Supporting Information

Photochemical C–H Arylation of Heteroarenes for DNA-Encoded Library Synthesis

Matthias Krumb‡,[a] Lisa Marie Kammer‡,[a] Shorouk O. Badir,[a] María Jesús Cabrera-Afonso,[a] Victoria E. Wu,[b] Minxue Huang,[b] Adam Csakai,[b] Lisa A. Marcaurelle,[b] and Gary A. Molander*[a]

[a]*Roy and Diana Vagelos Laboratories, Department of Chemistry, University of Pennsylvania, 231 South 34th Street, Philadelphia, PA 19104-6323, United States*

[b]*Encoded Library Technologies / NCE Molecular Discovery, R&D Medicinal Science and Technology, GlaxoSmithKline, 200 Cambridge Park Drive, Cambridge, MA 02140, United States*

‡*These authors contributed equally*

*To whom correspondence should be addressed. E-mail: gmolandr@sas.upenn.edu

1. General Considerations

1.1 General: All chemical transformations requiring inert atmospheric conditions were carried out using Schlenk line techniques with a 4- or 5-port dual-bank manifold. For blue light irradiation, two Kessil PR160-456 nm lamps (19 V DC 40 W Max) were placed 1.5 inches away from PCR tubes. NMR spectra (^{1}H, ^{13}C) were obtained at 298 ºK using 400 or 500 MHz spectrometers. ^{1}H NMR spectra were referenced to residual $CHCl_3$ (δ 7.26 ppm) in $CDCl_3$. ^{13}C NMR spectra were referenced to $CDCl_3$ (δ 77.3 ppm). Reactions were monitored by LC/MS, GC/MS, ^{1}H NMR, and/or TLC on silica gel plates (60 Å porosity, 250 μm thickness). TLC analysis was performed using hexanes/EtOAc as the eluent and visualized using ninhydrin, *p*-anisaldehyde stain, and/or UV light. Flash chromatography was accomplished using an automated system (CombiFlash®, UV detector, λ = 254 nm and 280 nm) with RediSep® R_f silica gel disposable flash columns (60 Å porosity, 40–60 μm) or RediSep R_f Gold® silica gel disposable flash columns (60 Å porosity, 20–40 μm). Accurate mass measurement analyses were conducted using electron ionization (EI) or electrospray ionization (ESI). The signals were mass measured against an internal lock mass reference of perfluorotributylamine (PFTBA) for EI-GCMS, and leucine enkephalin for ESI-LC/MS. IR spectra were recorded on an FT-IR using either neat oil or solid products. Solvents were purified with drying cartridges through a solvent delivery system. Melting points (°C) are uncorrected.

1.2 Chemicals: Deuterated NMR solvents were purchased and stored over 4Å molecular sieves. CH_2Cl_2, DMF, EtOAc, hexanes, and DMSO were used as purchased. DIPEA, Et_3N, HATU (*N*-[(dimethylamino)-1*H*-1,2,3-triazolo-[4,5-*b*]pyridin-1-ylmethylene]-*N*-methylmethanaminium hexafluorophosphate *N*-oxide), and (*E*)-3-(ethyldiazenyl)-*N*,*N*-dimethylpropan-1-amine were purchased from commercial suppliers and used without further purification. Synthesis of all new heteroarene derivatives and new on-DNA substrates is outlined here. Additional heteroarenes were purchased from commercial suppliers. The $Ir(ppy)_3$ catalyst was purchased from commercial suppliers and used without further manipulation. All other reagents were purchased commercially and used as received. Photoredox-catalyzed reactions were performed using PCR 8-strip tubes (Ref. Fisher 781320) with PCR strips of 8 caps (Ref. Fisher 781340). DMSO was purchased and used as received. HyPure™ Molecular Biology Grade Water was purchased and used as received without further manipulation.

1.3 Analysis of "on-DNA" reactions: Analysis of on-DNA reactions was performed by LC/MS: After reaction completion, an aliquot of the reaction mixture was diluted with H_2O to approximately 0.05–0.13 mM. At this point, 3 μL aliquots of the LC/MS sample was injected onto a reverse-phase chromatography column (Clarity 2.6 μm Oligo-MS 100 Å 2.1x50 mm) and eluted (10-90% B over 4 min at 0.5 mL/min flow rate; Solvent A: 0.75% v/v/ HFIP / 0.038% TEA / 5 μM EDTA in H_2O; Solvent B: 0.75% HFIP,

0.038% TEA, 5 μM EDTA in 90/10 MeOH/deionized H_2O) with no UV monitoring. Effluent was analyzed on a Waters SQ Detector 2 ACQUITY UPLC System in Thermo Exactive Plus LC-esiMS with a Vanquish UHPLC. For the functionalized headpiece samples (the on-DNA aryl halides), % conversion was determined based on reported peak intensities following deconvolution (between 3,000-10,000 Da) of the DNA charge states using Intact Mass™ by Protein Metrics Inc. (version 3.7-32x64). For the photoredox scope reactions, % conversion was determined using Intact Mass™ by Protein Metrics Inc. (version 3.7-32x64). Data was scanned between 0.3-2.2 min and deconvoluted between 4,000-6,000 Da, with a mass tolerance window of 2.5-3 Da, with 5% of base peak threshold was set for reporting. For conversion calculations for each example, the peaks annotated with colored dots were used for the calculations. Na, K, NH_4, and HFIP adducts were included in the product percentage. Detailed parameters can be found later in the Supporting Information.

1.4 Materials for "on-DNA" synthesis: DNA headpiece HP-NH_2(5'-/5Phos/GAGTCA/iSp9/iUniAmM/iSp9/TGACTCCC-3') was obtained from Biosearch Technologies, Novato, CA. The spacer-elongated AOP-Headpiece (Figure S1) was prepared via HATU coupling following the general procedure described later in this document with 5 equiv each of Fmoc-15-amino-4,7,10,13-tetraoxapentadecanoic acid (Fmoc-AOP), *i*-Pr_2NEt, and HATU. The lyophilized product of this reaction was then deprotected by exposure to a 10% piperidine in H_2O solution. After the reaction was deemed complete by LC/MS analysis, the reaction was precipitated following the EtOH protocol and is typically pure enough to be used without further purification.

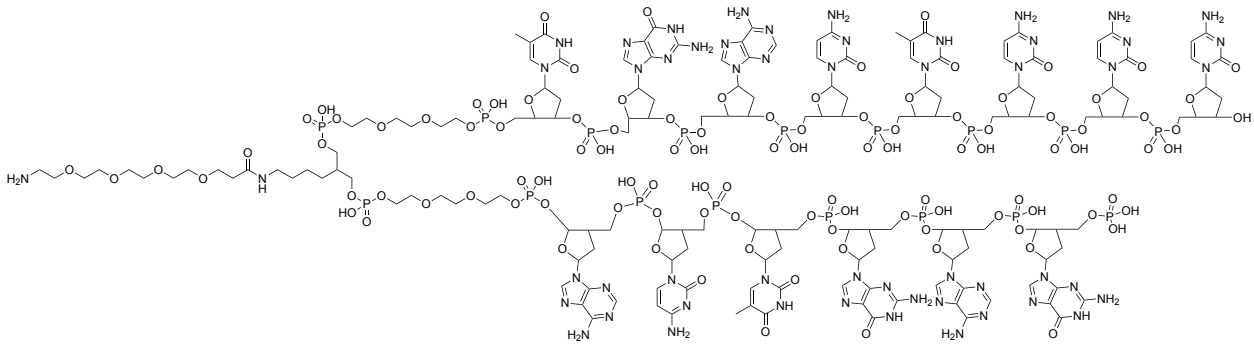
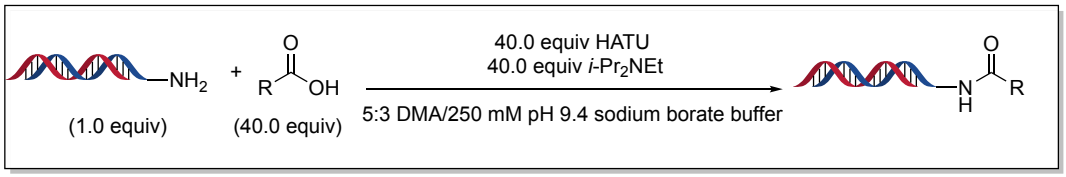


Figure S1. Sequence and structure of the AOP-headpiece (molecular weight = 5184.5220).

2. Preparation of on-DNA Substrates

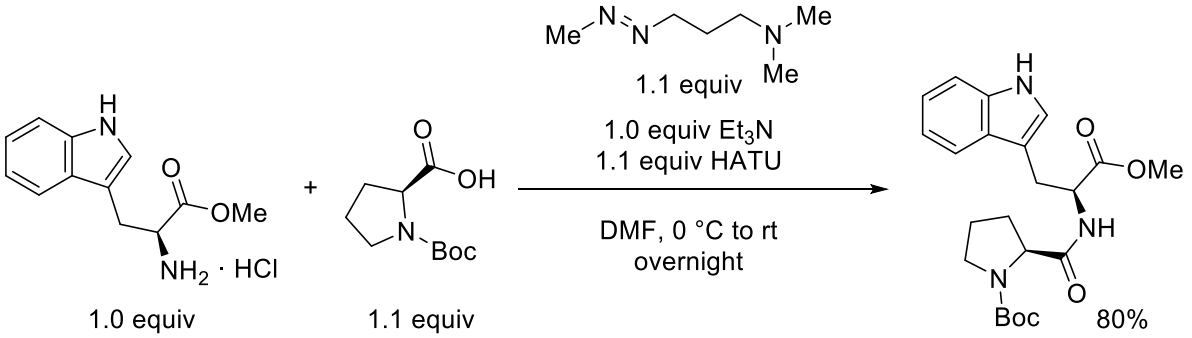


2.1 HATU premix protocol for acylation of DNA headpieces: The HATU (200 mM in DMA, 40.0 equiv), *i*-Pr_2NEt (200 mM in DMA, 40.0 equiv), and the corresponding carboxylic acid (200 mM in DMA, 40.0 equiv) solutions were individually cooled at 4 °C for 5 min. Once chilled, the acid, *i*-Pr_2NEt, and HATU solutions were added sequentially to a centrifuge tube, vortexed briefly, and allowed to react at 4 °C for 20 min. The oligomer solution (1 mM in 250 mM pH 9.4 sodium borate buffer) was then added, and the mixture was vortexed. The reaction was allowed to proceed at rt and monitored by LC/MS. Upon completion, the reaction was worked up following the EtOH precipitation protocol below.

2.2 EtOH precipitation protocol: The reaction mixture was transferred to a centrifuge tube where it filled at most 1/4 of the total volume. A volume of 5 M aq NaCl equal to 1/10 of the reaction volume was then added, followed by cold (–20 °C) EtOH equal to 2.5 reaction volumes. The resulting mixture was then left to stand in a –80 °C freezer for at least 1 h or overnight. The chilled mixture was then centrifuged for 30 min at 4 °C at 3,300 rpm. The supernatant was then decanted and allowed to dry under reduced pressure. The resulting pellet was re-dissolved in H_2O to give a theoretical concentration of 2 or 5 mM. Purity was assessed by LC/MS, and optical density was obtained via NanoDrop. For long term storage, solutions were frozen in liquid nitrogen and lyophilized to dryness to give a white solid. If purity was less than 90% by LC/MS, HPLC purification was performed: gradient of 95% A (50 mM TEAA, pH = 7.5)/5% B (1% H_2O in CH_3CN) to 60% A/40% B, through a Gemini C18 column (5 µm, 110 Å, 30x100 mm), with UV visualization at 260 nm.

3. Synthesis of Heteroarene Derivatives

*Methyl (tert-Butoxycarbonyl)-L-prolinyl-L-tryptophan (**2z**)*



Methyl (*tert*-Butoxycarbonyl)-L-prolinyl-L-tryptophan (2z). The compound was prepared according to a modified procedure from *Carpino et al.*[1] To a 100 mL flask equipped with a stir bar was added Boc-L-proline (651 mg, 3.0 mmol, 1.1 equiv), Et_3N (0.38 mL 2.75 mmol, 1.0 equiv), and L-tryptophan methyl-ester hydrochloride (0.70 g, 2.75 mmol, 1.0 equiv). The reaction mixture was dissolved in DMF (20 mL) and stirred for 5 min. At this point, HATU (1.15 g, 3.0 mmol, 1.1 equiv) was added. Afterwards, the soln was cooled to 0 °C with an ice bath, and carbodiimide (0.39 g, 3.0 mmol, 1.1 equiv) was added portionwise. The soln was kept at 0 °C for 1 h and then stirred overnight. Upon completion, the reaction was diluted with EtOAc (25 mL) and transferred to a separatory funnel. After extraction with 1 M HCl (3 × 25 mL) and a 10% aq soln of $NaHCO_3$ (3 × 25 mL), the combined organic layers were washed with brine (25 mL) and then dried (Na_2SO_4). The solvent was removed under reduced pressure, and the crude product was purified using flash chromatography (gradient hexanes to 100% EtOAc) to give the title compound (910 mg, 2.19 mmol, 80%) as a colorless oil. **^{1}H NMR** (400 MHz, $CDCl_3$), δ (ppm) = 8.74–8.66 (m, 1H), 7.51 (d, J = 8.1 Hz, 1H), 7.31 (d, J = 8.1 Hz, 1H), 7.15 (t, J = 7.5 Hz, 1H), 7.08 (t, J = 7.5 Hz, 1H), 7.01–6.97 (m, 1H), 6.52 (br s, 1H), 4.88 (br s, 1H), 4.28–4.11 (m, 1H), 3.66–3.63 (m, 3H), 3.47–3.08 (m, 3H), 1.99–1.91 (m, 3H), 1.84–1.59 (m, 1H), 1.38 (s, 9H). **^{13}C NMR** (101 MHz, $CDCl_3$), δ (ppm) = 171.7, 171.3, 154.6, 136.2, 127.6, 122.9, 122.1, 119.5, 118.3, 111.4, 109.7, 80.8, 60.5, 52.3, 30.7, 28.2 (3C), 27.8, 23.4, 21.3, 14.2. **FT-IR** (cm^{-1}, neat, ATR), $\tilde{\nu}$ = 3300, 2977, 1741, 1665, 1479, 1366, 1250, 850. **HRMS** (ESI) calc. for $C_{22}H_{30}N_3O_5$ $[M+H]^+$: 416.2185, found: 416.2178.

4. Reaction Workflow

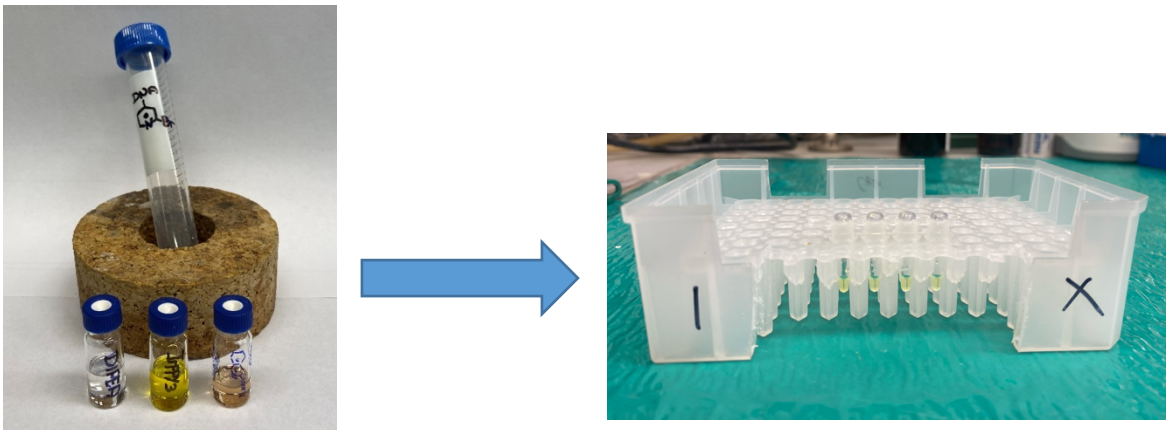


Figure S2. Photochemical reactions were conducted in PCR tubes under open-air conditions. Reagents were added as stock solutions.

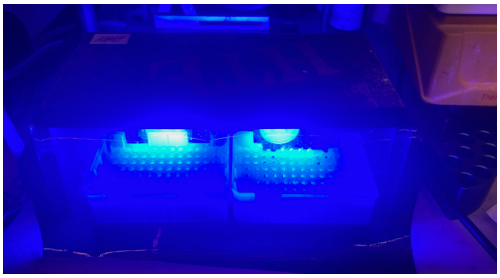


Figure S3. Reaction tubes were vortexed then positioned 1.5 inches away from two Kessil PR160 lamps (λ = 456 nm, 19 V DC 40 W Max) for the time designated for each experiment.

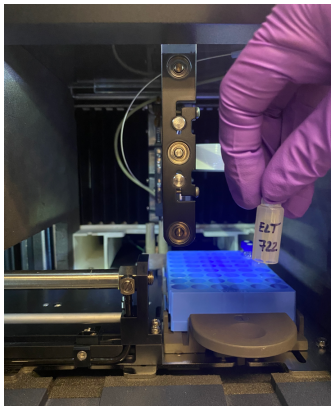
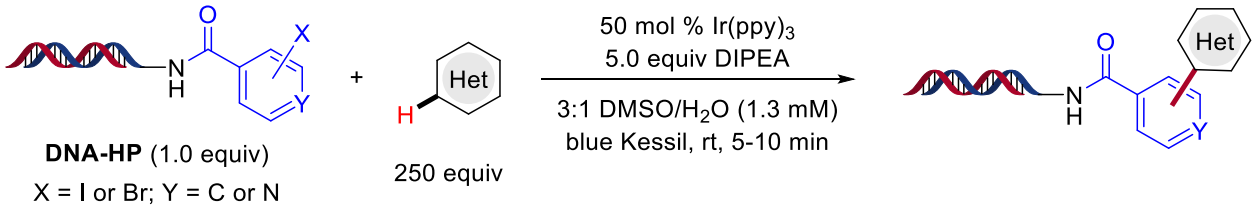


Figure S4. Upon completion, reactions were diluted with H_2O, filtered through MultiScreen® filter plates (Millipore) via centrifuge forces, and % conversion was determined using LC/MS analysis.

5. Procedures for Photoinduced Transformations

General Procedure I

Photoinduced on-DNA Minisci reaction via aryl radical formation



DNA-HP (1.0 equiv)
X = I or Br; Y = C or N

250 equiv

To a PCR Eppendorf tube was added DIPEA (2.5 µL of a 50 nmol/µL soln in DMSO, 125 nmol, 5 equiv), *fac*-Ir(ppy)$_3$ (2.5 µL of a 5 nmol/µL soln in DMSO, 12.5 nmol, 0.5 equiv), heteroarene (10 µL of a 625 nmol/µL soln in DMSO, 6250 nmol, 250 equiv), and DNA-tethered aryl halide (5 µL of a 5 nmol/µL soln in H_2O, 25 nmol, 1.0 equiv). The PCR tube was then capped, vortexed, and irradiated for 5-10 min with Kessil PR160 lamps at a distance of 1.5 inches. The reaction was then diluted with H_2O (150 µL), filtered through MultiScreen® filter plates via centrifuge forces, and subsequently analyzed by LC/MS.

6. qPCR, PCR, and Sequencing

4-Cycle tag mimic sequence

4-Cycle tag synthesis (aryl iodide)

The top and bottom strands (purchased from IDT as lyophilized powders) of a control 4-cycle tag were annealed by combining 1.2 µmol of each strand (2 mM in H_2O), heating to 95 °C for 5 min, then cooling to rt. The annealed tag solution (1.2 equiv) was then added to the 4-chloro-3-iodobenzoic acid headpiece (500 µL, 2 mM in H_2O), followed by 400 µL of 10x T4 ligation buffer, 8 mL of H_2O, and 40 µL of T4 DNA ligase purchased from Syngene. The ligation solution was vortexed and allowed to sit at rt overnight. The ligation was precipitated for 30 min at -80 °C following addition of 0.8 mL of 5 M NaCl (aq) and 20 mL of cold EtOH. The precipitated soln was then centrifuged at 3,300 rpm at 4 °C for 30 min, and the solvent was decanted to afford the DNA pellet, which was dried on a lyophilizer for 30 min. The crude pellet was resuspended in 8 mL of H_2O and split into two 30,000 molecular weight cut-off spin filters. The spin filters were put on the centrifuge for 15 min (20 °C, 3500 rpm), and the filtrate was collected. The original reaction flask was washed with 4 mL of H_2O and again split into the two-spin filter and put on the centrifuge for 15 min (20 °C, 3500 rpm). The filtrate was then collected, and the wash process was repeated two more times. Once complete, the product was collected and lyophilized overnight. The resulting white pellet was dissolved in 500 µL of H_2O, and a QC was taken showing 36% starting material and 47% desired product (this was not seen on the µTOF QC that was taken before). The ligation was pushed with additional annealed control tag (300 nmol, 1 mM in H_2O) followed by 200 µL of 10x T4 ligation buffer, 4 mL of H_2O, and 20 µL of T4 DNA ligase purchased from Syngene. The reaction was again capped, vortexed, and left to react at rt overnight. The ligation was precipitated for 30 min at -80 °C following addition of 0.8 mL of 5 M NaCl (aq) and 20 mL of cold EtOH. The precipitated soln was then centrifuged at 3,300 rpm at 4 °C for 30 min, and the solvent was decanted to afford the DNA pellet, which was dried on a lyophilizer for 30 min. The crude pellet was resuspended in 8 mL of H_2O and spit into two 30,000 molecular weight cut-off spin filters. The spin filters were put on the centrifuge for 15 min (20 °C, 3500 rpm), and the filtrate was collected. The original reaction flask was washed with 4 mL of H_2O and again split into the two-spin filter and put on the centrifuge for 15 min (20 °C, 3500 rpm). The filtrate was then collected, and the wash process was repeated two more times. Once complete, the product was collected and lyophilized overnight. The

lyophilized product was analyzed by optical density using a composite extinction coefficient of 1023700 L/(mol-cm) to determine isolated yield (715 nmol, 71.5%). LCMS calcd: 34,521, found: 34,521

Closing primer ligation on reacted material

Top Strand: 5'-/5Phos/ACG ATG CCC GGT CTA CNN NNN NNN NNN NCT GAT GGC GCG AGG GAG GC-3'

Bottom Strand: 5'-GTA GAC CGG GCA TCG TAA-3'

To each of the six exemplar reaction samples (2 nmol aliquot, 0.04 mM in H_2O) was added the closing primer (5 nmol, 1 mM in H_2O), 10X ligation buffer (10 μL), T4 DNA ligase (2 μL, 10 mg/mL), and H_2O (33 μL) for a final reaction volume of 100 μL. Ligations were allowed to proceed overnight at rt. Samples were analyzed by gel electrophoresis, and all were determined to have gone to sufficient completion.

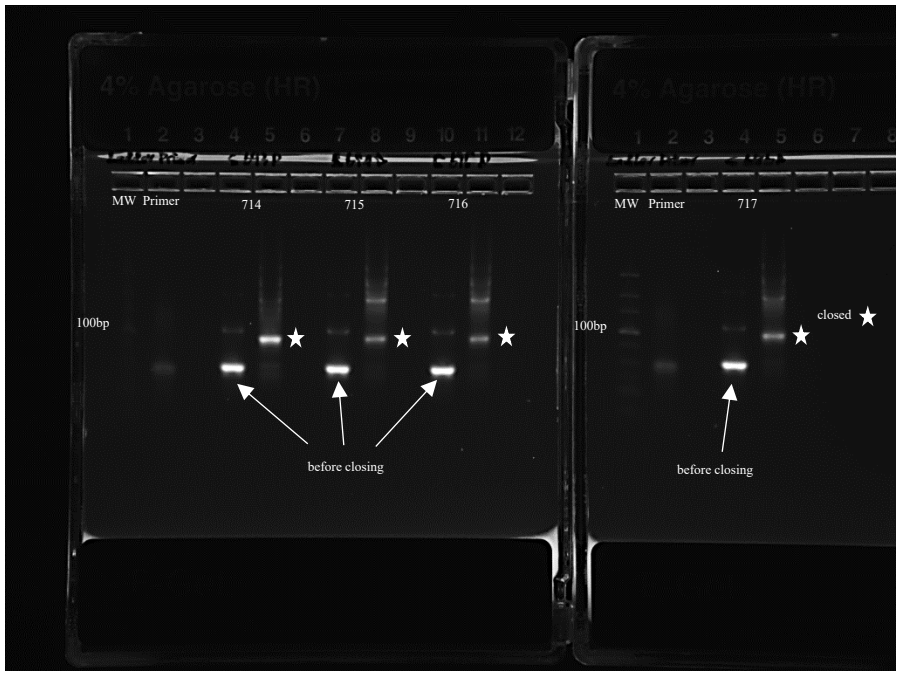


Figure S5. Gel electrophoresis of closing primer ligation on reacted material.

qPCR

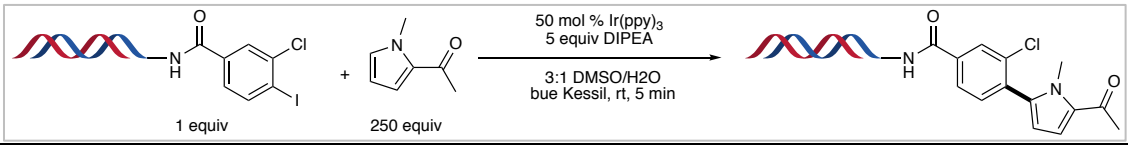
qFor: 5'-GCT ACC TCT GAC TCC CAA ATC GAT GT -3'

qRev: 5'-*ATA TTA GCC TCC CTC GCG CCA* TCA -3'

Quantitative PCR was performed on a Roche LightCycler 480 II PCR system with SYBR Green I as the detection dye. A bulk master mix solution was prepared by combining 1 mL of SYBR green, 60 µL of 10 µM PCR primer 565 Cla, 60 µL of 10 µM PCR primer 454 short, and 680 µL of H_2O. To 2 µL of sample was then added 18 µL of master mix. Samples were subjected to qPCR:

Stage	Temperature/Time	Number of Cycles
HotStart	95 °C / 5 min	1
Amplification	95 °C / 10 sec 55 °C / 15 sec 72 °C / 15 sec	40
Melt	95 °C / 1 sec 70 °C / 1 sec 95 °C	1
Cool	45 °C / 30 sec	1

Samples were then analyzed using the 2nd derivative maximum standard protocol on the instrument to determine how many molecules were present per µL sample. Samples achieved acceptable consistency across conditions in comparison to the no-light control sample ELT_717, suggesting that the conditions developed are not impacting the amount of amplifiable DNA present in a significant way.



Sample Name	Deviations	Molecules / µL sample
ELT_714	No deviation from standard conditions	1.13E+13
ELT_715	No DIPEA	8.56E+12
ELT_716	No photocatalyst	8.65E+12
ELT_717	No-light control	9.52E+12

Based on the Agilent TapeStation results following the PCR amplification and purification described above, an aliquot of each sample, representing approximately 1E8 molecules, were prepared for sequencing following the manufacturer's standard protocol with an Illumina MiSeq v3 kit, and sequenced on an Illumina MiSeq. Samples were subjected to 101 cycles for Read 1 and 9 cycles for Index 1. The resulting sequences were aligned to the 71 base reference sequence (below). The number of single base differences were counted and reported as a percentage of the total sequence count.

