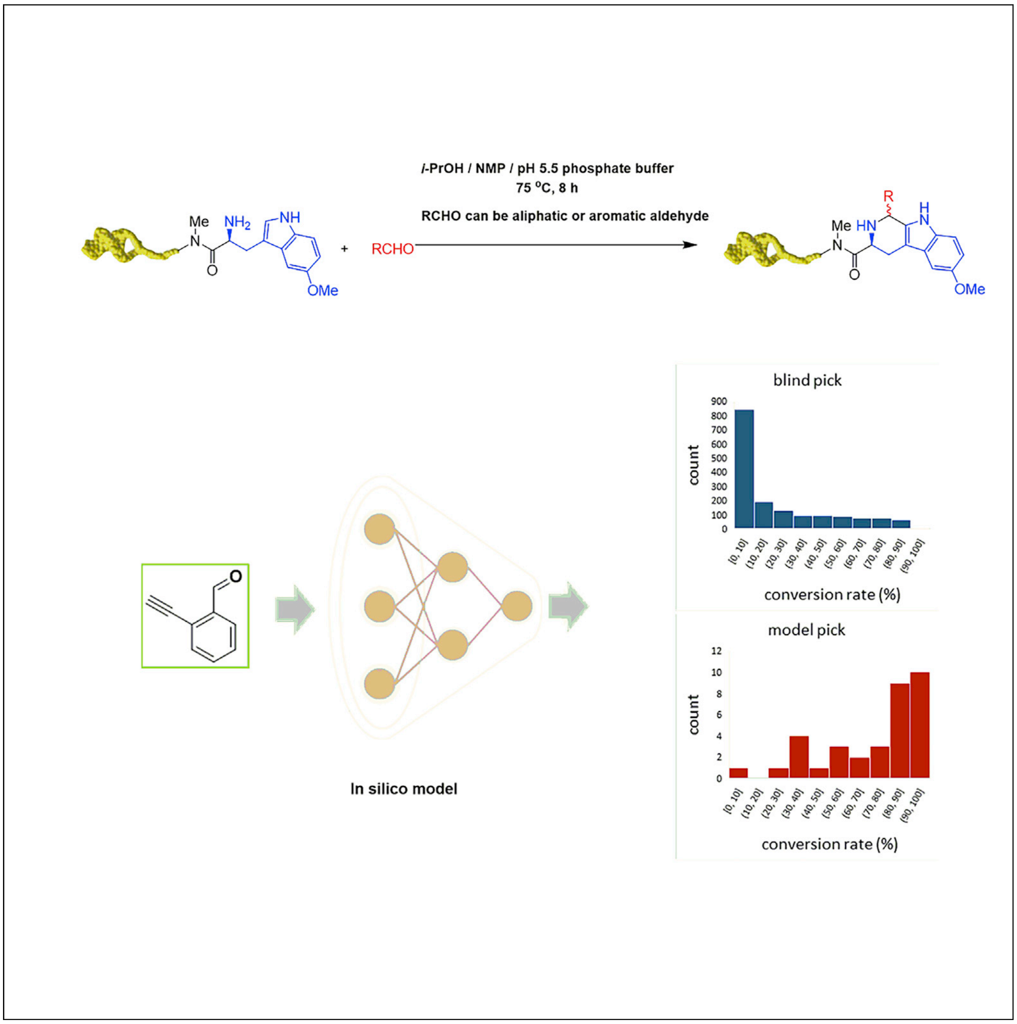
iScience

CellPress
OPEN ACCESS

Article

Solution-Phase DNA-Compatible Pictet-Spengler Reaction Aided by Machine Learning Building Block Filtering



Ke Li, Xiaohong Liu, Sixiu Liu, ..., Mingyue Zheng, Xuanjia Peng, Xiaojie Lu

myzheng@simm.ac.cn (M.Z.)
peng_xuanjia@wuxiapptec.com (X.P.)
xjlu@simm.ac.cn (X.L.)

HIGHLIGHTS

A mild solution-phase, plate applicable DNA-compatible Pictet-Spengler (PS) reaction

An efficient strategy for DNA-encoded diversified tryptoline libraries synthesis

A machine learning algorithm of building blocks filtering for DEL synthesis

An elegant application of machine learning for DNA-encoded library technology

Li et al., iScience *23*, 101142
June 26, 2020 © 2020 The Author(s).
https://doi.org/10.1016/j.isci.2020.101142

iScience

CellPress
OPEN ACCESS

Article

Solution-Phase DNA-Compatible Pictet-Spengler Reaction Aided by Machine Learning Building Block Filtering

Ke Li,[2] Xiaohong Liu,[3,4,5] Sixiu Liu,[1,5] Yulong An,[2] Yanfang Shen,[2] Qingxia Sun,[2] Xiaodong Shi,[2] Wenji Su,[2] Weiren Cui,[2] Zhiqiang Duan,[1,5] Letian Kuai,[2] Hongfang Yang,[2] Alexander L. Satz,[2] Kaixian Chen,[3,4,5] Hualiang Jiang,[3,4,5] Mingyue Zheng,[4,5,*] Xuanjia Peng,[2,*] and Xiaojie Lu[1,5,6,*]

SUMMARY

The application of machine learning toward DNA encoded library (DEL) technology is lacking despite obvious synergy between these two advancing technologies. Herein, a machine learning algorithm has been developed that predicts the conversion rate for the DNA-compatible reaction of a building block with a model DNA-conjugate. We exemplify the value of this technique with a challenging reaction, the Pictet-Spengler, where acidic conditions are normally required to achieve the desired cyclization between tryptophan and aldehydes to provide tryptolines. This is the first demonstration of using a machine learning algorithm to cull potential building blocks prior to their purchase and testing for DNA-encoded library synthesis. Importantly, this allows for a challenging reaction, with an otherwise very low building block pass rate in the test reaction, to still be used in DEL synthesis. Furthermore, because our protocol is solution phase it is directly applicable to standard plate-based DEL synthesis.

INTRODUCTION

DNA-encoded libraries (DELs) are collections of small molecules covalently linked to unique, structure-identifying DNA tags, which enable screens of a large pool of billions (even trillions) of library members for binders of disease-related biologically interesting targets (Clark et al., 2009; Ralph et al., 2011; Favalli et al., 2018; Neri and Lerner, 2018; Zhou et al., 2018; Faver et al., 2019; Reddavide et al., 2019; Dichson and Kodadek, 2019; Yuen et al., 2019). Compared with traditional combinatorial encoded methods, a distinctive and amplifiable DNA tag facilitates the decoding process and enables the screening of much larger libraries (trillions versus millions) (Buller et al., 2010; Encinas et al., 2014; Franzini and Randolph, 2016; Ottl et al., 2019). After affinity selection, the hit molecule's structural information is deciphered from the attached DNA via next-generation sequencing (Eidam and Satz, 2016; Roman et al., 2018).

High-quality DELs are the basis for the success of subsequent screening experiments, and quality includes high conversion rate for each building block (BB) used during library synthesis. Thousands of BBs are routinely reacted with a model DNA-conjugate to determine their appropriateness for use, in a particular reaction, prior to DNA-encoded library (DEL) synthesis. For all investigated reactions, a significant percentage of the acquired and tested BBs fail this validation step (generally a >50% conversion to desired product is required for a BB to ''pass'' the validation), greatly increasing reagent costs and library development time. Additionally, for particularly challenging reactions, the BB pass rate can be extremely low. Owing to the limited resources, it is not practical to pick high-conversion-rate BBs by experimentally determining the conversion rate of each commercially available BB, as a significant percentage of purchased BBs will fail to pass this validation step. To maximize the likelihood that purchased BBs will pass chemical validation, we envision the use of an informatics filter that could readily and inexpensively assess the likelihood of any particular BB to provide a high yield of desired product. Machine learning (ML) is a technology to build a mathematical model based on sample data, known as "training data," in order to make predictions or decisions without being explicitly programmed how to make the decision (Bishop, 2006). Great successes have been made with the method in the field of computer vision, natural language processing, and biological medicine (LeCun et al., 2015). However, no research regarding DEL reaction conversion rate ML

[1] State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 501 Haike Road, Zhang Jiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China

[2] DNA Encoded Library Platform, WuXi AppTec, 288 Fute Zhong Road, Waigaoqiao Free Trade Zone, Shanghai 200131, China

[3] Shanghai Institute for Advanced Immunochemical Studies, and School of Life Science and Technology, ShanghaiTech University, Shanghai, China

[4] Drug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, China

[5] University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, China

[6] Lead Contact

*Correspondence: myzheng@simm.ac.cn (M.Z.), peng_xuanjia@wuxiapptec.com (X.P.), xjlu@simm.ac.cn (X.L.)

https://doi.org/10.1016/j.isci.2020.101142

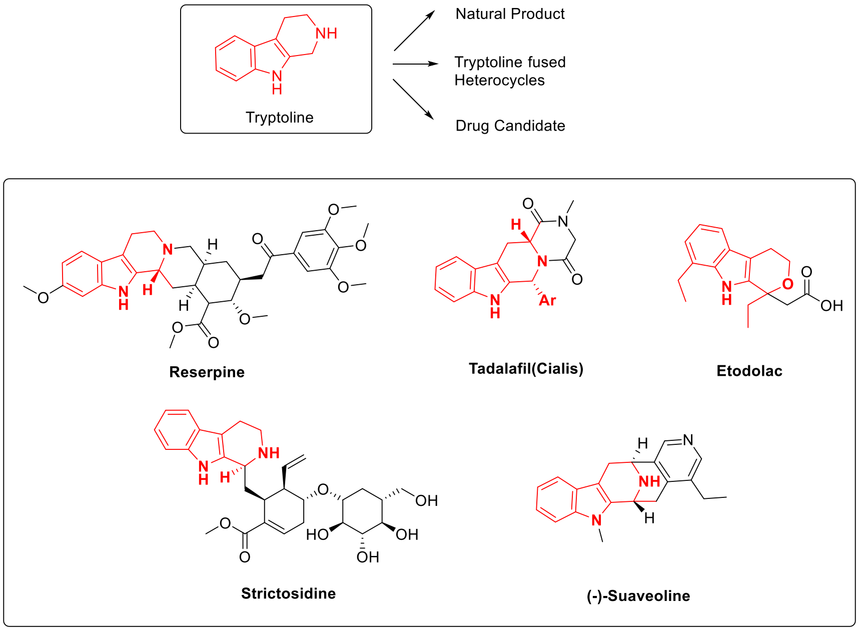


Figure 1. Bioactive C-1-Functionalized Tryptoline Derivatives through the Pictet-Spengler Cyclization Strategy

prediction has been reported. Studies in traditional organic synthesis have used ML for the estimation of catalytic performance (Kite et al., 1994; Omata and Yamada, 2004) and reaction success (Skoraczynski et al., 2017; Raccuglia et al., 2016). More recently, a study has applied descriptors obtained by quantum chemical calculation to predict reaction yield (Ahneman et al., 2018). However, quantum chemical calculation is a time-consuming process, which is not practical when applied to DEL reaction yield prediction because tens of thousands of BBs are needed to be evaluated in a library constructing process.

Furthermore, applying ML for BB filtering is particularly valuable for challenging DNA-compatible reactions owing to the expected low BB passing rate. Although DEL is successful for hit identification and widely used throughout the academic and industrial small molecule drug discovery community, it still suffers from a limited number of DNA-compatible reactions and thus limited access to desirable drug-like chemical space (Satz et al., 2015; Malone and Paegel, 2016; Lu et al., 2017a, 2017b; Wang et al., 2018a, 2018b; Li et al., 2018; Flood et al., 2019; Wang et al., 2019; Du et al., 2019; Lerner et al., 2019; Liu et al., 2019; Skopic et al., 2019; Xu et al., 2019). More DNA-compatible organic transformations, especially the challenge but highly valuable ones, are strongly desired to improve the chemical diversity of DNA-encoded libraries. We envisioned develop a new challenge DNA-compatible reaction and applied ML for the BB filtering for DEL synthesis is a rational strategy for DEL chemical space expansion especially for valuable privileged scaffolds based DELs.

We decided to focus on the DNA-compatible cyclization of highly functionalized and rigid rings. (Note that our laboratory has previously discussed the design and synthesis of orthogonally protected heterocyclic scaffolds for use in DEL synthesis [Gong et al., 2017]). Poly-substituted optically active tryptoline derivatives classified as nonisoprenoids are common structural motifs in indole-based alkaloids. As depicted in Figure 1, functionalization of the C-1 position of tryptoline derivatives is generally observed in natural-product-based indole alkaloids and commercial drugs such as tadalafil (Cialis) (Yamamura et al., 2017) and etodolac (LaPlante et al., 2013; Maity et al., 2019). The World Drug Index contains over 200 listings of this distinctive heterocycle, which is usually assembled by the (Pictet-Spengler) PS reaction (Maity et al., 2019). Unfortunately, owing to the acidic conditions required for the PS reaction, there were no DEL PS reactions reported in the literature until recently. Importantly, both reports demonstrate proof of concept only and would require non-trivial deviation from existing protocols to actually synthesize a DEL. One report details a high-throughput solid phase methodology to synthesize tryptoline-containing BBs (Zambaldo et al., 2019). After release of the desired products from resin, the tryptoline products were then conjugated to DNA oligomers in a 96-well plate-based format. The second method demonstrated that short DNA oligomers attached to resins were protected from mildly acidic conditions, such that the PS reaction could be

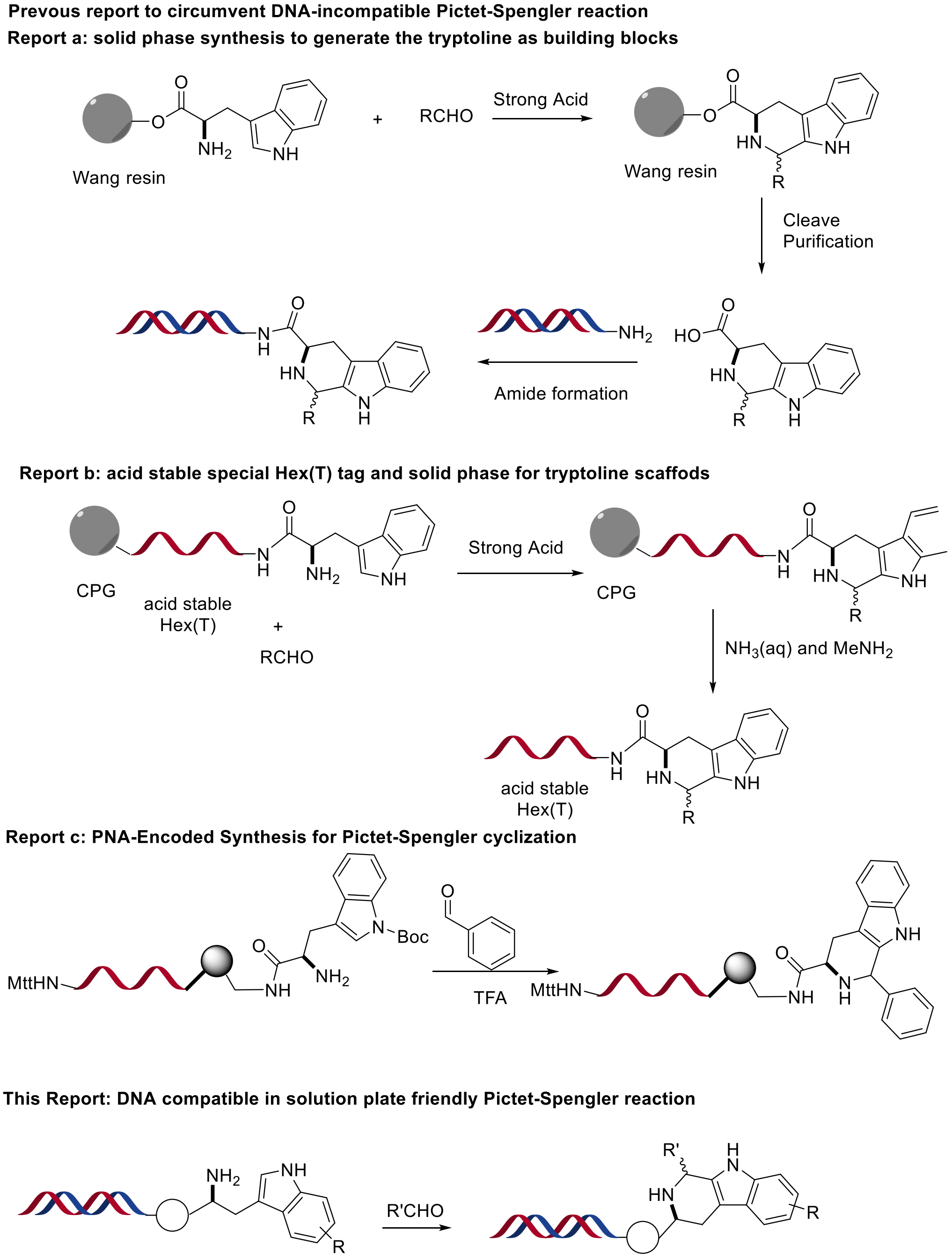
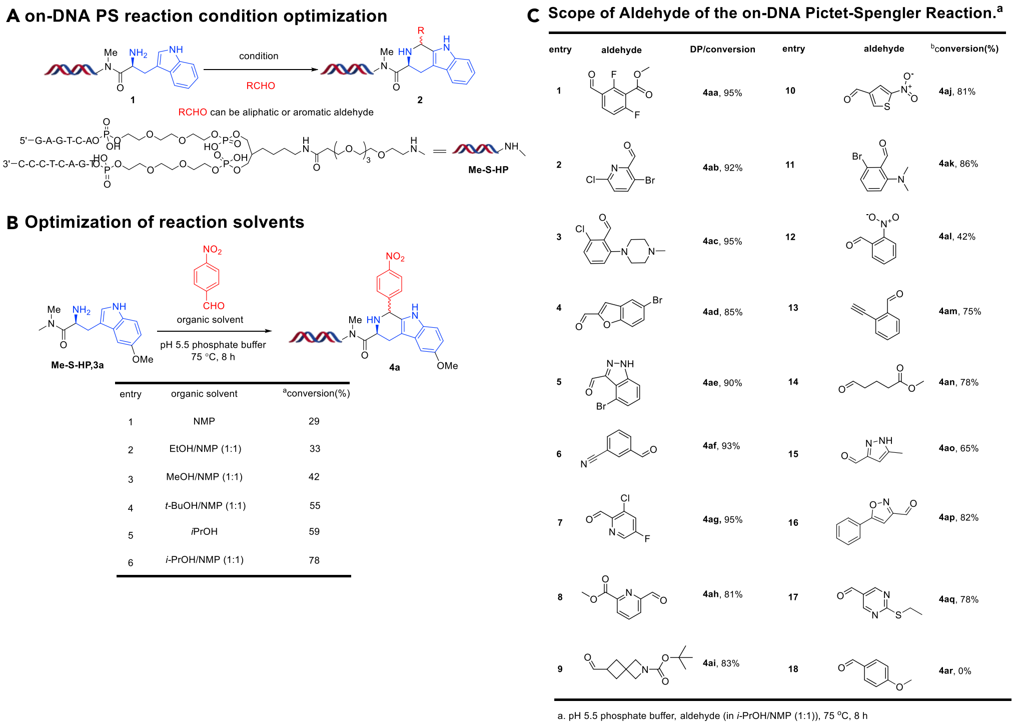

Prevous report to circumvent DNA-incompatible Pictet-Spengler reaction
Report a: solid phase synthesis to generate the tryptoline as building blocks
O
O
NH2
NH
Wang resin
+
RCHO
Strong Acid
O
O
HN
NH
R
Wang resin
Cleave
Purification
NH2
HO
O
HN
R
N
H
Amide formation
N
H
O
HN
R
N
H
Report b: acid stable special Hex(T) tag and solid phase for tryptoline scaffods
CPG
acid stable
Hex(T)
N
H
O
NH2
NH
+
RCHO
Strong Acid
CPG
N
H
O
HN
NH
R
NH3(aq) and MeNH2
acid stable
Hex(T)
N
H
O
HN
NH
R
Report c: PNA-Encoded Synthesis for Pictet-Spengler cyclization
MttHN
O
NH
NH2
N
Boc
O
TFA
MttHN
O
NH
N
H
NH
This Report: DNA compatible in solution plate friendly Pictet-Spengler reaction
NH2
NH
R
R'CHO
R'
H
N
HN
R

Figure 2. In-Solution Plate Friendly DNA-Compatible Pictet-Spengler Reaction

successfully accomplished in combination with electron-poor aldehydes (Figure 2) (Skopic et al., 2017). Besides, the preparation of PNA monomers with a protecting group combination (Mtt/Boc), which is orthogonal to Fmoc-based synthesis and compatible with PS reaction, was also elegantly demonstrated (Chouikhi et al., 2012). Despite the potential of the above methods, we believe there is still a clear need for a solution-phase PS reaction that is compatible with existing and proven DEL synthesis protocols (Figure 2).

Herein, we discuss the optimization (and reagent design) of a DNA-compatible and solution-phase PS reaction. The PS is a challenging reaction, as conditions that increase conversion rate often also increase DNA damage. Thus, under our optimized reaction conditions (which avoids DNA damage), a majority of


A on-DNA PS reaction condition optimization
Me
NH2
NH
condition
RCHO
1
2
R
HN
RCHO can be aliphatic or aromatic aldehyde
5'−G-A-G-T-C-A
3'−C-C-C-T-C-A-G-T
Me-S-HP
B Optimization of reaction solvents
NO2
CHO
organic solvent
pH 5.5 phosphate buffer
75 °C, 8 h
Me-S-HP,3a
OMe
4a
C Scope of Aldehyde of the on-DNA Pictet-Spengler Reaction.a

entry	organic solvent	[a]conversion(%)
1	NMP	29
2	EtOH/NMP (1:1)	33
3	MeOH/NMP (1:1)	42
4	*t*-BuOH/NMP (1:1)	55
5	*i*PrOH	59
6	*i*-PrOH/NMP (1:1)	78

entry	aldehyde	DP/conversion	entry	aldehyde	[b]conversion(%)
1		**4aa**, 95%	10		**4aj**, 81%
2		**4ab**, 92%	11		**4ak**, 86%
3		**4ac**, 95%	12		**4al**, 42%
4		**4ad**, 85%	13		**4am**, 75%
5		**4ae**, 90%	14		**4an**, 78%
6		**4af**, 93%	15		**4ao**, 65%
7		**4ag**, 95%	16		**4ap**, 82%
8		**4ah**, 81%	17		**4aq**, 78%
9		**4ai**, 83%	18		**4ar**, 0%

a. pH 5.5 phosphate buffer, aldehyde (in *i*-PrOH/NMP (1:1)), 75 °C, 8 h

Figure 3. On-DNA PS Reaction Condition Optimization

(A) On-DNA PS reaction condition optimization.

(B) Optimization of reaction solvents.

(C) Scope of aldehyde of the on-DNA Pictet-Spengler Reaction. [a]Conversions determined by liquid chromatography-mass spectrometry (LC-MS).

randomly chosen aldehyde BBs fail to give high conversation to desired cyclized products. To better filter commercial BBs prior to their purchase, we trained a deep neural network (DNN) model (a type of ML model) to predict the conversion rate of BBs in our challenging PS reaction. We then purchased a subset of these BBs and compared our model's predictions with experimental results.

RESULTS AND DISCUSSION

The Development of On-DNA PS Reaction

Developing a DNA-compatible solution-phase PS reaction using DNA-conjugated tryptamine substrates **1** and providing the desired products **2** (Figure 3A) (Zambaldo et al., 2019) is challenging since acidic conditions are typically required. To optimize a DNA-compatible reaction, we carried out parallel screening to test whether any acidic promoter could not only efficiently promote the PS reaction with simple aldehydes but also preserve the DNA without decomposition. Unfortunately, unlike reported literature procedures in traditional organic solvents, neither Lewis acids (Srinivasan and Ganesan, 2003) Sc(OTf)$_3$, In(OTf)$_3$, YbCl$_3$, YCl$_3$, Sm(OTf)$_3$ nor Brønsted acids H_3PO_4, HCOOH appeared to promote the PS reaction, and only DNA damage was observed. Unsurprisingly, basic conditions such as adding NaOH or employing pH12 buffer also did not promote the reaction, and using I_2 (Dipak and Mukut, 2008) also gave disappointing results.

Drawing upon existing literature reports and the known mechanism of the PS reaction, we hypothesized that the combination of an electronic-rich tryptamine derivative and an electronic-deficient aldehyde may have better reactivity compared with the previous substrates **1**. Thus, we chose to investigate a methoxy-substituted tryptamine-conjugated DNA substrate **3a**. Employing a pH5.5 phosphate buffer to maintain a weakly acidic condition, we observed ~29% conversion to the desired cyclized product **4a** and no obvious signs of DNA damage. Next, we screened a series of solvents as tabulated in Figure 3B. The solvent i-PrOH led to an increase to 59% conversion, despite the test aldehyde 4-Nitrobenzaldehyde being poorly soluble under these conditions (Entry 2, Figure 3B). Thus, we chose a mixture of NMP and i-PrOH to improve solubility and observed a further increase in conversion to 78% (Entry 3, Figure 3B).

Next, we explored the scope of our optimized conditions and gratifyingly saw the PS reaction proceed smoothly with a broad spectrum of aldehydes. We found that our optimized conditions tolerated different functional groups including halides (Figure 3C, entries 1, 2, 3, 4, 5, 7, 11, 15), esters (entries 1, 8, 14), alkynes (entry 13), t-butyloxy carbonyls (entry 9), and nitriles (entry 6). Most of the heterocyclic aryl aldehydes gave moderate to excellent conversion (entries 4, 5, 15, 16, 17); however, an electron-rich (OCH_3) aryl aldehyde even does not work (entry 18). And several aldehydes gave two different cyclization products because of stereo isomers (entries 1, 3, 5, 6, 7, 8, 10, 12, 16, 17) (for details see the Supplemental Information). To confirm that we were correctly assigning our DNA-conjugated products, we synthesized the corresponding off-DNA small molecule **4** as the free acid and then acylated this fully characterized cyclized molecule onto a DNA-conjugate. HPLC comparison of the two batches of **4** confirmed them to be the same (see the Supplemental Information).

DNN Model Construction and Validation

From the above exploration of aldehydes, 1,655 reaction records were collected (Table S1), based on which a DNN model was established (k-NearestNeighbor algorithm, KNN, a traditional machine learning methodwas also investigated as a baseline model, see Table S5). The extended connectivity fingerprints (ECPF4 [David and Mathew, 2010]) with a radius of two consecutive bonds and a length of 1,024 bits were used as input features, and conversion rates were used as the output task for learning (MACCS keys fingerprints were also tried; for details see Table S6).

To train the model, 20% of the data was randomly selected as an internal test dataset and the rest was selected as training dataset. Five-fold cross-validation was performed within the training dataset during the training process. In detail, training dataset was split into five folds and each fold is then used once as a validation, whereas the four remaining folds form the training set. For each fold, early stopping (Montavon et al., 2012) was applied and the training process was stopped when the mean square error (MSE) did not decrease for 50 epochs. The mean MSE of 5-fold cross-validation was chosen for searching the optimal k of KNN, and parameters of DNN and Adam optimizer (Kingma and Ba, 2014) were chosen for DNN parameters optimization (Table S2). In order to reduce the risk of overfitting, dropout (Srivastava et al., 2014) and weight decay (Kingma and Ba, 2014) were used for regularization. Bayesian optimization (BO) was also applied with pyGPGO (Jiménez and Ginebra, 2017) for DNN; however, no better hyper parameters were found. In the end, the DNN model with two hidden layers and the size 1,024 combined with ECFP4 showed the lowest mean MSE in 5-fold cross-validation and was chosen for further use.

Generally, only BBs showing high conversion rate in a test PS reaction (i.e. the validation) will be selected for DEL construction and those with low conversion rate will be discarded. Thus, a useful model must correctly identify BBs that latter give high conversion rate in the test PS reaction. To better quantify the performance of the model, BBs with conversion rates over 50% are labeled 1 and others are labeled 0. Following this definition, a precision-recall curve can be plotted (Figure 4A). The results indicate a satisfactory performance for selecting BBs with high conversion rates (the precision is 0.81), although some positive samples might be missed (i.e., the recall is 0.37, but this is of less concern when picking a small subset of BBs from a large list of commercial reagents and clustering could be further used to make that the final selection of BBs to be as diverse as possible).

We then carried out an external experimental validation of our ML model. All data containing aldehyde BB (Table S3) from the WuXi LabNetwork platform were collected, cleaned (remove ions), and evaluated by the model. The final predicted value of BB is the mean value of five best model in 5-fold cross-validation. In order to maximize the performance of the model, BBs were sorted in descending order according to predicted conversion rate and top 300 BBs were retained and those previously included in the train or internal test dataset were discarded (Table S4). After further filtering for price and real-time availability, 34 BBs were acquired for experimental evaluation (external test dataset, Table S4). The results show that the performance of the model on the external dataset was similar to that on the internal test dataset, with a precision for identifying BBs with conversion rate above 50% being 0.79 (0.81 for the internal test dataset). Compared with 1,655 blindly picked BBs, the model has a better performance to find high-conversion-rate BBs (high-conversion-rate BBs percentage: 18.4% versus 79.4%, Figure 4B). For parallel comparison of the

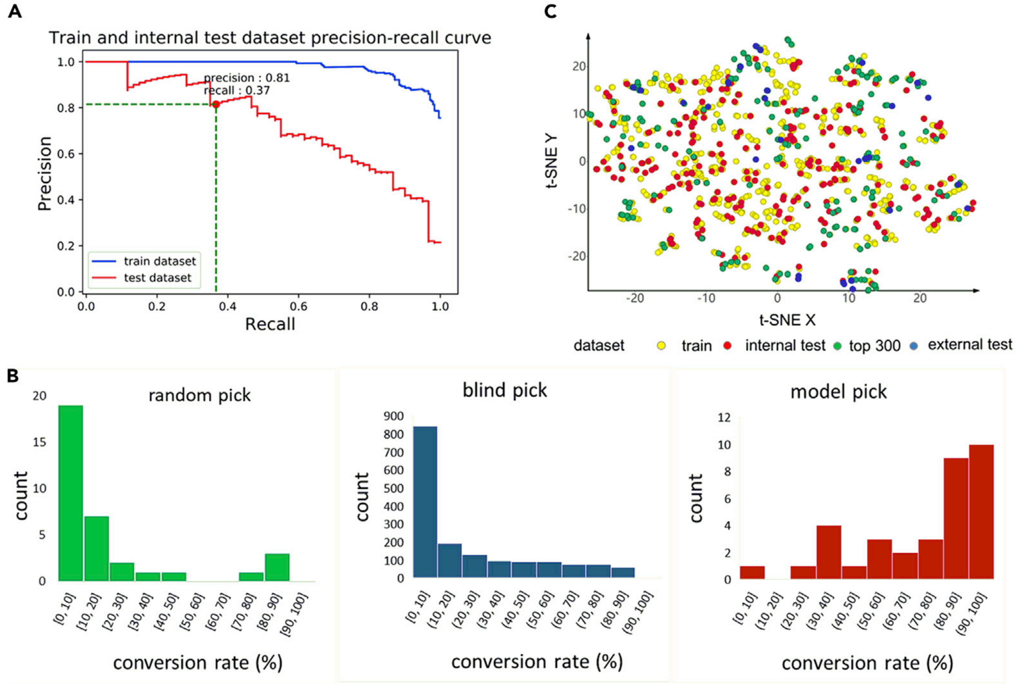


Figure 4. Performance of DNN and Comparison with Random Pick and Blind Pick

(A) Model performance on train dataset and internal test dataset.

(B) Left: conversion rate distribution of 34 randomly picked BBs. Middle: conversion rate distribution of 1,655 blindly picked BBs. Right: conversion rate distribution of 34 model picked BBs.

(C) Structure distribution of train dataset, internal test dataset, top 300 and external test dataset shown as t-SNE plot.

performance of ''random pick'' and ''model pick,'' a random-pick test was carried out, where subsets with the size of 34 BBs were randomly selected from the WuXi LabNetwork platform (BBs already in train and external dataset were discarded), and their conversion rates distribution was depicted in Table S4 and Figure 4B. The results verified that the performance of ''model pick'' is better than that of ''random pick.'' In addition, t-SNE result (Figure 4C) and structure clustering analysis (Table S7) showed that diverse of structures have been included in model-recommended BBs (top 300 and external test) compared with ''blind pick'' (train and internal test). In practice, more rigorous clustering selection can be made to make picked BBs as diverse as possible if enough BBs are available. We believe our above-described model serves as a proof of concept in how BBs can be filtered for purchase, particularly in cases of challenging reactions with otherwise low validation pass rates.

The Scope of Pictet-Spengler Reaction

After exploring the scope of aldehydes reacting with DNA-conjugated tryptamine **3a**, we further investigated differing DNA-conjugated tryptamines as shown in Figure 5A. We confirmed that the electronic effect of the substrate has an important influence on the reaction and unsurprisingly that the methoxy-substituted substrate (**3a**) gives the best result, whereas a bromosubstituted substrate gives almost no desired product. We then proceeded to investigate the reaction between a DNA-conjugated aryl aldehyde and different tryptamine substrates. At 80°C the reaction proceeded smoothly with good to excellent conversions (Figure 5B). For the on-DNA PS product **6c**, we have also carried out the amine capping with acetic acid and benzaldehyde, and both reactions provided the desired amine-capped products with acceptable conversions. This result indirectly demonstrates formation of the desired DNA-conjugated tryptamine PS product and also illustrates an interesting potential library design. (Note that, in contrast, capping of DNA-conjugate **4** with carboxylic acids or aldehydes is extremely difficult.)

In order to further explore the potential libraries employing this new on-DNA PS reaction condition, we designed DNA-conjugated indole amine **7**. The on-DNA PS reaction between **7** and 10 different aldehydes are

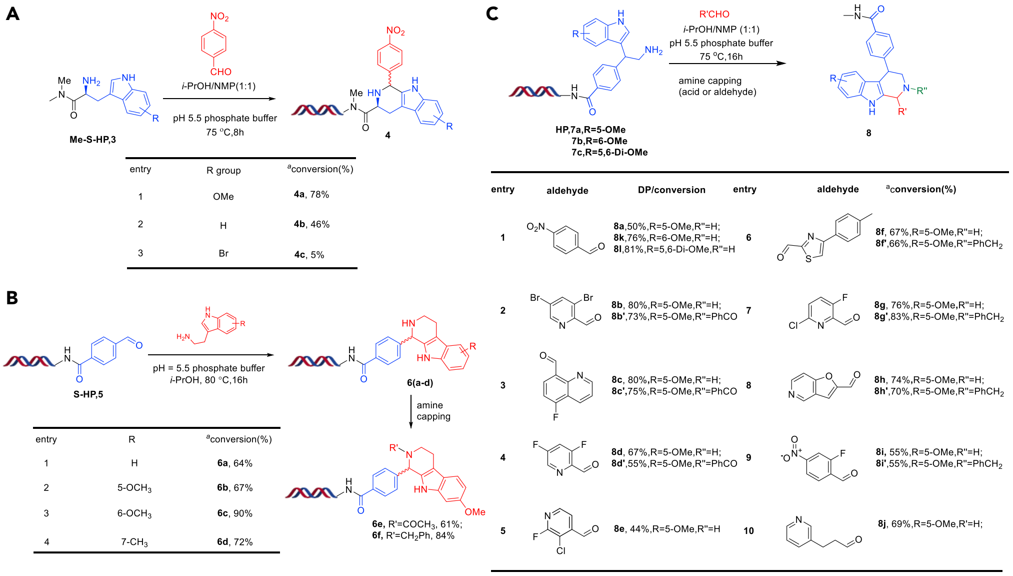

A
NO2
CHO
i-PrOH/NMP(1:1)
pH 5.5 phosphate buffer
75 °C,8h
Me NH2 NH
R
Me-S-HP,3
NO2
Me HN H N
R
4
entry | R group | aconversion(%)
1 | OMe | 4a, 78%
2 | H | 4b, 46%
3 | Br | 4c, 5%
B
H N R
H2N
pH = 5.5 phosphate buffer
i-PrOH, 80 °C,16h
S-HP,5
HN
HN
R
6(a-d)
amine capping
R'
N
HN
OMe
6e, R'=COCH3, 61%;
6f, R'=CH2Ph, 84%
entry | R | aconversion(%)
1 | H | 6a, 64%
2 | 5-OCH3 | 6b, 67%
3 | 6-OCH3 | 6c, 90%
4 | 7-CH3 | 6d, 72%
C
R'CHO
i-PrOH/NMP (1:1)
pH 5.5 phosphate buffer
75 °C,16h
amine capping
(acid or aldehyde)
HP,7a,R=5-OMe
7b,R=6-OMe
7c,R=5,6-Di-OMe
NH2
N-R''
R'
8
entry | aldehyde | DP/conversion | entry | aldehyde | aconversion(%)
1 | O2N | 8a,50%,R=5-OMe,R''=H; 8k,76%,R=6-OMe,R''=H; 8l,81%,R=5,6-Di-OMe,R''=H | 6 | N S | 8f, 67%,R=5-OMe,R''=H; 8f',66%,R=5-OMe,R''=PhCH2
2 | Br Br N | 8b, 80%,R=5-OMe,R''=H; 8b',73%,R=5-OMe,R''=PhCO | 7 | F Cl N | 8g, 76%,R=5-OMe,R''=H; 8g',83%,R=5-OMe,R''=PhCH2
3 | N F | 8c, 80%,R=5-OMe,R''=H; 8c',75%,R=5-OMe,R''=PhCO | 8 | N O | 8h, 74%,R=5-OMe,R''=H; 8h',70%,R=5-OMe,R''=PhCH2
4 | F F N | 8d, 67%,R=5-OMe,R''=H; 8d',55%,R=5-OMe,R''=PhCO | 9 | -O N+ O F | 8i, 55%,R=5-OMe,R''=H; 8i',55%,R=5-OMe,R''=PhCH2
5 | N F Cl | 8e, 44%,R=5-OMe,R''=H | 10 | N | 8j, 69%,R=5-OMe,R'=H;

Figure 5. Pictet-Spengler Reaction

(A) Electronic effect with Pictet-Spengler reaction.

(B) Pictet-Spengler reaction between on-DNA aldehyde and tryptamine substrates.

(C) Pictet-Spengler Reaction between DNA-conjugated indole-substituted amine and aldehyde.

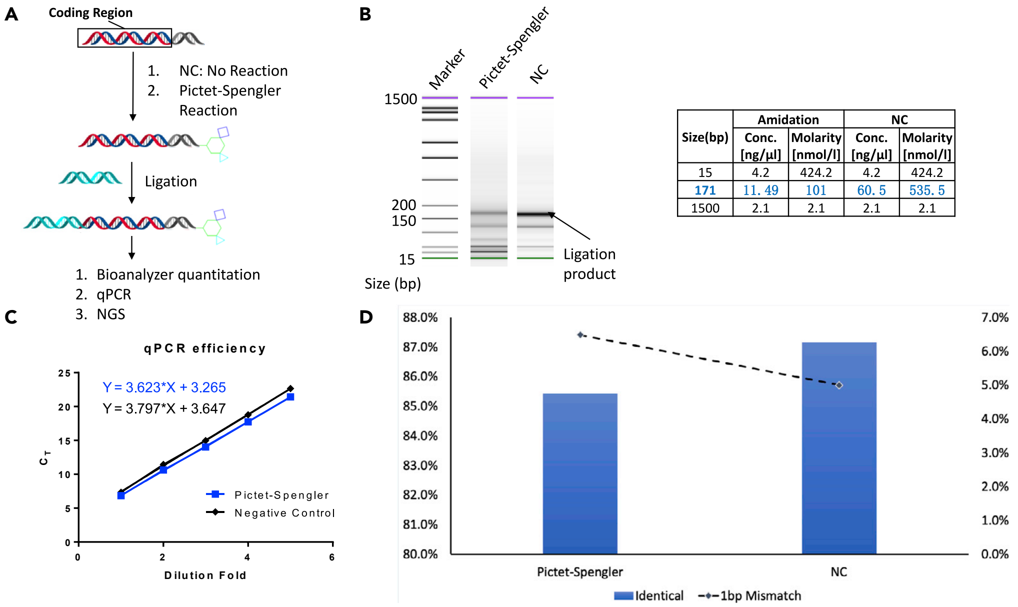
provided in Figure 5C. Again, the amine capping experiments were carried out employing both carboxylic acids and aldehydes; again, successful capping confirms the presence of the desired DNA-conjugated PS cyclized starting material and the ability to further diversity the scaffold during future library synthesis (Figure 5C).

DNA Damage Evaluation

A PS reaction was performed with a DNA-conjugated compound containing a double-stranded DNA coding region to mimic the library component, and the product was then ligated to an oligonucleotide to generate a full-length DNA fragment for qPCR analysis and next-generation sequencing (NGS) to assess the DNA integrity (Figure 6A). Ligation without any chemical reaction was used as a negative control separately. The ligation product was first examined with capillary electrophoresis (Bioanalyzer 2100, Agilent) to assess the size and quantity of the DNA strain. The result showed no shift of PS reaction product compared with the reactant, indicating no change of DNA size by PS reaction (Figure 6B). Then, the amplification efficiency was analyzed by qPCR (QuantStudio 7, Thermo Fisher) and the result suggested no significant change of DNA amplification efficiency when compared with the negative control groups. This result indicates that the PS reaction did not introduce unknown variation to the nucleotide, otherwise the efficiency will probably change between the two comparative groups (Figure 6C). In order to further assess the nucleotide-level integrity of the DNA, the ligation product was amplified and sequenced by the NGS. The NGS results suggested no significant modification of nucleotide from the PS reaction compared with the negative control NC (Figure 6D). In summary, the comprehensive DNA integrity assessment study demonstrated no damage to DNA by the PS reaction and, thus, could potentially be used for the DNA-encoded library construction (for details see Supplemental Information 5).

Potential DEL Library Synthetic Route

Lastly, we demonstrate proof-of-concept synthesis for two different three-cycle libraries (Figure 7A). For library 1, commercial aldehyde BBs would be purchased after filtering by our trained ML model. In library 2 (Figure 7B) we synthesize a different tryptoline scaffold, which is capable of being further diversified via acylation or alkylation. After on-DNA acylation to install the DNA-conjugated nitroalkene **13**, the addition reaction followed by nitro reduction yields the DNA-conjugated indole substituted amine **7**. The on-DNA PS reaction between **7** and a corresponding aldehyde provided the product **8** with good conversion (see the Supplemental Information). The amine capping of **8a** then provides further diversification of the THBC



Size(bp)	Amidation		NC	
	Conc. [ng/µl]	Molarity [nmol/l]	Conc. [ng/µl]	Molarity [nmol/l]
15	4.2	424.2	4.2	424.2
171	11. 49	101	60. 5	535. 5
1500	2.1	2.1	2.1	2.1

Figure 6. DNA Integrity was not Compromised by Pictet-Spengler Reaction

(A) Schematic illustration of the experiment design and workflow.

(B) Ligation and capillary electrophoresis, which suggest no DNA modification by Pictet-Spengler reaction.

(C) qPCR efficiency with CT as a function of dilution fold. No significant change of qPCR efficiency was observed on Pictet-Spengler reaction compared with the negative control groups.

(D) NGS data, suggesting no significant modification of DNA at the nucleotide level by Pictet-Spengler reaction.

scaffold (12 aldehydes were tested with 20%–95% conversions). These two diverse library designs demonstrate the potential of this novel on-DNA PS reaction.

In summary, with the rational design of the on-DNA indole substrates, we have developed the first DNA-compatible PS reaction for a variety of aldehydes under the optimized reaction conditions. Besides, suitable reaction conditions were identified for various combinations of PS reaction coupling partners. Moreover, a DNN model has been developed to make the prediction of the reaction conversion rate for the BBs, which was the first example of applying ML for the BB selections of the corresponding on-DNA reactions for DNA-encoded library synthesis. The detailed library production applying this new developed on-DNA PS reaction and selection results of the interesting biological interesting targets will be reported in due time.

Limitations of the Study

This reaction has certain limitations on the indole substrates; for example, the presence of methoxy group can make the PS reaction proceed smoothly. Besides, in order to have better prediction, a large set of reaction records need to be available.

Resource Availability

Lead Contact

XiaojieLu.

Materials Availability

See details in supplemental information.

Data and Code Availability

All the data has been attached in Supplemental information.

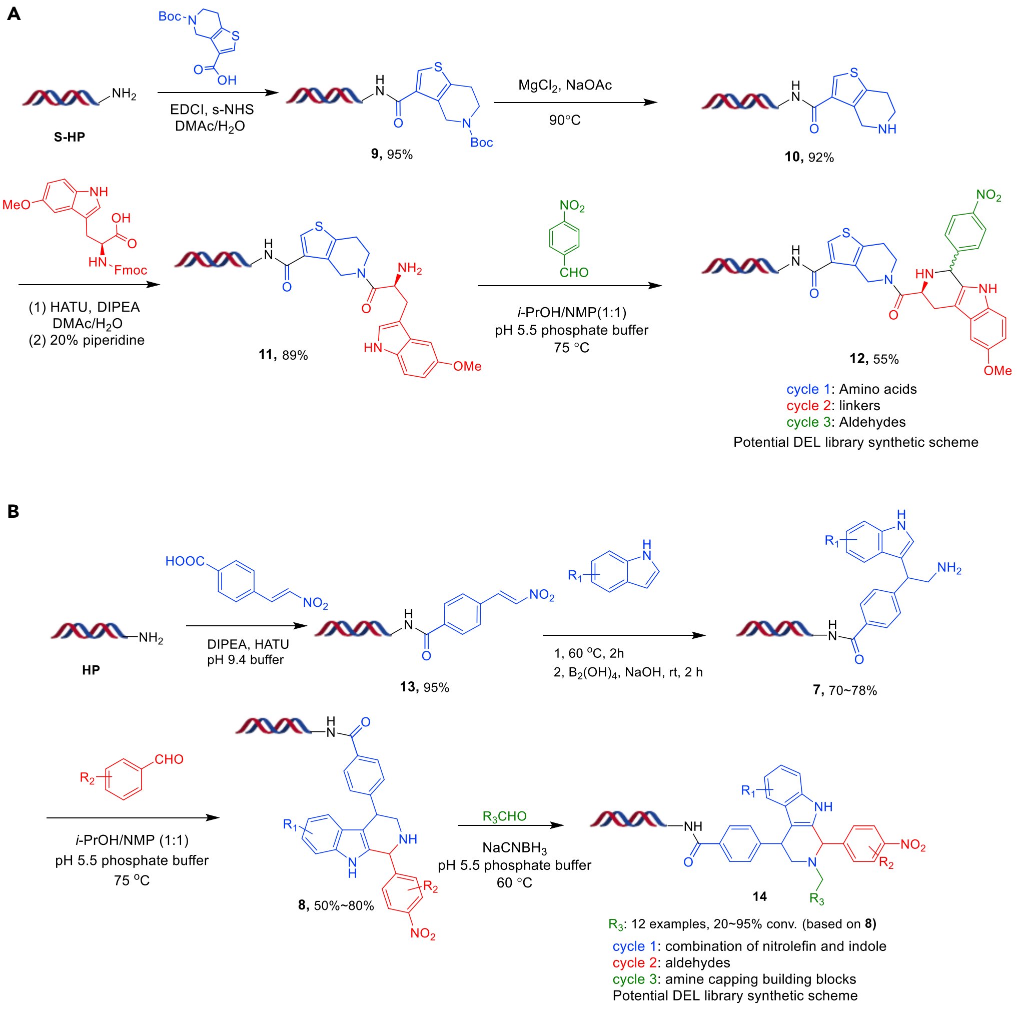


Figure 7. Potential DEL Library Synthetic Scheme

(A) Route for library 1.

(B) Route for library 2.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file. Full details of synthesis and LC-MS/MS analysis are provided in the Supplemental Information.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101142.

ACKNOWLEDGMENTS

We gratefully acknowledge financial support from the National Natural Science Foundation of China (81773634 to M.Z., 21877117 and 91953203 to X.L., 21907105 to Z.D.), National Science & Technology Major Project "Key New Drug Creation and Manufacturing Program," China (Number: 2018ZX09711002 to H.J., No. 2018ZX09711002-005 to X.L.), and "Personalized Medicines—Molecular Signature-based Drug Discovery and Development," Strategic Priority Research Program of the Chinese Academy of Sciences (XDA12050201 to M.Z.; XDA12040330 and XDA12050405 to X.L.), Shanghai Commission of Science and Technology (18431907100 to X.L.).

AUTHOR CONTRIBUTIONS

K.L., X. Liu, and S.L. contributed equally to this article. K.L. performed the PS reaction optimization, X. Liu. performed all the machine learning work, S.L. performed the validation and synthetic application; Y.A. and Y.S. performed the reaction optimization and validation; Q.S., X.S., W.S., and W.C. performed the DNA damage evaluation; Z.D. performed the synthetic application; L.K., H.Y., and A.S. guided the study and revised the manuscript; K.C. and H.J. guided the study; M.Z., X.P., and X. Lu conceived and designed the project and prepared the manuscript with feedback from all the authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 25, 2020
Revised: April 14, 2020
Accepted: May 4, 2020
Published: June 26, 2020

REFERENCES

Ahneman, D.T., Estrada, J.G., Lin, S., Dreher, S.D., and Doyle, A.G. (2018). Predicting reaction performance in C–N cross-coupling using machine learning. Science *360*, 186–190.

Bishop, C.M. (2006). Pattern Recognition and Machine Learning (Springer).

Buller, F., Mannocci, L., Scheuermann, J., and Neri, D. (2010). Drug discovery with DNA-encoded chemical libraries. Bioconjug. Chem. *21*, 1571–1580.

Chouikhi, D., Ciobanu, M., Zambaldo, C., Duplan, V., Barluenga, S., and Winssinger, N. (2012). Expanding the scope of PNA-encoded synthesis (pes): Mtt-protected PNA fully orthogonal to Fmoc chemistry and a broad array of robust diversity generating reactions. Chem. Eur. J. *18*, 12698–12704.

Clark, M.A., Charya, R.A., Arico-Muendel, C.C., Belyanskaya, S.L., Benjamin, D.R., Carlson, N.R., Centrella, P.A., Chiu, C.H., Creaser, S.P., Cuozzo, J.W., et al. (2009). Design, synthesis and selection of DNA-encoded small-molecule libraries. Nat. Chem. Biol. *5*, 647–654.

David, R., and Mathew, H. (2010). Extended-connectivity fingerprints. J. Chem. Inf. Model. *50*, 742–754.

Dichson, P., and Kodadek, T. (2019). Chemical composition of DNA-encoded libraries, past present and future. Org. Biomol. Chem. *17*, 4676–4688.

Dipak, P., and Mukut, G. (2008). Iodine-catalyzed highly effective Pictet–Spengler condensation: an efficient synthesis of tetrahydro-β-carbolines. Synth. Commun. *38*, 4426–4433.

Du, H.C., Bangs, M.C., Simmons, N., and Matzuk, M.M. (2019). Multistep synthesis of 1,2,4-oxadiazoles via DNA-conjugated aryl nitrile substrates. Bioconjug. Chem. *30*, 1304–1308.

Eidam, O., and Satz, A.L. (2016). Analysis of the productivity of DNA encoded libraries. Med. Chem. Commun. *7*, 1323–1331.

Encinas, L., O'Keefe, H., Neu, M., Remuinan, M.J., Patel, A.M., Guardia, A., Davie, C.P., Perez-Macias, N., Yang, H., Convery, M.A., et al. (2014). Encoded library technology as a source of hits for the discovery and lead optimization of a potent and selective class of bactericidal direct inhibitors of *Mycobacterium tuberculosis* InhA. J. Med. Chem. *57*, 1276–1288.

Favalli, N., Bassi, G., Scheuermann, J., and Neri, D. (2018). DNA-encoded chemical libraries-achievements and remaining challenges. FEBS Lett. *592*, 2168–2180.

Faver, J.C., Riehle, K., Lancia, D.R., Milbank, J.B.J., Kollmann, C.S., and Simmons, N. (2019). Quantitative comparison of enrichment from DNA-encoded chemical library selections. ACS Comb. Sci. *21*, 75–82.

Flood, D.T., Asai, S., Zhang, X., Wang, J., Yoon, L., Adams, Z.C., Dillingham, B.C., Sanchez, B.B., Vantourout, J.C., Flanagan, M.E., et al. (2019). Expanding reactivity in DNA-encoded library synthesis via reversible binding of DNA to an Inert quaternary ammonium support. J. Am. Chem. Soc. *141*, 9998–10006.

Franzini, R.M., and Randolph, C. (2016). Chemical space of DNA-encoded libraries. J. Med. Chem. *59*, 6629–6644.

Gong, Z., Hu, G., Li, Q., Liu, Z., Wang, F., Zhang, X., Xiong, J., Li, P., Xu, Y., Ma, R., et al. (2017). Compound libraries: recent advances and their applications in drug discovery. Curr. Drug Discov. Tech. *14*, 216–228.

Jiménez, J., and Ginebra, J. (2017). PyGPGO: Bayesian optimization for python. J. Open Source Softw. *2*, 431.

Kingma, D.P., and Ba, J. (2014). Adam: a method for stochastic optimization. arXiv e-prints, arXiv:1412.6980.

Kite, S., Hattorib, T., and Murakamib, Y. (1994). Estimation of catalytic performance by neural network-product distribution in oxidative dehydrogenation of ethylbenzene. Appl. Catal. *114*, 173–178.

LaPlante, S.R., Carson, R., Gillard, J., Aubry, N., Coulombe, R., Bordeleau, S., Bonneau, P., Little, M., O'Meara, J., and Beaulieu, P.L. (2013). Compound aggregation in drug discovery: implementing a practical NMR assay for medicinal chemists. J. Med. Chem. *56*, 5142–5150.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature *521*, 436–444.

Lerner, R.A., Ma, P., Xu, H., Li, J., Lu, F., Ma, F., Wang, S., Xiong, H., Wang, W., Buratto, D., et al. (2019). Functionality-independent DNA encoding of complex natural products. Angew. Chem. Int. Ed. *58*, 9254–9261.

Li, H., Sun, Z., Wu, W., Wang, X., Zhang, M., Lu, X., Zhong, W., and Dai, D. (2018). Inverse-Electron-Demand Diels–Alder reactions for the synthesis of pyridazines on DNA. Org. Lett. *20*, 7186–7191.

