Structure-based and knowledge-informed design of B-Raf inhibitors devoid of deleterious PXR binding

Melanie Schneider, Vanessa Delfosse, Muriel Gelin, Marina Grimaldi, Meritxell Granell, Laurène Heriaud, Jean-Luc Pons, Martin Cohen Gonsaud, Patrick Balaguer, William Bourguet, et al.

Structure-based and knowledge-informed design of B-Raf inhibitors devoid of deleterious PXR binding

Melanie Schneider[1], Vanessa Delfosse[1], Muriel Gelin[1], Marina Grimaldi[2], Meritxell Granell[1], Laurène Heriaud[1], Jean-Luc Pons[1], Martin Cohen Gonsaud[1]*, Patrick Balaguer[2]*, William Bourguet[1]*, Gilles Labesse[1]*

[1]Centre de Biologie Structurale (CBS), CNRS, INSERM, Univ Montpellier, F-34090 Montpellier, France

[2]Institut de Recherche en Cancérologie de Montpellier (IRCM), INSERM, Univ Montpellier, F-34090 Montpellier, France

*Corresponding Authors: martin.cohen-gonsaud@cbs.cnrs.fr, patrick.balaguer@inserm.fr, bourguet@cbs.cnrs.fr, labesse@cbs.cnrs.fr

Abstract:

Dabrafenib is an anticancer drug currently used in the clinics, alone or in combination. However, dabrafenib was recently shown to potently activate the human nuclear receptor pregnane X receptor (PXR). PXR activation increases the clearance of various chemicals and drugs, including dabrafenib itself. It may also enhance cell proliferation and tumor aggressiveness. So, there is a need for rational design of a potent protein-kinase B-Raf inhibitor devoid of binding to the secondary target PXR and resisting rapid metabolism. By solving the crystal structure of dabrafenib bound to PXR, and analyzing its binding mode into both PXR and its primary target, B-Raf-V600E, we were able to derive new compounds with nanomolar activity against B-Raf and no detectable affinity for PXR. The crystal structure of B-Raf in complex with our lead compound revealed a sub-domain swapping of the activation loop with potentially important functional implication for a prolonged inhibition of B-Raf V600E.

Introduction

Dabrafenib (DB) is an ATP-competitive inhibitor of the B-Raf serine/threonine kinase involved in the regulation of the mitogen-activated protein kinase (MAPK) pathway by phosphorylating MEK that in turn phosphorylates Erk-2[1]. It was developed to specifically target the V600E activating mutation of B-Raf, which is observed in more than 40% of melanom[2]. DB is currently approved for the treatment of advanced melanoma and metastatic non-small cell lung cancer with a B-Raf-V600E mutation[1]. It shows good initial response rates and improved overall survival of B-Raf-V600E mutant cancer patients, but unfortunately, resistance is rapidly acquired[3]. To limit resistance, combination with MEK inhibitors, such as trametinib, is now more frequently used in the clinics[4]. DB treatment also suffers from a so-called paradoxical effect[5], *i.e.* appearance of secondary tumors. Indeed, while inhibiting the growth of B-Raf-V600E harboring cell lines (e.g. A375), the drug induces an opposite behavior in

wild-type B-Raf cells (HL60). It is also promoting growth and metastasis of tumor cells with RAS mutations[3]. It may also play a role in skin lesions such as squamous cell carcinomas[6].

This poor overall outcome of DB treatment may be also related to low *in vivo* stability and its unfavorable fate within human bodies. Indeed, DB is known to be rapidly metabolized, first by CYP2C8 to form hydroxyl-DB that is further oxidized by CYP3A4 to produce carboxy-DB. The latter metabolite accumulates and slowly but spontaneously loses its carboxylate group to form desmethyl-DB[3]. This rapid metabolism of DB (half-life of 5h) severely decreases its effectiveness and implies its use at very high doses (150 mg twice daily). Accordingly, the level of DB in the body alternates between high concentrations likely impacting unwanted secondary targets such as other protein-kinases, and potentially low concentrations with only partial inhibition of its primary target B-Raf V600E, the wild-type B-Raf or the related C-Raf (IC_{50} of 0.7, 5.2, and 6.3 nM, respectively according to Rheault et al[7] but IC_{50} of 5.4, 2.7, and 1.5 nM by Zhang et al[5]). This alternation is expected to favor resistance appearance on a long-term treatment. Of note, partial inhibition of B-Raf could favor the paradoxical effect through the deleterious activation of B-Raf and/or C-Raf through allosteric regulation[8].

In addition, DB has been described as an inducer of major actors of drug metabolism such as several cytochromes P450 (CYP3A4, CYP2C9, CYP2C8, CYP2C19) and some UGT enzymes through an undisclosed mechanism[3]. We recently demonstrated that DB is a potent agonist of the human pregnane X receptor (PXR) and, therefore, a strong inducer of several P450s[9], two properties that clearly account for the poor pharmacokinetics profile of this anticancer drug. Indeed, PXR plays a critical role in liver and gastrointestinal metabolism in regulating most metabolizing enzymes (including, for example, CYP3A4) or transporters at the transcriptional level[10]. As such PXR is a key regulator of the mammalian xenobiotic response and drives clearance of many exogenous substances such as pollutants and drugs. Furthermore, activation of PXR is also involved in detrimental effects including drug-drug interactions (DDI), resistance to cancer therapy or accumulation of toxic metabolites.

Therefore, some drugs or drug candidates have been already shown to activate PXR and some examples of drug re-design have been described[11].

For long-term treatments as in cancer therapy, more detrimental impact may come into play. PXR has been shown to be involved in the modulation of tumor progression and can behave either as a tumor suppressor or as an enhancer[12]. PXR is expressed in various cancerous tissues where its activation has been shown, in some cases, to increase cell proliferation and tumor aggressiveness[13,14]. Specifically, PXR regulates the expression of a gene, *fgf19,* which is involved in colon cell proliferation[13]. In addition, activation of PXR can accelerate cell cycle progression *via* inhibition of some Foxo transcription factors[15]. However, in various cancers, including melanoma, higher PXR expression does not correlate with poorer overall survival at 1 or 2 years (see www.proteinaltas.org). On the contrary, higher expression of downstream cytochromes P450 (CYP3A4, CYP2C9 and CYP2C8) seems to decrease the survival rate at 1 or 2 years (www.proteinaltas.org). Accordingly, these conflicting results suggest that further studies are needed to properly decipher the impact of PXR activation.

In this context, it is therefore important to take into account the potential of PXR activation in the rational development of anticancer drugs and of their combinations to prevent premature drug metabolism, limit DDI and, maybe more importantly, reduce also the risk of adverse events such as induction of secondary tumors. In this regard, there is a growing interest in modifying and re-designing existing drugs (or drug-candidates) in order to reduce their deleterious side effects and acquired resistances, for instance by slowing down their fast metabolism rates.

In this study, we solved the crystal structure of DB bound to the PXR ligand binding domain (LBD). Structural analysis brought unprecedented highlights into the interactions of DB with PXR, and allowed the rational design of new derivatives remaining highly active against B-Raf but devoid of binding capacity to PXR. To limit organic synthesis and cell

assays to the most promising derivatives, we set-up an integrated approach combining virtual synthesis, docking and screening with the use of dedicated models based on machine learning. Additionally, the crystal structure of our lead compound bound to B-Raf V600E revealed an unanticipated structural rearrangement of the protein with putative implications for the pharmacological output.

RESULTS

Structure of PXR in complex with DB

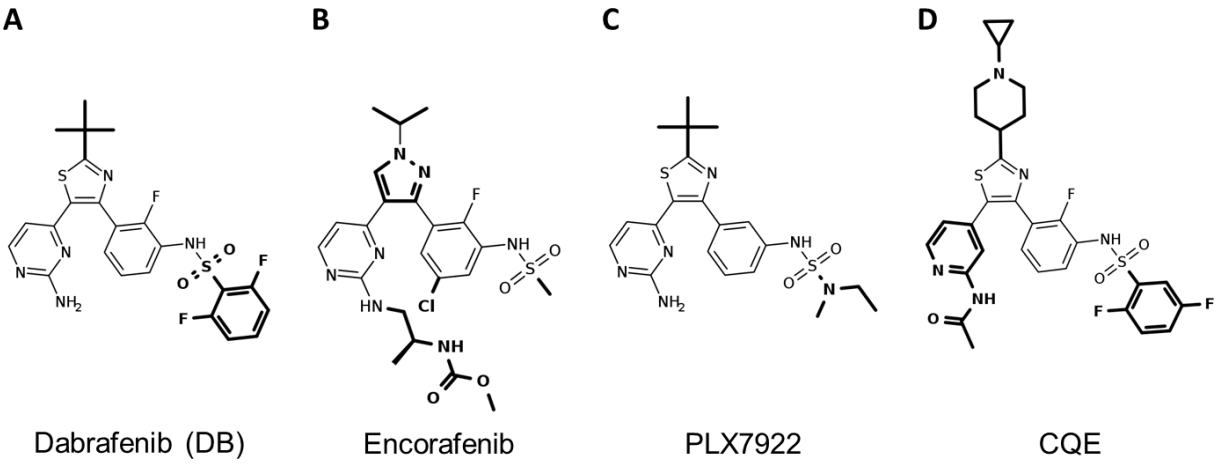


Figure 1. Chemical structures of original compounds and selected frameworks. Known inhibitors of B-Raf are displayed: A) Dabrafenib, B) Encorafenib, C) PLX7922 and D) bound ligand in PDB4CQE. For each framework specific substituents are shown in bold.

We recently identified DB (see Figure 1A) as a potent PXR agonist with an EC_{50} of 87 nM[9]. To gain atomic-level information on the mode of binding of DB to PXR, we solved the crystal structure of the PXR-LBD in complex with this compound at a resolution of 2.28 Å (Table S1). The PXR-LBD adopts the canonical active conformation of nuclear receptors, with its C-terminal activation helix H12 capping the ligand binding pocket (LBP) (Supporting Information Figure S1). DB could be positioned unambiguously in the electron density (Figure 2A) that was well defined for the entire compound, indicating that DB is well stabilized in the LBP, in line with its strong activity. DB forms mainly van der Waals interactions with LBP residues

and only one hydrogen bond is observed between the nitrogen atom of the sulfonamide moiety and Ser247 in helix H3 (Figure 2A). Surprisingly, none of the polar atoms of the sulfonamide moiety, the aminopyrimidine and thiazole rings appear to be engaged in a direct or water-mediated hydrogen bond with any LBP residues. DB adopts a U-shape conformation, with the aminopyrimidine and the difluorophenyl rings being within van der Waals distance (3.60 Å) to each other. The difluorophenyl ring and the sulfonamide moiety reside between helices H3 and H11, and closely adjacent to H12, whereas on the other side of the pocket, the *tert*-butyl group is nested into the so-called aromatic π-trap composed of Phe288, Trp299, and Tyr306[16]. Comparison of this structure with that of PXR in complex with the reference agonist SR12813 reveals that the binding sites of DB and SR12813 largely overlap, in agreement with their similar PXR agonistic activities (Figure 2B). These two ligands share a *tert*-butyl moiety that orients the same way in PXR. From this PXR-DB complex, several chemical routes for optimizations could be envisioned: decrease the hydrophobic interactions observed with the terminal difluorophenyl ring, the terminal *tert*-butyl group and/or promote van der Waals clashes.

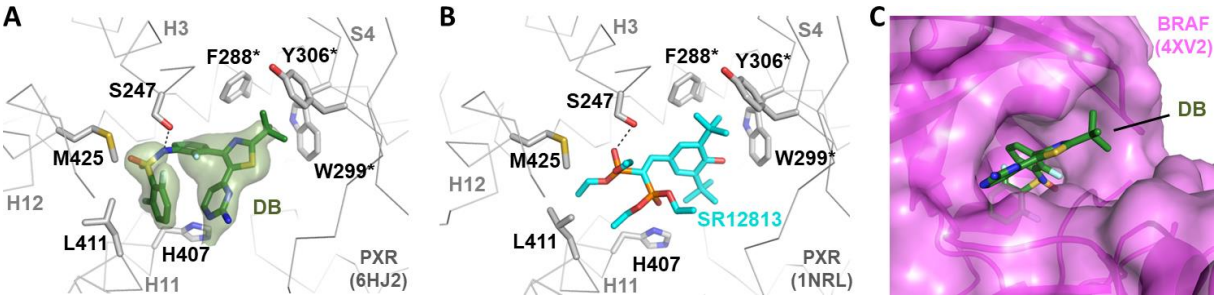


Figure 2. Interaction of dabrafenib with PXR and B-Raf. Close up view of the interactions of dabrafenib (DB) (A) and SR12813 (B) with residues of the LBP of PXR-LBD, in the same orientation, showing the similar space occupied by the two ligands. Residues of the aromatic π-trap are marked with an asterisk. The simulated annealing Fo-Fc omit map around the ligand is represented in A (in green). (C) Structure of B-RafV600E in complex with DB indicated that the *tert*-butyl group of the drug lies at the edge of the ATP-binding cavity and points towards the solvent. Oxygen, nitrogen, sulfur, phosphorous and fluorine atoms are colored in red, blue, yellow, orange and light blue, respectively.

Modeling of encorafenib binding into PXR

Because encorafenib (hereafter ECF; Figure 1B) is another B-Raf inhibitor which is both used in the clinics and related in structure to DB[1], we wondered if it could bind to PXR. Indeed, this compound was shown to be also an activator of PXR – although to a lower extent than DB – in our cell-based reporter assays (Table 1). Unfortunately, we failed so far to co-crystallize ECF-bound PXR. This prompted us to dock this drug and all its derivatives or analogues present in a focus dataset (PubChemAssay AID:1257566[17]) into the newly solved crystal structure of PXR bound to DB, using the program PLANTS[18] through an interface recently developed in the server @TOME (Docking results can be seen in: http://atome.cbs.cnrs.fr/EG/PXR-encorafenib). Those compounds share with DB, a common core made of three connected aromatic cycles harboring distinguishing substitutions. Two main binding modes were observed. One is similar to that of DB and is more frequently observed for the compounds harboring a large substituent on the sulfonamide group. An alternative binding mode was also predicted in which the aminopyrimidine ring and the central phenyl are swapped compared to the mode of binding of DB. This mainly occurred for compounds harboring a small substituent on the sulfonamide group (such as methyl, as in ECF). Hence, despite the loss of the large difluorodiphenyl moiety (replaced by a tiny methyl group) and the presence of a polar extension on the amino-pyrimidine ring, ECF was still predicted to bind to PXR to some extent. But, the lower hydrophobic surface of ECF suggested it would have a lower affinity for PXR with predictions in the sub-micromolar range in agreement with our experimental evidence. From this screening survey, several chemical routes for optimizations could be envisioned: modify the methyl sulfonamide, the isopropyl group attached to the thiazole ring and/or promote van der Waals clashes. In parallel, modification of the polar substituent on the aminopyrimidine appeared a complex issue (B-Raf affinity, cell penetration) while PXR seemed to readily accommodate such an extension. Finally, the chlorine atom added on the central phenyl ring seemed not a major change

affecting the interaction with PXR while it may impact favorably the interaction with B-Raf-V600E (see below).

An intermediate molecule in between DB and ECF

While the replacement of the large difluorophenyl moiety of DB by a tiny methyl group (as in ECF) would impact to some extent the affinity and possibly the mode of binding to PXR, it is also expected to drastically lower the affinity for the primary target B-Raf. Indeed, the corresponding difluorophenyl in DB points into a buried and hydrophobic sub-pocket. Hence, a compromise was envisioned following a previous example set up to lower the paradoxical effect[5]. In this study, the resulting compound harboring an ethylmethylamine group on the sulfonamide group (PLX7922; Figure 1C) was a significantly weaker inhibitor of B-Raf (12 nM *vs* 0.5 nM for DB). We thought that an isosteric but more hydrophobic version could be more effective. Nevertheless, it might not be sufficient to decrease drastically the binding to PXR according to the above virtual screening.

Another alternative to DB and ECF

Interestingly, a compound (Figure 1D; named hereafter CQE as in the PDB ID 4CQE and corresponding to compound **20**[19]) harboring a large and polar cyclopropylpiperidine moiety on the thiazole ring instead of the *tert*-butyl moiety of DB, has been reported to have a low-nanomolar activity against B-Raf V600E. The crystal structure of this compound bound to B-Raf V600E (PDB ID 4CQE) shows the alkylated piperidine pointing toward the solvent while contacting the glycine-rich loop and the the catalytic site. However, the interaction of this compound with PXR was not measured. Based on the binding mode of DB in PXR, the large cyclopropylpiperidine would clash too severely into the aromatic π-trap at the bottom of the PXR pocket, thus preventing such a binding mode. Alternative binding modes would require significant rearrangements of the PXR ligand binding pocket that are hard to predict. As a whole, these data suggested that the presence of such a large and polar substituent on the

thiazole group would be detrimental regarding PXR binding while it would be well accommodated by B-Raf V600E.

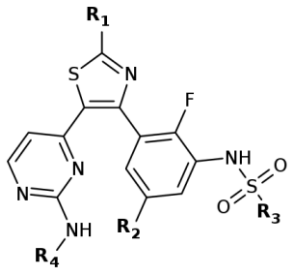
Selection of a common core for a focused optimization

The above results prompted us to further evaluate the overall chemical structure of DB and search for a combination of derivatizations that could prevent binding to PXR while maintaining a strong activity against the primary target B-Raf. Further comparisons of DB and the three other compounds was undertaken to search for feasible and promising modifications (Figure 1).

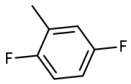
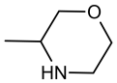
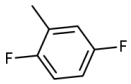
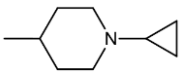
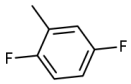
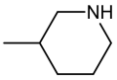
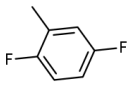
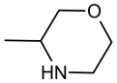
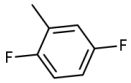
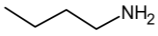
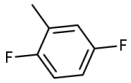
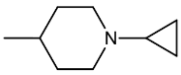
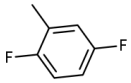
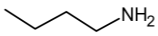
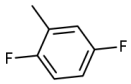
First of all, these four compounds share a common architecture with an isosteric core built from an aminopyrimidine (but in CQE having a closely related aminopyridine) directly connected to a 5-atom heterocycle (a thiazole but in ECF which harbors a pyrazole of similar shape), which in turn is attached directly to a fluorophenyl ring (with an additional chlorine in ECF) that harbors a sulfonamide group. More subtle variations were also detected such as the distinct position of one fluorine atom present on the difluorophenyl moiety in DB and CQE (Figure 1). CQE also harbors an acetyl group on its aminopyridine ring while DB and PLX7922 have no substituent on their aminopyrimidine moiety. As mentioned above, ECF differs from the other three compounds by the presence of a chlorine atom on the central phenyl ring and a polar extension on its aminopyrimidine. Interestingly, ECF[19], PLX7922[5] and CQE have all been proposed to promote lower paradoxical effect[20], and represent, therefore, promising alternative frameworks in order to produce better behaved derivatives of DB. However, for reason of synthesis accessibility and cost (of the aminopyridine or the pyrazol), the aminopyrimidine-thiazole pair connected to a central phenyl ring has been chosen as the most attractive scaffold for further elaborations. On this scaffold, various variations similar to those described above could be envisioned to bring the best combination of pharmaceutical behavior (high affinity for B-Raf and not PXR, possibly low paradoxical effect, cell penetration, good pharmacokinetical properties …). The most obvious ones, corresponded to

the addition of chlorine atom on the central phenyl ring (as in ECF; Figure 1B), variations on the sulfonamide group like a *sec*-butyl (to mimic PLX7922; Fig. 1C) or a o,m-difluorophenyl instead of a o,o-difluorophenyl (as in CQE; Figure 1D) and finally the addition of an acetyl group on the aminopyrimidine (to mimic CQE; Figure 1D). Finally, the most diverse substitutions of the C2 position of the thiazole were planned to replace the *tert*-butyl group.

Table 1. Structures and biological activities of DB and new compounds.



					B-Raf V600E		PXR	
					Inhib.	**Prolif.**	**Bind.**	**Activ.**
	R1	R2	R3	R4	IC50 (nM)	IC50 (nM)	ΔTm (°C)	EC50 (µM)
DB	*tert*-butyl	H	o,o-difluoroPh	H	0.7-5*	0.457	5.0	0.0872
1a	4-piperidine	H	*sec*-butyl	H	85% (1 µM)**	ND	ND	ND
2a	azetidine	H	o,m-difluoroPhl	H	2.1	32.8	ND	> 30
2b	pyrrolidine	H	o,m-difluoroPh	H	1.7	11.2	ND	9.79
2c	2-piperidine	H	o,m-difluoroPh	H	1.7	5.12	2.8	5.16
2d	2-piperazine	H	o,m-difluoroPh	H	5.9	127	0.43	> 30

3a	 *tert*-butyl	Cl	 o,m-difluoroPh	H	3.6	1.73	6.8	0.0814
3b	 5-morpholine	Cl	 o,m-difluoroPh	H	2.0	2.06	2.2	2.28
4a	 cyclopropylpiperidine	H	 o,m-difluoroPh	H	4.4	1.68	2.4	6.21
4b	 3-piperidine	H	 o,m-difluoroPh	H	4.5	44.1	ND	> 30
4c	 5-morpholine	H	 o,m-difluoroPh	H	6.0	3.77	1.2	11.5
4d	 aminopropyl	H	 o,m-difluoroPh	H	6.3	39.5	ND	> 30
5a	 cyclopropylpiperidine	H	 o,m-difluoroPh	 acetyl	ND	1.48	ND	10.8
5b	 aminopropyl	H	 o,m-difluoroPh	 acetyl	7.6	208	ND	> 30

Notes * refers to activity values in the litterature[5,7], ** refers to inhibition percentage at 1,000 nM.

Further analysis of the *tert*-butyl moiety for better optimization

Careful inspection of the crystal structures and above virtual screening results suggested that the replacement of the *tert*-butyl group on the thiazole by a larger and more hydrophilic moiety could impact the binding modes into PXR while maintaining the binding mode on B-Raf-V600E, as suggested by the above compound CQE Indeed, the structures of B-Raf or B-RafV600E in complex with DB (PDB IDs 5CSW and 4XV2, respectively) indicated that the

tert-butyl group of the drug lies at the edge of the ATP-binding cavity and points towards the solvent[5,21] (Figure 2C).

Furthermore, this *tert*-butyl group is subjected to chemical modifications by P450s so that its replacement by a more stable moiety would further decrease metabolism and excretion rate of the resulting drug. Importantly, the known metabolites of DB seem not to provide valuable alternatives according to virtual docking and affinity predictions by means of standard scoring function and also by molecular mechanics Poisson−Boltzmann surface area (MMPBSA) computations[22]. Using these approaches, two of the metabolites (hydroxymethyl- and desmethyl-DB) were predicted to have only slightly lower affinity for both B-Raf V600E and PXR, while the more polar carboxy-DB would no longer bind any of them. This new results match the above analysis regarding the van der Waals interactions deduced from the crystal structure survey. But, a precise and quicker evaluation of many other substitutions (hereafter labeled R1; Table 1) of the *tert*-butyl group would be required to select out a small subset of optimal derivatives.

Set up of a machine learning model to guide *tert*-butyl replacement

Virtual screening of a rather large number of potential substitutions was considered but required a fine ranking of the resulting compounds prior to moving toward organic synthesis and activity testing. Obviously, this could not be performed in a reasonable time by using MM-PBSA calculations, a too CPU-demanding approach. On the contrary, virtual screening allows rapid generation of plausible docking poses for many chemical compounds but at the expense of the quality of the affinity prediction. This drawback can be partially compensated by expert human/manual survey of the docking poses but at the expense of the overall throughput. So, we turned to combine classical docking simulation supplemented with machine learning technique to predict affinity as precisely as possible.

First, we docked into B-Raf, a large set of 2193 compounds with known IC50 against B-Raf V600E (extracted from BindingDB[23]) using the PLANTS module implemented in our

server @TOME[24]. This tool allows us to perform in parallel virtual docking in 20 distinct conformations of B-Raf (upon selection by ligand similarity). This led to satisfactory docking of most of the compounds. Then, we trained various affinity prediction models on generated structure-based descriptors, extended with global ligand-based molecular descriptors using the Random Forest algorithm and 10-fold cross-validation as previously described[24]. This led to predicted pIC50 values ranging from 7.0 to 8.5 (RMSE: 0.74, r_pearson: 0.72 and R^2: 0.52) showing Gaussian-like curves with little outliers. This suggested that the predictions were not very discriminative. To gain in precision and to alleviate any possible bias due to sub-optimal docking, we switched to a more focused dataset (derivatives of DB and ECF from one patent application; PubChemAssay AID:1257566) using a ligand-based approach. In this case, several machine learning models were trained using different type of fingerprints (MACCS, Pubchem …) as molecular descriptors and the two algorithms Random Forest (RF) and Support Vector Machine (SVM). Parameter optimization was performed and the 40 most important features from the different fingerprints were selected and used for a new round of training with these two algorithms. This led to two models, named, hereafter, 'RF-selected' and 'SVM-selected' showing significant improvement of both their RMSE (down to 0.54) and R^2 values (up to 0.68) in 10-fold cross-validation, while default fingerprints stand at higher RMSE (up to 0.8) and lower R^2 (down to 0.35). The 'RF-selected' predictions have the most pronounced discrimination and cover the largest affinity range (7.2 to 9.5), whereas the 'SVM-selected' predictions cover a smaller range (7.8 to 8.6) in lower agreement with experimental values. Interestingly, the two models appeared more discriminative with two small subsets of compounds (28 and 60 chemical entities, respectively) showing significantly higher affinity (RF model: pIC50 > 8.8 and 'SVM-model': pIC50 >8.3) while the remaining gathered around lower values ('RF-model': <pIC50> ~ 8.3 and 'SVM-model': <pIC50> ~ 8.1). Of note, the 28 best compounds for the 'SVM-selected' model also ranked among the best for the 'RF-selected' one. Furthermore, in these two models, the best combination of substitutions seemed to be those harboring the extra chlorine atom on the central

fluorophenyl (R2) by a margin while the worst came out to be the one with a *sec*-butyl (R3) instead of the difluorophenyl. These results suggested that we had in hand several tools to finely tune the selection of optimal combination of substituents.

In silico screening for optimal derivatizations

The simplest synthetic route to replace the *tert*-butyl group harbored by the thiazole ring of DB is to react a thioamide compound onto the desired scaffold. Accordingly, using a dedicated KNIME pipeline[25], a range of 179 thioamide compounds available from Enamine Chemical Supplier was virtually reacted onto the common scaffold in combination with focused substitutions on two other positions of interest (mainly *sec*-butyl or a difluorophenyl for R3 and H or Cl for R2, Table 1). The resulting compounds were docked into B-Raf binding sites using PLANTS through the interface in @TOME (see above). This allowed docking into multiple conformations of the protein-kinase and use of co-crystallized compound as a pharmacophoric shape restraint in each of the templates selected for virtual screening. This led to satisfactory docking of most of the compounds (http://atome.cbs.cnrs.fr/EG/BRaf-PO6). In parallel, docking into PXR was performed for counter-selection (http://atome.cbs.cnrs.fr/EG/PXR_human_P06). Then, the affinities were predicted using our best performing ML models, 'SVM-selected' and 'RF-selected'. For the R2 position, the chlorine atom appeared to bring slightly better affinity compared to a hydrogen atom. Similarly, the larger difluorophenyl in R3 increased the binding affinity to B-Raf compared to the smaller *sec*-butyl moiety. For the R1 position, 'SVM-selected' showed 28 substitutions (shown in Figure 3) lying out with significantly better predictions (pIC50 > 8.35 or 8.41) from the bulk (pIC50 < 8.25 or 8.27). Those substitutions were also among the best according to the model 'RF-selected' (pIC50 > 8.78 or 8.89 versus pIC50 < 8.40 or 8.71 for the bulk).

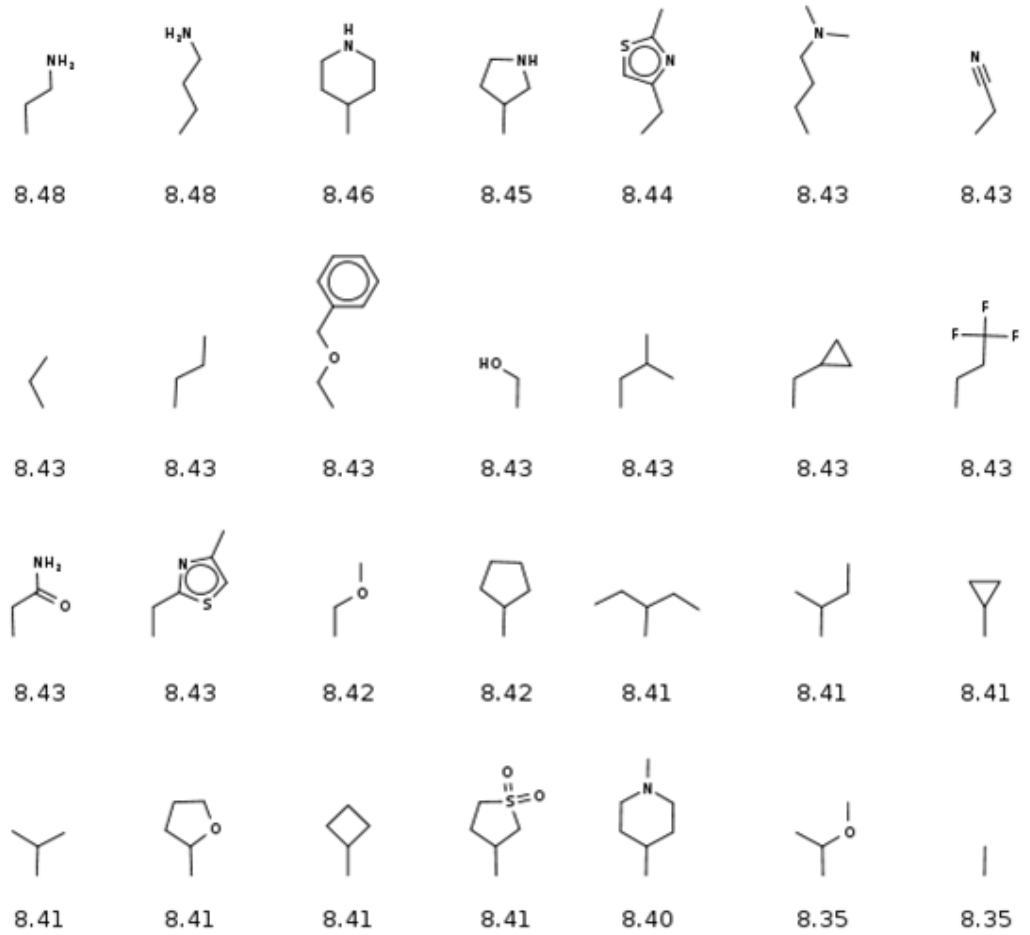


Figure 3. Chemical structures of 28 best substitutions for the *tert*-butyl. In silico generated molecules were docked and evaluated using a dedicated Random Forest model (RF-selected; see text).

In agreement with the above structural analysis, among the best B-Raf substituents R1 attached to the thiazole ring, eleven appeared to correspond to small and rather hydrophobic substituents (8.34 < pIC50 < 8.42 by the 'SVM-selected' model). Unfortunately, most of those substituents (such as the isopropyl chain) would perfectly fit the PXR binding site. Three aromatic substituents (8.42 < pIC50 < 8.45) were also found but the quality of their interactions with B-Raf was questioned. This was mainly due to seemingly unfavorable orientation of these aromatic rings and a lack of strongly favorable interactions while the presence of methylene group connecting these rings to the thiazole suggested an important and unfavorable entropy cost. The best substituents were four polar groups (8.44 < pIC50 < 8.49): two aminocycloalkyles (pyrazol and piperidine) group and two short aminoalkyl chains

(aminoethyl and aminopropyl). These four modifications seemed to be able to form favorable polar interactions with neighboring residues (N580, N581, or D594). Finally, a hydroxyl group, two methoxyalkyl chains and a pyrrole were also detected among the 28 best modifications. As representative of the four above substitutions, we selected the aminopropyl and the piperidine as promising substituents although we could wonder about their cell penetration (especially with the terminal amino group).

By analogy with the above results, we selected the cyclopropylpiperidine that was already present in CQE[20] as a reference substitution. We also selected a morpholine moiety attached to the thiazole ring by a carbon atom instead of the nearby nitrogen as usually observed in other drug compounds. This original connection would allow the formation of a hydrogen bond to the aspartate D594 (either directly or through a water molecule). Of note, the three selected heterocycloalkyl moieties (piperidine, morpholine and cyclopropylpiperidine) are well known to bring very favorable pharmacokinetics properties and are frequently encountered in drugs. MMPBSA was again applied to confirm the overall predictions from the machine learning and visualization survey for the four selected R1 substituents (aminopropyl, piperidine, morpholine, and cyclopropylpiperidine). Compared to DB (-380 kJ/mol), lower affinities for B-Raf were predicted with the following ranking: cyclopropylpiperidine (-300 kJ/mol) > piperidine (-270 kJ/mol) ~ morpholine (-285 kJ/mol) > aminopropyl (-250 kJ/mol).

Functional characterization of DB analogs

The first molecule synthesized and tested was based on the selected common scaffold harboring in R3 a *sec*-butyl (as for PLX7922, Figure 1C) and in R1 a piperidine instead of the *tert*-butyl (Table 1). However, this compound (**1a**) showed a sharply decreased activity against B-Raf (with only 85% inhibition at 1000 nM) compared to DB. Hence removing the halogenated phenyl ring in R3 was not further studied experimentally. We synthesized and tested a series of four compounds harboring as R1 group various aminocycloalkyl of 4 to 6

atom rings (**2a-2d**). While highly potent on the purified enzyme B-Raf-V600E (1.73-5.92 nM), they were much less active against A375 cell proliferation, probably due to cell penetration issue (Table 1 and Supporting Information Figure S3).

In order to evaluate the impact of chlorine atom for R2 on the central phenyl ring (as in ECF, Figure 1B) on the binding to B-Raf and PXR, two other molecules were synthesized and tested. One derives from DB itself with a *tert*-butyl in R1 (**3a**) while the second harbors a 5-morpholine group in R1 (**3b**). Interestingly, these two compounds were highly active on purified B-Raf V600E, with IC_{50} of 3.6 nM and 2.0 nM, respectively (Table 1) and also on the wild-type enzyme (1.7 and 1.6 nM, respectively). These experimental data confirmed the predicted benefit of adding a chlorine atom for binding to B-Raf. Of note, these compounds showed better activity against wild-type enzyme than mutated B-Raf V600E. As expected, **3a** is highly potent in activating PXR (EC_{50} = 81 nM), whereas **3b** was much less potent with EC_{50} of 2.3 µM. Thus, in agreement with our predictions, the chlorine atom is not sufficient to explain the differences between DB and ECF regarding their binding to PXR. As expected, common scaffold harboring a chlorine atom in R2 and a *tert*-butyl substituent in R1 still activates strongly PXR while the morpholine group in R1 prevents this in our cellular model. This confirms the importance of large and polar substituents on the thiazole group to disentangle B-Raf and PXR binding.

Then, we designed and synthesized another series of four molecules based on the common scaffold with a o,m-difluorophenyl in R3, and in R1, four different moieties sharing a cycloalkyl chain with one or two nitrogen atoms (e.g. pyrrolidine) (Table 1). The resulting compounds could be synthesized straightforwardly and were shown to be effective in the low nanomolar range (**4a**: 4.4 nM, **4b**: 4.5 nM; **4c**: 6.0 nM, and **4d**: 6.3 nM) against B-Raf V600E, placing them at an activity level similar to that of DB (0.8-5.4 nM), in perfect agreement with the various predictions described above. All these chemicals had a good antiproliferative activity in A375 cells (IC_{50} values are displayed in Table 1; see also Figure 4A). The most active compound was DB (IC_{50} 0.44 nM) while ECF and **4a** (cyclopiperidine) were slightly

less active (IC_{50}: 1.29-1.47 nM), and **4c** (morpholine) was the least active compound (IC_{50} of 3.8 nM). Having shown the activity of DB analogs on its primary target, we next evaluated the potential transactivation of PXR by DB and the designed analogues in a reporter cell line[26]. The maximal activities and EC_{50} values of the compounds are displayed in Table 1 (see also Figure 4A and Supporting Information Figure S2). They were categorized into three potency groups: weak (EC_{50} > 10 µM), moderate (10 µM > EC_{50} > 1 µM), and strong (EC_{50} < 1 µM) activators. DB is a strong activator (EC_{50} = 87 nM) while ECF is a moderate one in agreement with the loss of the difluorophenyl group favorably interacting with PXR.

As we wondered whether cell penetration might be an issue due to the four polar substitutions and to better match the related CQE, we synthesized a second series differing by an acetyl group attached to the amino group of the pyrimidine (substituent in R4). The acetylated version seemed less stable (in weeks) and only one IC_{50} could be determined on purified B-Raf V600E (7.6 nM for the aminopropyl substituted compound **5b**; Table 1). On cell culture, **5a** showed activity of 1.5 nM which is very similar to its unacetylated version **4a** (1.7 nM) while acetylation appeared too labile.

Finally, we checked direct PXR binding for a subset of representative ligands including DB and SR12813 using TSA experiment. It confirmed that ligand with low PXR activation actually bind less well to PXR-LBD (Figure 4C and Table 1). Noteworthy, the apparent binding correlates well with the hydrophobicity of the compounds as estimated using XLogP (Figure 4D).

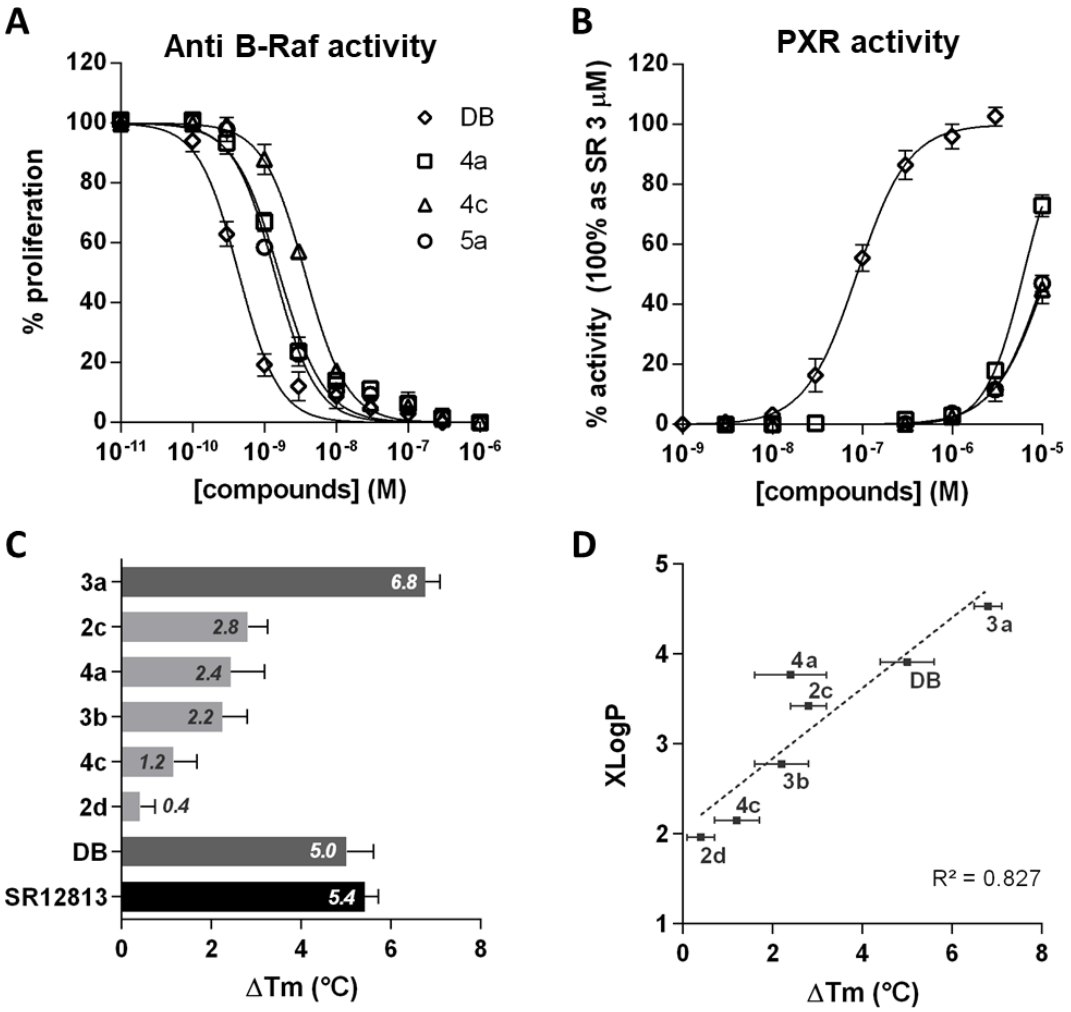


Figure 4. Cell assays for PXR activation and antiproliferative B-Raf activity. (A) Results of antiproliferative assays showing concentration response curves for DB, **4a**, **4c**, and **5a** in A375 melanoma cancer cells. A375 cells were treated with chemicals for 5 days and cell survival was measured using MTT assay. Results are expressed as percentage of the maximal response obtained with untreated cells. Data are the mean (SD) of three to five independent experiments. (B) Results of luciferase assays showing concentration response curves for DB, **4a**, **4c**, and **5a** in HG5LN GAL4hPXR cells. Results are expressed as percentage of the maximal response obtained in presence of 3 µM SR12813. Data are the mean (SD) of three to five independent experiments. (C) Relative differential melting temperatures (ΔTm) determined for PXR-LBD in presence of various compounds (**2c, 2d, 3a, 3b, 4a, 4c**, DB and SR12813). Data are represented as mean (SD) (n = 6). (D) Correlation curve between ΔTm and the XLogP (calculated with RDKit).

In conclusion, our compounds **4a**, **4c**, **5a**, and **5b** are the least potent at activating PXR, with EC_{50} in the 6-15 µM range. In conclusion, the corresponding chosen derivatizations successfully prevent activation of PXR in a cellular environment at the slight expense of the affinity for the primary target.

Crystal structure with DB derivative 4c

While our best compound **4a** was expected to perfectly mimic CQE with its cyclopropylpiperidine in R1, the second best compound, **4c** was predicted to bring a favorable interaction with a buried water molecule due to its differently oriented morpholine moiety in R1. So, we engaged in the determination of the structure of its complex with B-Raf-V600E. Recombinant B-Raf-V600E was successfully over-expressed in *E. coli* after several rounds of optimizations. The purified protein could be co-crystallized with one of our lead compounds **4c**. Crystals grew slowly (2-3 weeks) and diffracted to 2.3 Å resolution. Straightforward molecular replacement and subsequent refinement led to very good refinement statistics (Table S1). The overall structure looks very similar to the previously solved structures. Indeed, as seen in diverse crystallographic B-Raf structures, the newly determined structure presents the typical "back-to-back" interface with a second monomer from a symmetric unit cell. One molecule of inhibitor was detected in each of the two independent monomers (Figures 5A and 5B). The interactions observed between the ligand and the protein recapitulate most of those previously described for two related compounds: DB and CQE (Figure 5C). However, major changes are detected in the region of the morpholine moiety that points toward the solvent while being sandwiched between the Glycine-rich loop and the floor of the ATP-binding site near the 578-KxxN-581 motif and the nearby catalytic aspartate D594 (Figure 5D). Indeed, the nitrogen atom of the morpholine group points toward the aspartate side-chain and seems to interact with it through a water molecule that is also hydrogen-bonded to the nitrogen atom of the thiazole group. Accordingly, this structure matched our prediction and confirmed the favorable interactions that the morpholine group can form. Compared to *tert*-butyl in DB, the morpholine induced an opening of the binding site by pushing outward the glycine-rich loop (Figure 5D).

