MolecularLibraries
Pathways to Discovery

Probe Report

Title: Identification of a Selective Small-Molecule Inhibitor of Breast Cancer Stem Cells—Probe 2

Authors: Leigh C. Carmody,[1] Andrew Germain,[1] Barbara Morgan,[1] Lynn VerPlank,[1] Cristina Fernandez,[1] Yuxiong Feng,[2] Jose R. Perez[1], Sivaraman Dandapani,[1] Benito Munoz,[1], Michelle Palmer,[1] Eric S. Lander,[1] Piyush B. Gupta,[1,2] Stuart L. Schreiber[1]

[1]The Broad Institute Probe Development Center, Cambridge, MA;[2]The Whitehead Institute, Cambridge, MA

Corresponding author email: lcarmody@broadinstitute.org

Abstract:

Cancer stem cells (CSCs), which drive tumor growth, are known to be resistant to standard chemotherapy and radiation treatment. This raises a significant unmet need to find therapies that can target CSCs within tumors because these cells are responsible for recurrence, the primary cause of patient mortality. However, one of the technical challenges of working with CSCs is that they are not stable outside the tumor environment and are not easy to grow in culture media. Hence, stable sibling cell lines that were induced into epithelial-to-mesenchymal transdifferentiation (EMT) to stably propagate CSC-enriched populations were used to screen a library of 300,718 compounds from the Molecular Libraries Small Molecule Repository (MLSMR). Several classes of selective inhibitors of CSCs were identified. The use of isogenic control cell lines for the secondary validation assays minimized the probability of false hits advancing along the critical path to probe development. Of these, 19 compounds were chosen based on their selectivity, potency, and chemical tractability and were retested in the primary screen and secondary assays. Three scaffolds were prioritized for structure-activity relationship studies. In this report, we describe the development of the chemical probe ML245 that displayed greater than 14-fold selective inhibition of the breast CSC-like cell line (HMLE_sh_ECad, IC_{50}=0.536 µM) over the isogenic control cell line (HMLE_sh_GFP). Furthermore, the probe (ML245) was screened against a panel of 68 biological targets that are commonly used in drug discovery for lead profiling and eight were identified as binding targets in a primary biochemical assay at a single dose of 10 µM, Adenosine (A_1, A_{2A}, A_3), Opiate κ, and Serotonin 5-HT$_{2B}$ G-protein coupled receptors, norepinephrine and dopamine transporters, and GABA$_A$ receptor. To further characterize the targets of ML245, gene expression studies were conducted to identify potential target pathways and targets mediating the response. Gene expression profiling revealed that ML245 regulated the expression of several pro-apoptotic/mitochondrial maintenance factors (such as caspase recruitment domain family member 10, activating transcription factor 4, mitochondrial ribosomal protein MRPL12), and DNA-modifying enzymes

(such as Early Growth Response 1, zinc finger-containing proteins ZNF295 and CGRRF1, and TBP-associated factor 1D), in the CSC-like cell line and not the isogenic control cell line. Although the link to regulation of cancer is unknown, upregulation of the gene CGRRF1 by the probe (ML245) and other CSC-selective inhibitors is exciting due to its function of determining cell-cycle arrest and may have a role in inhibiting the progression of Burkitt's lymphoma. More studies are needed to understand the direct target of the probe (ML245) and its method of action for inhibiting CSC-like cells.

Probe Structure and Characteristics:
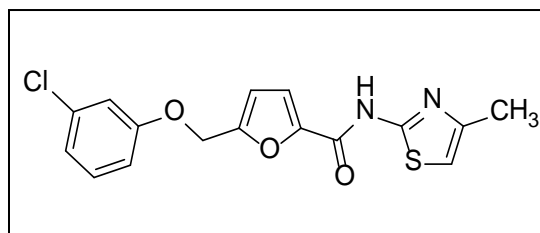
ML245

CID/ML No.	Target Name	EC$_{50}$ (µM) [SID, AID]	Anti-target	EC$_{50}$ (µM) [SID, AID]	Fold Selective	Secondary Assay Name: EC$_{50}$ (nM) [SID, AID]
CID 50904134/ ML245	Breast Cancer Stem Cell-like cell	0.536 [SID 110722981, AID 504449]	Mammary Epithelial Cell	7.87 [SID 110722981, AID 504450]	14.7	Tumorsphere, single point, Inactive [SID 10722981, AID 504859]

Recommendations for scientific use of the probe:

There is an obvious unmet need for developing breast cancer stem cell (CSC)-selective compounds. Breast CSCs are thought to be responsible for cancer recurrence and resistance to chemotherapeutic treatments. Developing probes against breast CSCs enables researchers to gain a better perspective and a pathway to targeting these cells in the clinic.

Nine small molecules have been identified to selectively inhibit breast CSCs. Only one molecule, salinomycin, has been identified to selectively inhibit this particular breast CSC-like cell line (HMLE_sh_ECad). Although it has a potency of approximately 1 µM against this cell line, it should be noted that salinomycin also shows toxicity against the control cell line (HMLE_sh_GFP) at 10 µM. Therefore, the goal of this project is to find a potent probe with less toxicity liability to control cell lines. In addition to the clinical value, these probes would have the potential to be used for novel target identification.

The current body of work identifies numerous genes whose expression is modified by treatment with the probe (ML245). Treatment of the HMLE_sh_ECad cell line identified changes in gene expression of several pro-apoptotic/mitochondrial maintenance factors (such as caspase recruitment domain family member 10 [CARD10], activating transcription factor 4 [ATF4], and mitochondrial ribosomal protein [MRPL12]) and DNA-modifying enzymes (such as early growth response 1 [EGR1], zinc finger-containing proteins [ZNF295, CGRRF1], and TBP-associated factor 1D [TAF1d]), in the CSC-like cell line and not the isogenic control cell line. Interestingly, the probe (ML245), an analog of ML245, and salinomycin upregulated the gene expression of RING finger protein CGRRF1, a protein that determines cell-cycle arrest. Furthermore, this increased expression of CGRRF1 correlated with a decreased division of Burkitt lymphoma cell lines (1), suggesting CGRRF1 may be an important target to investigate further. Furthermore, this probe (ML245) is currently being optimized for use in binding studies with the intent of identifying direct molecular targets within the cell. Moreover, the chemical probe (ML245) may also be used to investigate pathways regulating CSC development, lineage, self-renewal, and other basic research to develop a better understanding of breast CSCs.

1   Introduction

Scientific Rationale

Cancer stem cells (CSCs), which drive tumor growth, are known to be resistant to standard chemotherapy and radiation treatment (2-6). This resistance raises a significant unmet need to find therapies that can target CSCs within tumors because these cells are proposed to be responsible for recurrence, the primary cause of patient mortality.   In principle, it would be desirable to apply high-throughput screening (HTS) technologies to facilitate the identification of chemical compounds that specifically target CSCs.   Since CSCs generally comprise only a small minority within cancer cell populations, standard high-throughput cell viability assays applied to the broad populations of cancer cells cannot identify agents with CSC-specific toxicity (7). Accordingly, screening for agents that preferentially kill CSCs depends, at present, on the ability to obtain and to stably propagate highly pure CSC populations *in vitro*. Since CSCs cannot currently be stably propagated in culture, it has not been possible to leverage HTS to identify CSC-targeting compounds.

A novel HTS assay was recently developed to identify selective inhibitors of CSCs.   The approach used stable sibling cell lines that are induced into epithelial-to-mesenchymal transdifferentiation (EMT) as a method to stably propagate CSC-enriched populations (8).   The availability of isogenic control cell lines for the secondary validation assay of the primary screen, which is absent in most screens, minimized the probability of advancing spurious compound hits that are not selective for CSCs but rather target genetic differences between screened cell lines.

The objective of the experiments in this probe report is to identify chemical compounds that can selectively kill breast CSCs.   The chemical compounds identified by the experiments in this report would be promising new starting points for drug candidates for anticancer therapy and would also serve as useful probes to study CSC biology, which are currently lacking.

The current state of the art for probe molecules in cell culture assays against breast CSCs is salinomycin.   Salinomycin has a potency of about approximately 1 μM against HMLE_sh_ECad with approximately 10-fold selectivity over the control HMLE_sh_GFP (9). Since salinomycin shows toxicity at approximately 10 μM against HMLE_sh_GFP, this may cause gross toxicity to cells in general. A detailed comparison to existing art is provided in Section 4.1 and **Table A2** in **Appendix G**.

2 Materials and Methods

See subsections for a detailed description of the materials and methods used for each assay.

2.1 Assays

A summary listing of completed assays and corresponding PubChem AID numbers is provided in **Table A1** (**Appendix A**). Refer to **Appendix B** for the detailed assay protocols.

2.1.1 Primary HTS Using CellTiter-Glo (AID 2717)

HMLE_sh_ECad cells were prepared as previously described (8).

Procedures: Briefly, HMLE cells expressing either shRNA targeting E-cadherin (HMLE_sh_ECad) were propagated in 1:1 mixture of 10% fetal bovine serum (FBS; HyClone), 1% Penicillin/Streptomycin (Pen/Strep; Cellgro), 1% Glutamax-1 (Invitrogen), 70 nM Hydrocortisone (Sigma), 12 µg/ml Insulin (Sigma), 50 µg/ml Gentamicin (Sigma), 12.5 µg/ml Plasmocin (InVivogen), 10 ng/ml EGF in DMEM (Cellgro) with Mammary Epithelial Cell Growth Medium (MEGM complete medium; Lonza, Basel, Switzerland) at 37 ºC, 5% CO_2. For screening, the cells were counted and resuspended in complete media without serum. Next, 2,000 cells in 50 µl were plated per well in white, tissue culture-treated, 384-well plates (Corning). The cells were incubated at 37 ºC, 5% CO_2 for at least 4 hours and pinned with 100 nl of compounds. The cells were incubated approximately 72 hours, then 30 µl of CellTiter-Glo (Promega) diluted 1:3 with PBS was added to the well. The plates were read using the EnVision (PerkinElmer; Luminescence 0.1 sec/well) after 12 minutes.

2.1.2 Primary Retest (AID 449748)

Repeat of the primary screen at dose using CellTiter-Glo.

2.1.3 Primary Cell Viability Protocol with HMLE_sh_ECad Cells Using CellTiter-Glo (AID 493226, AID 493176, AID 504449, AID 504535, AID 504667, AID 504788)

HMLE_sh_ECad cells were prepared as stated in Section 2.1.1 and as previously described (8).

Procedures: Briefly, HMLE cells expressing shRNA targeting E-cadherin (HMLE_sh_ECad) were propagated in 1:1 mixture of 10% FBS (HyClone), 1% Pen/Strep (Cellgro), 1% Glutamax-1

(Invitrogen), 70 nM Hydrocortisone (Sigma), 12 µg/ml Insulin (Sigma), 50 µg/ml Gentamicin (Sigma), 12.5 µg/ml Plasmocin (InVivogen), 10 ng/ml EGF in DMEM (Cellgro) with Mammary Epithelial Cell Growth Medium (MEGM complete medium; Lonza, Basel, Switzerland) at 37 ⁰C, 5% CO_2. For screening, the cells were counted and resuspended in complete media without serum. Next, 2,000 cells in 50 µl were plated per well in white, tissue culture-treated, 384-well plates (Corning). The cells were incubated at 37 ⁰C, 5% CO_2 for at least 4 hours and pinned with 100 nl of compounds. The cells were incubated approximately 72 hours, then 30 µl of CellTiter-Glo (Promega) diluted 1:3 with PBS was added to the well. The plates were read using the EnVision (PerkinElmer; Luminescence 0.1 sec/well) after 12 minutes.

2.1.4   Secondary Cell Viability Protocol with HMLE_sh_GFP Cells Using CellTiter-Glo (AID 504325, AID 493196, AID 504450, AID 504533, AID 504666, AID 504789)

HMLE_sh_GFP cells were prepared as previously described (8). Refer to the procedures as described in Section 2.1.3 and **Appendix B**.

2.1.5   Assay Provider Cell Viability Protocol with CellTiter-Glo

HMLE cells expressing shRNA targeting Twist (HMLE_sh_Twist) were prepared as previously described (8). Refer to the procedures as described in Section 2.1.3 and **Appendix B**.

2.1.6.   Secondary Orthogonal Assay for *In Vitro* Inhibition of Tumorspheres with Sum159 HMLER Cells (AID 504623)

Sum159 cells were provided by the Assay Provider as previously described (8).

Procedures: Tumorsphere assays were performed as previously described (10) with minor changes. Sum159 cells were propagated in 5% FBS (HyClone), 1% Pen/Strep, 1% Glutamax-1, 12 µg/ml Insulin, 50 µg/ml Gentamicin in F12/DMEM (Cellgro). Next, 96-well, ultra-low adhesion plates (Costar) were plated with 100 µl media respective to cell type. Then, 400 nl of compounds were pinned into the media. The cells were harvested, counted, and resuspended in their propagation media with 1% methylcellulose (ES-CultM3120, Stem Cell Technologies). Then, 100 µl of resuspended cells were added to the plates containing media with compound for a final count of 2000 cells/well in 200 µl with 0.5% methylcellulose. Tumorspheres were allowed to form for 9 days incubated at 37 °C, 5% CO_2. Tumorspheres were imaged using a 2X objective on the ImageXpress Micro (Molecular Devices, Sunnyvale, CA). Cell clusters greater than 100 µM in diameter were identified using MetaXpress software (version 3.1; Molecular Devices).

2.1.7   Gene Expression Assays with HMLE_sh_ECad and HMLE_sh_GFP Cells

Procedures: Three compounds (ML245, CID50904149, and salinomycin) selectively inhibit a breast CSC-like cell line (HMLE_sh_ECad) over the isogenic control cell line (HMLE_sh_GFP) in 72-hour toxicity assays. The IC_{50} concentrations derived from the toxicity experiments were used for the gene expression experiments below.

The full protocol and data are published on PubChem (http://www.ncbi.nlm.nih.gov):

- AID 493226 (http://www.ncbi.nlm.nih.gov/pcassay?term=493226),
- AID 504535 (http://www.ncbi.nlm.nih.gov/sites/entrez?db=pcassay&term=504535), and
- AID 50449 (http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?q=&aid=504449).

Replicate samples (3-4) of HMLE_sh_ECad and HMLE_sh_GFP cells were treated with vehicle (DMSO), ML245, CID50904149, or salinomycin at the derived IC_{50} dose derived from above for 24 hours before isolation of RNA. Total RNA was isolated using the RNeasy Protect Mini Kit (Qiagen). Quality control (QC) processing of the RNA samples and the gene expression analysis were performed by the Genome Analysis Platform (GAP) at the Broad Institute.

Briefly, RNA samples were analyzed for quality using Agilent Bioanalyzer Chips. cRNA synthesis from the total RNA samples passing QC was prepared for analysis on a HumanHT-12 Expression BeadChip (Illumina, San Diego, CA) according to the manufacturer's instructions. QC checks and analyses were confirmed with GenomeStudio (version 2010.3; Illumina, San Diego, CA). The entire data set was normalized using the quantile module available in open source program GenePattern (http://genepattern.broadinstitute.org/). After the samples were normalized, gene expression was compared between the DMSO-treated and compound-treated cells. Those genes with a significant difference (p>0.005) and had 1.5-fold change in expression were selected.

2.2 Probe Chemical Characterization

After preparation as described in Section 2.3, the probe (ML245) was analyzed by UPLC, ^{1}H and ^{13}C NMR spectroscopy, and high-resolution mass spectrometry. The data obtained from NMR and mass spectroscopy were consistent with the structure of the probe, and UPLC analysis indicated an isolated purity of >95%. Characterization data (^{1}H NMR spectra and UPLC chromatograms) of the probe (ML245) are provided in **Appendix E.**

The solubility of the probe (ML245) was experimentally determined to be 0.07 µM in Phosphate buffered saline (PBS; pH 7.4, 23 °C) solution. Plasma protein binding (PPB) studies showed that the probe was 99.7% bound in human plasma. The probe was stable in human plasma, with approximately 100% remaining after a 5-hour incubation period. The compound was found to be stable in the presence of glutathione (GSH) with 92% remaining after 48 hours.

The stability of the probe (ML245) in PBS (0.1% DMSO) was measured over 48 hours. We noticed that the concentration of the probe fluctuated between 37.5% and 131% through the course of the assay (data not shown). We believe that this fluctuation is due to the low

solubility of the probe. Thus, we decided to determine the total amount of the probe present in the well after the probe was treated with PBS alone for a given length of time. We added acetonitrile at various time points to wells containing the probe in PBS and measured the total amount of the probe. This result is shown in **Figure 1**. From these results, the probe seems to be stable in PBS since more than 89% is still present after 48 hours of incubation. **Table 1** summarizes known probe properties as displayed in PubChem.

Figure 1. Stability Data for the Probe (ML245) in PBS Buffer (pH 7.4, 23 °C) Plus Acetonitrile Over 48 Hours.
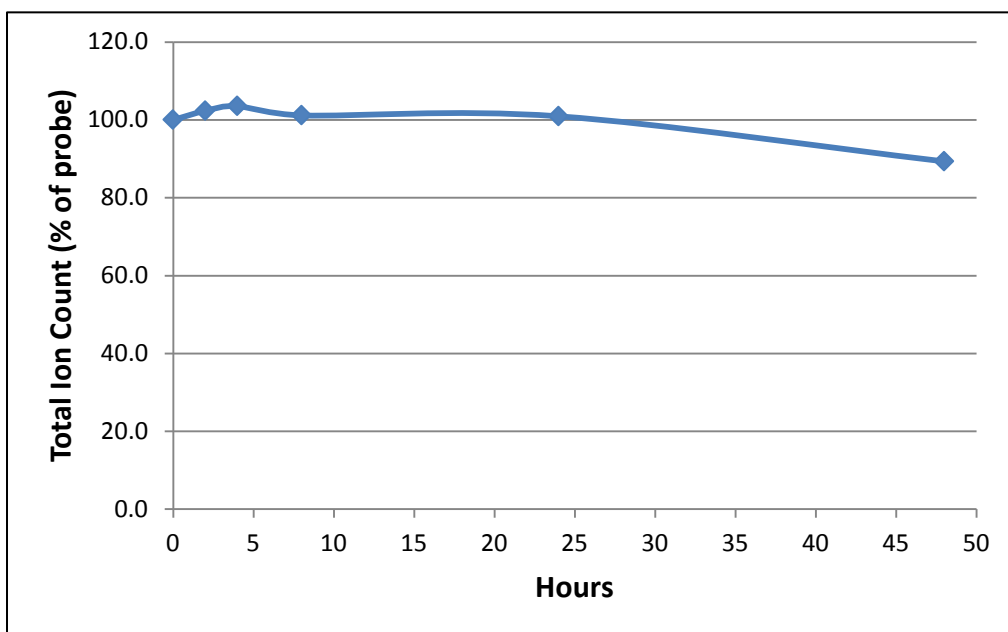
Table 1. Summary of Known Probe Properties in PubChem.

IUPAC Chemical Name	5-((3-chlorophenoxy)methyl)-N-(4-methylthiazol-2-yl)furan-2-carboxamide
PubChem CID	50904134
Molecular Weight	348.80 [g/mol]
Molecular Formula	$C_{16}H_{13}ClN_2O_3S$
XLogP3-AA	4
H-Bond Donor	1
H-Bond Acceptor	3
Rotatable Bond Count	5
Exact Mass	348.03
Topological Polar Surface	92.6

2.3   Probe Preparation

The probe (ML245) was synthesized from 5-(hydroxymethyl)furan-2-carboxylic acid **1** in three steps (see **Scheme 1**).   Conversion of 5-(hydroxymethyl)furan-2-carboxylic acid **1** into the dichloride with thionyl chloride, followed by amidation provided furanyl chloride **2**. Alkylation with 3-chlorophenol provided the probe 5-((3-chlorophenoxy) methyl)-*N*-(4-methylthiazol-2-yl)furan-2-carboxamide in 47% yield over three steps.

Scheme 1.   Synthesis of the Probe (ML245)
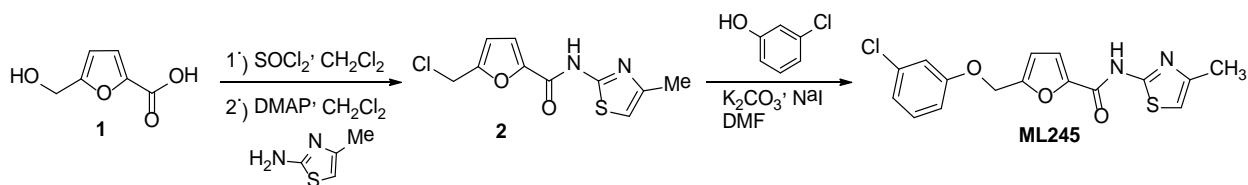
Experimental procedures for the synthesis of the probe (ML245) are provided in **Appendix C**.

2.4   Additional Analytical Analysis

The probe (ML245) was found to be stable in human and mouse plasma with 100% remaining after 5 hours in each case. Experimental procedures for the analytical assays are provided in **Appendix D**.

3 Results

Probe attributes:

- IC_{50} = 0.536 µM against HMLE_sh_ECad
- IC_{50} against HMLE_sh_GFP > 14 x IC_{50} against HMLE_sh_ECad

3.1 Summary of Screening Results

Figure 2 displays the critical path for probe development.   To explore SAR, a number of analogs were synthesized and tested.   Selected results are shown in **Tables 2-10** in Section 3.4.

A high-throughput screen (HTS) of 300,718 compounds (PubChem AID 2717) assessing the viability of the breast CSC-like population (HMLE_sh_ECad) was completed. HMLE_sh_ECad cells have undergone an EMT by introduction of a short hairpin RNA (shRNA) targeting the *CDH1* gene, which encodes E-cadherin into human mammary epithelial cells (HMLEs). These cells provide a uniform population of breast CSC-like cells with which to conduct a large-scale screen (8). The signal was normalized to neutral (DMSO) and positive (Puromycin) control, and a 75% inhibition cutoff at a screening concentration of 3.75 µM average was used to define a hit. Next, 3,190 hits were identified as inhibitors of HMLE_sh_ECad. The list of active hits was narrowed by discarding compounds that hit in >10% of MLPCN assays or scored positive in prior mammalian cell toxicity screens. Compounds were ranked in order of their promiscuity in MLPCN assays and the top 2,500 compounds were requested. Of these, 2,244 compounds were available from the compound supplier (i.e., BioFocus).

These cherry-picked compounds were retested in 9-point dose in the primary assay using the cell lines HMLE_sh_ECad and HMLE_sh_GFP (control). A shRNA targeting of *enhanced GFP* (eGFP) gene, which is not endogenously expressed in the HMLE cells, was introduced to the cell line to create the control mammary epithelial cell line. These cells do not undergo EMT. From the 2,244 compounds retested at dose concentrations (range: 160 nM-19.5 µM; **Table A1**, Appendix A), 97% (2181 compounds) re-confirmed in this assay (AID 449748). In parallel, these 2,244 compounds were included in HMLE_sh_GFP viability assays (AID 463074) to determine if these compounds affected the viability of the control cell line. Taken

together, 26 compounds were identified as 25-fold more potent against HMLE_sh_ECad versus HMLE_sh_GFP.

After completing the retests and a secondary assay at dose from DMSO stocks, 19 exact or similar compounds were ordered from commercial vendors. These 19 compounds were tested in the primary screen at dose and the secondary screen to determine if the Quality Controlled dry powder compounds retained selectivity. From this set of dry powder compounds, a furan containing compound (CID4791237, **Table 2, Entry 1**) retained selectivity, and this scaffold was prioritized for probe development (AID 493176 and AID 493196).

Figure 2. Critical Path for Probe Development.
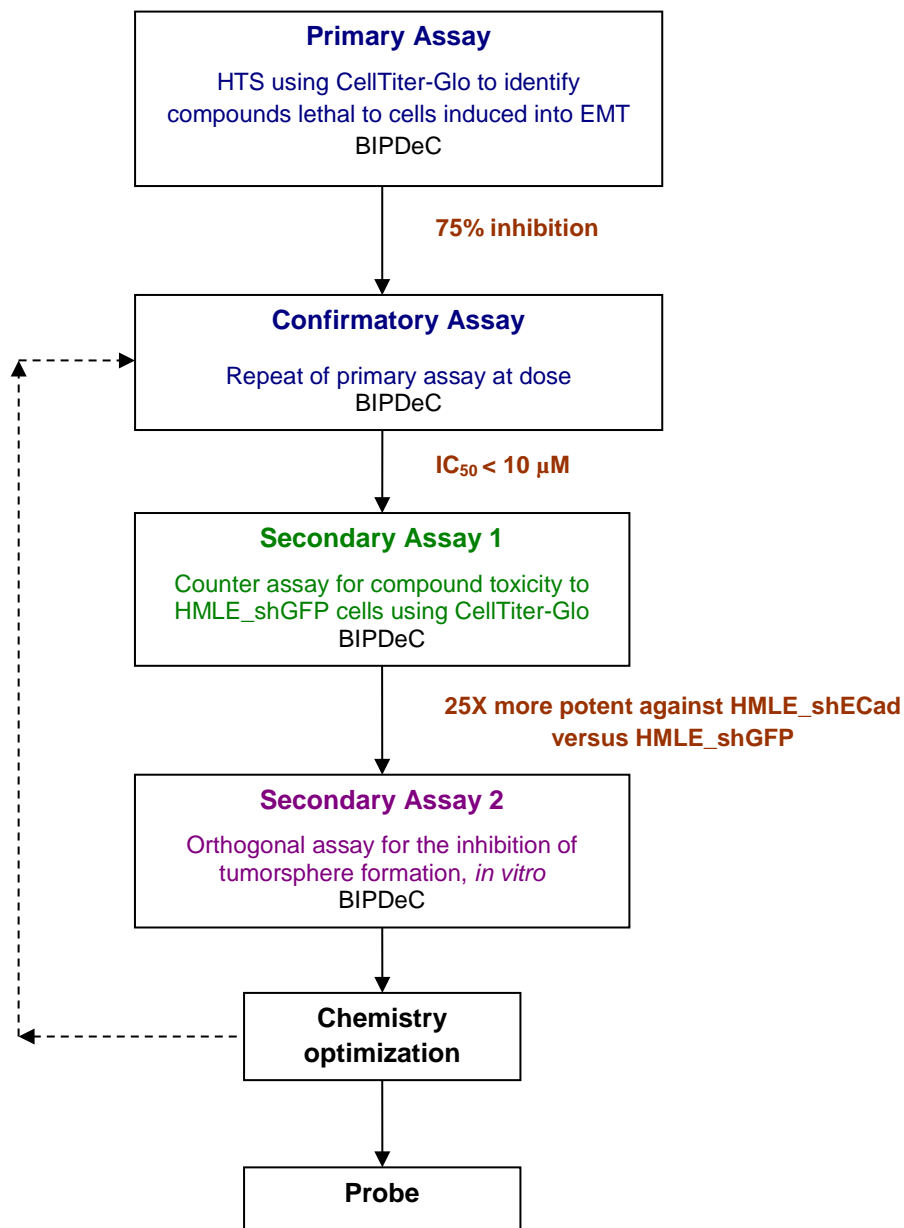
3.2   Dose Response Curves for the Probe

Figure 3.   Dose-dependent Activity of the Probe (ML245).
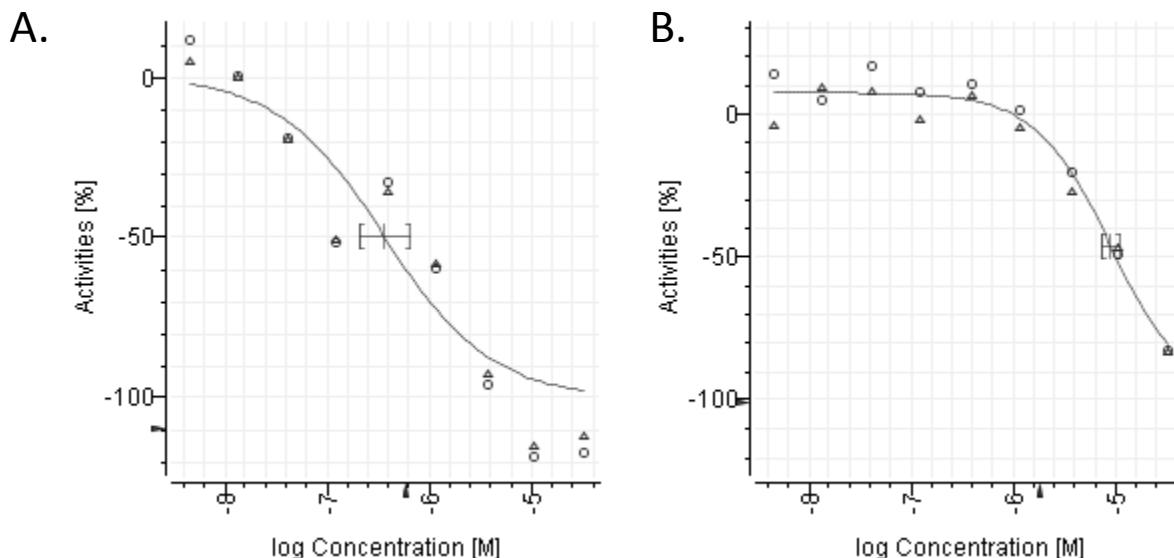
Figure 3.   Dose-dependent activity of the probe (ML245) in Target and Counterscreen Assays: Primary screen from dry powders on HMLE_sh_ECad (IC$_{50}$ = 0.536 µM; AID 504535, 504667) (**A**); control cell line HMLE_sh_GFP (7.87 µM) Toxicity (AID 504533, 504666) (**B**).

3.3   Scaffold/Moiety Chemical Liabilities

A search of PubChem for the hit compound from the primary screen (CID 4791237, **Table 2, Entry 1**) revealed that it had been tested in 276 BioAssays and was confirmed as active only in one assay other than the assays described in this report. The assay was a screen for Niemann-Pick C1 promoter activators (1 µM).

3.4   SAR Analysis

Screening results for the primary assay are described in **Tables 2-10**. The results of SAR at the amide are shown in **Table 2**. The free acid (**Table 2, Entry 2**) was found to be inactive. Methylation of the amide (**Table 2, Entry 3**) led to a drop in potency and selectivity. Replacement of the thiazole with substituted thiazoles (**Table 2, Entry 4 and Entry 5**) or removal of the methyl substitution (**Table 2, Entry 6**) resulted in a significant drop in potency and selectivity in all cases. Replacing the methyl thiazole with benzothiazole (**Table 2, Entry 7**) led to loss of activity. Substitution with other heterocyclic amides (**Table 2, Entry 8 and Entry 9**)

also resulted in inactive compounds. Replacement of the thiazole with aliphatic amides (**Table 2**, **Entry 10 and Entry 11**) also resulted in inactive compounds. **Table A4** in **Appendix H** provides data on all other amide analogs that were tested. Based on the data in **Table 2** and **Table A4**, we concluded that the 2-methyl thiazole amide was the optimum group for the eastern portion of the molecule for potency and selectivity against the HMLE_shEcad cell line. For all of the subsequent SAR studies we retained the 2-methyl thiazole on the eastern side.

Tables 3, 4 and 5 illustrate the effect of substitution on the phenoxy group on the western portion of the molecule. **Table 3** contains data on a set of compounds that evaluates the effect on electron withdrawing and electron donating groups on the *ortho*-position. Substitution with a chloro (**Table 3**, **Entry 3**), a fluoro (**Table 3**, **Entry 4**), a methyl (**Table 3**, **Entry 7**) and a trifluoromethoxy (**Table 3**, **Entry 6**) led to potent compounds, however, with a loss in selectivity.

Table 4 describes the results from substitution at the *meta*-position. As with the *ortho*-position, substitution with a chloro (**Table 4, Entry 2**), a fluoro (**Table 4**, **Entry 3**), a methyl (**Table 4**, **Entry 6**) and a trifluoromethoxy (**Table 4**, **Entry 5**) resulted in compounds that were around 200-500 nM while, however, compounds substituted at the *meta*-position retained between 7- and 15-fold selectivity. This finding led us to further investigate the *meta*-position, and these results are summarized in **Table A5** in **Appendix H**. As can be seen in **Table A5**, substitution at the *meta*-position with a variety of alkyl and alkoxy groups result in compounds that maintain potency but with decreased selectivity compared to the compounds in **Table 3**.

Substitution at the *para*-position of the phenyl group, as shown in **Table 5**, results mostly in compounds with a slight decrease in activity, but a significant drop in selectivity.

Having investigated the effects of a range of common substituents individually on the *ortho*-, *meta*-, and *para*- positions, the effects of a set of carefully designed di-substituted analogs were evaluated. Compared to the unsubstituted phenyl, the *ortho*-chloro analog (**Table 3**, **Entry 3**) was 10-fold more potent but less selective. The *meta*-chloro analog (**Table 4**, **Entry 2**) was less potent but retained the selectivity (15-fold) inherent in the parent unsubstituted analog. Hence, a collection of di-chloro substituted analogs were synthesized, and these results are summarized in **Table 6**.

Several of the compounds were found to have potencies of less than 1 μM; however, most lacked reasonable selectivity. The 3,5-dichloro substituted compound (**Table 6, Entry 7**) was

the exception. It was extremely potent with 312 nM, while also retaining selectivity against the control cell line at greater than 14-fold. **Table A7** in **Appendix H** contains data on all other di- and tri-substituted compounds that were tested.

Table 7 contains data for compounds in which the phenyl group has been replaced with pyridyl groups. Intriguingly, replacement of the phenyl with 2-pyridyl (**Table 7, Entry 2**) retains some activity while being completely inactive against the control cell line and is significantly more soluble. This finding has led us to synthesize a group of compounds in which the phenyl is replaced with pyridyls in an effort to take advantage of the selectivity and solubility. However, these analogs did not improve potency. Data for these compounds can be found in **Table 7** (**Entries 2-4**). The phenoxy group was also replaced with several alkoxy groups; however, this replacement led to mostly inactive compounds (**Table A6** in **Appendix H**).

Modification of the western portion of the molecule was carried out, and the data is shown in **Table 8.** Replacement of the ether oxygen with a carbon (**Table 8, Entry 2**) results in similar activity but a slight loss in selectivity. Replacement with a nitrogen (**Table 8, Entry 3**) results in an increase in potency but a slight loss in selectivity. Conversion of the ether to a sulfone (**Table 8, Entry 4**) results in a loss of activity. The addition of a methylene between the phenyl group and the ether oxygen results in a loss of activity and selectivity (**Table 8, Entry 5**). Replacement of the phenyl ether with 2-aminopyridine (**Table 8, Entry 6**) results in a slight loss in activity but an increase in selectivity and solubility. This finding has led us to synthesize a group of compounds in which the phenoxy is replaced with *ortho*- and *meta*-substituted aminopyridines in an effort to take advantage of the selectivity and solubility of the aminopyridine while trying to increase potency. As in the case of hydroxypyridine, substitution did not lead to an increase in potency. Data can be found in **Table A10** in **Appendix H**.

Table 9 illustrates the effect of substitution in furan and replacement of the furan with other aromatic groups. Substitution on the furan with a methyl group (**Table 9**, **Entry 2**) results in a loss of activity and selectivity. Interestingly, replacement of the furan with a thiophene (**Table 9, Entry 3**) results in a compound that is much more potent towards the control cell line. Replacement with a phenyl (**Table 9, Entry 4**) results in a slight loss of potency and selectivity and replacement with a pyridine results in a significant loss in potency and selectivity.

Several analogs were made based on the phenyl replacement in an effort to increase potency and selectivity as seen in **Table 10.** Substitution with halogens and methyl at the *ortho*- or *meta*-positions results in a slight increase in potency and selectivity; however, these compounds are significantly less potent than their furan equivalents. Several aminopyridine analogs of the phenyl series were synthesized; however, while increasing solubility and selectivity these compounds were much less potent (**Table A11, Appendix H**).

All five of the other possible regioisomers of the furan were synthesized and tested. All of the other isomers either suffered from a loss in selectivity or potency. This data can be found in **Table A9** in **Appendix H**.

Based on the SAR described previously, the 3-chloro-substituted analog (**Table 4**, **Entry 2**) was chosen as the probe due to its excellent potency and selectivity.

Figure 4. Dose-dependent Activity of Select Analogs in the Primary Assay.
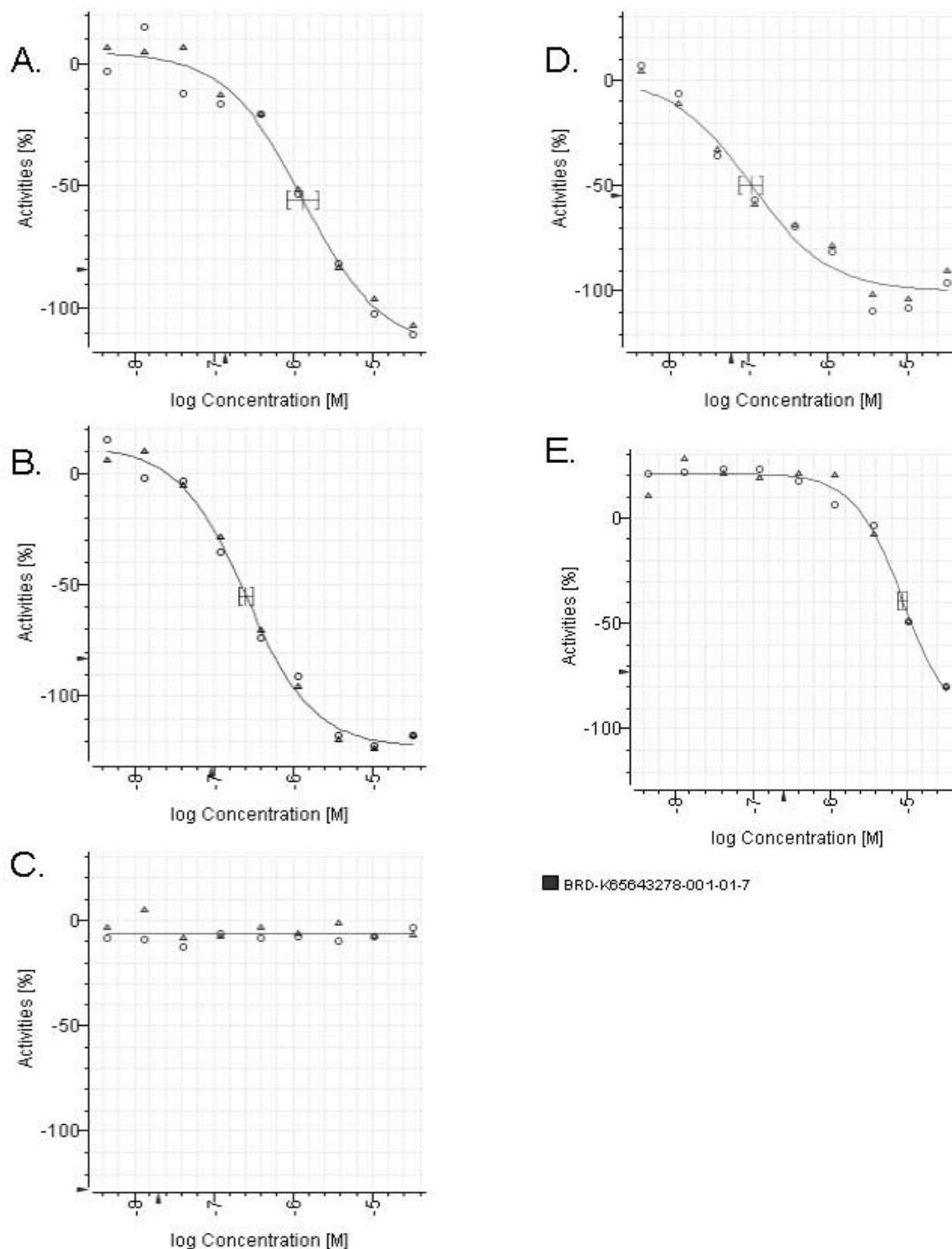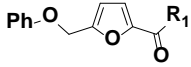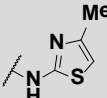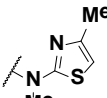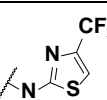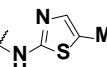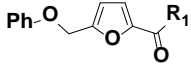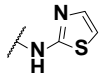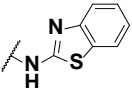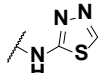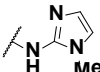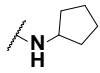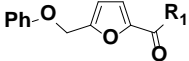
Figure 4. Four curves of analogs tested in the primary screen at dose: CID4791237 (**A; Table 2,** Entry 1), CID51003702 (**B; Table 6, Entry 6**), CID9214159 (**C; Table 2, Entry 10**), CID50944077 (**D; Table A7, Entry 9**), CID50904149 (**E; Table 4, Entry 10**).

Table 2. SAR of the Amide.
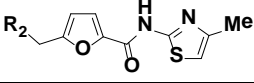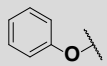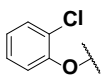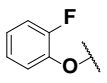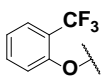
	SAR Analysis		Structure		Target Activity		Anti-target Activity	
Entry No.	CID SID Broad ID	*	Ph-O—[furan]—R_1, R_1	n	HMLE_sh_ECad EC_{50} (µM)	n	HMLE_sh_GFP EC_{50} (µM)	Selectivity (HMLE_sh_GFP/ HMLE_sh_ECad)
1	4791237 104170322 BRD-K39514895	S	[4-Me-thiazol-2-yl-amino] (Me)	7	1.64	8	24.59	15.0
	Solubility (PBS): 2.6 µM							Purity 91%
2	580491 104170315 BRD-K55148929	S	OH	5	Inactive	6	Inactive	NA
	Solubility (PBS): 496 µM							Purity >95%
3	51003714 117687937 BRD-K65452642	S	[4-Me-thiazol-2-yl, N-Me]	3	22.37	3	22.15	0.99
	Solubility (PBS): <1 µM							Purity: >95%
4	38728154 112208899 BRD-K07901599	S	[4-CF_3-thiazol-2-yl-amino]	7	8.38	7	22.60	2.70
	Solubility (PBS): <1 µM							Purity 88%
5	41549965 110722997 BRD-K32184050	S	[5-Me-thiazol-2-yl-amino]	5	19.58	5	Inactive	NA
	Solubility (PBS): <1 µM							Purity >95%
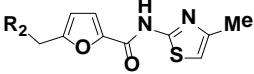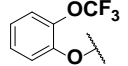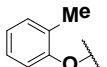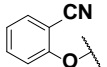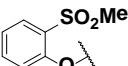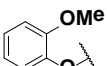
SAR Analysis			Structure		Target Activity		Anti-target Activity		
Entry No.	CID SID Broad ID	*	Ph-O $\diagup$ R$_1$	n	HMLE_sh_ECad EC$_{50}$ (µM)	n	HMLE_sh_GFP EC$_{50}$ (µM)	Selectivity (HMLE_sh_GFP/ HMLE_sh_ECad)	
6	2967498 110723036 BRD-K17874125	S		5	Inactive	5	Inactive	NA	
	Solubility (PBS): <1 µM							Purity >95%	
7	8882461 110722999 BRD-K65855402	S		5	Inactive	5	57.50	NA	
	Solubility (PBS): <1 µM							Purity >95%	
8	29369457 112208915 BRD-K12643219	S		7	Inactive	7	Inactive	NA	
	Solubility (PBS): <1 µM							Purity >95%	
9	50904165 110723022 BRD-K41590661	S		5	Inactive	5	Inactive	NA	
	Solubility (PBS): 61.9 µM							Purity >95%	
10	9214159 104170323 BRD-K24240386	S		5	Inactive	6	Inactive	NA	
	Solubility (PBS): 29.2 µM							Purity 88%	

SAR Analysis			Structure	Target Activity		Anti-target Activity		
Entry No.	CID SID Broad ID	*	Ph-O furan R$_1$ O	n	HMLE_sh_ECad EC$_{50}$ (µM)	n	HMLE_sh_GFP EC$_{50}$ (µM)	Selectivity (HMLE_sh_GFP/ HMLE_sh_ECad)
11	50910550 112208926 BRD-K98413806	S	N–Me / Me	7	Inactive	7	Inactive	NA
	Solubility (PBS): 355 µM							Purity >95%

NA = not applicable; *S = synthesized
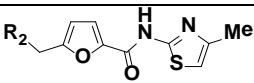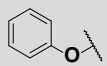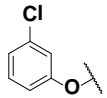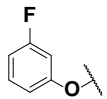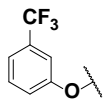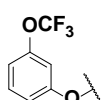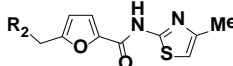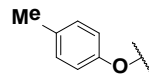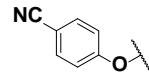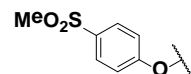
Table 3. SAR of the Phenyl Group (*ortho*-substitution).
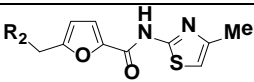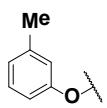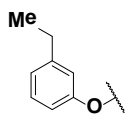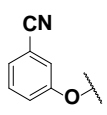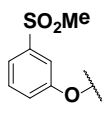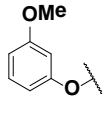
SAR Analysis			Structure		Target Activity		Anti-target Activity		
Entry No.	CID SID Broad ID	*	R_2—furan—C(O)—NH—thiazole—Me (R_2)	n[a]	HMLE_sh_ECad EC$_{50}$ (µM)	n	HMLE_sh_GFP EC$_{50}$ (µM)	Selectivity	
1	4791237 104170322 BRD-K39514895	S	phenyl-O	7	1.64	8	24.59	15.0	
	Solubility (PBS): 2.6 µM								Purity 91%
2	50904141 110722988 BRD-K63778830	S	HO	3	Inactive	3	Inactive	NA	
	Solubility (PBS): ND								Purity >95%
3	50904162 110723013 BRD-K53368809	S	Cl-phenyl-O	3	0.145	3	0.211	1.46	
	Solubility (PBS): <1 µM								Purity >95%
4	50904129 110722985 BRD-K31780183	S	F-phenyl-O	4	0.013	4	0.120	9.22	
	Solubility (PBS): <1 µM								Purity >95%
5	50904156 110722977 BRD-K47599362	S	CF$_3$-phenyl-O	3	0.170	3	0.051	0.30	
	Solubility (PBS): <1 µM								Purity 92%

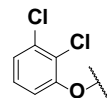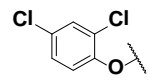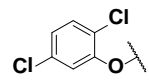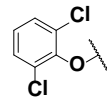| SAR Analysis			Structure	Target Activity		Anti-target Activity		
Entry No.	CID SID Broad ID	*	R_2 furan C(=O)NH-thiazole-Me	n^a	HMLE_sh_ECad EC_{50} (μM)	n	HMLE_sh_GFP EC_{50} (μM)	Selectivity
6	50904164 110723018 BRD-K90017492	S	OCF_3	5	0.547	5	0.750	1.37
	Solubility (PBS): <1 μM							Purity >95%
7	50910527 112208922 BRD-K64073927	S	Me	4	0.191	4	0.455	2.38
	Solubility (PBS): <1 μM							Purity >95%
8	50904167 110723006 BRD-K78153215	S	CN	3	1.43	3	3.56	2.49
	Solubility (PBS): <1 μM							Purity 90%
9	50910522 112208931 BRD-K20970172	S	SO_2Me	4	28.48	4	Inactive	28.48 vs. inactive
	Solubility (PBS): 19.6 μM							Purity >95%
10	50904146 110722994 BRD-K43005923	S	OMe	3	2.03	3	14.15	6.97
	Solubility (PBS): <1 μM							Purity >95%

ND = not determined; *S = synthesized
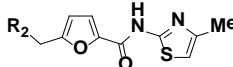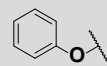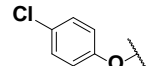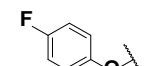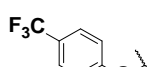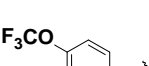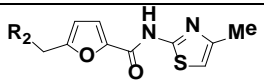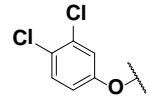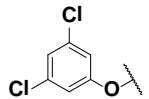
Table 4. SAR of the Phenyl Group (*meta*-substitution).
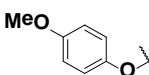
SAR Analysis			Structure		Target Activity		Anti-target Activity		
Entry No.	CID SID Broad ID	*	R_2 — (structure) R_2	n^a	HMLE_sh_ECad EC$_{50}$ (µM)	n	HMLE_sh_GFP EC$_{50}$ (µM)	Selectivity	

Entry No.	CID SID Broad ID	*	R$_2$	n^a	HMLE_sh_ECad EC$_{50}$ (µM)	n	HMLE_sh_GFP EC$_{50}$ (µM)	Selectivity
1	4791237 104170322 BRD-K39514895	S	(phenyl)	7	1.64	8	24.59	15.0
	Solubility (PBS): 2.6 µM							Purity 91%
2	50904134 110722981 BRD-K59019422	S	Cl	4	0.536	4	7.87	14.7
	Solubility (PBS): 0.07 µM							Purity >95%
3	50904145 110722990 BRD-K93748343	S	F	7	0.536	7	2.70	5.03
	Solubility (PBS): <1 µM							Purity >95%
4	50904153 110723010 BRD-K27111643	S	CF$_3$	3	6.78	3	39.07	5.76
	Solubility (PBS): <1 µM							Purity >95%
5	50904144 110723001 BRD-K77185929	S	OCF$_3$	6	0.217	6	1.96	9.03
	Solubility (PBS): <1 µM							Purity 92%

SAR Analysis			Structure		Target Activity		Anti-target Activity		
Entry No.	CID SID Broad ID	*	R_2—furan—C(O)—NH—thiazole—Me	n^a	HMLE_sh_ECad EC$_{50}$ (µM)	n	HMLE_sh_GFP EC$_{50}$ (µM)	Selectivity	
6	50910527 112208928 BRD-K64073927	S	Me (3-methylphenoxy)	4	0.464	2	3.19	6.86	
	Solubility (PBS): <1 µM							Purity >95%	
7	51003708 117687956 BRD-K66606544	S	Me-CH2 (phenoxy)	6	1.61	6	8.09	5.03	
	Solubility (PBS): <1 µM							Purity: >95%	
8	50904135 110722973 BRD-K91666083	S	CN (phenoxy)	3	3.64	3	19.4	5.33	
	Solubility (PBS): <1 µM							Purity >95%	
9	50910522 112208931 BRD-K62248673	S	SO$_2$Me (phenoxy)	4	Inactive	4	Inactive	NA	
	Solubility (PBS): 0.4 µM							Purity >95%	
10	50904149 110723014 BRD-K65643278	S	OMe (phenoxy)	4	6.70	4	20.05	3.00	
	Solubility (PBS): <1 µM							Purity 95%	

NA = not applicable; *S = synthesized
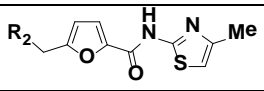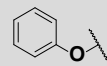
Table 5. SAR of the Phenyl Group (*para*-substitution).