71 base sequence

5’ – GTA GAC CGG GCA TCG TAA CGA CCT TTC TCC GCT TAC CAG GCT TCT TGC GGA ACA CAT CGA TTT GGG AGT CA – 3’

Minisci Arylation: **Standard conditions** (ELT 714)

# of matched bases (out of 71)	% of total sequences (n = 2730559)
71	**85.68**
70	11.84
69	1.49
68	0.37
67	0.17
66	0.10
65	0.07
64	0.07
63	0.05
62	0.04
61	0.02
60	0.02
59	0.02
55	0.01
56	0.01
57	0.01
58	0.01
54	0.01
53	< 0.01
51	< 0.01
52	< 0.01
50	< 0.01
49	< 0.01
48	< 0.01
47	< 0.01
46	< 0.01
45	< 0.01
44	< 0.01
43	< 0.01
42	< 0.01
34	< 0.01

71 base sequence

5’ – GTA GAC CGG GCA TCG TAA CGA CCT TTC TCC GCT TAC CAG GCT TCT TGC GGA ACA CAT CGA TTT GGG AGT CA – 3’

Minisci Arylation: **No DIPEA** (ELT 715)

# of matched bases (out of 71)	% of total sequences (n = 2562964)
71	**85.31**
70	12.20
69	1.53
68	0.36
67	0.16
66	0.11
65	0.08
64	0.06
63	0.04
62	0.03
61	0.02
56	0.02
57	0.01
60	0.01
58	0.01
59	0.01
55	0.01
54	0.01
52	< 0.01
53	< 0.01
51	< 0.01
42	< 0.01
43	< 0.01
41	< 0.01
50	< 0.01
46	< 0.01
48	< 0.01
47	< 0.01
49	< 0.01
40	< 0.01
45	< 0.01
44	< 0.01
38	< 0.01
39	< 0.01
36	< 0.01
34	< 0.01
33	< 0.01
27	< 0.01

71 base sequence

5’ – GTA GAC CGG GCA TCG TAA CGA CCT TTC TCC GCT TAC CAG GCT TCT TGC GGA ACA CAT CGA TTT GGG AGT CA – 3’

Minisci Arylation: **No photocatalyst** (ELT 716)

# of matched bases (out of 71)	% of total sequences (n = 2516811)
71	**85.40**
70	12.13
69	1.53
68	0.37
67	0.16
66	0.12
65	0.07
64	0.05
63	0.04
62	0.03
61	0.02
57	0.01
60	0.01
58	0.01
56	0.01
59	0.01
55	0.01
54	0.01
53	< 0.01
52	< 0.01
50	< 0.01
51	< 0.01
47	< 0.01
46	< 0.01
49	< 0.01
48	< 0.01
44	< 0.01
45	< 0.01
24	< 0.01

71 base sequence

5’ – GTA GAC CGG GCA TCG TAA CGA CCT TTC TCC GCT TAC CAG GCT TCT TGC GGA ACA CAT CGA TTT GGG AGT CA – 3’

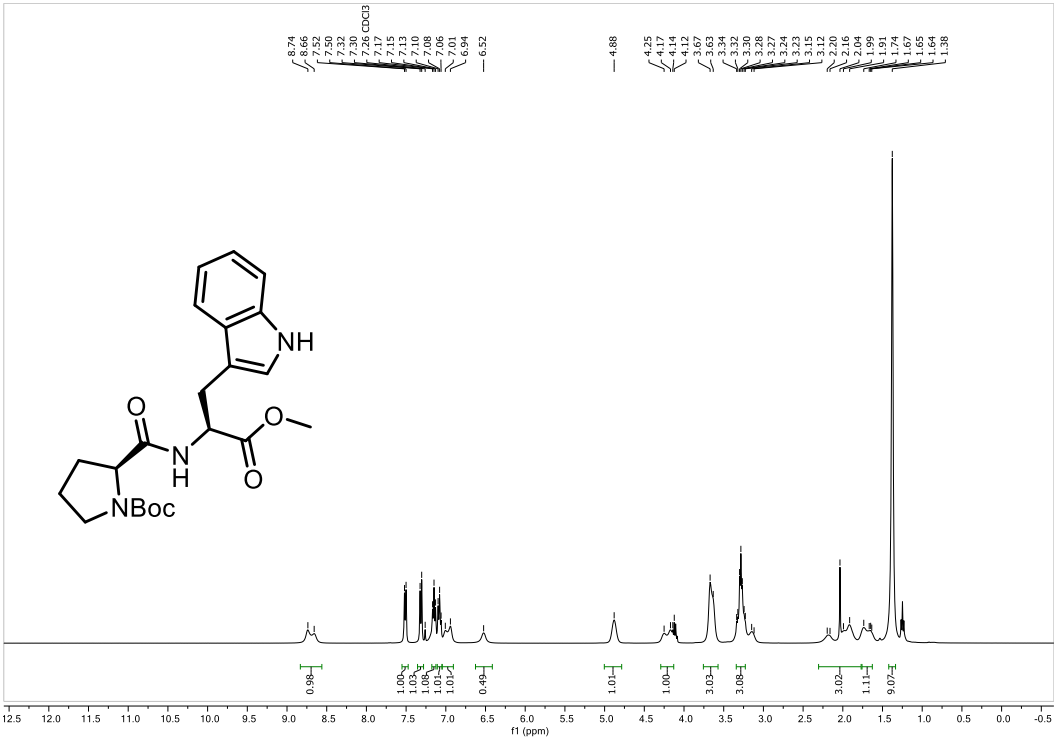
Minisci Arylation: **No-light control** (ELT 717)

# of matched bases (out of 71)	% of total sequences (n = 2641537)
71	**85.42**
70	12.15
69	1.50
68	0.36
67	0.16
66	0.11
65	0.08
64	0.05
63	0.04
62	0.03
61	0.02
60	0.02
56	0.01
59	0.01
55	0.01
57	0.01
58	0.01
54	0.01
53	< 0.01
52	< 0.01
51	< 0.01
50	< 0.01
47	< 0.01
49	< 0.01
46	< 0.01
45	< 0.01
44	< 0.01
43	< 0.01
42	< 0.01
40	< 0.01
34	< 0.01

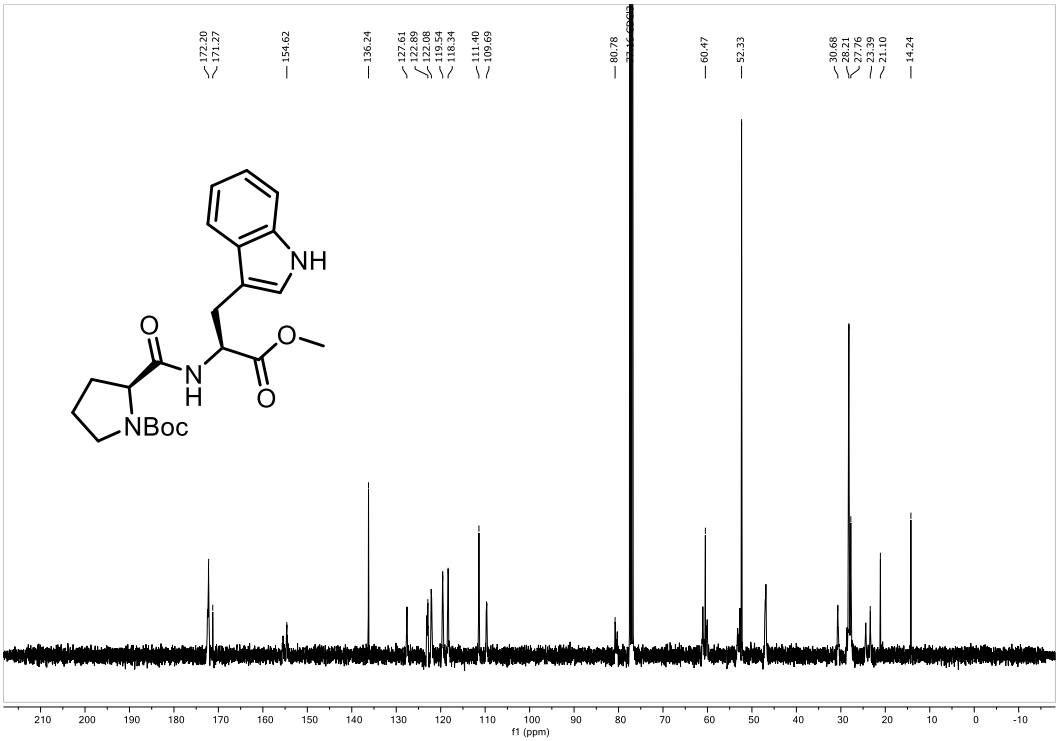
7. References

1. L. A. Carpino, *J. Am. Chem. Soc.,* 1993, **115**, 4397–4398.

8. NMR Spectra



^{1}H NMR (400 MHz, $CDCl_3$) of compound **2z**.



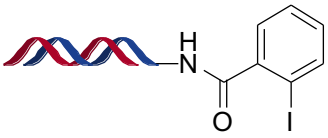
^{13}C NMR (101 MHz, $CDCl_3$) of compound **2z**.

9. UPLC/MS Spectra

Functionalized DNA headpieces

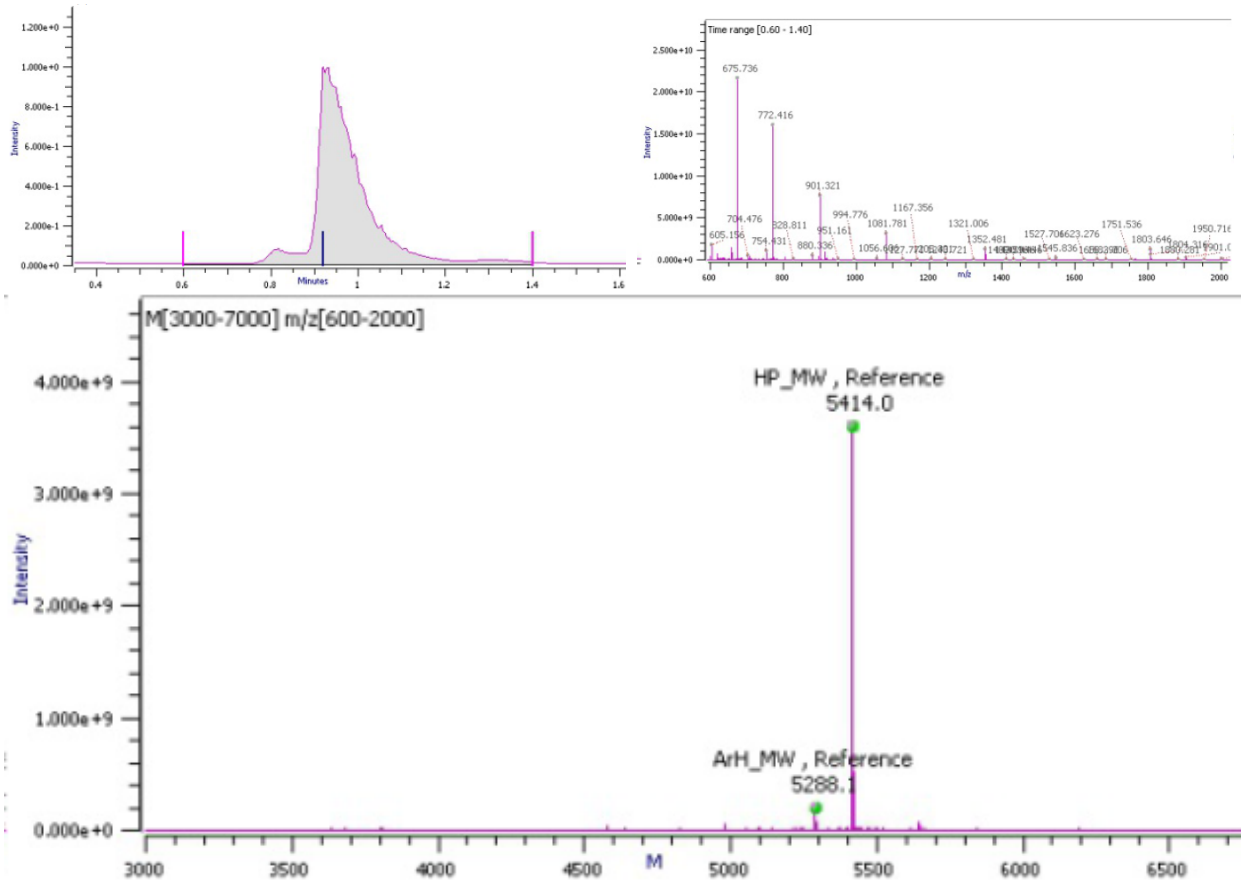
The synthesis of on-DNA aryl halides **1A-L** was previously reported in *J. Am. Chem. Soc.,* 2019, **141**, 3723–3732; *Org. Lett.,* 2020, **22**, 1046–105; and S. O. Badir and G. A. Molander et al., *Chem. Sci.,* 2021, **12**, 12036–12045.

Headpiece 1D

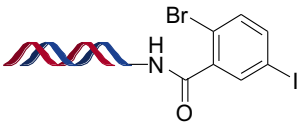


Molecular Weight: 5414.53

ELTM_423, TW-N78441-9-423 - (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)

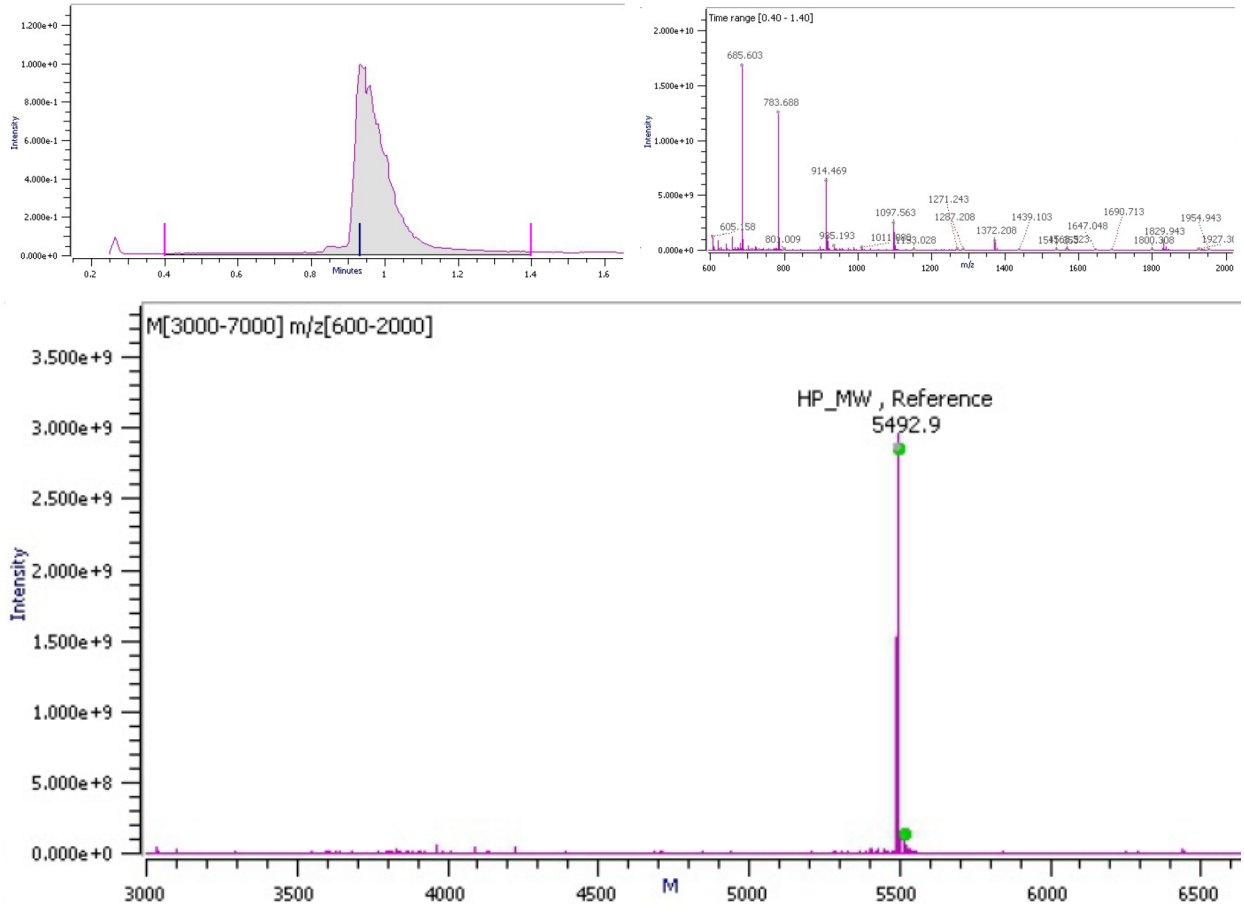


Headpiece 1J



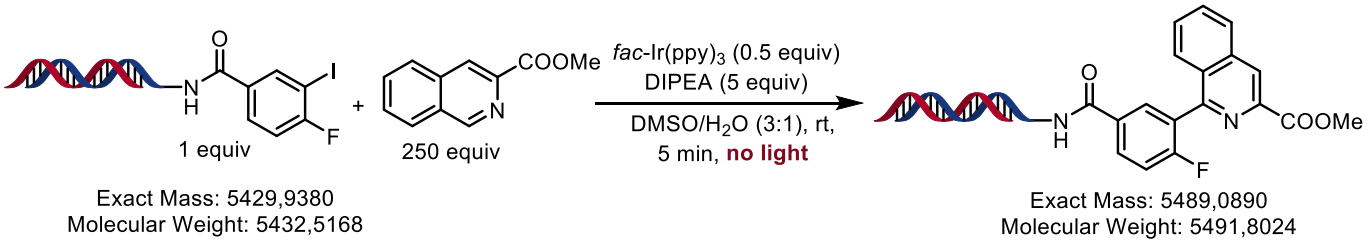
Molecular Weight: 5493.42

ELTM_503, TW-N78441-10-503 (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)

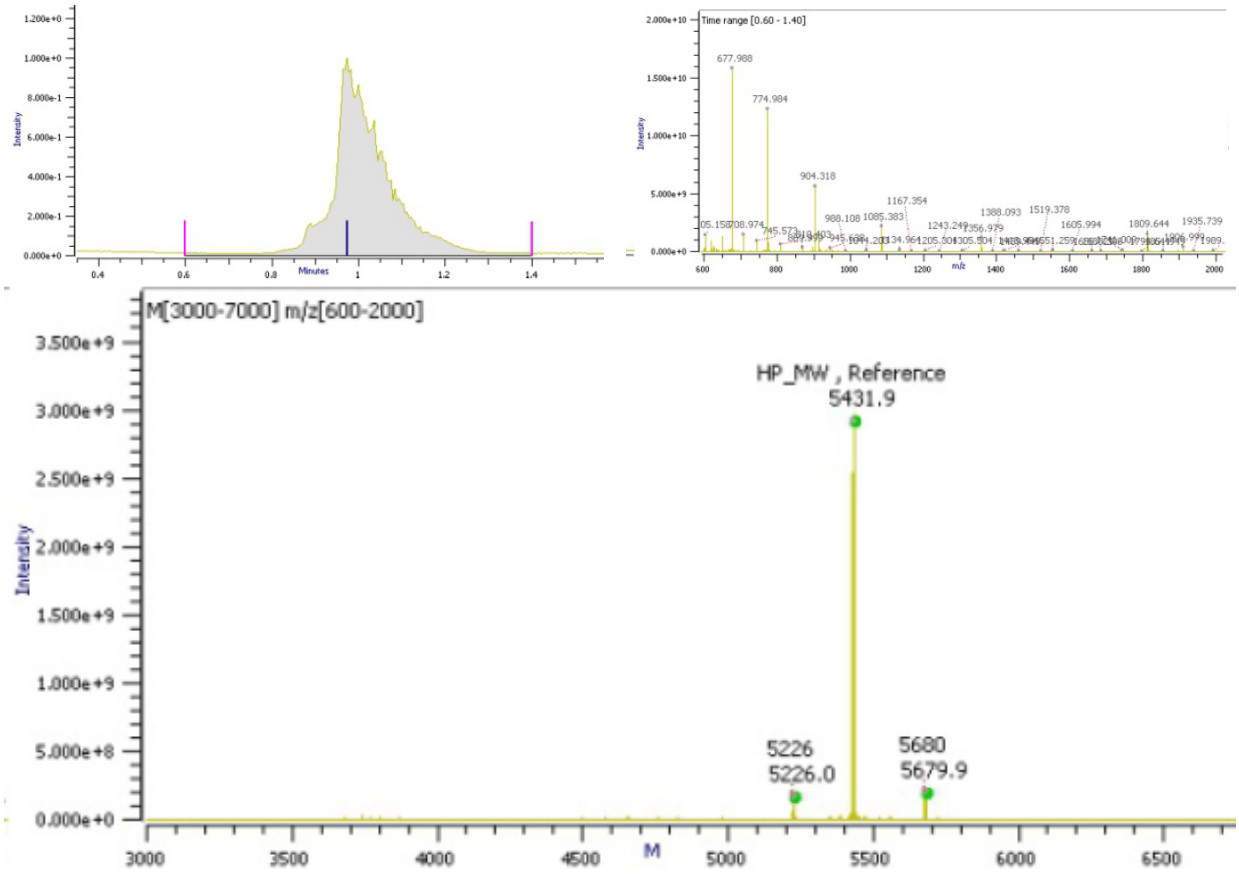


Control experiments for On-DNA reactions

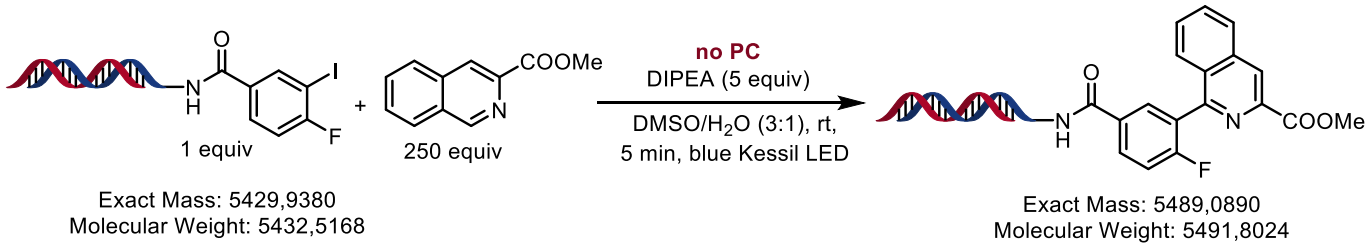
No light: No product formation (**4a**) is detected



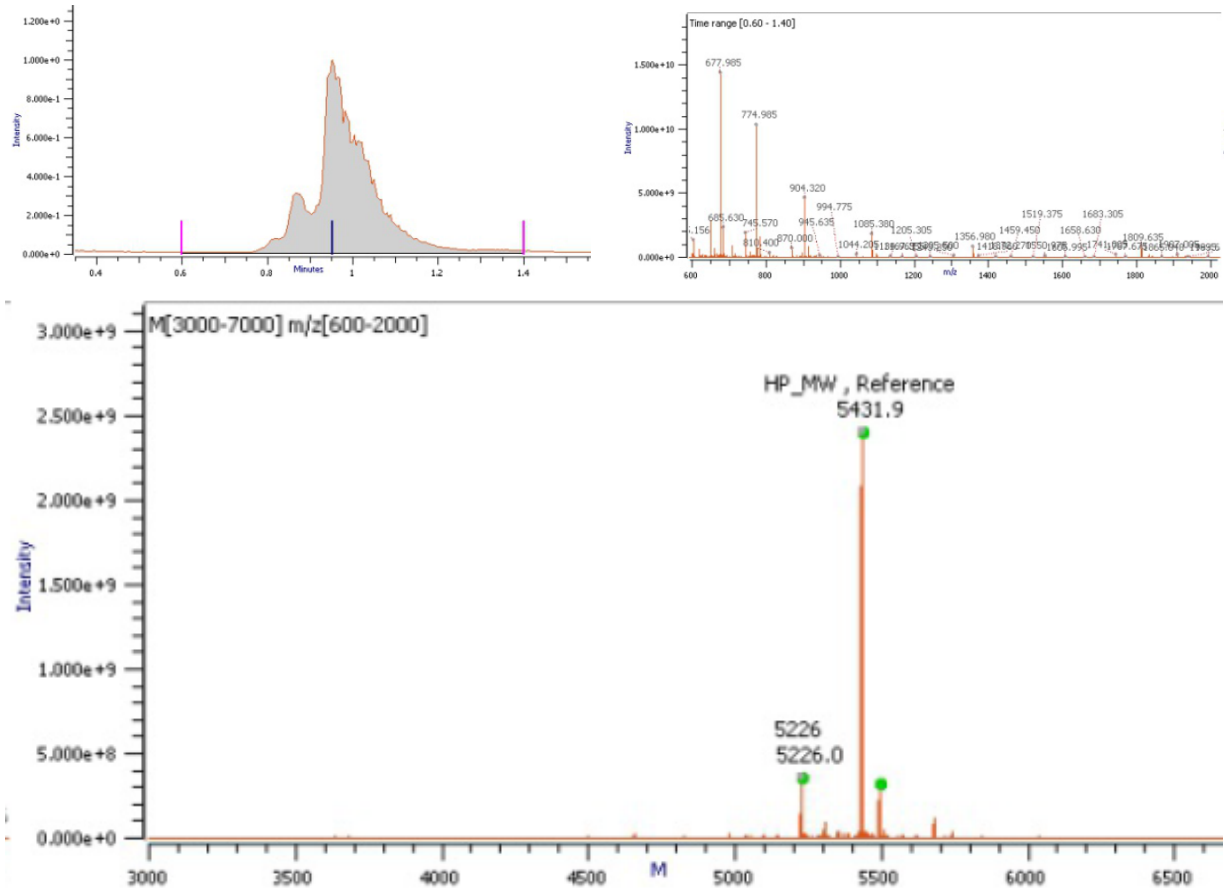
ELTM_422, TW-N78441-9-422 (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



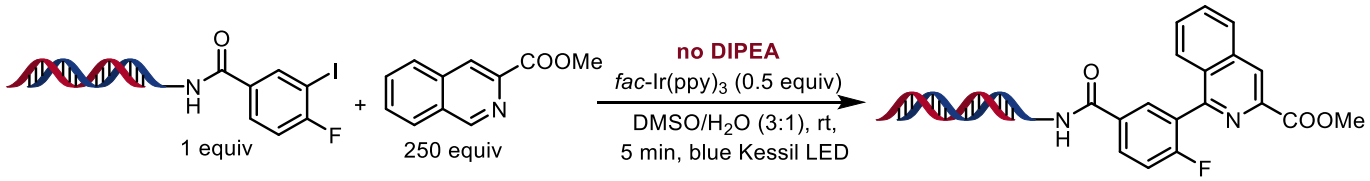
No *fac*-Ir(ppy)$_3$: No product formation (**4a**) is detected



ELTM_414, TW-N78441-9-414 (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



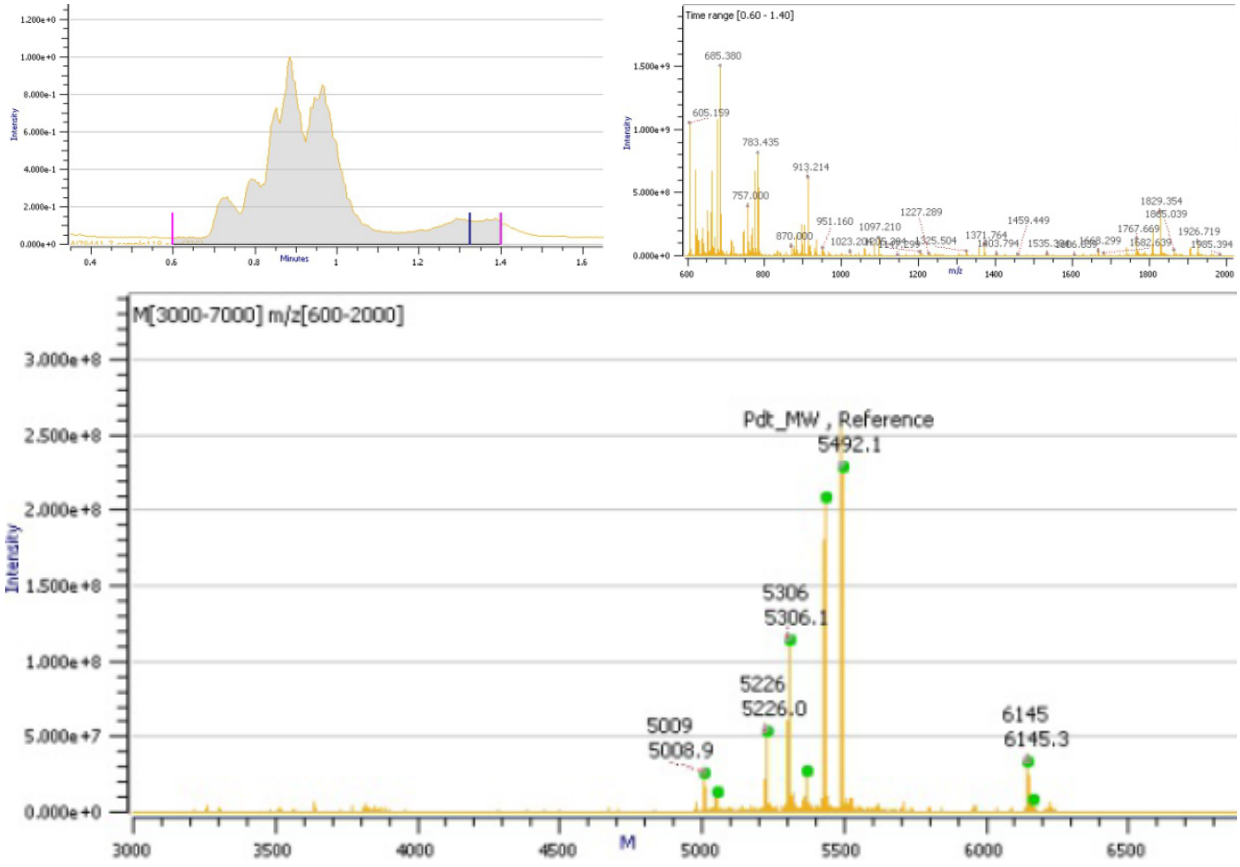
No DIPEA: Product **4a**, 34% yield



Exact Mass: 5429,9380
Molecular Weight: 5432,5168

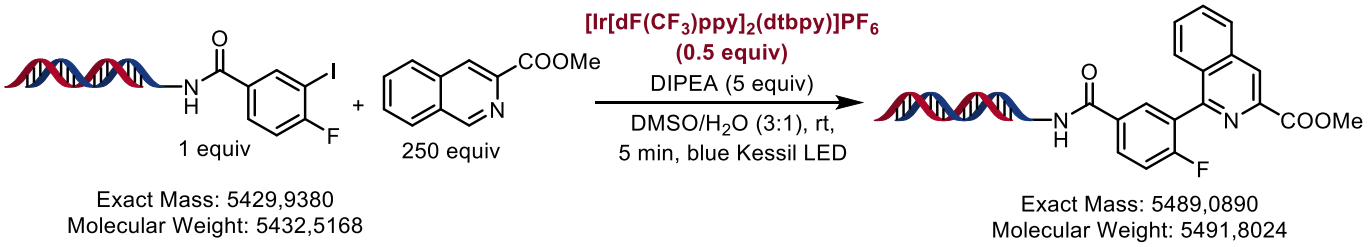
Exact Mass: 5489,0890
Molecular Weight: 5491,8024