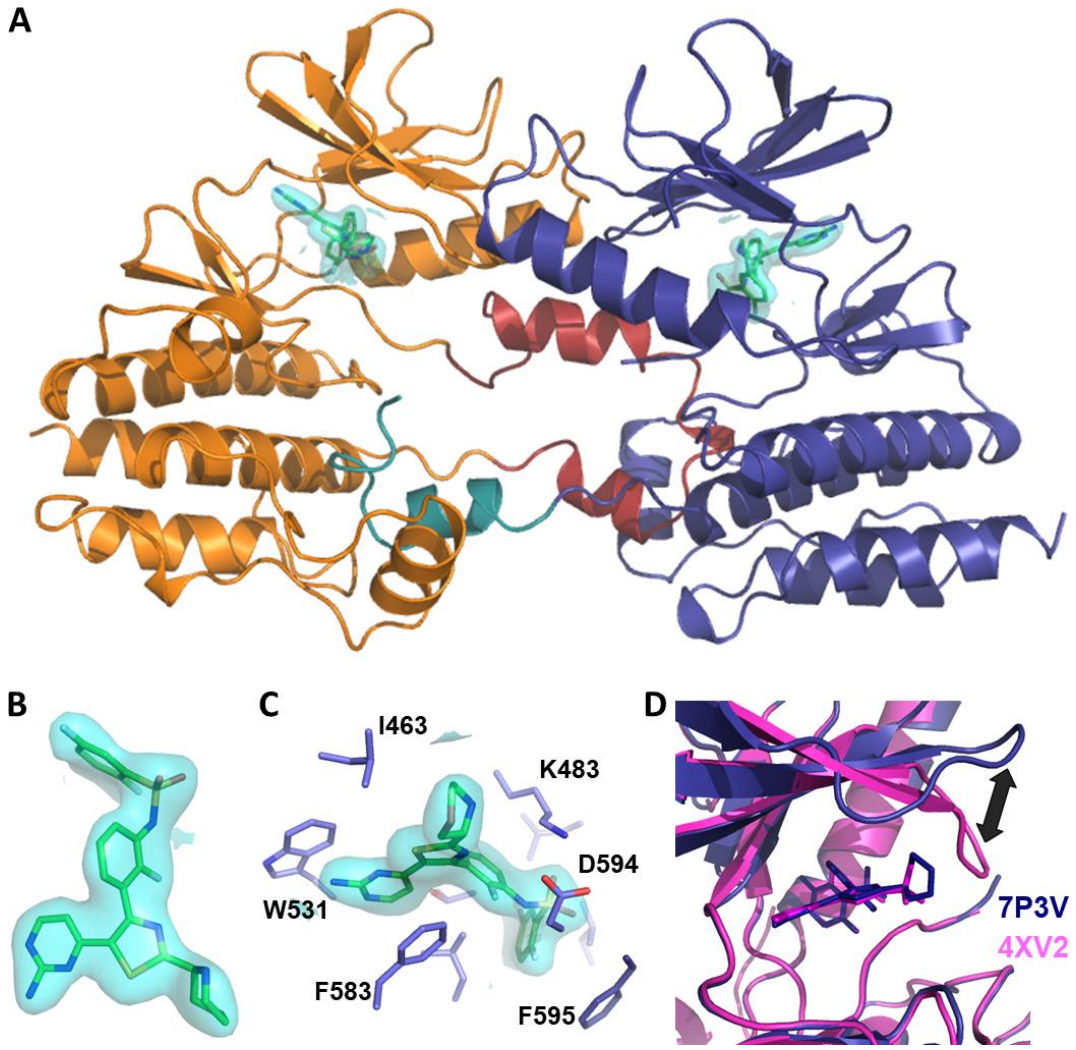


Figure 5. Crystal structure of B-Raf V600E in complex with lead compound **4c**. (A) The refined crystallographic B-Raf V600E structure (chain A in orange but for the swapped segment in firebrick, and chain B in blue but for the swapped segment in cyan) with the designed ligand **4c** (electron density in cyan). The dimeric structure shows a domain swap of the activation loop. Residues in the vicinityof **4c** are shown in line representation. (B) The co-crystallized ligand **4c**. Ligand is shown as stick and its electron density (omit map) is shown in cyan. (C) As in B. but for the presence of the surrounding protein residues in blue sticks. (D) Protein-protein superposition of B-Raf liganded with either DB (PDB ID 4XV2, in magenta) or compound **4c** (this study; PDB ID 7P3V, in blue).

Interestingly, it became apparent during the final steps of the refinement process that a sub-domain swapping involving the region 597-631 has occurred. Indeed, the activation loop (in cyan or firebrick; Figure 5A) now protrudes far into the helical C-terminal domain (C-lobe) of the neighboring protomer. Additionally, the two facing αC helices interact through a new and large area, and with their respective G-rich loop. The latter seems to be stabilized in a

more ‘open’ conformation due to the presence of the morpholine group. However, the ligand is not in direct contact with any residue of the partner protomer.

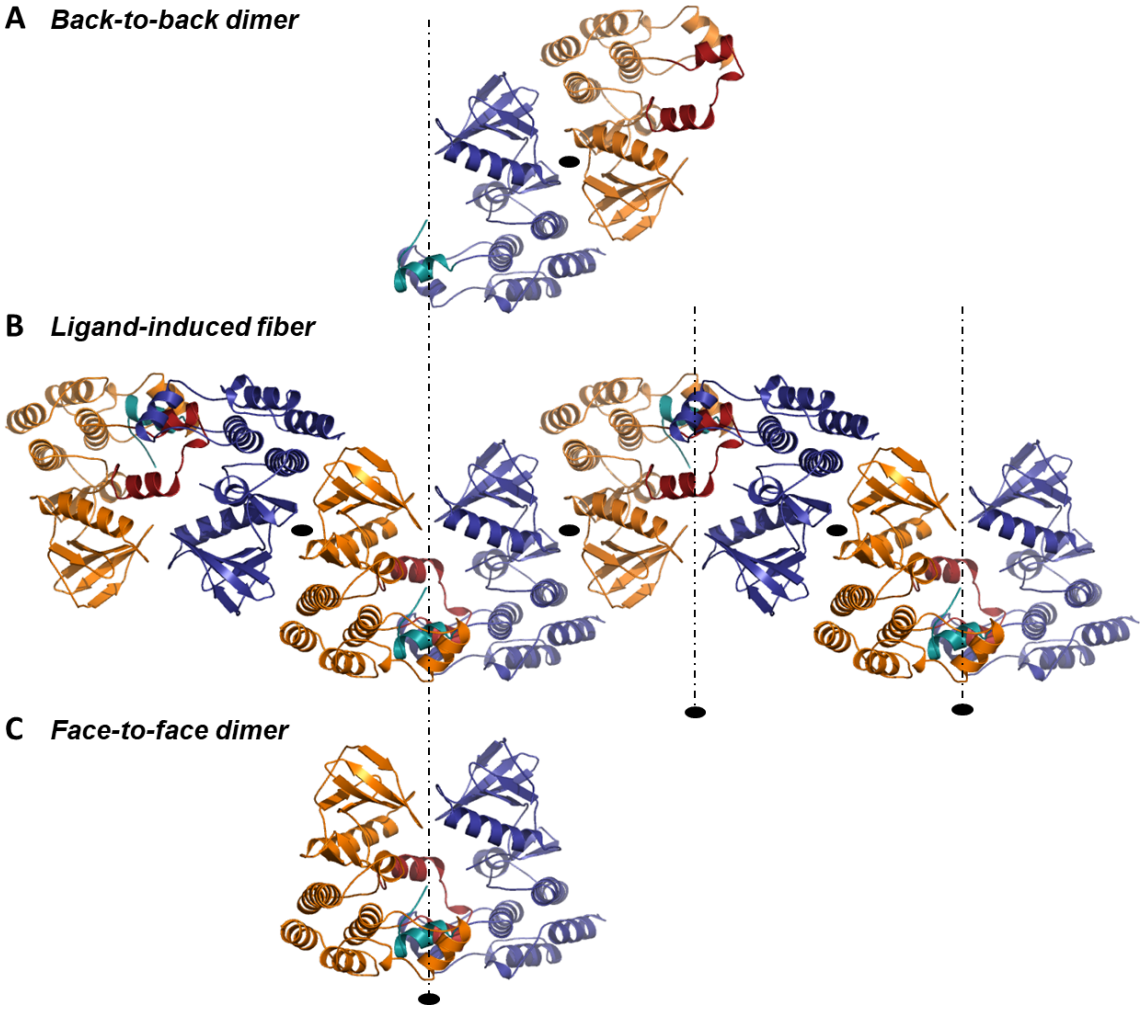


Figure 6. The B-Raf/4c complex forms fiber in crystals. (A) Classical back-to-back B-raf dimer. (B) B-Raf V600E fiber deduced from the two complementary dimer interfaces (back-to-back and face-to-face). (C) Face-to-face dimer of B-Raf V600E induced by the new compound 4c and the resulting domain swapping of the activation loop (the color code is the same as in Figure 5). The 2-fold symmetry axes are depicted by black dashed lines and ovals.

Due to this sub-domain exchange, the new structure showed another dimeric interface that can be described as a "face-to-face" interface (Figures 5A and 6C) in contrast to the "back-to-back" interface of the classical B-Raf dimer (Figure 6A). At this new interface, the two protomers are highly interlaced (Figure 6B). In the B-Raf/**4c** structure, the activation loop

is partially structured in chain A and it adopts a completely well-defined loop in chain B. In this chain, we observed clearly an α-helix starting from W604 and running up to Q612, followed by a one-turn helical structure, and again another α-helix from P622 to Q628 (Figure 5A, in cyan). In chain A, residues 598-613 are not well defined, but the residue stretching from S614 to A621 form a helical turn followed by a clear α-helix from P622 to D629. The latter α-helix is swapped with the facing protomer and located at the exact same position as in a non-swapped conformation (Figure 5A, in firebrick). As a result, a rather large portion of the total solvent-accessible area of the two chains is identified at this interface (15.6% and 15.8%), and the portion of residues involved is as large as 23.9%. In addition, 32 hydrogen bonds and 9 salt bridges are established between the two chains in the crystal structure. Thus, the interface is estimated to be stable in solution. As both dimer interfaces (the classical "back-to-back" (Figure 6A) and the new "face-to-face" (Figure 6C) are large in size and are expected to be stable in solution, a long fiber-like macromolecular structure is expected to form (Figure 6B). This particular feature may have important implication for the functioning of B-Raf upon ligand-induced swapping. A possible impact of the new "face-to-face" domain-swap dimer on the downstream signaling cascade, especially the activation of the protein kinase MEK (being a direct substrate) was investigated by structural superposition of the new dimer structure with the crystallographic complex of B-Raf and MEK (PDB ID 6U2G[27]). Clearly, the structure of the crystallographic B-Raf domain-swap dimer is not compatible with MEK binding, as the MEK binding site overlaps partially with the second B-Raf domain-swap protomer. Of note, re-refinement of the crystal structure of Braf-V600E complex with CQE (PDB ID 4CQE) revealed an identical domain swapping. This original structural property may participate in the lower paradoxical effect of CQE and, therefore, we predict that this feature may be extended to our lead compound although further experimental validation is necessary to fully demonstrate this point.

DISCUSSION AND CONCLUSION

Our previous study[9] demonstrated for the first time that DB is a potent activator of PXR in various cell lines over-expressing this nuclear receptor. We further demonstrated that DB-mediated PXR activation was associated with enhanced expression of several PXR-target genes, which is in accordance with increased expression of CYP2B6 and CYP3A4 previously observed in hepatocytes upon DB treatment[3]. While DB usually inhibits the proliferation of cancer cells by targeting B-Raf-V600E, it was also shown to induce the proliferation of tumor cell lines expressing wild-type B-Raf and an oncogenic mutant of RAS, a small G-protein regulating RAF protein[5].

Though its precise mechanism is not completely understood, PXR activation can, in some cases, result in accelerated cell cycle progression *via* the inhibition of Foxo transcription factors[15], or the increase of *fgf19* expression[13]. Conversely, PXR may also act as a tumor suppressor[12]. Accordingly, tacking into account potential PXR activation, while maintaining the activity against B-Raf, appears to be an important dual-purpose strategy as it would decrease too rapid clearance, unfavorable DDI and possibly also decrease some undesirable or paradoxical effects.

Our current study revealed the binding mode of DB to the PXR LBD thanks to the first crystal structure of this complex. DB is a rather large ligand that makes numerous contacts with the residues of the PXR-LBP and fills a volume similar to that occupied by the reference agonist SR12813. These data not only support the strong PXR agonist property of DB but also provide invaluable information for the development of DB analogues with dissociated activities on B-Raf and PXR.

In this project, we designed novel inhibitors of B-Raf devoid of binding and activation of PXR. Many affinity prediction techniques, from virtual screening to machine learning, were set up to guide the design and synthesis of a small set of novel compounds. These compounds were successfully evaluated on the purified protein-kinase B-Raf as well as on cell lines A375 showing inhibitory activity equivalent to the parent compound. In parallel, it

was also demonstrated that those new compounds differ from DB as they showed little binding as well as weak or no activation of PXR.

The binding mode of the designed drug candidate **4c** in its primary target B-Raf was confirmed by X-ray crystallography. The resulting crystal structure also revealed an original domain swapping never described before for B-Raf protein-kinase while it has been described for several other serine/threonine kinases such as Chk2 and other serine/threonine kinases[28]. Non-classical auto-phosphorylation has been predicted to result from such swapped dimer which brings the activation loop from one monomer to the entrance of the second protomer. In our case, serine S614 would be brought in the vicinity of the ATP binding site and the catalytic center. Interestingly, a phosphorylation on this site has been detected and displays an inhibitory activity[29]. If confirmed this would add another favorable twist to this new type of compounds.

In addition, the swapped dimer being compatible with the standard back-to-back dimerization of B-Raf, the new compound could induce formation of fibers. The size of these interfaces (back-to-back and face-to-face) suggests that they could occur in a cellular context. Interestingly, such a domain swapping masks the interaction surface of B-Raf with MEK and would therefore prevent downstream signaling. As the same sub-domain swapping can be detected into the previously published structure PDB ID 4CQE, we wonder if it could play a role in the lower paradoxical effect noticed for the co-crystallized compound (CQE).

Altogether, our results are of clinical importance as DB dramatically impacts its own metabolism or that of other drugs or medications, in particular MEK inhibitors such as trametinib, which is used in combination with DB for the treatment of metastatic melanoma. Thus, these results strongly argue for assessing the ability of newly developed drugs to activate PXR, in order to limit their detoxification or the potential increase in cancer cell proliferation.

In conclusion, a rational drug design approach was successfully set up to improve a drug already in the clinic but suffering from high PXR activation. Our results suggest also

novel facets to the so-called paradoxical effect in link with both PXR activation and fiber formation due to subdomain swapping in the primary target B-Raf-V600E.

EXPERIMENTAL SECTION

Reagents and chemicals

SR12813 was purchased from Sigma-Aldrich (Saint Quentin Fallavier, France). Dabrafenib was purchased from Euromedex (Souffelweyersheim, France). Drugs were solubilized in DMSO and stored at -20°C.

Cell lines

A375 cells have been ordered to ATCC. The HG5LN GAL4-hPXR cell line was already described[26]. Briefly, Hela cells were stably transfected with the GAL4RE5-Glob-Luc-SVNeo plasmid and the pSG5-GAL4(DBD)-hPXR(LBD)-puro plasmids leading to the HG5LN GAL4-hPXR cell line.

A375 cells were cultured at 37°C under humidified 5% CO2 atmosphere in phenol red (DMEM)-F12 medium (Thermofisher, Villebon sur Yvette, France) supplemented with 1% penicillin/streptomycin, 5% fetal calf serum (FCS). HG5LN GAL4-hPXR were cultured at 37°C under humidified 5% CO2 atmosphere in the same medium supplemented with 1 mg/ml G418 and 0.25 µg/ml puromycin.

PXR transactivation assays

HG5LN GAL4hPXR cells were seeded in 96-well white opaque flat bottom plates at 25,000 cells per well in 100 µL of DMEM-F12 without phenol red supplemented with penicillin/streptomycin (1%) and dextran-coated charcoal-treated fetal calf serum (DCC-FCS) (5%) (test medium). Compounds were added 24h later using automated workstation (Biomek 3000, Beckman Coulter, Villepinte, Paris) and cells were incubated at 37°C for 16h. Then,

test medium was removed and 50 µL of test medium containing luciferin at 0.3 mM were added per well. Luciferase activity was measured for 2s in intact living cells after 10 min stabilization using a MicroBetaWallac luminometer (PerkinElmer). Each compound was tested at various concentrations in three independent experiments at least. For each experiment, tests were performed in quadruplicates for each concentration, and data are expressed by means values with standard deviations. Individual agonist concentration response curves were fitted using the sigmoid concentration response function (log(agonist) *vs*. response-variable slope) of a graphics and statistics software program (GraphPad Prism, version 5.0). EC_{50} (effective concentration for half-maximal luciferase activity) values were calculated from equations used to fit the data in this graphic software. To analyze significances, we compared individual compound treatments with controls using one-way analysis of variance (ANOVA) with the help of GraphPad Prism software.

Cell proliferation assays

A375 cells are provided from ATCC. Cell proliferation was assessed using the standard MTT assay as previously described[30]. Briefly, A375 cells were seeded in 96-well plates (1,000 cells per well) in the same test medium than HG5LN GAL4hPXR cells and treated with increasing concentrations of the different compounds. The medium containing the chemicals was renewed at day 3. At day 5, cells were washed with PBS three times and 0,1 ml of MTT solution (0.5 mg/ml) was added to each well. After incubation (4h), viable cells cleaved the MTT tetrazolium ring into a dark blue formazan reaction product, whereas dead cells remained colorless. The MTT-containing medium was gently removed and DMSO was added to each well. After shaking, the plates were read in absorbance at 540 nm. Results were expressed as percentage of proliferation with respect to the chemical-free control (100%). Data were obtained by concentration response curves plotted as percentage of proliferation versus concentration of the products. Individual concentration response curves

were fitted using the sigmoid concentration response function of a graphics and statistics software program (GraphPad Prism, version 5.0). IC_{50} (half maximal inhibitory concentration) values were calculated from equations used to fit the data in this graphic software. To analyze significances, we compared individual compound treatments with controls using one-way analysis of variance (ANOVA) with the help of GraphPad Prism software.

PXR Crystal structure

The human PXR-LBD (130-434) was co-produced with a fragment of the steroid receptor coactivator-1 (SRC-1, 623-710) to enhance PXR stability, and purified as described previously[31]. Crystals of the hPXR-LBD/dabrafenib complex were obtained by dry co-crystallization[32]. 1 µL of protein (3 mg/mL) was mixed with 1 µl of precipitant (100 mM imidazole pH 7.0, 10% (v/v) isopropanol) on a dry drop of concentrated ligand (0.7 µL at $10^{-}4$ M), and let equilibrated against a reservoir of 500 µL of precipitant. Crystals appear in 24h. Diffraction data were collected on the ID30B beamline at the European Synchrotron Radiation Facilities (λ = 0.9793 Å, 100K), Grenoble, France. Data were processed and scaled with XDS and XSCALE[33]. Crystals belong to space group $P4_32_12$. The structure was solved and refined using Phenix[34] and COOT[35]. The percentage of residues located in the favored Ramachandran plot region is 97.4% (calculated with MolProbity[36]). Data collection and refinement statistics are summarized in Supplementary Table 3. Figures were prepared with PyMOL (http://pymol.org/).

PXR thermal shift assays

To evaluate the binding of compounds (DB, SR12813, **2c**, **2d**, **3a**, **3b**, **4a**, **4c**), the relative melting temperatures (Tm) of PXR-LBD in an empty form (DMSO) or loaded with a ligand were determined by measuring the unfolding of the protein as function of increasing temperature with a Tycho NT.6 (Nanotemper). The changes in the intrinsic fluorescence of the protein, detected at both 350 nm and 330 nm from tryptophan and tyrosine residues, are

measured during the thermal slope, from 35 to 95°C. PXR-LBD complexes were prepared by mixing 3 µl of pure protein at 40 µM (20 µM final) with 3 µl of ligand at 200 µM (100 µM final) in a buffer containing Tris 50 mM pH 8.5, NaCl 150 mM, Glyc 5%, DTT 2 mM, and incubated for 30 min at room temperature before measurements. The Tms were obtained by fitting the fluorescence data (350 nm/330 nm ratio) from 6 independent experiments with a Boltzmann model using the GraphPad Prism software. The ΔTms were obtained by subtracting the Tm of the empty form from the Tm of the bound forms.

Small molecule datasets

In this study, two ligand datasets were used for model training that differ in size, molecular diversity and overall data quality:

• BindingDB BRAF-V600E (2018) with annotated IC_{50} affinity measures - 2193 molecules

• PubChemAssay AID:1257566 with annotated IC_{50} affinity measures (produced by one laboratory and labelled as confirmatory) - 103 molecules.

Note that the BindingDB BRAF-V600E dataset contains also the PubChemAssay molecules.

Substituent variation as thioamides

The different extensions attached to the various scaffolds are purchasable fragments that contain the reaction entity (thioamide) for addition to the scaffolds (from Enamine Chemical Supplier: 179 fragments). The complete molecules were synthesized in silico using KNIME pipelines using RDKit nodes, which additionally removes protecting groups BOC, phthalimide and benzyl carbamate.

Investigation on applicability domain

To ensure that the training datasets are eligible for prediction, we first investigate on the chemical space that is covered by the different molecular datasets. Here, the chemical space s is captured by performing principal component analysis on the PubChem fingerprints of the

two training sets BindingDB BRAF-V600E and PubChemAssay AID: 1257566. In order to ensure that these sets can be trustfully used to predict the new molecules, the chemical series based on our scaffold (also represented in PubChem dataset) are projected onto the PCs of the two training datasets. Additionally, the small PubChemAssay training set was also projected onto the PCs of the larger BindingDB BRAF-V600E dataset. Noteworthy, the molecules are contained in the much larger BindingDB dataset. All those subsets have a very similar location on the PC scatter plots spanned by the BindingDB BRAF-V600E dataset. Furthermore, all molecules based on our scaffold are covered by the PC scatter plot spanned by the PubChemAssay. Additionally, the affinity ranges covered by the two training sets are large and similar for both datasets, although, the distribution of the PubChemAssay is more skewed to higher affinities than the large BindingDB BRAF-V600E dataset.

Machine learning for affinity prediction

For all analysis, calculations and machine learning the R language (version 3.4.4) and RStudio are used. First, to obtain an overview of the data exploratory data analysis is performed. For the training of all machine learning models in this study mainly the R package caret (version 6.0-81) is used. In order to avoid over-fitting of the models 10-fold cross validation repeated 10 times is used for all models. Training of machine learning algorithms in regression mode:

In a first approach structural data from a docking campaign of the BindingDB BRAF-V600E (2018) dataset on our @TOME server[24] is employed. Here, every molecule is docked into 20 different protein structures in parallel using PLANTS[18] and the 20 co-crystallized ligands as molecular shape restraints. Subsequently, the generated complexes are evaluated by different metrics on the server. The whole procedure is repeated for the same set of molecules with different initial 3D conformations and different partial charges. Finally, global ligand-based molecular descriptors (e.g.: molecular weight, number of rotatable bonds, XlogP,..) are added and engineered. This forms the final dataset that is used for the broad

structure-based affinity prediction. The complete procedure is explained in more details in a previous publication[24].

Tailored ligand-based affinity prediction

PubChemAssay AID:1257566 entitled ”Raf/Mek amplified luminescence proximity homogeneous assay” is used to set up the tailored ligand-based prediction models. The tailored approach follows a rather traditional QSAR approach, as it is based on different types of fingerprints (FPs) computed through the rcdk package[36] on the PubChemAssay dataset composed of 103 molecules. The FP types are:

• MACCS - 166 bit MACCS keys described by MDL from RDKit: Open-source cheminformatics (http://www.rdkit.org).

• PubChem - 881 bit FPs defined by PubChem

• extended - hashed FPs, with a default length of 1,024 bits and default search depth of 6, considers paths of a given length and takes rings and atomic properties into account

• graph - hashed FPs, with a default length of 1,024 bits and default search depth of 6, considers connectivity

• shortestpath - hashed FPs, with a default length of 1,024 bits and default search depth of 6, based on the shortest paths between pairs of atoms and takes into account ring systems, charges...

• circular - implementation of the ECFP6 fingerprint, with a length of 1,024 bits and default search depth of 6. Choices for initial bit length and search depth are made based on recommendations provided by ChemAxon (https://docs.chemaxon.com/display/docs/Chemical+Fingerprints). For all six FP types bits with zero variance across the 103 molecules are removed, which leads to reduced FP bit sizes of 51 for MACCS, 141 for PubChem, 238 for extended, 156 for graph, 691 for shortestpath, and 428 for circular FPs. In order to exploit the FP data in an exhaustive way,

two different, widely used algorithms, Support Vector Machine (SVM) with a radial kernel, and Random Forest (RF) are employed on all six FP sets individually.

Investigation on variable importance

During model training optimal parameters were selected by caret's automatic grid search with 10 values per parameter (tuneLength = 10). In the case of RF variable importance is tracked during training as the mean decrease in node impurity. Moreover, quantile Random Forest (qRF) is employed on all six FP types with variable importance tracking. qRF is a generalisation of random forest. It gives a non-parametric way of estimating conditional quantiles for high-dimensional predictor variables. The trained qRF models show lower cross-validation performances than the RF models. Therefore, qRF is only used for variable importance confirmation, not for affinity prediction. Additionally, two further algorithms, Multivariate Adaptive Regression Splines (MARS) and Boruta, an all-relevant feature selection wrapper algorithm, are employed to identify important variables for all six FP types. MARS is a non-parametric regression technique that automatically models nonlinearities and interactions between variables. It can be seen as an extension of linear models. Boruta iteratively compares importance of attributes with importance of shadow attributes, created by shuffling original ones. It does a sharp classification of features rather than ordering. Being an all-relevant method, it aims to find all features connected with the decision and therefore, it also includes redundant features. By default, the ranger package Random Forest implementation is used.

In analogy to the problem-solving principle of "Occam's Razor" we seek a model with the smallest number of descriptors that yield a reasonable model. Therefore, the results of the four algorithms employed for investigating variable importance are compared and the most distinct features, the most highly ranked by the algorithms are extracted for every FP type and subsequently combined as custom 'selection' comprising 40 FPs. On this selection new 'SVM- and 'RF-models' are trained. For comparison, 'SVM- and 'RF-models' are also trained

on the combination of all generated FPs (not including the ones with zero variance). The 'selected' models and 'all' models are compared with the previously trained models on single FP types by performance on cross-validation during training. Remarkably, the 'selected' 'SVM- and 'RF-models' show the best performance, with lowest RMSE and highest squared R values. They are followed by the 'SVM all model', whereas the 'RF all model' is situated much lower, in the midfield among all trained 'SVM- and 'RF-models'. In general, the SVM algorithm shows slightly improved performances over RF on the same FPs.

MM-PBSA calculation

For MM-PBSA calculation the 50-ns MD production trajectories were reduced to 501 frames each, by extracting a frame every 100 ps. The resulting snapshots of the MD simulations were utilized for post-processing free energies by the single-trajectory MM-PBSA method implemented in g_mmpbsa[38]. Different dielectric constants (ε = 2,4,6,8,12,20) were tested for the binding pocket, while the solution dielectric constant was kept constant at ε = 80. Calculations are performed based on a homogeneous medium with a range of dielectric constants for the solute, an ionic strength of 150 mM, an ionic radius of 0.95 Å for positive charged ions and 1.81 Å for negative charged ions, and a solvent probe radius of 1.4 Å. An example configuration file for g_mmpbsa is provided within the supplements of Section 3.7.1. Other parameters influencing the grid dimensions of the calculation, such as 'cfac', 'gridspace' and 'fadd' were varied from suggested defaults (1.5, 0.5, and 10, respectively) showing only marginal variations in the results and therefore not further changed.

Analysis and visualization are performed with Gromacs tools, PyMol, VMD, Chimera, Python scripts, and provided scripts from the g_mmpbsa package.

Chemical synthesis and enzymatic assay

Molecules were synthesized, purified and characterized by HPLC, NMR and MS by AGV discovery (See Supplement Information). High purity (above 95% but for **1a** and **4a** which were obtained at 93% purity) were checked using HPLC using a Waters alliance 2695 HPLC system with autosampler and Waters 2996 diode array detector. A Macherey-Nagel Nucleoshell RP18 plus column (5 µm, 4 mm x 100 mm) was used at 40°C and a flow rate of 1mL/min. Elution solvents were A (H2O 99.9%, H2CO2 0.1%); B (CH3CN 99.9%, H2CO2 0.1%) with a rising gradient of B. Detection was performed on a 210-400nm range.The LanthaScreen kinase activity (ThermoFisher, USA) assay (10 point titration) was used to monitor *in vitro* activity on recombinant protein B-Raf-V600E (5 nM) and B-Raf wild-type (5 nM) for all the chemical compounds.

Expression and purification of B-Raf

A Pet28a(+) vector with DNA encoding the B-Raf kinase domain residues 448-723 containing the V600E mutation, 16 solubilizing mutations (I543A, I544S, I551K, Q562R, L588N, K630S, F667E, Y673S, A688R, L706S, Q709R, S713E, L716E, S720E, P722S, and K723G - permitting kinase domain over-expression in bacteria), encoding as well an N-terminal His tag, and a thrombin cleavage site between the protein and the His tag, was provided by Dr. Michael Grasso (Marmorstein Lab, Department of Chemistry, University of Pennsylvania, Philadelphia, USA). The protein was expressed in *E. coli* BL21(DE3) cells, with an overnight pre-culture at 37°C in LB, followed by 6h at 37 and an overnight incubation at 25°C on an auto-inductive kanamycin medium, spun down the next medium, spun down the next day, lysed in lysis buffer (buffer A supplemented with lysozyme), frozen, thawed, and sonicated. The lysate was then spun down at 18,000 rpm, and the supernatant was incubated on a His-trap nickel column at 4°C for 1h. The supernatant was then eluted, the column washed with buffer A, and the B-Raf proteins eluted with buffer B (wich is buffer A supplemented with 300 mM imidazole). Protein was then dialyzed into buffer C (wich does not contain imidazole)

and applied to a 16/60 Superdex 75 gel filtration column in a final buffer D. Protein was frozen and stored for future use.

Used buffers are:

• Buffer A (50 mM Tris, pH 7.0, 250 mM NaCl, 5% glycerol, 2 mM β-Mercaptoethanol),

• Buffer B (Buffer A and 300 mM imidazole),

• Buffer C (25 mM Tris, pH 7.0, 75 mM NaCl, 5% glycerol, 1 mM EDTA, 10 mM DTT),

• Buffer D (20 mM HEPES at pH 7.0, 150 mM NaCl, 5% glycerol, 10 mM DTT).

B-Raf crystallization and data collection

B-RafV600E-16M at 9 mg/mL was mixed with 10% of a 10 mM DMSO inhibitor solution, and after initial screenings using the commercial kits "PACT", "PEGs-I", and "PEGs-II" from QIAGEN on 96-Well plates, trays were set up screening around a crystallization condition of 100 mM BisTrisPropane at pH 7.0-8.0, 20% PEG monomethyl ether 2000/3350/4000, and 100-350 mM Na-formate using the hanging-drop vapor diffusion method at 18°C. (a) Crystallization condition of 100 mM BisTrisPropane at pH 8.0, 20% PEG 3350, and 250 mM Na-formate. (b) Crystallization condition of 100 mM BisTrisPropane at pH 8.0, 20% PEG 3350, and 300 mM Na-formate. Crystals were flash frozen in liquid nitrogen. X-ray diffraction data was collected at a wavelength of 0.979 Å and a beam size fitted to the crystal dimensions at the synchrotron ALBA (Barcelona, Spain), at beamline BL13-XALOC.

Crystallographic structure determination

The structure was determined by molecular replacement in PHENIX[34] using Phaser using PDB ID 5ITA as a search model. The molecular replacement search model was used as monomer and had its ligand removed. Model building and refinement were performed using Coot[35] and PHENIX[34]. NCS was used, as two B-Raf monomers were present in the asymmetric unit. The CIF file for the inhibitor was generated using the Grade Web Server (at http://grade.globalphasing.org). The atomic dimeric structure is refined to a final resolution of

2.37 Å without very good refinement statistic after modelling of a sub-domain swap. Indeed, after several refinement steps, there were several discrepancies apparent between the model and the observed data (as shown by the Fobs - Fcalc difference map in Figure 3.41a) and the 2Fobs - Fcalc density map indicated a continuation of the structure from one protomer to the neighboring one. After modelling the domain swap these discrepancies vanished.

ASSOCIATED CONTENT

Supporting Information

Table with the data collection and refinement statistics; depiction of the whole structure of the DB-bound PXR LBD; curves of the cell assays for PXR activation; curves of the cell assays for antiproliferative B-Raf activity; 1H-NMR spectrum, the HPLC elution profile, the mass spectrometry and the UV-CD spectra of compounds **1a**, **2a**, **2b**, **2c**, **2d**, **3a**, **3b**, **4a**, **4b**, **4c**, **4d**, **5a** and **5b**.

Molecular formula strings (CSV).

Accession Codes

PXR-DB PDB ID 6HJ2, and B-Raf-V600E-**4c** PDB ID 7P3V. Authors will release the atomic coordinates upon article publication.

AUTHOR INFORMATION

Corresponding Authors

Martin Cohen Gonsaud - Centre de Biologie Structurale (CBS), CNRS, INSERM, Univ Montpellier, Montpellier, France ; Phone +33 (0)4 67 41 79 06 ; martin.cohen-gonsaud@cbs.cnrs.fr

William Bourguet - Centre de Biologie Structurale (CBS), INSERM, CNRS, Univ Montpellier, Montpellier, France ; Phone +33 (0)4 67 41 77 12 ; bourguet@cbs.cnrs.fr

Gilles Labesse - Centre de Biologie Structurale (CBS), CNRS, INSERM, Univ Montpellier, Montpellier, France ; Phone +33 (0)4 67 41 77 12 ; labesse@cbs.cnrs.fr

Patrick Balaguer - Institut de Recherche en Cancérologie de Montpellier (IRCM), INSERM, Univ Montpellier, Montpellier, France; Phone +33 (0)4 67 61 24 09 ; patrick.balaguer@inserm.fr

Author Contributions

Participated in research design: G.L., M.C.G., P.B. and W.B. Conducted experiments: M.S., V.D., M.G., M.G., M.G., L.H., P.B. and G.L. Performed data analysis: M.S., V.D., M.G., M.C.G., P.B. and G.L. Wrote the manuscript: M.S., V. D., M.G., P.B., W.B. and G.L. Revised the manuscript: V. D., M.G., P.B. and G.L.

Notes

The authors declare no conflict of interest but a patent application on the use of the corresponding compounds in cancer therapy (WO 2021/110997).

FUNDING SOURCES

This work was supported by grants from the Cancéropôle Grand Sud-Ouest, project GSO-BRAFvsPXR (A.E., W.B., P.P., and P.B.), a Grant Emergence from the Cancéropôle Grand Sud-Ouest (GL), the FRM [FDT201904008322] (MS). This work was supported by the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INBS-05.

ACKNOWLEGMENTS

We acknowledge the experimental assistance from the staff of the European Synchrotron Radiation Facility (ESRF, Grenoble, France) and the synchrotron ALBA (Barcelona, Spain) during data collection. The views expressed in the article do not necessarily reflect the views of the funding agencies.

ABBREVIATIONS

B-Raf, wild-type protein-kinase B-Raf; B-Raf-V660E, mutant V600E protein-kinase B-Raf; DB, Dabrafenib; DDI, Drug-Drug Interaction; ECF, Encorafenib; Fps, fingerprints; LBD, Ligand Binding Domain; LBP, Ligand Binding Pocket; PXR, Pregnane X Receptor; RF, Random Forest; SVM, support vector machine

REFERENCES

(1) Roskoski R Jr. Targeting oncogenic Raf protein-serine/threonine kinases in human cancers. *Pharmacol Res*. **2018,** *135*, 239-258. doi: 10.1016/j.phrs.2018.08.013.

(2) Holderfield M, Deuker MM, McCormick F, McMahon M. Targeting RAF kinases for cancer therapy: BRAF-mutated melanoma and beyond. *Nat Rev Cancer*. **2014**; *14*, 455–467.

(3) Puszkiel A, Noé G, Bellesoeur A, Kramkimel N, Paludetto MN, Thomas-Schoemann A, Vidal M, Goldwasser F, Chatelut E, Blanchet B. Clinical pharmacokinetics and pharmacodynamics of dabrafenib. *Clin Pharmacokinet*. **2019**, *58*, 451-467. doi:10.1007/s40262-018-0703-0

(4) Dhillon, S. Dabrafenib plus Trametinib: A review in advanced melanoma with a BRAF (V600) mutation. *Target. Oncol.* **2016**, *11*, 417–428

(5) Zhang C, Spevak W, Zhang Y, Burton EA, Ma Y, Habets G, Zhang J, Lin J, Ewing T, Matusow B, Tsang G, Marimuthu A, Cho H, Wu G, Wang W, Fong D, Nguyen H, Shi S, Womack P, Nespi M, Shellooe R, Carias H, Powell B, Light E, Sanftner L, Walters J, Tsai J, West BL, Visor G, Rezaei H, Lin PS, Nolop K, Ibrahim PN, Hirth P, Bollag G. Raf inhibitors that evade paradoxical MAPK activation. *Nature*. **2015**, *526*, 583-586. doi: 10.1038/nature14982

(6) Anforth RM, Blumetti TC, Kefford RF, Sharma R, Scolyer RA, Kossard S, Long GV, Fernandez-Peñas P. Cutaneous manifestations of dabrafenib (GSK2118436): a selective inhibitor of mutant BRAF in patients with metastatic melanoma. *Br J Dermatol.* **2012**, *167*, 1153-1160. doi: 10.1111/j.1365-2133.2012.11155.x.

(7) Rheault TR, Stellwagen JC, Adjabeng GM, Hornberger KR, Petrov KG, Waterson AG, Dickerson SH, Mook RA Jr, Laquerre SG, King AJ, Rossanese OW, Arnone MR, Smitheman KN, Kane-Carson LS, Han C, Moorthy GS, Moss KG, Uehling DE. Discovery of Dabrafenib: A selective inhibitor of Raf kinases with antitumor activity against B-Raf-driven tumors. *ACS Med Chem Lett*. **2013**, *4*, 358-362. doi: 10.1021/ml4000063. PMID: 24900673; PMCID: PMC4027516.

(8) Poulikakos PI, Zhang C, Bollag G, Shokat KM, Rosen N. RAF inhibitors transactivate RAF dimers and ERK signalling in cells with wild-type BRAF. *Nature*. **2010**, *464*, 427-430. doi: 10.1038/nature08902.

(9) Creusot N, Gassiot M, Alaterre E, Chiavarina B, Grimaldi M, Boulahtouf A, Toporova L, Gerbal-Chaloin S, Daujat-Chavanieu M, Matheux A, Rahmani R, Gongora C, Evrard A, Pourquier P, Balaguer P. The anti-cancer drug Dabrafenib is a potent activator of the human pregnane X receptor. *Cells*. **2020**, *9*, 1641. doi:10.3390/cells9071641

(10) Chai SC, Wright WC, Chen T. Strategies for developing pregnane X receptor antagonists: Implications from metabolism to cancer. *Med Res Rev.* **2020**, *40*, 1061-1083. doi: 10.1002/med.21648.

(11) Hall A, Chanteux H, Ménochet K, Ledecq M, Schulze MED. Designing out PXR activity on drug discovery projects: a review of structure-based methods, Empirical and Computational Approaches. *J Med Chem*. **2021**, *64*, 6413-6522. doi: 10.1021/acs.jmedchem.0c02245. PMID: 34003642.

(12) Oladimeji PO, Chen T. PXR: more than just a master xenobiotic receptor. *Mol Pharmacol.* **2018**, *93*, 119-127. doi: 10.1124/mol.117.110155. PMID: 29113993; PMCID: PMC5767680.

(13) Wang H, Venkatesh M, Li H, Goetz R, Mukherjee S, Biswas A, Zhu L, Kaubisch A, Wang L, Pullman J, Whitney K, Kuro-o M, Roig AI, Shay JW, Mohammadi M, Mani S. Pregnane X receptor activation induces FGF19-dependent tumor aggressiveness in humans and mice. *J Clin Invest.* **2011**, *121*, 3220-3232. doi: 10.1172/JCI41514

(14) Qiao E, Ji M, Wu J, Ma R, Zhang X, He Y, Zha Q, Song X, Zhu LW, Tang J. Expression of the PXR gene in various types of cancer and drug resistance. *Oncol Lett.* **2013**, *5*, 1093-1100. doi: 10.3892/ol.2013.1149. PMID: 23599746; PMCID: PMC3628904.

(15) Shizu R, Abe T, Benoki S, Takahashi M, Kodama S, Miayata M, Matsuzawa A, Yoshinari K. PXR stimulates growth factor-mediated hepatocyte proliferation by cross-talk with the FOXO transcription factor. *Biochem J.* **2016**, *473*, 257-266. doi: 10.1042/BJ20150734

(16) Delfosse V, Huet T, Harrus D, Granell M, Bourguet M, Gardia-Parège C, Chiavarina B, Grimaldi M, Le Mével S, Blanc P, Huang D, Gruszczyk J, Demeneix B, Cianférani S, Fini JB, Balaguer P, Bourguet W. Mechanistic insights into the synergistic activation of the RXR-PXR heterodimer by endocrine disruptor mixtures. *Proc Natl Acad Sci U S A*. **2021**, *118*, e2020551118. doi: 10.1073/pnas.2020551118.

(17) Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res*. **2021**, *49*, D1388-D1395. doi: 10.1093/nar/gkaa971.

(18) Korb, O.; Stützle, T.; Exner, T. E. "PLANTS: application of ant colony optimization to structure-based drug design" (**2006**) *Lecture Notes in Computer Science* 4150, 247-258.

(19) Adelmann CH, Ching G, Du L, Saporito RC, Bansal V, Pence LJ, Liang R, Lee W, Tsai KY. Comparative profiles of BRAF inhibitors: the paradox index as a predictor of clinical

toxicity. *Oncotarget*. **2016**, *7*, 30453-30460. doi: 10.18632/oncotarget.8351. PMID: 27028853; PMCID: PMC5058692.

(20) Pulici M, Traquandi G, Marchionni C, Modugno M, Lupi R, Amboldi N, Casale E, Colombo N, Corti L, Fasolini M, Gasparri F, Pastori W, Scolaro A, Donati D, Felder E, Galvani A, Isacchi A, Pesenti E, Ciomei M. Optimization of diarylthiazole B-raf inhibitors: identification of a compound endowed with high oral antitumor activity, mitigated hERG inhibition, and low paradoxical effect. *ChemMedChem*. **2015**, *10*, :276-295. doi:10.1002/cmdc.201402424

(21) Waizenegger IC, Baum A, Steurer S, Stadtmüller H, Bader G, Schaaf O, Garin-Chesa P, Schlattl A, Schweifer N, Haslinger C, Colbatzky F, Mousa S, Kalkuhl A, Kraut N, Adolf GR. A novel RAF Kinase inhibitor with DFG-Out-binding mode: high efficacy in BRAF-mutant tumor xenograft models in the Absence of Normal Tissue Hyperproliferation. Mol Cancer Ther. **2016**, *15*, 354-365. doi: 10.1158/1535-7163.

(22) Schneider, M. & Labesse, G. MM-PBSA and the importance of the dielectric constant for kinase drug design. **2020** ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.13347368.v1

(23) Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* **2016**, *44*, D1045-D1053. doi: 10.1093/nar/gkv1072.

(24) Schneider M, Pons JL, Bourguet W, Labesse G. Towards accurate high-throughput ligand affinity prediction by exploiting structural ensembles, docking metrics and ligand similarity. *Bioinformatics*. **2020**, *36*, 160-168. doi: 10.1093/bioinformatics/btz538. PMID: 31350558; PMCID: PMC6956784.

(25) Beisken S, Meinl T, Wiswedel B, de Figueiredo LF, Berthold M, Steinbeck C. KNIME-CDK: Workflow-driven cheminformatics. *BMC Bioinformatics.* **2013**, 257. doi: 10.1186/1471-2105-14-257. PMID: 24103053; PMCID: PMC3765822.

(26) Lemaire G, Mnif W, Pascussi JM, Pillon A, Rabenoelina F, Fenet H, Gomez E, Casellas C, Nicolas JC, Cavaillès V, Duchesne MJ, Balaguer P. Identification of new human pregnane X receptor ligands among pesticides using a stable reporter cell system. *Toxicol Sci*. **2006**, *91*, 501-509. doi: 10.1093/toxsci/kfj173.

(27) Haling JR, Sudhamsu J, Yen I, Sideris S, Sandoval W, Phung W, Bravo BJ, Giannetti AM, Peck A, Masselot A, Morales T, Smith D, Brandhuber BJ, Hymowitz SG, Malek S. Structure of the BRAF-MEK complex reveals a kinase activity independent role for BRAF in MAPK signaling. *Cancer Cell.* **2014**, *26*, 402-413. doi: 10.1016/j.ccr.2014.07.007. PMID: 25155755.

(28) Oliver AW, Knapp S, Pearl LH. Activation segment exchange: a common mechanism of kinase autophosphorylation? *Trends Biochem Sci*. **2007**, *32*, 351-356. doi: 10.1016/j.tibs.2007.06.004.

(29) Dernayka L, Rauch N, Jarboui MA, Zebisch A, Texier Y, Horn N, Romano D, Gloeckner CJ, Kriegsheim Av, Ueffing M, Kolch W, Boldt K. Autophosphorylation on S614 inhibits the activity and the transforming potential of BRAF. *Cell Signa*l. **2016**, *28*, 1432-1439. doi: 10.1016/j.cellsig.2016.06.016.

(30) Delfosse V, Grimaldi M, Pons JL, Boulahtouf A, le Maire A, Cavailles V, Labesse G, Bourguet W, Balaguer P. Structural and mechanistic insights into bisphenols action provide guidelines for risk assessment and discovery of bisphenol A substitutes. *Proc Natl Acad Sci U S A*. **2012**, *109*, 14930-14935. doi: 10.1073/pnas.1203574109. PMID: 22927406; PMCID: PMC3443136.

(31) Delfosse V, Dendele B, Huet T, Grimaldi M, Boulahtouf A, Gerbal-Chaloin S, Beucher B, Roecklin D, Muller C, Rahmani R, Cavaillès V, Daujat-Chavanieu M, Vivat V, Pascussi JM, Balaguer P, Bourguet W. Synergistic activation of human pregnane X receptor by binary cocktails of pharmaceutical and environmental compounds. *Nat Commun.* **2015**, *6*, 8089-8095. doi: 10.1038/ncomms9089.