Liu, F., Wang, H., Li, S., Bare, G.A.L., Chen, X., Wang, C., Moses, J.E., Wu, P., and Sharpless, K.B. (2019). Biocompatible SuFEx click chemistry: thionyl tetrafluoride (SOF4)-derived connective hubs for bioconjugation to DNA and proteins. Angew. Chem. Int. Ed. *58*, 8029–8033.

Lu, X., Fan, L., Phelps, C.B., Davie, C.P., and Donahue, C.P. (2017a). Ruthenium promoted On-DNA ring-closing metathesis and cross-metathesis. Bioconjug. Chem. *28*, 1625–1629.

Lu, X., Roberts, S., Franklin, G.J., and Davie, C. (2017b). On-DNA Pd and Cu promoted C–N cross-coupling reactions. Med. Chem. Commun. *8*, 1614–1617.

Maity, P., Adhikari, D., and Jana, A.K. (2019). An overview on synthetic entries to tetrahydro-β-carbolines. Tetrahedron *75*, 965–1028.

Malone, M.L., and Paegel, B.M. (2016). What is a "DNA-compatible" Reaction? ACS Comb. Sci. *18*, 182–187.

Montavon, G., Orr, G.B., and Müller, K.R. (2012). Neural Networks: Tricks of the Trade, Second Edition (Springer).

Neri, D., and Lerner, R.A. (2018). DNA-encoded chemical libraries: a selection system based on endowing organic compounds with amplifiable information. Annu. Rev. Biochem. *87*, 479–502.

Omata, K., and Yamada, M. (2004). Prediction of effective additives to a Ni/Active carbon catalyst for vapor-phase carbonylation of methanol by an artificial neural network. Ind. Eng. Chem. Res. *43*, 6622–6625.

Ottl, J., Leder, L., Schaefer, J.V., and Dumelin, C.E. (2019). Encoded library technologies as Integrated lead finding platforms for drug discovery. Molecules *24*, 1629–1650.

Raccuglia, P., Elbert, K.C., Adler, P.D., Falk, C., Wenny, M.B., Mollo, A., Zeller, M., Friedler, S.A., Schrier, J., and Norquist, A.J. (2016). Machine-learning-assisted materials discovery using failed experiments. Nature *533*, 73–76.

Ralph, E.K., Christoph, E.D., and David, R.L. (2011). Small-molecule discovery from DNA-encoded chemical libraries. Chem. Soc. Rev. *40*, 5707–5717.

Reddavide, F.V., Cui, M., Lin, W., Fu, N., Heiden, S., Andrade, H., Thompson, M., and Zhang, Y. (2019). Second generation DNA-encoded dynamic combinatorial chemical libraries. Chem. Commun. *55*, 3753–3756.

Roman, J.P., Haro, R., Blas, J.D., Jessop, T.C., and Castanon, J. (2018). Design and development of a technology platform for DNA-encoded library production and affinity selection. SLAS Discov. *23*, 387–396.

Satz, A.L., Cai, J., Chen, Y., Goodnow, R., Gruber, F., Kowalczyk, A., Petersen, A., Naderi-Oboodi, G., Orzechowski, L., and Strebel, Q. (2015). DNA compatible multistep synthesis and applications to DNA encoded libraries. Bioconjug. Chem. *26*, 1623–1632.

Skopic, M.K., Salamon, H., Bugain, O., Jung, K., Gohla, A., Doetsch, L.J., Dos Santos, D., Bhat, A., Wagner, B., and Brunschweiger, A. (2017). Acid- and Au(i)-mediated synthesis of hexathymidine-DNA-heterocycle chimeras, an efficient entry to DNA-encoded libraries inspired by drug structures. Chem. Sci. *8*, 3356–3361.

Skoraczynski, G., Dittwald, P., Miasojedow, B., Szymkuc, S., Gajewska, E.P., Grzybowski, B.A., and Gambin, A. (2017). Ror2 signaling regulates golgi structure and transport through IFT20 for tumor invasiveness. Sci. Rep. *7*, 3582, 3542.

Srinivasan, N., and Ganesan, A. (2003). A highly efficient Lewis acid-catalysed Pictet–Spengler reactions discovered by parallel screening. Chem. Commun. *7*, 916–917.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. *15*, 1929–1958.

Skopic, M.K., Götte, K., Gramse, C., Dieter, M., Pospich, S., Raunser, S., Weberskirch, R., and Brunschweiger, A. (2019). Micellar Bronsted acid mediated synthesis of DNA-tagged heterocycles. J. Am. Chem. Soc. *141*, 10546–10555.

Wang, J., Lundberg, H., Asai, S., Martin-Acosta, P., Chen, J.S., Brown, S., Farrell, W., Dushin, R.G., O'Donnell, C.J., Ratnayake, A.S., et al. (2018a). Kinetically guided radical-based synthesis of C (sp^3)-C (sp^3) linkages on DNA. Proc. Natl. Acad. Sci. U S A *115*, E6404–E6410.

Wang, X., Sun, H., Liu, J., Dai, D., Zhang, M., Zhou, H., Zhong, W., and Lu, X. (2018b). Ruthenium-promoted C−H activation reactions between DNA conjugated acrylamide and aromatic Acids. Org. Lett. *20*, 4764–4768.

Wang, X., Sun, H., Liu, J., Dai, D., Zhang, M., Zhou, H., Zhong, W., and Lu, X. (2019). Palladium-promoted DNA-compatible Heck reaction. Org. Lett. *21*, 719–723.

Xu, H., Ma, F., Wang, N., Hou, W., Xiong, H., Lu, F., Li, J., Wang, S., Ma, P., Yang, G., and Lerner, R.A. (2019). DNA-Encoded Libraries: aryl fluorosulfonates as versatile electrophiles enabling facile On-DNA Suzuki, Sonogashira, and Buchwald reactions. Adv. Sci. *23*, 1901551–1901556.

Yamamura, A., Fujitomi, E., Ohara, N., Tsukamoto, K., Sato, M., and Yamamura, H. (2017). Tadalafil induces antiproliferation, apoptosis, and phosphodiesterase type 5 downregulation in idiopathic pulmonary arterial hypertension in vitro. Eur. J. Pharm. *810*, 44–50.

Yuen, L.H., Dana, S., Liu, Y., Bloom, S.I., Thorsell, A.G., Neri, D., Donato, A.J., Kireev, D.B., Schuler, H., and Franzini, R.M. (2019). A focused DNA-encoded chemical library for the discovery of Inhibitors of NAD^+-Dependent enzymes. J. Am. Chem. Soc. *141*, 5169–5181.

Zambaldo, C., Geigle, S.N., and Satz, A.L. (2019). High-throughput solid-phase building block synthesis for DNA-encoded libraries. Org. Lett. *21*, 9353–9357.

Zhou, Y., Li, C., Peng, J., Xie, L., Meng, L., and Li, Q. (2018). DNA-encoded dynamic chemical library and its applications in Ligand Discovery. J. Am. Chem. Soc. *140*, 15859–15867.

iScience, Volume *23*

Supplemental Information

Solution-Phase DNA-Compatible Pictet-Spengler Reaction Aided by Machine Learning Building Block Filtering

Ke Li, Xiaohong Liu, Sixiu Liu, Yulong An, Yanfang Shen, Qingxia Sun, Xiaodong Shi, Wenji Su, Weiren Cui, Zhiqiang Duan, Letian Kuai, Hongfang Yang, Alexander L. Satz, Kaixian Chen, Hualiang Jiang, Mingyue Zheng, Xuanjia Peng, and Xiaojie Lu

Solution Phase DNA-Compatible Pictet-Spengler Reaction Aided By Machine Learning Building Block Filtering

Ke Li, [2] Xiaohong Liu, [1, 3, 4, 5] Sixiu Liu, [1, 5] Yulong An, [2] Yanfang Shen, [2] Qingxia Sun,[2] Xiaodong Shi, [2] Wenji Su, [2] Weiren Cui, [2] Zhiqiang, Duan, [1,5] Letian Kuai, [2] Hongfang Yang, [2] Alexander L. Satza, [2] Kaixian Chen, [1,3,4,5] Hualiang Jiang, [1,3,4,5] Mingyue Zheng, [1,4,5*] Xuanjia Peng,[2*] Xiaojie Lu[1,5*]

[1]State Key Laboratory of Drug Research, Shanghai Institute of MateriaMedica, Chinese Academy of Sciences, 501 Haike Road, Zhang Jiang Hi-Tech Park, Pudong, Shanghai, P. R. China 201203

[2]DNA Encoded Library Platform, WuXi AppTec, 288 FuteZhong Road, Waigaoqiao Free Trade Zone, Shanghai 200131, China

[3]School of Life Science and Technology, ShanghaiTechUniversity, Shanghai, China,and Shanghai Institute for Advanced Immunochemical Studies, and School of Life Science and Technology, ShanghaiTech University

[4]Drug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of MateriaMedica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, China

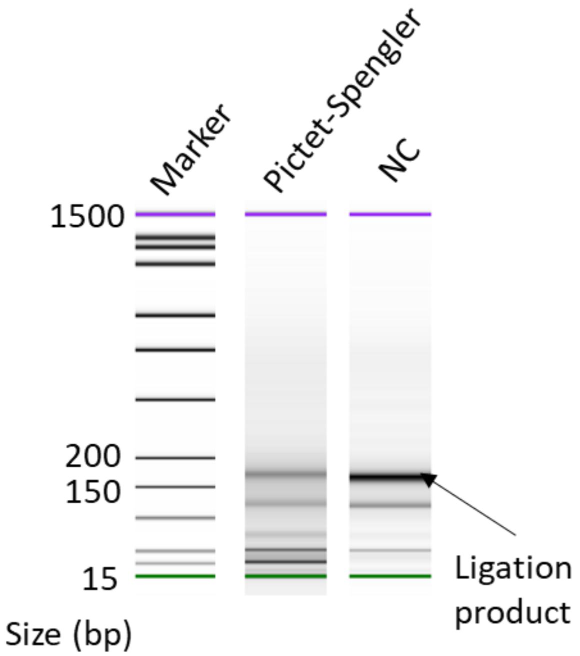
[5]University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, China

Supporting Information

Content

Figures

Figure S1 Capillary gel electrophoresis result of ligations, related to Figure 6.



Size(bp)	Pictet-Spengler		NC	
	Conc. [ng/µl]	Molarity [nmol/l]	Conc. [ng/µl]	Molarity [nmol/l]
15	4.2	424.2	4.2	424.2
171	**11.49**	**101**	**60.5**	**535.5**
1500	2.1	2.1	2.1	2.1

Fig. S1 Capillary gel electrophoresis result of ligations

Figure S2 qPCR data summary of the concentration check group, related to Figure 6.

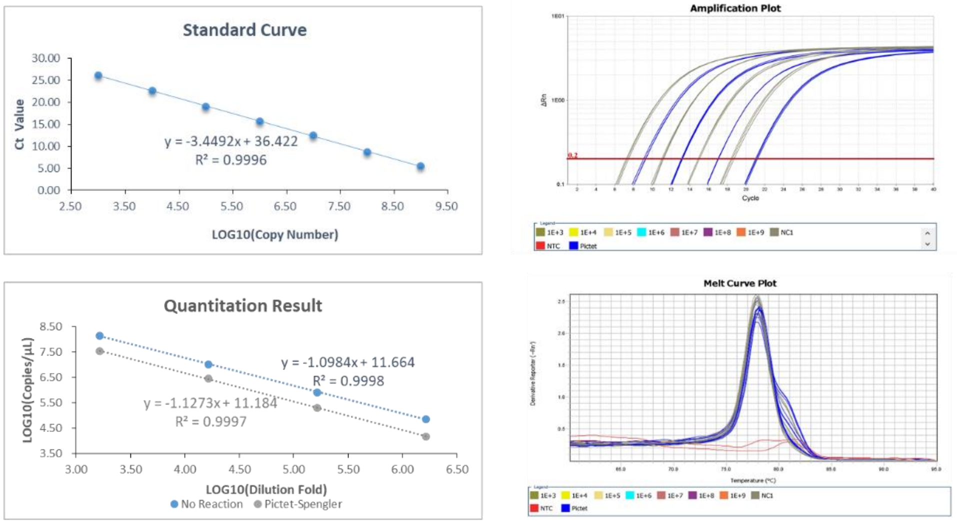
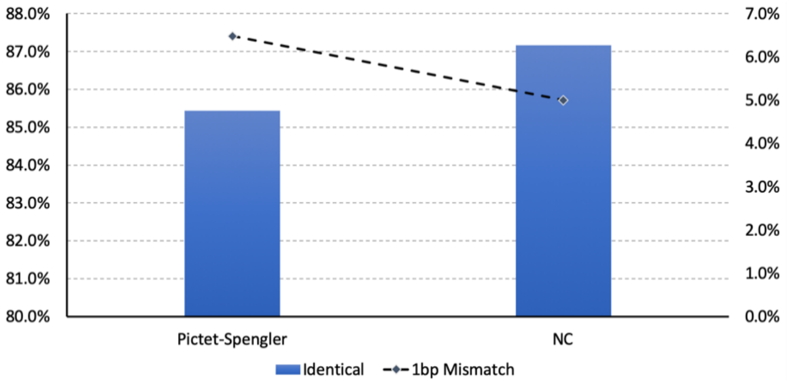


Fig. S2 qPCR data summary of the concentration check group

Figure S3 Statistics of next-generation sequencing results. The left Y-axis is the fraction of identical reads from perfect match, while the right Y-axis is the fraction of 1bp mismatch, related to Figure 6.

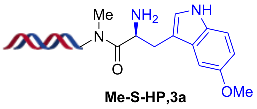


Condition	sequenced reads	perfect match	1bp mismatch
Pictet-Spengler	141,183,054	120,623,714	9,151,620
NC	121,974,103	106,321,624	6,097,462

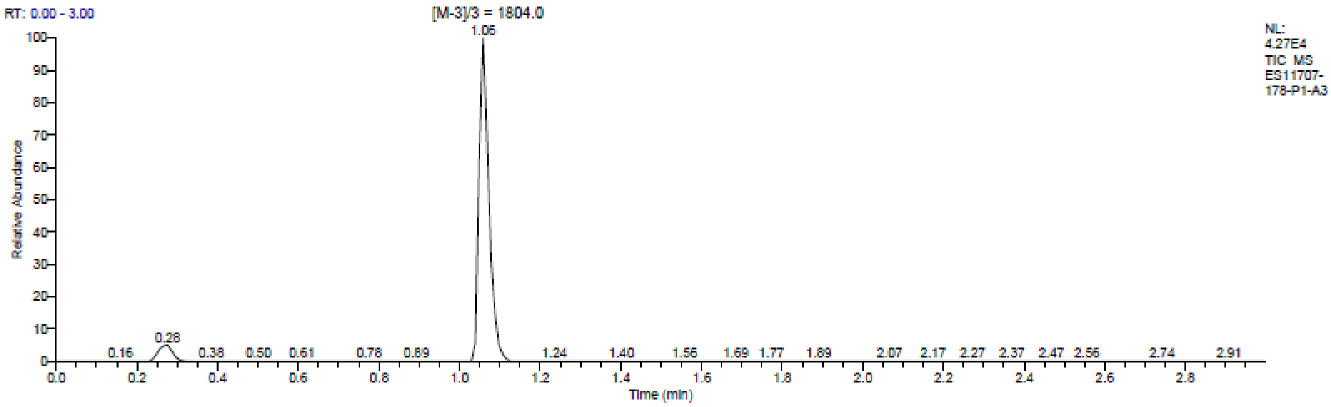
Fig. S3 Statistics of next-generation sequencing results. The left Y-axis is the fraction of identical reads from perfect match, while the right Y-axis is the fraction of 1bp mismatch.

Figure S4, Trace and Mass of 3a, related to Figure 3.

Following **General Procedure 1**
Purity: >99.00%
Exact mass: 5414.97
Triply charged mass [M-3]/3, calculated: 1803.99; observed:1804.0



Me-S-HP,3a



RT	Peak Height	Peak Width	Peak Area	Area %
1.06	28560.25	0.12	72829.86	100.00

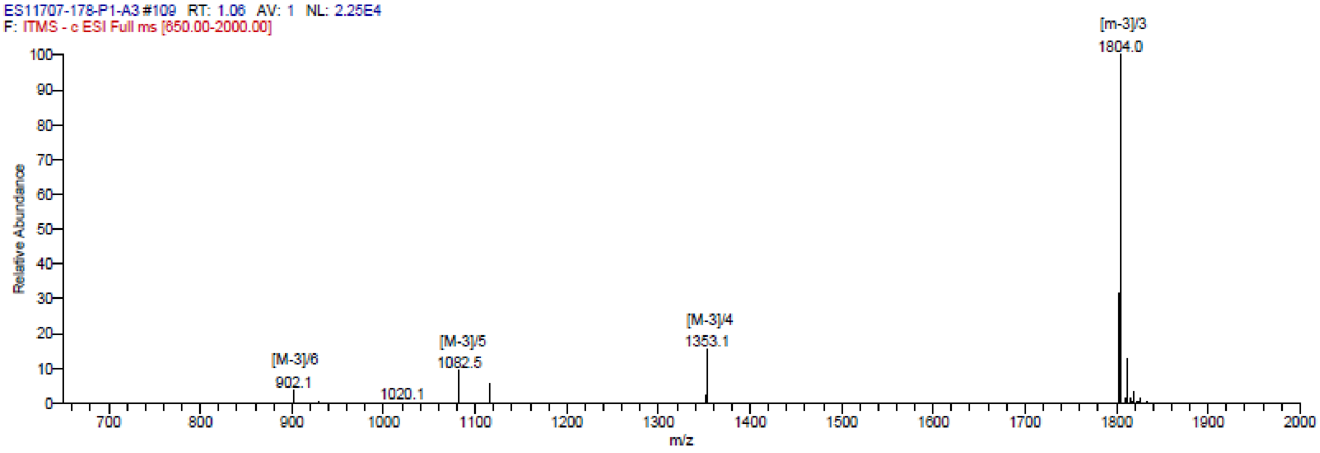


Fig. S4. LC trace and mass of **3a**.

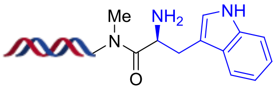
Figure S5, Trace and Mass of 3b, related to Figure 3.

Following **General Procedure 1**

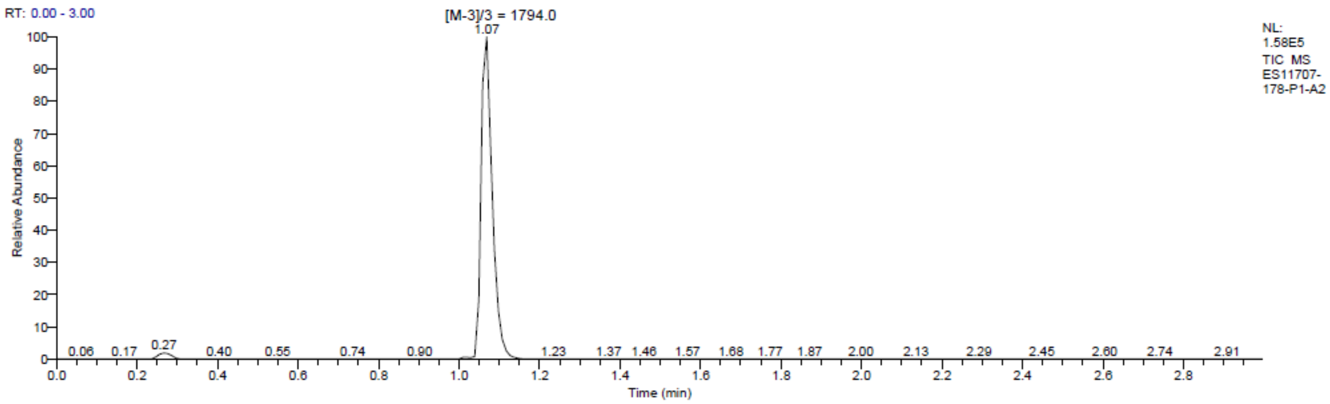
Purity: >99.00%

Exact mass: 5384.95

Triply charged mass [M-3]/3, calculated: 1793.98; observed:1794.0



Me-S-HP,3b



RT	Peak Height	Peak Width	Peak Area	Area %
1.07	115477.47	0.14	307138.72	100.00

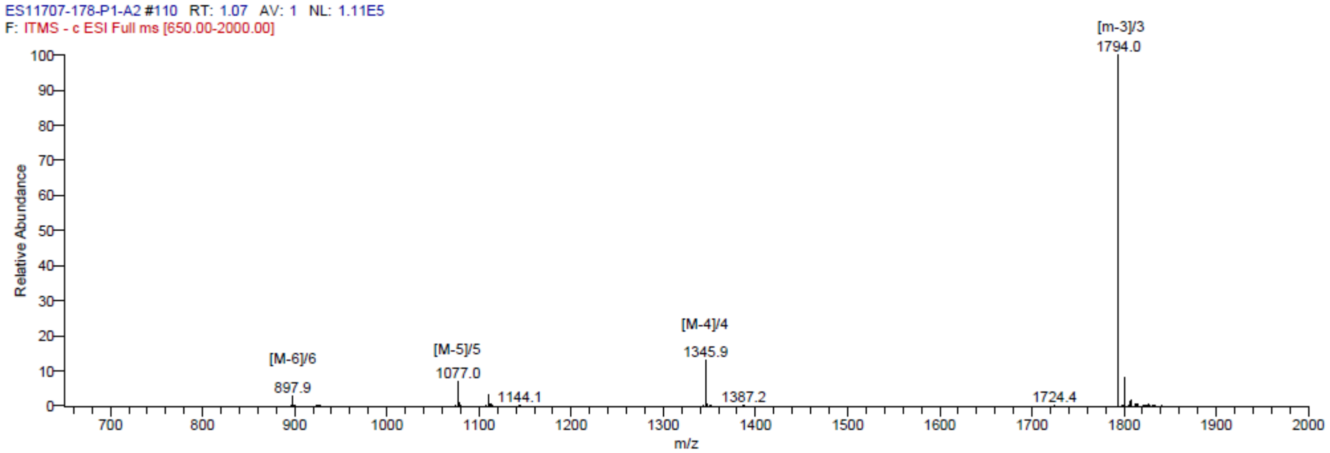


Fig. S5. LC trace and mass of **3b**.

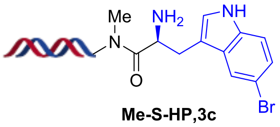
Figure S6, Trace and Mass of 3c, related to Figure 3.

Following **General Procedure 1**

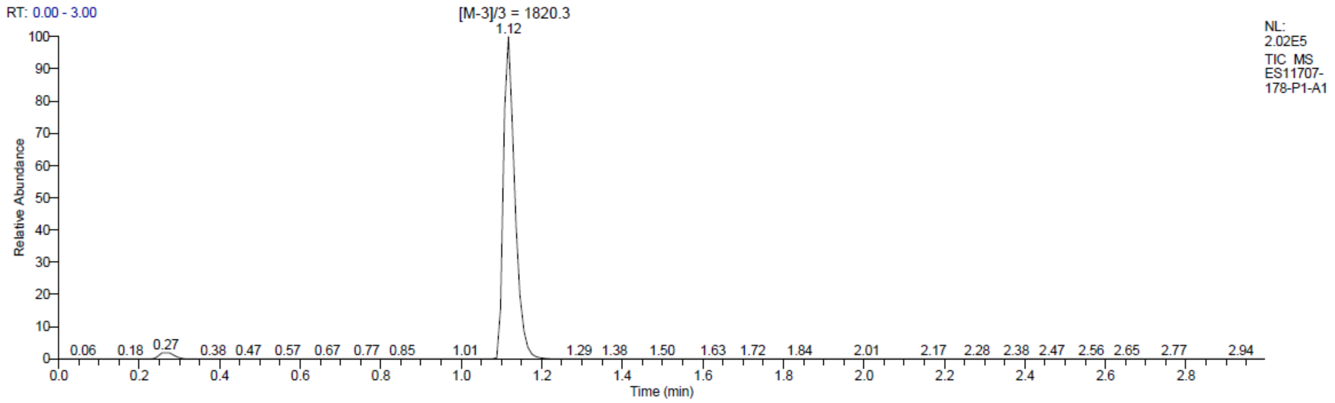
Purity: >99.00%

Exact mass: 5463.84

Triply charged mass [M-3]/3, calculated: 1820.28; observed:1820.3



Me-S-HP,3c



RT	Peak Height	Peak Width	Peak Area	Area %
1.12	147851.50	0.13	400119.20	100.00

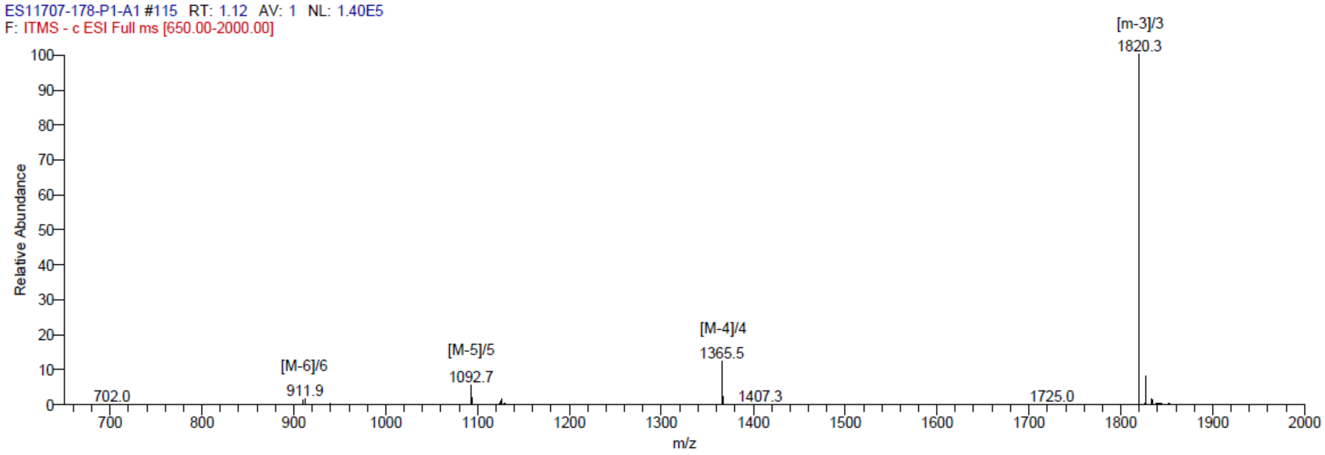
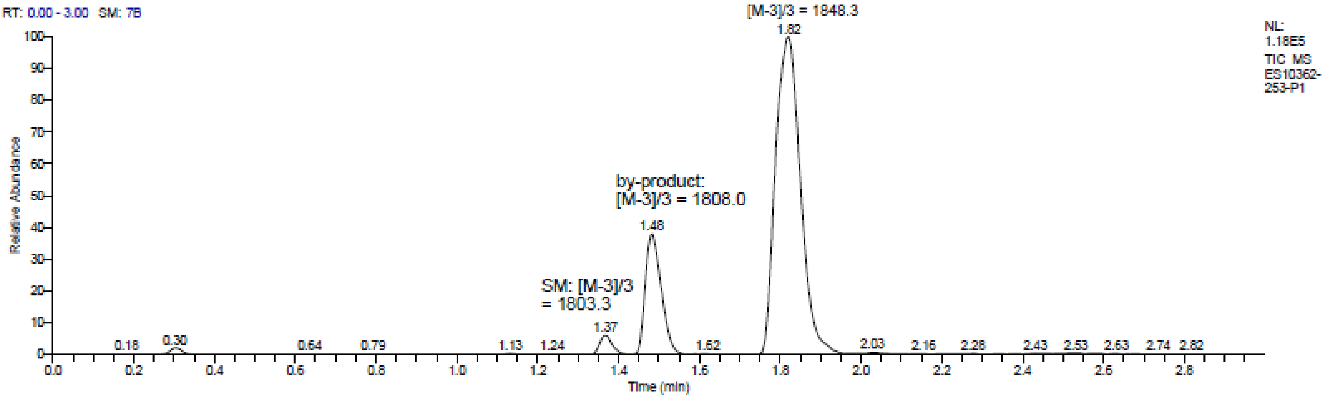


Fig. S6. LC trace and mass of **3c**.

Figure S7, Trace and Mass of 4a, related to Figure 3.

Following **General Procedure 2**
Percent conversion: 78.54%
Exact mass: 5548.09
Triply charged mass [M-3]/3, calculated: 1848.36; observed:1848.3



RT	Peak Height	Peak Width	Peak Area	Area %
1.37	7522.86	0.10	14770.99	2.30
1.48	45900.65	0.14	122931.83	19.16
1.82	119469.26	0.22	503847.65	78.54

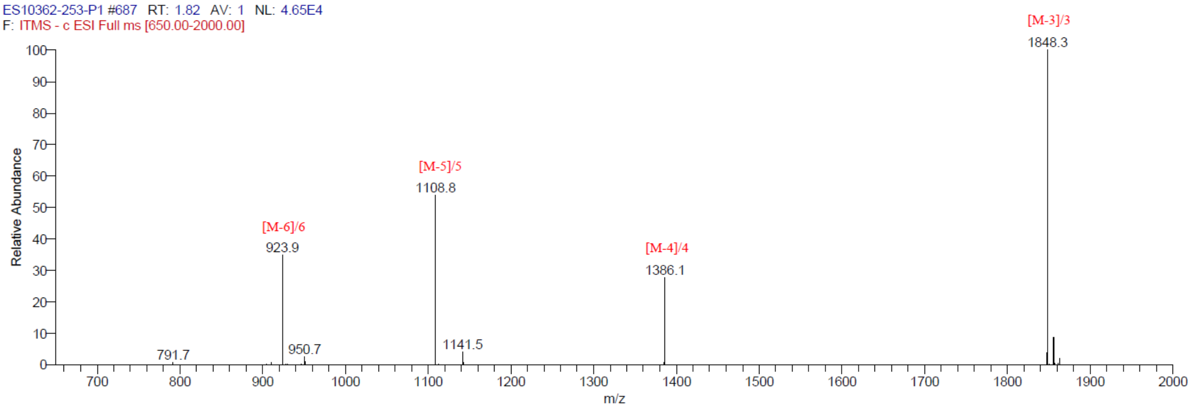


Fig. S7. LC trace and mass of **4a**.

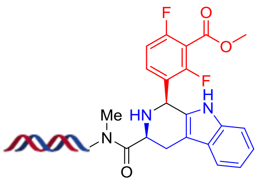
Figure S8, Trace and Mass of 4aa and 4aa’, related to Figure 3.

Following **General Procedure 2**

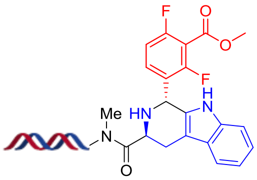
Percent conversion: 56.77%& 38.02%, totally 94.79%

Exact mass: 5596.84

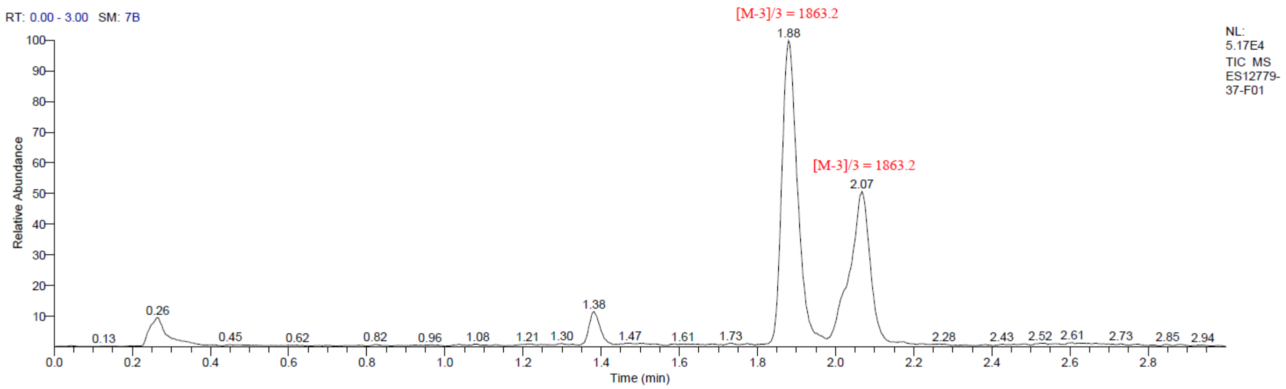
Triply charged mass [M-3]/3, calculated: 1864.61; observed:1863.2&1863.2



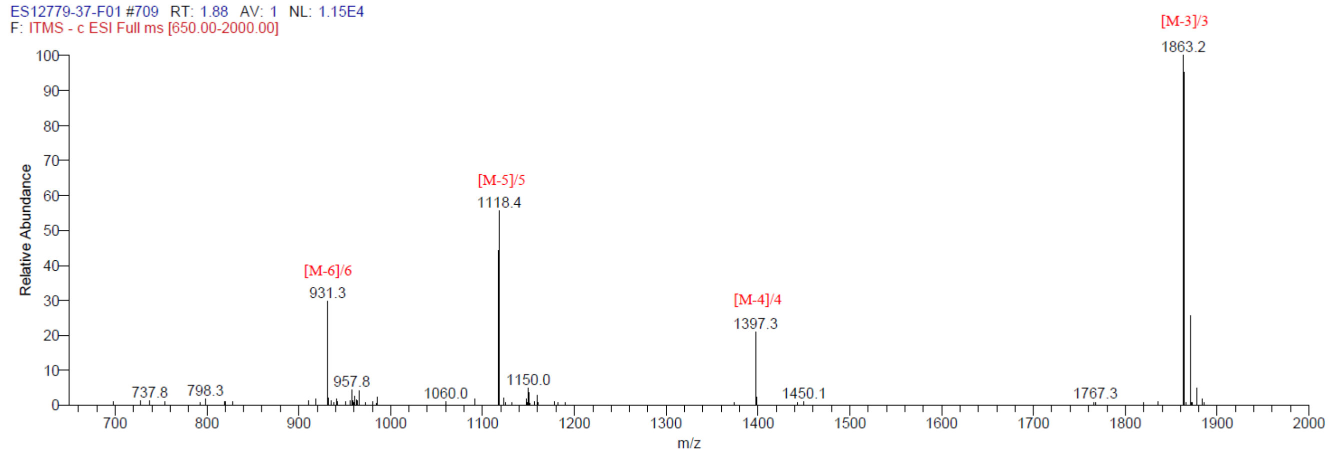
Me-S-HP,4aa



Me-S-HP,4aa'



RT	Peak Height	Peak Width	Peak Area	Area %
1.38	5996.54	0.20	13113.83	5.21
1.88	53406.53	0.15	142914.51	56.77
2.07	26972.42	0.22	95694.35	38.02



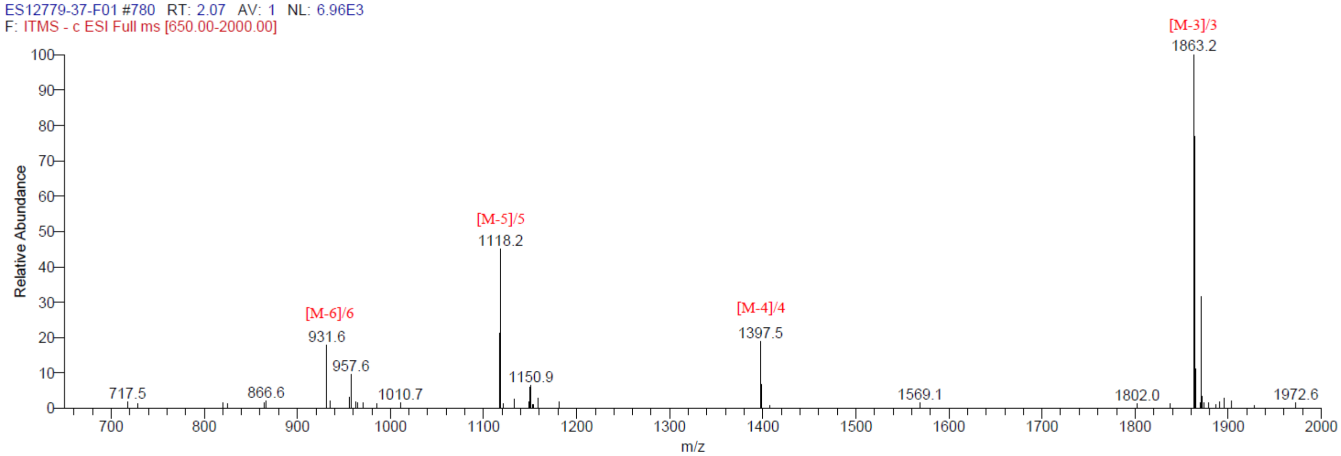
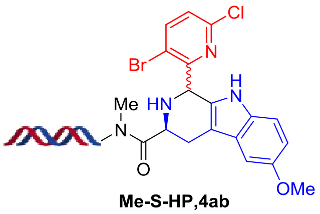


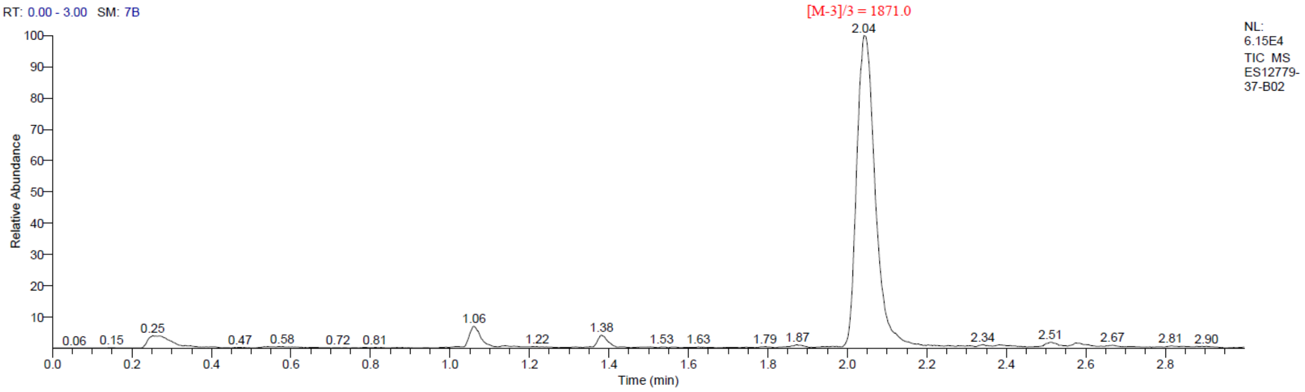
Fig. S8. LC trace and mass of **4aa and 4aa'**.

Figure S8, Trace and Mass of 4ab, related to Figure 3.

Following **General Procedure 2**
Percent conversion: 91.52%
Exact mass: 5617.42
Triply charged mass [M-3]/3, calculated: 1871.47; observed:1871.0



Me-S-HP,4ab



RT	Peak Height	Peak Width	Peak Area	Area %
1.06	4230.85	0.28	10472.04	4.84
1.38	2757.17	0.10	4460.32	2.06
2.05	62186.68	0.24	197967.26	91.52
2.51	981.50	0.08	2021.16	0.93
2.58	993.09	0.05	1397.73	0.65

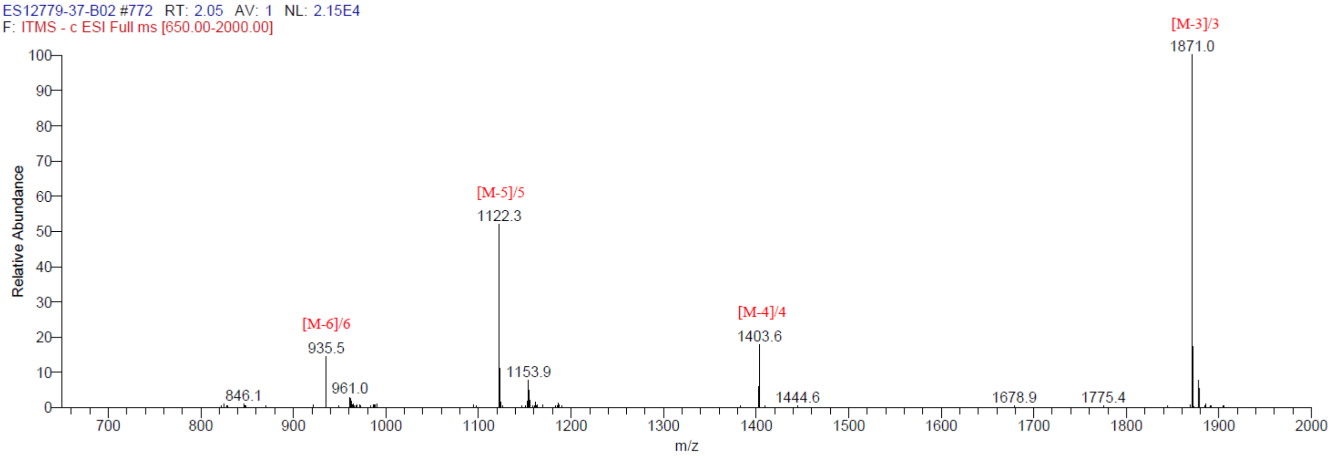


Fig. S9. LC trace and mass of **4ab**

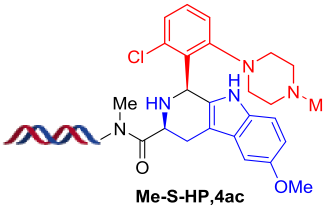
Figure S10, Trace and Mass of 4ac and 4ac', related to Figure 3.

Following **General Procedure 2**

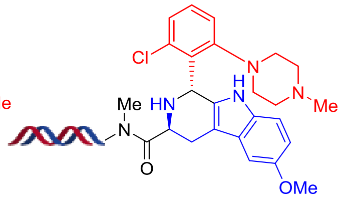
Percent conversion: 25.47% & 69.54%, totally 95.01%

Exact mass: 5635.69

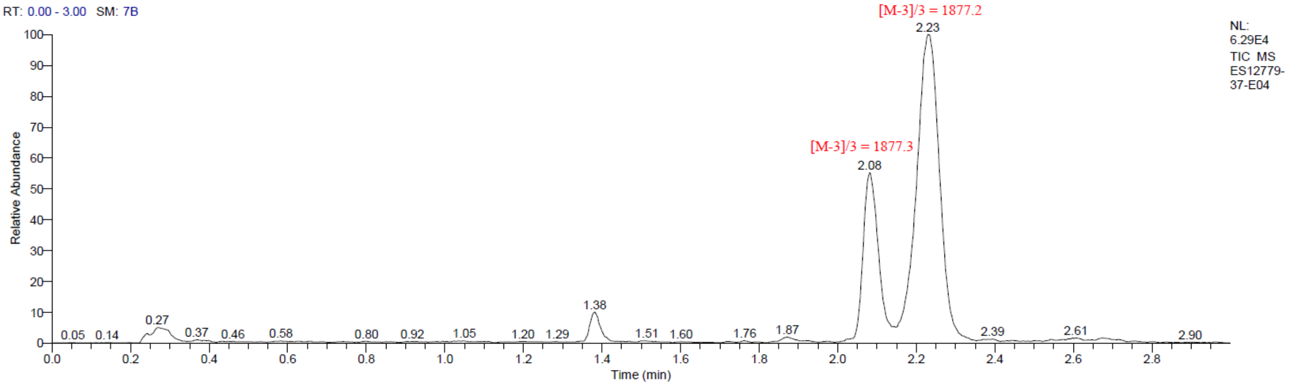
Triply charged mass [M-3]/3, calculated: 1877.56; observed:1877.3&1877.2



Me-S-HP,4ac



Me-S-HP,4ac'



RT	Peak Height	Peak Width	Peak Area	Area %
1.38	6465.13	0.13	12013.33	3.26
1.87	1201.29	0.12	3413.85	0.93
2.08	35553.80	0.11	93989.70	25.47
2.24	62824.80	0.20	256594.28	69.54
2.67	856.85	0.10	2970.67	0.81

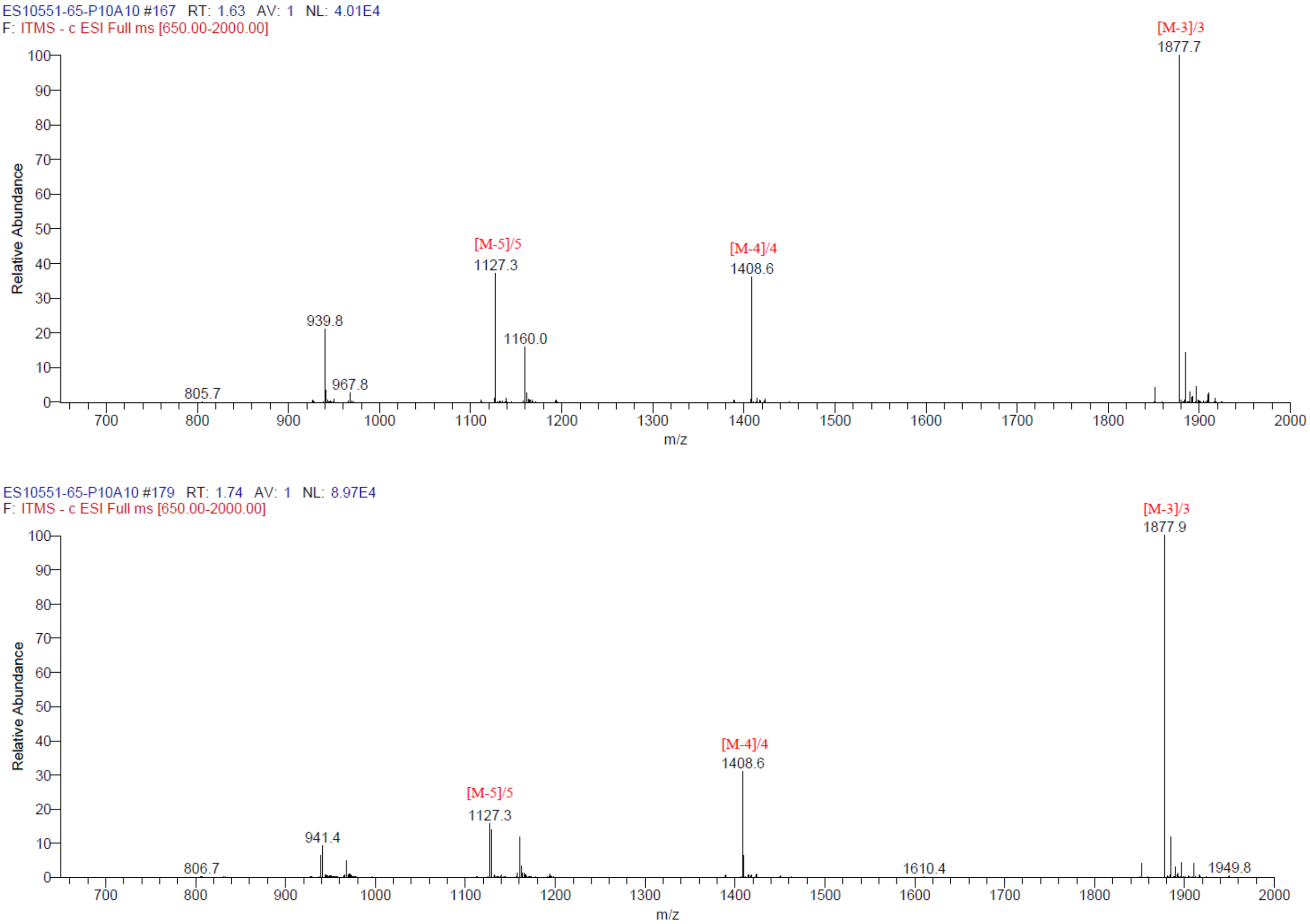
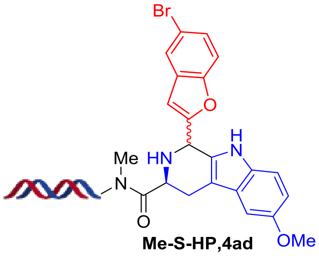


Fig. S10. LC trace and mass of **4ac and 4ac'**

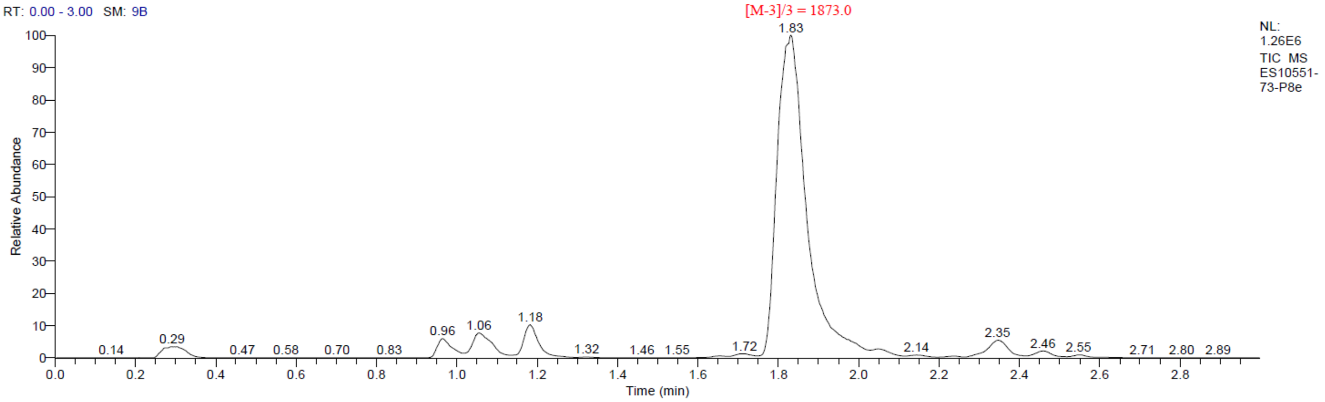
Figure S11, Trace and Mass of 4ad, related to Figure 3.

Following **General Procedure 2**
Percent conversion: 85.01%
Exact mass: 5622.01

Triply charged mass [M-3]/3, calculated: 1873.0; observed:1873.0



Me-S-HP,4ad



RT	Peak Height	Peak Width	Peak Area	Area %
0.96	84020.35	0.16	203731.35	2.63
1.06	109847.24	0.11	340623.44	4.39
1.18	145603.79	0.26	364129.24	4.69
1.83	1307271.23	0.36	6596635.55	85.01
2.34	70043.20	0.16	254270.02	3.28

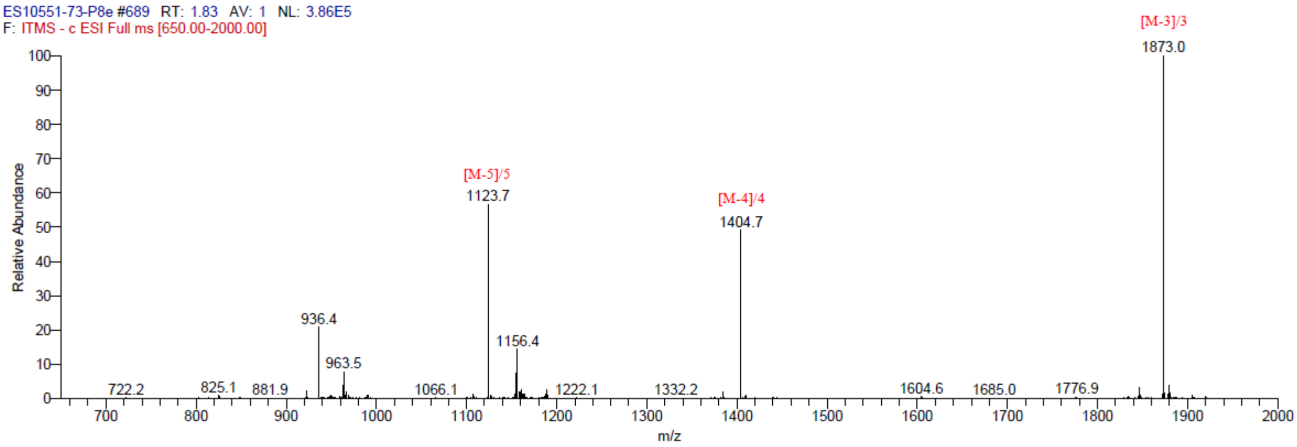
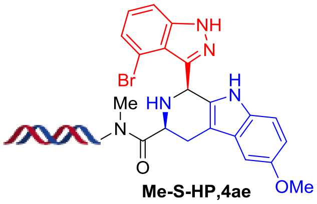


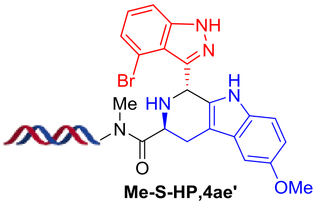
Fig. S11. LC trace and mass of **4ad**

Figure S12, Trace and Mass of 4ae and 4ae’, related to Figure 3.