ELTM_118, N78441-7-118 (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)

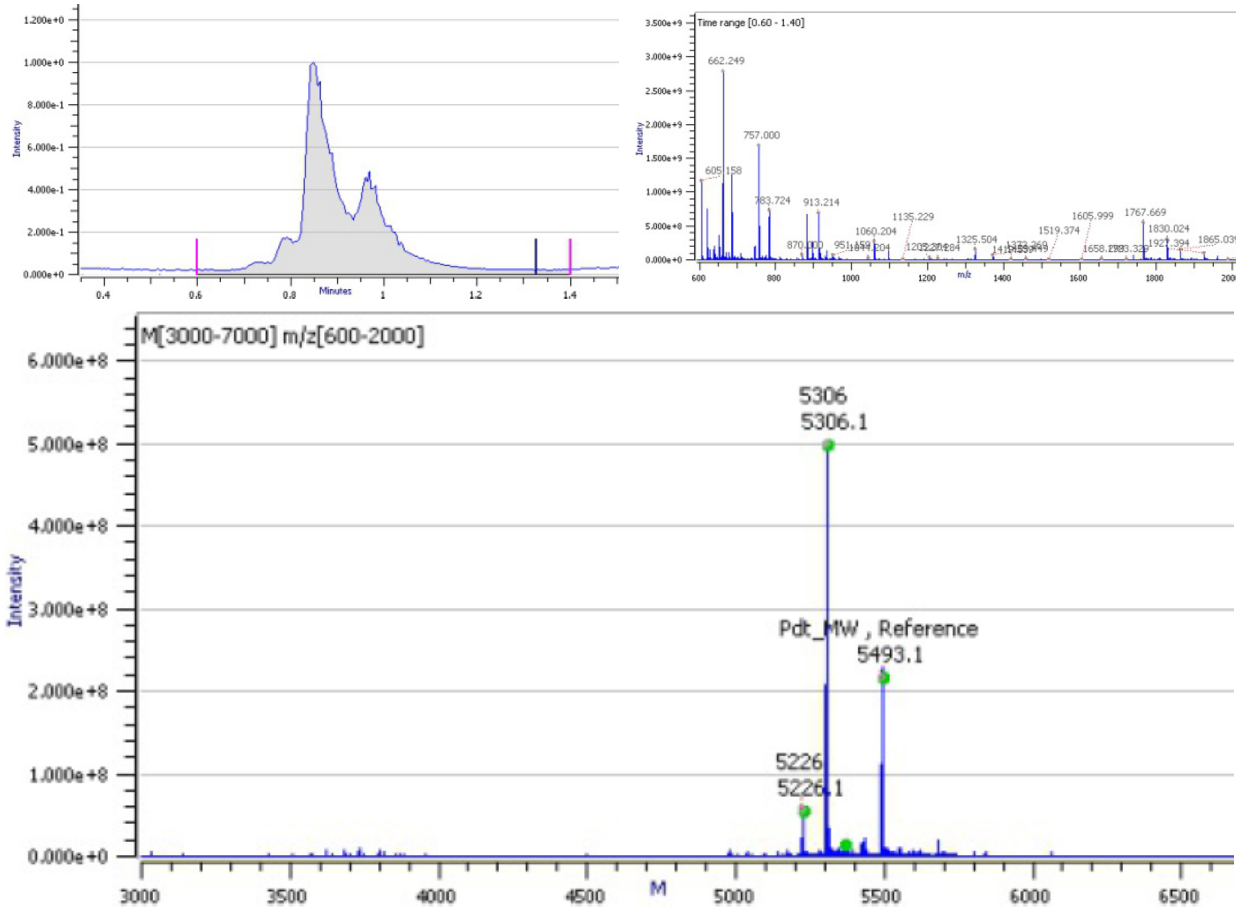


Optimization studies: Catalyst screen

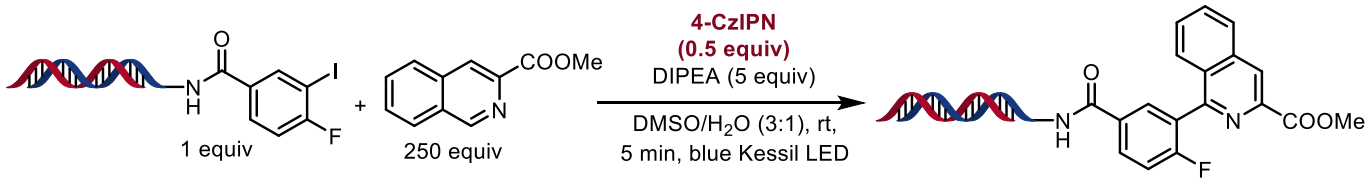
[Ir[dF(CF$_3$)ppy]$_2$(dtbpy)]PF$_6$: Product **4a**, 30%



ELTM_98, TW-N78441-7-98 (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



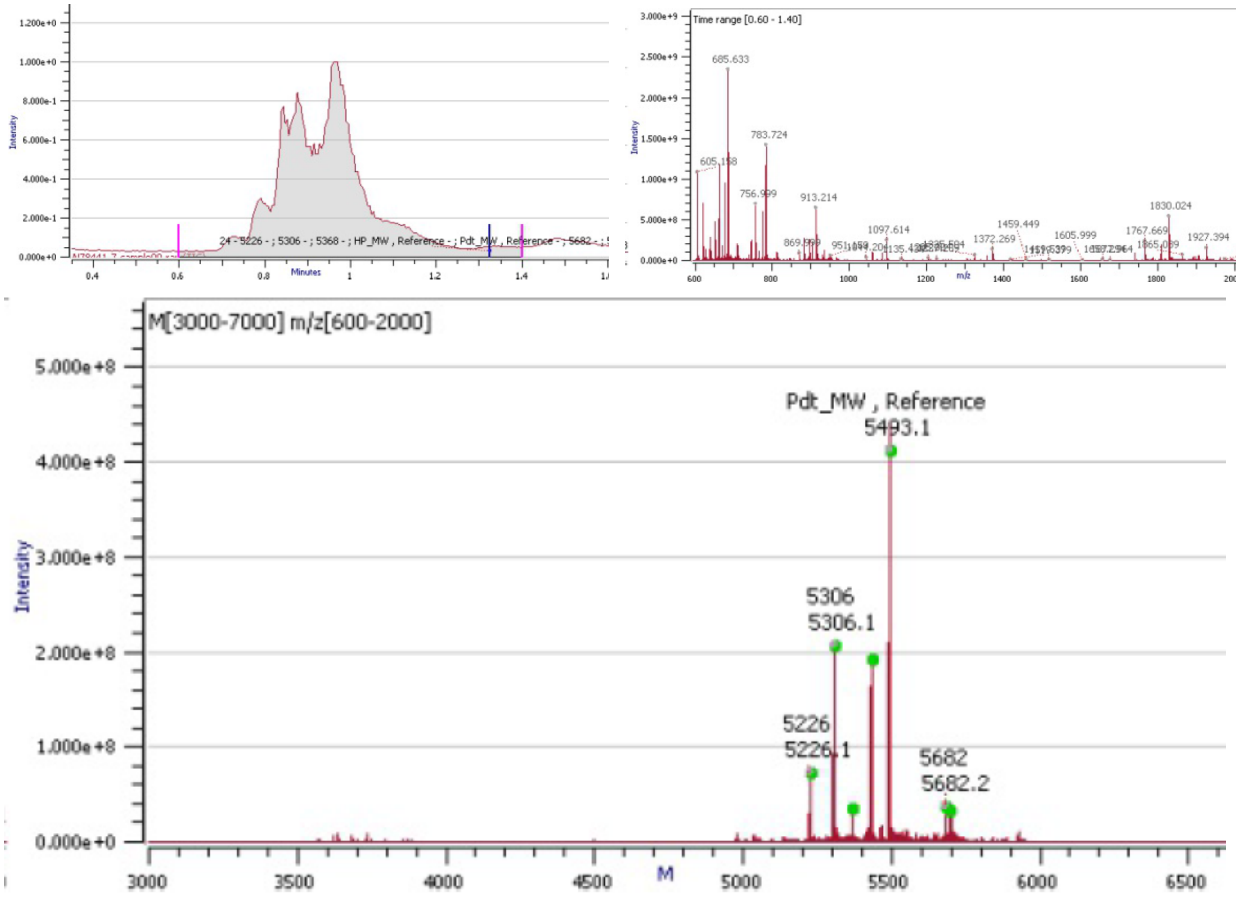
4-CzIPN: Product **4a**, 45%



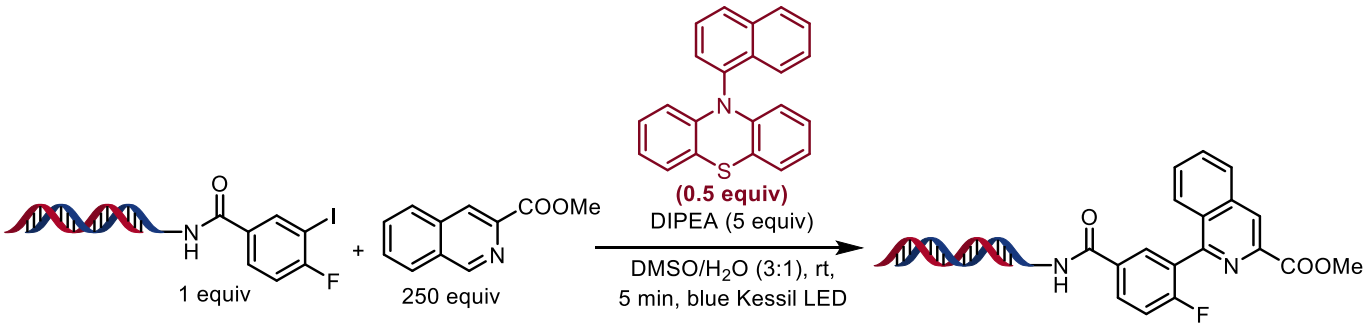
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5489,0890
Molecular Weight: 5491,8024

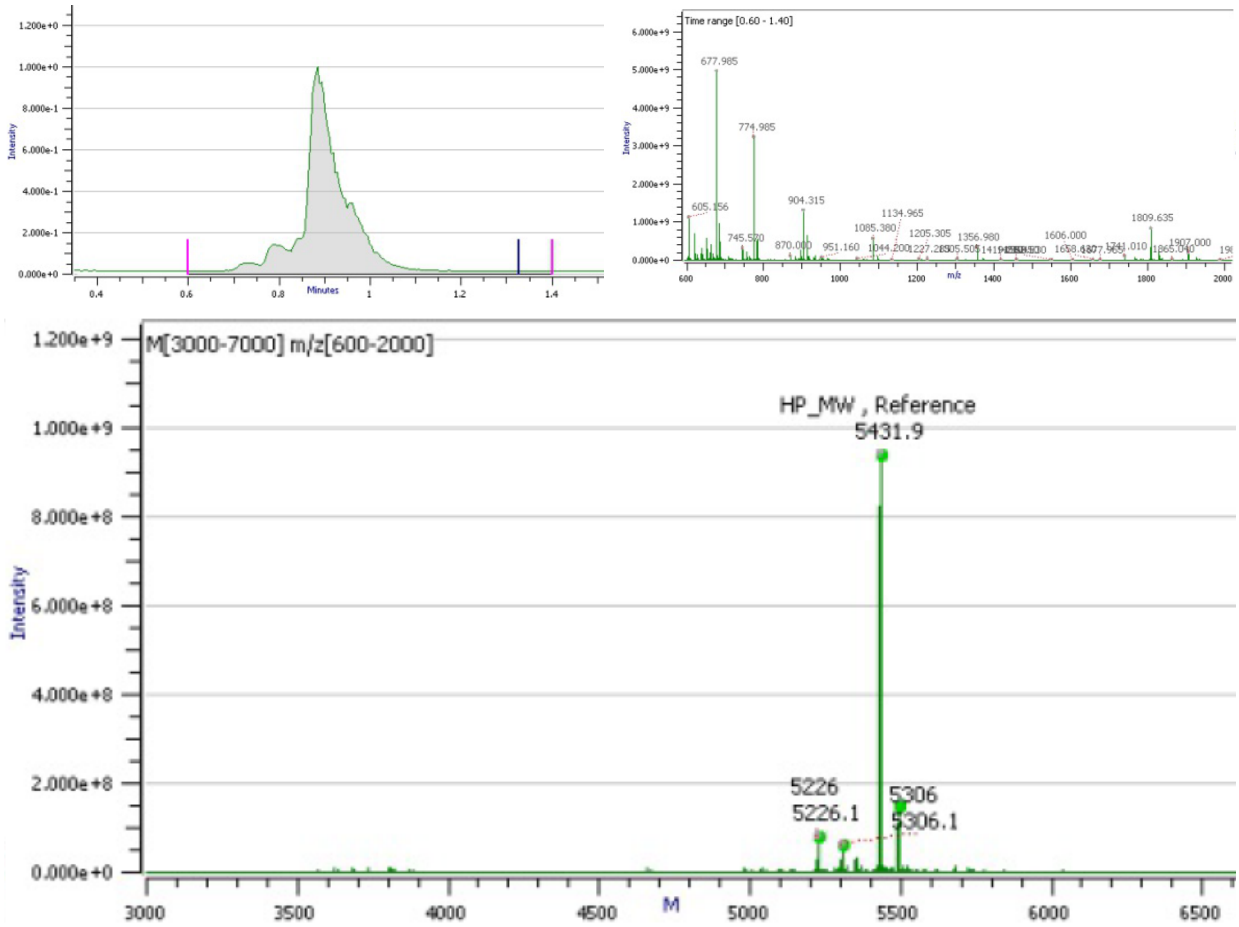
ELTM_99, TW-N78441-7-99 (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



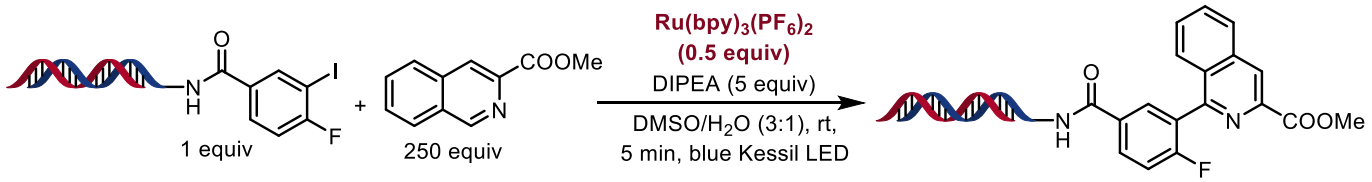
Naphthyl-phenothiazine: Product **4a**, 13%



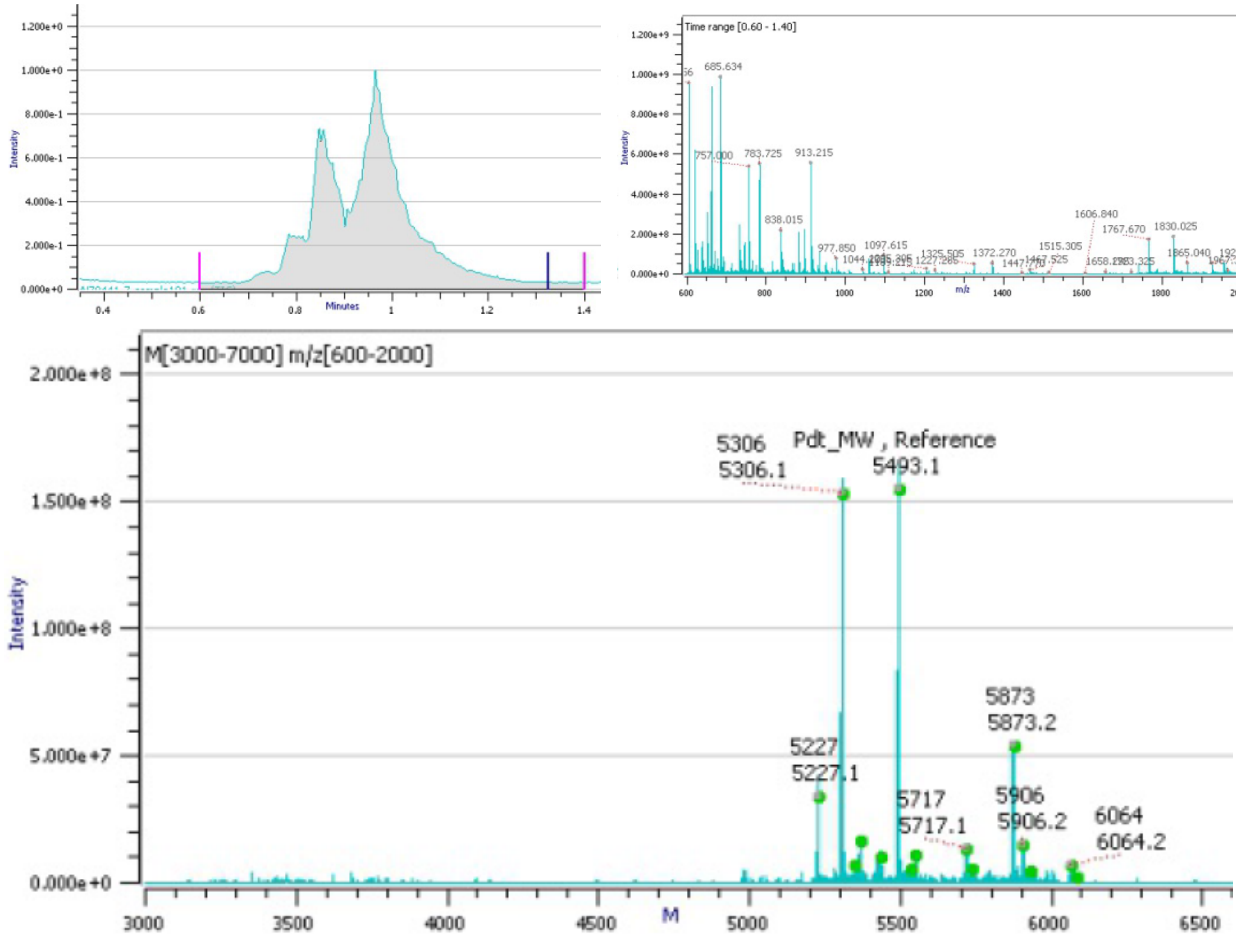
ELTM_100, TW-N78441-7-100 (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



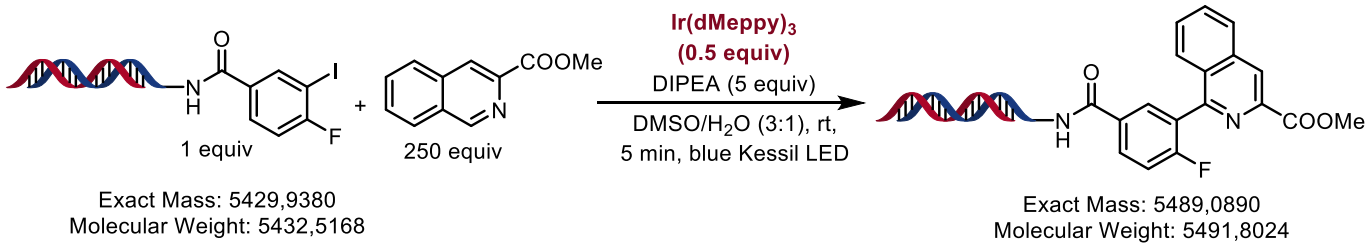
$Ru(bpy)_3(PF_6)_2$: Product **4a**, 34%



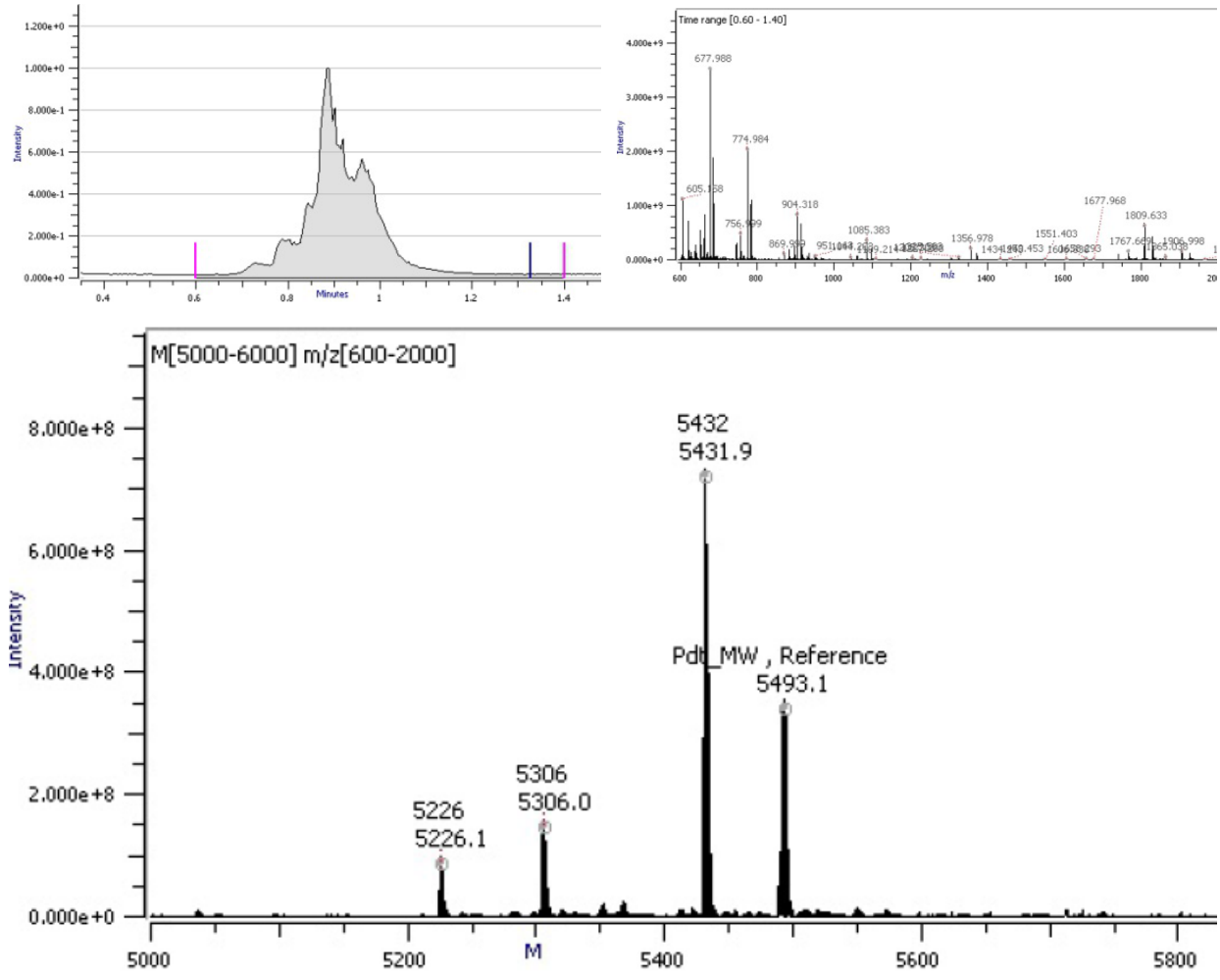
ELTM_101, TW-N78441-7-101 (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



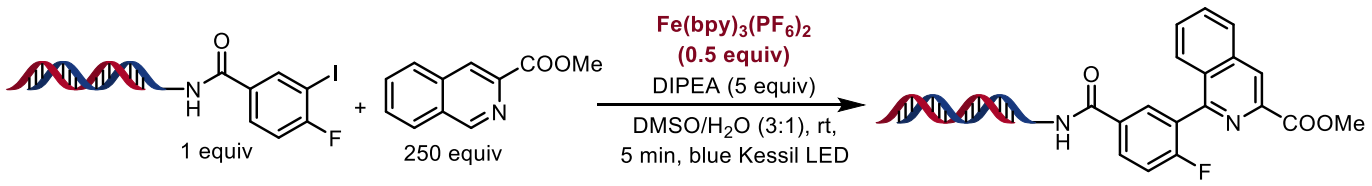
Ir(dMeppy)$_3$: Product **4a**, 28%



ELTM_102, TW-N78441-7-102 (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



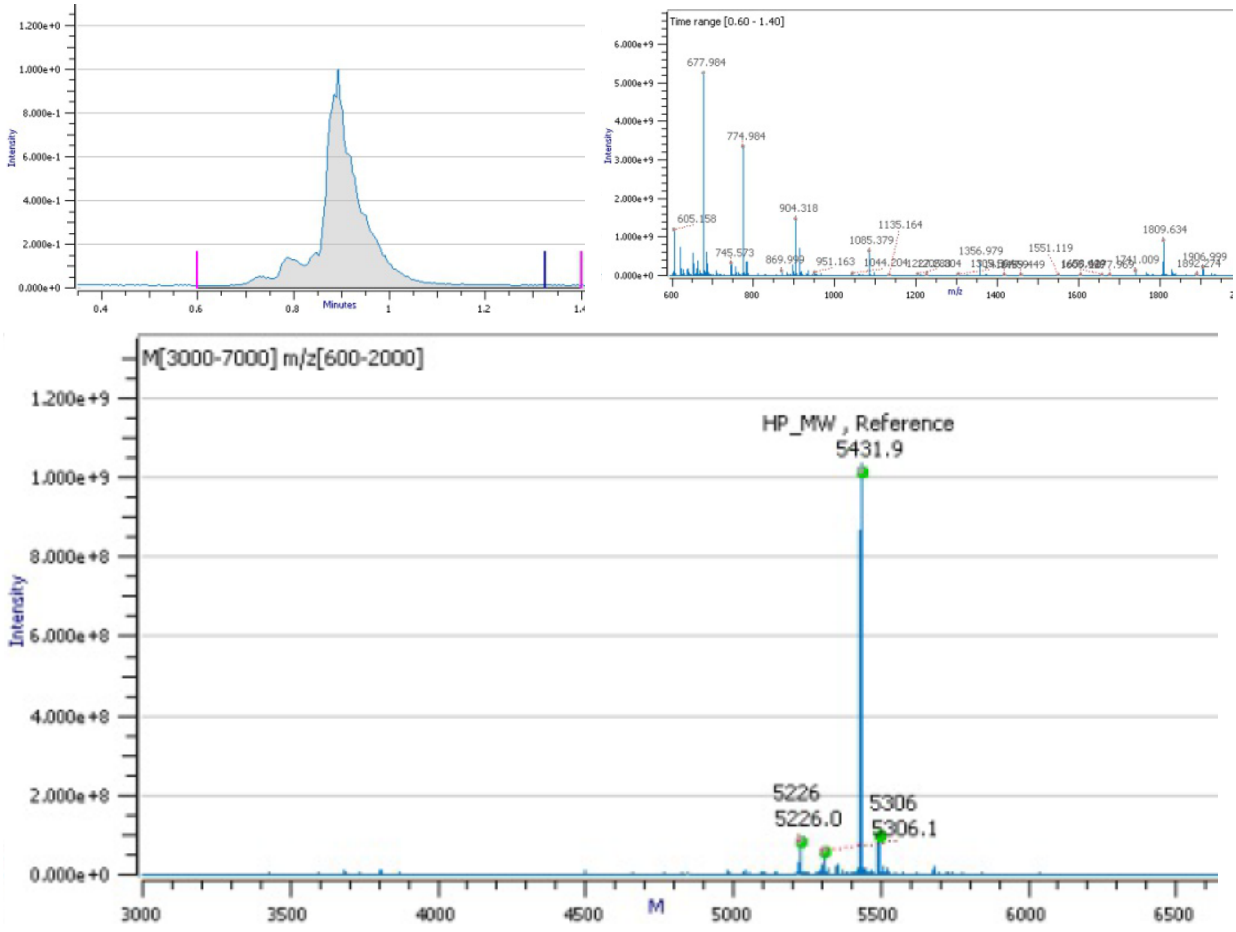
$Fe(bpy)_3(PF_6)_2$: Product **4a**, 8%



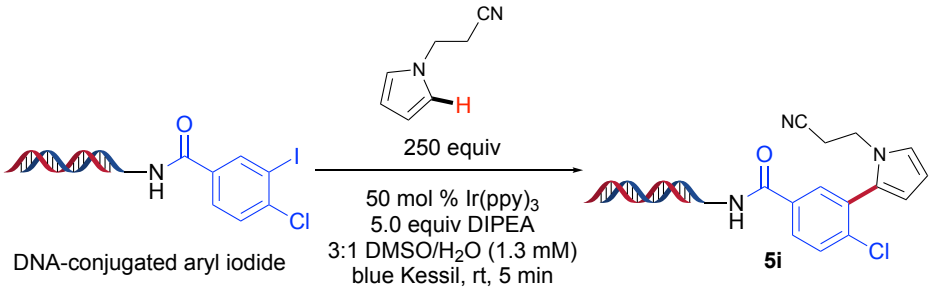
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