(32) Gelin M, Delfosse V, Allemand F, Hoh F, Sallaz-Damaz Y, Pirocchi M, Bourguet W, Ferrer JL, Labesse G, Guichou JF. Combining 'dry' co-crystallization and *in situ* diffraction to facilitate ligand screening by X-ray crystallography. *Acta Crystallogr D Biol Crystallogr.* **2015**, *71*, 1777-1787. doi: 10.1107/S1399004715010342.

(33) Kabsch, W. XDS. *Acta Crystallogr., D: Biol. Crystallogr.* **2010**, *66*, 125-132, DOI: 10.1107/S0907444909047337

(34) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: A comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 213-221, DOI: 10.1107/S0907444909052925

(35) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486-501, DOI: 10.1107/s0907444910007493

(36) Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 12– 21,

(37) Guha, R. Chemical informatics functionality in R. *J Stat Softw.* **2007**, 18, 1–16.

(38) Kumari, R., Kumar, R.; Open Source Drug Discovery Consortium, Lynn A. g_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model*. **2014**, *54*, 1951-1962. doi: 10.1021/ci500020m. PMID: 24850022.

Supporting Information

Structure-based and knowledge-informed design of B-Raf inhibitors devoid of deleterious PXR binding

Melanie Schneider[1], Vanessa Delfosse[1], Muriel Gelin[1], Marina Grimaldi[2], Meritxell Granell[1], Laurène Heriaud[1], Jean-Luc Pons[1], Martin Cohen Gonsaud[1]*, Patrick Balaguer[2]*, William Bourguet[1]*, Gilles Labesse[1]*

[1]Centre de Biologie Structurale (CBS), CNRS, Inserm, Univ Montpellier, F-34090 Montpellier, France
[2]Institut de Recherche en Cancérologie de Montpellier (IRCM), Inserm, Univ Montpellier, F-34090 Montpellier, France

*Corresponding Authors: martin.cohen-gonsaud@cbs.cnrs.fr, patrick.balaguer@inserm.fr, william.bourguet@cbs.cnrs.fr, gilles.labesse@cbs.cnrs.fr

Table of contents

Table S1. Data collection and refinement statistics. p.S2

Table S2. SMILES strings of molecules used in this work**.** p.S3

Figure S1. Whole structure of the DB-bound PXR LBD. p.S4

Figure S2. Cell assays for PXR activation. p.S5

Figure S3. Cell assays for antiproliferative B-Raf activity. p.S6

Figure S4. Molecules characterizations (1D-NMR, LC-MS, UV). p.S7

Table S1. Data collection and refinement statistics.

	hPXR/Dabrafenib (6HJ2)	**B-Raf/4c (7P3V)**
Data Collection		
Space group	$P\,4_3\,2_1\,2$	$P\,2_1\,2_1\,2_1$
Cell dimensions		
a, *b*, *c* (Å)	91.39, 91.39, 85.86	52.63, 104.85, 109.56
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	45.69 – 2.28 (2.36 – 2.28)	47.44 – 2.37 (2.46 – 2.37)
R_{pim}	0.066 (0.532)	0.054 (0.909)
$CC_{1/2}$	0.999 (0.947)	0.998 (0.370)
$I / \sigma I$	22.6 (4.4)	9.8 (0.7)
Completeness (%)	99.9 (99.9)	99.7 (97.9)
Redundancy	12.5 (12.7)	6.5 (5.7)
Refinement		
Resolution (Å)	45.69 – 2.28	47.44 – 2.37
No. reflections	17,132	47,443
R_{work} / R_{free}	0.189 / 0.232	0.195 / 0.247
No. atoms		
Protein	2,157	4,191
Ligand/ion	39	112
Water	78	125
B-factors		
Protein	48.98	70.03
Ligand/ion	55.99	54.40
Water	51.43	63.14
R.m.s. deviations		
Bond lengths (Å)	0.009	0.008
Bond angles (°)	0.973	0.990

*Values in parentheses are for highest-resolution shell.

Table S2. SMILES strings of molecules used in this work.

The molecular formula strings and the associated biochemical and biological data are provided as a separate CSV file.

Figure S1. Whole structure of the DB-bound PXR-LBD. Secondary structure elements are annotated, in particular the H12 C-terminal helix (in deep cyan). Dabrafenib (DB) is represented in green sticks. Oxygen, nitrogen, sulfur, and fluorine atoms are colored in red, blue, yellow, and light blue, respectively.

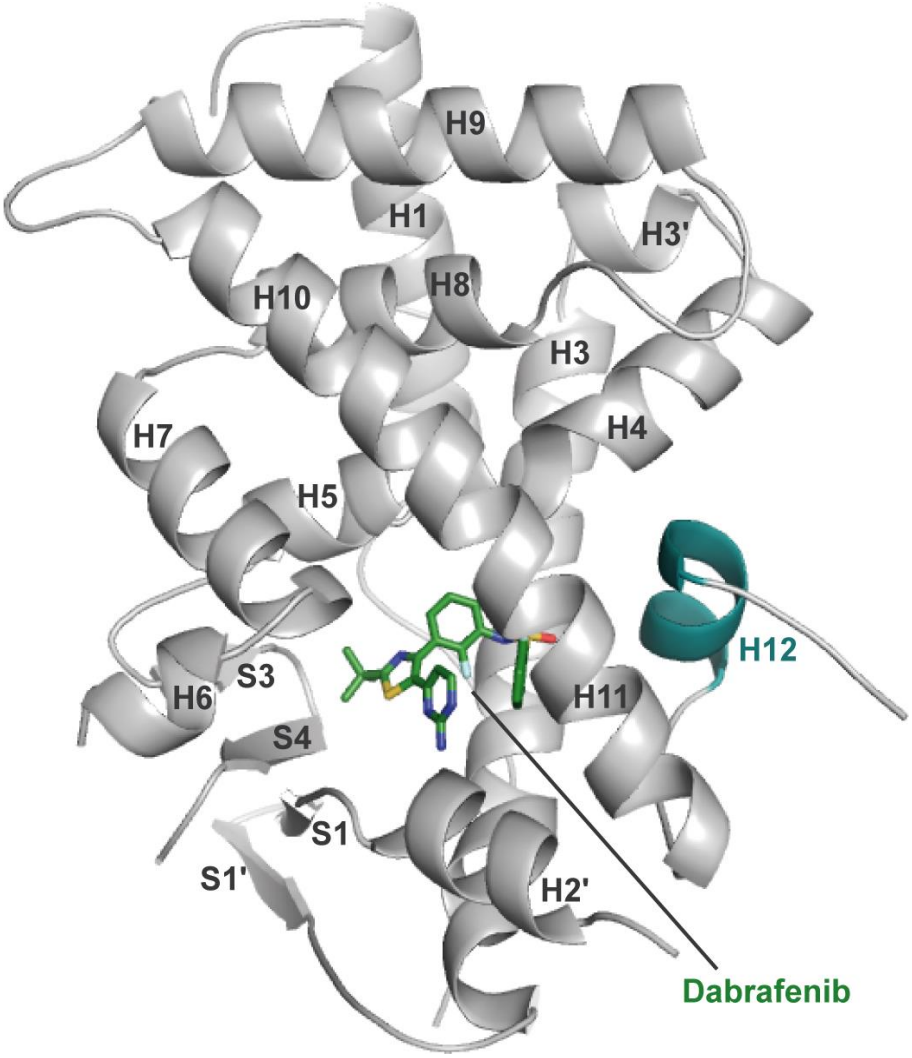


Figure S2. Cell assays for PXR activation. Results of luciferase assays showing dose-response curves for encorafenib, dabrafenib (DB) and its analogues in HG5LN GAL4-hPXR cells. Results are expressed as percentage of the maximal response obtained in presence of 3 µM SR12813. Data are the mean (SD) of three to five independent experiments. The x-axis represents the concentration of ligand in molar.

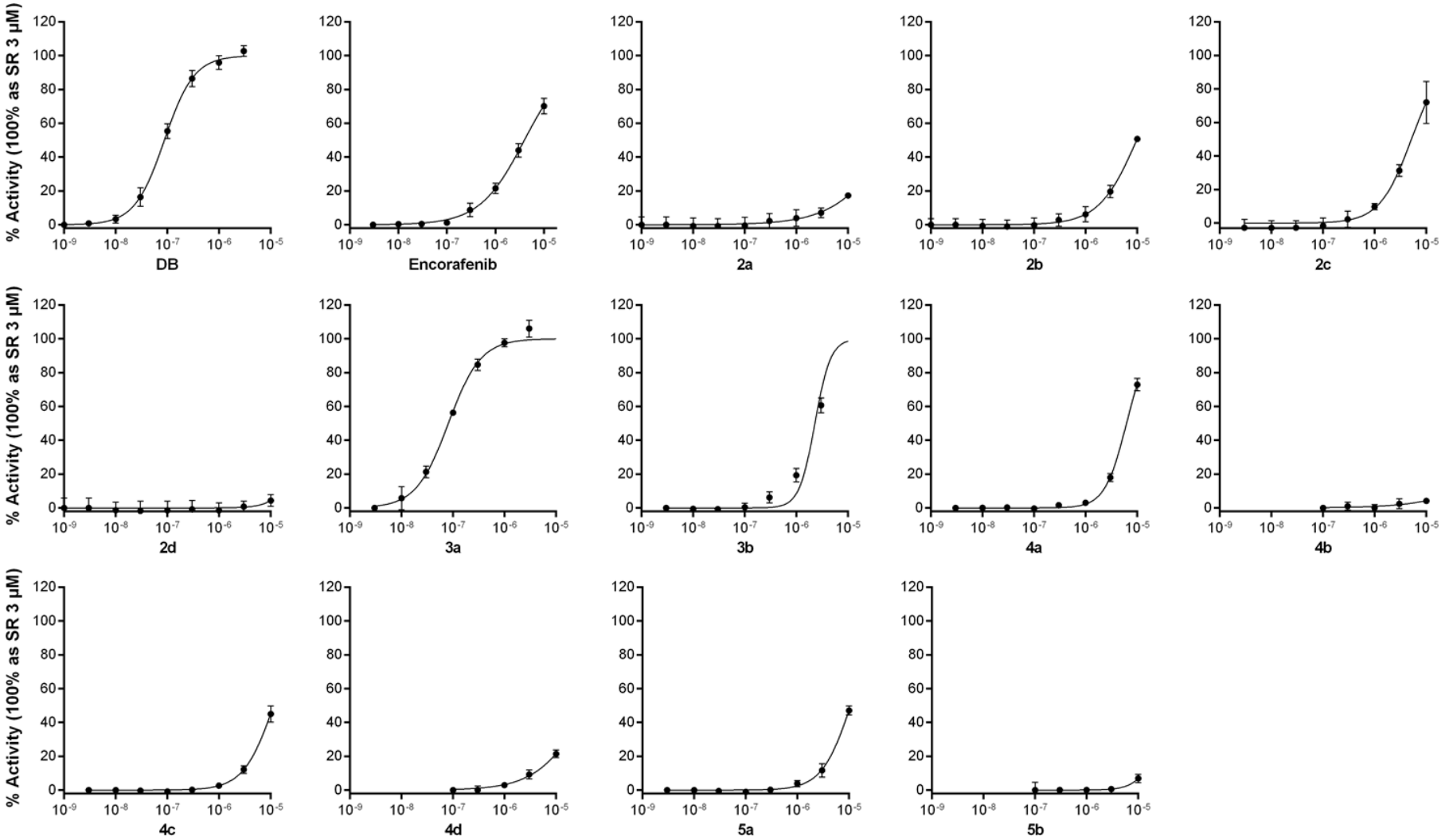


Figure S3. Cell assays for antiproliferative B-Raf activity. Results of antiproliferative assays showing dose-response curves for encorafenib, dabrafenib (DB) and its analogues in A375 melanoma cancer cells. A375 cells were treated with chemicals for 5 days and cell survival was measured using MTT assay. Results are expressed as percentage of the maximal response obtained with untreated cells. Data are the mean (SD) of three to five independent experiments. The x-axis represents the concentration of ligand in molar.

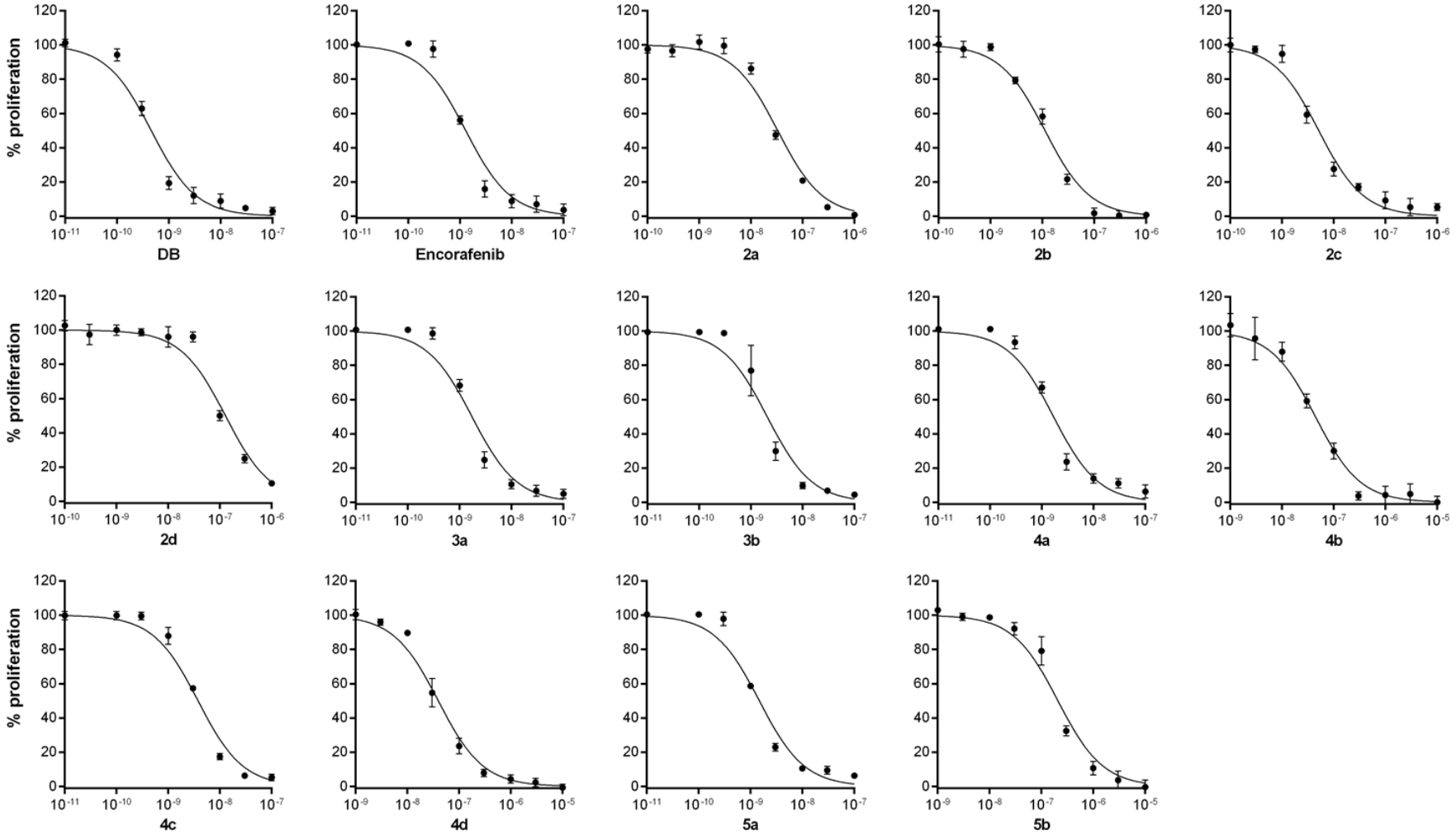
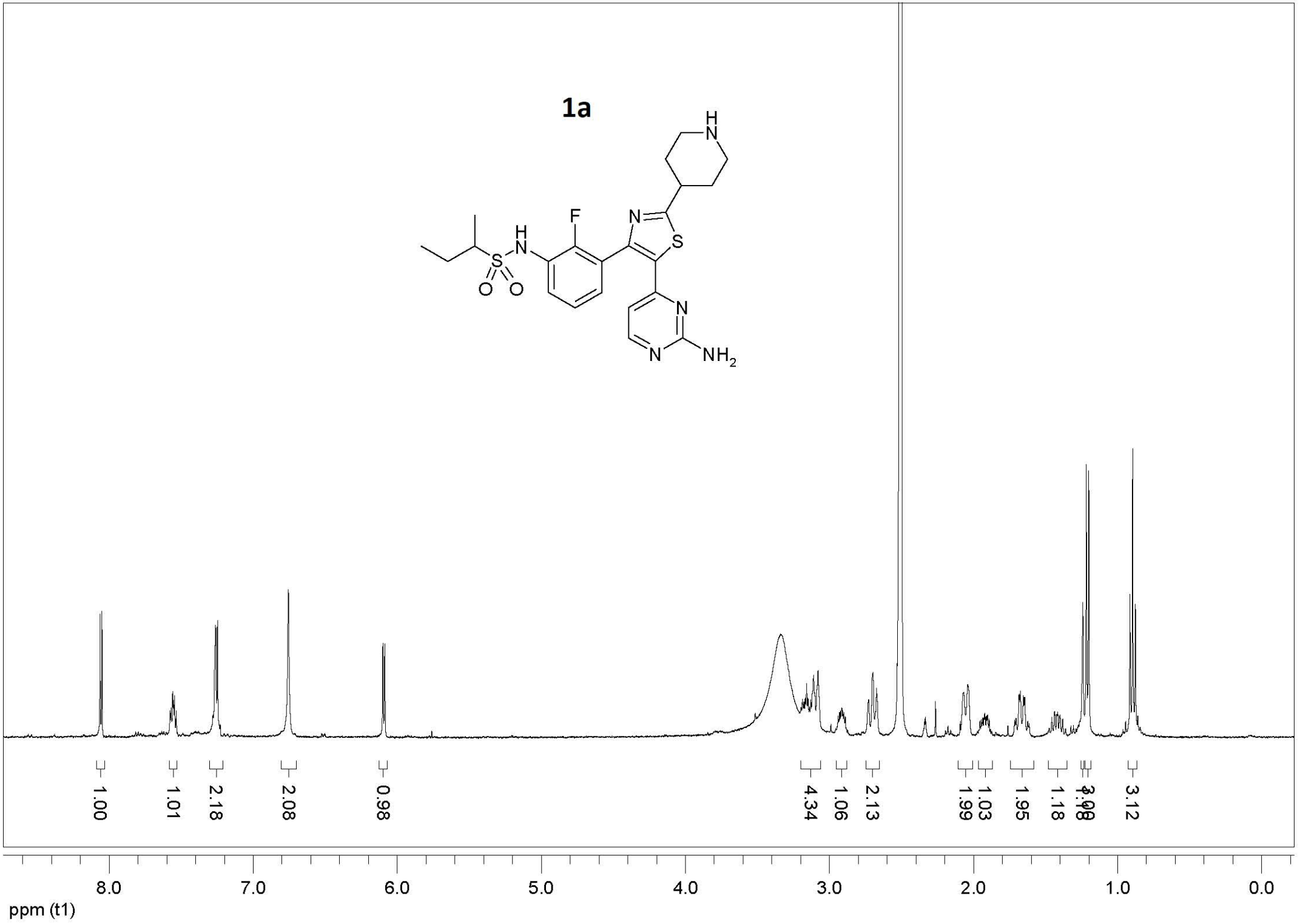
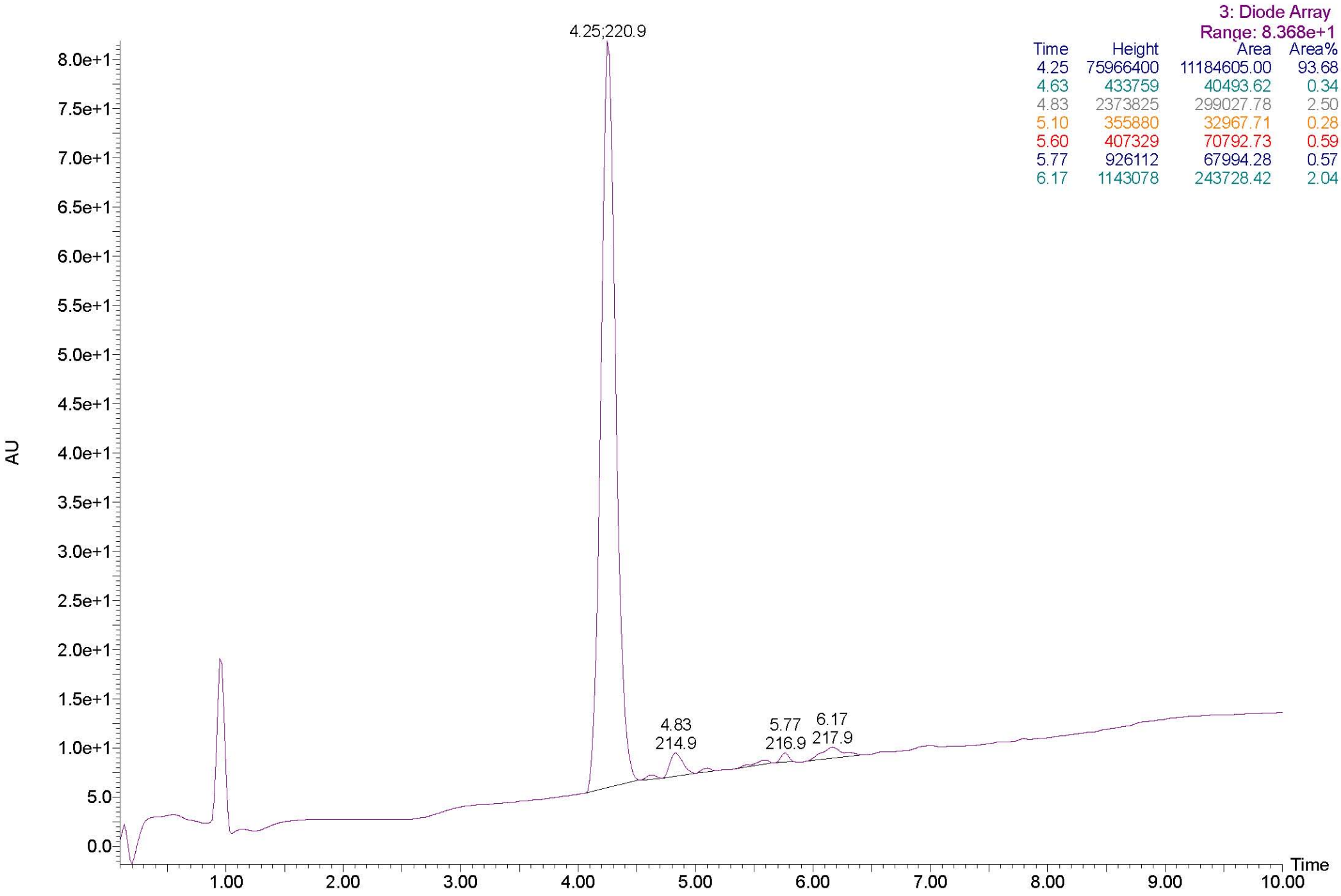


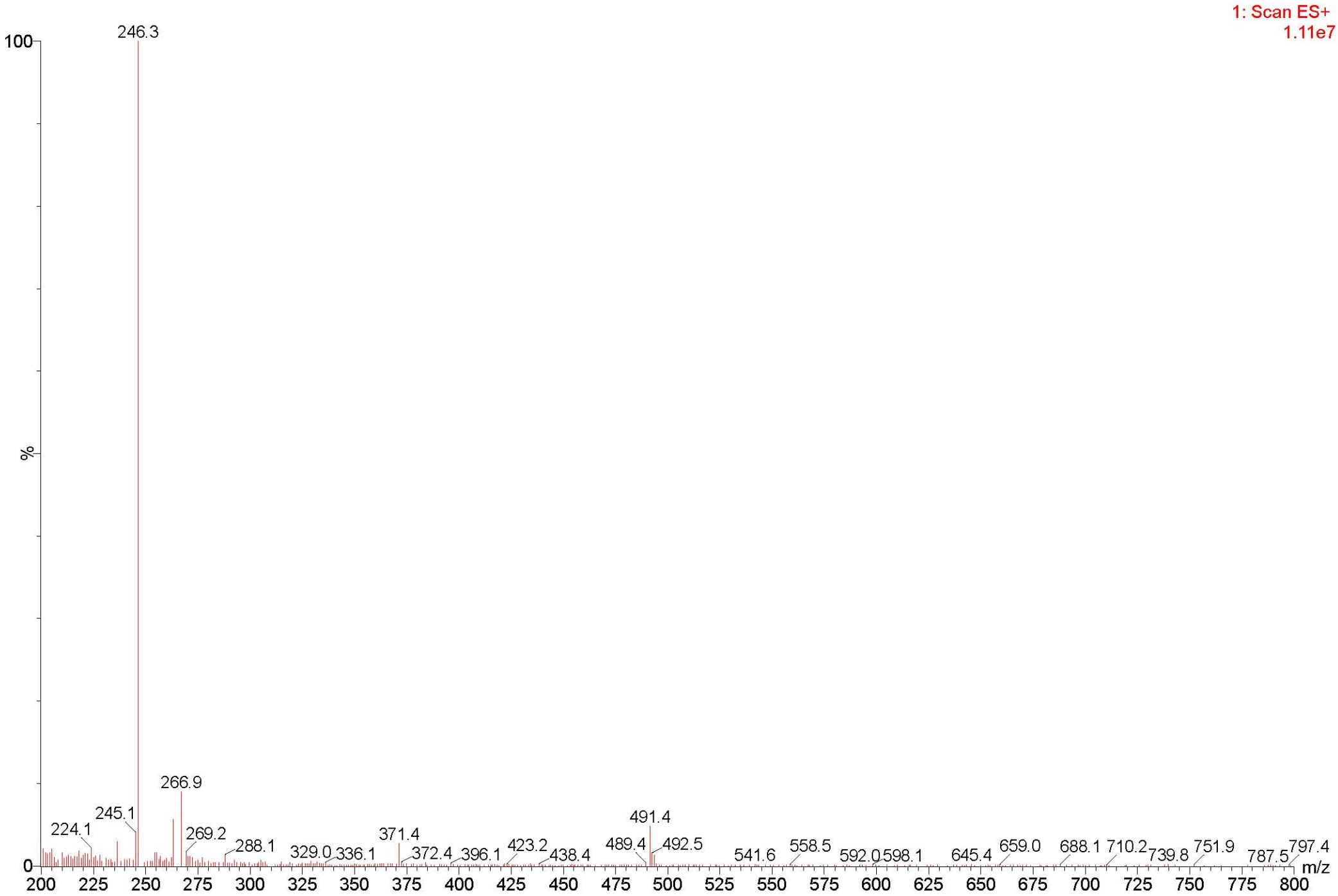
Figure S4. Molecules characterizations. Hereafter, for each compound (**1a**, **2a**, **2b**, **2c**, **2d**, **3a**, **3b**, **4a**, **4b**, **4c**, **4d**, **5a** and **5b**) we provide the 1H-NMR spectrum, the HPLC elution profile, the mass spectrometry and the UV-CD spectra.


1a
HN
F
N
S
H
N
S
O
O
N
N
NH2
1.00
1.01
2.18
2.08
0.98
4.34
1.06
2.13
1.99
1.03
1.95
1.18
3.12
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
ppm (t1)

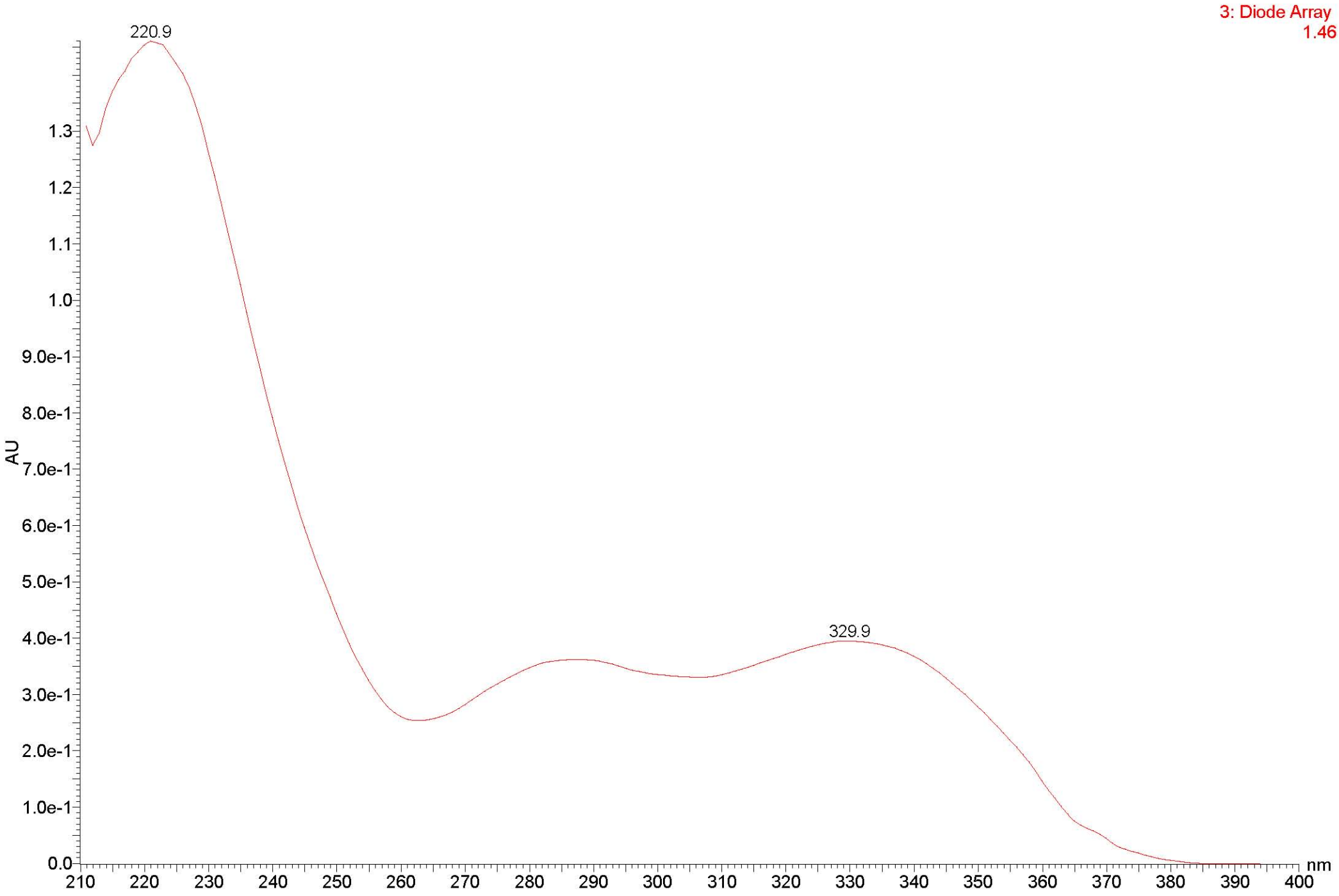
1a

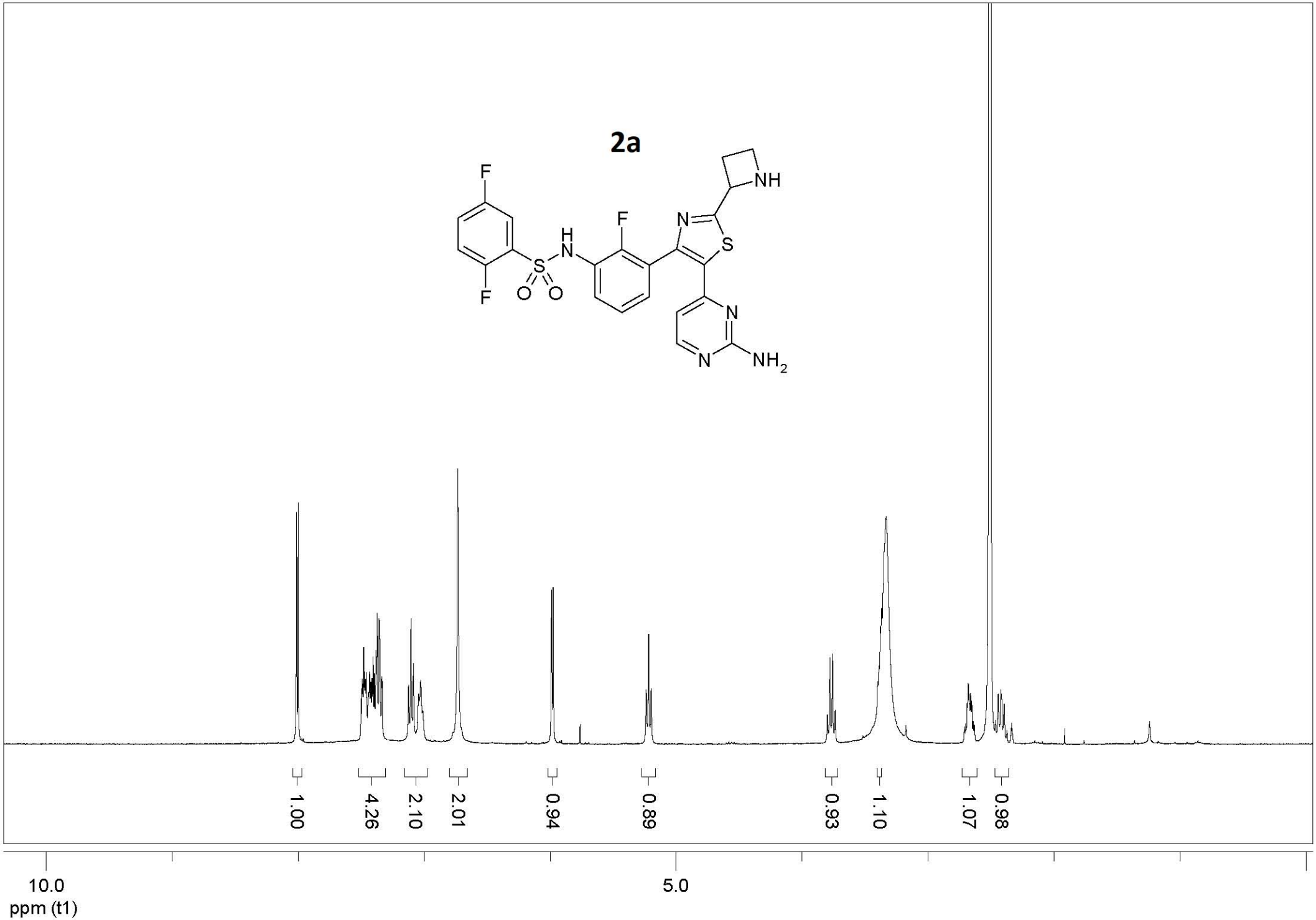

3: Diode Array
Range: 8.368e+1
Time Height Area Area%
4.25 75966400 11184605.00 93.68
4.63 433759 40493.62 0.34
4.83 2373825 299027.78 2.50
5.10 355880 32967.71 0.28
5.60 407329 70792.73 0.59
5.77 926112 67994.28 0.57
6.17 1143078 243728.42 2.04
4.25;220.9
4.83
214.9
5.77
216.9
6.17
217.9
AU
Time
8.0e+1
7.5e+1
7.0e+1
6.5e+1
6.0e+1
5.5e+1
5.0e+1
4.5e+1
4.0e+1
3.5e+1
3.0e+1
2.5e+1
2.0e+1
1.5e+1
1.0e+1
5.0
0.0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

1a

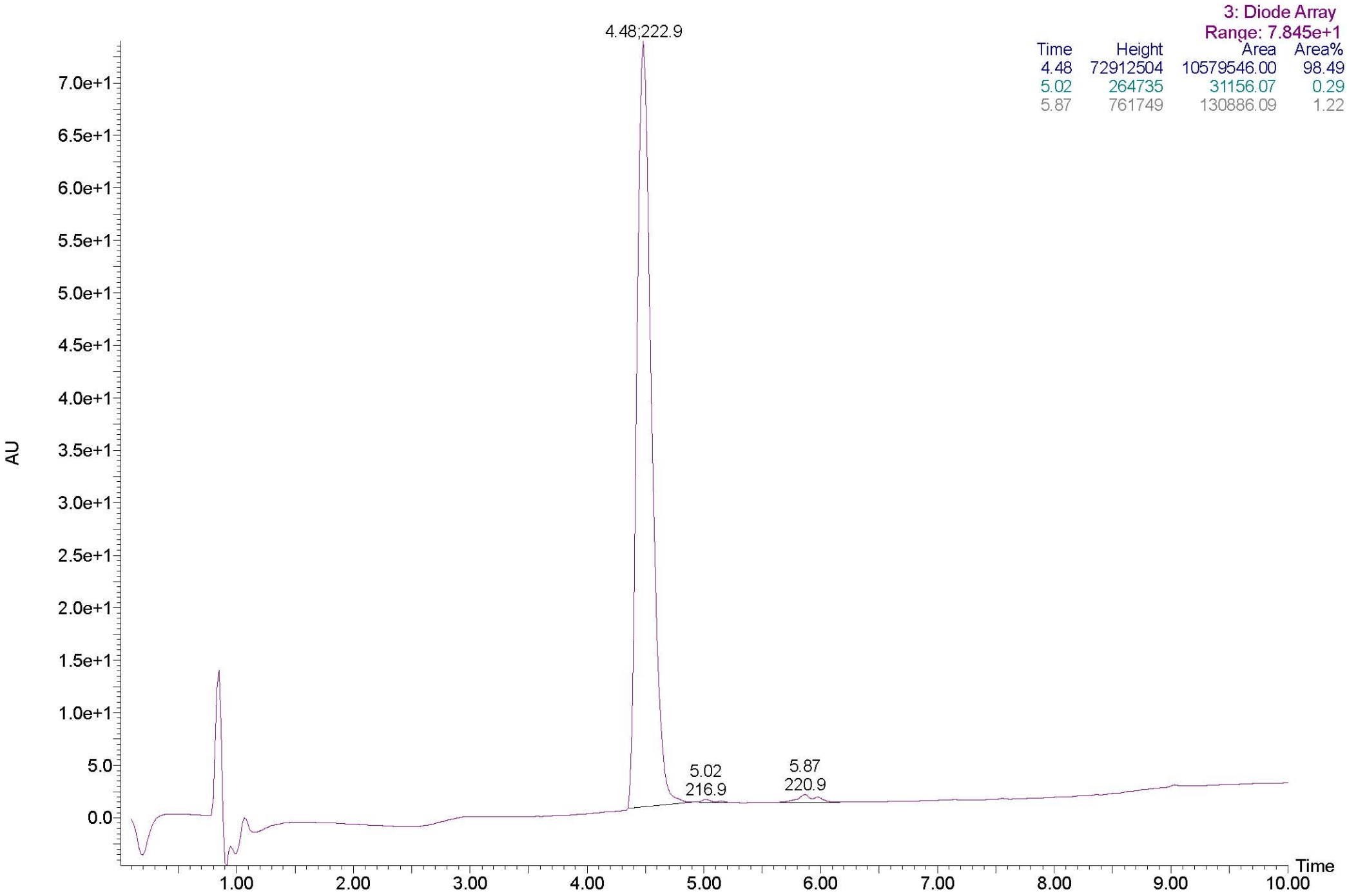

1: Scan ES+
1.11e7
100
%
0
246.3
266.9
245.1
224.1
269.2
288.1
329.0
336.1
371.4
372.4
396.1
423.2
438.4
489.4
491.4
492.5
541.6
558.5
592.0
598.1
645.4
659.0
688.1
710.2
739.8
751.9
787.5
797.4
m/z
200 225 250 275 300 325 350 375 400 425 450 475 500 525 550 575 600 625 650 675 700 725 750 775 800

1a

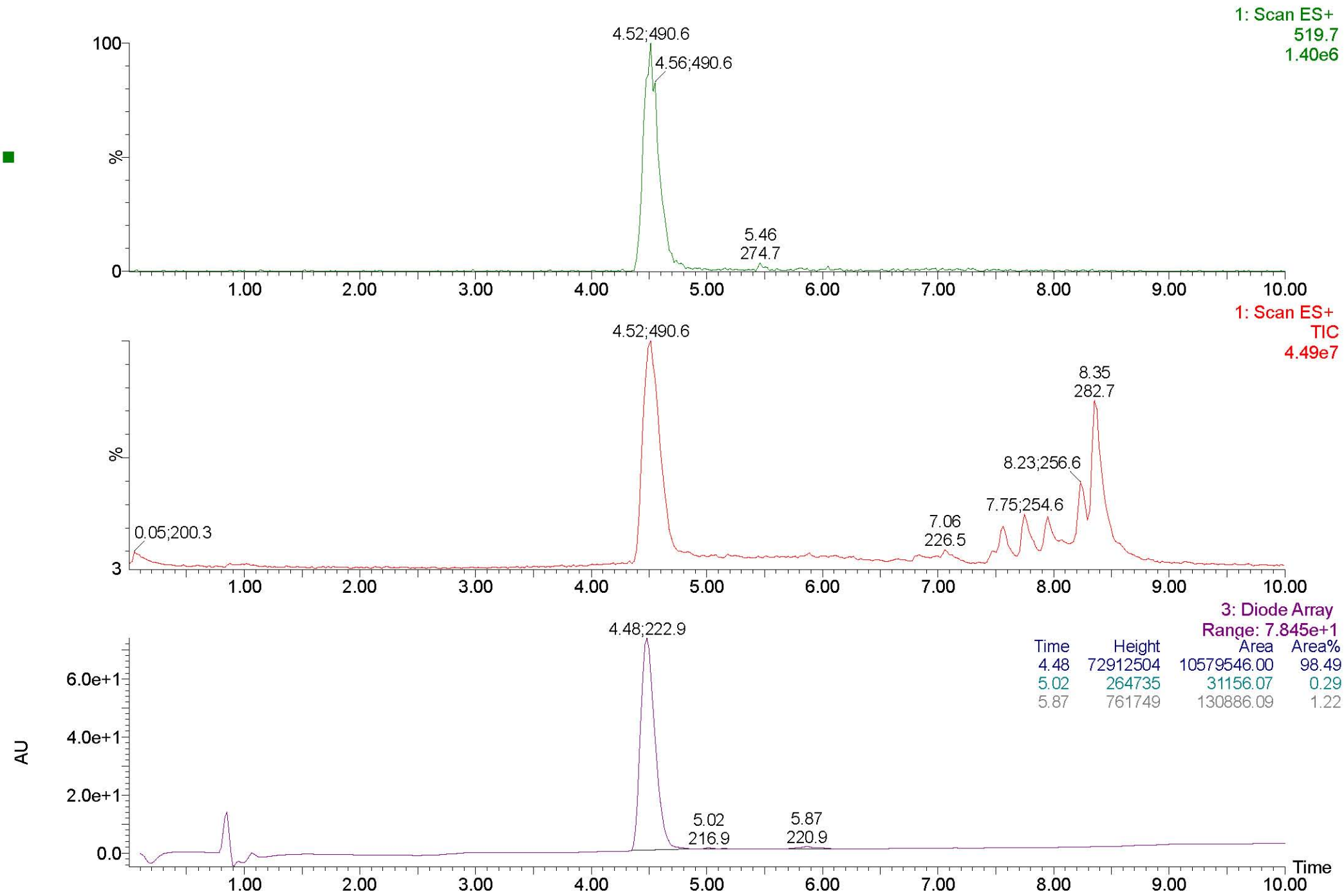

3: Diode Array
1.46
220.9
329.9
AU
1.3
1.2
1.1
1.0
9.0e-1
8.0e-1
7.0e-1
6.0e-1
5.0e-1
4.0e-1
3.0e-1
2.0e-1
1.0e-1
0.0
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400
nm


2a
F
F
S
O
O
H
N
F
N
S
NH
N
N
NH2
1.00
4.26
2.10
2.01
0.94
0.89
0.93
1.10
1.07
0.98
10.0
ppm (t1)
5.0