Entry No.	CID SID Broad ID	*	R₂	n [a]	HMLE_sh_ECad EC₅₀ (µM)	n	HMLE_sh_GFP EC₅₀ (µM)	Selectivity
	SAR Analysis		**Structure**		**Target Activity**		**Anti-target Activity**	
1	4791237 104170322 BRD-K39514895	S		7	1.64	8	24.59	15.0
	Solubility (PBS): 2.6 µM							Purity 91%
2	35763151 110722989 BRD-K58889983	S		3	5.05	3	19.39	3.84
	Solubility (PBS): <1 µM							Purity >95%
3	26362833 110723005 BRD-K66460692	S		3	9.97	3	28.55	2.86
	Solubility (PBS): <1 µM							Purity >95%
4	50904147 110723009 BRD-K52642310	S		3	4.07	3	12.18	2.99
	Solubility (PBS): <1 µM							Purity 87%
5	50904142 110722983 BRD-K05208185	S		3	8.71	3	20.20	2.32
	Solubility (PBS): ND							Purity >95%

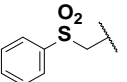| SAR Analysis			Structure	Target Activity		Anti-target Activity		
Entry No.	CID SID Broad ID	*	R_2 furan amide thiazole-Me structure	n^a	HMLE_sh_ECad EC_{50} (µM)	n	HMLE_sh_GFP EC_{50} (µM)	Selectivity
6	38046061 112208936 BRD-K60920175	S	Me—phenyl—O	4	2.27	4	21.2	9.35
	Solubility (PBS): <1 µM							Purity >95%
7	50904133 110722993 BRD-K49571296	S	NC—phenyl—O	5	14.04	5	Inactive	NA
	Solubility (PBS): <1 µM							Purity >95%
8	50910525 112208914 BRD-K26644939	S	MeO_2S—phenyl—O	7	Inactive	7	Inactive	NA
	Solubility (PBS): 3.3 µM							Purity >95%
9	35763156 110723002 BRD-K51022414	S	MeO—phenyl—O	3	38.93	2	Inactive	38.93 vs. inactive
	Solubility (PBS): <1 µM							Purity >95%

NA = not applicable; ND = not determined; *S = synthesized
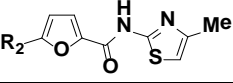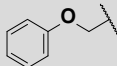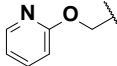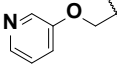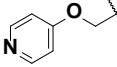
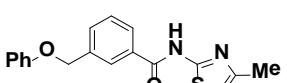**Table 6. SAR of the Phenyl Group (*di*-substitution).**

SAR Analysis			Structure	Target Activity		Anti-target Activity		
Entry No.	CID SID Broad ID	*	R_2 structure (with core: 5-R_2-furan-2-carboxamide of N-(4-Me-thiazol-2-yl))	n^a	HMLE_sh_ECad EC$_{50}$ (µM)	n	HMLE_sh_GFP EC$_{50}$ (µM)	Selectivity
1	4791237 104170322 BRD-K39514895	S	phenyl-O	7	1.64	8	24.59	15.0
	Solubility (PBS): 2.6 µM							Purity 91%
2	51003713 117687928 BRD-K25232195	S	2,3-dichlorophenyl-O	5	0.127	5	0.186	1.46
	Solubility (PBS): <1 µM							Purity: >95%
3	1977298 124360004 BRD-K71104906	S	2,4-dichlorophenyl-O	5	0.260	5	0.750	2.89
	Solubility (PBS): ND							Purity: >95%
4	50944059 115950043 BRD-K35890657	S	2,5-dichlorophenyl-O	7	0.032	7	0.096	3.00
	Solubility (PBS): <1 µM							Purity >95%
5	51003712 117687952 BRD-K18254459	S	2,6-dichlorophenyl-O	6	4.34	6	10.53	2.43
	Solubility (PBS): <1 µM							Purity: >95%

SAR Analysis			Structure		Target Activity		Anti-target Activity		
Entry No.	CID SID Broad ID	*	R_2—furan—C(=O)—NH—thiazole(S,N)—Me	n^a	HMLE_sh_ECad EC_{50} (µM)	n	HMLE_sh_GFP EC_{50} (µM)	Selectivity	
6	51003702 124360022 BRD-K16234721	S	(3,4-dichlorophenoxymethyl)	5	0.388	5	0.126	0.32	
	Solubility (PBS): ND							Purity: >95%	
7	51003715 117687942 BRD-K51193239	S	(3,5-dichlorophenoxymethyl)	6	0.312	6	5.25	16.8	
	Solubility (PBS): <1 µM							Purity: >95%	
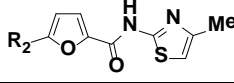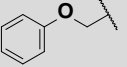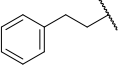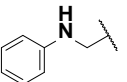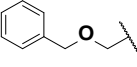
ND = not determined;*S = synthesized
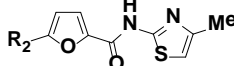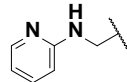
Table 7. SAR of Phenyl Replacement.

SAR Analysis			Structure		Target Activity		Anti-target Activity		
Entry No.	CID SID Broad ID	*	R_2	n^a	HMLE_sh_ECad EC$_{50}$ (μM)	n	HMLE_sh_GFP EC$_{50}$ (μM)	Selectivity	
1	4791237 104170322 BRD-K39514895	S	(phenyl-O structure)	7	1.64	8	24.59	15.0	
	Solubility (PBS): 2.6 μM						Purity 91%		
2	50910552 112208918 BRD-K28966749	S	(pyridin-2-yl-O structure)	7	13.6	7	Inactive	13.6 vs. inactive	
	Solubility (PBS): 36.3 μM						Purity >95%		
3	51351633 121269754 BRD-K21379016	S	(pyridin-3-yl-O structure)	3	Inactive	3	Inactive	NA	
	Solubility (PBS): 247 μM						Purity: >95%		
4	51351622 121269749 BRD-K51223627	S	(pyridin-4-yl-O structure)	6	Inactive	6	Inactive	NA	
	Solubility (PBS): ND						Purity: 89%		

NA = not applicable; ND = not determined; *S = synthesized
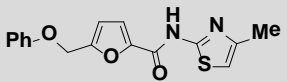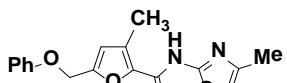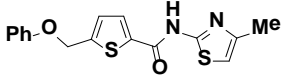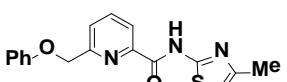
Table 8. SAR of the Linker.

SAR Analysis			Structure		Target Activity	Anti-target Activity		
Entry No.	CID SID Broad ID	*	R_2	n [a]	HMLE_sh_ECad EC$_{50}$ (µM)	n	HMLE_sh_GFP EC$_{50}$ (µM)	Selectivity
1	4791237 104170322 BRD-K39514895	S	(phenoxy structure)	7	1.64	8	24.59	15.0
	Solubility (PBS): 2.6 µM							Purity 91%
2	50910539 112208925 BRD-K08255553	S	(phenethyl structure)	7	1.84	7	17.06	9.27
	Solubility (PBS): <1 µM							Purity >95%
3	50910533 112208906 BRD-K90478238	S	(anilino structure)	7	0.481	8	2.99	6.22
	Solubility (PBS): 2.0 µM							Purity >95%
4	41548804 112208921 BRD-K46740330	S	(sulfonyl structure)	8	Inactive	8	Inactive	NA
	Solubility (PBS): <1 µM							Purity >95%
5	50910536 112208901 BRD-K02605476	S	(benzyloxy structure)	4	5.48	4	15.08	2.75
	Solubility (PBS): <1 µM							Purity >95%

SAR Analysis			Structure		Target Activity		Anti-target Activity		
Entry No.	CID SID Broad ID	*	R_2-furan-C(=O)-NH-thiazole-Me	n^a	HMLE_sh_ECad EC_{50} (µM)	n	HMLE_sh_GFP EC_{50} (µM)	Selectivity	
6	50944065 115950042 BRD-K51782431	S		7	9.01	7	Inactive	9.01 vs. inactive	
	Solubility (PBS): 266 µM							Purity: >95%	
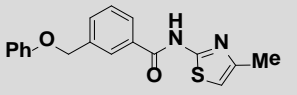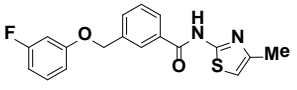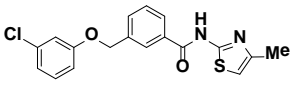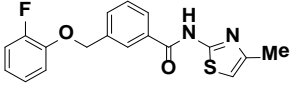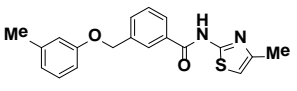
NA = not applicable; *S = synthesized

Table 9. Replacement of the Furan.

Entry No.	CID SID Broad ID	*	Structure**	n^a	HMLE_sh_ECad EC$_{50}$ (µM)	n	HMLE_sh_GFP EC$_{50}$ (µM)	Selectivity
	SAR Analysis		**Structure****		**Target Activity**		**Anti-target Activity**	
1	4791237 104170322 BRD-K39514895	S		7	1.64	8	24.59	15.0
	Solubility (PBS): 2.6 µM							Purity 91%
2	51003686 124360008 BRD-K46391914	S		5	6.41	5	20.87	3.26
	Solubility (PBS): ND							Purity: 92%
3	51003688 124360018 BRD-K04437141	S		5	0.424	5	0.089	0.21
	Solubility (PBS): ND							Purity: >95%
4	51003701 117687945 BRD-K20316594	S		3	4.76	3	16.52	3.47
	Solubility (PBS): <1 µM							Purity: >95%
5	51003726 117687932 BRD-K91131025	S			Inactive		Inactive	NA
	Solubility (PBS): <1 µM							Purity: >95%

NA = not applicable; *S = synthesized; **No general structure as whole structure for each compound is displayed.

Table 10. SAR on Furan Replacement (Phenyl).

	SAR Analysis		Structure[**]	n [a]	Target Activity	n	Anti-target Activity	
Entry No.	CID SID Broad ID	*			HMLE_sh_ECad EC_{50} (µM)		HMLE_sh_GFP EC_{50} (µM)	Selectivity
1	51003701 117687945 BRD-K20316594	S		3	4.76	3	16.52	3.47
	Solubility (PBS): <1 µM							Purity: >95%
2	51351634 121269761 BRD-K51453538	S		6	0.481	6	2.50	5.19
	Solubility (PBS): <1 µM							Purity: 94%
3	51351627 121269777 BRD-K56362326	S		6	1.41	6	9.51	6.74
	Solubility (PBS): <1 µM							Purity: 92%
4	51351617 121269770 BRD-K28820698	S		3	1.82	3	12.67	6.96
	Solubility (PBS): <1 µM							Purity: >95%
5	51351618 121269753 BRD-K44916431	S		6	2.50	6	17.33	6.93
	Solubility (PBS): <1 µM							Purity: >95%

*S = synthesized; **No general structure as whole structure for each compound is displayed.

In summary, 112 synthesized compounds were tested in both primary (AID 493226, AID 493176, AID 504449, and AID 504535) and secondary assays (AID 504325, AID 493196, AID 504450, AID 504450, and AID 504533).

3.5   Cellular Activity

To confirm results obtained by the probe Screening Center, several additional experiments were conducted by the Assay Provider. The probe (CID 50910523/ML245) and five additional compounds (CID 4791237, CID 9214159, CID 50904129, CID 50904149, and CID 50944065) were investigated for their effects on viability in the HMLE_sh_ECad and HMLE_sh_GFP cell lines. Although varying absolute IC_{50} values were obtained from the collaborators, ML245 had the greatest selectivity of all the compounds tested.

Two additional transformed cell lines, HMLE_sh_Twist and MDA231, were used to corroborate the findings. HMLE_sh_Twist, a second upstream model of EMT-induced breast CSC-like cells, was also significantly inhibited by the probe (ML245) (**Figure 5, Table 11**). Although ML245 is toxic to HMLE_Twist, the selectivity was only approximately 8-fold. This finding supports that the compound is not specific to the generated cell line HMLE_sh_ECad and can inhibit other breast CSC-like cells.

Furthermore, activity of ML245 and five analogs were tested against a breast cancer cell line (MDA231; **Table 11**). All of the compounds only weakly targeted the MDA231 cell line, which has been characterized as a mixed population of breast CSCs and more differentiated cells. This may explain the reduced potency since the probe would not be toxic to more differentiated cells. Additional testing of cancer cell lines will be required to see if this scaffold can inhibit other breast cancer lines. Since ML245 is selective for breast CSCs, any further development towards drug discovery should be considered in conjunction with chemotherapeutic compounds that target the differentiated cells in the tumor.

Figure 5. Dose Inhibition Assay of HMLE_sh_GFP, HMLE_sh_ECad, and HMLE_sh_Twist by the Probe (ML245).
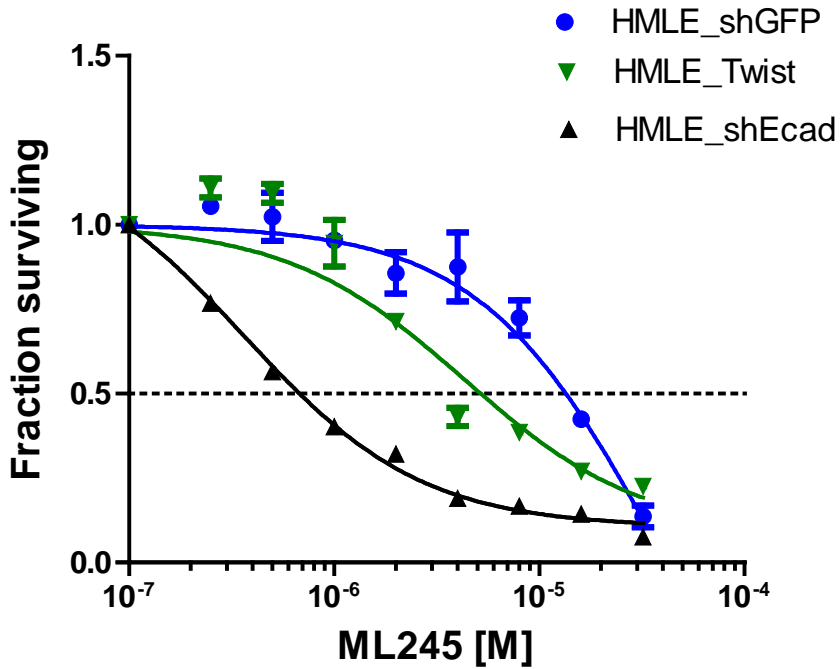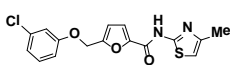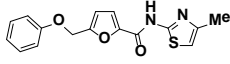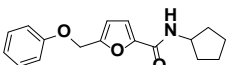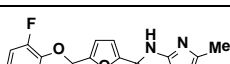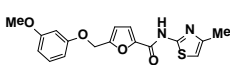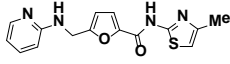
Figure 5. Dose curve data provided by the Gupta laboratory showing probe (ML245) inhibition of HMLE_sh_GFP (blue●), HMLE_sh_ECad (black▲), and HMLE_sh_Twist (green▼).

Table 11. Activity Data of HMLE_sh_GFP, HMLE_sh_ECad, and HMLE_sh_Twist by ML245 and Five Analogs.

CID SID Broad ID	Structure	HMLE_sh_ECad EC$_{50}$ (µM)	HMLE_sh_TWIST EC$_{50}$ (µM)	HMLE_sh_GFP EC$_{50}$ (µM)	MDA231 EC$_{50}$ (µM)	Selectivity (HMLE_sh_ECad vs. HMLE_sh_GFP)	Tumor-sphere assay
50904134 110722981 BRD-K59019422		0.625	4.3	37.7	262	60-fold	IA
4791237 104170322 BRD-K39514895		2.435	18.22	100	IA	41-fold	INC
9214159 104170323 BRD-K24240386		800	580	IA	IA	800 vs IA	IA
50904129 110722985 BRD-K31780183		0.07	0.13	0.36	780	5.1-fold	IA
50904149 110723014 BRD-K65643278		0.3	1.7	4.4	0.48	15-fold	ND
50944065 115950042 BRD-K51782431		IA	600	IA	62	IA	ND

IA= inactive; INC=inconclusive; ND=not determined

In addition, ML245 and three additional analogs (CID 4791237, CID 9214159, CID 50904129) were tested in a tumorsphere assay (AID 504623; **Table 11**). In this assay, Sum159 cells (a human breast cancer cell line) were grown in the presence or absence of compound for 9 days in a low adhesion environment. Under these conditions, Sum159 cells, which contain a mixture of breast CSCs and differentiated breast cancer cells, form tumorspheres. The number of tumorspheres was counted in each of the conditions. The signal (number of tumorspheres) was normalized to neutral (DMSO) and positive (Puromycin) control, and a 30% inhibition cut-off at an average screening concentration of 20 µM was used to define a hit.

Neither the probe (ML245) nor any of the analogs significantly and consistently inhibited tumorsphere formation. In previous scaffolds, nonselective analogs (CID 24816775, CID 16189574, and CID 6262187) that were toxic to all cell lines did inhibit tumorsphere formation (**Table 11)**. In this series, CID 50904129, which is similarly toxic/nonselective, did not similarly inhibit tumorsphere formation. It is possible that the viscous methylcellulose/media used to keep tumorspheres from growing together also inhibited this compound series from inhibiting tumorsphere formation. Another possibility is that these compounds are not effective against SUM159 cells. Further work needs to be done to determine if these compounds can inhibit tumor formation.

Although the probe did not inhibit tumor formation, previous probes ML239 and ML243 also did not inhibit tumor formation. Salinomycin, the only other small molecule that is known to selectively inhibit HMLE_sh_ECad also does not inhibit SUM159 tumorsphere formation(8). Additional assays and cell lines will be required to address these issues. Taken together, the probe (ML245) clearly meets the established probe criteria in whole cells specified for this project (**Figure 2, Figure 4,** and **Table 12**); experimental details are provided in Section 2.1.2**.**

Table 12. Comparison of the Probe (ML245) to Project Criteria

No.	Property	CPDP Requirement	Probe
1	Potency	EC_{50} HMLE_sh_ECad < 10 µM	EC_{50} HMLE_sh_ECad = 0.536 µM
2	Selectivity	EC_{50} HMLE_sh_GFP < 10 x IC_{50} HMLE_sh_ECad	EC_{50} HMLE_sh_GFP < 14 x IC_{50} HMLE_sh_ECad

3.6 Profiling Assays

To gain insight into the possible mechanism or signaling pathways that the probe (ML245) could be targeting, gene expression studies were conducted. HMLE_sh_ECad and HMLE_sh_GFP cells were treated with the probe (ML245), analog CID50904149, or salinomycin for 24 hours at the IC_{50} identified after 3-day treatment in the primary assay (0.5 µM, 6 µM, and 1 µM, respectively). In addition, control samples were included where DMSO was added for 24 hours. cRNA synthesis from the total RNA samples were analyzed on a Human HT-12 Expression BeadChip (Illumina). All samples passed QC standards for gene expression and hybridization.

The samples were normalized, and gene expression of the triplicate samples was compared using the Comparative Marker Selection Module in GenePattern (Open source: http://genepattern.broadinstitute.org/). Gene expression that had p ≤ 0.005 and a change at least 150% (± 50%) change in expression was selected. First, the control conditions between the cell lines were compared and 624 genes were differentially expressed. Comparable results were observed previously (11). Moreover, many of the genes confirmed the results observed by the Weinberg laboratory using these cell lines (12). Specifically, of the 19 genes indicated by Onder *et al.* as statistically significantly different, 14 repeated in our assay (11, 12). Moreover, Onder *et al.* identified several proteins with altered expression, including vimentin, beta-catenin, gamma-catenin, CK-8, and fibronectin (12). We observed similar changes in our gene expression assay, except for fibronectin, where we did not observe any change in gene expression. However, gene expression does not always correlate with alterations in protein expression. Further characterization is required to determine if those results varied.

Next, we compared differences in gene expression after 24-hour treatment with ML245, analog CID 50904149 (**Table 4, Entry 10**), or salinomycin (CID 3085092) in HMLE_sh_ECad or HMLE_sh_GFP (**Figure 6 and Figure 7**) to yield a better understanding of genes and/or pathways regulated by the breast CSC-like selective compounds. Using the same criteria as above, 349 genes were identified as changing when comparing RNA-isolated, DMSO-treated versus ML245-treated HMLE_sh_ECad cells (**Figures 6A, 6C, 6D**). However, only two genes (CYP1B1 and DHRS3) had altered expression by the same criteria after ML245 treatment in the control line, HMLE_sh_GFP (**Figure 6D**). This strongly suggests that ML245 is highly selective for the HMLE_sh_ECad cells.

To gain a better insight to which genes and/or pathways are affected by this scaffold, both the CSC-like cell line (HMLE_sh_ECad) and control (HMLE_sh_GFP) cell lines were treated with analog CID 50904149. After treatment with CID 50904149, 238 genes were alternately expressed in HMLE_sh_ECad, while 160 genes were altered in HMLE_sh_GFP (**Figures 6B, 6C, 6D**). The higher number of genes affected in HMLE_sh_GFP cell line may be expected since CID 50904149 is less selective than ML245 (3-fold versus 14.7-fold, respectively). More critically, when comparing the genes altered by ML245 and CID 50904149 in the HMLE_sh_ECad line, 84 genes overlapped in both groups (**Figure 6A-D**).

To compare our compounds with the prior art (i.e., salinomycin), we conducted parallel studies with ML245 (**Figure 7**). Again, we compared salinomycin-treated cells to DMSO-treated cells using the same criteria as above. Salinomycin altered the expression of 683 genes in the CSC-like cell line. Interestingly, salinomycin had a greater impact on the control line (HMLE_sh_GFP) with 1021 genes having differential gene expression (**Figure 6**). There were 84 genes with altered expression by salinomycin that were also affected by ML245 (**Figure 6D**). Interestingly, *all* of the 84 genes identified when comparing the effects of ML245 and CID5094149 were identified again as having altered gene expression with treatment of salinomycin. were the same 84 genes found when comparing ML245 and CID 5094149 (**Figure 6** and **Figure 7**). These 84 genes suggest ML245, CID 50904149, and salinomycin could be killing CSCs via a similar mechanism.

These 84 genes were, therefore, prioritized as genes of interest. A portion of the genes were transcription factors or DNA-modifying enzymes. At least 12 of the 84 fell into this category, including five zinc finger-containing genes (EGR1, ZCCHC7, ZNF295, CGRRF1), leucine zipper proteins (LZTFL1, TSC22D3), several additional transcription factors (ATF4, LMO4, CEBPG, NUPR1, TAF1D), and DNA-modifying enzymes such as GADD45A. Taken together, these 12 genes do not represent a clear signaling pathway; however, it may be possible that the compound directly targets nuclear protein gene set.

One particular gene of interest is the RING finger protein CGRRF1, which is a protein that determines cell-cycle arrest. An increased expression of CGRRF1 correlated with a decreased division of Burkitt lymphoma cell lines (1), and a potential way to stop the proliferation of lymphomas. The probe (ML245), CID 5094149, and salinomycin increased the expression of this particular gene in our gene expression studies. Therefore, this gene will be a prioritized gene of interest.

Furthermore, several genes included in the 84 prioritized genes are involved in apoptotic signaling and mitochondrial maintenance. These include caspase recruitment domain family member 10 (CARD10), Activating Transcription Factor 4 (ATF4) (13), ATPase subunit C targeting peptide (ATP5G1) (14), cytochrome P450, family 20 (CYP20A1) (15), mitochondrial ribosomal protein (MRPL12) (16). Since the expression of these five genes was altered after treatment with CSC-selective compounds, the mechanism of selective death may be increasing

pro-apoptotic signaling. Future studies will be conducted to address the role of the probe (ML245) role in DNA regulation and apoptotic signaling.

Figure 6.   Heat Map Comparisons of Differential Gene Expression by ML245 and CID50904149.
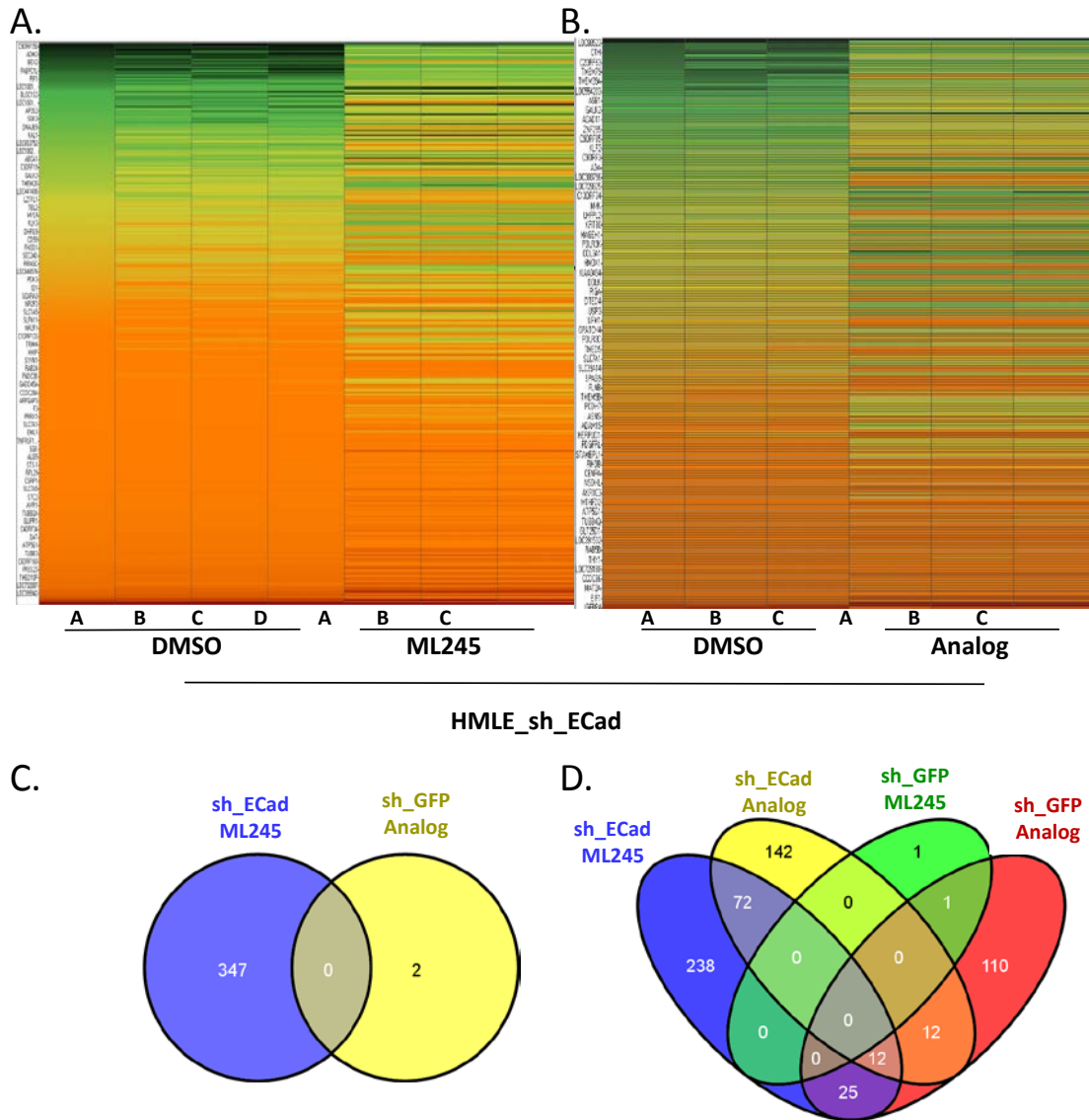
Figure 6. Genes that were significantly differentially expressed (p <0.005) and had greater than 1.5-fold difference in expression were identified. Heat map representations of gene expression changes HMLE_shE_Cad (sh_ECad) cells were treated with DMSO vs. ML245 (**A**) or DMSO vs. CID50904149 (Analog; **Table 4**, **Entry 10**) (**B**) are shown. 349 and 238 genes were found to be alternately regulated by ML245 or CID 50904149, respectively. Of those genes, 84 overlapped in both conditions as represented in a Venn diagram (**C, D**). Gene comparisons were also made between DMSO vs. ML245 or DMSO vs. CID 50904149 (Analog) in HMLE_sh_GFP (sh_GFP) cell lines (**D**).

Figure 7.　Heat Map Comparisons of Differential Gene Expression by Salinomycin

A.

| A | B | C | D | A | B | C | D |
| DMSO | | | | Salinomycin | | | |

HMLE_shE_Cad

B.

| A | B | C | A | B | C | D |
| DMSO | | | Salinomycin | | | |

HMLE_sh_GFP

C.

shECad_Sal　　shGFP_Sal

421　262　759

D.

sh_ECad_ML245　　sh_ECad_Analog

160　0　4

84

103　150

489
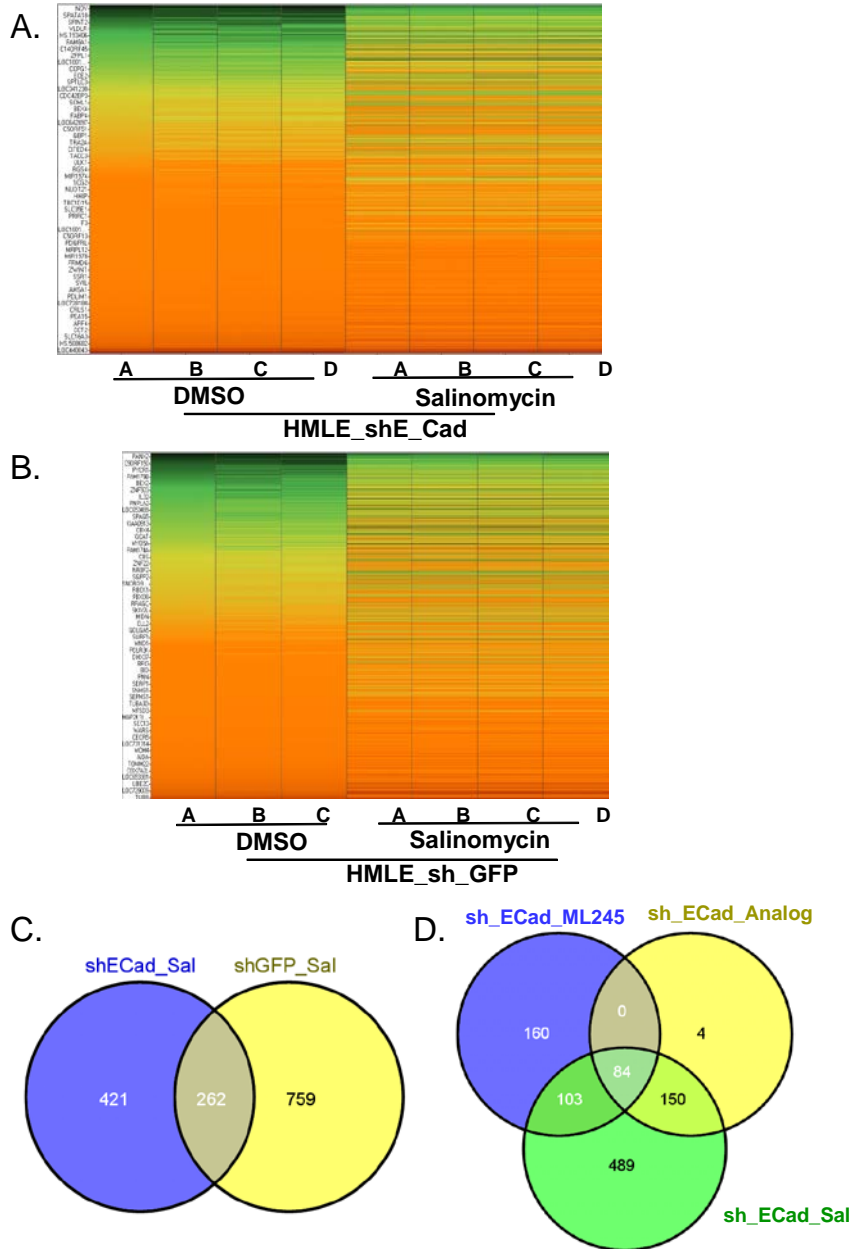
sh_ECad_Sal

Figure 7. Gene expression profile of prior art, salinomycin. HMLE_sh_ECad (shECad) (**A**) and HMLE_sh_GFP (sh_GFP) (**B**) were treated with DMSO or salinomycin (Sal). Genes that had significantly different expression (p≤ 0.005) and had at least 1.5-fold change in expression were plotted in a heat map. After Sal treatment, 683 and 1,021 genes were altered in HMLE_sh_ECad or HMLE_sh_GFP line, respectively. Of those, 262 genes overlapping were identified in both conditions (**C**). Comparing HMLE_sh_ECad selective compounds ML245, salinomycin, and CID 50904149 (Analog), 84 genes were identified to be regulated by all three compounds.

The probe (ML245) was screened against a panel of 68 targets in radioligand binding assays that are commonly used in drug discovery for lead profiling. The assays were done by Ricerca Biosciences, LLC and include targets from various areas such as, GPCRs, ion channels, and transporters. The probe (ML245) was active in eight of the 68 assays and displayed inhibition of adenosine A_1, A_{2A}, and A_3 receptors at 79%, 88%, and 94% at 10 µM, respectively. The probe (ML245) also displayed inhibition against two transporters, Norepinephrine (NET) and Dopamine (DAT) at 98% and 101%, respectively. The other three targets that the probe (ML245) inhibited were the $GABA_A$ (93%), Opiate κ (87%), and serotonin (5-Hydroxytryptamine) $5\text{-}HT_{2B}$ (78%) receptors. These assays only assess binding to the above mentioned targets and the functional consequence of that binding will need to be assessed.

4   Discussion

Beginning from the hit from HTS (**Table 2, Entry 1**), we investigated the SAR of three regions of the molecule through the synthesis of 110 compounds (**Figure 8**), the eastern amide portion, the western aryl ring, and the furan moiety. Variation from the 2-methyl thiazole amide or modification of the furan was not tolerated. Through SAR of the phenyl ring, it was found that substitution at the 3-position was optimal for activity and selectivity. The probe (ML245, **Table 4, Entry 2**) was identified as much more potent than the hit compound.

Figure 8.   SAR Summary

Mono and di- substituted analogs
60 compounds evaluated
meta-chloro substituent is optimal

Thiazole analogs
Other heterocycles
26 compounds evaluated
Methyl thiazole is optimal

Furan Replacement and linker Modification
24 compounds evaluated
Furan with one carbon linker is optimal
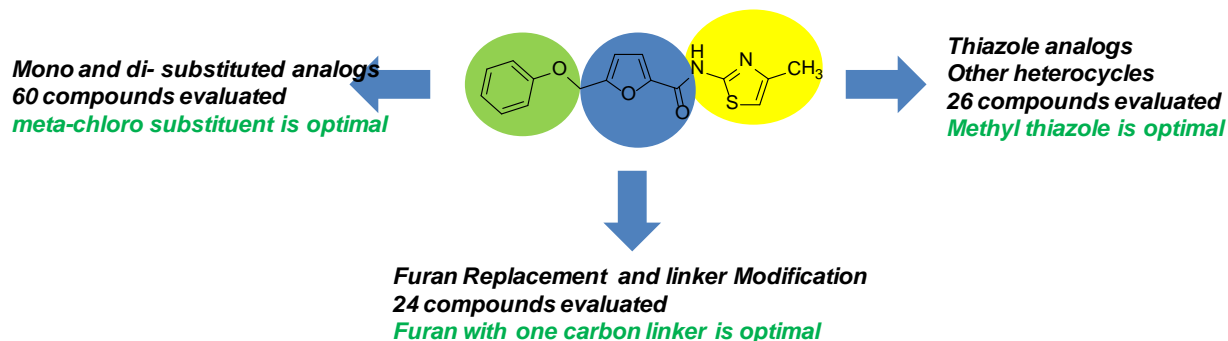
Overall, the data supports the identification of a potent (i.e., 0.536 µM) and selective (greater than 14-fold) probe (ML245) that differentially targets the breast CSC-like cells (see **Figure 3** and **Figure 5**; **Table 11**). In addition to the assays outlined in the CPDP, we have conducted gene expression profiling studies to gain insight into the probe's mechanism of action.

Examination of the HMLE_sh_GFP control cell line did not identify any genes that were differentially regulated after treatment with the probe (ML245) for 24 hours. After treatment of the HMLE_sh_ECad line for 24 hours, 558 genes were identified as having greater than 150% difference in expression.

Interestingly, a significant number of genes affected by ML245 are transcription factors or DNA-modifying enzymes including zinc finger-containing genes (ZFAND2A, ZCCHC7, ZNF295, CGRRF1, TRIM39), leucine zipper (LZTFL1, TSC22D3), and several additional transcription factors (ATF3, ATF4, LMO4, CEBPG). Although the mechanism is unclear, there is potential that ML245 is selectively targeting nuclear proteins.

4.1   Comparison to Existing Art and How the New Probe is an Improvement

The current state-of-the-art for the probe molecule salinomycin exhibits a potency of about 1 µM against breast CSCs (8); however, it should be noted that salinomycin shows toxicity against the control cell line (HMLE_sh_GFP) at 10 µM. Whereas the new probe (ML245) exhibits a similar potency at 0.536 µM, this probe is more selective and more attractive structurally allowing for rapid analog synthesis.

In the investigation of prior art relevant to the breast CSC project, the following databases were used: SciFinder, Entrez, PubMed, PubChem, and PatentLens. The search terms applied and hit statistics for the prior art search are provided in **Table A2 (Appendix G)**. Abstracts were obtained for all references returned and were analyzed for relevance to the current project. The searches were performed on, and are current as of, September 22, 2011.

Literature and patent searches revealed nine small molecules capable of selectively inhibiting breast CSCs (**Figure 9**). It is important to note that salinomycin is the only small molecule shown to be effective in the cell line HMLE_sh_ECad used in the primary assay. The effectiveness of salinomycin in HMLE_sh_ECad cells was determined by the Assay Provider, in which they developed an HTS screen to discover compounds that were selectively toxic for breast CSCs. Salinomycin was shown to provide a reduction in the proportion of CSCs with an IC_{50} value of approximately 1 µM against HMLE_sh_ECad with approximately 10-fold selectivity over the control HMLE_sh_GFP (8). In addition, *ex vivo* and *in vivo* treatment of mice with intraperitoneal injection of salinomycin inhibited mammary tumor growth and increased

differentiation of epithelial tumor cells. To date, this is the only report of selectively active compounds against the HMLE_sh_ECad cell line.

Figure 9.   Known Small Molecules Inhibitors of Breast Cancer Stem Cells

Repertaxin (CXCR1 inhibitor)

Parthenolide (NF⁻κβ inhibitor)

8⁻Quinolinol (NF⁻κβ inhibitor)

Sulforaphane

Metformin

Berberine

Salinomycin

Piperine (Wnt Inhibitor)

Curcumin (Wnt Inhibitor)
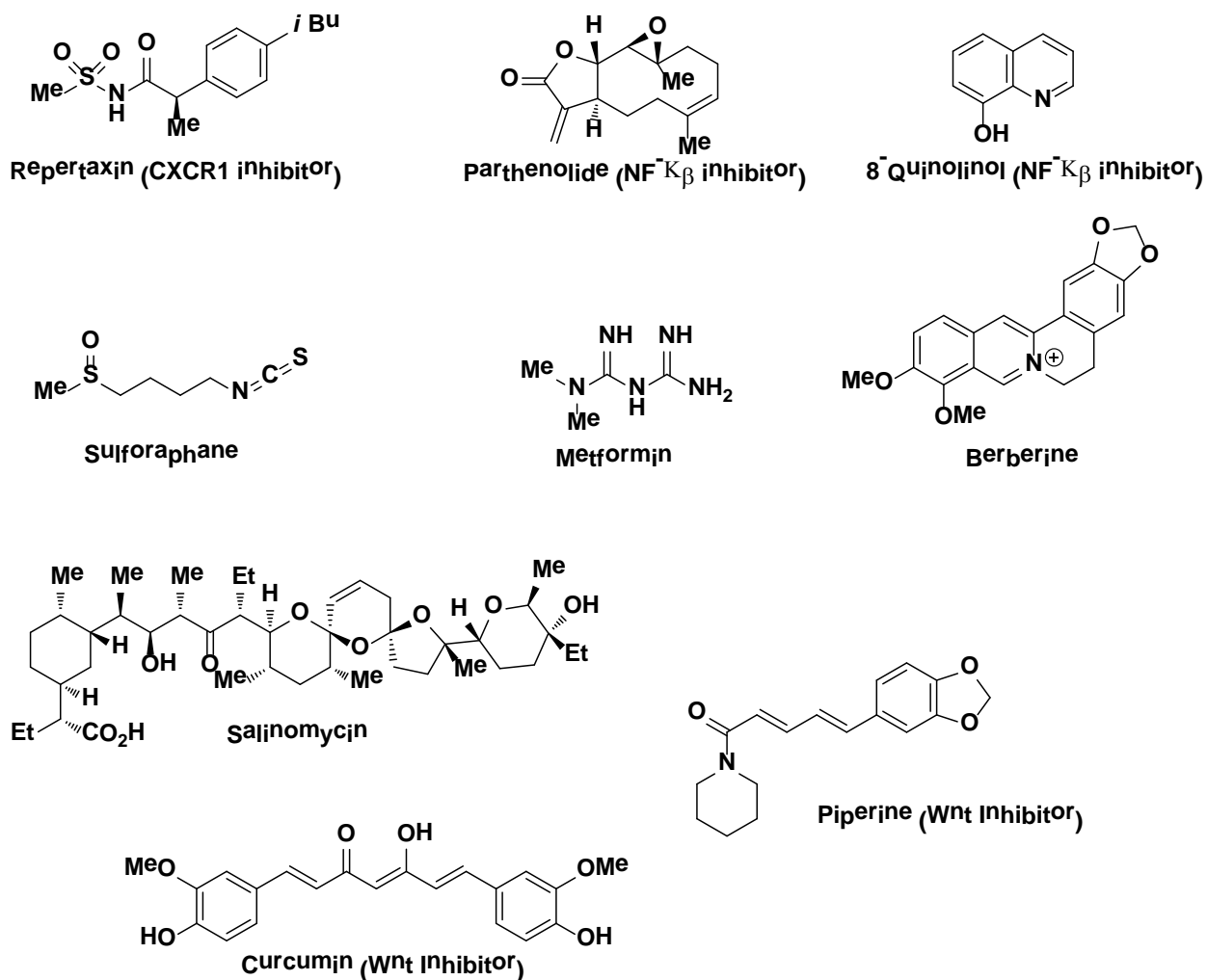
There has been a recent increase in the research of selective killing of breast CSCs. Though interesting, the cells used in this report are not susceptible to these agents. In one approach, researchers targeted the breast CSC pathways involved in self-renewal and survival. These pathways include NF-κB, Notch, Hedgehog, and Wnt. The NF-κB inhibitors parthenolide and 8-quinolinol (8Q) were reported by Zhou et al. to inhibit breast CSC-dependent MCF7 sphere cell growth selectively over bulk MCF7 cell growth (17). They reported that parthenolide and

pyrrolidinedithiocarbamate inhibited sphere MCF7 cell growth by 33% and 51%, respectively, and showed no obvious growth inhibition of bulk MCF7 cells. This work was followed by a 2009 publication, in which Zhou et al. reported that 8Q inhibited the sphere cell proliferation by 86% at 5 μM and inhibited the bulk cells by 30% (18). 8Q also showed some antitumor activity in both MCF7 and MDAMB-435 (breast cancer cell line) xenograft models, but it had a much better effect when combined with paclitaxel than either agent administered alone. Kakarala et al. reported that curcumin and piperine, respectively, inhibited Wnt signaling, ALDH+ cells (breast CSC biomarker) and mammosphere formation by 50% at 5 μM and completely at 10 μM (19). In 2010, Li et al. showed that sulforaphane down-regulated the Wnt-pathway and could inhibit ALDH+ cells *in vitro* by 65%-80% at 1-5 μM, but had minimal effect on the population of bulk breast cancer cell lines. In addition, sulforaphane inhibited mammosphere formation and the cancer cells from sulforaphane-treated mice failed to produce any tumors in recipient mice up to 33 days after implantations (20).
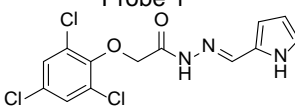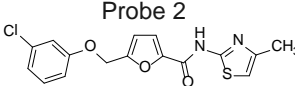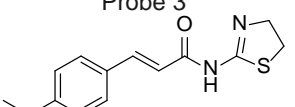
Metformin, a known oral hypoglycemic agent, was reported to selectively kill breast CSCs through an unknown mechanism. In this study, the combination of metformin and doxorubicin had a synergistic effect in reducing tumor mass and preventing relapse in xenograft mice (20). In addition, Jiralerspong et al. reported a study of 2529 patients undergoing neoadjuvant chemotherapy, in which the pathologic complete response rate following surgery was 24% for diabetic patients taking metformin, 8% for diabetic patients not taking metformin, and 16% for nondiabetic patients (21). Furthermore, using both CXCR1-blocking antibodies and the small molecule inhibitor, repertaxin, demonstrated that CXCR1 selectively decreased breast CSCs *in vitro* and in NOD/SCID xenograft models, respectively (20).

Berberine was investigated to determine its effects on the enriched stem cell-like population of MCF-7 cells. After 72 hours, the viability of MCF-7 cells was reduced to 96%, 84%, 79%, 65%, and 56%, and the stem cell-like population was reduced to 1.4% (1.5% of MCF7 cells were the stem cell-like population cells), 1.1%, 0.9%, 0.5%, and 0.2% by 10 nM, 100 nM, 1 μM, 10 μM, and 20 μM of berberine, respectively (22).

All nine compounds identified through this project are registered with the NIH Molecular Libraries, and the respective compound information is provided in **Table A3 (Appendix G).**

To date, we have identified 3 probe compounds, ML239 (Probe 1), ML245 (Probe 2), and ML243 (Probe 3). The three probes are potent inhibitors (0.5- 2uM) of the Breast Cancer Stem Cell-like cells, sufficiently selective (14.7-32 fold) against HMLE_sh_ECad, and meet the probe criteria. ML245 is the most potent against the primary target cell line, while ML243 has the greatest selectivity compared to HMLE_sh_ECad (see **Table 13**).

Table 13.   Comparison of Breast Cancer Stem Cells: Probes 1, 2, and 3

CID/ML No./Probe #	Target Name	EC$_{50}$ (µM) [SID, AID]	Anti-target	EC$_{50}$ (µM) [SID, AID]	Fold Selective	Secondary Assay Name: EC$_{50}$ (nM) [SID, AID]
CID 49843203/ ML239/ Probe 1	Breast Cancer Stem Cell-like cell	1.18 µM [SID 104170348, AID 493226]	Mammary Epithelial Cell	27.6 [SID 104170348, AID 504325]	23.4-fold	tumorsphere assay No activity [SID 114279607, AID 504623]
CID 50904134/ ML245/ Probe 2	Breast Cancer Stem Cell-like cell	0.536 [SID 110722981, AID 504449]	Mammary Epithelial Cell	7.87 [SID110722 981, AID504450]	14.7	Tumorsphere, single point, Inactive [SID 10722981, AID 504859]
CID 50910523/ ML243/ Probe 3	Breast Cancer Stem Cell-like cell	2.0 µM [SID 112208912, AID 504535]	Mammary Epithelial Cell	63.98 µM [SID 112208912, AID 504533]	32	Tumorsphere, single point, Inactive [SID 112208912, AID 504859]

4.2   Mechanism of Action Studies

We have included gene expression assays that indicate pathways that are regulated by addition of the probe (ML245). Future studies will be required to determine the target of the probe (ML245).

4.3 Planned Future Studies

To date, the probe (ML245) has been used in gene expression profiling studies, where HMLE_sh_ECad and HMLE_sh_GFP have been treated for 24 hours (see **Figure 6**). Treatment of the breast CSC-like cell line (HMLE_sh_ECad) by ML245 yielded interesting results identifying genes in pro-apoptotic signaling pathways as well as numerous transcription factors as potential targets. Furthermore, numerous transcription factor genes were regulated in these studies, suggesting that ML245 either directly or indirectly regulates transcription.

Although these initial gene profiling experiments suggest that these pathways are critical to the probe's selective killing of breast CSCs, follow-up studies are required to confirm these results. Since we have identified two additional probes (**Table 13**) that confer selective killing of the HMLE_sh_ECad cell line, we plan to revisit the gene expression profiling studies to determine if the combined data sets provides greater insight into critical signaling pathways and potential targets within breast CSCs.

5   References

1.   De Falco G, Leucci E, Lenze D, Piccaluga PP, Claudio PP, Onnis A, Cerino G, Nyagol J, Mwanda W, Bellan C, Hummel M, Pileri S, Tosi P, Stein H, Giordano A, Leoncini L. Gene-expression analysis identifies novel RBL2/p130 target genes in endemic Burkitt lymphoma cell lines and primary tumors. Blood. 2007 Aug 15;110(4):1301-7. Epub 2007 May 7. PMID:17485552

2.   Bao S, Wu Q, McLendon RE, Hao Y, Shi Q, Hjelmeland AB, Dewhirst MW, Bigner DD, Rich JN. Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. Nature 2006 Dec 7;444(7120):756-60. Epub 2006 Oct 18. PubMed PMID: 17051156

3.   Diehn M, Cho RW, Lobo NA, Kalisky T, Dorie MJ, Kulp AN, Qian D, Lam JS, Ailles LE, Wong M, et al. Association of reactive oxygen species levels and radioresistance in cancer stem cells. Nature 2009 Apr 9;458(7239):780-3. PubMed PMID: 19194462; PMCID: PMC2778612

4.   Diehn M, Clarke MF. Cancer stem cells and radiotherapy: new insights into tumor radioresistance. J Natl Cancer Inst 2006 Dec 20;98(24):1755-7. PubMed PMID: 17179471

5.   Eyler CE, Rich JN. Survival of the fittest: cancer stem cells in therapeutic resistance and angiogenesis. J Clin Oncol 2008 Jun 10;26(17):2839-45. PubMed PMID: 18539962; PMCID: PMC2739000

6.   Li X, Lewis MT, Huang J, Gutierrez C, Osborne CK, Wu MF, Hilsenbeck SG, Pavlick A, Zhang X, Chamness GC, et al. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy J Natl Cancer Inst 2008 May 7;100(9):672-9. Epub 2008 Apr 29. PubMed PMID: 18445819

7.   Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell 2008 May 16;133(4):704-15. PubMed PMID: 18485877; PMCID: PMC2728032

8.   Gupta PB, Onder TT, Jiang G, Tao K, Kuperwasser C, Weinberg RA, Lander ES. Identification of selective inhibitors of cancer stem cells by high-throughput screening. Cell 2009 Aug 21;138(4):645-59. Epub 2009 Aug 13. PubMed PMID: 19682730

9.   Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C, Fröhling S, Chan EM, Sos ML, Michel K, Mermel C, Silver SJ, Weir BA, Reiling JH, Sheng Q, Gupta PB, Wadlow RC, Le H, Hoersch S, Wittner BS, Ramaswamy S, Livingston DM, Sabatini DM, Meyerson M, Thomas RK, Lander ES, Mesirov JP, Root DE, Gilliland DG, Jacks T, Hahn WC. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009 Nov 5;462(7269):108-12. Epub 2009 Oct 21. PMID:19847166; PMCID: PMC2783335

10.   Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ, Wicha MS. In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev 2003 May 15;17(10):1253-70. PubMed PMID: 12756227; PMCID: PMC196056

11.   Carmody LC, Germain A, Morgan B, Burhlage S, VerPlank L, Munoz B, Feng Y, Gupta PB, Lander ES, Palmer M, Schreibe, SL. Identification of selective breast cancer stem cell inhibitors. Probe Reports from the NIH Molecular Libraries Program. National Center for Biotechnology Information, U.S. National Library of Medicine Epub 2011. Bookshelf ID: NBK47352; PMID: 21433368

12. Onder TT, Gupta PB, Mani SA, Yang J, Lander ES, Weinberg RA. Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways. Cancer Res. 2008 May 15;68(10):3645-54. PMID:18483246

13. Selimovic D, Ahmad M, El-Khattouti A, Hannig M, Haïkel Y, Hassan M. Apoptosis-related protein-2 triggers melanoma cell death by a mechanism including both endoplasmic reticulum stress and mitochondrial dysregulation. Carcinogenesis. 2011 Aug;32(8):1268-78. Epub 2011 Jun 21. PMID:21693538

14. Vives-Bauza C, Magrané J, Andreu AL, Manfredi G. Novel role of ATPase subunit C targeting peptides beyond mitochondrial protein import. Mol Biol Cell. 2010 Jan 1;21(1):131-9. Epub 2009 Nov 4. PMID:19889836; PMCID: PMC2801706

15. Stark K, Wu ZL, Bartleson CJ, Guengerich FP. mRNA distribution and heterologous expression of orphan cytochrome P450 20A1. Drug Metab Dispos. 2008 Sep;36(9):1930-7. Epub 2008 Jun 9. PMID:18541694

16. Wang Z, Cotney J, Shadel GS. Human mitochondrial ribosomal protein MRPL12 interacts directly with mitochondrial RNA polymerase to modulate mitochondrial gene expression. J Biol Chem. 2007 Apr 27;282(17):12610-8. Epub 2007 Mar 2. PMID:17337445; PMCID: PMC2606046

17. Zhou J, Zhang H, Gu P, Bai J, Margolick JB, Zhang Y. NF-kappaB pathway inhibitors preferentially inhibit breast cancer stem-like cells. Breast Cancer Res Treat 2008 Oct;111(3):419-27. Epub 2007 Oct 27. PubMed PMID: 17965935

18. Zhou J, Zhang H, Gu P, Margolick JB, Yin D, Zhang Y. Cancer stem/progenitor cell active compound 8-quinolinol in combination with paclitaxel achieves an improved cure of breast cancer in the mouse model. Breast Cancer Res Treat. 2009 May;115(2):269-77. Epub 2008 May 28. PubMed PMID: 18506619

19. Kakarala M, Brenner DE, Korkaya H, Cheng C, Tazi K, Ginestier C, Liu S, Dontu G, Wicha MS. Targeting breast stem cells with the cancer preventive compounds curcumin and piperine. Breast Cancer Res Treat. 2010 Aug;122(3):777-85. Epub 2009 Nov 7. PubMed PMID: 19898931; PMCID: PMC3039120

20. Li Y, Zhang T, Korkaya H, Liu S, Lee HF, Newman B, Yu Y, Clouthier SG, Schwartz SJ, Wicha MS, Sun D. Sulforaphane, a dietary component of broccoli/broccoli sprouts, inhibits breast cancer stem cells. Clin Cancer Res. 2010 May 1;16(9):2580-90. Epub 2010 Apr 13. PubMed PMID: 20388854; PMCID: PMC2862133

21. Jiralerspong S, Palla SL, Giordano SH, Meric-Bernstam F, Liedtke C, Barnett CM, Hsu L, Hung MC, Hortobagyi GN, Gonzalez-Angulo AM. Metformin and pathologic complete responses to neoadjuvant chemotherapy in diabetic patients with breast cancer. J Clin Oncol. 2009 Jul 10;27(20):3297-302. Epub 2009 Jun 1. PubMed PMID: 19487376; PMCID: PMC2736070

22. Kim JB, Ko E, Han W, Shin I, Park SY, Noh DY. Berberine diminishes the side population and ABCG2 transporter expression in MCF-7 breast cancer cells. Planta Med. 2008 Nov;74(14):1693-700. Epub 2008 Oct 24. PubMed PMID: 18951337

Appendix A: Assay Summary Table

Table A1. Summary of Completed Assays and AIDs

PubChem AID No.	Type	Target	Concentration Range (µM)	Samples Tested
2717	Primary	Breast Cancer Stem Cell-like HMLE_sh_ECad	Average 3.75 µM	300,625
449748	Primary	Breast Cancer Stem Cell-like HMLE_sh_ECad	160 nM-19.5 µM	2243
493176	Primary	Breast Cancer Stem Cell-like HMLE_sh_ECad	80 nM-19.5 µM	19
493226	Primary	Breast Cancer Stem Cell-like HMLE_sh_ECad	4.6 nM-30 µM	65
504449	Primary	Breast Cancer Stem Cell-like HMLE_sh_ECad	4.6 nM-30 µM	76
504535	Primary	Breast Cancer Stem Cell-like HMLE_sh_ECad	4.6 nM-30 µM	45
504667	Primary	Breast Cancer Stem Cell-like HMLE_sh_ECad	4.6 nM-30 µM	45
504788	Primary	Breast Cancer Stem Cell-like HMLE_sh_ECad	4.6 nM-30 µM	45
463074	Secondary	Control Cells HMLE_sh_GFP	160 nM-19.5 µM	2243
493196	Secondary	Control Cells HMLE_sh_GFP	80 nM-19.5 µM	19
504325	Secondary	Control Cells HMLE_sh_GFP	4.6 nM-30 µM	65
504450	Secondary	Control Cells HMLE_sh_GFP	4.6 nM-30 µM	76
504533	Secondary	Control Cells HMLE_sh_GFP	4.6 nM-30 µM	45
504666,	Secondary	Control Cells HMLE_sh_GFP	4.6 nM-30 µM	45
504789	Secondary	Control Cells HMLE_sh_GFP	4.6 nM-30 µM	45
504623	Orthogonol	Sum159 Tumorsphere formation	Single dose; Imaging	16
504859	Orthogonol	Sum159 Tumorsphere formation	Single dose; Imaging	16
2721	Summary	NA	NA	NA

NA= not applicable

Appendix B: Detailed Assay Protocols

Primary HTS Using CellTiter-Glo (AID 2717)

Propagation Media (CSC media: complete media with serum)

Using already filtered/sterile components, add: 440 ml DMEM (Cellgro 10-013-CM), 50 ml FBS (HyClone SH30071.03),5 ml Pen/Strep, 5 ml Glutamax-1 (Invitrogen 35050-061), 700 µl 50 µM Hyrdocortisone (Sigma H039K8402), 600 µl 10 mg/ml Insulin (Sigma I9278), 500 µl 50 mg/ml Gentamicin (Sigma G1397), 250 µl 25 mg/ml Plasmocin (Invivogen ant-mpt), 50 µl 100 mg/ml EGF (Sigma E9644) resuspended in 2% serum/PBS, plus 500 ml Mammary Epithelial Cell Growth Medium (MEGM Complete Medium; Lonza CC-3051) or MEGM Bullet Kit with components CC-3150,1 ml BPE (Lonza CC-4009); Makes 1 liter.