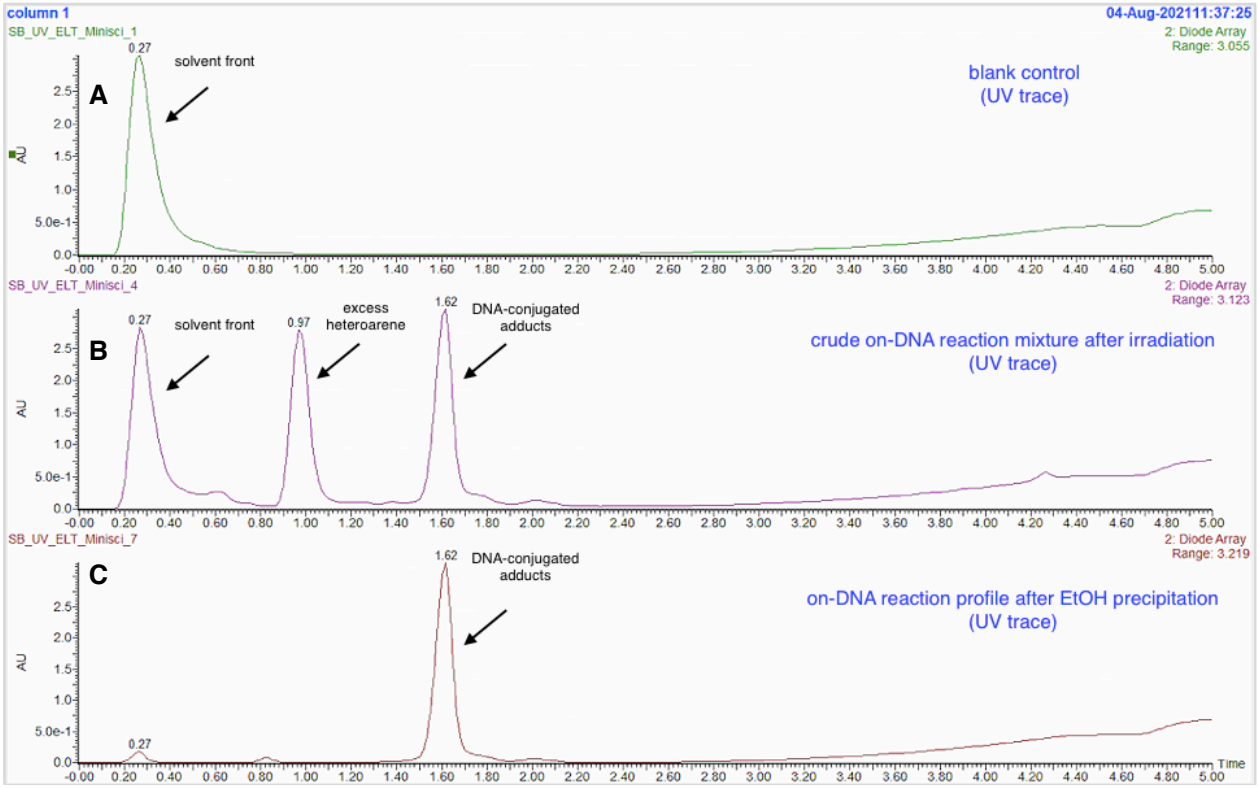
Exact Mass: 5489,0890
Molecular Weight: 5491,8024

ELTM_103, TW-N78441-7-103 (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



*UV trace profile for exemplary substrate **5i** before and after ethanol precipitation:*



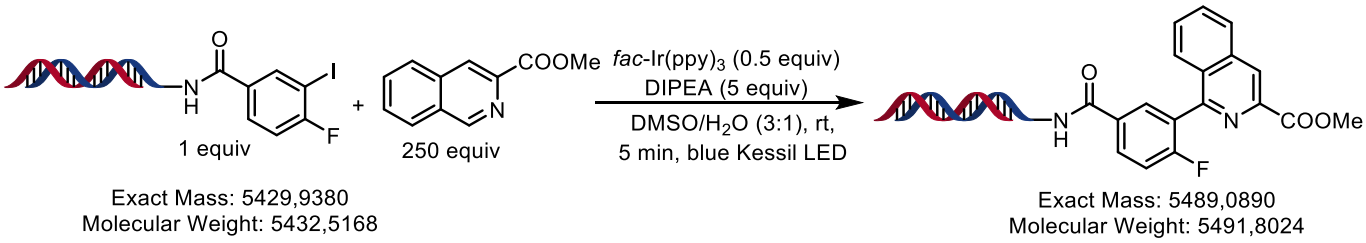


As demonstrated from UV traces **B** & **C** for exemplary substrate **5i**, the organic byproducts can be removed by standard DEL purification procedures (following the ethanol precipitation protocol described in Section 2.2). Furthermore, the exemplary sample submitted to EtOH precipitation has a UV chromatogram that is consonant with that of the DNA-adducts post purification.

Determination of yields for On-DNA reactions

Variation of heteroarenes

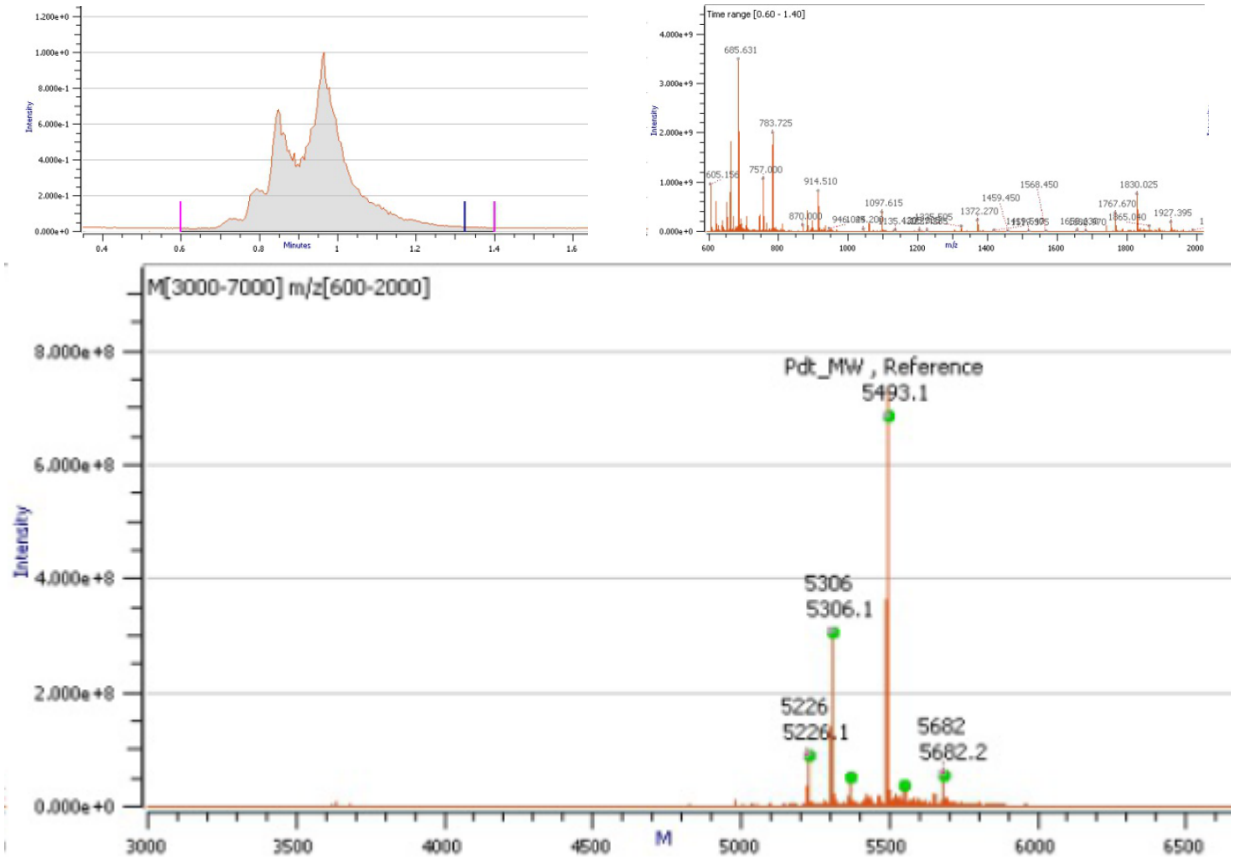
Product **4a**, 60% yield



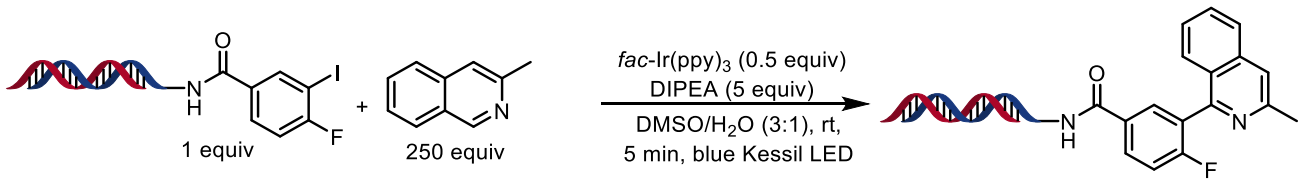
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5489,0890
Molecular Weight: 5491,8024

ELTM_117, TW-N78441-7-117.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



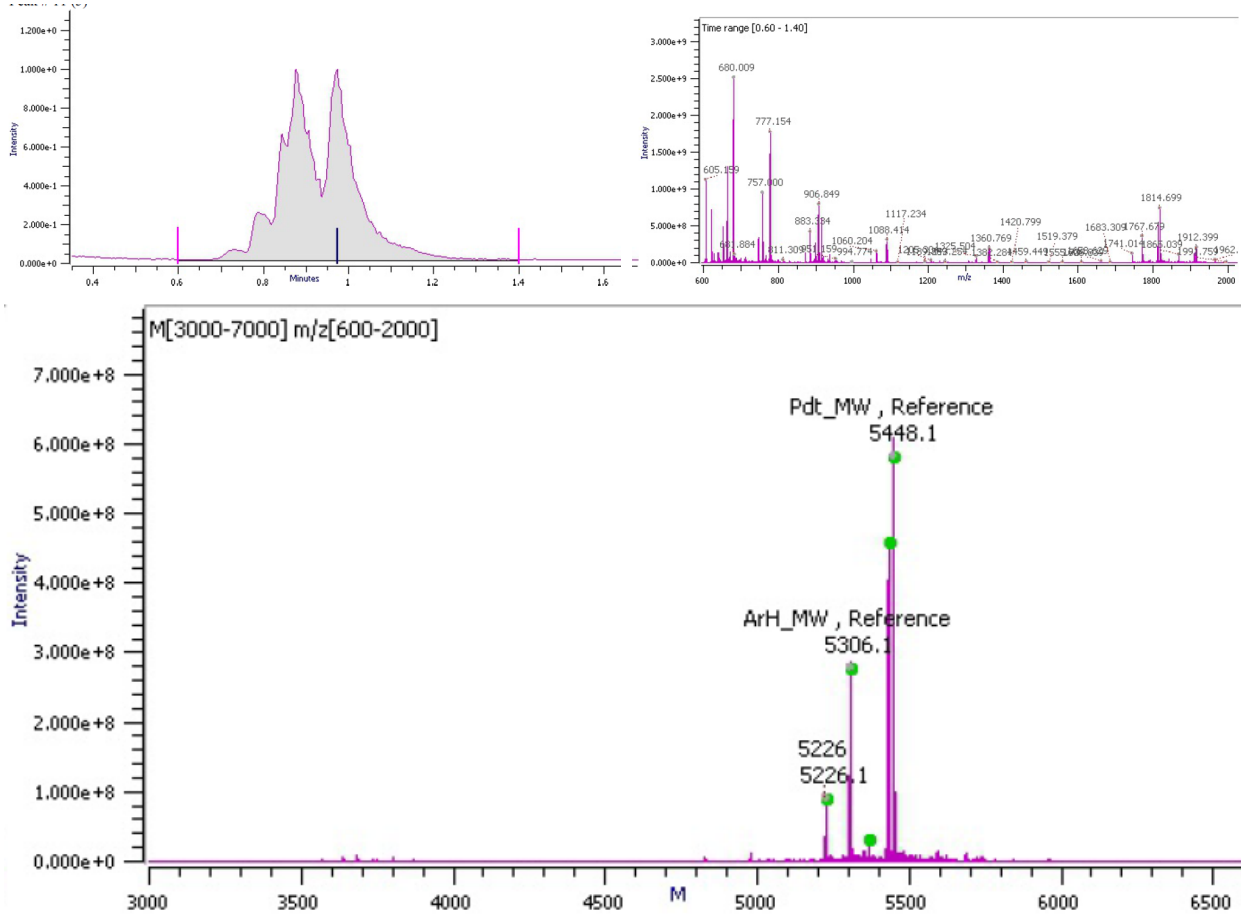
Product **4b**, 44% yield



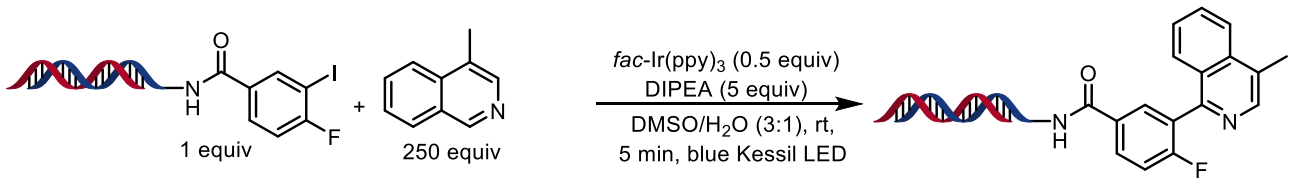
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5445,0992
Molecular Weight: 5447,7934

ELTM_206, TW-N78441-8-206.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



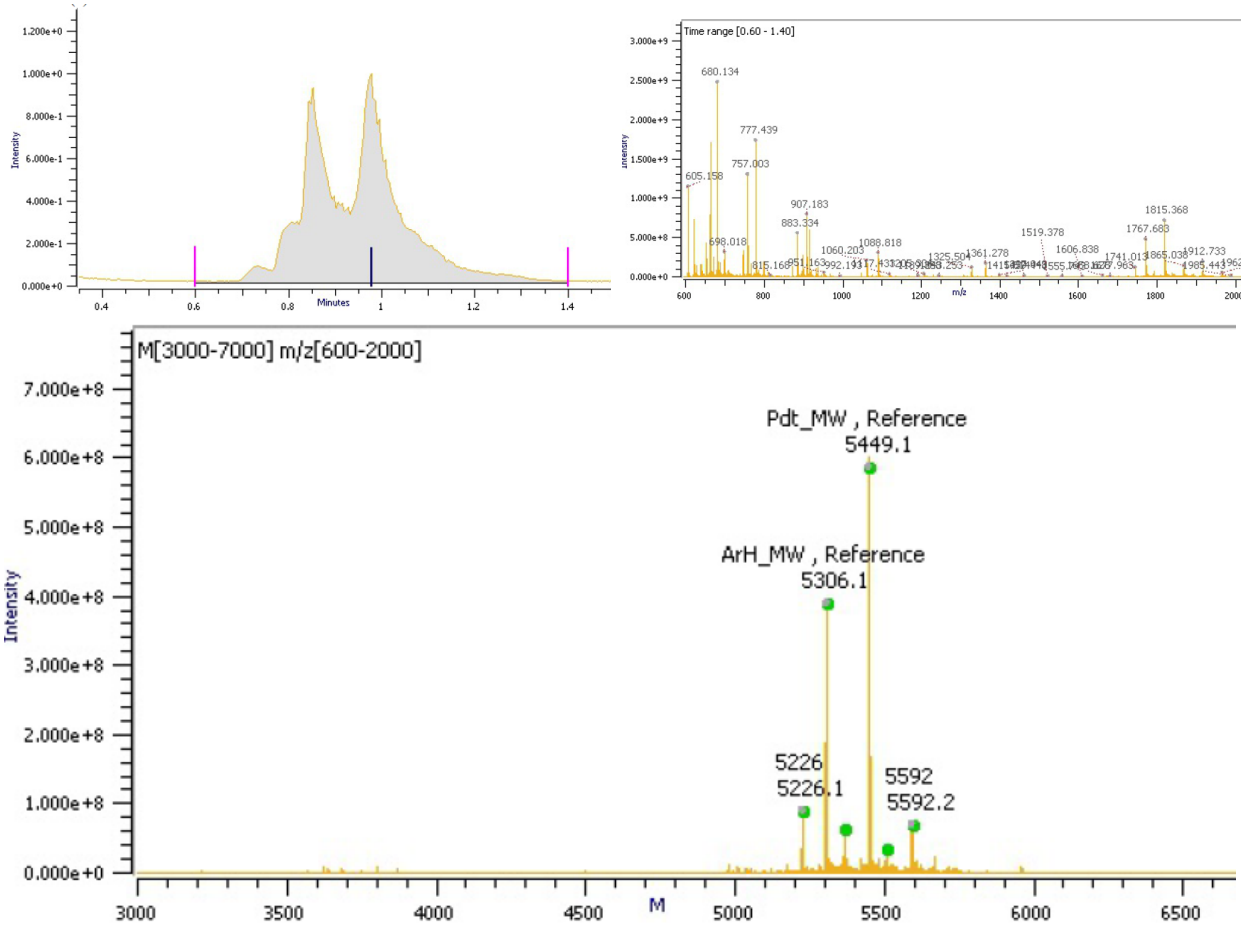
Product **4c**, 52% yield



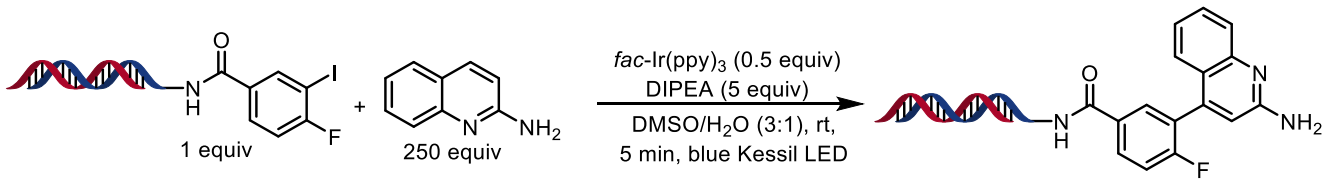
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5445,0992
Molecular Weight: 5447,7934

ELTM_167, TW-N78441-8-167.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



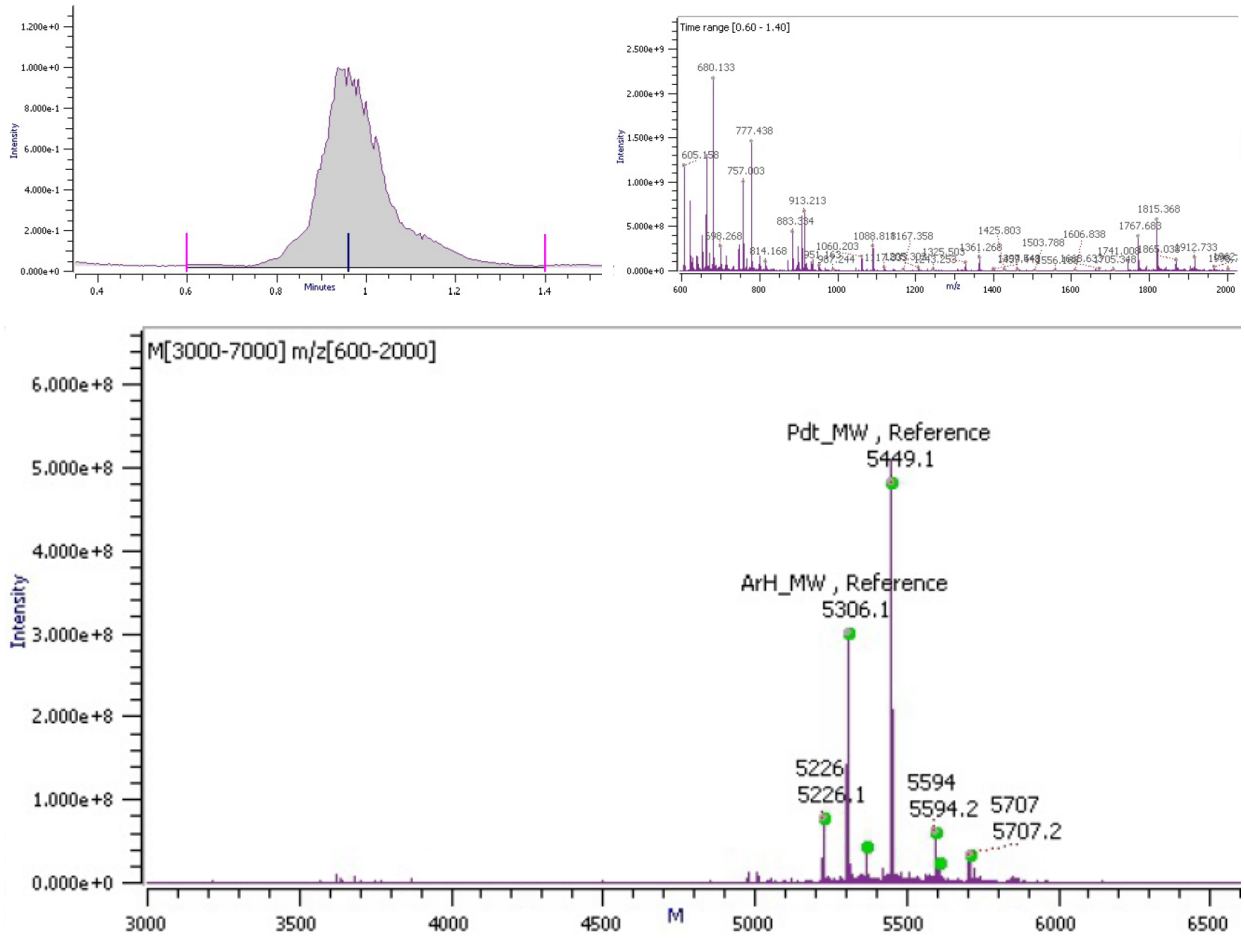
Product **4d**, 52% yield



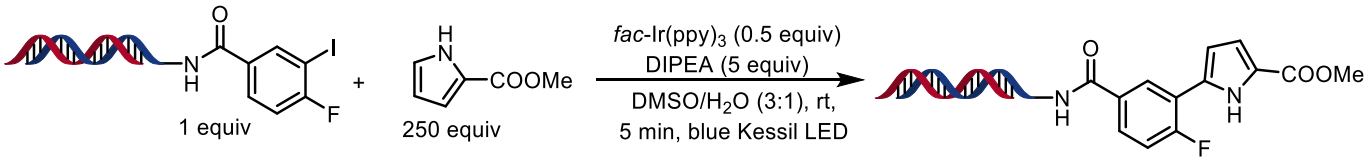
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5446,0944
Molecular Weight: 5448,7814

ELTM_171, TW-N78441-8-171.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



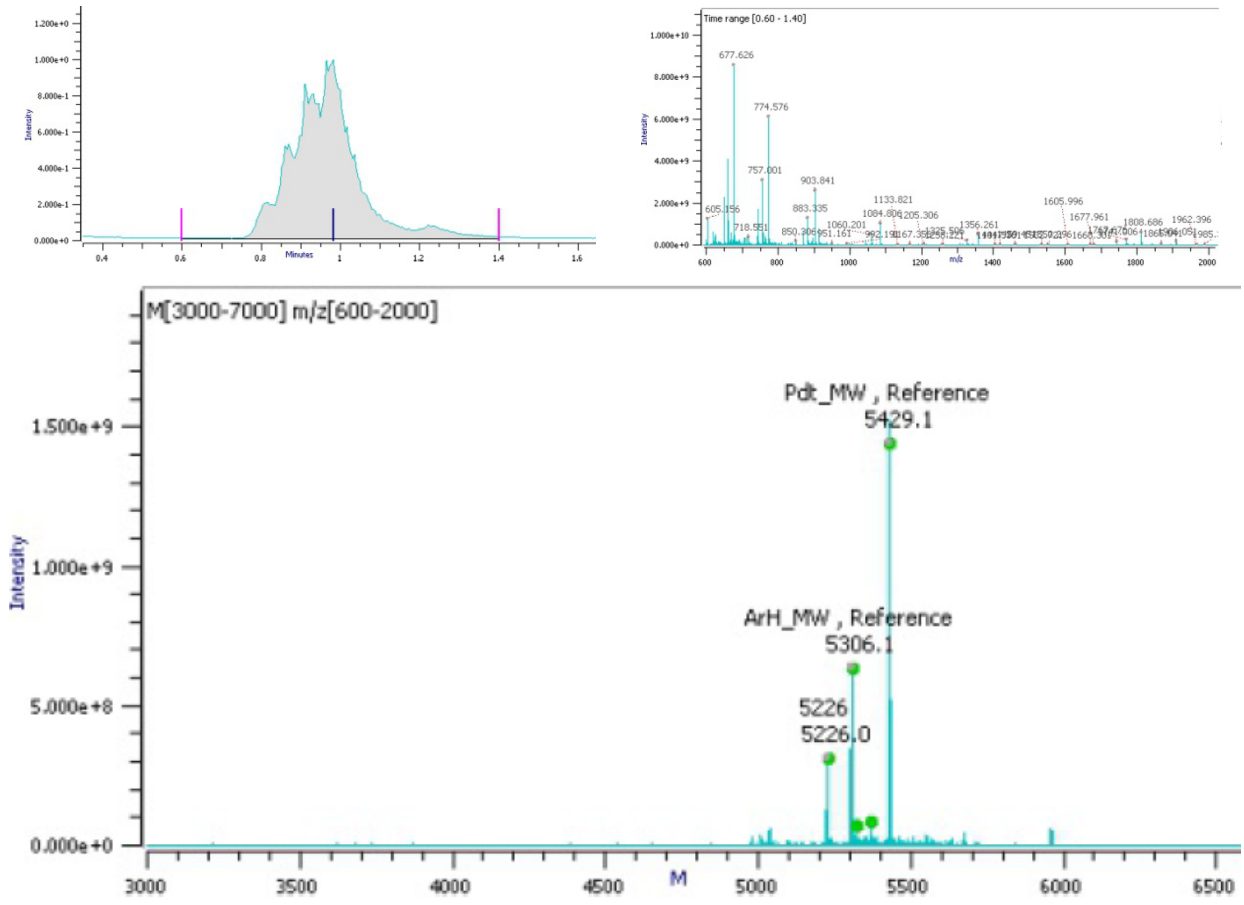
Product **4e**, 67% yield



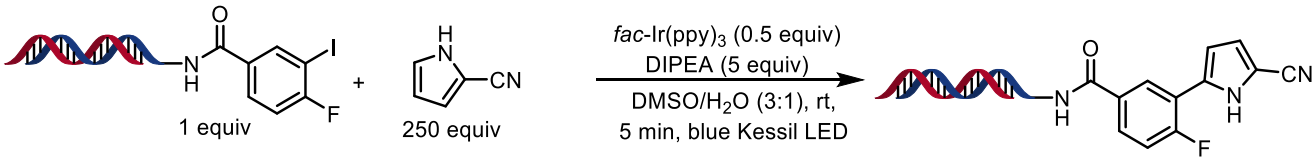
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5427,0733
Molecular Weight: 5429,7314

