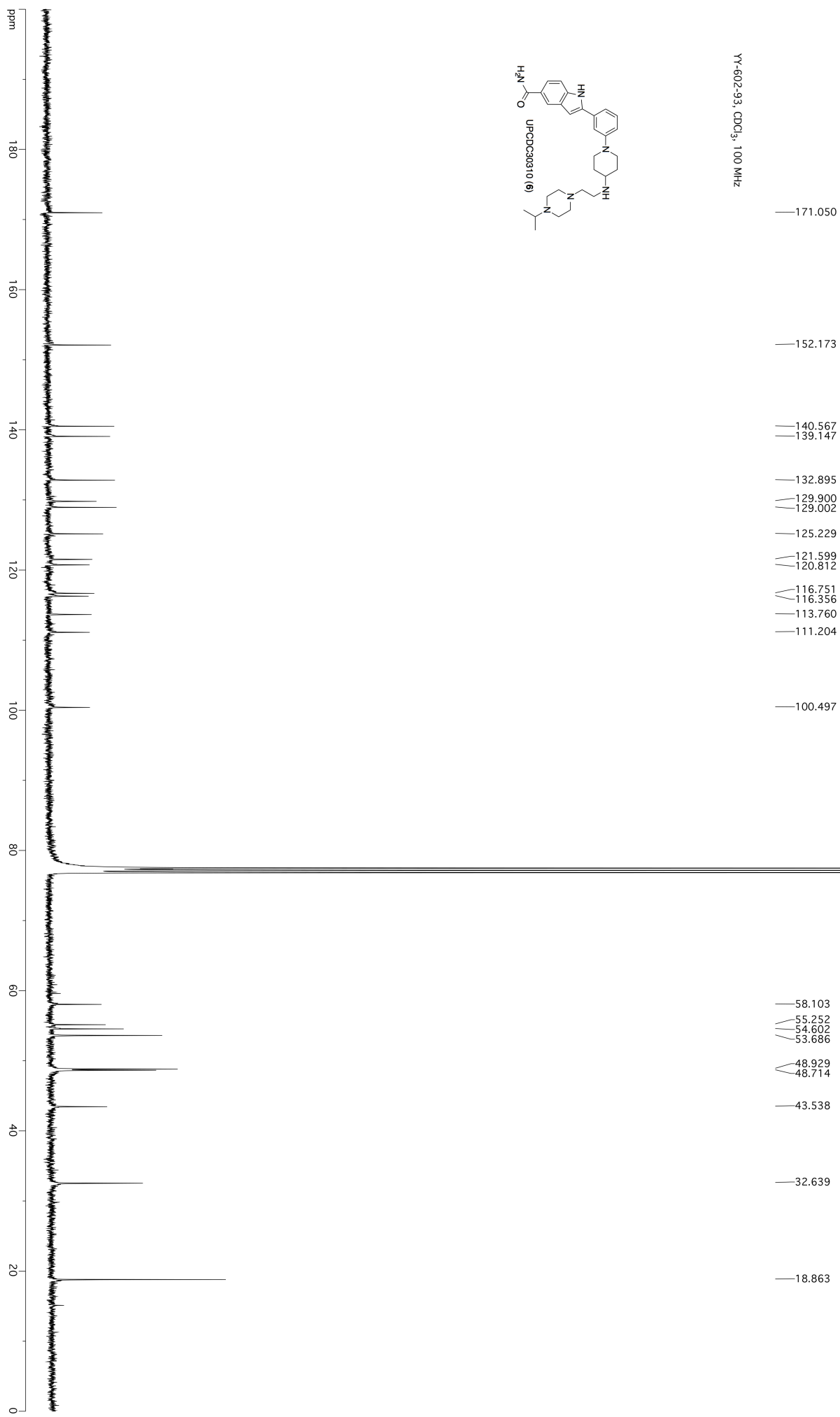


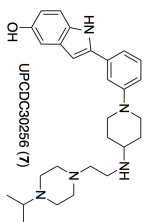
- 7.977
- 7.975
- 7.529
- 7.507
- 7.396
- 7.392
- 7.375
- 7.371
- 7.193
- 7.171
- 7.116
- 7.109
- 6.958
- 6.951
- 6.937
- 6.930
- 6.642
- 6.640
- 3.713
- 3.681
- 2.778
- 2.770
- 2.762
- 2.745
- 2.714
- 2.709
- 2.661
- 2.645
- 2.635
- 2.629
- 2.612
- 2.609
- 2.597
- 2.592
- 2.582
- 2.556
- 2.532
- 2.516
- 2.499
- 2.374
- 2.013
- 1.986
- 1.567
- 1.558
- 1.536
- 1.528
- 1.508
- 1.499
- 1.478
- 1.468
- 1.080
- 1.064



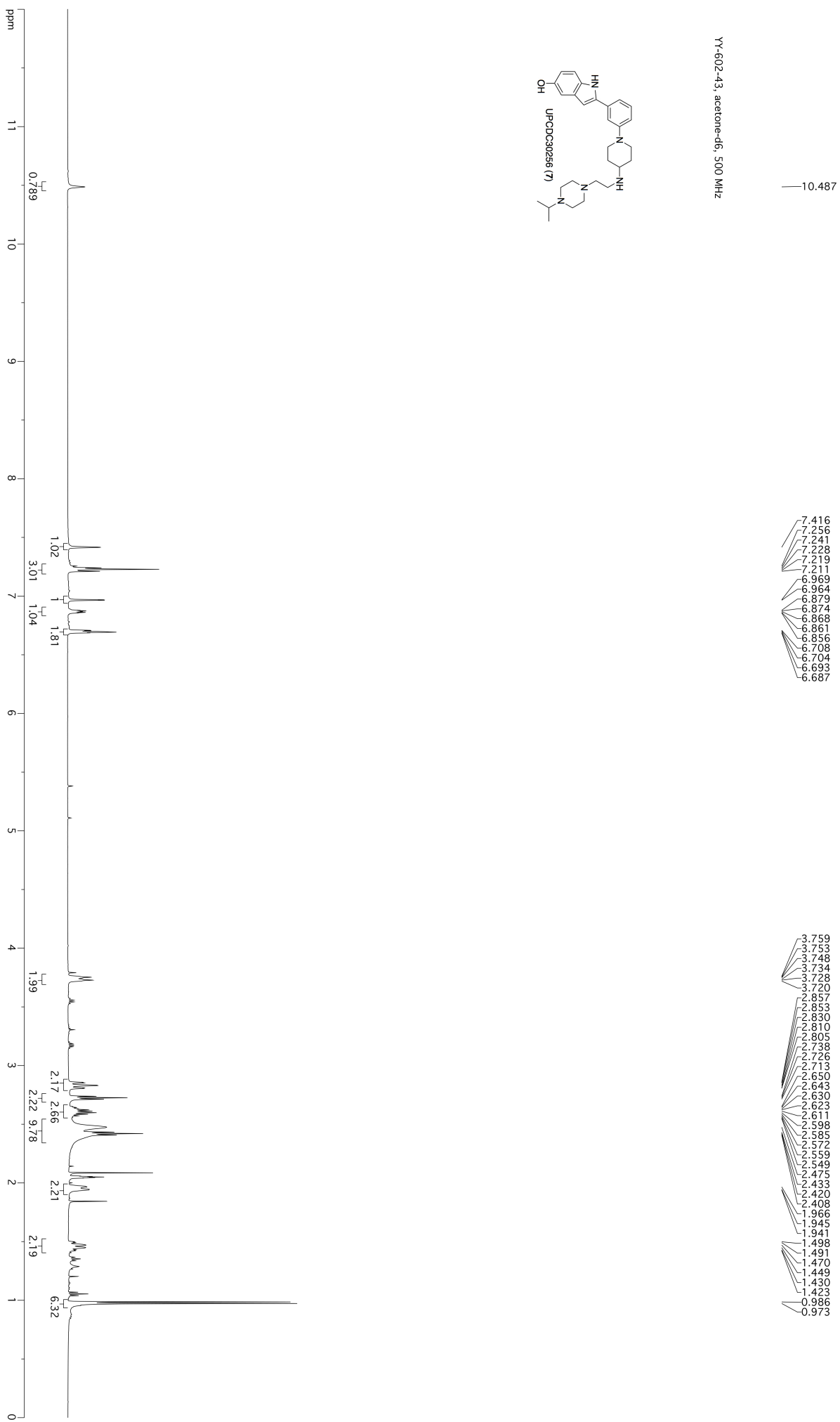




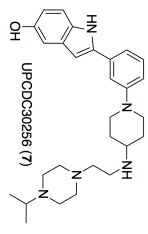
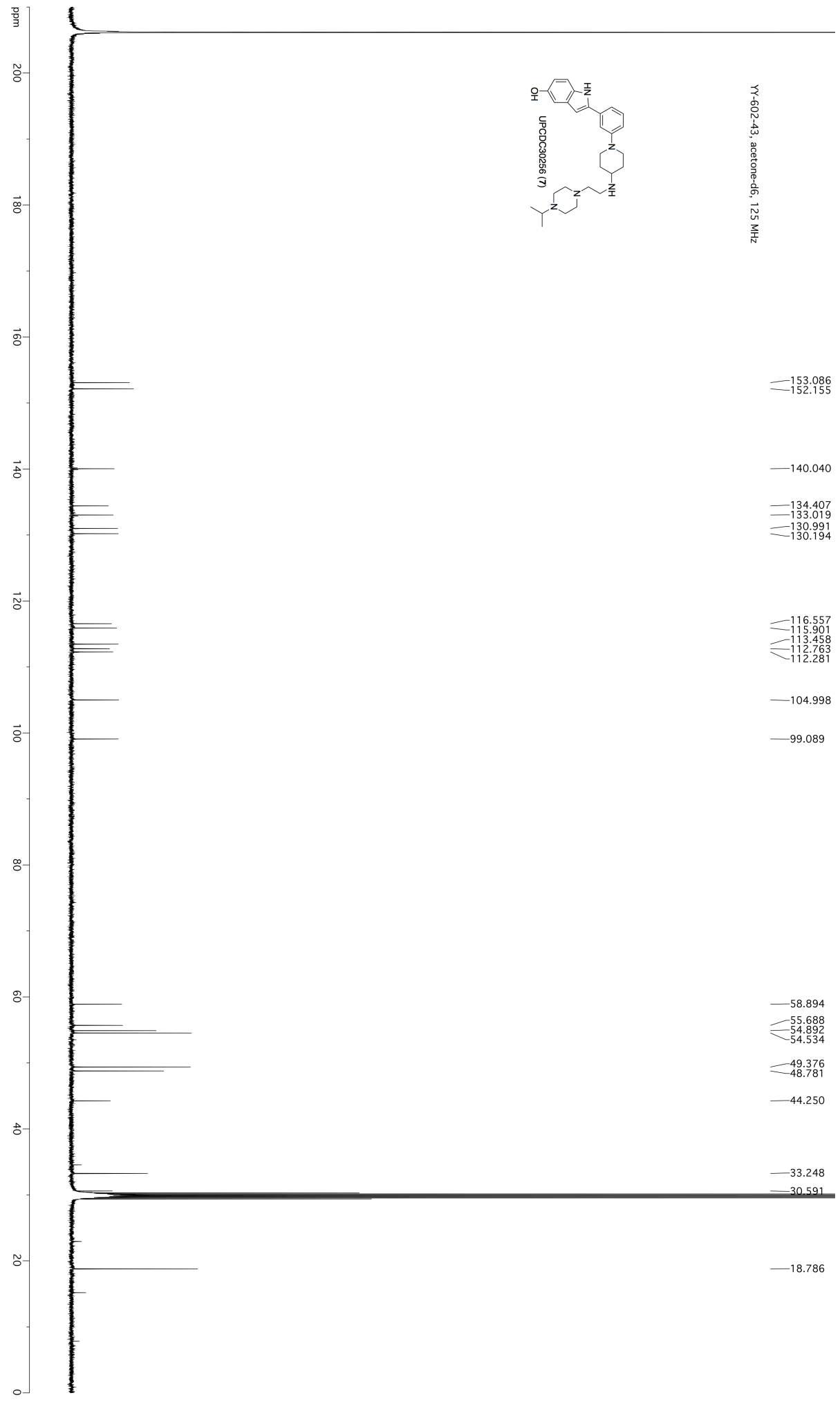




YY-602-43, acetone-d6, 500 MHz







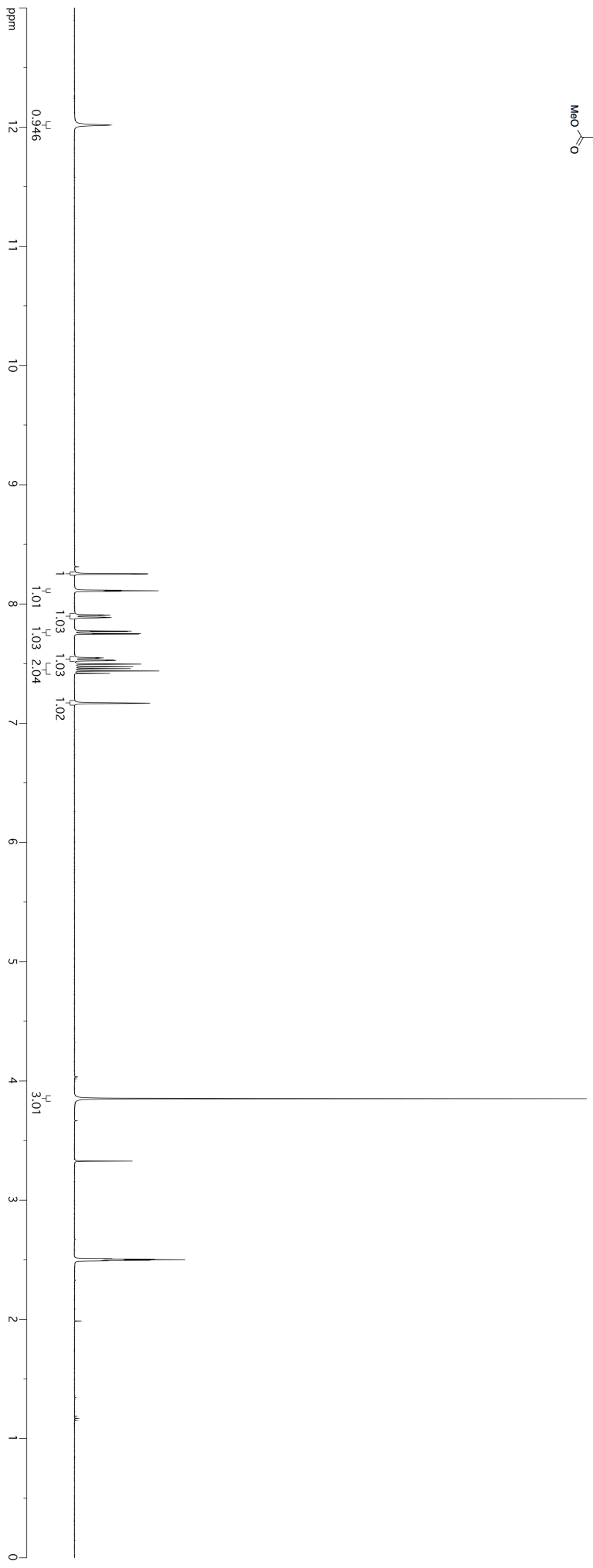
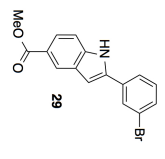
YY-602-43, acetone-d6, 125 MHz

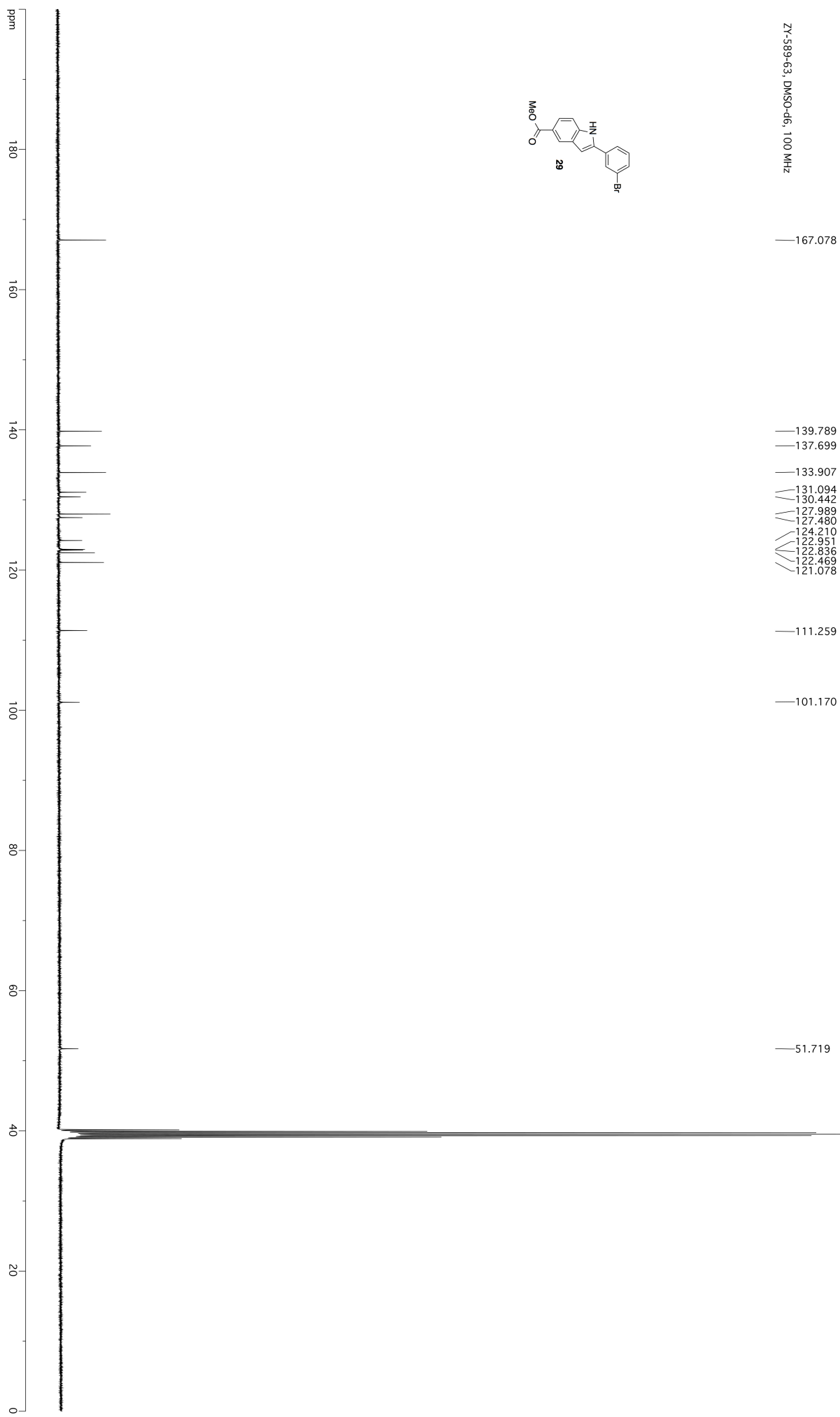
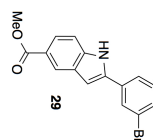
— 12.017

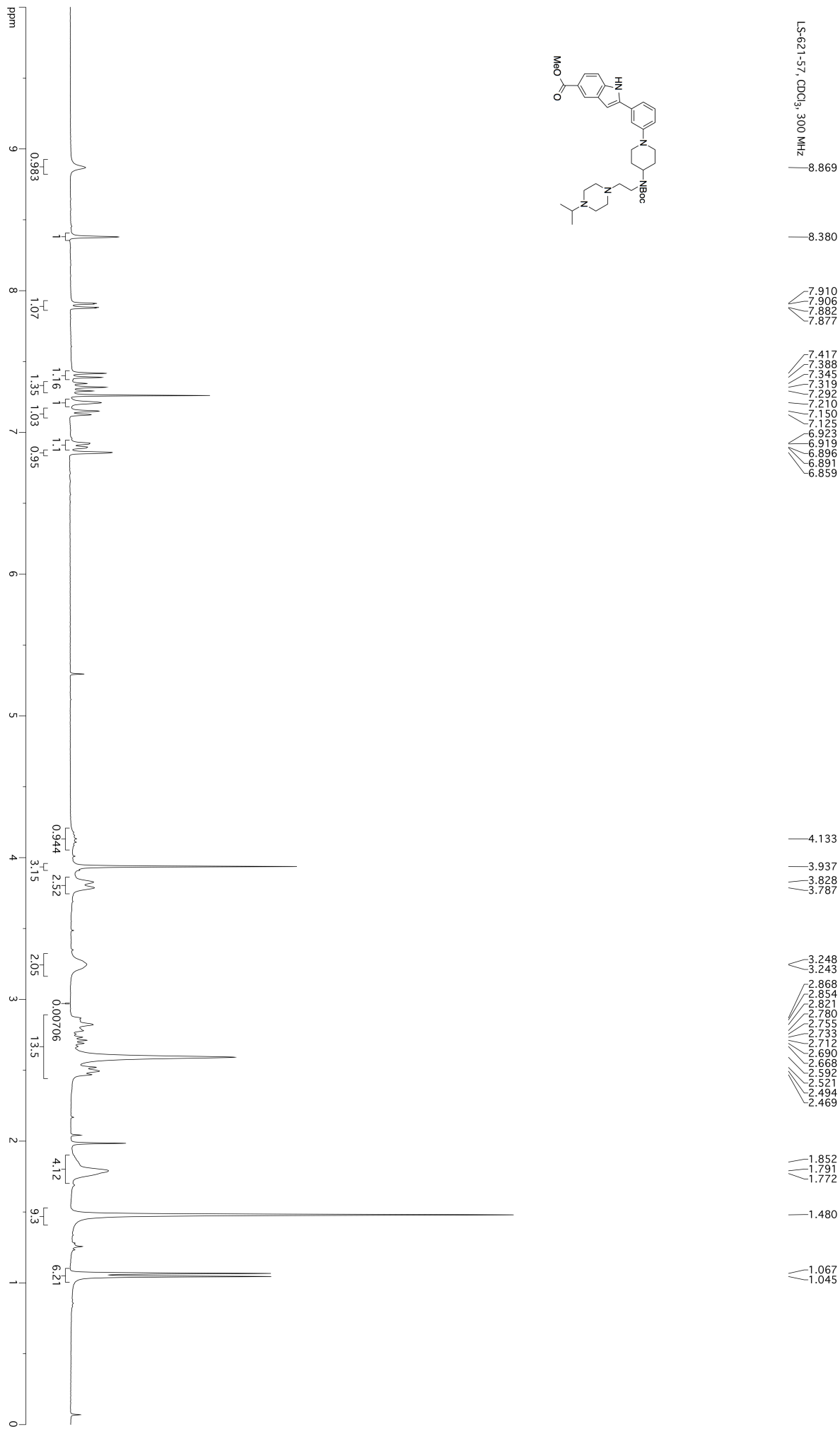
8.254  
8.251  
8.116  
8.111  
8.107  
7.909  
7.907  
7.902  
7.890  
7.887  
7.886  
7.883  
7.773  
7.769  
7.752  
7.748  
7.550  
7.548  
7.546  
7.543  
7.530  
7.528  
7.526  
7.523  
7.496  
7.474  
7.458  
7.438  
7.418  
7.168

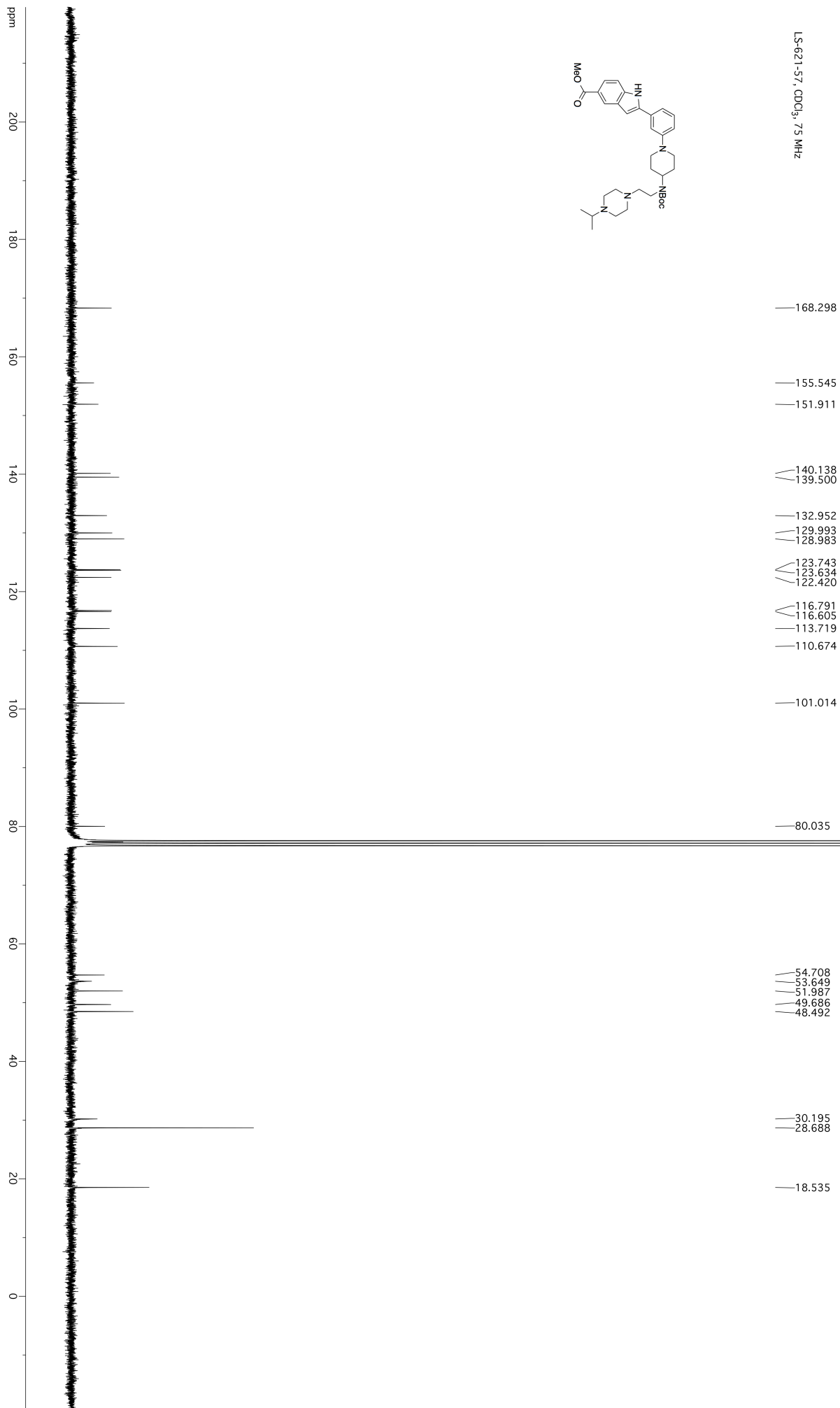
— 3.851

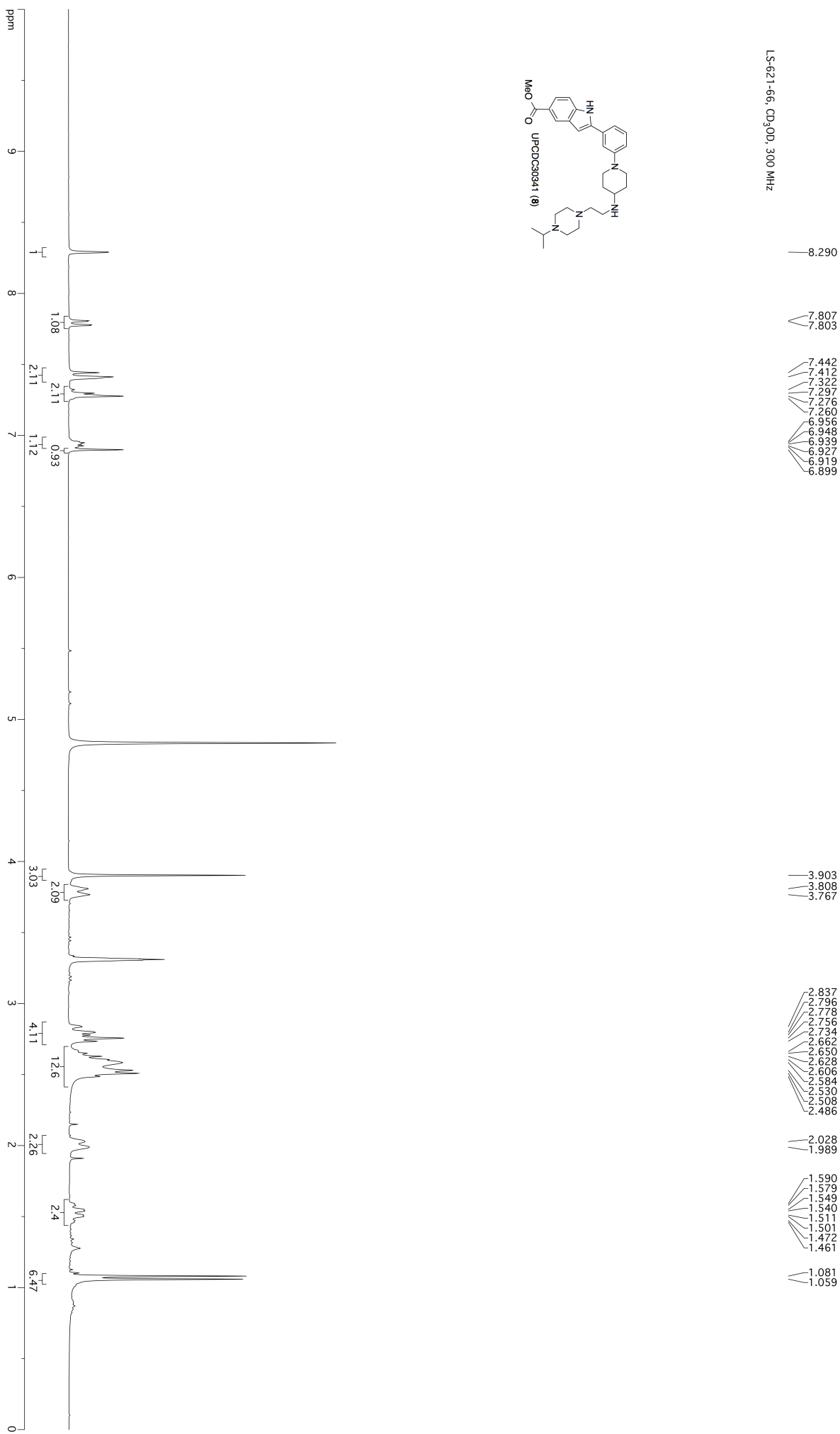
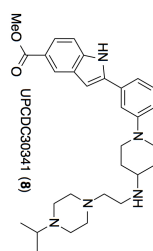
ZY-589-63-1H, DMSO-d6, 400 MHz