Screening Media (CSC media: complete media without serum)

Using already filtered/sterile components, add: 490 ml DMEM (Cellgro 10-013-CM), 5 ml Pen/Strep, 5 ml Glutamax-1 (Invitrogen 35050-061), 700 µl 50 µM Hyrdocortisone (Sigma H039K8402),600 µl 10 mg/ml Insulin (Sigma I9278), 500 µl 50 mg/ml Gentamicin (Sigma G1397), 250 µl 25 mg/ml Plasmocin (Invivogen ant-mpt), 50 µl 100 mg/ml EGF (Sigma E9644) resuspended in 2% serum/PBS, 500 ml MEGM Complete Medium (Lonza CC-3051) or MEGM Bullet Kit with components CC-3150, 1 ml BPE (Lonza CC-4009); Makes 1 liter.

Using HMLE_sh_ECad cells provided by the Assay Providers, 10 million cells/vial were frozen down in CSC media (complete media with serum and10% DMSO) and stored in liquid nitrogen.

Cell Propagation

1. Two to 3 days prior to screening, thaw vials and plate into T225 Falcon flasks with 40 ml CSC complete media with serum.
2. Incubate at 37 °C, 5% CO_2 for approximately 16 hours.
3. Wash cells with 1x PBS; add 5 ml Trypsin and EDTA to cells for 2-5 minutes.
4. Resuspend cells in 25 ml CSC complete media with serum;spin at 1400 rpm for 4 minutes.
5. Aspirate media and resuspend cells in CSC complete media with serum.
6. Plate cells in T225 flasks (as above) at 1:4 or 1:8 density and allow to grow 1-2 days.

Cell Screening

1. Trypsinize cells and spin at 1400 rpm for 4 minutes.
2. Resuspend pellets in CSC complete without serum.
3. Add 50 µL cell supension to 50 µL trypan blue and count using a Cellometer (Nexcelom Bioscience).
4. Dilute cells to 40,000cells/ml in CSC complete without serum.

5. Using a standard sized cassette, dispense 2000 cells/50 µl into tissue-culture treated, 384-well plates (Corning 3850) using a Multidrop Combi (Thermo Scientific).
6. Incubate the cells at 37 °C, 5% CO_2 for at least 4 hours.
7. Pin the cells with 100 nl compounds; incubate at 37 °C, 5% CO_2 for approximately 70-74 hours.
8. Prepare CellTiterGlo per supplier (Promega) instructions and and dilute 1:3 with 1x PBS.
9. Incubate assay plates at room temperature for 30 minutes.
10. Dispense 30 µl CellTiterGlo using a standard sized cassette by a Multidrop Combi (Thermo Scientific).
11. Incubate the plates at room temperature for 10 minutes and read using the EnVision (Luminescence 1 second)

Primary Retest (AID 449748)

Repeat of the primary screen at dose using CellTiter-Glo.

Primary Cell Viability Protocol with HMLE_sh_ECad Cells Using CellTiter-Glo (AID 493226, AID 493176, AID 504449, AID 504535)

HMLE_sh_ECad cells were prepared as previously described (8).

1. Propagate HMLE cells expressing either shRNA targeting E-cadherin (sh_ECad), control shRNA targeting eGFP (sh_GFP), or shRNA targeting Twist (sh_Twist) in a 1:1 mixture of 10% fetal bovine serum (FBS; HyClone), 1% Penicillin/Streptomycin (Pen/Strep; Cellgro), 1% Glutamax-1 (Invitrogen), 70 nM Hydrocortisone (Sigma), 12 µg/ml Insulin (Sigma), 50 µg/ml Gentamicin (Sigma), 12.5 µg/ml Plasmocin (InVivogen), 10 ng/ml EGF in DMEM (Cellgro) with Mammary Epithelial Cell Growth Medium (MEGM complete medium; Lonza, Basel, Switzerland).
2. Incubate at 37 ºC, 5% CO_2.
3. For screening, count the cells and resuspend in complete media without serum.
4. Plate 2,000 cells in 50 µl per well in white, tissue culture-treated, 384-well plates (Corning).
5. Incubate the cells at 37 ºC, 5% CO_2 for at least 4 hours and pin with 100 nl of compounds.
6. Incubate the cells approximately 72 hours, then add 30 µl of CellTiter-Glo (Promega) diluted 1:3 with PBS to the well.
7. Read the plates using the EnVision (PerkinElmer; Luminescence 0.1 sec/well) after 12 minutes.

Secondary Cell Viability Protocol with HMLE_sh_GFP Cells Using CellTiter-Glo (AID 504325, AID 493196, AID 504450, AID 504533)

HMLE_sh_GFP cells were prepared as previously described (8). Refer to the procedures as described in the Primary Cell Viability Protocol.

Assay Provider Cell Viability Protocol with CellTiter-Glo

HMLE_sh_Twist cells were prepared as previously described (8). Refer to the procedures as described in the Primary Cell Viability Protocol.

Secondary Orthogonal Assay for *In Vitro* Inhibition of Tumorspheres with Sum159 Cells (AID 504623)

Sum159 cells were provided by Dr. Gupta as previously described (8).

1. Perform tumorsphere assays were performed as previously described (10) with minor changes.
2. Propagate Sum159 cells in 5% FBS (HyClone), 1 % Pen/Strep, 1% Glutamax-1, 12 µg/mlInsulin, 50 µg/ml Gentamicin in F12/DMEM (Cellgro).
3. Plate 100 µl media respective to cell type in 96-well, ultra-low adhesion plates (Costar).
4. Pin 400 nl of compounds into the media.
5. Harvest the cells, count, and resuspend in their propagation media with 1% methylcellulose (ES-CultM3120, Stem Cell Technologies).
6. Add 100 µl of resuspended cells to the plates containing media with compound for a final count of 2000 cells/well in 200 µl with 0.5% methylcellulose.
7. Allow tumorspheres to form for 9 days incubated at 37 °C, 5% CO_2.
8. Image tumorspheres using a 2X objective on the ImageXpress Micro (Molecular Devices). Identify cell clusters greater than 100 µM in diameter were identified using MetaXpress software (version 3.1; Molecular Devices, Sunnyvale, CA).

Gene Expression Assays with HMLE_sh_ECad and HMLE_sh_GFP Cells

1. ML245, CID50904149, Salinomycin/CID23682228 are selectively toxic to HMLE_shECad cell lines (over HMLE_shGFP cell lines) in 3-day toxicity assays. The IC_{50} concentration of toxicity towards HMLE_shECad cell line in these 3-day toxicity assays is the same concentration used for experiments below (i.e., 0.54 µM, 6.7 µM, and 1 µM, respectively). 300,000 cells HMLE_shECad or HMLE_shGFP cells were plated in a 6-well dish and allowed to adhere for 4 hours.
2. Replicate samples (3-4) of HMLE_shECad and HMLE_shGFP cells were treated with vehicle, ML245, CID50904149, or Salinomycin at the above concentrations for 24 hours prior to isolation of RNA. 30
3. Isolation of total RNA was conducted using the protocols for the RNeasy Protect Mini Kit (Qiagen).
4. **Genome Analysis Platform (GAP) at the Broad Institute** analyzed RNA samples for quality using Aglient Bioanalyzer Chips, and prepared cRNA synthesis from the total RNA samples passing QC for analysis on a HumanHT-12 Expression BeadChip (Illumina, San Diego, CA) according to the manufacturer's instructions. Results were confirmed using GenomeStudio (version 2010.3; Illumina, San Diego, CA).
5. All of the samples were run in replicates (3-4) and the entire data set was normalized using the quantile module available in open source program GenePattern (http://genepattern.broadinstitute.org/).
6. Gene expression was compared between the DMSO-treated and compound-treated using the ComparitiveSelection Module in GenePattern.
7. Those genes with a significant difference (p>0.005) and had 1.50-fold difference in expression were selected.

Appendix C: Experimental Procedures for the Synthesis of the Probe

Chemistry Experimental Methods

General details.   All oxygen and/or moisture-sensitive reactions were carried out under nitrogen (N_2) atmosphere in glassware that had been flame-dried under vacuum (approximately 0.5 mm Hg) and purged with N_2 prior to use.   All reagents and solvents were purchased from commercial vendors and used as received, or synthesized according to methods already reported.   NMR spectra were recorded on a Bruker 300 (300 MHz ^{1}H, 75 MHz ^{13}C) or Varian UNITY INOVA 500 (500 MHz ^{1}H, 125 MHz ^{13}C) spectrometer.   Proton and carbon chemical shifts are reported in ppm (δ) referenced to the NMR solvent.   Data are reported as follows: chemical shifts, multiplicity (br = broad, s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet; coupling constant(s) in Hz).

Unless otherwise indicated, NMR data were collected at 25 °C.   Flash chromatography was performed using 40-60 μm Silica Gel (60 Å mesh) on a Teledyne Isco Combiflash R$_f$.   Tandem Liquid Chromatography/Mass Spectrometry (LC/MS) was performed on a Waters 2795 separations module and 3100 mass detector.   Analytical thin layer chromatography (TLC) was performed on EM Reagent 0.25 mm silica gel 60-F plates.   Visualization was accomplished with ultraviolet (UV) light and aqueous potassium permanganate ($KMnO_4$) stain followed by heating. High-resolution mass spectra were obtained at the MIT Mass Spectrometry Facility (Bruker Daltonics APEXIV 4.7 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer).

5-(hydroxymethyl)furan-2-carboxylic acid (1.15 g, 8.09 mmol) was diluted with Benzene (81 ml). Thionyl chloride (3.54 ml, 48.6 mmol) was added and then the reaction is heated to reflux. The reaction was stirred overnight. The reaction was concentrated and carried on directly to the next step. The dichloride was dissolved in dichloromethane (112 ml) and 4-methylthiazol-2-amine (1.302 g, 11.17 mmol) was added followed by 4-Dimethylaminopyridine (DMAP;0.138 g, 1.117 mmol) and triethylamine (3.89 ml, 27.9 mmol). The reaction was stirred at room temperature until complete by LC-MS analysis. The reaction was concentrated and purified by silica gel chromatography to yield the furanyl chloride (1.61 g) in 56% yield over two steps. The furanyl chloride (63.1 mg, 0.246 mmol) was dissolved in dimethylformamide (2.5 ml). Sodium iodide (3.68 mg, 0.025 mmol), potassium carbonate (51.0 mg, 0.369 mmol), and 3-chlorophenol (31.6 mg, 0.246 mmol) were added and the reaction was stirred at room temperature until complete by LC-MS analysis. The reaction was concentrated under vacuum and purified by silica gel chromatography to provide 5-((3-chlorophenoxy)methyl)-N-(4-methylthiazol-2-yl)furan-2-carboxamide (56 mg) (CID50910523/ML243) in 65% yield.

^{1}H NMR (300 MHz, CDCl3) δ 7.19 (s, 2H), 6.91 (s, 1H), 6.85 – 6.74 (m, 2H), 6.50 (s, 2H), 5.28 – 5.19 (m, 2H), 2.30 (s, 3H),; **^{13}C NMR (75 MHz, CDCl3)** δ 157.38, 155.83, 153.78, 151.82, 146.26, 144.93, 133.81, 129.33, 120.72, 116.44, 114.20, 112.05, 111.37, 107.34, 61.33, 15.77. calculated [M$^+$H] 349.0408 experimental [M$^+$H] 349.0408.

Appendix D: Experimental Procedures for Analytical Assays

Solubility. Solubility was determined in phosphate buffered saline (PBS) pH 7.4 with 1% DMSO. Each compound was prepared in duplicate at 100 μM in both 100% DMSO and PBS with 1% DMSO. Compounds were allowed to equilibrate at room temperature with a 250 rpm orbital shake for 24 hours. After equilibration, samples were analyzed by UPLC-MS (Waters, Milford, MA) with compounds detected by SIR detection on a single quadrupole mass spectrometer. The DMSO samples were used to create a two point calibration curve to which the response in PBS was fit.

PBS Stability. Stability was determined in the presence of PBS pH 7.4 with 0.1% DMSO. Each compound was prepared in duplicate on six separate plates and allowed to equilibrate at room temperature with a 250 rpm orbital shake for 48 hours. One plate was removed at each time point (0, 2, 4, 8, 24, and 48 hours). An aliquot was removed from each well and analyzed by UPLC-MS (Waters, Milford, MA) with compounds detected by SIR detection on a single quadrupole mass spectrometer. Additionally, to the remaining material at each time point, acetonitrile was added to force dissolution of compound (to test for recovery of compound). An aliquot of this was also analyzed by UPLC-MS.

GSH Stability. Stability was determined in the presence of PBS pH 7.4 μM and 50 μM glutathione with 0.1% DMSO. Each compound was prepared in duplicate on six separate plates and allowed to equilibrate at room temperature with a 250 rpm orbital shake for 48 hours. One plate was removed at each time point (0, 2, 4, 8, 24, and 48 hours). An aliquot was removed from each well and analyzed by UPLC-MS (Waters, Milford, MA) with compounds detected by SIR detection on a single quadrupole mass spectrometer. Additionally, to the remaining material at each time point, acetonitrile was added to force dissolution of compound (to test for recovery of compound). An aliquot of this was also analyzed by UPLC-MS.

Plasma Protein Binding. Plasma protein binding was determined by equilibrium dialysis using the Rapid Equilibrium Dialysis (RED) device (Pierce Biotechnology, Rockford, IL) for both human and mouse plasma. Each compound was prepared in duplicate at 5 μM in plasma (0.95% acetonitrile, 0.05% DMSO) and added to one side of the membrane (200 μl) with PBS pH 7.4 added to the other side (350 μl). Compounds were incubated at 37º C for 5 hours with a 250 rpm orbital shake. After incubation, samples were analyzed by UPLC-MS (Waters, Milford, MA) with compounds detected by SIR detection on a single quadrupole mass spectrometer.

Plasma Stability. Plasma stability was determined at 37º C at 5 hours in both human and mouse plasma. Each compound was prepared in duplicate at 5 μM in plasma diluted 50/50 (v/v) with PBS pH 7.4 (0.95% acetonitrile, 0.05% DMSO). Compounds were incubated at 37º C for 5 hours with a 250 rpm orbital shake with time points taken at 0 and 5 hours. Samples were analyzed by UPLC-MS (Waters, Milford, MA) with compounds detected by SIR detection on a single quadrupole mass spectrometer.

Appendix E: Chemical Characterization Data for Probe and Analogs

^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Probe CID 50904134/ ML245
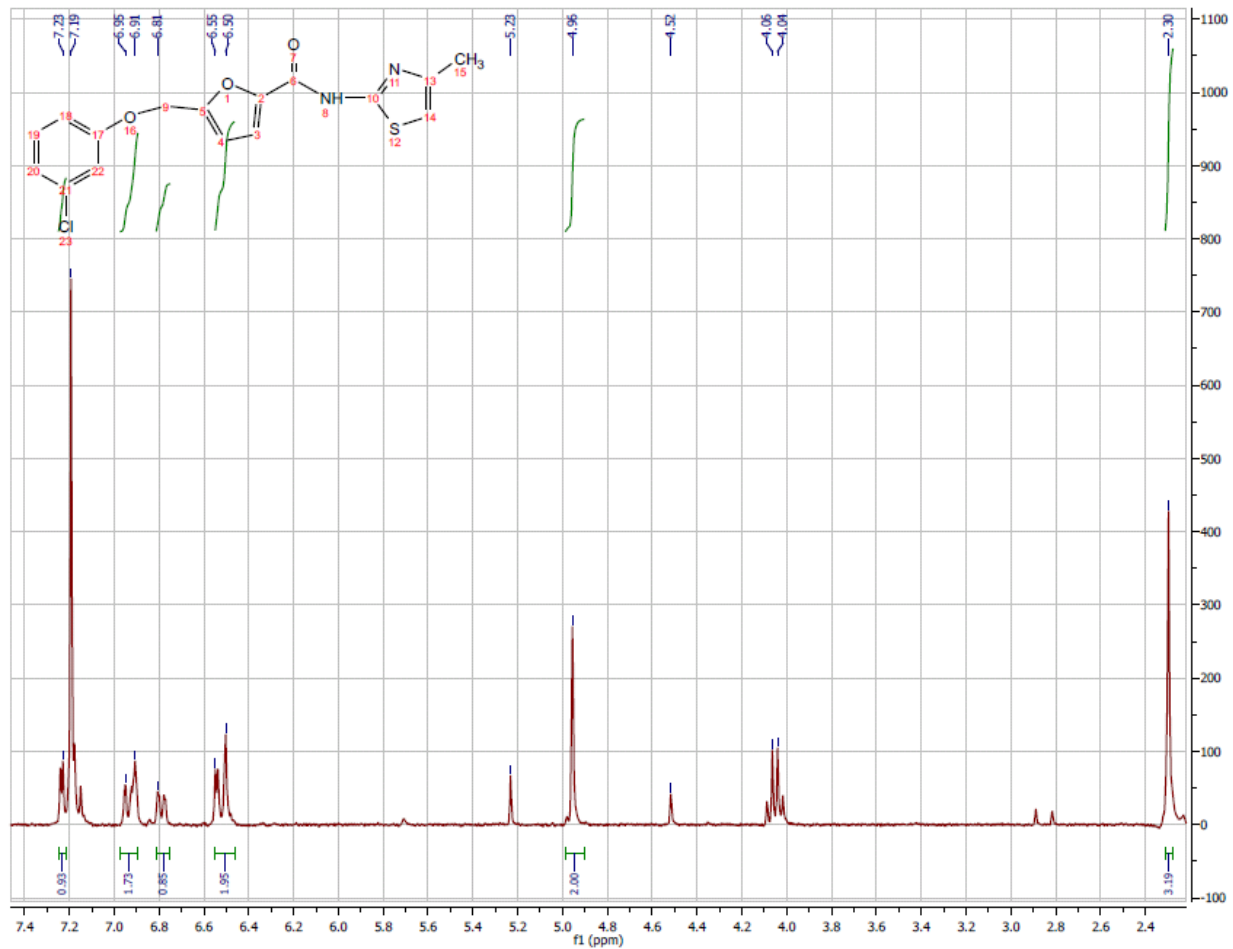
^{13}C NMR Spectrum (300 MHz, CDCl$_3$) of Probe CID 50904134/ ML245
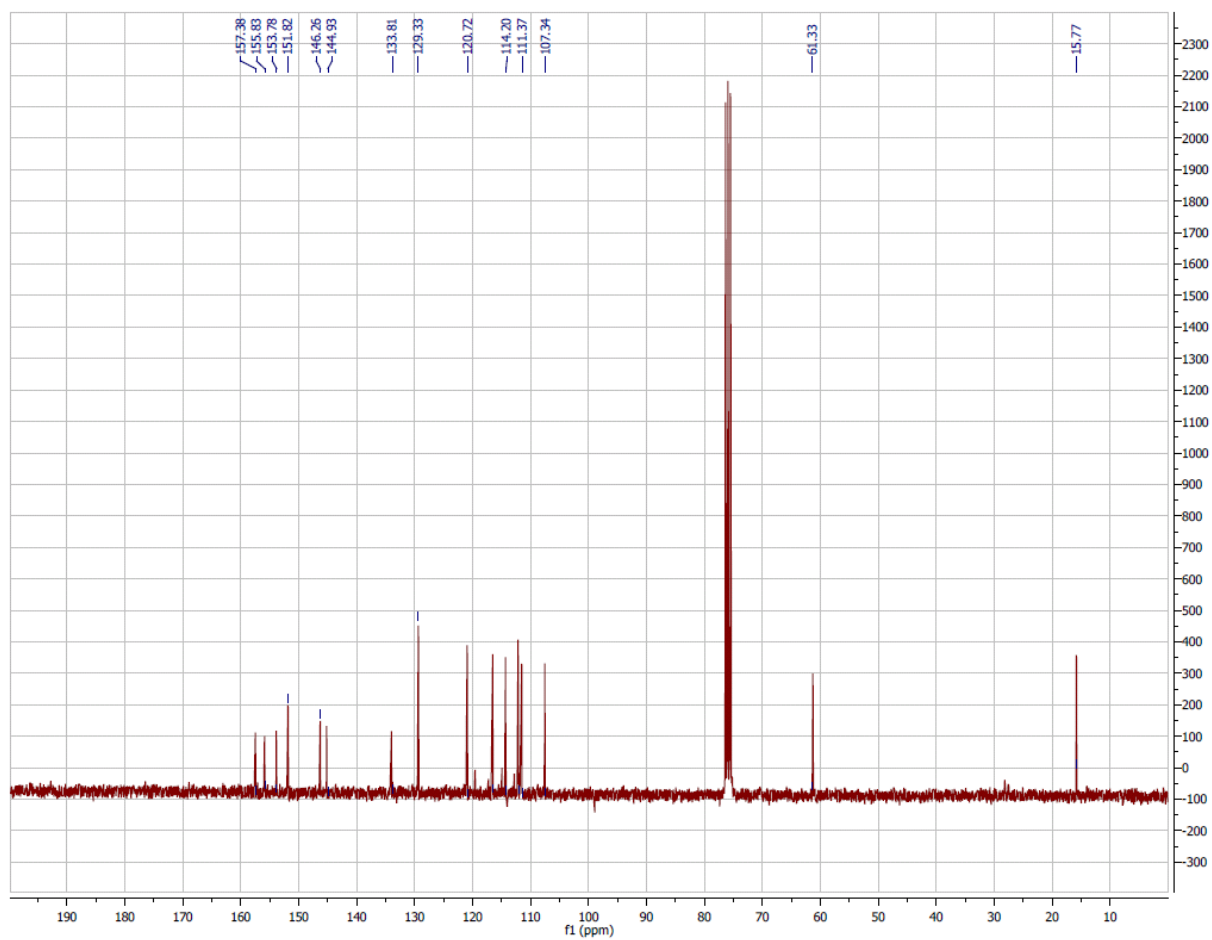
UPLC-MS Chromatogram of Probe CID 50904134/ ML245
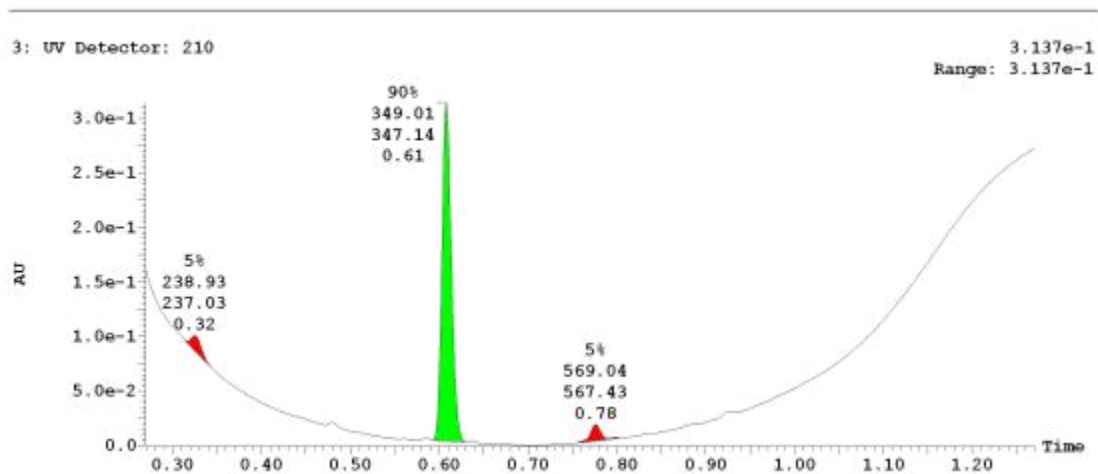
HRMS of **Probe CID 50904134/ ML245**
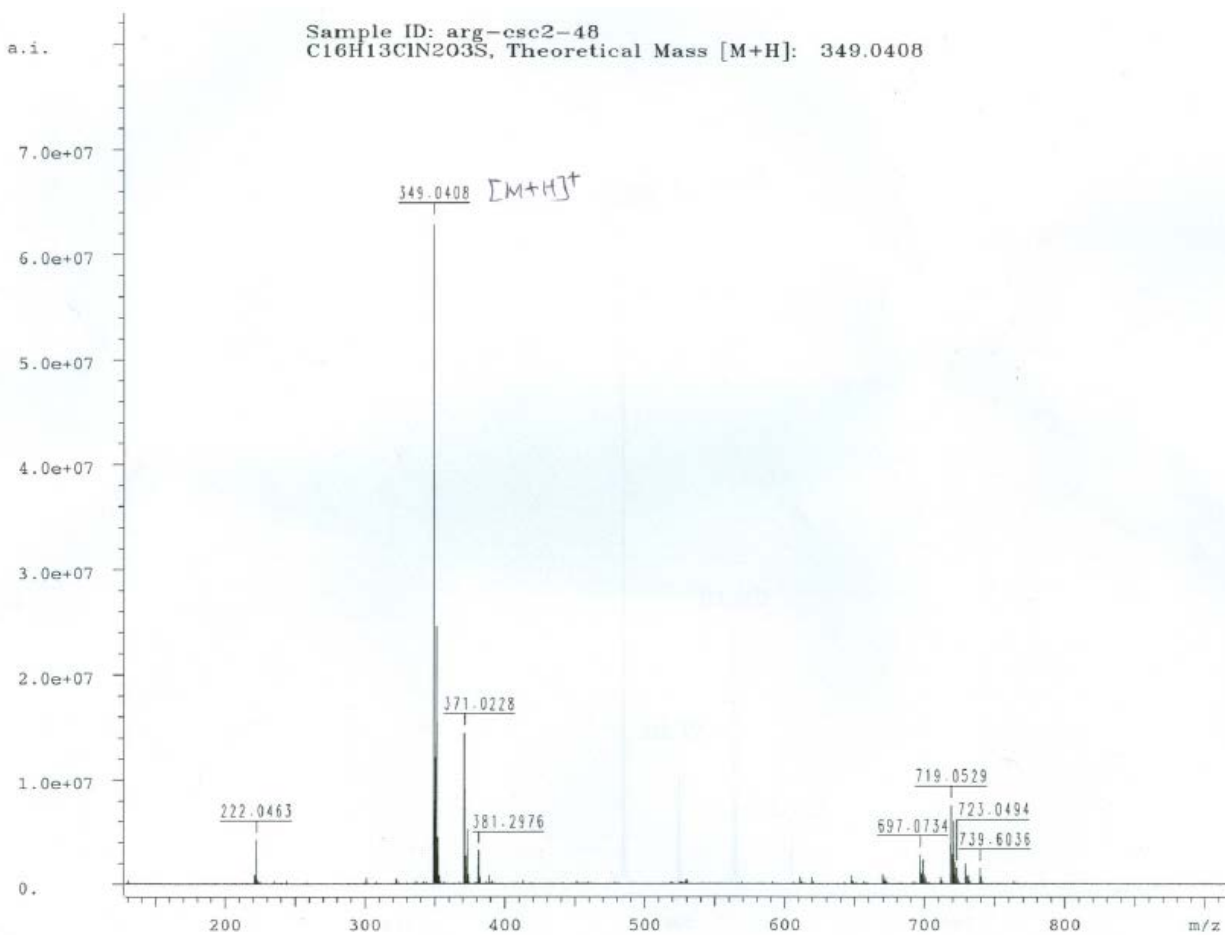
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 4791237
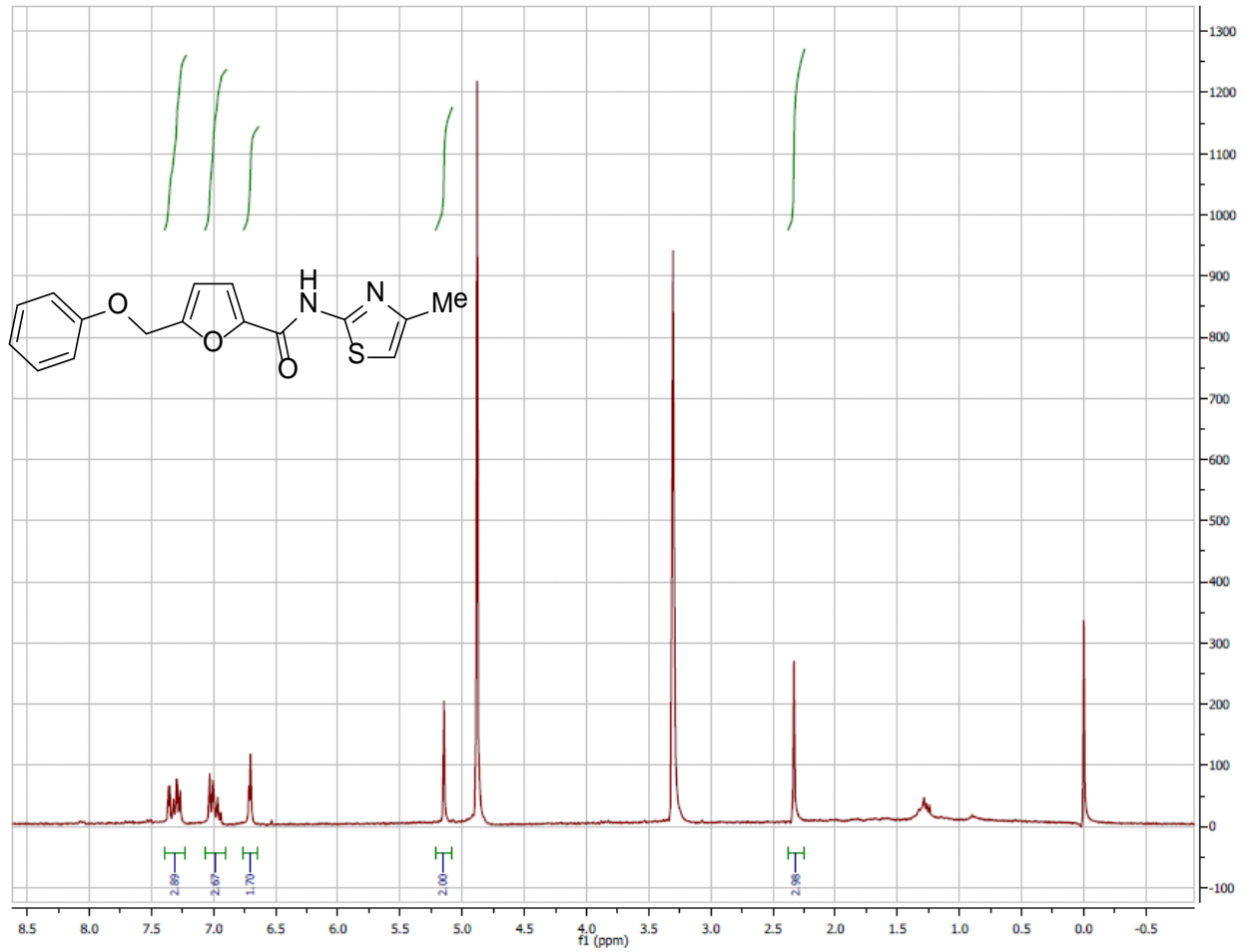
UPLC-MS Chromatogram of Analog CID 4791237

3: UV Detector: 210

5.921e-1
Range: 5.921e-1

91%
314.97
313.17
0.57

1%
343.99
342.18
0.45

5%
225.09
223.26
0.66

3%
173.05
213.11
0.73
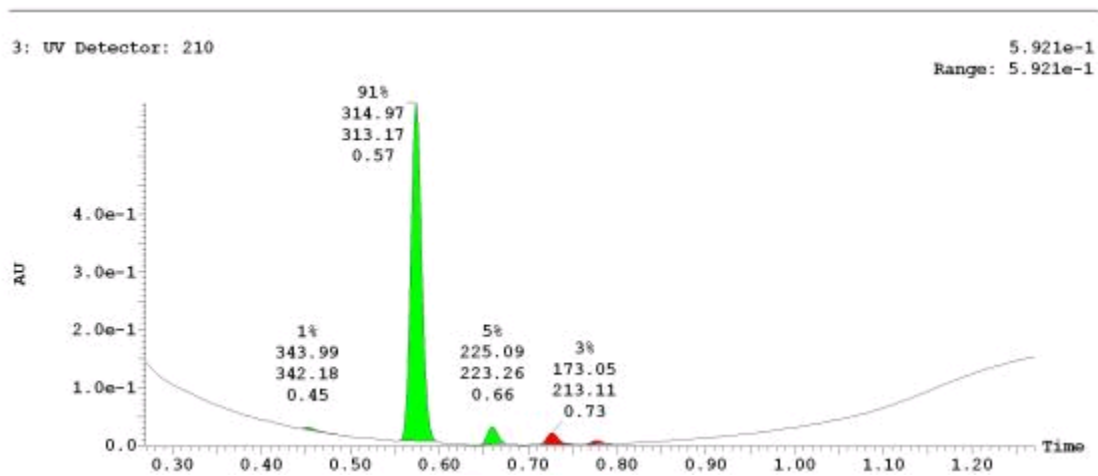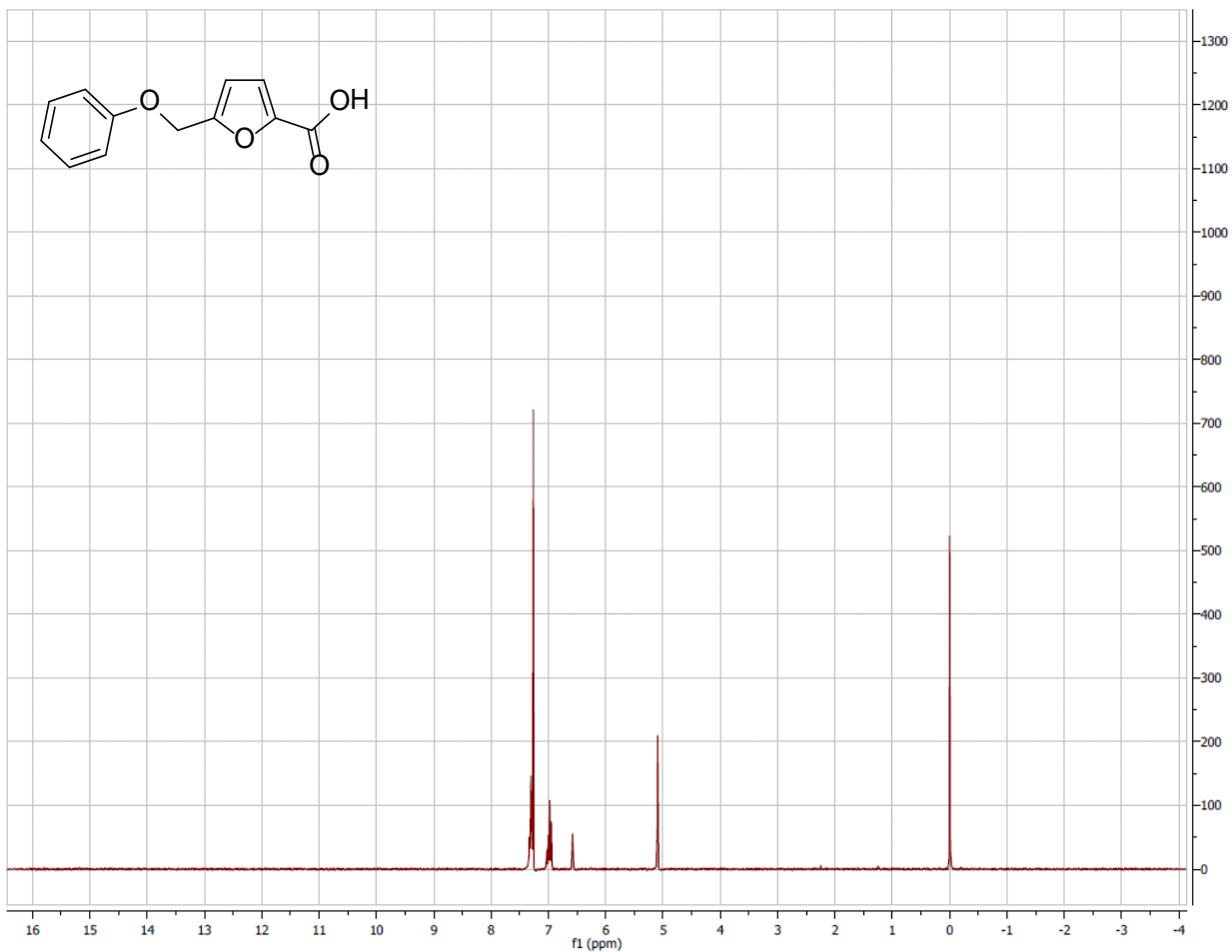
UPLC-MS Chromatogram of Analog CID 580491
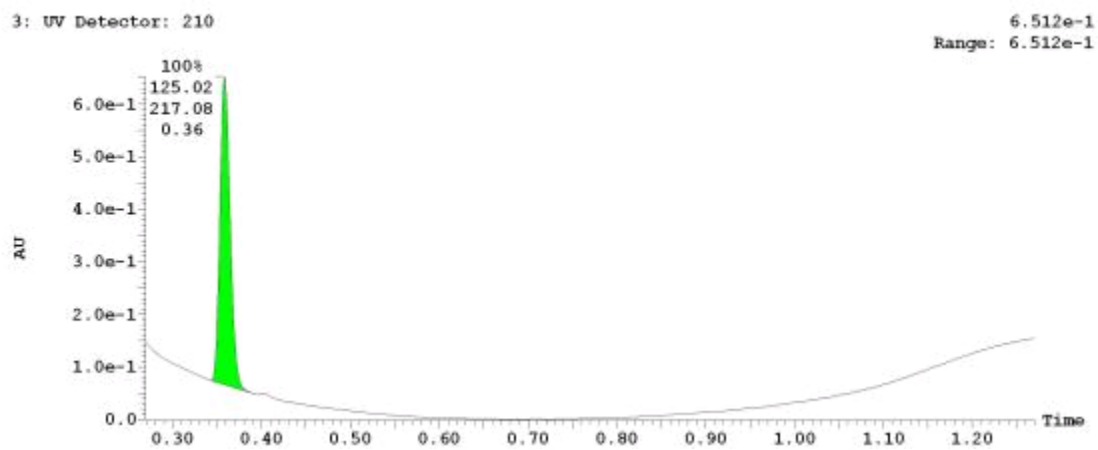
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 51003714
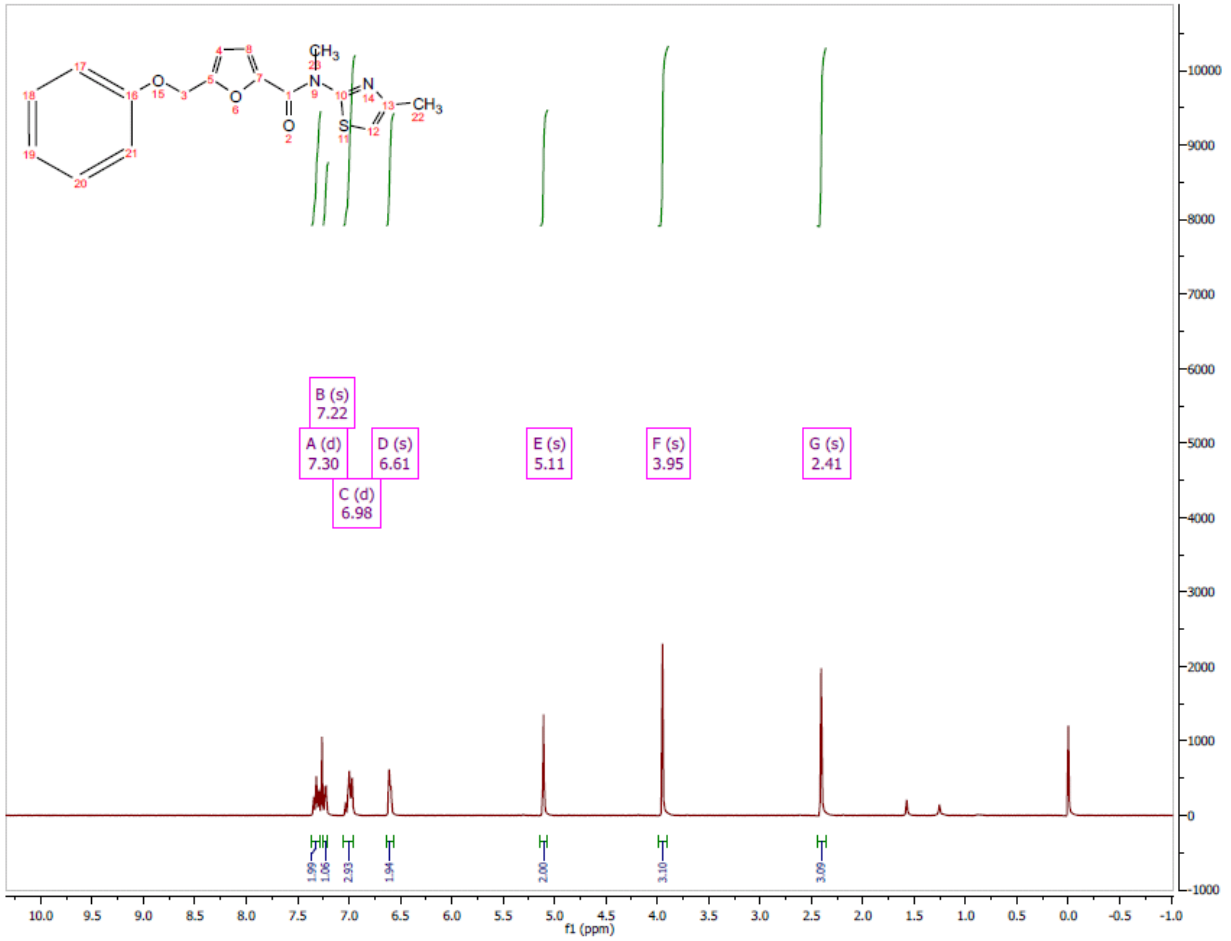
UPLC-MS Chromatogram of Analog CID 51003714

BROAD INSTITUTE
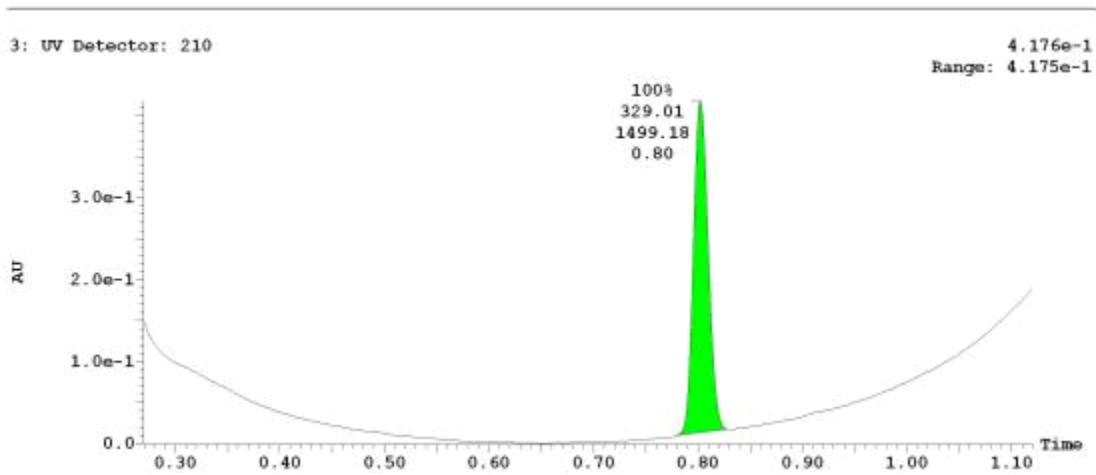
UPLC-MS Chromatogram of Analog CID 38728154
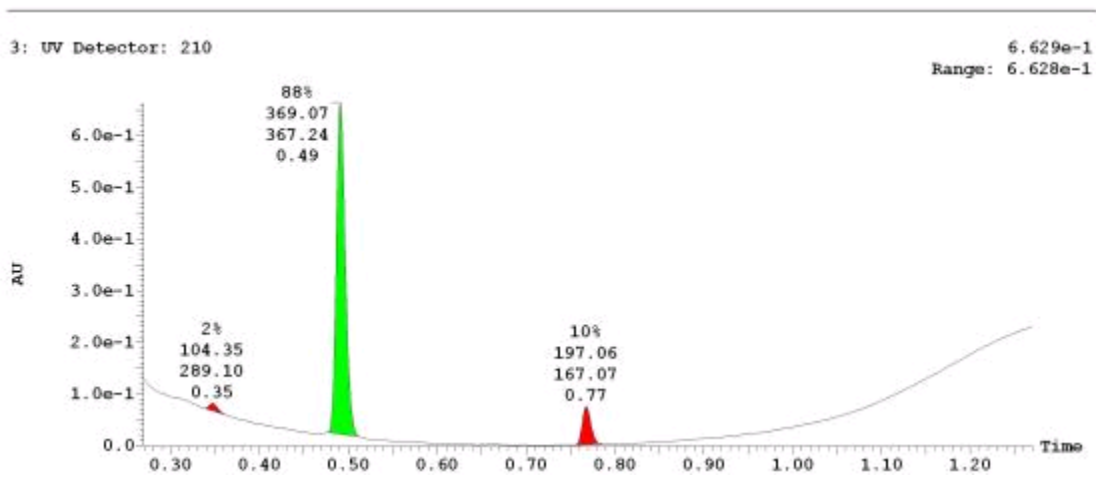
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 41549965