ELTM_370, TW-N78441-9-370.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



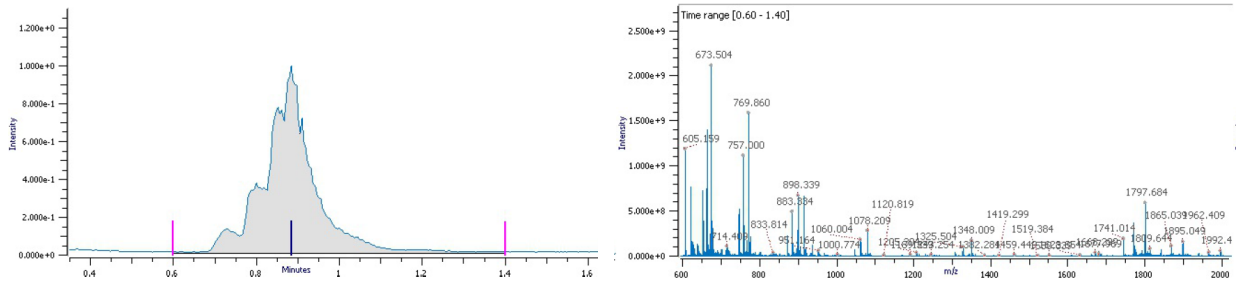
Product **4f**, 51% yield

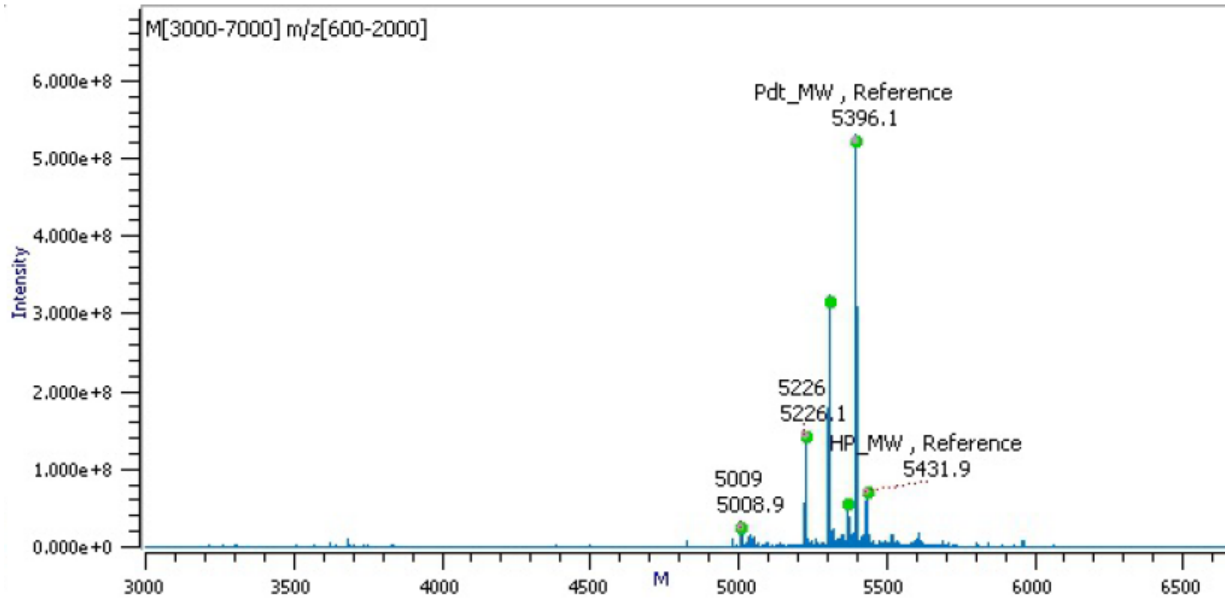


Exact Mass: 5429,9380
Molecular Weight: 5432,5168

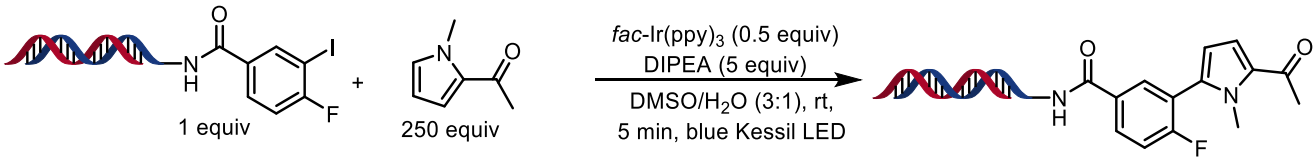
Exact Mass: 5394,0631
Molecular Weight: 5396,7054

ELTM_224, TW-N78441-8-224.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)





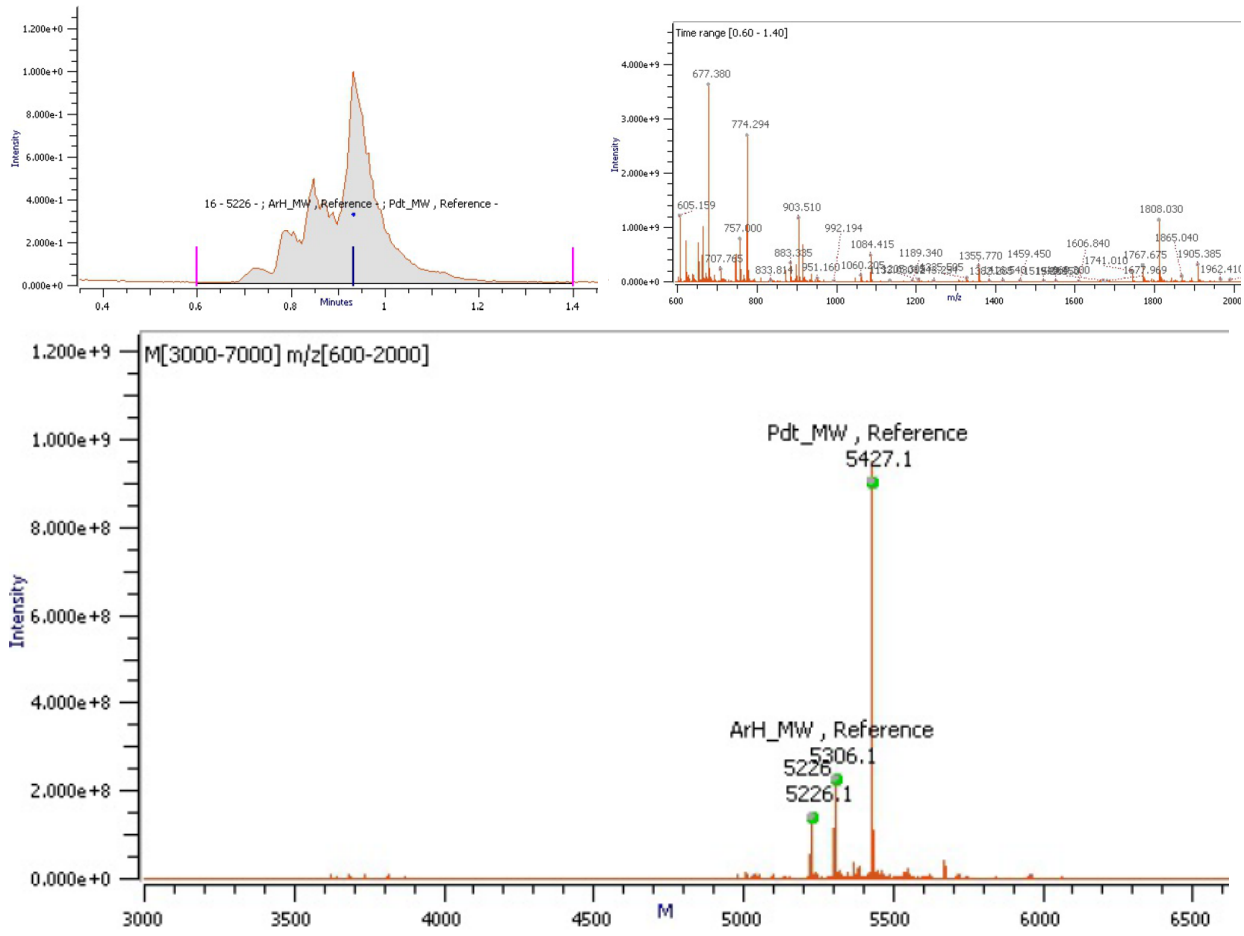
Product **4g**, 78% yield



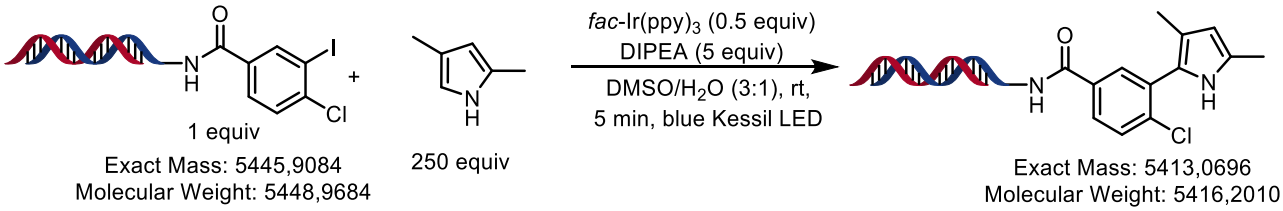
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5425,0941
Molecular Weight: 5427,7594

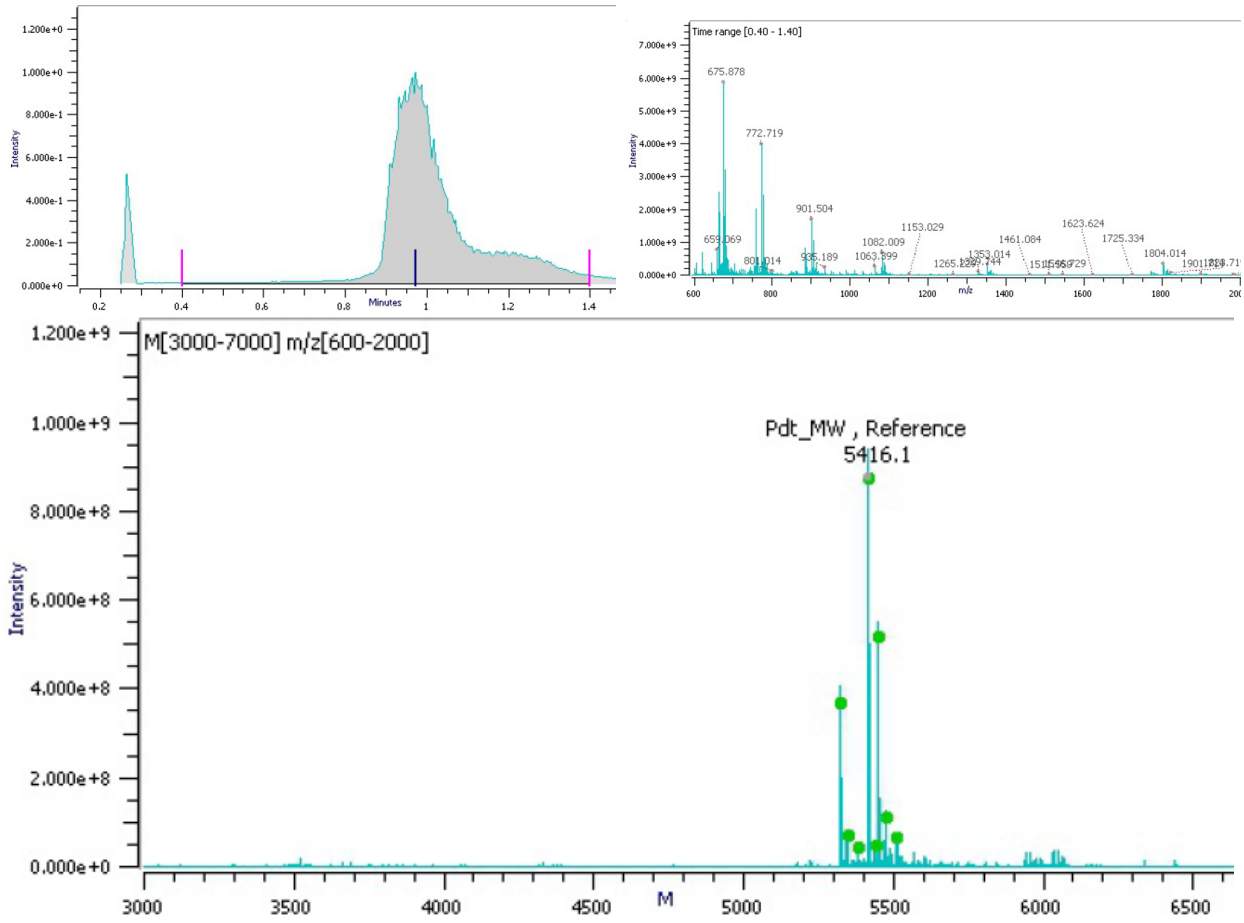
ELTM_228, TW-N78441-8-228.raw - (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



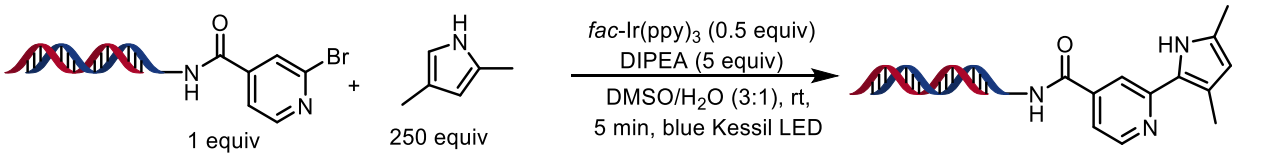
Product **5h** 48% yield



ELTM_544, TW-N78441-10-544.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



Product **6h**, 67% yield



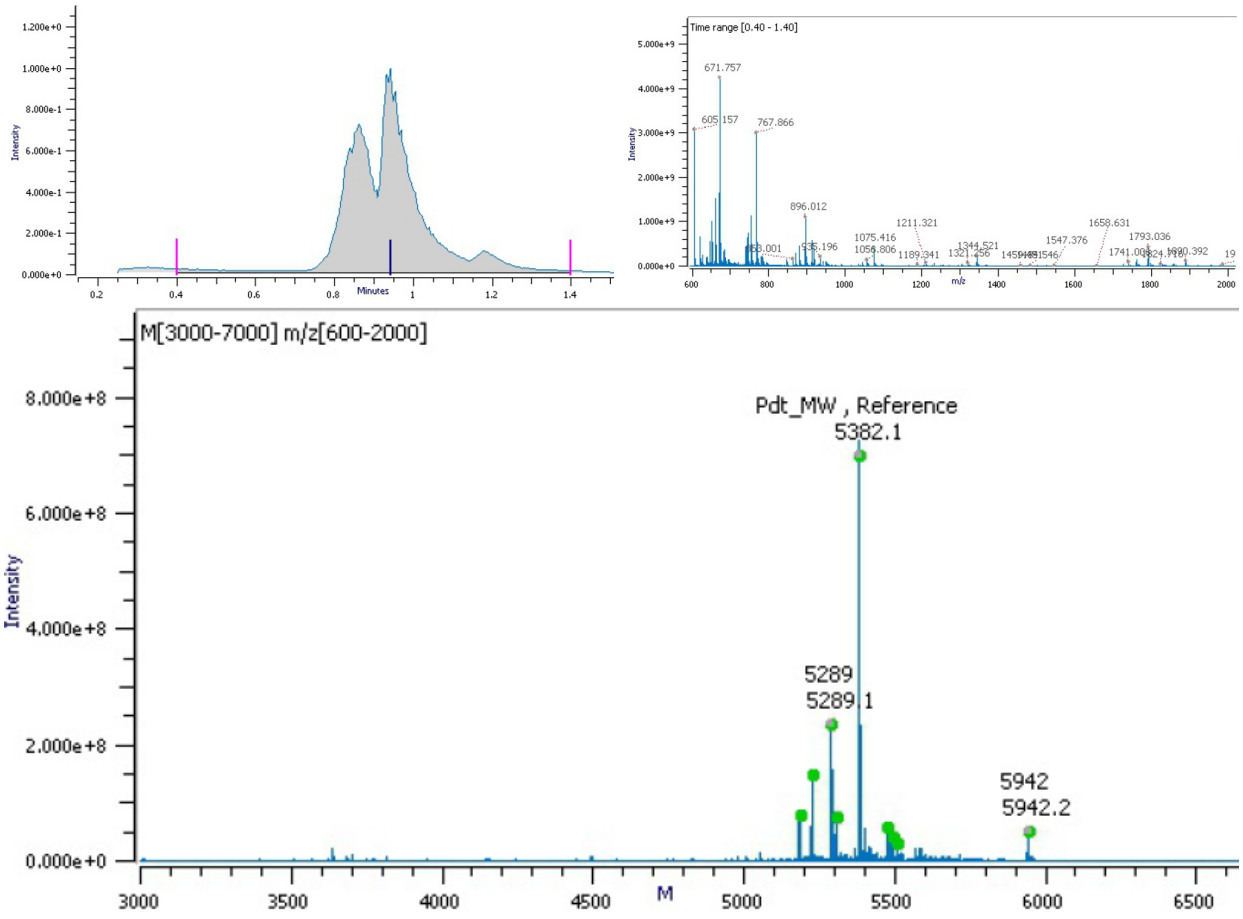
1 equiv

Exact Mass: 5364,9565
Molecular Weight: 5368,5140

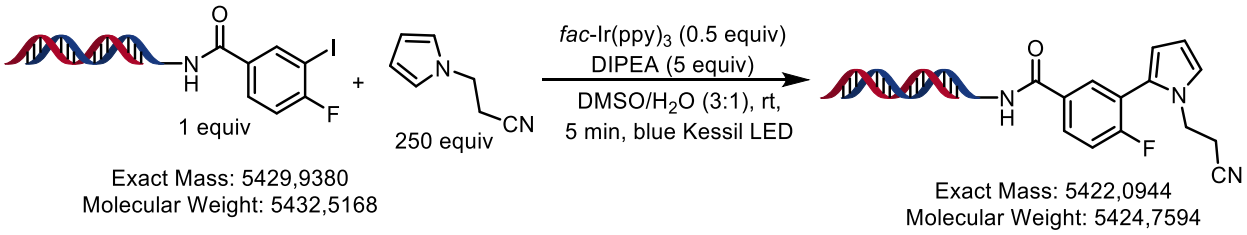
250 equiv

Exact Mass: 5380.1038
Molecular Weight: 5382,7470

ELTM_481, TW-N78441-10-481.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



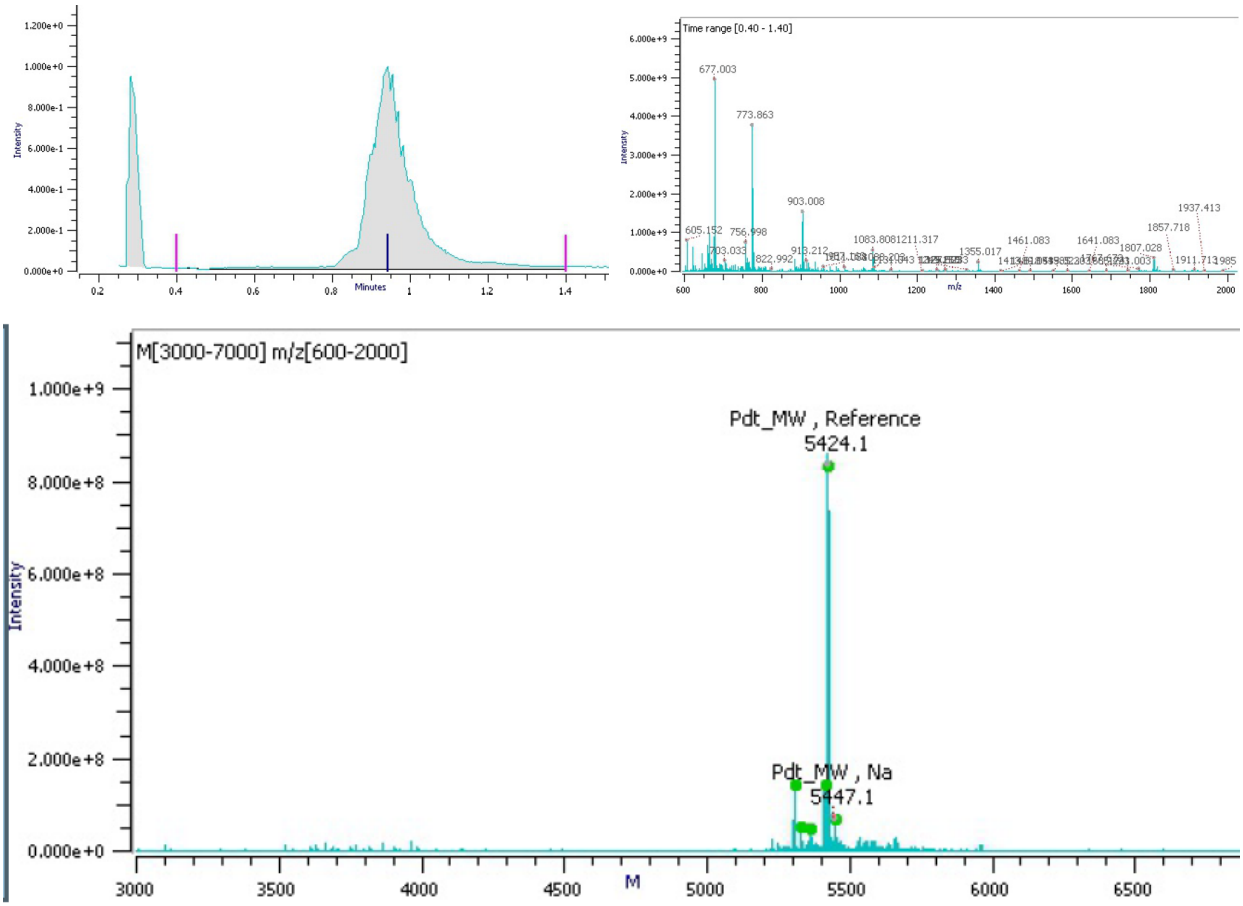
Product **4i**, 76% yield



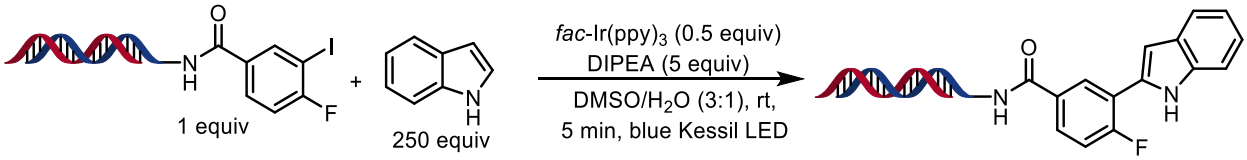
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5422,0944
Molecular Weight: 5424,7594

ELTM_648, TW-N78441-14-648.raw - (Plot Elution Peak - Plot MS1- Plot Deconvoluted Mas spectrum)



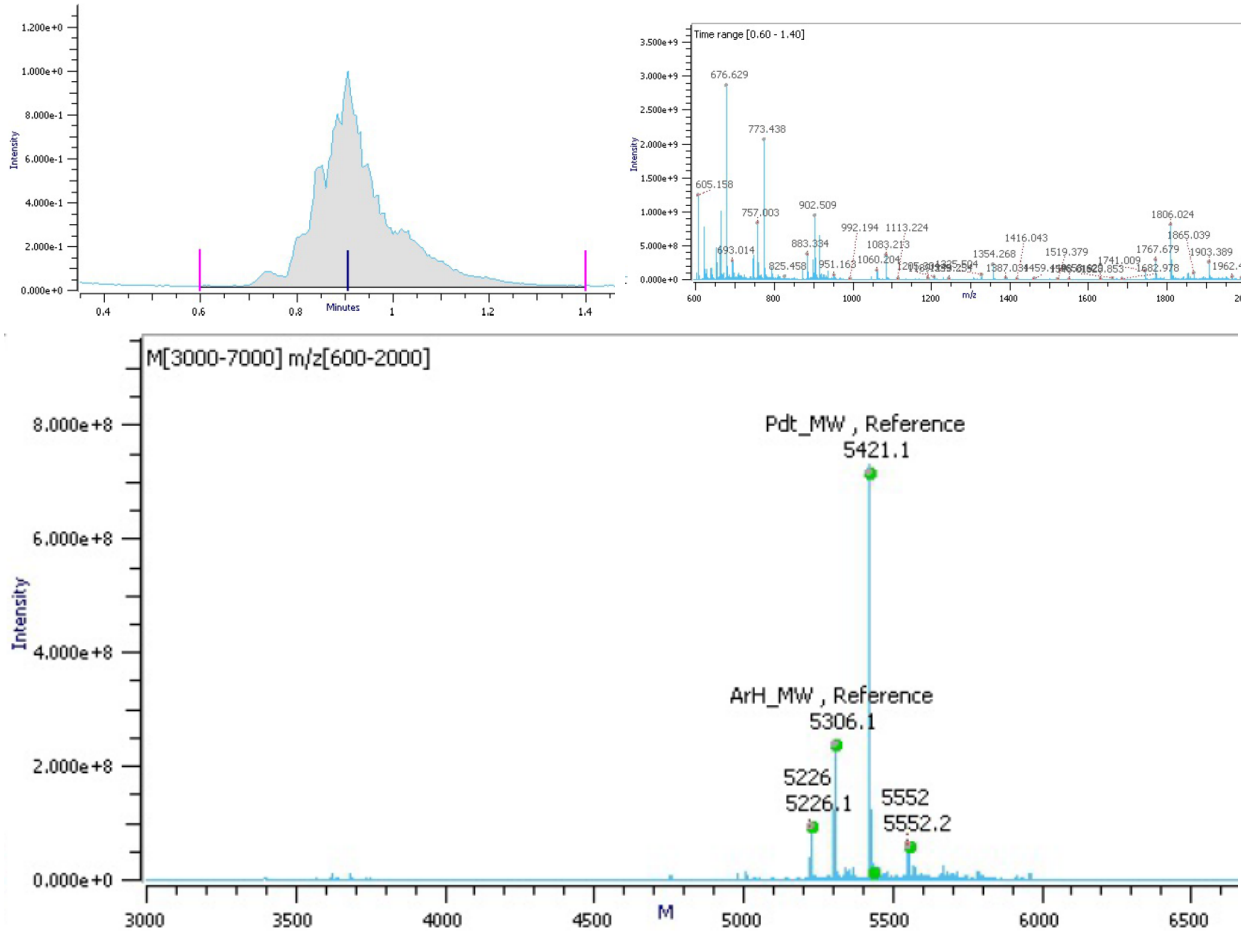
Product **4j**, 70% yield



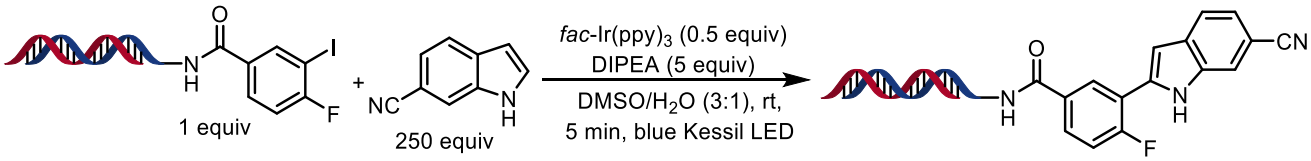
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5419,0835
Molecular Weight: 5421,7554