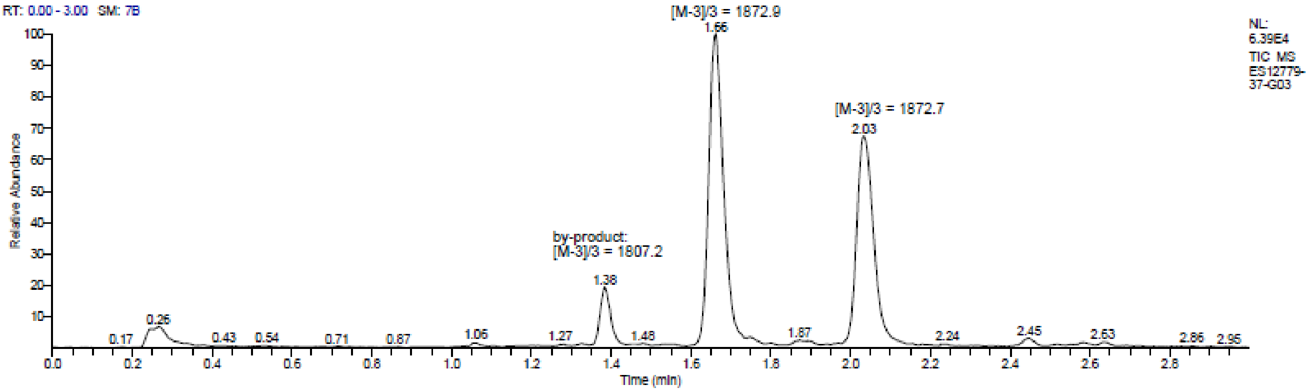
Following **General Procedure 2**
Percent conversion: 50.58% & 39.12%, totally 89.70%
Exact mass: 5622.02
Triply charged mass [M-3]/3, calculated: 1873.01; observed:1872.9&1872.7



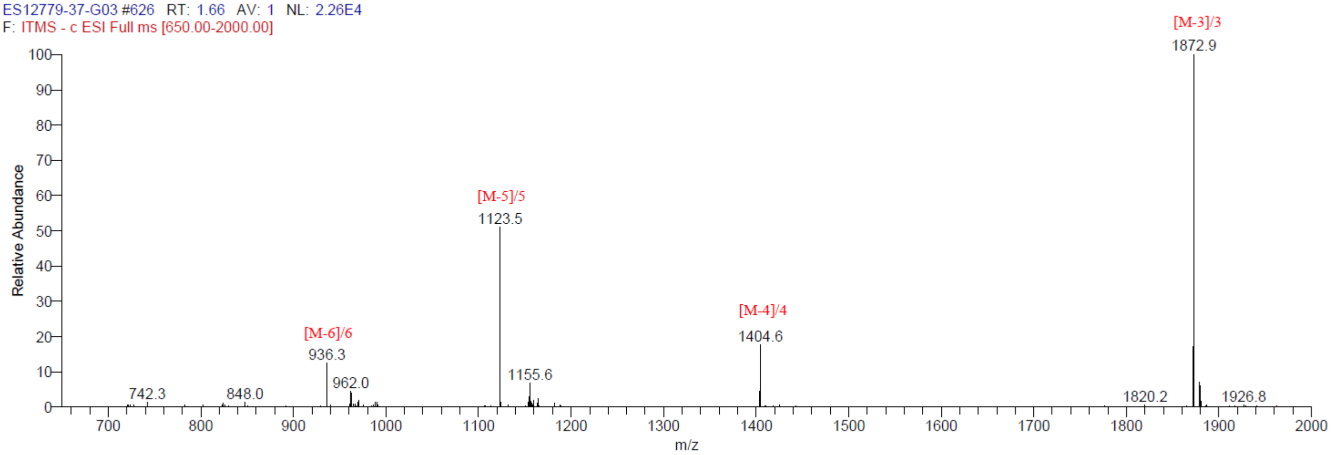
Me-S-HP,4ae



Me-S-HP,4ae'



RT	Peak Height	Peak Width	Peak Area	Area %
1.38	12397.34	0.31	26670.70	7.96
1.66	66787.65	0.25	169467.46	50.58
1.90	1294.92	0.08	3906.43	1.17
2.04	43767.81	0.29	131090.13	39.12
2.44	1862.73	0.09	3925.55	1.17



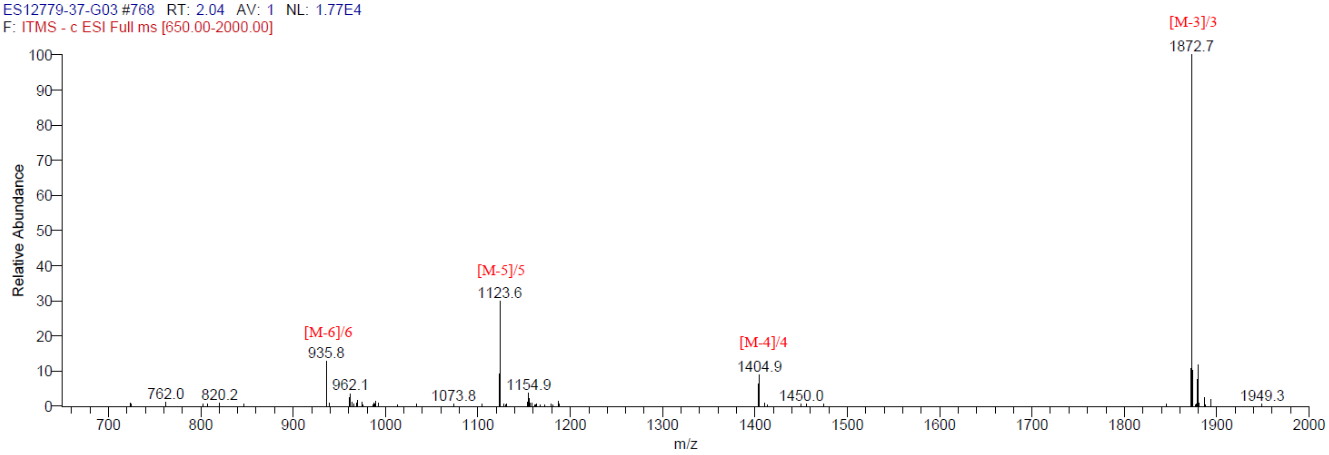
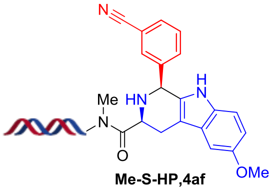


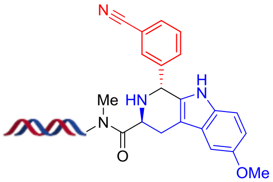
Fig. S12. LC trace and mass of **4ae and 4ae'**

Figure S13, Trace and Mass of 4af and 4af', related to Figure 3.

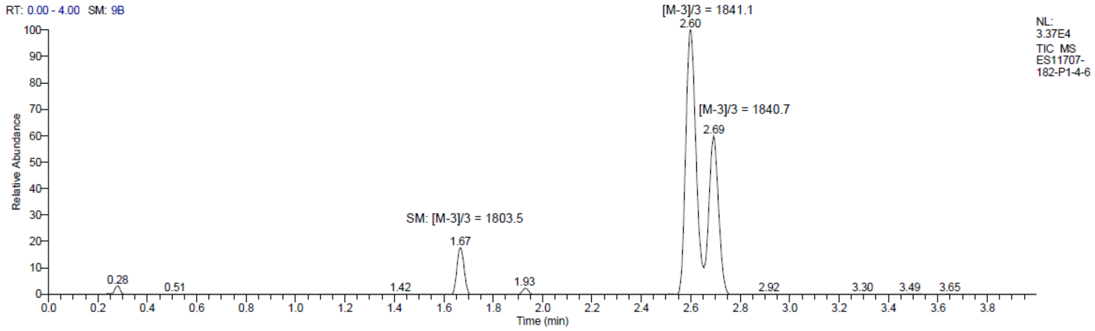
Following **General Procedure 2**
Percent conversion: 59.17% & 33.65%, totally 92.82%
Exact mass: 5528.10
Triply charged mass [M-3]/3, calculated: 1841.7; observed: 1841.1&1840.7



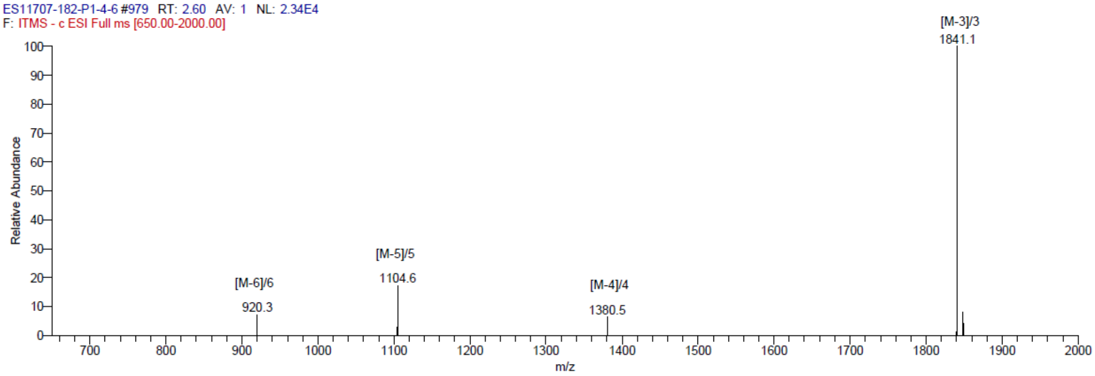
Me-S-HP,4af



Me-S-HP,4af'



RT	Peak Height	Peak Width	Peak Area	Area %
1.67	6701.44	0.06	11699.80	7.18
2.60	35171.56	0.10	96409.10	59.17
2.69	21512.70	0.09	54830.57	33.65



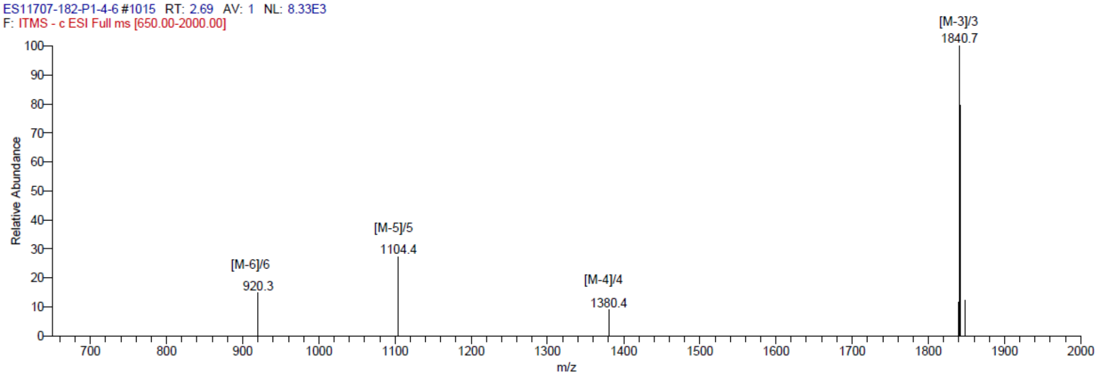
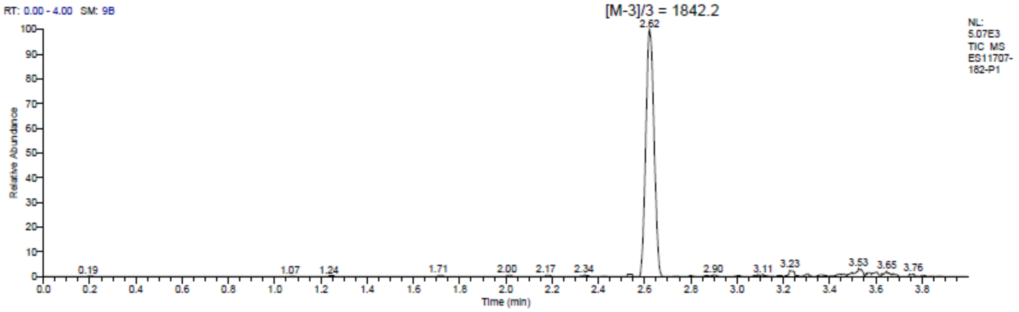


Fig. S13. LC trace and mass of **4af and 4af’**

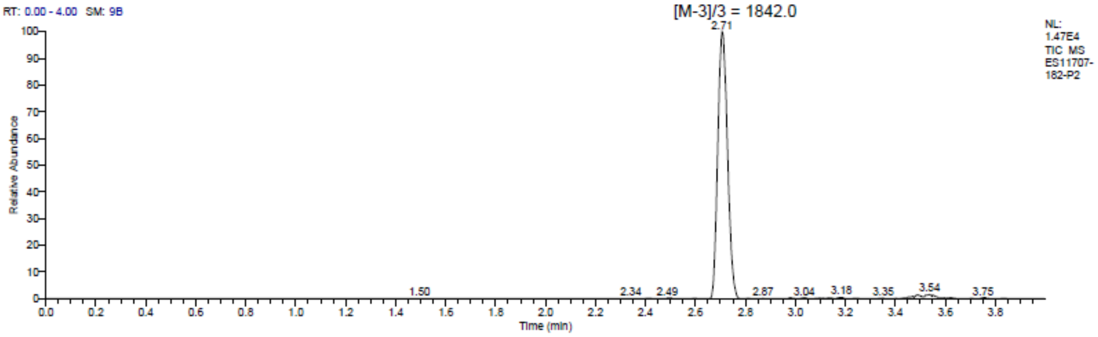
Figure S14, 4af and 4af’ were separated by HPLC, related to Figure 3.

Retain time = 2.62



RT	Peak Height	Peak Width	Peak Area	Area %
2.62	5672.18	0.08	12528.58	98.89
3.24	256.21	0.02	140.09	1.11

Retain time = 2.71



RT	Peak Height	Peak Width	Peak Area	Area %
2.70	15284.30	0.10	39618.11	100.00

Fig. S14. LC trace of **4af and 4af’**

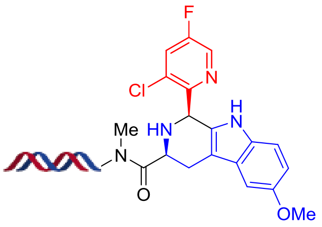
Figure S15, Trace and Mass of 4ag and 4ag’, related to Figure 3.

Following **General Procedure 2**

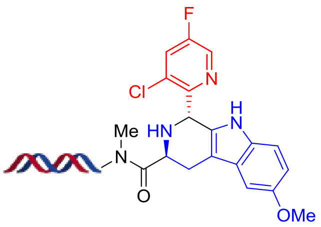
Percent conversion: 64.78% and 30.30%, totally 95.08%

Exact mass: 5617.94

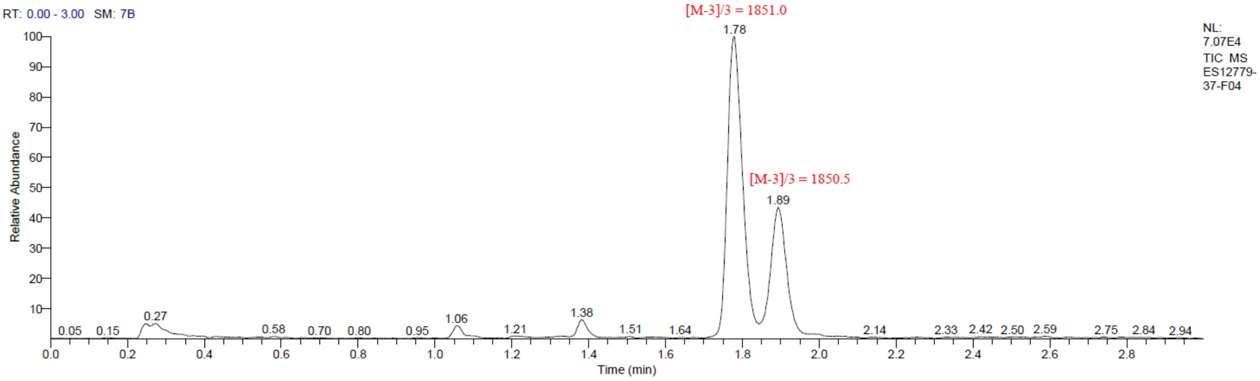
Triply charged mass [M-3]/3, calculated: 1871.66; observed:1871.70



Me-S-HP,4ag



Me-S-HP,4ag'



RT	Peak Height	Peak Width	Peak Area	Area %
1.06	3419.32	0.06	5250.78	1.76
1.10	596.77	0.04	805.29	0.27
1.38	4666.00	0.21	9396.50	3.15
1.78	74009.31	0.14	193113.16	64.78
1.89	31004.09	0.17	89520.43	30.03

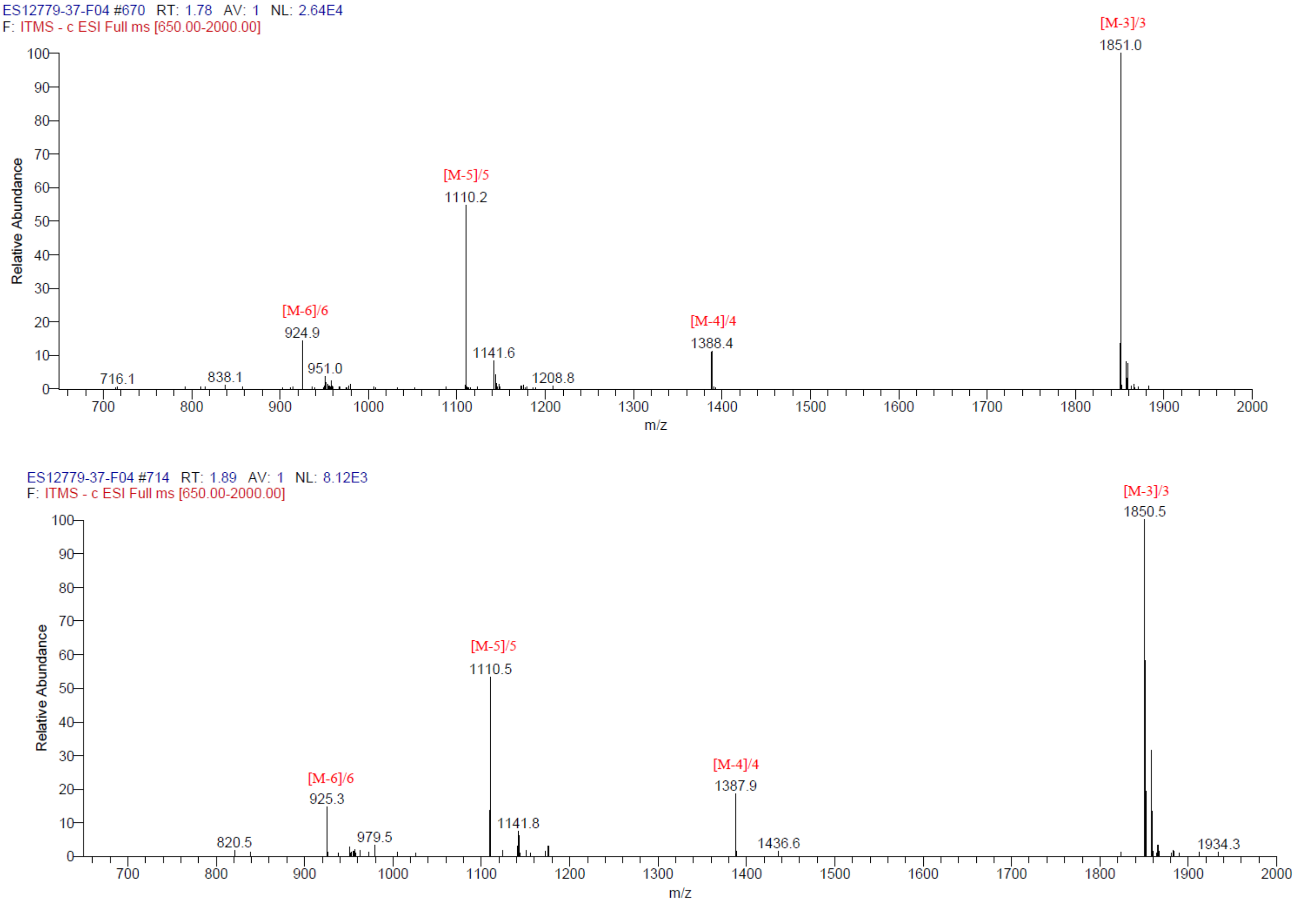


Fig. S15. LC trace and mass of **4ag and 4ag'**

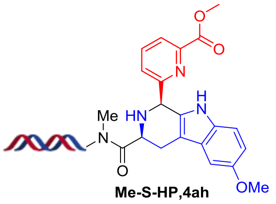
Figure S16, Trace and Mass of 4ah and 4ah', related to Figure 3.

Following **General Procedure 2**

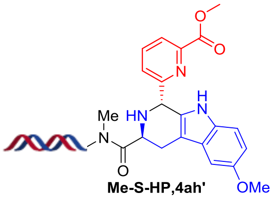
Percent conversion: 25.95% & 54.87%, totally 80.82%

Exact mass: 5562.12

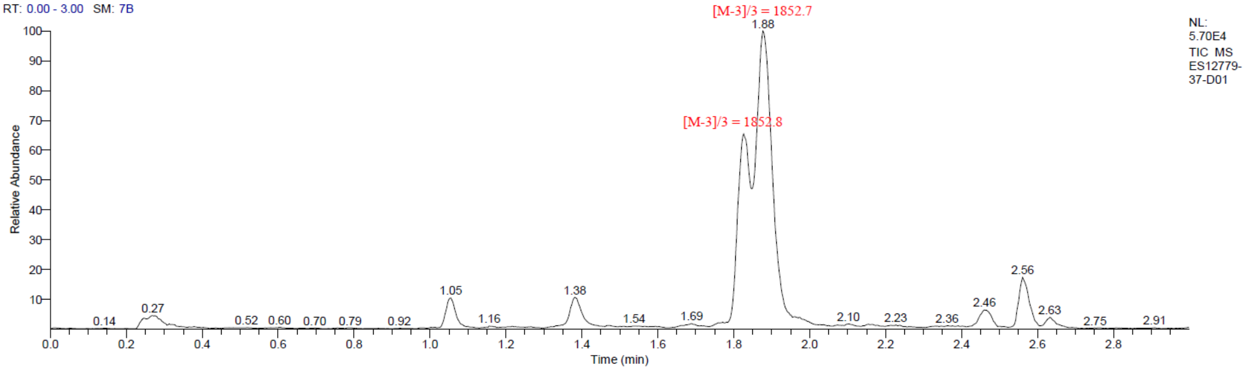
Triply charged mass [M-3]/3, calculated: 1853.04; observed:1852.8&1852.7



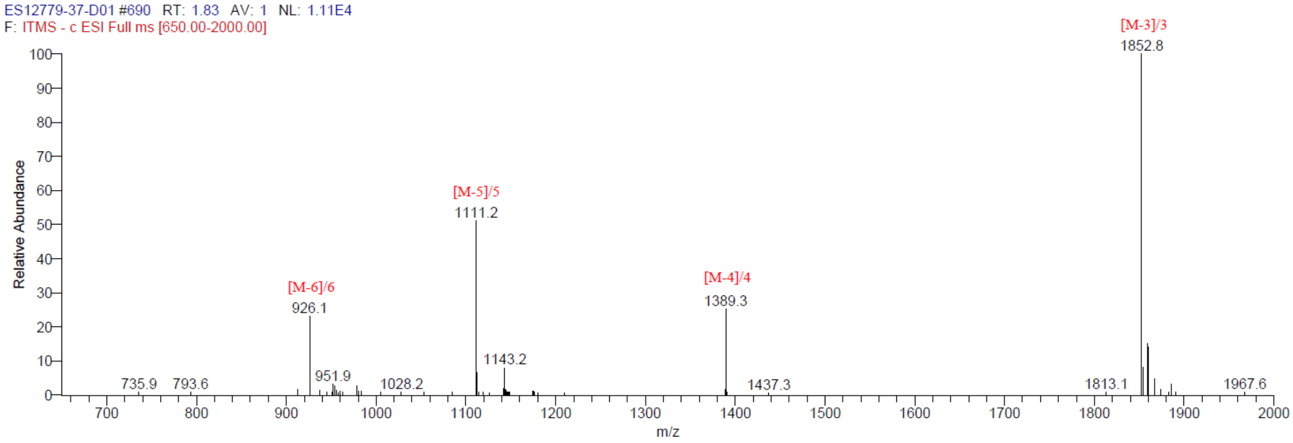
Me-S-HP,4ah



Me-S-HP,4ah'



RT	Peak Height	Peak Width	Peak Area	Area %
1.05	6239.64	0.13	10935.10	3.45
1.38	6209.66	0.31	17139.13	5.41
1.83	38485.46	0.07	82142.41	25.95
1.88	58533.04	0.20	173670.18	54.87
2.47	3452.89	0.22	9999.99	3.16
2.56	9969.86	0.08	18715.71	5.91
2.63	2129.45	0.09	3930.05	1.24



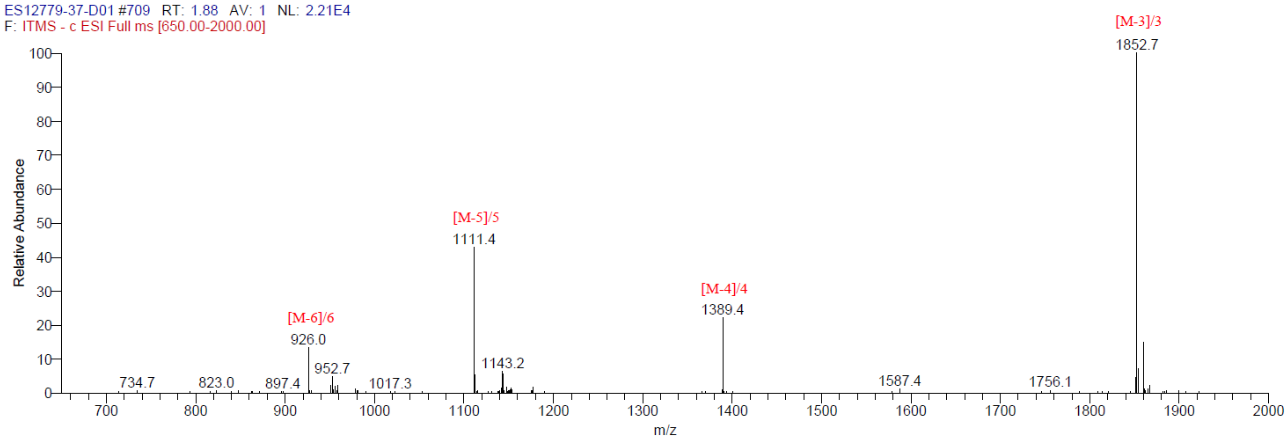
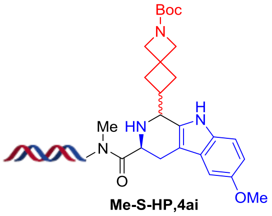


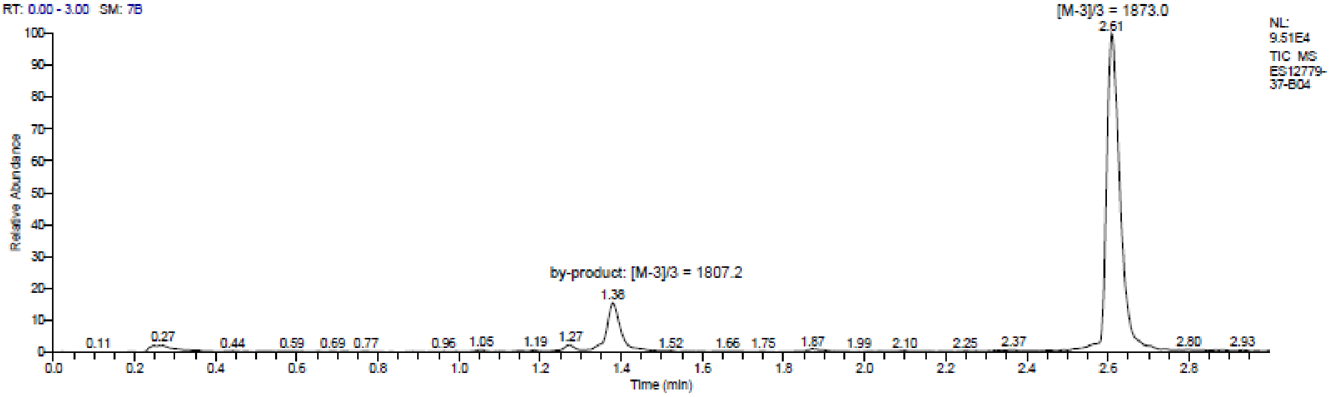
Fig. S16. LC trace and mass of **4ah and 4ah'**

Figure S17, Trace and Mass of 4ai, related to Figure 3.

Following **General Procedure 2**
Percent conversion: 82.58%
Exact mass: 5622.26
Triply charged mass [M-3]/3, calculated: 1873.09; observed:1873.0



Me-S-HP,4ai



RT	Peak Height	Peak Width	Peak Area	Area %
1.27	2212.28	0.10	4354.08	1.59
1.38	15302.15	0.23	37225.31	13.62
1.87	947.79	0.05	1204.93	0.44
2.61	99230.41	0.28	225662.29	82.58
2.79	739.04	0.22	4823.96	1.77

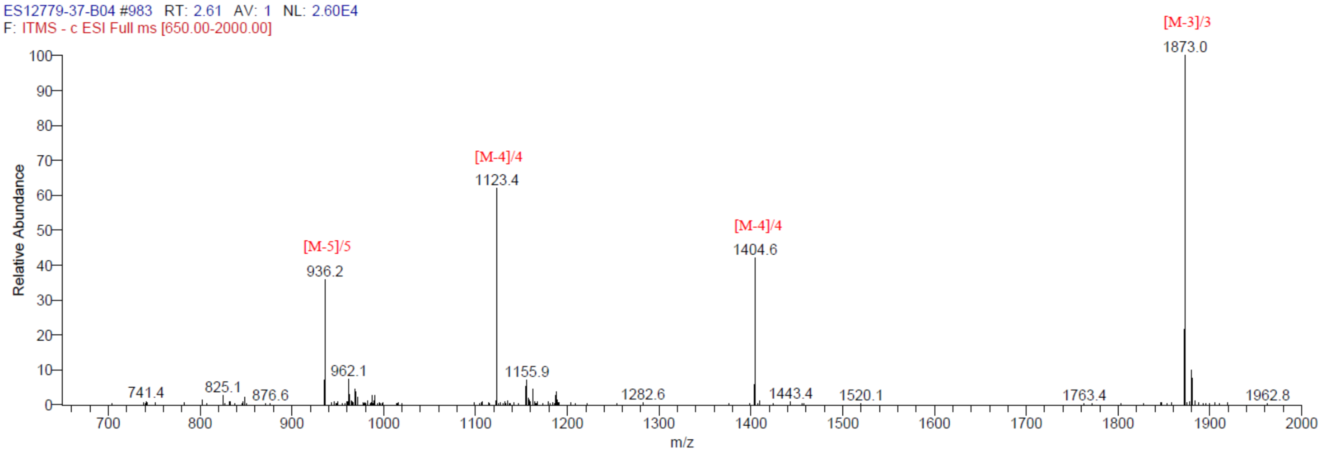
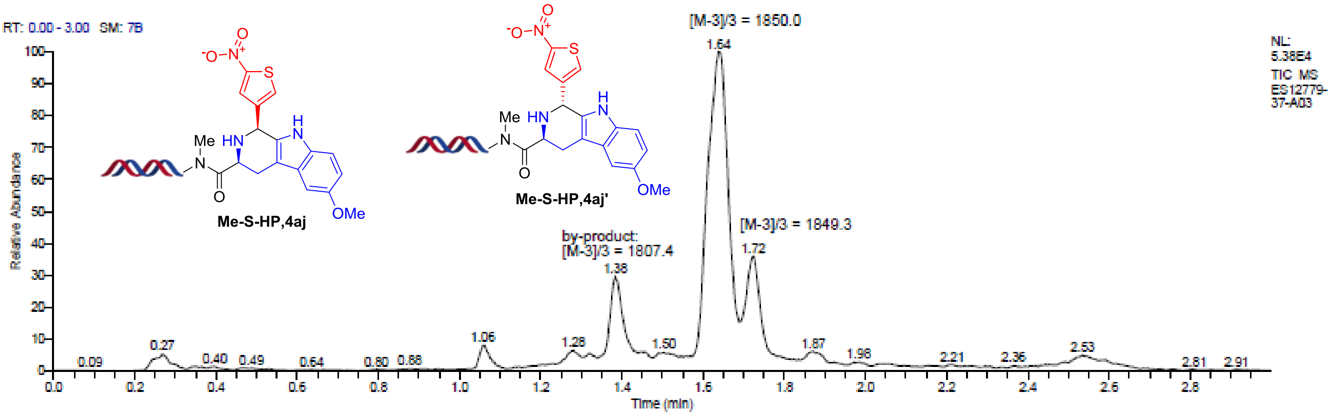


Fig. S17. LC trace and mass of **4ai**

Figure S18, Trace and Mass of 4aj and 4aj’, related to Figure 3.

Following **General Procedure 2**
Percent conversion: 63.28% & 17.23%, totally 80.51%
Exact mass: 5554.12

Triply charged mass [M-3]/3, calculated: 1850.37; observed:1850.0&1849.3



RT	Peak Height	Peak Width	Peak Area	Area %
1.38	15190.19	0.14	35388.36	12.28
1.64	54024.18	0.11	182279.76	63.28
1.72	18594.61	0.14	49628.25	17.23
1.87	2265.47	0.13	7698.09	2.67
2.54	2290.97	0.23	13079.71	4.54

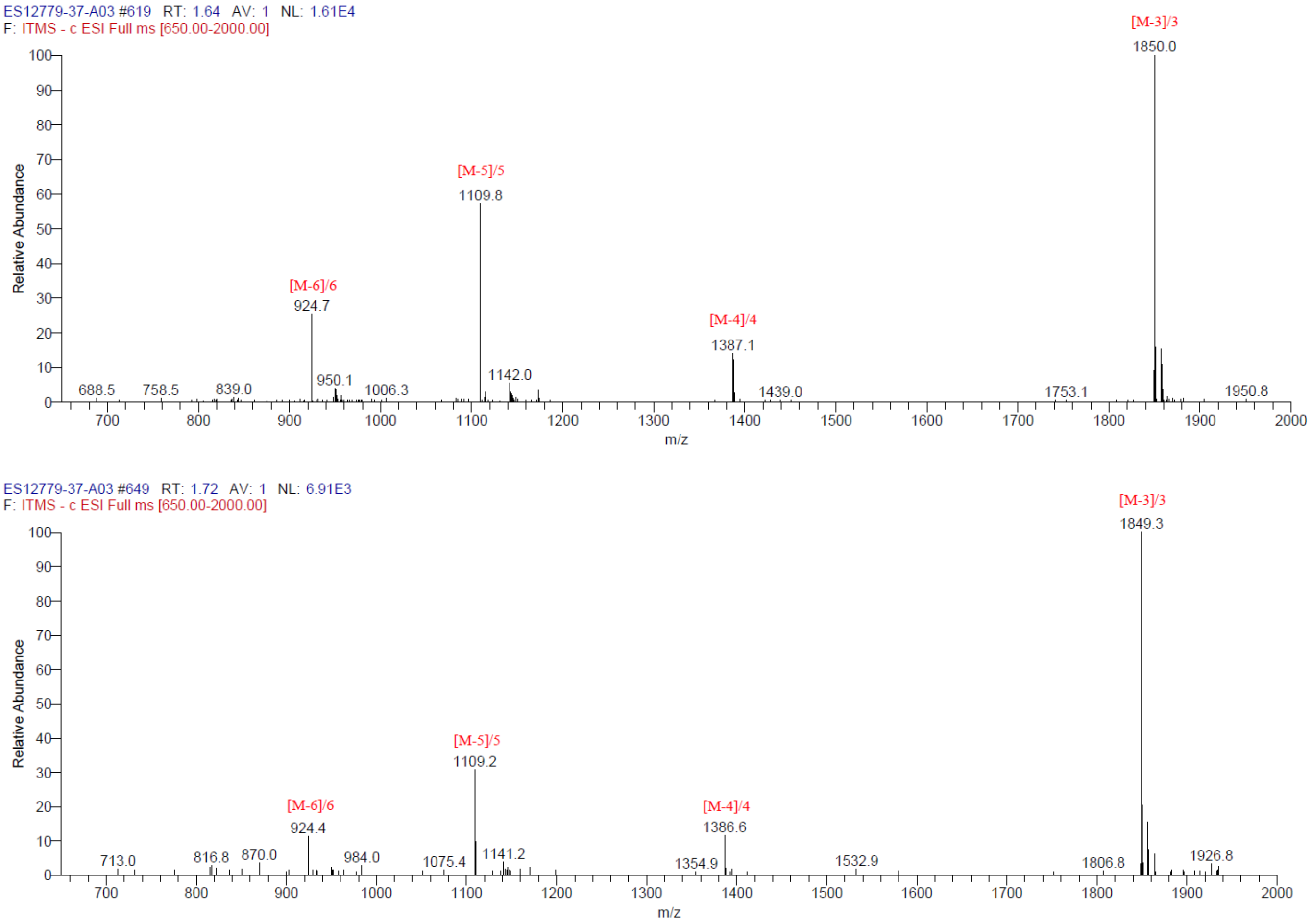


Fig. S18. LC trace and mass of **4aj and 4aj'**

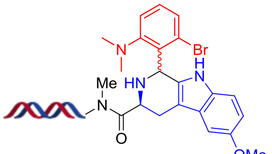
Figure S19, Trace and Mass of 4ak, related to Figure 3.

Following **General Procedure 2**

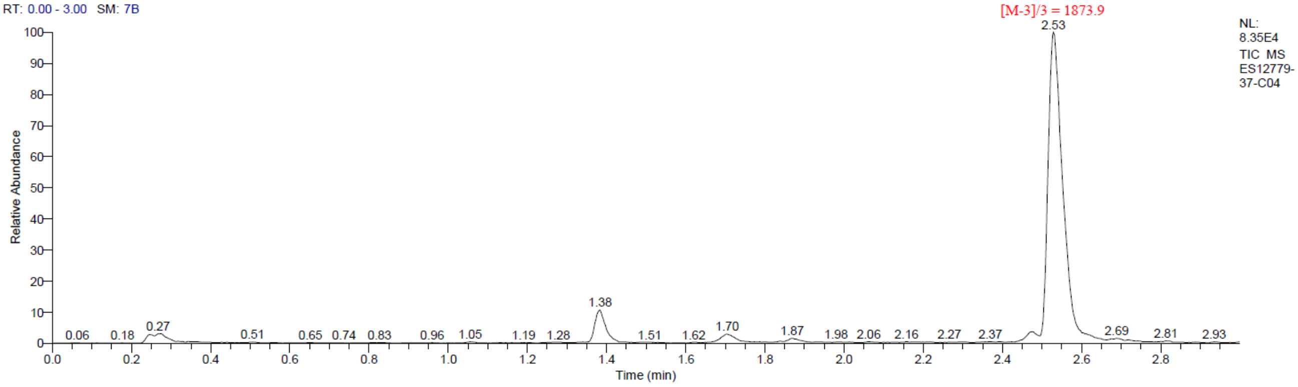
Percent conversion: 85.85%

Exact mass: 5625.06

Triply charged mass [M-3]/3, calculated: 1874.02; observed:1873.9



Me-S-HP,4ak



RT	Peak Height	Peak Width	Peak Area	Area %
1.38	9513.96	0.32	19463.58	7.72
1.71	2433.60	0.18	7122.57	2.82
1.87	1181.36	0.15	2965.93	1.18
2.47	3095.64	0.08	6136.89	2.43
2.53	84558.80	0.16	216499.12	85.85

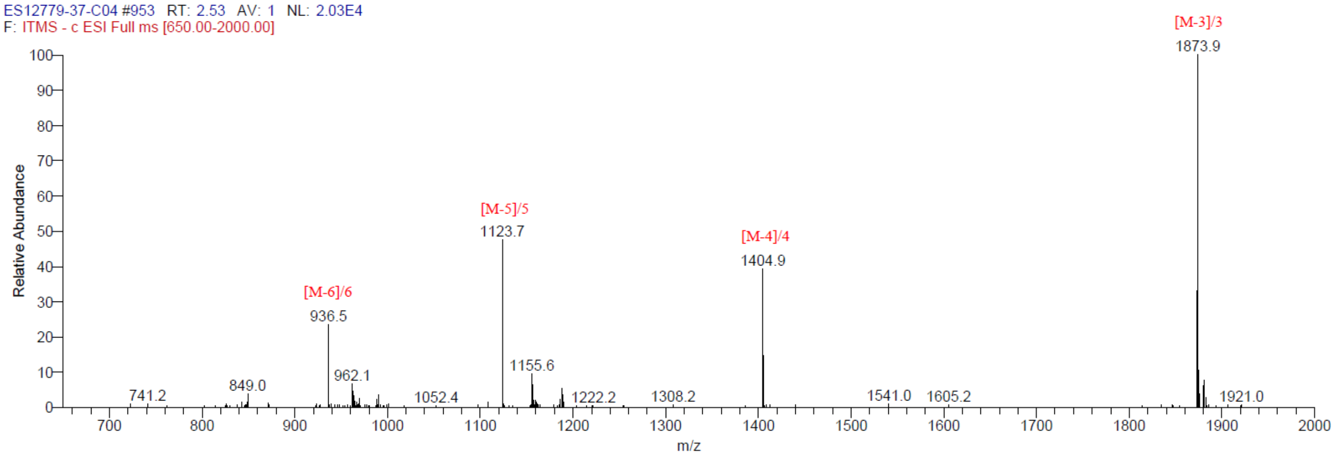


Fig. S19. LC trace and mass of **4ak**

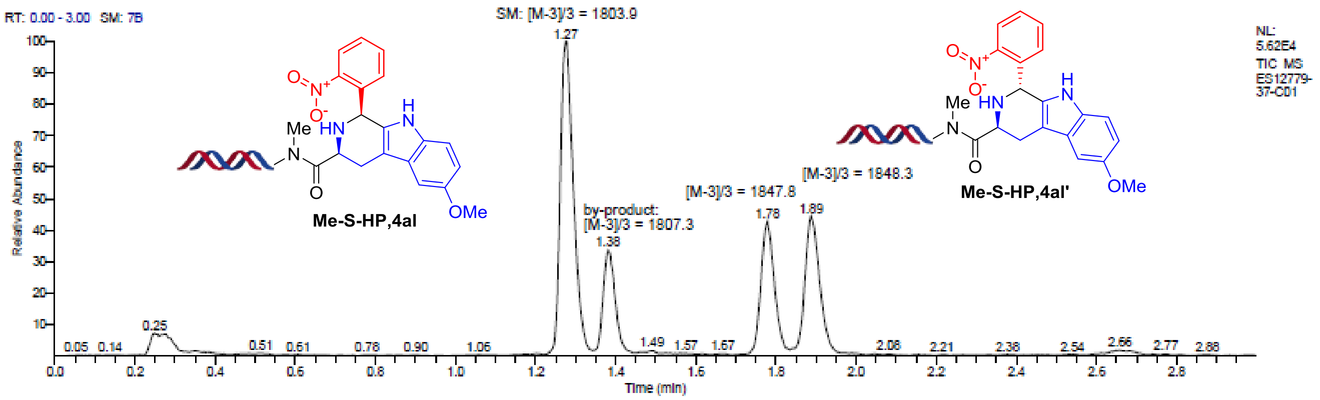
Figure S20, Trace and Mass of 4al and 4al’, related to Figure 3.

Following **General Procedure 2**

Yield: 19.15% & 22.35%, totally 41.50%

Exact mass: 5548.09

Triply charged mass [M-3]/3, calculated: 1848.36; observed:1847.8&1848.3



RT	Peak Height	Peak Width	Peak Area	Area %
1.27	60115.72	0.12	126688.35	43.65
1.38	18959.38	0.10	39168.61	13.50
1.78	24337.33	0.11	55580.53	19.15
1.89	25076.06	0.18	64867.29	22.35
2.66	1005.22	0.14	3903.55	1.35

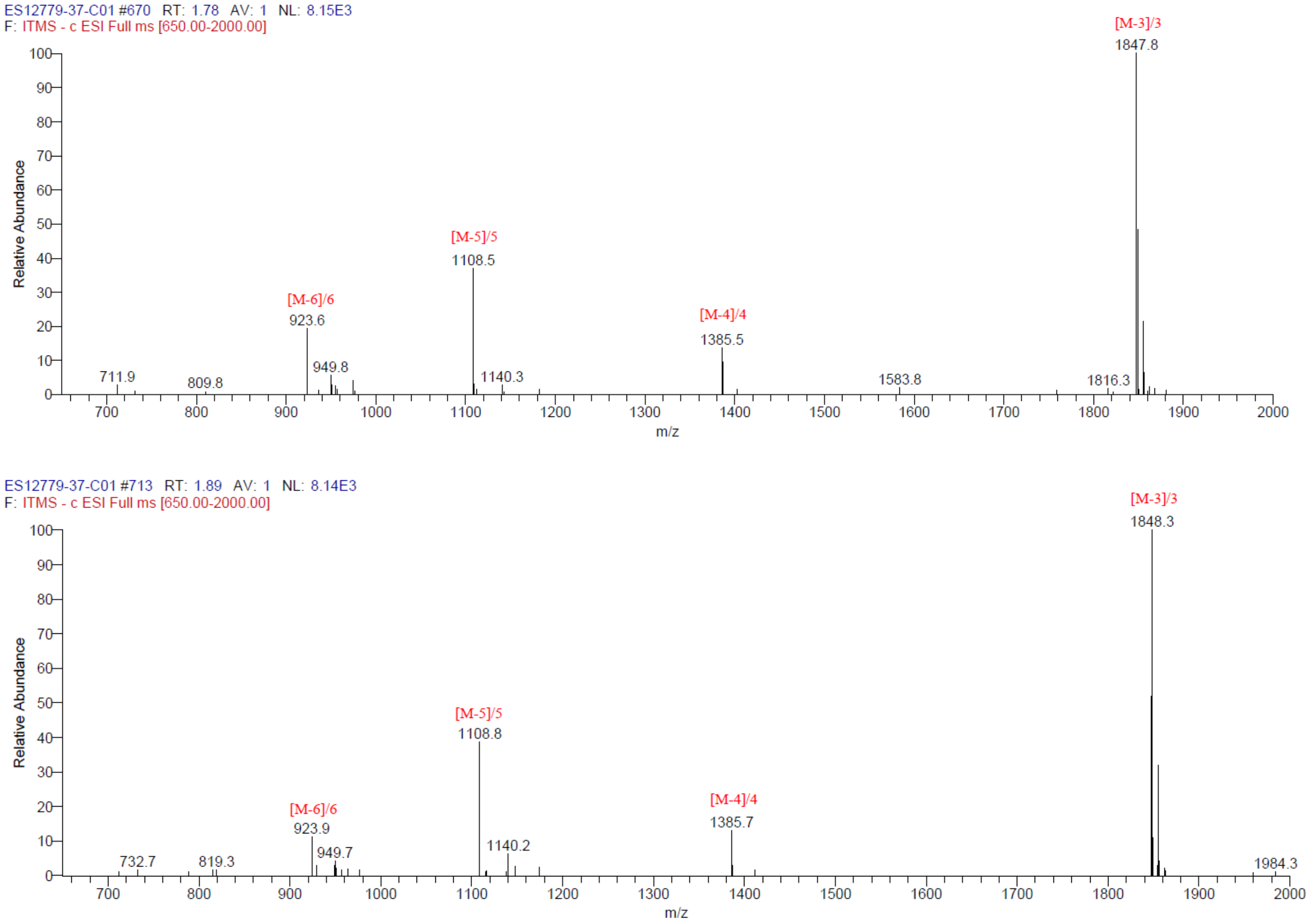


Fig. S20. LC trace and mass of **4al and 4al'**

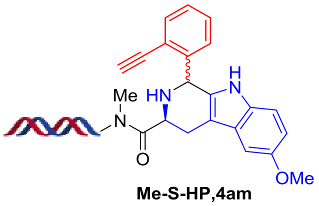
Figure S21, Trace and Mass of 4am, related to Figure 3.

Following **General Procedure 2**

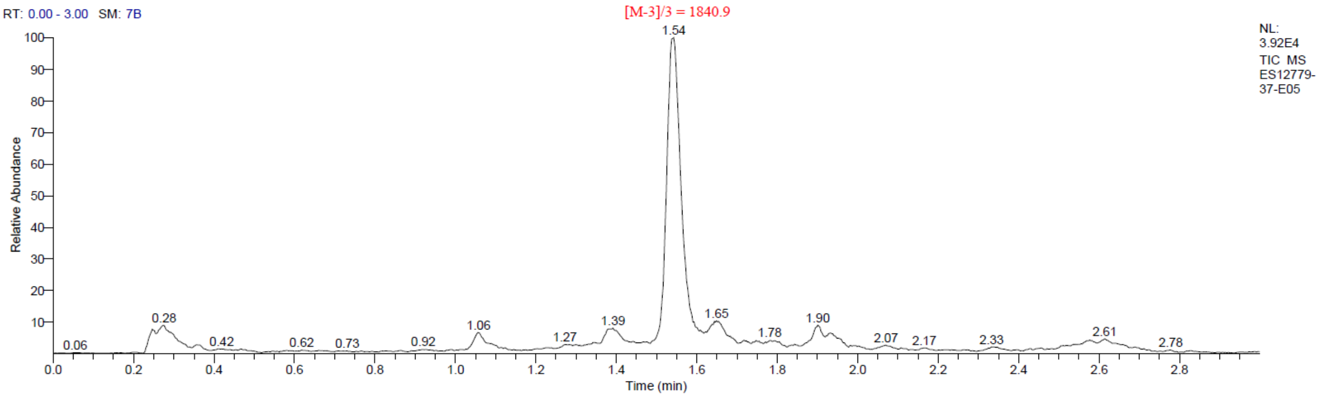
Percent conversion: 74.53%

Exact mass: 5527.12

Triply charged mass [M-3]/3, calculated: 1841.37; observed:1840.9



Me-S-HP,4am



RT	Peak Height	Peak Width	Peak Area	Area %
1.38	2736.38	0.10	8895.09	6.48
1.54	40191.83	0.16	102327.31	74.53
1.65	3401.66	0.08	10085.29	7.35
1.90	2992.37	0.15	9393.42	6.84
2.58	1405.86	0.18	6599.88	4.81

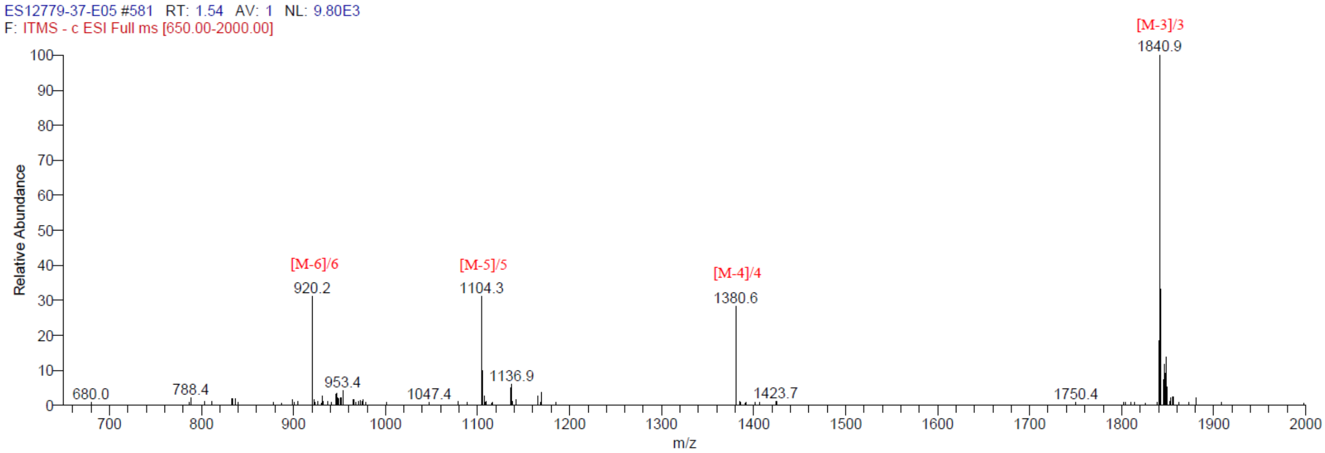


Fig. S21. LC trace and mass of **4am**

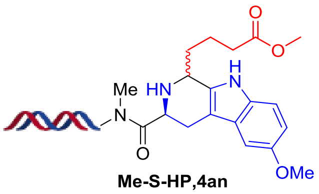
Figure S22, Trace and Mass of 4an, related to Figure 3.

Following **General Procedure 2**

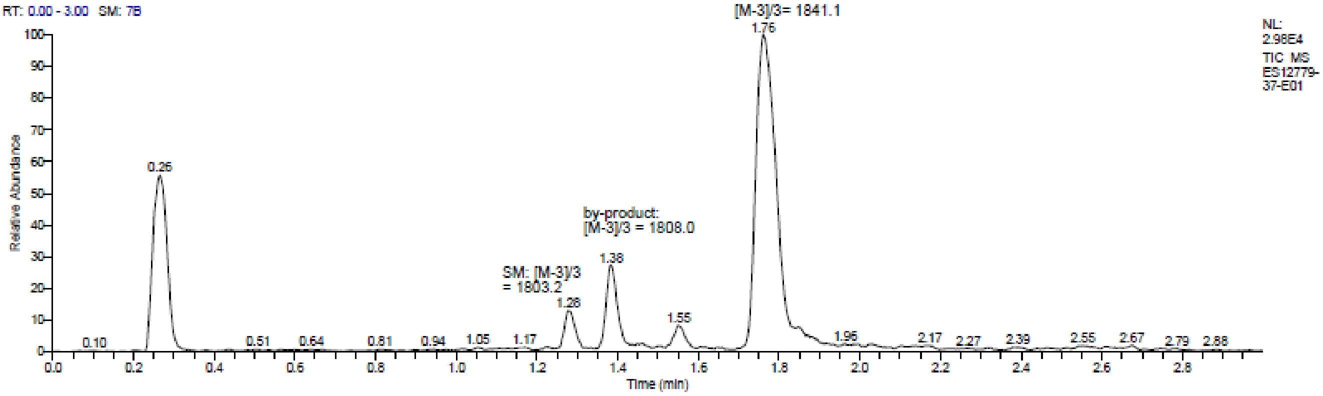
Percent conversion: 77.61%

Exact mass: 5527.11

Triply charged mass [M-3]/3, calculated: 1841.37; observed:1841.1



Me-S-HP,4an



RT	Peak Height	Peak Width	Peak Area	Area %
1.28	4151.99	0.14	7376.16	5.36
1.38	8840.46	0.10	15881.45	11.54
1.55	2499.27	0.14	5172.97	3.76
1.77	29575.75	0.21	106807.96	77.61
1.92	659.57	0.10	2388.20	1.74

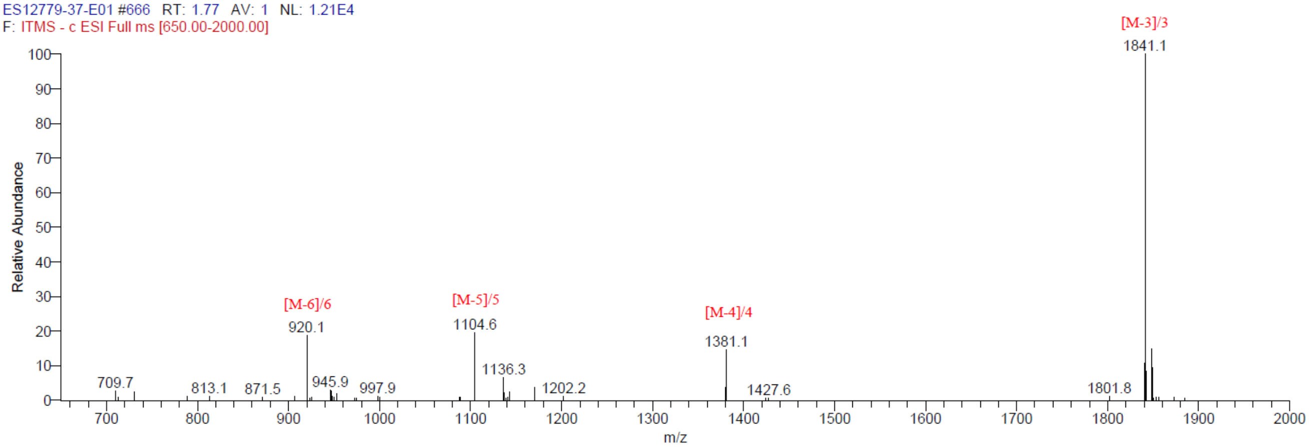


Fig. S22. LC trace and mass of **4an**

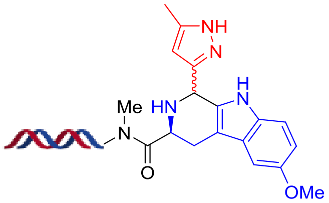
Figure S23, Trace and Mass of 4ao, related to Figure 3.