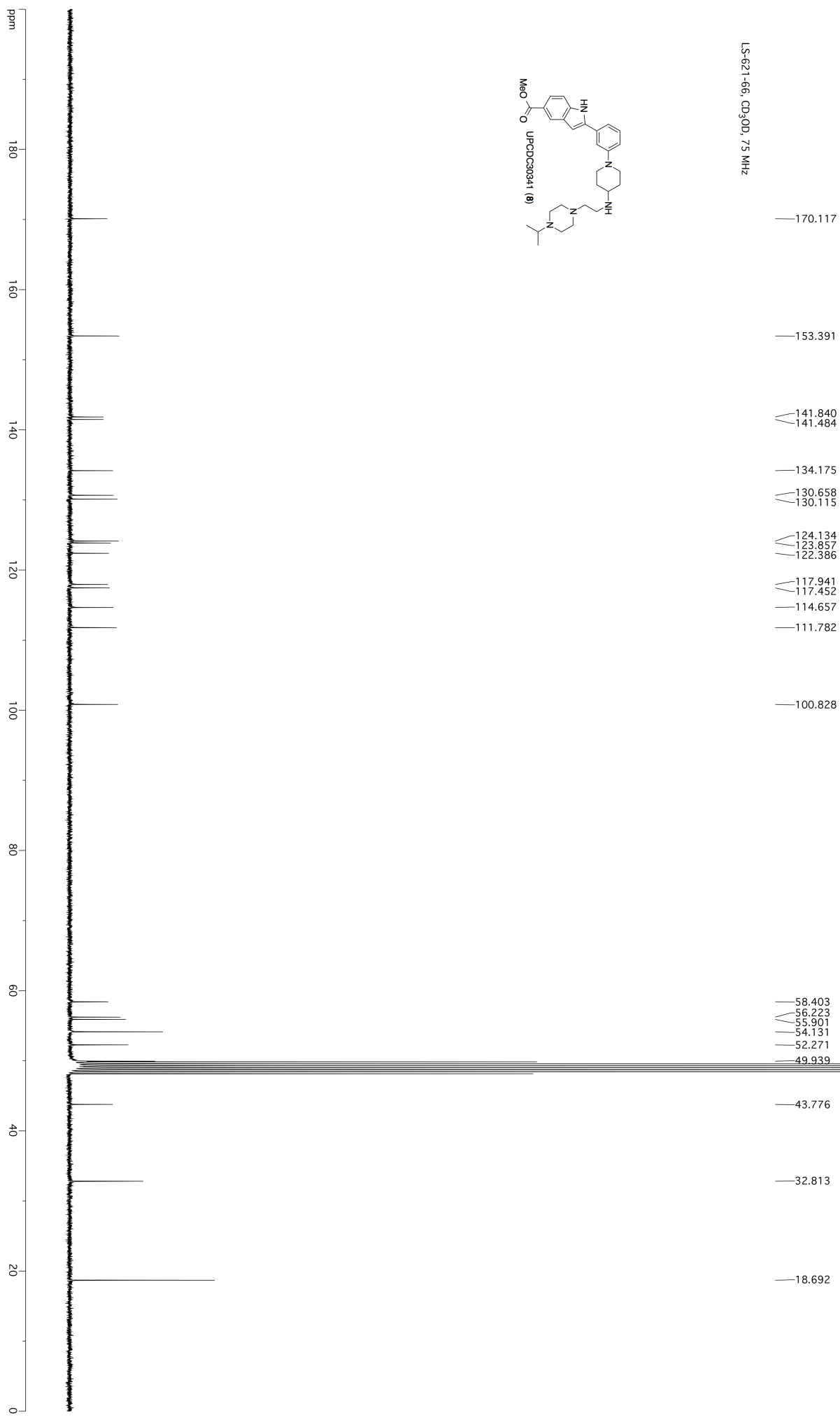


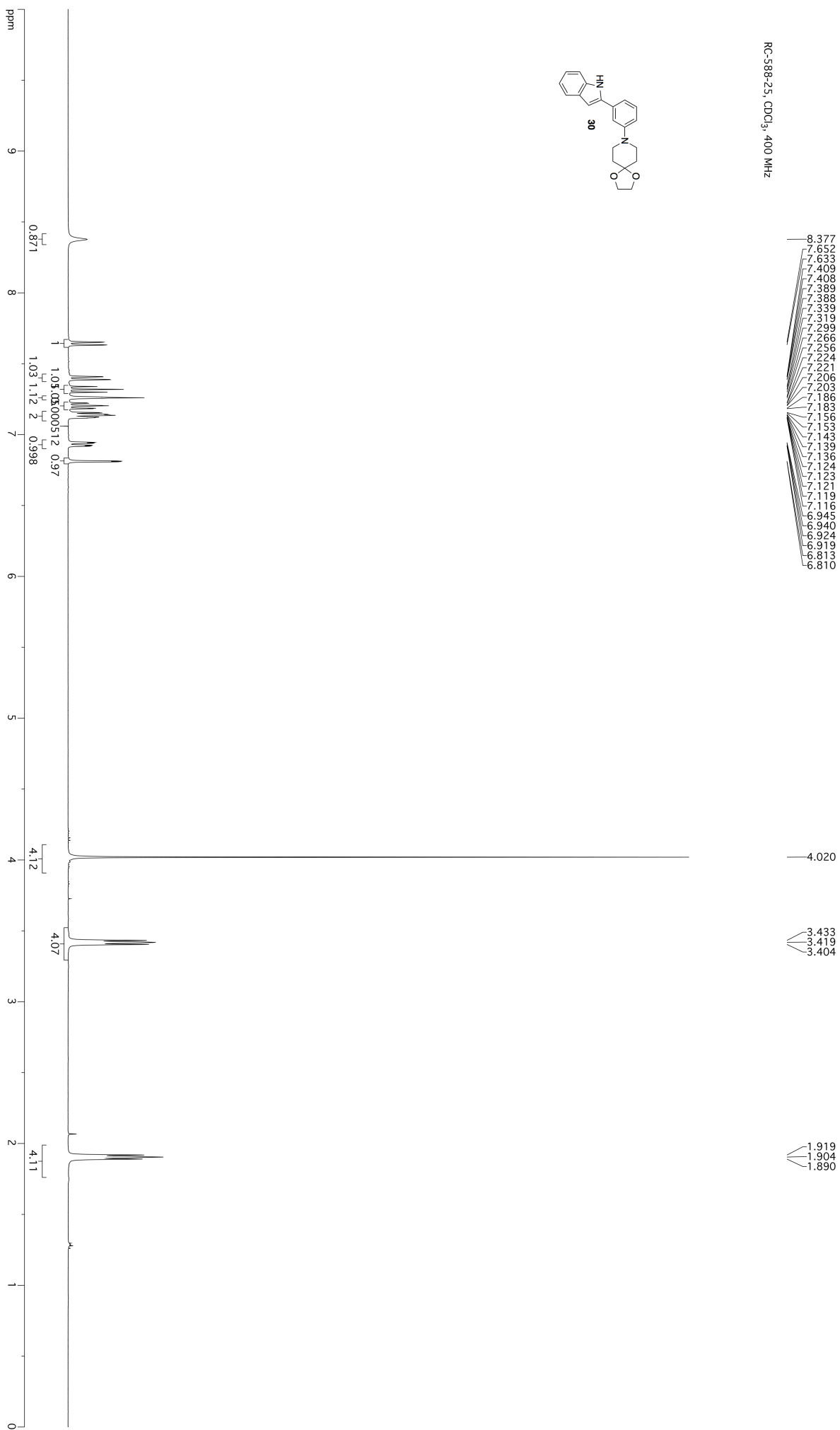
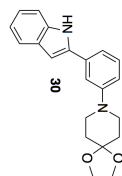
ZY-589-63, DMSO-d<sub>6</sub>, 100 MHz



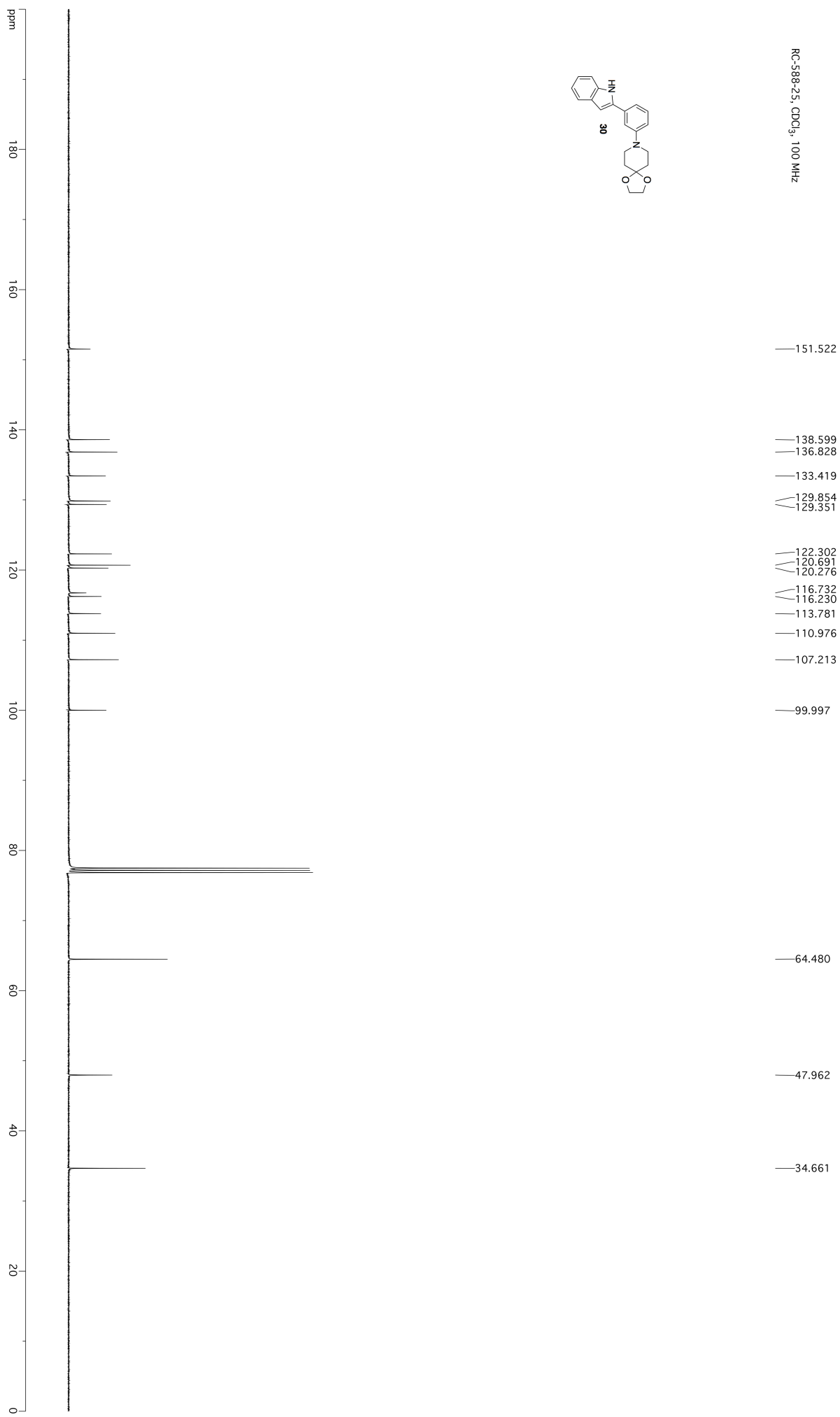
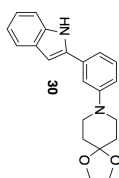


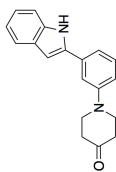
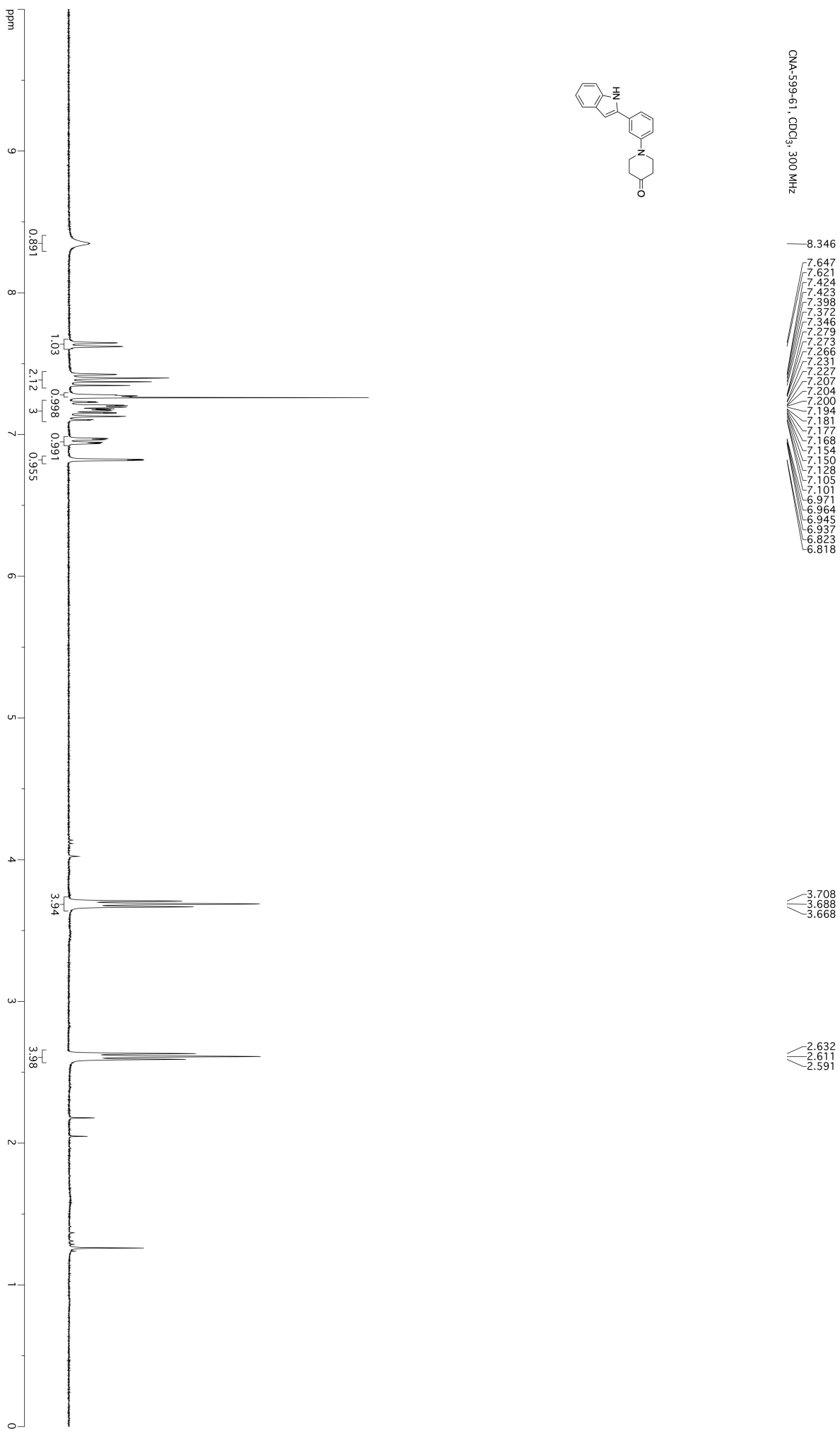
LS-621-66, CD<sub>3</sub>OD, 300 MHz

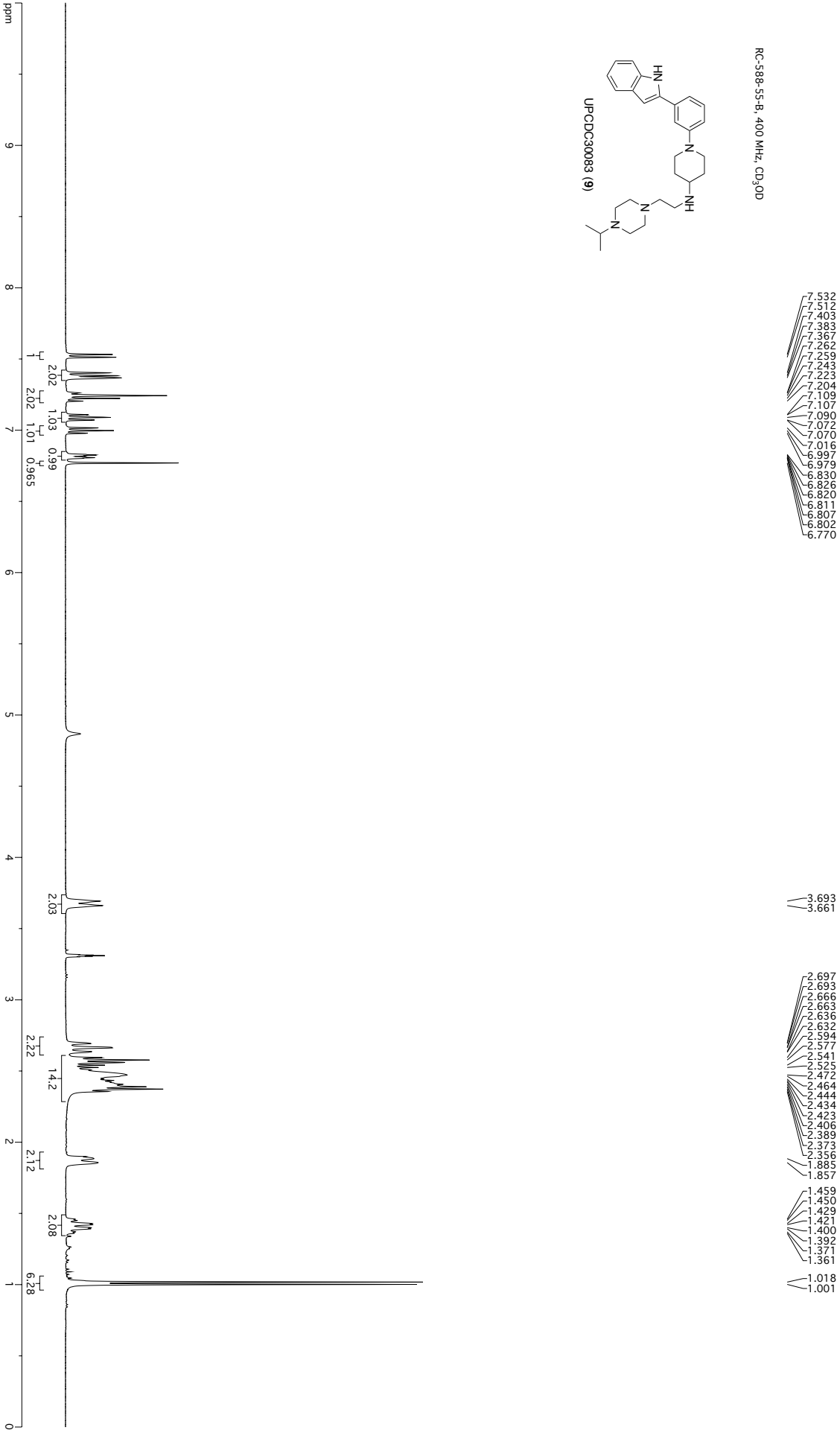
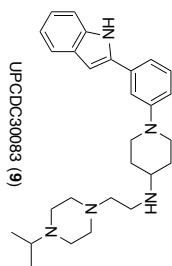