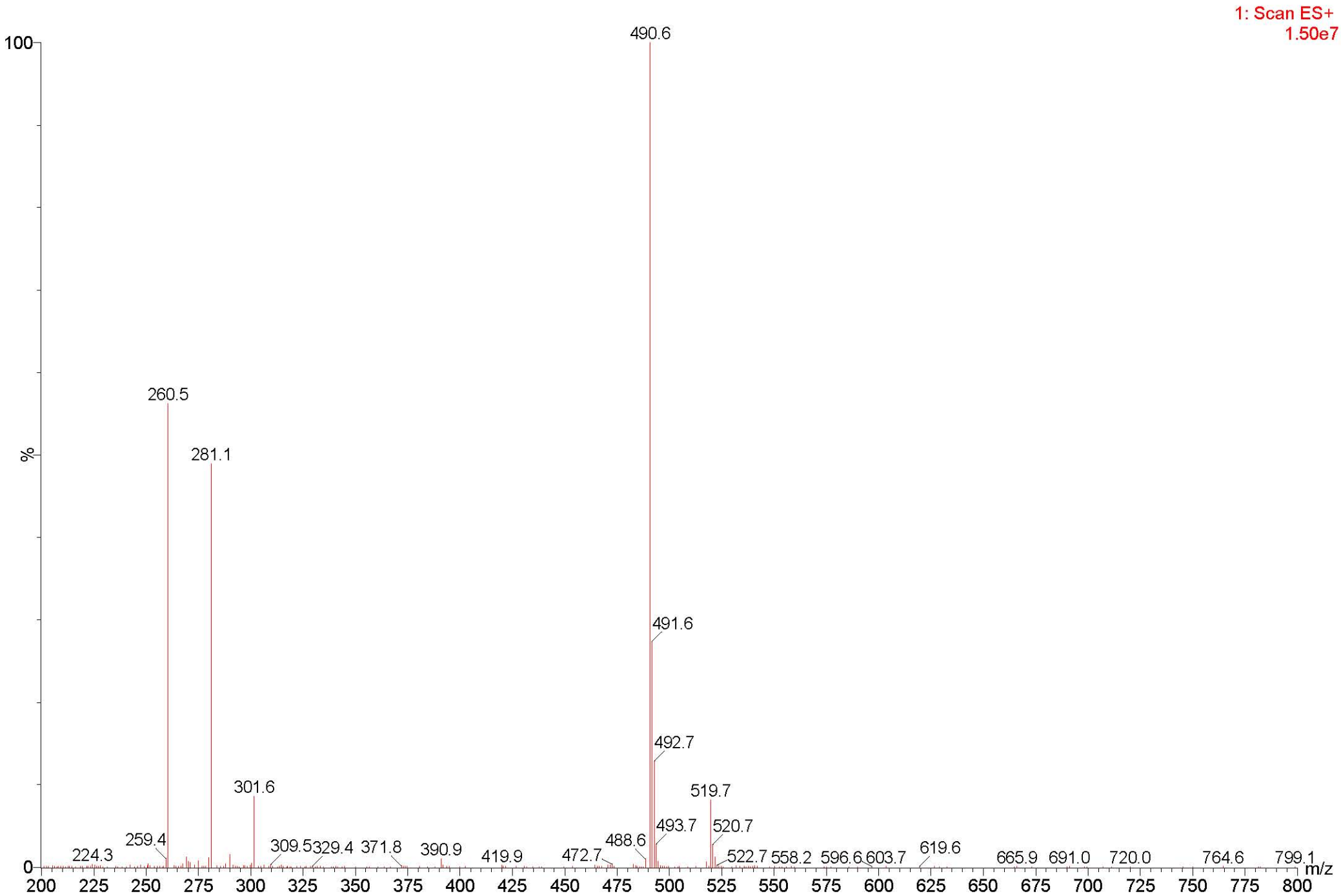
2a


3: Diode Array
Range: 7.845e+1
Time Height Area Area%
4.48 72912504 10579546.00 98.49
5.02 264735 31156.07 0.29
5.87 761749 130886.09 1.22
4.48;222.9
5.02
216.9
5.87
220.9
AU
Time
7.0e+1
6.5e+1
6.0e+1
5.5e+1
5.0e+1
4.5e+1
4.0e+1
3.5e+1
3.0e+1
2.5e+1
2.0e+1
1.5e+1
1.0e+1
5.0
0.0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

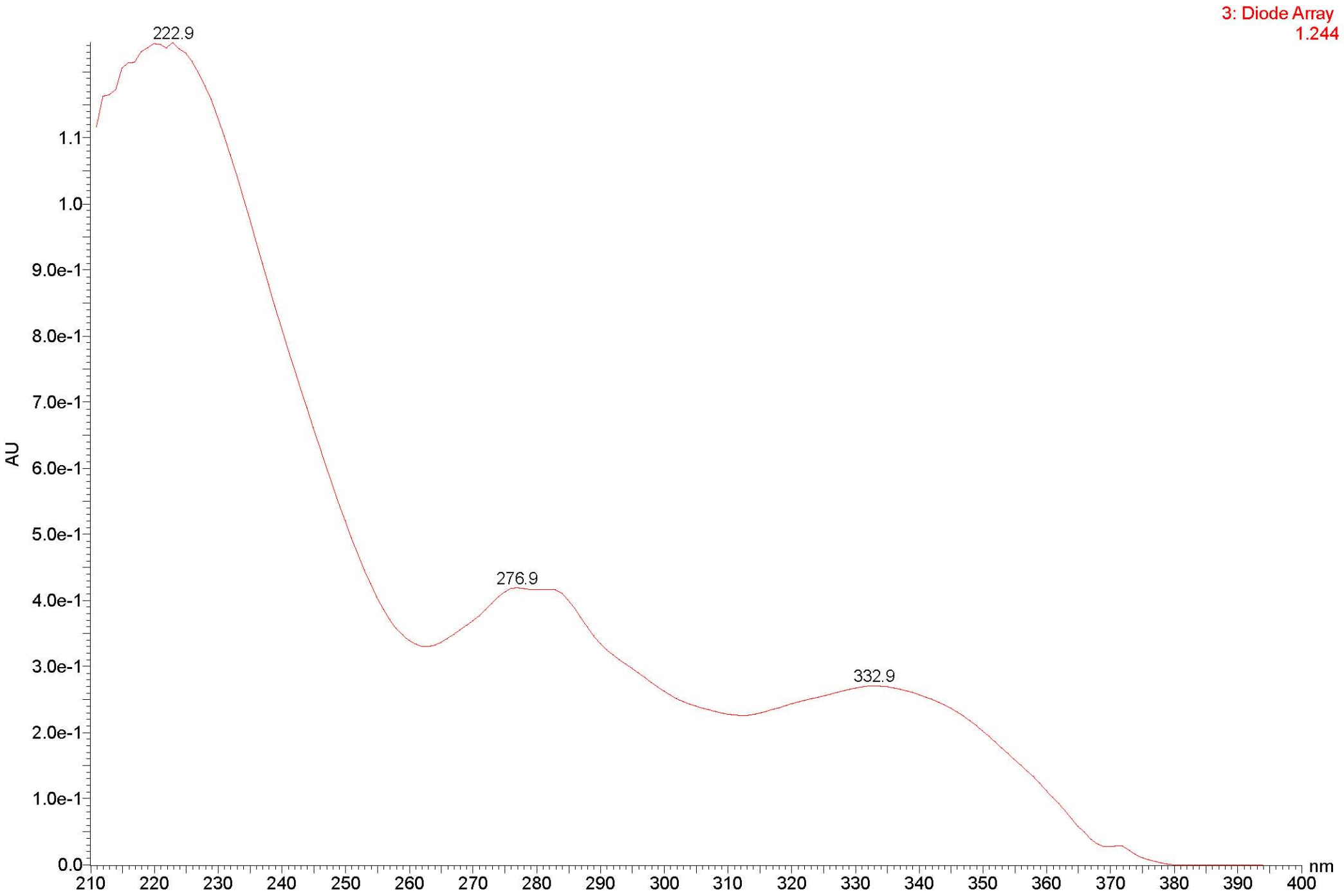
2a

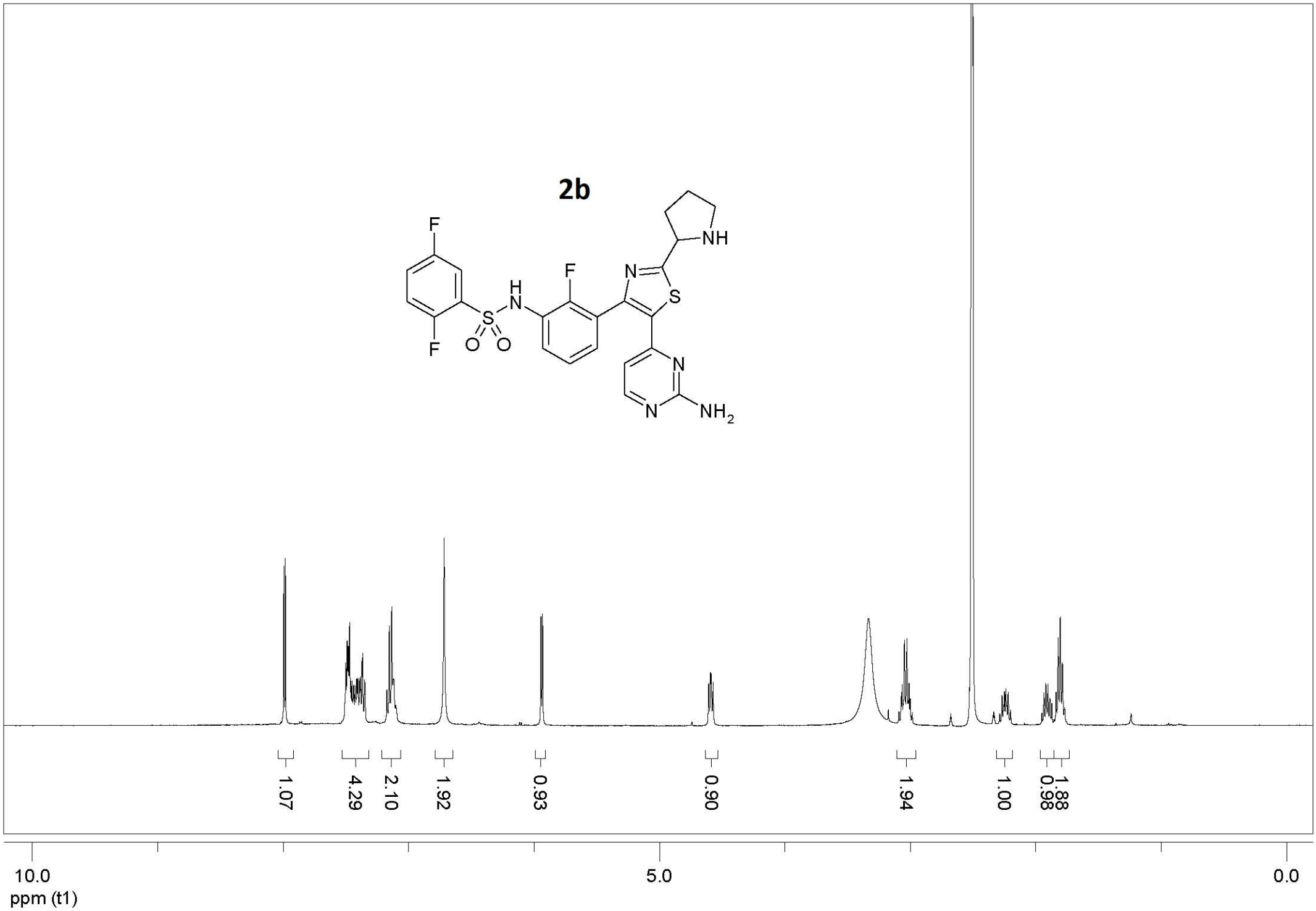

1: Scan ES+
519.7
1.40e6
4.52;490.6
4.56;490.6
5.46
274.7
100
%
0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
1: Scan ES+
TIC
4.49e7
4.52;490.6
8.35
282.7
8.23;256.6
7.75;254.6
7.06
226.5
0.05;200.3
%
3
3: Diode Array
Range: 7.845e+1
Time Height Area Area%
4.48 72912504 10579546.00 98.49
5.02 264735 31156.07 0.29
5.87 761749 130886.09 1.22
4.48;222.9
5.02
216.9
5.87
220.9
AU
6.0e+1
4.0e+1
2.0e+1
0.0
Time

2a

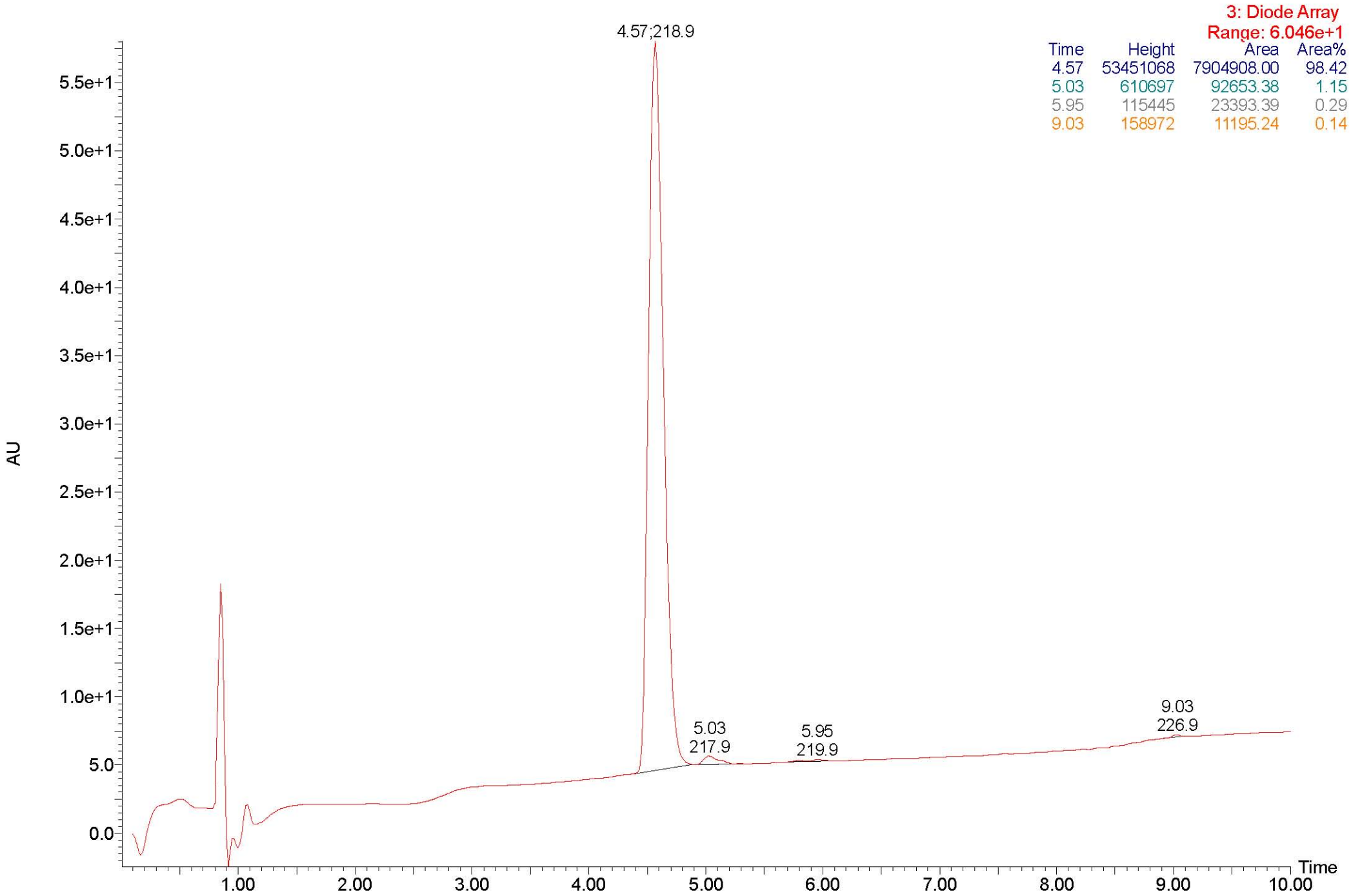

1: Scan ES+
1.50e7
100
%
0
490.6
260.5
281.1
301.6
491.6
492.7
519.7
493.7
520.7
259.4
224.3
309.5
329.4
371.8
390.9
419.9
472.7
488.6
522.7
558.2
596.6
603.7
619.6
665.9
691.0
720.0
764.6
799.1
m/z
200 225 250 275 300 325 350 375 400 425 450 475 500 525 550 575 600 625 650 675 700 725 750 775 800

2a

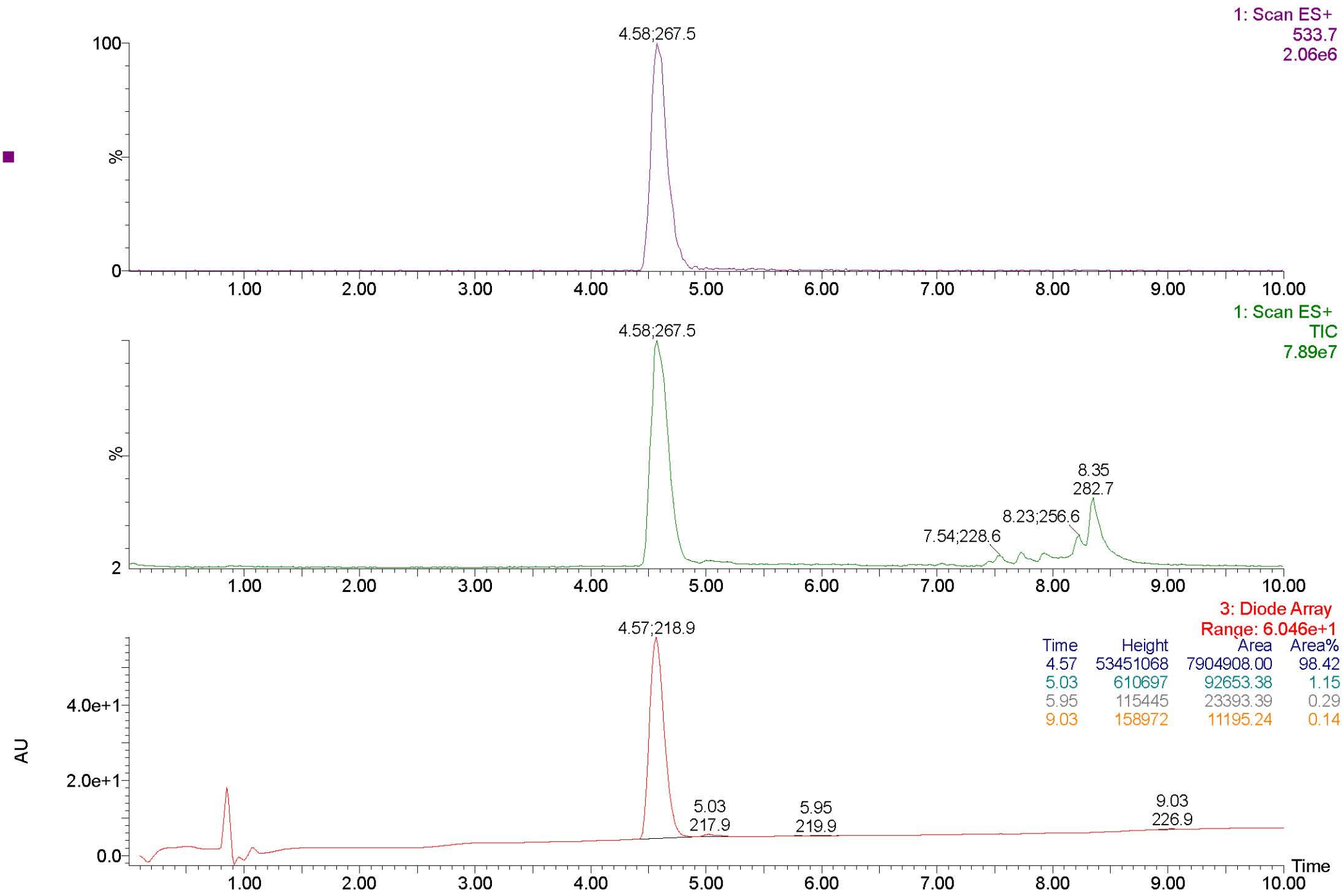

3: Diode Array
1.244
222.9
276.9
332.9
AU
1.1
1.0
9.0e-1
8.0e-1
7.0e-1
6.0e-1
5.0e-1
4.0e-1
3.0e-1
2.0e-1
1.0e-1
0.0
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400
nm


2b
F
F
S
O
O
H
N
F
N
S
NH
N
N
NH2
1.07
4.29
2.10
1.92
0.93
0.90
1.94
1.00
0.98
1.88
10.0
ppm (t1)
5.0
0.0

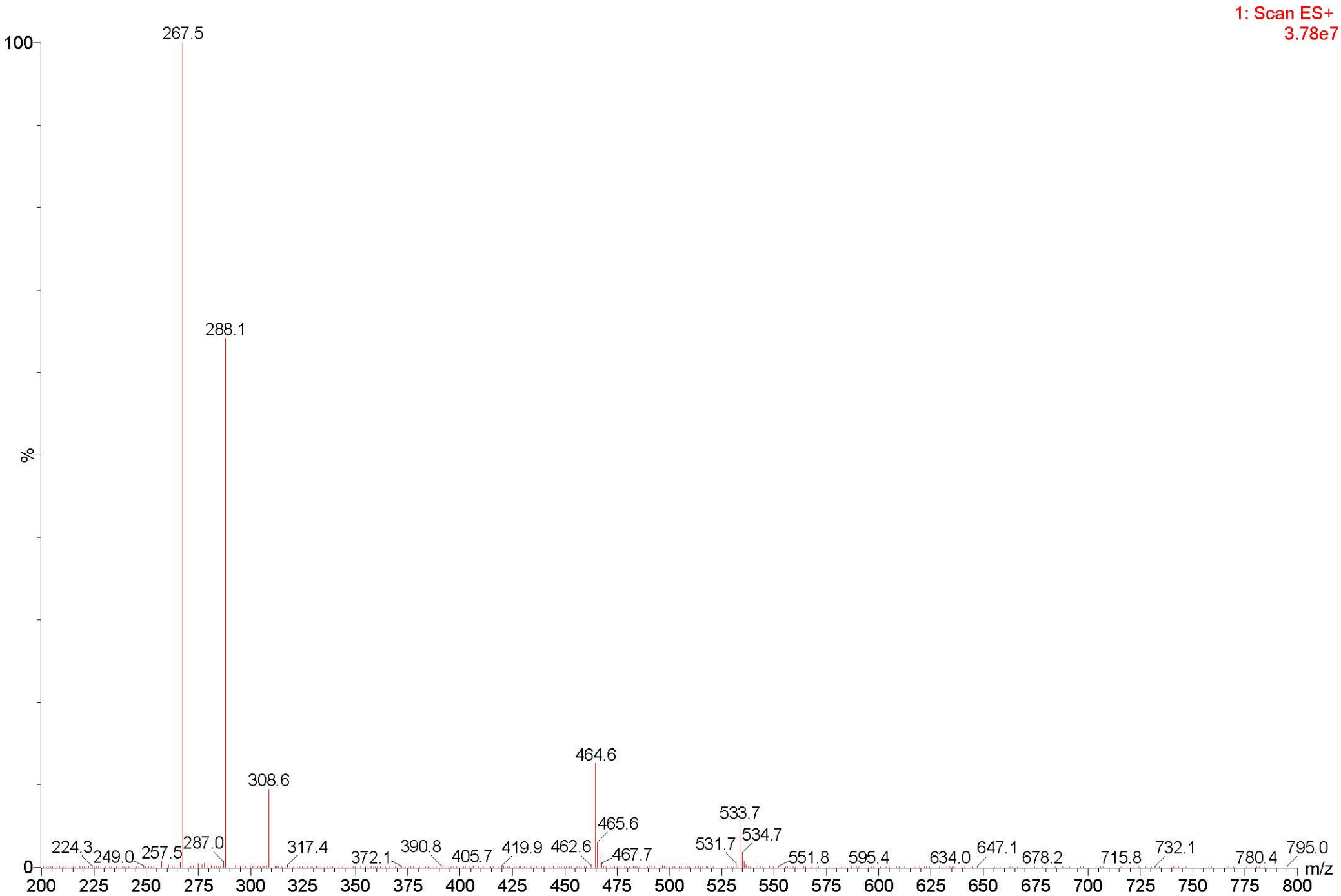
2b


3: Diode Array
Range: 6.046e+1
Time Height Area Area%
4.57 53451068 7904908.00 98.42
5.03 610697 92653.38 1.15
5.95 115445 23393.39 0.29
9.03 158972 11195.24 0.14
4.57;218.9
5.03
217.9
5.95
219.9
9.03
226.9
AU
Time
5.5e+1
5.0e+1
4.5e+1
4.0e+1
3.5e+1
3.0e+1
2.5e+1
2.0e+1
1.5e+1
1.0e+1
5.0
0.0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

2b

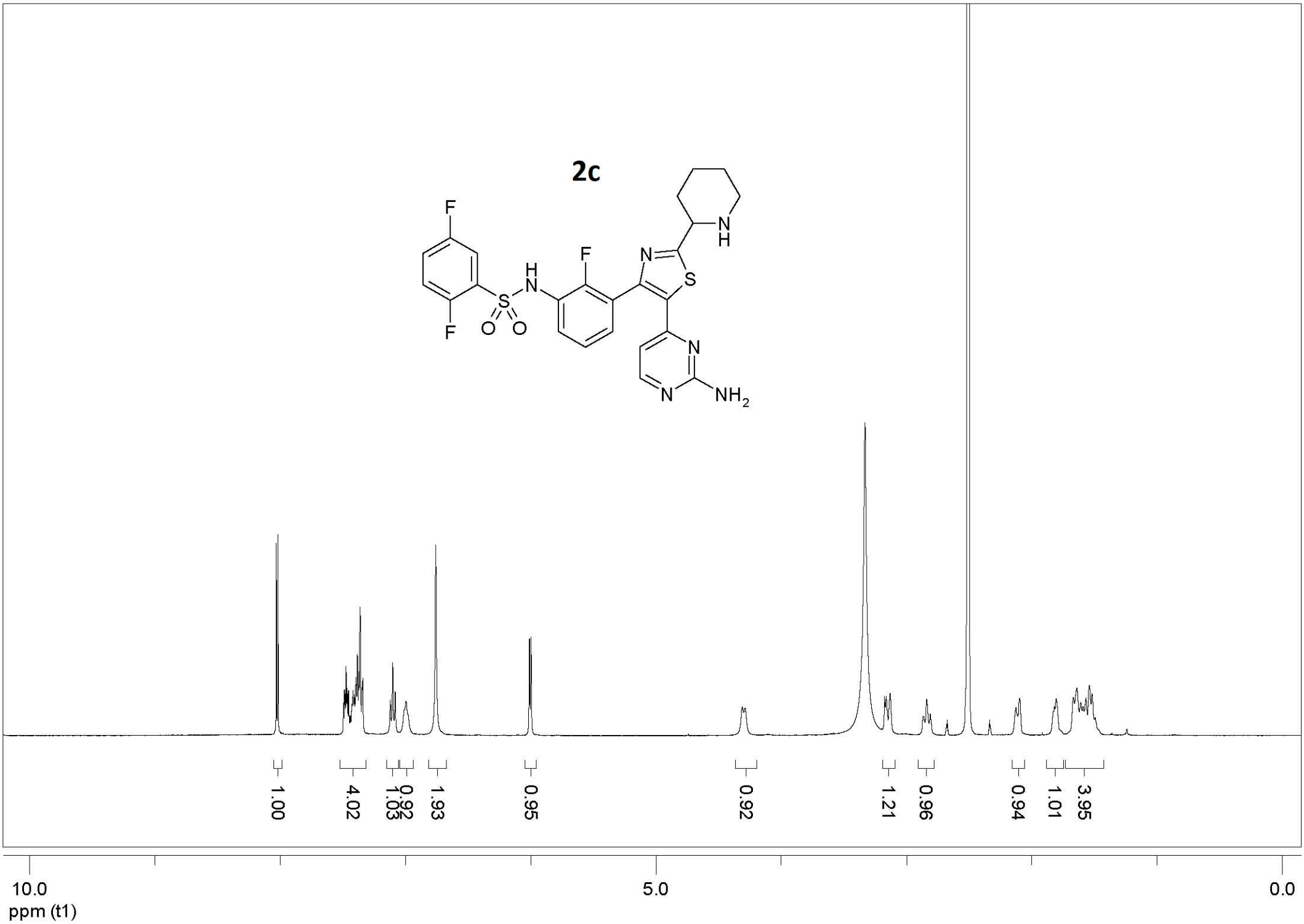

1: Scan ES+
533.7
2.06e6
4.58;267.5
100
%
0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
1: Scan ES+
TIC
7.89e7
4.58;267.5
8.35
282.7
8.23;256.6
7.54;228.6
2
3: Diode Array
Range: 6.046e+1
4.57;218.9
Time Height Area Area%
4.57 53451068 7904908.00 98.42
5.03 610697 92653.38 1.15
5.95 115445 23393.39 0.29
9.03 158972 11195.24 0.14
AU
4.0e+1
2.0e+1
0.0
5.03
217.9
5.95
219.9
9.03
226.9
Time

2b

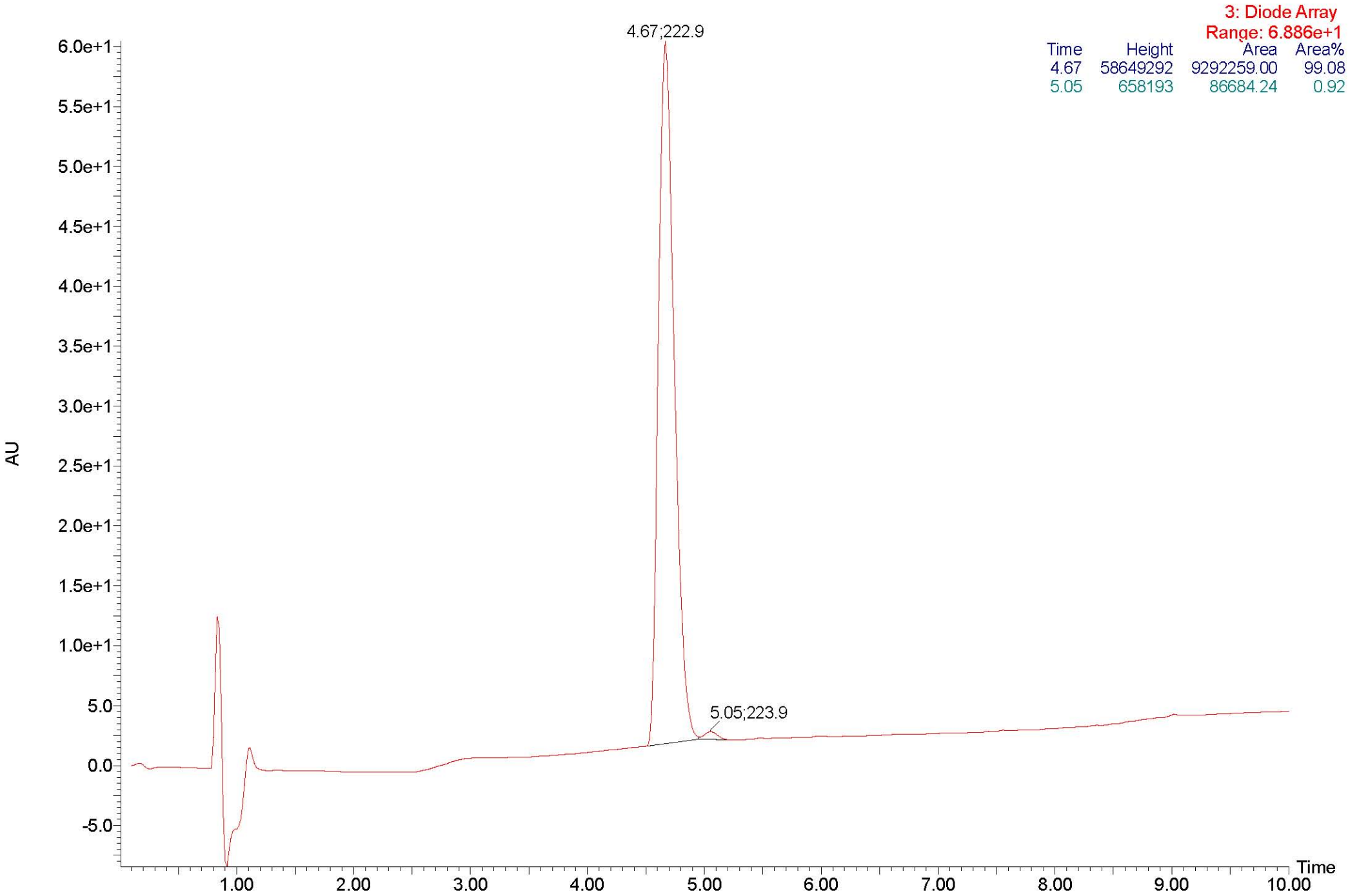

1: Scan ES+
3.78e7
100
%
0
267.5
288.1
308.6
464.6
465.6
533.7
534.7
224.3
249.0
257.5
287.0
317.4
372.1
390.8
405.7
419.9
462.6
467.7
531.7
551.8
595.4
634.0
647.1
678.2
715.8
732.1
780.4
795.0
m/z
200
225
250
275
300
325
350
375
400
425
450
475
500
525
550
575
600
625
650
675
700
725
750
775
800

2b

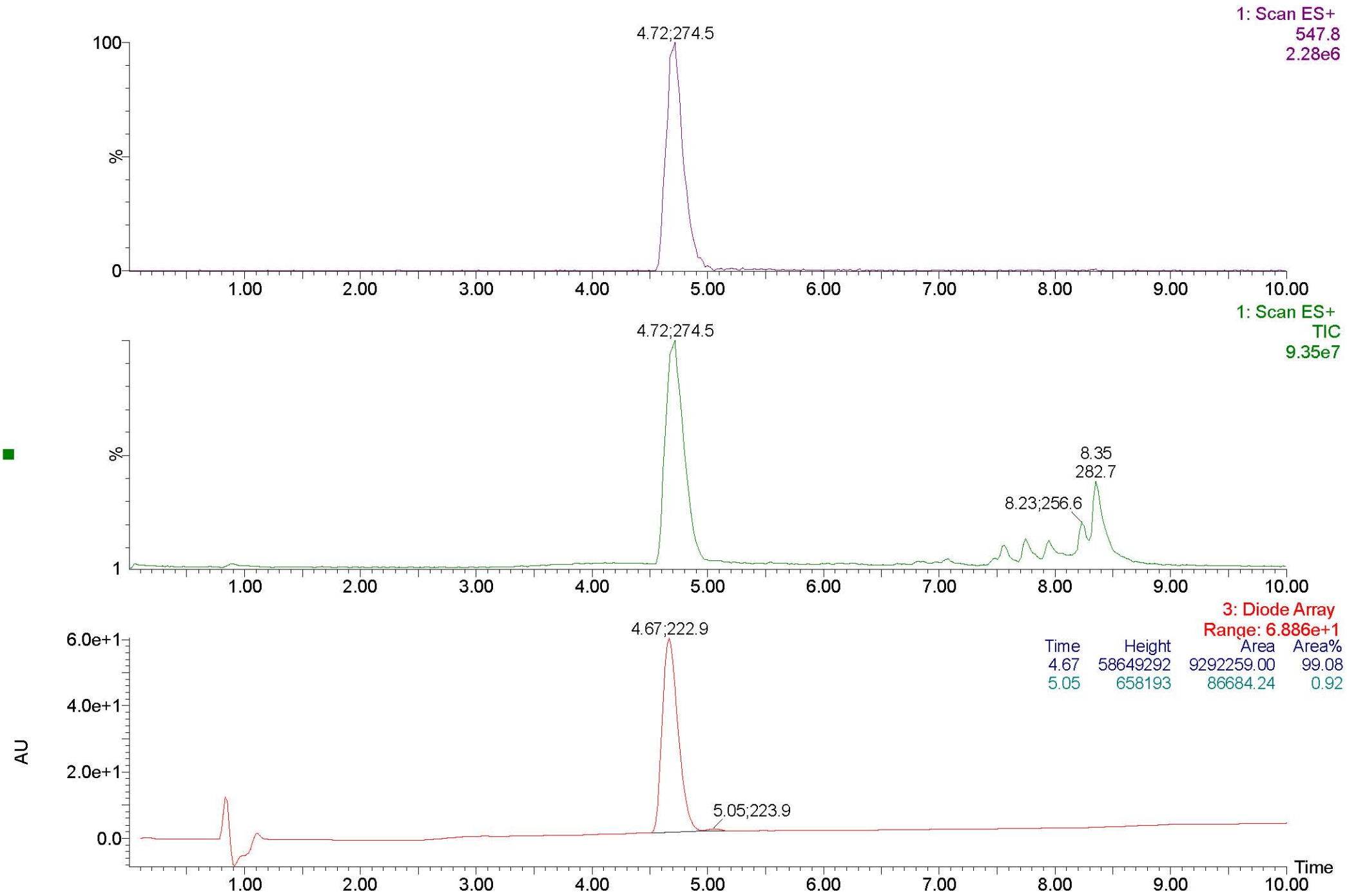

3: Diode Array
1.009
218.9
281.9
AU
9.5e-1
9.0e-1
8.5e-1
8.0e-1
7.5e-1
7.0e-1
6.5e-1
6.0e-1
5.5e-1
5.0e-1
4.5e-1
4.0e-1
3.5e-1
3.0e-1
2.5e-1
2.0e-1
1.5e-1
1.0e-1
5.0e-2
0.0
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400
nm


2c
F
F
S
O
O
H
N
F
N
S
N
H
N
N
NH2
1.00
4.02
1.03
0.92
1.93
0.95
0.92
1.21
0.96
0.94
1.01
3.95
10.0
5.0
0.0
ppm (t1)

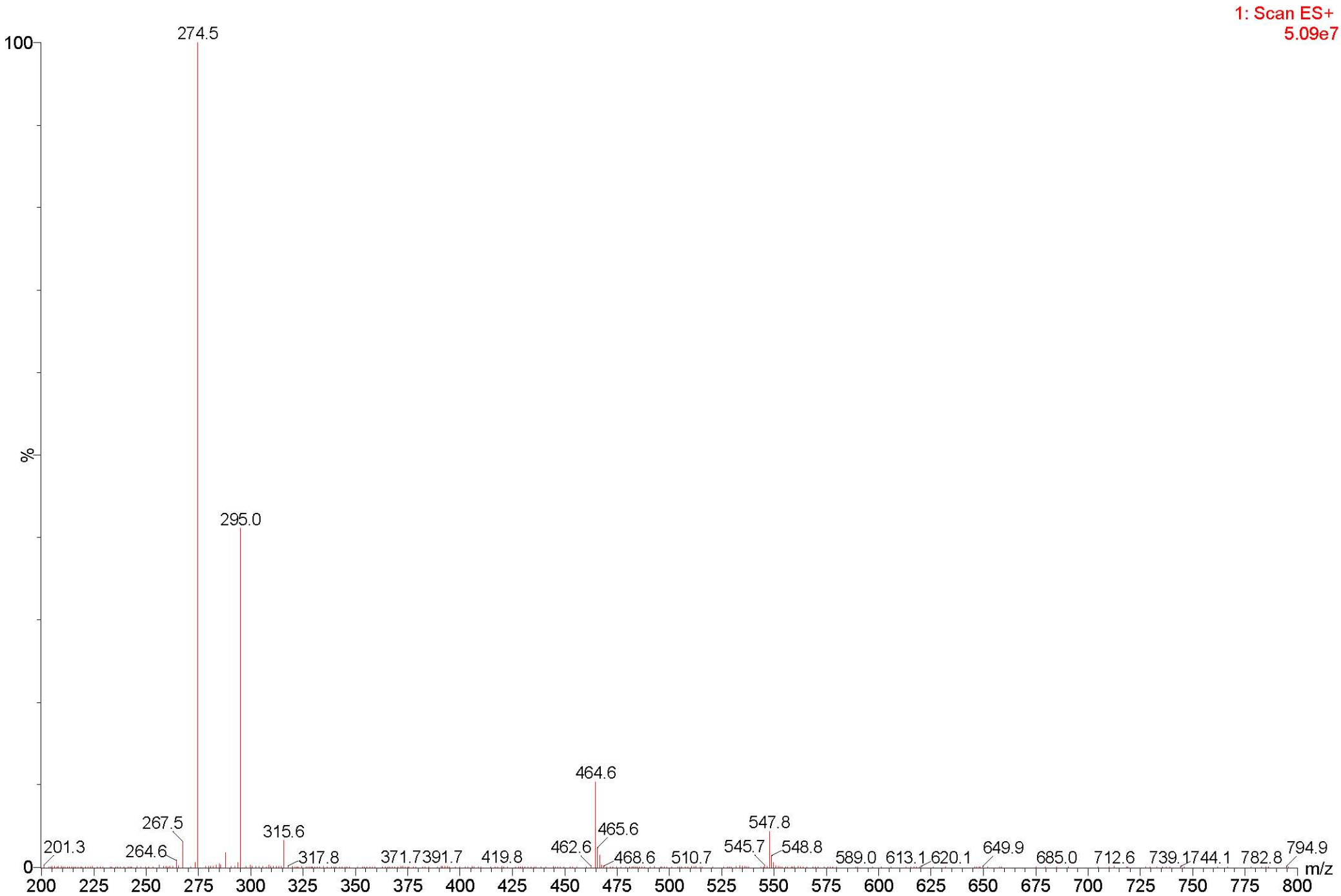
2c


3: Diode Array
Range: 6.886e+1
Time Height Area Area%
4.67 58649292 9292259.00 99.08
5.05 658193 86684.24 0.92
4.67;222.9
5.05;223.9
AU
Time
6.0e+1
5.5e+1
5.0e+1
4.5e+1
4.0e+1
3.5e+1
3.0e+1
2.5e+1
2.0e+1
1.5e+1
1.0e+1
5.0
0.0
-5.0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

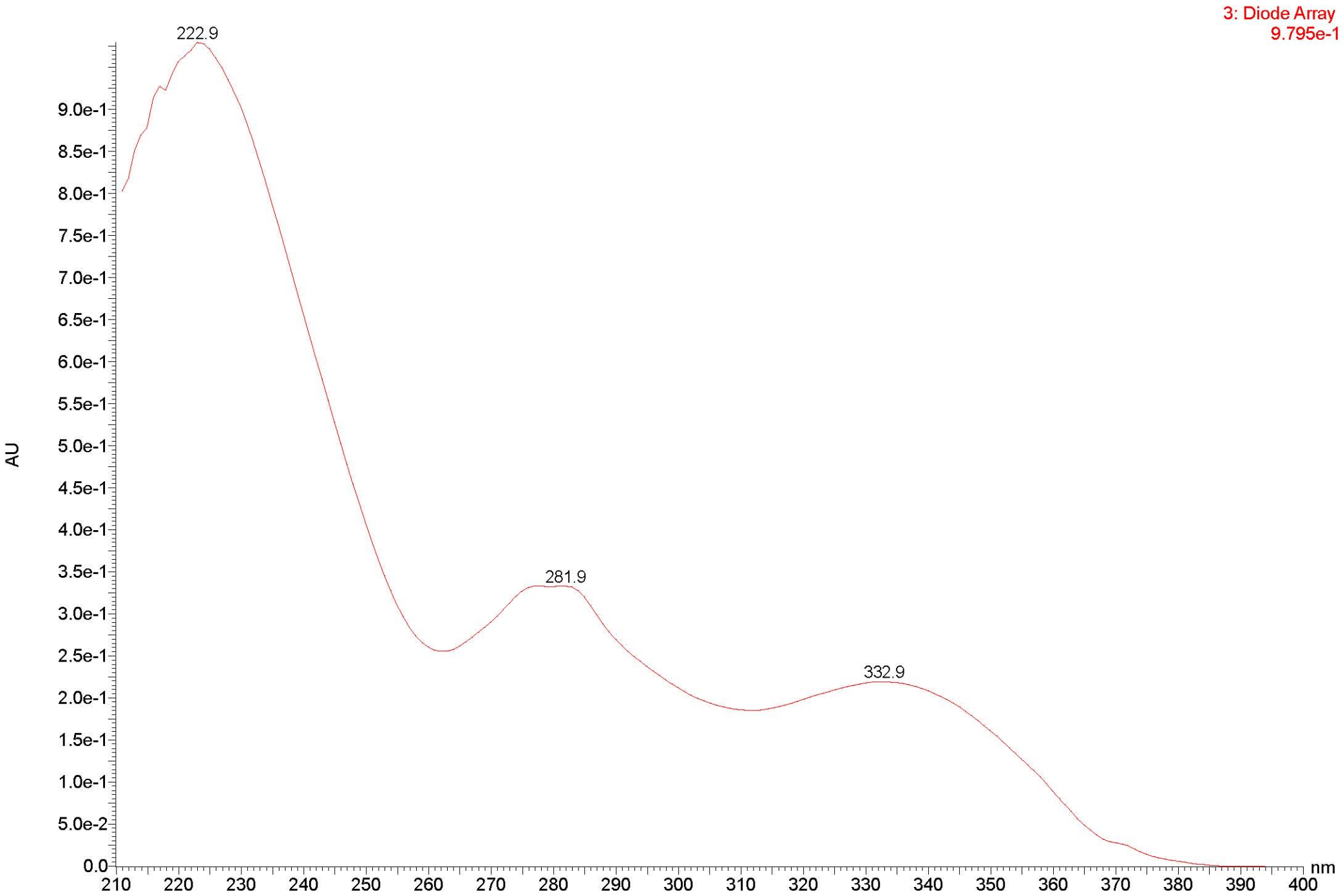
2c

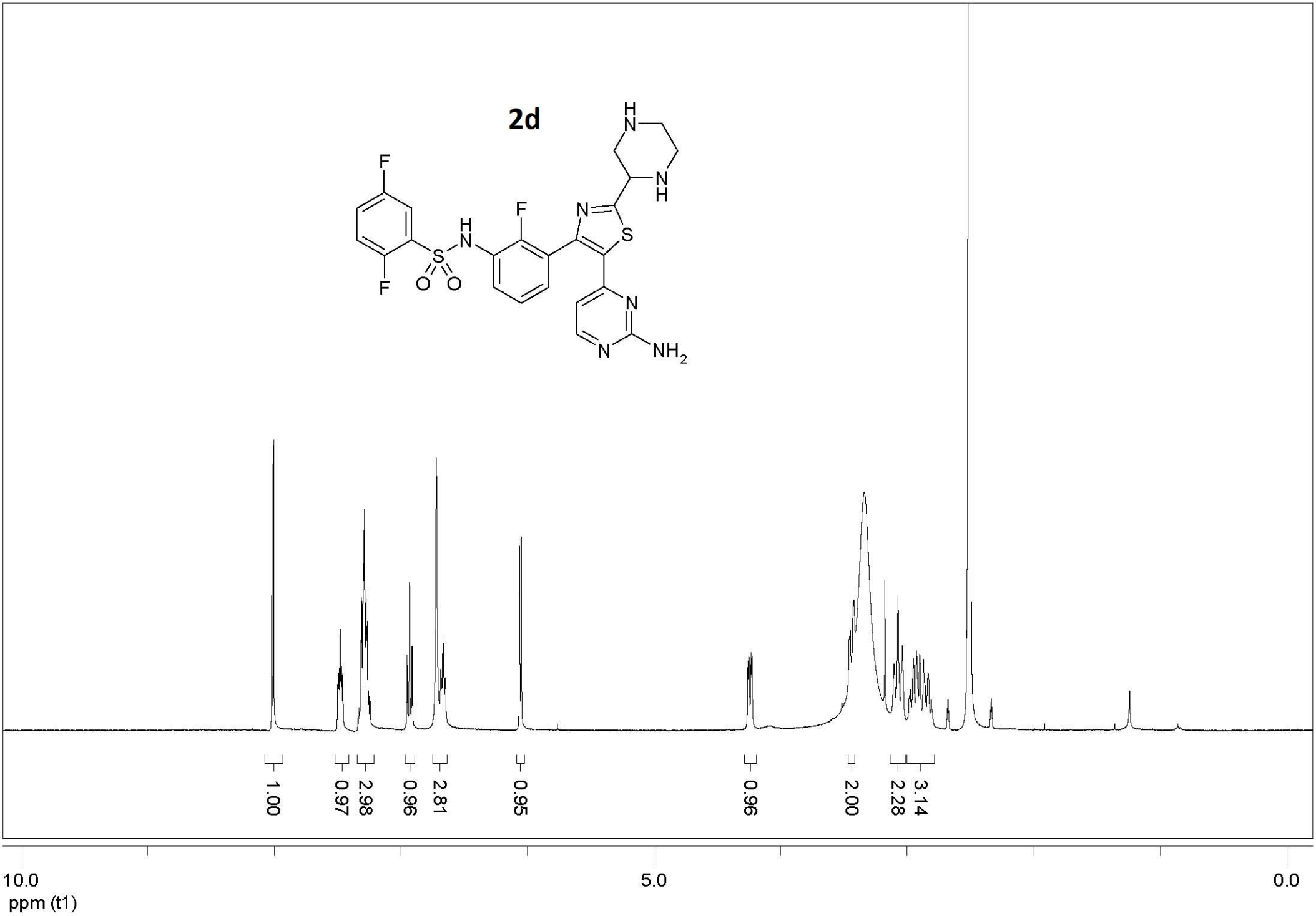

1: Scan ES+
547.8
2.28e6
4.72;274.5
100
%
0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
1: Scan ES+
TIC
9.35e7
4.72;274.5
8.35
282.7
8.23;256.6
%
1
3: Diode Array
Range: 6.886e+1
4.67;222.9
5.05;223.9
6.0e+1
4.0e+1
2.0e+1
0.0
AU
Time
Time Height Area Area%
4.67 58649292 9292259.00 99.08
5.05 658193 86684.24 0.92