UPLC-MS Chromatogram of Analog CID 41549965

3: UV Detector: 210

7.452e-1
Range: 7.452e-1

100%
315.00
313.15
0.59
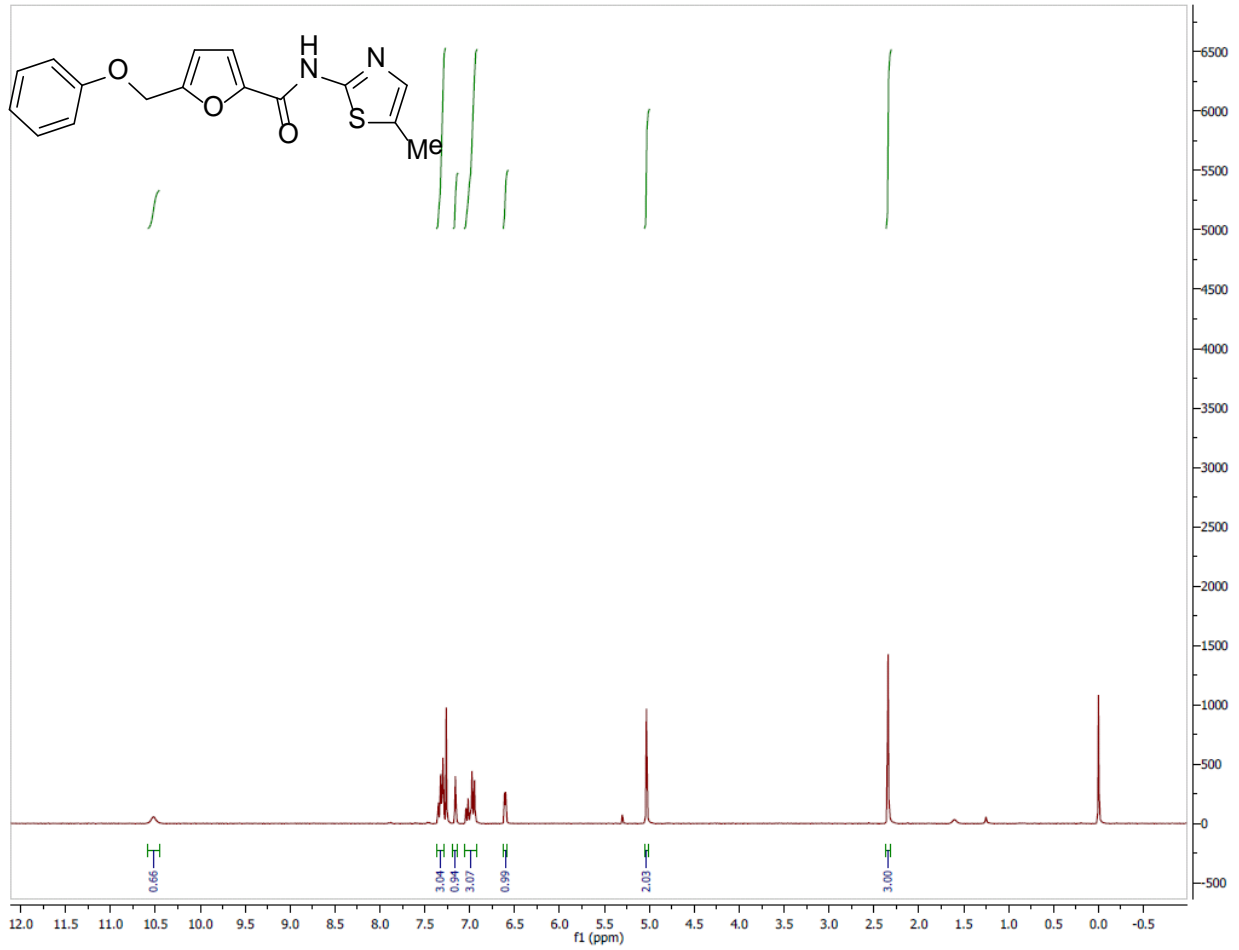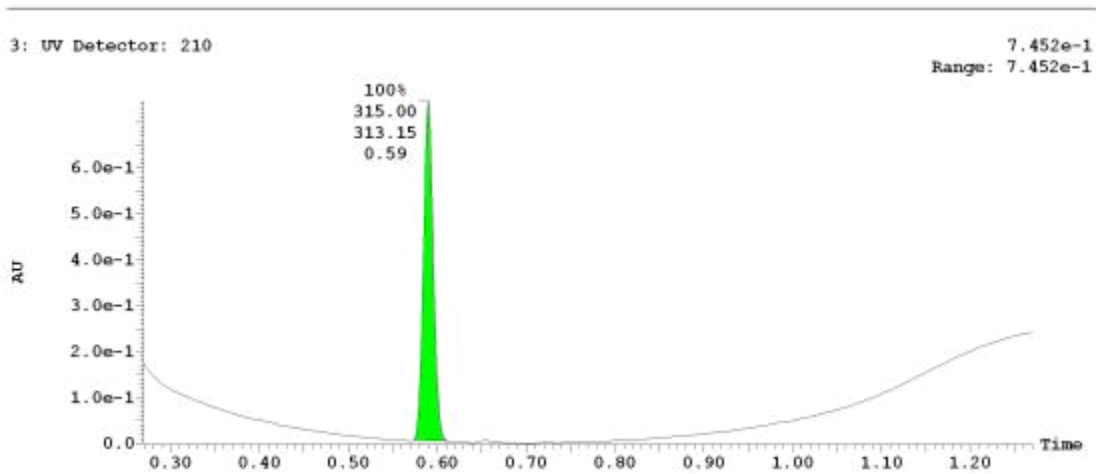
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 2967498

UPLC-MS Chromatogram of Analog CID 2967498

3: UV Detector: 210

8.694e-1
Range: 8.694e-1

100%
301.03
299.15
0.52
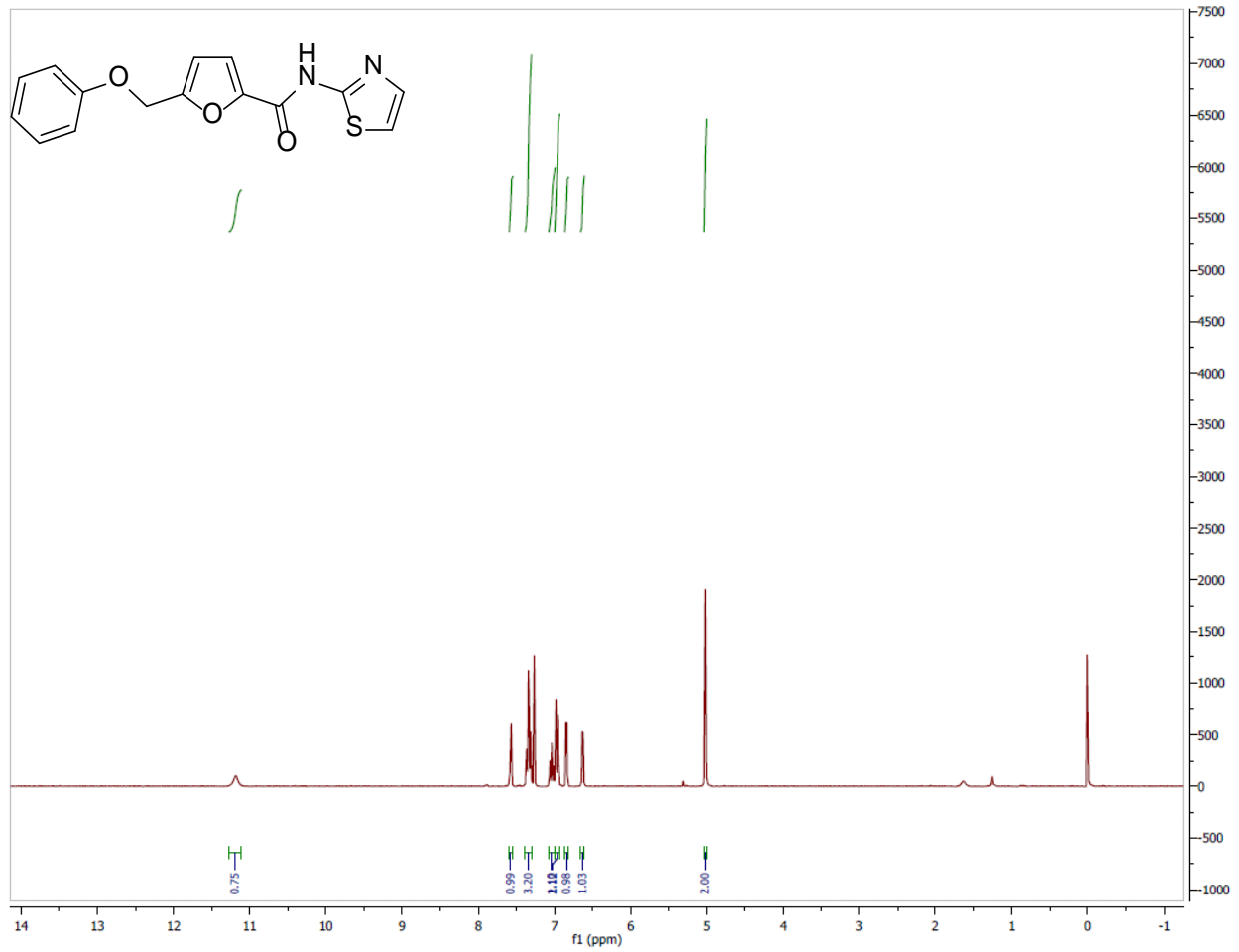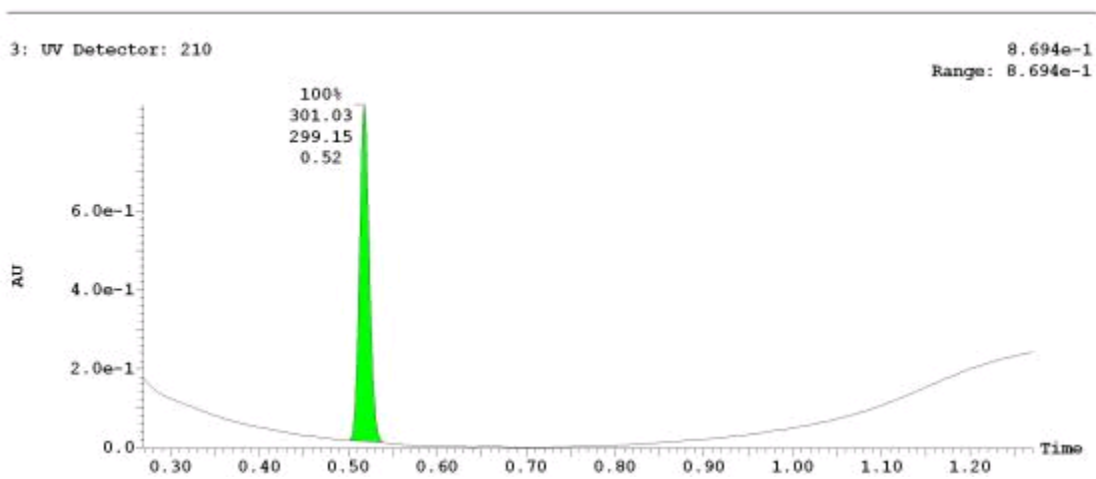
¹H NMR Spectrum (300 MHz, CDCl₃) of Analog CID 8882461

UPLC-MS Chromatogram of Analog CID 8882461

3: UV Detector: 210

1.119
Range: 1.119

97%
351.08
349.23
0.55
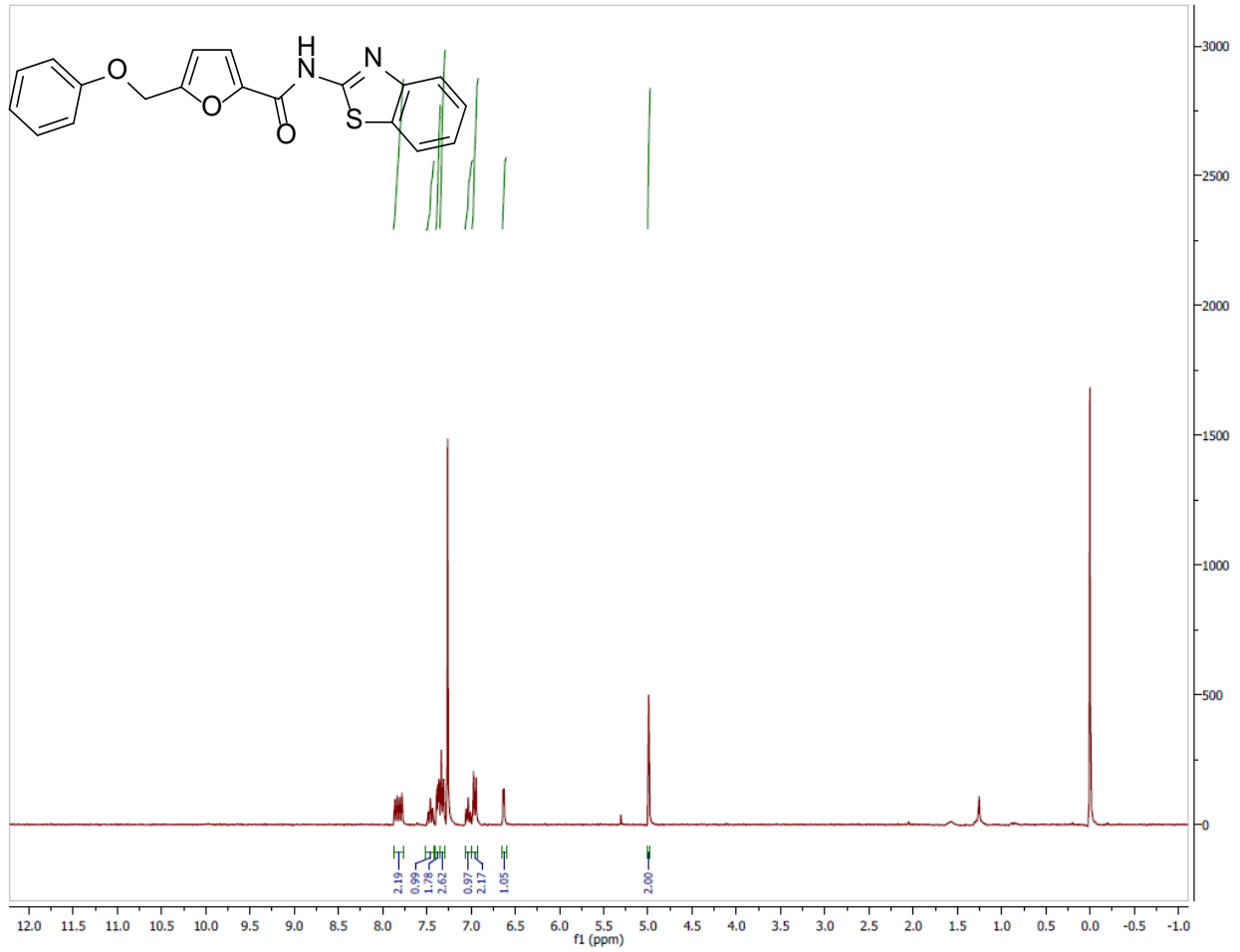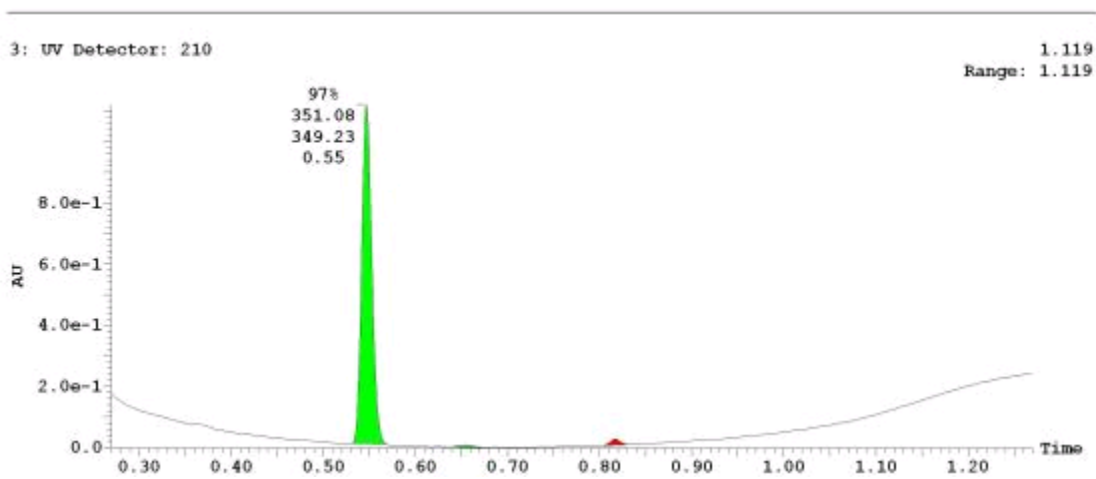
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 29369457
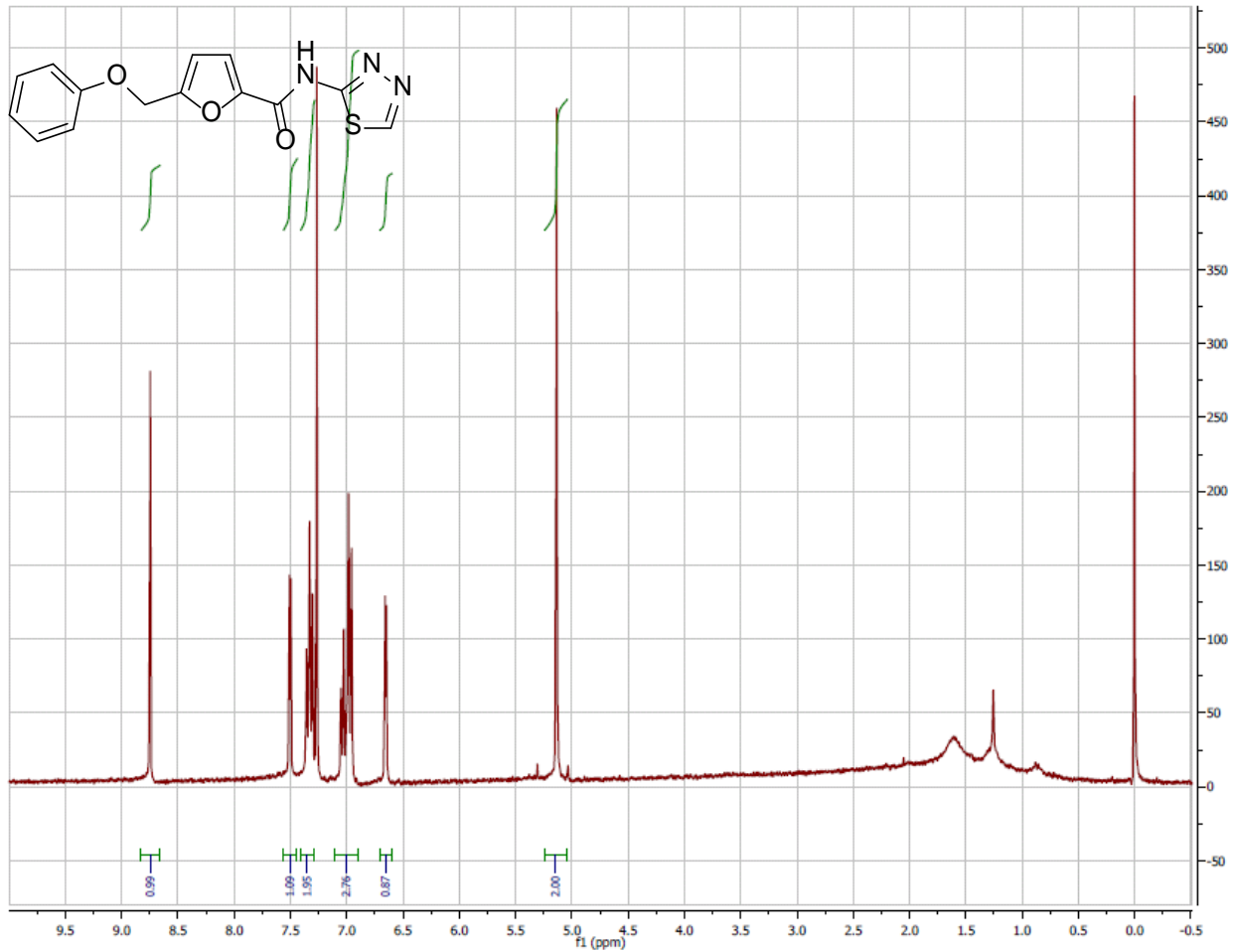
UPLC-MS Chromatogram of Analog CID 29369457
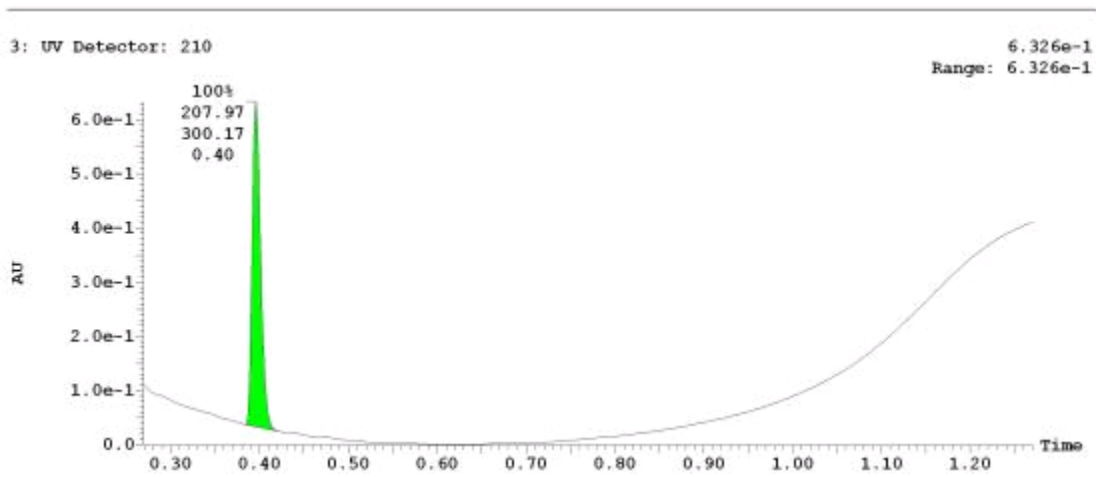
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 50904165
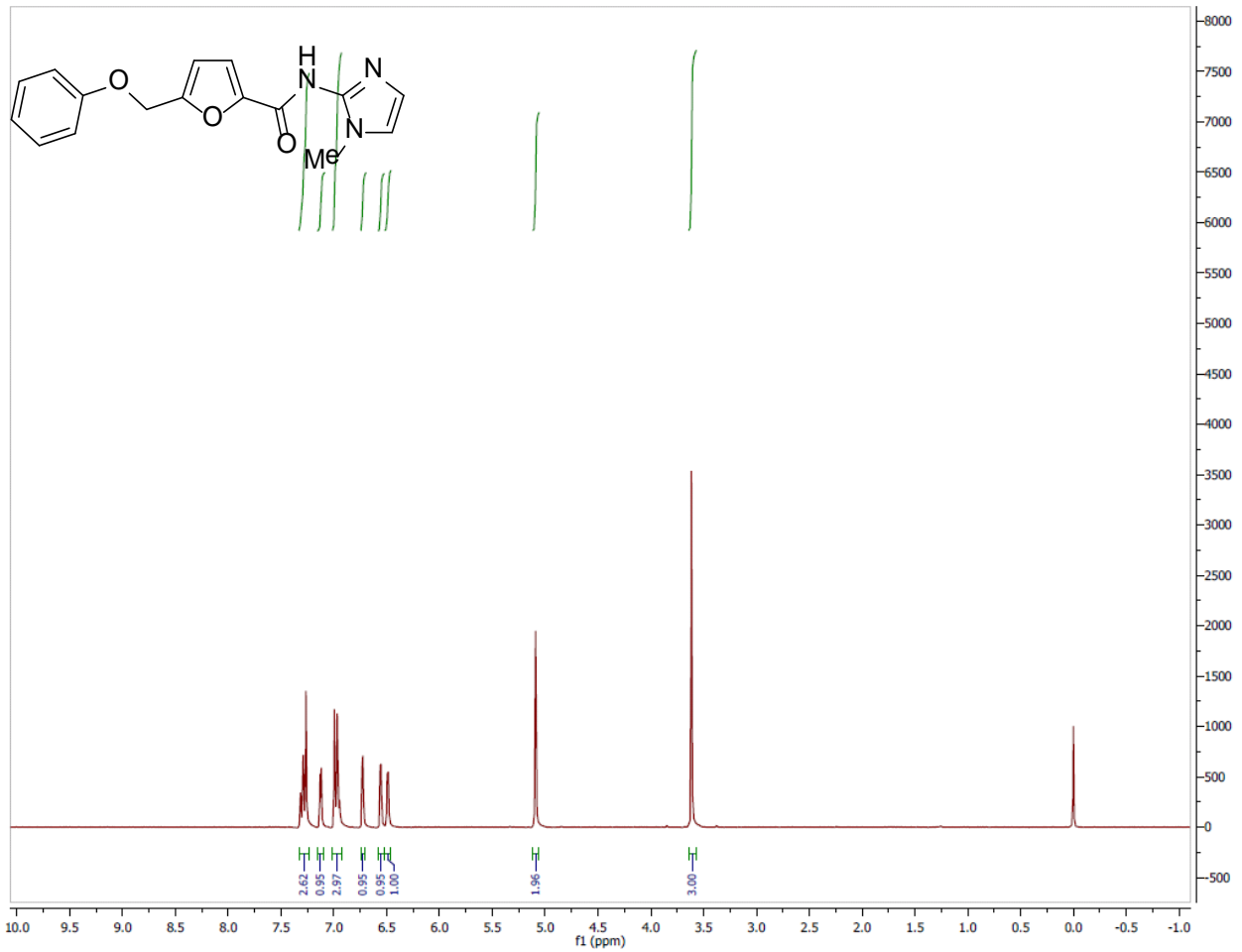
UPLC-MS Chromatogram of Analog CID 50904165

3: UV Detector: 210

4.844e-1
Range: 4.416e-1

100%
298.03
112.96
0.52
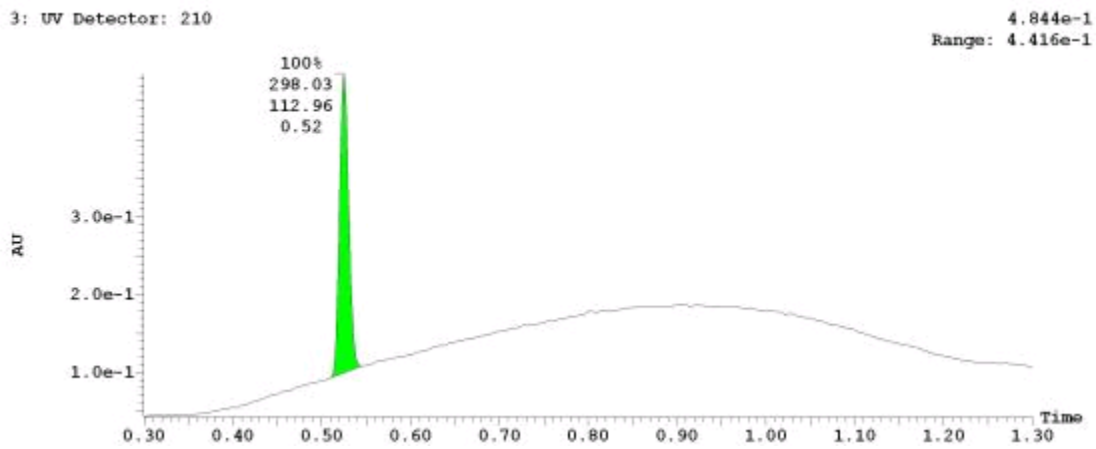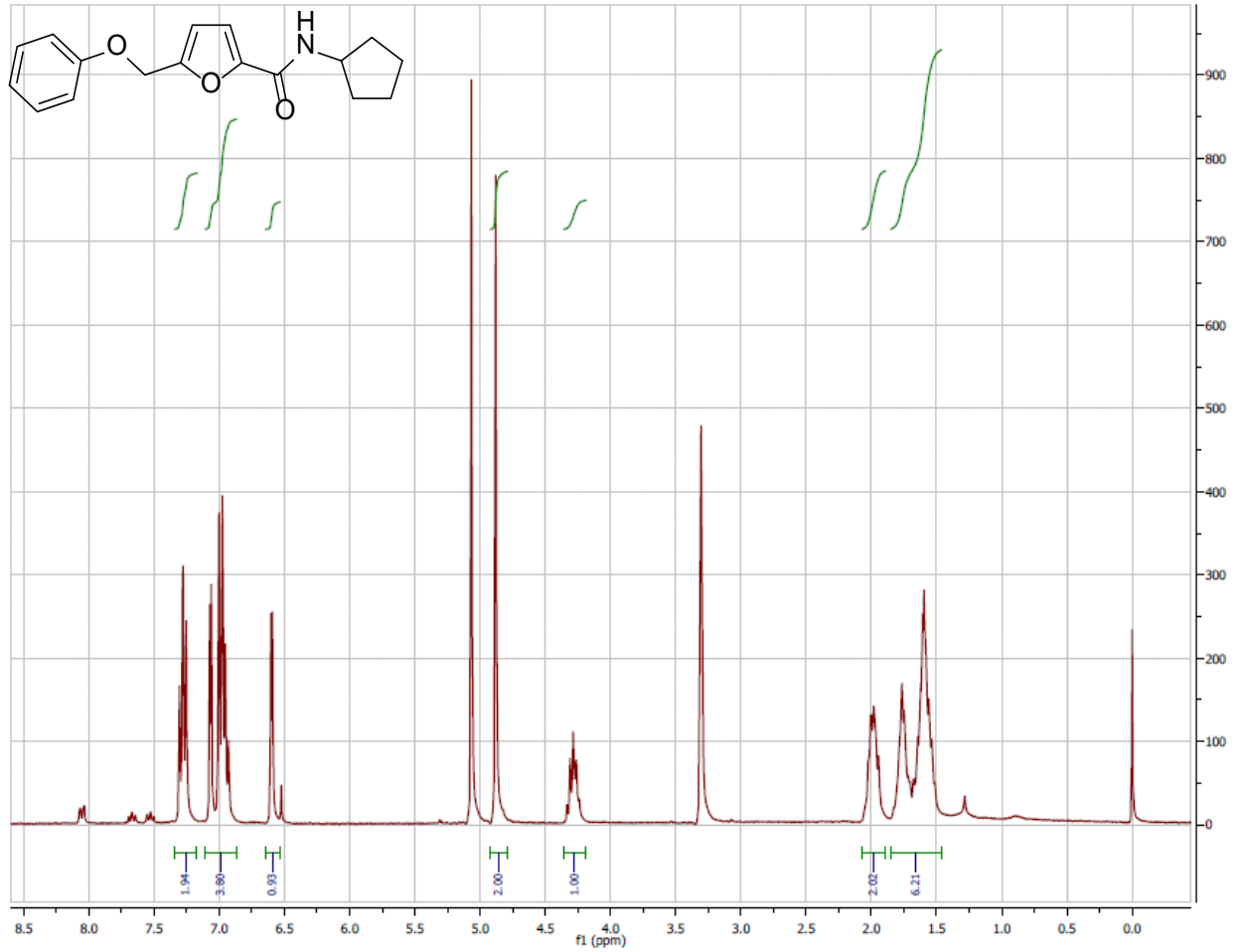
UPLC-MS Chromatogram of Analog CID 9214159

3: UV Detector: 210

6.532e-1
Range: 6.532e-1

88%
285.99
284.18
0.74

12%
147.96
283.35
0.66
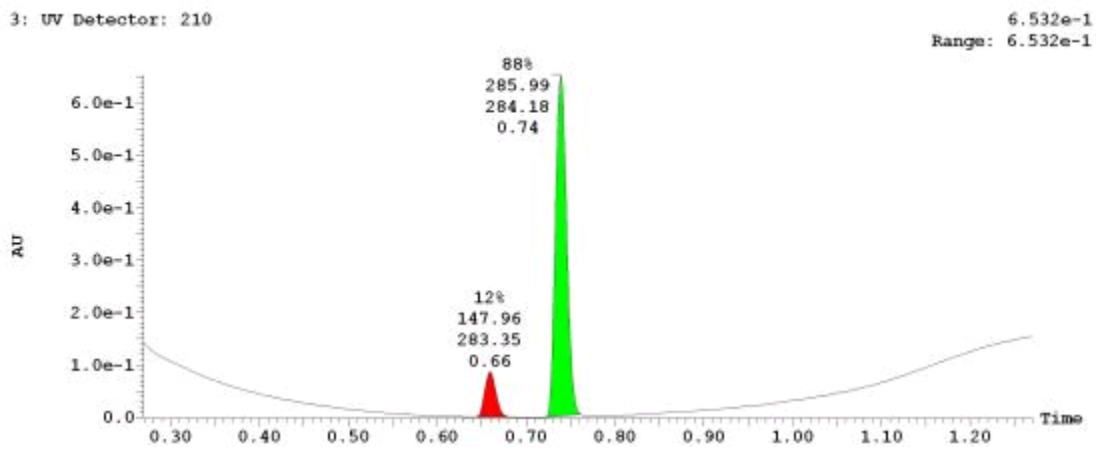
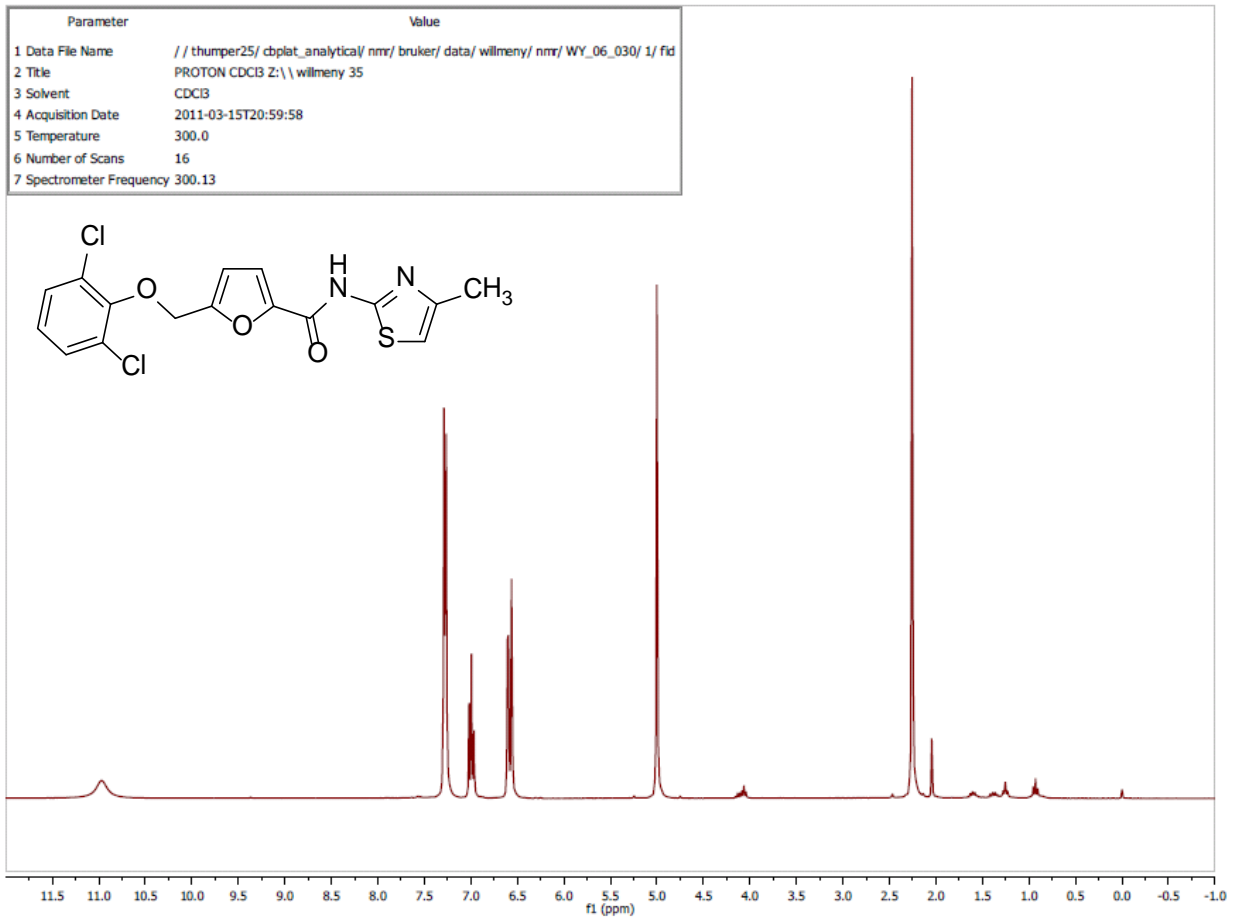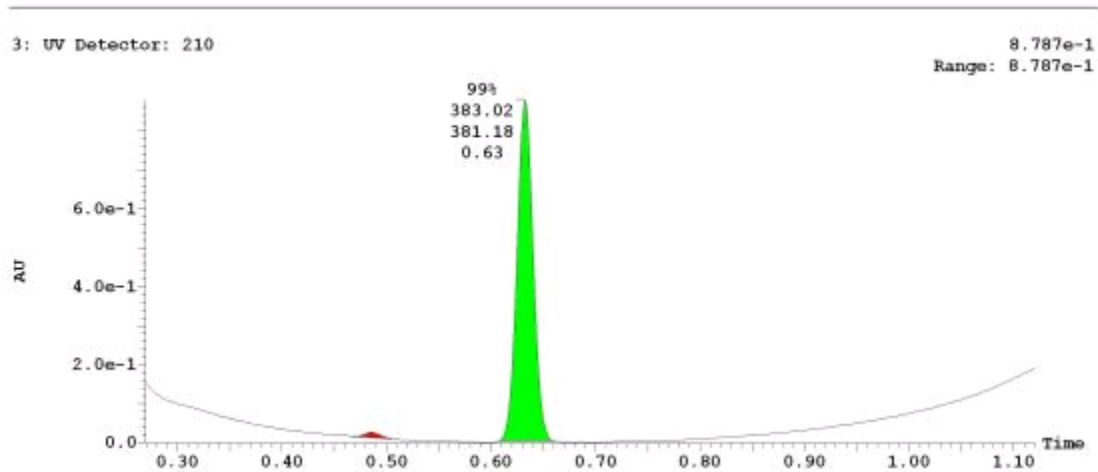# ¹H NMR Spectrum (300 MHz, CDCl₃) of Analog CID 50910550
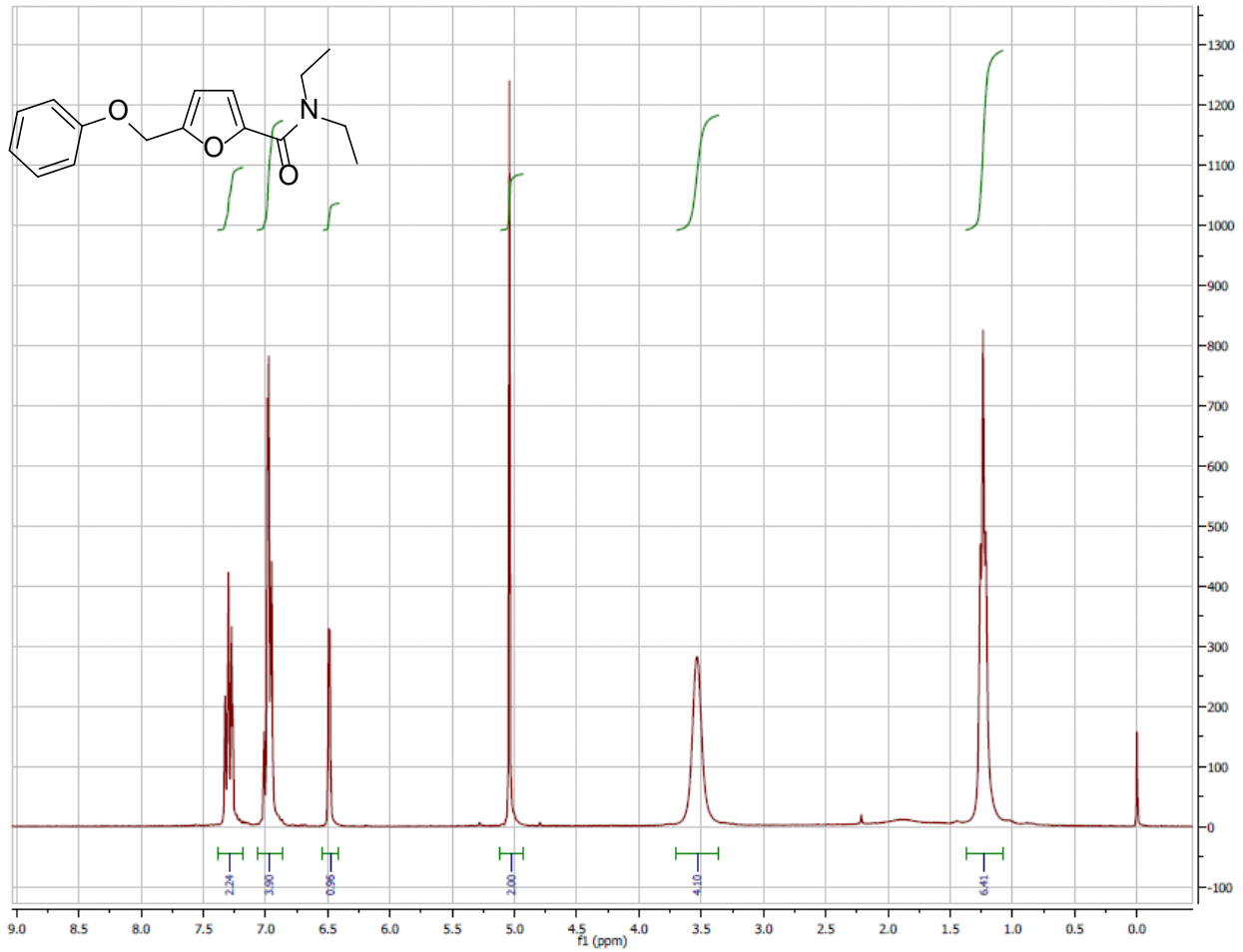
UPLC-MS Chromatogram of Analog CID 50910550
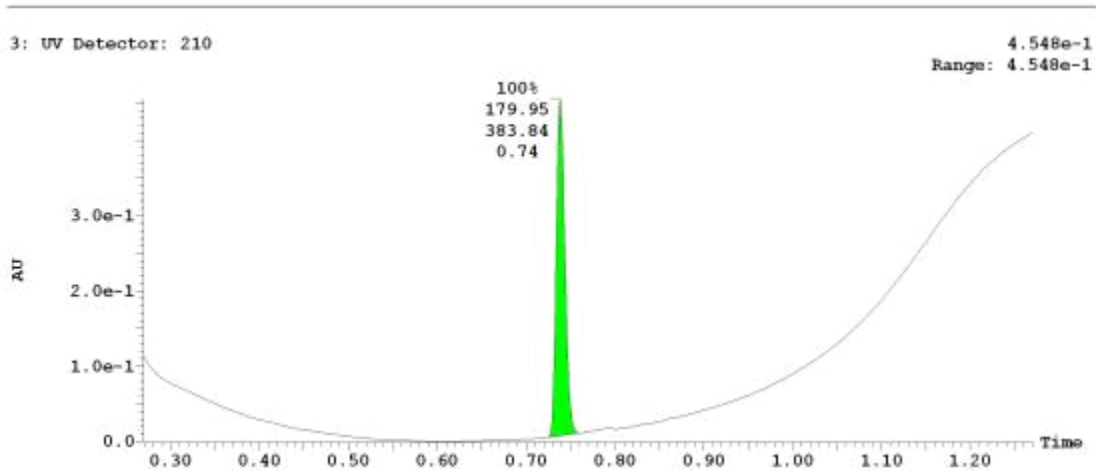
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 50904141
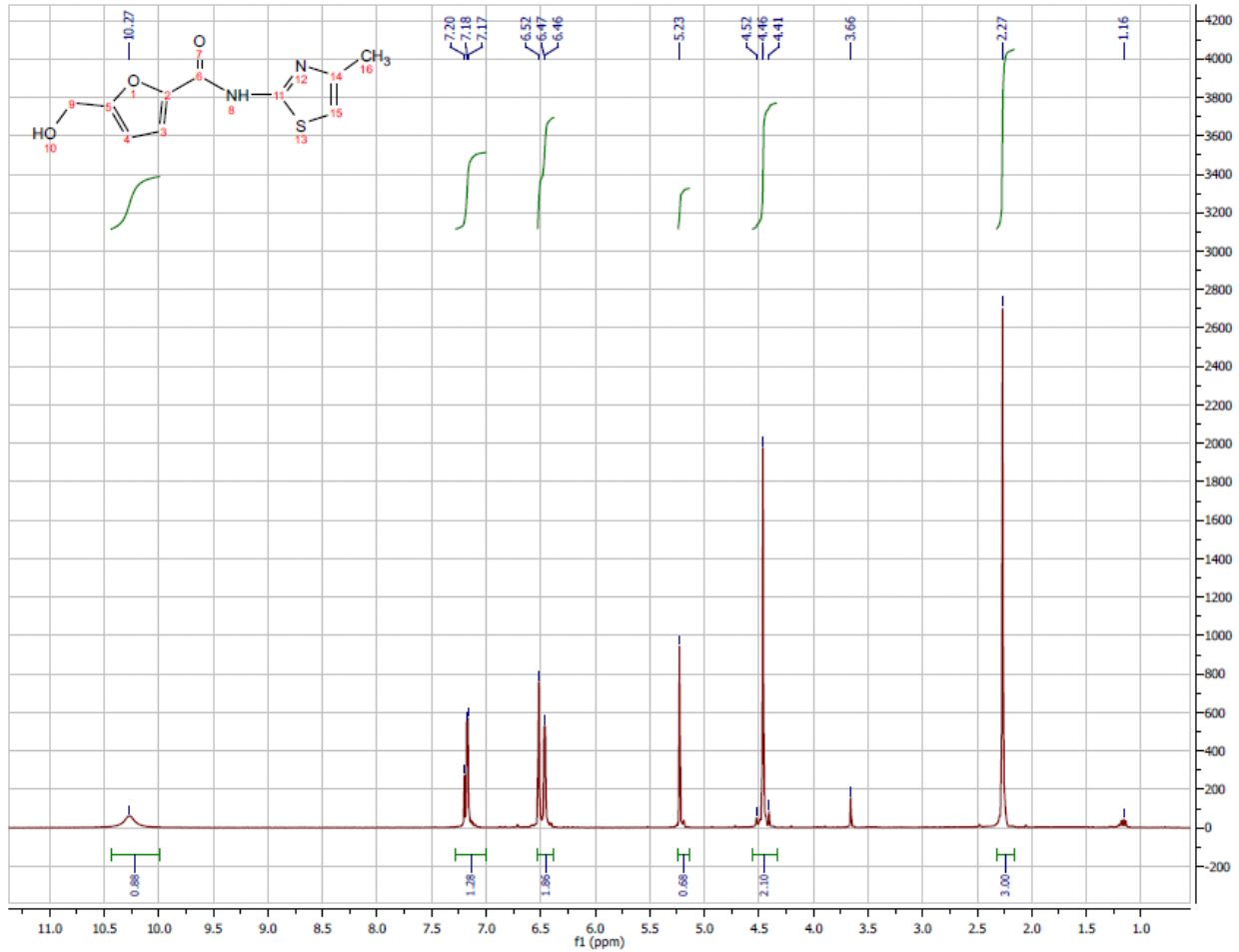
UPLC-MS Chromatogram of Analog CID 50904141 (While the compound is pure as can be seen by NMR, low UV activity results in a poor UPLC trace.)
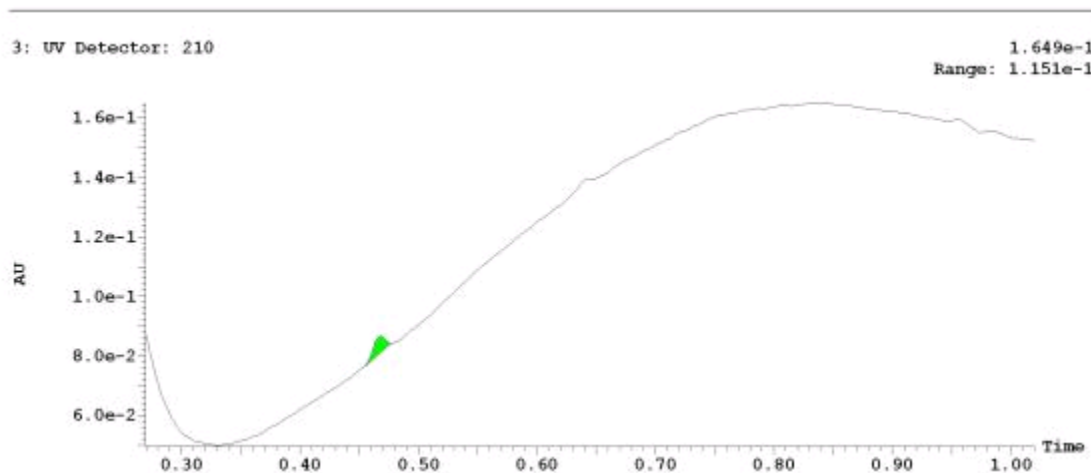
UPLC-MS Chromatogram of Analog CID 50904162
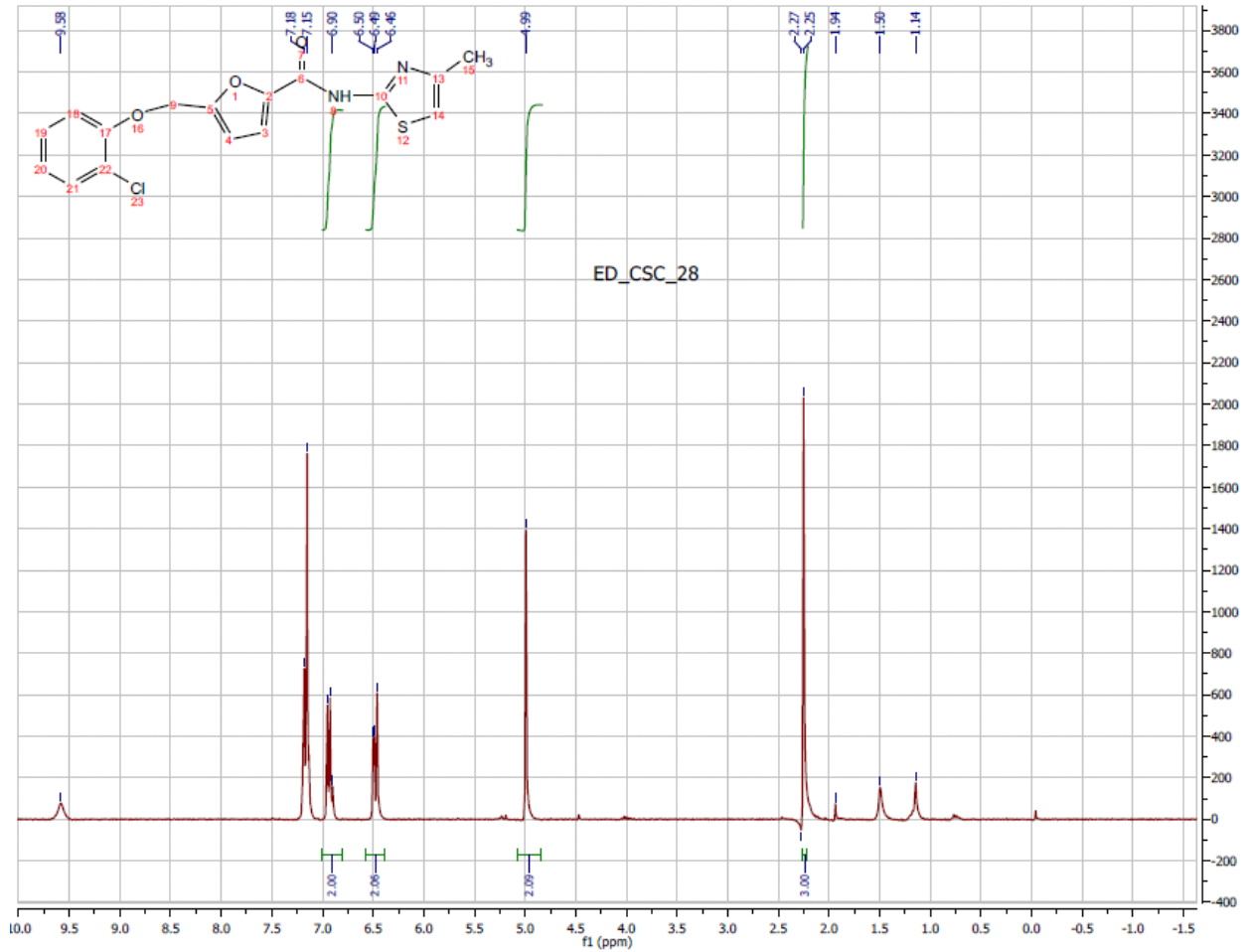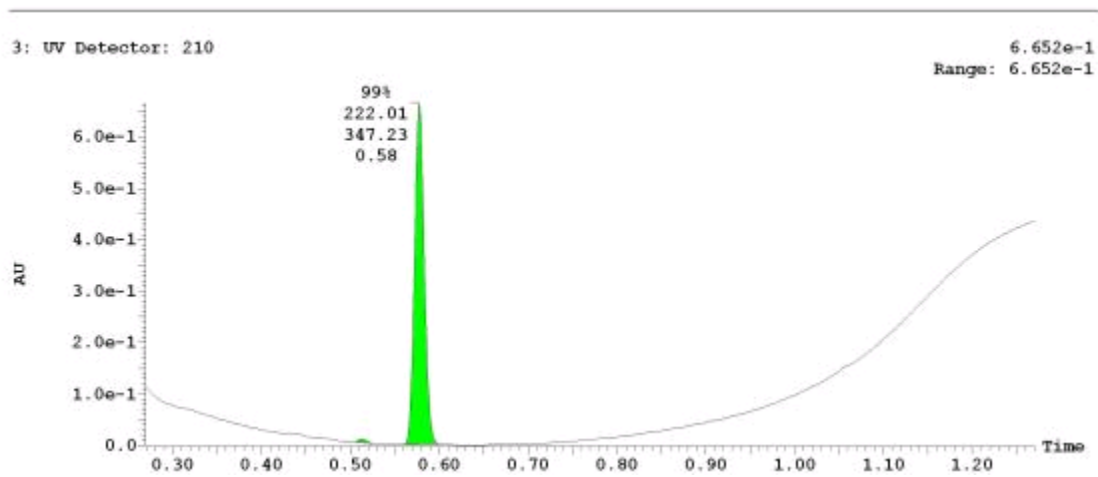
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 50904129
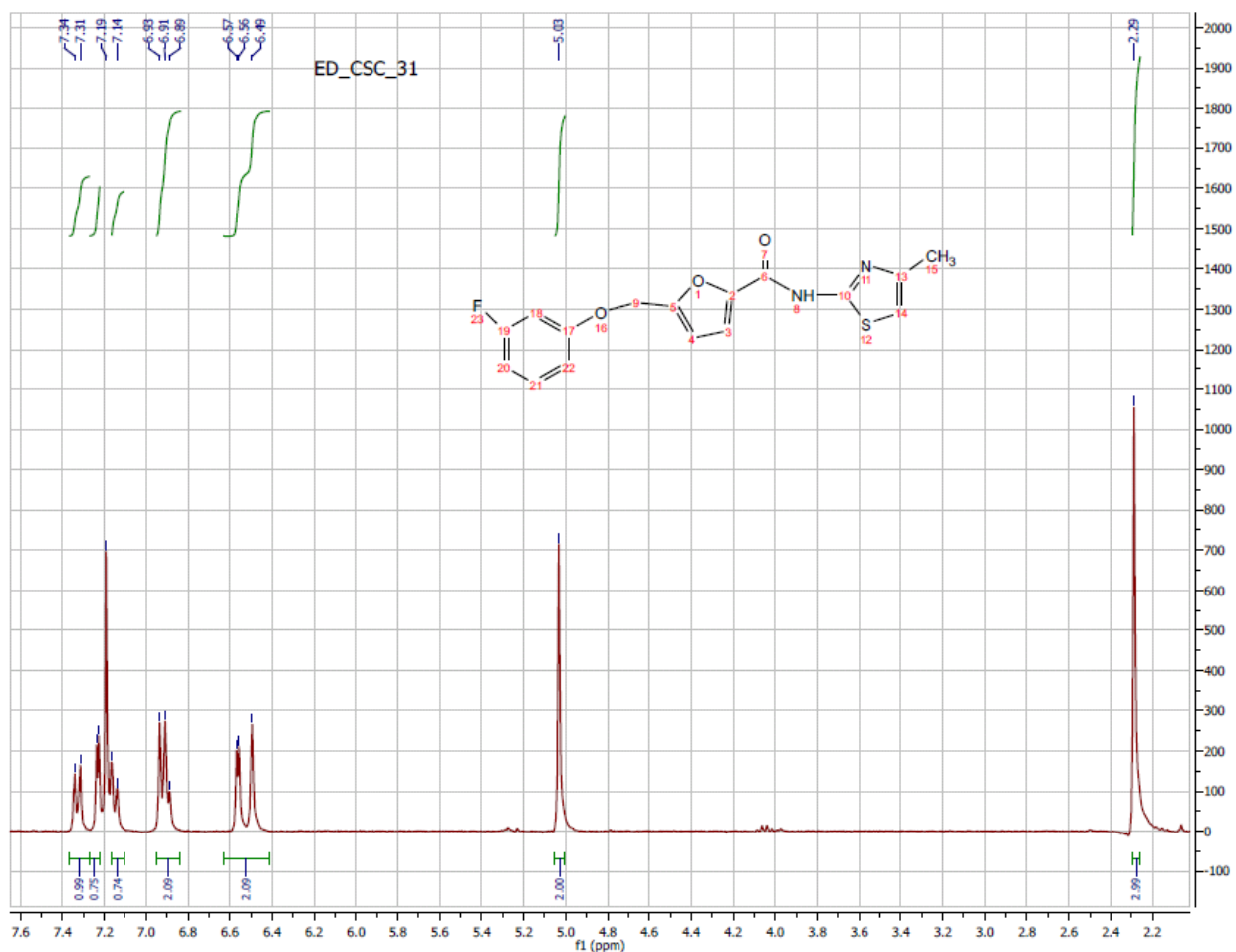
ED_CSC_31