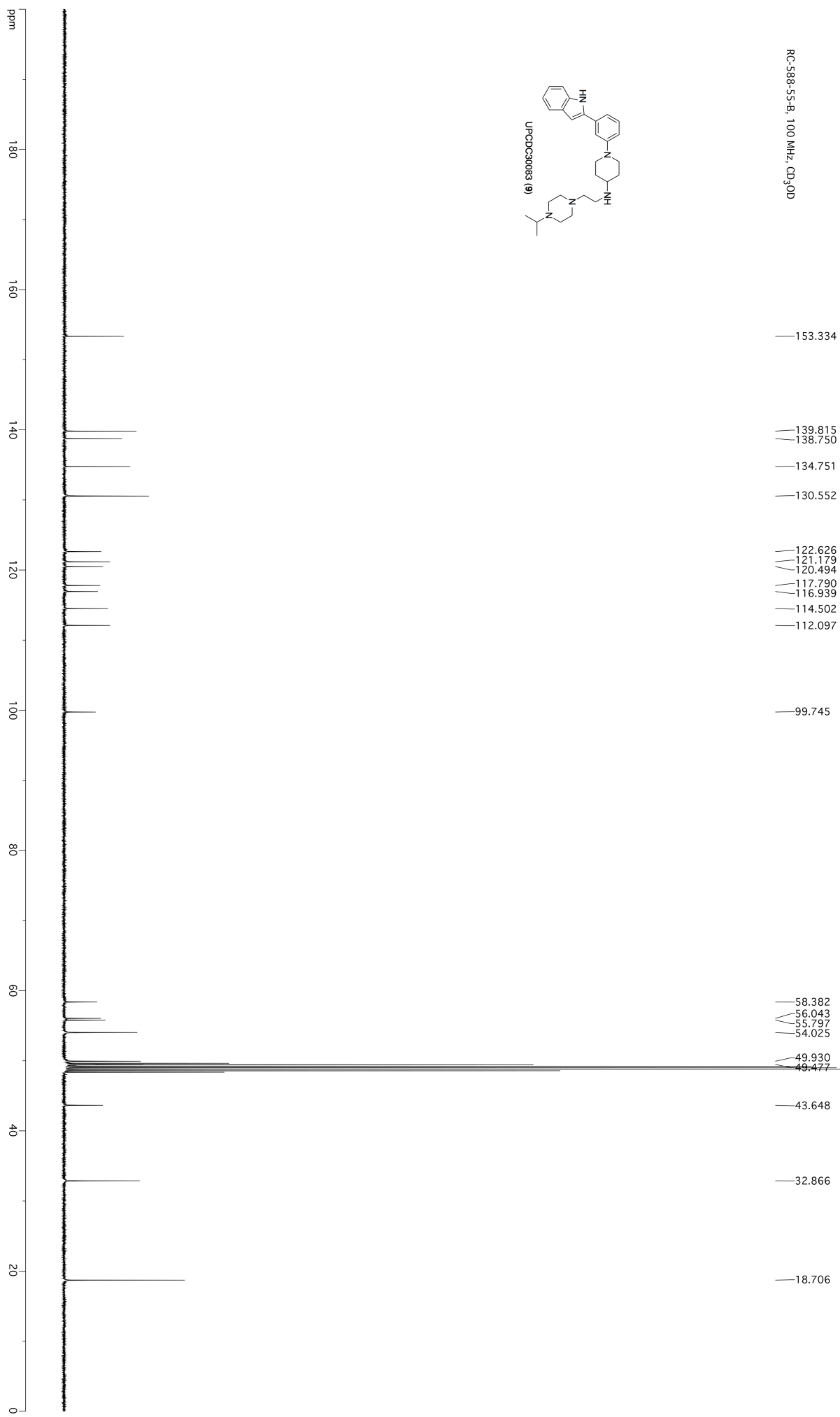
RC-588-25, CDCl<sub>3</sub>, 400 MHz

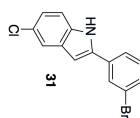


RC-588-25, CDCl<sub>3</sub>, 100 MHz

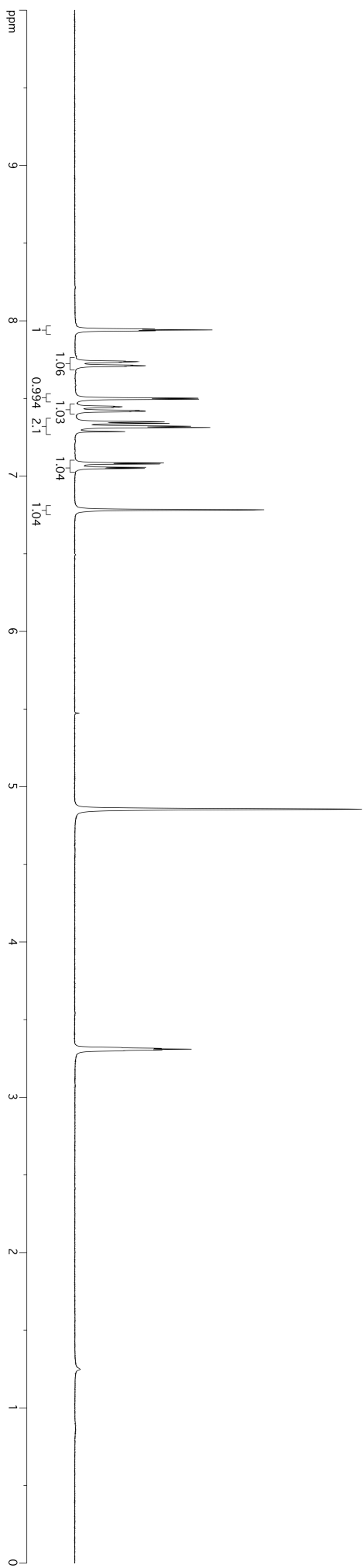
GNA-599-61, CDCl<sub>3</sub>, 300 MHz

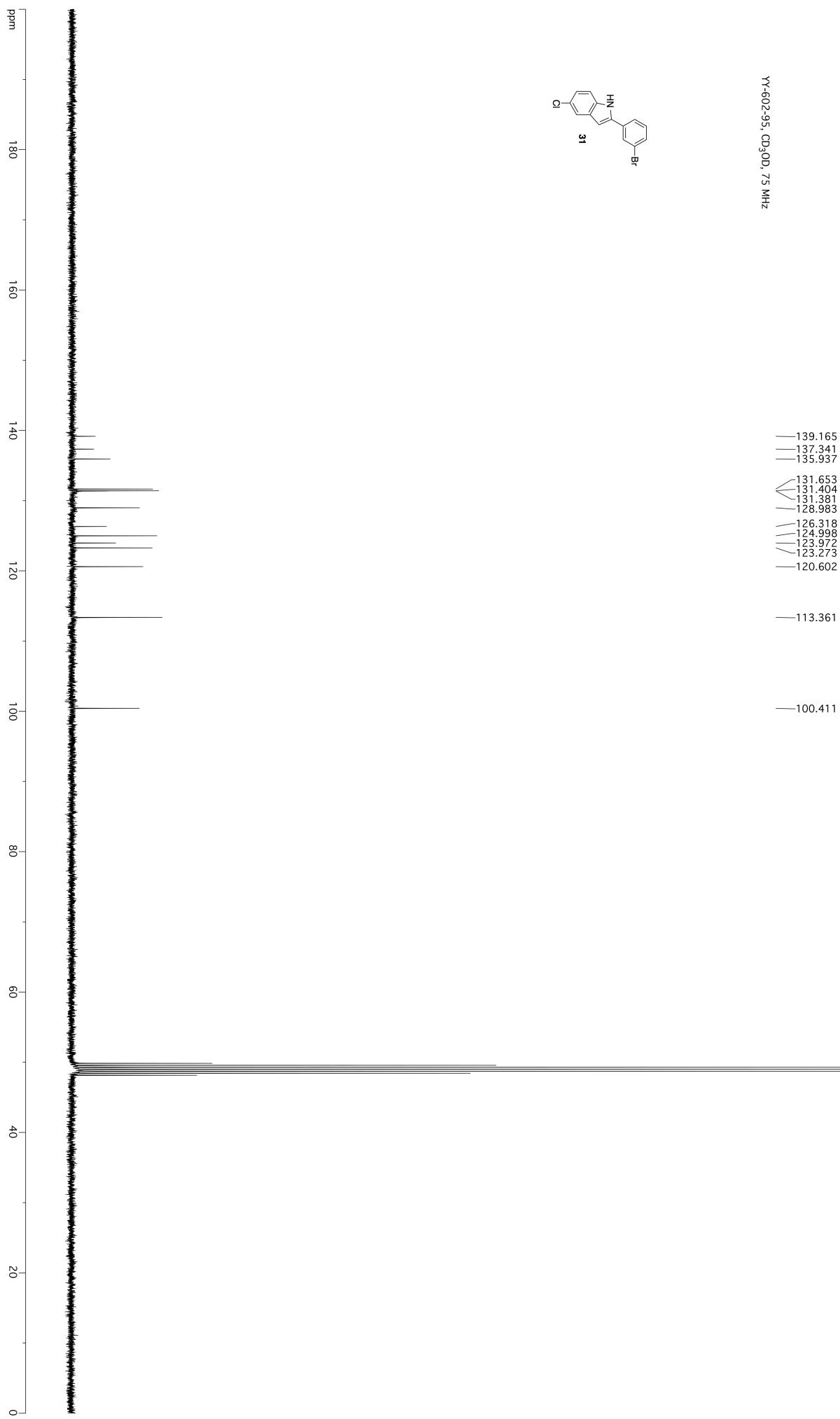
RC-588-55-B, 400 MHz, CD<sub>3</sub>OD



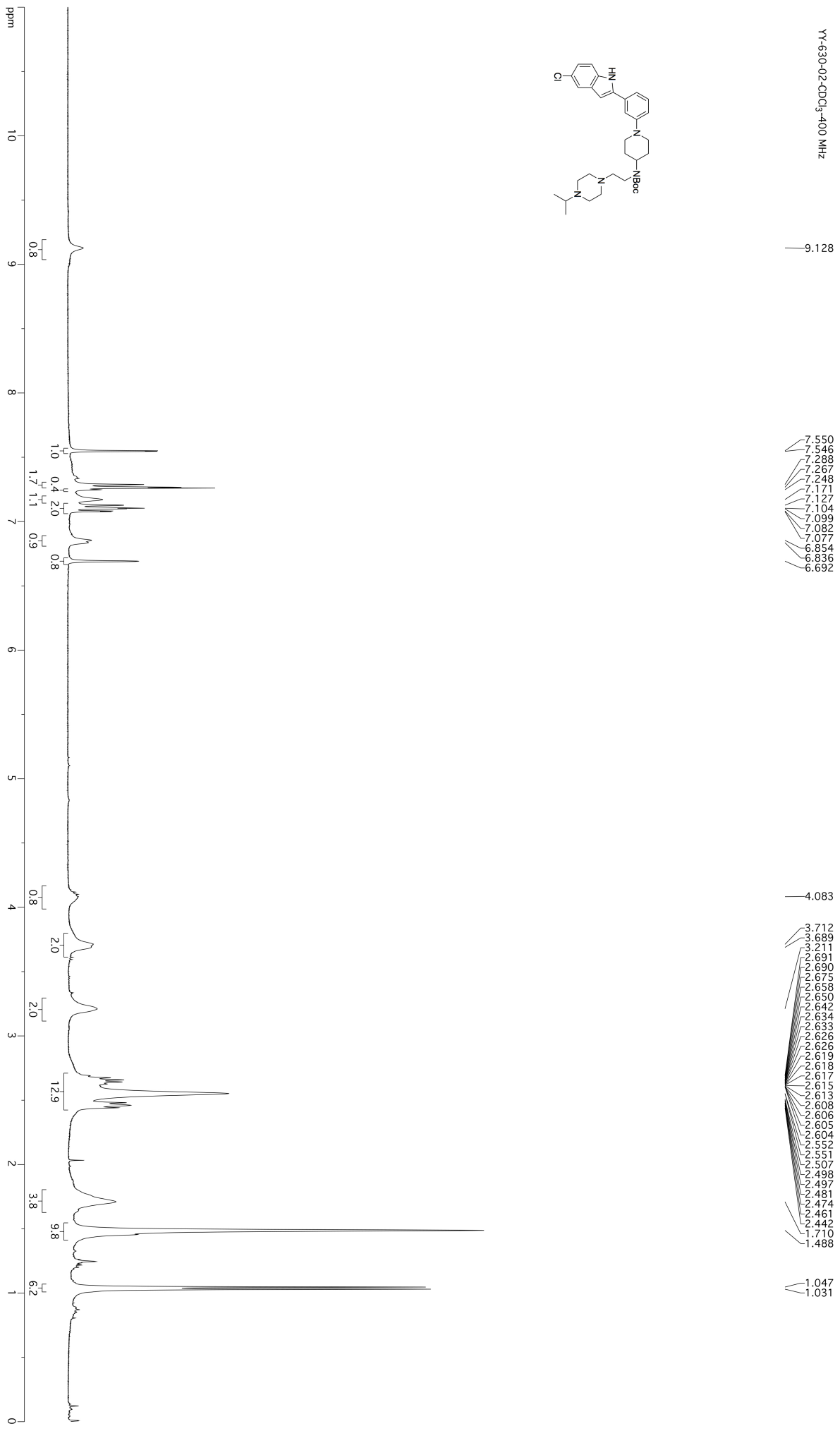
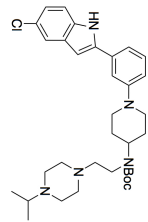
YY-602-95, CD<sub>3</sub>OD, 300 MHz

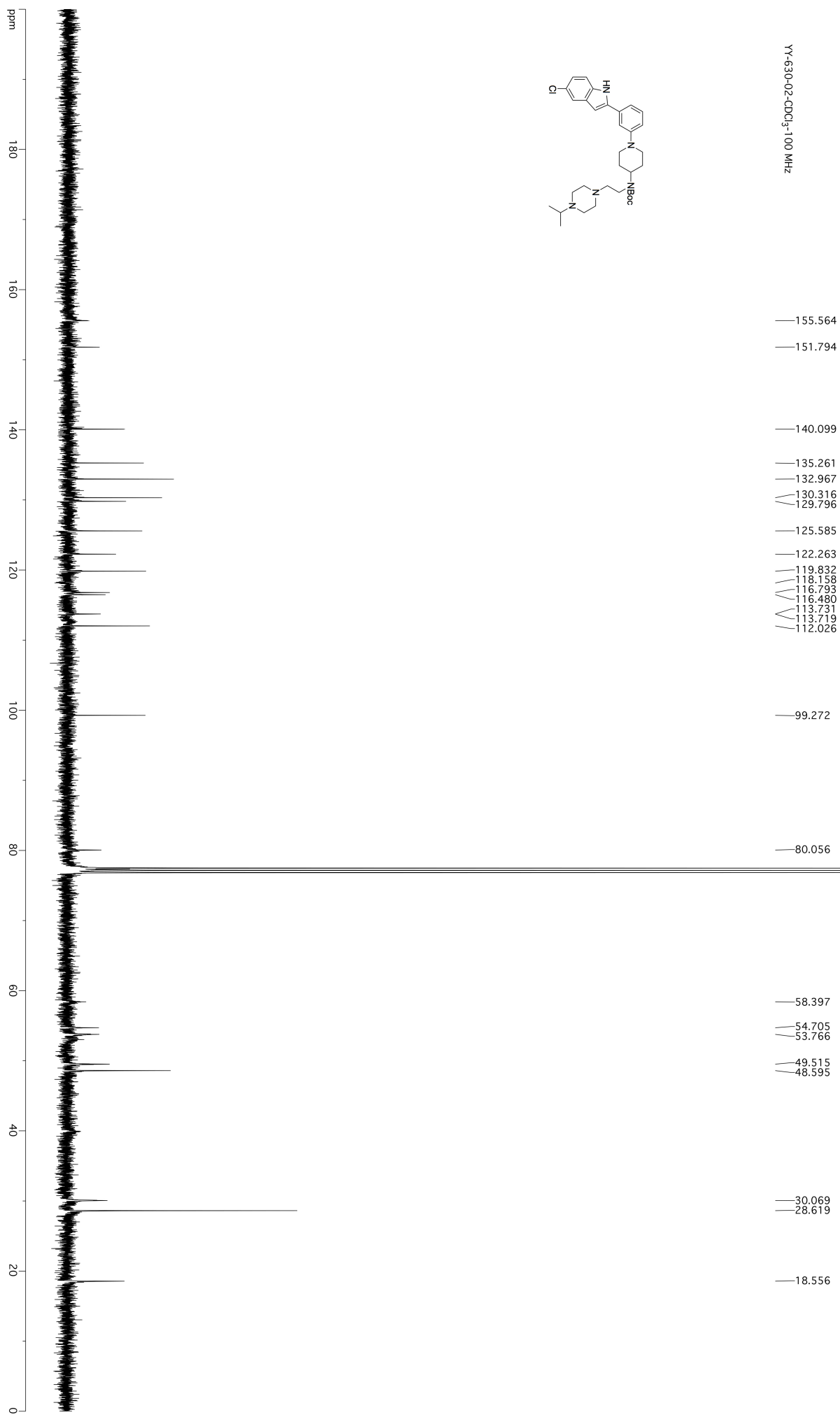
7.947  
7.941  
7.936  
7.741  
7.737  
7.732  
7.715  
7.711  
7.706  
7.502  
7.496  
7.450  
7.447  
7.444  
7.441  
7.424  
7.420  
7.418  
7.414  
7.350  
7.321  
7.313  
7.287  
7.085  
7.078  
7.056  
7.049  
6.783



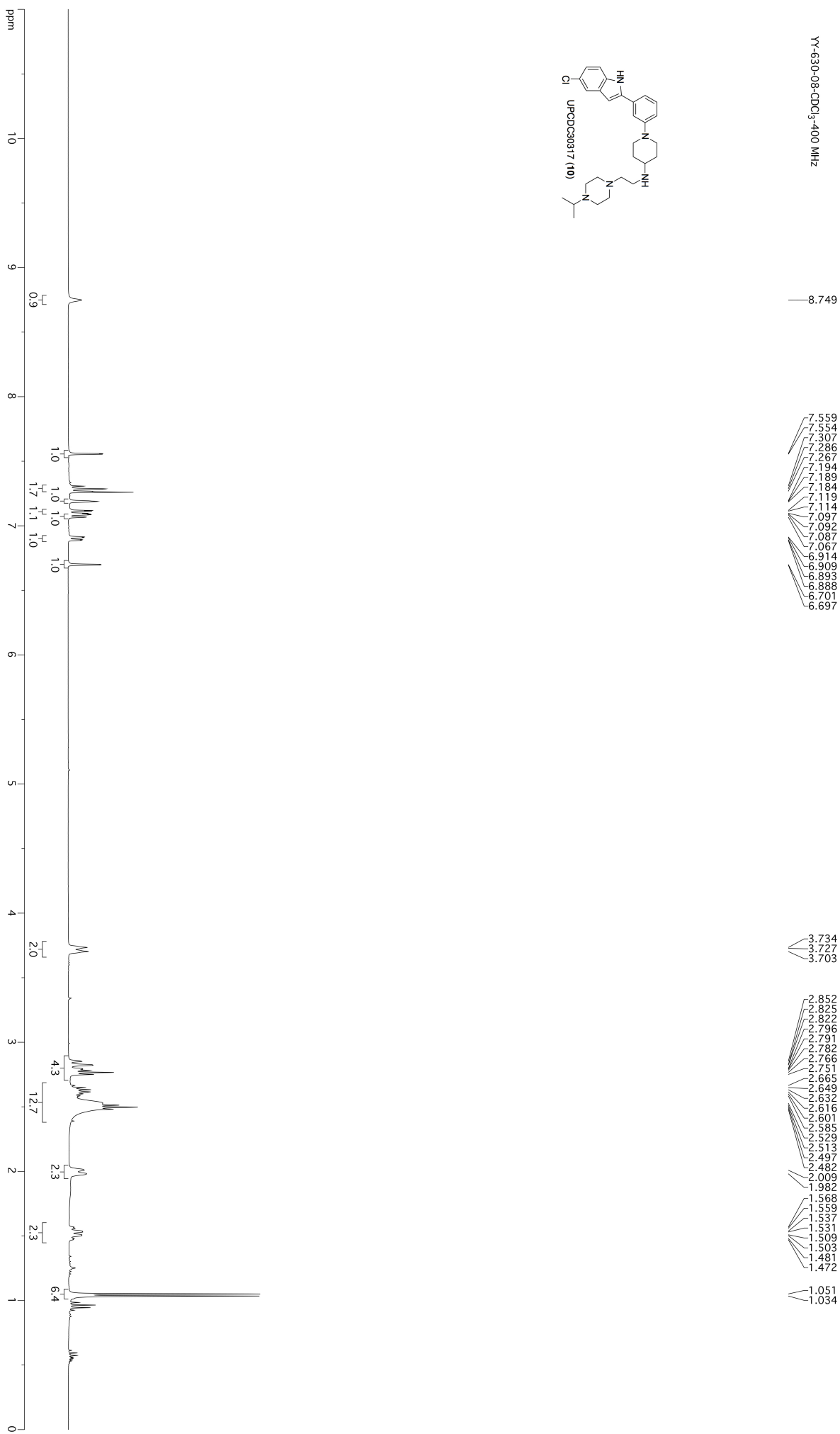
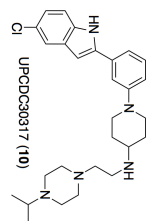


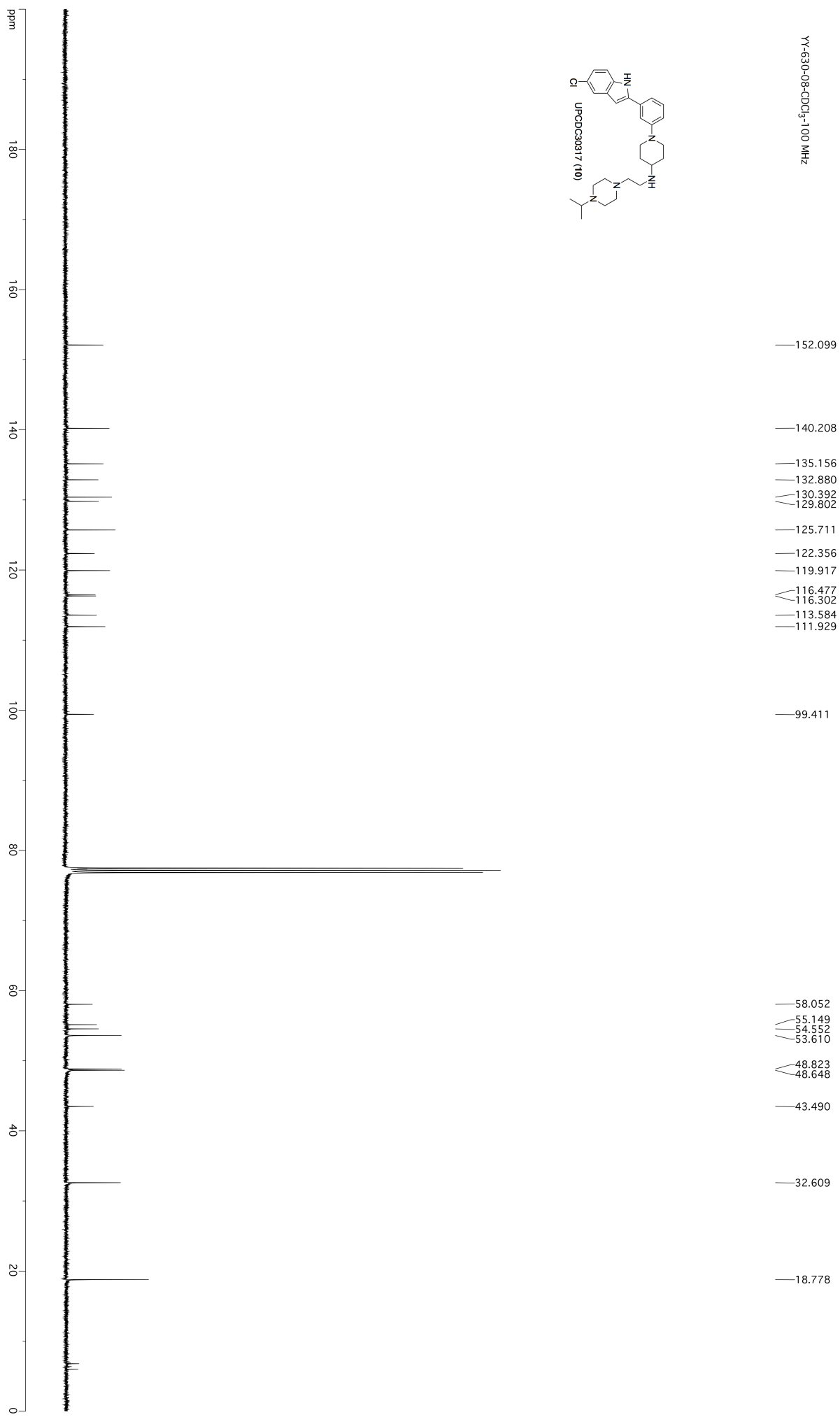
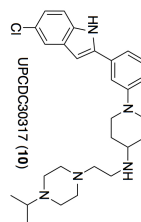
YY-630-02-CDCl<sub>3</sub>-400 MHz

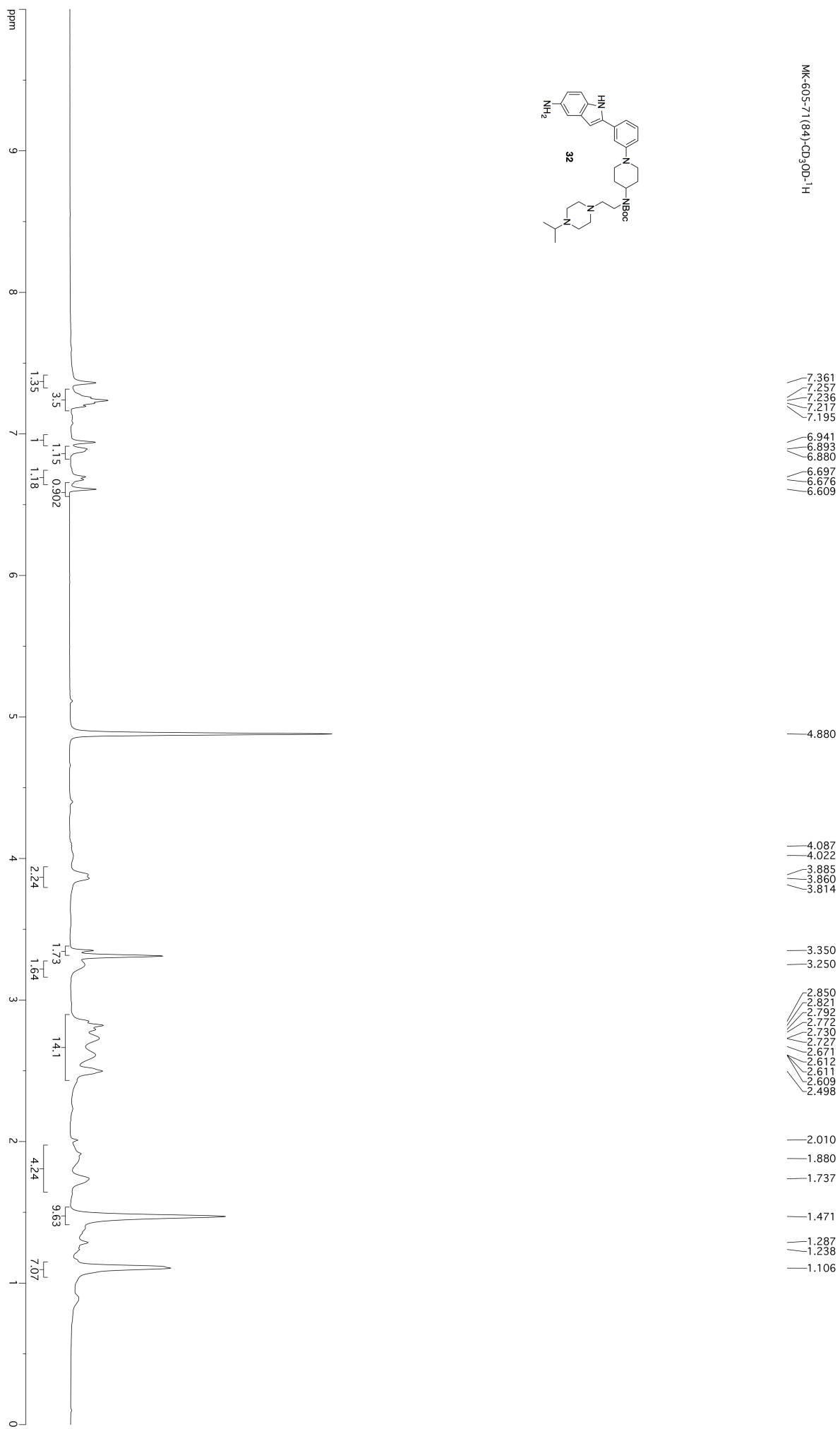
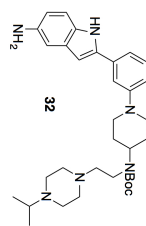


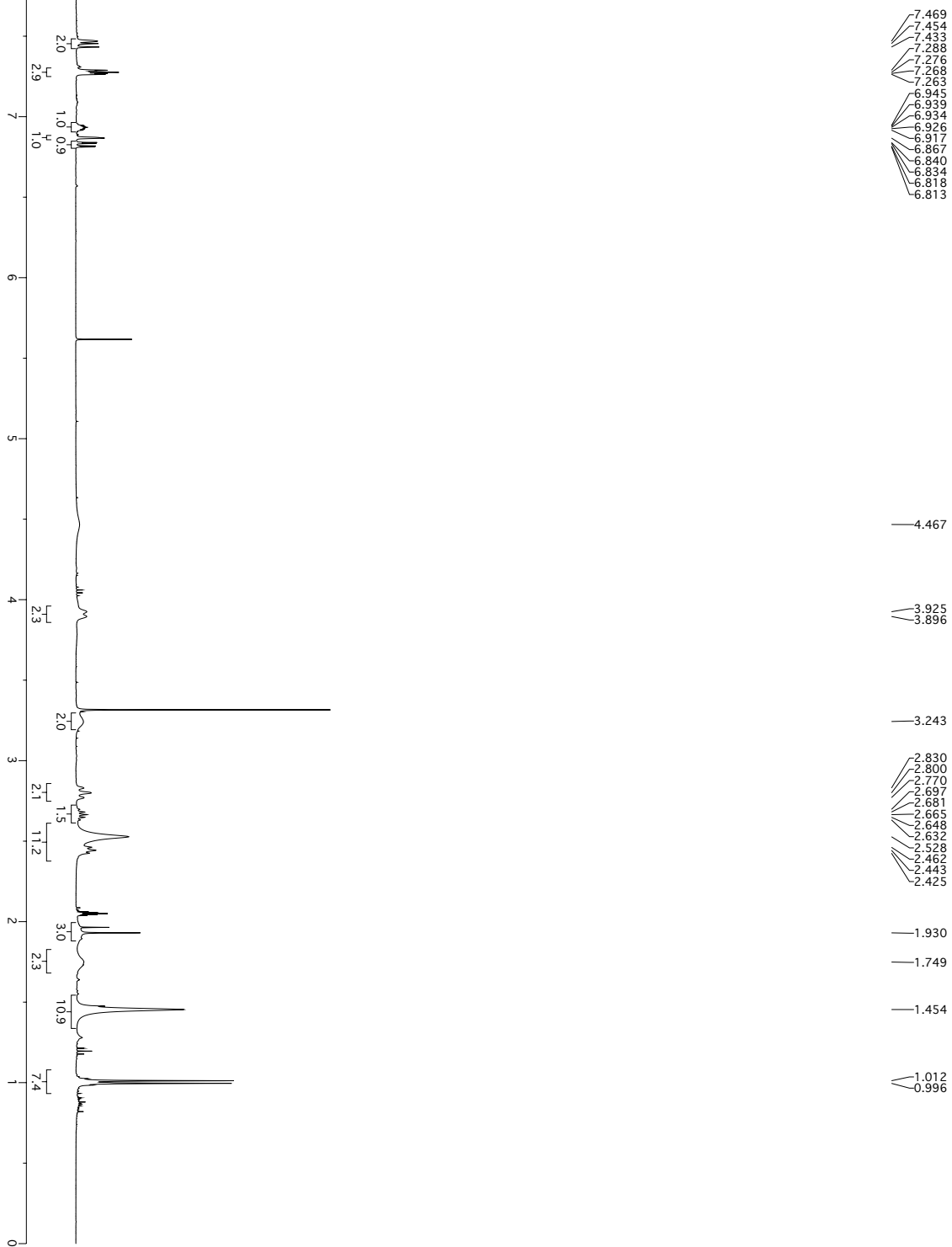
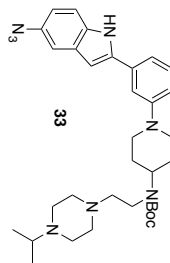




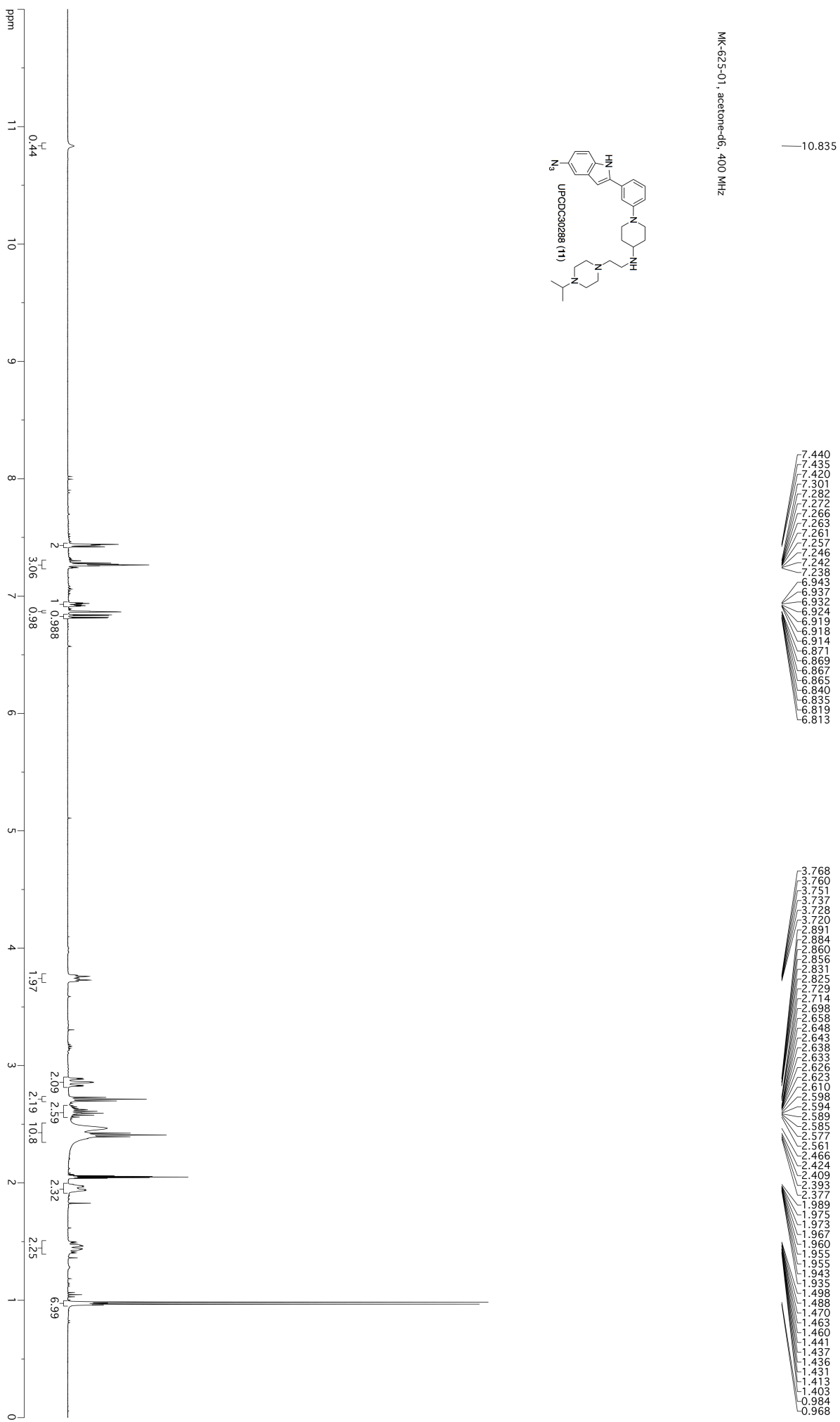
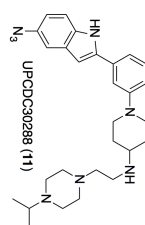
YY-630-08-CDCl<sub>3</sub>-400 MHz