ELTM_179, TW-N78441-8-179.raw - (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



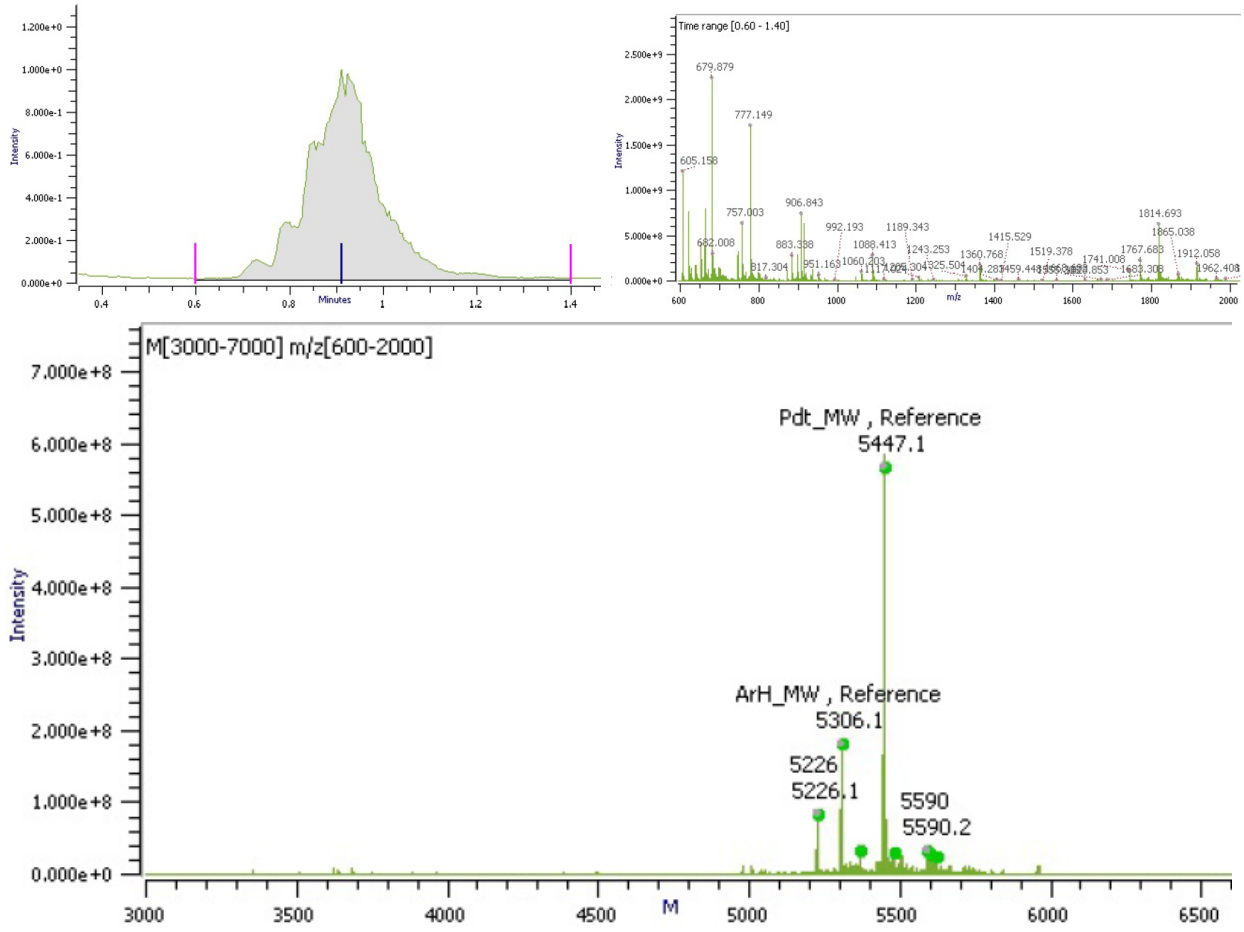
Product **4k**, 63% yield



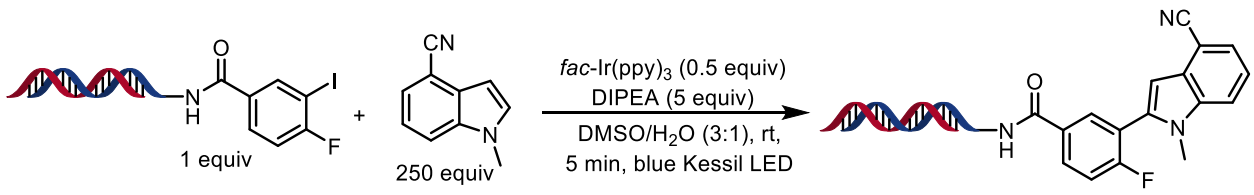
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5444,0788
Molecular Weight: 5446,7654

ELTM_176, TW-N78441-8-176.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



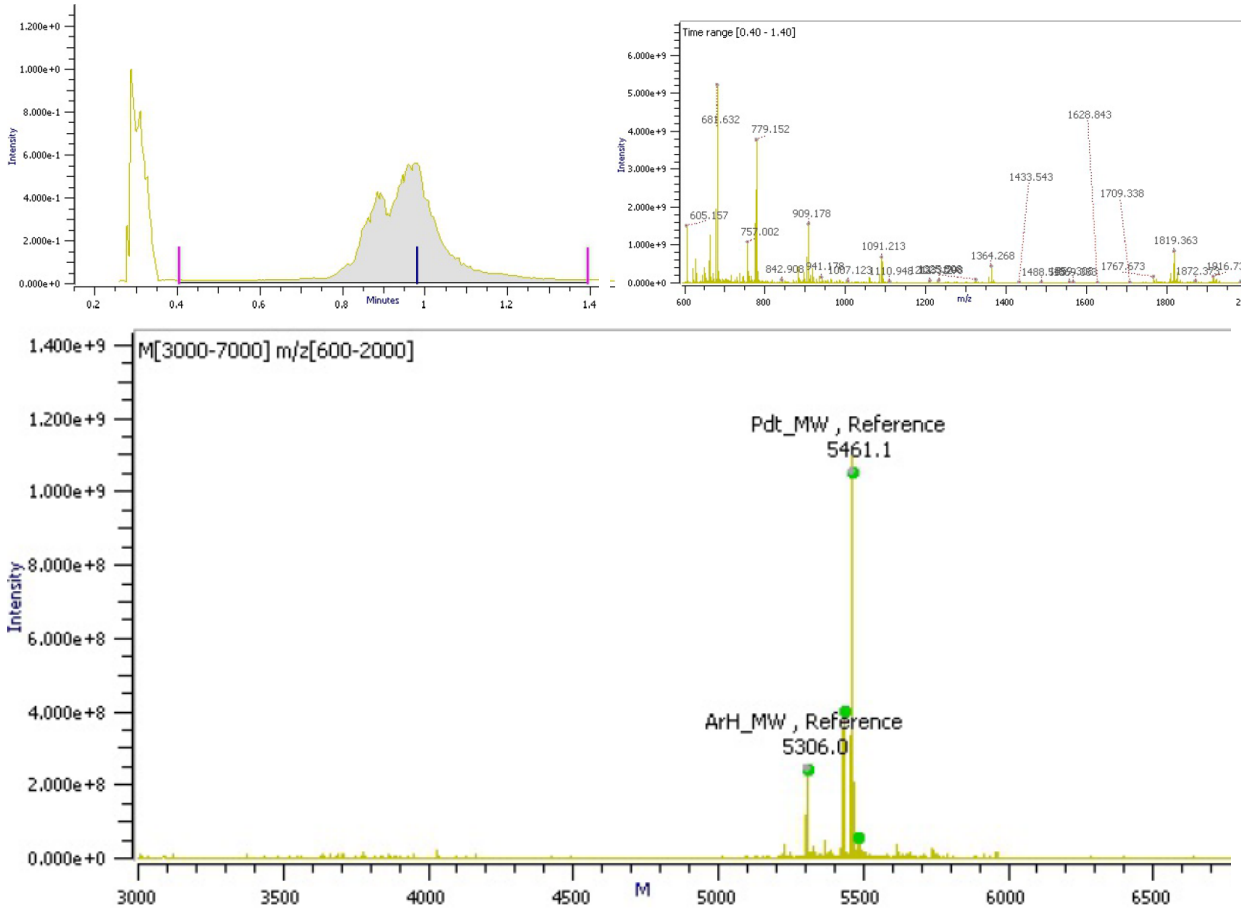
Product **4l**, 72% yield



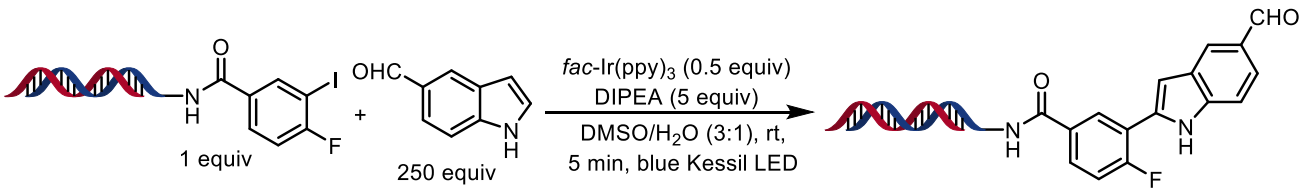
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5458,0944
Molecular Weight: 5460,7924

ELTM_660, TW-N78441-15-660.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



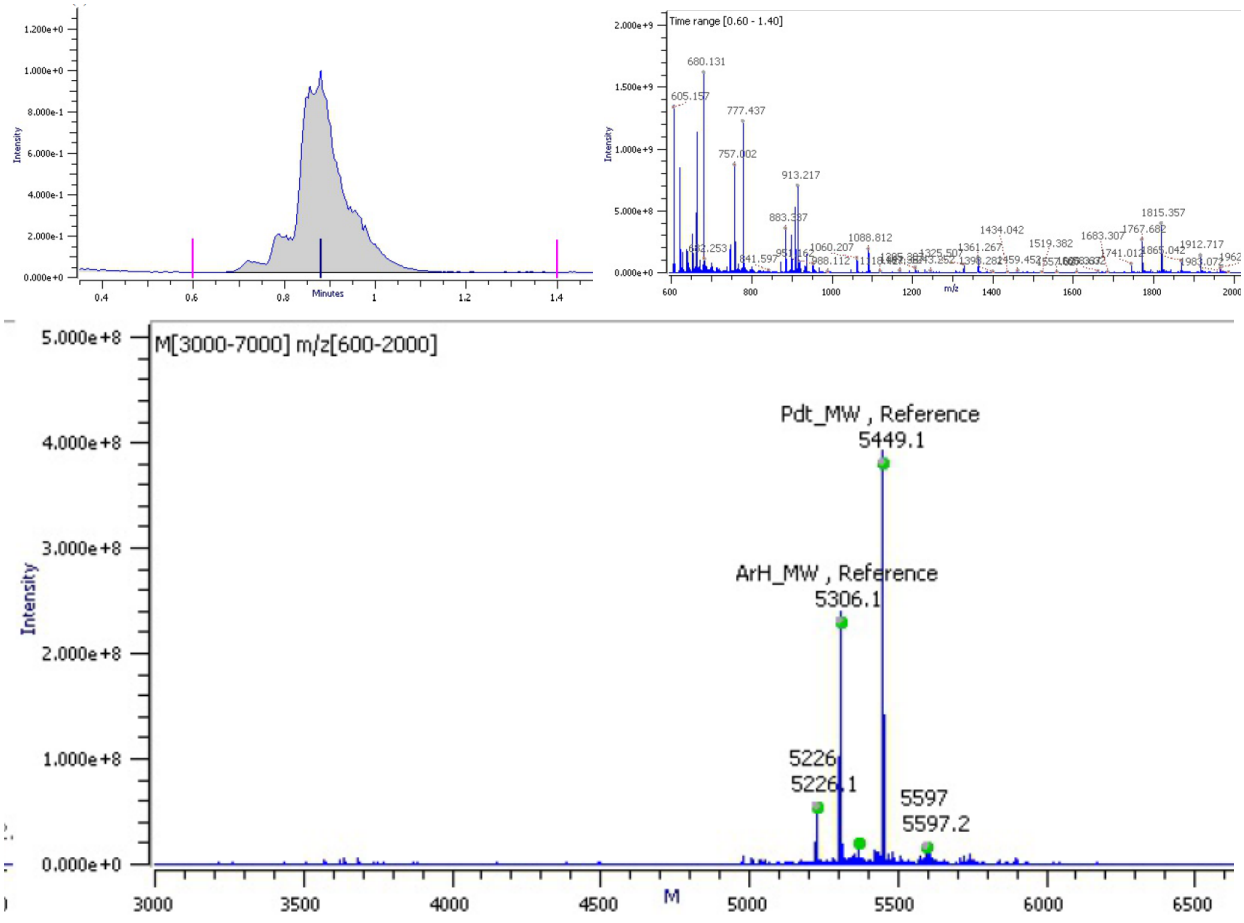
Product **4m**, 60% yield



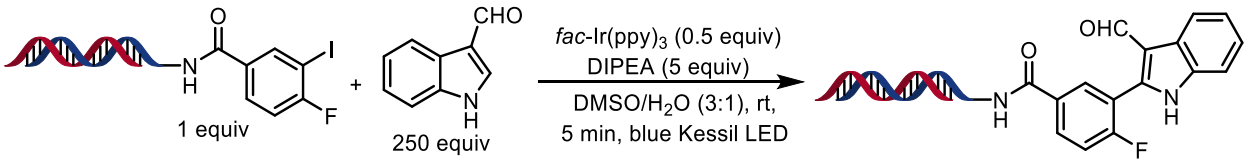
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5447,0784
Molecular Weight: 5449,7654

ELTM_180, TW-N78441-8-180.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



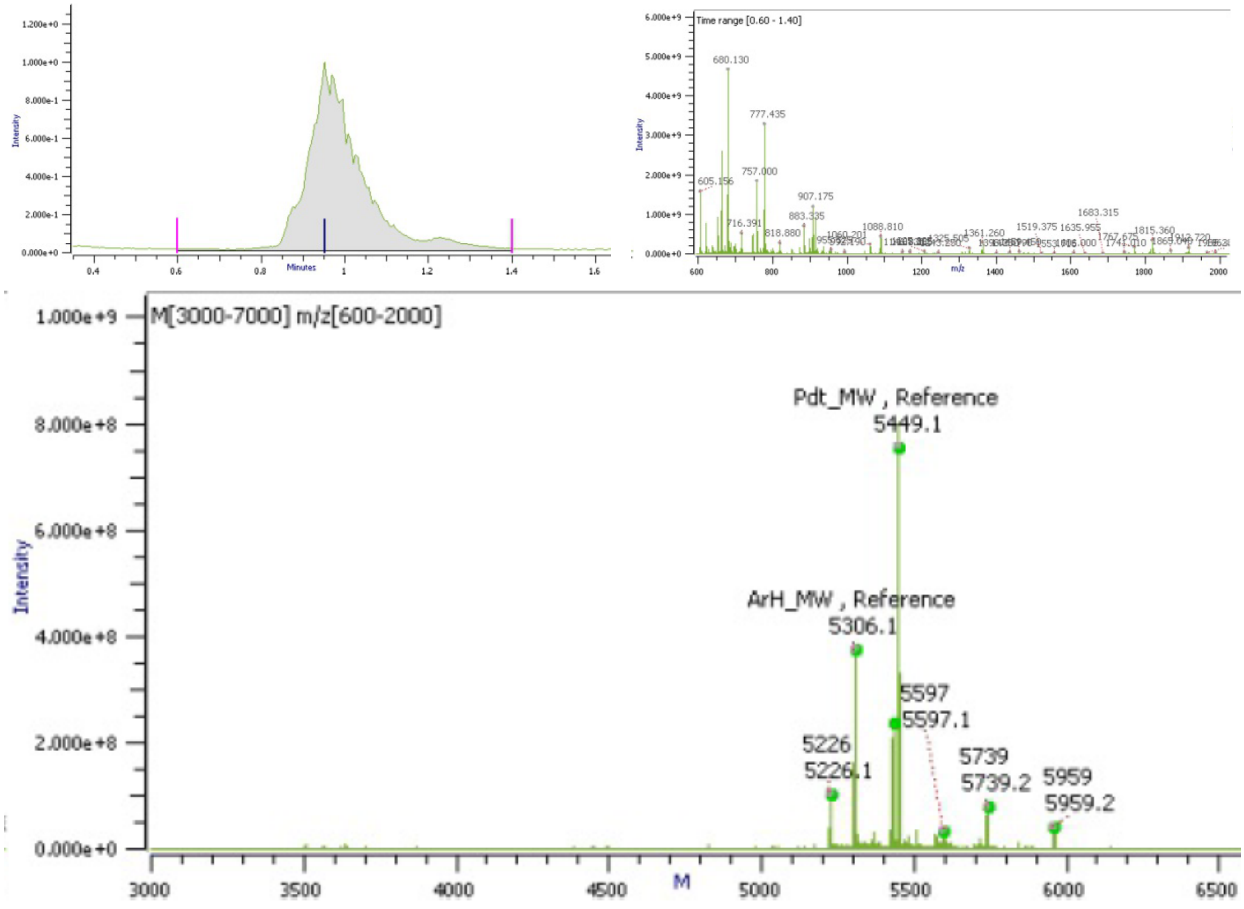
Product **4n**, 55% yield



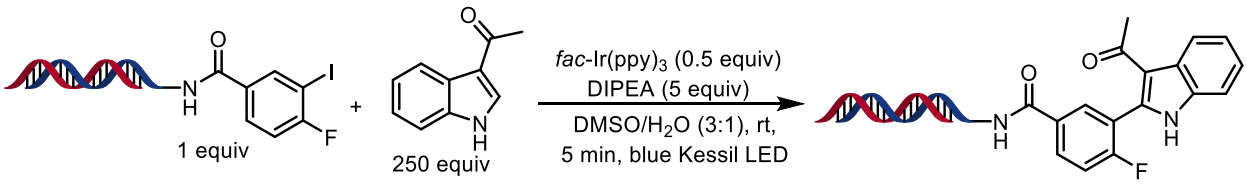
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5447,0784
Molecular Weight: 5449,7654

ELTM_346, TW-N78441-9-346.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



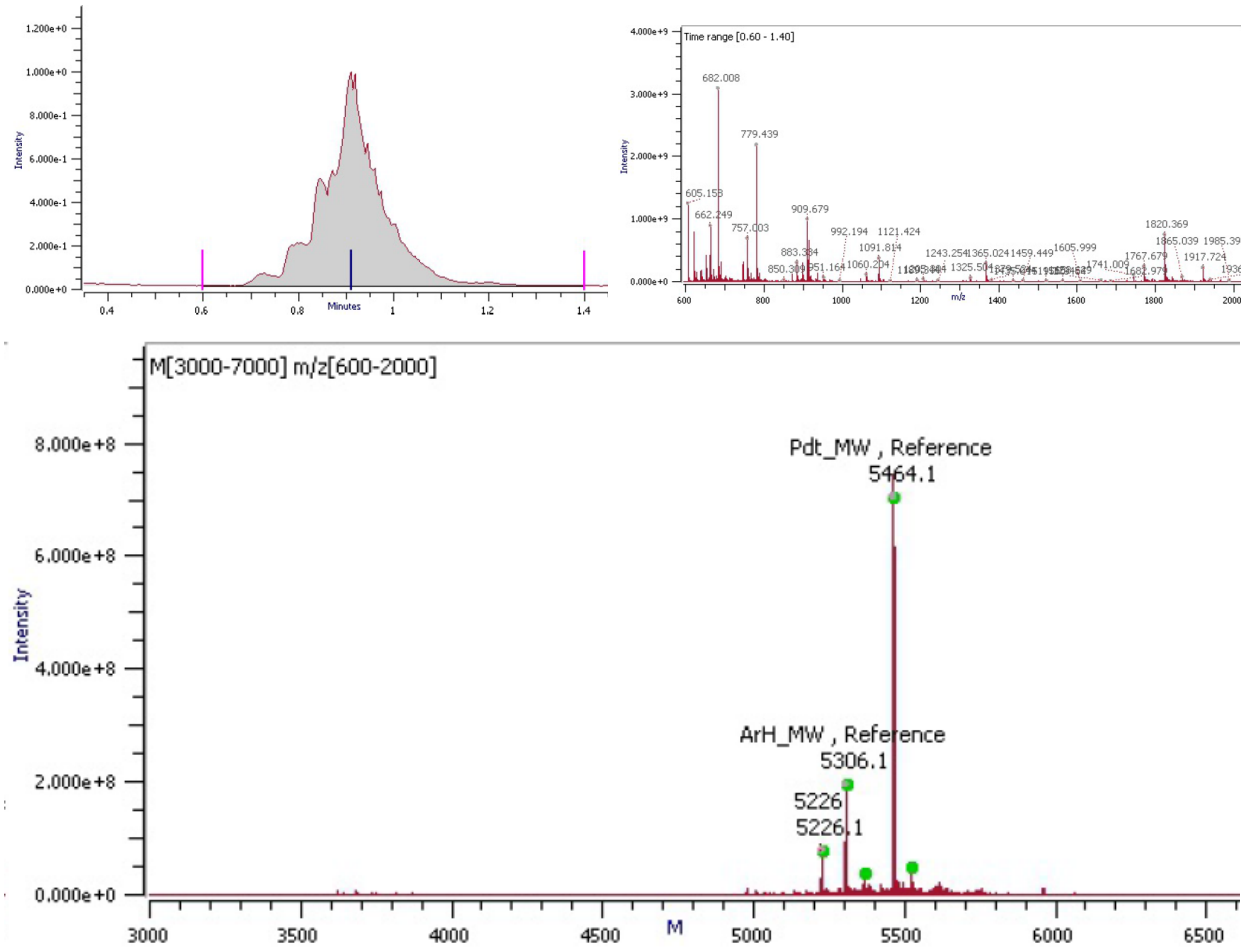
Product **4o**, 73% yield



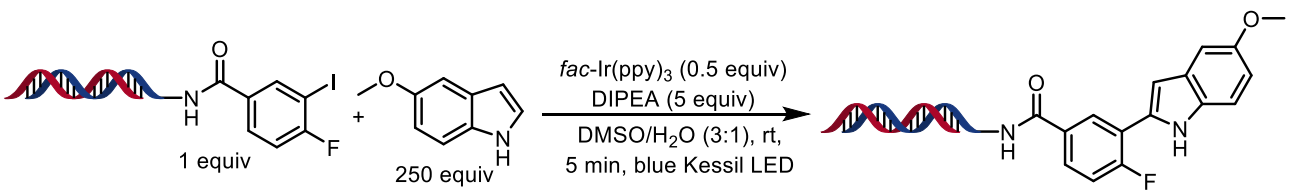
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5461,0941
Molecular Weight: 5463,7924

ELTM_181, TW-N78441-8-181.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



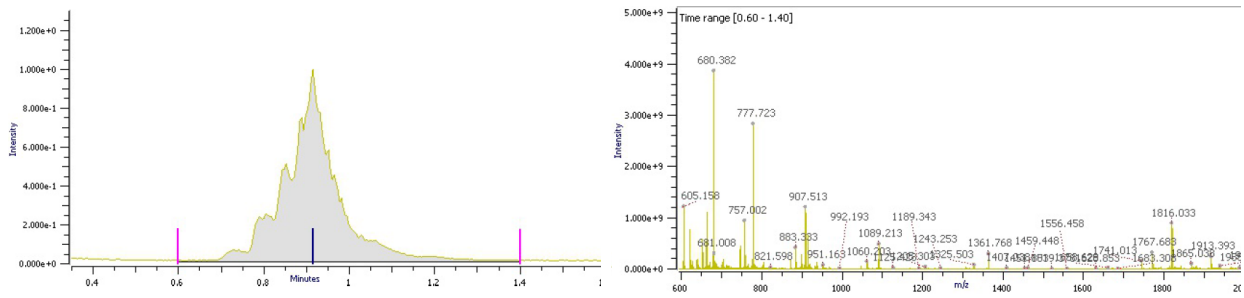
Product **4p**, 72% yield

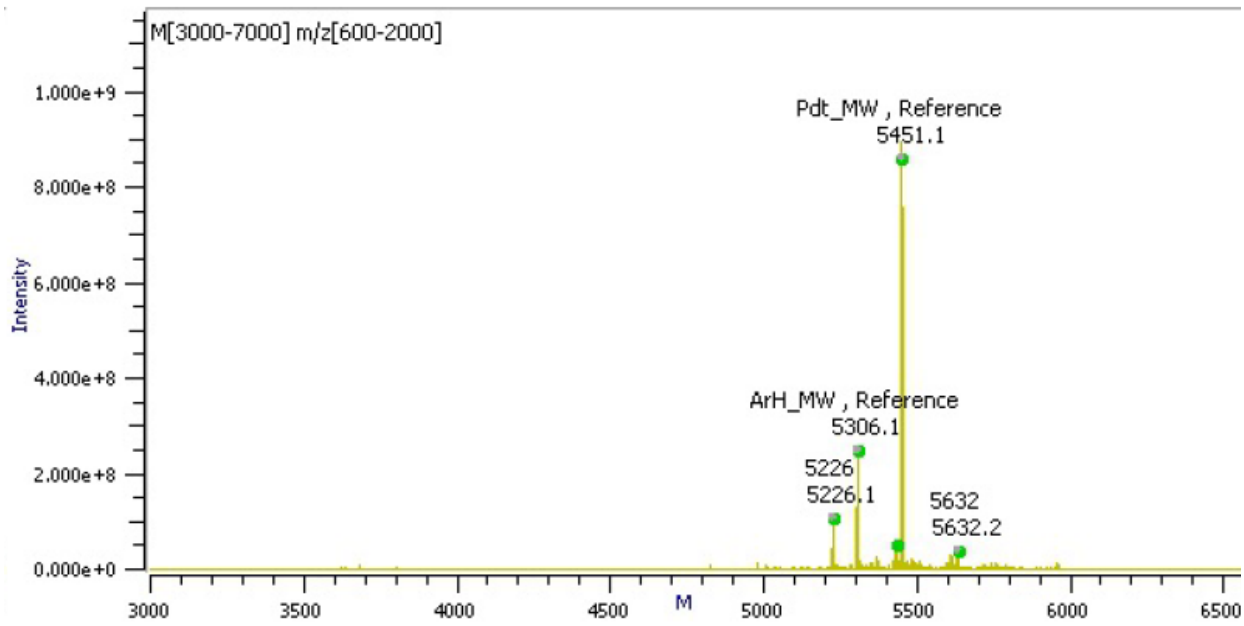


Exact Mass: 5429,9380
Molecular Weight: 5432,5168

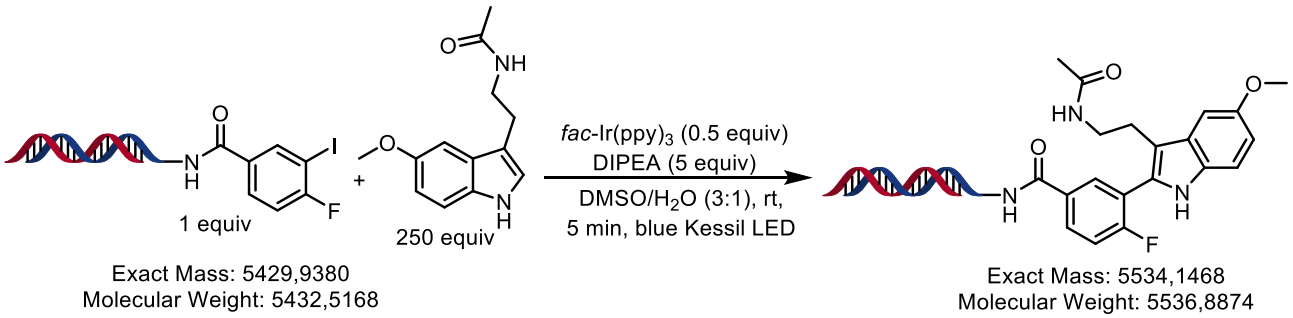
Exact Mass: 5449,0941
Molecular Weight: 5451,7814