UPLC-MS Chromatogram of Analog CID 50904129

3: UV Detector: 210

4.438e-1
Range: 4.437e-1

97%
222.02
347.21
0.58

3%
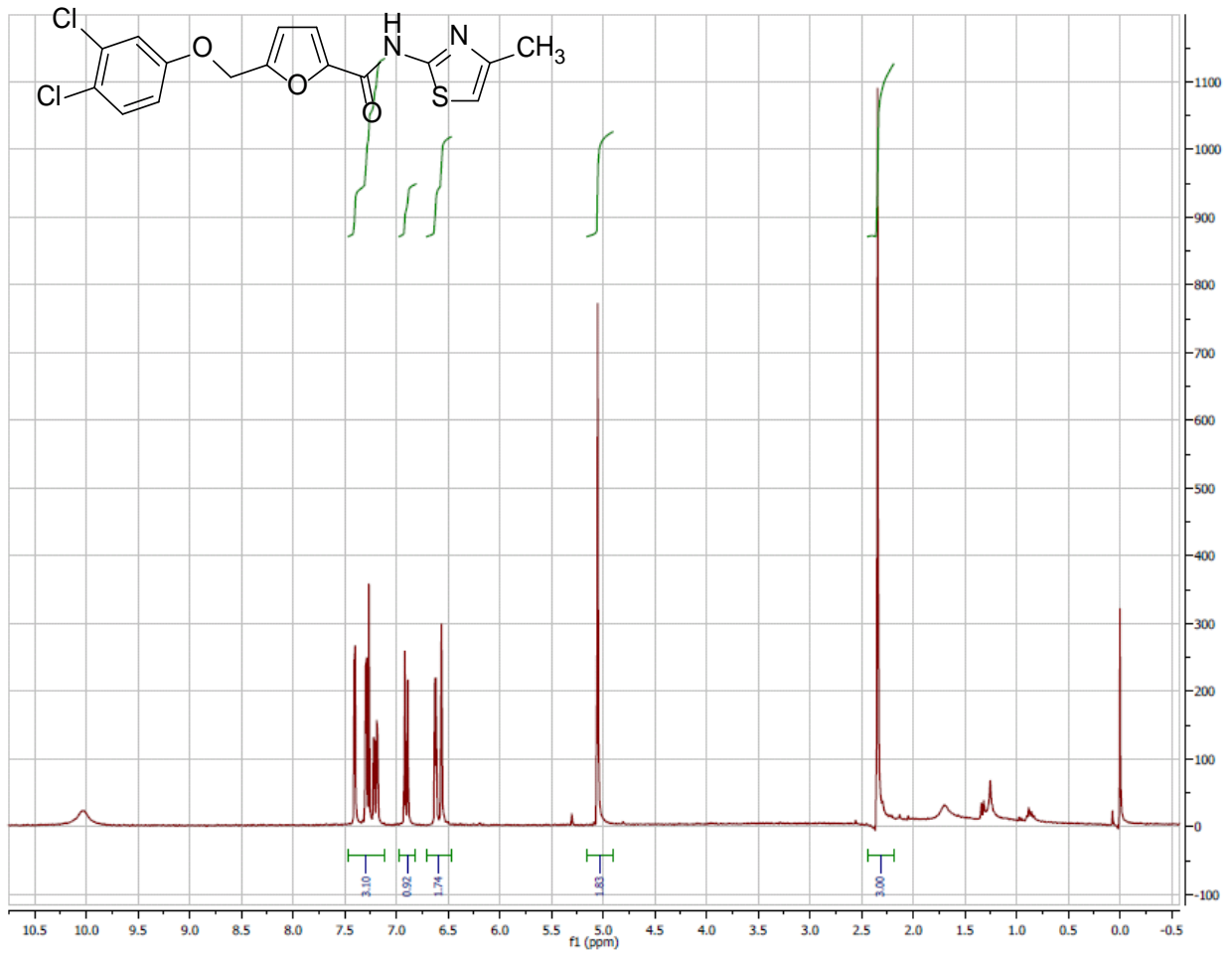349.06
347.22
0.44
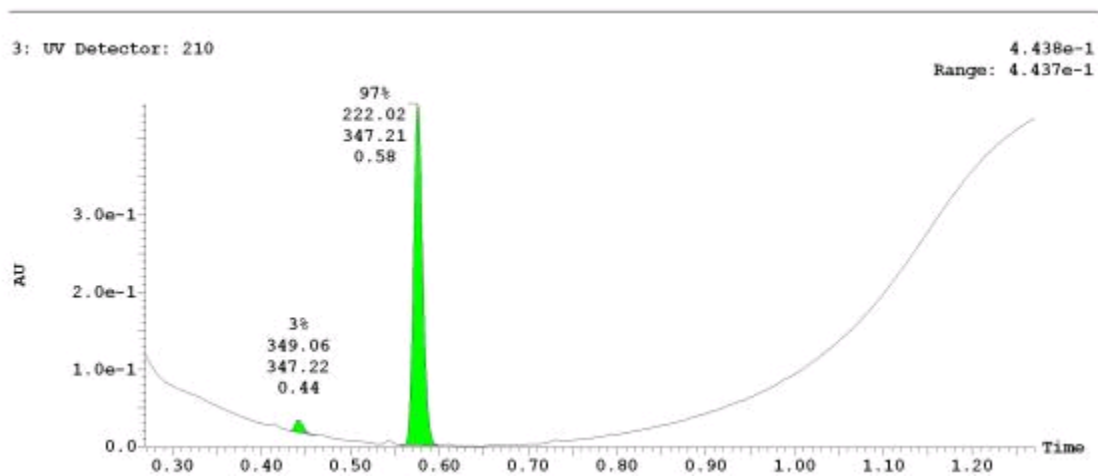
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 50904156
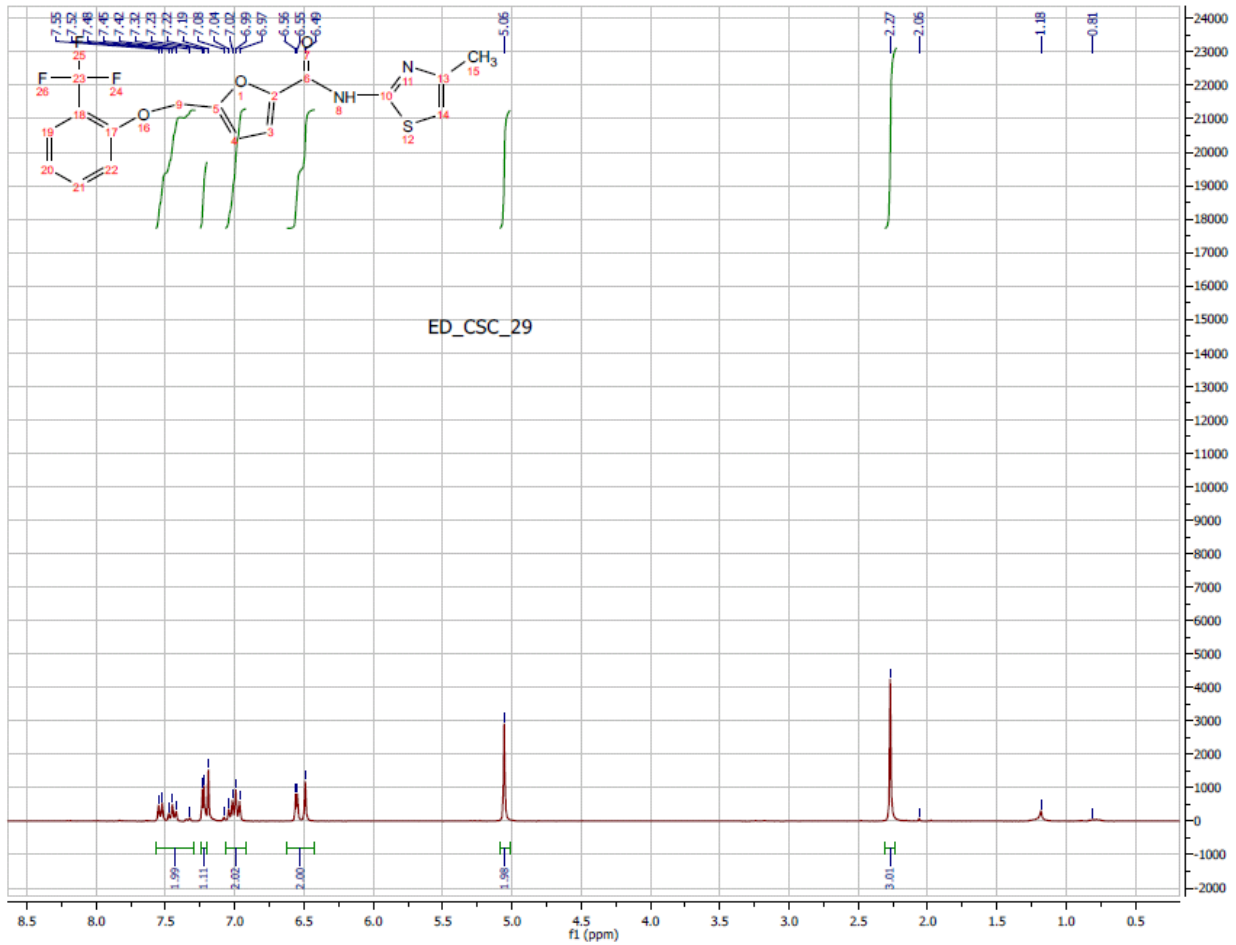
UPLC-MS Chromatogram of Analog CID 50904156
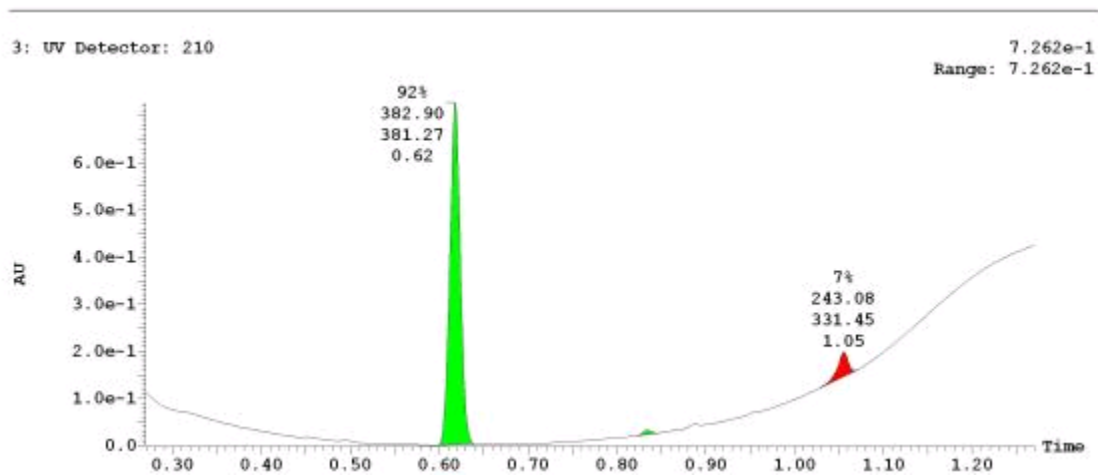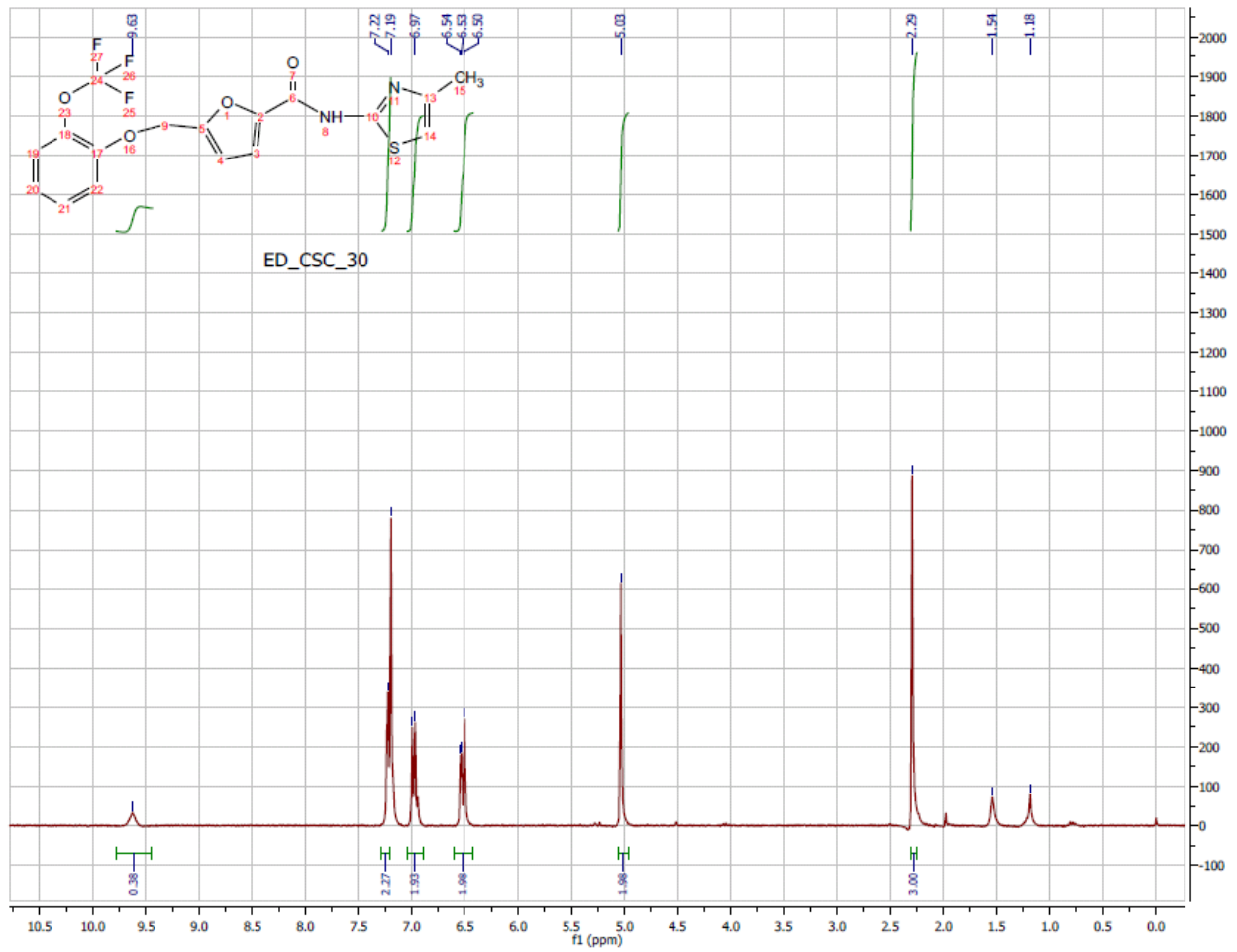
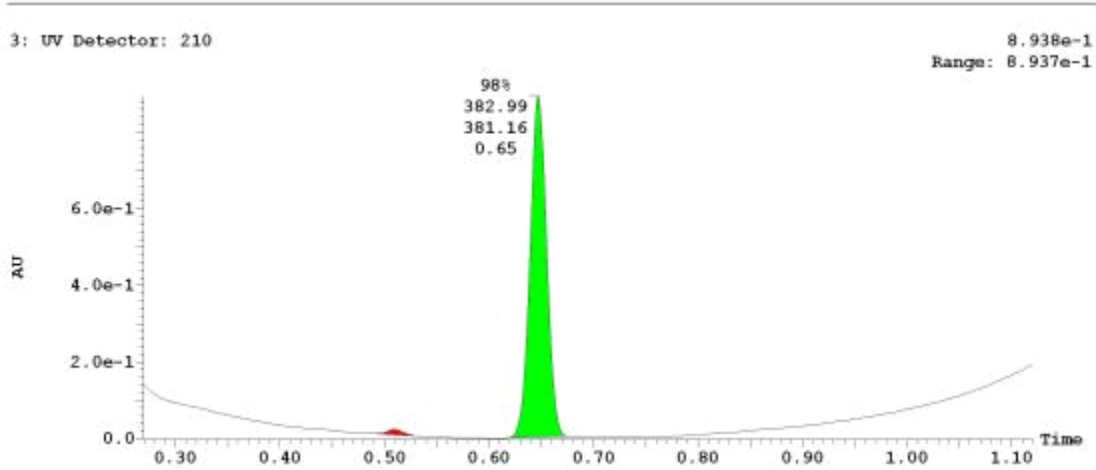# ¹H NMR Spectrum (300 MHz, CDCl₃) of Analog CID 50904164

 # ^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 50904164

ED_CSC_30

UPLC-MS Chromatogram of Analog CID 50904164

3: UV Detector: 210

4.61e-1
Range: 4.61e-1

100%
222.01
397.28
0.63

^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 50910527

ED_CSC_51

UPLC-MS Chromatogram of Analog CID 50910527

3: UV Detector: 210

9.687e-1
Range: 9.686e-1

100%
220.96
327.24
0.59

¹H NMR Spectrum (300 MHz, CDCl₃) of Analog CID 50910522

^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 50910522

UPLC-MS Chromatogram of Analog CID 50910522
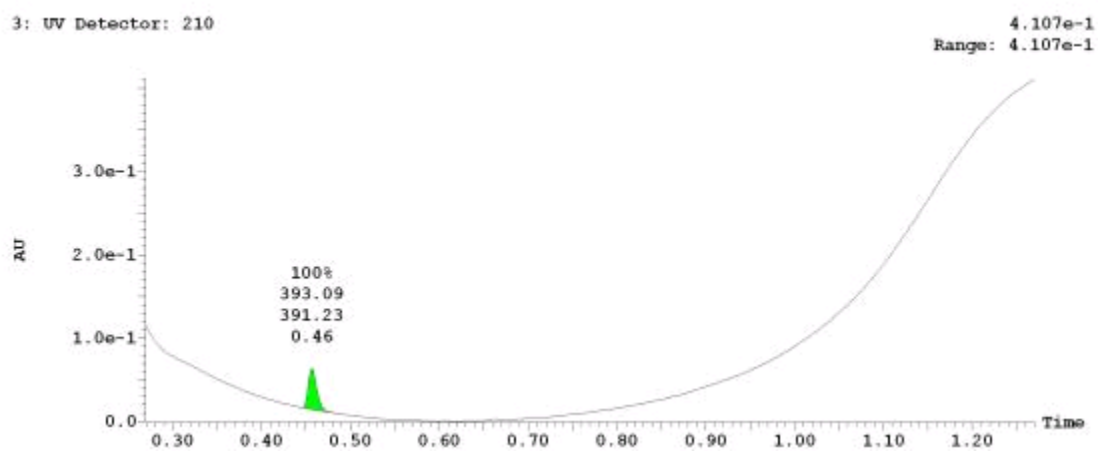
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 50904146
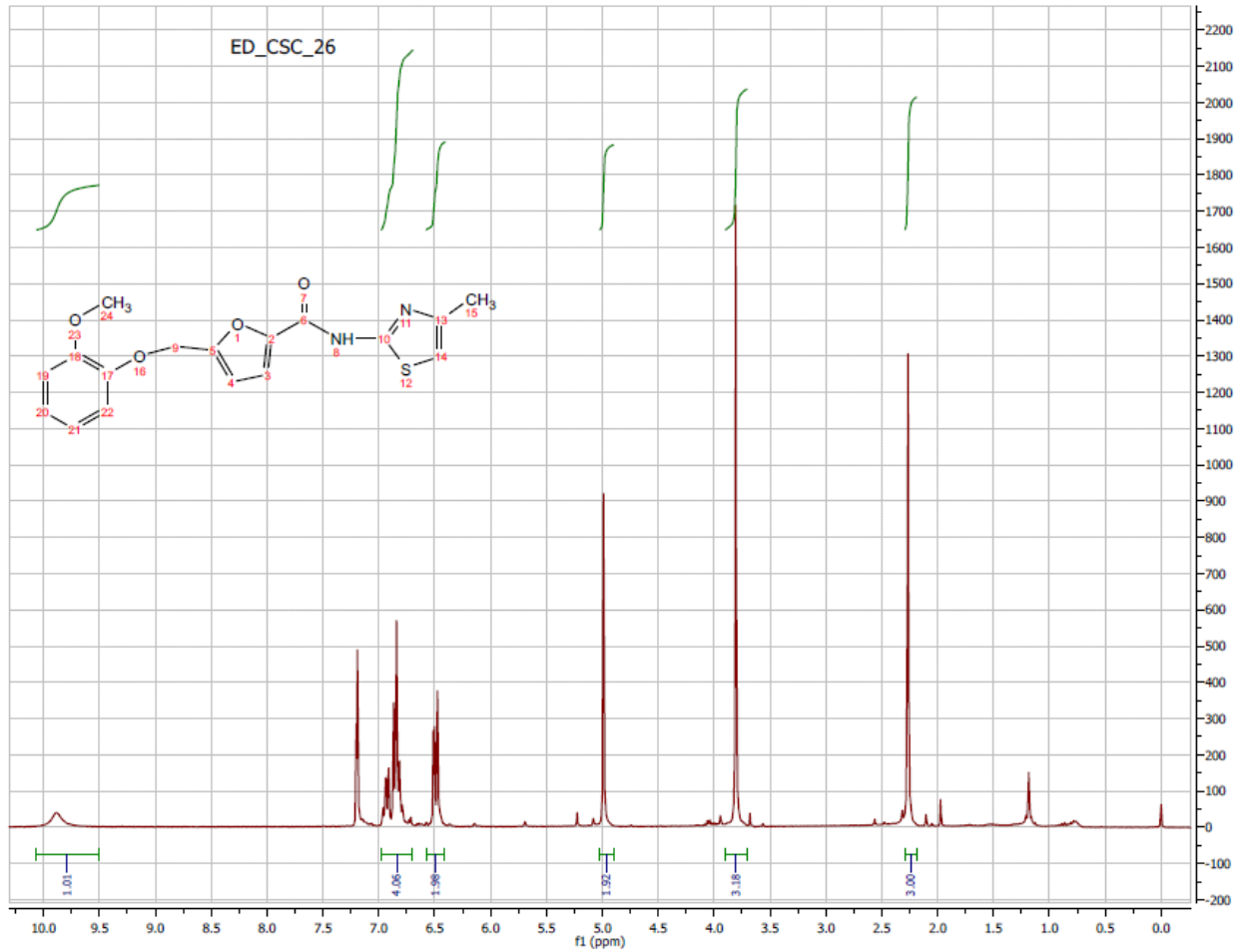
UPLC-MS Chromatogram of Analog CID 50904146
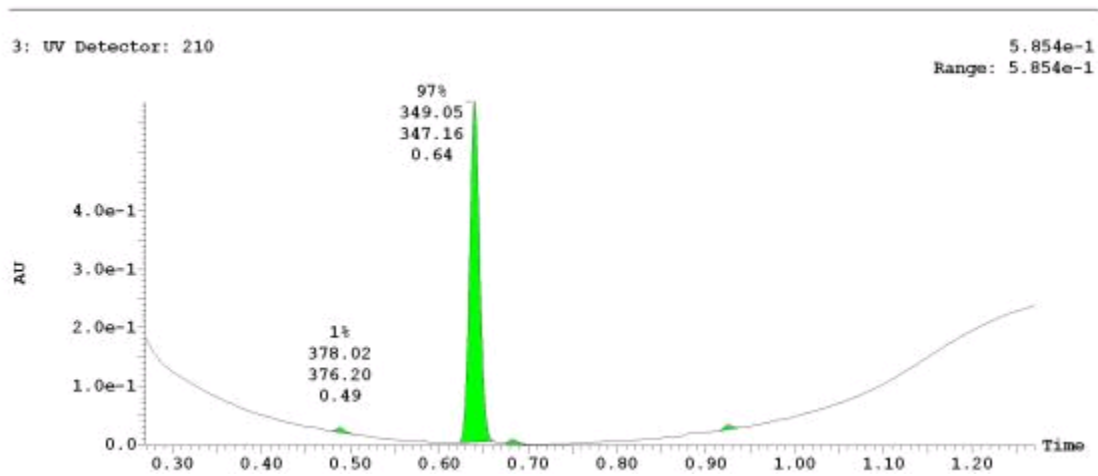
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 50904145
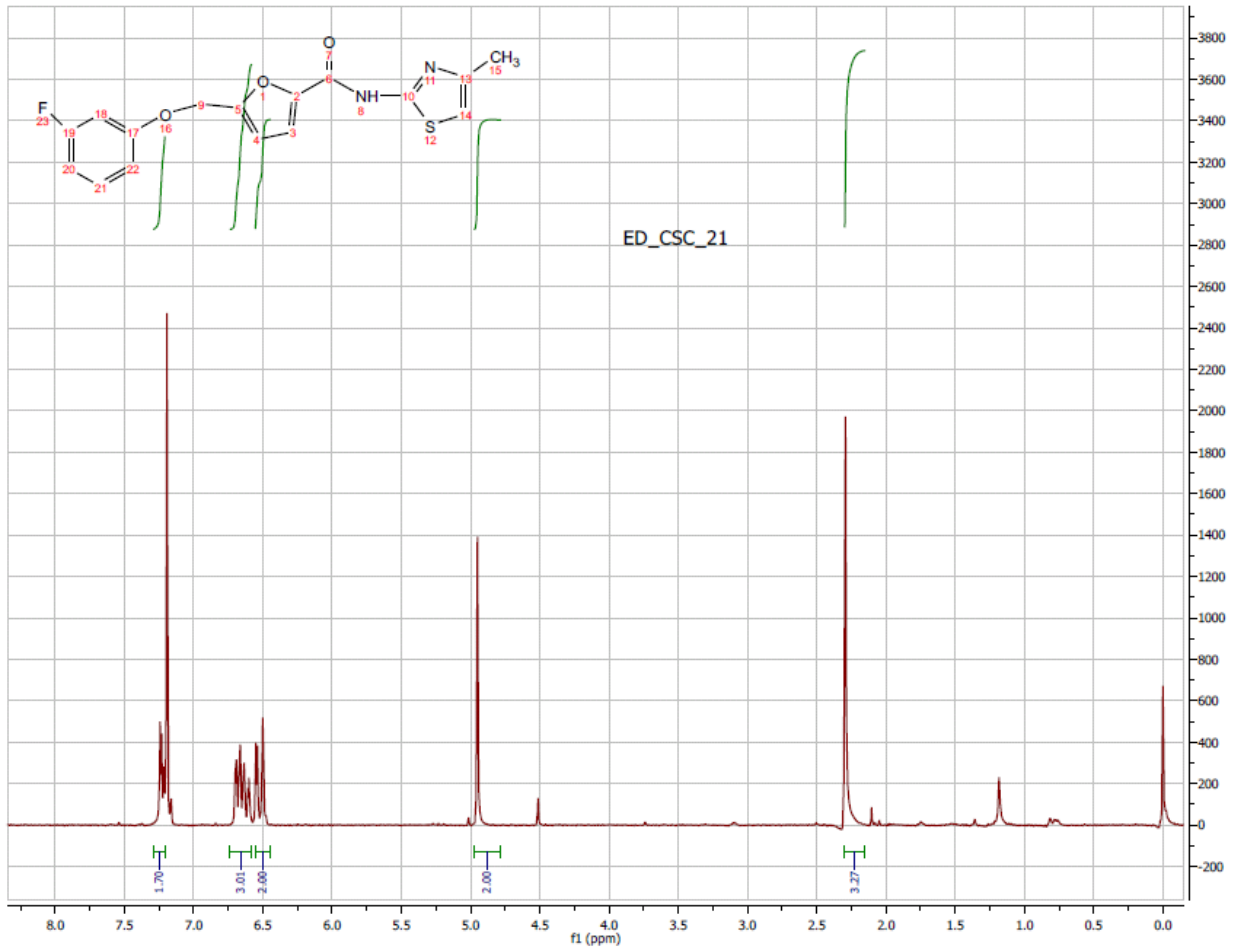
ED_CSC_21

UPLC-MS Chromatogram of Analog CID 50904145

3: UV Detector: 210

95%
332.97
331.21
0.55

5%
239.03
237.14
0.33

4.216e-1
Range: 4.215e-1
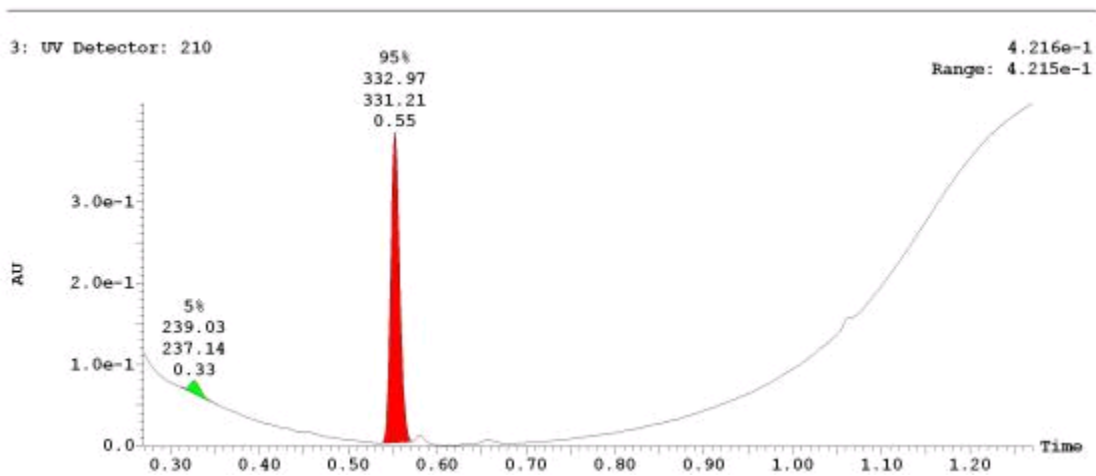
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 50904153

UPLC-MS Chromatogram of Analog CID 50904153
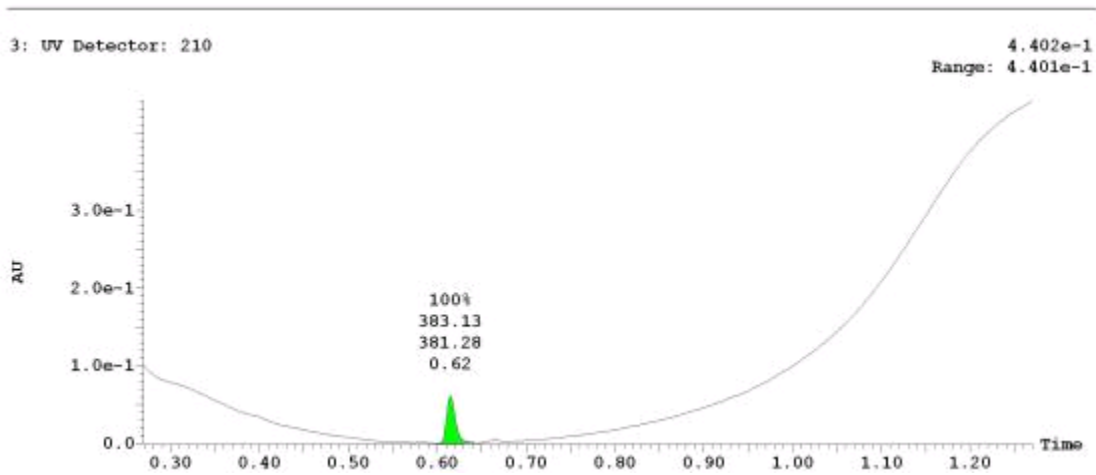
3: UV Detector: 210

4.402e-1
Range: 4.401e-1

100%
383.13
381.28
0.62
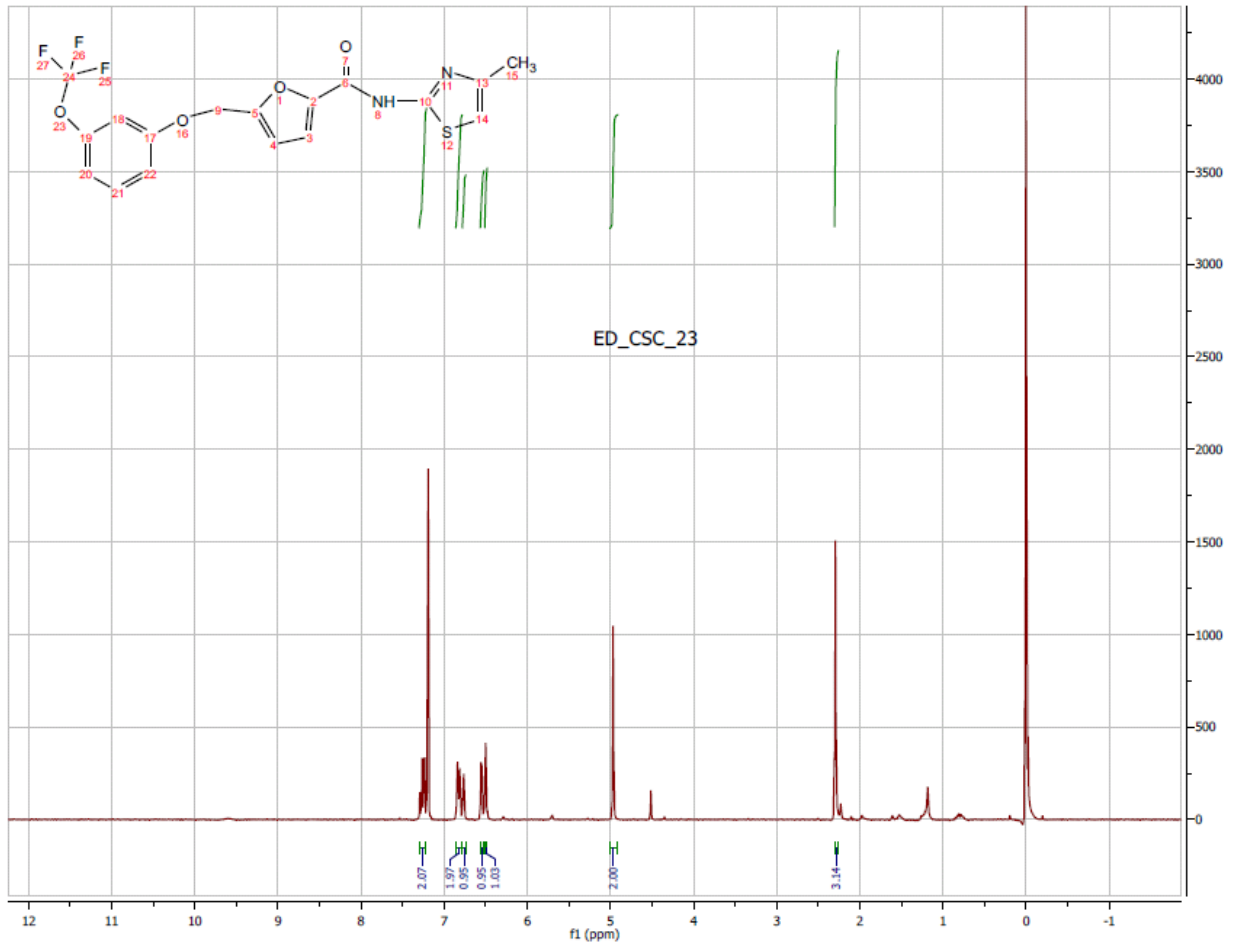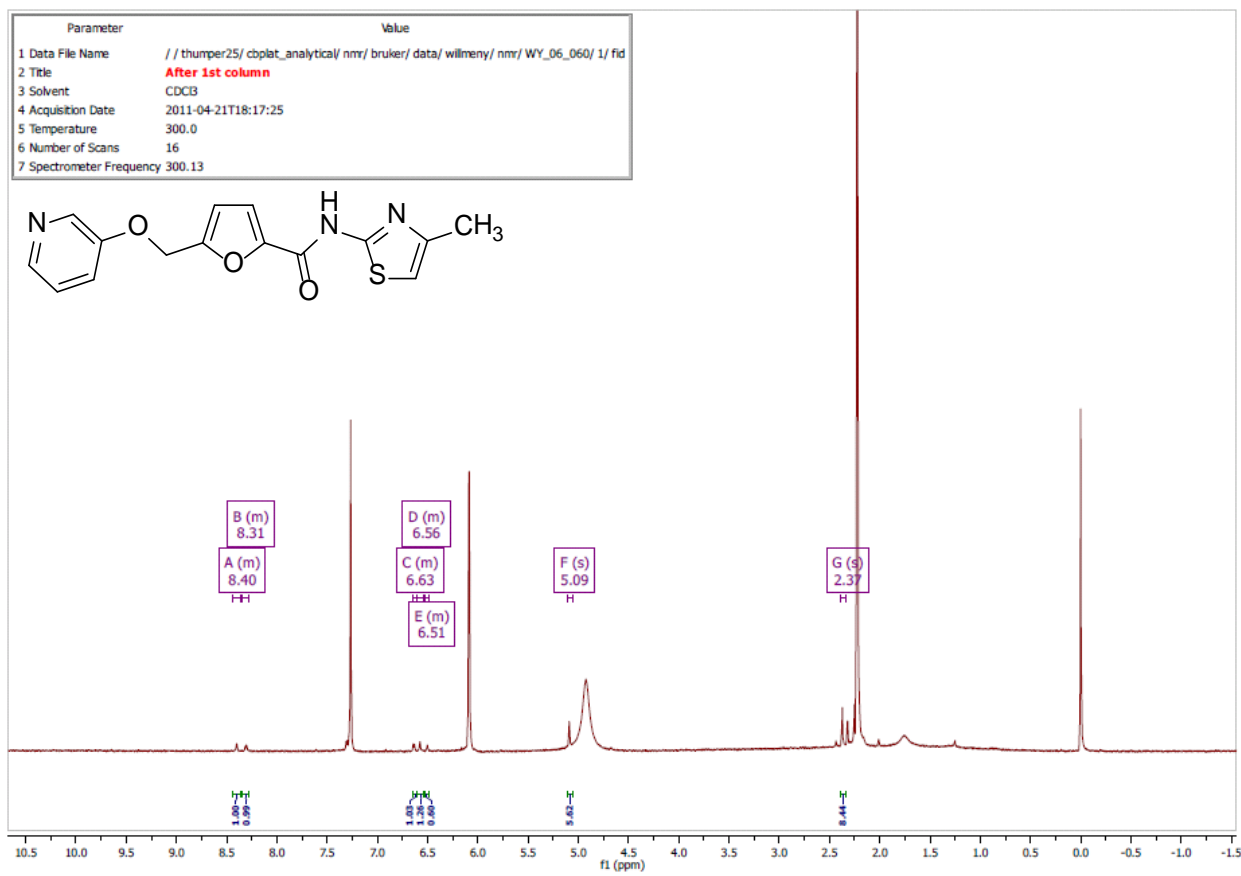
UPLC-MS Chromatogram of Analog CID 50904144
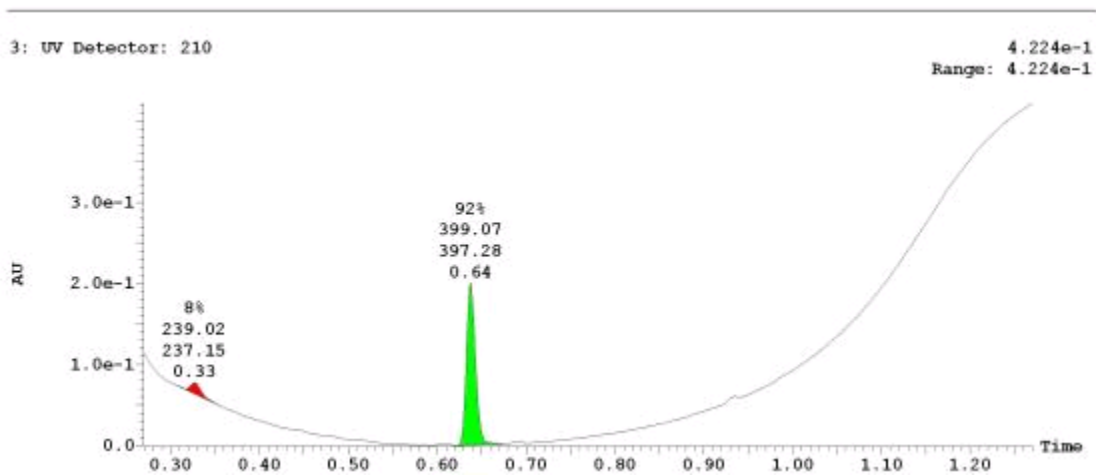
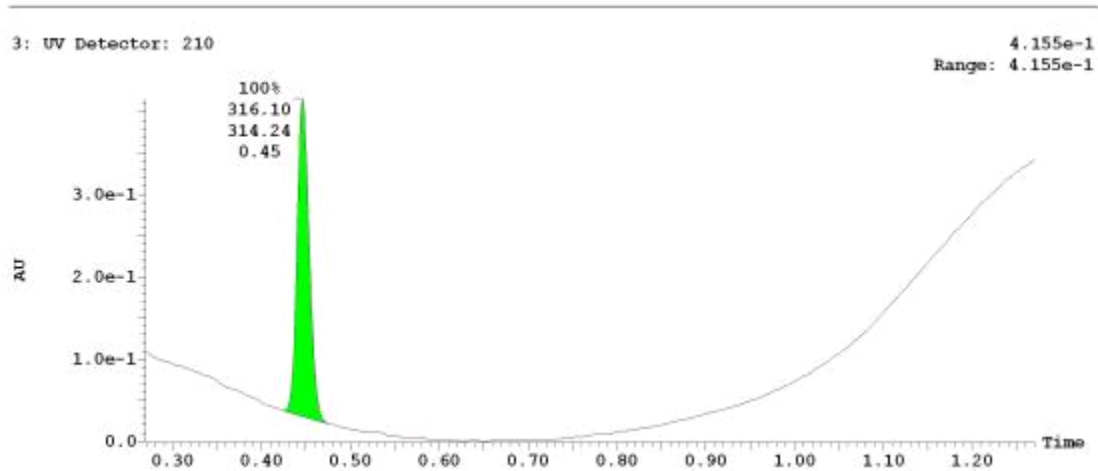**¹H NMR Spectrum (300 MHz, CDCl₃) of Analog CID 50910527**
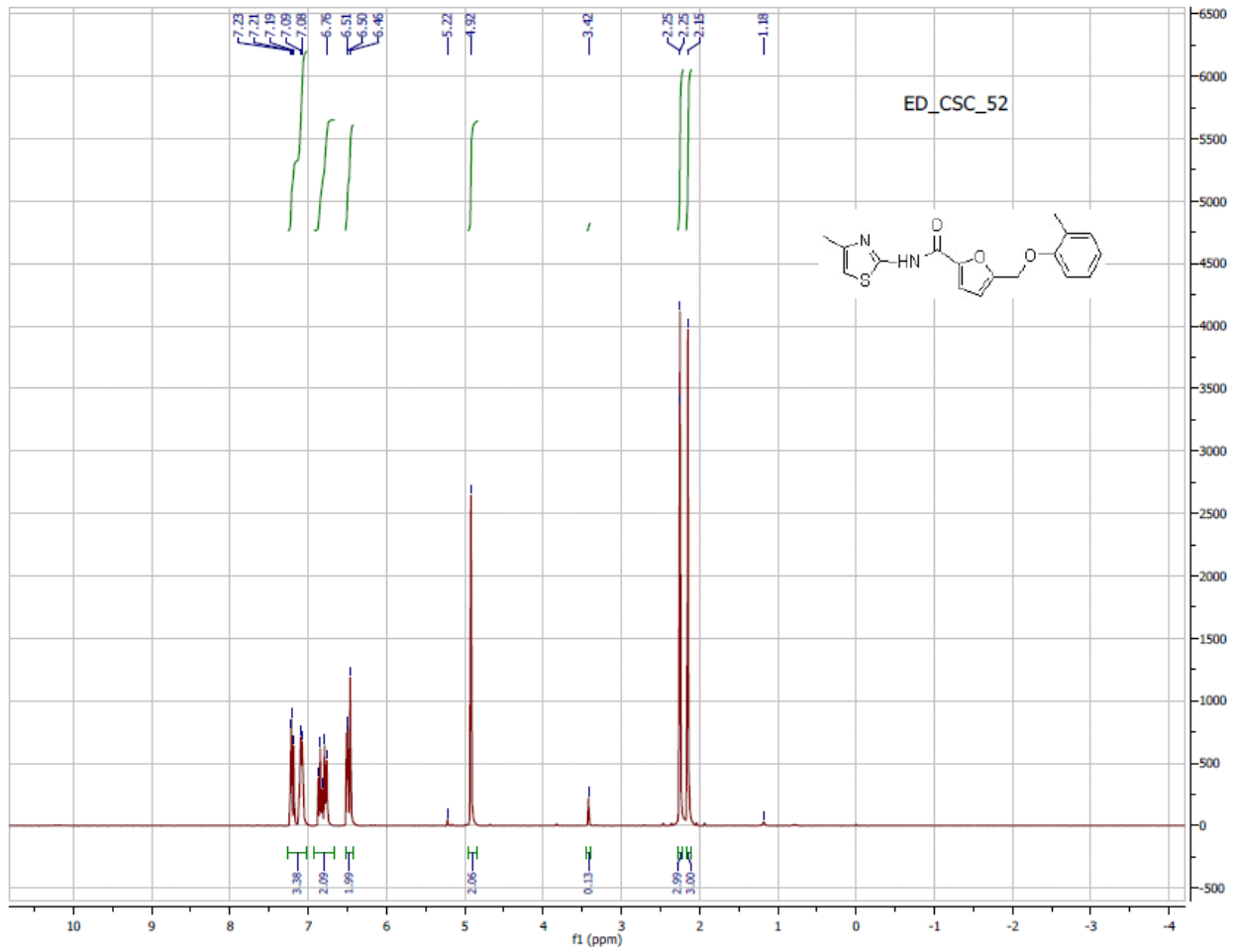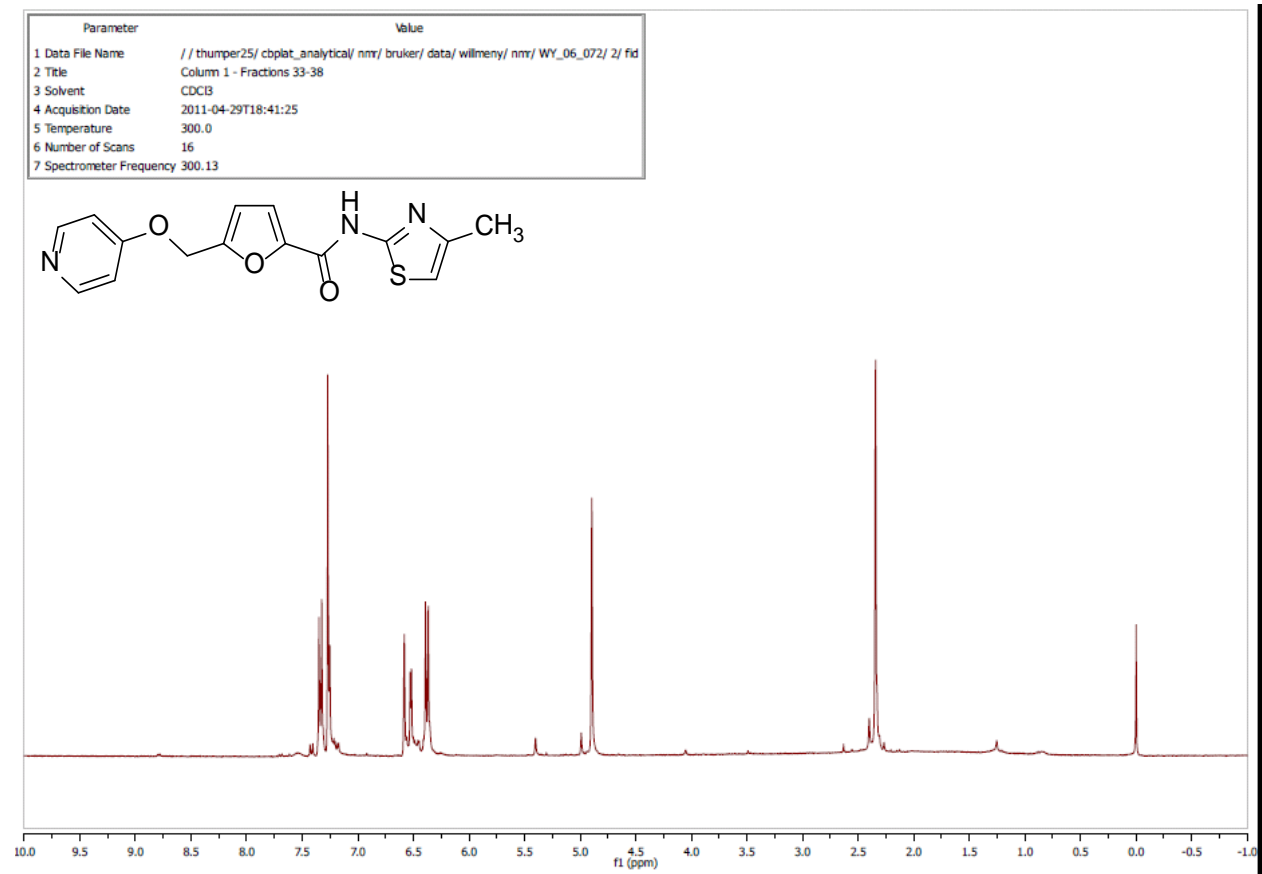