2c

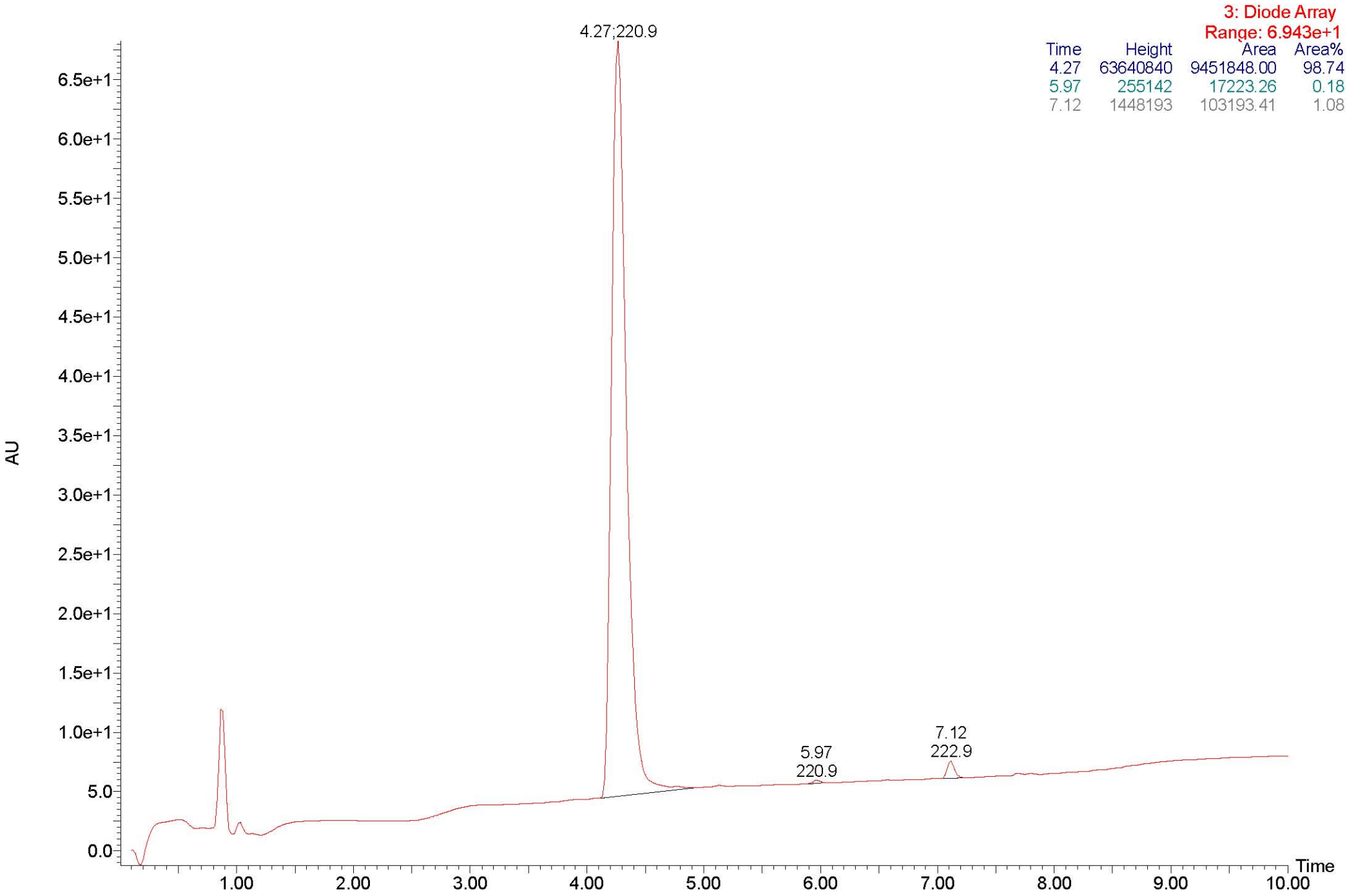

1: Scan ES+
5.09e7
100
%
0
274.5
295.0
464.6
267.5
315.6
465.6
547.8
201.3
264.6
317.8
371.7391.7
419.8
462.6
468.6
510.7
545.7
548.8
589.0
613.1620.1
649.9
685.0
712.6
739.1744.1
782.8
794.9
m/z
200 225 250 275 300 325 350 375 400 425 450 475 500 525 550 575 600 625 650 675 700 725 750 775 800

2c

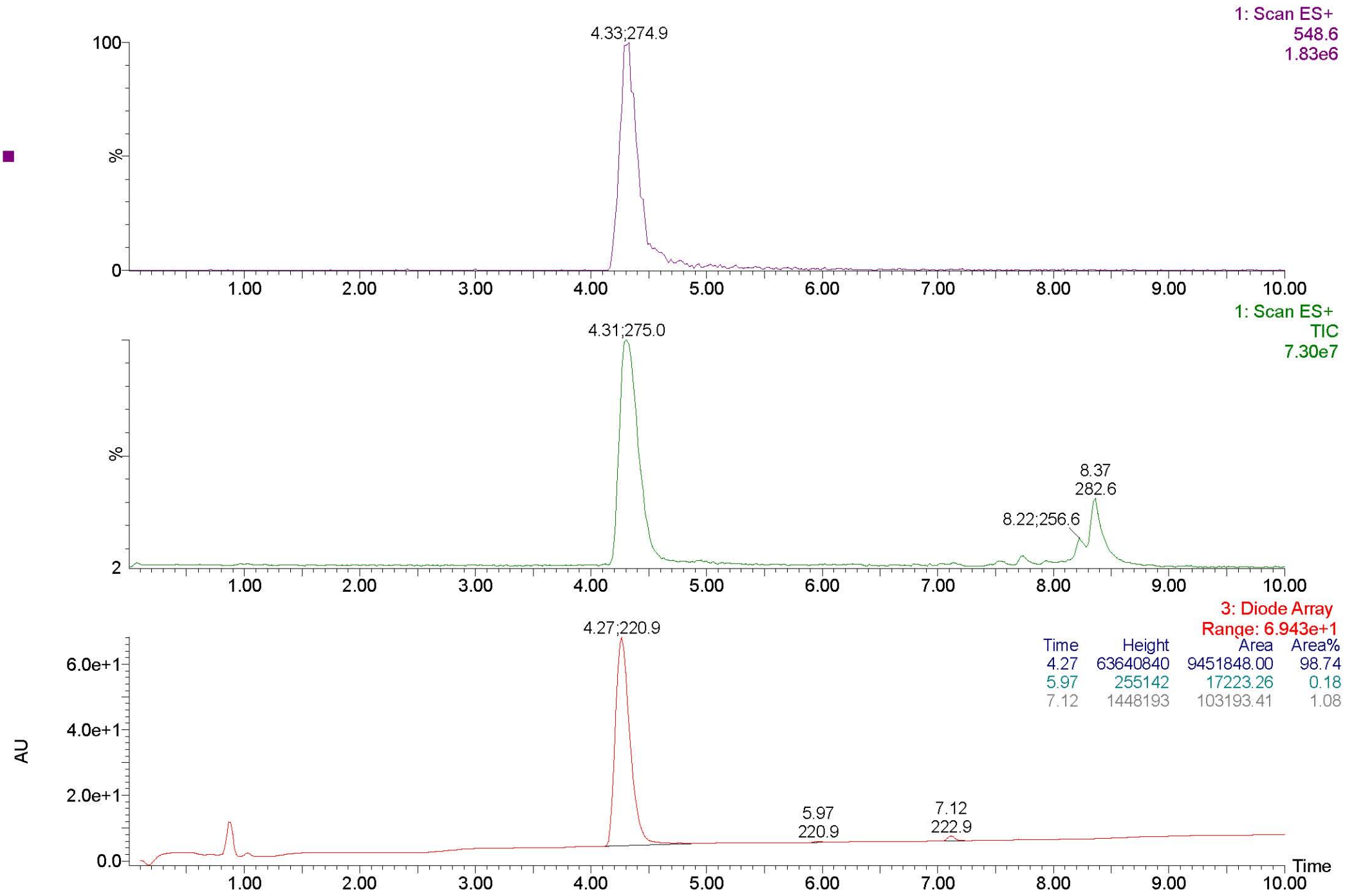

3: Diode Array
9.795e-1
222.9
281.9
332.9
AU
9.0e-1
8.5e-1
8.0e-1
7.5e-1
7.0e-1
6.5e-1
6.0e-1
5.5e-1
5.0e-1
4.5e-1
4.0e-1
3.5e-1
3.0e-1
2.5e-1
2.0e-1
1.5e-1
1.0e-1
5.0e-2
0.0
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400
nm


2d
F
F
S
O
O
H
N
F
N
S
H
N
N
N
N
NH2
1.00
0.97
2.98
0.96
2.81
0.95
0.96
2.00
2.28
3.14
10.0
5.0
0.0
ppm (t1)

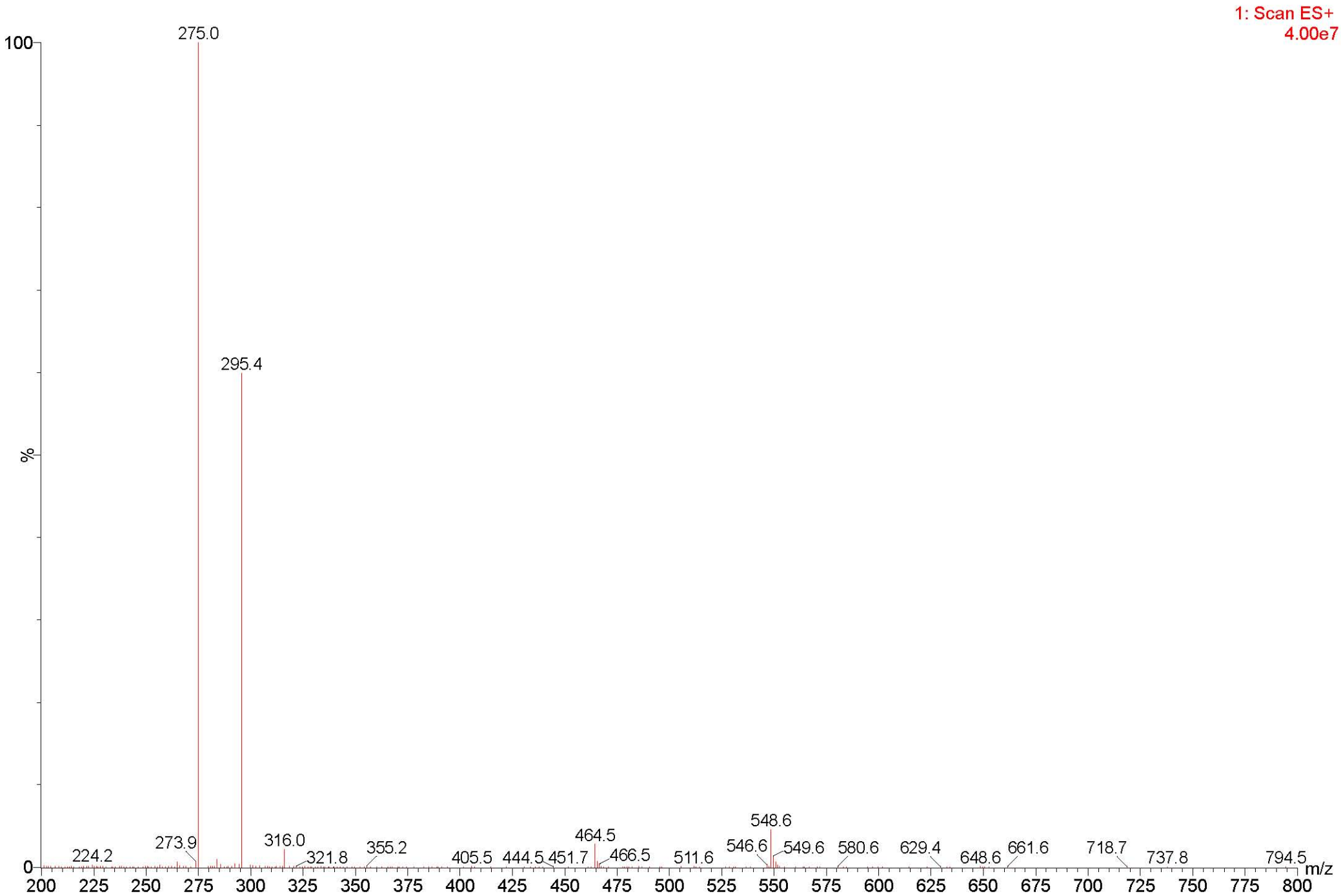
2d


3: Diode Array
Range: 6.943e+1
Time Height Area Area%
4.27 63640840 9451848.00 98.74
5.97 255142 17223.26 0.18
7.12 1448193 103193.41 1.08
4.27;220.9
5.97
220.9
7.12
222.9
AU
Time
6.5e+1
6.0e+1
5.5e+1
5.0e+1
4.5e+1
4.0e+1
3.5e+1
3.0e+1
2.5e+1
2.0e+1
1.5e+1
1.0e+1
5.0
0.0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

2d

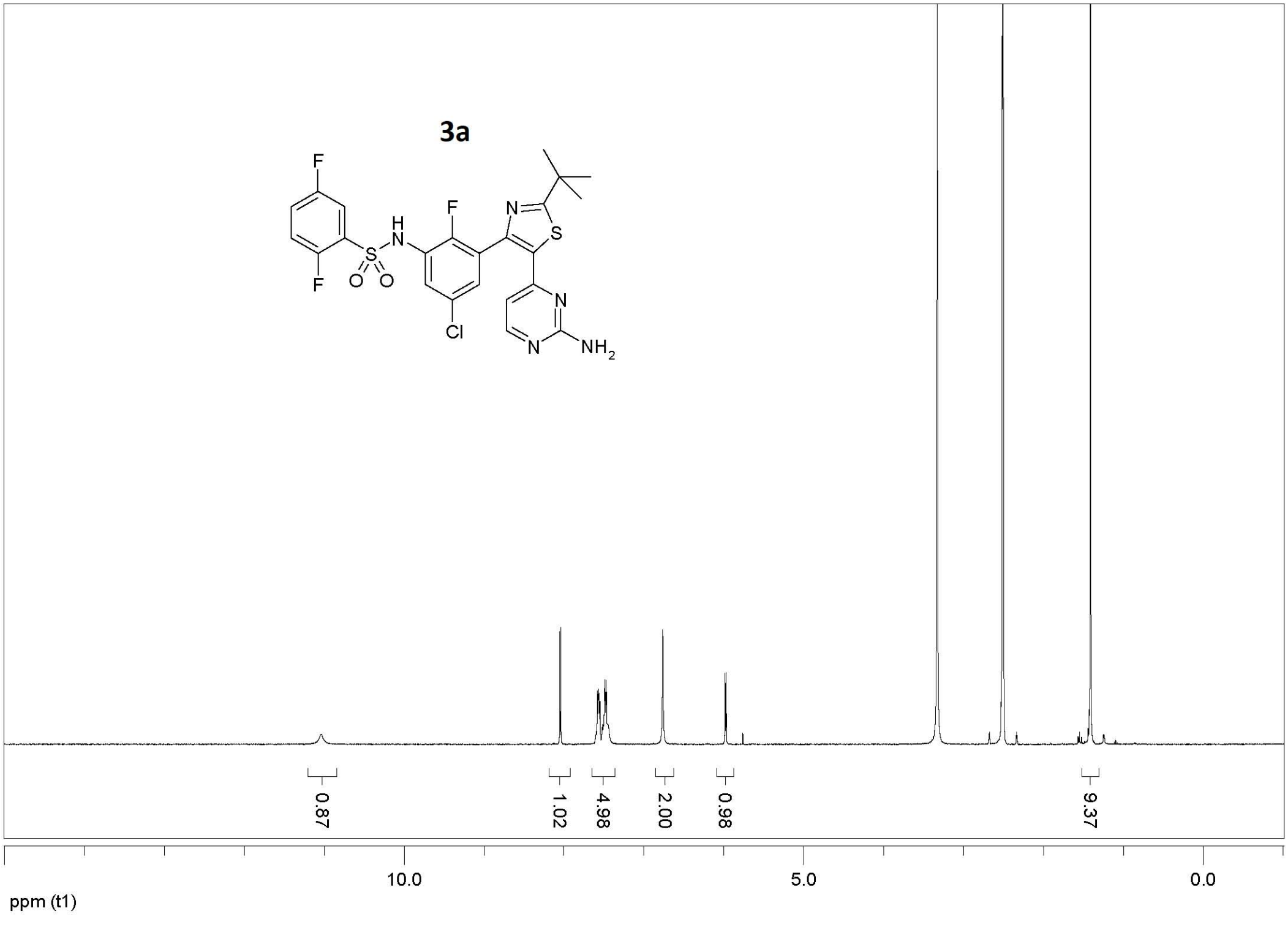

1: Scan ES+
548.6
1.83e6
4.33;274.9
100
%
0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
1: Scan ES+
TIC
7.30e7
4.31;275.0
8.37
282.6
8.22;256.6
%
2
3: Diode Array
Range: 6.943e+1
4.27;220.9
Time Height Area Area%
4.27 63640840 9451848.00 98.74
5.97 255142 17223.26 0.18
7.12 1448193 103193.41 1.08
6.0e+1
4.0e+1
2.0e+1
0.0
AU
5.97
220.9
7.12
222.9
Time

2d

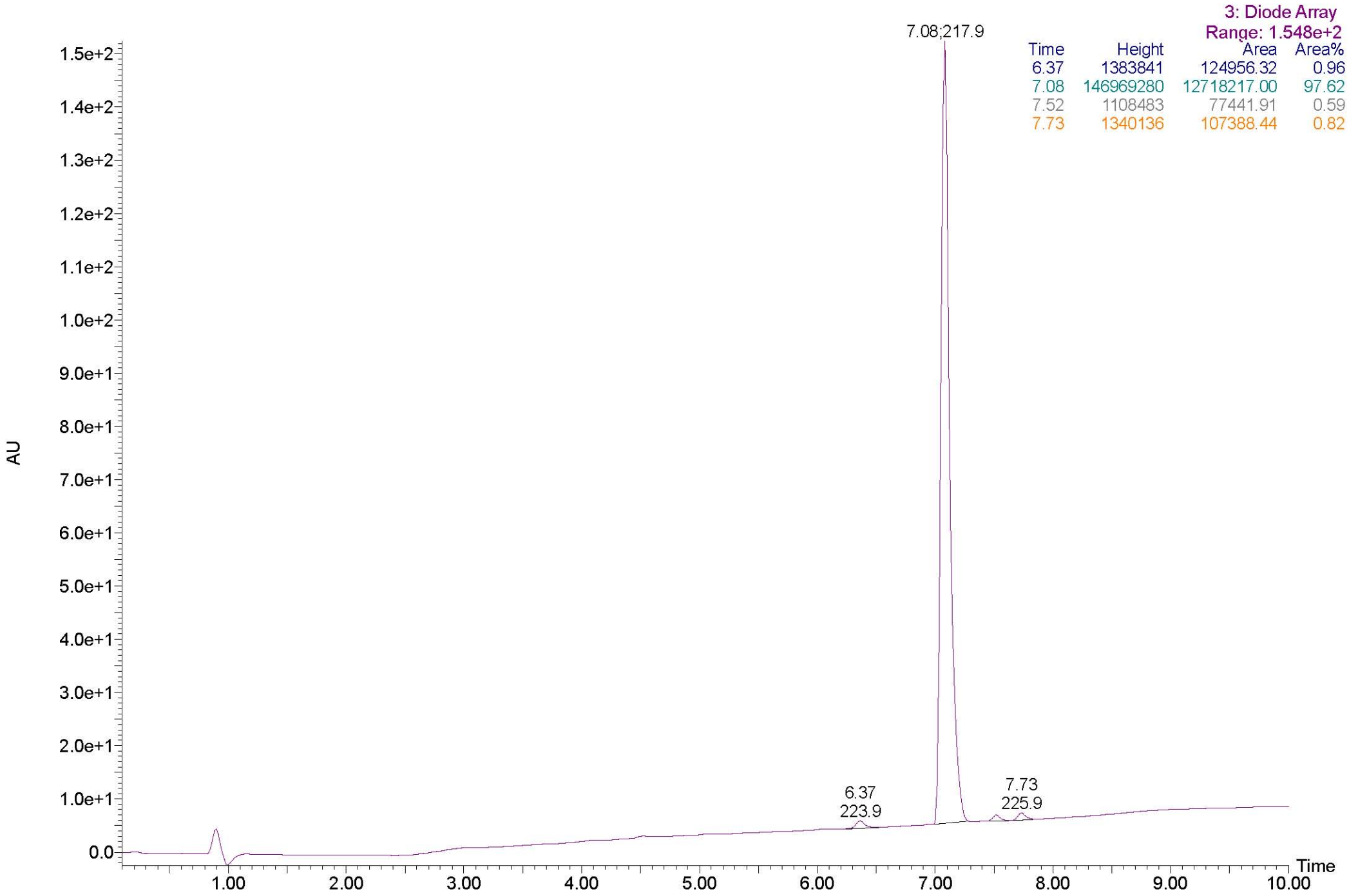

1: Scan ES+
4.00e7
100
%
0
275.0
295.4
224.2
273.9
316.0
321.8
355.2
405.5
444.5
451.7
464.5
466.5
511.6
546.6
548.6
549.6
580.6
629.4
648.6
661.6
718.7
737.8
794.5
m/z
200
225
250
275
300
325
350
375
400
425
450
475
500
525
550
575
600
625
650
675
700
725
750
775
800

2d

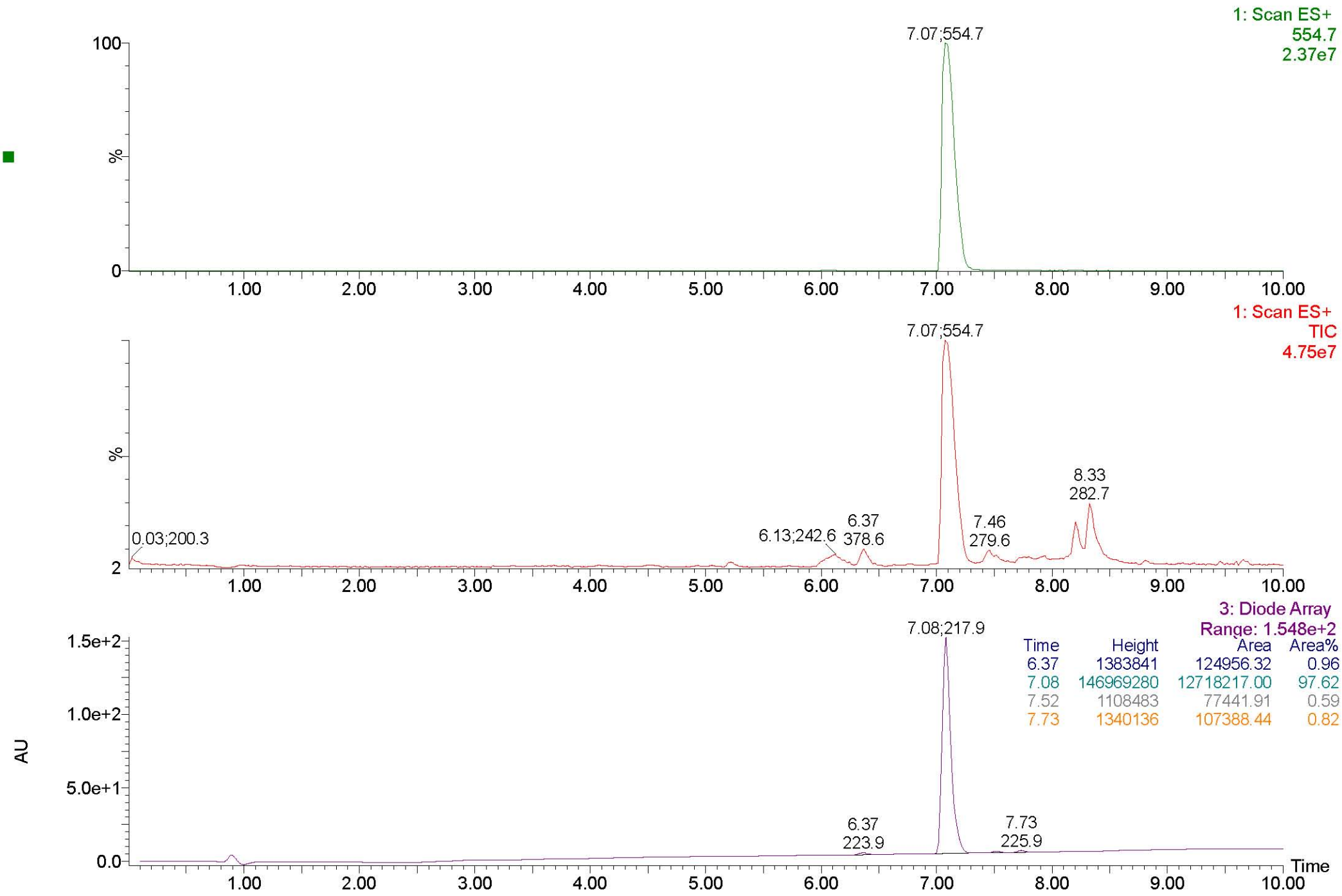

3: Diode Array
1.172
220.9
281.9
330.9
AU
1.1
1.0
9.0e-1
8.0e-1
7.0e-1
6.0e-1
5.0e-1
4.0e-1
3.0e-1
2.0e-1
1.0e-1
0.0
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400
nm


3a
F
F
F
H
N
S
O
O
N
S
N
Cl
N
NH2
0.87
1.02
4.98
2.00
0.98
9.37
10.0
5.0
0.0
ppm (t1)

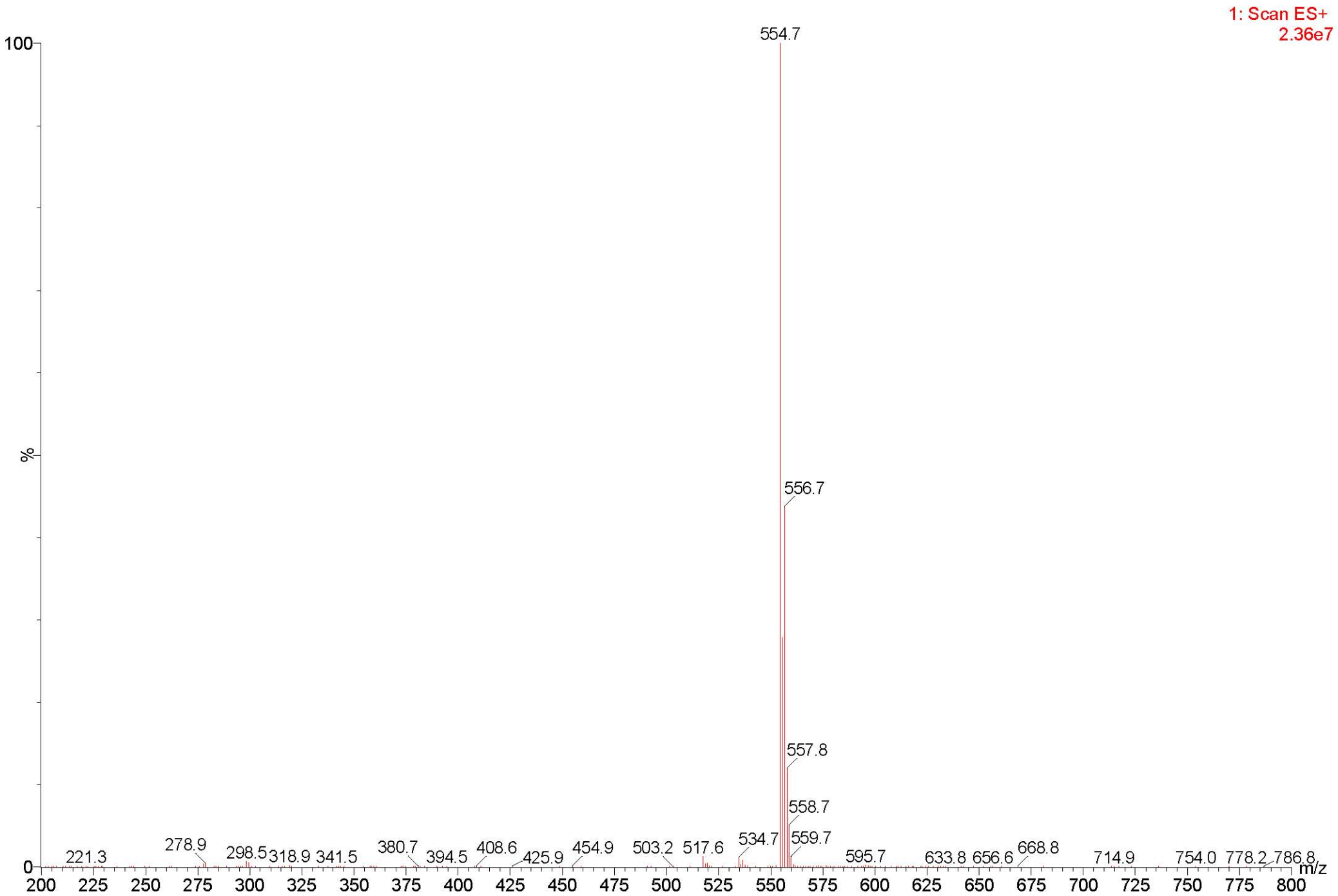
3a

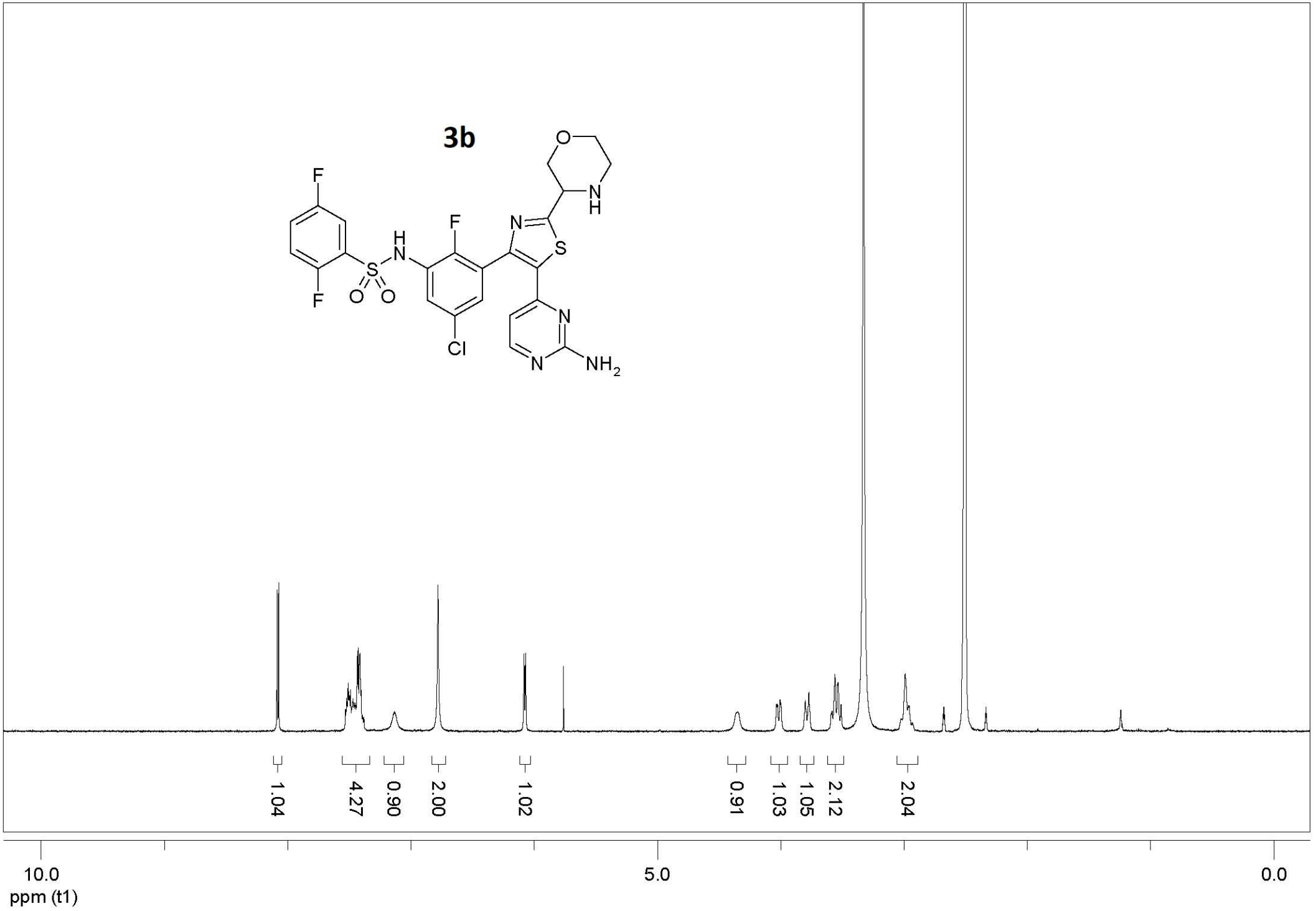

3: Diode Array
Range: 1.548e+2
Time Height Area Area%
6.37 1383841 124956.32 0.96
7.08 146969280 12718217.00 97.62
7.52 1108483 77441.91 0.59
7.73 1340136 107388.44 0.82
7.08;217.9
6.37
223.9
7.73
225.9
AU
Time
1.5e+2
1.4e+2
1.3e+2
1.2e+2
1.1e+2
1.0e+2
9.0e+1
8.0e+1
7.0e+1
6.0e+1
5.0e+1
4.0e+1
3.0e+1
2.0e+1
1.0e+1
0.0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

3a

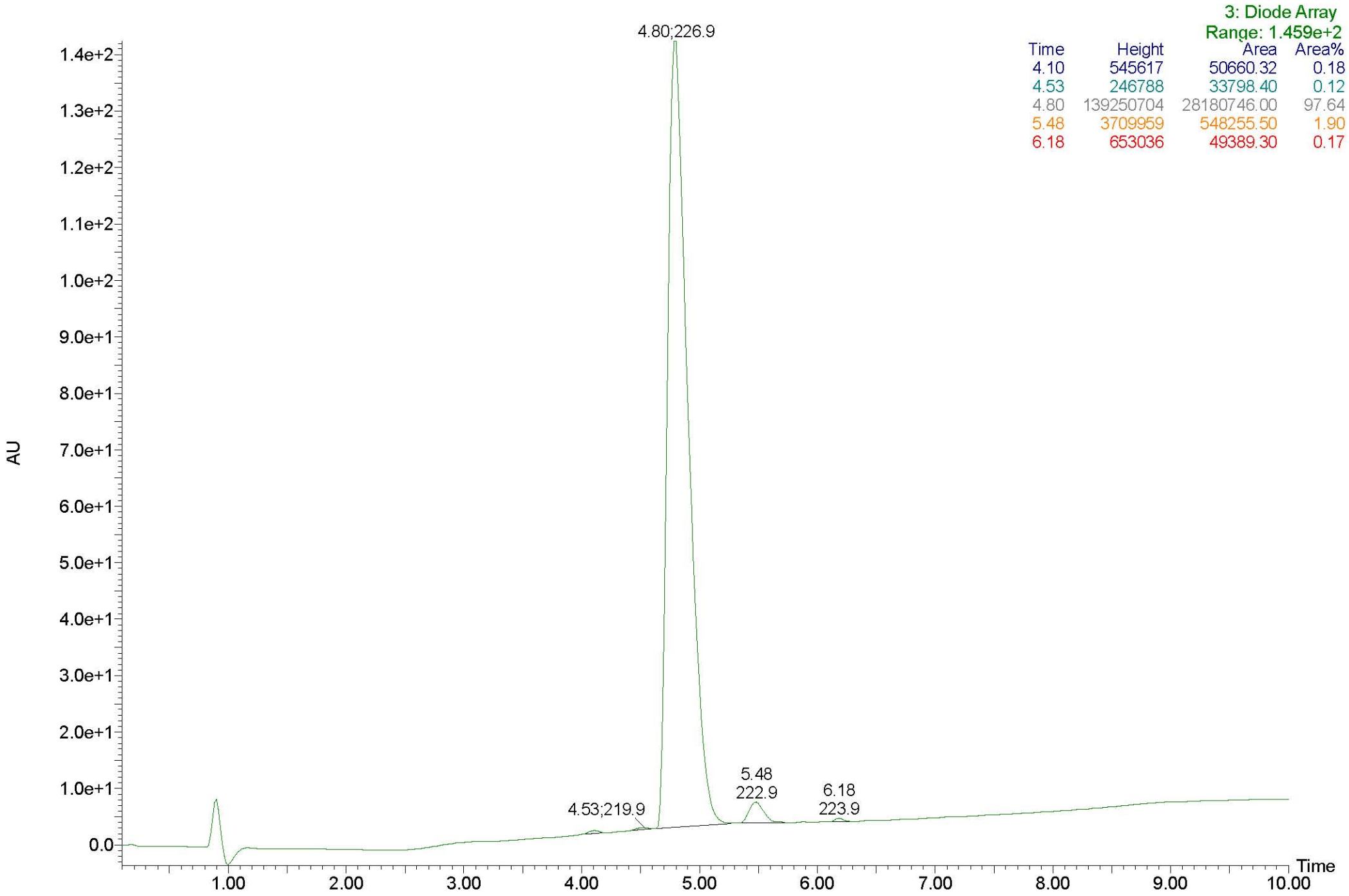

1: Scan ES+
554.7
2.37e7
7.07;554.7
100
%
0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
1: Scan ES+
TIC
4.75e7
7.07;554.7
8.33
282.7
6.37
378.6
7.46
279.6
6.13;242.6
0.03;200.3
2
3: Diode Array
Range: 1.548e+2
7.08;217.9
Time Height Area Area%
6.37 1383841 124956.32 0.96
7.08 146969280 12718217.00 97.62
7.52 1108483 77441.91 0.59
7.73 1340136 107388.44 0.82
1.5e+2
1.0e+2
5.0e+1
0.0
AU
6.37
223.9
7.73
225.9
Time

3a

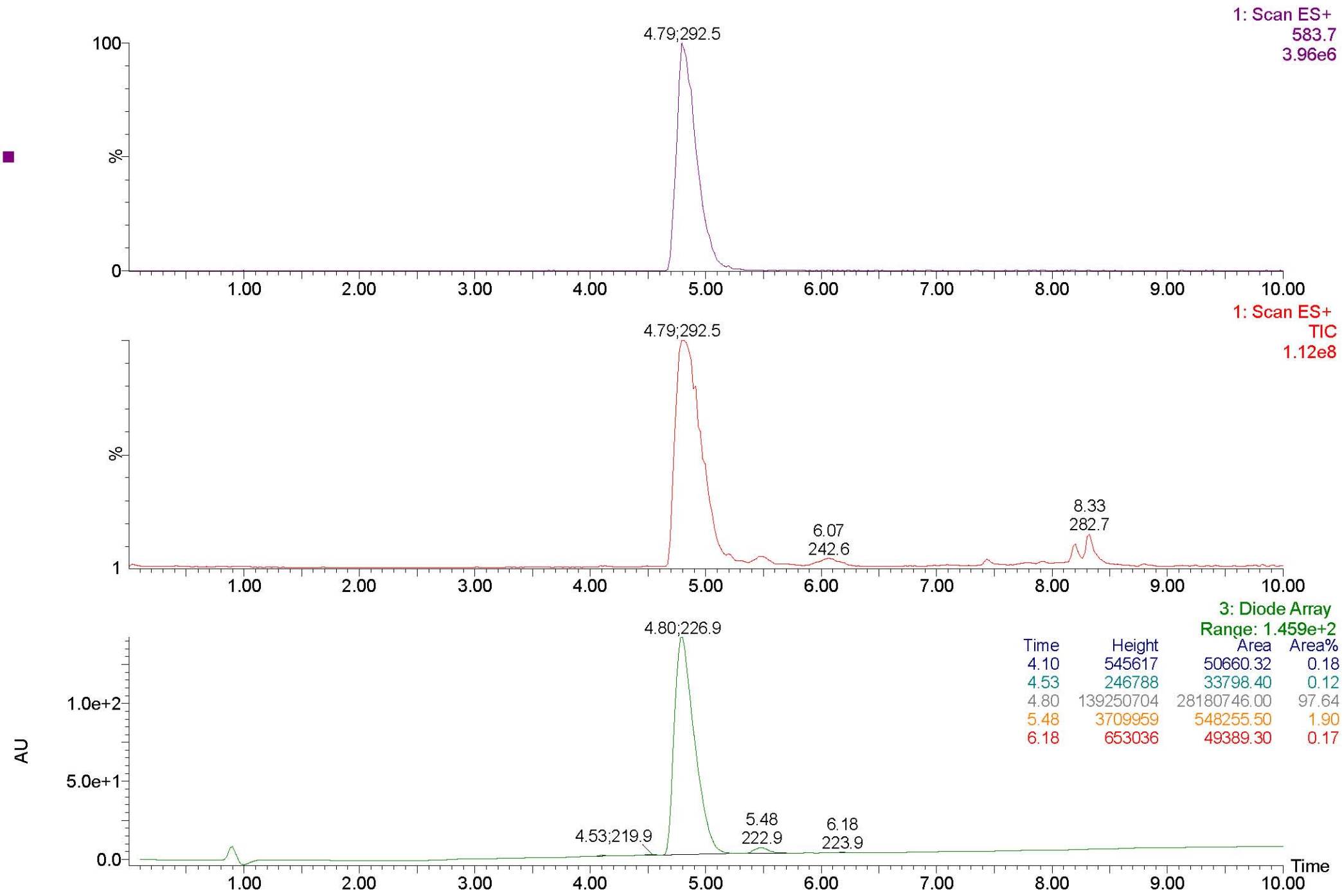

1: Scan ES+
2.36e7
100
%
0
554.7
556.7
557.8
558.7
559.7
221.3
278.9
298.5
318.9
341.5
380.7
394.5
408.6
425.9
454.9
503.2
517.6
534.7
595.7
633.8
656.6
668.8
714.9
754.0
778.2
786.8
m/z
200
225
250
275
300
325
350
375
400
425
450
475
500
525
550
575
600
625
650
675
700
725
750
775
800


3b
F
F
F
S
O
O
H
N
F
N
S
O
N
H
N
Cl
N
NH2
1.04
4.27
0.90
2.00
1.02
0.91
1.03
1.05
2.12
2.04
10.0
ppm (t1)
5.0
0.0