Following **General Procedure 2**

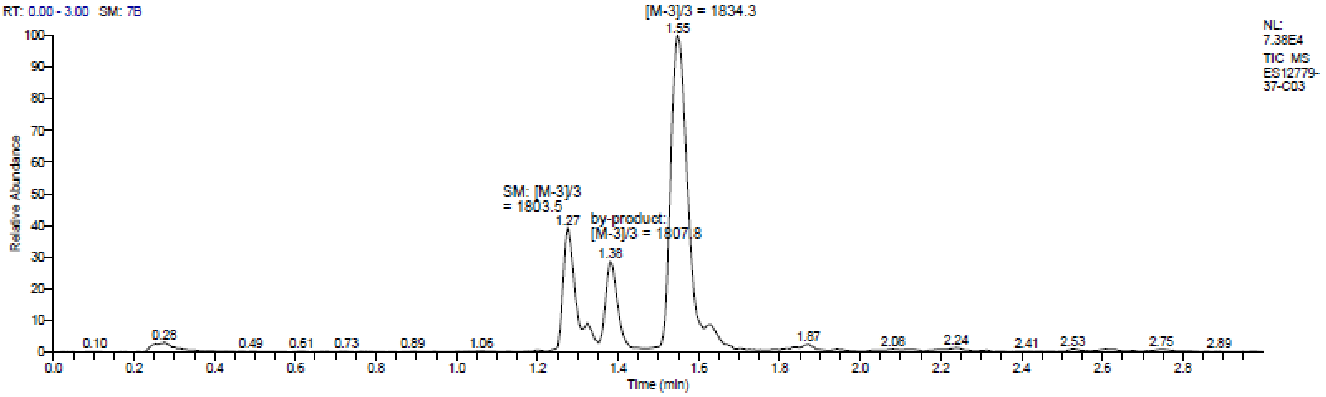
Percent conversion: 65.33%

Exact mass: 5506.81

Triply charged mass [M-3]/3, calculated: 1834.60; observed:1834.3



Me-S-HP,4ao



RT	Peak Height	Peak Width	Peak Area	Area %
1.27	30103.49	0.13	66503.50	19.08
1.38	22332.29	0.13	45980.14	13.19
1.55	75894.09	0.26	227705.91	65.33
1.88	1451.28	0.12	4269.15	1.22
2.24	863.62	0.16	4082.19	1.17

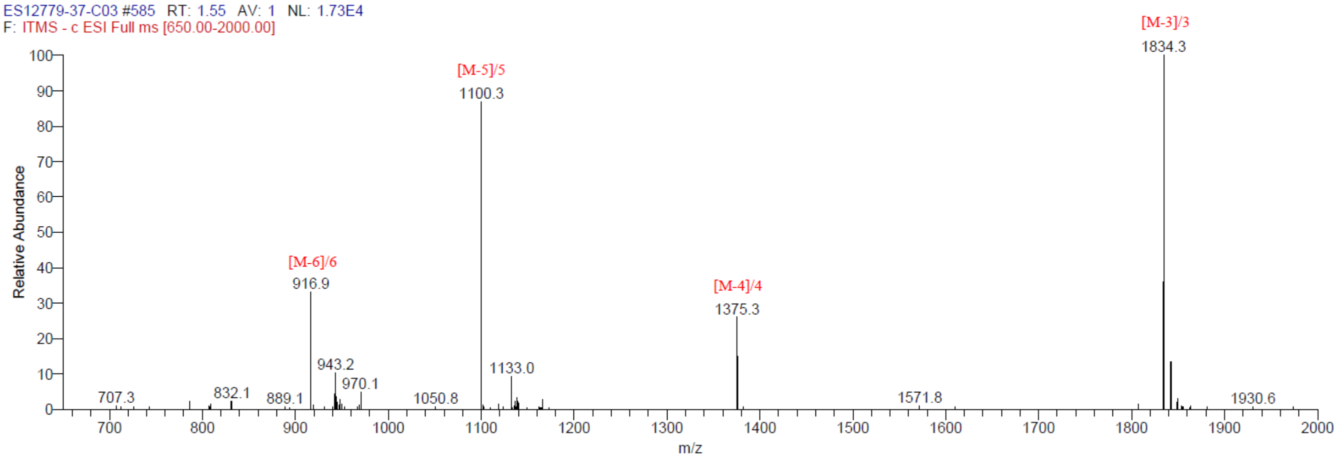


Fig. S23. LC trace and mass of **4ao**

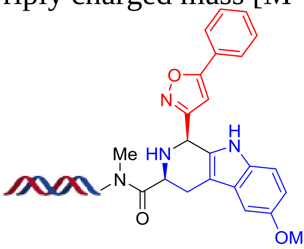
Figure S24, Trace and Mass of 4ap and 4ap’, related to Figure 3.

Following **General Procedure 2**

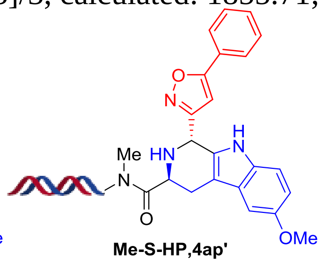
Percent conversion: 53.85% & 27.93%, totally 81.78%

Exact mass: 5570.14

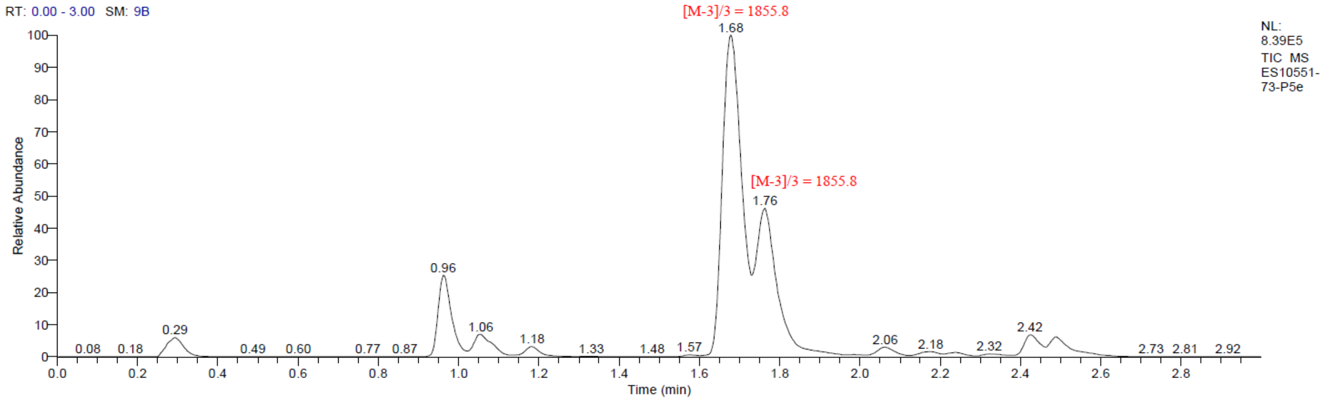
Triply charged mass [M-3]/3, calculated: 1855.71; observed:1855.8



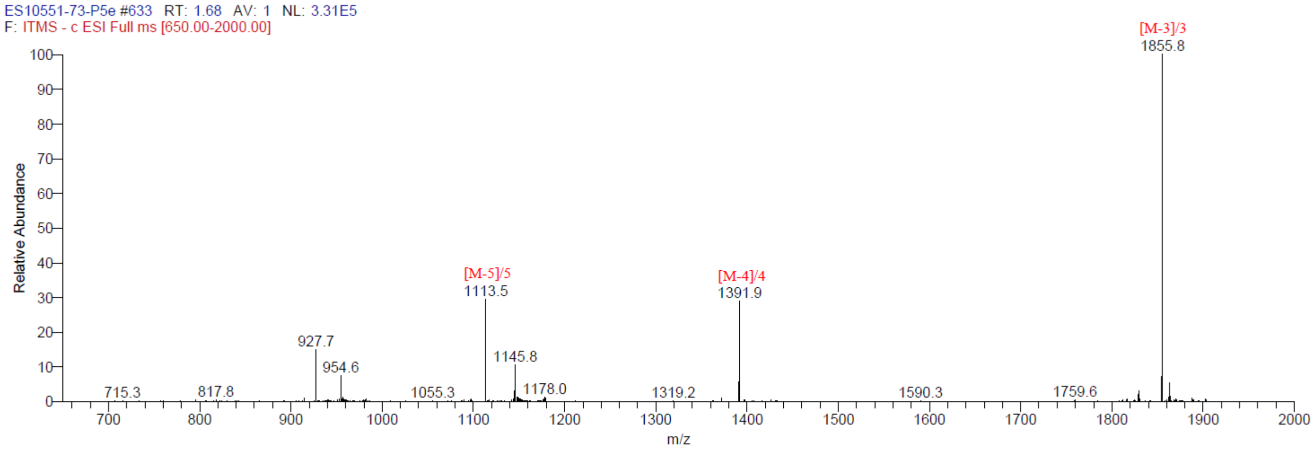
Me-S-HP,4ap



Me-S-HP,4ap'



RT	Peak Height	Peak Width	Peak Area	Area %
0.96	244725.74	0.13	502628.70	9.67
1.05	67679.69	0.10	193983.67	3.73
1.68	867050.88	0.13	2799586.85	53.85
1.76	401511.75	0.29	1451920.10	27.93
2.42	57372.51	0.17	250393.95	4.82



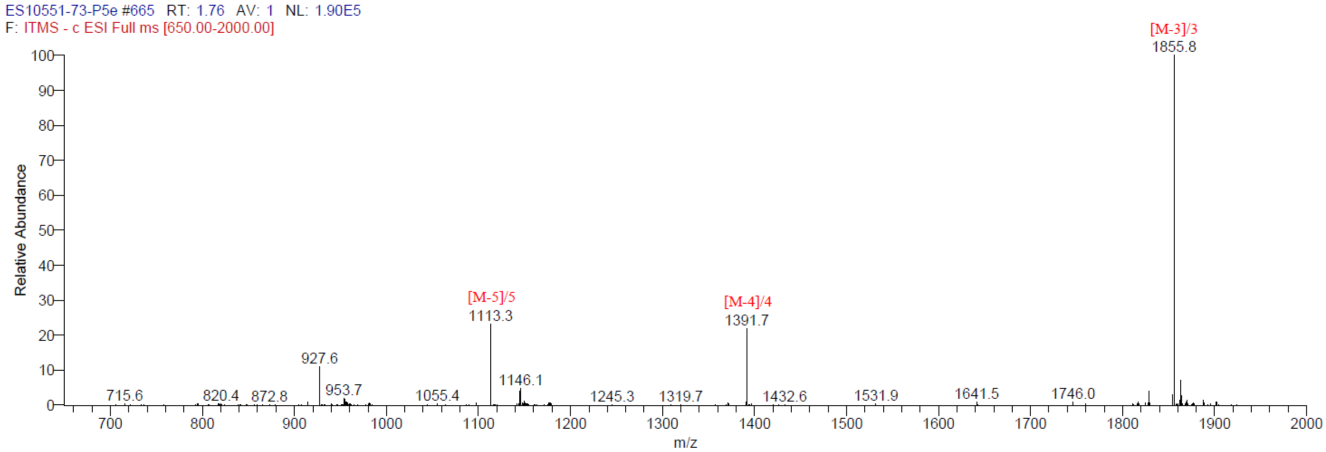
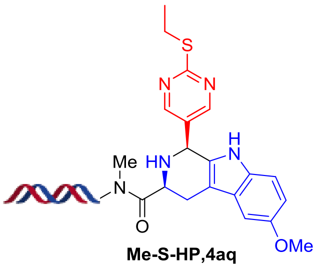


Fig. S24. LC trace and mass of **4ap and 4ap'**

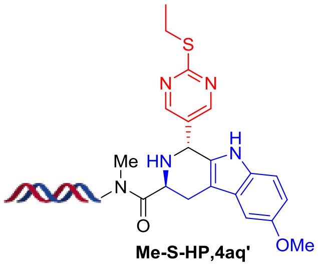
Figure S25, Trace and Mass of 4aq and 4aq', related to Figure 3.

Following **General Procedure 2**
Percent conversion: 23.84% & 53.99%, totally 77.83%
Exact mass: 5565.19

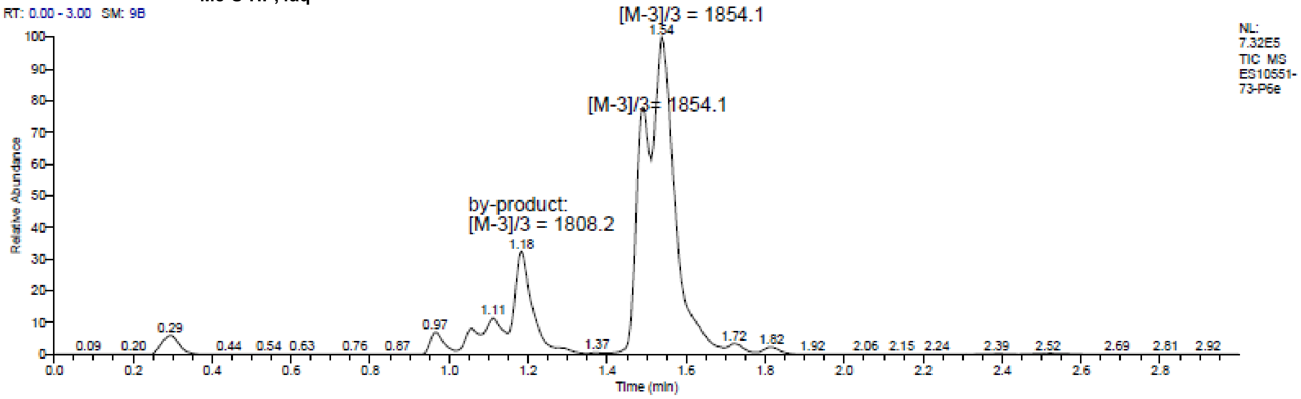
Triply charged mass [M-3]/3, calculated: 1854.06; observed:1854.1



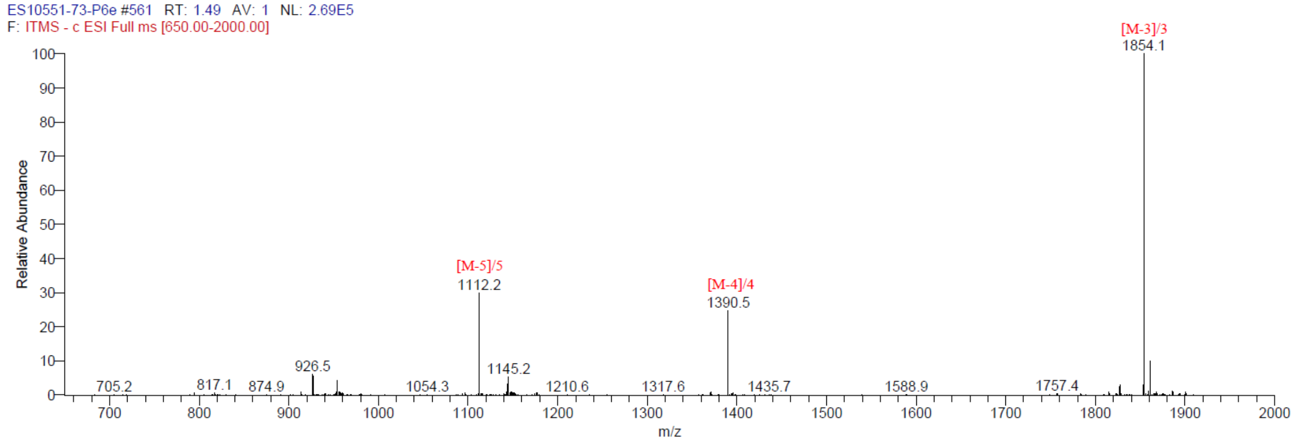
Me-S-HP,4aq



Me-S-HP,4aq'



RT	Peak Height	Peak Width	Peak Area	Area %
0.97	58516.51	0.15	129060.76	2.43
1.11	90505.79	0.08	298007.28	5.61
1.18	255398.20	0.19	749718.65	14.12
1.49	612901.23	0.11	1265538.53	23.84
1.54	749602.53	0.27	2866059.89	53.99



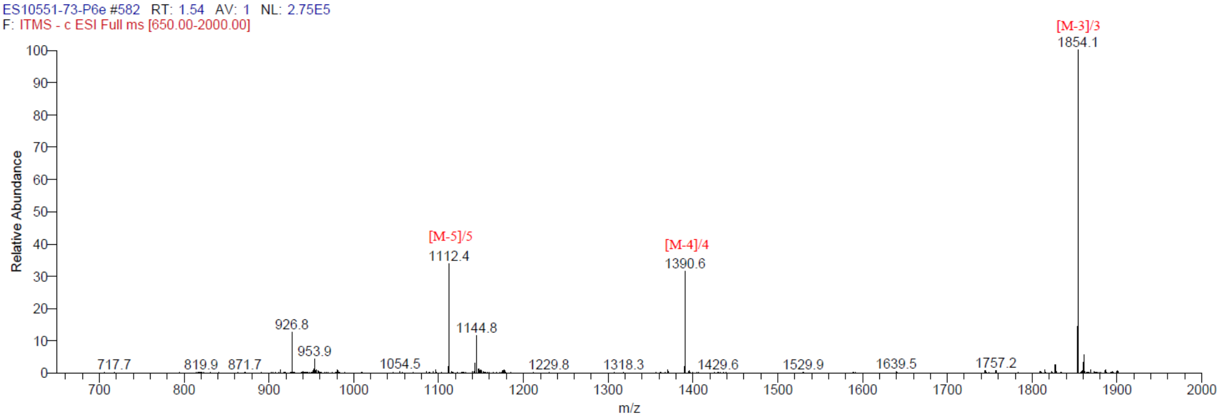
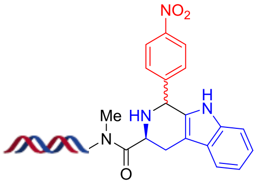


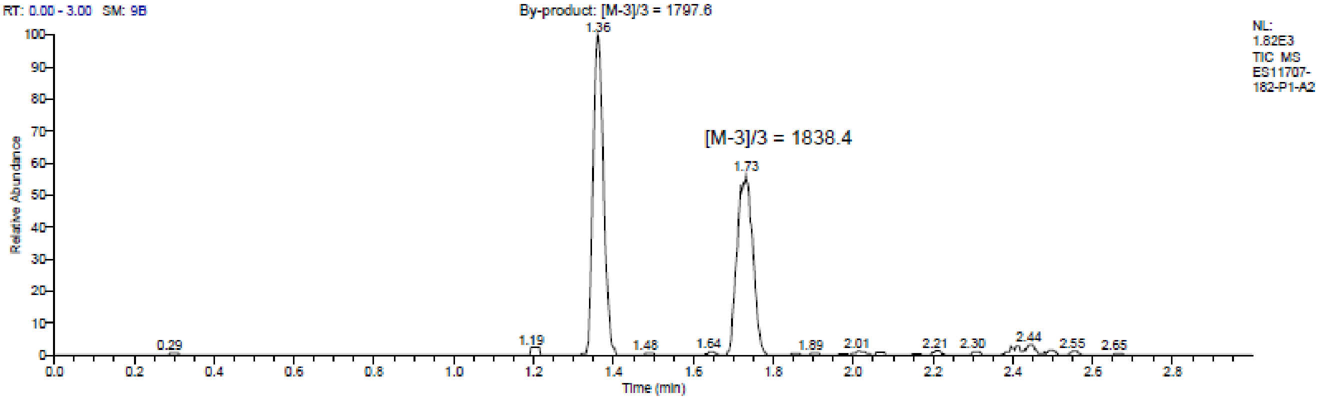
Fig. S25. LC trace and mass of **4aq and 4aq’**

Figure S26, Trace and Mass of 4b, related to Figure 5.

Following **General Procedure 2**
Percent conversion: 45.97%
Exact mass: 5518.07
Triply charged mass [M-3]/3, calculated: 1838.36; observed:1838.4



Me-S-HP,4b



RT	Peak Height	Peak Width	Peak Area	Area %
1.36	2199.16	0.10	3286.43	54.03
1.73	1145.73	0.11	2796.18	45.97

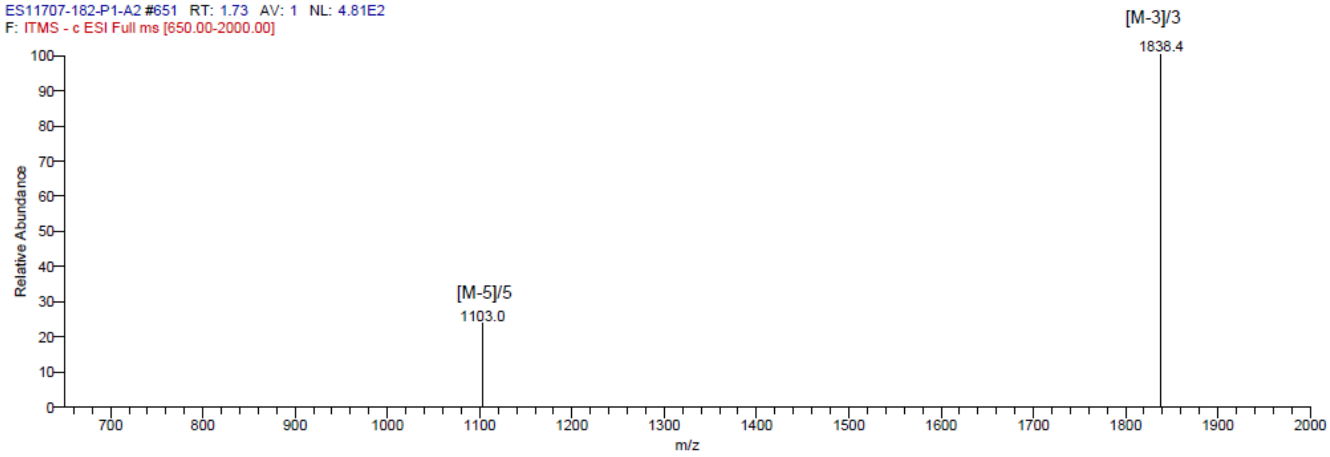
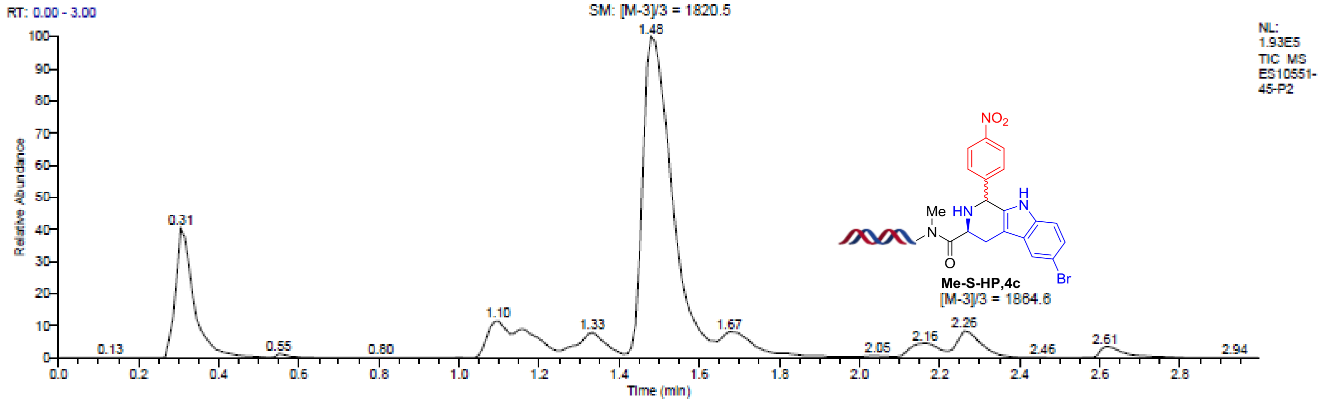


Fig. S26. LC trace and mass of **4b**.

Figure S27, Trace and Mass of 4c, related to Figure 5.

Following **General Procedure 2**
Percent conversion: 5.16%
Exact mass: 5596.96

Triply charged mass [M-3]/3, calculated: 1864.65; observed:1864.6



RT	Peak Height	Peak Width	Peak Area	Area %
1.10	20412.54	0.12	82941.93	5.68
1.16	16491.42	0.11	76927.03	5.27
1.33	13968.92	0.17	83428.48	5.71
1.49	181670.50	0.22	987404.57	67.62
1.68	15236.92	0.33	110163.62	7.54
2.16	8545.71	0.14	43981.62	3.01
2.27	14607.89	0.29	75282.50	5.16

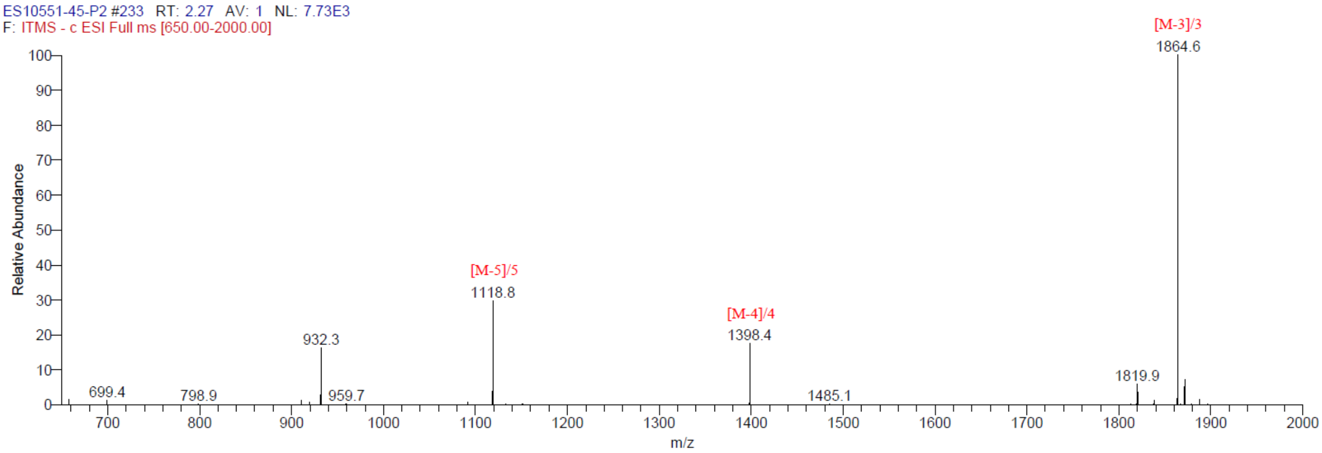


Fig. S27. LC trace and mass of **4c**.

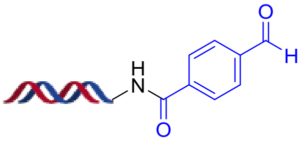
Figure S28, Trace and Mass of 5, related to Figure 5.

Following **General Procedure 1**

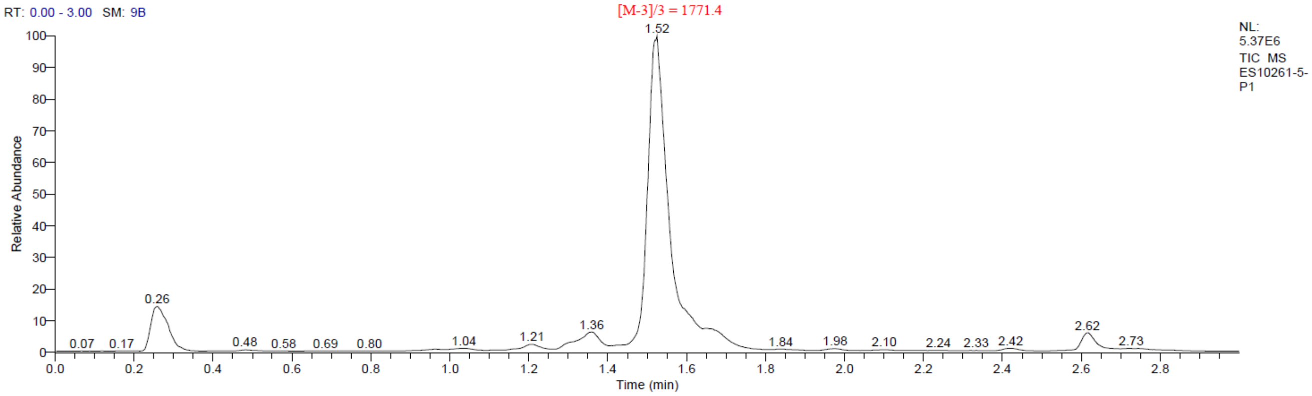
Purity: 90.83%

Exact mass: 5316.61

Triply charged mass [M-3]/3, calculated: 1771.2; observed:1771.4



S-HP,5



RT	Peak Height	Peak Width	Peak Area	Area %
1.03	41986.16	0.10	148617.42	0.75
1.21	100941.19	0.13	278799.36	1.40
1.36	248195.90	0.10	608614.27	3.05
1.52	5604412.58	0.22	18100890.92	90.83
2.61	325012.89	0.18	791440.10	3.97

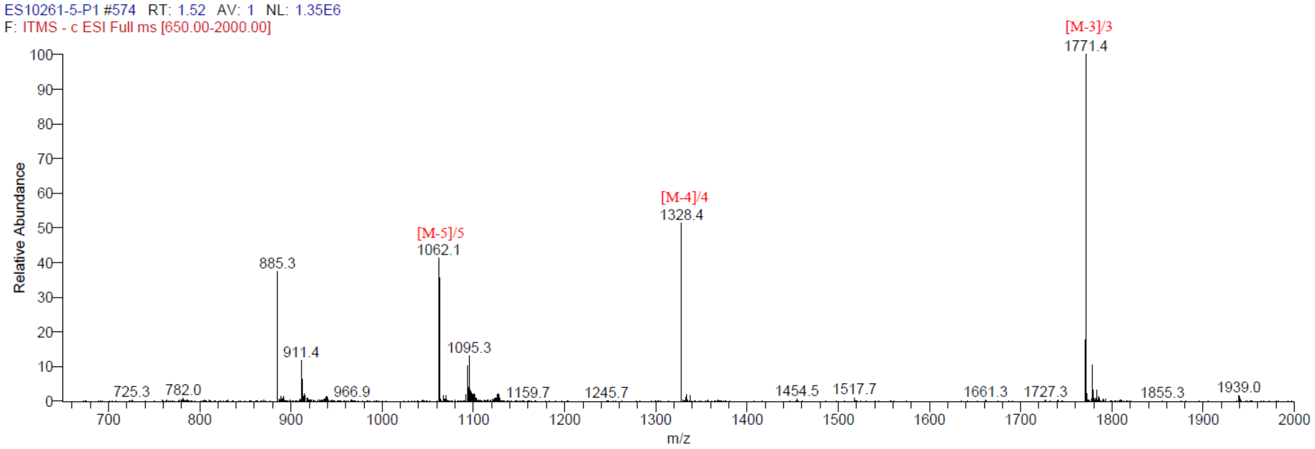
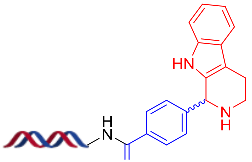


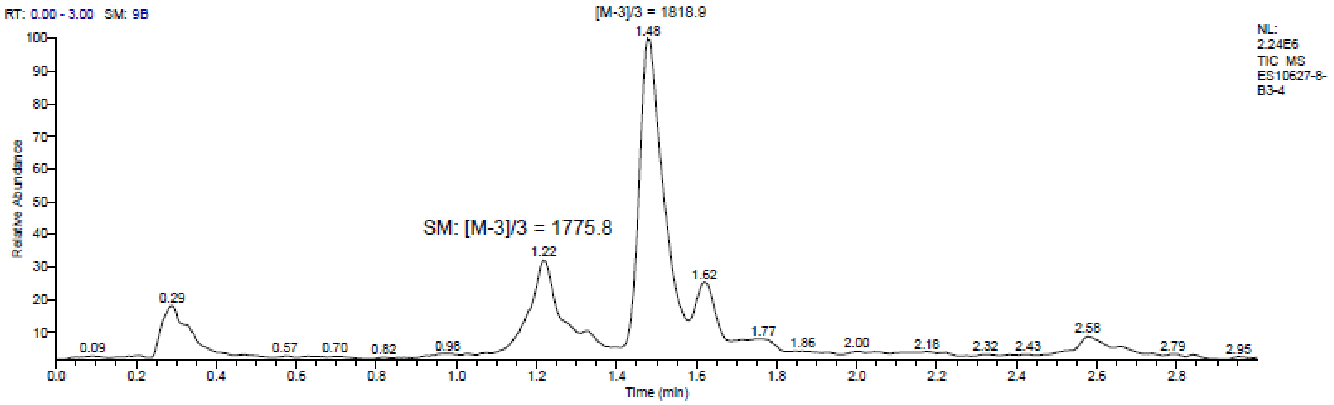
Fig. S28. LC trace and mass of **5.**

Figure S29, Trace and Mass of 6a, related to Figure 5.

Following **General Procedure 3**
Percent conversion: 64.42%
Exact mass: 5458.79
Triply charged mass [M-3]/3, calculated: 1818.60; observed:1818.9



S-HP,6a



RT	Peak Height	Peak Width	Peak Area	Area %
1.22	641282.46	0.21	2933774.86	21.80
1.33	120328.96	0.07	283924.03	2.11
1.48	2209307.88	0.17	8670371.09	64.42
1.62	449223.63	0.10	1397600.11	10.38
1.71	30393.19	0.10	172945.78	1.29

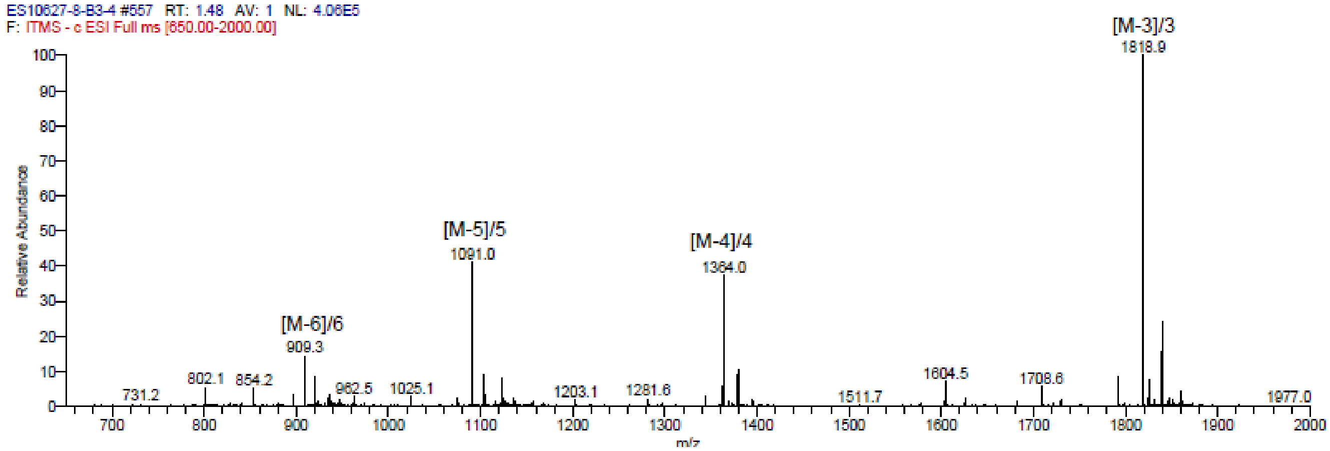


Fig. S29. LC trace and mass of **6a**

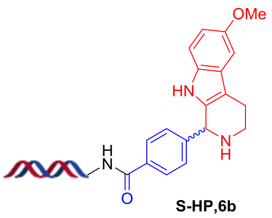
Figure S30, Trace and Mass of 6b, related to Figure 5.

Following **General Procedure 3**

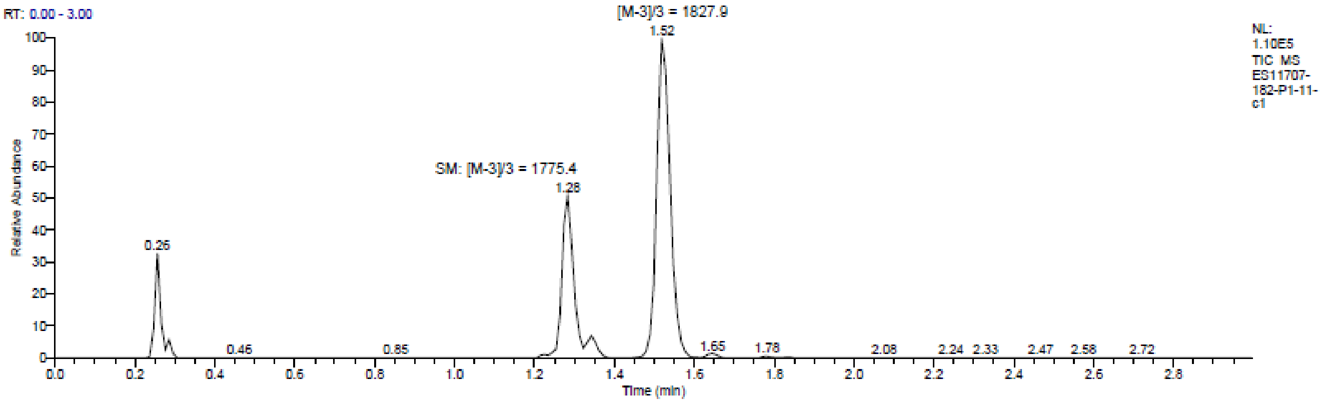
Percent conversion: 66.67%

Exact mass: 5488.89

Triply charged mass [M-3]/3, calculated:1828.63; observed:1827.90



S-HP,6b



RT	Peak Height	Peak Width	Peak Area	Area %
1.28	40881.23	0.23	125893.36	32.48
1.52	84858.02	0.17	258459.05	66.67
1.65	1054.33	0.07	2445.48	0.63
1.78	372.92	0.08	861.06	0.22

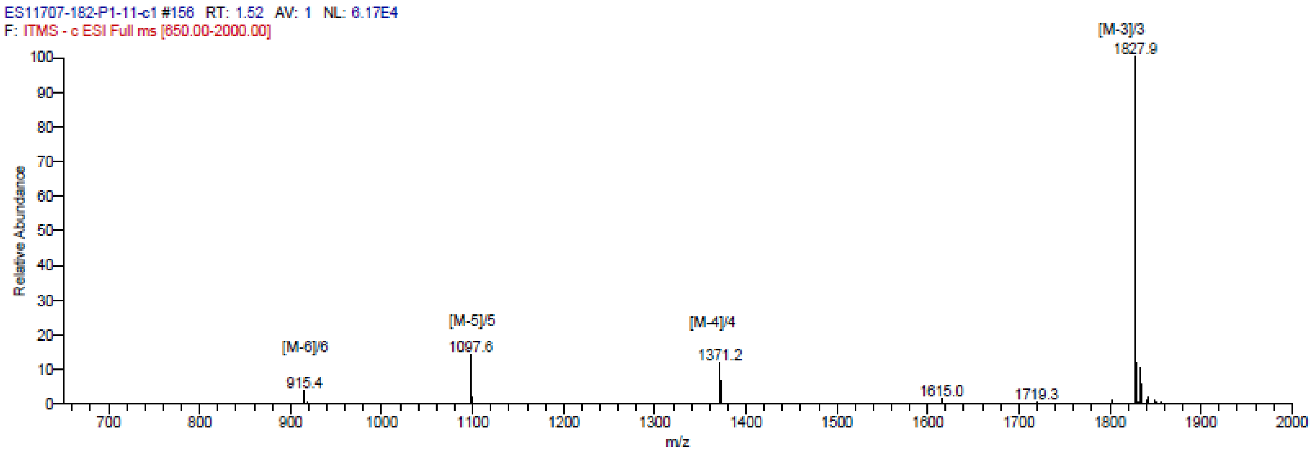


Fig. S30. LC trace and mass of **6b**

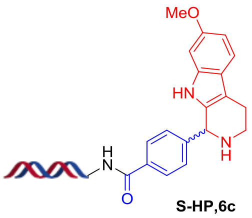
Figure S31, Trace and Mass of 6c, related to Figure 5.

Following **General Procedure 3**

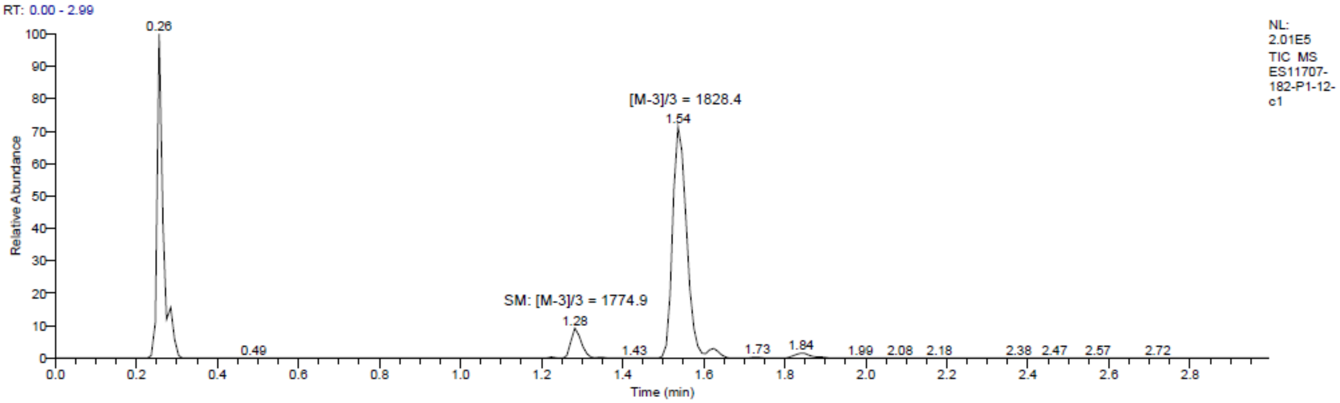
Percent conversion: 89.64%

Exact mass: 5488.86

Triply charged mass [M-3]/3, calculated:1828.62; observed:1828.40



S-HP,6c



RT	Peak Height	Peak Width	Peak Area	Area %
1.22	303.99	0.04	437.66	0.11
1.28	12591.48	0.12	32813.77	8.31
1.54	112961.87	0.22	353818.78	89.64
1.73	426.03	0.06	479.66	0.12
1.84	2384.38	0.10	7163.25	1.81

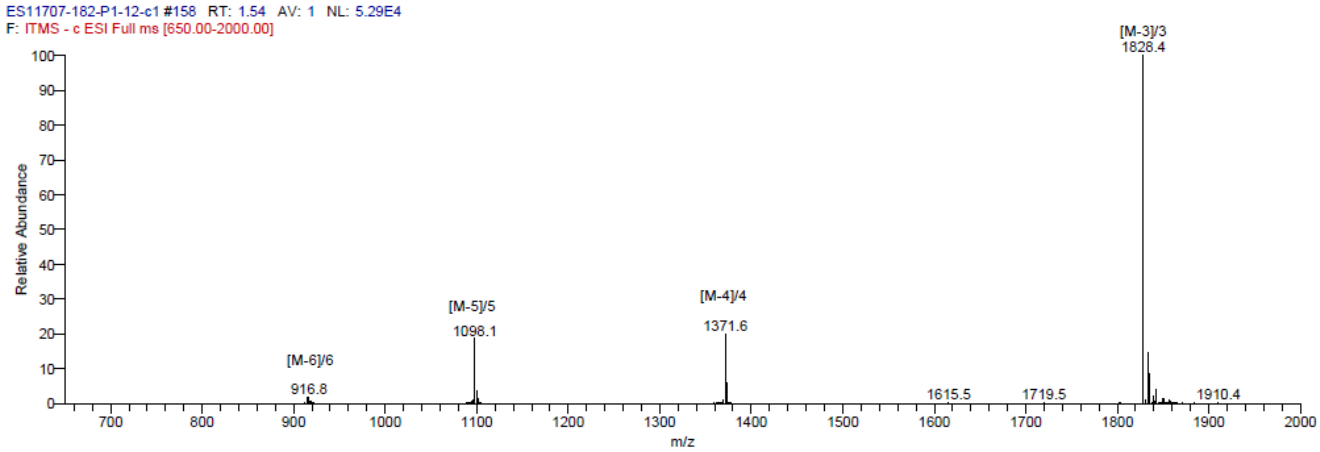


Fig. S31. LC trace and mass of **6c**

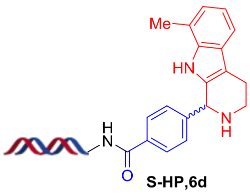
Figure S32, Trace and Mass of 6d, related to Figure 5.

Following **General Procedure 3**

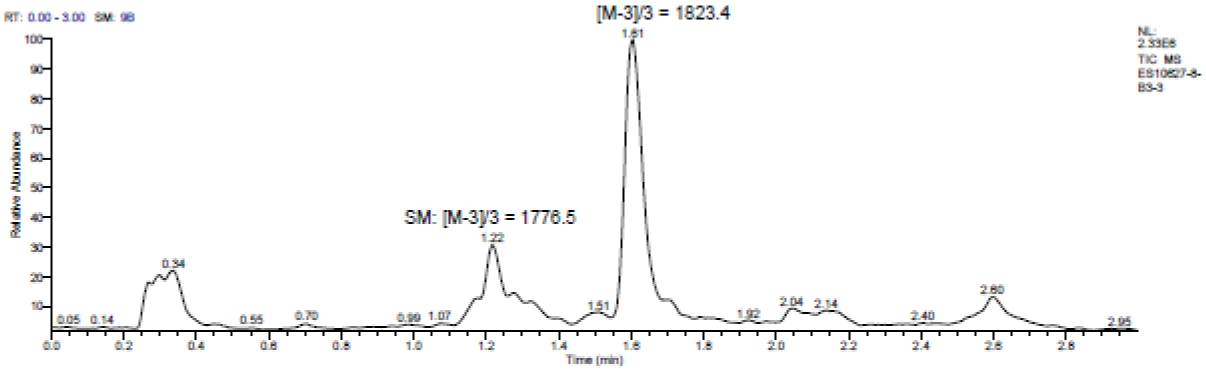
Percent conversion: 71.85%

Exact mass: 5472.86

Triply charged mass [M-3]/3, calculated: 1823.86; observed:1823.4



S-HP,6d



RT	Peak Height	Peak Width	Peak Area	Area %
1.18	189392.01	0.07	469782.06	4.21
1.22	620567.28	0.07	1557324.58	13.96
1.27	208641.03	0.10	823460.94	7.38
1.50	80281.16	0.11	291082.35	2.61
1.60	2296921.82	0.19	8017762.08	71.85

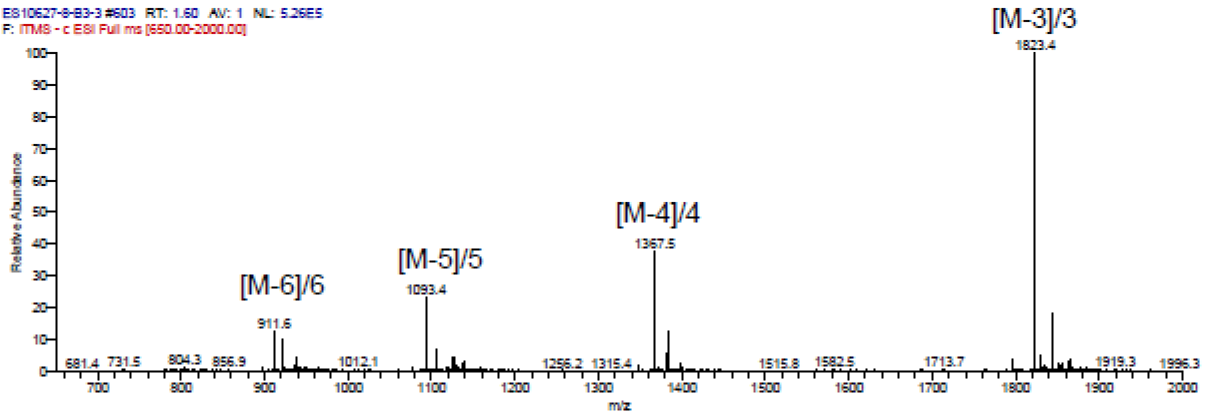


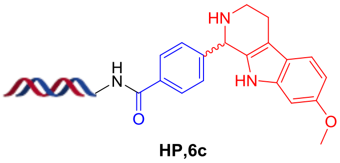
Fig. S32. LC trace and mass of **6d**

Figure S33, Mass Spectrum of 6c, related to Figure 5.

Following **General Procedure 3**

Percent conversion: 83.33%

Exact mass: 5241.59, observed mass:5241.95



HP,6c

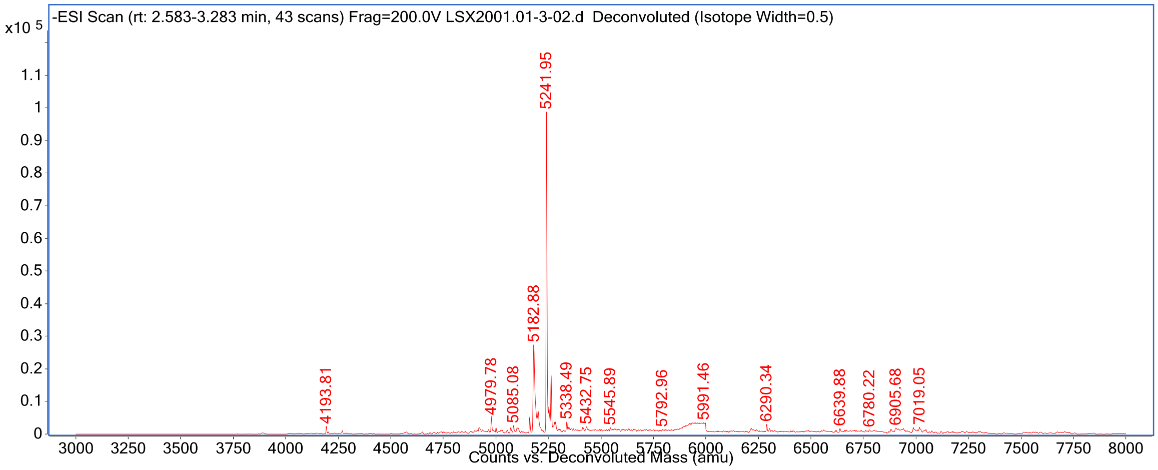


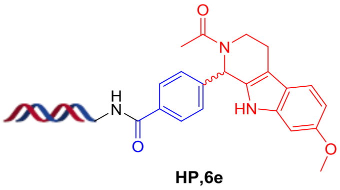
Fig. S33. Deconvoluted mass of **6c**

Figure S34, Mass Spectrum of 6e, related to Figure 5.

Following **General Procedure 4**

Percent conversion: 61.02%

Exact mass: 5283.62, observed mass:5284.02



HP,6e

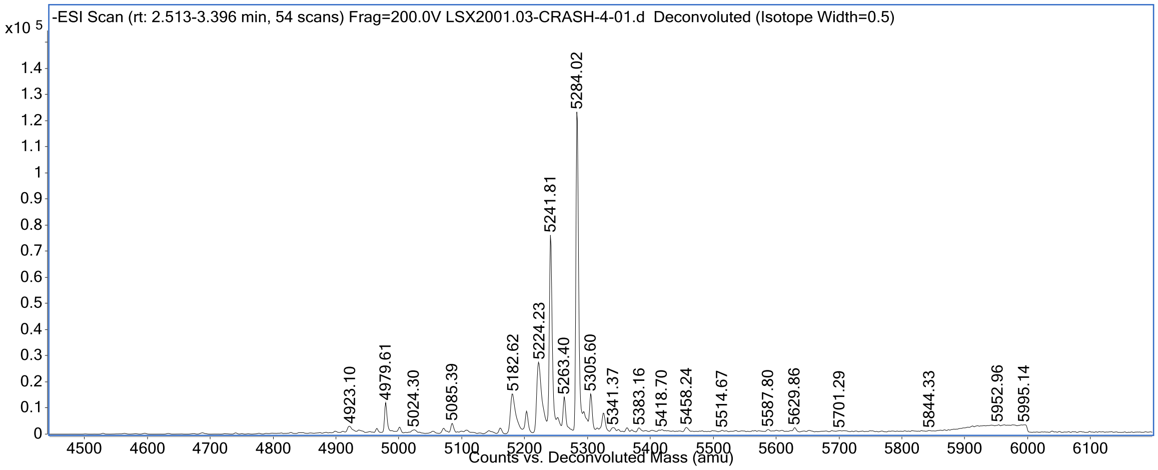


Fig. S34. Deconvoluted mass of **6e**

Figure S35, Mass Spectrum of 6f related to Figure 5.