ELTM_182, TW-N78441-8-182.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)





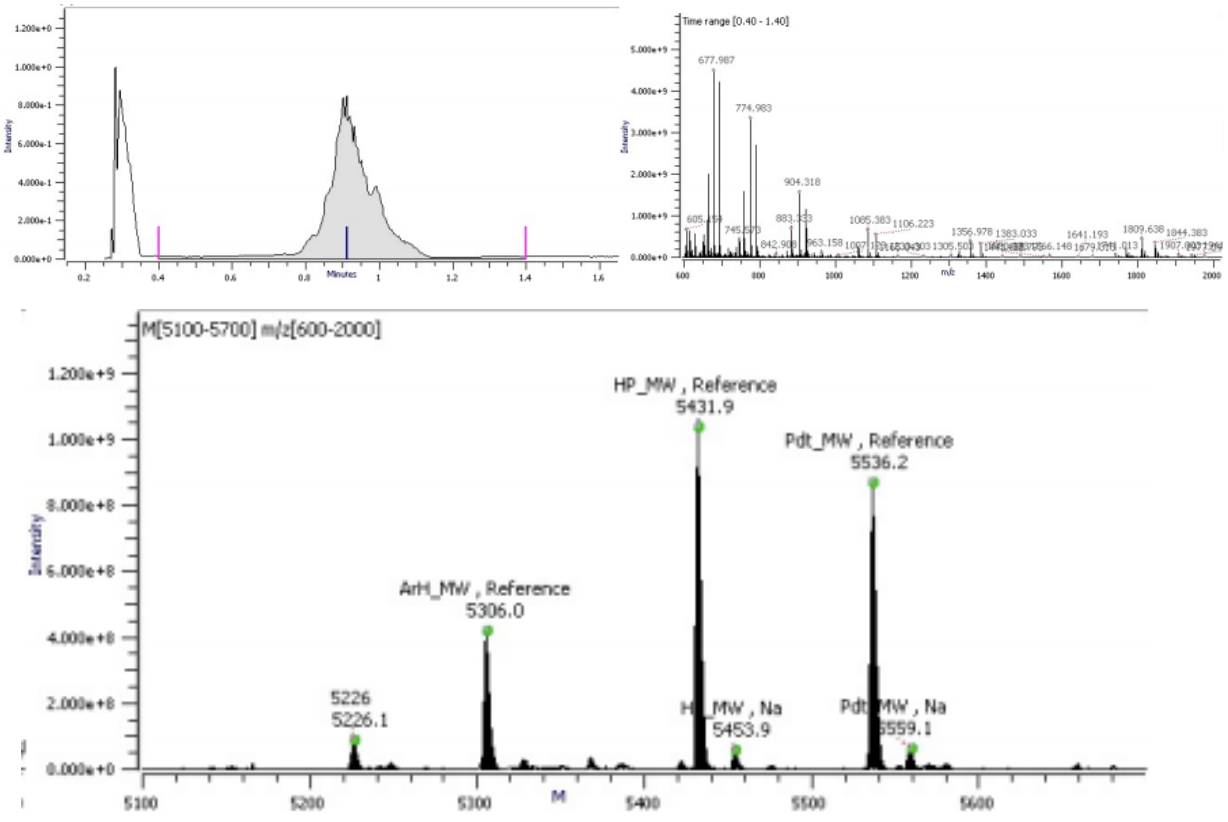
Product **4q**, 37% yield



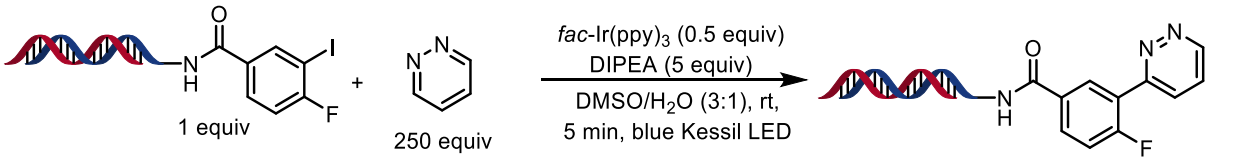
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5534,1468
Molecular Weight: 5536,8874

TW-N78441-15-651 - (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



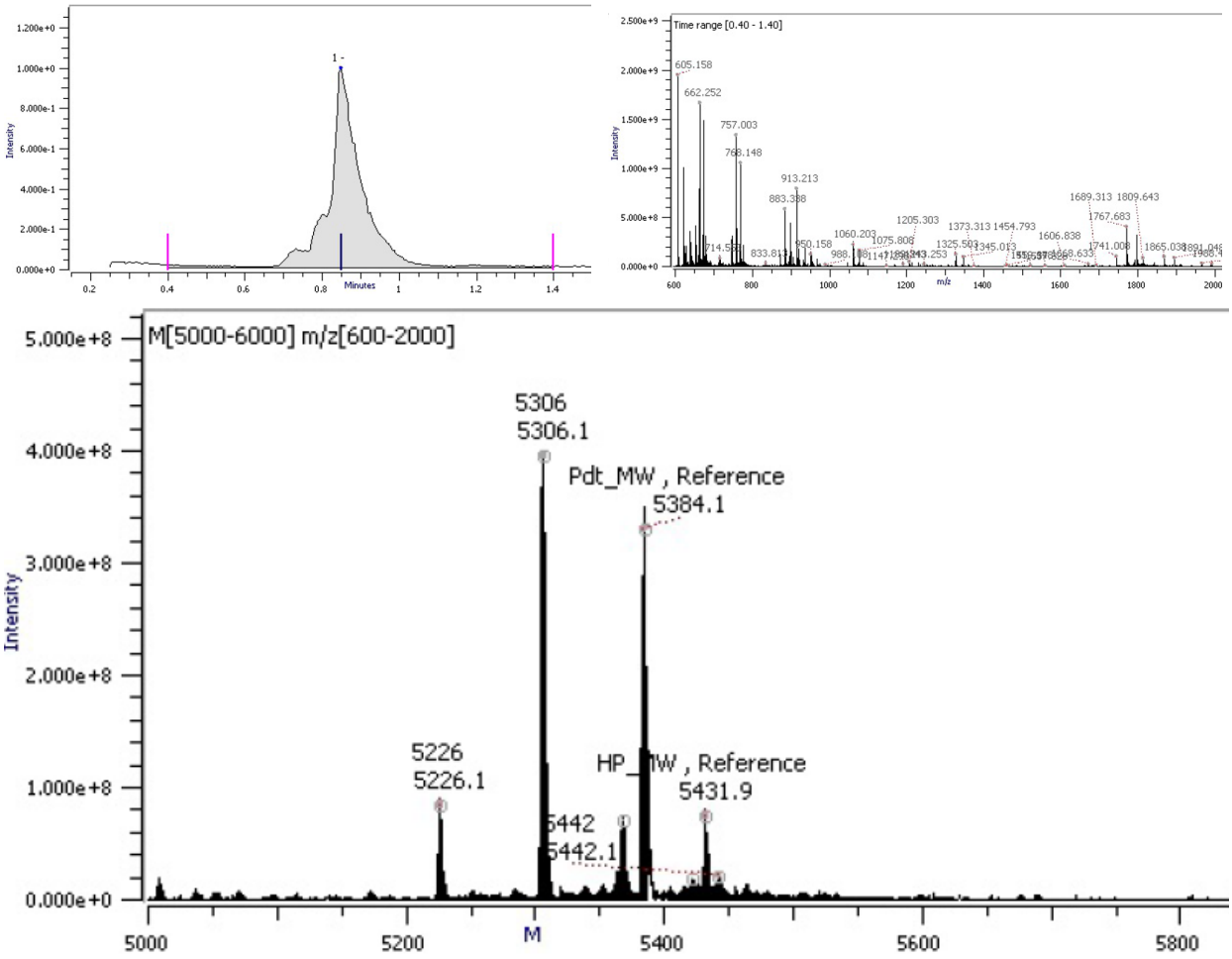
Product **4r**, 37% yield



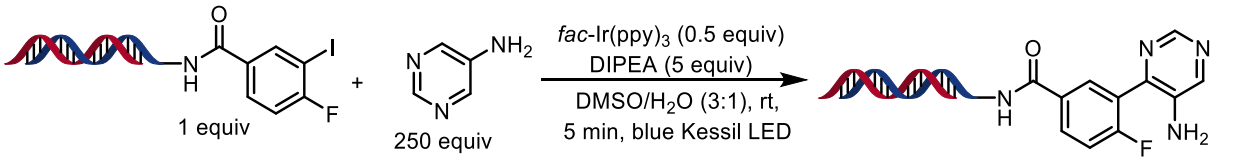
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5382,0631
Molecular Weight: 5384,6944

ELTM_146, TW-N78441-8-146.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



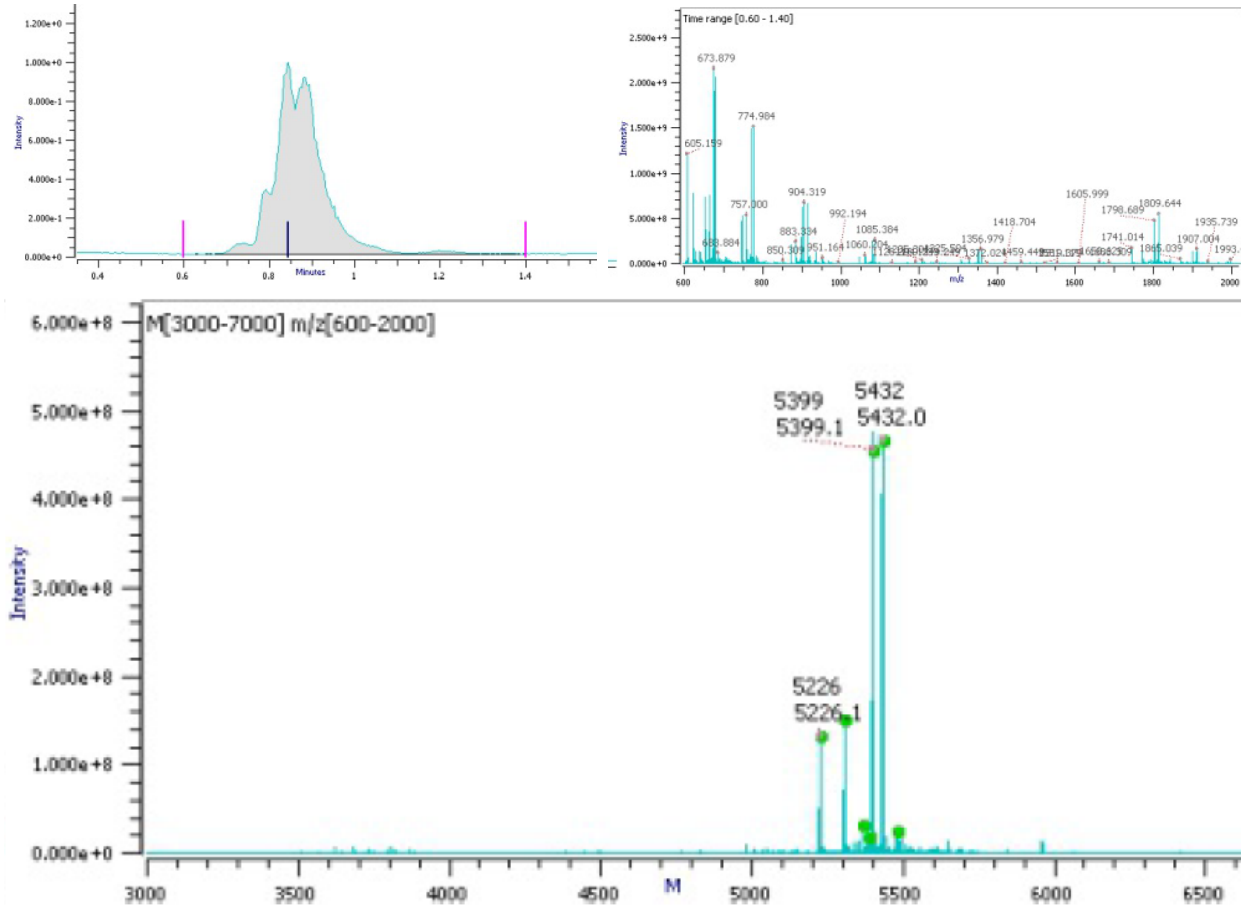
Product **4s**, 40% yield



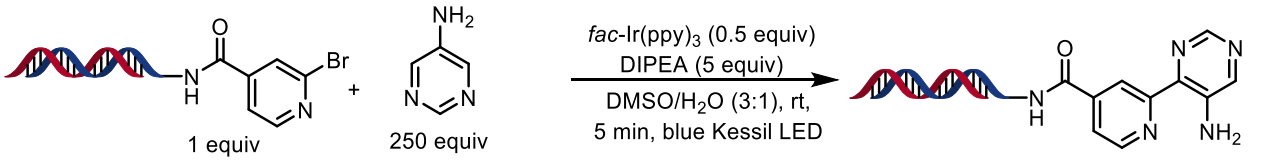
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5397,0740
Molecular Weight: 5399,7094

ELTM_225, TW-N78441-8-225.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



Product **6s**, 89% yield

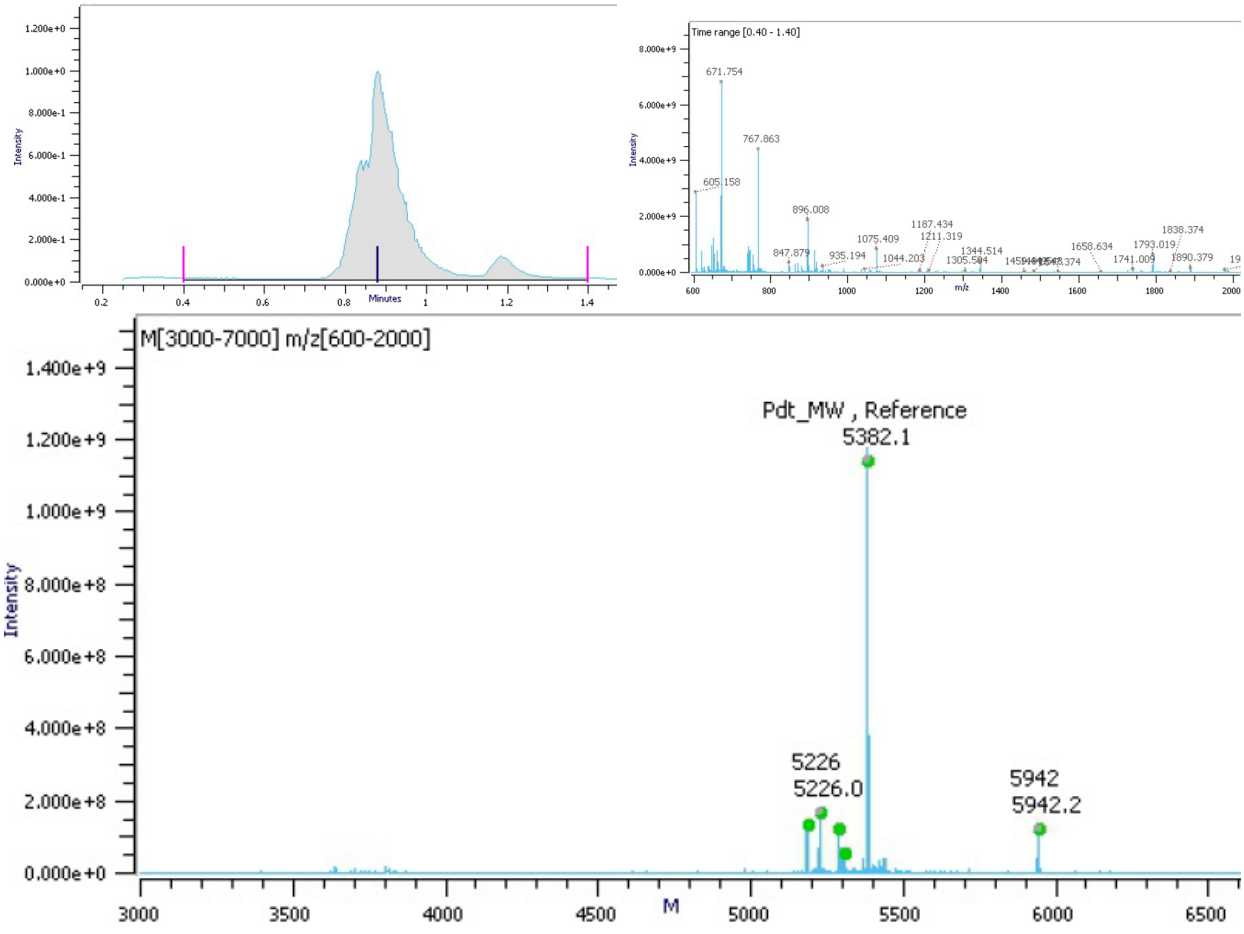


1 equiv
Exact Mass: 5364,9565
Molecular Weight: 5368,5140

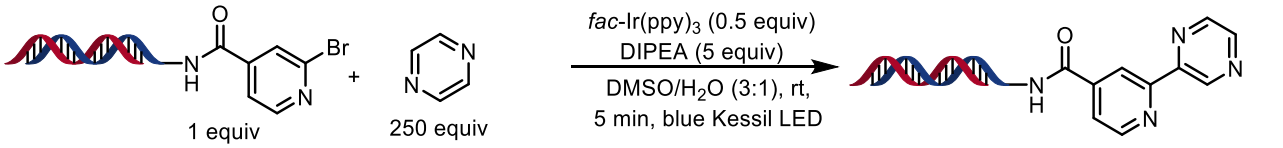
250 equiv

Exact Mass: 5380,0787
Molecular Weight: 5382,7070

ELTM_447, TW-N78441-10-447.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



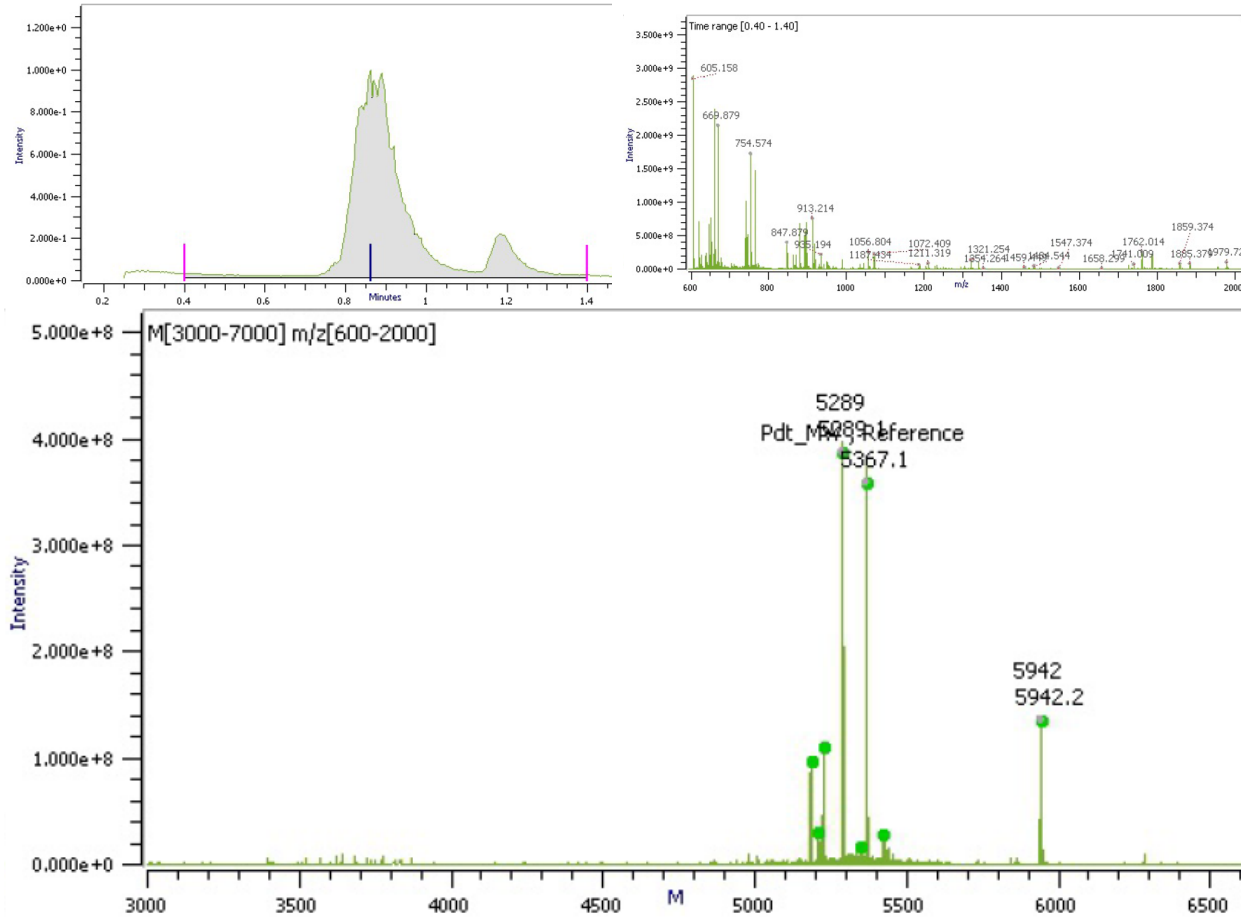
Product **6t**, 42% yield



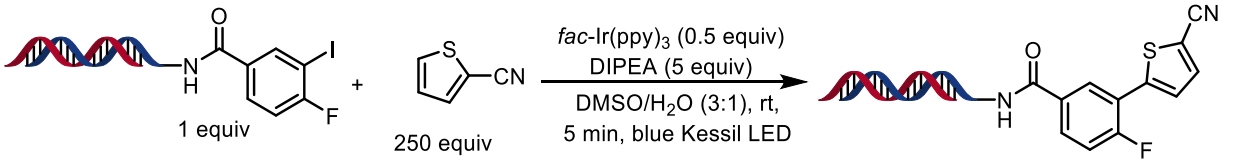
Exact Mass: 5364,9565
Molecular Weight: 5368,5140

Exact Mass: 5365,0678
Molecular Weight: 5367,6920

ELTM_446, TW-N78441-10-446.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



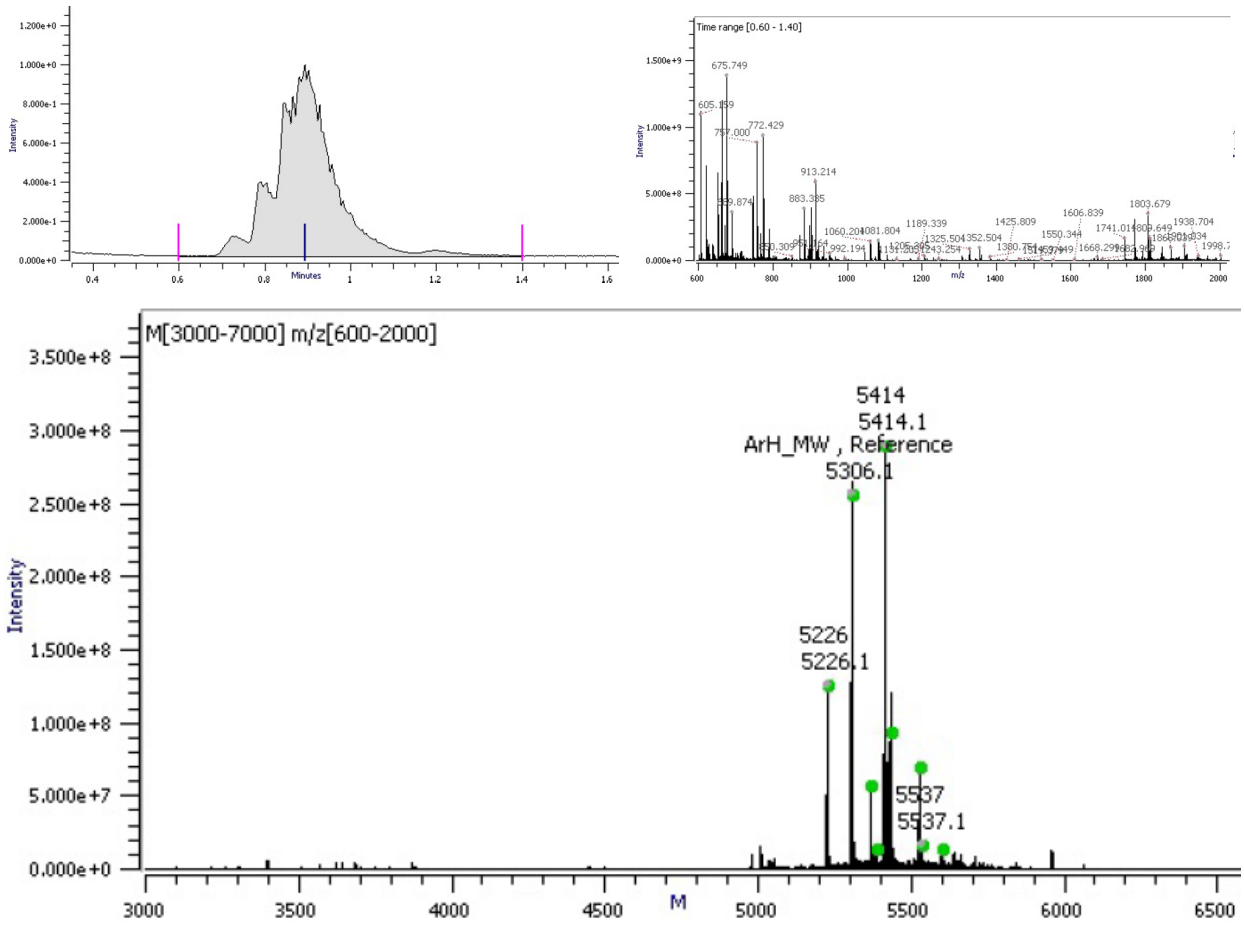
Product **4u**, 34% yield



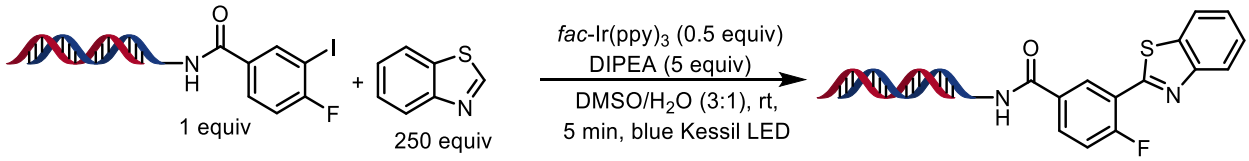
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5411,0243
Molecular Weight: 5413,7504