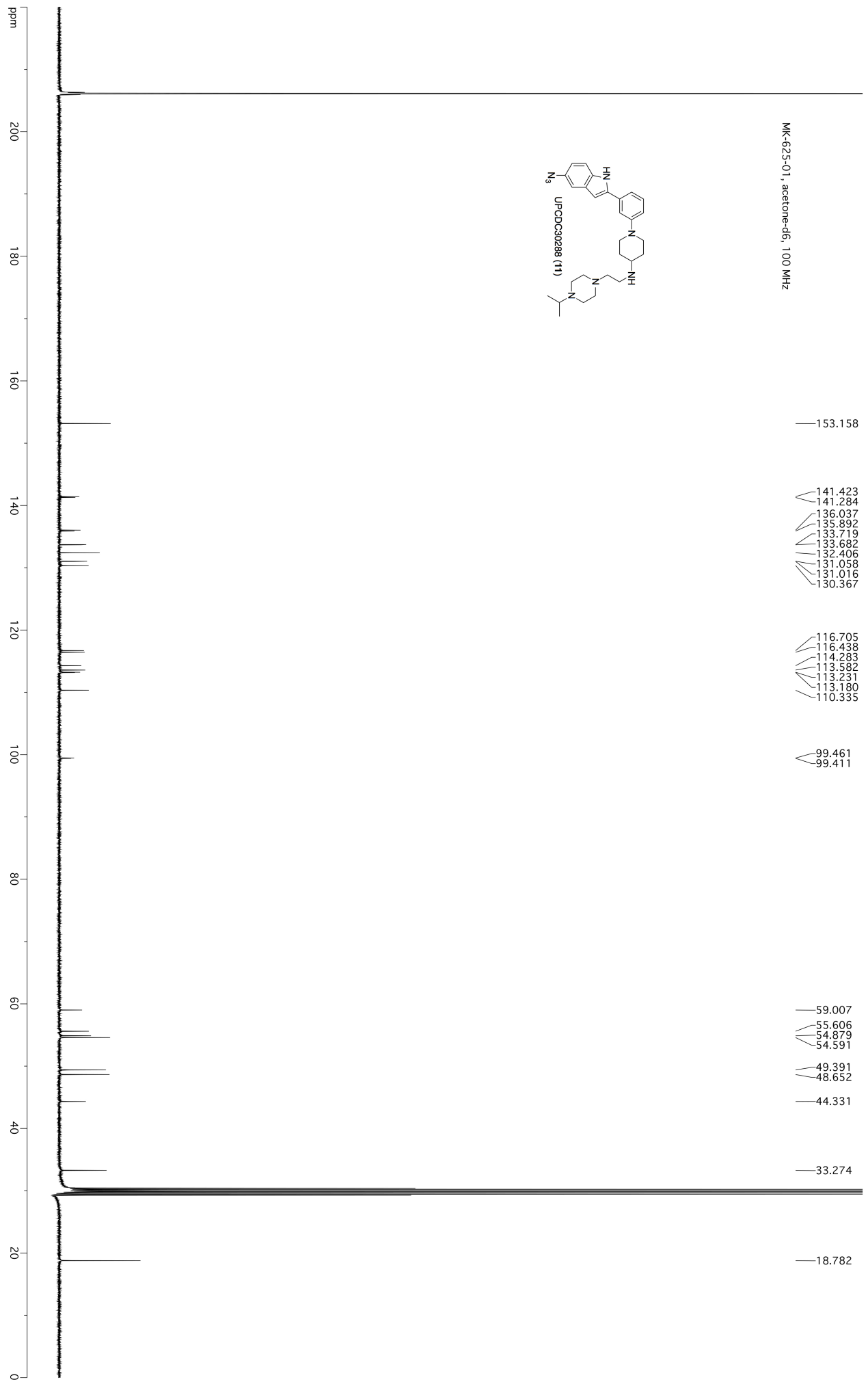
YY-630-08-CDCl<sub>3</sub>-100 MHz

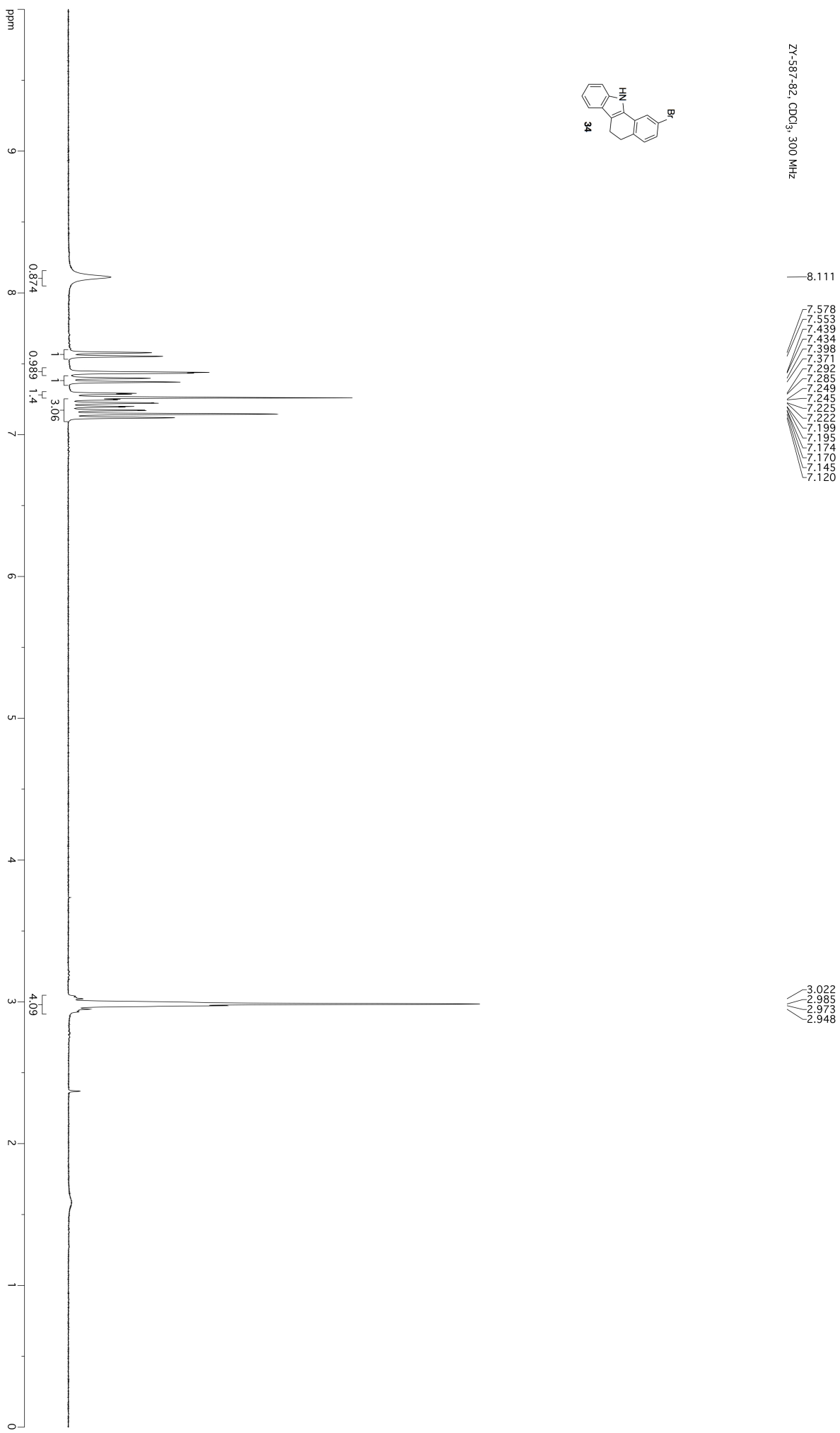
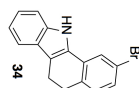
MK-605-71 (84)-CD<sub>3</sub>OD-1H

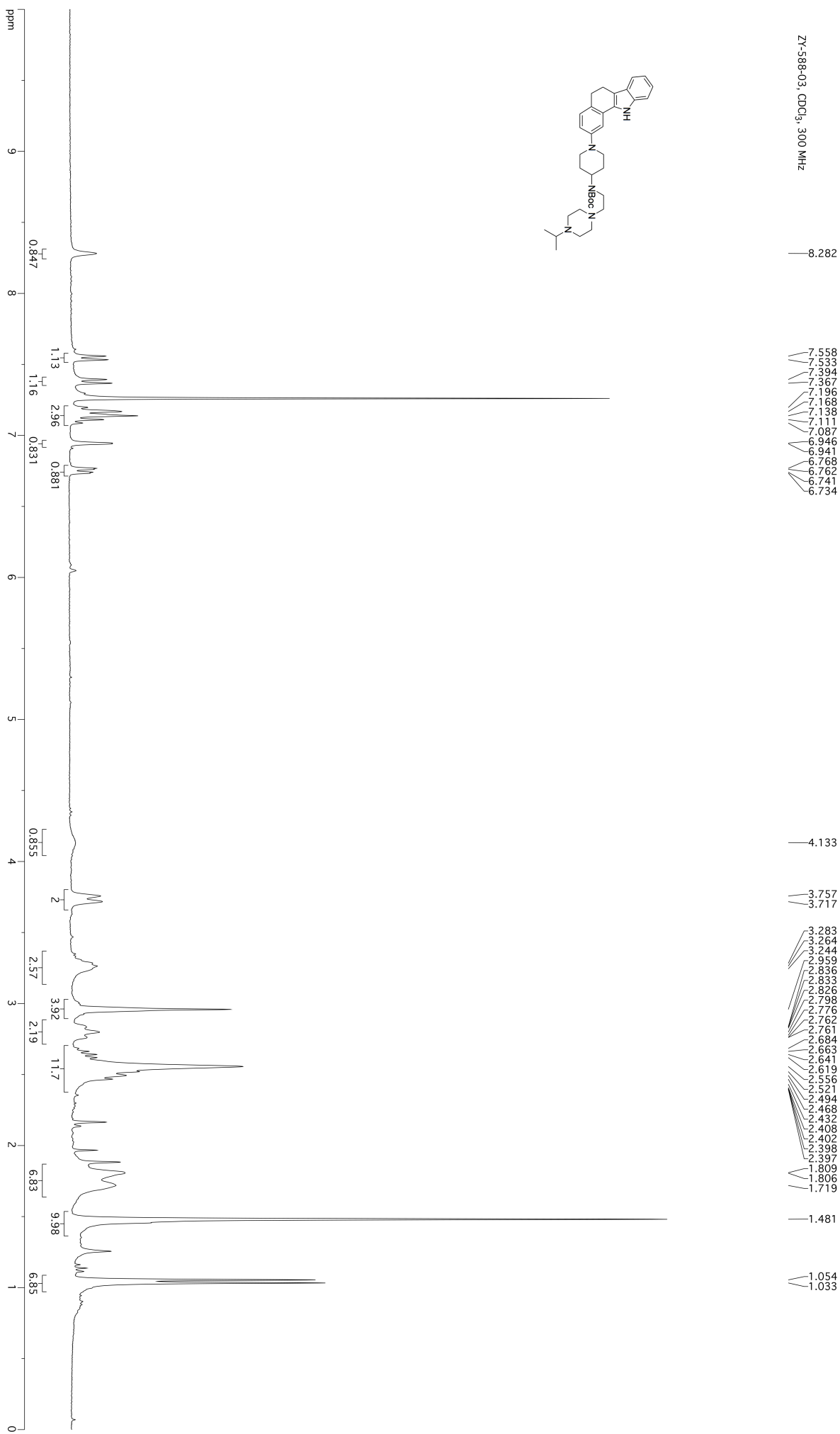
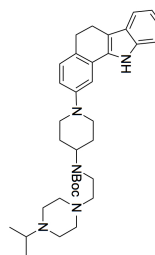
MK-605-94, acetone-d<sub>6</sub>, 400 MHz

MK-625-01, acetone-d6, 400 MHz

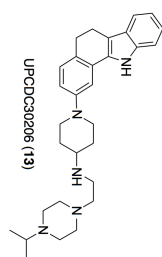




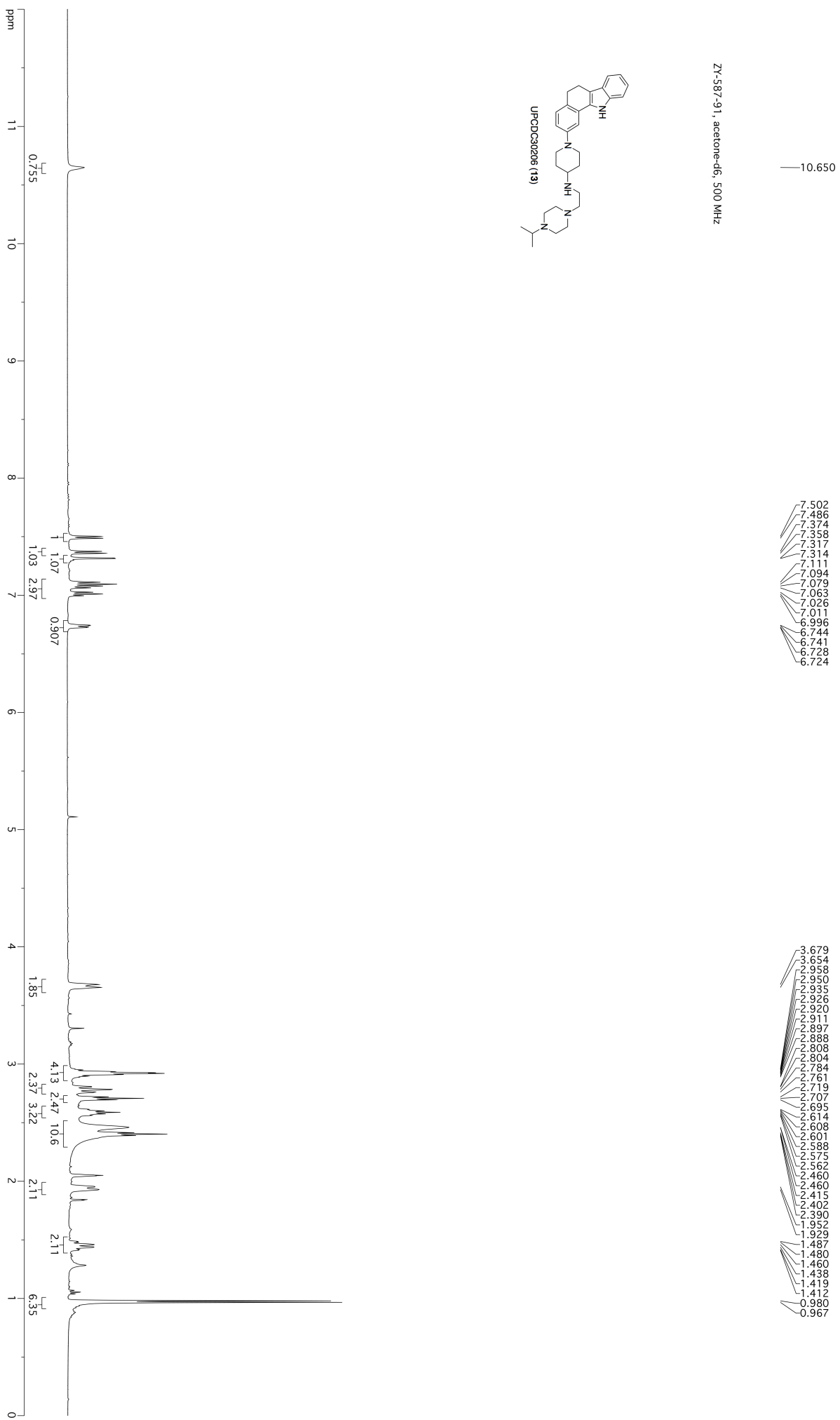
ZY-587-82, CDCl<sub>3</sub>, 300 MHz

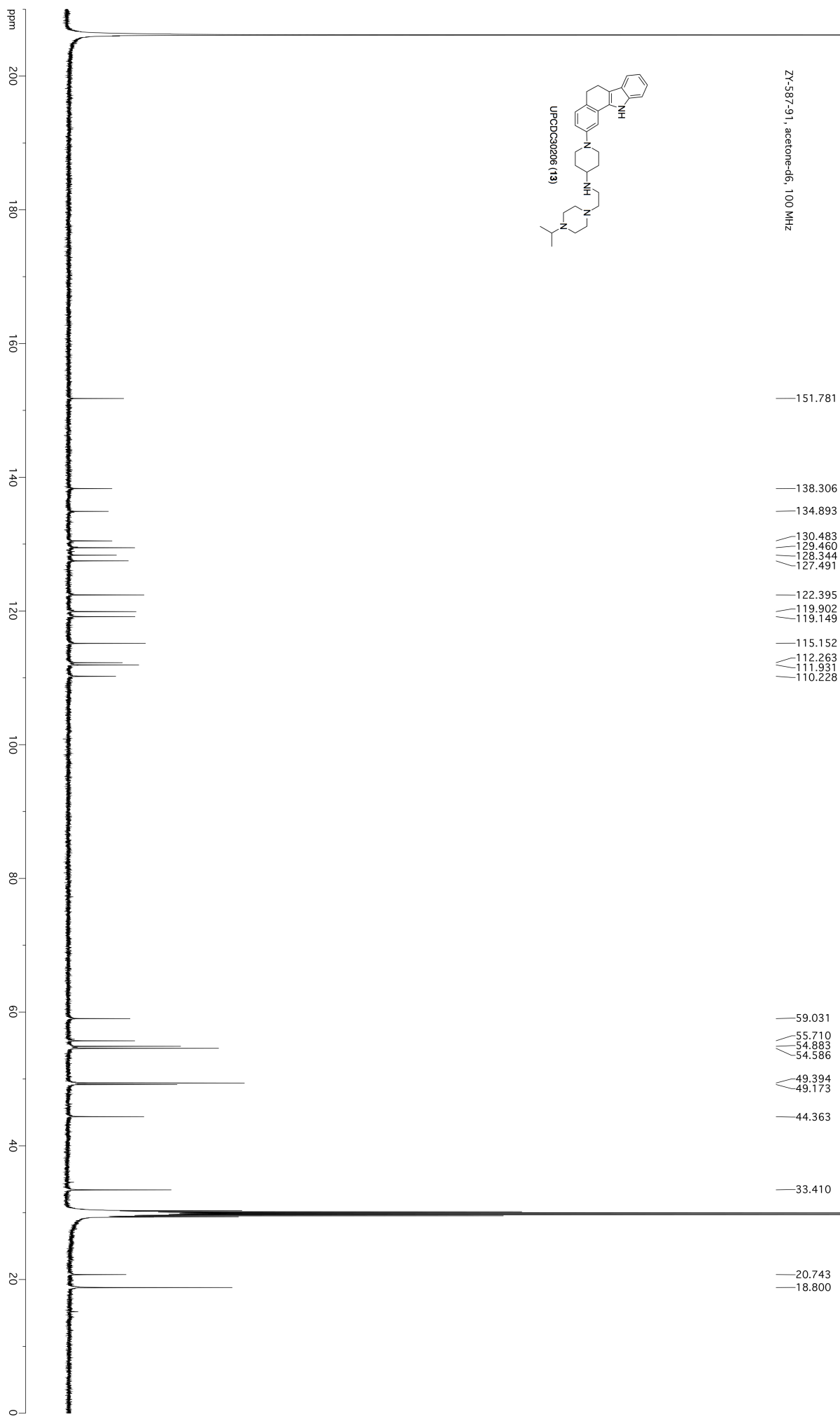
ZY-588-03, CDCl<sub>3</sub>, 300 MHz

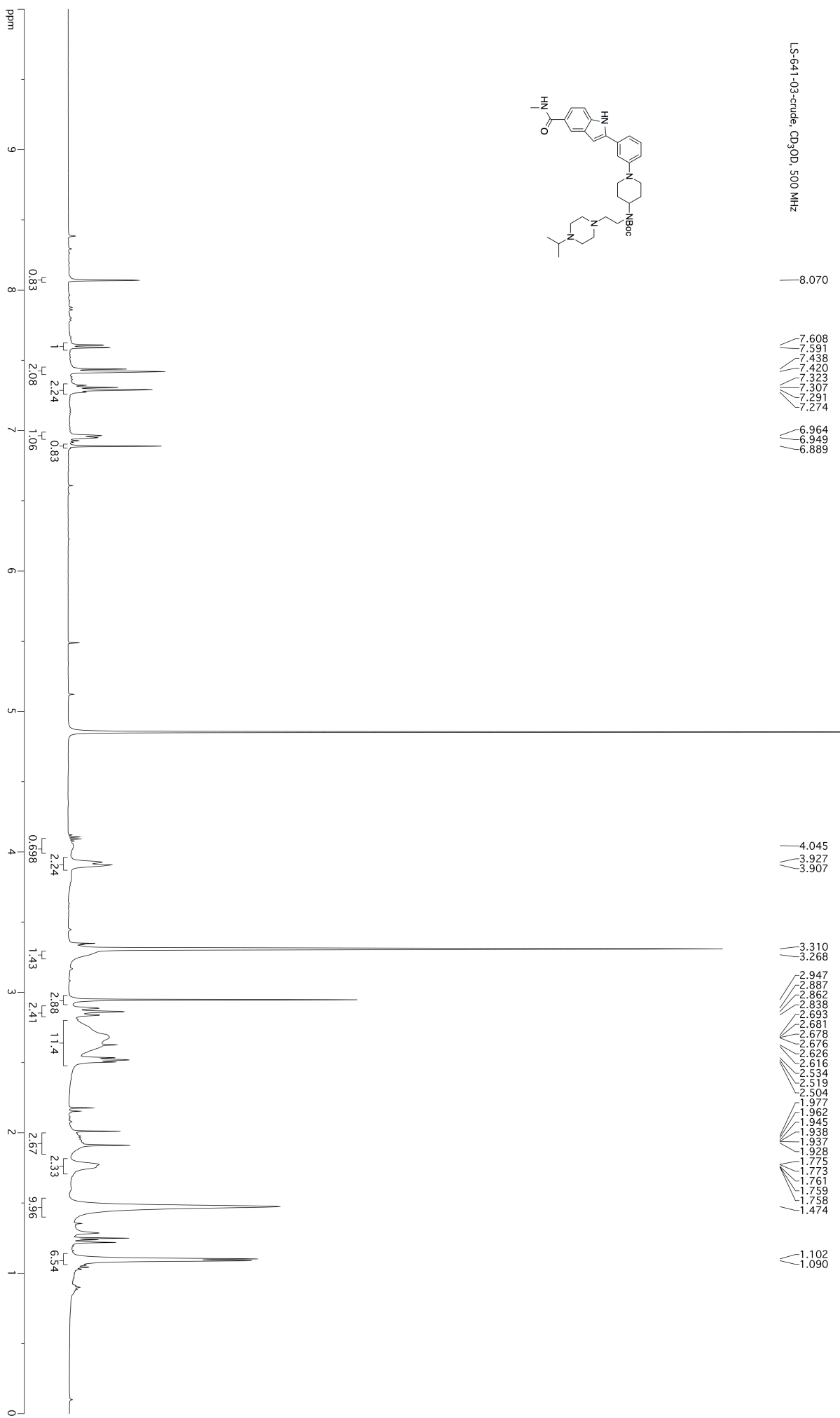


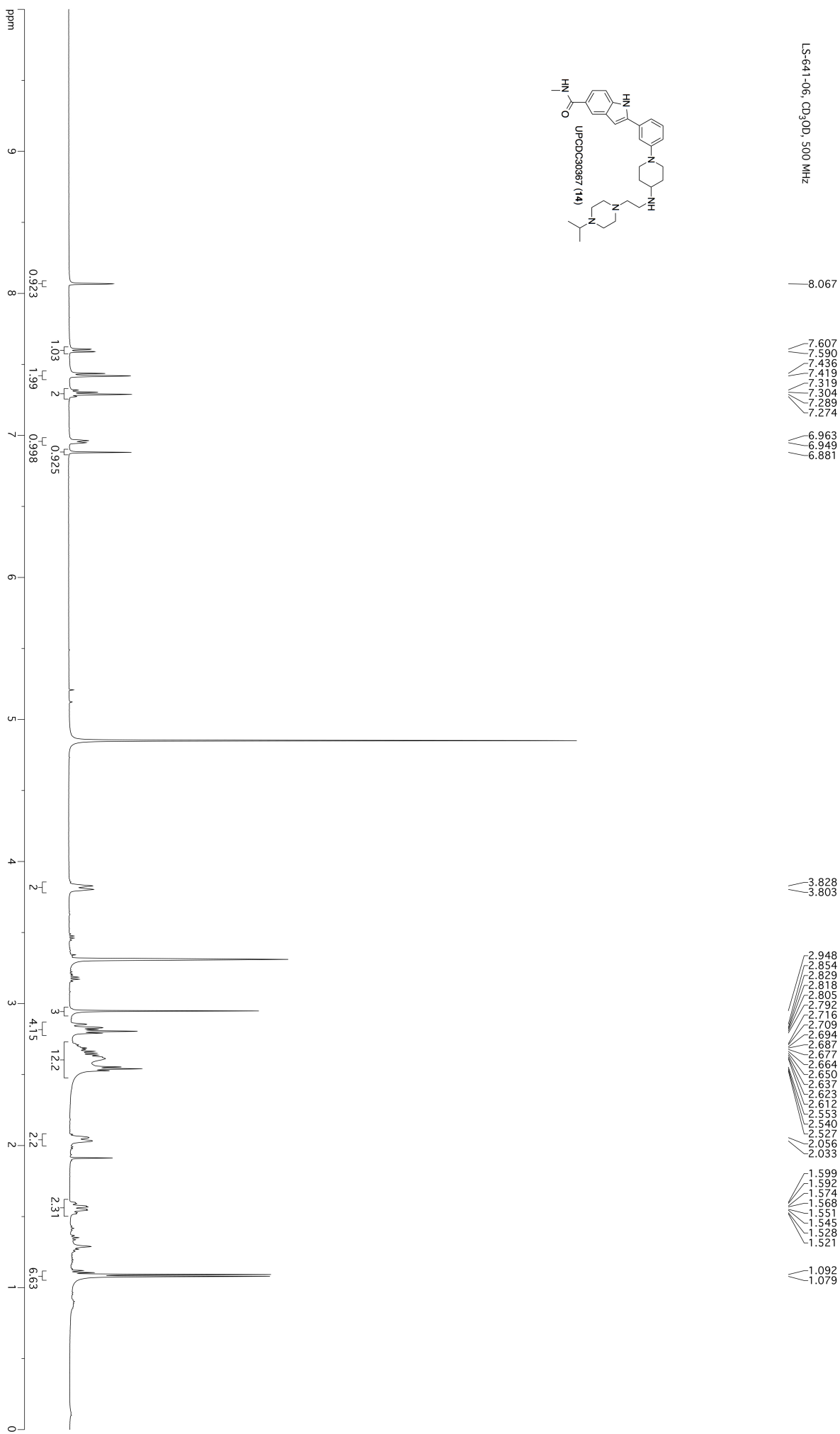
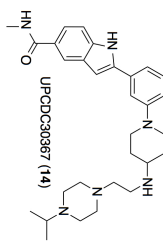