Following **General Procedure 4**

Percent conversion: 84.06%

Exact mass: 5331.74, observed mass:5332.14

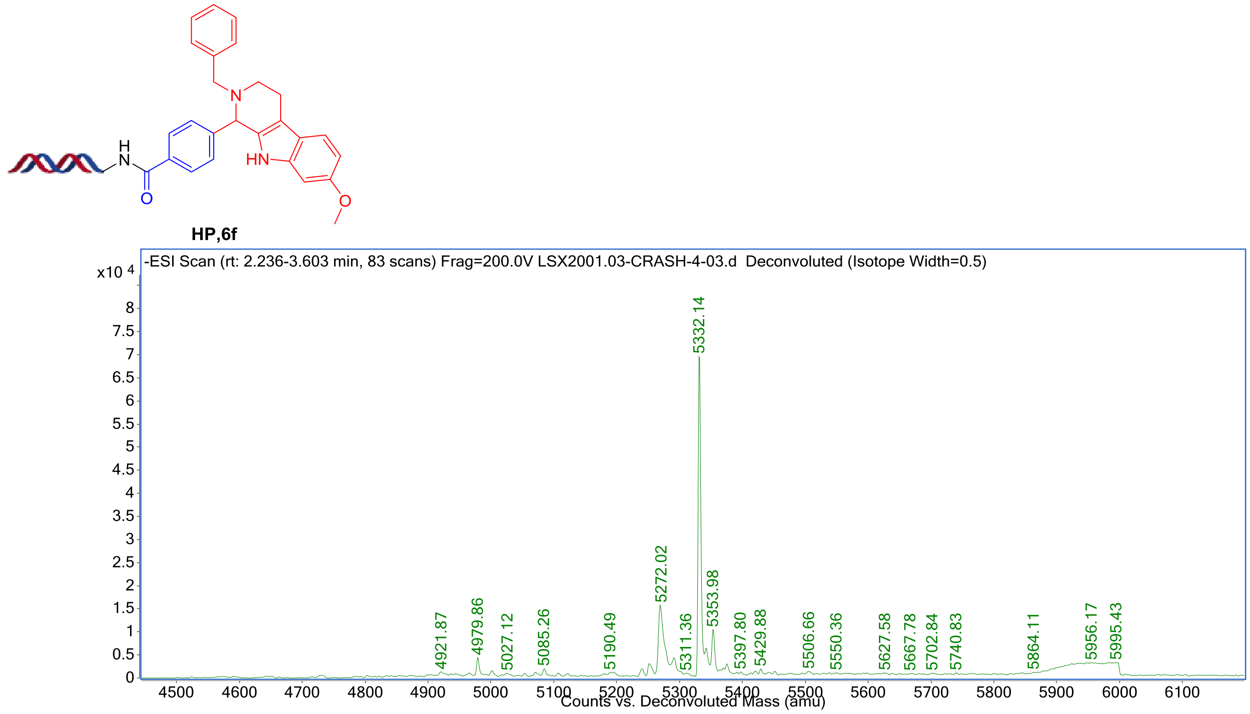


Fig. S35. Deconvoluted mass of **6f**

Figure S36, Mass Spectrum of 7a, related to Figure 5.

Percent conversion: 78%
Exact mass: 5229.57
Observed: 5230.09

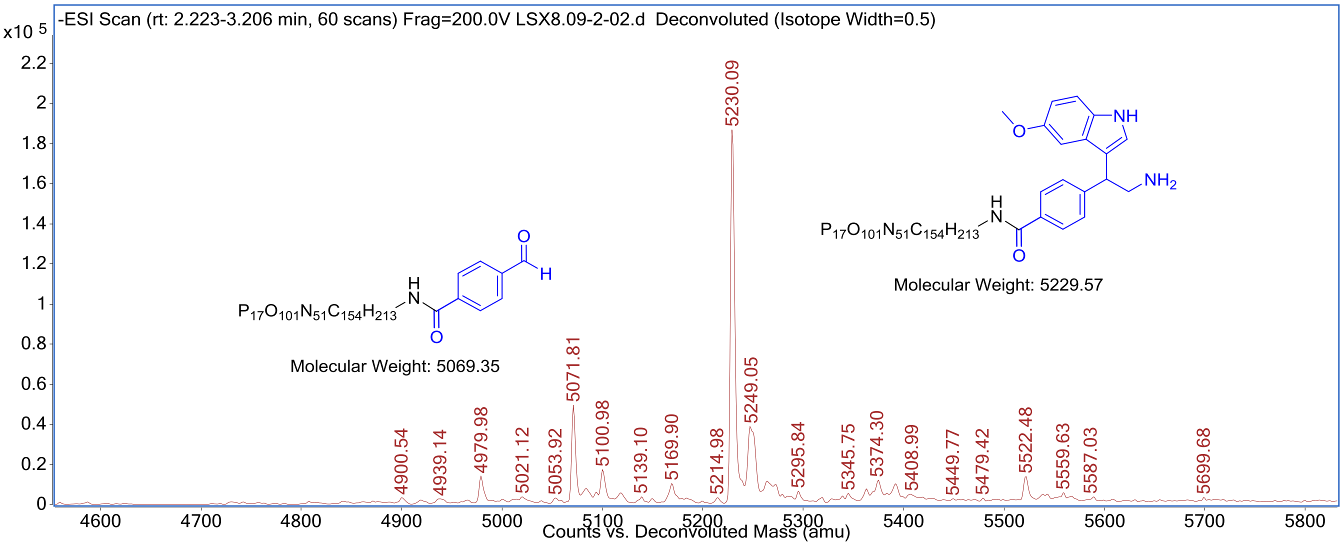


Fig.S36. Deconvoluted mass of **7a**

Figure S37, Mass Spectrum of 7b, related to Figure 5.

Percent conversion: 74.19%
Exact mass: 5229.57
Observed: 5230.13

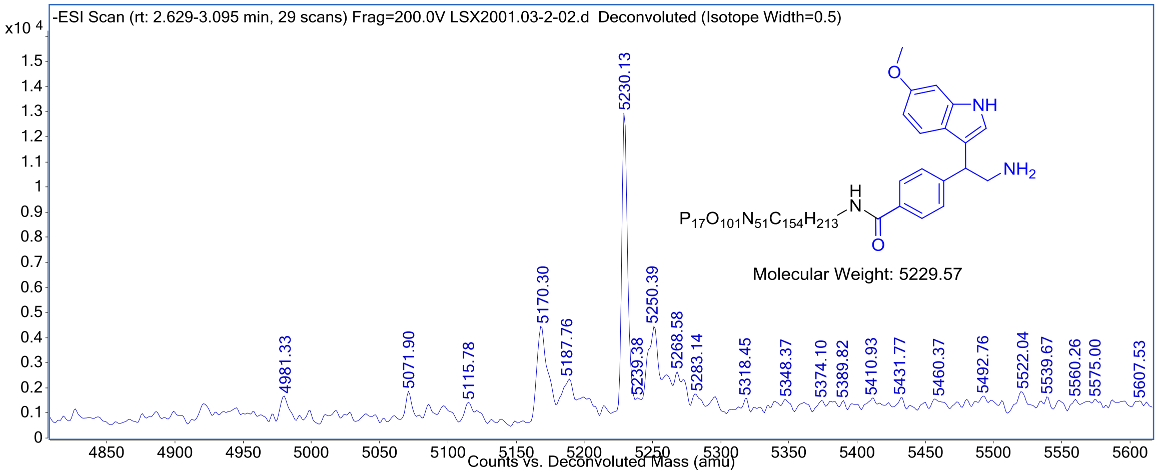


Fig. S37. Deconvoluted mass of **7b**

Figure S38, Mass Spectrum of 7c, related to Figure 5.

Percent conversion: 75.41%

Exact mass: 5259.59

Observed: 5260.14

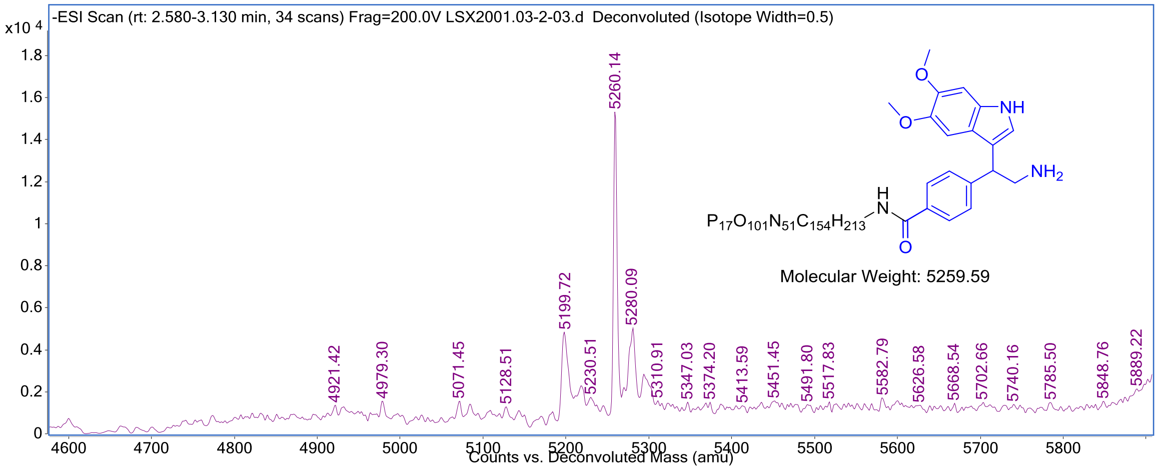


Fig. S38. Deconvoluted mass of **7c**

Figure S39, Mass Spectrum of 8a, related to Figure 5.

Percent conversion: 50%

Exact mass: 5362.67

Observed: 5362.87

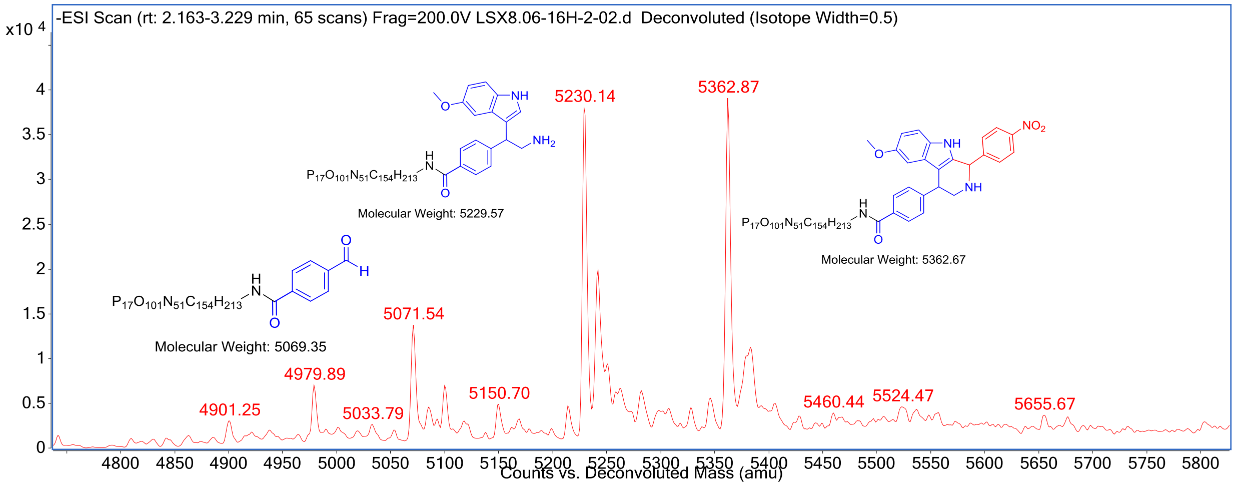


Fig. S39. Deconvoluted mass of **8a**

Figure S40, Mass Spectrum of 8b, related to Figure 5.

Percent conversion: 80.52%
Exact mass: 5476.46
Observed: 5477.12

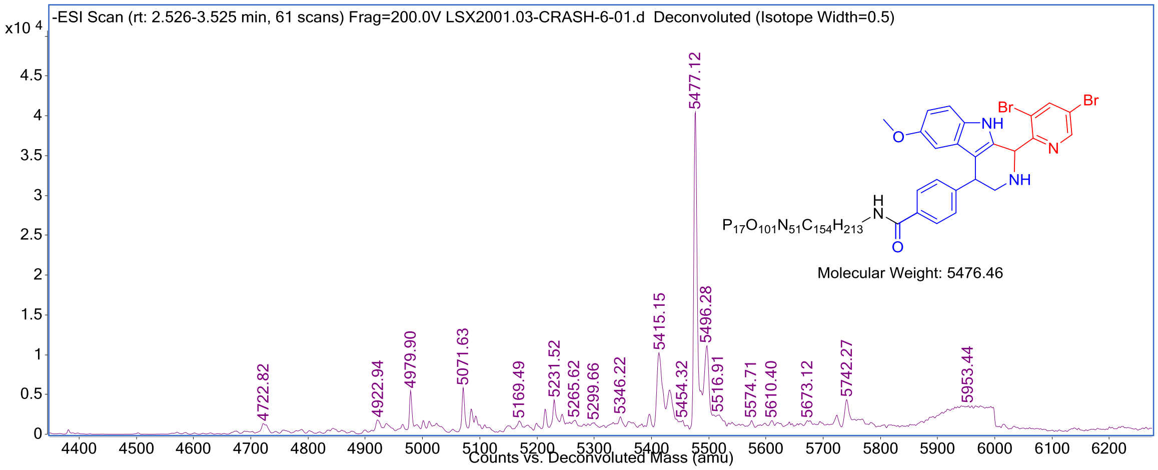


Fig. S40. Deconvoluted mass of **8b**

Figure S41, Mass Spectrum of 8b', related to Figure 5.

Percent conversion: 72.53%
Exact mass: 5580.57
Observed: 5581.35

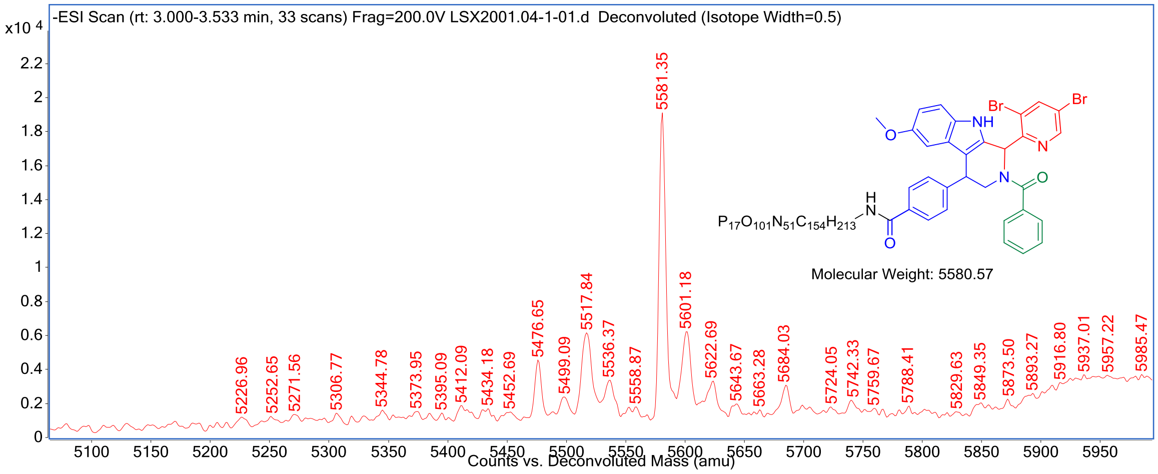


Fig. S41. Deconvoluted mass of **8b'**

Figure S42, Mass Spectrum of 8c, related to Figure 5.

Percent conversion: 80.52%
Exact mass: 5385.73
Observed: 5387.27

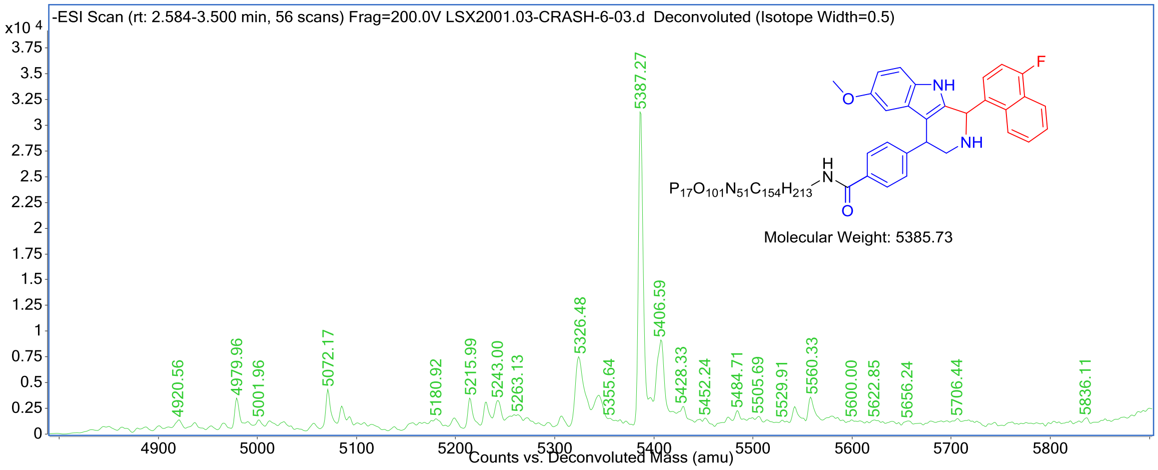


Fig. S42. Deconvoluted mass of **8c**

Figure S43, Mass Spectrum of 8c', related to Figure 5.

Percent conversion: 74.70%
Exact mass: 5489.84
Observed: 5491.59

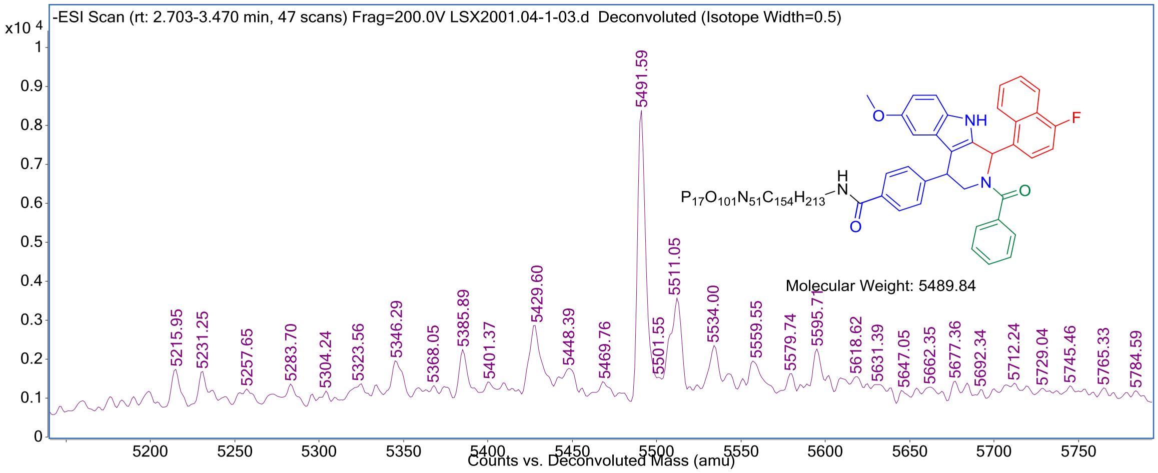


Fig. S43. Deconvoluted mass of **8c’**

Figure S44, Mass Spectrum of 8d, related to Figure 5.

Percent conversion: 66.99%
Exact mass: 5354.65
Observed: 5355.15

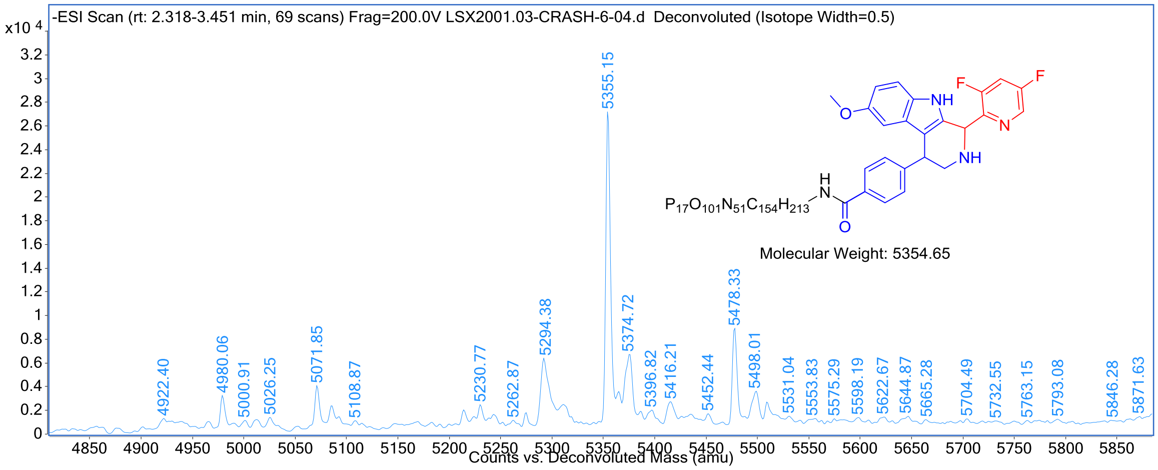


Fig. S44. Deconvoluted mass of **8d**

Figure S45, Mass Spectrum of 8d’, related to Figure 5.

Percent conversion: 52.94%
Exact mass: 5458.75
Observed: 5459.34

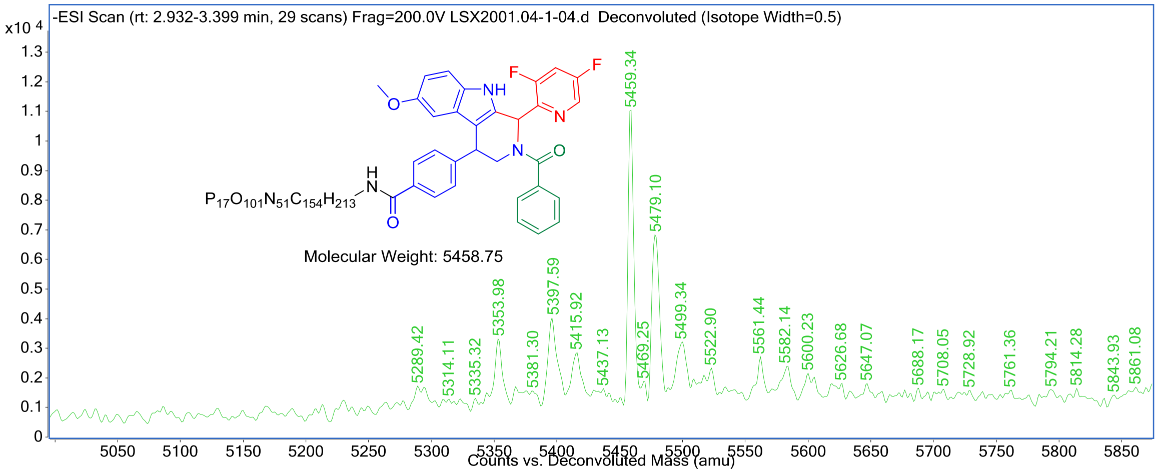


Fig. S45. Deconvoluted mass of **8d'**

Figure S46, Mass Spectrum of 8e, related to Figure 5.

Percent conversion: 43.95%
Exact mass: 5371.10
Observed: 5371.62

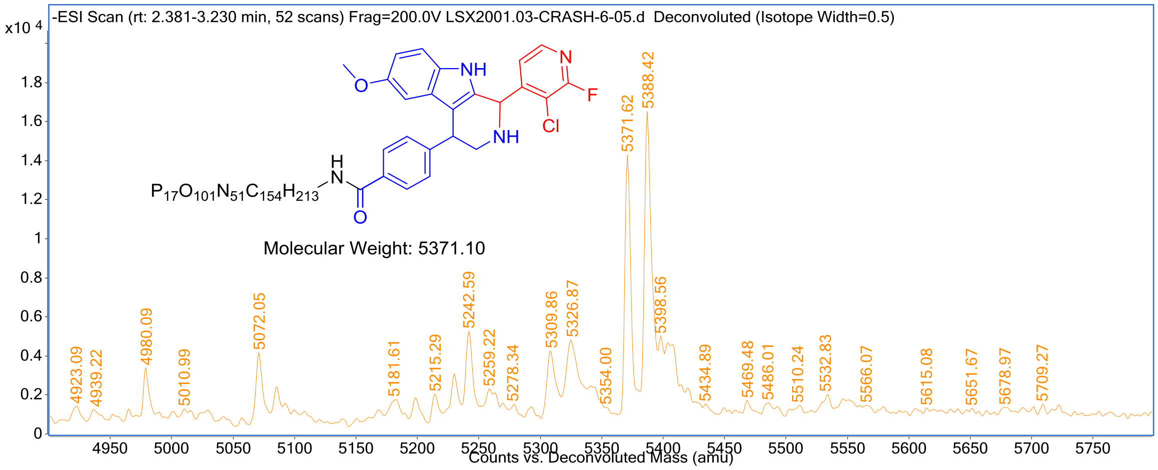


Fig. S46. Deconvoluted mass of **8e**

Figure S47, Mass Spectrum of 8f, related to Figure 5.

Percent conversion: 66.99%
Exact mass: 5414.81
Observed: 5414.79

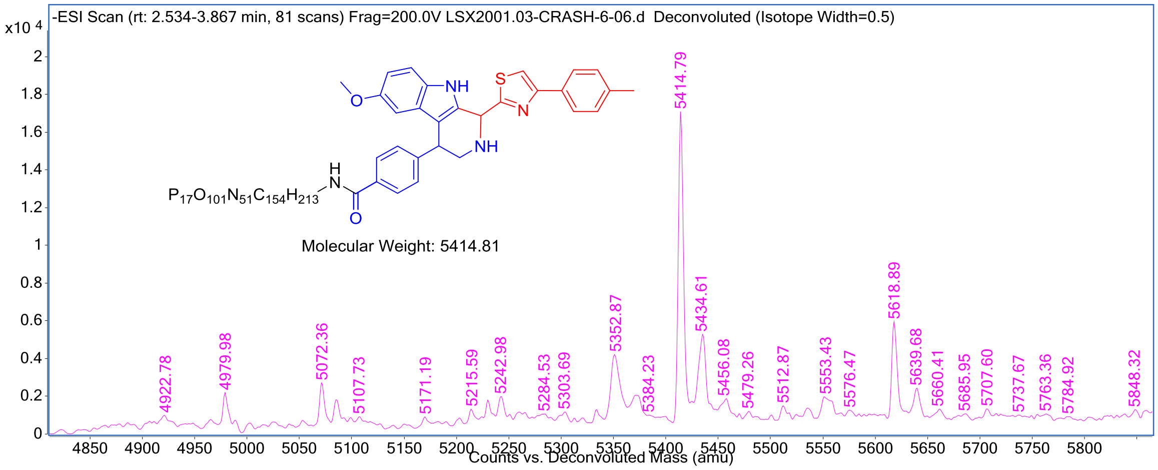


Fig. S47. Deconvoluted mass of **8f**

Figure S48, Mass Spectrum of 8f’, related to Figure 5.

Percent conversion: 66.30%
Exact mass: 5504.94
Observed: 5505.49

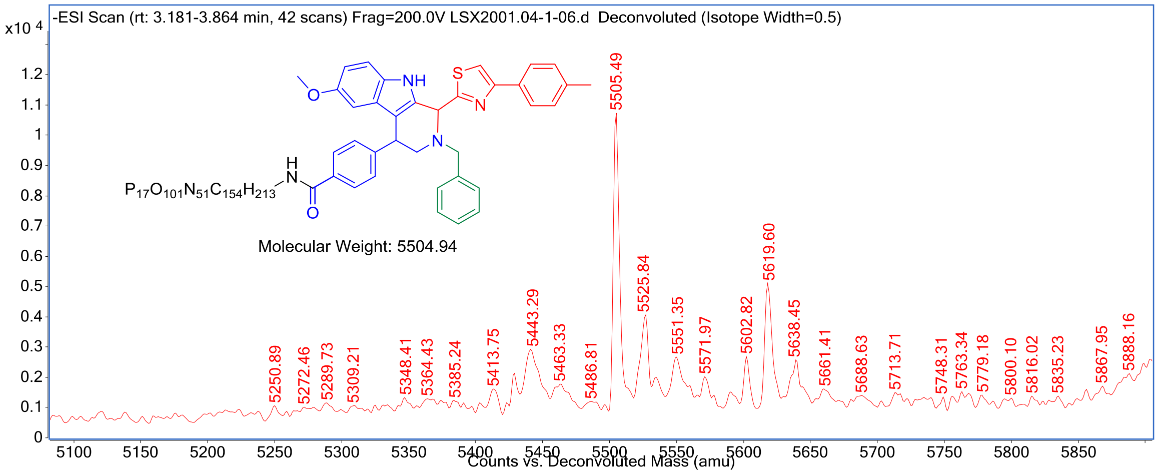


Fig. S48. Deconvoluted mass of **8f’**

Figure S49, Mass Spectrum of 8g, related to Figure 5.

Percent conversion: 75.96%
Exact mass: 5371.10
Observed: 5371.59

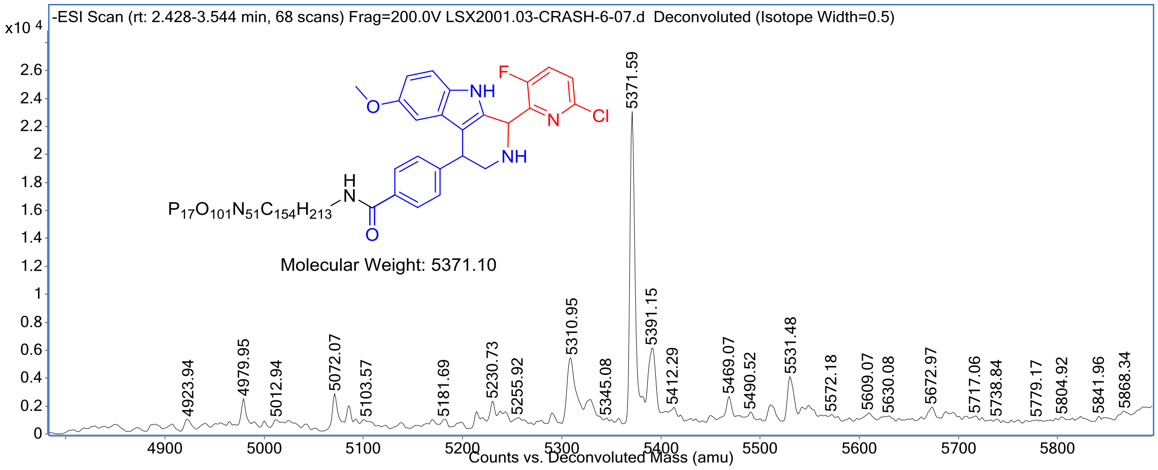


Fig. S49. Deconvoluted mass of **8g**

Figure S50, Mass Spectrum of 8g’, related to Figure 5.

Percent conversion: 83.56%
Exact mass: 5461.22
Observed: 5461.86

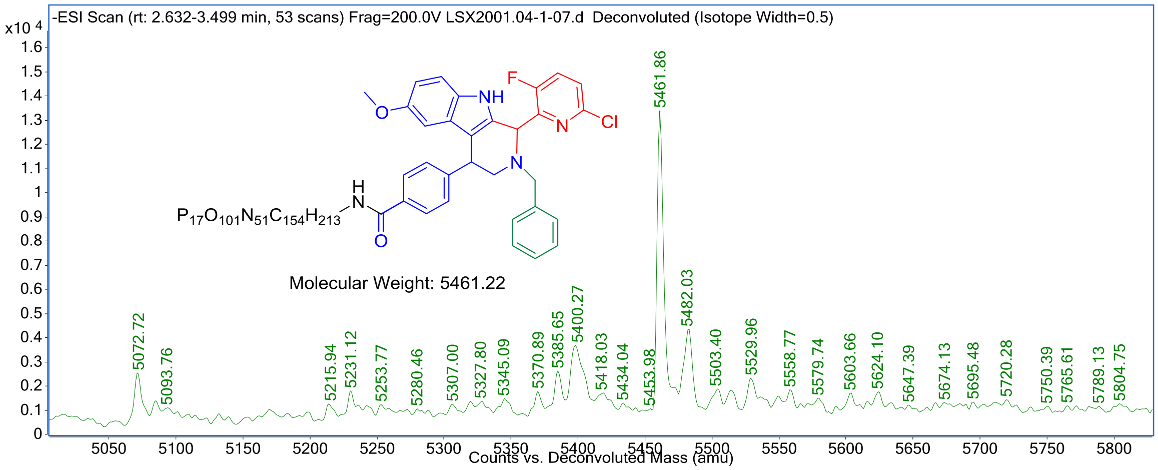


Fig. S50. Deconvoluted mass of **8g’**

Figure S51, Mass Spectrum of 8h, related to Figure 5.

Percent conversion: 73.58%
Exact mass: 5358.69
Observed: 5359.30

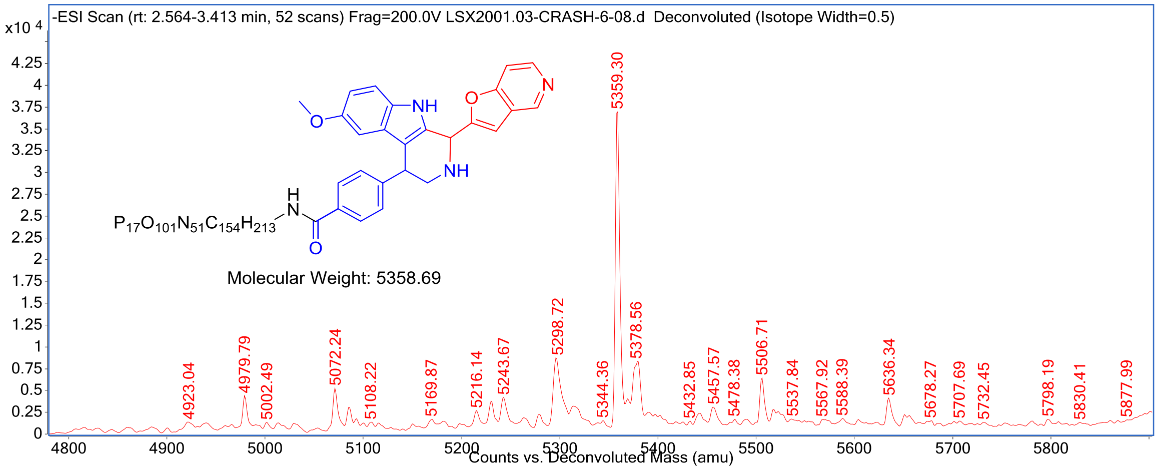


Fig. S51. Deconvoluted mass of **8h**

Figure S52, Mass Spectrum of 8h', related to Figure 5.

Percent conversion: 70.45%
Exact mass: 5448.81
Observed: 5449.44

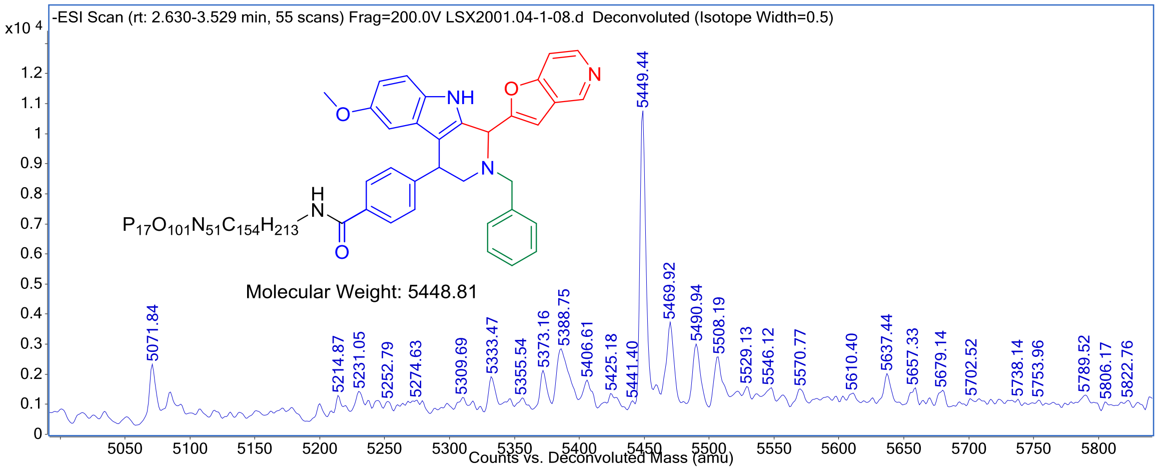


Fig. S52. Deconvoluted mass of **8h'**

Figure S53, Mass Spectrum of 8i, related to Figure 5.

Percent conversion: 54.66%
Exact mass: 5380.67
Observed: 5380.94

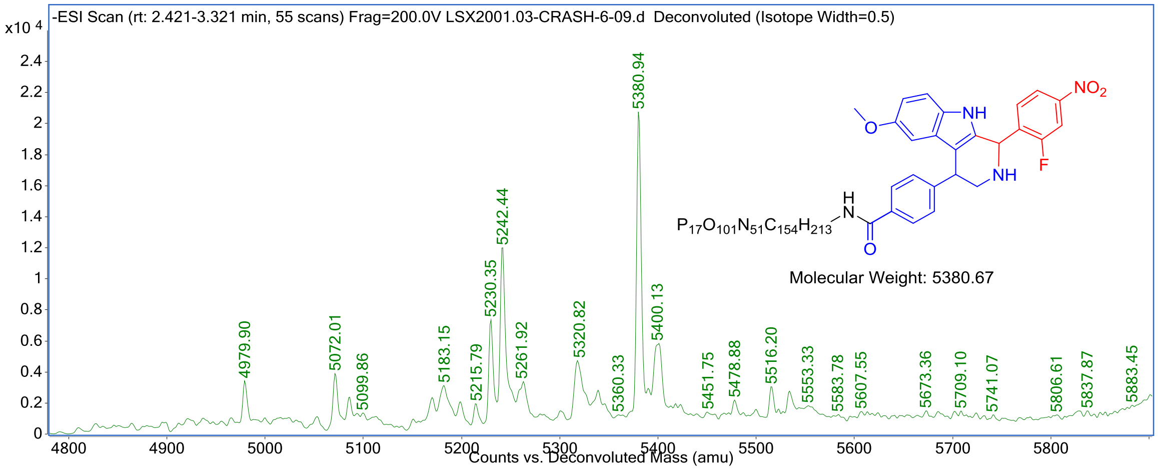


Fig. S53. Deconvoluted mass of **8i**

Figure S54, Mass Spectrum of 8i’, related to Figure 5.

Percent conversion: 54.78%
Exact mass: 5470.79
Observed: 5471.55

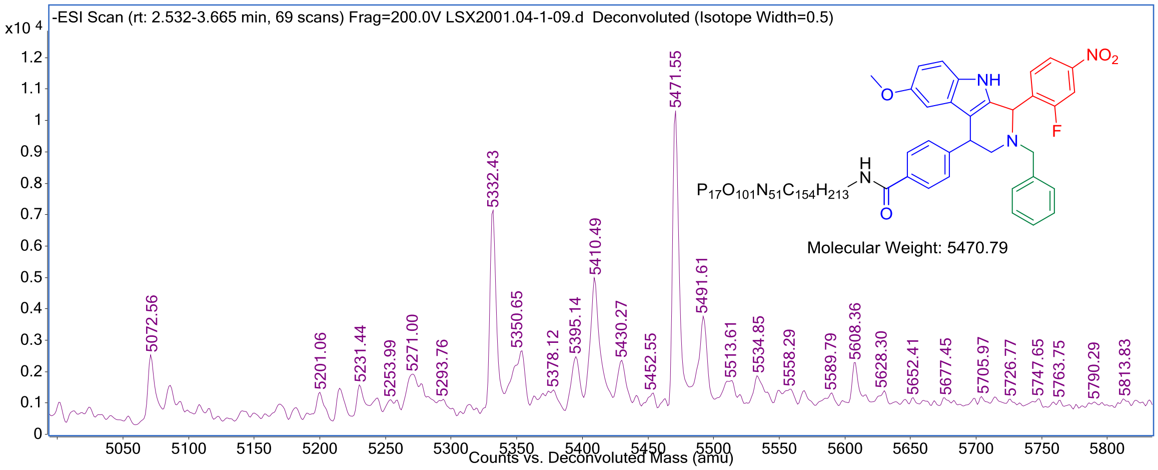


Fig. S54. Deconvoluted mass of **8i’**

Figure S55, Mass Spectrum of j, related to Figure 5.

Percent conversion: 69.07%
Exact mass: 5346.72
Observed: 5347.18

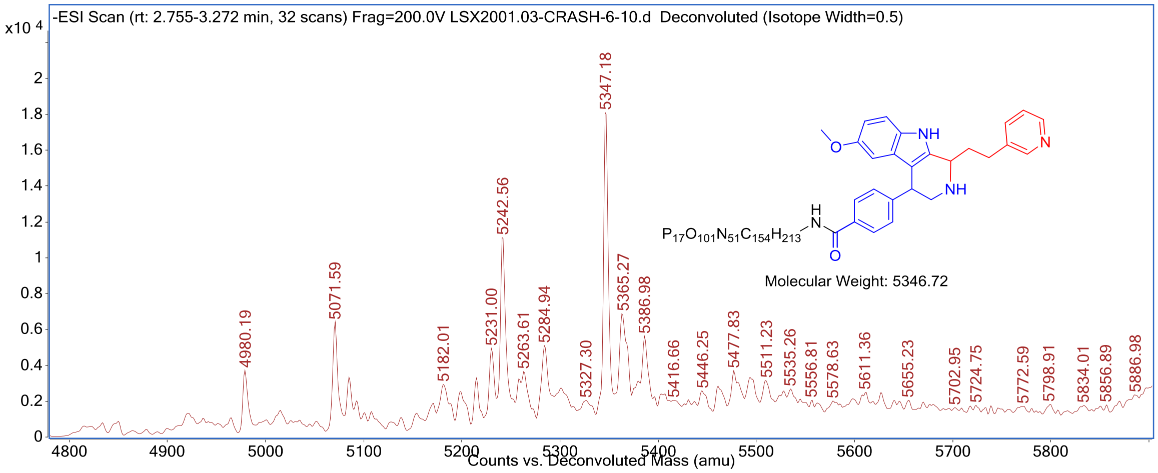


Fig. S55. Deconvoluted mass of **8j**

Figure S56, Mass Spectrum of 8k, related to Figure 5.

Percent conversion: 75.81%
Exact mass: 5362.67
Observed: 5361.87

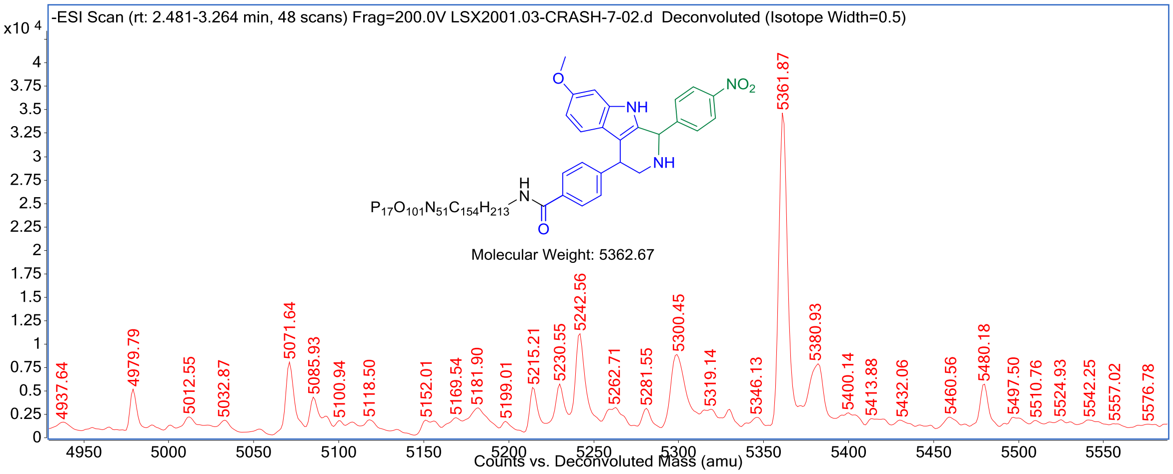


Fig. S56. Deconvoluted mass of **8k**

Figure S57, Mass Spectrum of 8l, related to Figure 5.

Percent conversion: 81.48%
Exact mass: 5392.70
Observed: 5391.74

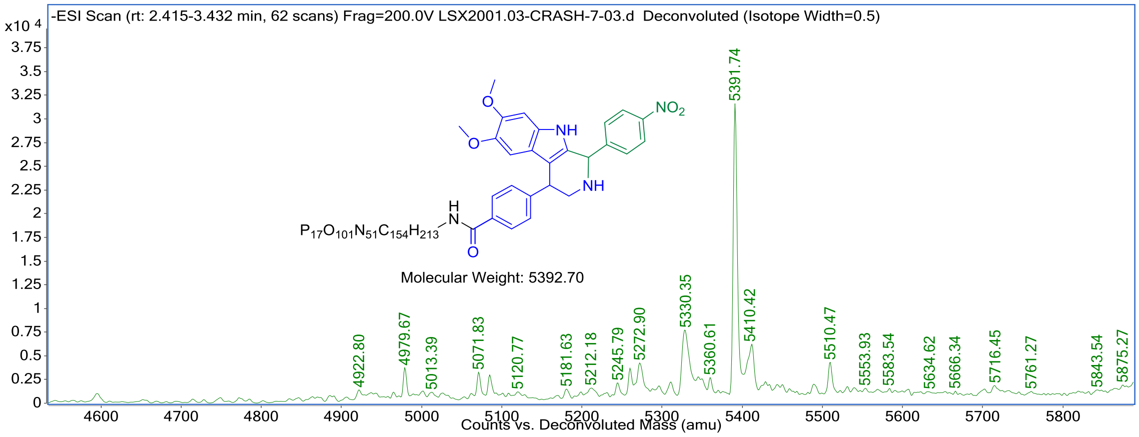


Fig. S57. Deconvoluted mass of **8l**

Figure S58, Mass Spectrum of 3a, related to Figure 3.

Exact mass: 5414.79

Observed: 5415.15

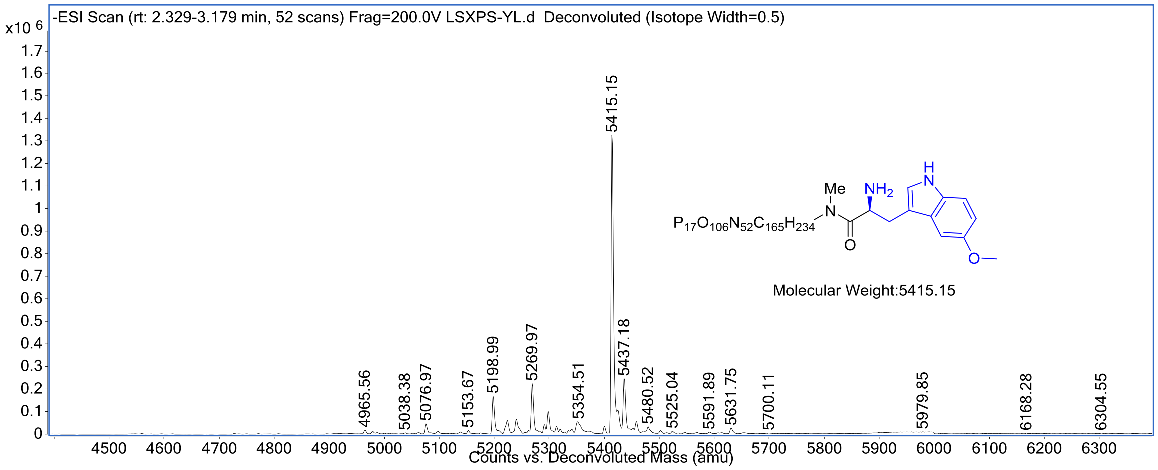


Fig. S58. Deconvoluted mass of on-DNA product **3a**

Figure S59, Mass Spectrum of 4ea, related to Figure 4.

Percent conversion: 83.84%

Exact mass: 5545.90

Observed: 5546.20

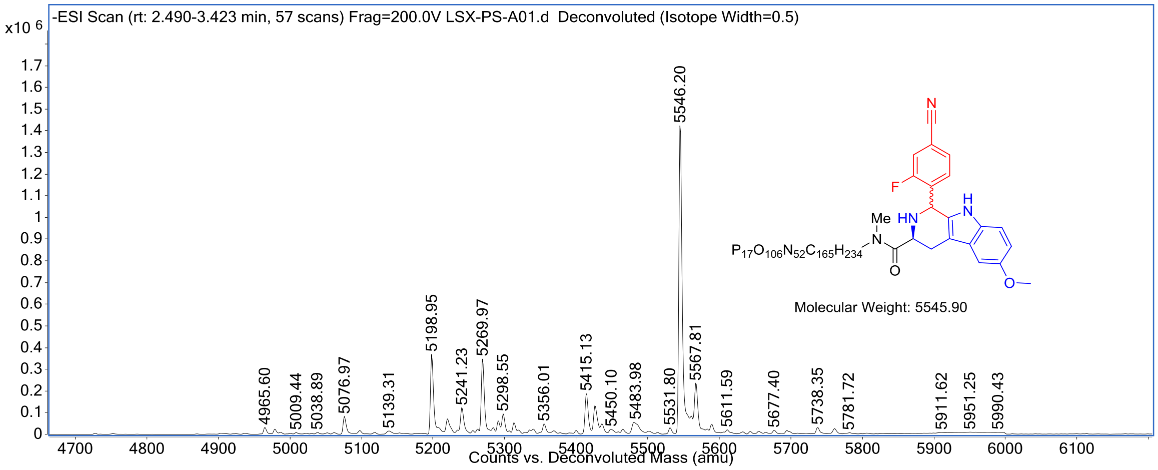


Fig. S59. Deconvoluted mass of **4ea**

Figure S60, Mass Spectrum of 4eb, related to Figure 4.

Percent conversion: 88.04%
Exact mass:5620.45
Observed: 5620.75

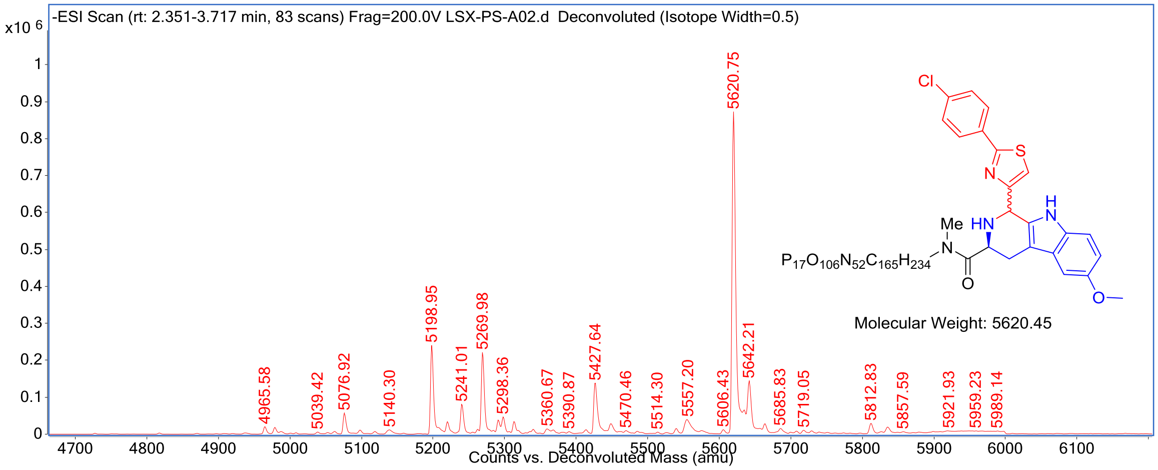


Fig. S60. Deconvoluted mass of **4eb**

Figure S61, Mass Spectrum of 4ec, related to Figure 4.

Percent conversion: 84.47%
Exact mass: 5542.92
Observed: 5542.90

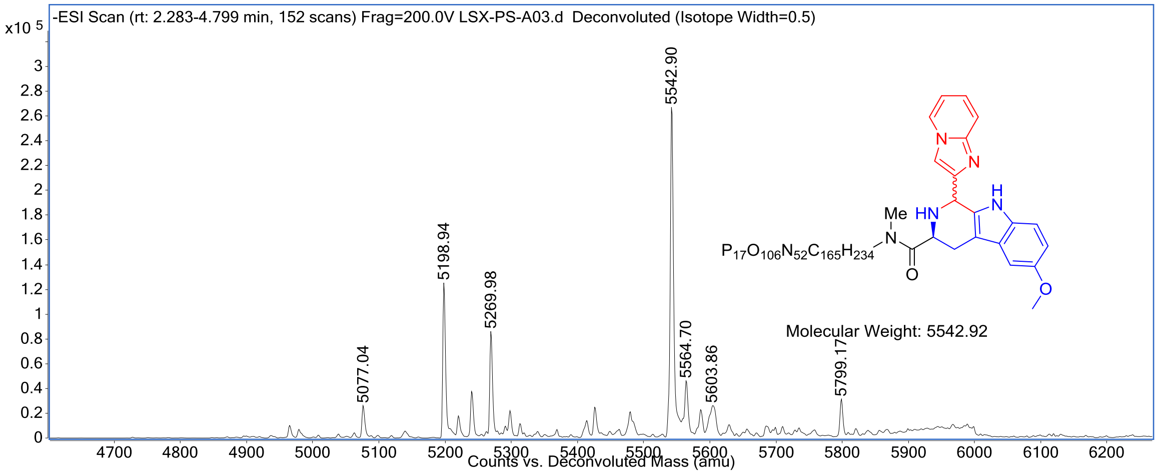


Fig. S61. Deconvoluted mass **4ec**

Figure S62, Mass Spectrum of 4ed, related to Figure 4.