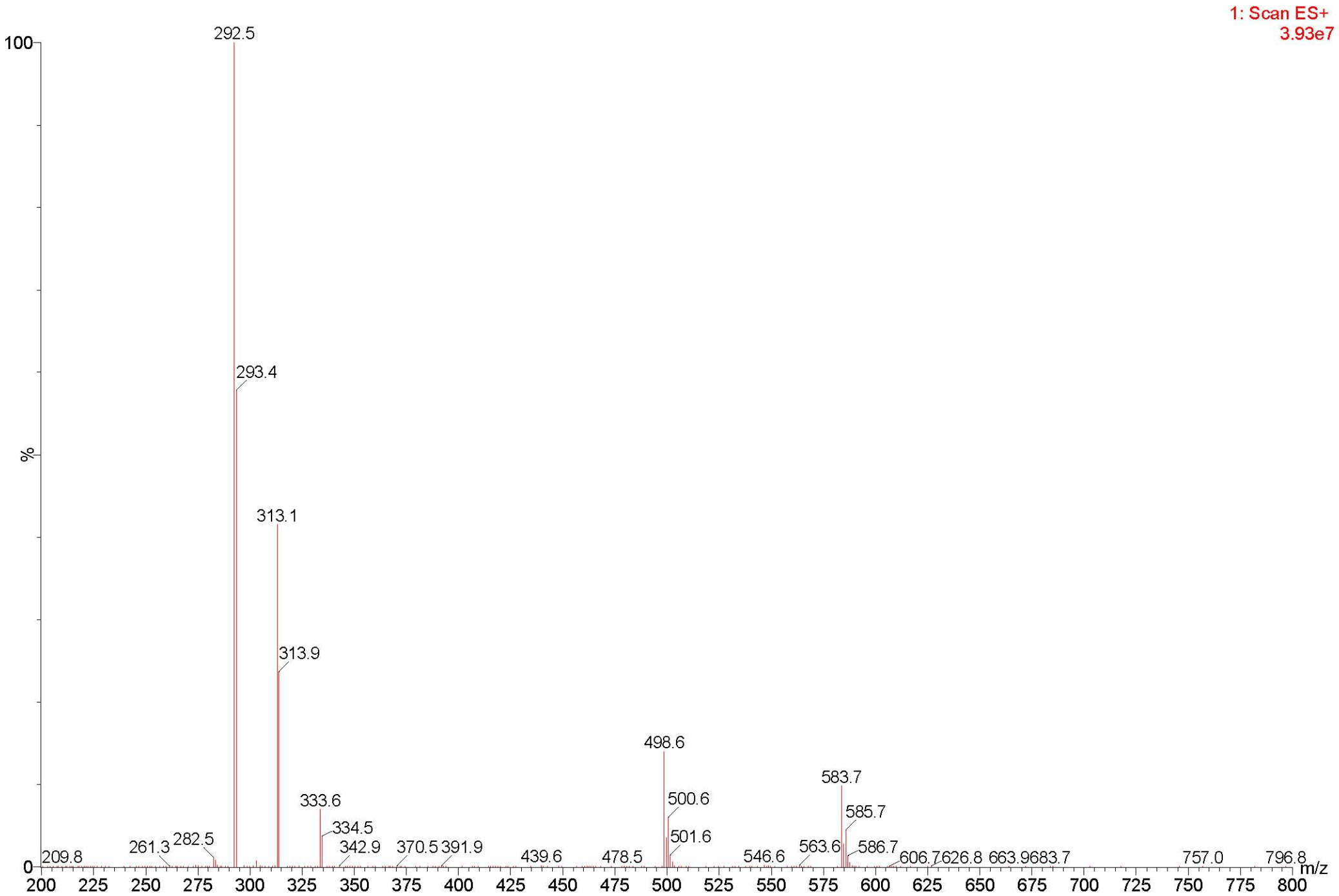
3b

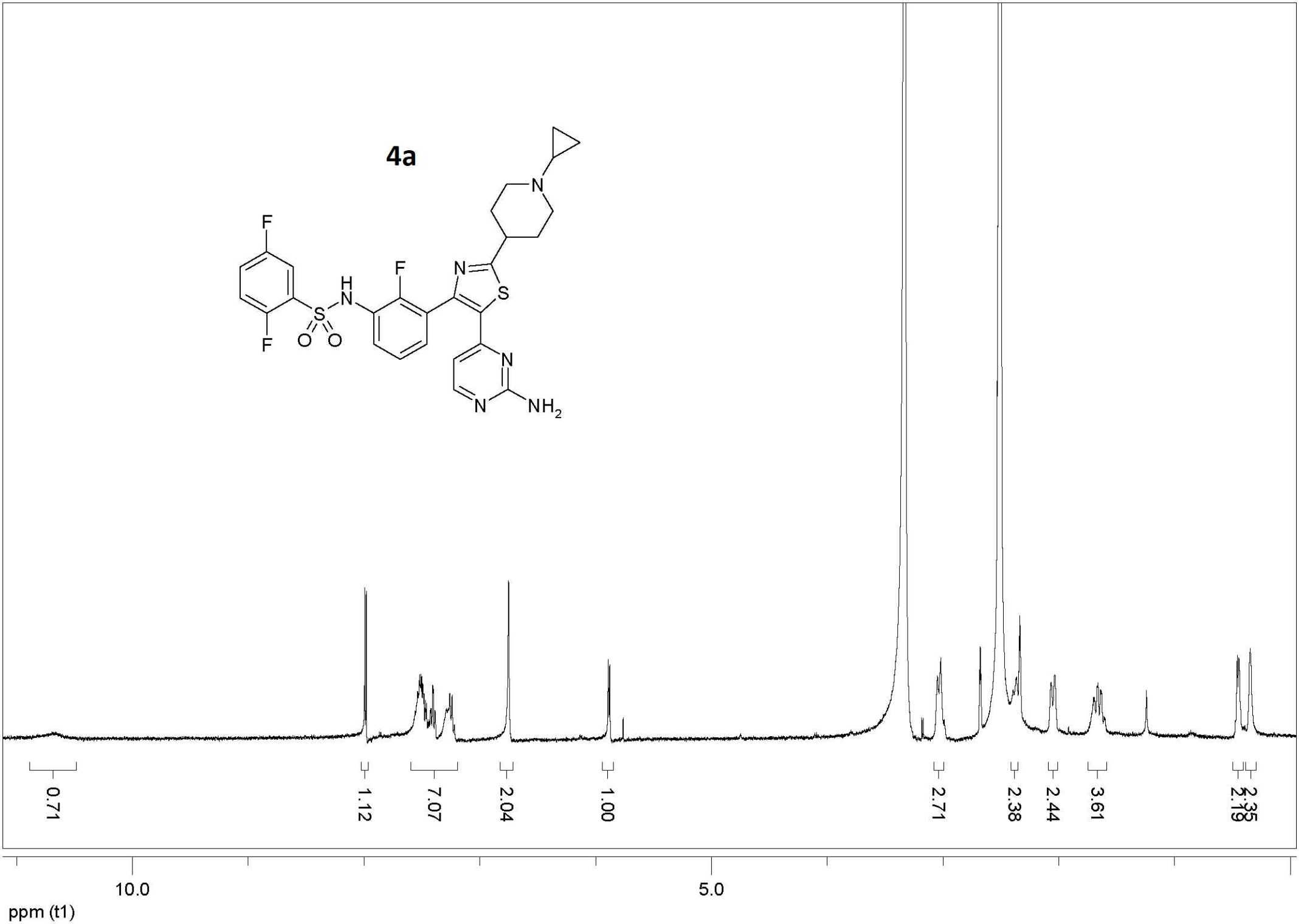

3: Diode Array
Range: 1.459e+2
Time Height Area Area%
4.10 545617 50660.32 0.18
4.53 246788 33798.40 0.12
4.80 139250704 28180746.00 97.64
5.48 3709959 548255.50 1.90
6.18 653036 49389.30 0.17
4.80;226.9
4.53;219.9
5.48
222.9
6.18
223.9
AU
Time
1.4e+2
1.3e+2
1.2e+2
1.1e+2
1.0e+2
9.0e+1
8.0e+1
7.0e+1
6.0e+1
5.0e+1
4.0e+1
3.0e+1
2.0e+1
1.0e+1
0.0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

3b

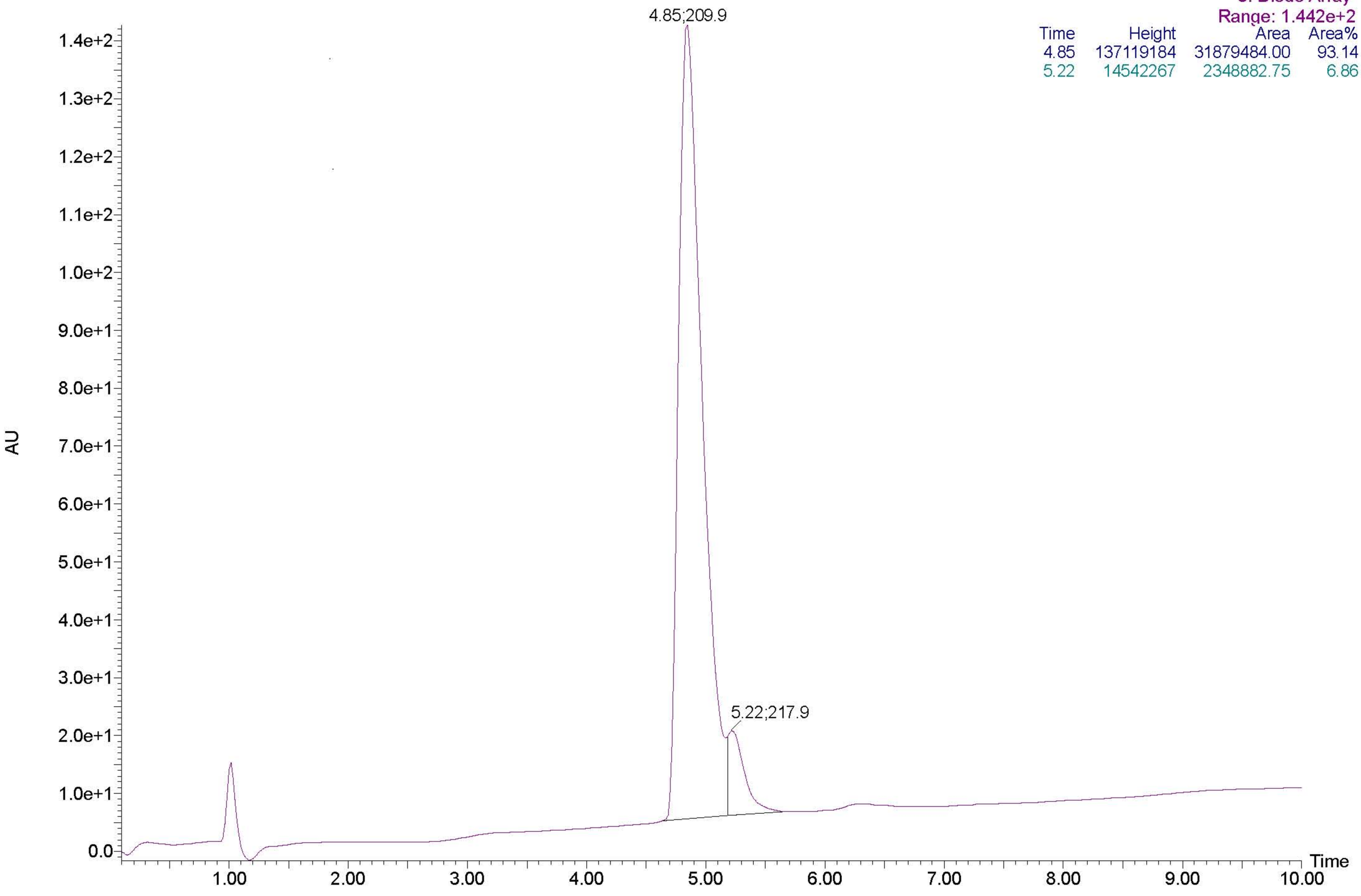

1: Scan ES+
583.7
3.96e6
4.79;292.5
100
%
0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
1: Scan ES+
TIC
1.12e8
4.79;292.5
6.07
242.6
8.33
282.7
1
3: Diode Array
Range: 1.459e+2
4.80;226.9
4.53;219.9
5.48
222.9
6.18
223.9
1.0e+2
5.0e+1
0.0
AU
Time
Time Height Area Area%
4.10 545617 50660.32 0.18
4.53 246788 33798.40 0.12
4.80 139250704 28180746.00 97.64
5.48 3709959 548255.50 1.90
6.18 653036 49389.30 0.17

3b

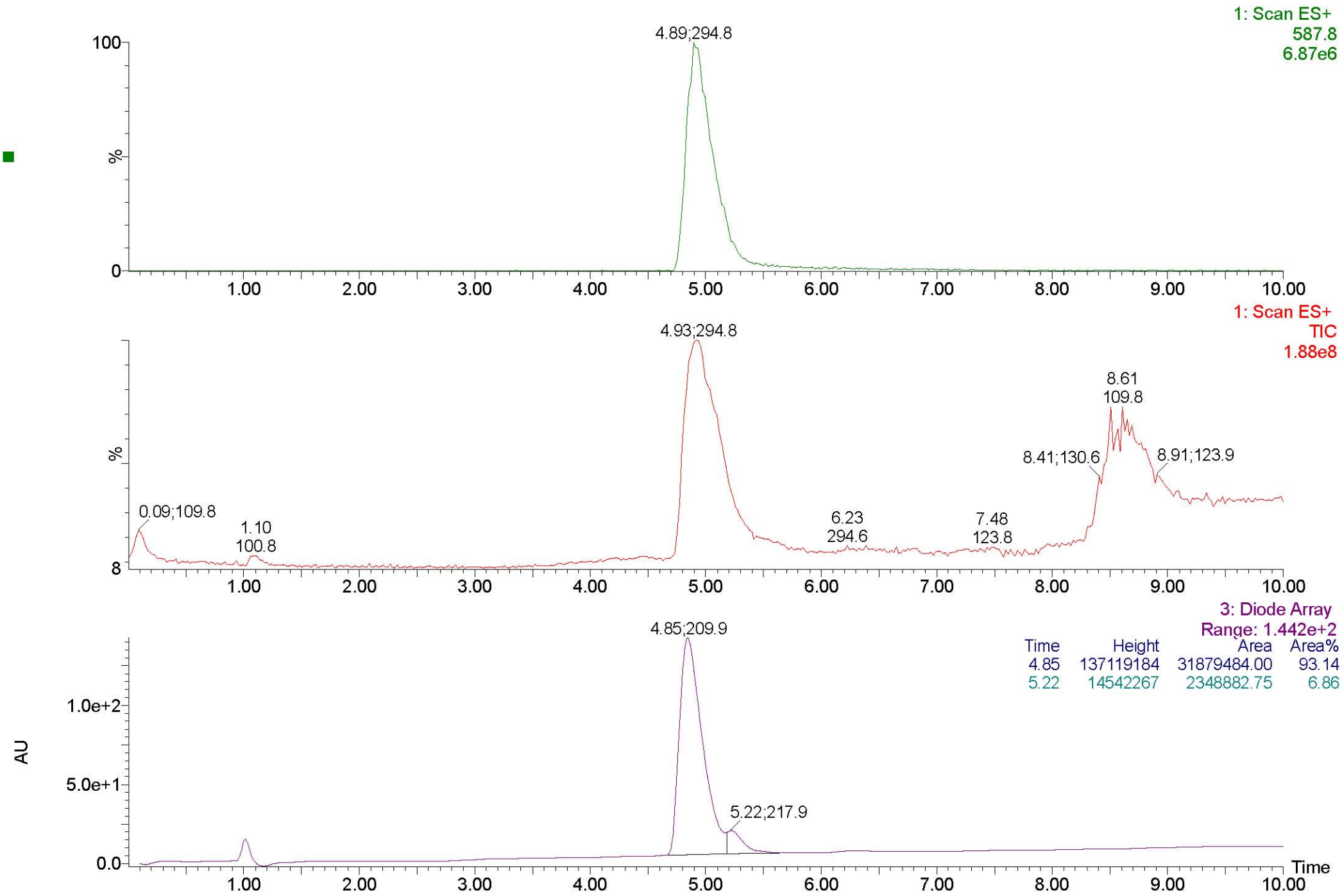

1: Scan ES+
3.93e7
100
%
0
292.5
293.4
313.1
313.9
333.6
334.5
342.9
370.5
391.9
439.6
478.5
498.6
500.6
501.6
546.6
563.6
583.7
585.7
586.7
606.7
626.8
663.9
683.7
757.0
796.8
209.8
261.3
282.5
m/z
200
225
250
275
300
325
350
375
400
425
450
475
500
525
550
575
600
625
650
675
700
725
750
775
800


4a
F
F
F
H
N
S
O
O
N
S
N
N
N
NH2
0.71
1.12
7.07
2.04
1.00
2.71
2.38
2.44
3.61
2.19
2.35
10.0
5.0
ppm (t1)

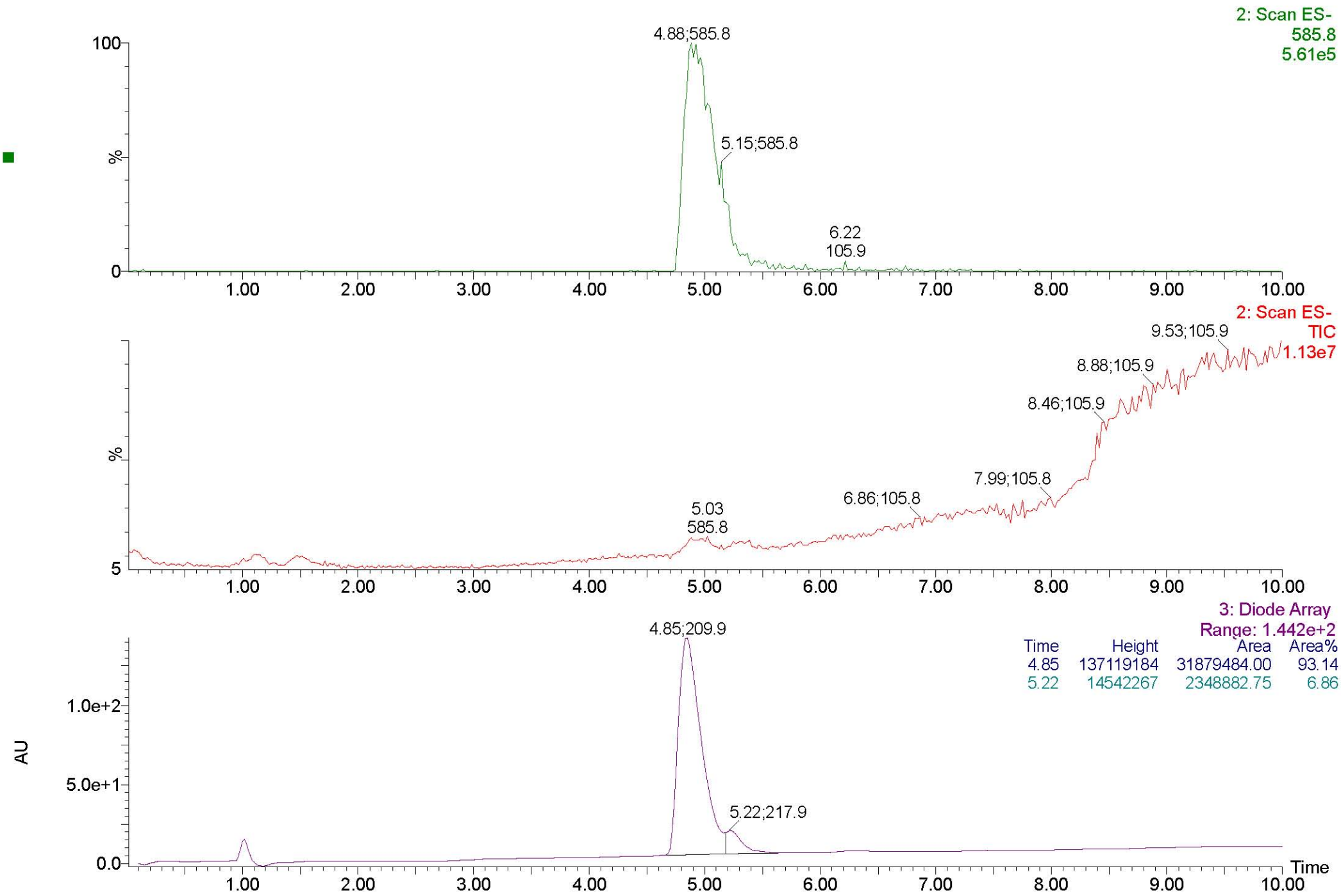
4a


3: Diode Array
Range: 1.442e+2
Time Height Area Area%
4.85 137119184 31879484.00 93.14
5.22 14542267 2348882.75 6.86
4.85;209.9
5.22;217.9
AU
Time
1.4e+2
1.3e+2
1.2e+2
1.1e+2
1.0e+2
9.0e+1
8.0e+1
7.0e+1
6.0e+1
5.0e+1
4.0e+1
3.0e+1
2.0e+1
1.0e+1
0.0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

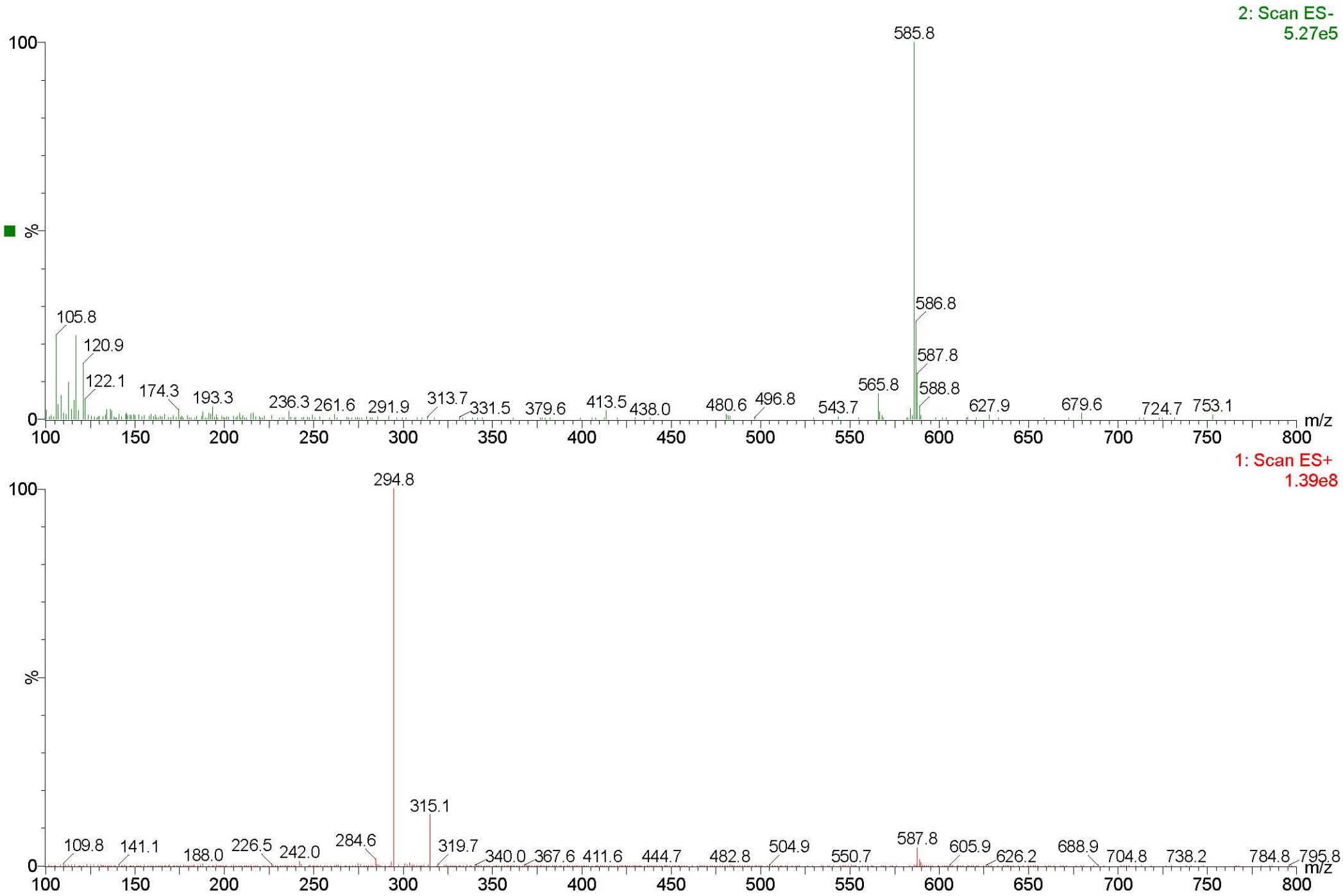
4a

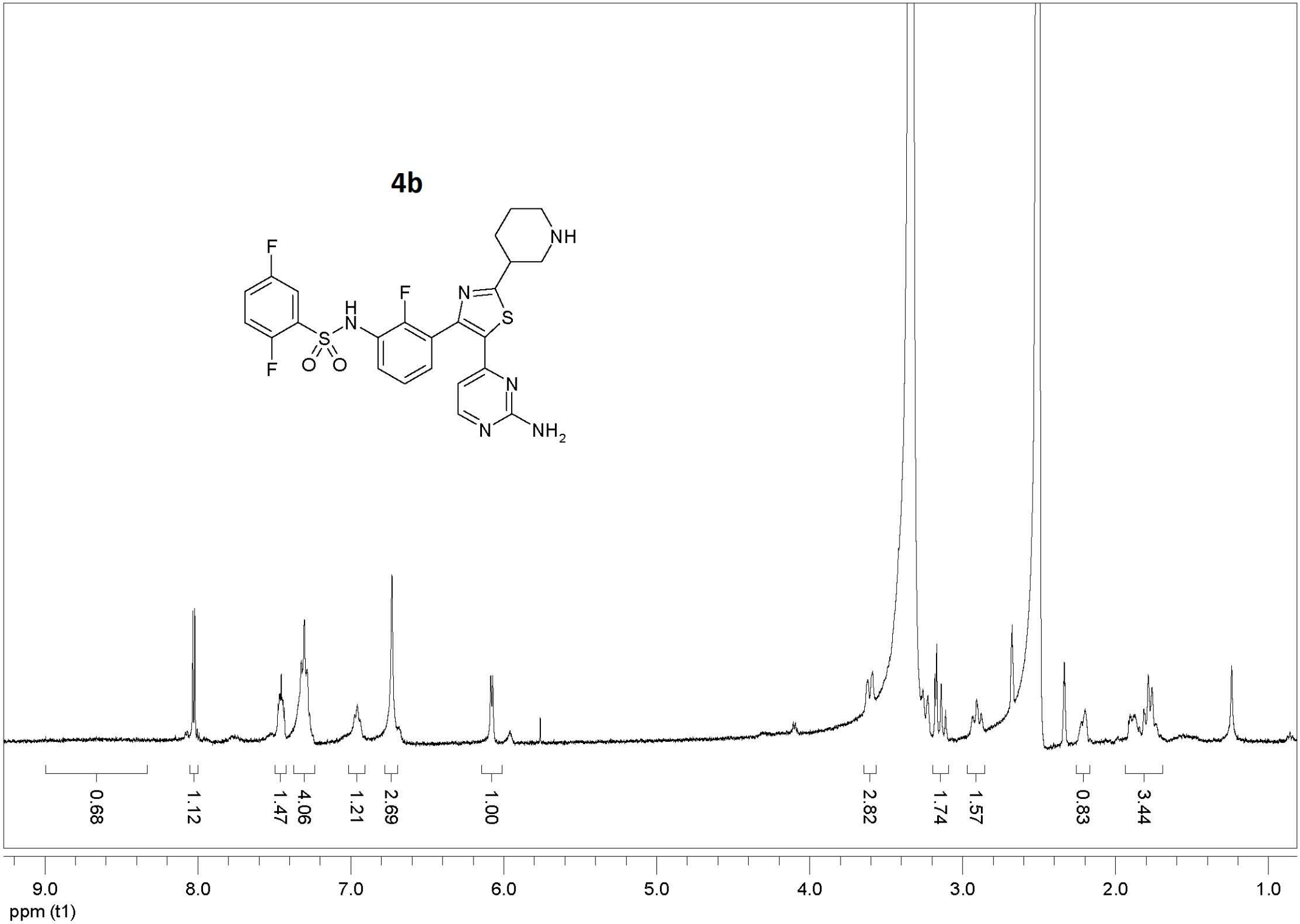

1: Scan ES+
587.8
6.87e6
4.89;294.8
100
%
0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
1: Scan ES+
TIC
1.88e8
4.93;294.8
8.61
109.8
8.41;130.6
8.91;123.9
0.09;109.8
1.10
100.8
6.23
294.6
7.48
123.8
8
3: Diode Array
Range: 1.442e+2
Time Height Area Area%
4.85 137119184 31879484.00 93.14
5.22 14542267 2348882.75 6.86
4.85;209.9
5.22;217.9
AU
1.0e+2
5.0e+1
0.0
Time

4a

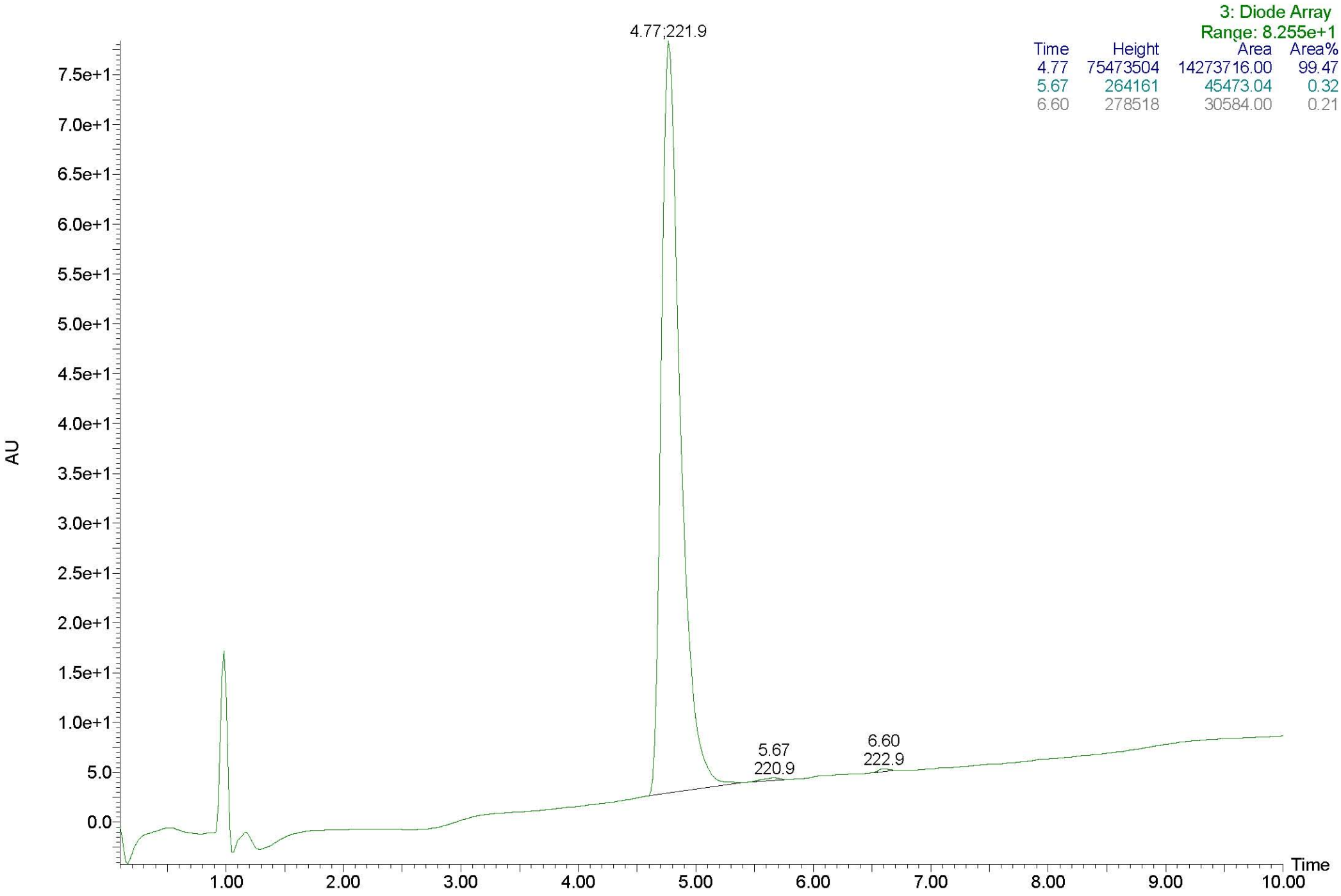

2: Scan ES-
585.8
5.61e5
100
%
0
4.88;585.8
5.15;585.8
6.22
105.9
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
2: Scan ES-
TIC
1.13e7
9.53;105.9
8.88;105.9
8.46;105.9
7.99;105.8
6.86;105.8
5.03
585.8
5
3: Diode Array
Range: 1.442e+2
Time Height Area Area%
4.85 137119184 31879484.00 93.14
5.22 14542267 2348882.75 6.86
AU
1.0e+2
5.0e+1
0.0
4.85;209.9
5.22;217.9
Time

4a

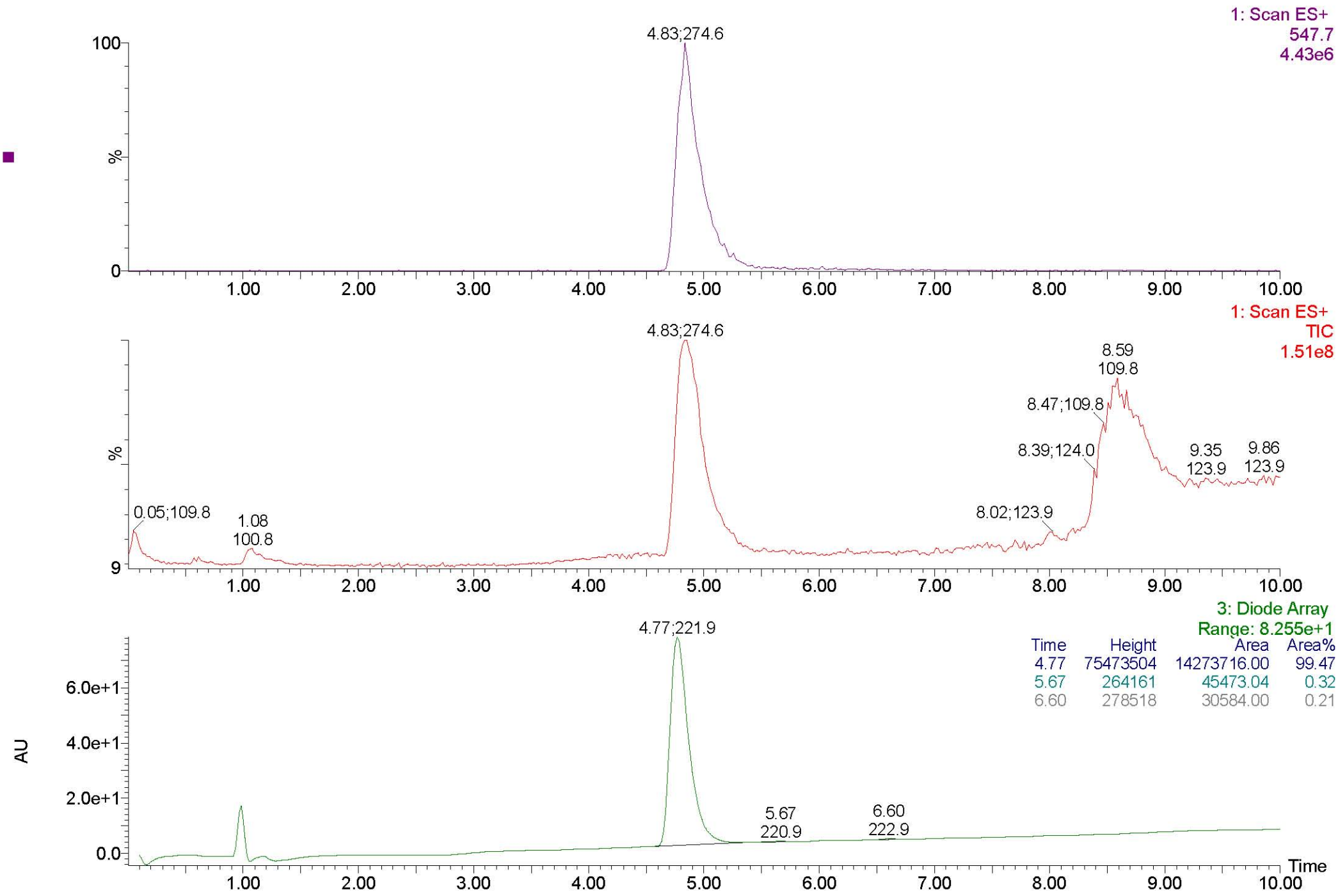

2: Scan ES-
5.27e5
100
%
0
585.8
586.8
587.8
588.8
105.8
120.9
122.1
174.3
193.3
236.3
261.6
291.9
313.7
331.5
379.6
413.5
438.0
480.6
496.8
543.7
565.8
627.9
679.6
724.7
753.1
m/z
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
1: Scan ES+
1.39e8
100
%
0
294.8
315.1
109.8
141.1
188.0
226.5
242.0
284.6
319.7
340.0
367.6
411.6
444.7
482.8
504.9
550.7
587.8
605.9
626.2
688.9
704.8
738.2
784.8
795.8
m/z
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800


4b
F
F
S
O
O
H
N
F
N
S
NH
N
N
NH2
0.68
1.12
1.47
4.06
1.21
2.69
1.00
2.82
1.74
1.57
0.83
3.44
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
ppm (t1)

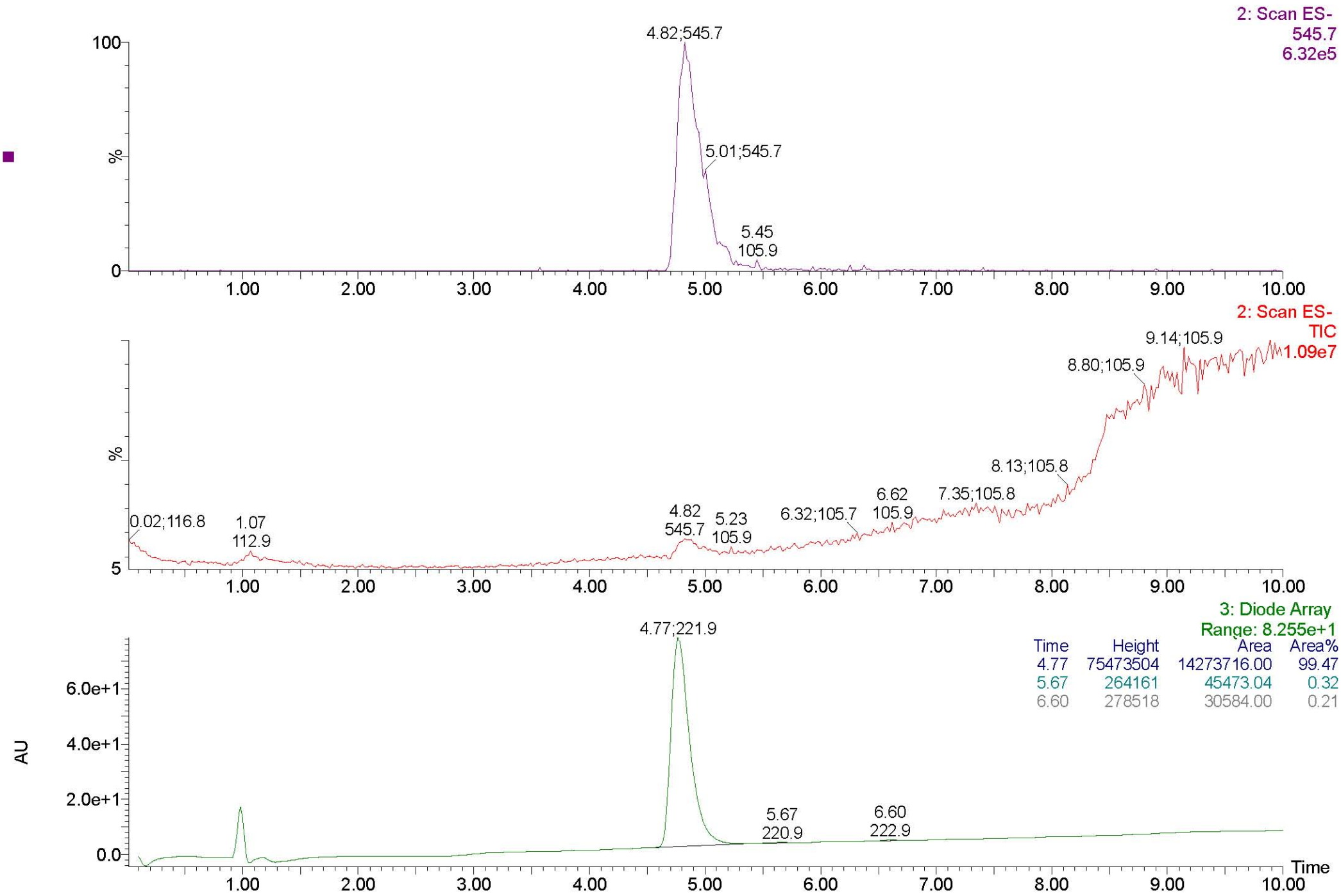
4b


3: Diode Array
Range: 8.255e+1
Time Height Area Area%
4.77 75473504 14273716.00 99.47
5.67 264161 45473.04 0.32
6.60 278518 30584.00 0.21
4.77;221.9
5.67
220.9
6.60
222.9
AU
Time
7.5e+1
7.0e+1
6.5e+1
6.0e+1
5.5e+1
5.0e+1
4.5e+1
4.0e+1
3.5e+1
3.0e+1
2.5e+1
2.0e+1
1.5e+1
1.0e+1
5.0
0.0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