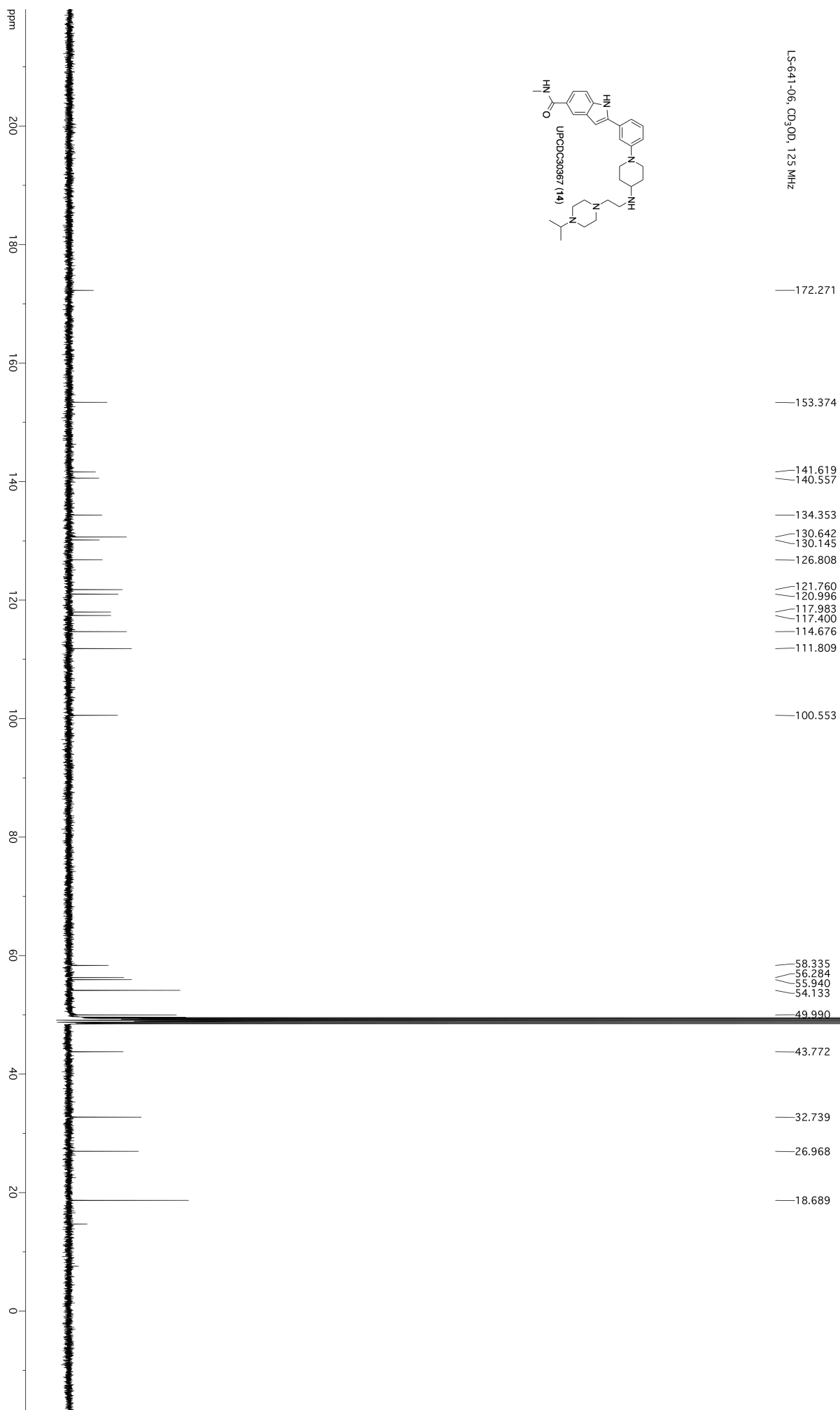
ZY-587-91, acetone-d<sub>6</sub>, 500 MHz

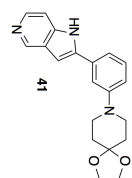




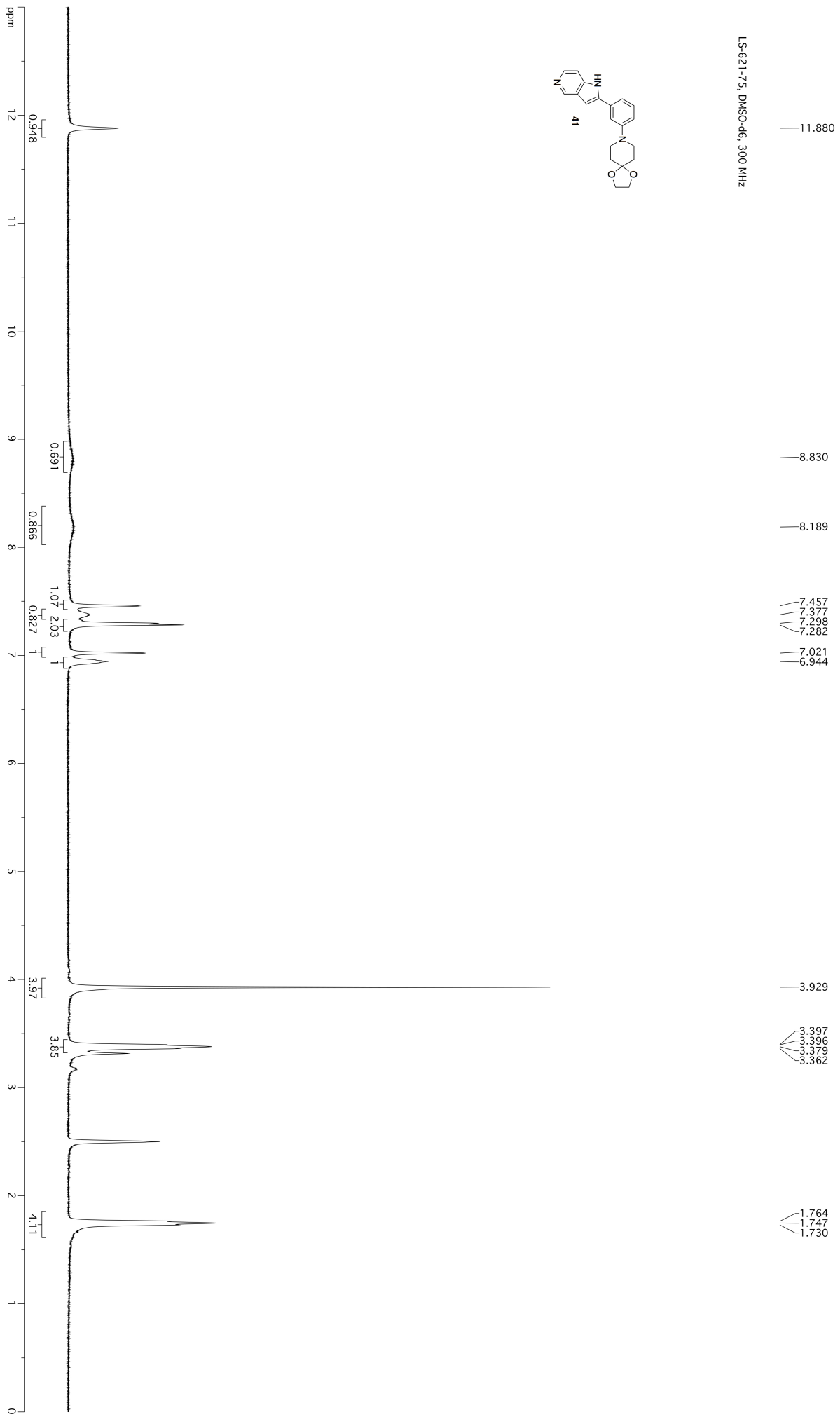


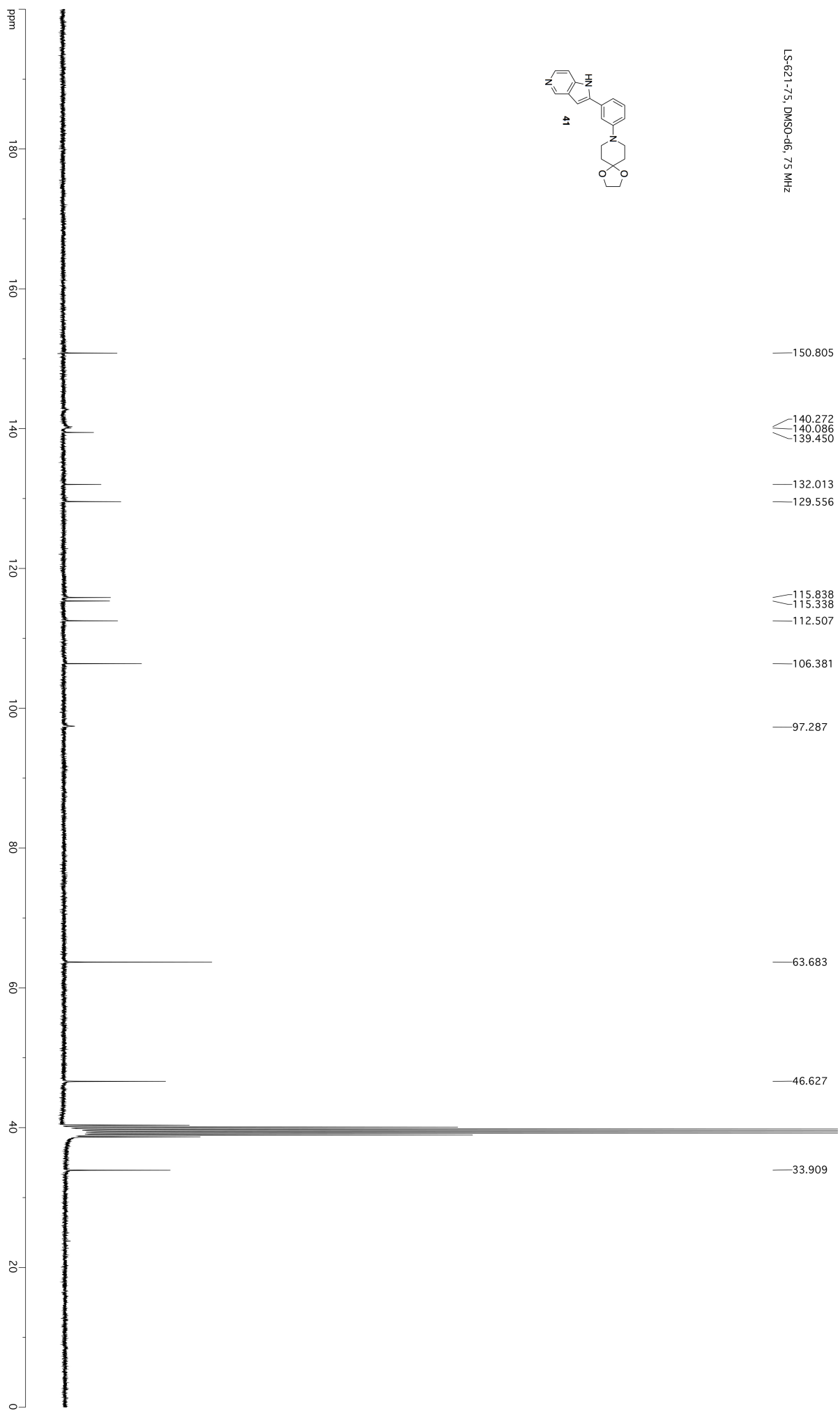
LS-641-06, CD<sub>3</sub>OD, 500 MHz

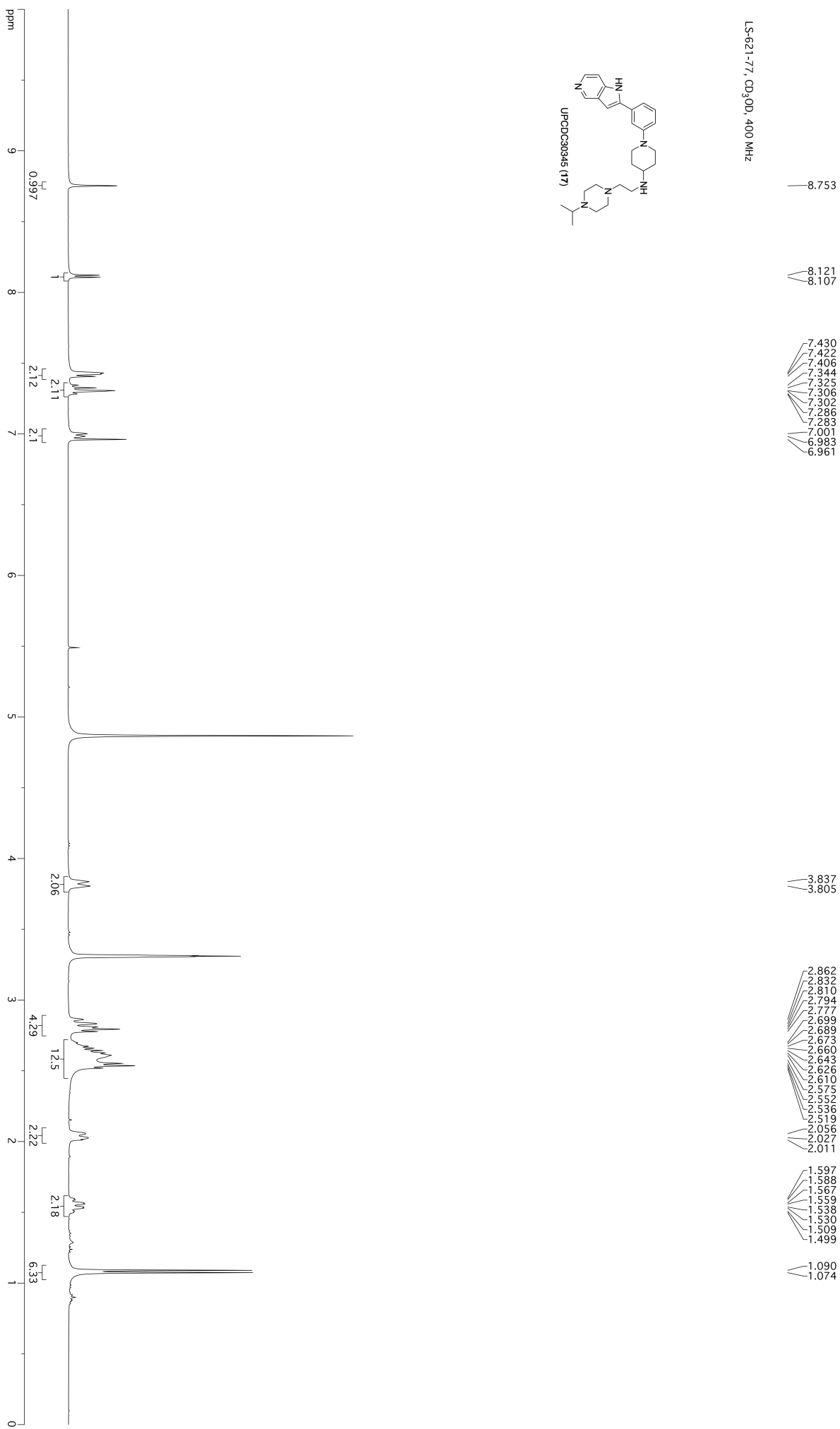
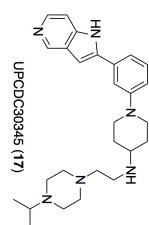




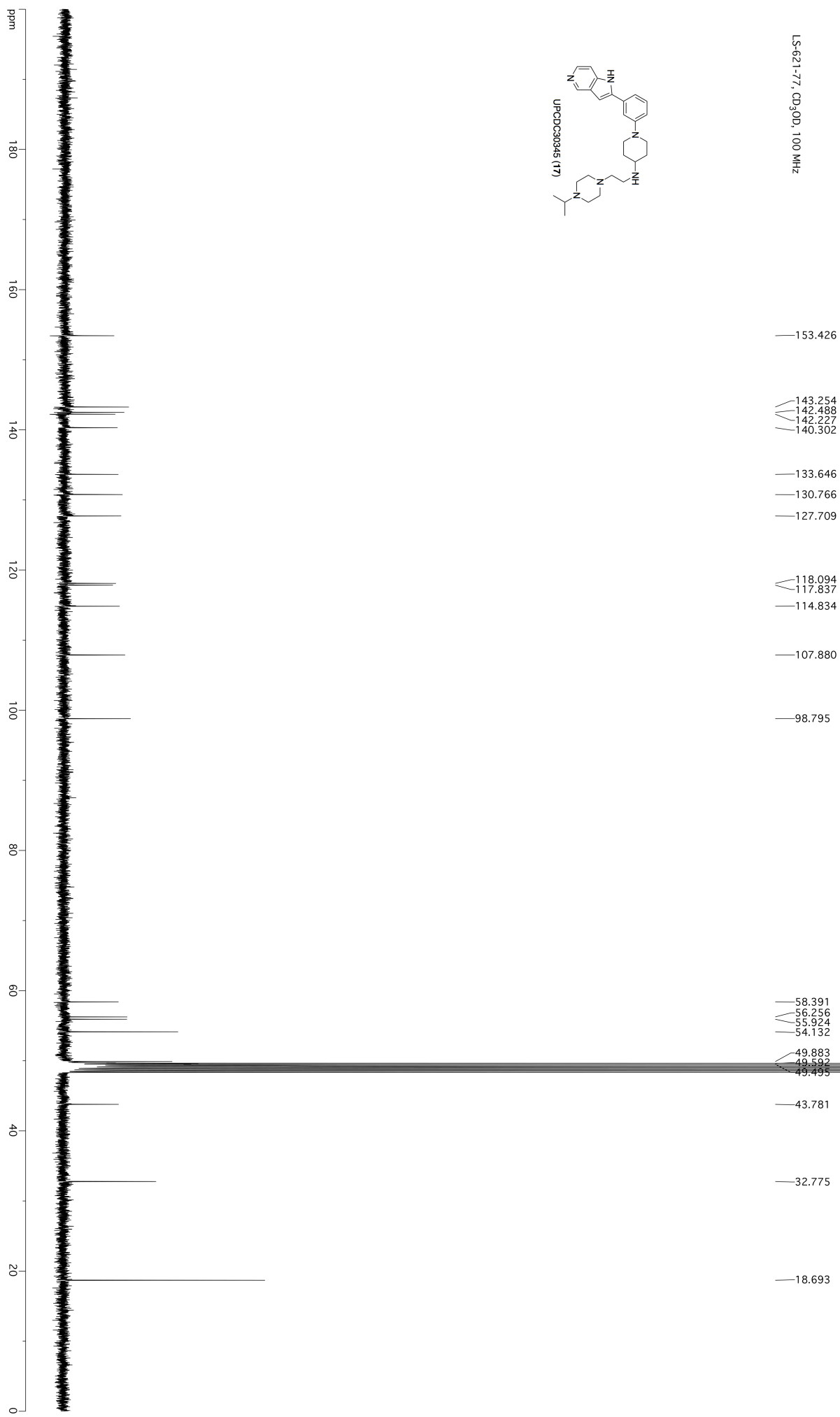
LS-621 -7.5, DMSO-d6, 300 MHz

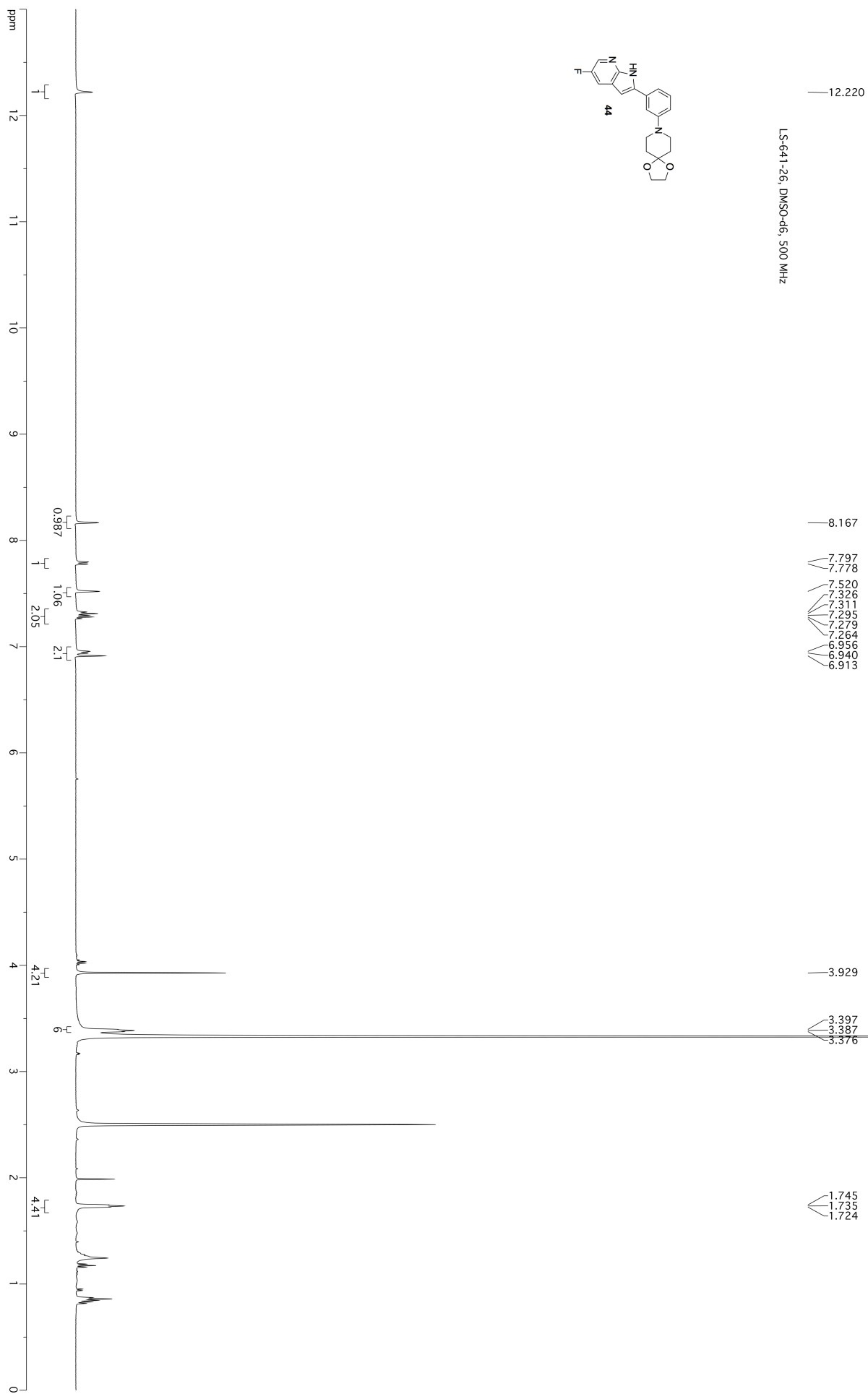


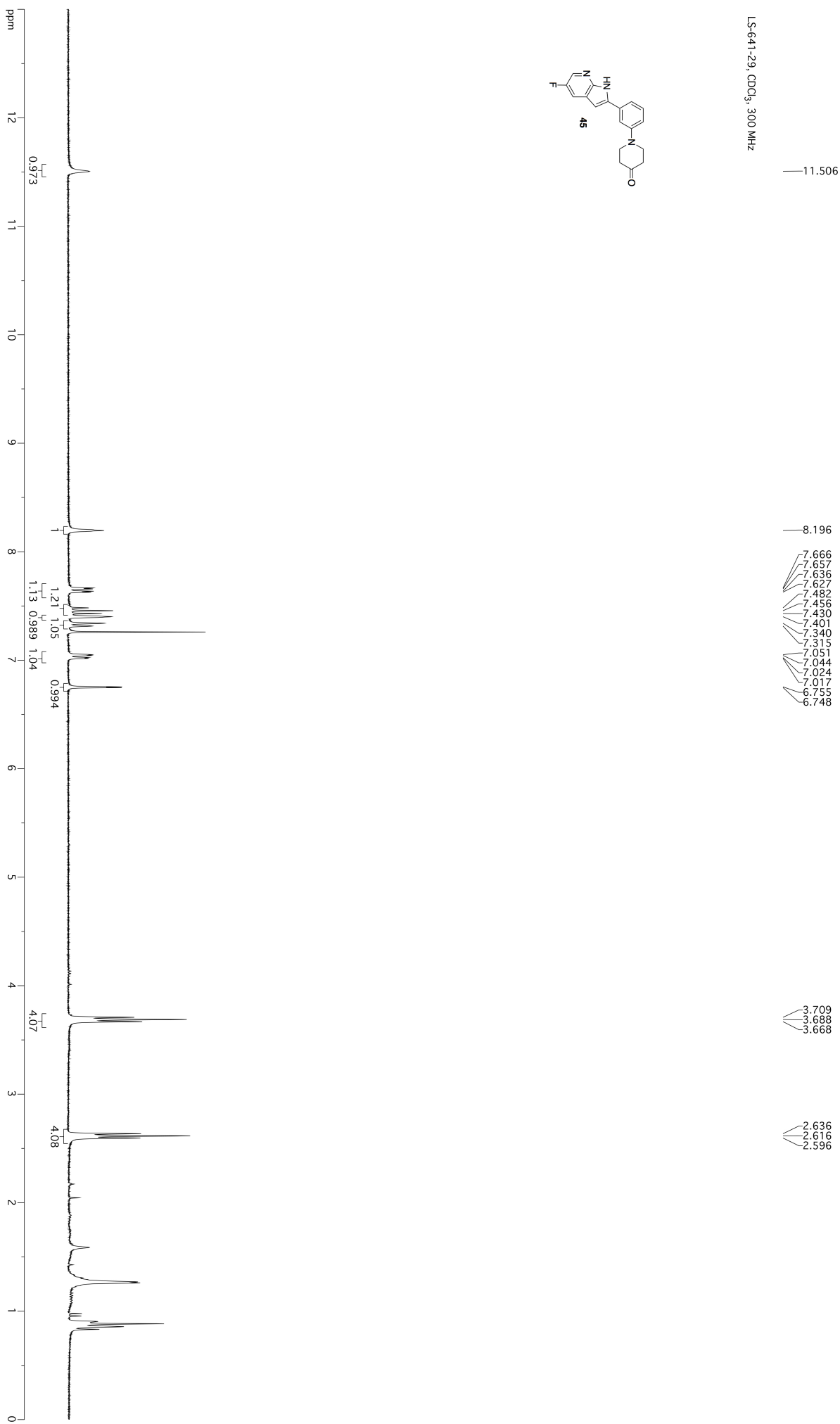
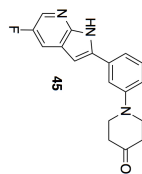