Percent conversion: 91.51%
Exact mass: 5661.68
Observed: 5662.16

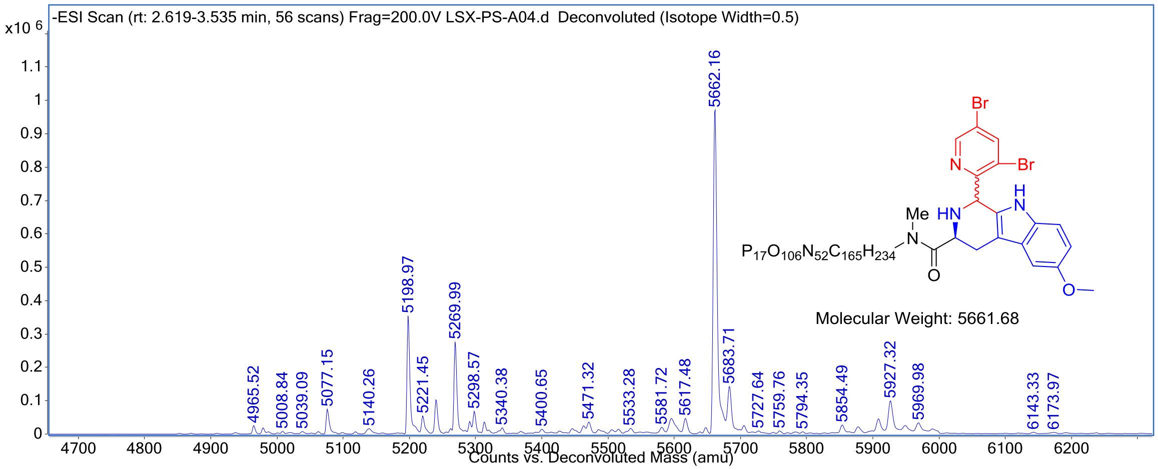


Fig. S62. Deconvoluted mass of **4ed**

Figure S63, Mass Spectrum of 4ee, related to Figure 4.

Percent conversion: 77.91%
Exact mass: 5604.00
Observed: 5604.31

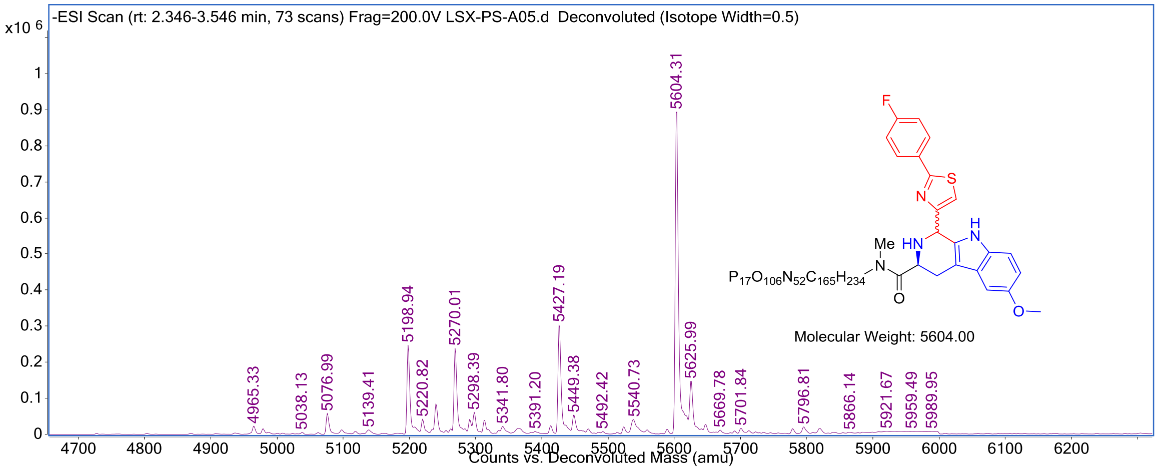


Fig. S63. Deconvoluted mass of **4ee**

Figure S64, Mass Spectrum of 4ef, related to Figure 4.

Percent conversion: 35.04%
Exact mass: 5572.77
Observed: 5573.14

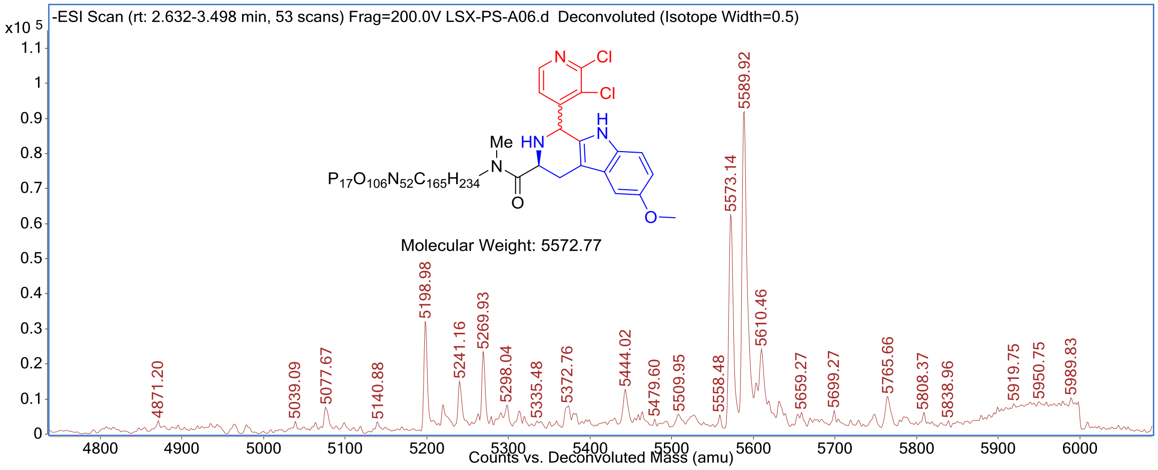


Fig. S64. Deconvoluted mass of **4ef**

Figure S65, Mass Spectrum of 4eg, related to Figure 4.

Percent conversion: 87.51%
Exact mass: 5592.03
Observed: 5592.36

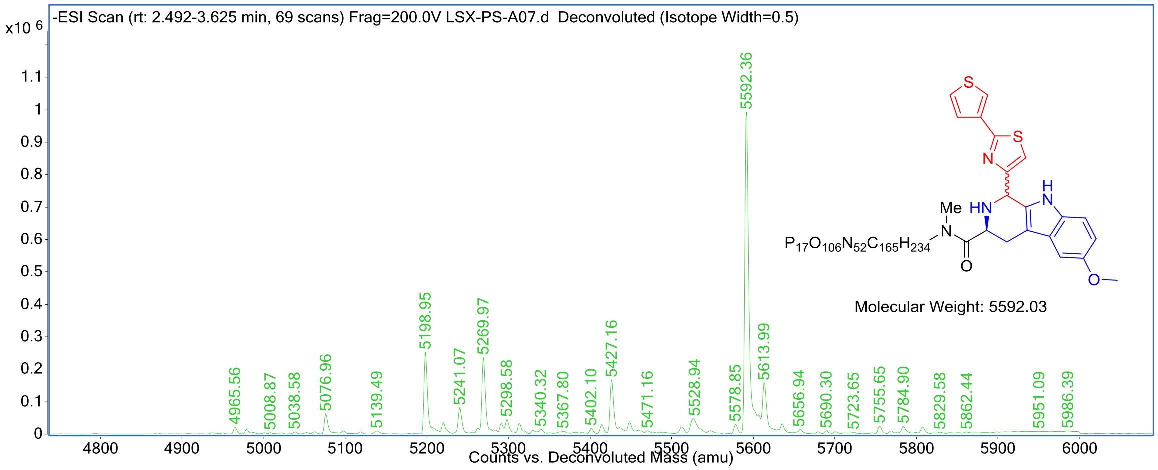


Fig. S65. Deconvoluted mass of **4eg**

Figure S66, Mass Spectrum of 4eh, related to Figure 4.

Percent conversion: 98%
Exact mass: 5577.91
Observed: 5578.28

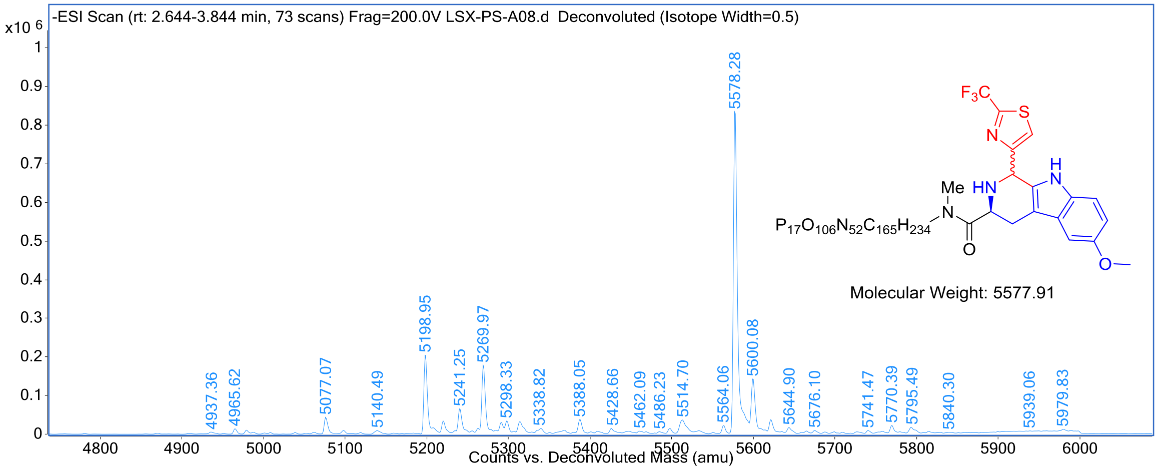


Fig. S66. Deconvoluted mass of **4eh**

Figure S67, Mass Spectrum of 4ei, related to Figure 4.

Percent conversion: 30.22%
Exact mass: 5617.22
Observed: 5617.60

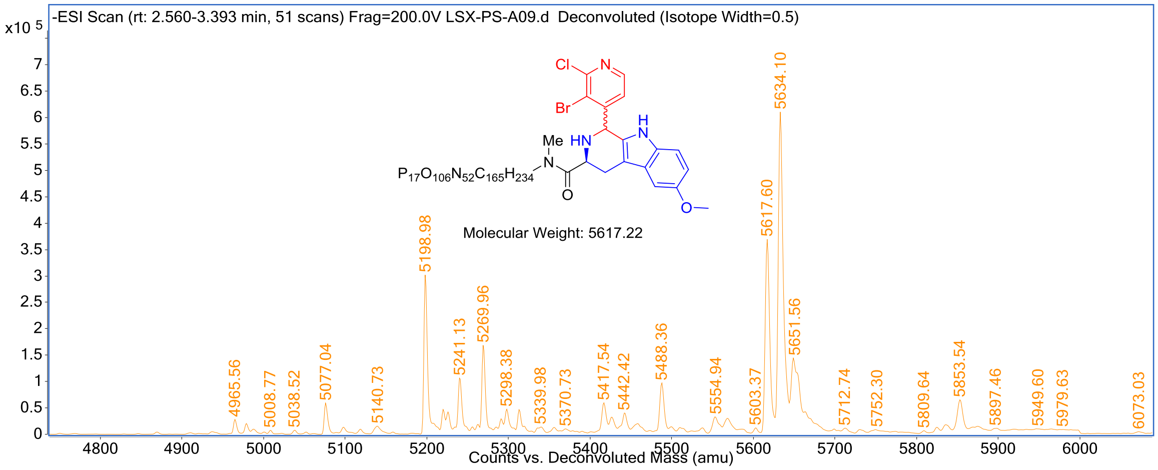


Fig. S67. Deconvoluted mass of **4ei**

Figure S68, Mass Spectrum of 4ej, related to Figure 4.

Percent conversion: 98%
Exact mass: 5571.94
Observed: 5572.25

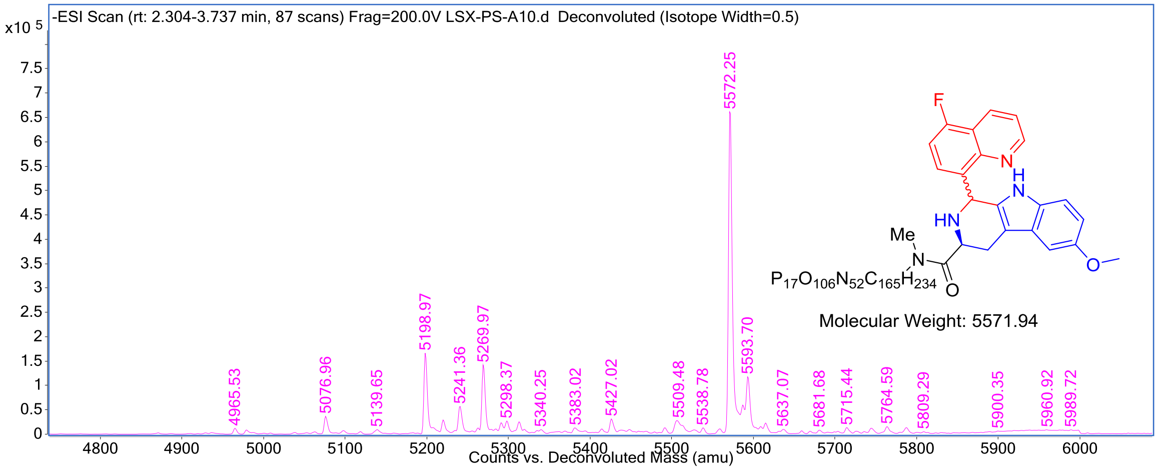


Fig. S68. Deconvoluted mass of **4ej**

Figure S69, Mass Spectrum of 4ek, related to Figure 4.

Percent conversion:98%
Exact mass: 5539.87
Observed: 5540.21

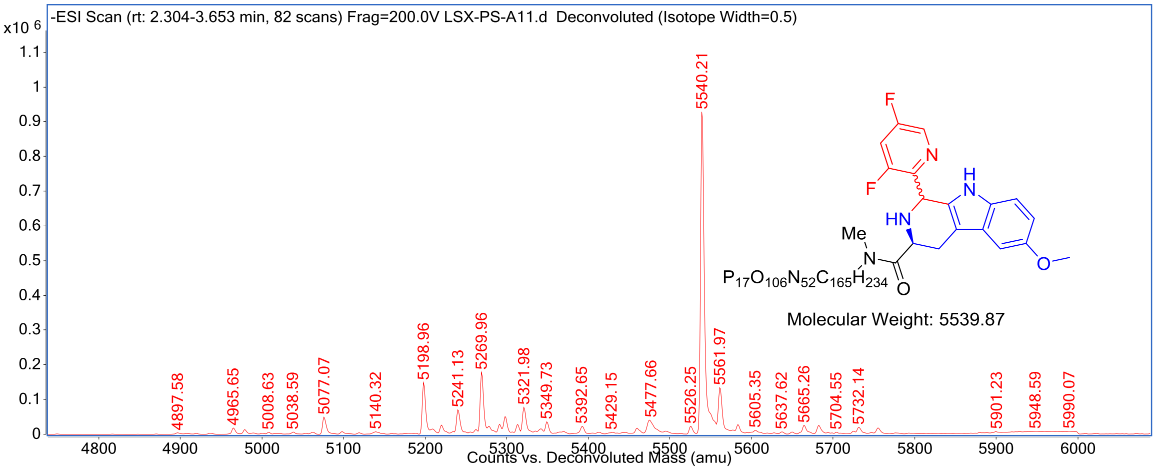


Fig. S69. Deconvoluted mass of **4ek**

Figure S70, Mass Spectrum of 4el, related to Figure 4.

Percent conversion: 82.34%
Exact mass: 5556.95
Observed: 5557.30

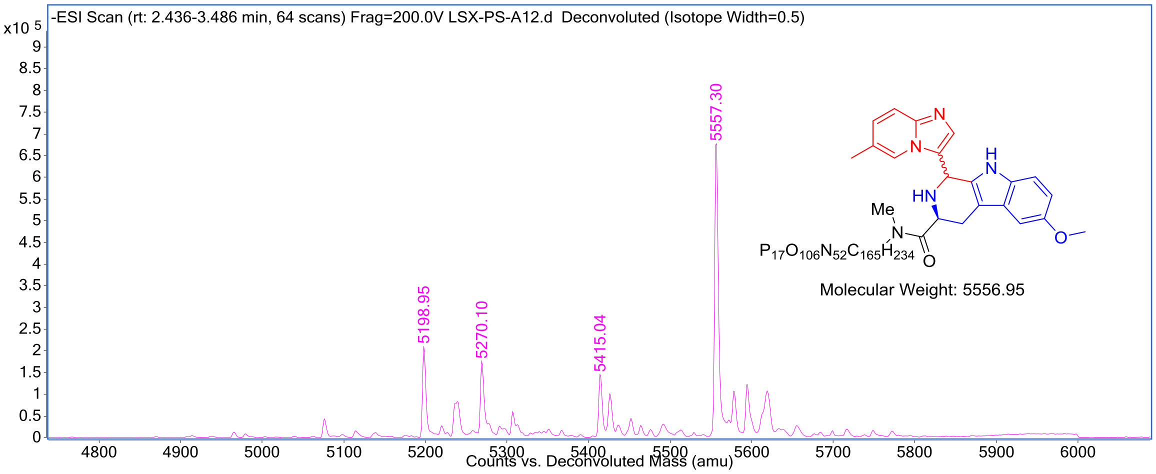


Fig. S70. Deconvoluted mass of **4el**

Figure S71, Mass Spectrum of 4em, related to Figure 4.

Percent conversion: 91.69%
Exact mass: 5531.94
Observed: 5532.25

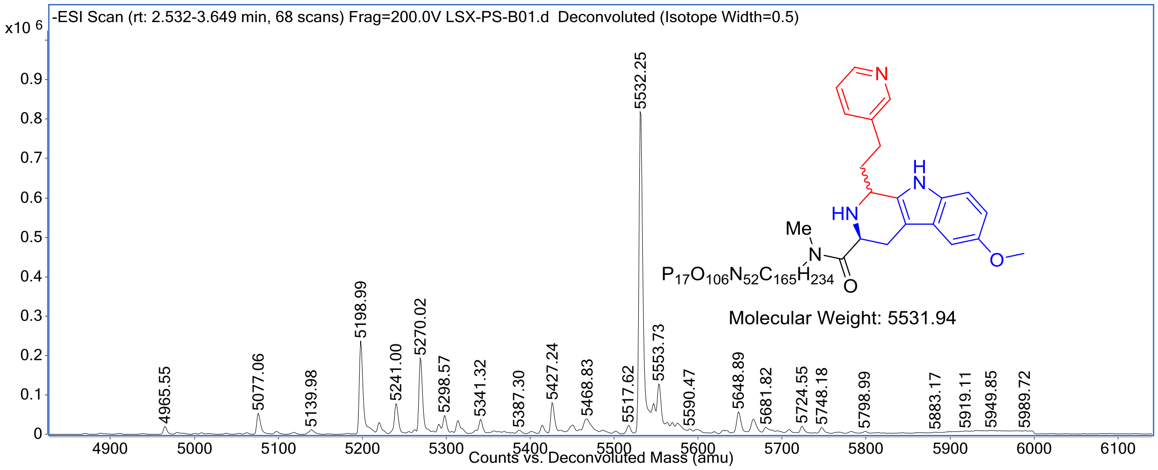


Fig. S71. Deconvoluted mass of **4em**

Figure S72, Mass Spectrum of 4en, related to Figure 4.

Percent conversion: 39.19%
Exact mass: 5544.35
Observed: 5544.35

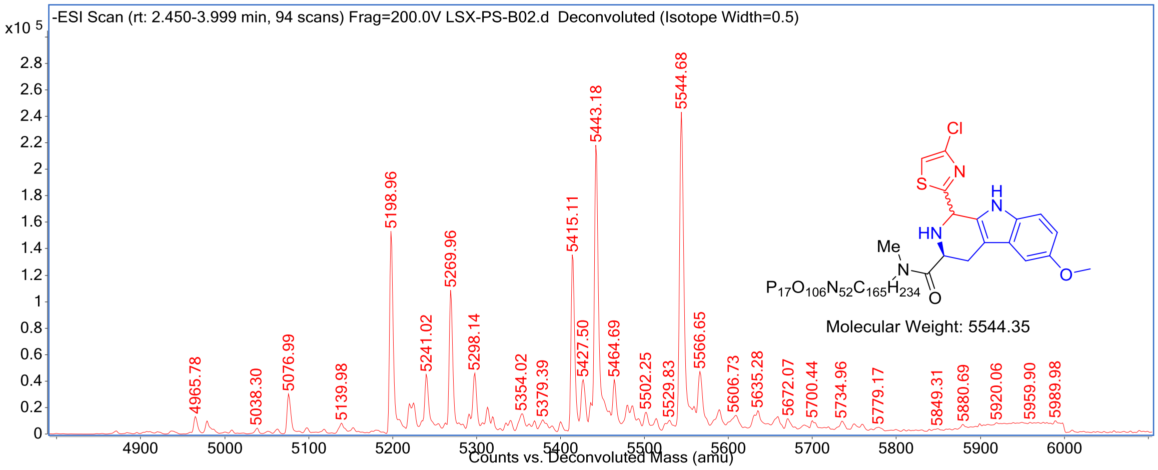


Fig. S72. Deconvoluted mass of **4en**

Figure S73, Mass Spectrum of 4eo, related to Figure 4.

Percent conversion: 88.90%
Exact mass: 5583.88
Observed: 5582.50

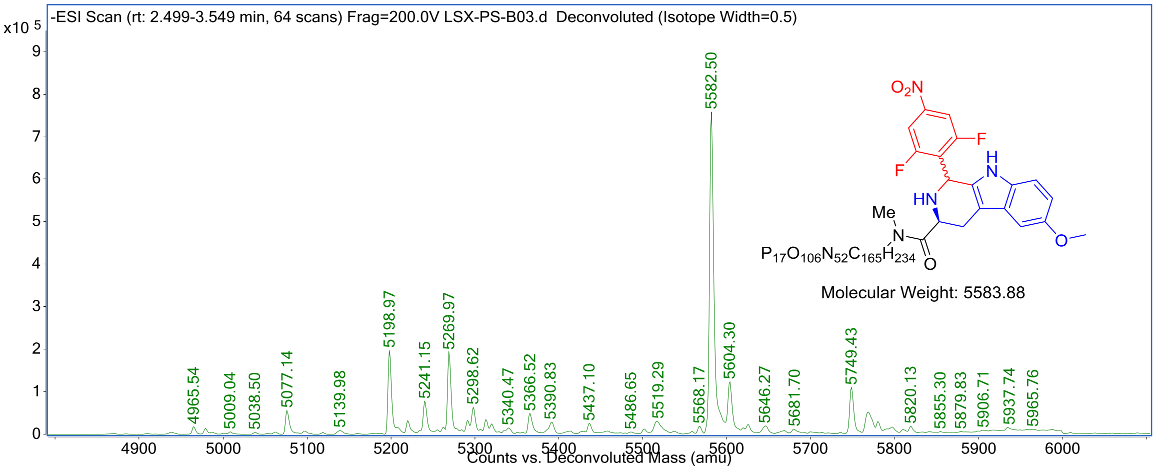


Fig. S73. Deconvoluted mass of **4eo**

Figure S74, Mass Spectrum of 4ep, related to Figure 4.

Percent conversion: 95.00%
Exact mass: 5565.89
Observed: 5566.22

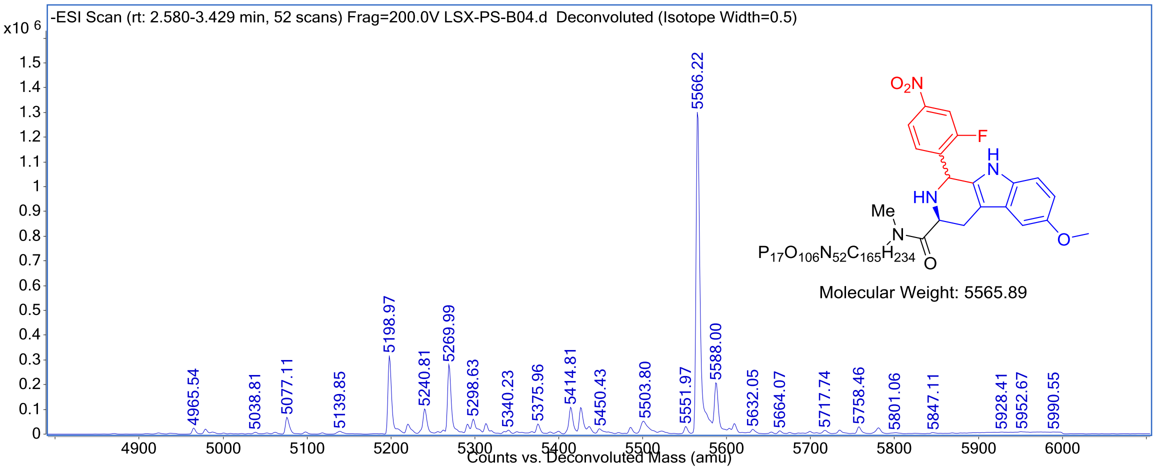


Fig. 74. Deconvoluted mass of **4ep**

Figure S75, Mass Spectrum of 4eq, related to Figure 4.

Percent conversion: 83.67%
Exact mass: 5542.92
Observed: 5543.01

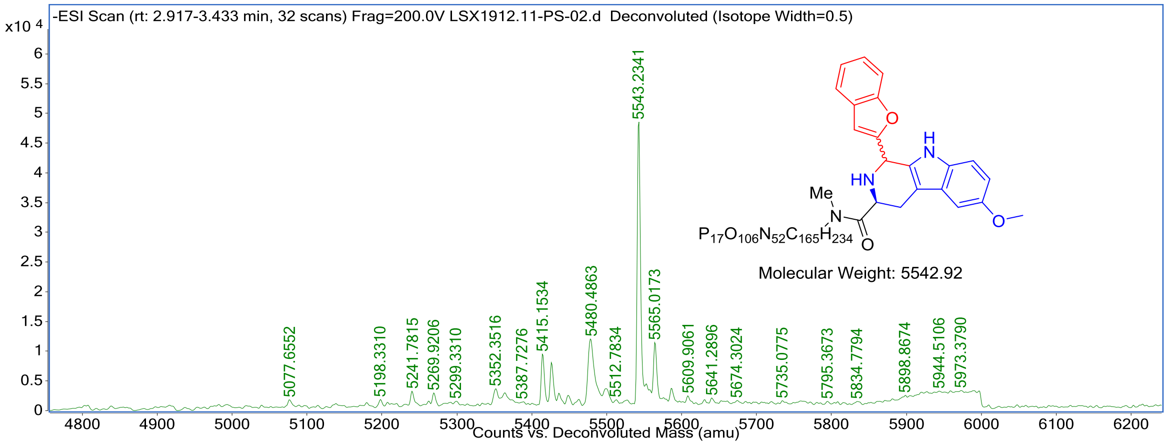


Fig. S75. Deconvoluted mass of **4eq**

Figure S76, Mass Spectrum of 4er, related to Figure 4.

Percent conversion: 83.50%
Exact mass: 5588.80
Observed: 5588.73

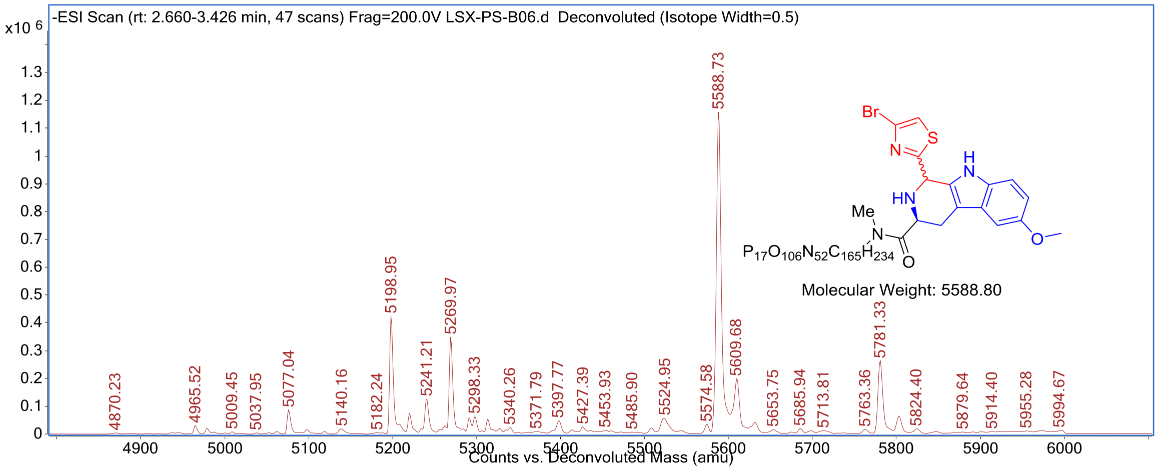


Fig. S76. Deconvoluted mass of **4er**

Figure S77, Mass Spectrum of 4es, related to Figure 4.

Percent conversion: 27.74%
Exact mass: 5616.03
Observed: 5616.45

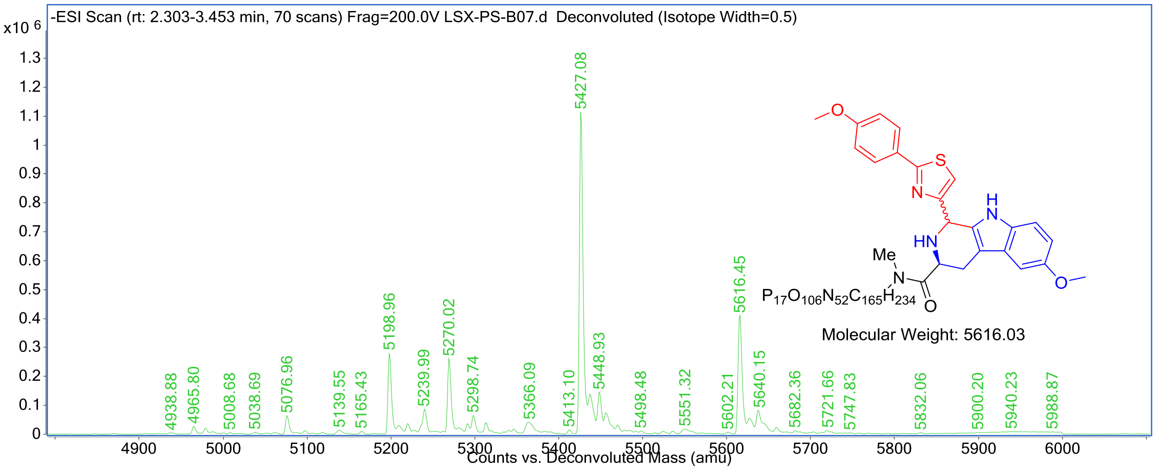


Fig. S77. Deconvoluted mass of **4es**

Figure S78, Mass Spectrum of 4et, related to Figure 4.

Percent conversion: 76.96%
Exact mass: 5592.89
Observed: 5593.27

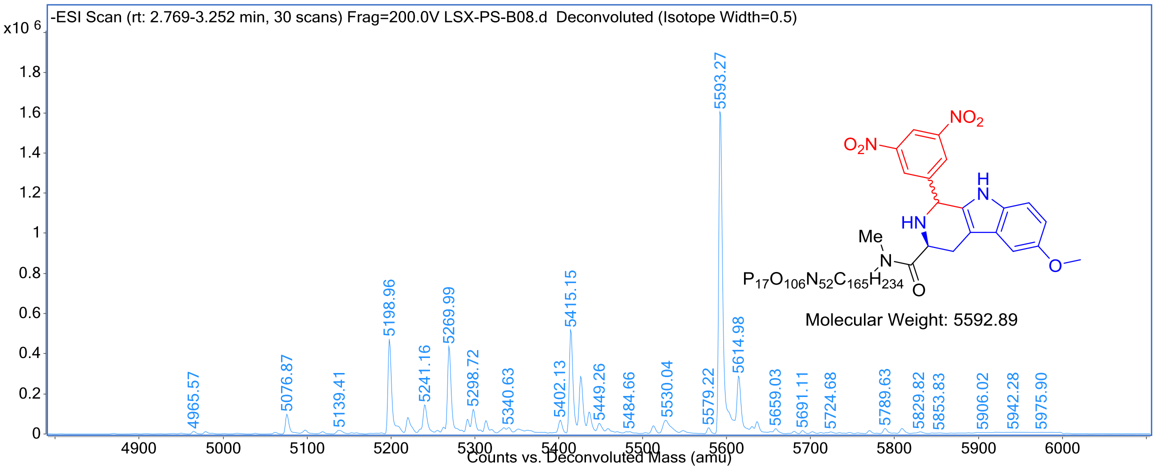


Fig. S78. Deconvoluted mass of **4et**

Figure S79, Mass Spectrum of 4eu, related to Figure 4.

Percent conversion: 98%
Exact mass: 5543.91
Observed: 5593.27

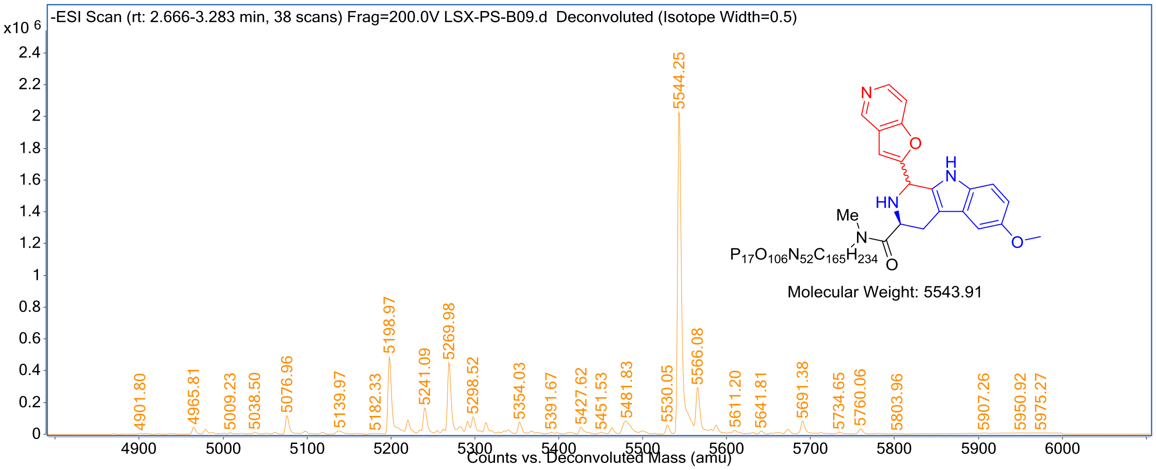


Fig. S79. Deconvoluted mass of **4eu**

Figure S80, Mass Spectrum of 4ev, related to Figure 4.

Percent conversion: 98%
Exact mass: 5556.32
Observed: 5556.68

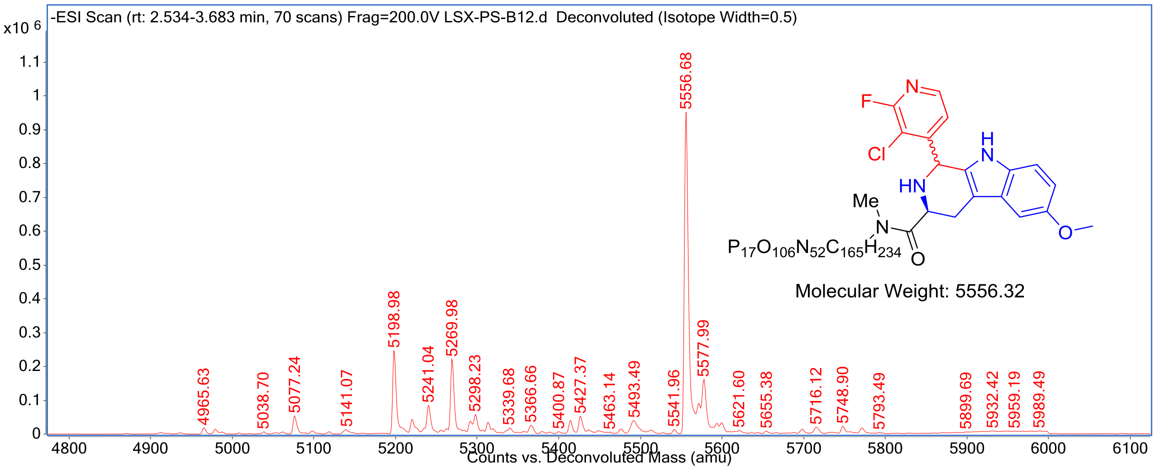


Fig. S80. Deconvoluted mass of **4ev**

Figure S81, Mass Spectrum of 4ew, related to Figure 4.

Percent conversion: 68.87%
Exact mass: 5617.22
Observed: 5617.64

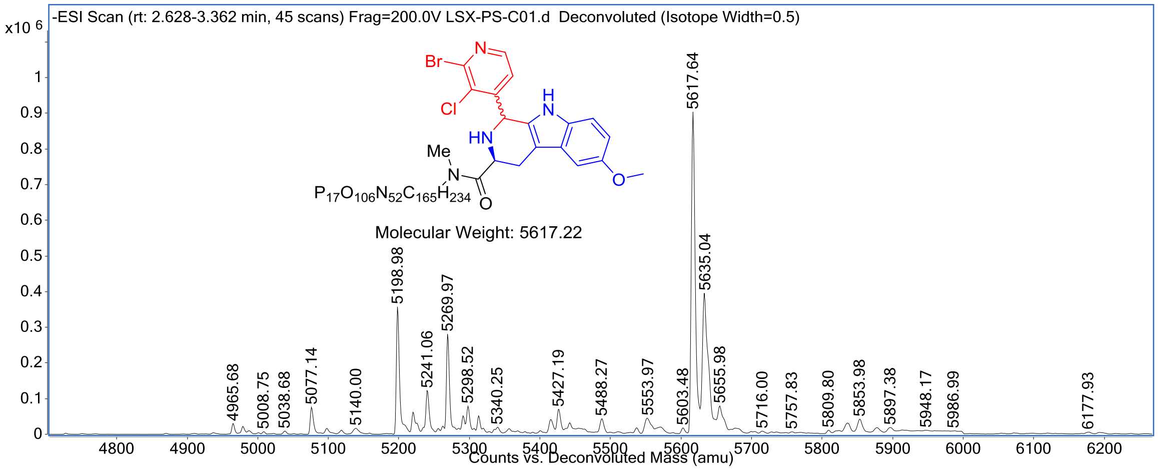


Fig. S81. Deconvoluted mass of **4ew**

Figure S82, Mass Spectrum of 4ex, related to Figure 4.

Percent conversion: 98%
Exact mass: 5600.03
Observed: 5600.33

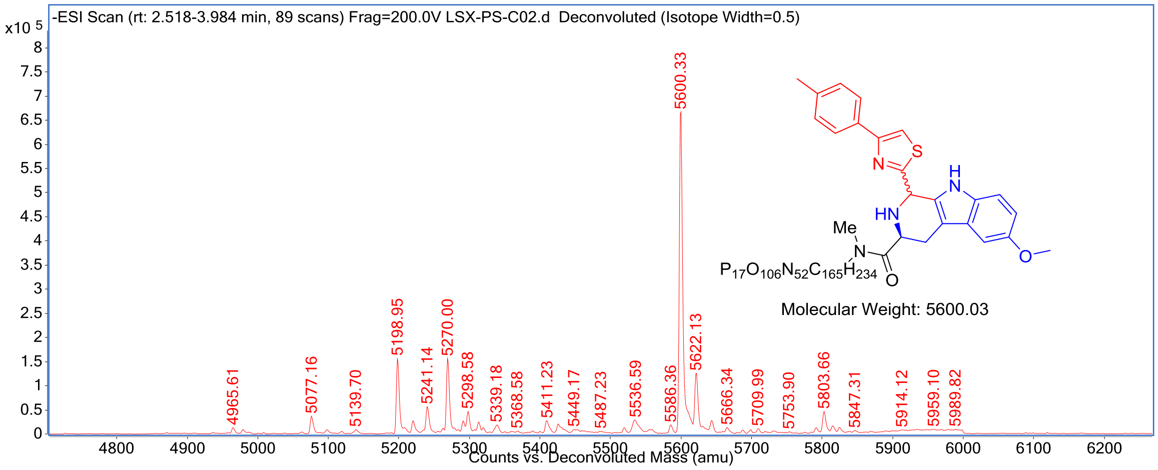


Fig. S82. Deconvoluted mass of **4ex**

Figure S83, Mass Spectrum of 4ey, related to Figure 4.

Percent conversion: 72.60%
Exact mass: 5565.39
Observed: 5565.71

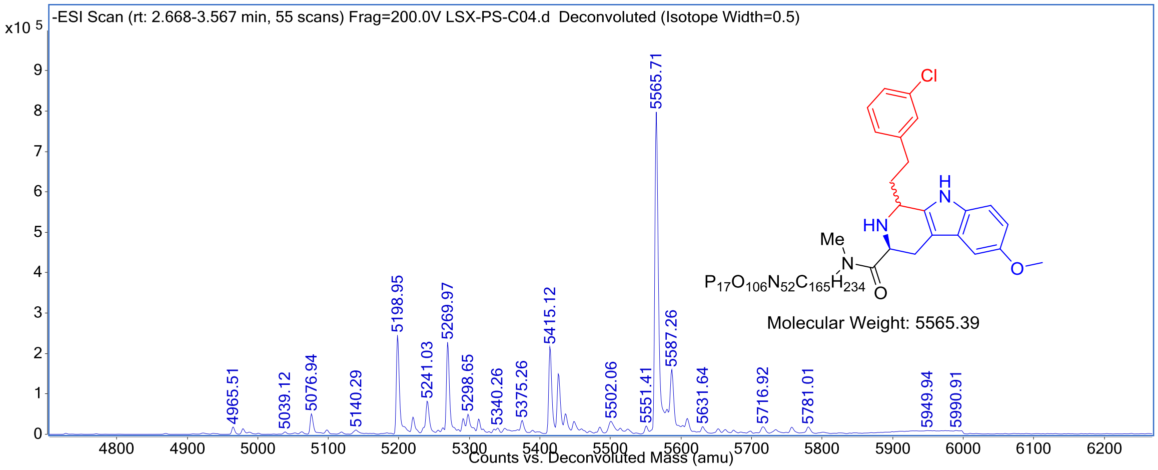


Fig. S83. Deconvoluted mass of **4ey**

Figure S84, Mass Spectrum of 4ez, related to Figure 4.

Percent conversion: 98%
Exact mass: 5556.32
Observed: 5556.65

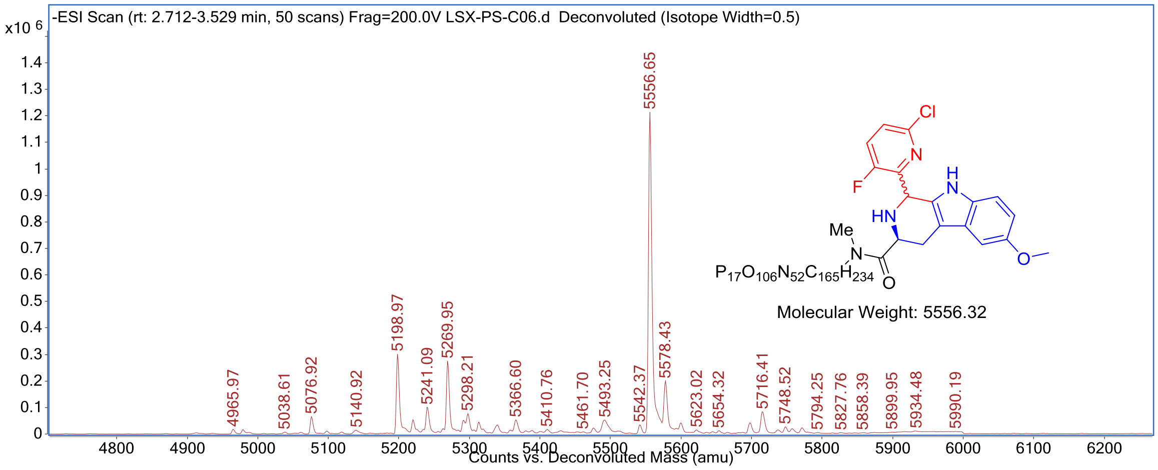


Fig. S84. Deconvoluted mass of **4ez**

Figure S85, Mass Spectrum of 4fa, related to Figure 4.

Percent conversion: 55.85%
Exact mass: 5626.79
Observed: 5626.67

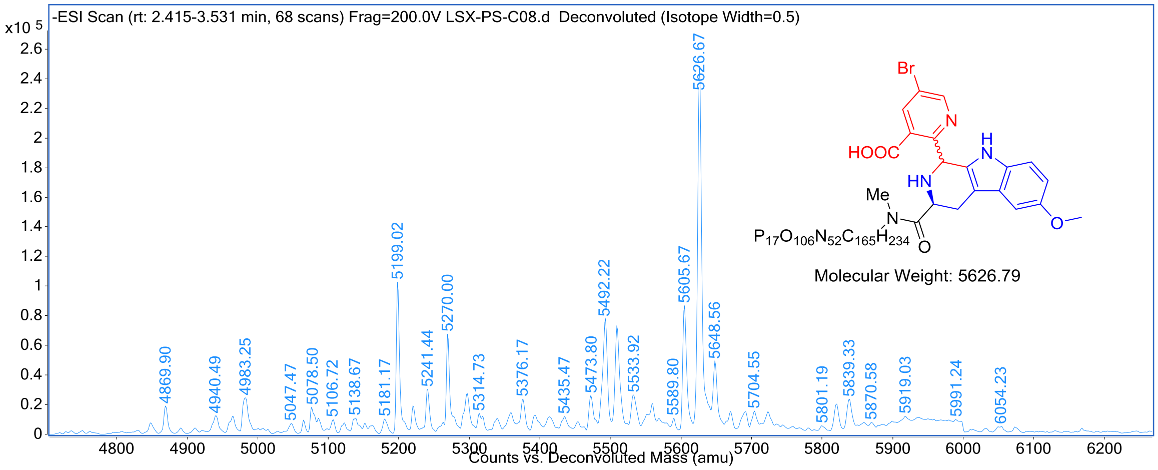


Fig. S85. Deconvoluted mass of **4fa**

Figure S86, Mass Spectrum of 4fb, related to Figure 4.

Percent conversion: 0
Exact mass: 5739.04
Observed: NO

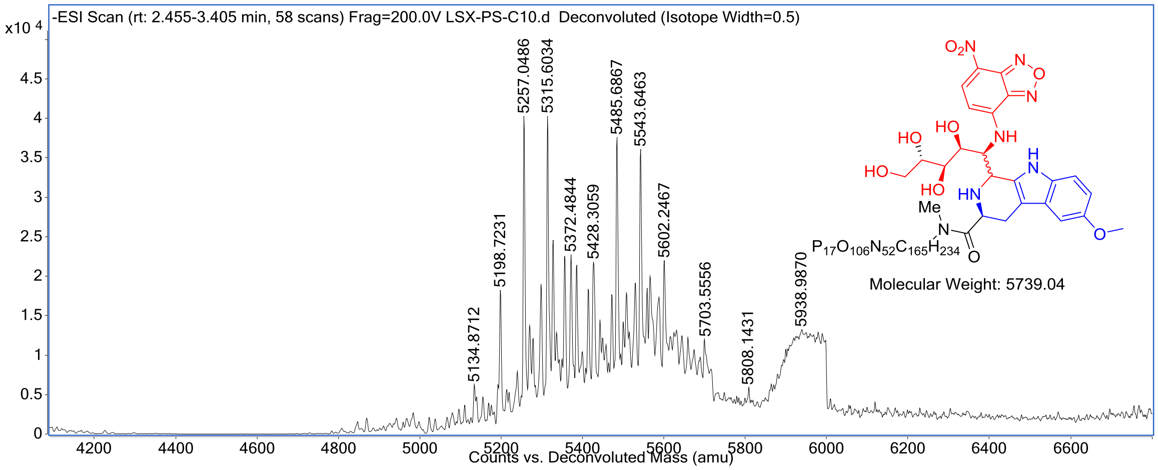


Fig. S86. Deconvoluted mass of **4fb**

Figure S87, Mass Spectrum of 4fc, related to Figure 4.

Percent conversion: 66.67%
Exact mass: 5562.35
Observed: 5562.5587

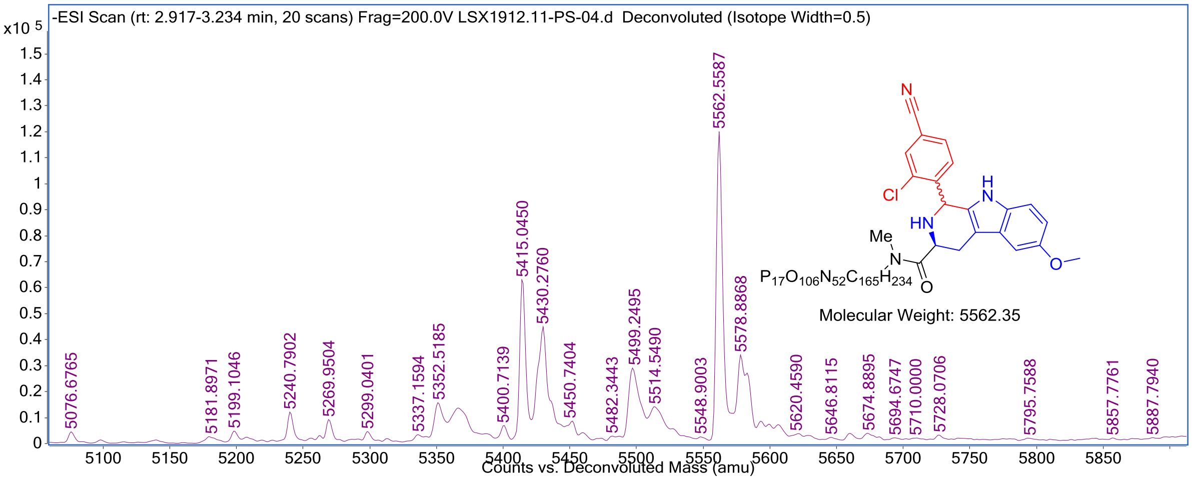


Fig. S87. Deconvoluted mass of **4fc**

Figure S88, Mass Spectrum of 4fd, related to Figure 4.

Percent conversion: 81.43%
Exact mass: 5530.95
Observed: 5531.2255

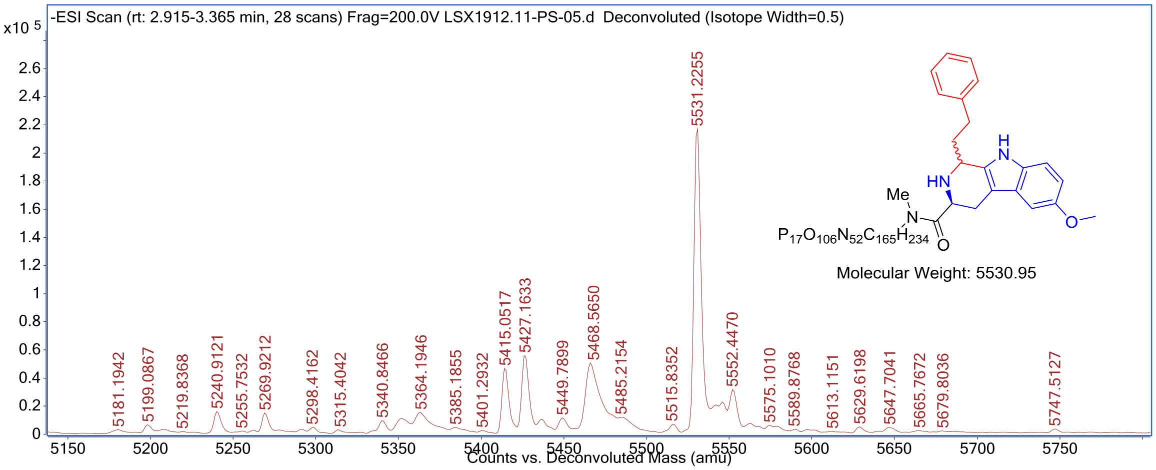


Fig. S88. Deconvoluted mass of **4fd**

Figure S89, Mass Spectrum of 4fe, related to Figure 4.

Percent conversion: 51.43%
Exact mass: 5626.79
Observed: 5627.0993

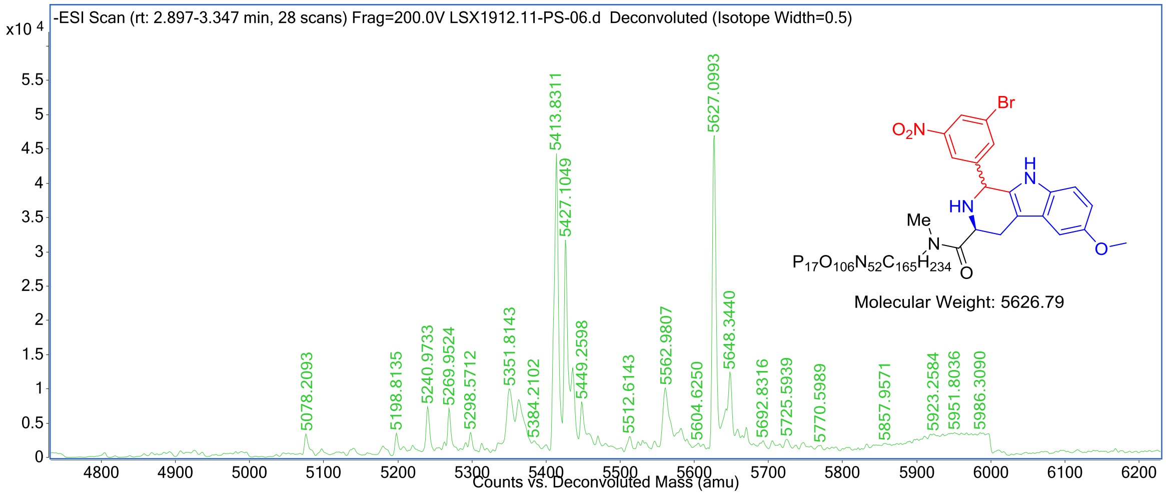


Fig. S89. Deconvoluted mass of **4fe**

Figure S90, Mass Spectrum of 4ff, related to Figure 4.