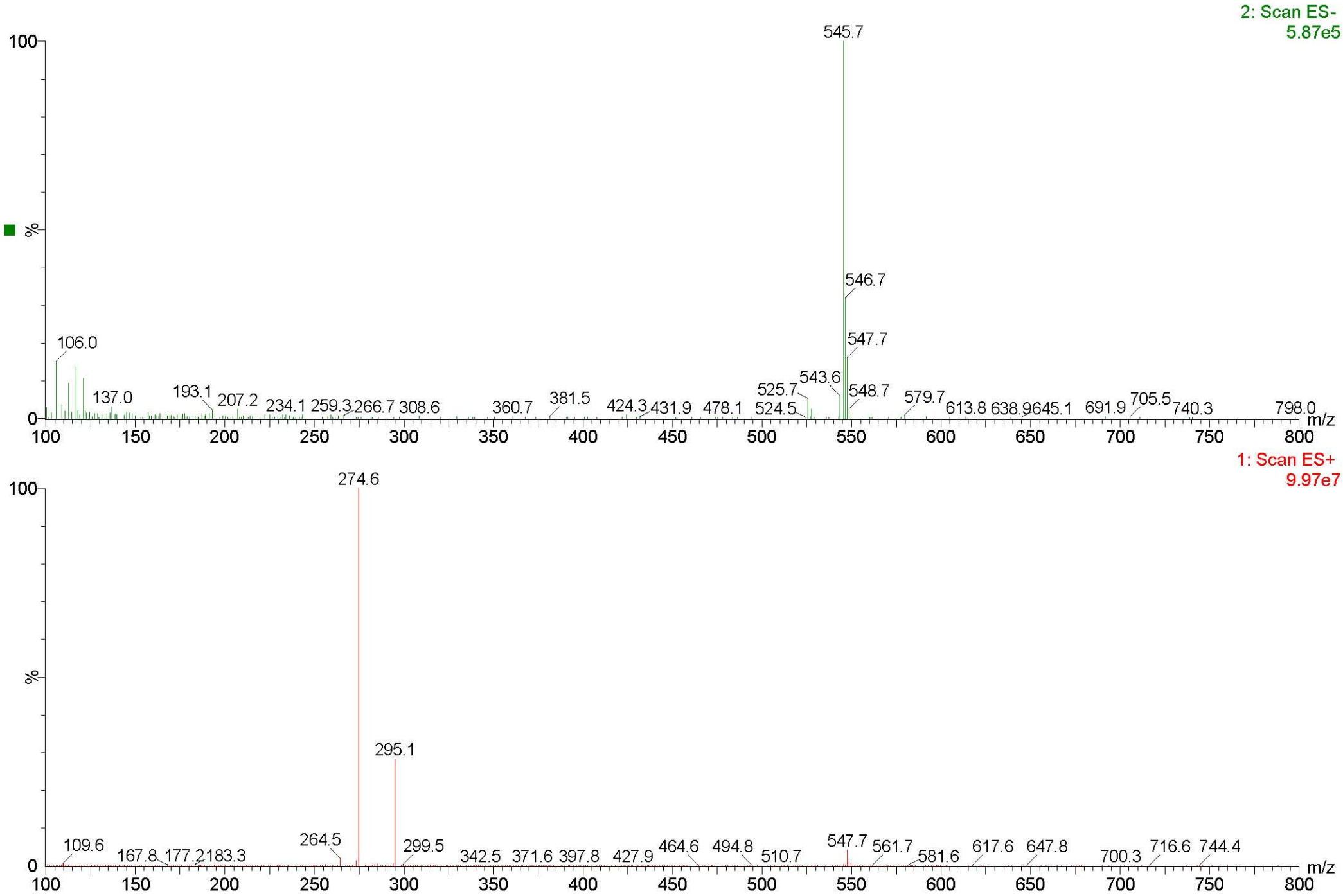
4b

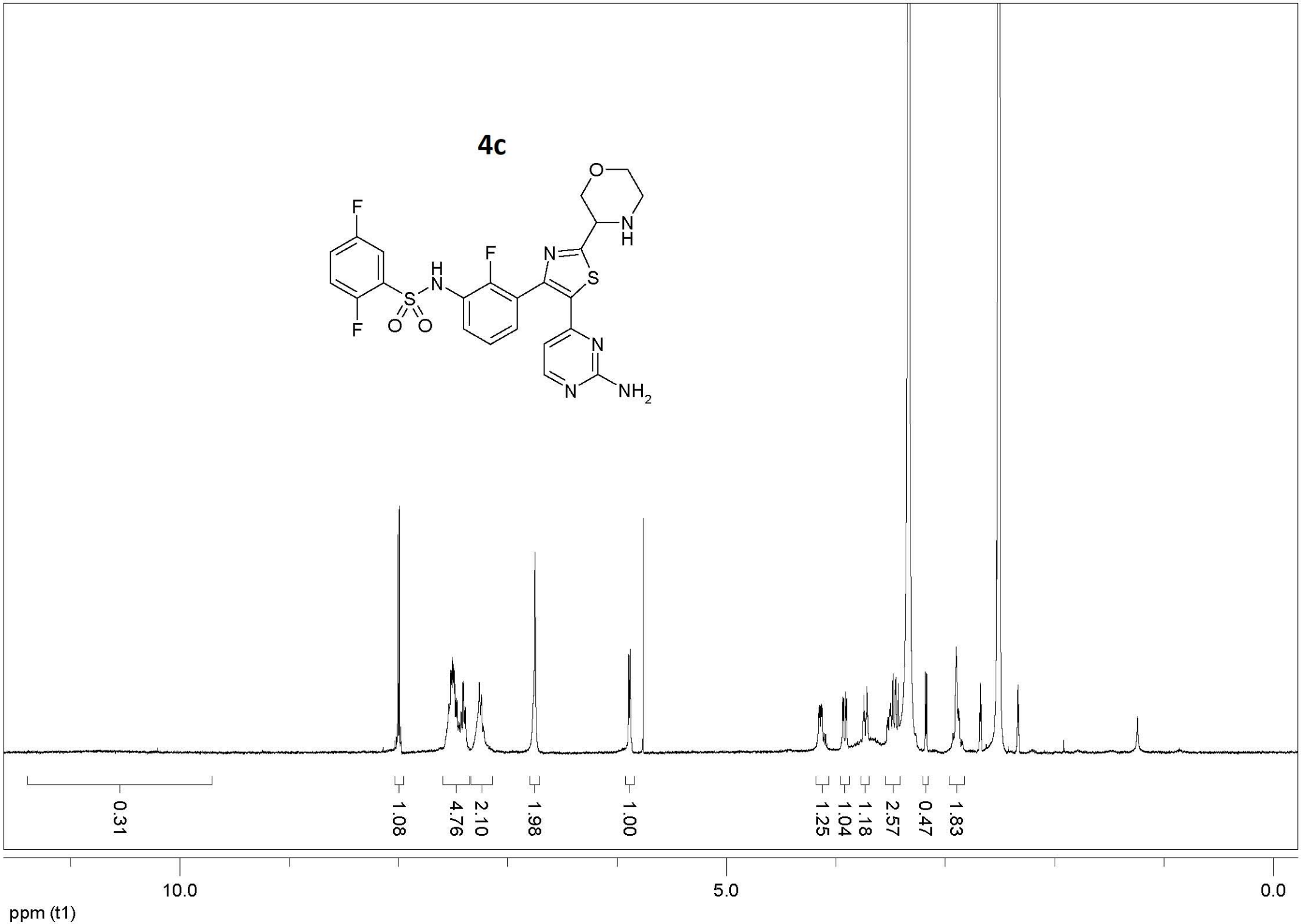

1: Scan ES+
547.7
4.43e6
100
%
0
4.83;274.6
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
1: Scan ES+
TIC
1.51e8
%
9
4.83;274.6
8.59
109.8
8.47;109.8
8.39;124.0
9.35
123.9
9.86
123.9
0.05;109.8
1.08
100.8
8.02;123.9
3: Diode Array
Range: 8.255e+1
Time Height Area Area%
4.77 75473504 14273716.00 99.47
5.67 264161 45473.04 0.32
6.60 278518 30584.00 0.21
AU
6.0e+1
4.0e+1
2.0e+1
0.0
4.77;221.9
5.67
220.9
6.60
222.9
Time

4b

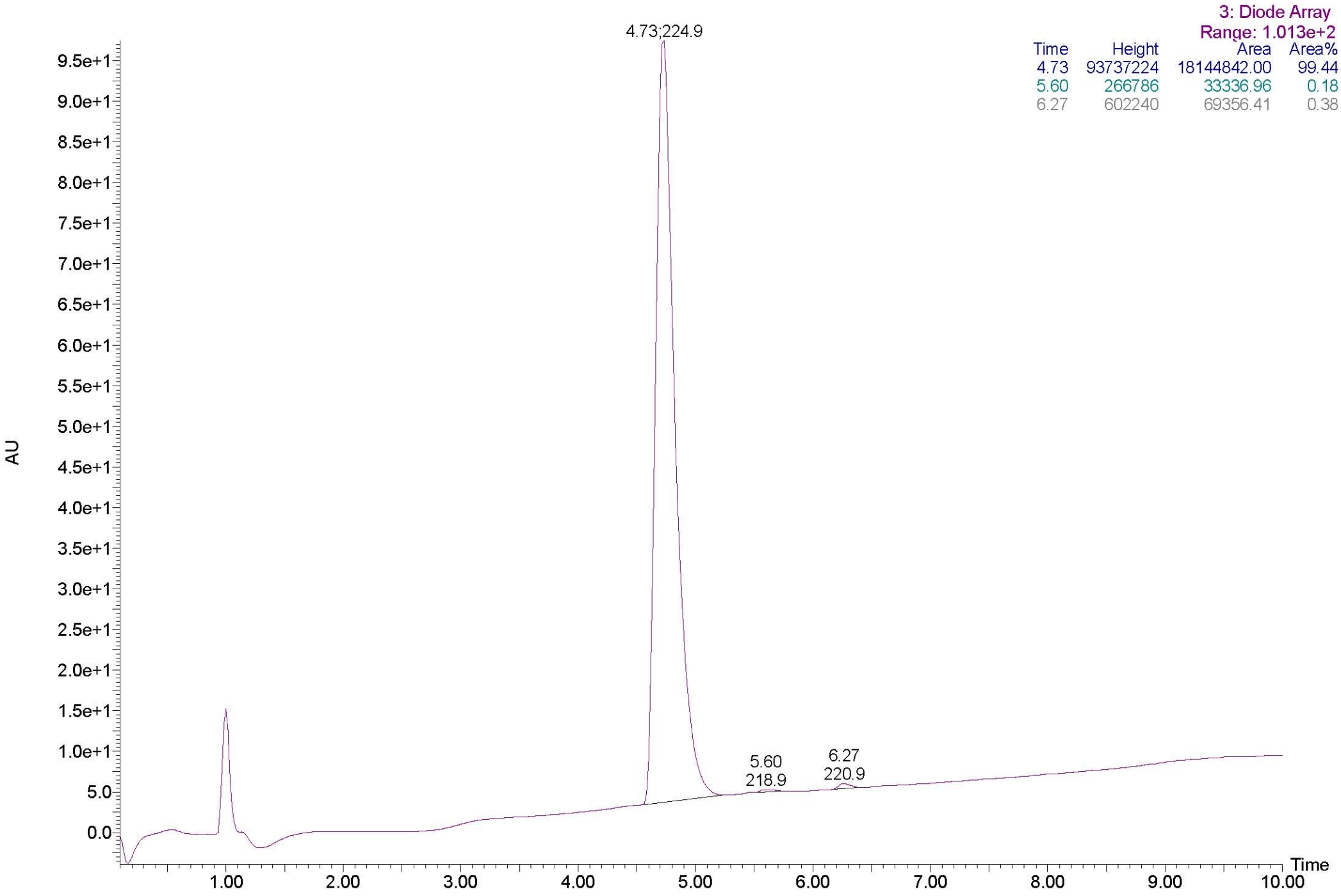

2: Scan ES-
545.7
6.32e5
4.82;545.7
5.01;545.7
5.45
105.9
100
%
0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
2: Scan ES-
TIC
1.09e7
9.14;105.9
8.80;105.9
8.13;105.8
7.35;105.8
6.62
105.9
6.32;105.7
4.82
545.7
5.23
105.9
0.02;116.8
1.07
112.9
5
3: Diode Array
Range: 8.255e+1
Time Height Area Area%
4.77 75473504 14273716.00 99.47
5.67 264161 45473.04 0.32
6.60 278518 30584.00 0.21
4.77;221.9
5.67
220.9
6.60
222.9
AU
6.0e+1
4.0e+1
2.0e+1
0.0
Time

4b

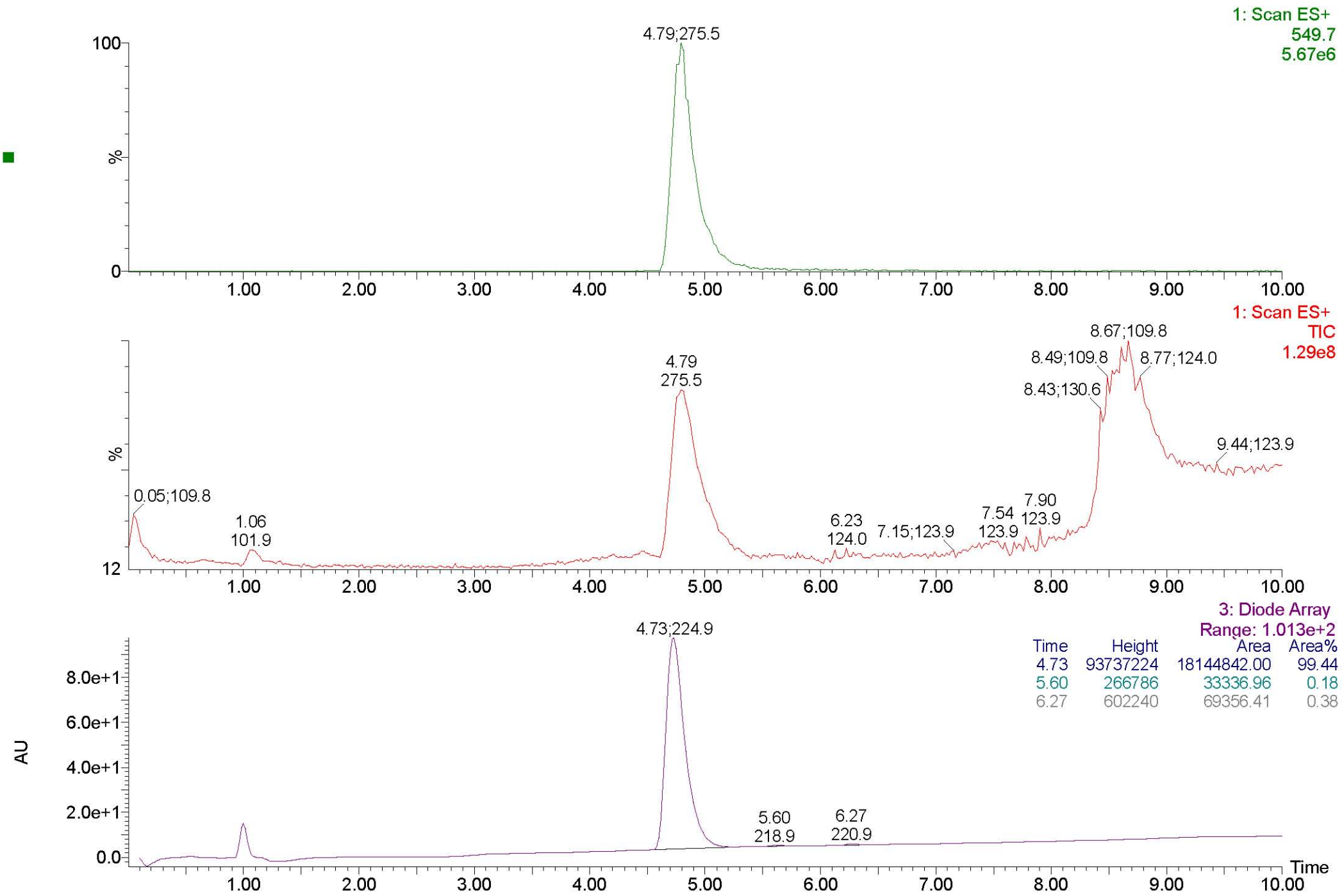

2: Scan ES-
5.87e5
100
%
0
545.7
546.7
547.7
543.6
548.7
106.0
137.0
193.1
207.2
234.1
259.3
266.7
308.6
360.7
381.5
424.3
431.9
478.1
525.7
524.5
579.7
613.8
638.9
645.1
691.9
705.5
740.3
798.0
m/z
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
1: Scan ES+
9.97e7
100
%
0
274.6
295.1
109.6
167.8
177.2
183.3
264.5
299.5
342.5
371.6
397.8
427.9
464.6
494.8
510.7
547.7
561.7
581.6
617.6
647.8
700.3
716.6
744.4
m/z
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800


4c
F
F
S
O
O
H
N
F
N
S
O
N
H
N
N
NH2
0.31
1.08
4.76
2.10
1.98
1.00
1.25
1.04
1.18
2.57
0.47
1.83
10.0
5.0
0.0
ppm (t1)

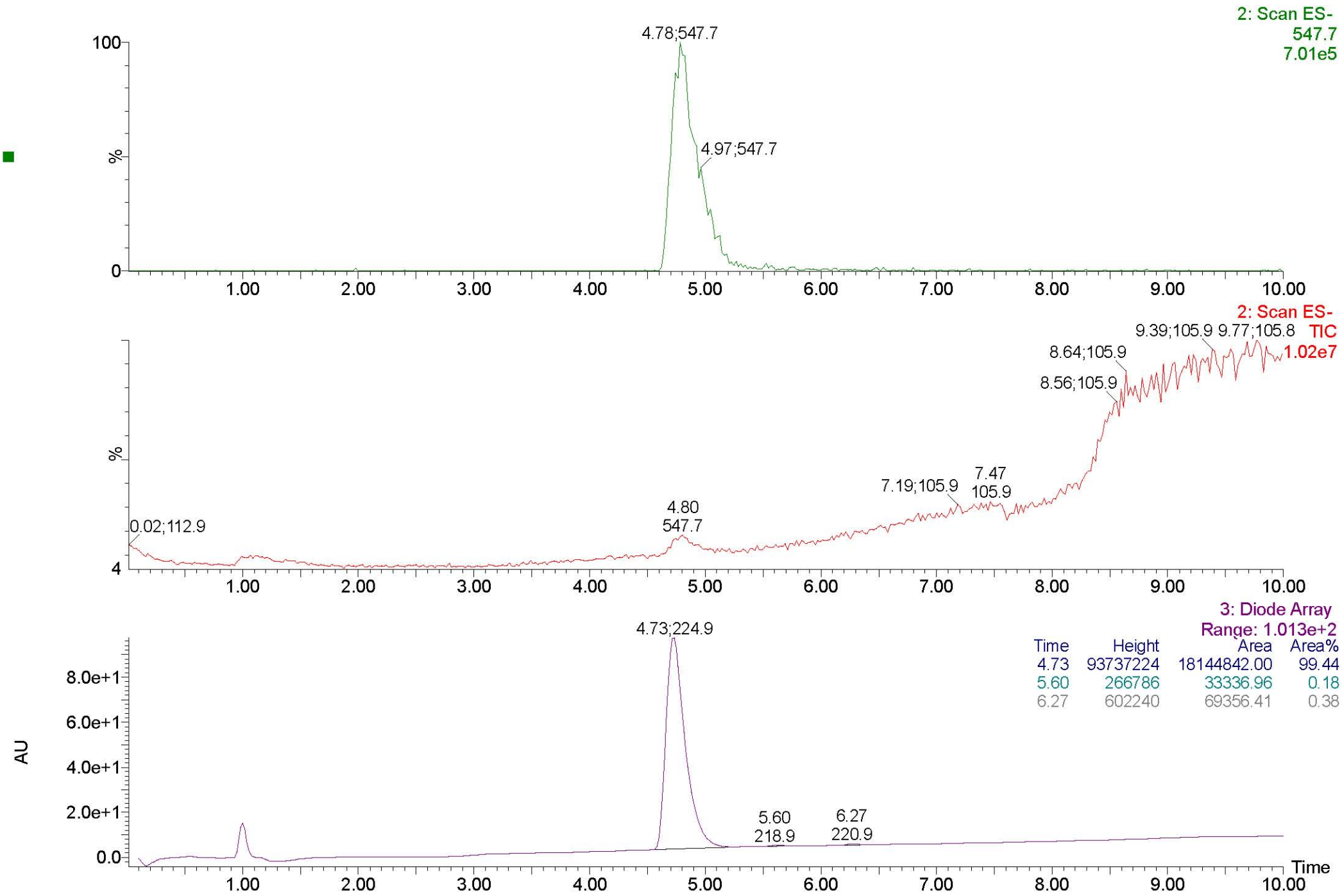
4c


3: Diode Array
Range: 1.013e+2
Time Height Area Area%
4.73 93737224 18144842.00 99.44
5.60 266786 33336.96 0.18
6.27 602240 69356.41 0.38
4.73;224.9
5.60
218.9
6.27
220.9
AU
Time
9.5e+1
9.0e+1
8.5e+1
8.0e+1
7.5e+1
7.0e+1
6.5e+1
6.0e+1
5.5e+1
5.0e+1
4.5e+1
4.0e+1
3.5e+1
3.0e+1
2.5e+1
2.0e+1
1.5e+1
1.0e+1
5.0
0.0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

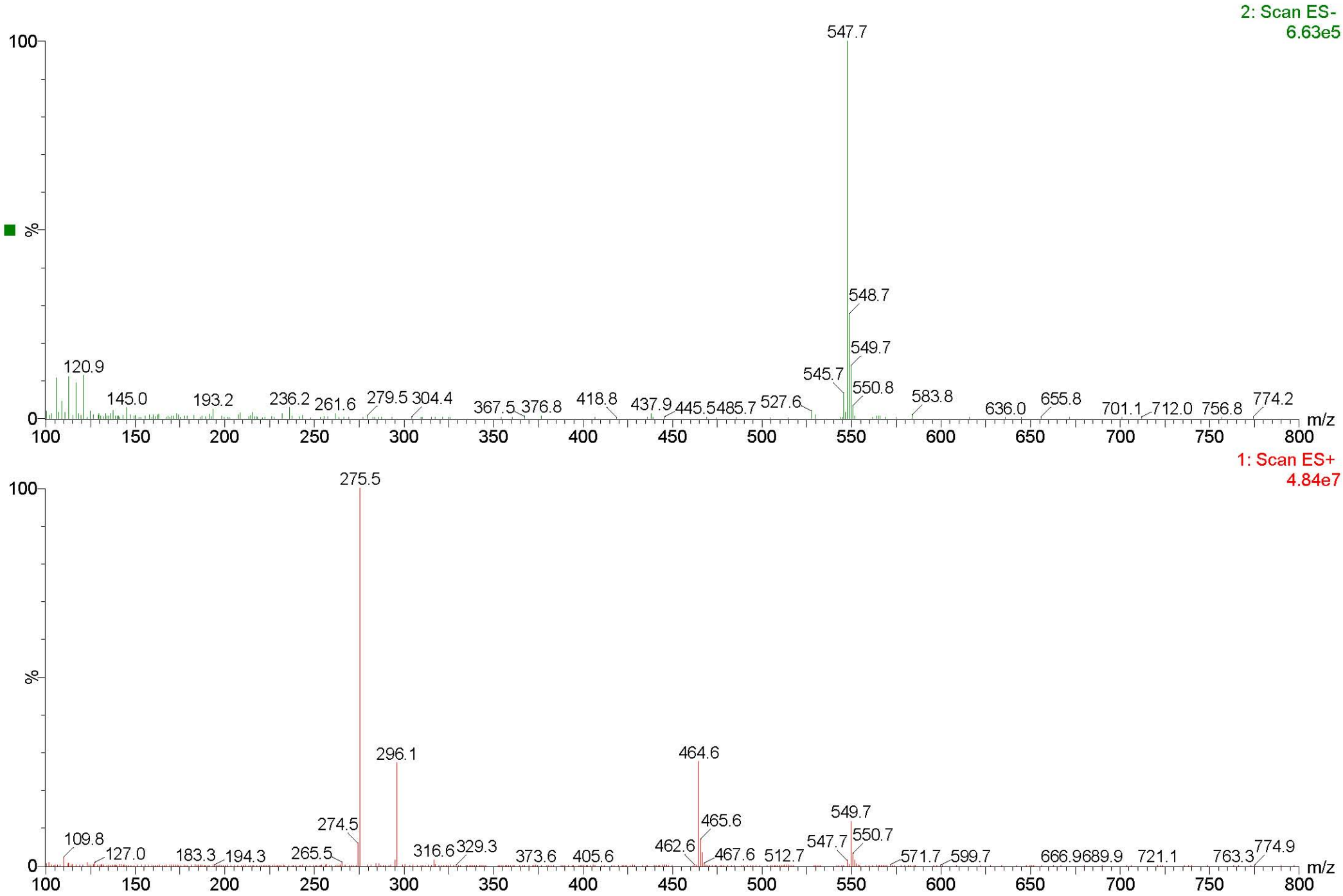
4c

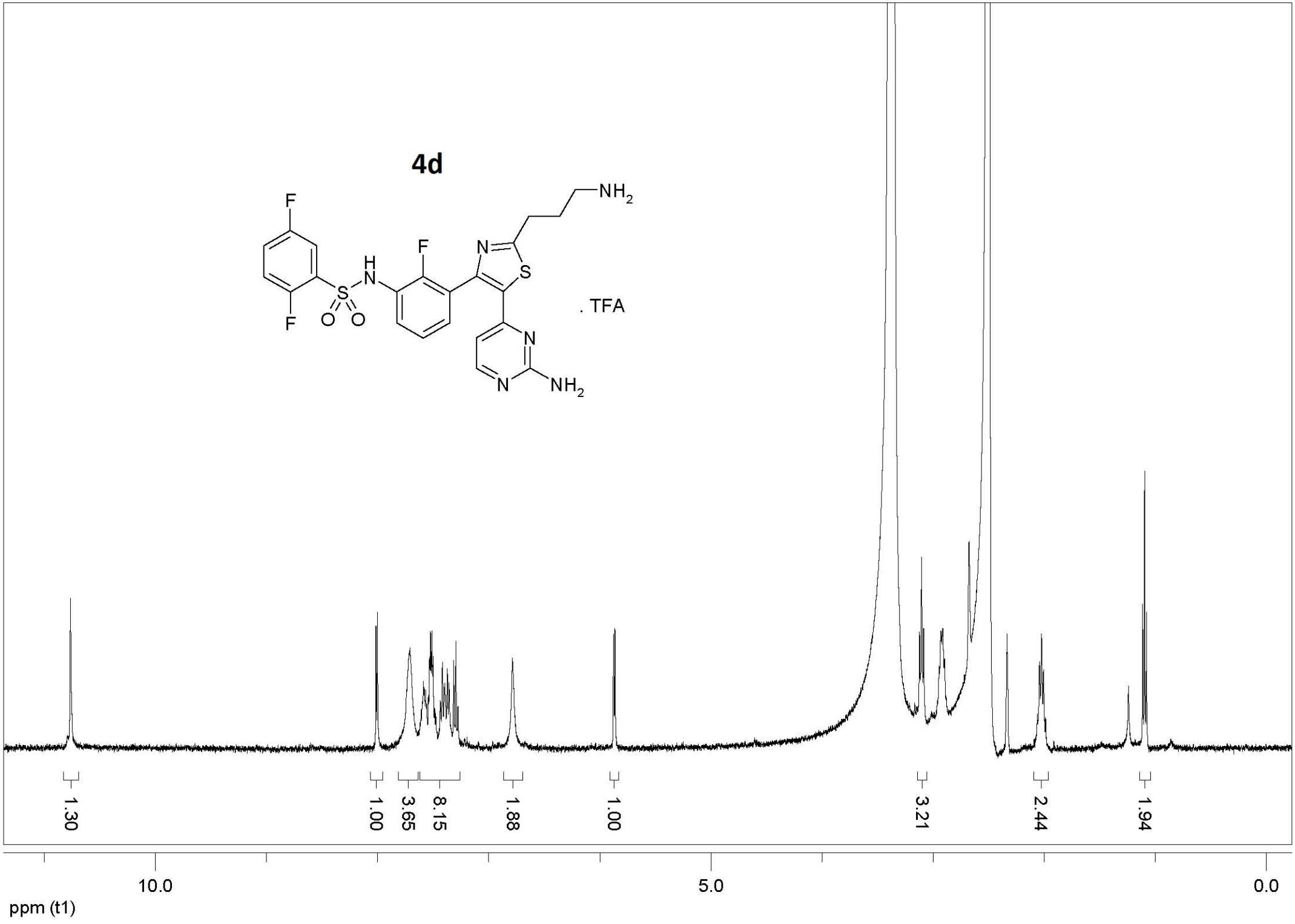

1: Scan ES+
549.7
5.67e6
4.79;275.5
100
%
0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
1: Scan ES+
TIC
1.29e8
8.67;109.8
8.49;109.8
8.77;124.0
8.43;130.6
4.79
275.5
9.44;123.9
0.05;109.8
1.06
101.9
6.23
124.0
7.15;123.9
7.54
123.9
7.90
123.9
12
3: Diode Array
Range: 1.013e+2
4.73;224.9
Time Height Area Area%
4.73 93737224 18144842.00 99.44
5.60 266786 33336.96 0.18
6.27 602240 69356.41 0.38
AU
8.0e+1
6.0e+1
4.0e+1
2.0e+1
0.0
5.60
218.9
6.27
220.9
Time

4c

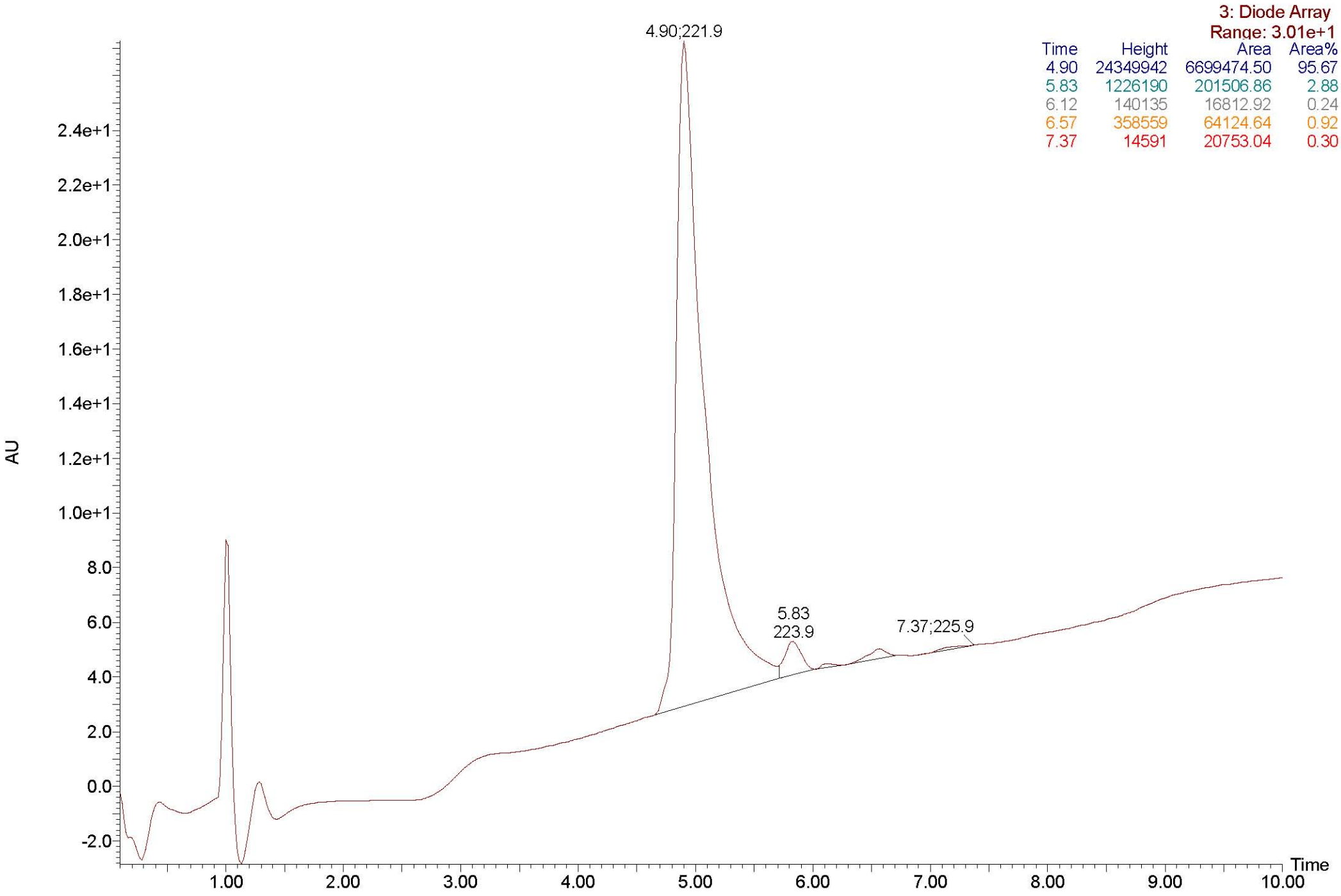

2: Scan ES-
547.7
7.01e5
4.78;547.7
4.97;547.7
100
%
0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
2: Scan ES-
TIC
1.02e7
9.39;105.9
9.77;105.8
8.64;105.9
8.56;105.9
7.19;105.9
7.47
105.9
4.80
547.7
0.02;112.9
%
4
3: Diode Array
Range: 1.013e+2
Time Height Area Area%
4.73 93737224 18144842.00 99.44
5.60 266786 33336.96 0.18
6.27 602240 69356.41 0.38
4.73;224.9
5.60
218.9
6.27
220.9
AU
8.0e+1
6.0e+1
4.0e+1
2.0e+1
0.0
Time

4c

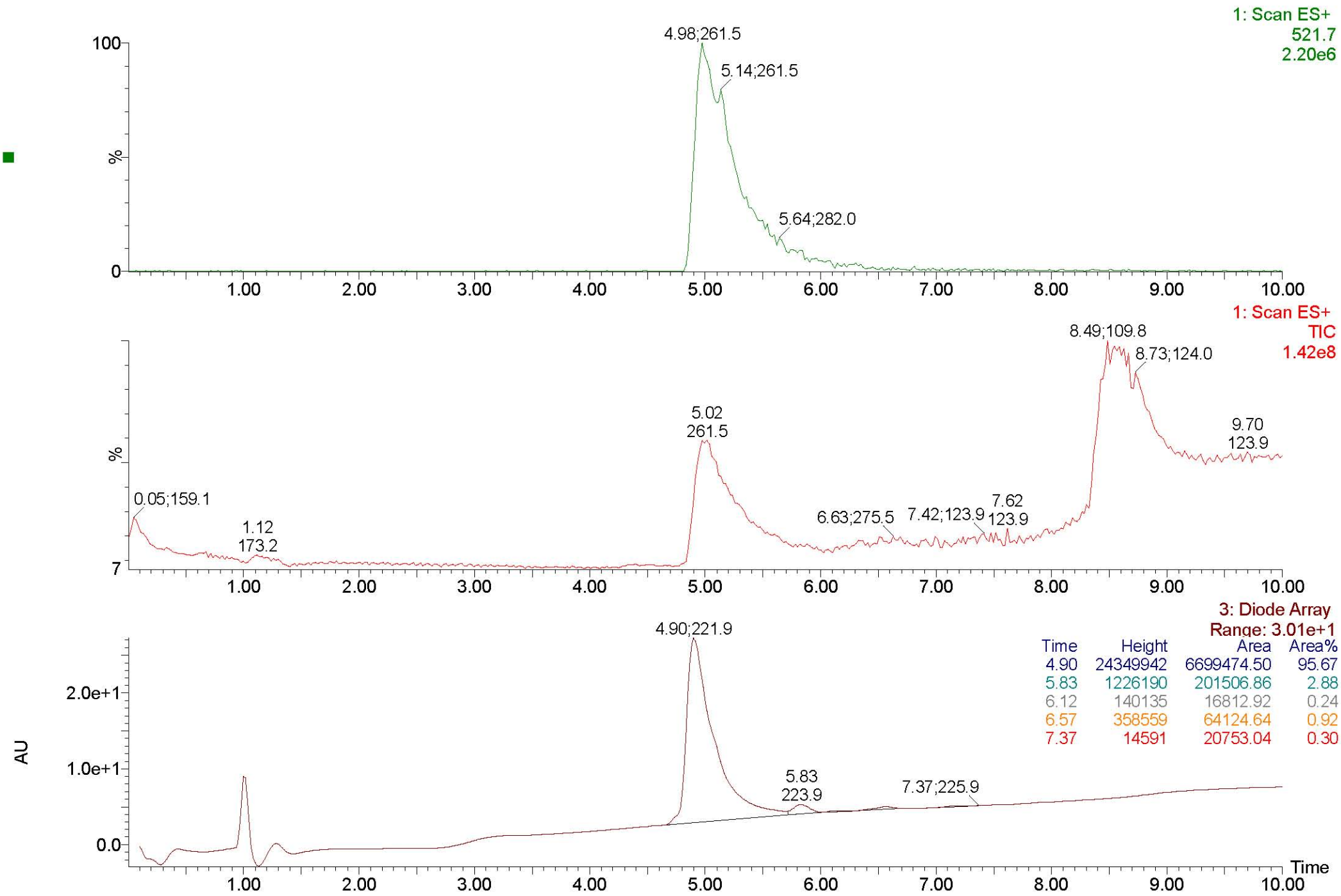

2: Scan ES-
6.63e5
100
%
0
547.7
548.7
549.7
545.7
550.8
120.9
145.0
193.2
236.2
261.6
279.5
304.4
367.5
376.8
418.8
437.9
445.5
485.7
527.6
583.8
636.0
655.8
701.1
712.0
756.8
774.2
m/z
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
1: Scan ES+
4.84e7
100
%
0
275.5
296.1
464.6
274.5
465.6
549.7
550.7
547.7
109.8
127.0
183.3
194.3
265.5
316.6
329.3
373.6
405.6
462.6
467.6
512.7
571.7
599.7
666.9
689.9
721.1
763.3
774.9
m/z
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800


4d
F
F
F
H
N
S
O
O
N
S
N
NH2
. TFA
N
N
NH2
1.30
1.00
3.65
8.15
1.88
1.00
3.21
2.44
1.94
10.0
5.0
0.0
ppm (t1)

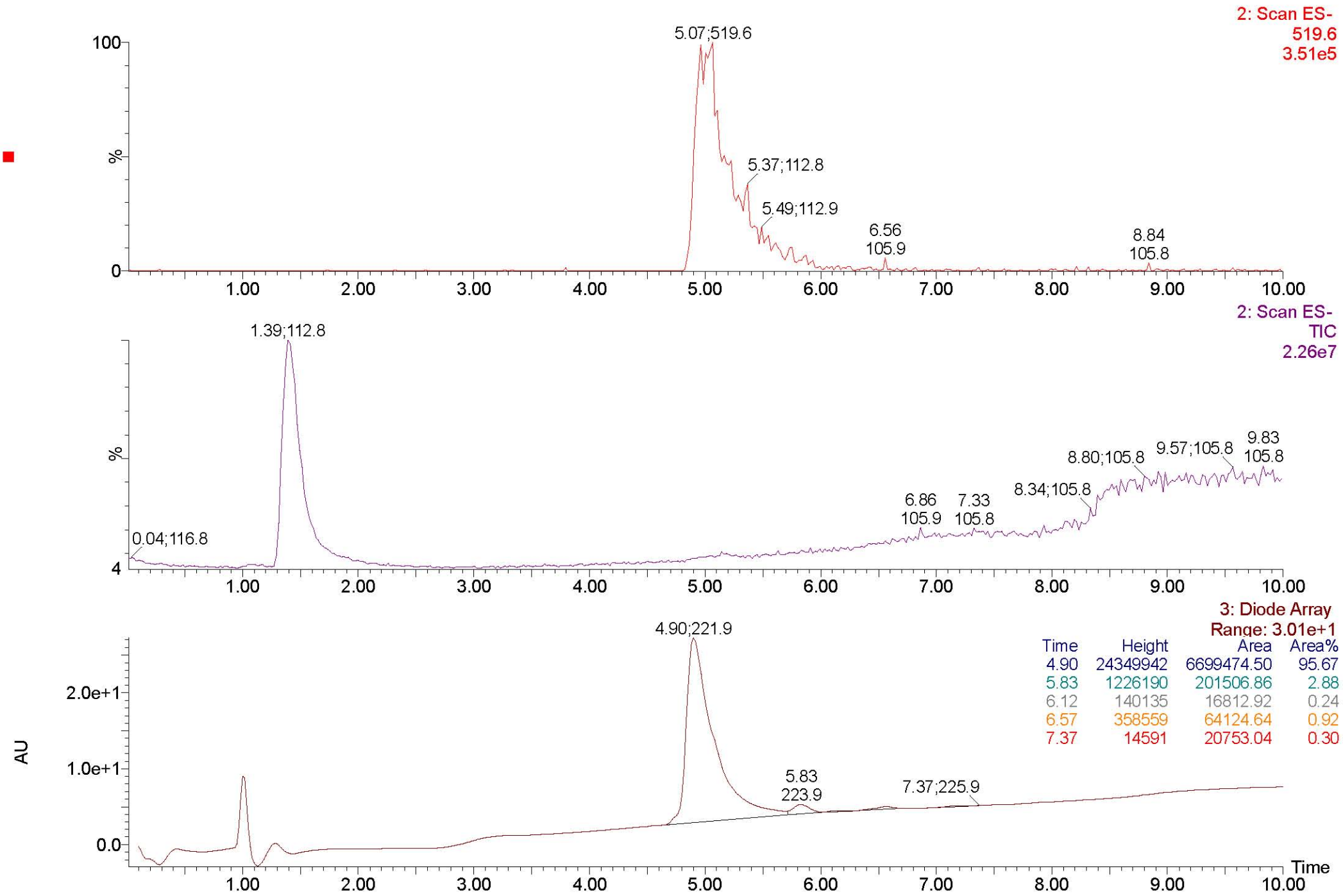
4d


3: Diode Array
Range: 3.01e+1
Time Height Area Area%
4.90 24349942 6699474.50 95.67
5.83 1226190 201506.86 2.88
6.12 140135 16812.92 0.24
6.57 358559 64124.64 0.92
7.37 14591 20753.04 0.30
4.90;221.9
5.83
223.9
7.37;225.9
AU
Time
2.4e+1
2.2e+1
2.0e+1
1.8e+1
1.6e+1
1.4e+1
1.2e+1
1.0e+1
8.0
6.0
4.0
2.0
0.0
-2.0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

4d

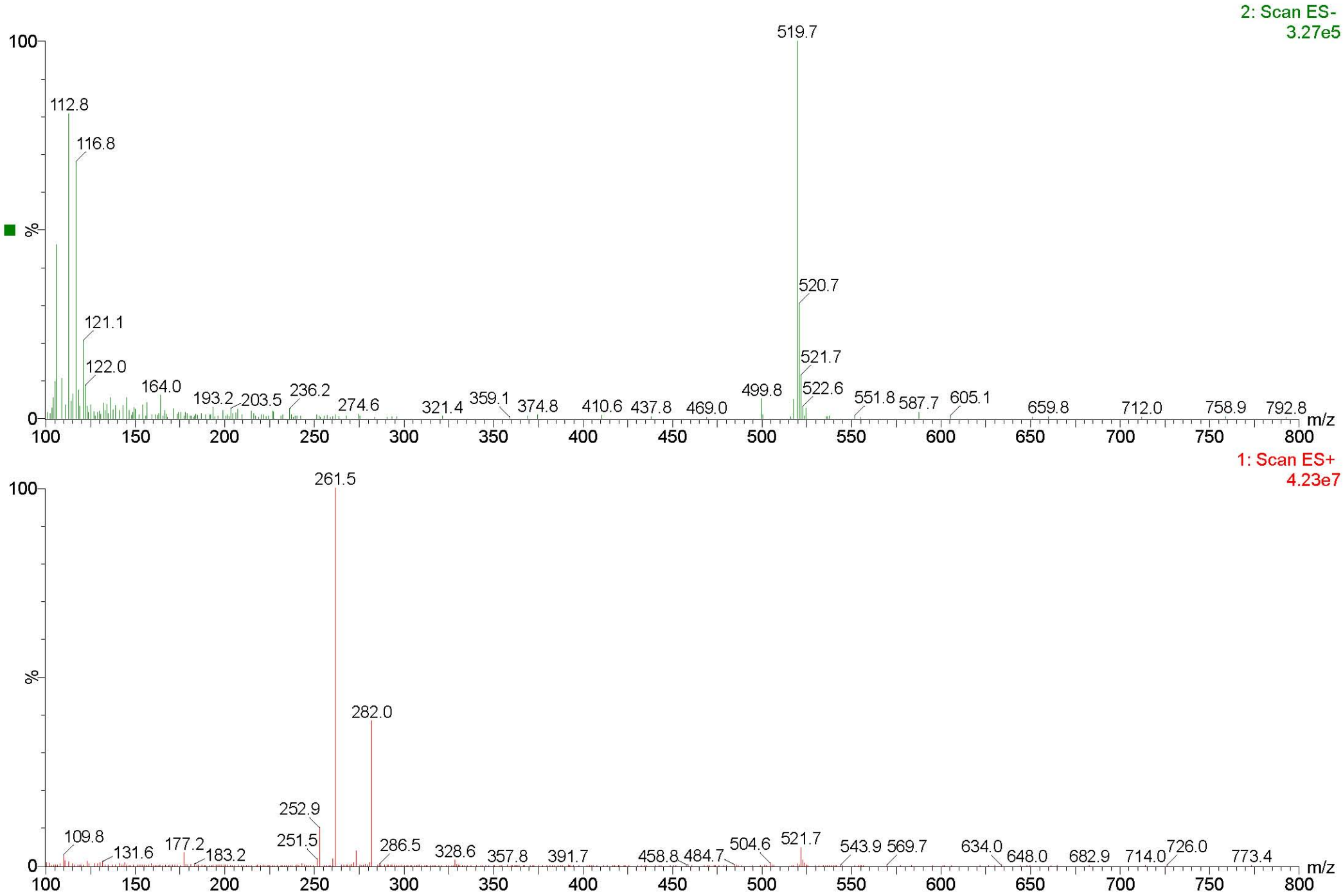

1: Scan ES+
521.7
2.20e6
4.98;261.5
5.14;261.5
5.64;282.0
100
%
0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
1: Scan ES+
TIC
1.42e8
8.49;109.8
8.73;124.0
5.02
261.5
9.70
123.9
0.05;159.1
1.12
173.2
6.63;275.5
7.42;123.9
7.62
123.9
%
7
3: Diode Array
Range: 3.01e+1
4.90;221.9
Time Height Area Area%
4.90 24349942 6699474.50 95.67
5.83 1226190 201506.86 2.88
6.12 140135 16812.92 0.24
6.57 358559 64124.64 0.92
7.37 14591 20753.04 0.30
5.83
223.9
7.37;225.9
AU
2.0e+1
1.0e+1
0.0
Time