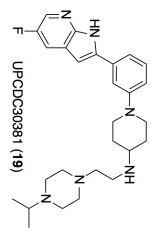
LS-621-77, CD<sub>3</sub>OD, 400 MHz



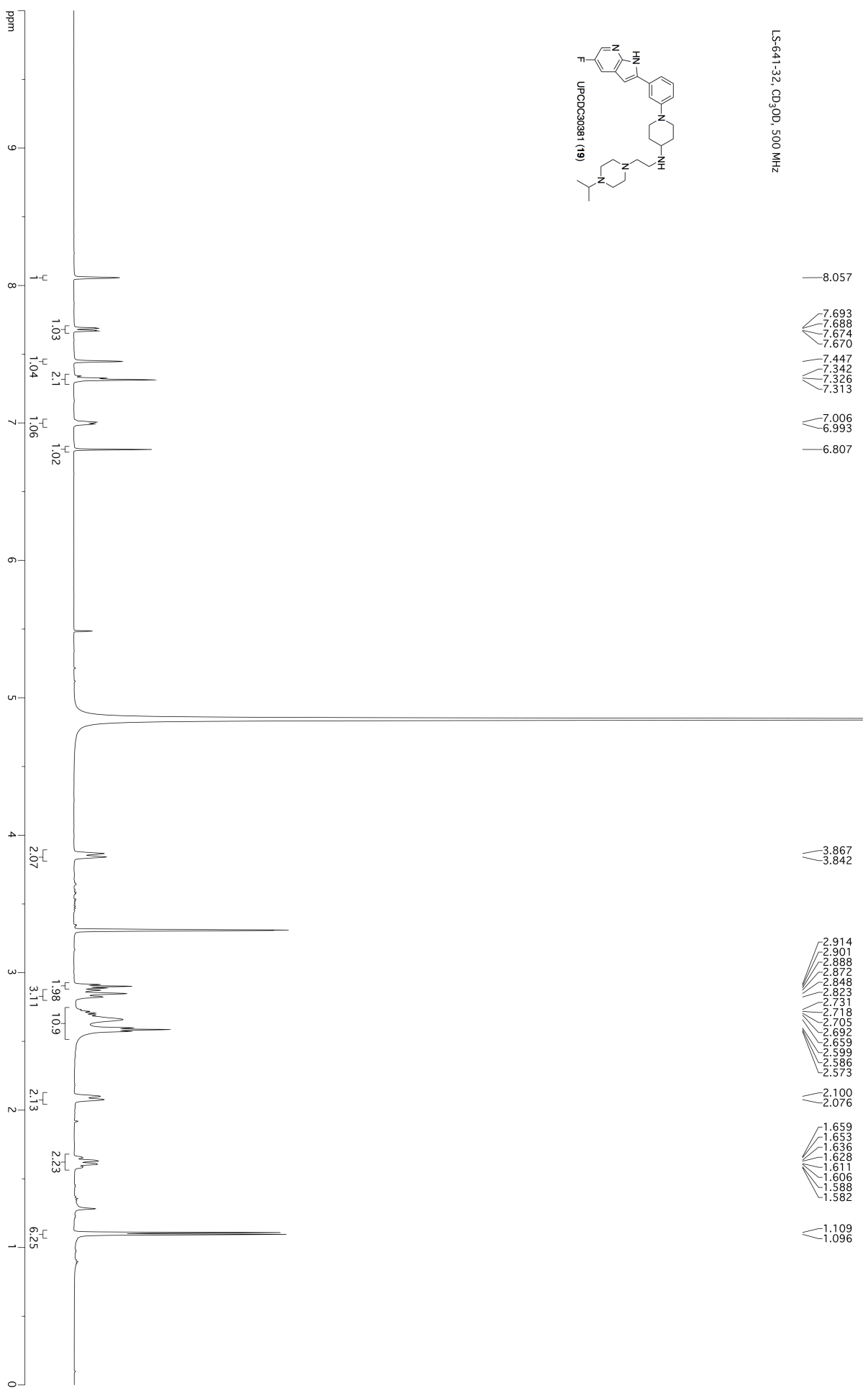


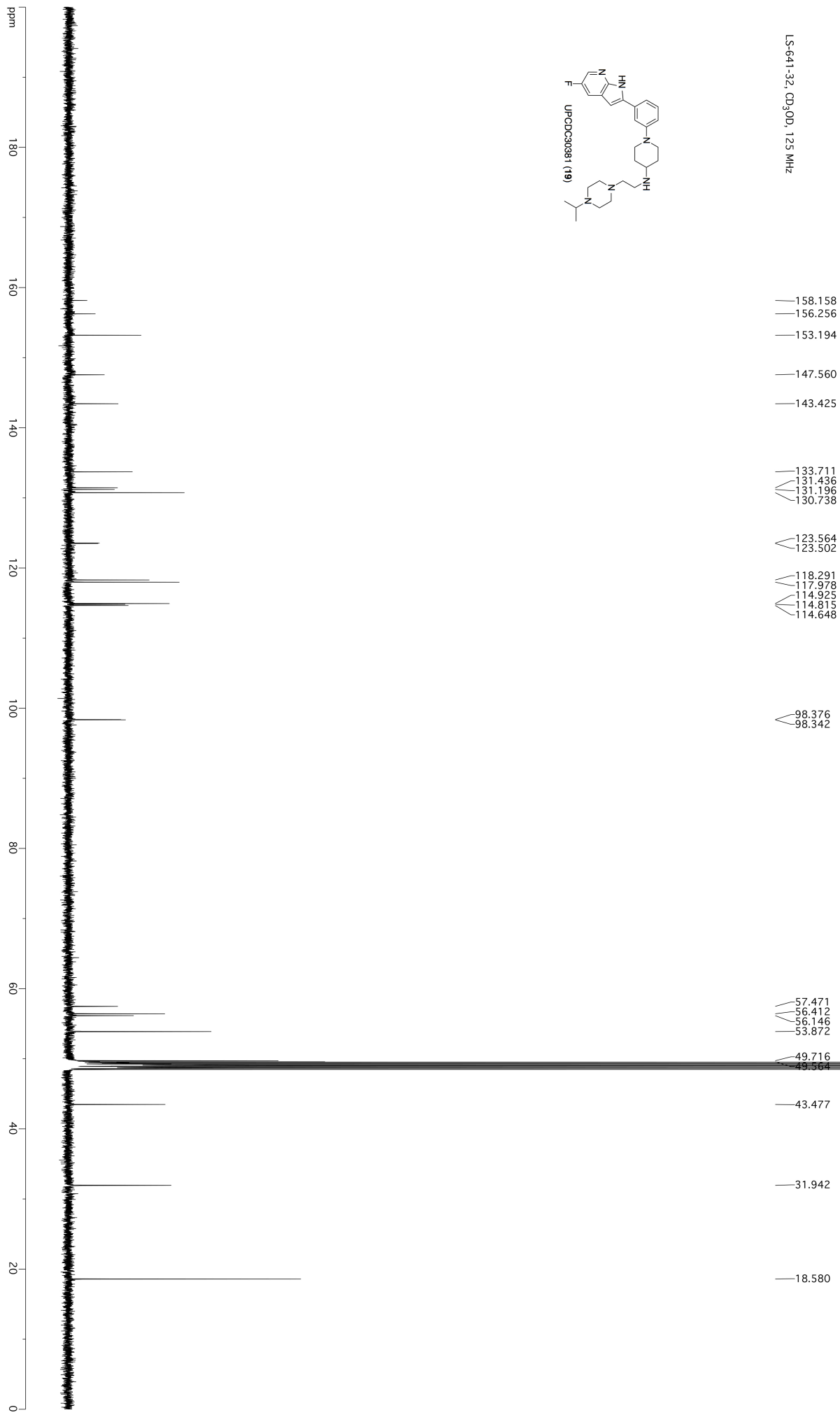


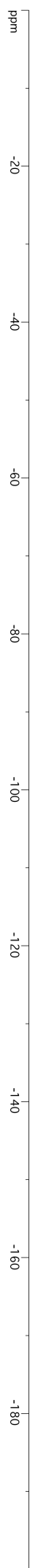
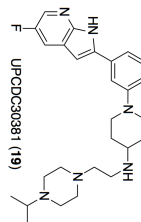
LS-641-29, CDCl<sub>3</sub>, 300 MHz



LS-641-32, CD<sub>3</sub>OD, 500 MHz

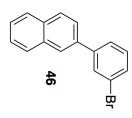




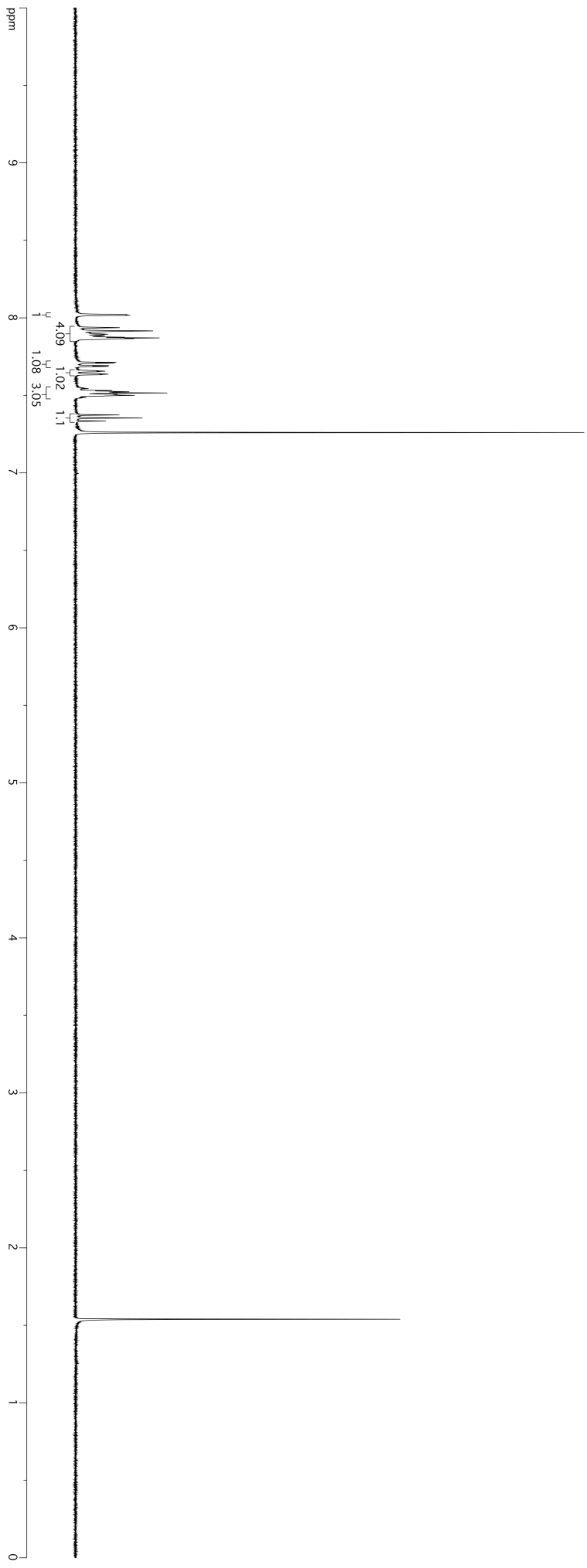
LS-641-32; CD<sub>3</sub>OD, 376 MHz

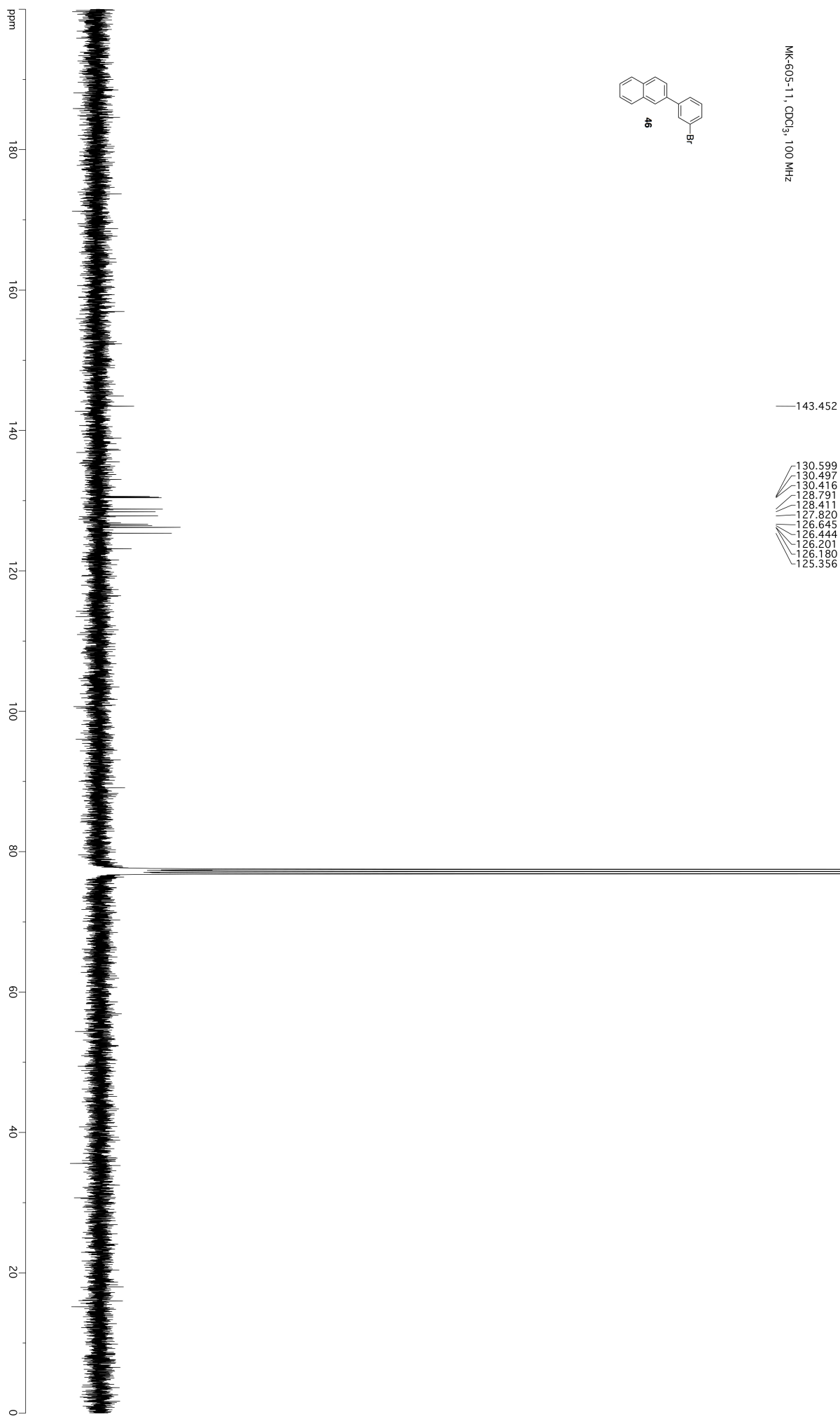
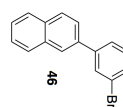
—141.407

MK-605-1, CDCl<sub>3</sub>, 400 MHz



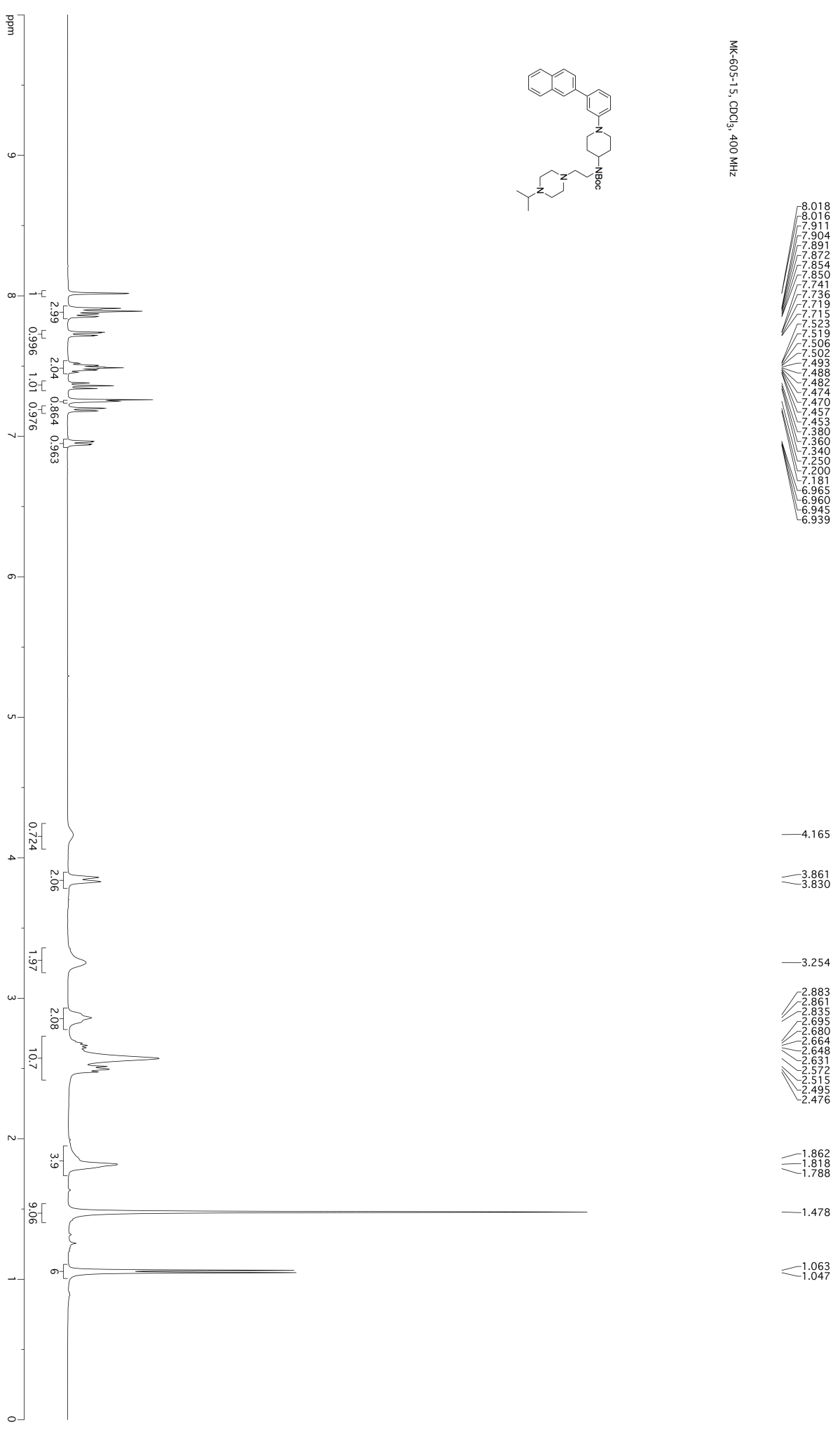
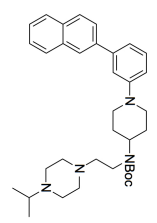
- 8.020
- 8.017
- 8.012
- 7.937
- 7.916
- 7.906
- 7.903
- 7.902
- 7.901
- 7.894
- 7.886
- 7.884
- 7.884
- 7.883
- 7.875
- 7.871
- 7.866
- 7.863
- 7.713
- 7.708
- 7.692
- 7.687
- 7.659
- 7.656
- 7.653
- 7.653
- 7.640
- 7.637
- 7.634
- 7.548
- 7.543
- 7.531
- 7.523
- 7.519
- 7.515
- 7.507
- 7.505
- 7.500
- 7.495
- 7.488
- 7.487
- 7.483
- 7.483
- 7.374
- 7.354
- 7.335

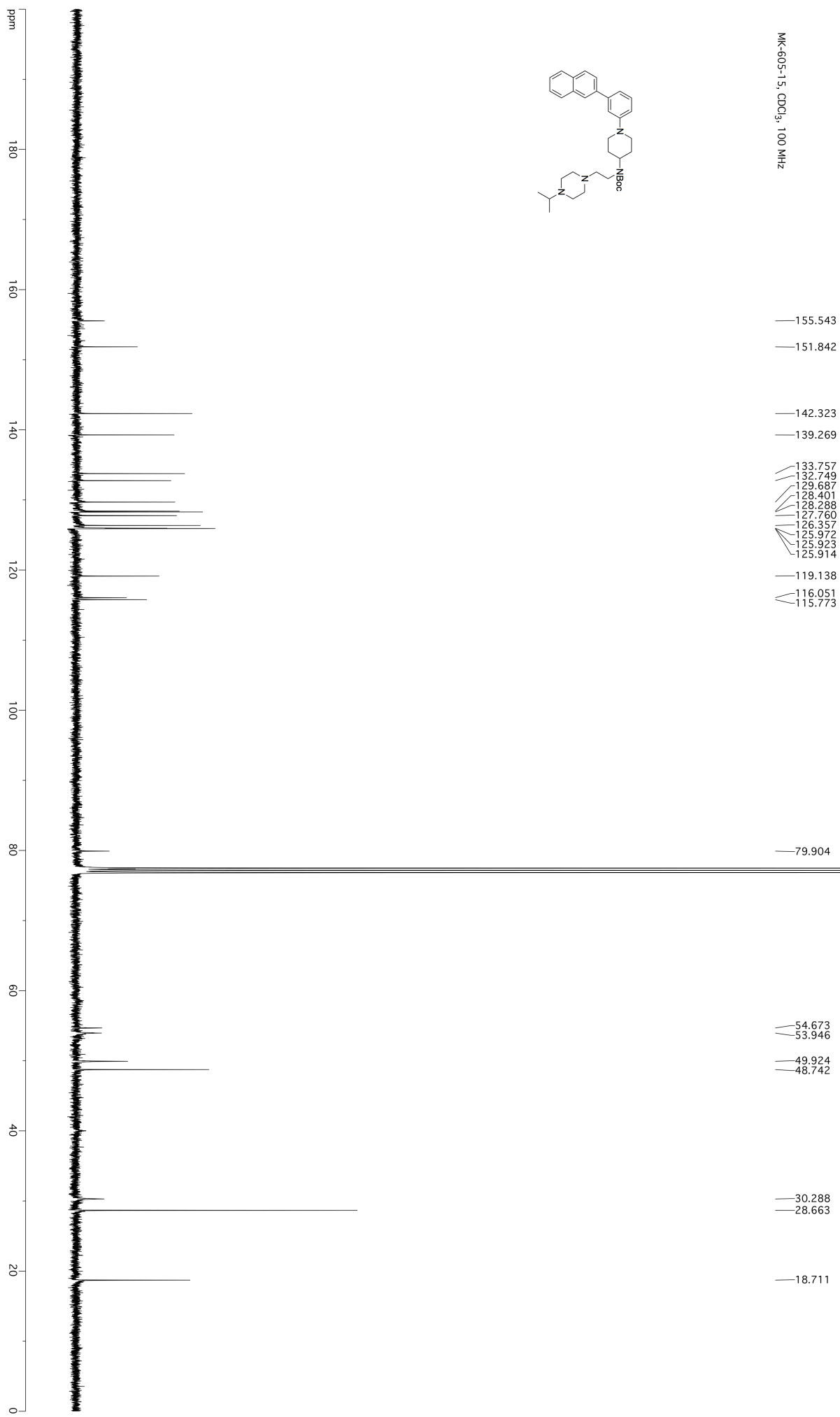


MK-605-11, CDCl<sub>3</sub>, 100 MHz

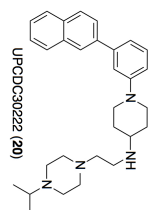


MK-605-15, CDCl<sub>3</sub>, 400 MHz





MK-605-19, acetone-d6, 500 MHz



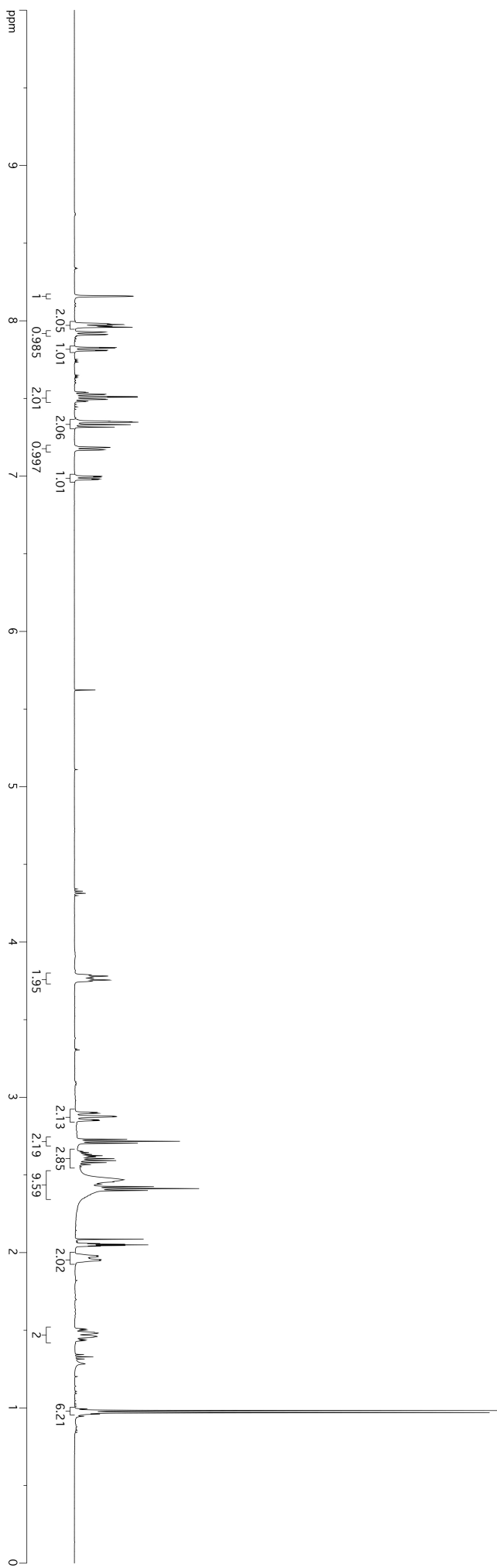
8.159  
8.157  
7.980  
7.976  
7.966  
7.959  
7.928  
7.912  
7.828  
7.824  
7.811  
7.807  
7.541  
7.538  
7.527  
7.525  
7.512  
7.508  
7.496  
7.493  
7.482  
7.480  
7.354  
7.350  
7.347  
7.332  
7.316  
7.185  
7.184  
7.170  
7.169  
6.999  
6.994  
6.982  
6.978

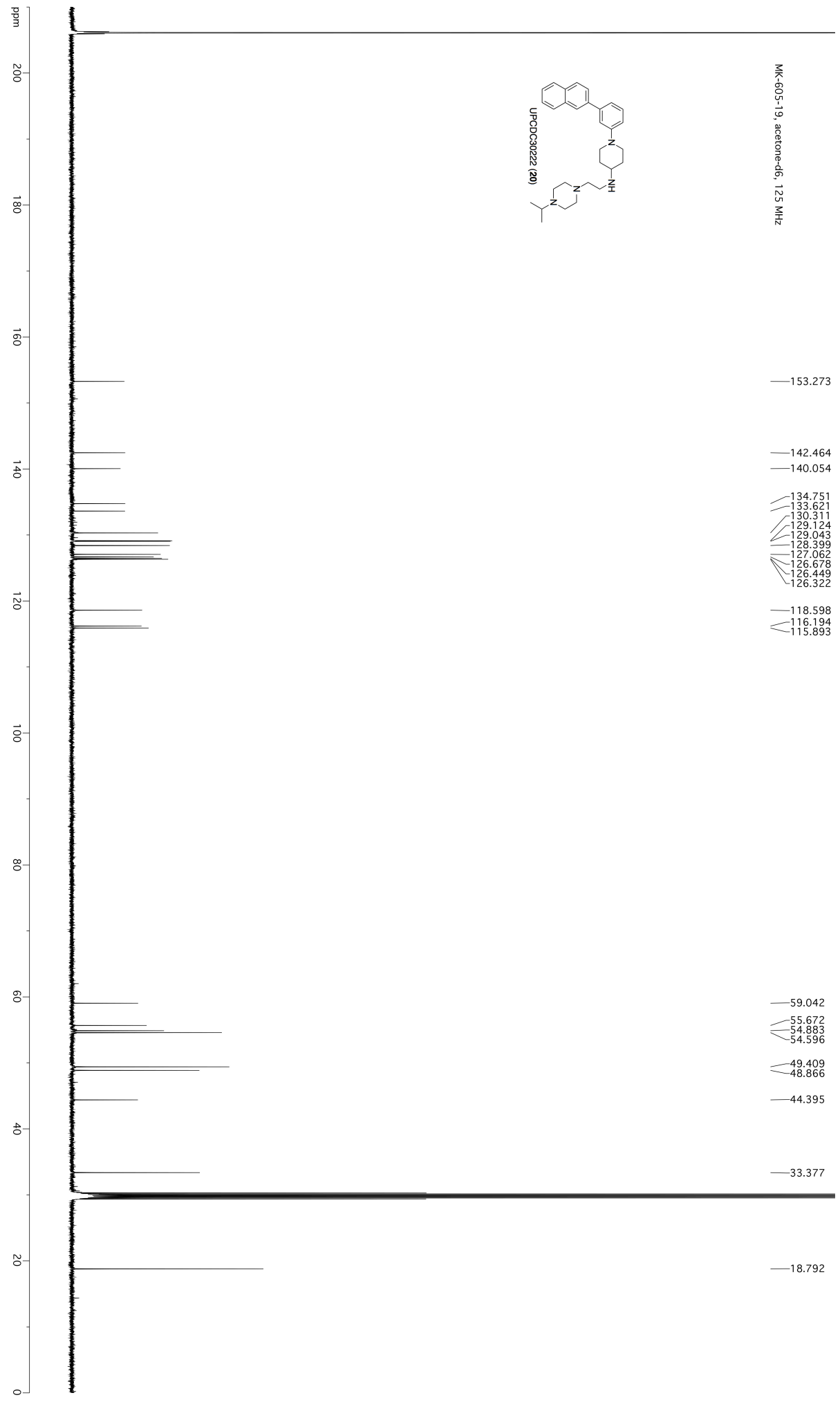
3.787  
3.781  
3.773  
3.762  
3.755  
3.748

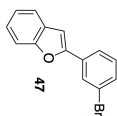
2.903  
2.897  
2.878  
2.875  
2.855  
2.850  
2.729  
2.717  
2.705  
2.652  
2.644  
2.636  
2.632  
2.624  
2.618  
2.613  
2.605  
2.593  
2.579  
2.566  
2.553

1.980  
1.974  
1.964  
1.955  
1.946  
1.509  
1.501  
1.486  
1.481  
1.459  
1.456  
1.441  
1.433