Percent conversion: 45.74%
Exact mass: 5547.89
Observed: 5546.4868

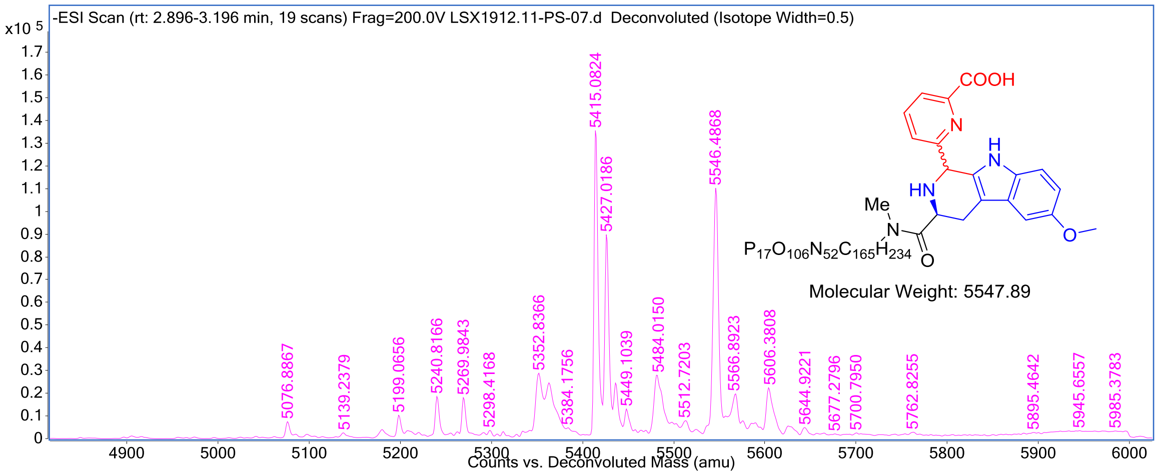


Fig. S90. Deconvoluted mass of **4ff**

Figure S91, Mass Spectrum of 4fg, related to Figure 4.

Percent conversion: 58.70%
Exact mass: 5615.89
Observed: 5615.2219

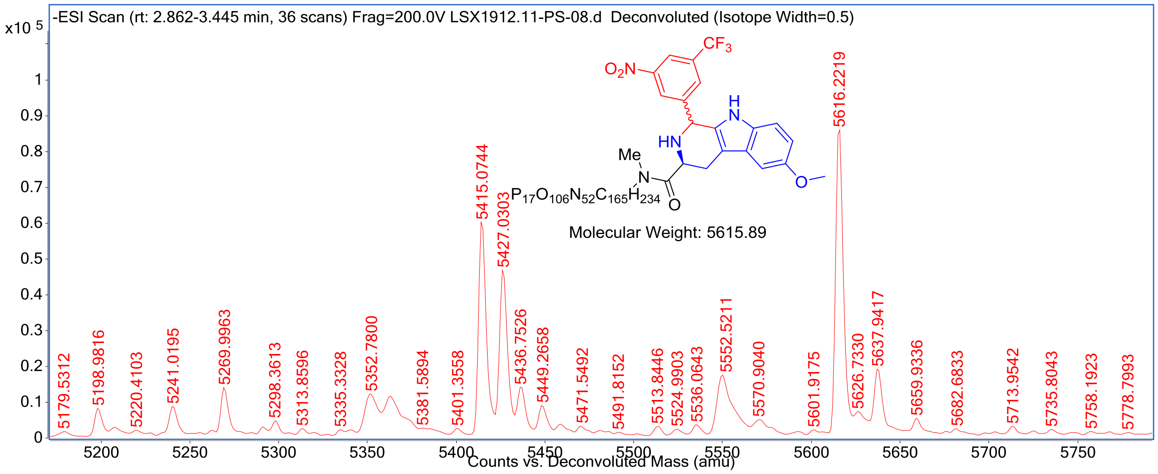


Fig. S91. Deconvoluted mass of **4fg**

Figure S92, Mass Spectrum of 4fh, related to Figure 4.

Percent conversion: 37.04%
Exact mass: 5531.70
Observed: 5530.1939

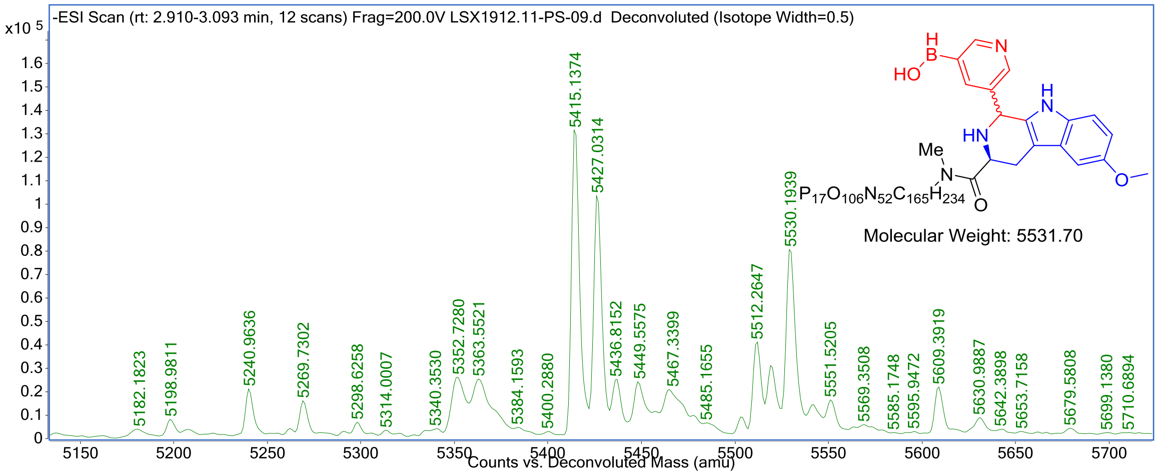
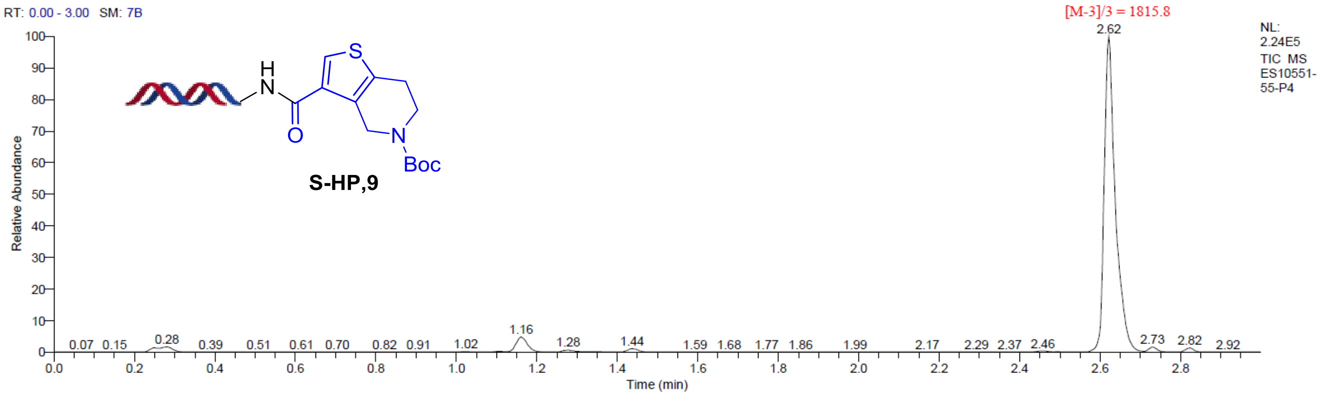


Fig. S92. Deconvoluted mass of **4fh**

Figure S93, LC Trace and Mass of 9, related to Figure 7.

Following General Procedure 5
Percent conversion: 95.29%
Exact mass: 5449.83
Triply charged mass [M-3]/3, calculated: 1815.61; observed:1815.8



RT	Peak Height	Peak Width	Peak Area	Area %
1.16	10889.55	0.16	21645.79	4.71
2.62	245583.85	0.21	437934.39	95.29

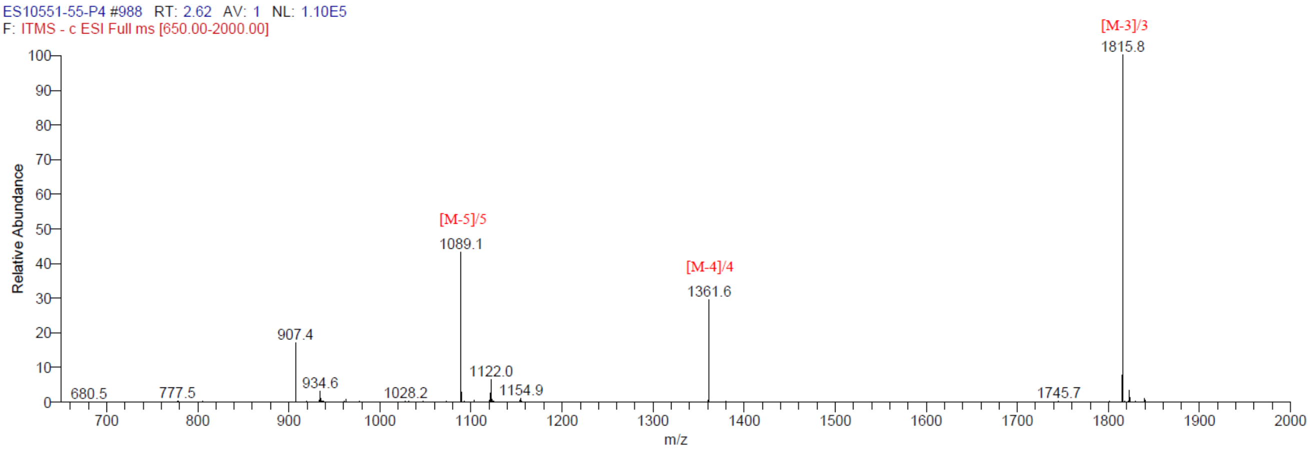


Fig. S93. LC trace and mass of **9**

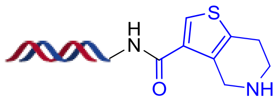
Figure S94, LC Trace and Mass of 10, related to Figure 7.

Following **General Procedure 5**

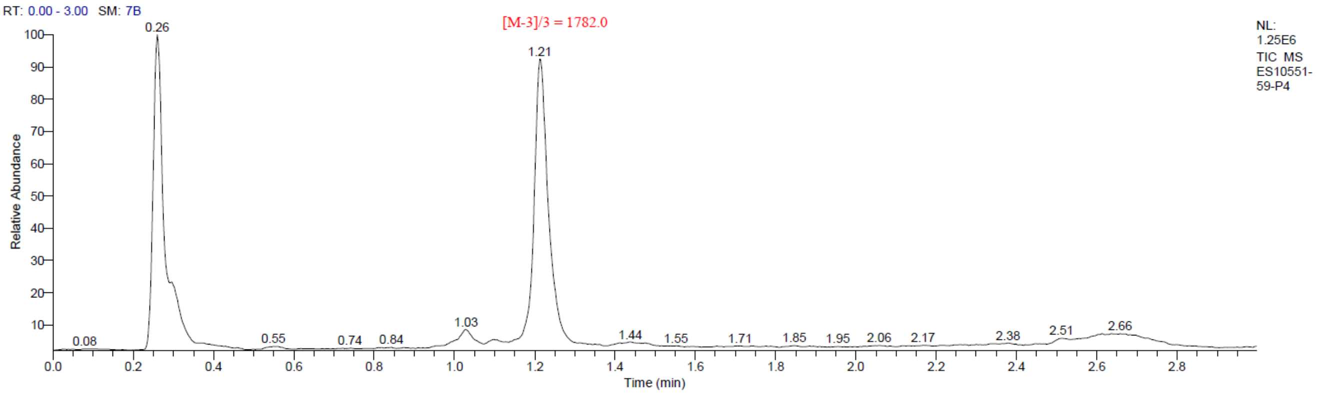
Percent conversion: 92.45%

Exact mass: 5349.83

Triply charged mass [M-3]/3, calculated: 1782.28; observed:1782.0



S-HP,10



RT	Peak Height	Peak Width	Peak Area	Area %
1.03	75201.39	0.18	243060.60	7.55
1.21	1172769.32	0.28	2975952.47	92.45

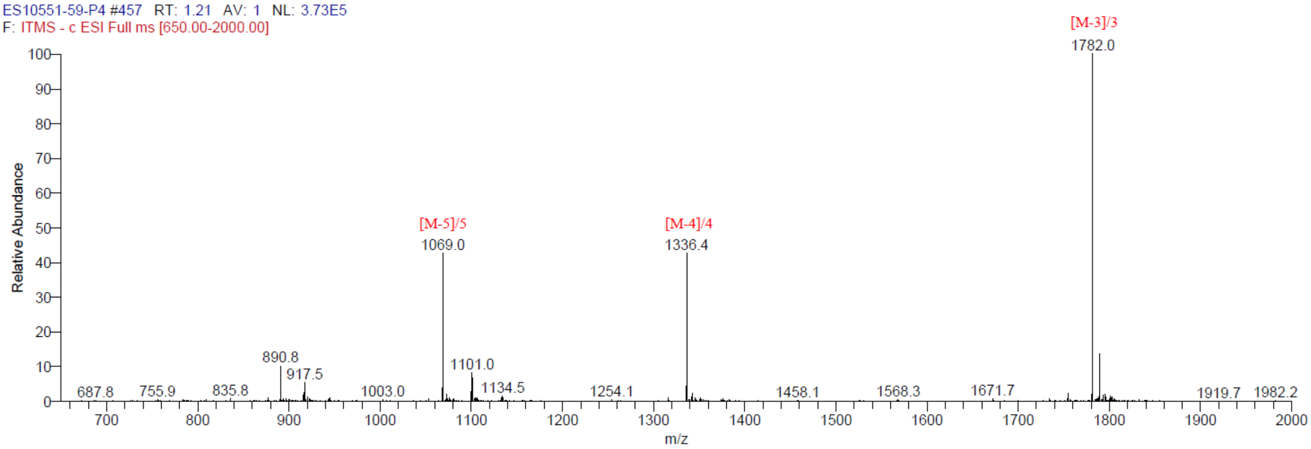


Fig. S94. LC trace and mass of **10**

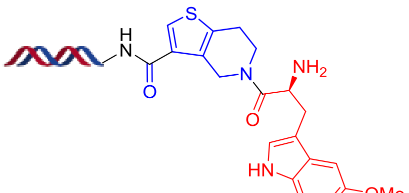
Figure S95, LC Trace and Mass of 11, related to Figure 7.

Following **General Procedure 1**

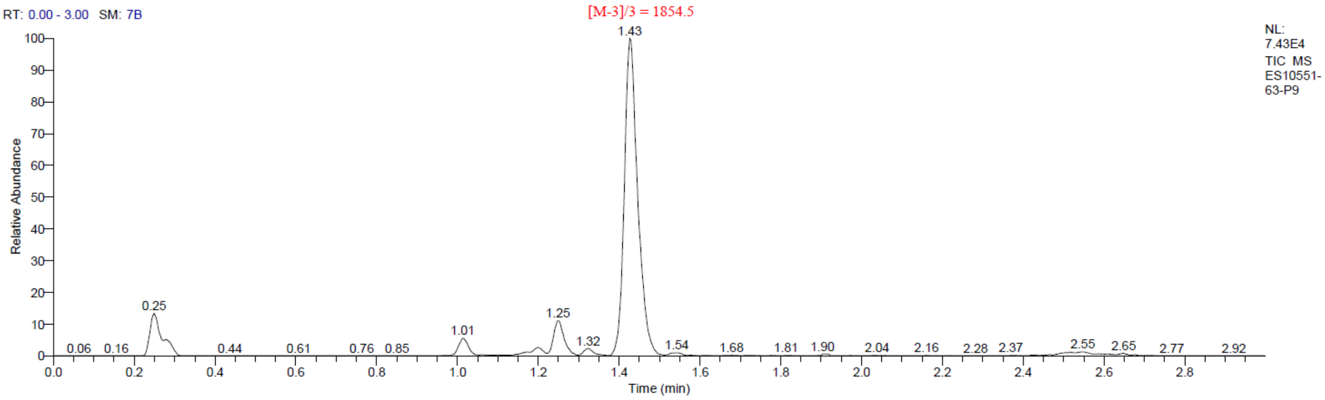
Percent conversion: 88.95%

Exact mass: 5566.32

Triply charged mass [M-3]/3, calculated: 1854.44; observed:1854.5



S-HP,11



RT	Peak Height	Peak Width	Peak Area	Area %
1.25	8969.07	0.22	21303.06	11.05
1.43	80179.27	0.14	171406.83	88.95

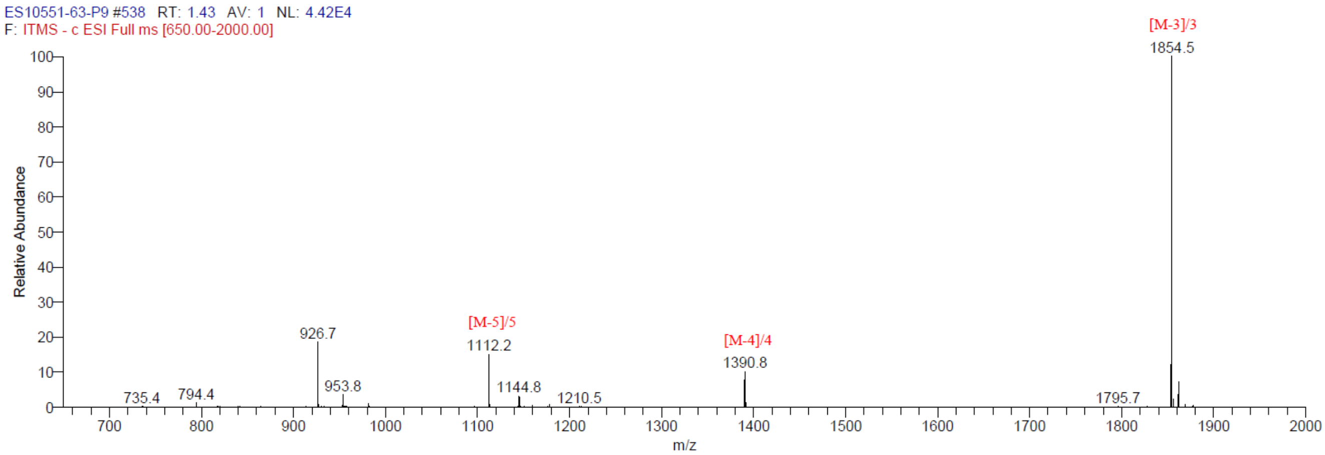


Fig. S95. LC trace and mass of **11**

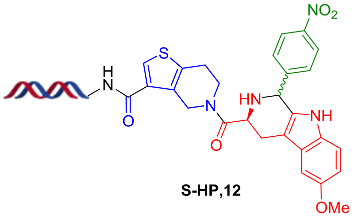
Figure S96, LC Trace and Mass of 12, related to Figure 7.

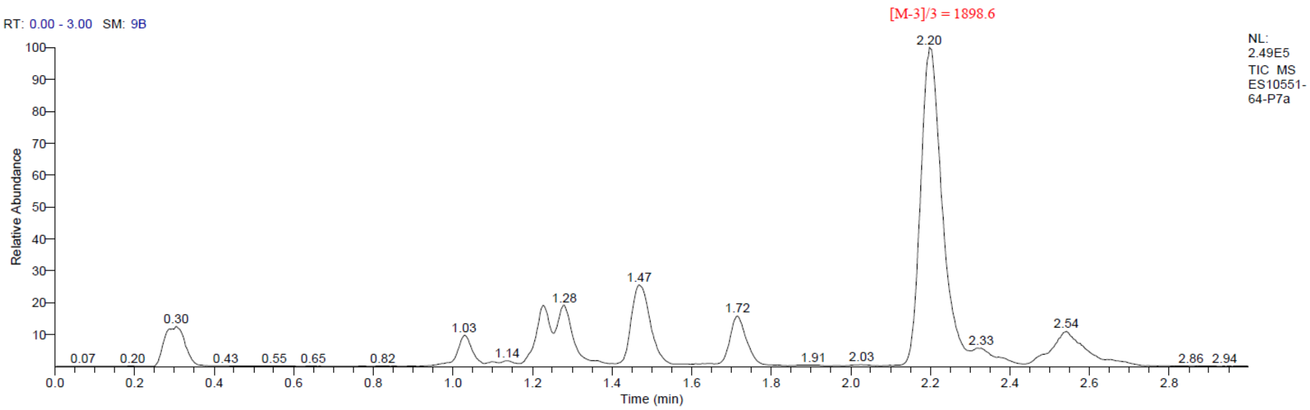
Following **General Procedure 2**

Percent conversion: 55.32%

Exact mass: 5699.44

Triply charged mass [M-3]/3, calculated: 1898.81 observed:1898.6





RT	Peak Height	Peak Width	Peak Area	Area %
1.23	53660.90	0.33	284509.43	15.68
1.47	65770.35	0.22	223109.96	12.29
1.72	42742.23	0.32	124866.96	6.88
2.20	264922.45	0.37	1004065.09	55.32
2.54	28023.39	0.48	178383.38	9.83

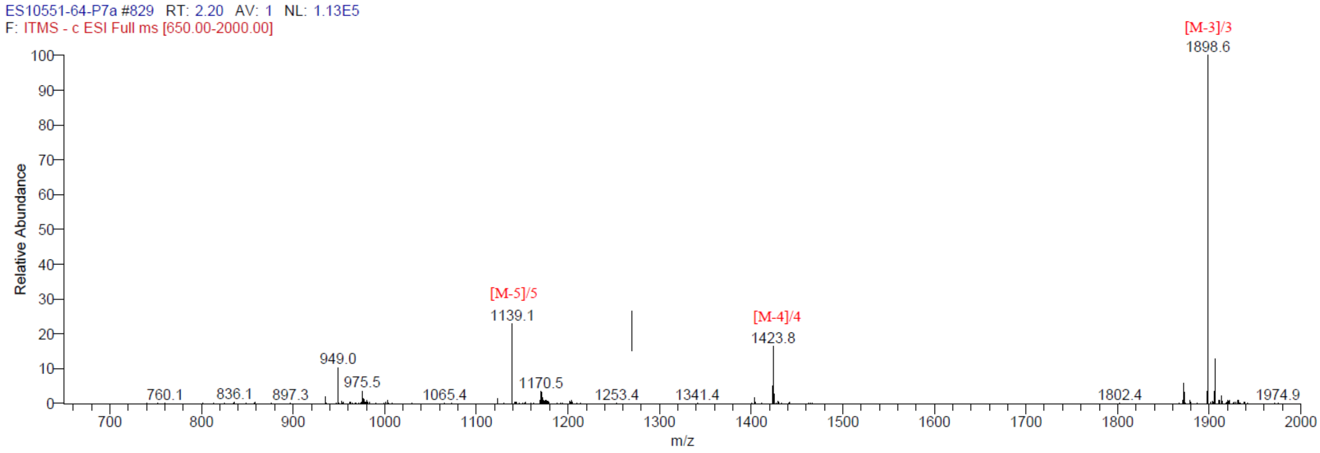


Fig. S96. LC trace and mass of **12**

Figure S97, Mass Spectrum of 13, related to Figure 7.

Percent conversion: 95%
Exact mass: 5112.37
Observed: 5112.8975

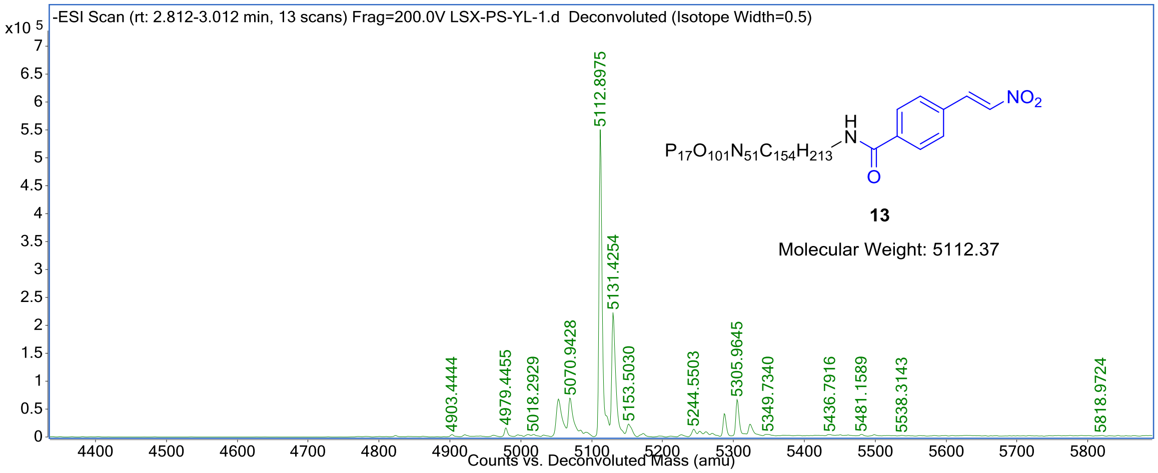
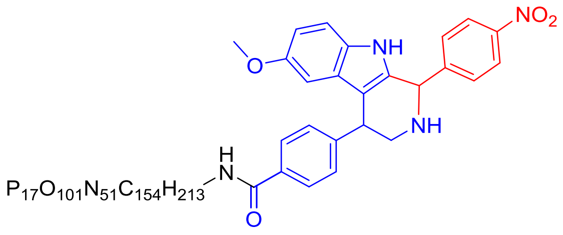


Fig. S97. Deconvoluted mass of **13**

Figure S98, Mass Spectrum of 8a, related to Figure 7.

Percent conversion: 50%
Exact mass: 5362.67
Observed: 5362.87



Molecular Weight: 5362.67

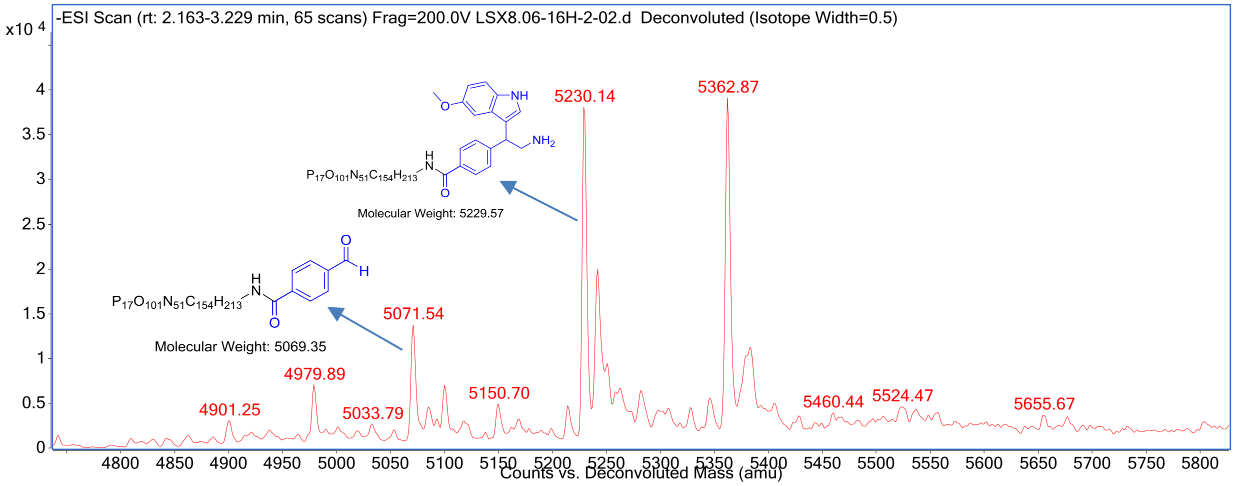
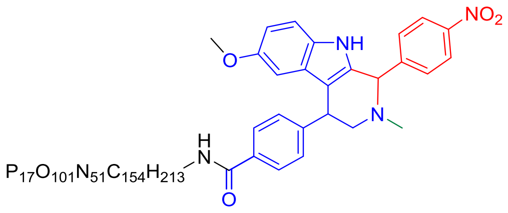


Fig. S98. Deconvoluted mass of **8a**

Figure S99, Mass Spectrum of 14a, related to Figure 7.

Percent conversion: 95%
Exact mass: 5376.70
Observed: 5377.3636



Molecular Weight: 5376.70

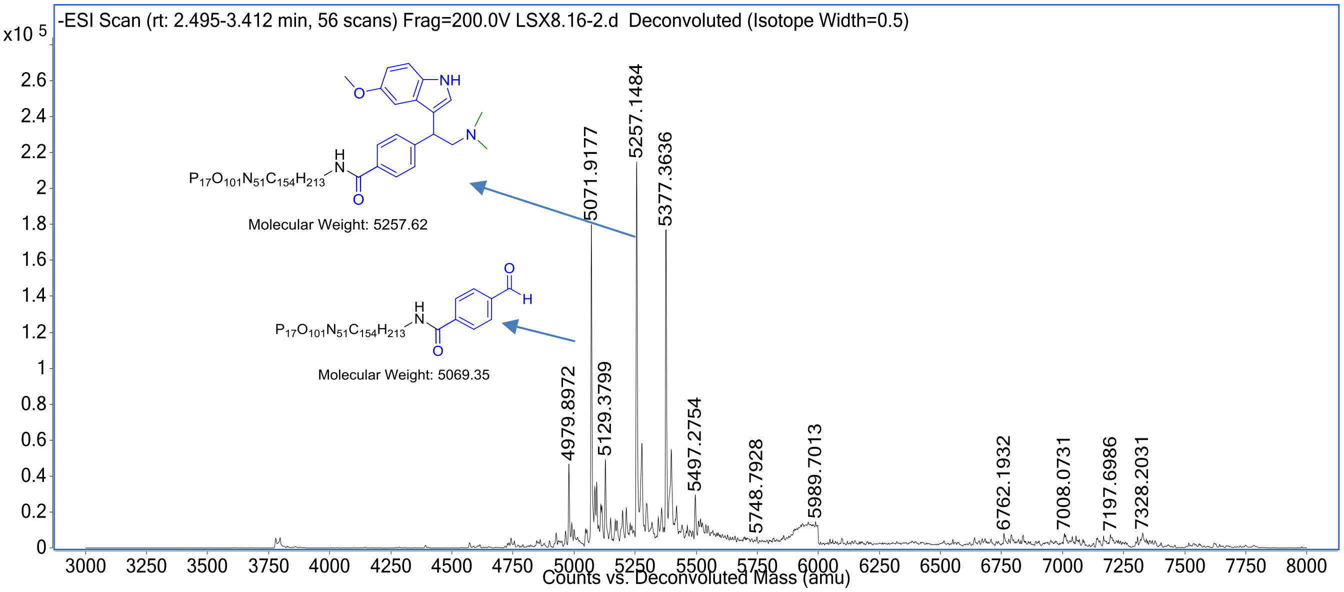
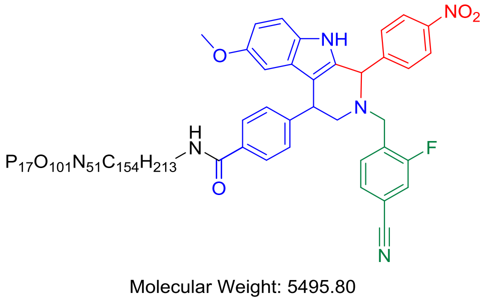


Fig. S99. Deconvoluted mass of **14a**

Figure S100, Mass Spectrum of 14b, related to Figure 7.

Percent conversion: 95%
Exact mass: 5495.80
Observed: 5496.4340


NO_2
NH
O
N
F
H
N
O
$P_{17}O_{101}N_{51}C_{154}H_{213}$
N
Molecular Weight: 5495.80

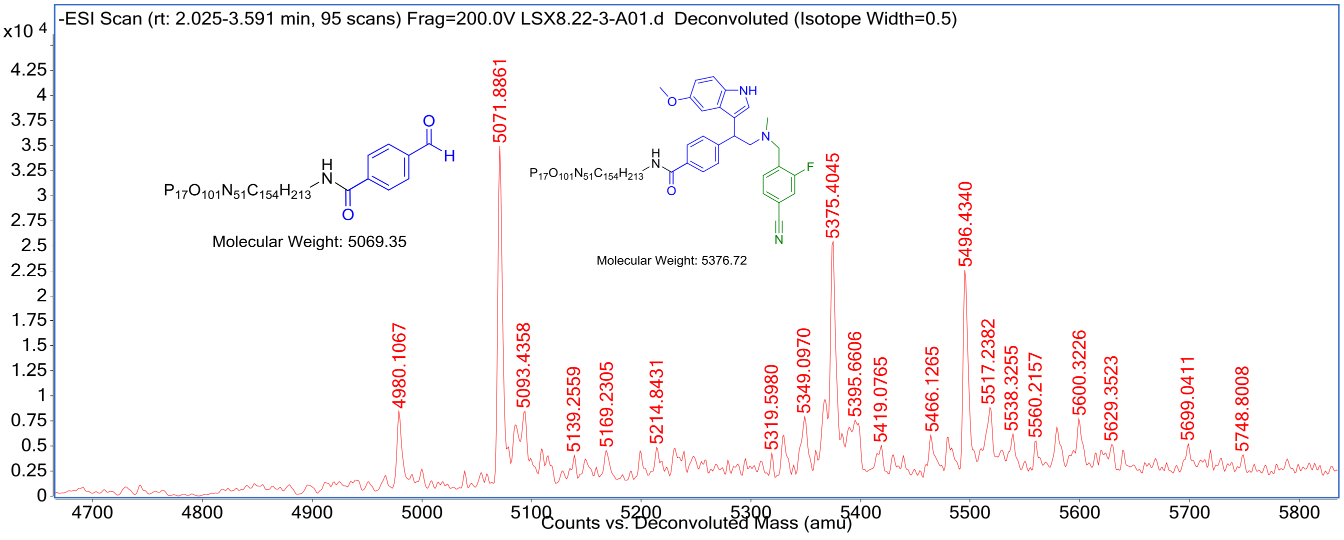
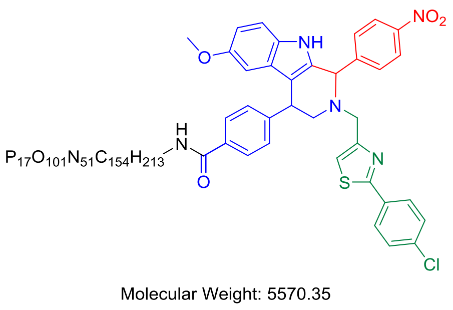


Fig. S100. Deconvoluted mass of **14b**

Figure S101, Mass Spectrum of 14c, related to Figure 7.

Percent conversion: 68.97%
Exact mass: 5570.35
Observed: 5570.8703



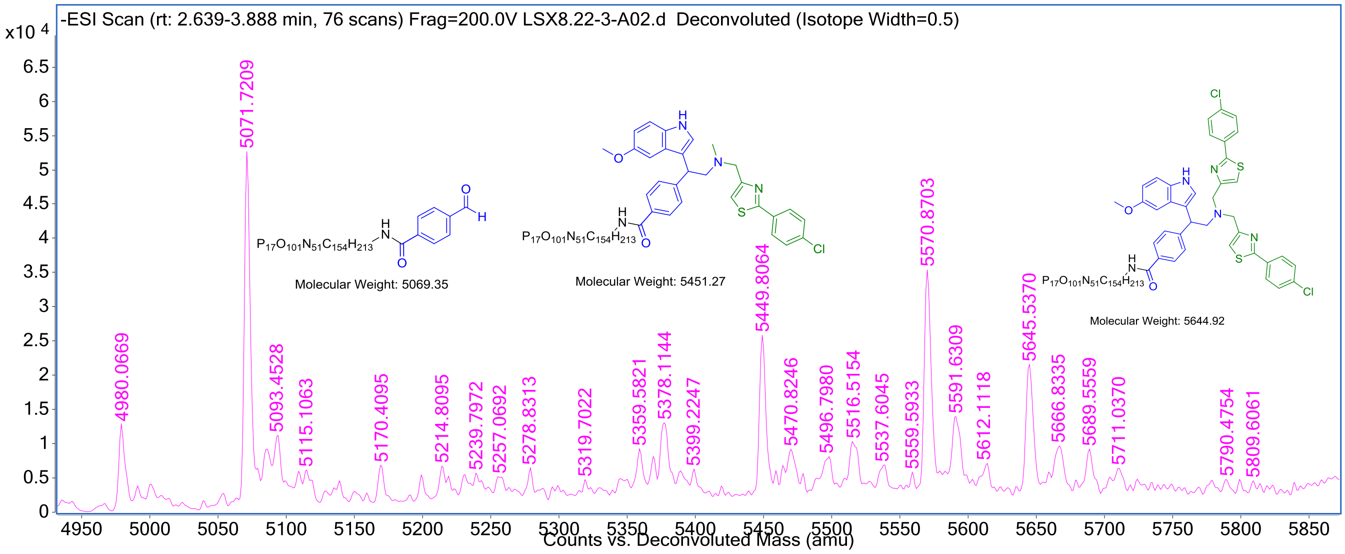
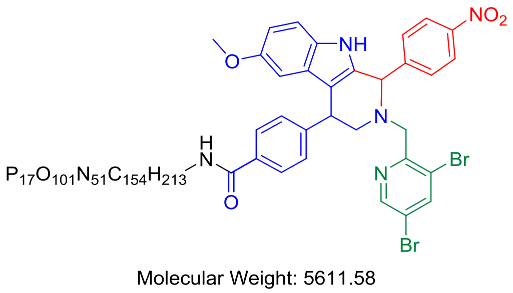


Fig. S101. Deconvoluted mass of **14c**

Figure S102, Mass Spectrum of 14d, related to Figure 7.

Percent conversion: 60.98%
Exact mass: 5611.58
Observed: 5612.3937



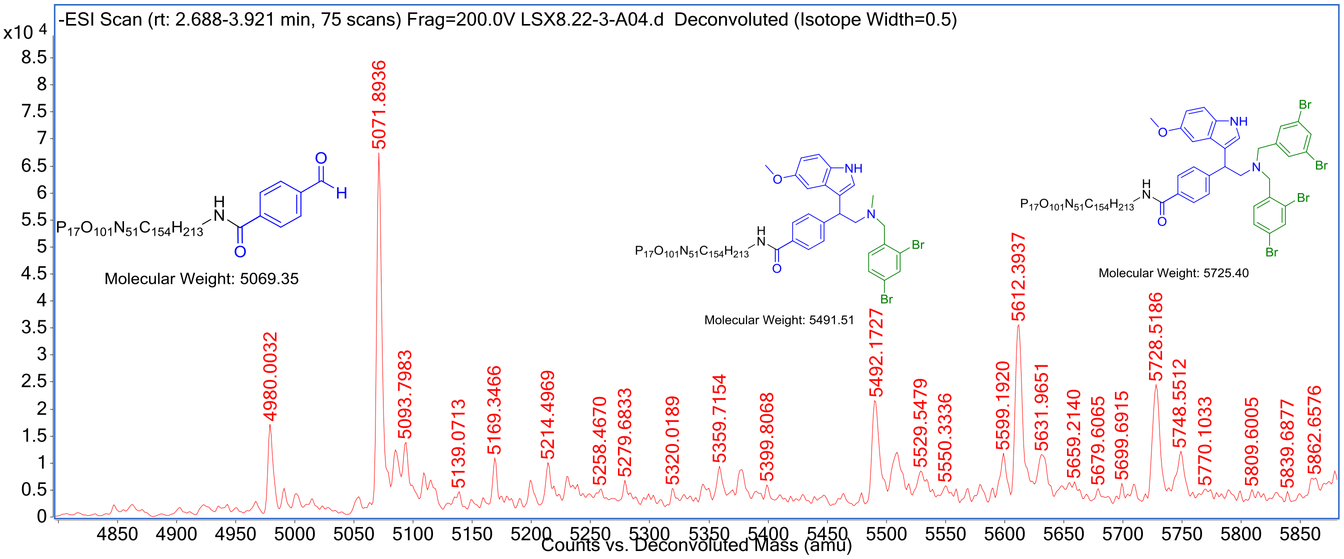
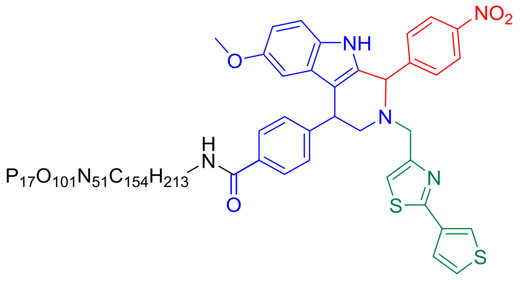


Fig. S102. Deconvoluted mass of **14d**

Figure S103, Mass Spectrum of 14e, related to Figure 7.

Percent conversion: 56.67%
Exact mass: 5541.93
Observed: 5542.3905



Molecular Weight: 5541.93

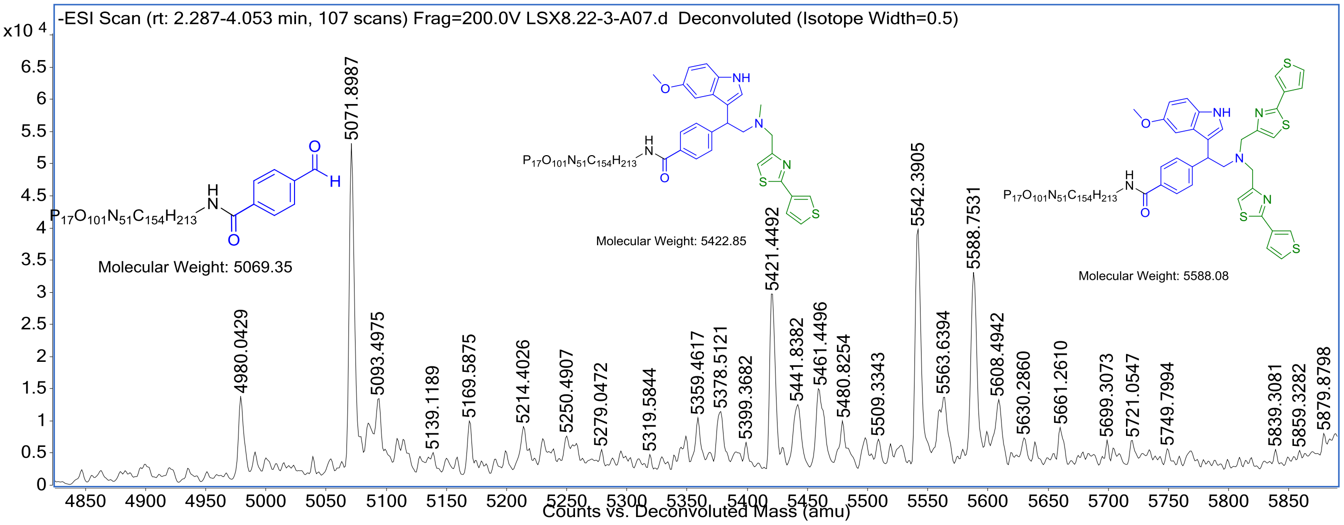
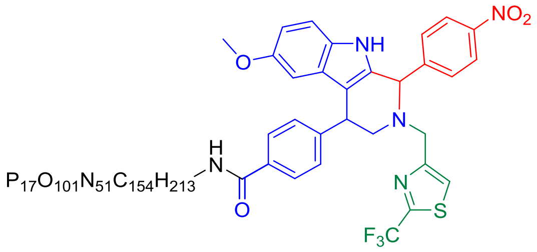


Fig. S103. Deconvoluted mass of **14e**

Figure S104, Mass Spectrum of 14f, related to Figure 7.

Percent conversion: 58.33%
Exact mass: 5527.81
Observed: 5528.3546



Molecular Weight: 5527.81

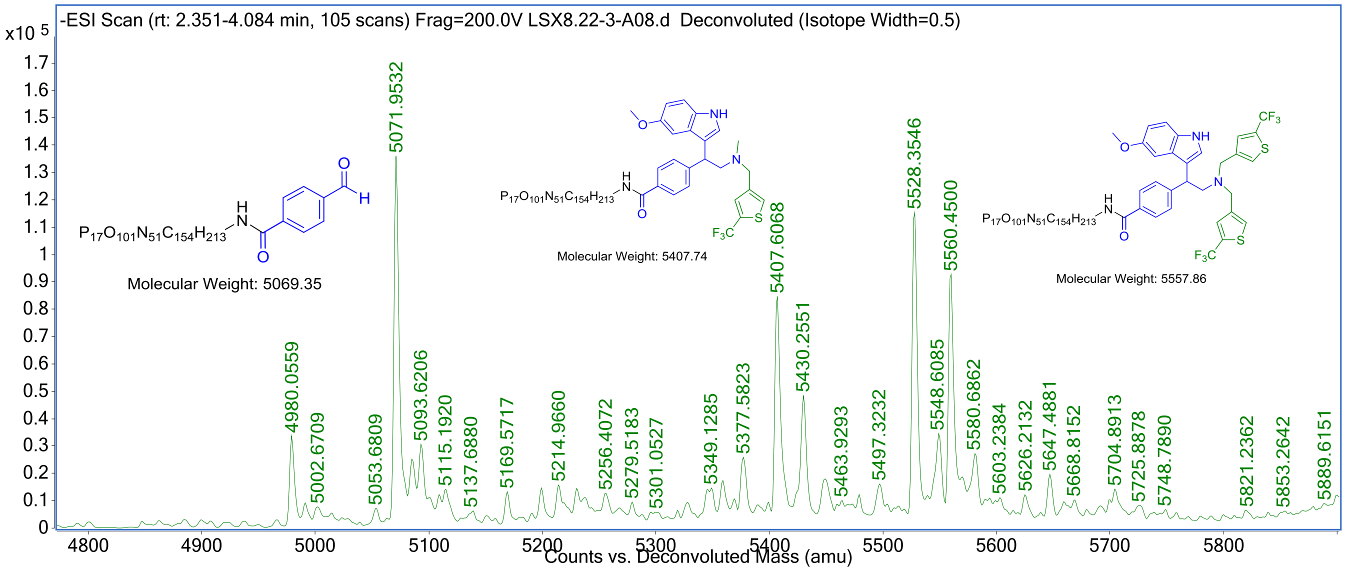
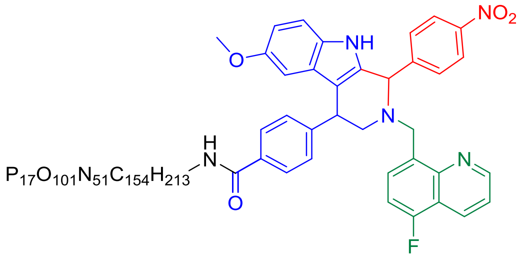


Fig. S104. Deconvoluted mass of **14f**

Figure S105, Mass Spectrum of 14g, related to Figure 7.

Percent conversion: 54.55%
Exact mass: 5521.84
Observed: 5522.4649



Molecular Weight: 5521.84

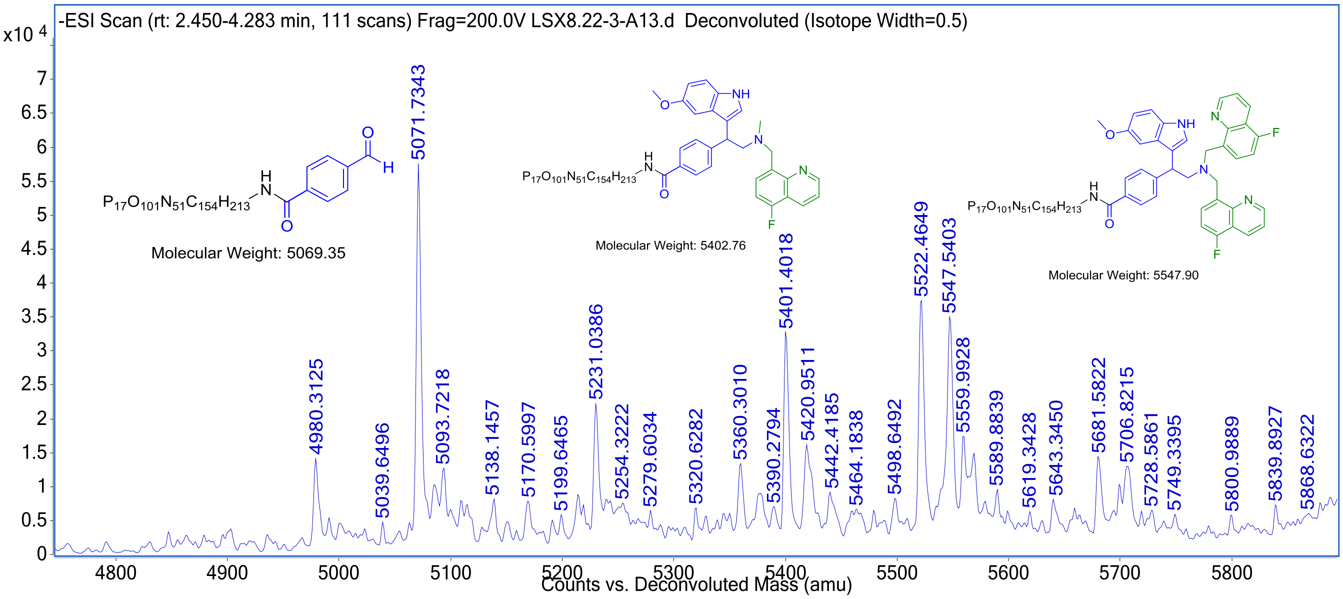
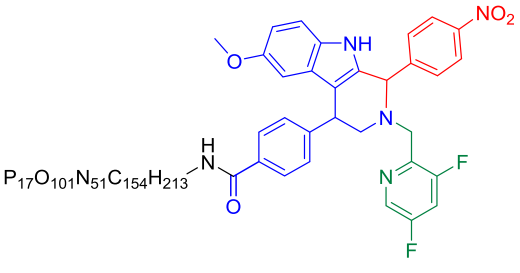


Fig. S105. Deconvoluted mass of **14g**

Figure S106, Mass Spectrum of 14h, related to Figure 7.

Percent conversion: 59.32%
Exact mass: 5489.77
Observed: 5489.4860



Molecular Weight: 5489.77

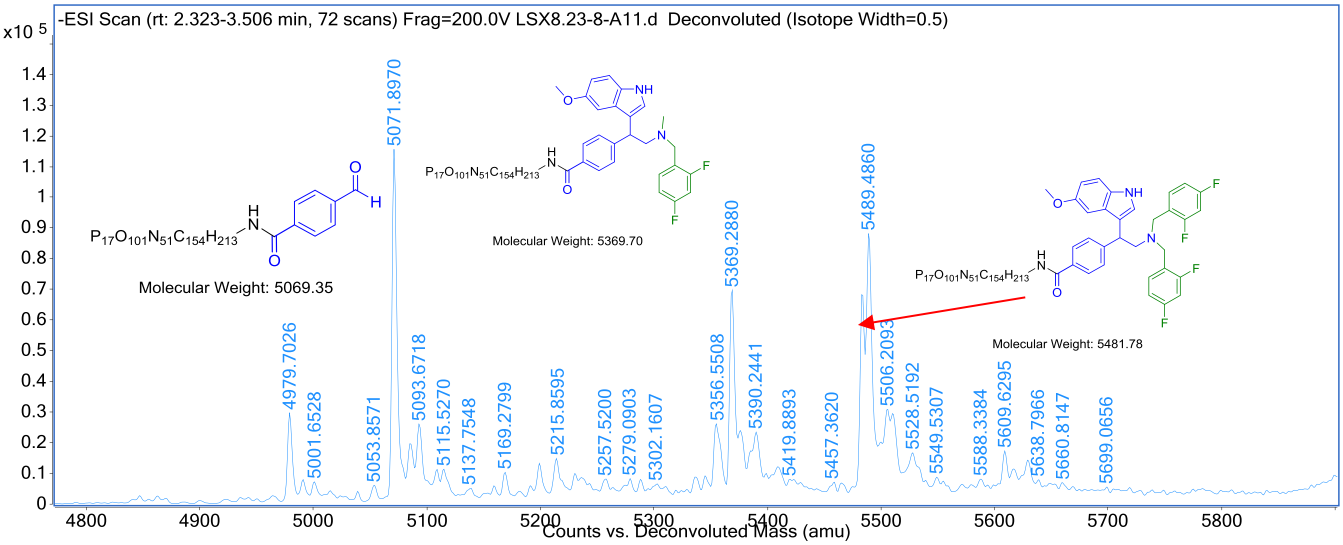
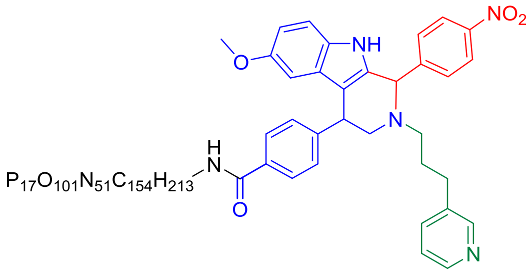


Fig. S106. Deconvoluted mass of **14h**

Figure S107, Mass Spectrum of 14i, related to Figure 7.