UPLC-MS Chromatogram of Analog CID 50910527
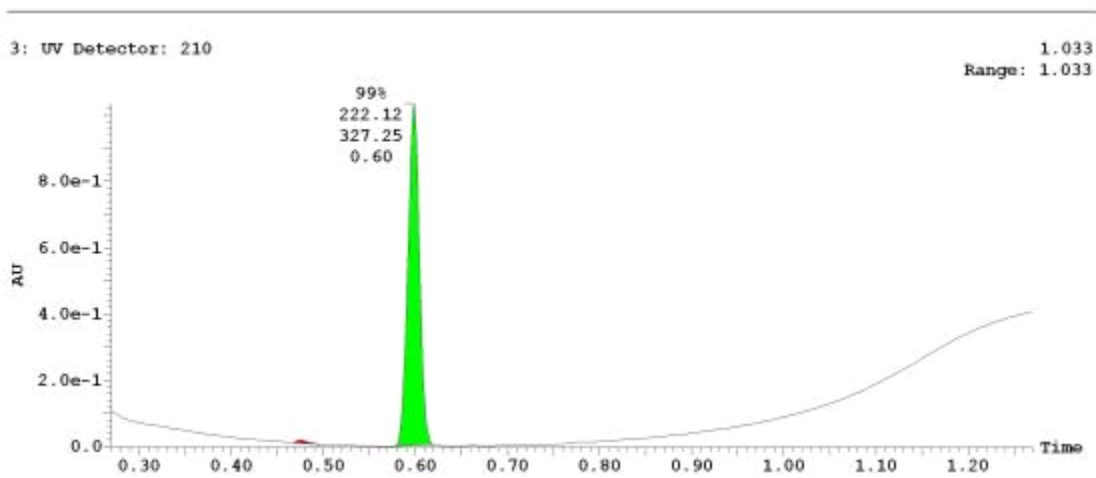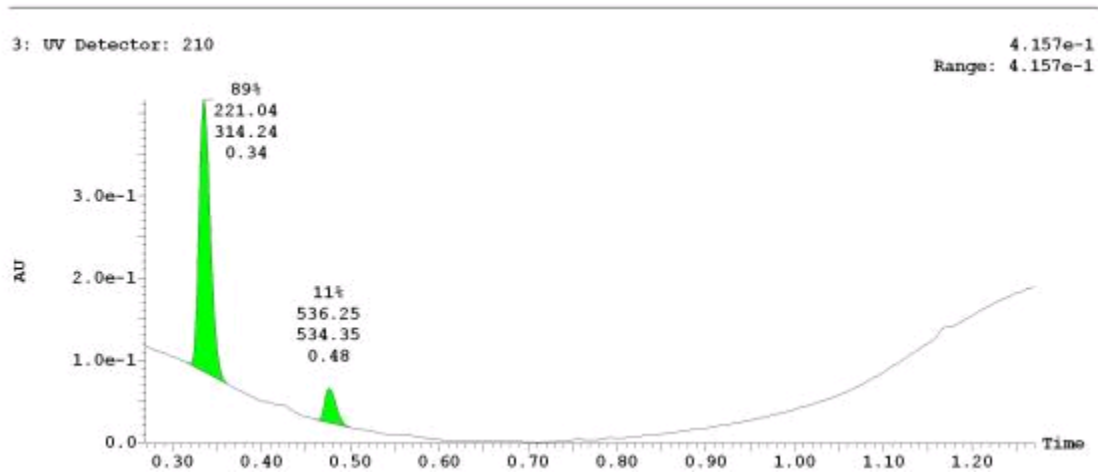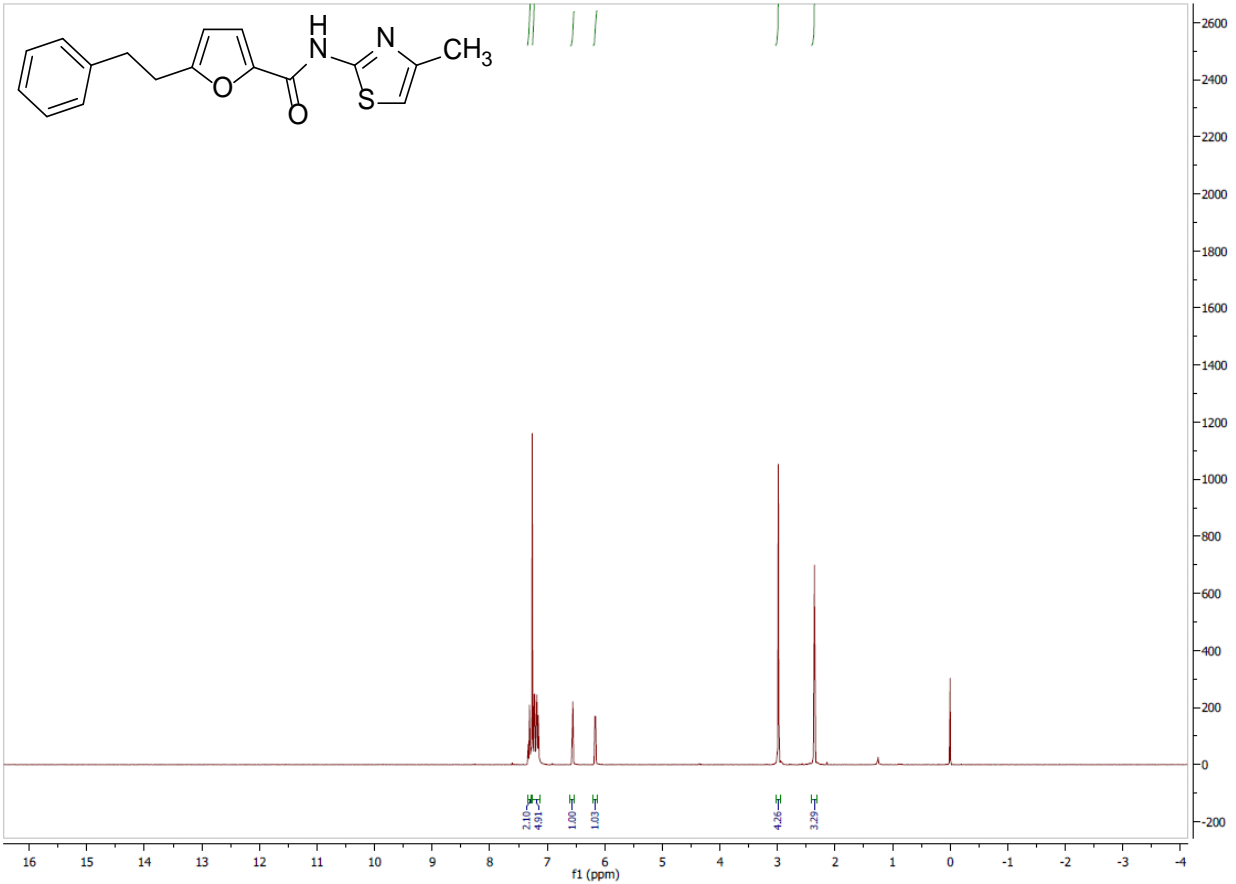
¹H NMR Spectrum (300 MHz, CDCl₃) of Analog CID 51003708
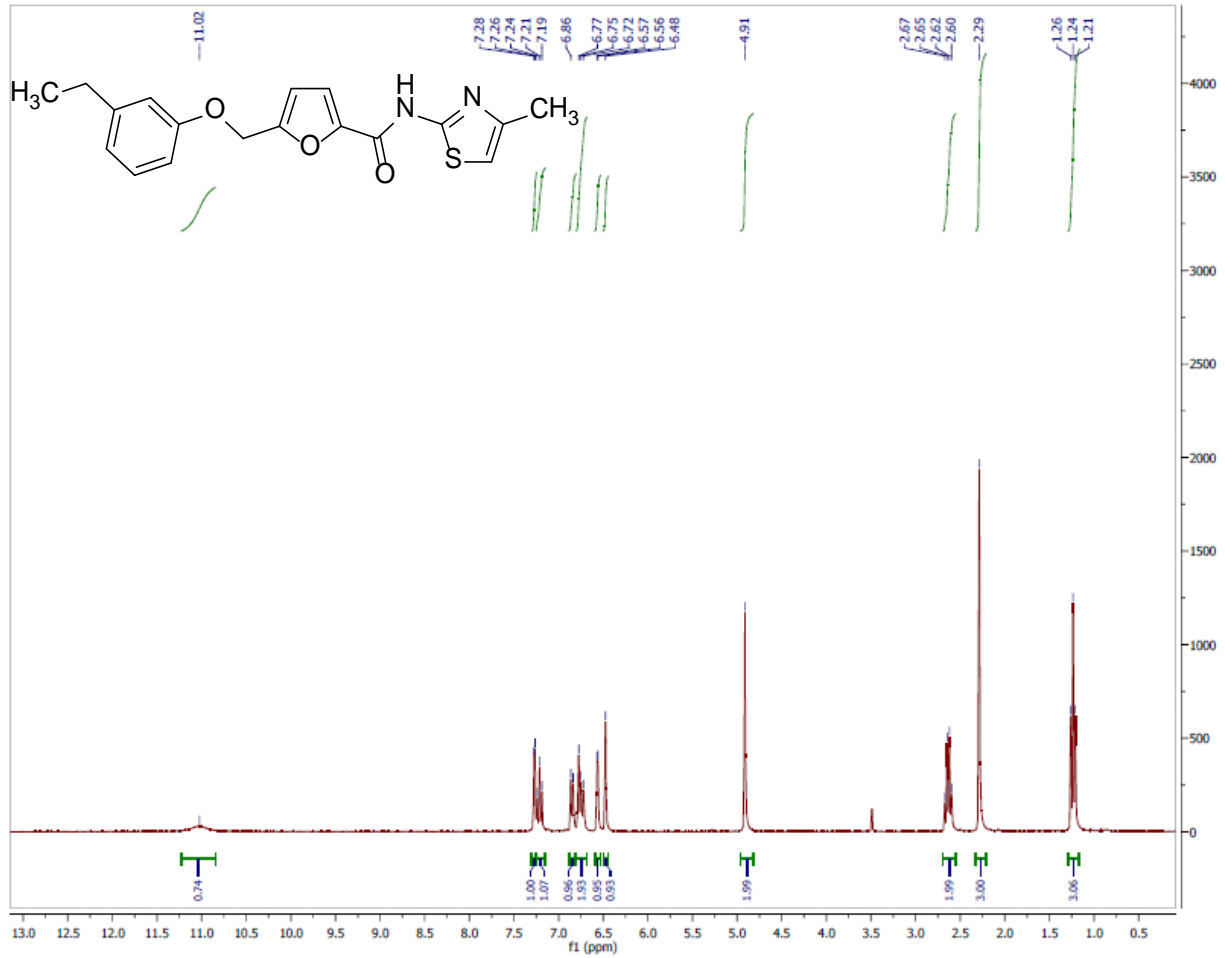
UPLC-MS Chromatogram of Analog CID 51003708
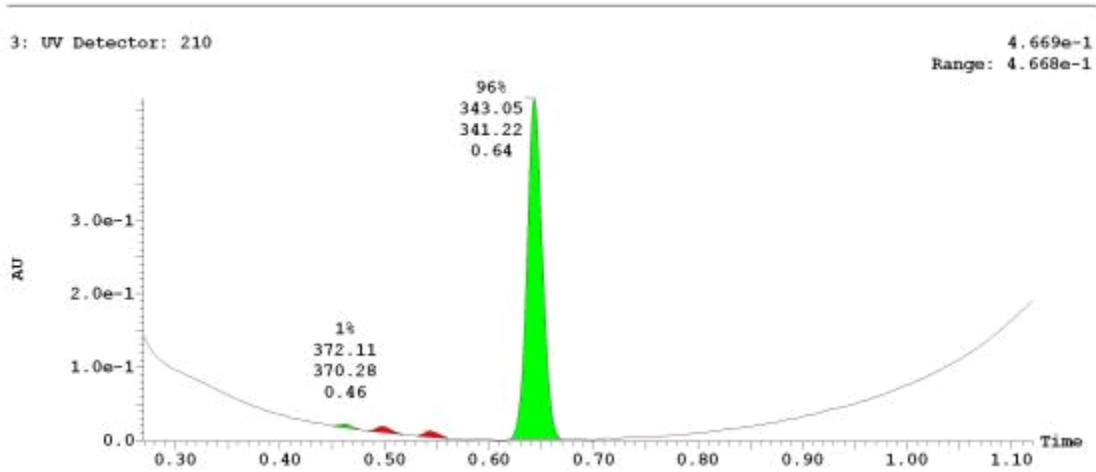
UPLC-MS Chromatogram of Analog CID 50904135

UPLC-MS Chromatogram of Analog CID 50910522

3: UV Detector: 210

6.234e-1
Range: 6.233e-1

97%
393.00
391.24
0.47

2%
115.12
420.38
0.36
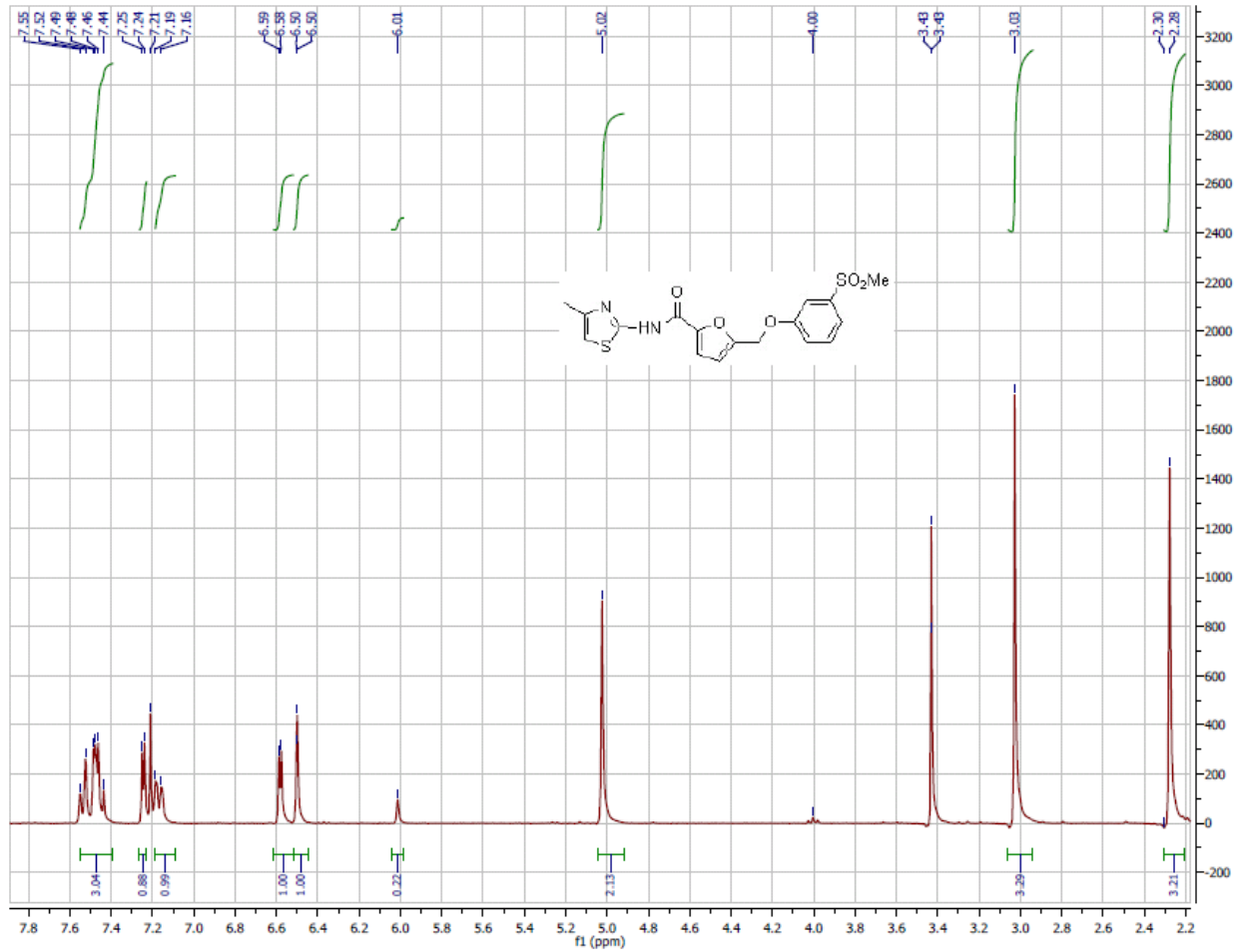
UPLC-MS Chromatogram of Analog CID 50904149

¹H NMR Spectrum (300 MHz, CDCl₃) of Analog CID 35763151
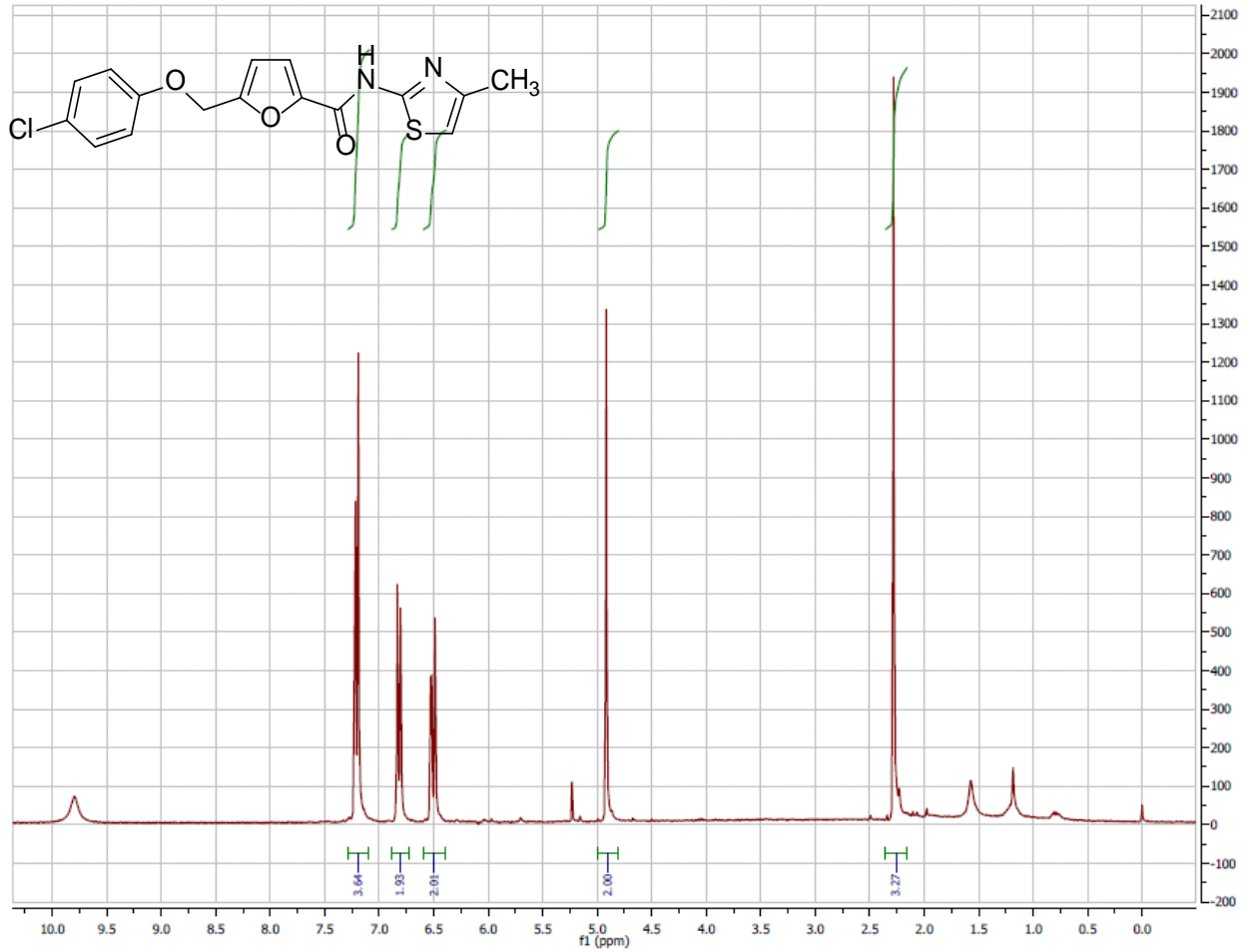
UPLC-MS Chromatogram of Analog CID 35763151
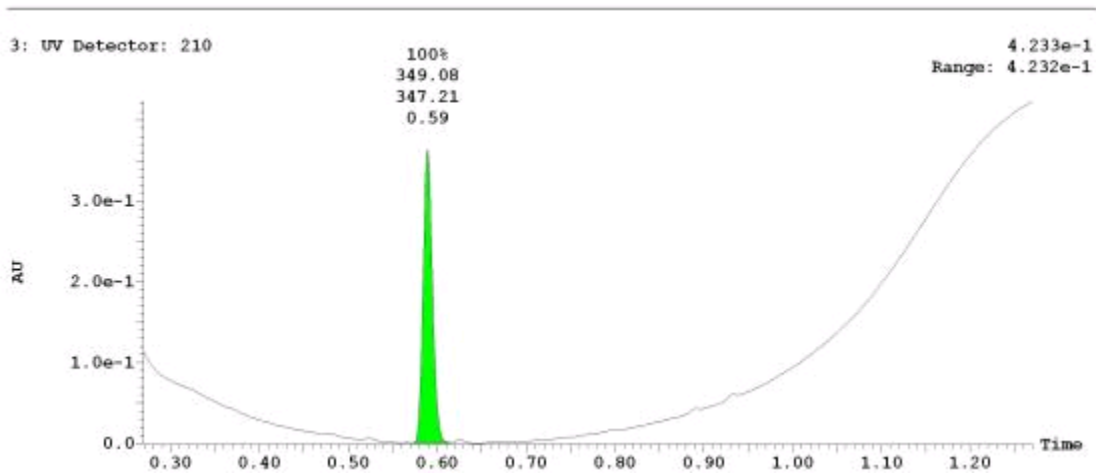
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 26362833
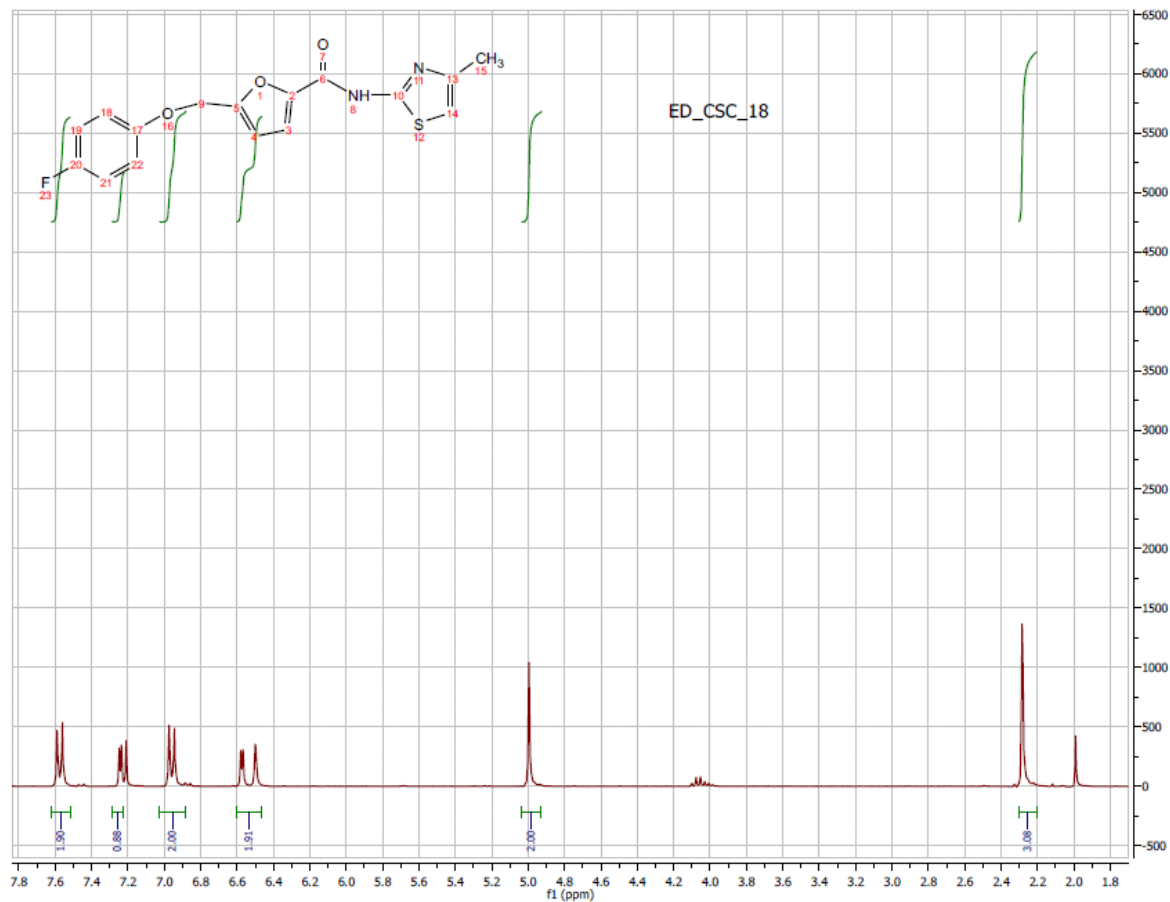
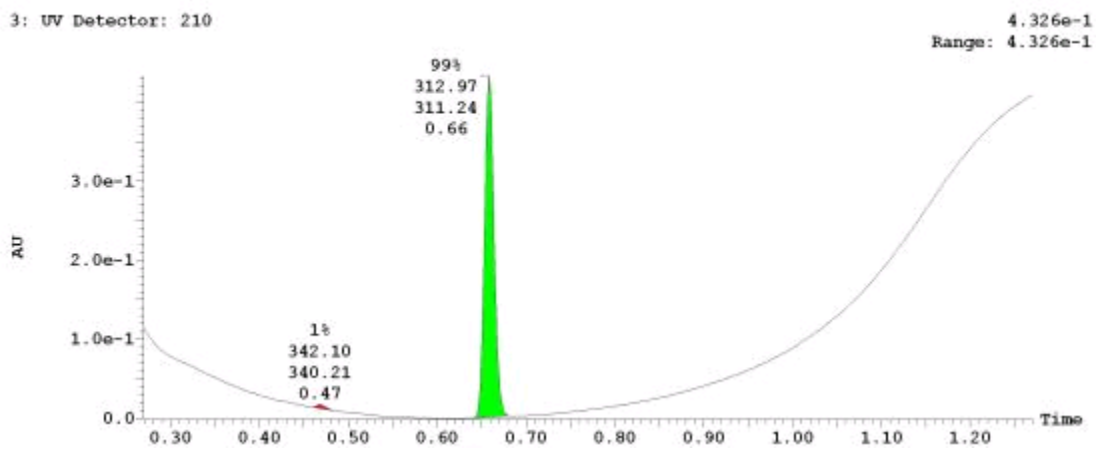# UPLC-MS Chromatogram of Analog CID 26362833
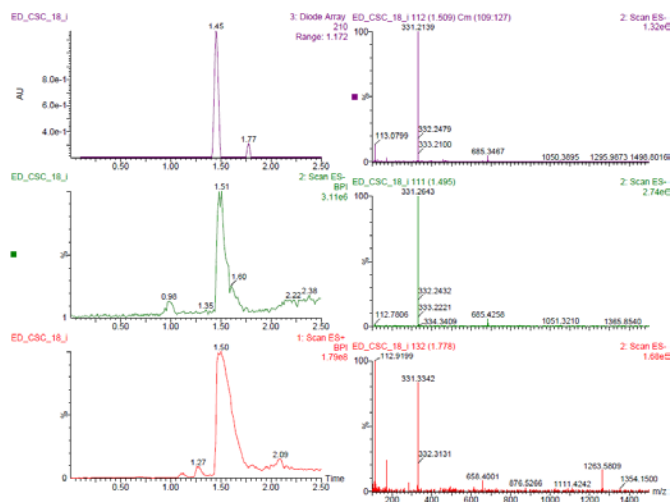
UPLC-MS Chromatogram of Analog CID 50904147
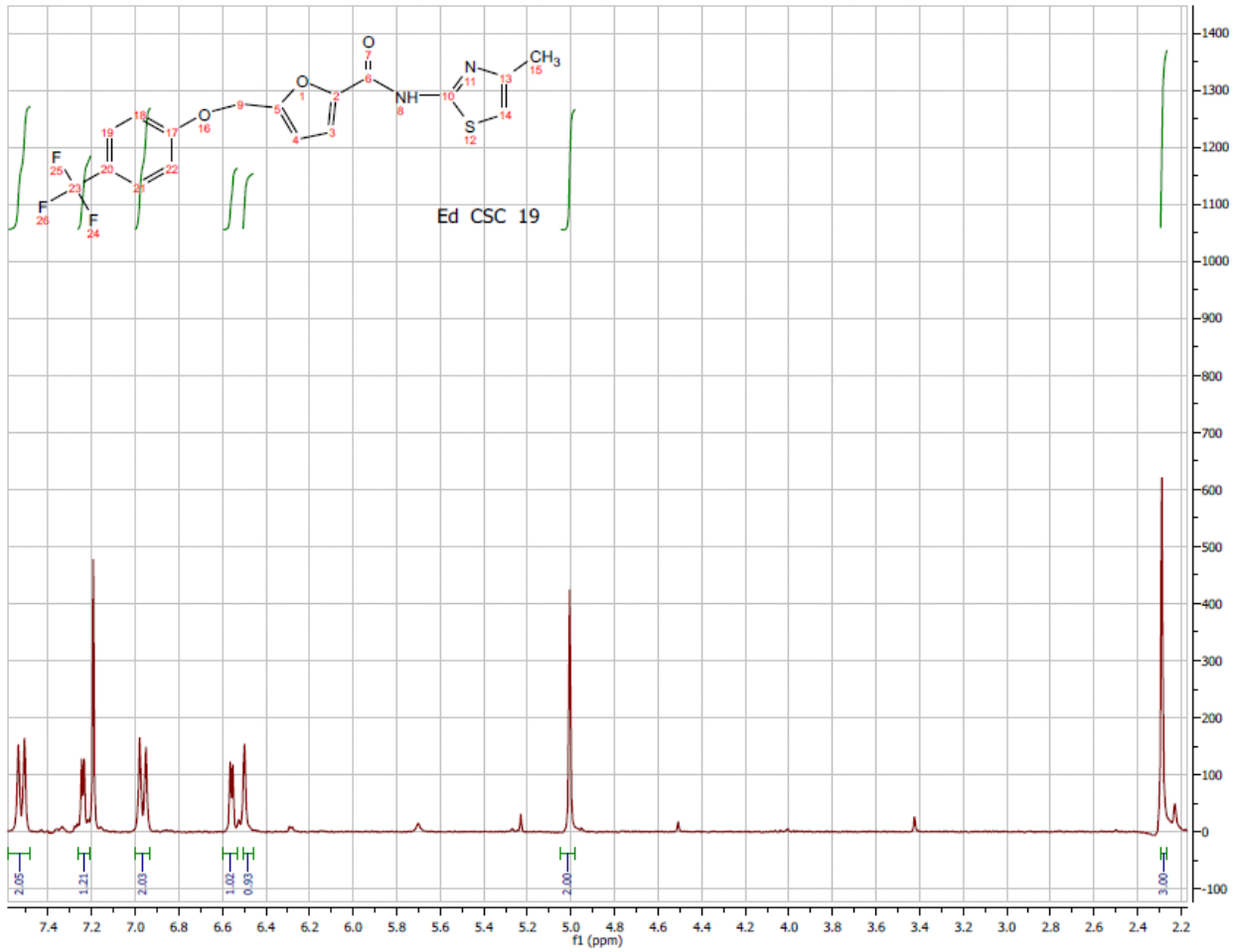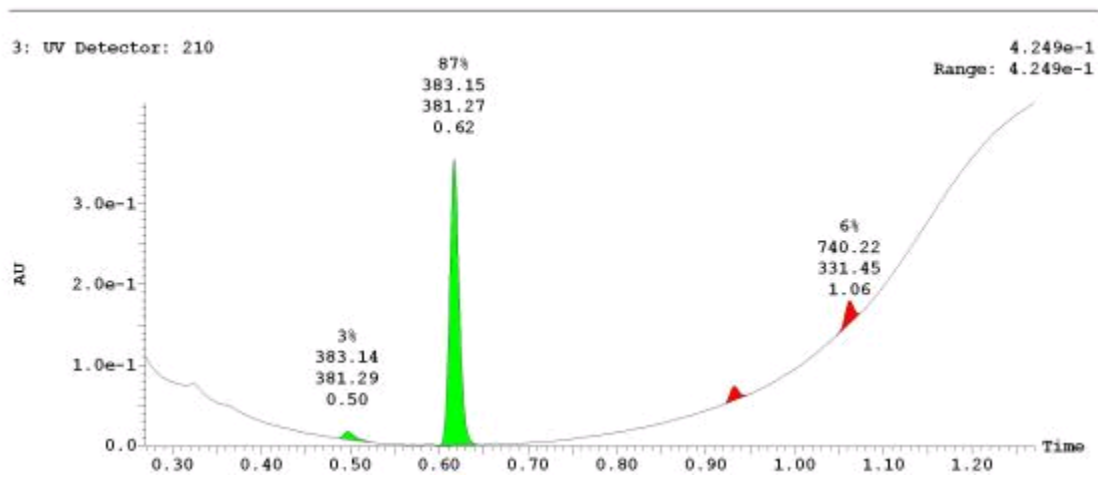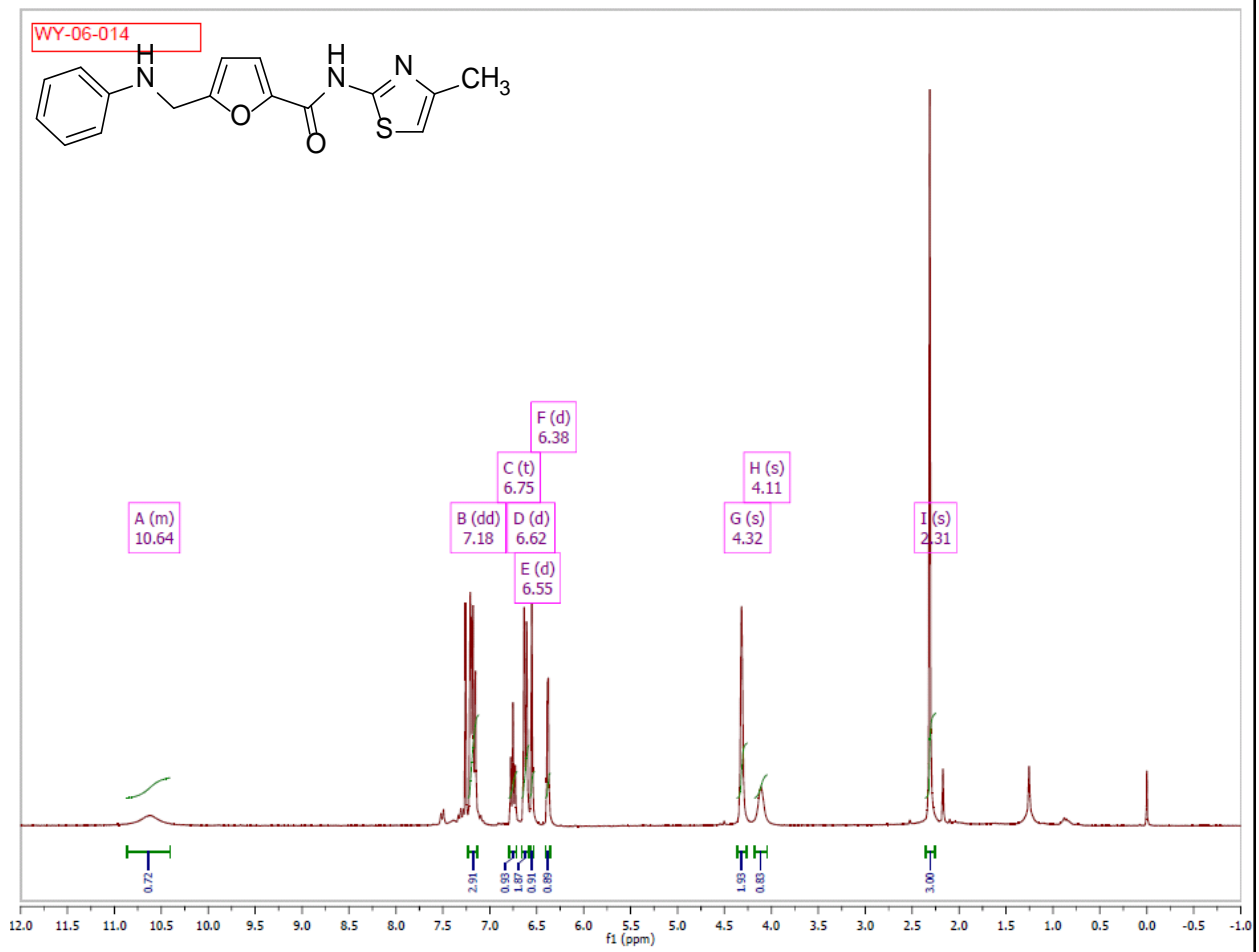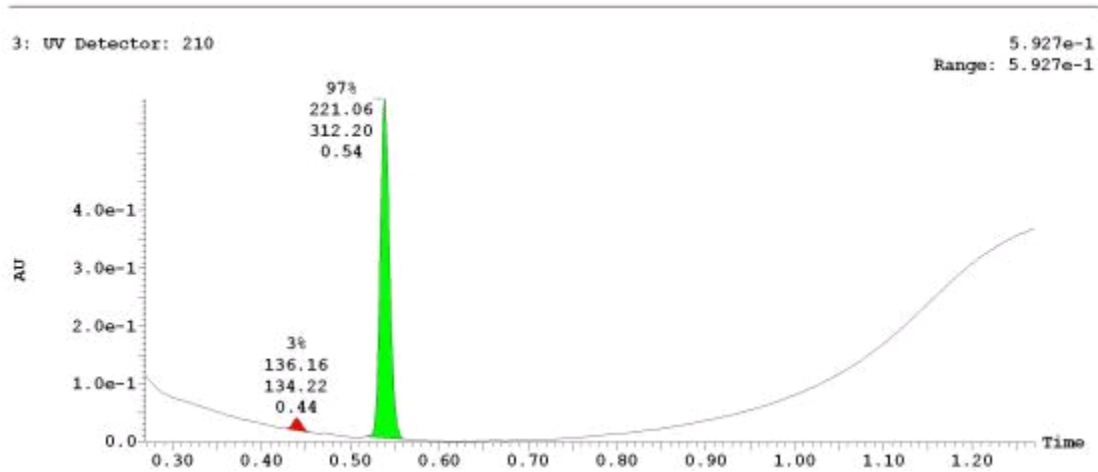
¹H NMR Spectrum (300 MHz, CDCl₃) of Analog CID 50904142
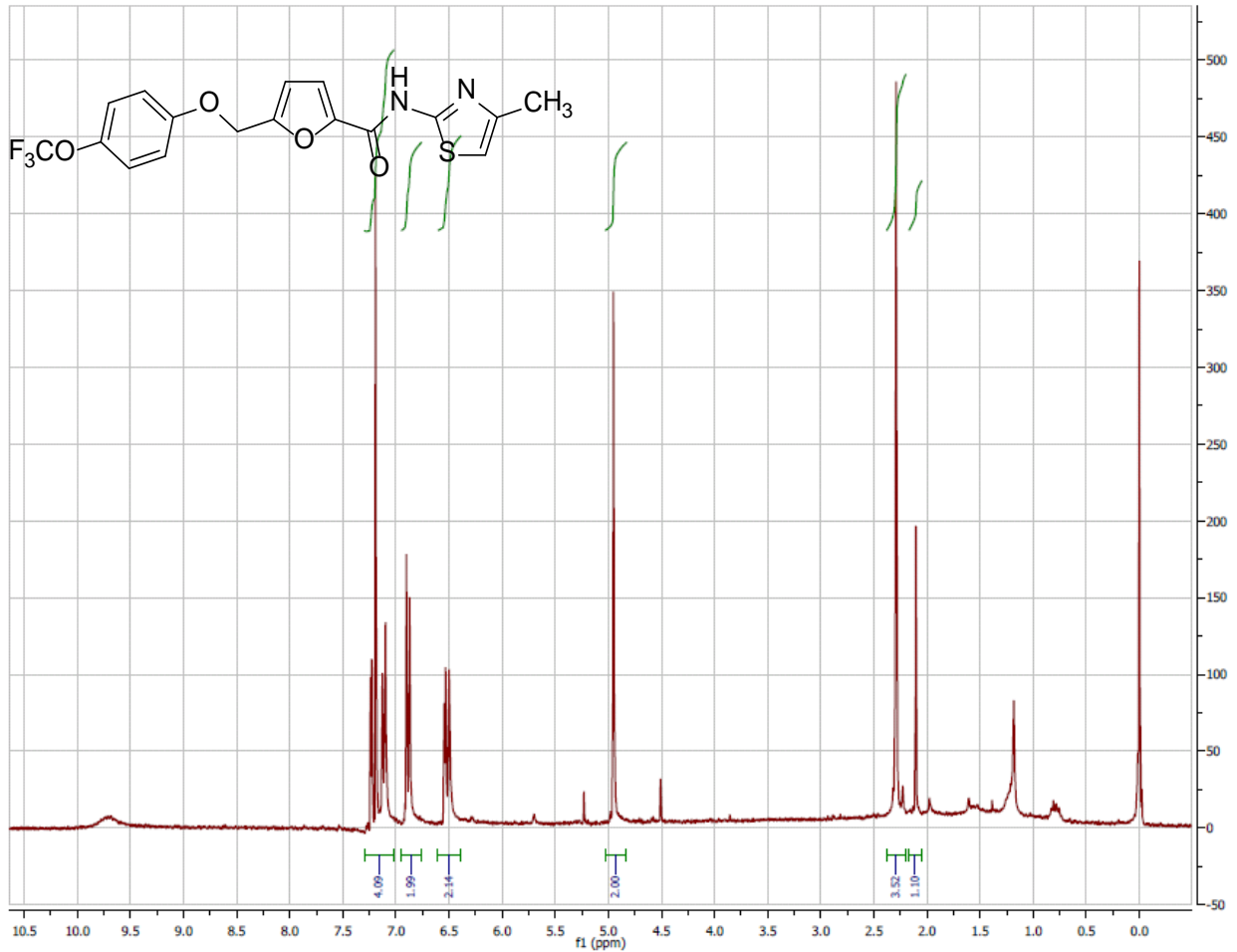
UPLC-MS Chromatogram of Analog CID 50904142
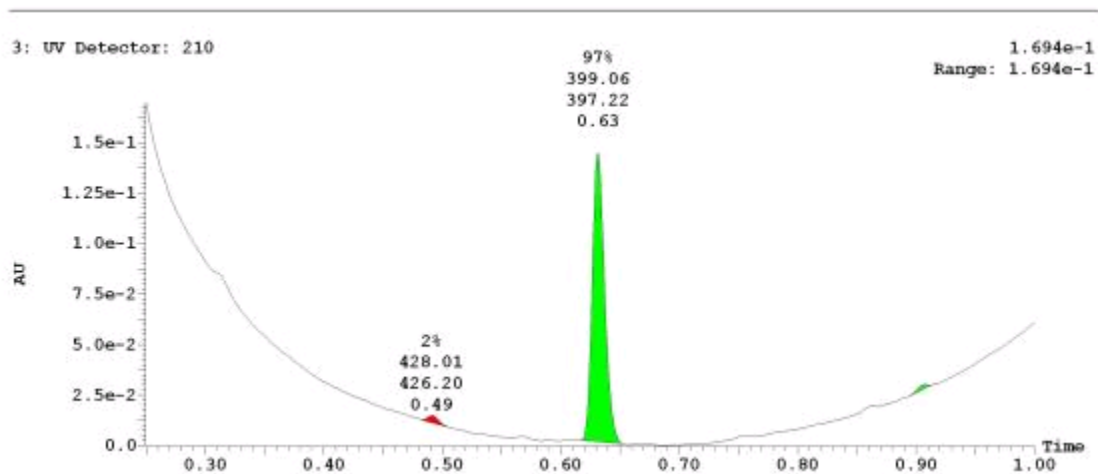
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 38046061
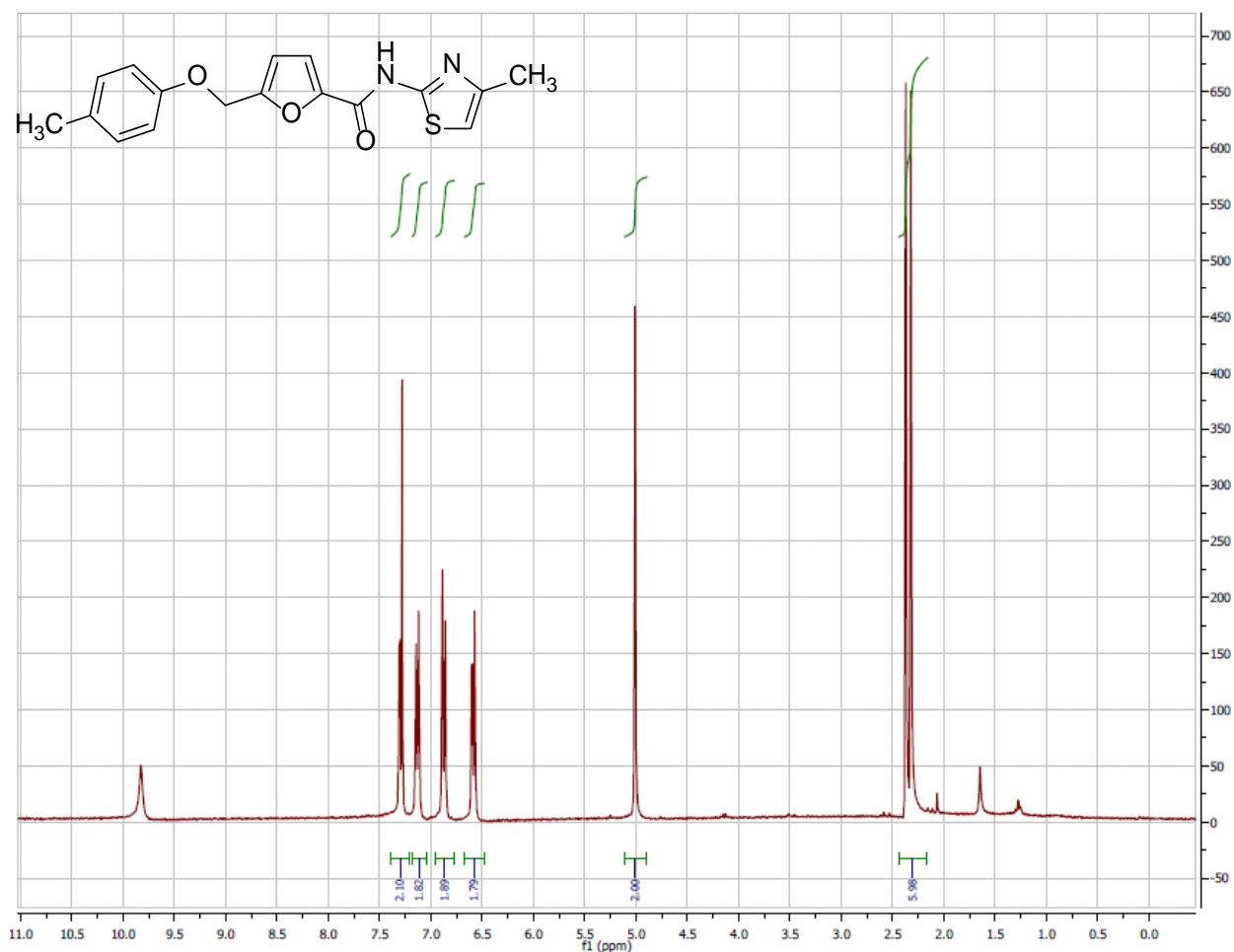
UPLC-MS Chromatogram of Analog CID 38046061
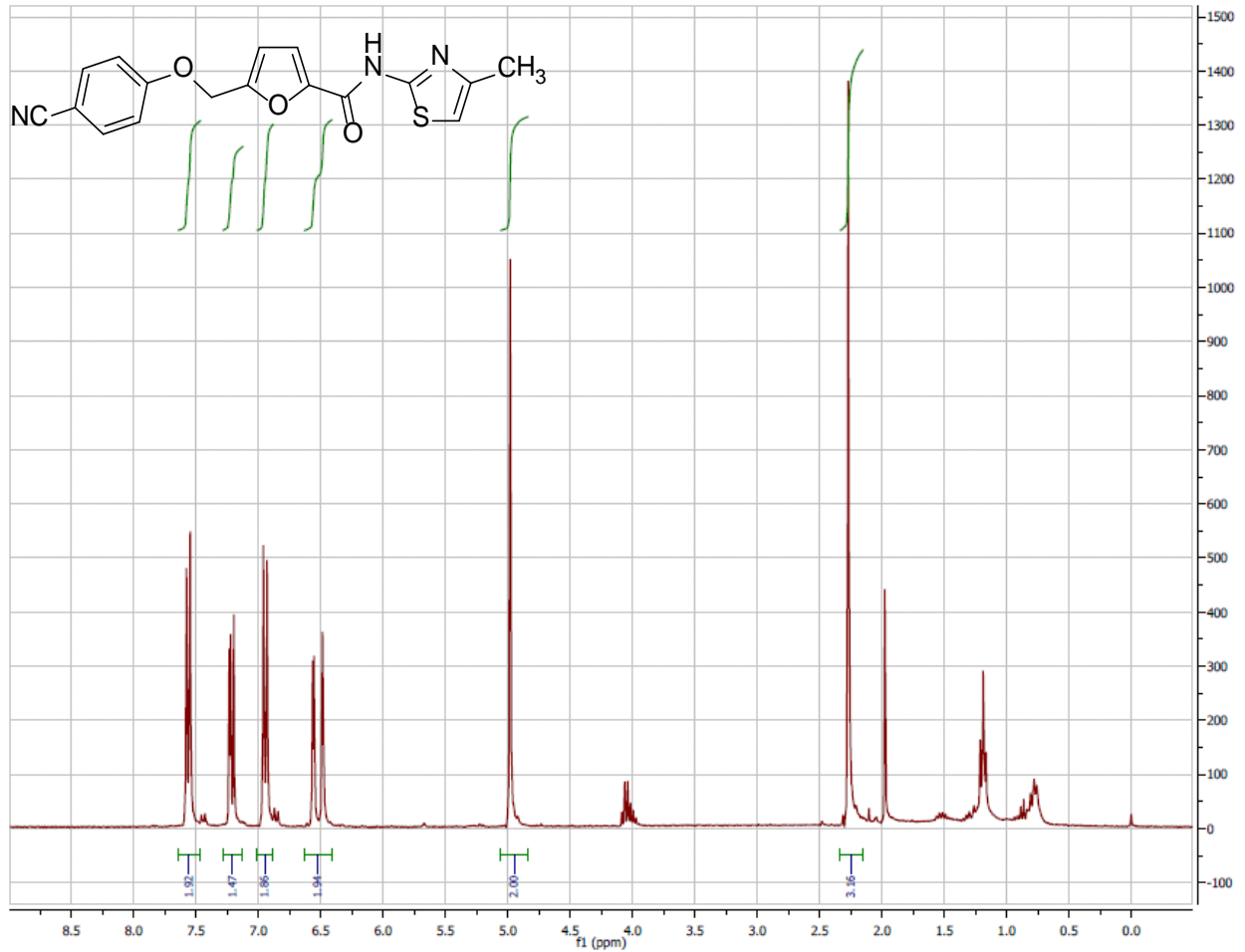
3: UV Detector: 210