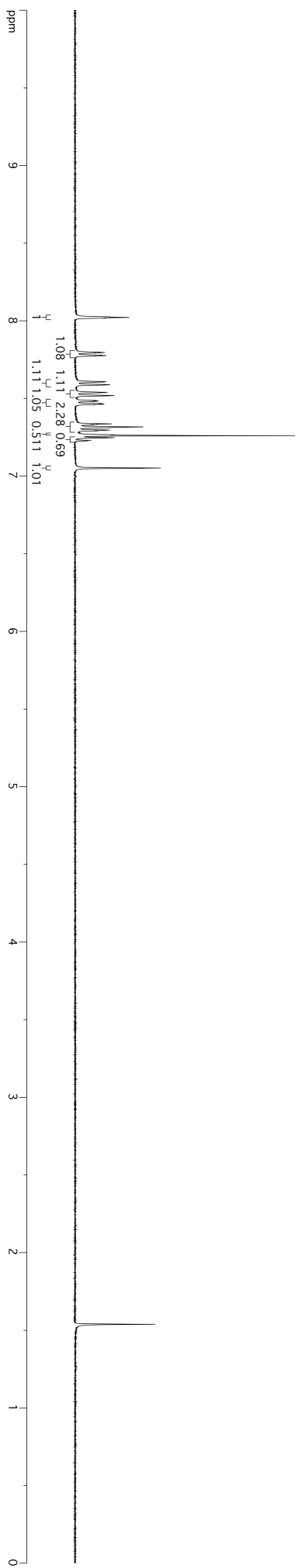
0.983  
0.970

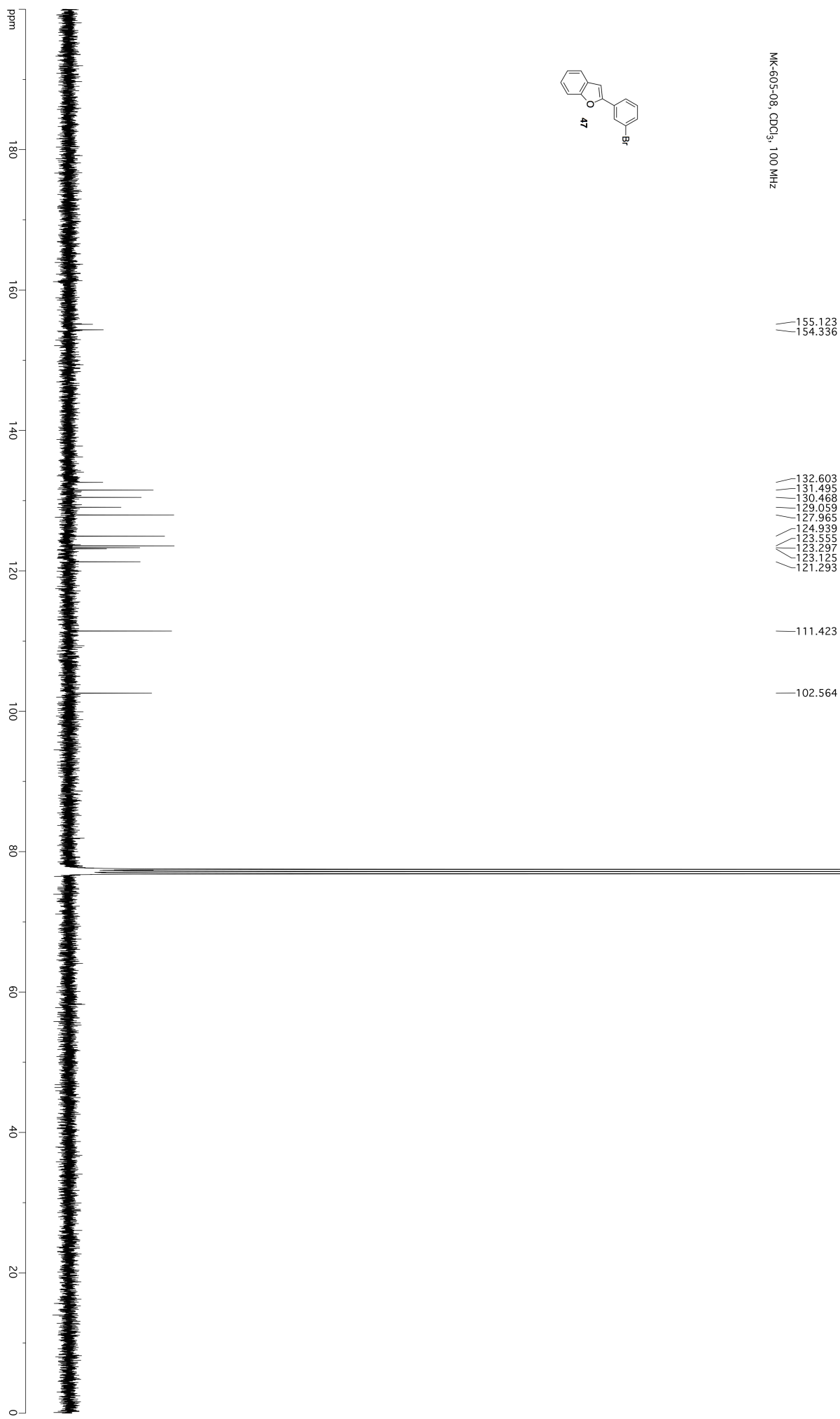




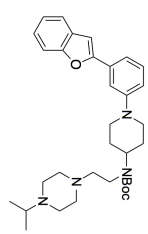
NK-605-08, CDCl<sub>3</sub>, 400 MHz

8.025  
8.022  
8.017  
7.998  
7.996  
7.995  
7.992  
7.779  
7.776  
7.775  
7.773  
7.609  
7.608  
7.606  
7.604  
7.590  
7.589  
7.587  
7.585  
7.583  
7.538  
7.519  
7.517  
7.487  
7.485  
7.482  
7.480  
7.467  
7.465  
7.462  
7.460  
7.337  
7.333  
7.330  
7.317  
7.312  
7.309  
7.296  
7.295  
7.291  
7.267  
7.247  
7.246  
7.229  
7.227  
7.052  
7.050

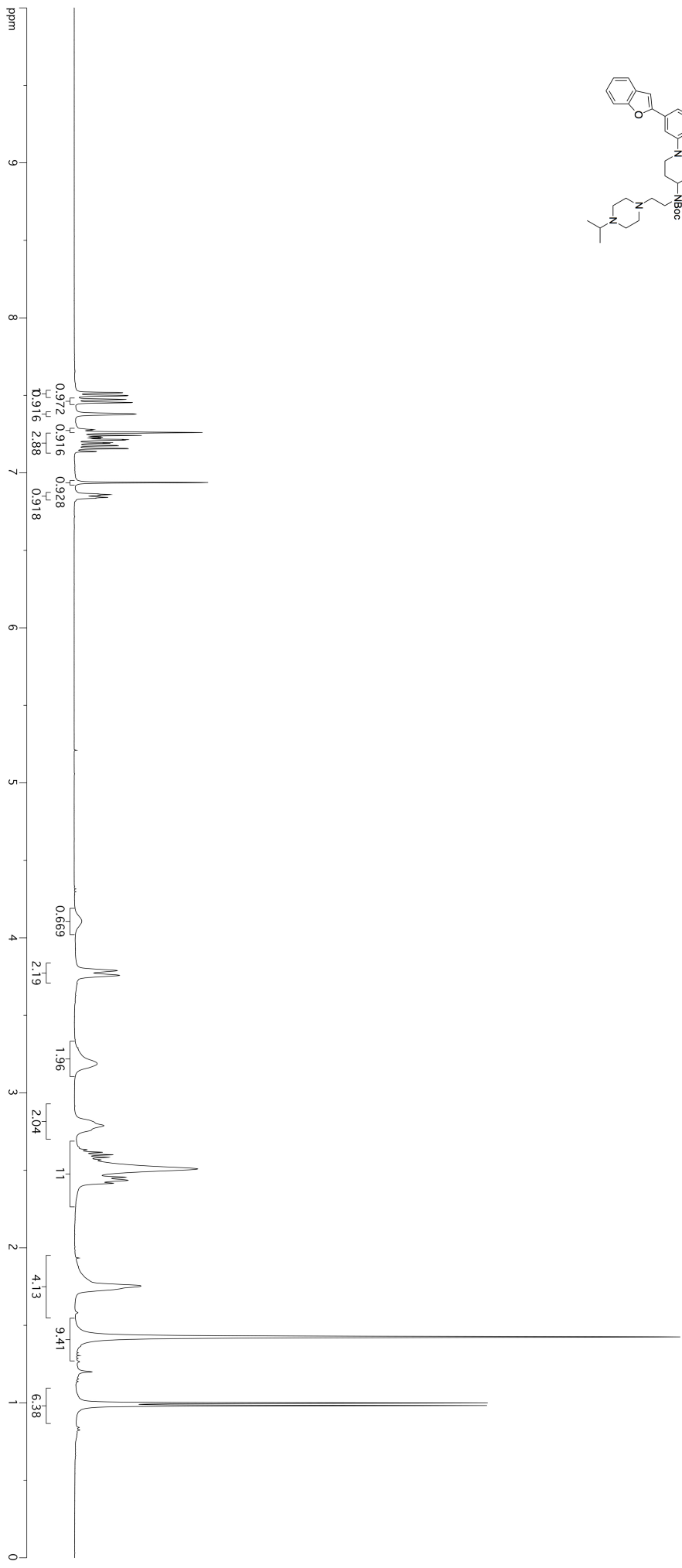




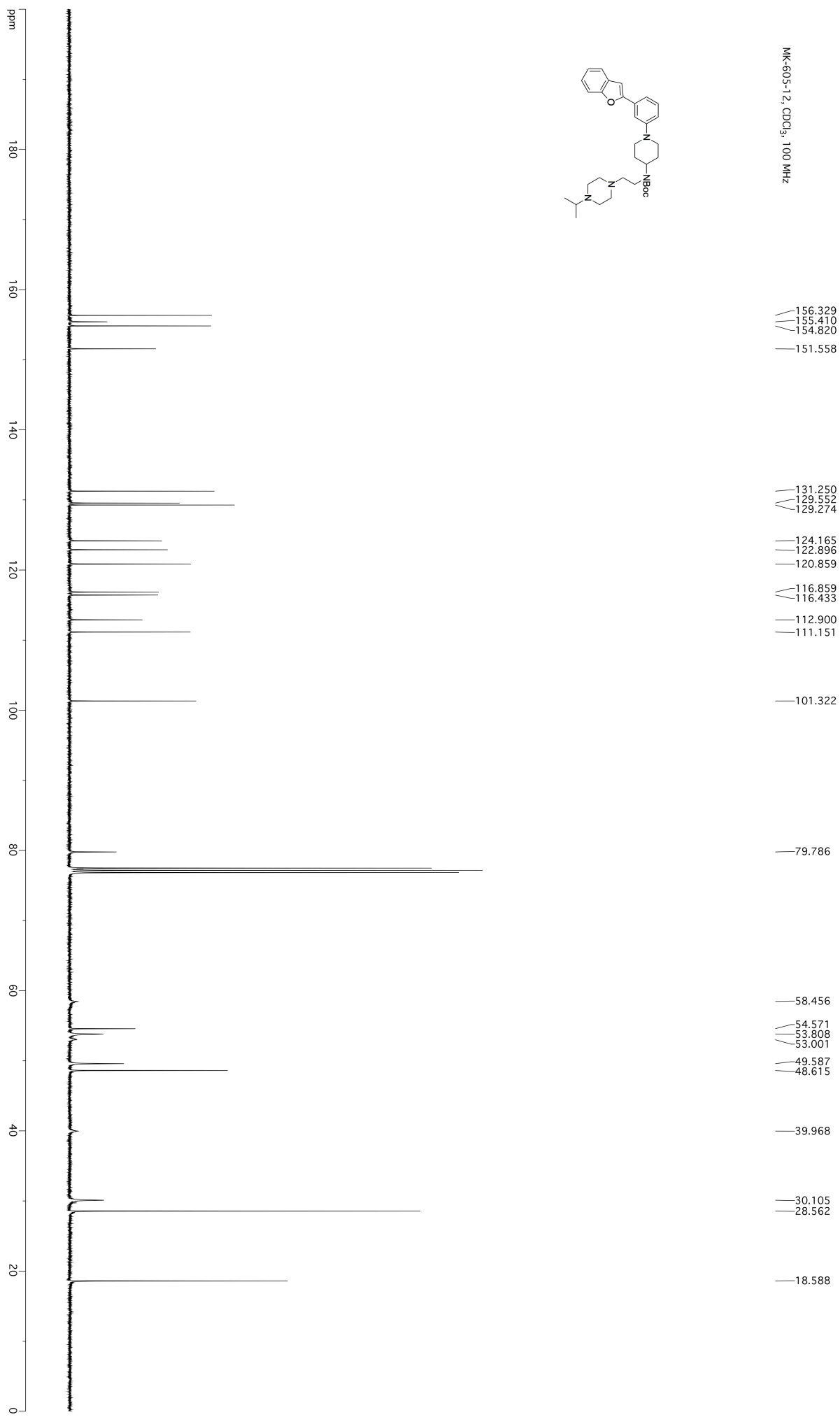
MK-605-12, CDCl<sub>3</sub>, 400 MHz



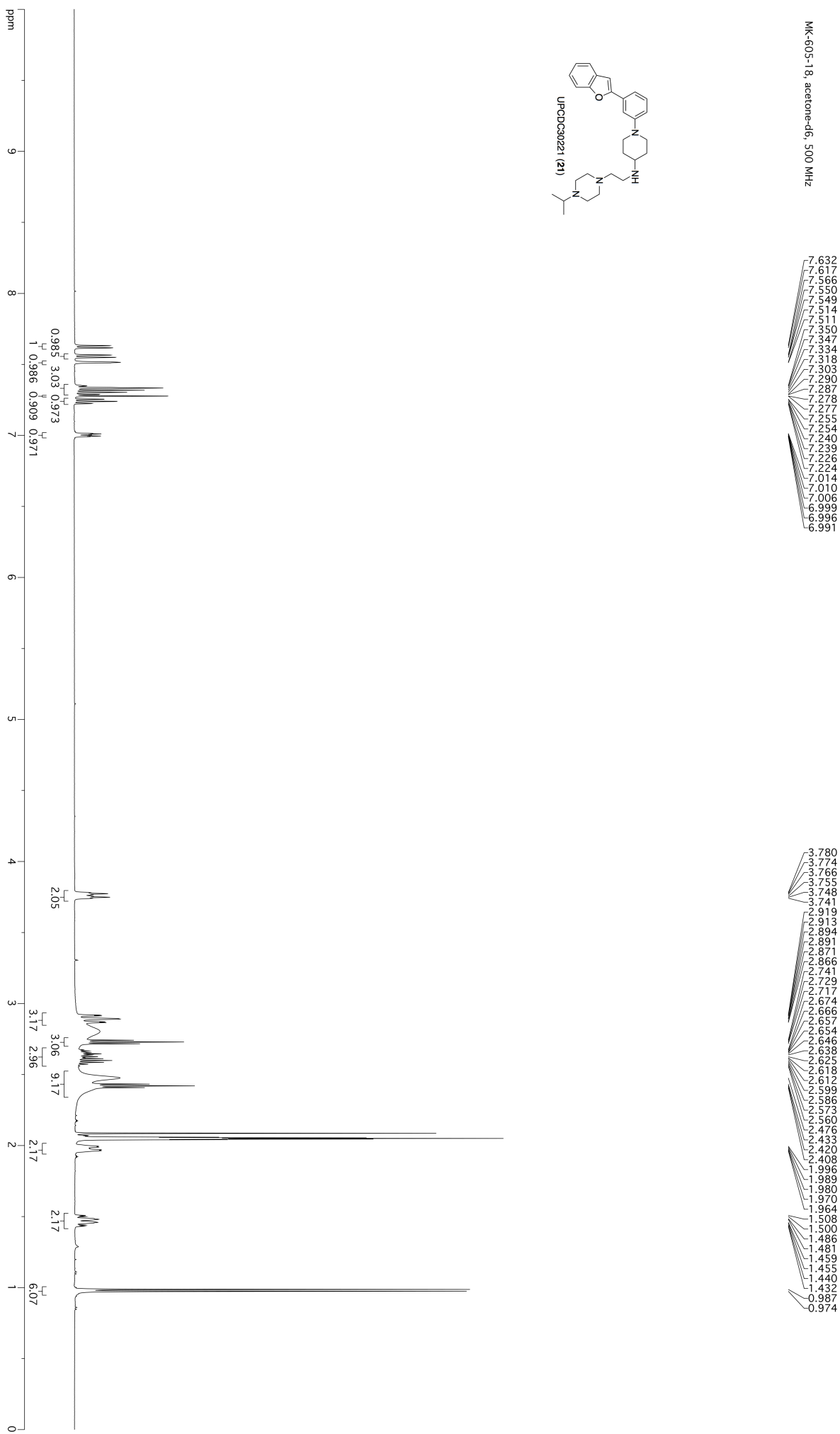
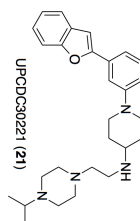
- 7.518
- 7.516
- 7.499
- 7.497
- 7.474
- 7.454
- 7.453
- 7.379
- 7.279
- 7.276
- 7.241
- 7.232
- 7.229
- 7.222
- 7.216
- 7.214
- 7.211
- 7.194
- 7.191
- 7.177
- 7.174
- 7.158
- 7.156
- 7.139
- 7.137
- 6.938
- 6.936
- 6.864
- 6.860
- 6.854
- 6.846
- 6.842
- 6.836

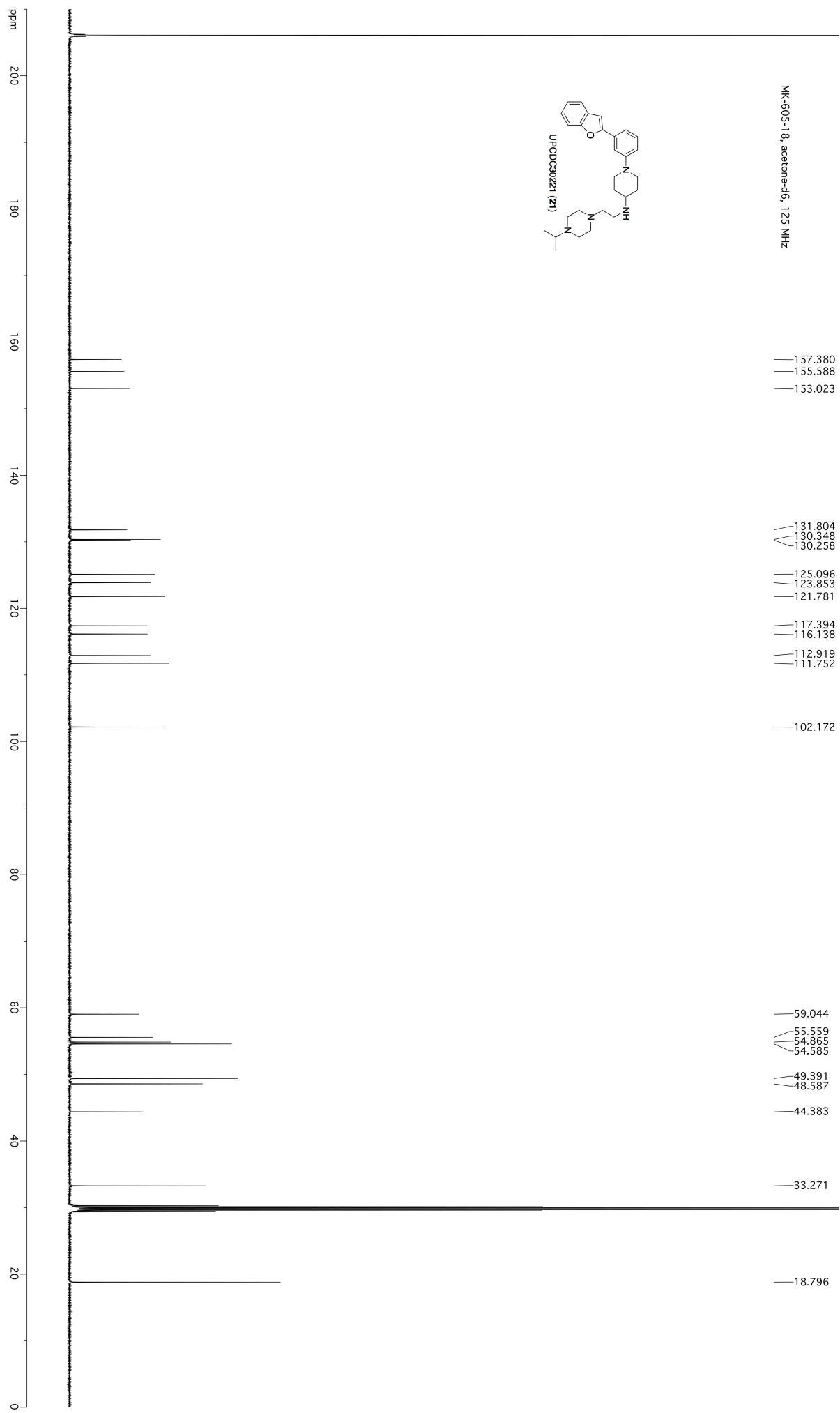


- 4.107
- 3.788
- 3.758
- 3.188
- 2.788
- 2.633
- 2.627
- 2.617
- 2.600
- 2.584
- 2.567
- 2.509
- 2.456
- 2.436
- 2.416
- 1.754
- 1.425
- 0.999
- 0.982

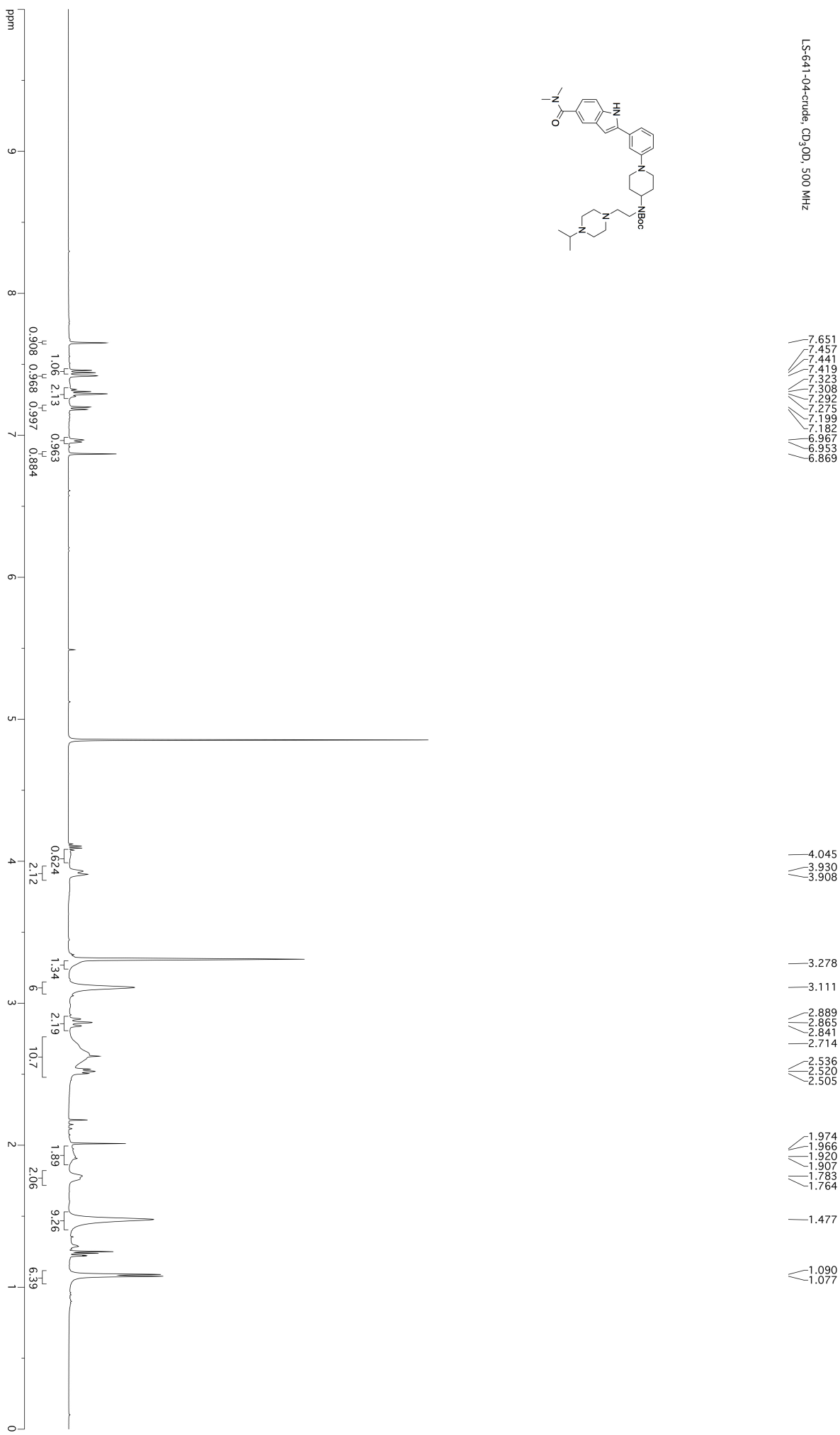
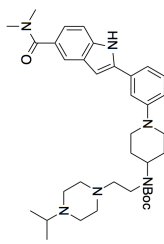


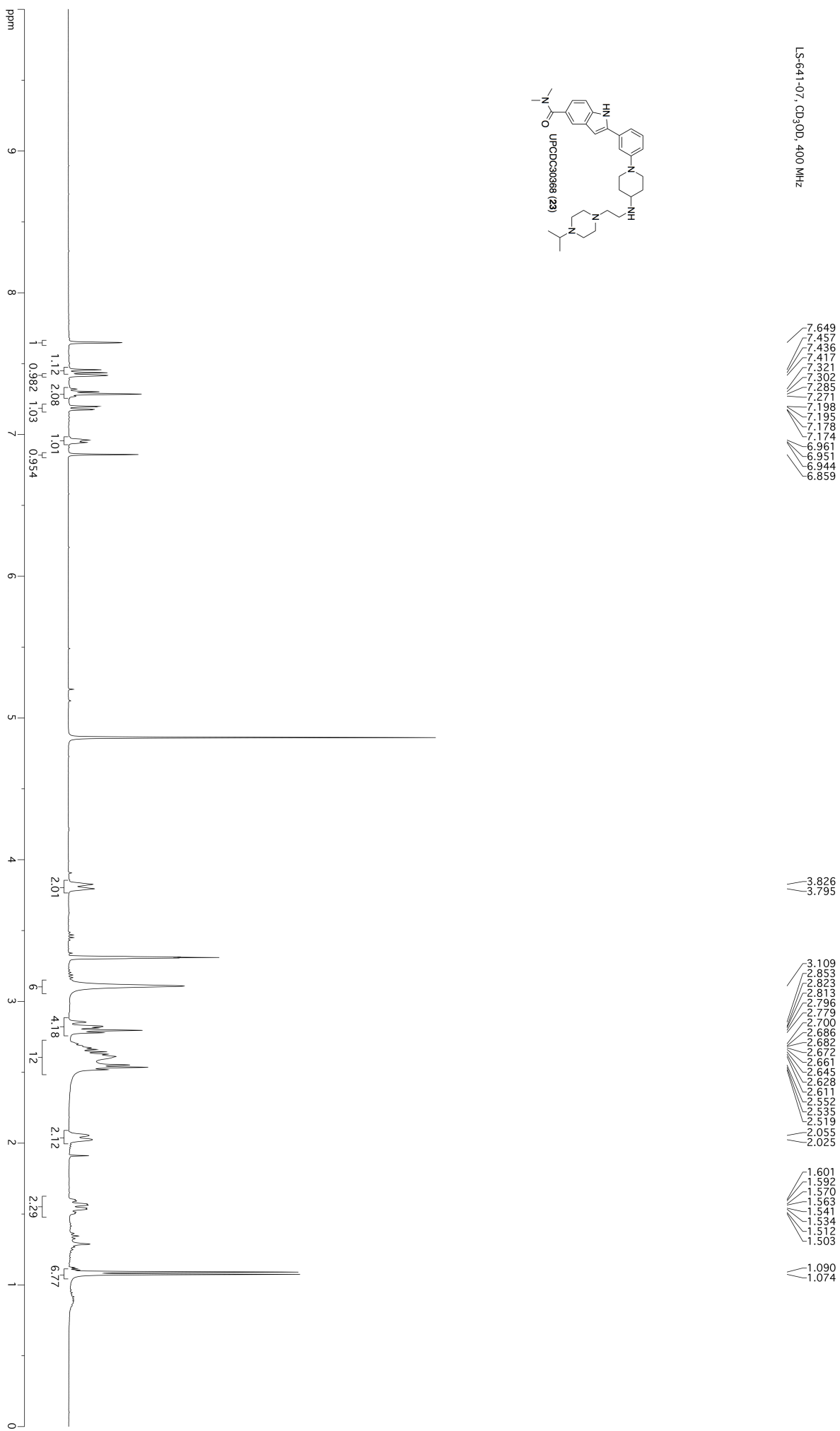
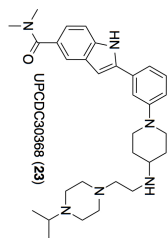