4d

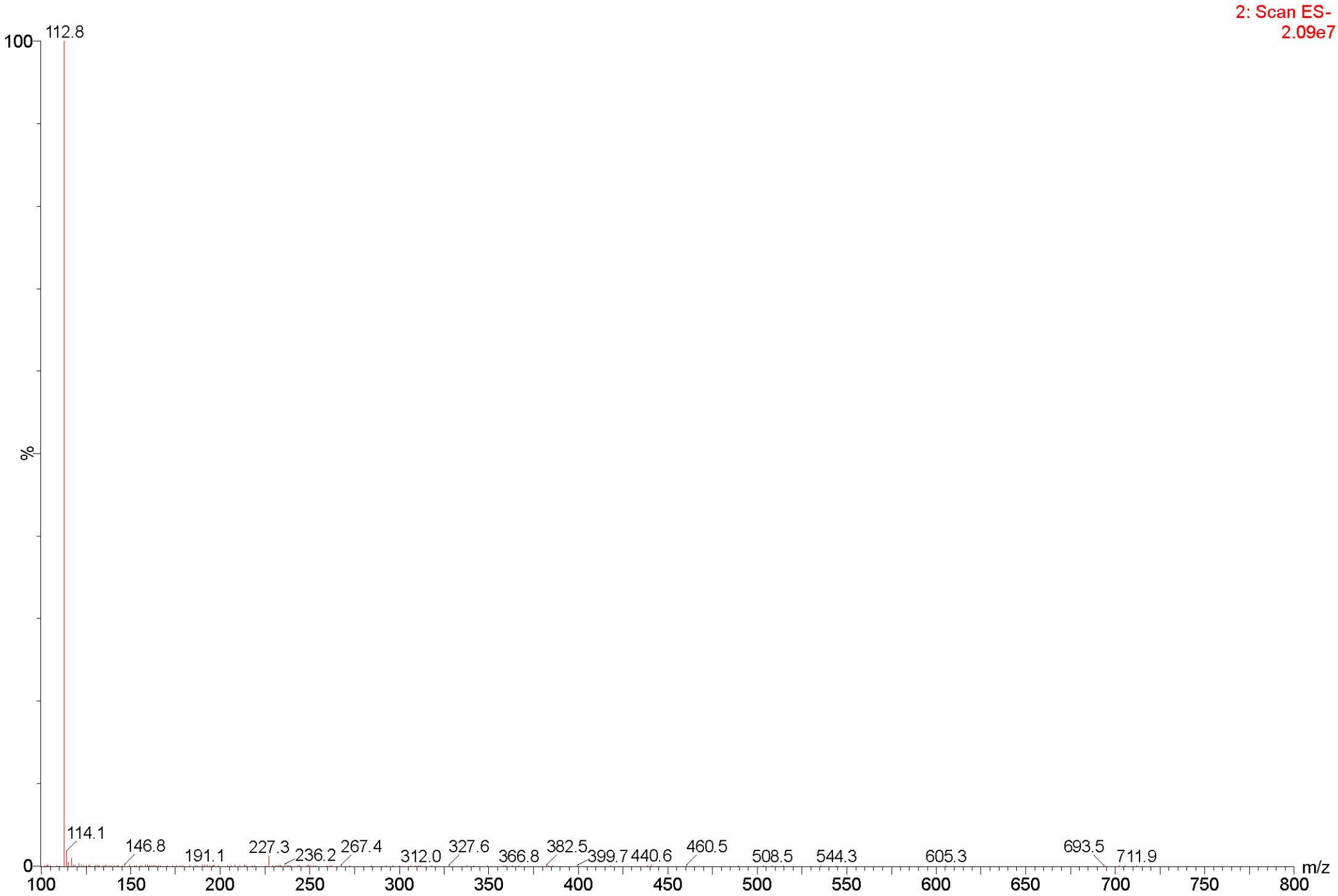


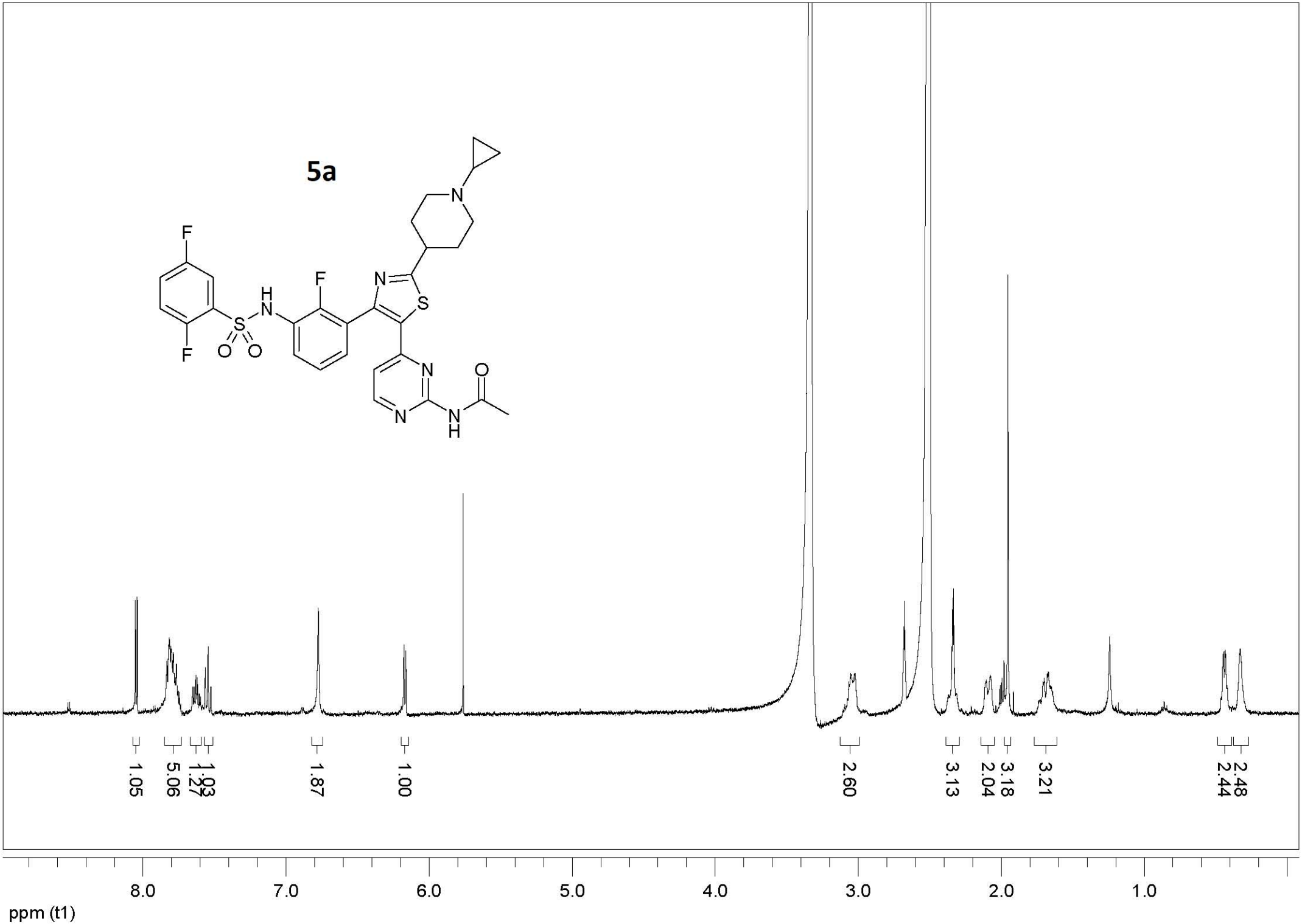
Time	Height	Area	Area%
4.90	24349942	6699474.50	95.67
5.83	1226190	201506.86	2.88
6.12	140135	16812.92	0.24
6.57	358559	64124.64	0.92
7.37	14591	20753.04	0.30

4d

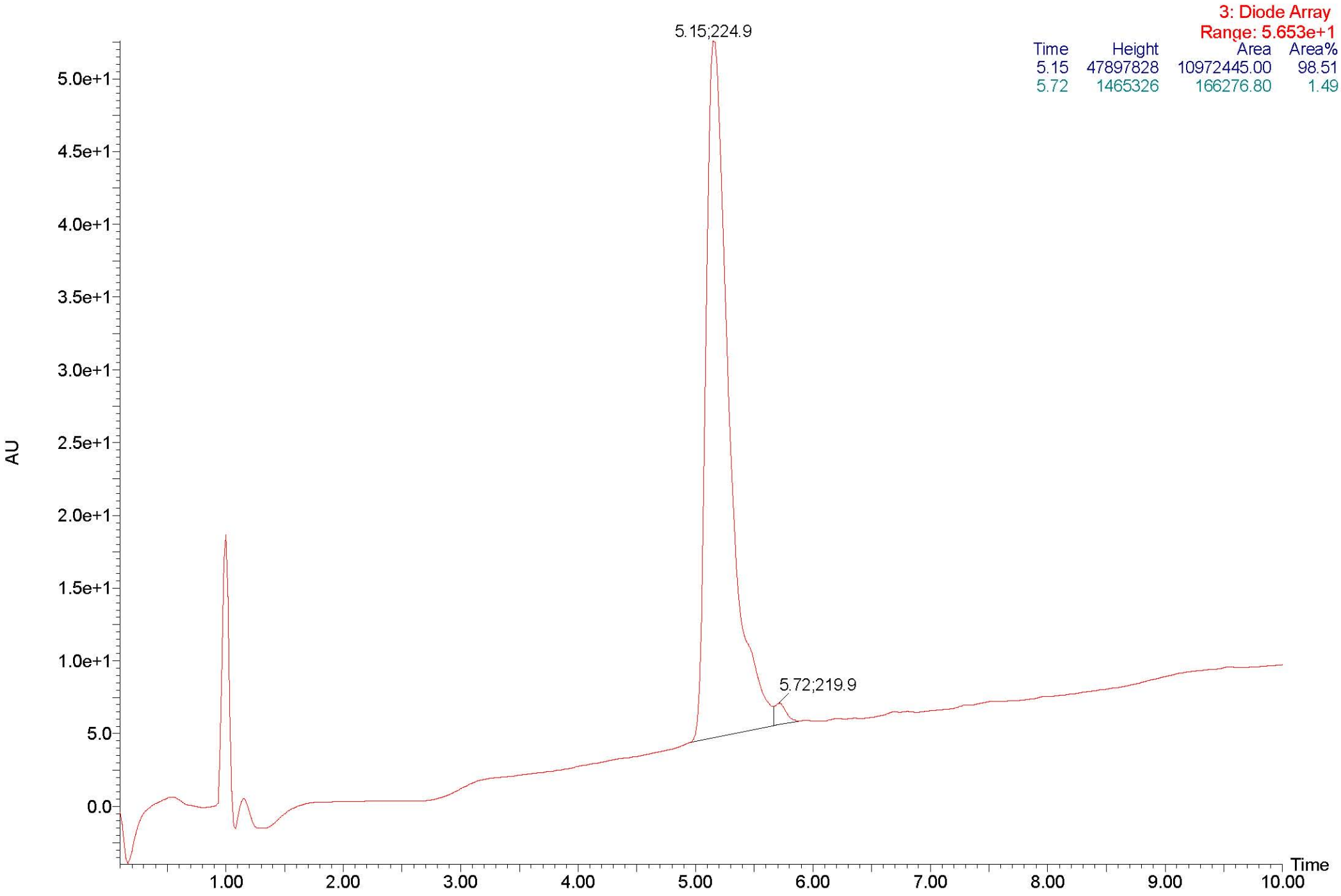

2: Scan ES-
3.27e5
519.7
112.8
116.8
121.1
122.0
164.0
193.2
203.5
236.2
274.6
321.4
359.1
374.8
410.6
437.8
469.0
499.8
520.7
521.7
522.6
551.8
587.7
605.1
659.8
712.0
758.9
792.8
%
m/z
1: Scan ES+
4.23e7
261.5
282.0
252.9
251.5
109.8
131.6
177.2
183.2
286.5
328.6
357.8
391.7
458.8
484.7
504.6
521.7
543.9
569.7
634.0
648.0
682.9
714.0
726.0
773.4

4d

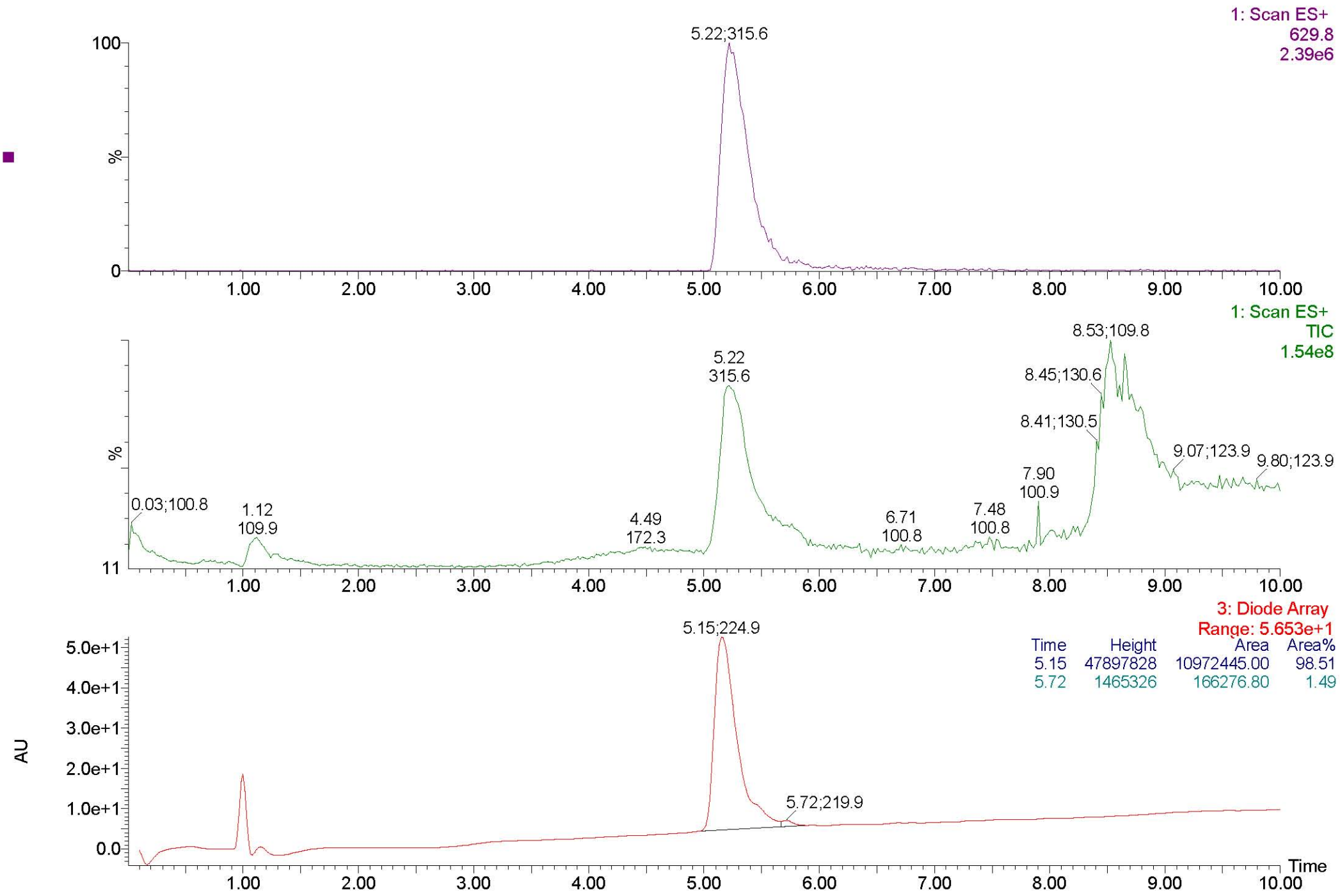

2: Scan ES-
2.09e7
100
%
0
112.8
114.1
146.8
191.1
227.3
236.2
267.4
312.0
327.6
366.8
382.5
399.7
440.6
460.5
508.5
544.3
605.3
693.5
711.9
m/z
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800


5a
F
F
F
H
N
S
O
O
N
S
N
N
N
H
O
1.05
5.06
1.27
1.03
1.87
1.00
2.60
3.13
2.04
3.18
3.21
2.44
2.48
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
ppm (t1)

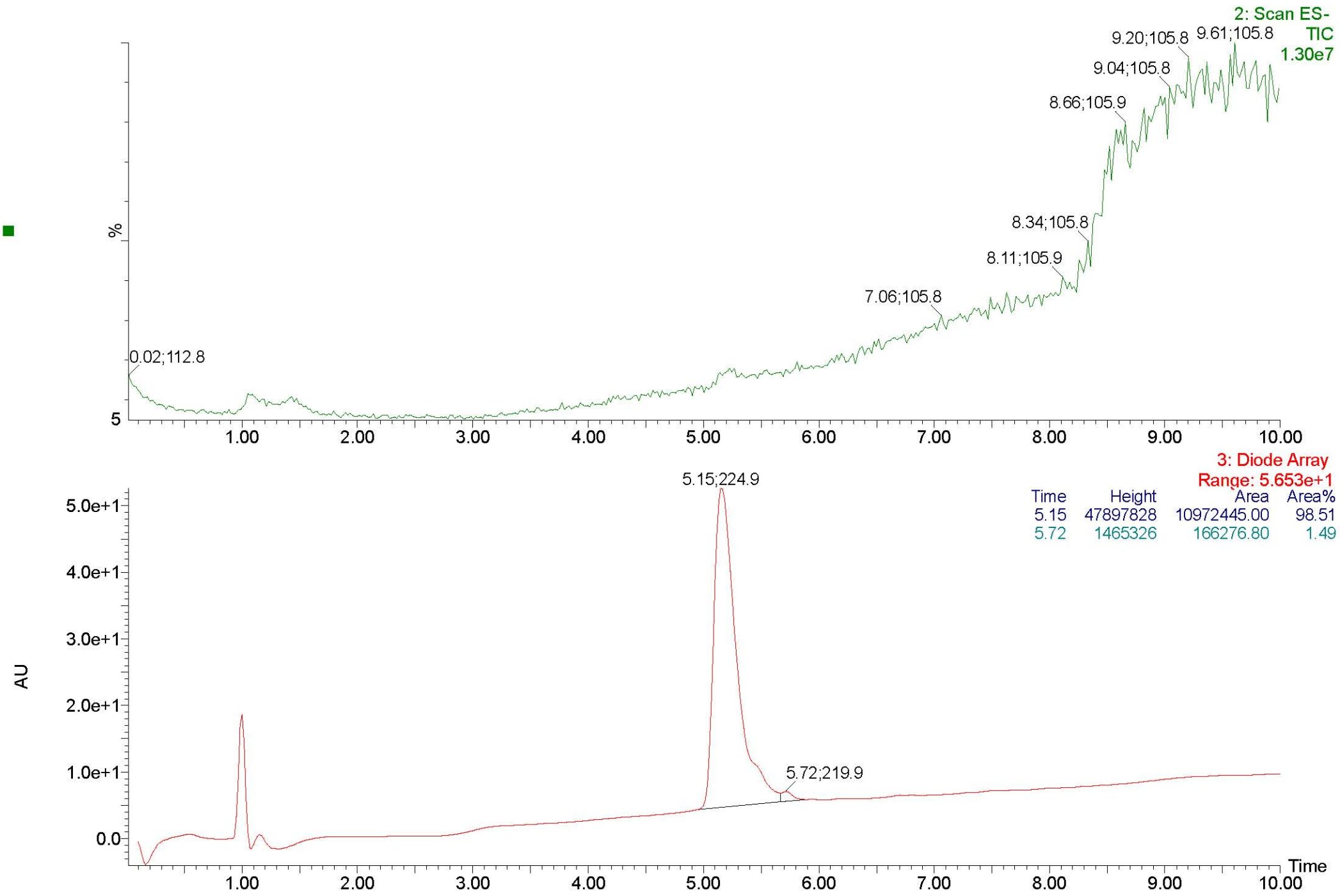
5a


3: Diode Array
Range: 5.653e+1
Time Height Area Area%
5.15 47897828 10972445.00 98.51
5.72 1465326 166276.80 1.49
5.15;224.9
5.72;219.9
AU
Time
5.0e+1
4.5e+1
4.0e+1
3.5e+1
3.0e+1
2.5e+1
2.0e+1
1.5e+1
1.0e+1
5.0
0.0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

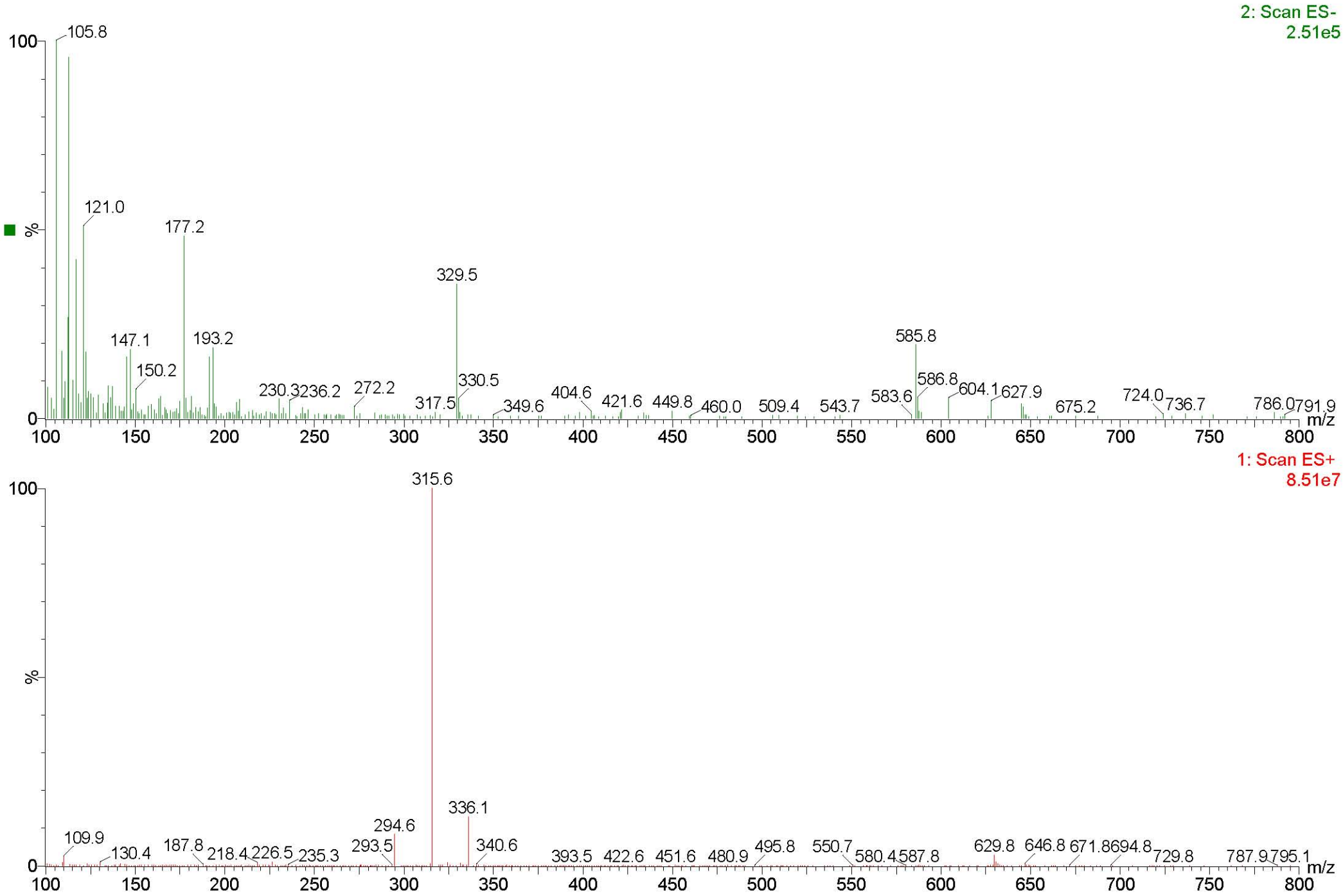
5a

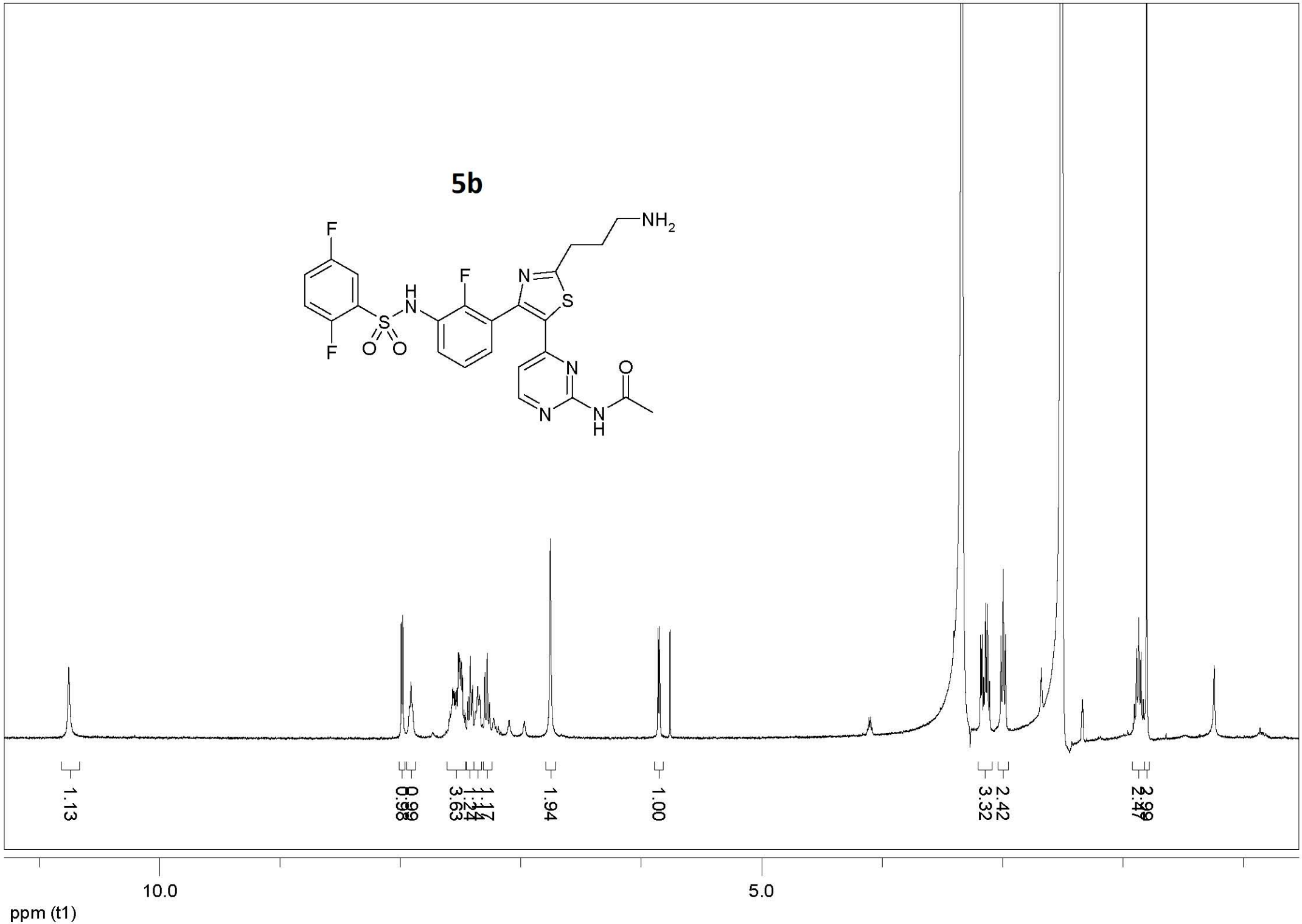

1: Scan ES+
629.8
2.39e6
5.22;315.6
100
%
0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
1: Scan ES+
TIC
1.54e8
8.53;109.8
5.22
315.6
8.45;130.6
8.41;130.5
9.07;123.9
9.80;123.9
7.90
100.9
0.03;100.8
1.12
109.9
4.49
172.3
6.71
100.8
7.48
100.8
11
3: Diode Array
Range: 5.653e+1
5.15;224.9
Time Height Area Area%
5.15 47897828 10972445.00 98.51
5.72 1465326 166276.80 1.49
AU
5.0e+1
4.0e+1
3.0e+1
2.0e+1
1.0e+1
0.0
5.72;219.9
Time

5a

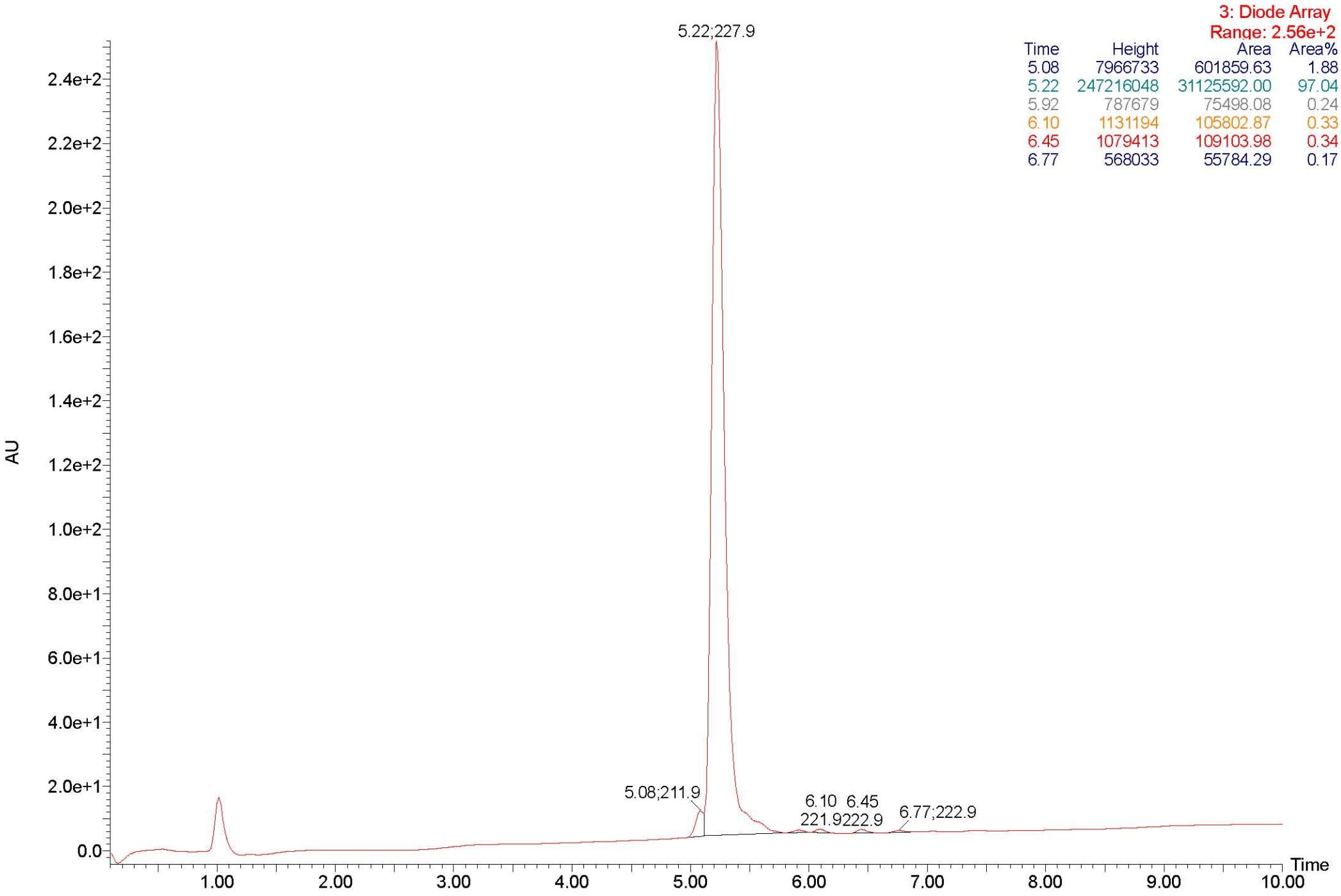

2: Scan ES-
TIC
1.30e7
9.20;105.8
9.61;105.8
9.04;105.8
8.66;105.9
8.34;105.8
8.11;105.9
7.06;105.8
0.02;112.8
%
5
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
3: Diode Array
Range: 5.653e+1
Time Height Area Area%
5.15 47897828 10972445.00 98.51
5.72 1465326 166276.80 1.49
5.15;224.9
5.72;219.9
AU
5.0e+1
4.0e+1
3.0e+1
2.0e+1
1.0e+1
0.0
Time

5a

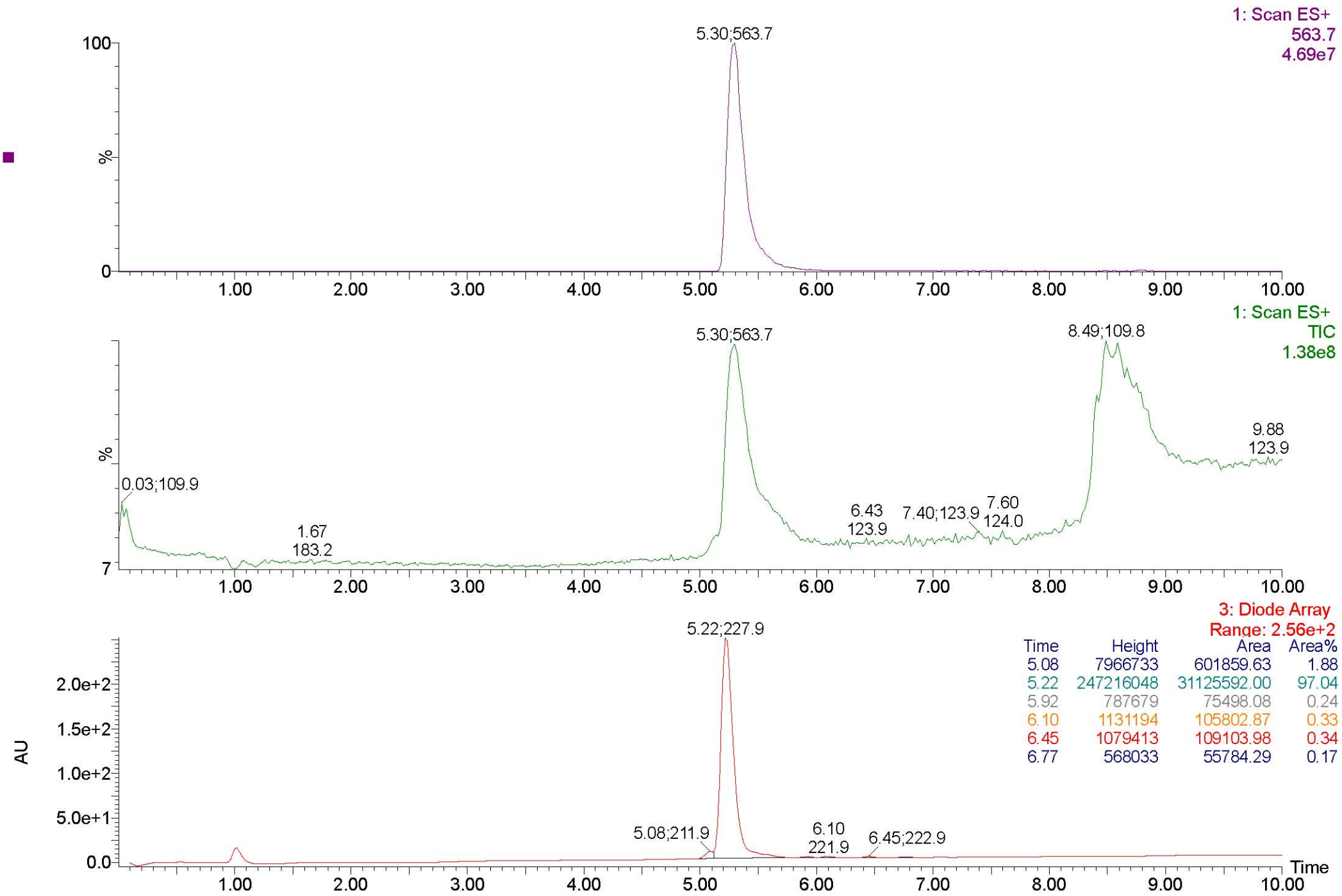

2: Scan ES-
2.51e5
100
%
0
105.8
121.0
177.2
329.5
147.1
193.2
150.2
585.8
230.3
236.2
272.2
330.5
317.5
349.6
404.6
421.6
449.8
460.0
509.4
543.7
583.6
586.8
604.1
627.9
675.2
724.0
736.7
786.0
791.9
m/z
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
1: Scan ES+
8.51e7
100
%
0
315.6
336.1
294.6
109.9
130.4
187.8
218.4
226.5
235.3
293.5
340.6
393.5
422.6
451.6
480.9
495.8
550.7
580.4
587.8
629.8
646.8
671.8
694.8
729.8
787.9
795.1
m/z
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800


5b
F
F
F
H
N
S
O
O
N
S
N
NH2
N
O
N
H
1.13
0.98
0.99
3.63
1.24
1.17
1.94
1.00
3.32
2.42
2.47
2.99
10.0
5.0
ppm (t1)

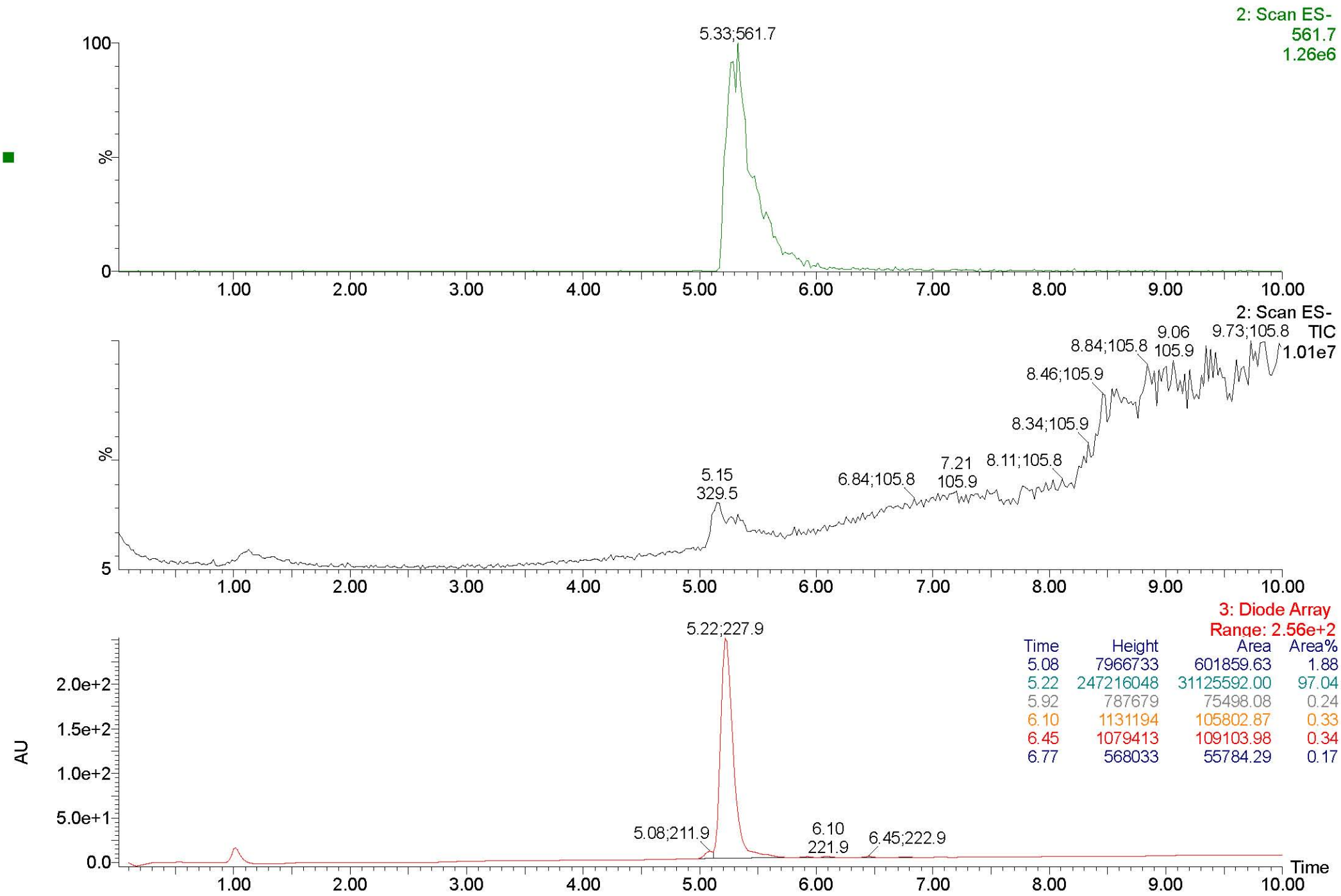
5b


3: Diode Array
Range: 2.56e+2
Time Height Area Area%
5.08 7966733 601859.63 1.88
5.22 247216048 31125592.00 97.04
5.92 787679 75498.08 0.24
6.10 1131194 105802.87 0.33
6.45 1079413 109103.98 0.34
6.77 568033 55784.29 0.17
5.22;227.9
5.08;211.9
6.10 6.45
221.9222.9
6.77;222.9
AU
Time
2.4e+2
2.2e+2
2.0e+2
1.8e+2
1.6e+2
1.4e+2
1.2e+2
1.0e+2
8.0e+1
6.0e+1
4.0e+1
2.0e+1
0.0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00

5b

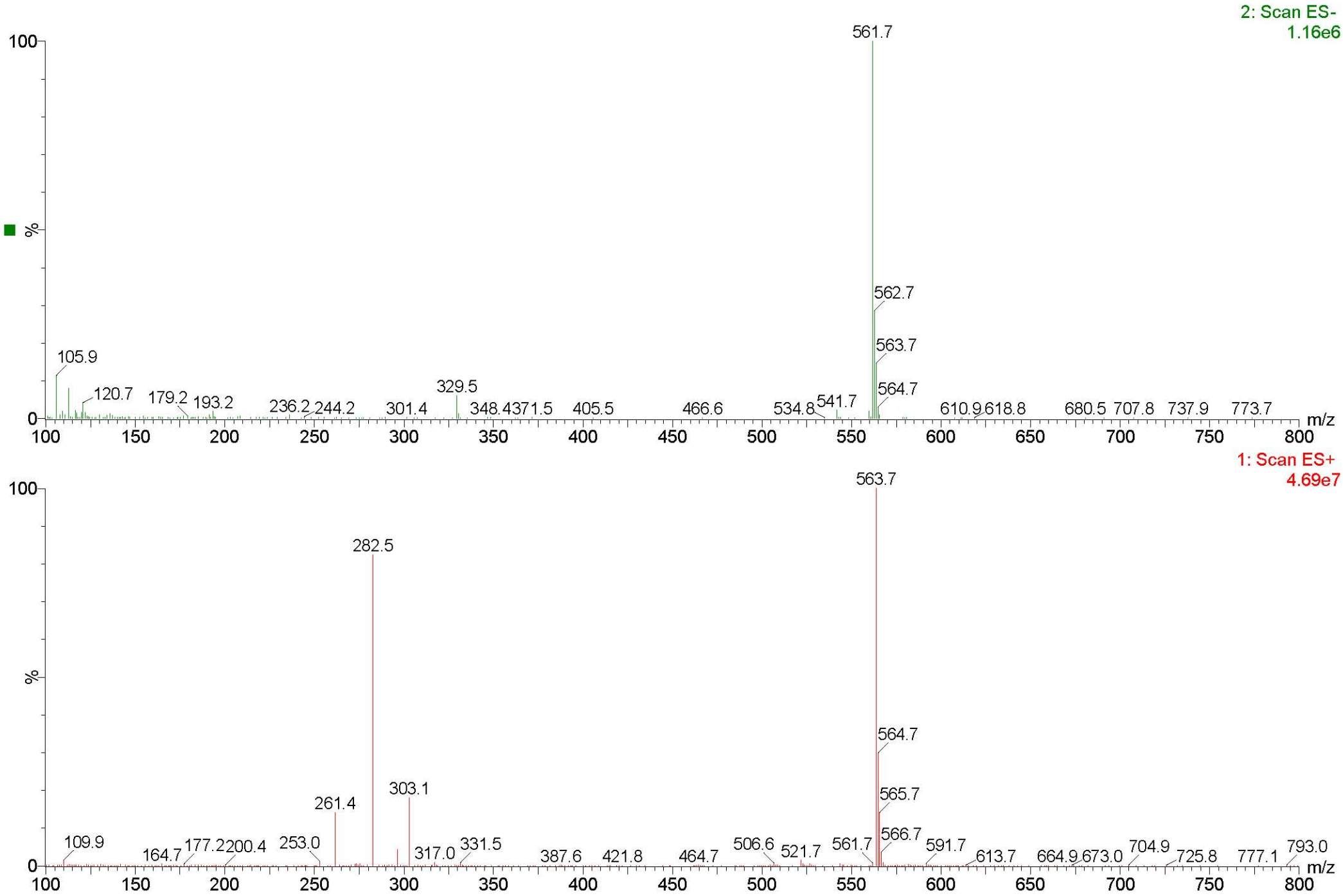

1: Scan ES+
563.7
4.69e7
5.30;563.7
100
%
0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
1: Scan ES+
TIC
1.38e8
5.30;563.7
8.49;109.8
9.88
123.9
0.03;109.9
1.67
183.2
6.43
123.9
7.40;123.9
7.60
124.0
7
3: Diode Array
Range: 2.56e+2
5.22;227.9
AU
2.0e+2
1.5e+2
1.0e+2
5.0e+1
0.0
5.08;211.9
6.10
221.9
6.45;222.9
Time
Time Height Area Area%
5.08 7966733 601859.63 1.88
5.22 247216048 31125592.00 97.04
5.92 787679 75498.08 0.24
6.10 1131194 105802.87 0.33
6.45 1079413 109103.98 0.34
6.77 568033 55784.29 0.17

5b



Time	Height	Area	Area%
5.08	7966733	601859.63	1.88
5.22	247216048	31125592.00	97.04
5.92	787679	75498.08	0.24
6.10	1131194	105802.87	0.33
6.45	1079413	109103.98	0.34
6.77	568033	55784.29	0.17

5b


2: Scan ES-
1.16e6
100
%
0
561.7
562.7
563.7
564.7
105.9
120.7
179.2
193.2
236.2
244.2
301.4
329.5
348.4
371.5
405.5
466.6
534.8
541.7
610.9
618.8
680.5
707.8
737.9
773.7
m/z
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
1: Scan ES+
4.69e7
100
%
0
563.7
282.5
564.7
565.7
566.7
303.1
261.4
109.9
164.7
177.2
200.4
253.0
317.0
331.5
387.6
421.8
464.7
506.6
521.7
561.7
591.7
613.7
664.9
673.0
704.9
725.8
777.1
793.0
m/z
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800