6.736e-1
Range: 6.736e-1

100%
221.97
327.23
0.59

¹H NMR Spectrum (300 MHz, CDCl₃) of Analog CID 50904133

UPLC-MS Chromatogram of Analog CID 50904133

3: UV Detector: 210

1.362
Range: 1.362

97%
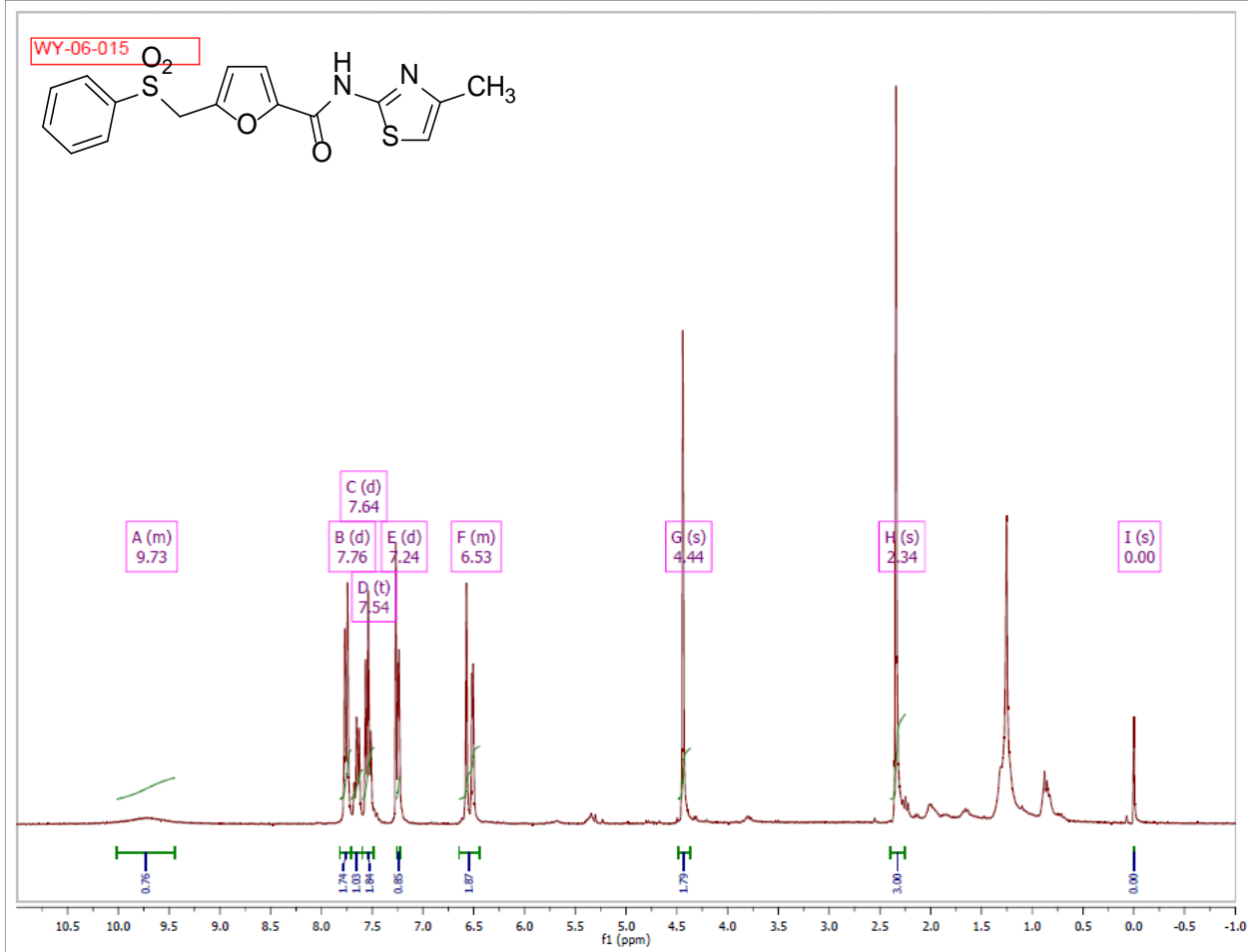340.05
338.18
0.51

0%
368.98
367.14
0.38
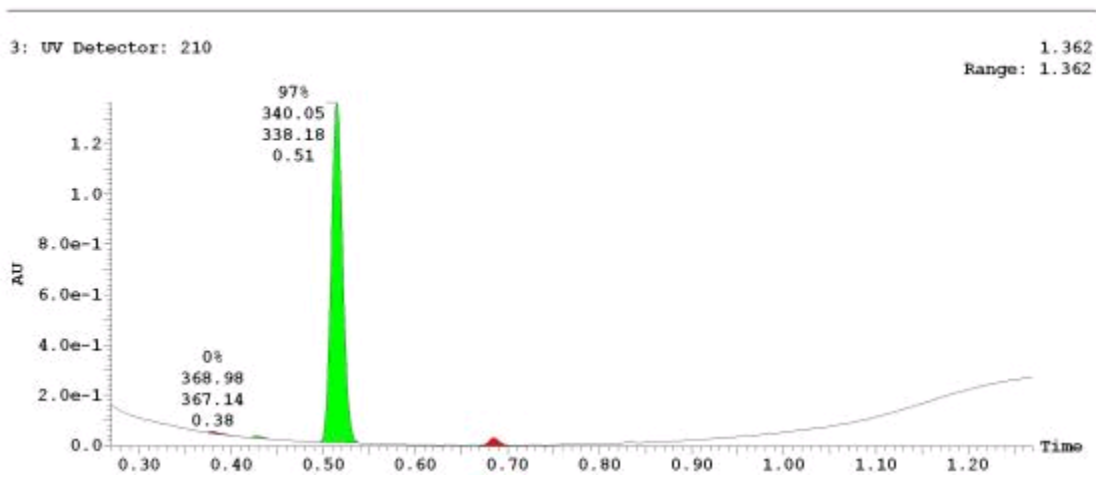
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 50910525
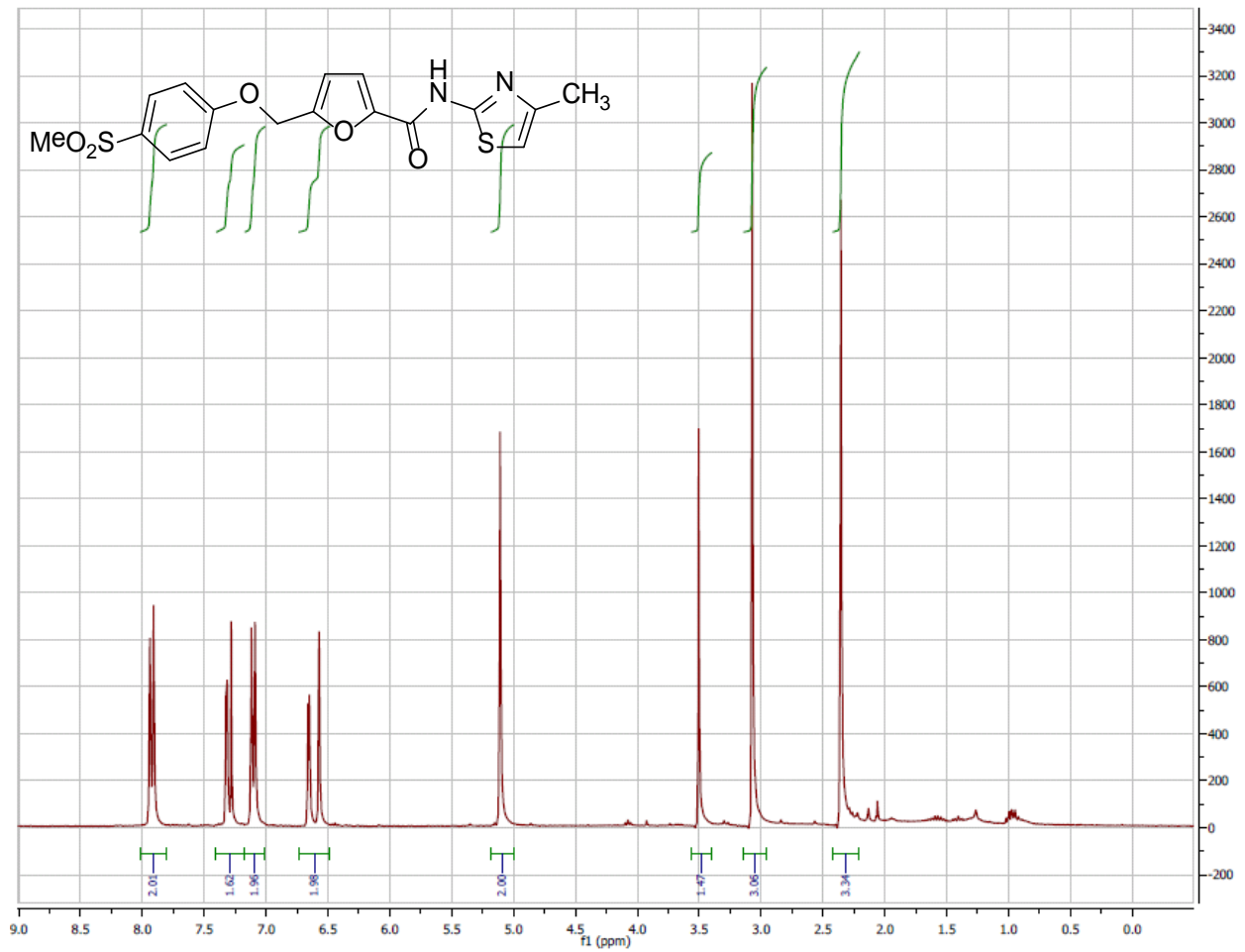
UPLC-MS Chromatogram of Analog CID 50910525
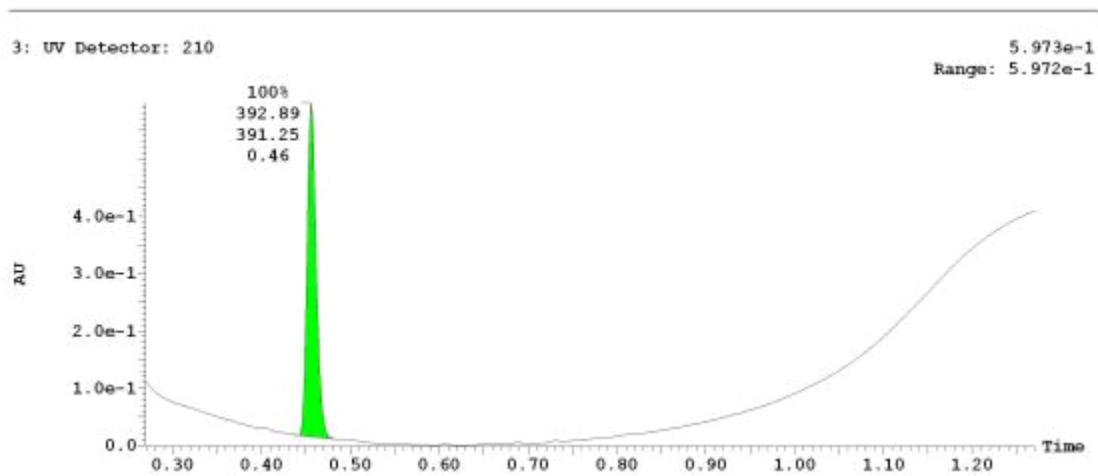
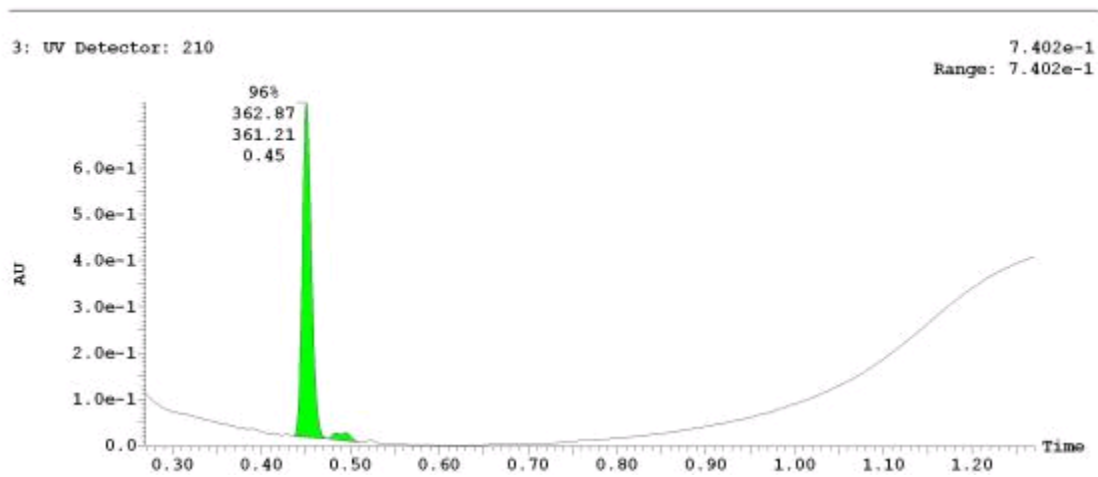**¹H NMR Spectrum (300 MHz, CDCl₃) of Analog CID 35763156**
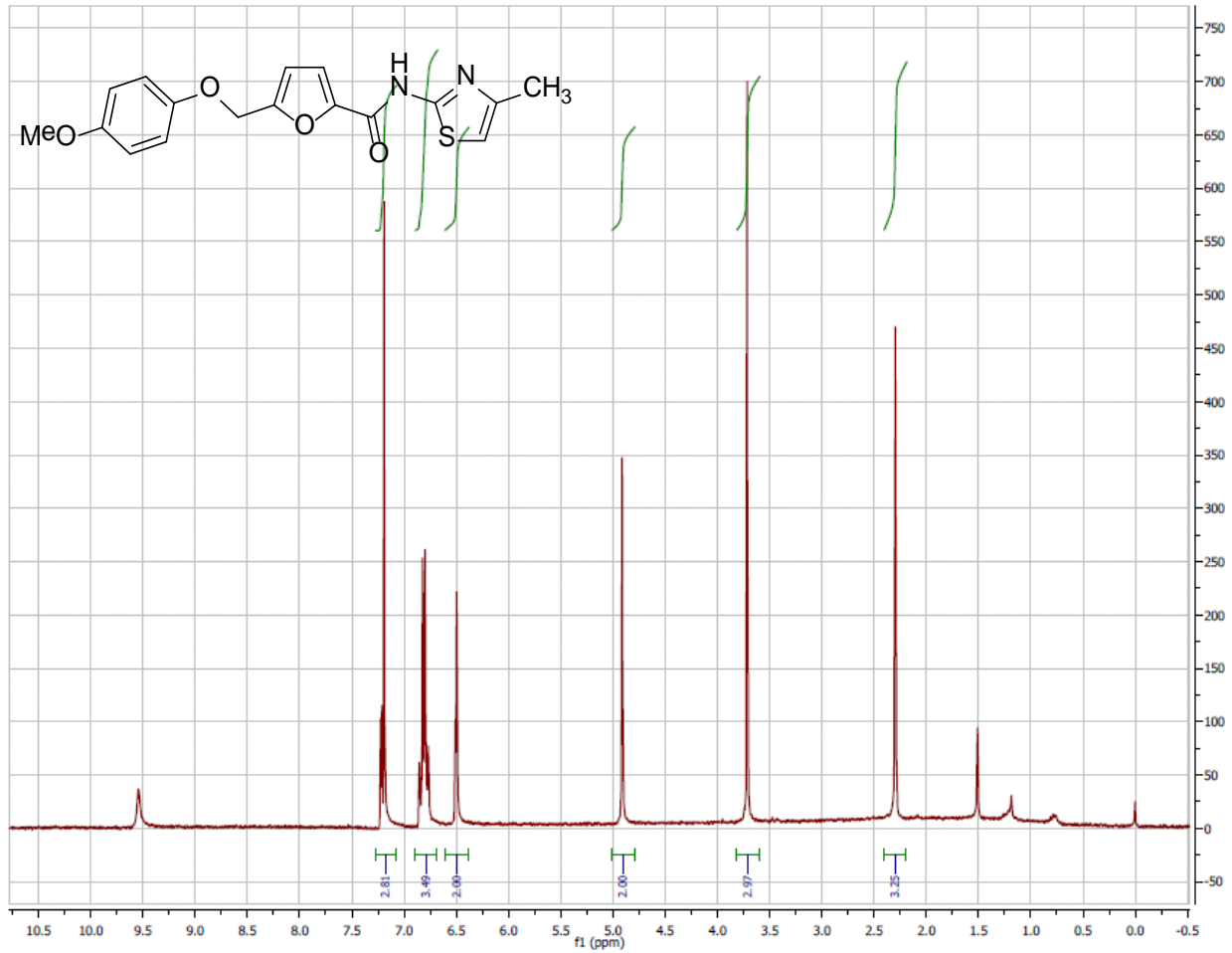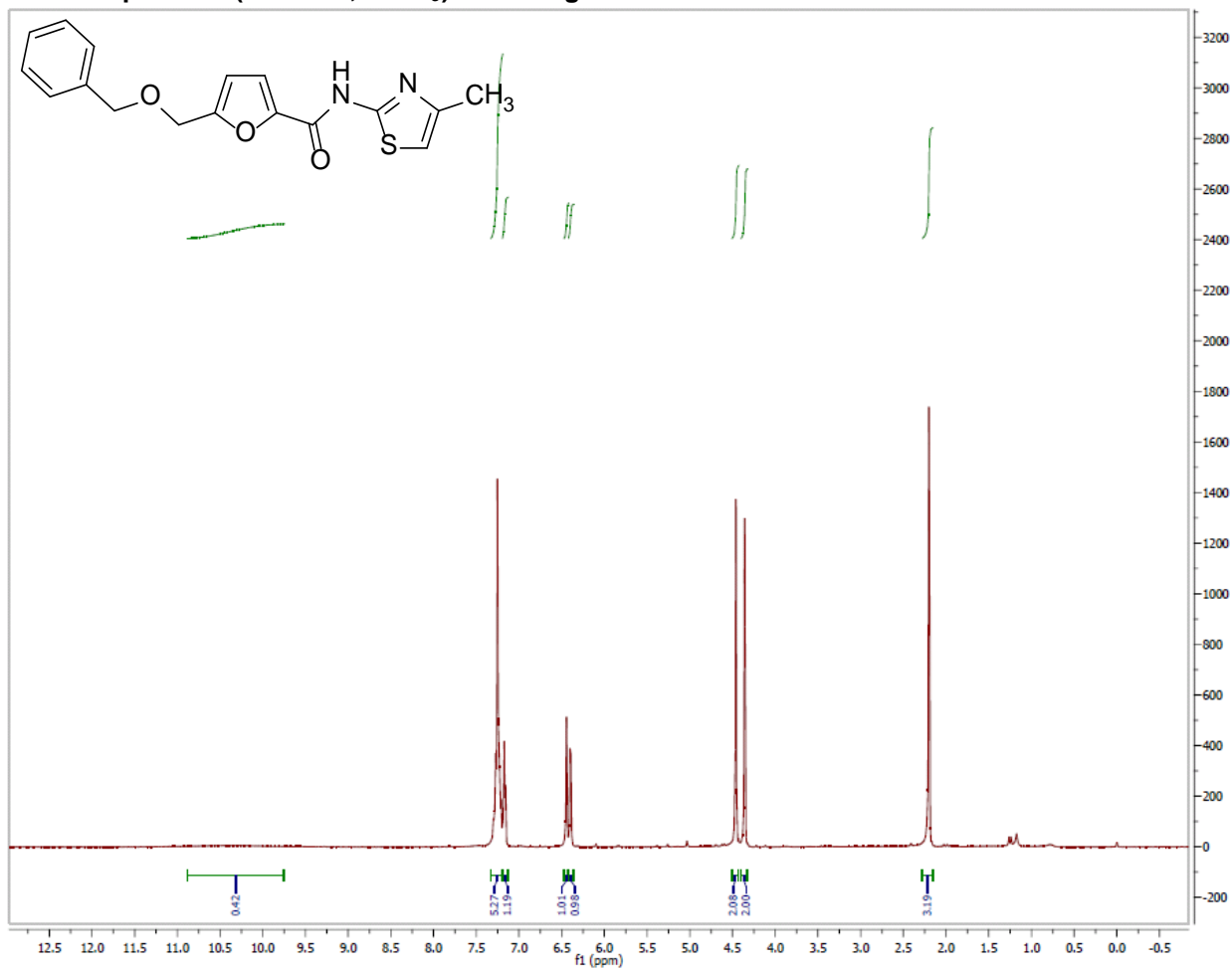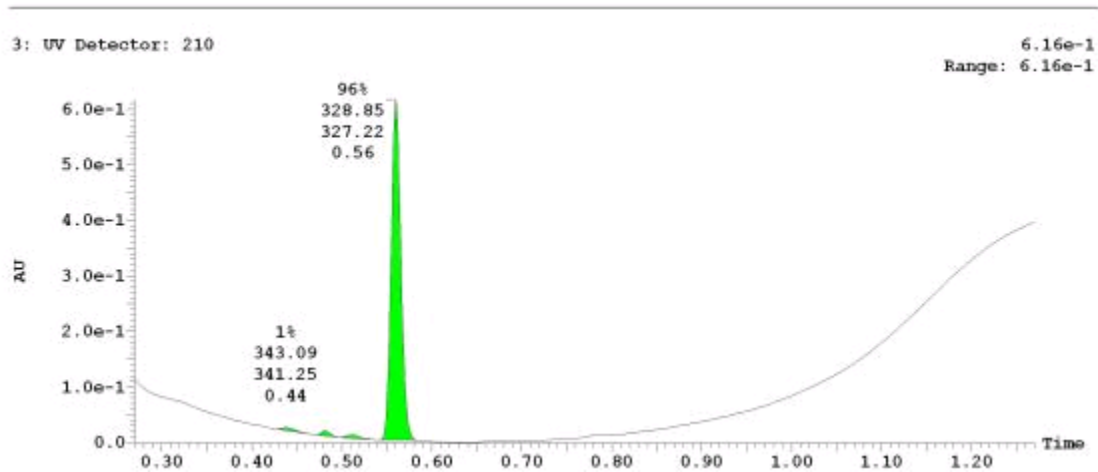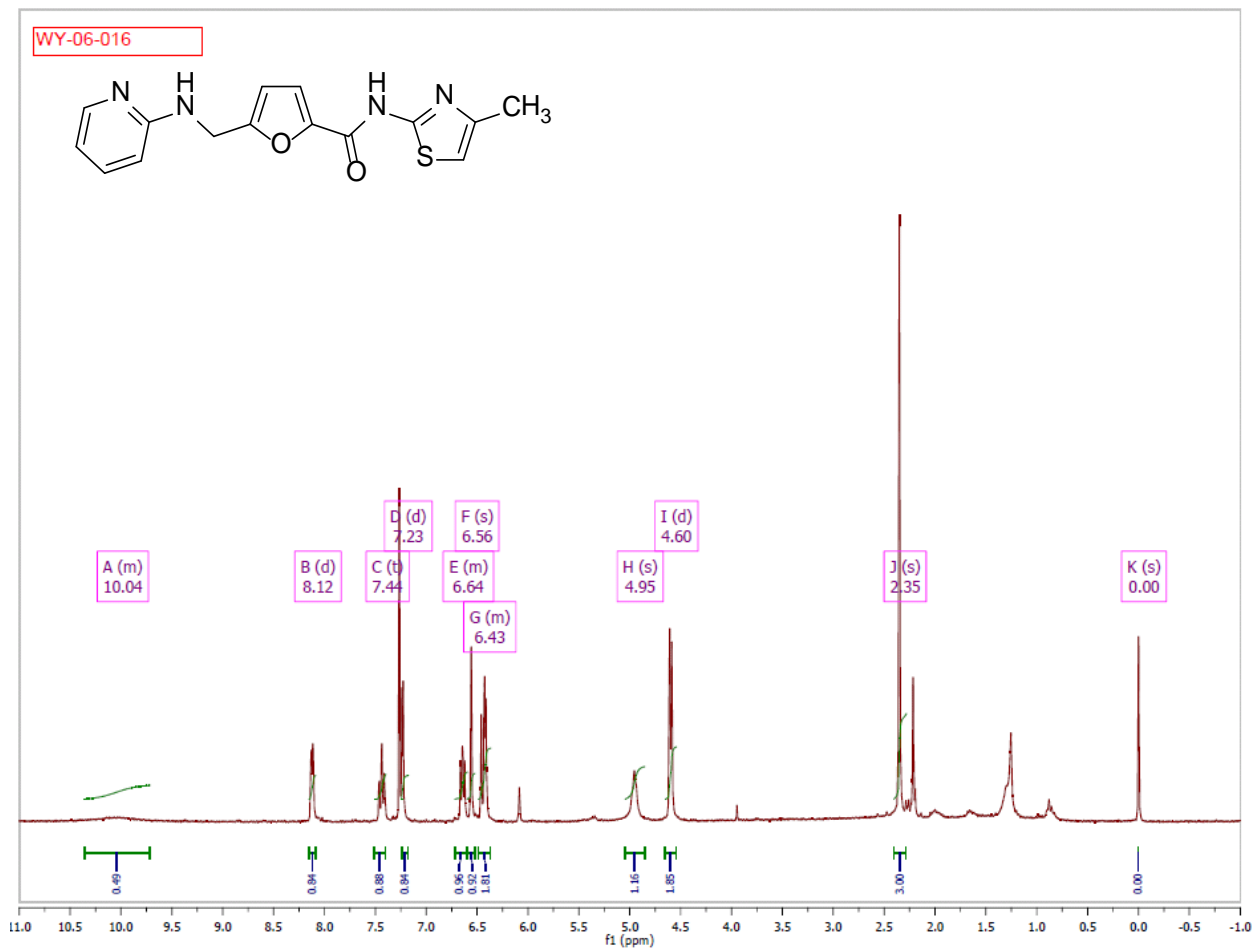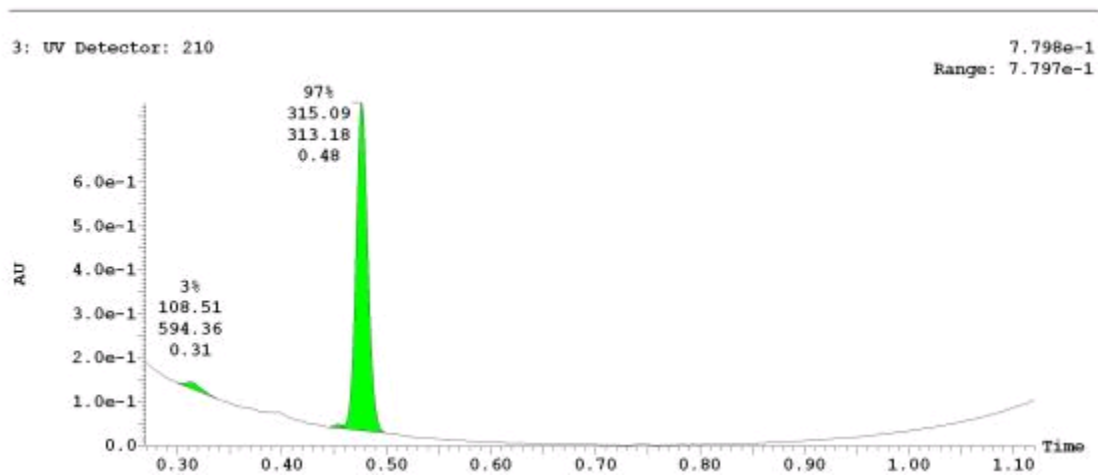

UPLC-MS Chromatogram of Analog CID 35763156

3: UV Detector: 210

4.354e-1
Range: 4.353e-1

100%
345.14
343.23
0.54
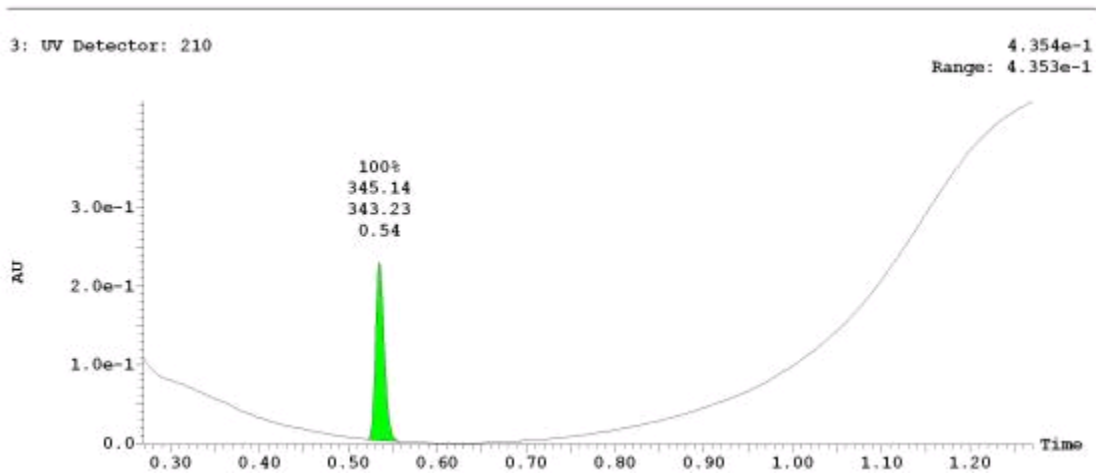
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 51003713

	Parameter	Value
1	Data File Name	//thumper25/cbplat_analytical/nmr/bruker/data/willmeny/nmr/WY_06_031/2/fid
2	Title	After HCl wash PROTON CDCl3 Z:\\willmeny 25
3	Solvent	CDCl3
4	Acquisition Date	2011-03-18T22:57:52
5	Temperature	300.0
6	Number of Scans	16
7	Spectrometer Frequency	300.13
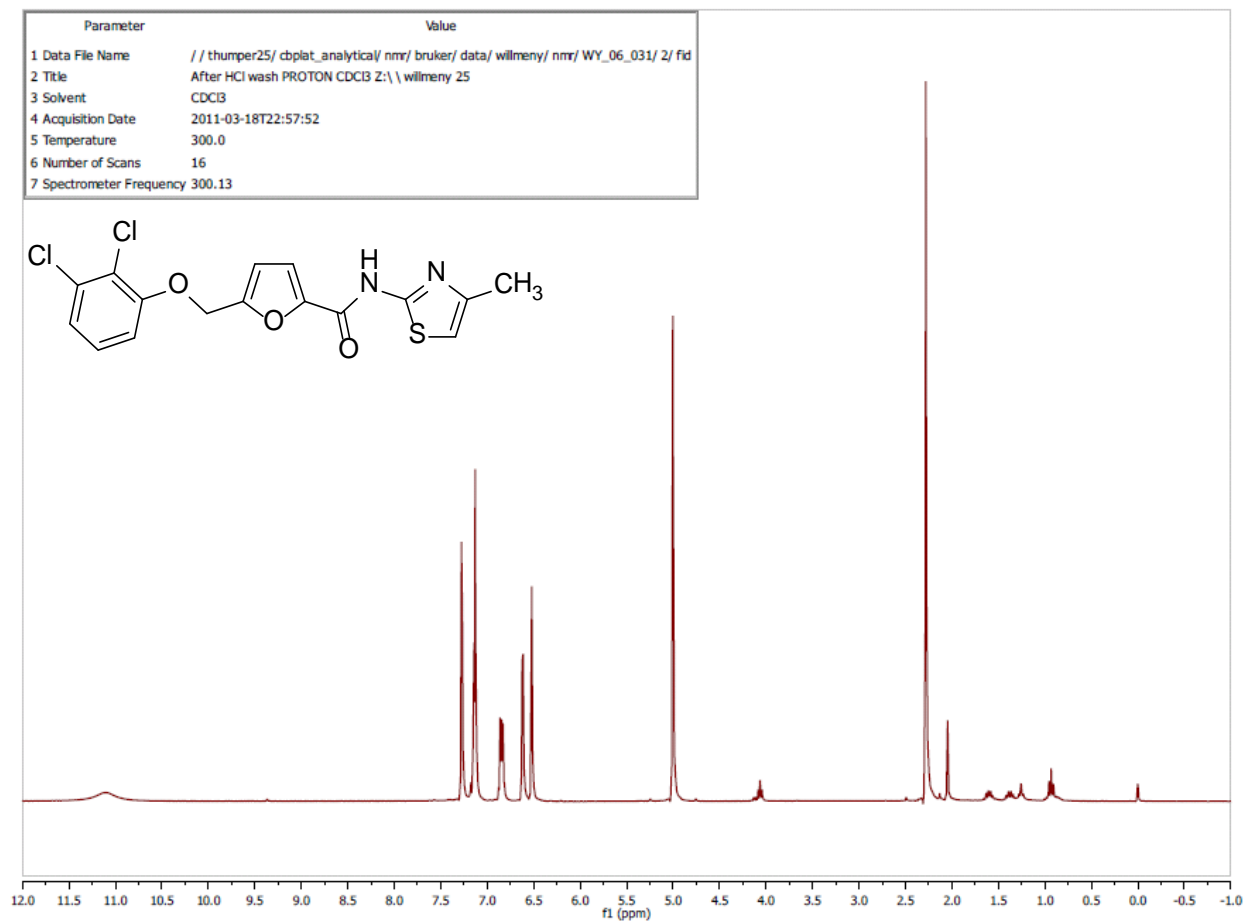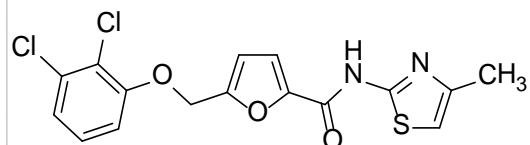
UPLC-MS Chromatogram of Analog CID 51003713
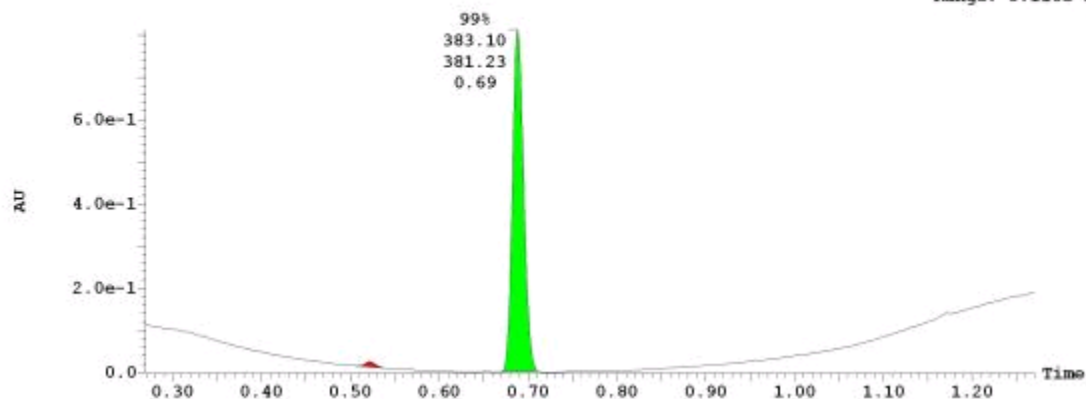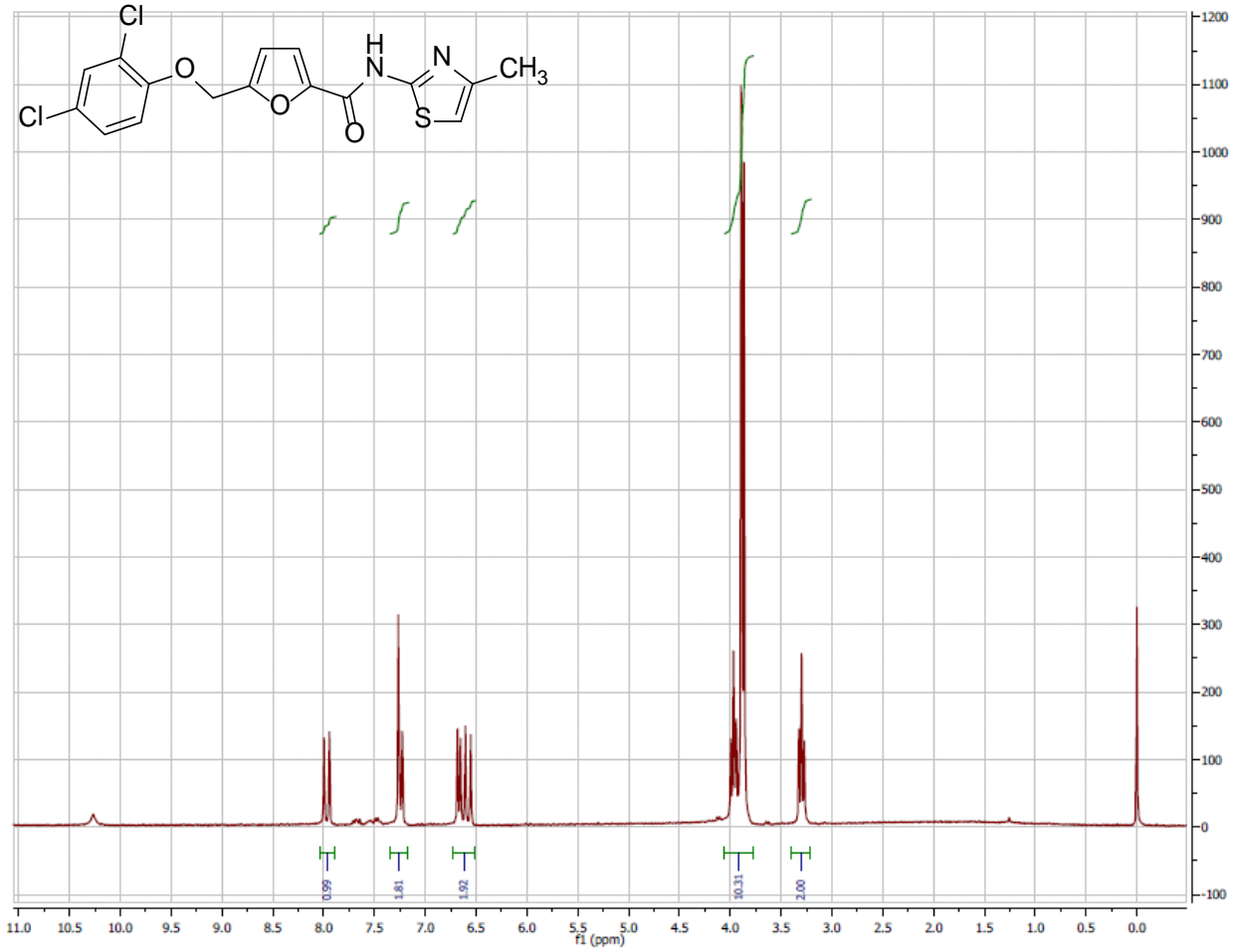
UPLC-MS Chromatogram of Analog CID 1977298
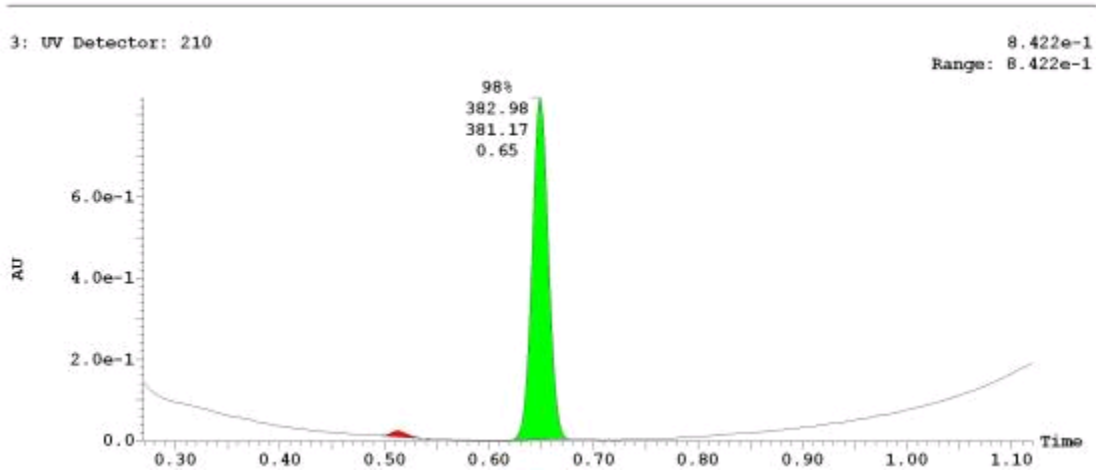
UPLC-MS Chromatogram of Analog CID 50944059
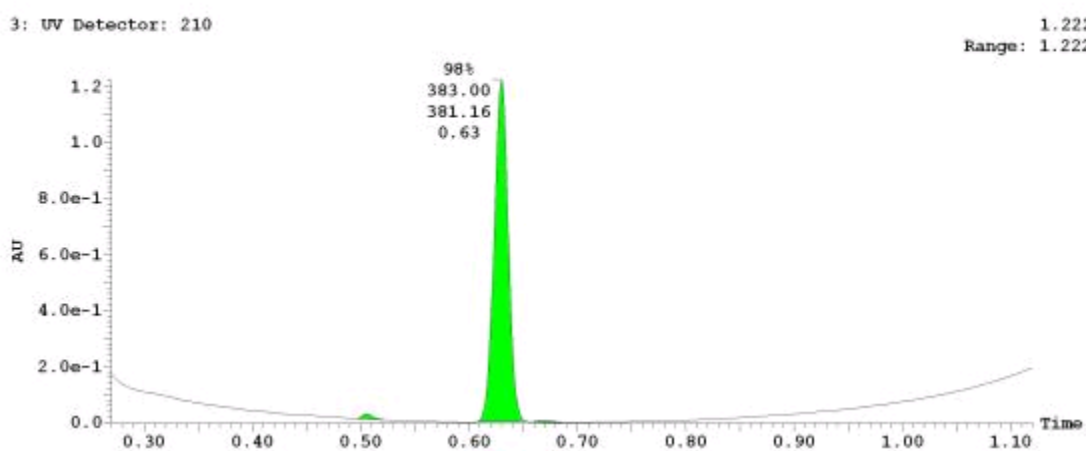
¹H NMR Spectrum (300 MHz, CDCl₃) of Analog CID 51003712

	Parameter	Value
1	Data File Name	//thumper25/cbplat_analytical/nmr/bruker/data/willmeny/nmr/WY_06_030/1/fid
2	Title	PROTON CDCl3 Z:\\willmeny 35
3	Solvent	CDCl3
4	Acquisition Date	2011-03-15T20:59:58
5	Temperature	300.0
6	Number of Scans	16
7	Spectrometer Frequency	300.13

UPLC-MS Chromatogram of Analog CID 51003712

3: UV Detector: 210

8.787e-1
Range: 8.787e-1

99%
383.02
381.18
0.63

^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 51003702

UPLC-MS Chromatogram of Analog CID 51003702

3: UV Detector: 210

8.938e-1
Range: 8.937e-1

98%
382.99
381.16
0.65

Parameter	Value
1 Data File Name	//thumper25/cbplat/analytical/nmr/bruker/data/cfernad/nmr/CVF_CSC_34/1/fid
2 Title	CVF_CSC_34
3 Comment	PROTON MeOD Z:\\cfernad 34
4 Origin	Bruker BioSpin GmbH
5 Owner	nmr
6 Site	
7 Spectrometer	spect
8 Author	
9 Solvent	MeOD
10 Temperature	300.0
11 Pulse Sequence	zg30
12 Experiment	1D
13 Number of Scans	16
14 Receiver Gain	1448
15 Relaxation Delay	1.0000
16 Pulse Width	11.5000
17 Acquisition Time	5.3085
18 Acquisition Date	2011-03-17T09:34:00
19 Modification Date	2011-03-17T10:48:47
20 Spectrometer Frequency	300.13
21 Spectral Width	6172.8
22 Lowest Frequency	-770.5
23 Nucleus	1H
24 Acquired Size	32768
25 Spectral Size	65536

UPLC-MS Chromatogram of Analog CID 51003715

3: UV Detector: 210

9.798e-1
Range: 9.798e-1

98%
383.10
381.24
0.73

¹H NMR Spectrum (300 MHz, CDCl₃) of Analog CID 50910552

^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 50910552

WY-06-013

Label	Shift
A (m)	10.31
B (d)	8.17
C (dd)	7.60
D (m)	7.28
F (d)	6.78
E (m)	6.93
G (d)	6.60
H (s)	6.53
I (s)	5.33
J (s)	2.33
K (s)	0.00

UPLC-MS Chromatogram of Analog CID 50910552

3: UV Detector: 210
7.215e-1
Range: 7.214e-1

100%
221.04
314.19
0.49

^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 51351633

Parameter	Value
1 Data File Name	//thumper25/cbplat_analytical/nmr/bruker/data/willmeny/nmr/WY_06_060/1/fid
2 Title	After 1st column
3 Solvent	CDCl3
4 Acquisition Date	2011-04-21T18:17:25
5 Temperature	300.0
6 Number of Scans	16
7 Spectrometer Frequency	300.13

UPLC-MS Chromatogram of Analog CID 51351633

^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 51351622

Parameter	Value
1 Data File Name	//thumper25/cbplat_analytical/nmr/bruker/data/willmeny/nmr/WY_06_072/2/fid
2 Title	Column 1 - Fractions 33-38
3 Solvent	CDCl3
4 Acquisition Date	2011-04-29T18:41:25
5 Temperature	300.0
6 Number of Scans	16
7 Spectrometer Frequency	300.13

UPLC-MS Chromatogram of Analog CID 51351622

3: UV Detector: 210

4.157e-1
Range: 4.157e-1

89%
221.04
314.24
0.34

11%
536.25
534.35
0.48

^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 50910539

UPLC-MS Chromatogram of Analog CID 50910539

3: UV Detector: 210

4.326e-1
Range: 4.326e-1

99%
312.97
311.24
0.66

1%
342.10
340.21
0.47

^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 50910533

UPLC-MS Chromatogram of Analog CID 50910533

^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 41548804

UPLC-MS Chromatogram of Analog CID 41548804

¹H NMR Spectrum (300 MHz, CDCl₃) of Analog CID 50910536

UPLC-MS Chromatogram of Analog CID 50910536

BROAD INSTITUTE

^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 50944065

UPLC-MS Chromatogram of Analog CID 50944065

¹H NMR Spectrum (300 MHz, CDCl₃) of Analog CID 51003686
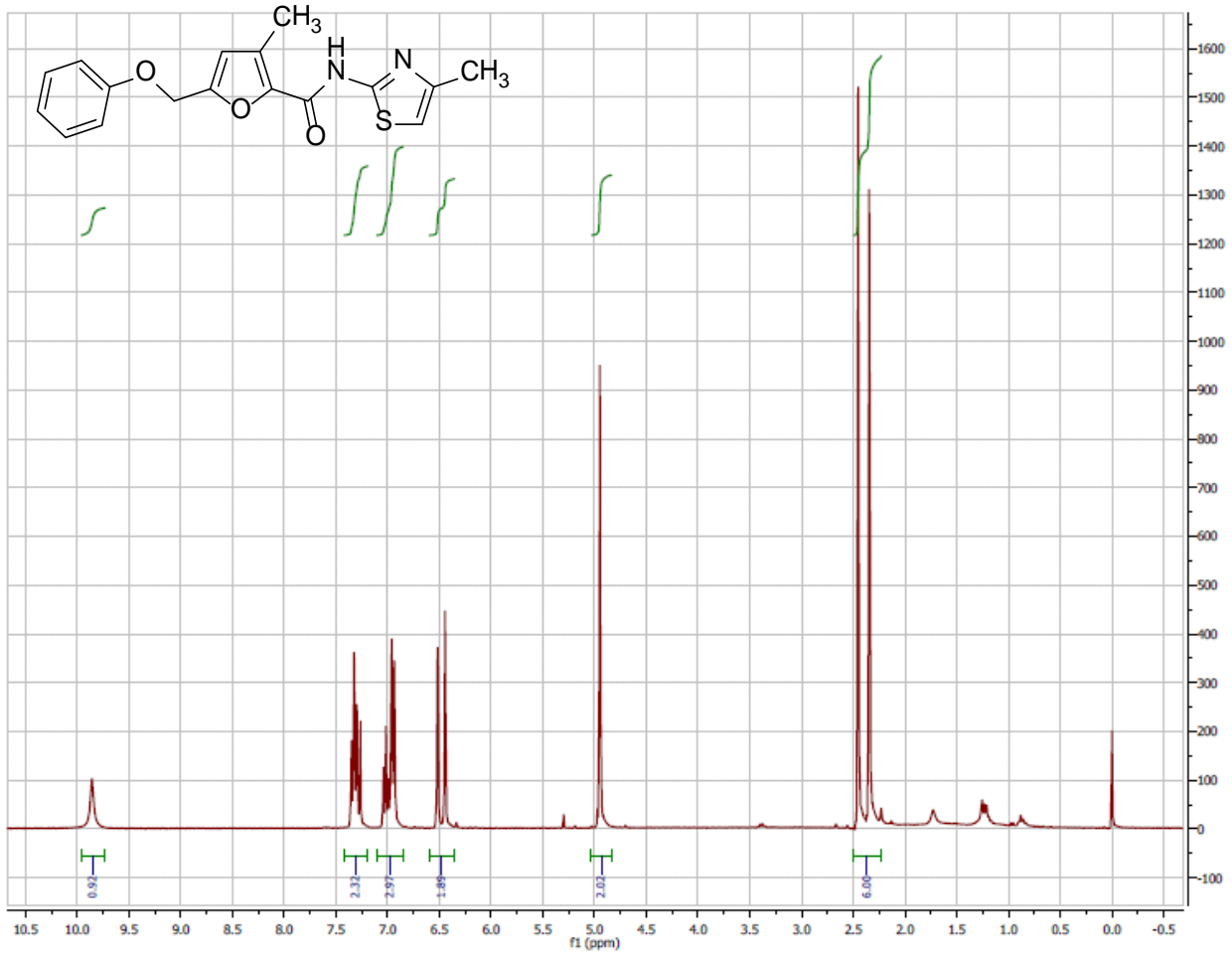
UPLC-MS Chromatogram of Analog CID 51003686
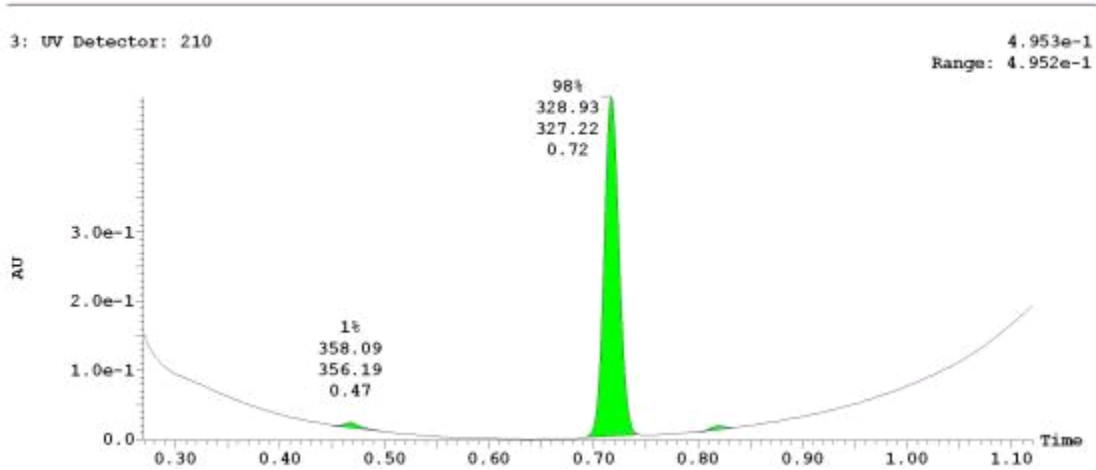
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 51003688
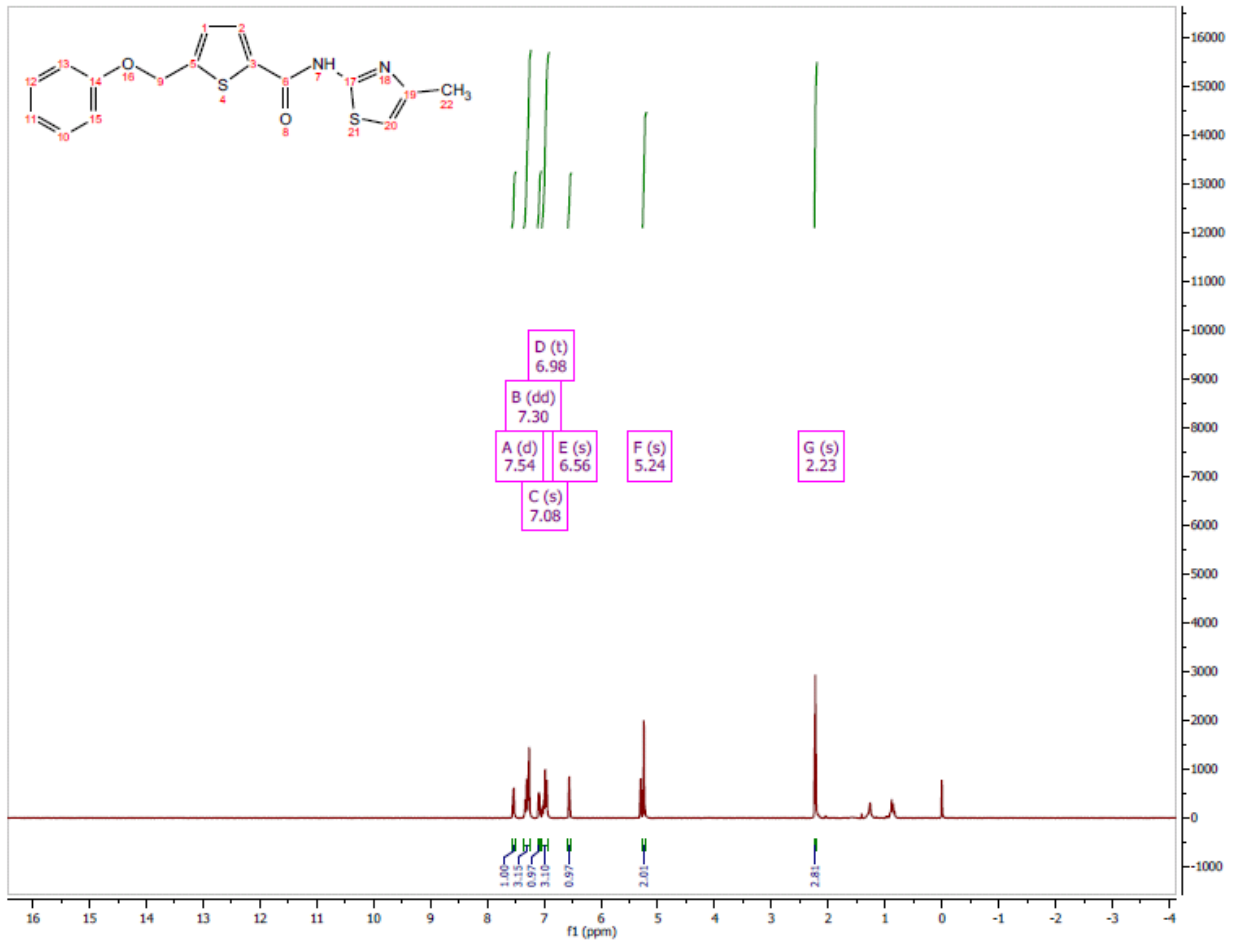
UPLC-MS Chromatogram of Analog CID 51003688