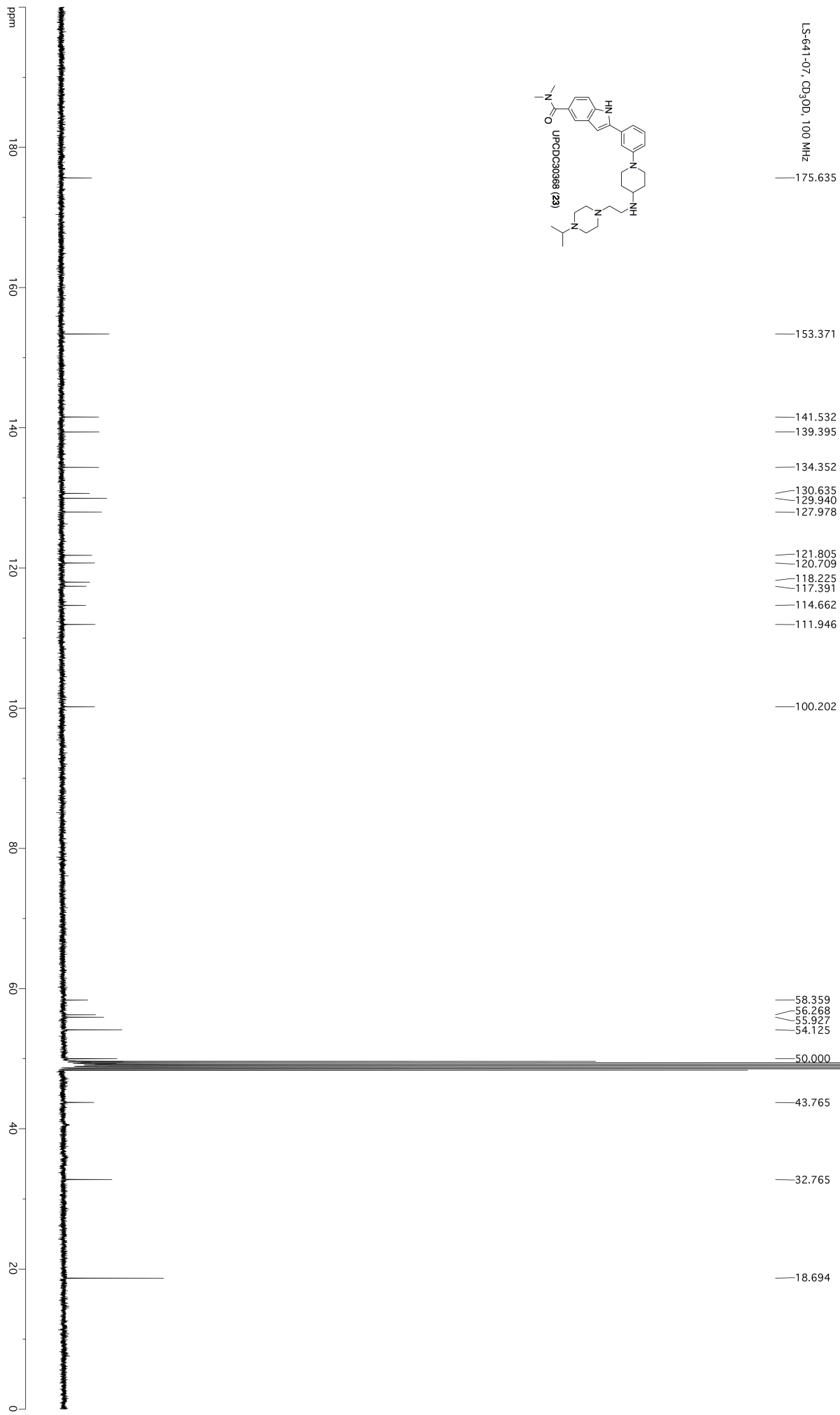
MK-605-18, acetone-d<sub>6</sub>, 500 MHz

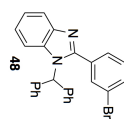


LS-641-04-crude, CD<sub>3</sub>OD, 500 MHz

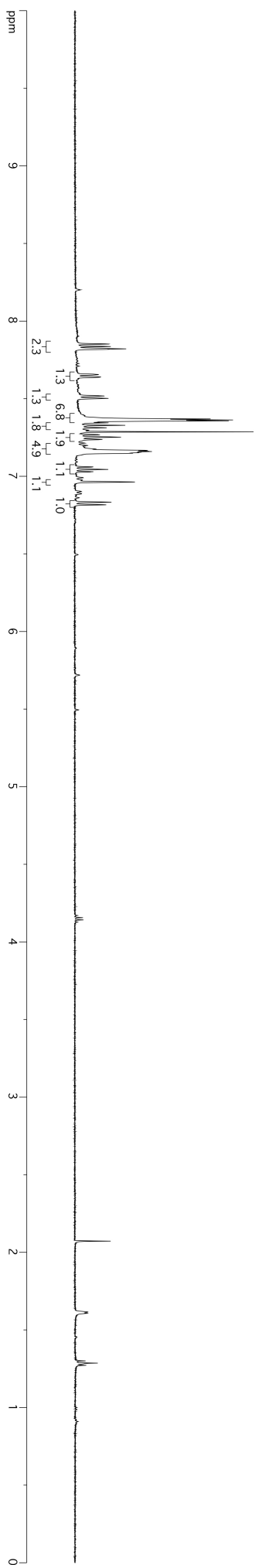


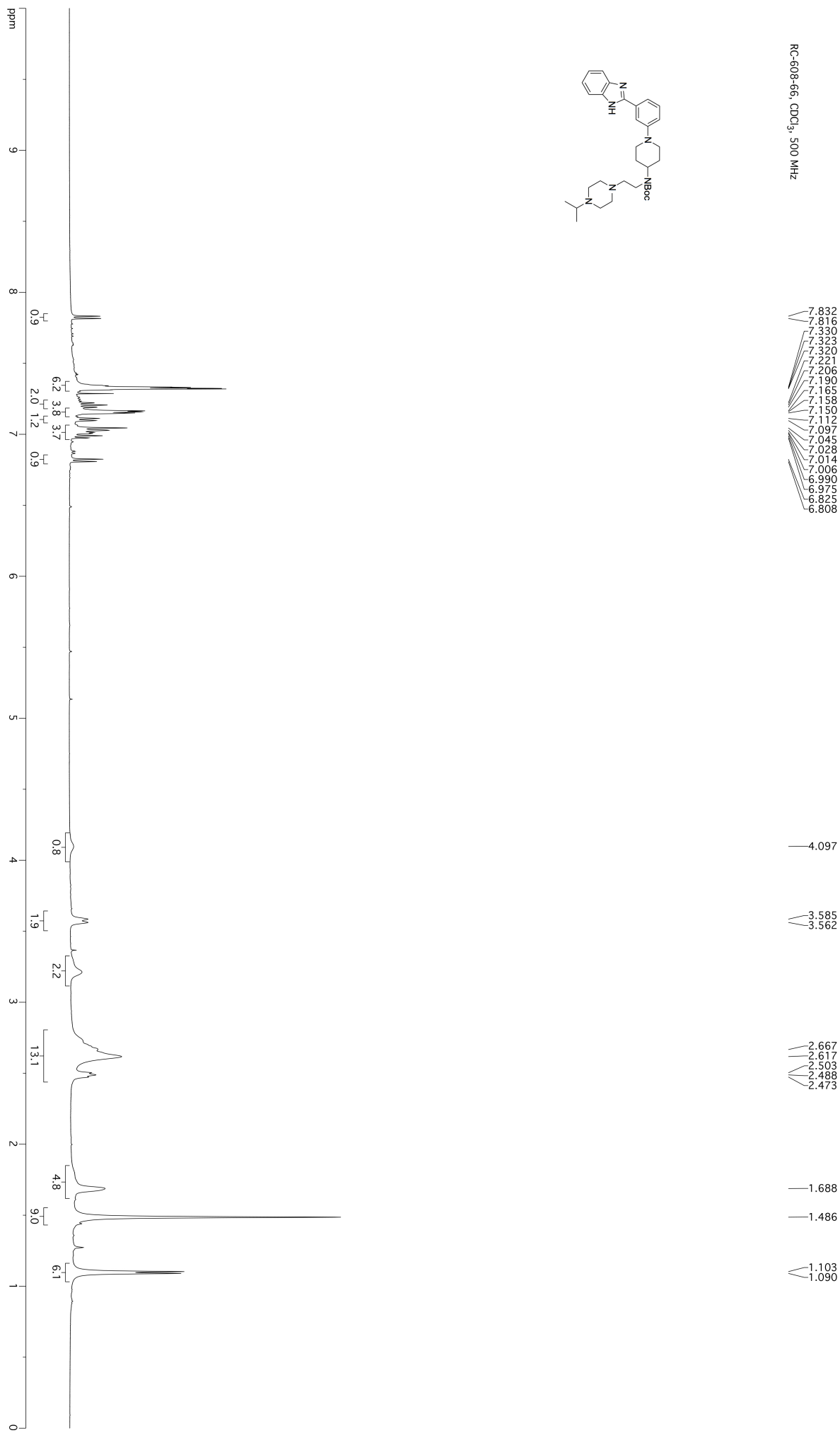
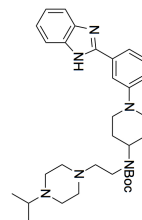
LS-641-07, CD<sub>3</sub>OD, 400 MHz



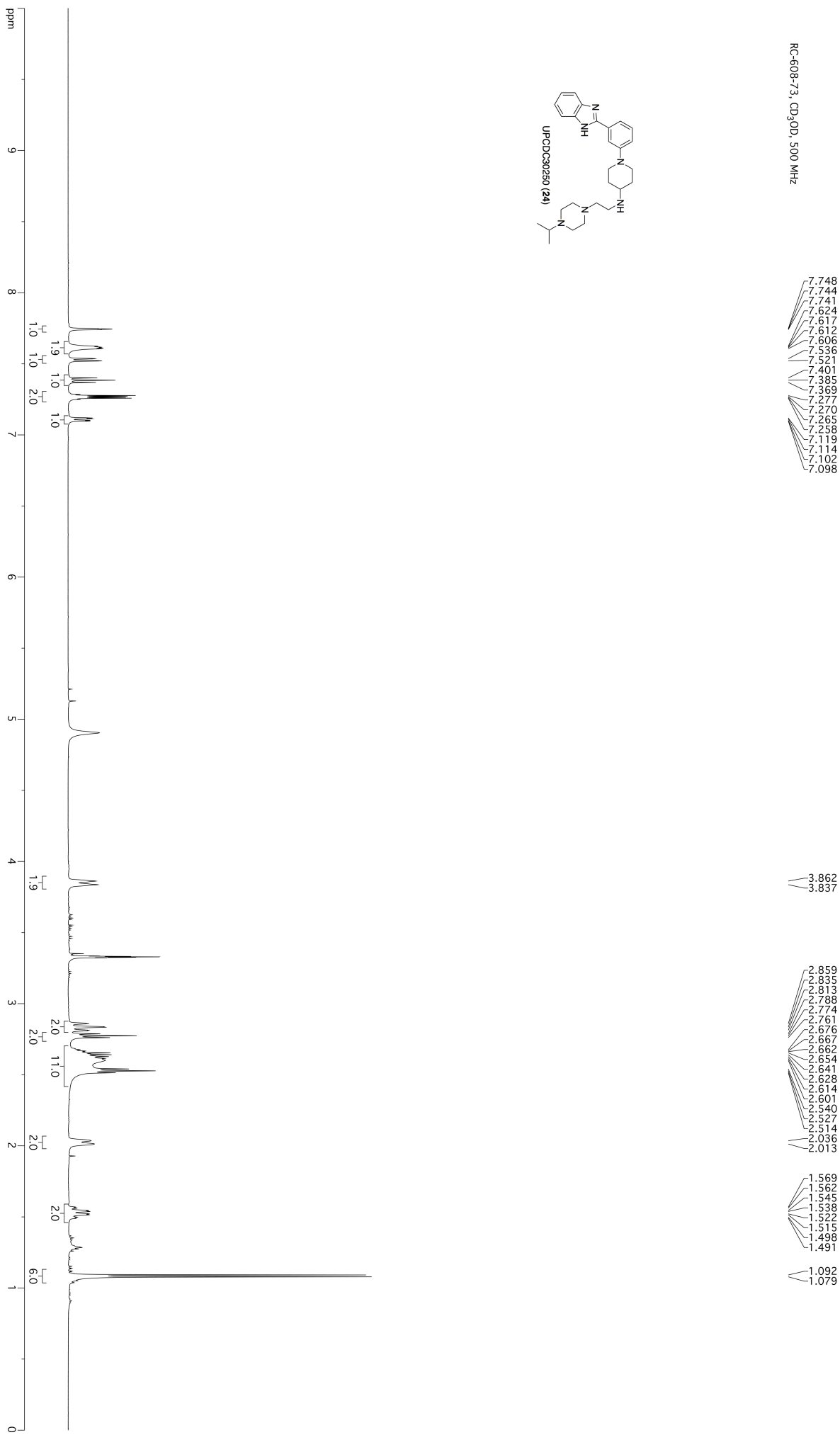
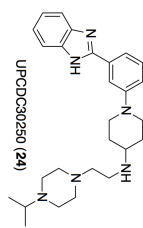
RC-608-59, CDCl<sub>3</sub>, 500 MHz

7.852  
7.836  
7.821  
7.818  
7.816  
7.654  
7.653  
7.640  
7.638  
7.517  
7.502  
7.370  
7.363  
7.357  
7.349  
7.344  
7.328  
7.313  
7.268  
7.267  
7.252  
7.238  
7.236  
7.176  
7.175  
7.175  
7.166  
7.161  
7.159  
7.155  
7.155  
7.152  
7.148  
7.061  
7.059  
7.044  
7.030  
7.028  
6.963  
6.833  
6.816

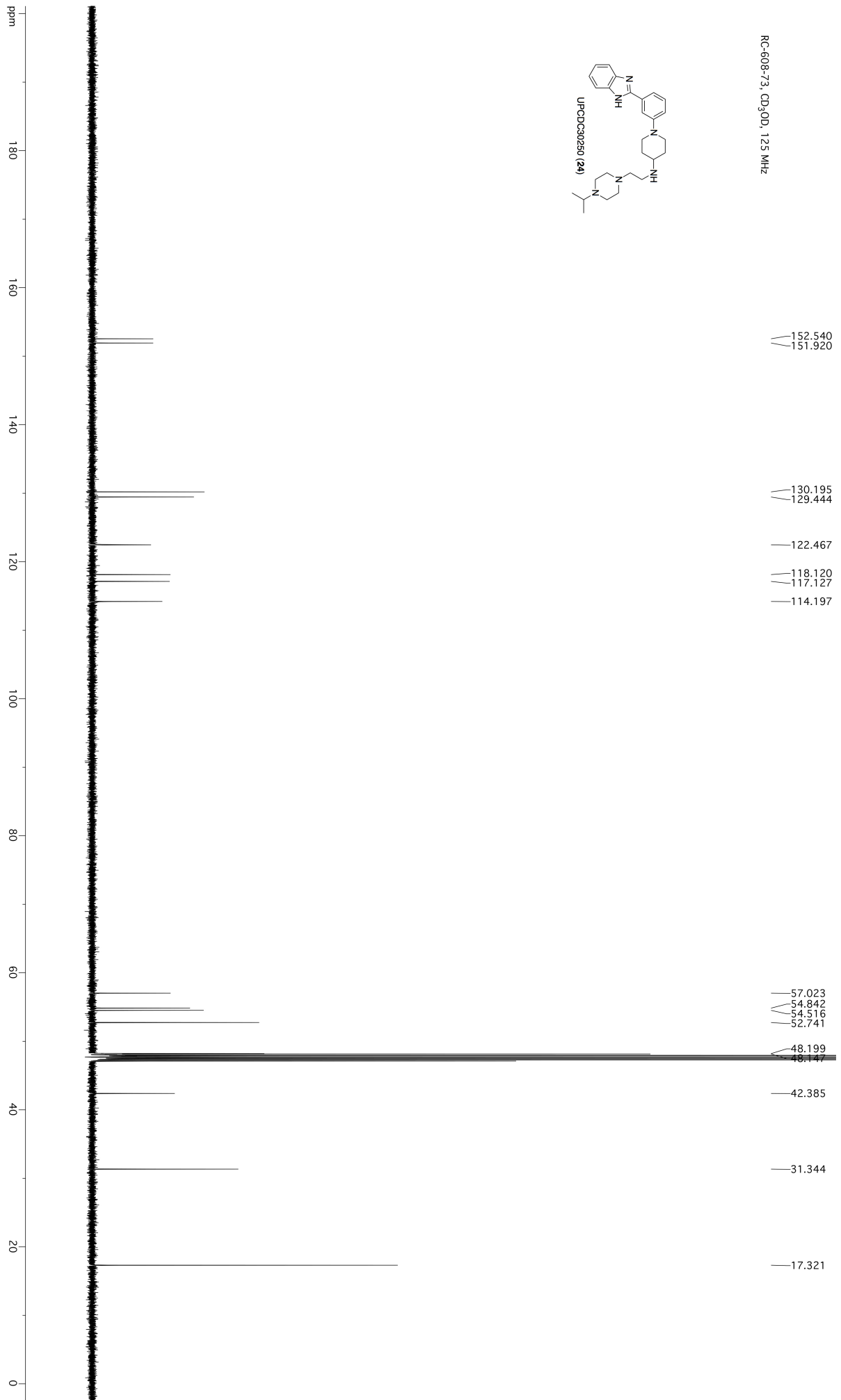


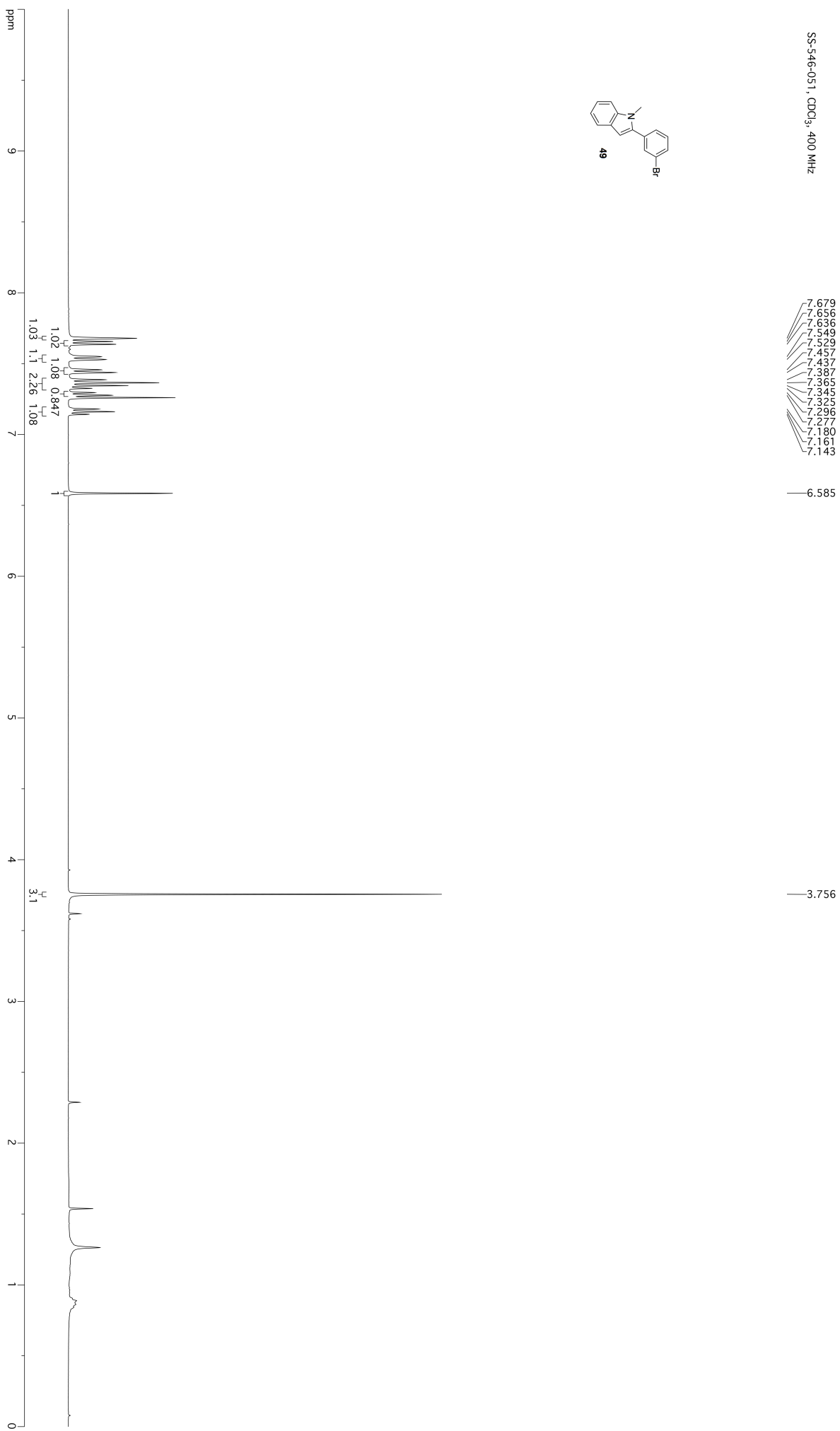
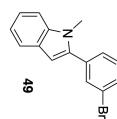
RC-608-66, CDCl<sub>3</sub>, 500 MHz

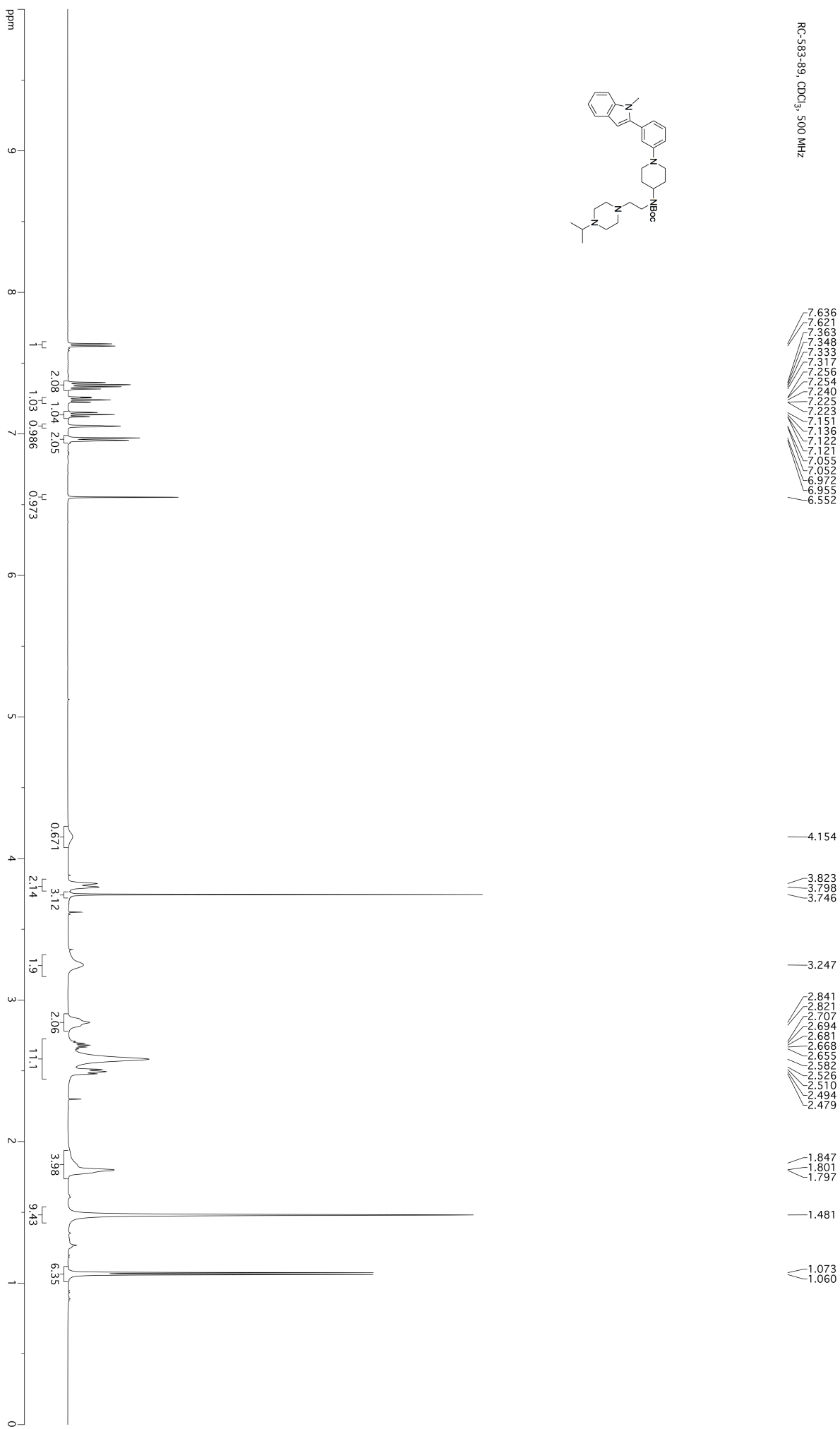
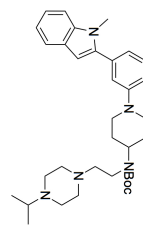
RC-608-r3, CD<sub>3</sub>OD, 500 MHz

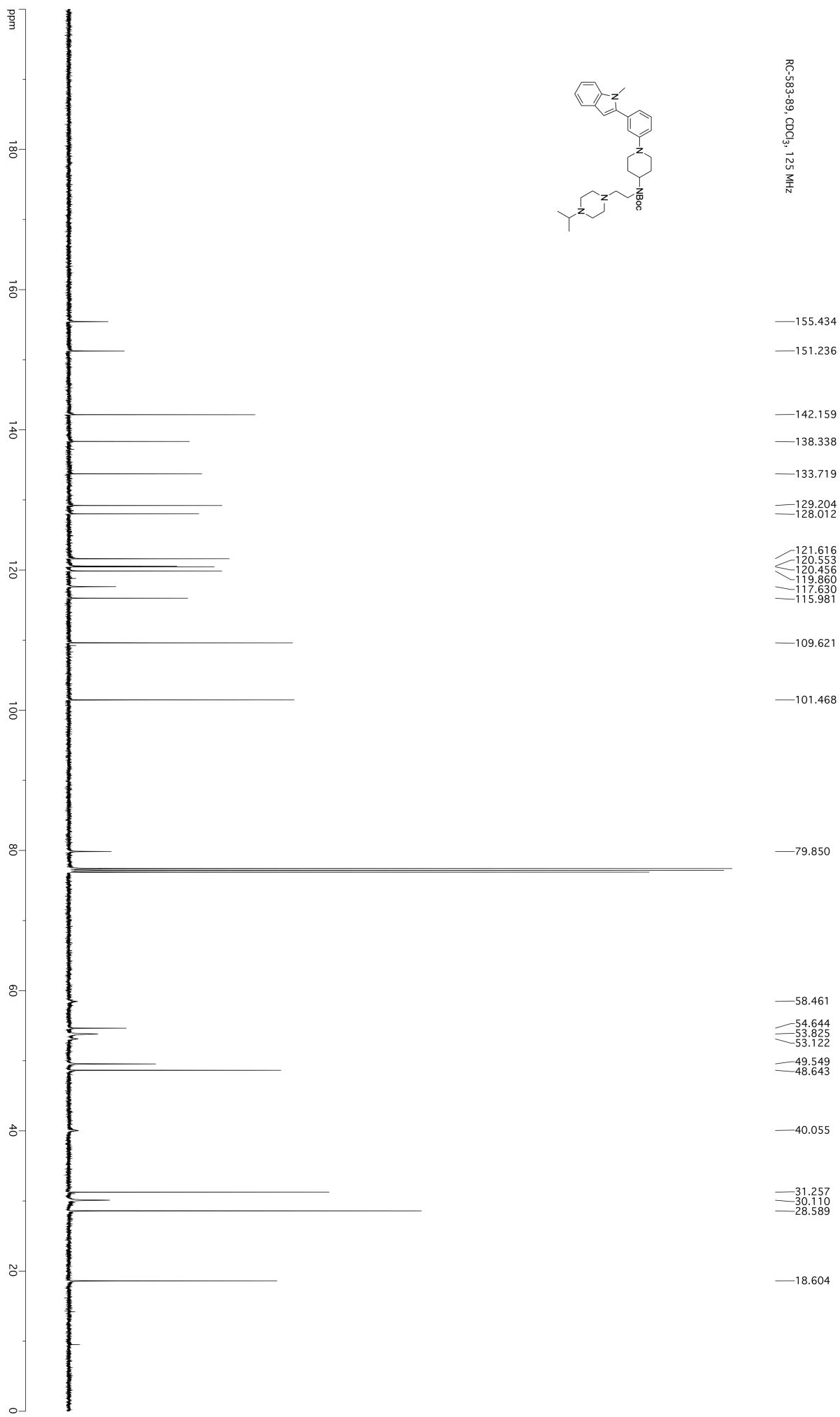






SS-546-051, CDCl<sub>3</sub>, 400 MHz

RC-583-89, CDCl<sub>3</sub>, 500 MHz



RC-583-93, acetone-d6, 400 MHz