Percent conversion: 36.97%
Exact mass: 5481.84
Observed: 5482.4366



Molecular Weight: 5481.84

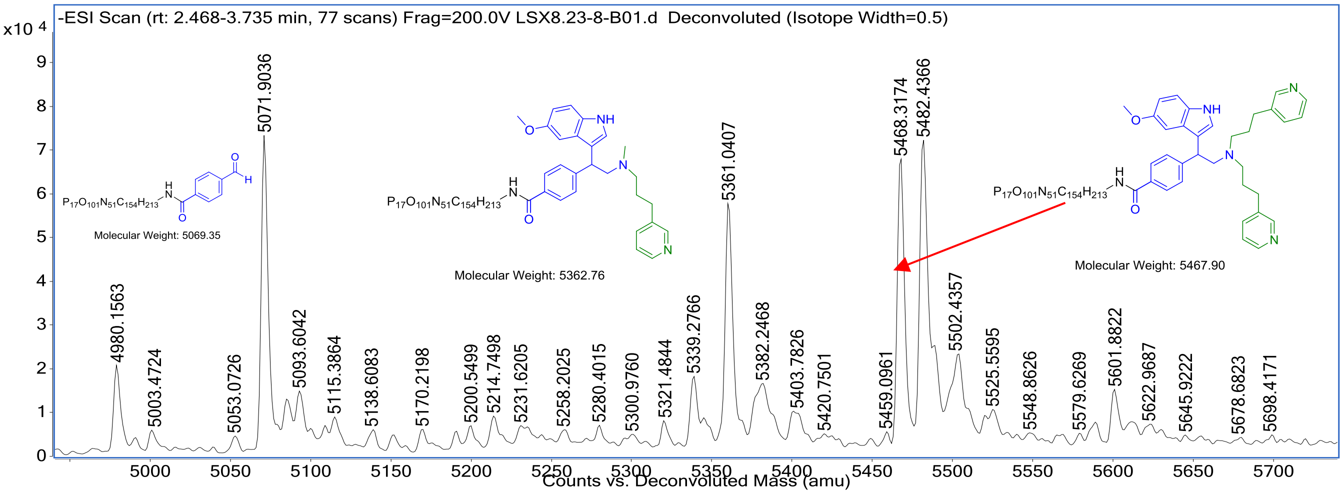
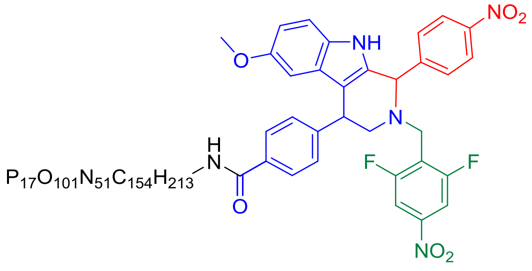


Fig. S107. Deconvoluted mass of **14i**

Figure S108, Mass Spectrum of 14j, related to Figure 7.

Percent conversion: 44.16%
Exact mass: 5533.78
Observed: 5534.1750



Molecular Weight: 5533.78

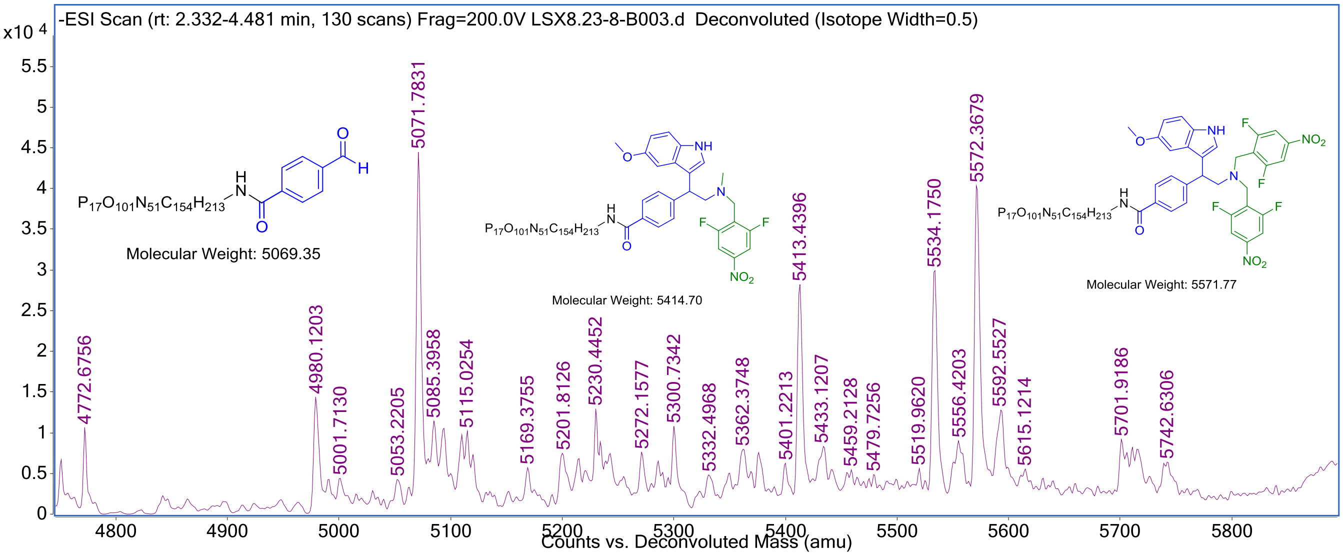
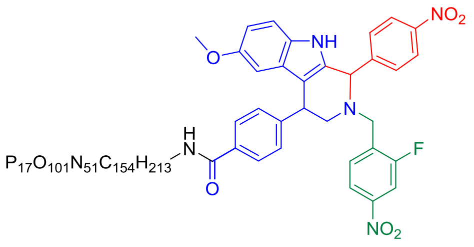


Fig. S108. Deconvoluted mass of **14j**

Figure S109, Mass Spectrum of 14k, related to Figure 7.

Percent conversion: 20.17%
Exact mass: 5515.79
Observed: 5516.4274



Molecular Weight: 5515.79

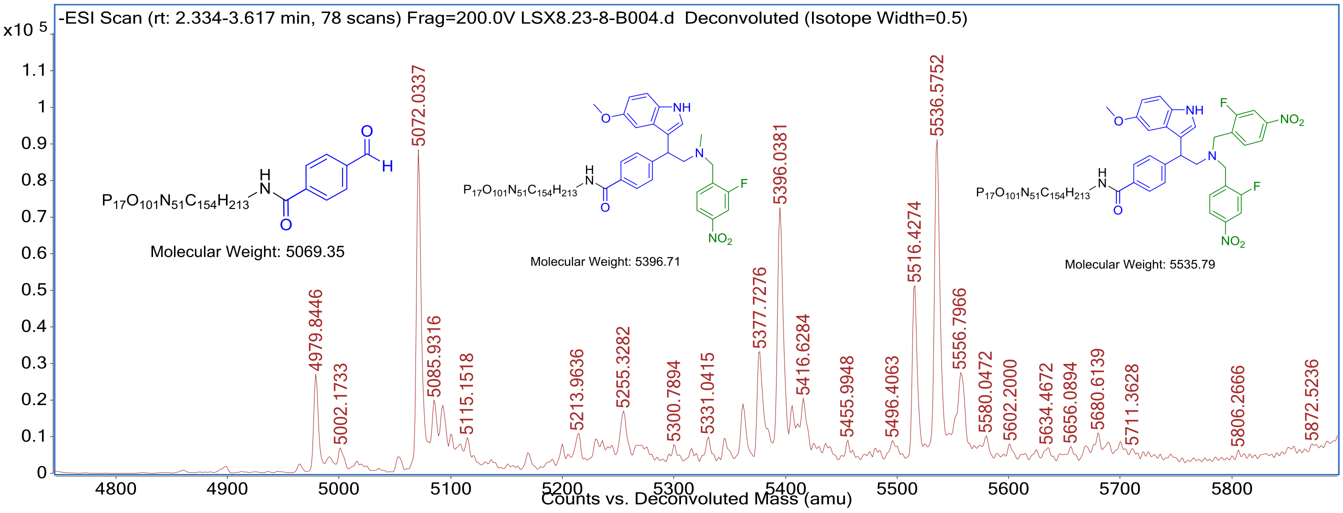
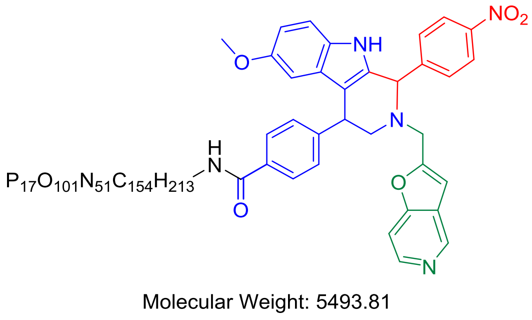


Fig. S109. Deconvoluted mass of **14k**

Figure S110, Mass Spectrum of 14l, related to Figure 7.

Percent conversion: 57.38%
Exact mass: 5493.81
Observed: 5492.5954


NO2
NH
O
N
O
N
H
N
$P_{17}O_{101}N_{51}C_{154}H_{213}$
O
Molecular Weight: 5493.81

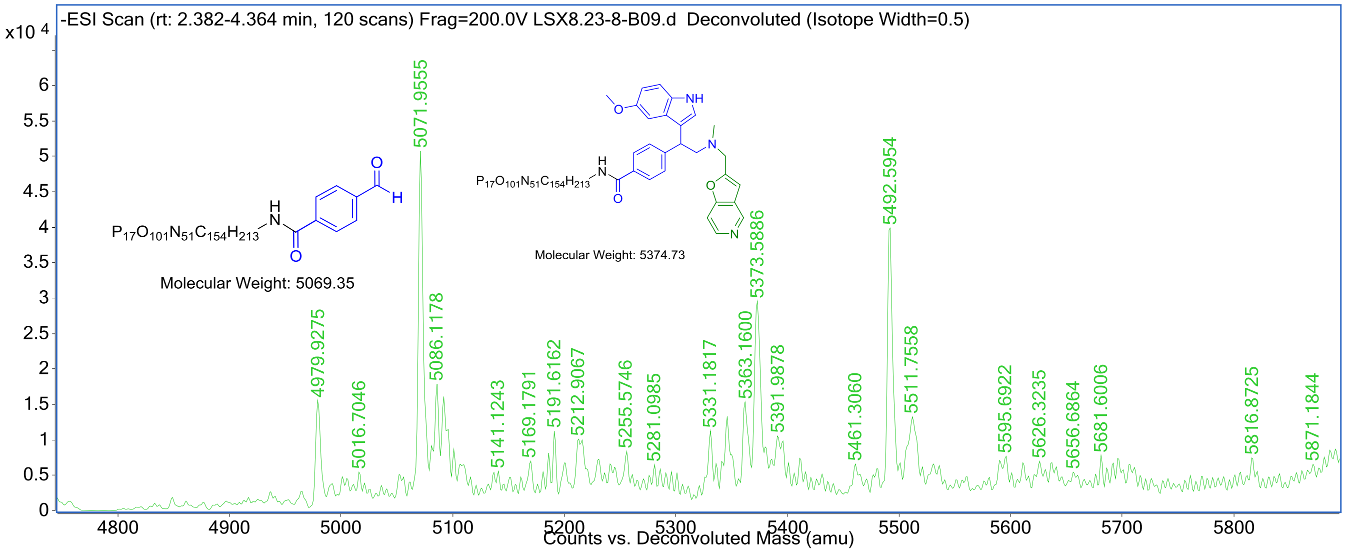


Fig. S110. Deconvoluted mass of **14l**

Tables

Table S5. DNN and KNN comparison, related to Figure 4.

In order to compare the performance of DNN and conventional similarity-based ML methods, a KNN model was implemented based on the same training dataset, and its optimal parameter K=9 (searching range 1-50) was determined by 5-fold cross validation as the same way as DNN. As summarized in the following table, both precision and recall of DNN are higher than the values of KNN on internal test dataset. (ECFP4 as fingerprint)

model	DNN		KNN	
metrics	precision	recall	precision	recall

Internal test dataset	0.81	0.37	0.6	0.2

Table S6. ECFP4 and MACCS comparison, related to Figure 4.

ECFP4 and MACCS keys were separately taken as input of DNN and trained with the same procedures. The performance of two fingerprints on internal test dataset is summarized as following, where both precision and recall of model trained with MACCS are lower than that with ECFP4 on internal test dataset.

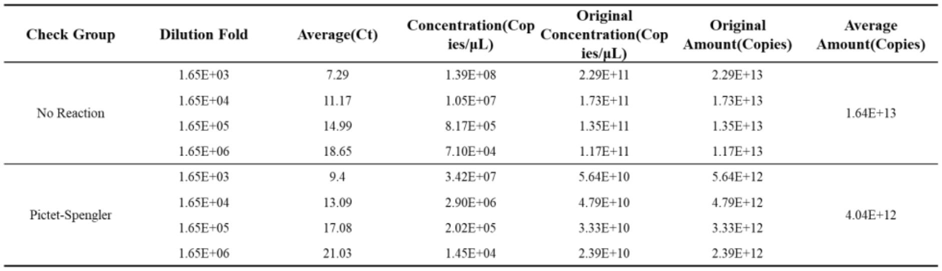
fingerprint	ECFP4		MACCS	
metrics	precision	recall	precision	recall
Internal test dataset	0.81	0.37	0.78	0.23

Table S7. The number of clusters: the number of structures with different threshold of the train dataset, internal test dataset, top300 candidates and external test dataset, related to Figure 4.

Table S7. The number of clusters: the number of structures with different threshold

Threshold	0.4	0.5	0.6	0.7
Train	283:1325	520:1324	900:1324	1178:1324
	(1:4.68)	(1:2.55)	(1:1.47)	(1:1.12)
Internal test	135:331	220:331	300:331	324:331
	(1:2.45)	(1:1.50)	(1:1.10)	(1:1.02)
Top300	58:300	94:300	159:300	259:300
	(1:5.17)	(1:3.19)	(1:1.89)	(1:1.16)
External test	13:34	18:34	25:34	33:34
	(1:2.62)	(1:1.89)	(1:1.36)	(1:1.03)

Table S8 Quantitation Result of the concentration check group, related to Figure 6.



Check Group	Dilution Fold	Average(Ct)	Concentration(Copies/μL)	Original Concentration(Copies/μL)	Original Amount(Copies)	Average Amount(Copies)
No Reaction	1.65E+03	7.29	1.39E+08	2.29E+11	2.29E+13	1.64E+13
	1.65E+04	11.17	1.05E+07	1.73E+11	1.73E+13	
	1.65E+05	14.99	8.17E+05	1.35E+11	1.35E+13	
	1.65E+06	18.65	7.10E+04	1.17E+11	1.17E+13	
Pictet-Spengler	1.65E+03	9.4	3.42E+07	5.64E+10	5.64E+12	4.04E+12
	1.65E+04	13.09	2.90E+06	4.79E+10	4.79E+12	
	1.65E+05	17.08	2.02E+05	3.33E+10	3.33E+12	
	1.65E+06	21.03	1.45E+04	2.39E+10	2.39E+12	

Transparent Methods

SI-1 Machine learning model

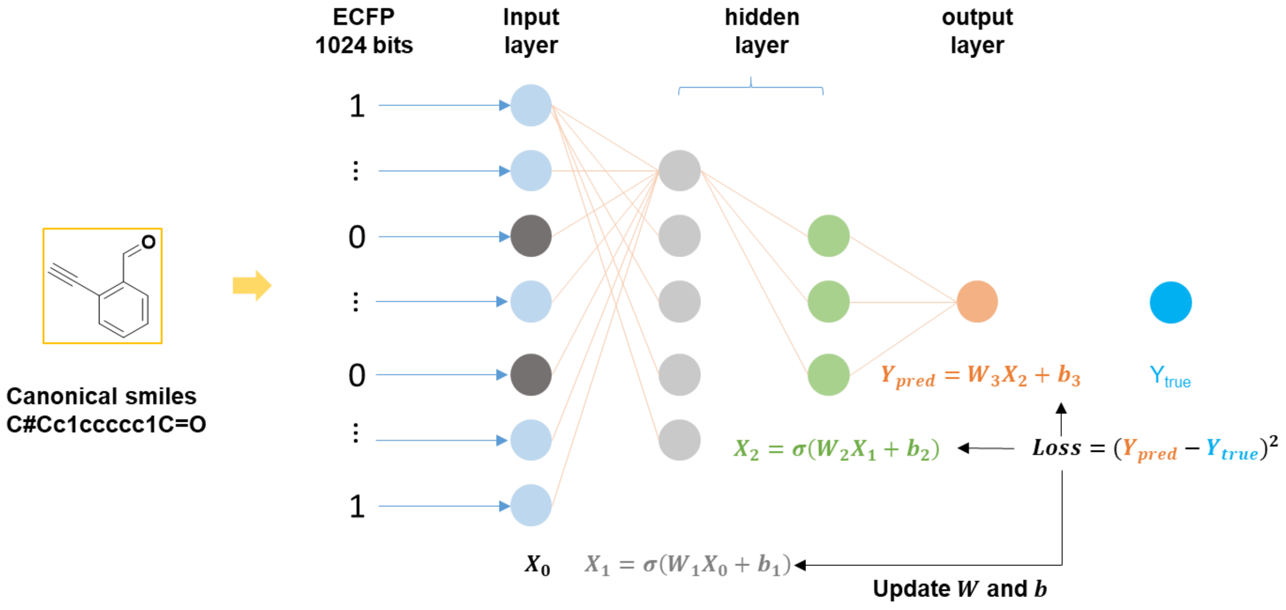
When training the model, molecules initially represented by ECFP4 fingerprints were fed into multiple hidden layers defined below:

$$X_L = \sigma(W_L X_{L-1} + b_L)\ (L \geq 1)$$

where L represents the L-th layer of the model, X_L represents the L-th representation of the molecule. When L=1, X_0 is the input feature ECFP. W_L and b_L represent the weight matrix and bias for the L-th layer. σ is a function for nonlinear transformation. The cost function J(Θ) of the model is as following:

$$\mathrm{J}(\Theta) = \left(Y_{pred} - Y_{true}\right)^2$$

Cost function (here is the mean squared error, MSE) applied to a batch of all training data is minimized with respect to the model parameters Θ. Given predicted Y_{pred} and the true Y_{true}, Θ is updated according to the gradient of the prediction.

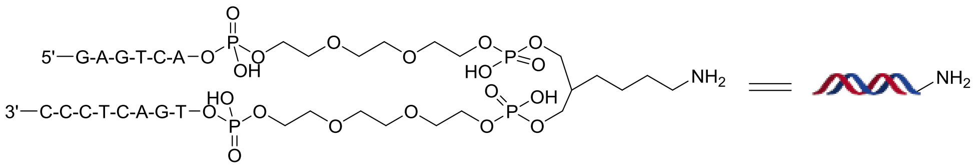


SI-2 General Experimental

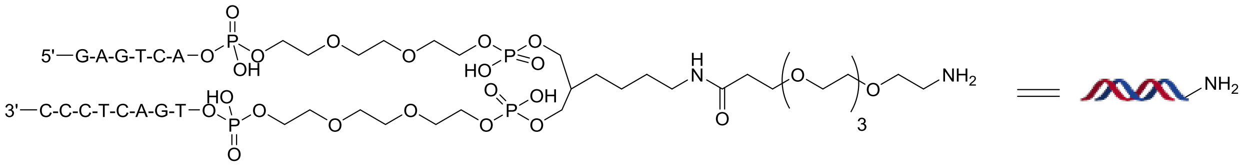
Dimethylsulfoxide (DMSO), 1-methyl-2-pyrrolidinone (NMP), and 2-Propanol (*i*-PrOH) and *N*,*N*-dimethylacetamide (DMAc), EtOH were purchased from Sigma-Aldrich. HATU (CAS: 148893-10-1), *N*,*N*-Diisopropylethylamine (DIPEA), NaCl, NaOAc were purchased from TCI. The $MgCl_2$ was purchased from *J&K*. The ddH_2O was obtained by passing the Milli-Q Direct. The buffer was purchased from

Vazyme. On-DNA reaction yields were determined by UV traces of LC/MSanalysis. The centrifugeinstruments including Allegra X-15R, eppendorf-5424R.

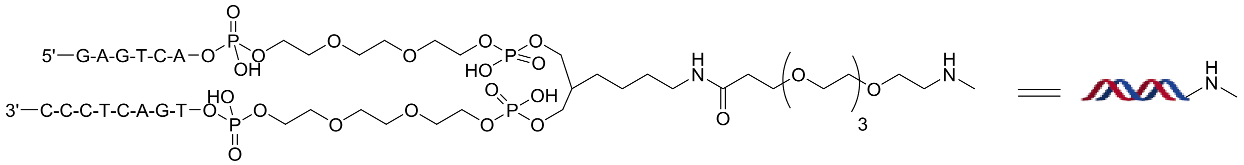
SI-3 HP, S-HP and Me-S-HP Material



HP (Exact Mass: 4937.23)



S-HP (Exact Mass: 5184.48)



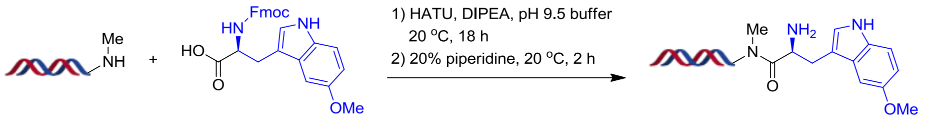
Me-S-HP (Exact Mass: 5198.48)

SI-4 General Procedure

SI-4-1 EtOH Precipitation for DNA substrate

To a DNA reaction mixture was added 10% (V/V) 5 M NaCl solution and 2.5−3 folds the volume of absolute ethanol. The colloidal solution was then allowed to stand at −80 °C for 2 h. The solutions were centrifuged at 4°C for 30 min at 4000 g; the supernatants were discarded. And the DNA pellet was dried at 30°C for 1 h in vacuo. General, ethanol precipitation was performed after each chemical reaction.

SI-4-2 General Procedure 1 for DNA-conjugated tryptamine



Me-S-HP **BB-1** **3a**

1) Acylation of Me-S-HP

To a 15 mL tube was added HATU (200 mM in DMSO, 500 μL, 100 eq.), DIPEA (200 mM in DMSO, 500μL, 100 eq.) and amino acid (200 mM in DMAc, 300 μL, 60 eq.). This solution was eddied, then centrifuged and stood at 20 °C for 15 min to make the activated ester.

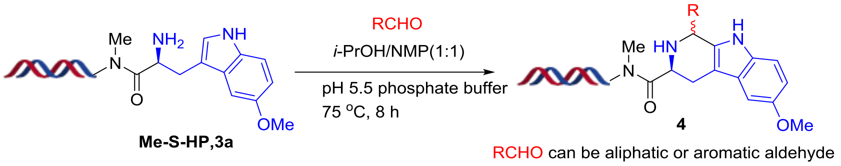
Next, the freshly prepared active ester solution was transferred to the Me-S-HP solution (1 mM in pH 9.5 sodium borate buffer, 1.00 mL, 1 eq.). After addition, the solution was eddied, centrifuged and stood at

20 °C for 2 h. Then the reaction mixture were treated with the second addition of the activated ester solution. The tube was centrifuged, eddied, re-centrifuged and stood at 20 °C for 16 h. After reaction, ethanol precipitation was done.

2) De-Fmoc

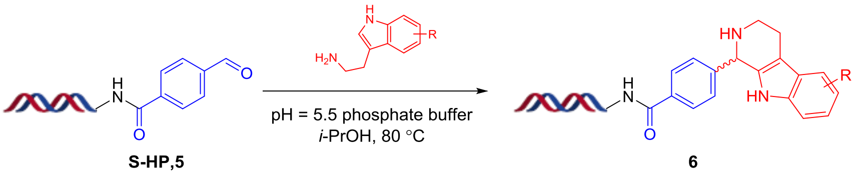
The solid of DNA substrate that from acylation was dissolved in 500 μL ddH$_2$O to make the 1 mM solutions in 15 mL tube. Then to the DNA solutionwas added 20% piperidine (500 μL). The tube was eddied, centrifuged and stood at 20 °C for 2 hr. After reaction, ethanol precipitation was done.

SI-4-3 General Procedure 2 for DNA-compatible Pictet-Spengler reaction



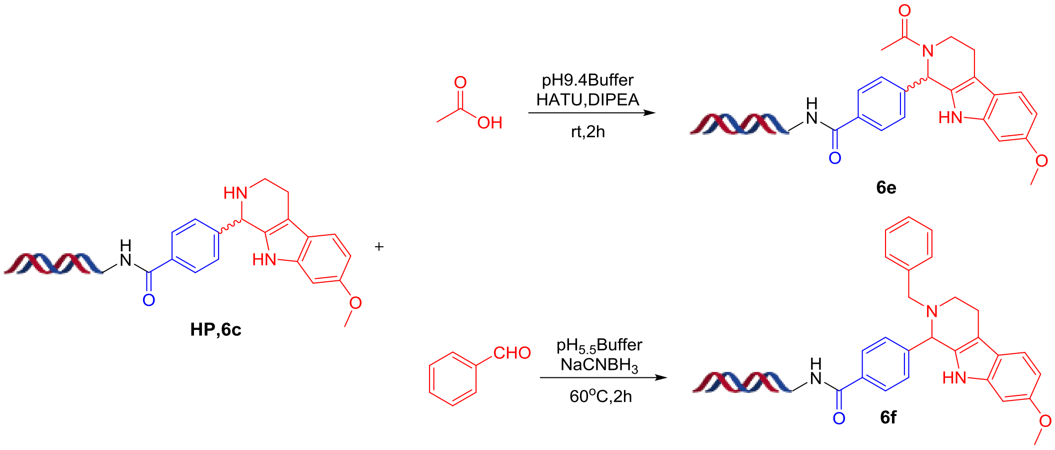
To the solution of DNA-conjugated tryptamine substrate **3**(1 mM in pH 5.5 sodium phosphate buffer, 9.00 μL, 1 *eq.*) was added aldehyde solution (400 mM in NMP, 4.0 μL, 180 *eq.*) in a 96-well plate. The plate was centrifuged, eddied and re-centrifuged. Then the pure *i*-PrOH (4.0 μL) was added to the mixed solution. The mixture was heated in PCR at 75 °C for 8 hr. After then, ethanol precipitation was done.

SI-4-4 General Procedure 3 for DNA-compatible Pictet-Spengler reaction



To the solution of DNA-conjugated aldehyde substrate **5**(1 mM in pH 5.5 sodium phosphate buffer, 10.0 μL, 1 *eq.*) was added tryptamines (400 mM in *i*-PrOH, 5.0 μL, 200 *eq.*) in a 96-well plate. The plate was centrifuged, eddied and re-centrifuged. Then the pure *i*-PrOH (5.0 μL) was added to the mixed solution. The mixture was heated in PCR at 80 °C for 16 hr. After then, ethanol precipitation was done.

SI-4-5 General Procedure 4 for amine capping



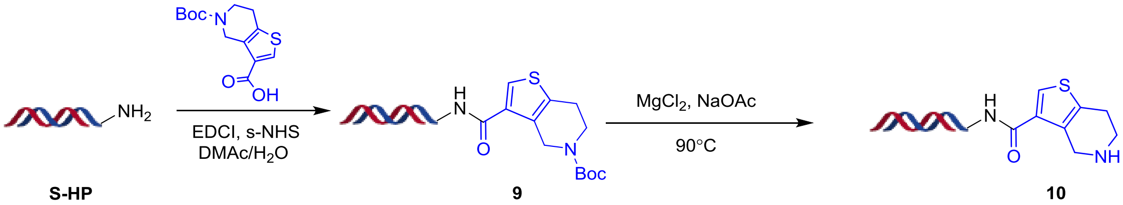
1) Acylation of substrate **6c**

To a 600 μL tube was added HATU (200 mM in DMA, 5 μL, 200 eq.), DIPEA (200 mM in DMA, 5 μL, 200 eq.) and acetic acid (200 mM in DMA, 5 μL, 200 eq.). This solution was mixed by vortex, then centrifuged and stand at 20 °C for 10 min to make the activated ester. Next, the freshly prepared active ester solution was transferred to the HP solution (2.5uL, 2 mM in water, 1 eq.), which was added 2.5uL pH 9.4 buffer solution. After addition, the solution was vortex, centrifuged and stood at 20°C for 2 h. After reaction, ethanol precipitation was done.

2) Reductive amination of **6c**

To the solution of DNA-conjugated amine substrate **6c** (1 mM in pH 5.5 sodium phosphate buffer, 5.00 μL, 1 eq.) was added aldehyde solution (200 mM in DMA, 5 *μ*L, 200 eq.) and NaCNBH$_3$ solution (400 mM in water, 2.5 *μ*L, 200 eq.) in a 250 uL tube. Then the mixture was heated in PCR at 60 °C for 2 hr. After then, ethanol precipitation was done.

SI-4-6 General Procedure 5 for DNA-conjugated amino acids synthesis



1) Acylation of S-HP

To a 15 mL tube was added the solution of EDCI (200 mM in DMSO, 125 *μ*L, 50 *eq.*), *s*-NHS (200 mM in DMSO/ddH$_2$O=1/1, 75 *μ*L, 30 *eq.*) and BocN-amino acid. This solution was eddied, then centrifuged and stood at 20 °C for 15 min to make the activated ester.

Next, the freshly prepared active ester solution was transferred to the S-HP solution (1 mM in pH 9.5 sodium borate buffer, 500 μL, 1 *eq.*). After addition, the solution was eddied, centrifuged and stood at 20 °C for 2 h. Then the reactions were treated with the second addition of the activated ester solutions. The tube was centrifuged, eddied, re-centrifuged and stood at 20 °C for 16 h. After then, ethanol precipitation was done.

2) De-Boc

The solid of DNA substrate **9** was dissolved in 500 μL ddH$_2$O to make the 1 mM solution in 15 mL tube. Then to the DNA solution was added 500 μL NaOAc aq. (75 mM in ddH$_2$O, 75 eq.) and 250 μL MgCl$_2$ aq. (1 mM in ddH$_2$O, 0.5 *eq.*). The solution was mixed and stood at 90 °C for 16 hr. After then, ethanol precipitation was done.

SI-5 DNA Damage Evaluation

Pictet-Spengler reaction was performed with a DNA conjugated compound with a double stranded DNA coding region to mimic the library component. The product was then ligated to an oligonucleotide to generate a full-length DNA fragment and examined by bioanalyzer (Figure 3). The concentration of DNA was affected by Pictet-Spengler reaction somehow, because there was some liquid phase change and an EtOH precipitation process in the Pictet-Spengler reaction, which could result some loss of the DNA during these steps. The length of the product has no change compared to the control group on the other hand, indicating that the Pictet-Spengler reaction did not affect the maneuverability of the DNA ligation.

SI-5-1 QPCR Test

The concentration and the amplification efficiency of the ligation products were assessed by qPCR after ethanol precipitation. Two parallel experimental groups were set up to determine 1) if the Pictet-Spengler reaction affects the ligation efficiency or the remaining DNA quantity 2) if the Pictet-Spengler reaction affects the amplification efficiency by PCR. qPCR was performed with the SYBR Green Master Mix kit (Thermo) on a Real-Time PCR System (QuantStudio 7 Flex). All samples were run in triplicates and subjected to PCR cycles as follows: 95 °C heat activation for 5 min followed by 40 cycles of 95 °C denaturation (10 seconds each), 55 °C annealing (15 seconds each), and 72 °C extension (30 seconds each). The result showed a slight difference in the starting concentration of the template, suggesting possible degradation or loss of DNA during the process of reaction, consistent with the observation by Bioanalyzer (**Fig. 4**). To further assess the amplification efficiency, the quantity of the full length DNA templates was first normalized based on the Bioanalyzer result and qPCR with serial dilution was performed. Linear fitting was then calculated respectively based on the CT values. The slope, which dictates the amplification efficacy, was compared between the experimental groups. No significant difference was observed between the Pictet-Spengler reaction group and the negative control group, indicated no obvious impact on PCR efficiency by the reaction. Moreover, melting curves of the qPCR products were examined and no peak shift or multiple peaks were observed, suggesting no significant alteration of DNA species after the reaction. Thus, in summary, the DNA remained in good integrity after the Pictet-Spengler reaction.

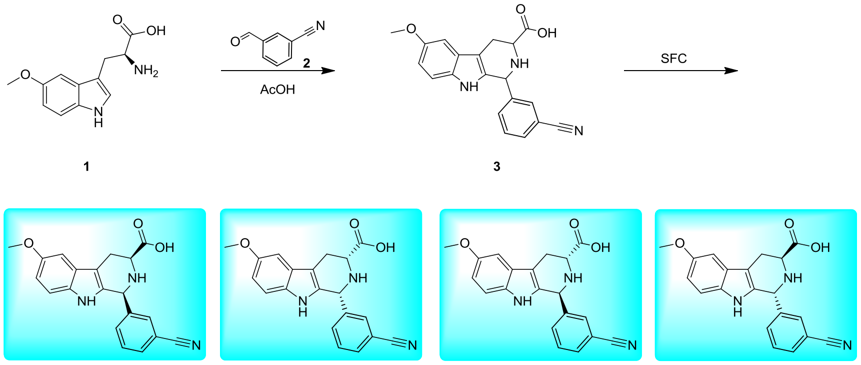
SI-5-2 Next-generation sequencing.

2 μL of the 1.65e+5 folds dilution sample was used as a template for PCR amplification. To a PCR tube was added diluted sample (2 μL), 10x high fidelity PCR buffer (5 μL), 50 mM $MgSO_4$ (2 μL), 10 mM dNTP mix (1 μL), Platinum Taq DNA Polymerase (0.2 μL), 10 μM forward primer (2 μL), 10 μM reverse primer (2 μL), and nuclease-free water (35.8 μL). The PCR products were purified by the Agencourt AMPure XP Beads method. The purified samples were sent for next-generation sequencing (Illumina NovaSeq). Bowtie2 was used to map the sequenced reads by local alignment. The detailed mapping identity were extracted from CIGAR string and XM flag in the SAM format. The results of NGS showed that all samples retained the right sequence as expected (**Figure 6**), indicating that the chemical reactions did not affect the encodability of DNA tags.

In conclusion, our data revealed that the Pictet-Spengler reactions used in this paper caused no damage to DNA, and thus could potentially be used for the encoded library construction.

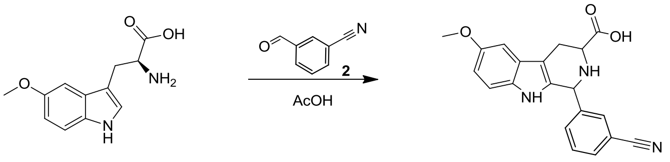
SI-6 Off-DNA Validation of PS Reaction

SI-6-1 Synthetic Scheme



DP1-Peak1, 30.0 mg, 95% purity DP1-Peak2, 30.0 mg, 95% purity DP2-Peak1, 50.0 mg, 95% purity

SI-6-2 General procedure for preparation of intermediate 3



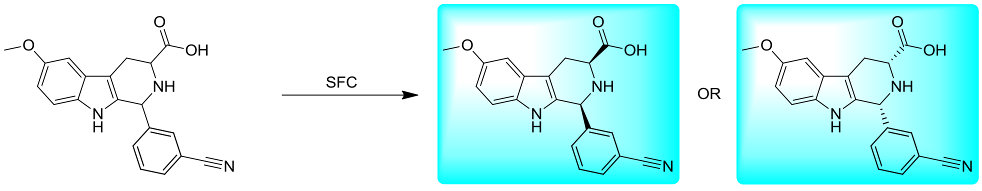
1 3

To a solution of **Compound 1** (500 mg, 2.13 mmol, 1.00 eq) in AcOH (5.00 mL) was added **Compound 2** (307.88 mg, 2.35 mmol, 1.10 eq). The mixture was stirred at 110 °C for 5 hrs. LCMS (EW22081-2-P1A1, RT1=0.700, RT2=0.735) showed **Compound 1** was consumed completely and two main peaks with desired was detected. The mixture was concentrated under reduced pressure to give a residue. **Compound 3** contained two parts (triturated product and prep-HPLC product). The crude product was triturated with CH_3CN at 25 °C for 30 min and obtained **Compound 3** (300 mg, 100% purity) as a yellow solid, which was confirmed by LCMS (EW22081-2-P1A3), HPLC (EW22081-2-P1H1), H NMR (EW22081-2-P1R3), SFC (EW22081-2-P1S2_c2). And the mother liquor was purified by prep-HPLC (column: Phenomenex luna C18 15*40mm*15um; mobile phase: (water (0.1%TFA)-CAN); B%: 10%-40%, 10min) and got another **Compound 3** (320 mg, 69.8% purity) as a yellow solid, which was confirmed by LCMS (EW22081-2-P1A7), SFC (EW22081-2-P1S4_d1), SFC (EW22081-2-P1S4_d2).

^{1}H NMR: EW22081-2-P1R3, (400 MHz, MeOD)

δ 7.89 (t, J = 15.6 Hz, 2 H), 7.79 (d, J = 8.0 Hz, 1 H), 7.68 (t, J =15.6 Hz, 1H), 7.15 (d, J =8.8 Hz, 1H), 7.06 (d, J = 2.0 Hz, 1H), 6.80 (dd, J_1 = 8.8 Hz, J_2 = 2.4 Hz, 1H), 5.86 (s, 1H), 4.15 (dd, J_1 = 12 Hz, J_2 = 5.2 Hz, 1H), 3.83 (s, 3H), 3.50 (dd, J_1 = 15.6 Hz, J_2 = 4.4 Hz, 1H), 3.20 - 3.10 (m, 1H).

SI-6-3 General procedure for preparation of DP1-Peak1 and DP1-Peak2



3

DP1-Peak1, 30.0 mg, 95% purity

DP1-Peak2, 30.0 mg, 95% purity

Compound 3 was purified by prep-SFC (column: DAICEL CHIRALCEL OJ (250mm*30mm, 10um); mobile phase: (0.1%NH3H2O ETOH); B%: 40%-40%, 3.6 min; 180 min) to go two product. DP1-Peak1 (0.08 g, 24.5% yield, 98.0% purity) was obtained as a off-white solid, which was confirmed by H NMR (EW22081-3-P1R3), LCMS (EW22081-3-P1A1), HPLC (EW22081-3-P1H1), SFC (EW22081-3-P1S1_c1), NOE (EW22081-3-P1N1), C NMR (EW22081-3-P1C3). DP1-Peak2 (0.10 g, 30.3% yield, 97.0% purity) was obtained as a yellow solid, which was confirmed by H NMR (EW22081-3-P1R4), LCMS (EW22081-3-P1A2), HPLC (EW22081-3-P1H2), SFC (EW22081-3-P1S2_c1), NOE (EW22081-3-P1N2), C NMR (EW22081-3-P1C4).

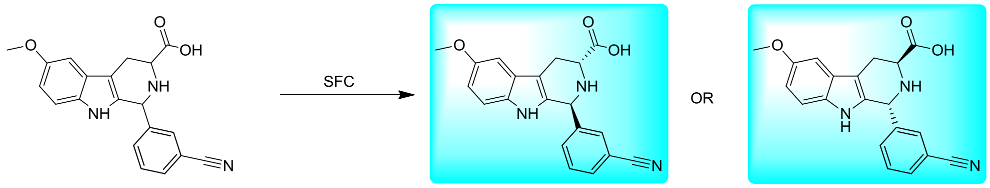
^{1}H NMR: EW22081-3-P1R3, (400 MHz, DMSO-d_6)

δ 10.23 (s, 1H), 7.82 (d, J = 6.8 Hz, 2H), 7.72 (d, J = 8.0 Hz, 1H), 7.58 (t, J = 15.6 Hz, 1H), 7.08 (d, J = 8.8 Hz, 1H), 6.97 (d, J = 2.2 Hz, 1H), 6.66 (dd, J_1 = 8.8 Hz, J_2 = 2.0 Hz, 1H), 5.35 (s, 1H), 3.75 (s, 3H), 3.72 (s, 1H), 3.06 (d, J = 14.8 Hz, 1H), 2.81 (t, J = 26.0 Hz, 1H).

^{1}H NMR: EW22081-3-P1R4, (400 MHz, DMSO-d_6)

δ 10.24 (s, 1H), 7.82 (d, J = 6.4 Hz, 2H), 7.72 (d, J = 8.0 Hz, 1H), 7.58 (t, J = 16.0 Hz, 1H), 7.08 (d, J = 8.8 Hz, 1H), 6.97 (d, J = 2.4 Hz, 1H), 6.66 (dd, J_1 = 8.4 Hz, J_2 = 2.0 Hz, 1H), 5.36 (s, 1H), 3.75 (s, 3H), 3.72 (s, 1H), 3.07 (d, J = 15.2 Hz, 1H), 2.81 (t, J = 24.8 Hz, 1H).

SI-6-4 General procedure for preparation of DP2-Peak1



3

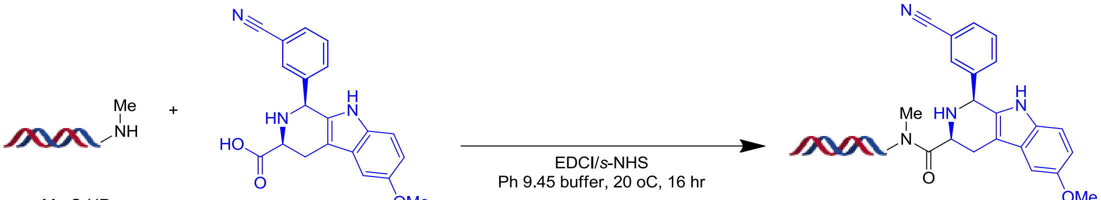
DP2-Peak1, 50.0 mg, 95% purity

Compound 3 was purified by prep-SFC (column: DAICEL CHIRALPAK IG (250mm*30mm, 10um); mobile phase: (0.1%NH3H2O ETOH); B%: 45%-45%, 4.6min; 50min) to go two product. DP2-Peak1 (0.15 g, 394.68 umol, 45.70% yield, 91.4% purity) was obtained as a yellow solid, which was confirmed by H NMR (EW22081-4-P1R3), LCMS (EW22081-4-P1A1), HPLC (EW22081-4-P1H1), NOE (EW22081-4-P1E1), C NMR (EW22081-4-P1C1), SFC (EW22081-4-P1S2_d11).

^{1}H NMR: EW22081-4-P1R3, EW22081-4-P1R2, (400MHz, DMSO-d_6)

δ 10.57 (s, 1H), 7.76 (t, J = 20.8 Hz, 2H), 7.61 (d, J = 7.6 Hz, 1H), 7.55 (t, J = 15.2 Hz, 1H), 7.14 (d, J = 8.8 Hz, 1H), 6.99 (d, J = 2.4 Hz, 1H), 6.70 (dd, J_1 = 8.8 Hz, J_2 = 2.4 Hz, 1H), 5.53 (s, 1H), 3.76 (s, 3H), 3.73 (d, J = 6.4 Hz, 1H), 3.12 (dd, J_1 = 15.2 Hz, J_2 = 5.2 Hz, 1H), 2.94 (dd, J 1= 15.2 Hz, J2 = 7.6, 1H).

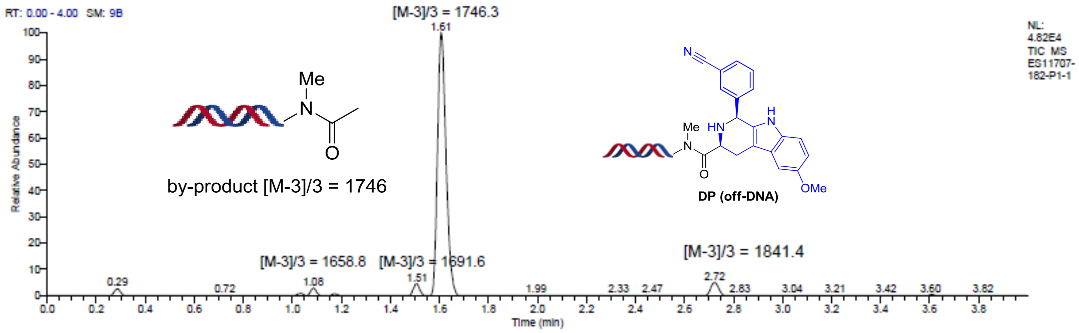
SI-6-5 LC Trace and Mass of DP



Me-S-HP

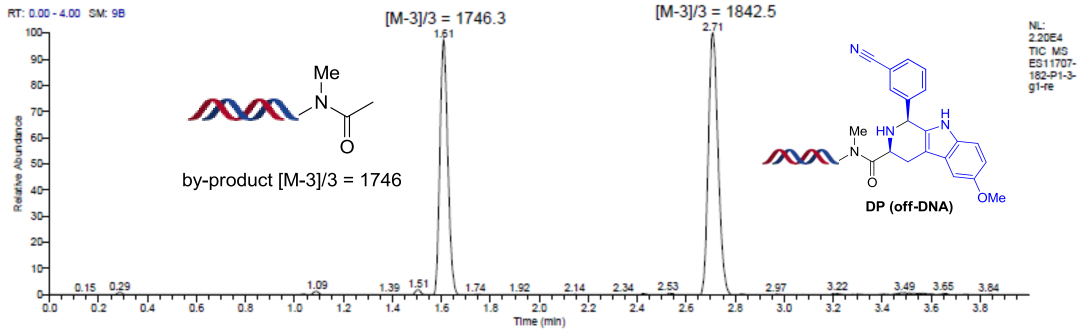
DP (off-DNA)

Following **General Procedure 1**
Yield: 4.14%
Exact mass: 5528.10
Triply charged mass [M-3]/3, calculated: 1841.7; observed: 1841.4



RT	Peak Height	Peak Width	Peak Area	Area %
1.08	1792.03	0.04	2141.68	1.69
1.50	2568.62	0.05	3770.00	2.97
1.61	53830.47	0.10	115768.49	91.20
2.72	2614.04	0.06	5257.59	4.14

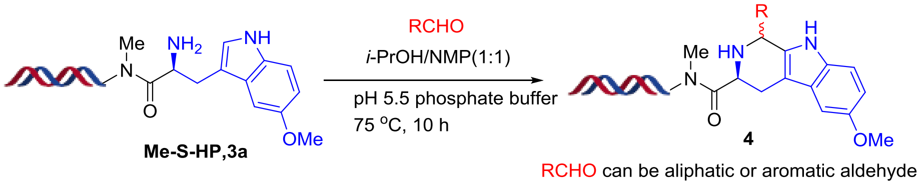
DP (off-DNA)+4af (retain time = 2.71)



RT	Peak Height	Peak Width	Peak Area	Area %
1.61	23531.58	0.09	48002.13	44.13
2.71	23648.51	0.11	60762.34	55.87

Fig.40. LC trace and mass of **DP**

SI-7 Mass Spectrum of 34 Aldehyde Building Blocks



Materials

Product **3a**: 1 mM in sodium phosphate buffer (250 mM, pH = 5.5)

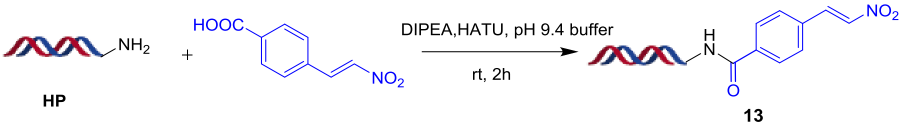
Aldehyde: 400 mM in NMP

Sodium phosphate buffer, pH = 5.5, 250 mM

Procedure

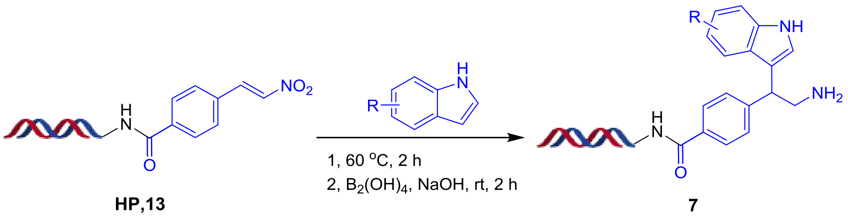
To **3a** solution (5 nmol, 5 μL), was added to a solution of aldehyde (400 mM in NMP, 2.25 μL, 180 eq). Then the *i*-PrOH (2.25 μL) was added to the mixed solution. The mixture was vortexed. Heat the reaction mixture in PCR at 75 °C for 10 h. After the reaction, add 40.5 μL water, then add 5 M NaCl solution (10 % by volume) and cold ethanol (2.5 times by volume, ethanol stored at -20 °C). The mixture was stored at a -80 °C freezer for more than 30 minutes. Centrifuge the sample for around 30 minutes at 4 °C in a micro-centrifuge at 10000 rpm. The above supernatant was removed and the pellet (precipitate) was cooled in liquid nitrogen and then placed on a lyophilizer. After lyophilization, the dry pellet was recovered.

SI-8 General Procedure for DNA-conjugated nitroalkene 13



To a 600 μL tube was added HATU (200 mM in DMA, 50 μL, 100 eq.), DIPEA (200 mM in DMA, 50 μL, 100 eq.) and 4-(2-nitrovinyl) benzoic acid (200 mM in DMA, 100 μL, 200 eq.). This solution was mixed by vortex, then centrifuged and stand at 20 °C for 10 min to make the activated ester. Next, the freshly prepared active ester solution was transferred to the HP solution (50 μL, 2 mM in water, 1 eq.), which was added 50 μL pH 9.4 buffer solution. After addition, the solution was vortex, centrifuged and stood at 20 °C for 2 h. After reaction, ethanol precipitation was done.

SI-9 General Procedure for DNA conjugated indole substituted amine 7



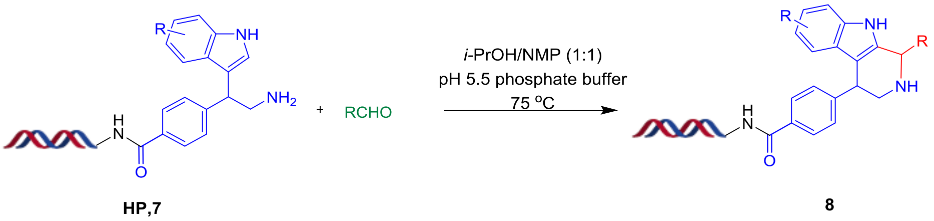
1) Addition of 6-methoxy-1*H*-indole

To the solution of DNA-conjugated nitroalkene **13** (1 mM in water, 50.00 μL, 1 eq.) was added 6-methoxy-1*H*-indole solution (200 mM in DMA, 50.0 μL, 200 eq.) in a 250 μL tube. The mixture was vortex. Then the mixture was heated in PCR at 60 °C for 2 hr. After then, ethanol precipitation was done.

2) Nitro reduction

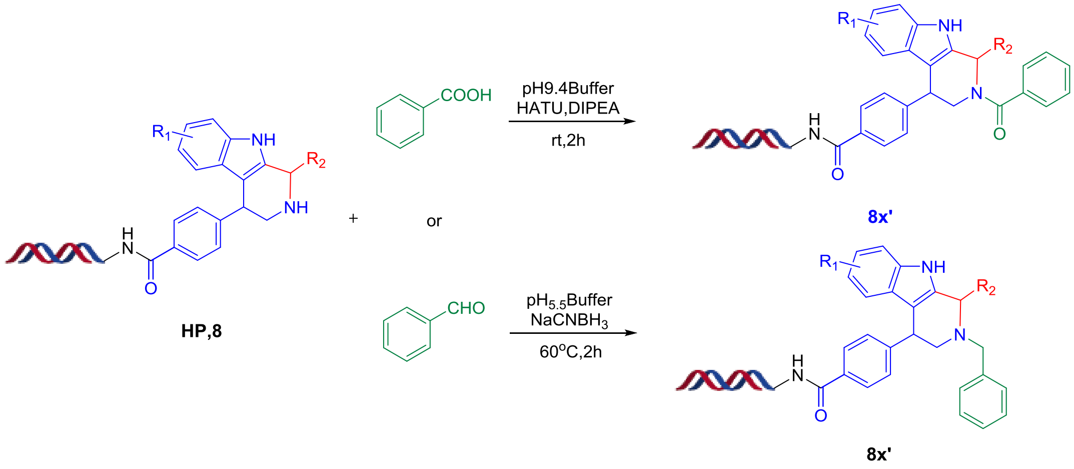
To the solution of DNA-conjugated substrate (1 mM in water, 50.00 μL, 1 eq.) was added $B_2(OH)_4$ solution (100 mM in water, 50.0 μL, 100 eq.) in a 250 μL tube. The mixture was vortex. After addition, the solution was vortex, centrifuged and stood at 20 °C for 2 h. After reaction, ethanol precipitation was done.

SI-10 General Procedure for DNA-compatible Pictet-Spengler reaction



To the solution of DNA-conjugated indole substituted amine **7** (1 mM in pH 5.5 sodium phosphate buffer, 5.00 μL, 1 eq.) was added 4-nitrobenzaldehyde solution (400 mM in NMP, 2.25 μL, 180 eq.) in a 250 μL tube. Then the pure *i*-PrOH (4.0 μL) was added to the mixed solution. The mixture was heated in PCR at 75 °C for 16 hr. After then, ethanol precipitation was done.

SI-10-1 General Procedure for amine capping



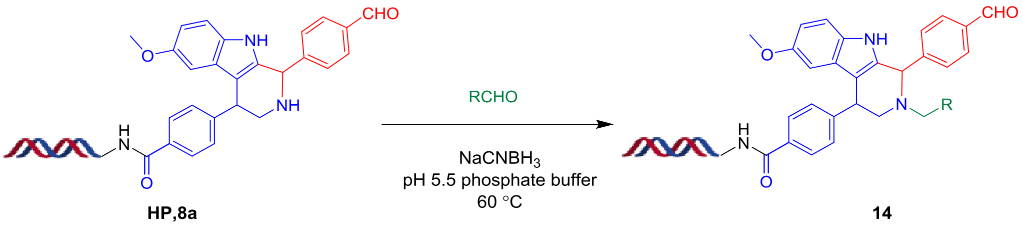
1) Acylation of substrate **8**

To a 600 µL tube was added HATU (200 mM in DMA, 5 µL, 200 eq.), DIPEA (200 mM in DMA, 5 µL, 200 eq.) and benzoic acid (200 mM in DMA, 5 µL, 200 eq.). This solution was mixed by vortex, then centrifuged and stand at 20 °C for 10 min to make the activated ester. Next, the freshly prepared active ester solution was transferred to the HP solution (2.5uL, 2 mM in water, 1 eq.), which was added 2.5uL pH 9.4 buffer solution. After addition, the solution was vortex, centrifuged and stood at 20°C for 2 h. After reaction, ethanol precipitation was done.

2) Reductive amination of **8**

To the solution of DNA-conjugated amine substrate **8** (1 mM in pH 5.5 sodium phosphate buffer, 5.00 µL, 1 eq.) was added aldehyde solution (200 mM in DMA, 5 μL, 200 eq.) and $NaCNBH_3$ solution (400 mM in water, 2.5 μL, 200 eq.) in a 250 uL tube. Then the mixture was heated in PCR at 60 °C for 2 hr. After then, ethanol precipitation was done.

SI-11 General Procedure for amine capping



HP,8a

14

To the solution of DNA-conjugated amine substrate **8a** (1 mM in pH 5.5 sodium phosphate buffer, 5.00 μL, 1 eq.) was added aldehyde solution (400 mM in NMP, 2.5 μL, 200 eq.) and $NaCNBH_3$ solution (400 mM in water, 2.5 μL, 200 eq.) in a 250 uL tube. Then the mixture was heated in PCR at 60 °C for 2 hr. After then, ethanol precipitation was done.