3: UV Detector: 210

4.408e-1
Range: 4.408e-1

98%
331.04
329.19
0.58

2%
217.01
358.20
0.47
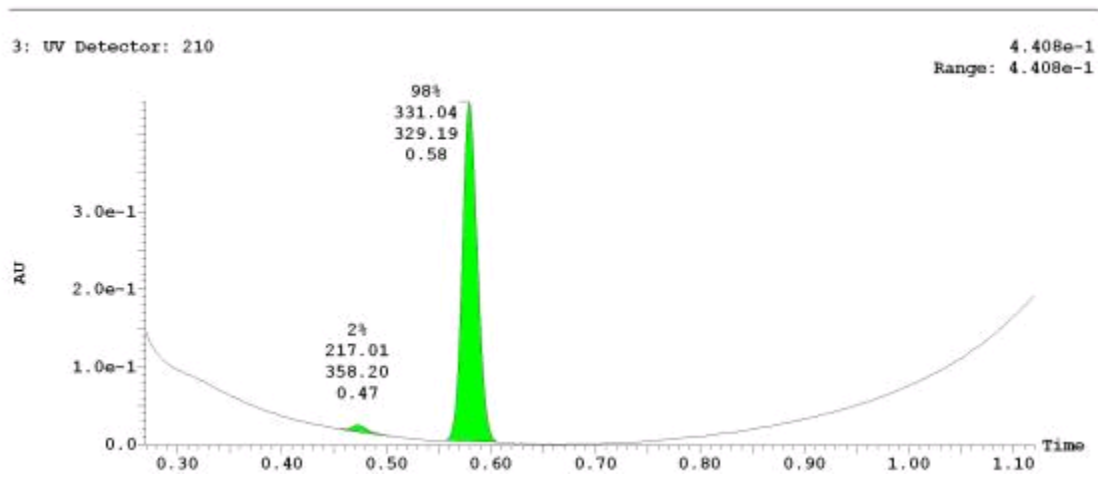
¹H NMR Spectrum (300 MHz, CDCl₃) of Analog CID 51003701

UPLC-MS Chromatogram of Analog CID 51003701

Page 114 of 130
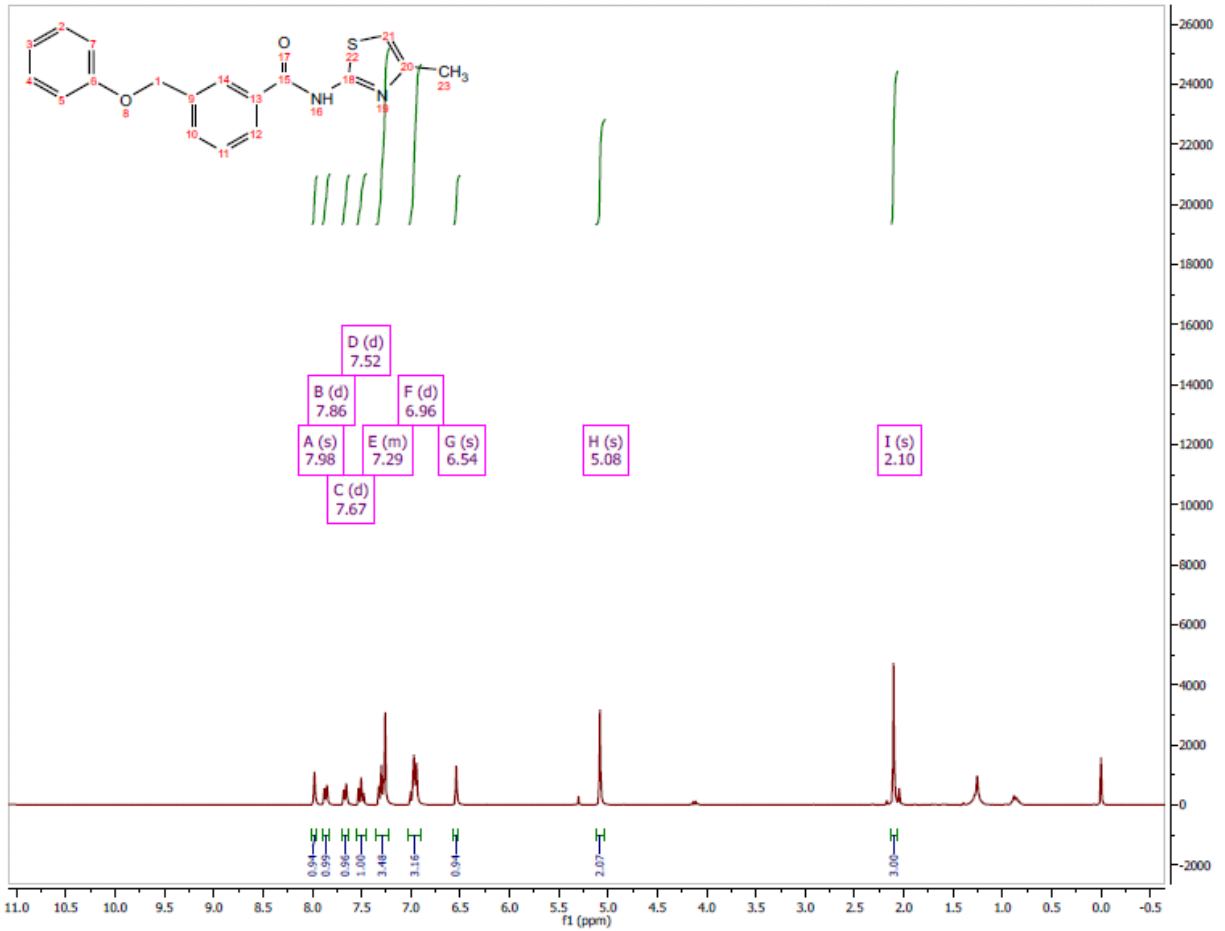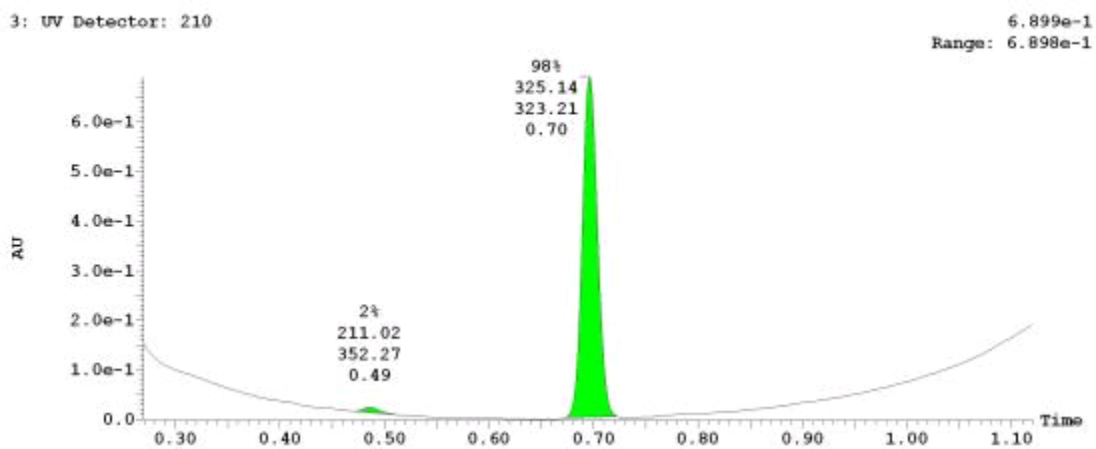
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 51003726

UPLC-MS Chromatogram of Analog CID 51003726
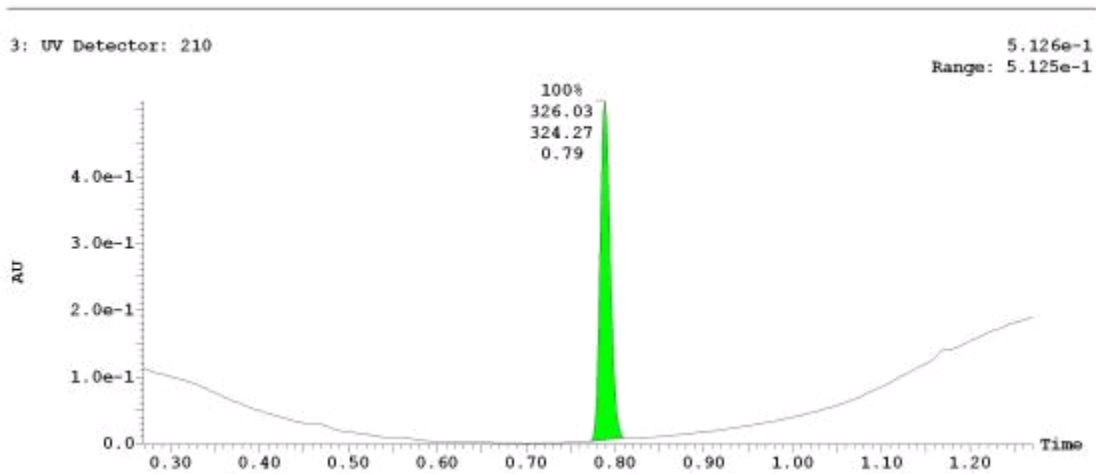
BROAD

I N S T I T U T E
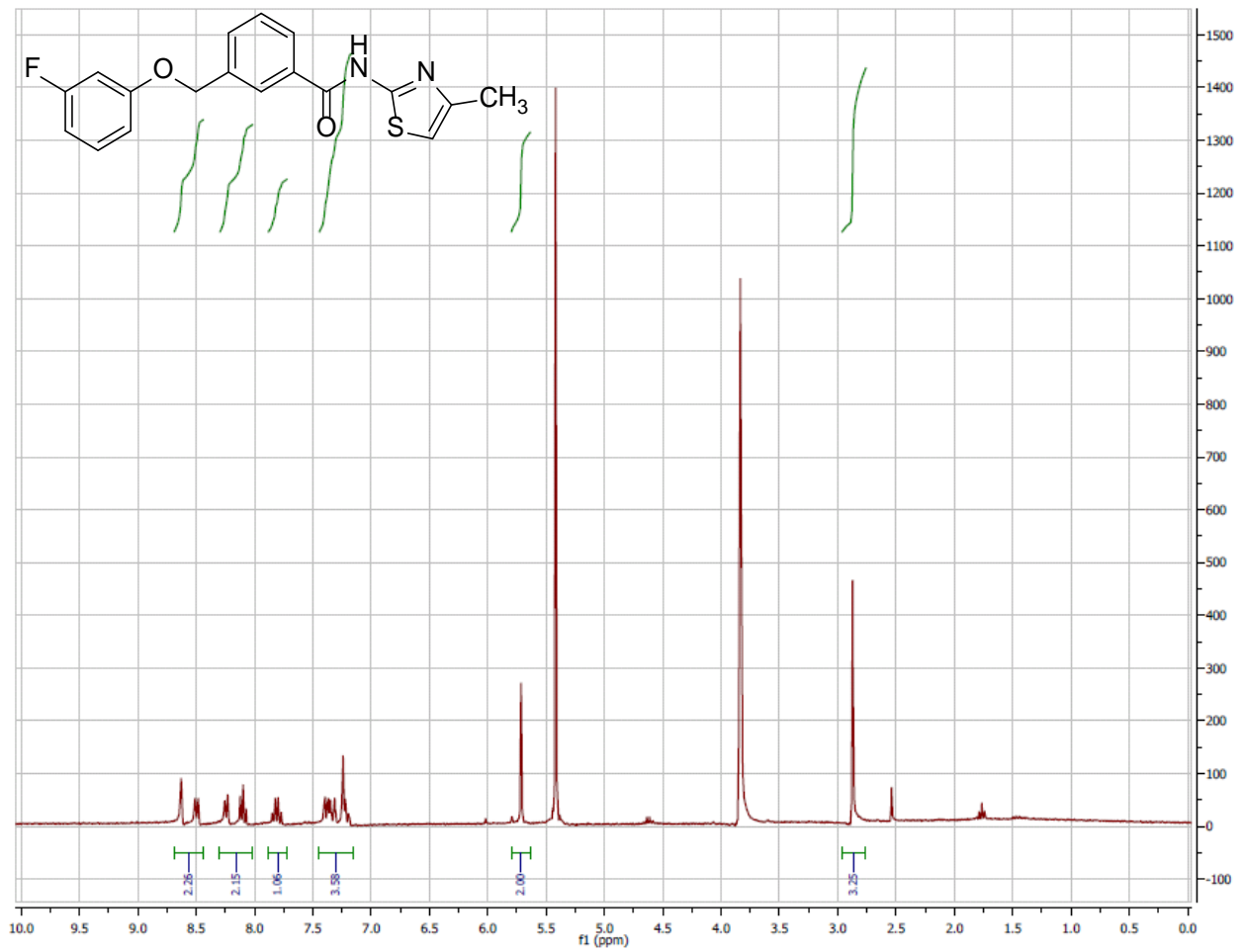
¹H NMR Spectrum (300 MHz, CDCl₃) of Analog CID 51351634
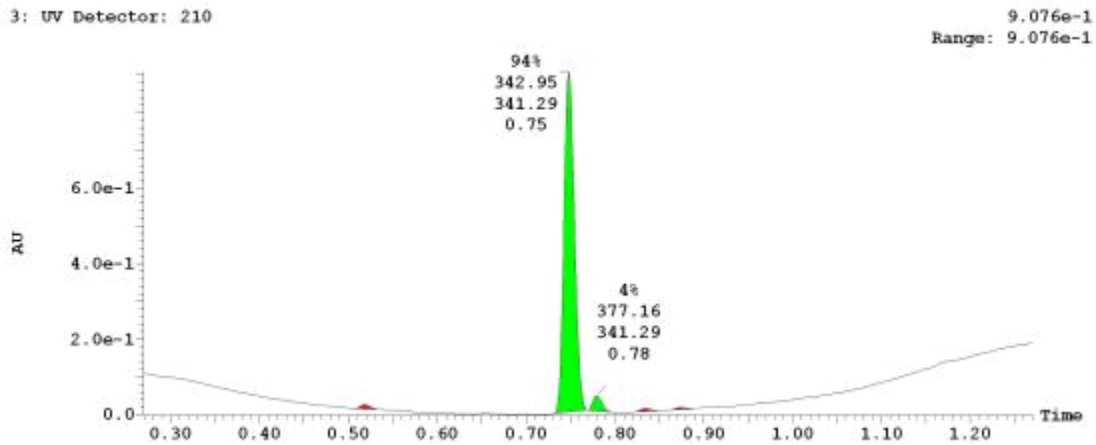
UPLC-MS Chromatogram of Analog CID 51351634

UPLC-MS Chromatogram of Analog CID 51351627
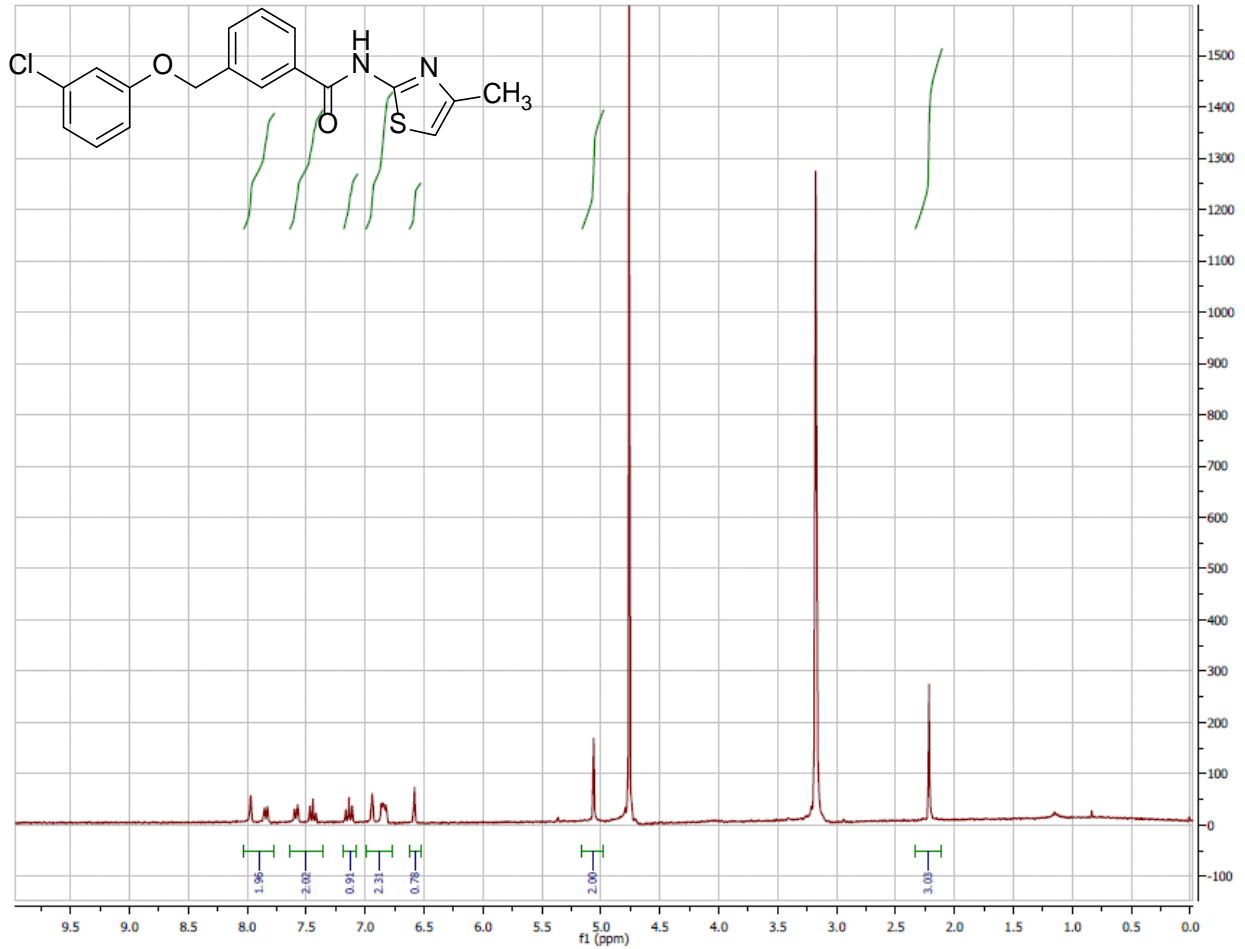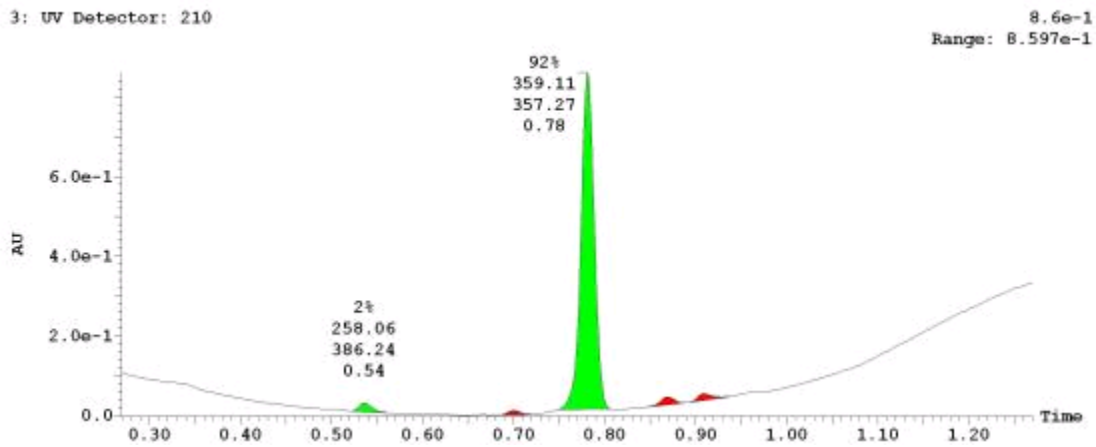
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 51351617
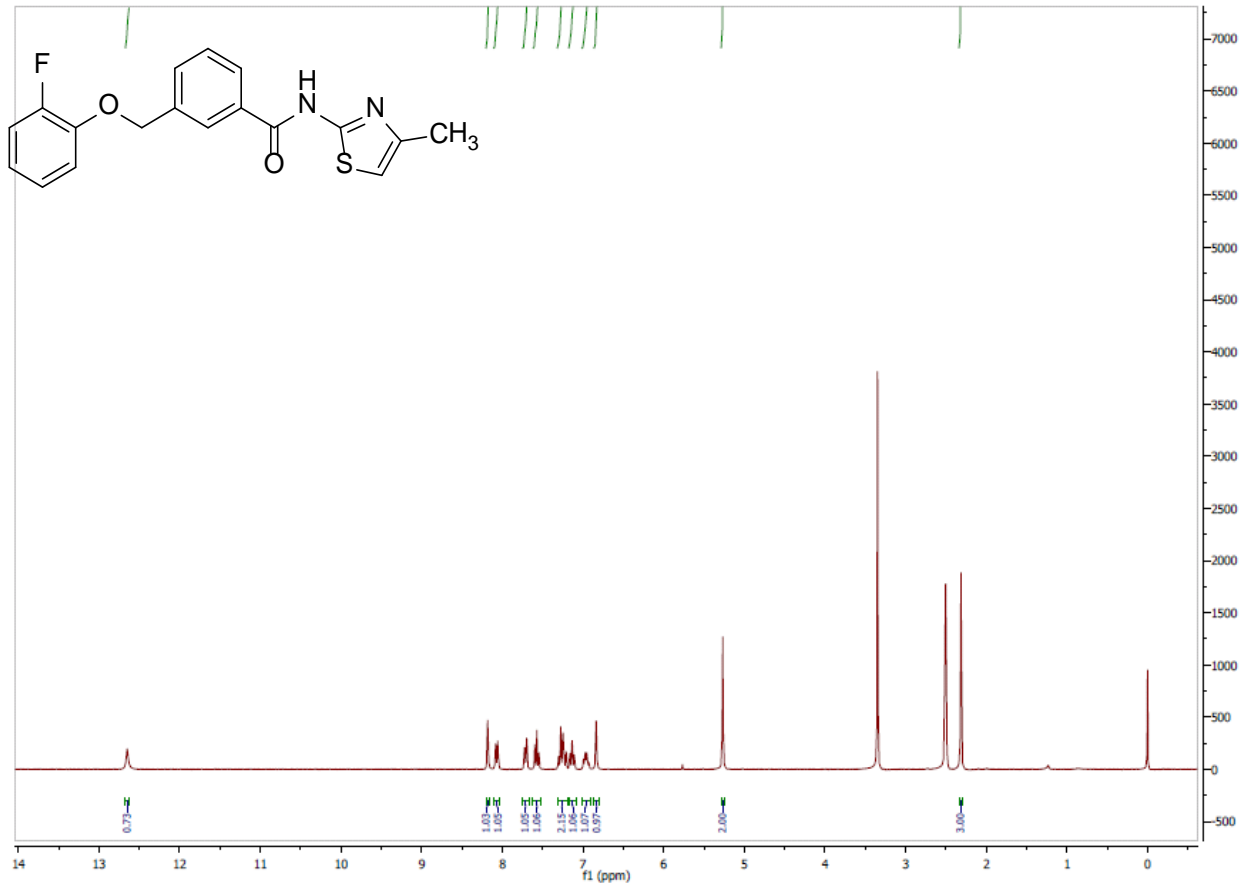
UPLC-MS Chromatogram of Analog CID 51351617
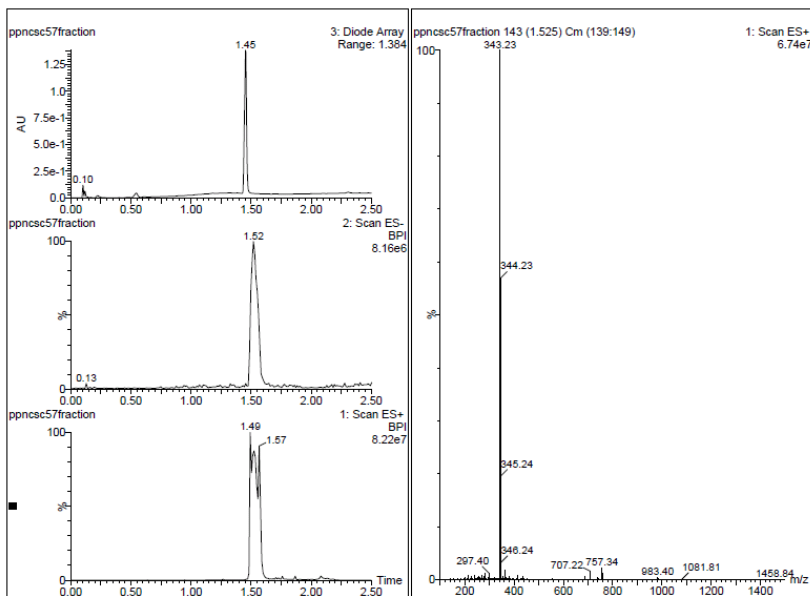
^{1}H NMR Spectrum (300 MHz, CDCl$_3$) of Analog CID 51351618
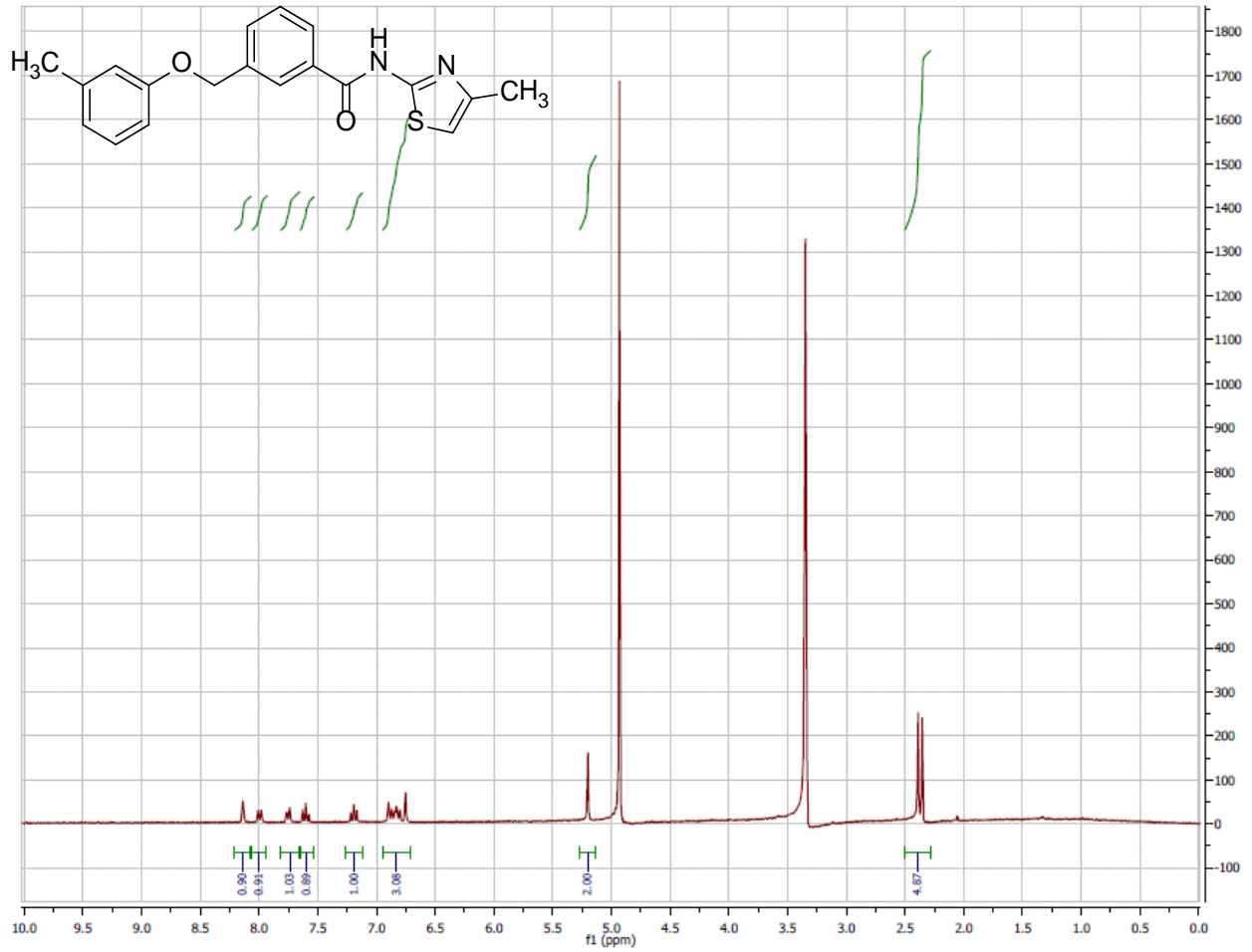
UPLC-MS Chromatogram of Analog CID 51351618

3: UV Detector: 210

9.907e-1
Range: 9.907e-1

97%
339.03
337.30
0.78
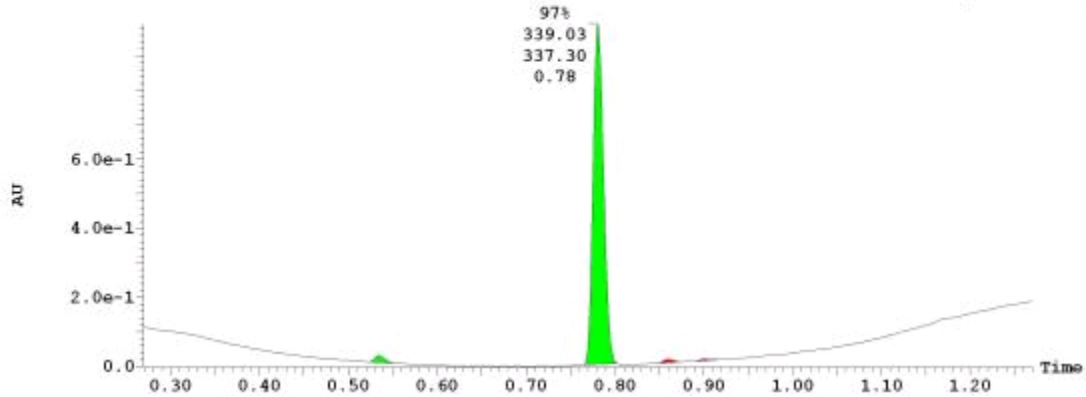
Appendix G: Prior Art Search

Table A2.   Search Strings and Databases Employed in the Prior Art Search
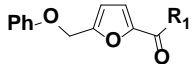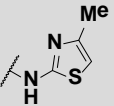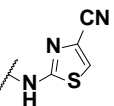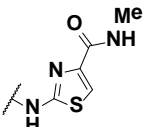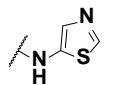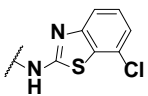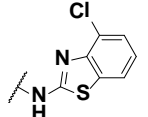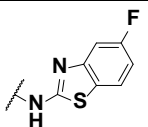
| Search String	Database	Hits Found
"breast cancer stem cells"	Sci Finder	384
"breast cancer AND stem cells"	Patent Lens	1264
"breast cancer stem cells"	Entrez (PubChem Bioassay)	3151 (35)
"salinomycin"	SciFinder	1413
"NF-κβ inhibitors AND BCSC"	SciFinder	18
"Wnt inhibitors AND BCSC"	SciFinder	30
"CXCR inhibitors AND BCSC"	SciFinder	9
"Notch inhibitors AND BCSC"	SciFinder	26
"Hedgehog inhibitors AND BCSC"	SciFinder	8
"MCF7 AND stem cells"	Entrez (PubChem Bioassay)	93 (5)
"HMLE"	Entrez (PubChem Bioassay)	7 (4)
"HMLE AND stem cells"	SciFinder	49
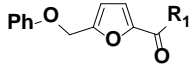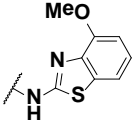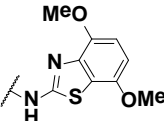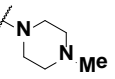
Table A3.   Prior Art Substance Identifiers

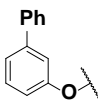| Compound Name	CAS No.	PubChem CID
Repertaxin 266359	266359-83-5	9932389
Parthenolide 20554	20554-84-1	5420804
8- Quinolinol	148-24-3	1923
Sulforaphane	4478-93-7	9577379
Metformin	657-24-9	4091
Berberine	633-66-9	2353
Salinomycin	53003-10-4	23682228
Curcumin	458-37-7	969516
Piperine	94-62-2	638024

Appendix H: SAR Analysis
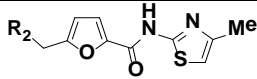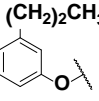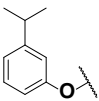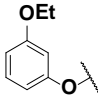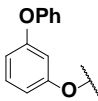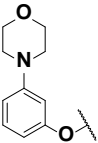
Table A4 SAR of the Amide (cont'd)

SAR Analysis			Structure	Target Activity		Anti-target Activity		
Entry No.	CID SID Broad ID	*	Ph–O–[furan]–R₁ / R₁	n [a]	HMLE_sh _ECad EC₅₀ (µM)	n	HMLE_sh _Control EC₅₀ (µM)	Selectivity
1	4791237 104170322 BRD-K39514895	S	(Me thiazole structure)	7	1.64	8	24.59	15.0
	Solubility (PBS): 2.6 µM							Purity 91%
2	50910544 112208903 BRD-K90712915	S	(CN thiazole structure)	7	12.15	7	35.14	2.89
	Solubility (PBS): <1 µM							Purity >95%
3	50910520 112208907 BRD-K63469263	S	(O=C–NH–Me thiazole structure)	5	Inactive	5	Inactive	NA
	Solubility (PBS): <1 µM							Purity >95%
4	51003697 117687954 BRD-K90461858	S	(thiazole structure)	3	16.93	3	56.2	3.31
	Solubility (PBS): <1 µM							Purity: 94%
5	50904130 110723038 BRD-K43168048	S	(Cl benzothiazole structure)	5	11.71	5	17.20	1.47
	Solubility (PBS): <1 µM							Purity >95%
6	50944074 115950044 BRD-K42965569	S	(Cl benzothiazole structure)	7	Inactive	7	Inactive	NA
	Solubility (PBS): <1 µM							Purity >95%
7	50944057 115950035 BRD-K80182713	S	(F benzothiazole structure)	7	Inactive	7	Inactive	NA
	Solubility (PBS): 1.3 µM							Purity >95%

| SAR Analysis			Structure	Target Activity		Anti-target Activity		
Entry No.	CID SID Broad ID	*	Ph-O—[furan]—R$_1$ (=O)	n [a]	HMLE_sh _ECad EC$_{50}$ (µM)	n	HMLE_sh _Control EC$_{50}$ (µM)	Selectivity
8	51003707 124360016 BRD-K33760127	S	MeO [benzothiazole] N H S	5	Inactive	5	Inactive	NA
	Solubility (PBS): ND							Purity: >95%
9	50944053 115950040 BRD-K87410972	S	MeO [benzothiazole] N H S OMe	4	Inactive	4	Inactive	NA
	Solubility (PBS): <1 µM							Purity >95%
10	8180118 104170324 BRD-K21355271	S	N [thiazoline] N H S	5	30.24	5	Inactive	30.24 vs inactive
	Solubility (PBS): 23.7 µM							Purity >95%
11	50904132 110723040 BRD-K27575760	S	[pyridine] N H N	3	Inactive	3	29.65	NA
	Solubility (PBS): <1 µM							Purity >95%
12	7454731 104170325 BRD-K96816443	S	[phenyl] N H	5	78.10	6	75.54	0.97
	Solubility (PBS): <1 µM							Purity >95%
13	50904131 110722992 BRD-K62334199	S	N [morpholine] O	5	Inactive	5	Inactive	NA
	Solubility (PBS): 458 µM							Purity >95%
14	50904155 110723034 BRD-K25954141	S	N [piperazine] N Me	3	Inactive	3	Inactive	NA
	Solubility (PBS): 482 µM							Purity >95%
15	7678045 112208923 BRD-K96502119	S	Me N H Me	7	Inactive	7	Inactive	NA
	Solubility (PBS): 418 µM							Purity 95%
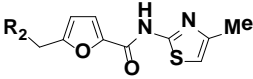
NA = not applicable; *S = synthesized

Table A5 SAR of the Phenyl Group (*meta*-substitution) cont'd

SAR Analysis			Structure		Target Activity		Anti-target Activity		
Entry No.	CID SID Broad ID	*	R_2—[furan]—C(O)NH—[thiazole]—Me R_2		n^a	HMLE_sh_ECad EC$_{50}$ (µM)	n	HMLE_sh_Control EC$_{50}$ (µM)	Selectivity
1	4791237 104170322 BRD-K39514895	S	[phenyl-O]		7	1.64	8	24.59	15.0
	Solubility (PBS): 2.6 µM								Purity 91%
2	51003691 117687950 BRD-K71337125	S	(CH$_2$)$_2$CH$_3$ [phenyl-O]		6	0.867	6	2.61	3.01
	Solubility (PBS): <1 µM								Purity: >95%
3	51003692 117687949 BRD-K79248962	S	[isopropyl-phenyl-O]		6	1.25	6	2.82	2.25
	Solubility (PBS): <1 µM								Purity: >95%
4	51003696 117687922 BRD-K97223794	S	Ph [phenyl-O]		6	1.25	6	2.82	2.25
	Solubility (PBS): <1 µM								Purity: >95%
5	50944051 115950036 BRD-K89661009	S	OEt [phenyl-O]		7	1.59	7	30.71	19.31
	Solubility (PBS): <1 µM								Purity: >95%
6	51003692 117687949 BRD-K72154374	S	OPh [phenyl-O]		8	0.627	8	1.58	2.52
	Solubility (PBS): <1 µM								Purity: >95%
7	51003703 117687929 BRD-K39494786	S	[morpholine-N-phenyl-O]		5	13.48	5	12.74	0.95
	Solubility (PBS): <1 µM								Purity: 93%
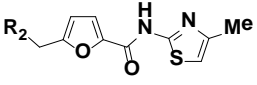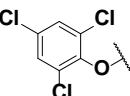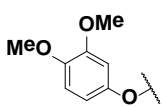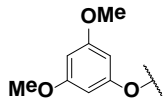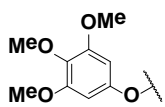
*S = synthesized
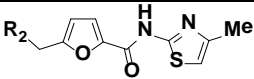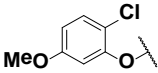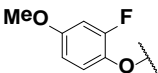
Table A6 SAR of Phenyl Replacements

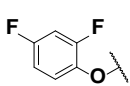| Entry No.	CID SID Broad ID	*	Structure R₂	n[a]	HMLE_sh_ECad EC₅₀ (µM)	n	HMLE_sh_Control EC₅₀ (µM)	Selectivity
			SAR Analysis / **Structure** / **Target Activity** / **Anti-target Activity**					
1	4791237 104170322 BRD-K39514895	S		7	1.64	8	24.59	15.0
	Solubility (PBS): 2.6 µM							Purity 91%
2	50910540 112208934 BRD-K35546108	S	MeO	4	Inactive	4	Inactive	NA
	Solubility (PBS): 493 µM							Purity >95%
3	42589576 112208913 BRD-K40093526	S	Me Me O	4	Inactive	4	Inactive	NA
	Solubility (PBS): 445 µM							Purity >95%
4	50910524 112208905 BRD-K21807997	S	Me Me Me O	4	Inactive	4	Inactive	NA
	Solubility (PBS): 246 µM							Purity 93%
5	50910521 112208909 BRD-K98322867	S		4	43.88	4	Inactive	>50
	Solubility (PBS): 90.8 µM							Purity 94%
6	50910531 112208917 BRD-K37617520	S		4	15.28	4	Inactive	>50
	Solubility (PBS): 12.5 µM							Purity >95%

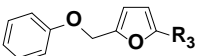NA = not applicable; *S = synthesized
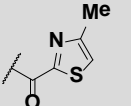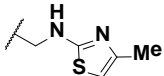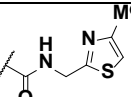
Table A7 SAR of the Phenyl Group (*di-* and *tri-*substitution) cont'd

Entry No.	CID SID Broad ID	*	Structure R₂	n ᵃ	HMLE_sh_ECad EC₅₀ (µM)	n	HMLE_sh _Control EC₅₀ (µM)	Selectivity
			SAR Analysis → **Structure** (R₂N–furan–C(O)–NH–thiazole–Me)		**Target Activity**		**Anti-target Activity**	
1	4791237 104170322 BRD-K39514895	S	*(phenoxy)*	7	1.64	8	24.59	15.0
	Solubility (PBS): 2.6 µM							Purity 91%
2	51003721 117687927 BRD-K06142398	S	*(trichloro-phenoxy)*	5	0.043	5	0.054	1.25
	Solubility (PBS): <1 µM							Purity: >95%
3	51003724 117687943 BRD-K15732868	S	*(dimethoxy-phenoxy)*	3	19.75	3	31.40	1.59
	Solubility (PBS): <1 µM							Purity: >95%
4	51003706 117687925 BRD-K17052053	S	*(dimethoxy-phenoxy)*	3	inactive	3	inactive	NA
	Solubility (PBS): <1 µM							Purity: >95%
5	51003709 117687919 BRD-K76335092	S	*(dimethoxy-phenoxy)*	5	3.73	5	55.94	15.0
	Solubility (PBS): <1 µM							Purity: >95%
6	51003704 117687935 BRD-K23011031	S	*(dimethoxy-phenoxy)*	3	Inactive	3	Inactive	NA
	Solubility (PBS): <1 µM							Purity: >95%
7	51003710 117687921 BRD-K51245976	S	*(trimethoxy-phenoxy)*	3	22.77	3	24.37	1.07
	Solubility (PBS): <1 µM							Purity: >95%
8	51003689 117687938 BRD-K81807319	S	*(difluoro-phenoxy)*	5	0.893	5	14.64	16.4
	Solubility (PBS): <1 µM							Purity: >95%

SAR Analysis			Structure	Target Activity		Anti-target Activity		
Entry No.	CID SID Broad ID	*	R_2 furan-amide-thiazole-Me core	n^a	HMLE_sh_ECad EC_{50}	n	HMLE_sh_Control EC_{50} (µM)	Selectivity
9	50944077 115950041 BRD-K24288847	S	Cl / MeO / O substituted benzene	4	0.108	4	0.339	3.14
	Solubility (PBS): <1 µM							Purity: >95%
10	50944062 115950037 BRD-K62520235	S	MeO / F / O substituted benzene	4	3.16	4	18.93	5.99
	Solubility (PBS): <1 µM							Purity: >95%
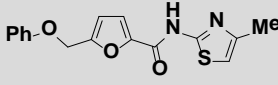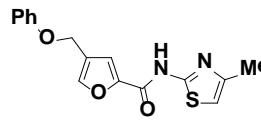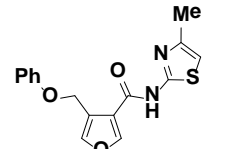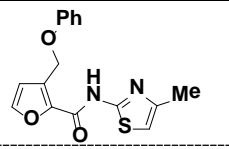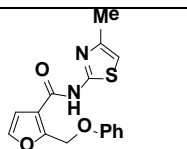
NA = not applicable; *S = synthesized

Table A8 SAR of the Eastern Portion

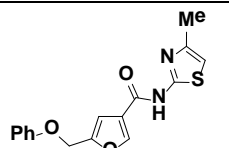| SAR Analysis			Structure phenyl-O-CH2-furan-R_3	Target Activity		Anti-target Activity		
Entry No.	CID SID Broad ID	*	R_3	n^a	HMLE_sh_ECad EC_{50} (µM)	n	HMLE_sh_Control EC_{50} (µM)	Selectivity
1	4791237 104170322 BRD-K39514895	S	Me / thiazole-carbonyl	7	1.64	8	24.59	15.0
	Solubility (PBS): 2.6 µM							Purity 91%
2	50910538 112208919 BRD-K47466502	S	H-N-thiazole-Me	8	Inactive	8	Inactive	NA
	Solubility (PBS): <1 µM							Purity: >95%
3	50904163 110722987 BRD-K92512370	S	Me / thiazole-CH2-NH-carbonyl	5	Inactive	5	Inactive	NA
	Solubility (PBS): 253 µM							Purity >95%
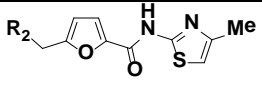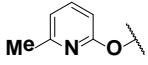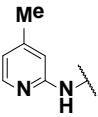
NA = not applicable; *S = synthesized

Table A9 Regioisomers of the Furan

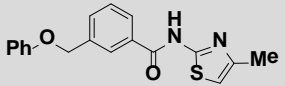| Entry No.	CID SID Broad ID	*	Structure	n [a]	HMLE_sh _ECad EC$_{50}$ (µM)	n	HMLE_sh _Control EC$_{50}$ (µM)	Selectivity
	SAR Analysis		**Structure**	**Target Activity**		**Anti-target Activity**		
1	4791237 104170322 BRD-K39514895	S		7	1.64	8	24.59	15.0
	Solubility (PBS): 2.6 µM							Purity 91%
2	51003687 117687933 BRD-K55772369	S		3	0.574	3	0.368	0.64
	Solubility (PBS): <1 µM							Purity: >95%
3	50944060 115950047 BRD-K16376786	S		4	2.05	4	3.95	1.92
	Solubility (PBS): <1 µM							Purity >95%
4	50944066 115950034 BRD-K75808851	S		4	Inactive	4	Inactive	NA
	Solubility (PBS): <1µM							Purity >95%
5	50944072 117687939 BRD-K58357915	S		6	23.62	6	Inactive	>50
	Solubility (PBS): <1 µM							Purity >95%
6	50944072 115950045 BRD-K58357915	S		4	17.35	4	29.65	1.71
	Solubility (PBS): <1 µM							Purity >95%

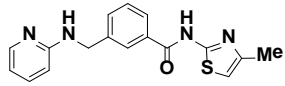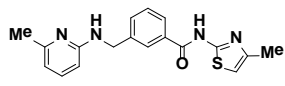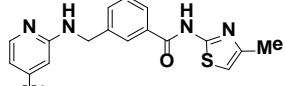NA = not applicable; *S = synthesized

Table A10 SAR of the Phenyl Group Substituted Hydroxy- and Aminopyridines

SAR Analysis			Structure	Target Activity		Anti-target Activity		
Entry No.	CID SID Broad ID	*	 R_2	n^a	HMLE_sh_ECad EC_{50} (uM)	n	HMLE_sh_Control EC_{50} (uM)	Selectivity
1	4791237 104170322 BRD-K39514895	S		7	1.64	8	24.59	15.0
	Solubility (PBS): 2.6 µM							Purity 91%
2	51351616 121269757 BRD-K05411689	S		3	21.87	3	Inactive	21.87 vs. inactive
	Solubility (PBS): 11.3 µM							Purity: >95%
3	51351636 121269746 BRD-K54708098	S		3	17.87	3	inactive	17.87 vs. inactive
	Solubility (PBS): <1 µM							Purity >95%
4	51351638 121269742 BRD-K53999057	S		3	33.2	3	Inactive	33.2 vs. inactive
	Solubility (PBS): 4.9 µM							Purity: 86%
5	50944062 115950037 BRD-K22613736	S		3	42.63	3	Inactive	42.63 vs. inactive
	Solubility (PBS): ND							Purity: >95%

ND = not determined; *S = synthesized

Table A11 SAR on Furan Replacement (Phenyl) cont'd

Entry No.	CID SID Broad ID	*	Structure[**]	n[a]	HMLE_sh_ECad EC$_{50}$ (µM)	n	HMLE_sh_Control EC$_{50}$ (µM)	Selectivity
	SAR Analysis		**Structure[**]**	**Target Activity**		**Anti-target Activity**		
1	51003701 117687945 BRD-K20316594	S		3	4.76	3	16.52	3.47
	Solubility (PBS): <1 µM							Purity: >95%
2	51351635 121269747 BRD-K47798215	S		3	34.7	3	Inactive	34.7 vs. inactive
	Solubility (PBS): 2.5 µM							Purity: 79%
3	51351611 121269763 BRD-K07759637	S		3	20.2	3	inactive	20.2 vs. inactive
	Solubility (PBS): <1 µM							Purity: >95%
4	51351620 121269755 BRD-K66028472	S		3	17.2	3	Inactive	17.2 vs. inactive
	Solubility (PBS): <1 µM							Purity: 84%

*S = synthesized; **No general structure as whole structure for each compound is displayed.

Appendix I: Compounds Submitted to Evotec
Table A12. Probe and Analog Information

BRD	SID	CID	P/A	MLSID	ML
BRD-K59019422-001-03-9	124404435	50904134	P	MLS003672589	ML245
BRD-K39514895-001-08-1	124404432	4791237	A	MLS003672590	NA
BRD-K16234721-001-02-6	124360022	51003702	A	MLS003672591	NA
BRD-K31780183-001-02-0	124404436	50904129	A	MLS003672593	NA
BRD-K65643278-001-03-3	124404434	50904149	A	MLS003672594	NA
BRD-K24240386-001-02-9	124404433	9214159	A	MLS003672592	NA

A = analog; NA= not applicable; P = probe