ELTM_223, TW-N78441-8-223.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



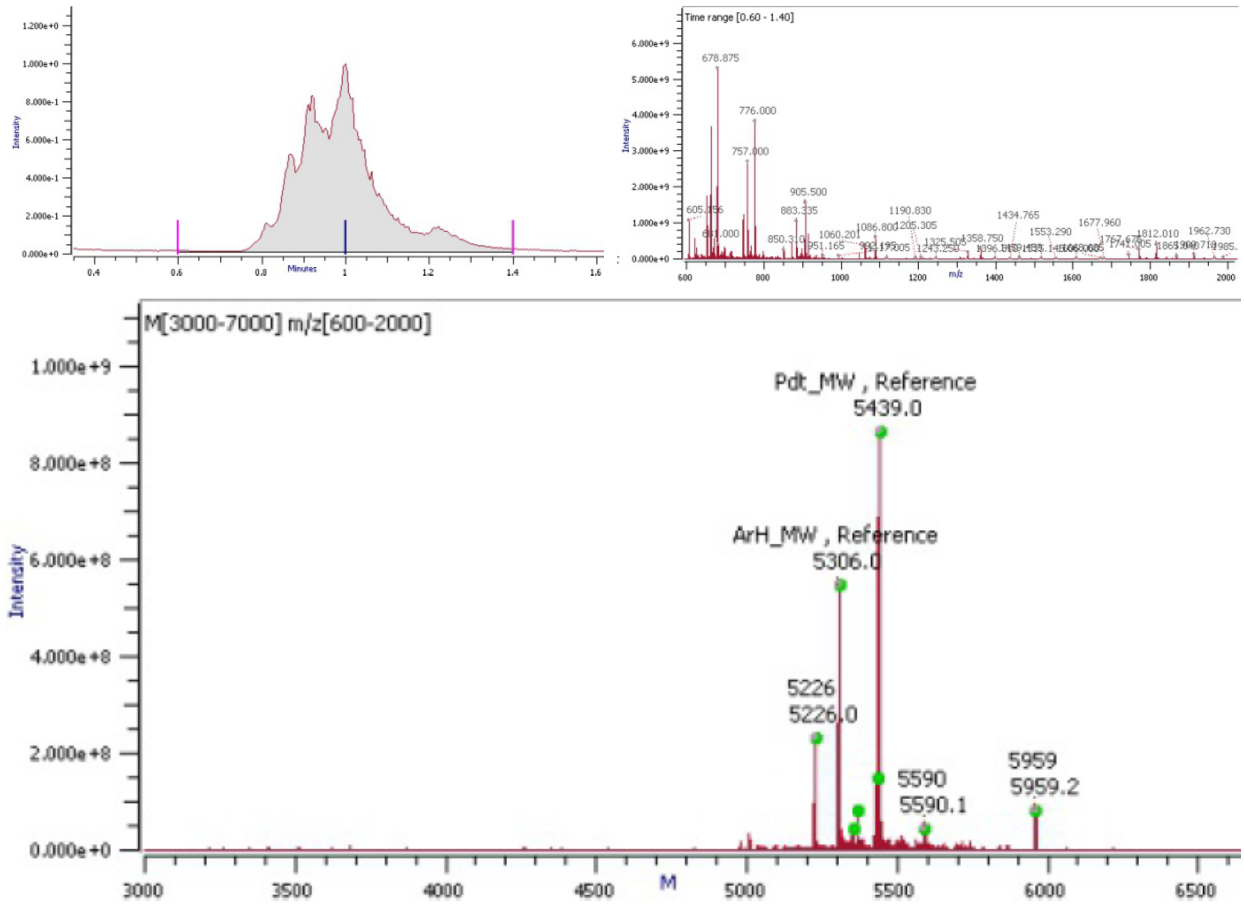
Product **4v**, 50% yield



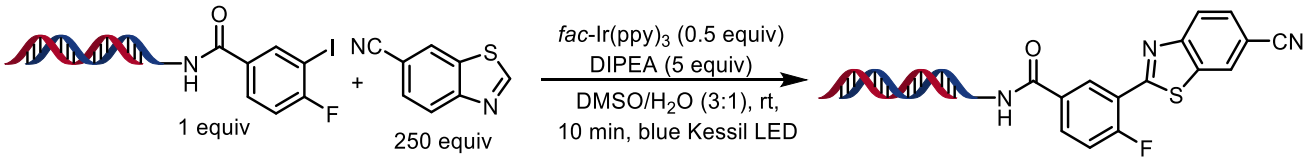
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5437,0399
Molecular Weight: 5439,7884

ELTM_390, TW-N78441-9-390.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



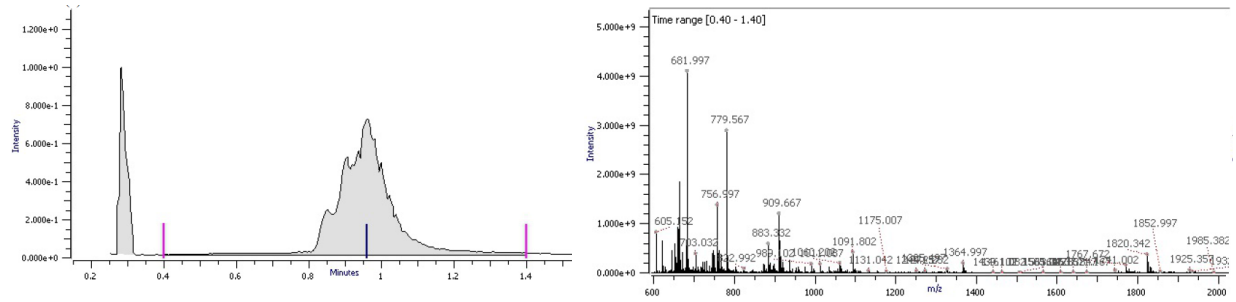
Product **4w**, 61% yield

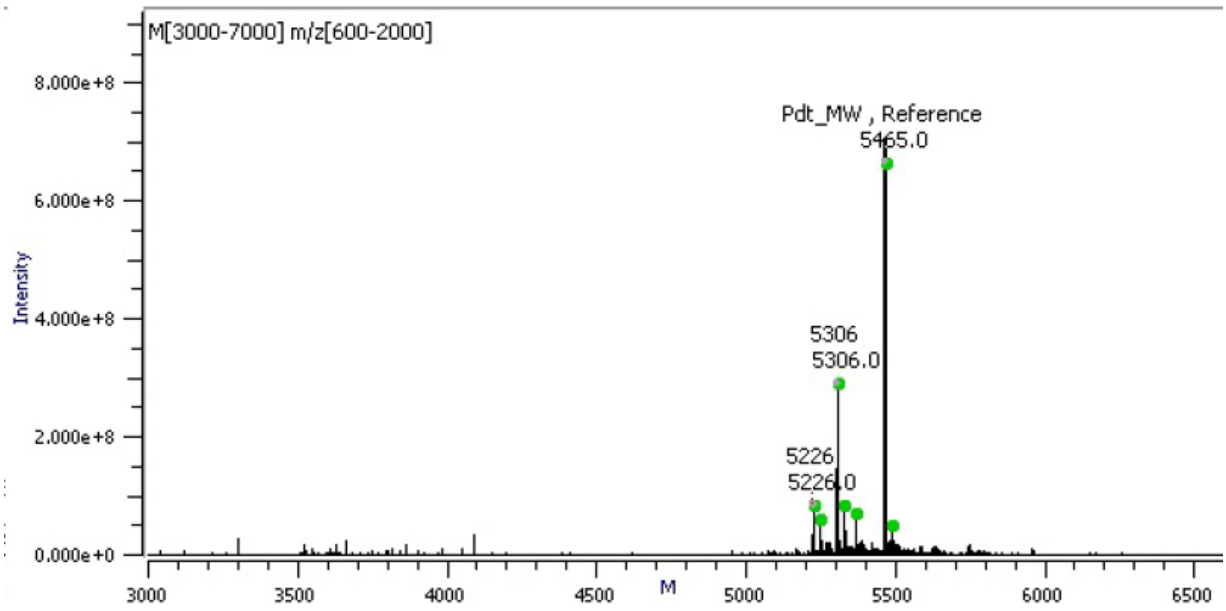


Exact Mass: 5429,9380
Molecular Weight: 5432,5168

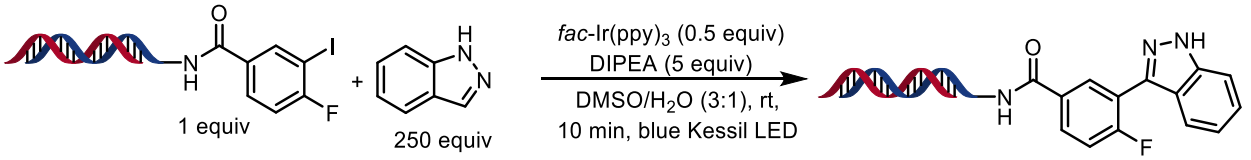
Exact Mass: 5462,0352
Molecular Weight: 5464,7984

ELTM_636, TW-N78441-14-636.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)





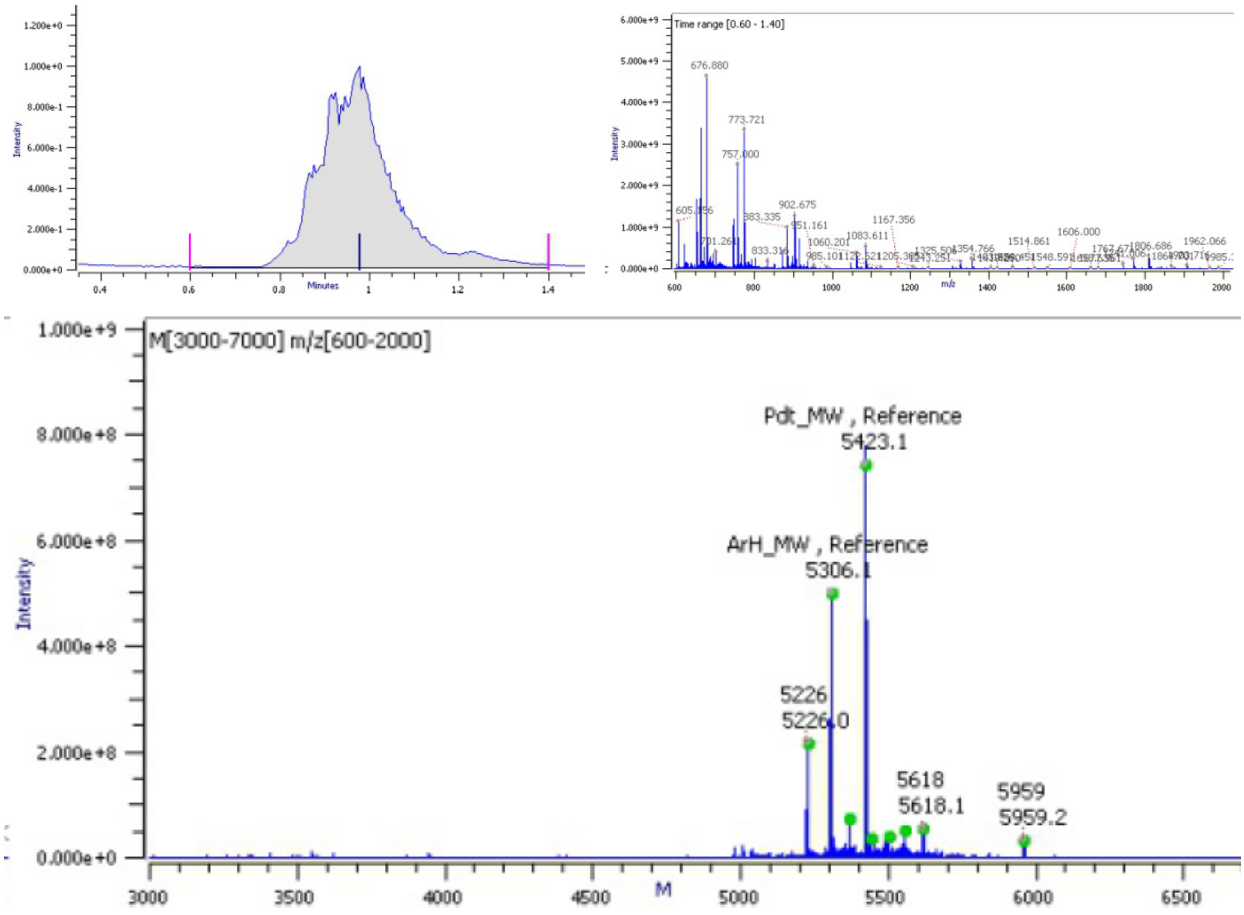
Product **4x**, 51% yield



Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5420,0788
Molecular Weight: 5422,7434

ELTM_355, TW-N78441-9-355.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)

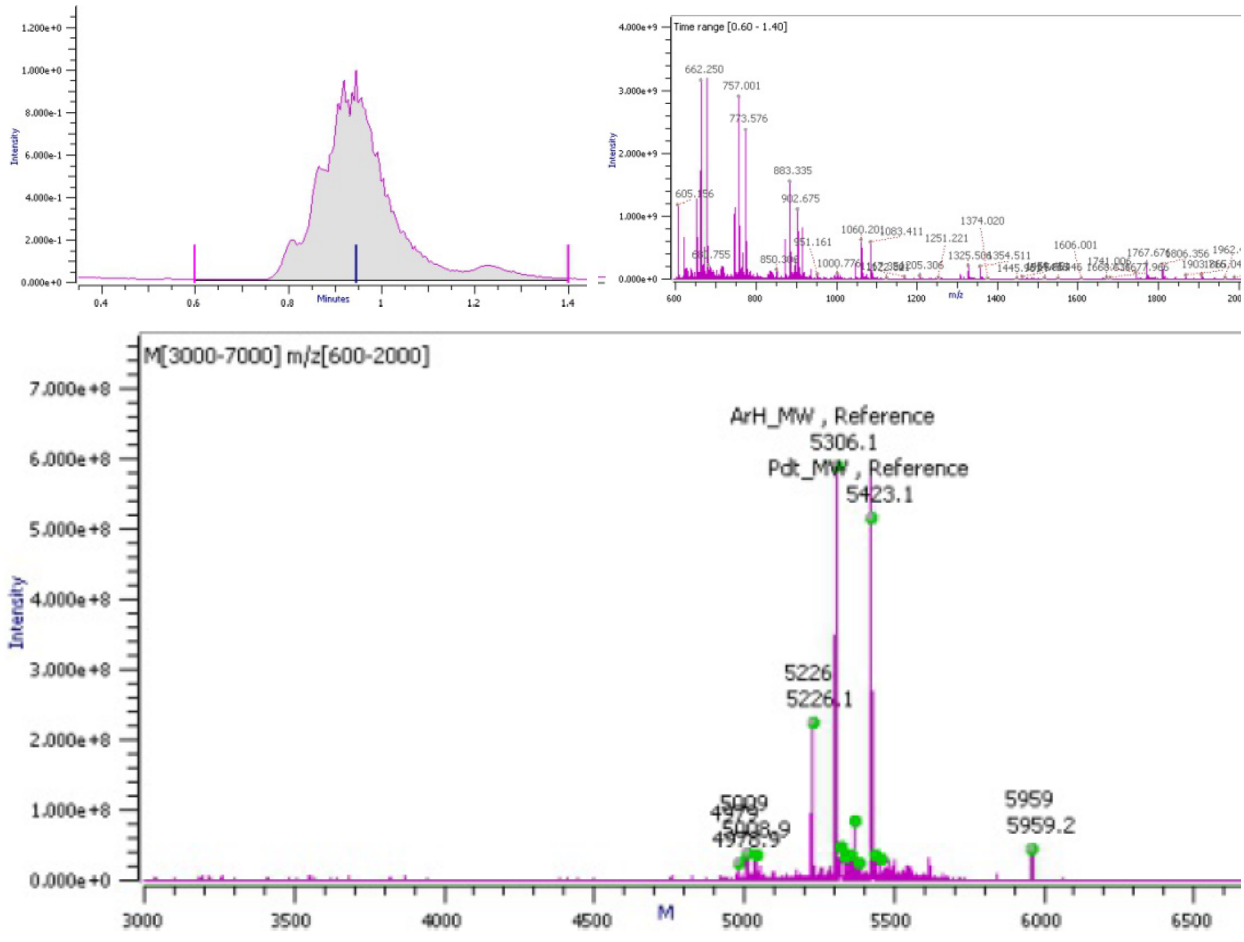


Product **4y**, 35% yield

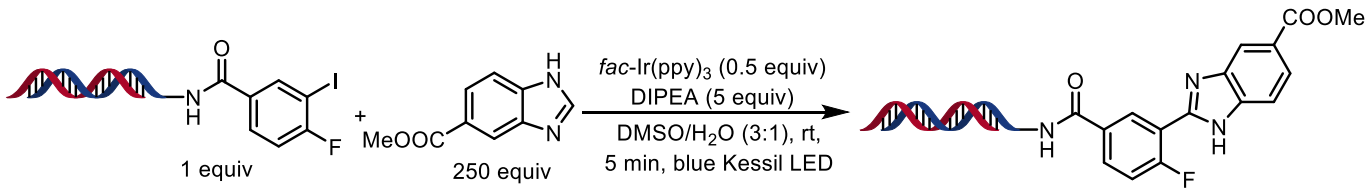
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5420,0788
Molecular Weight: 5422,7434

ELTM_364, TW-N78441-9-364.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



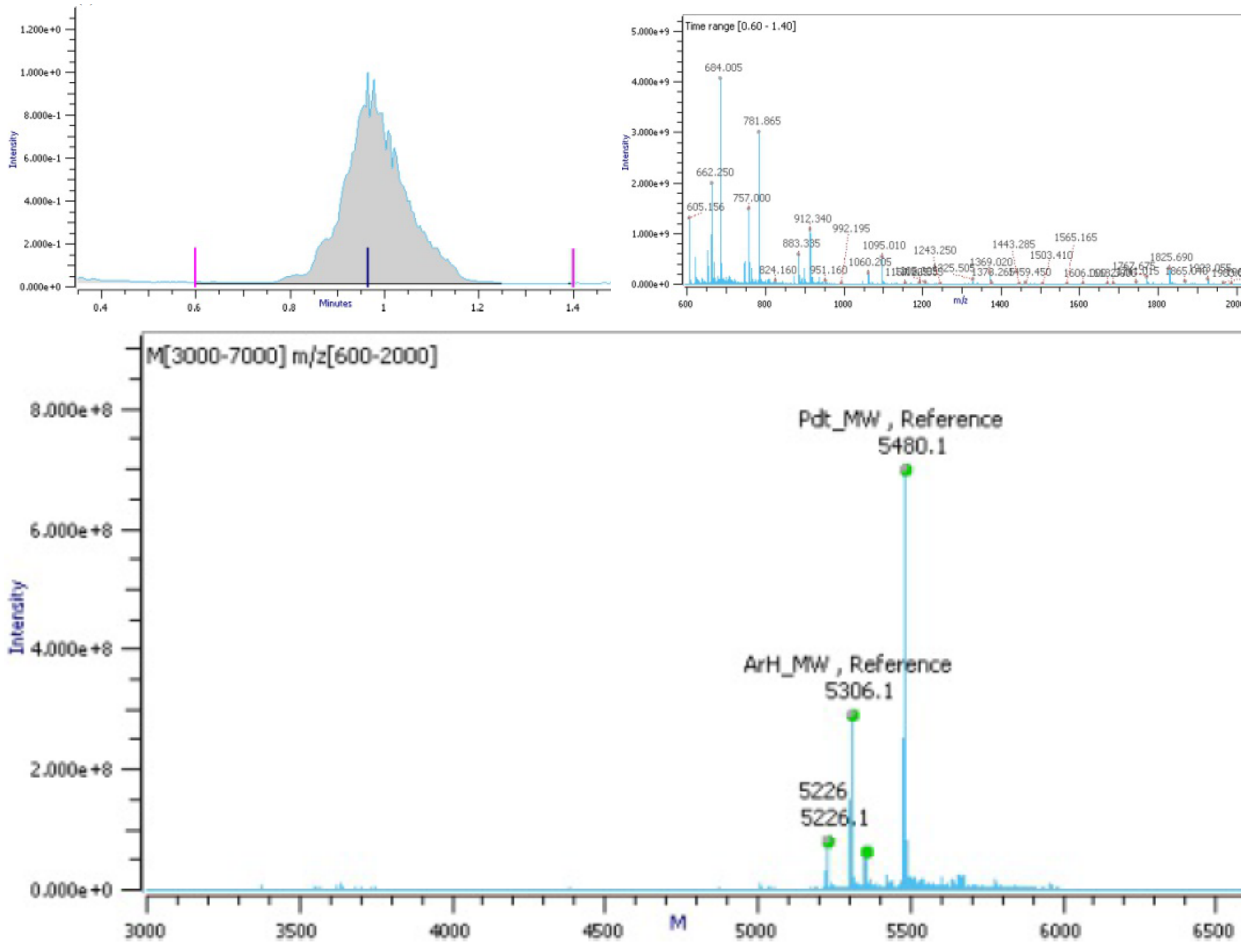
Product **4z**, 73% yield



Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5478,0842
Molecular Weight: 5480,7794

ELTM_349, TW-N78441-9-349.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)

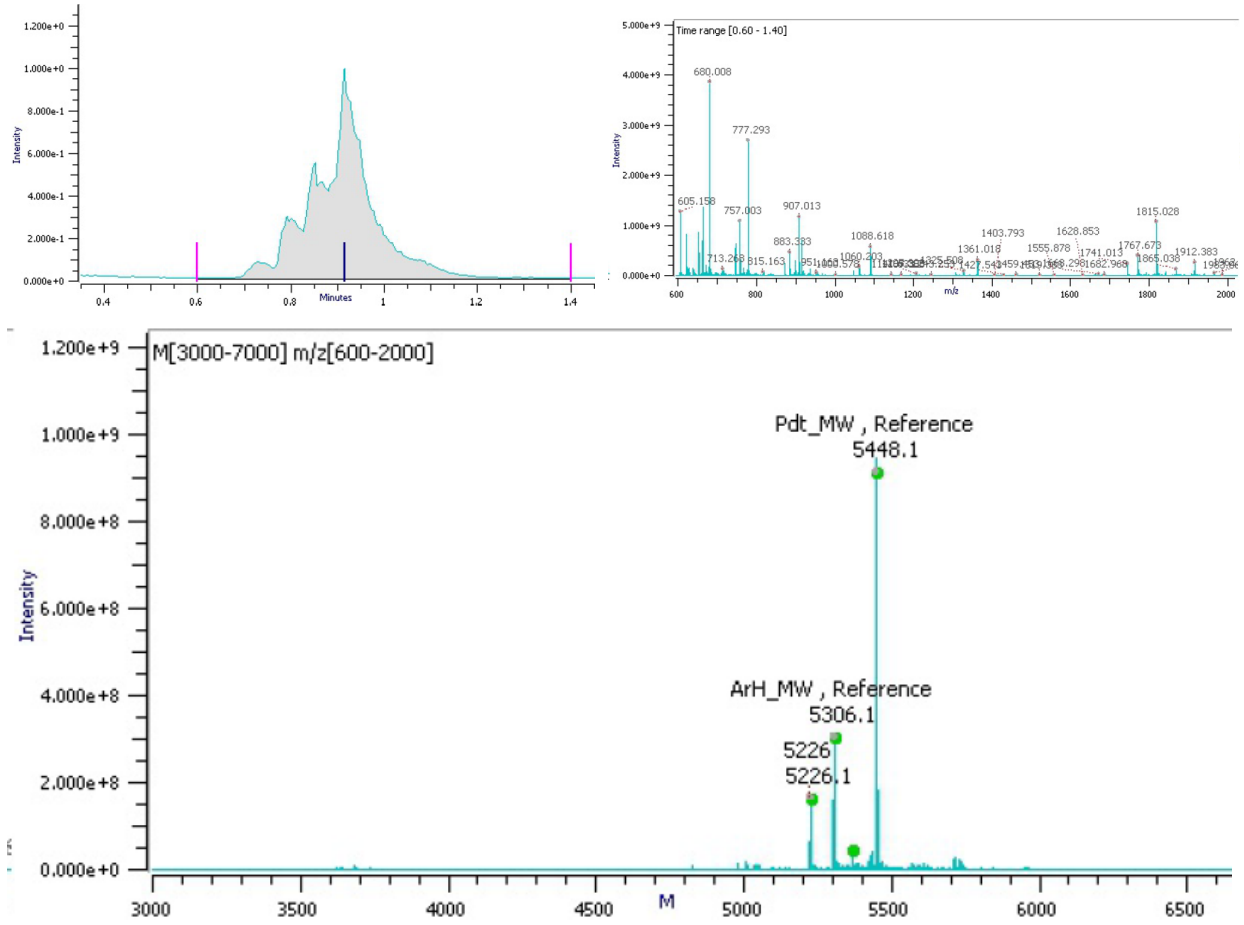


Product **4aa**, 70% yield

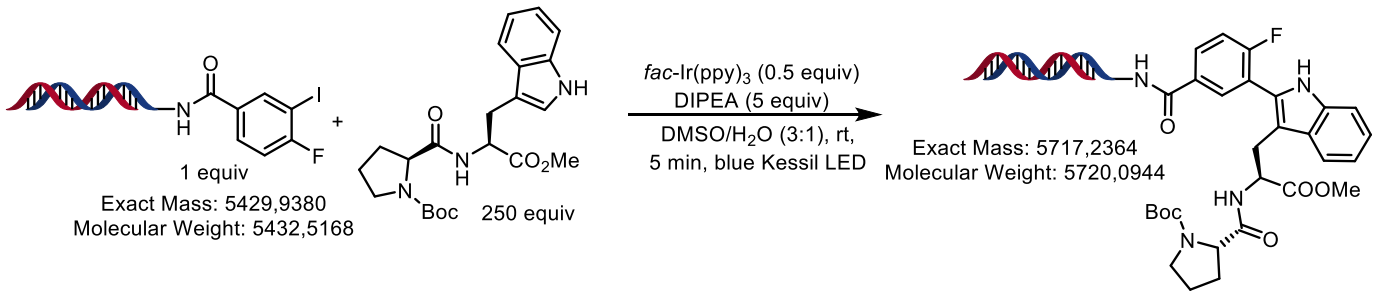
Exact Mass: 5429,9380
Molecular Weight: 5432,5168

Exact Mass: 5446,0944
Molecular Weight: 5448,7814

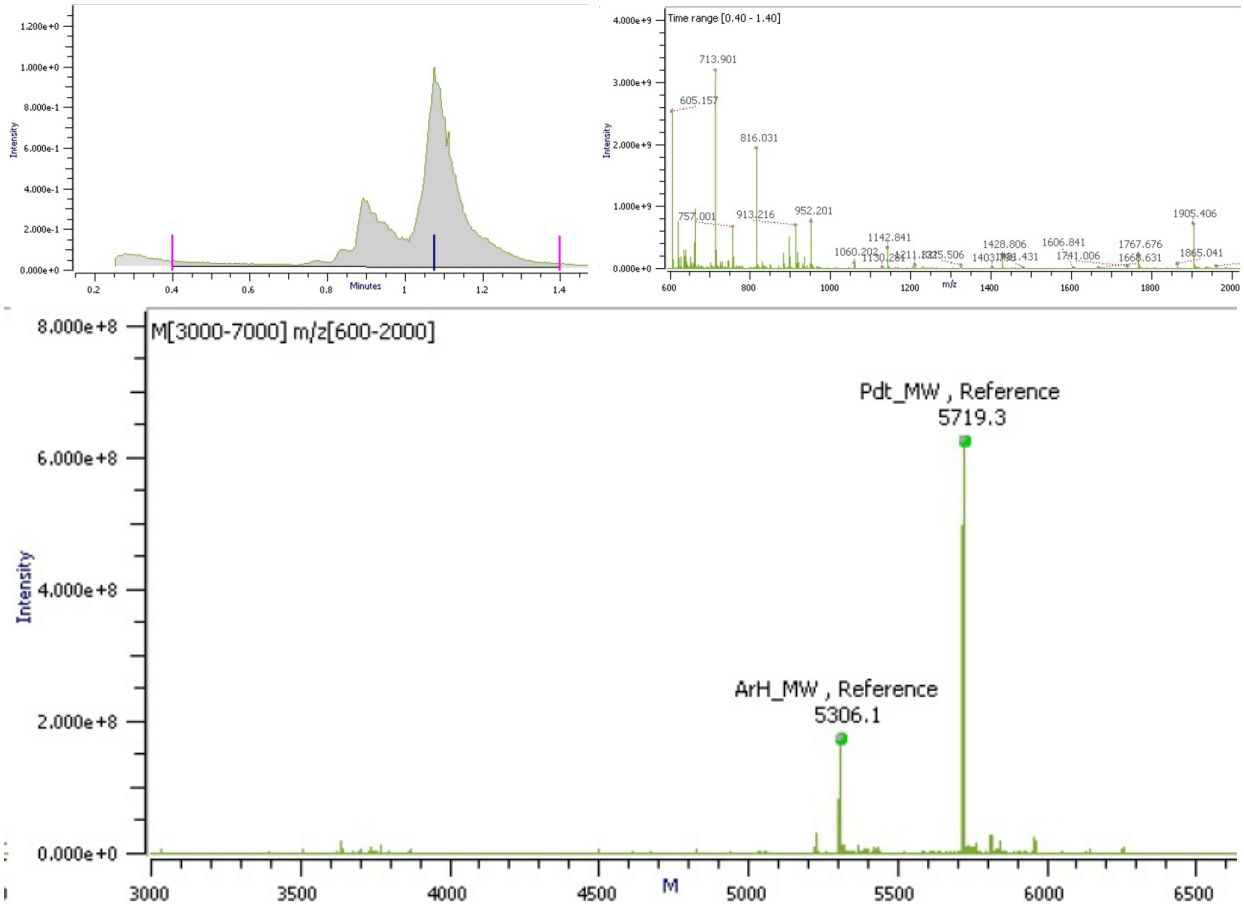
ELTM_222, TW-N78441-8-222.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



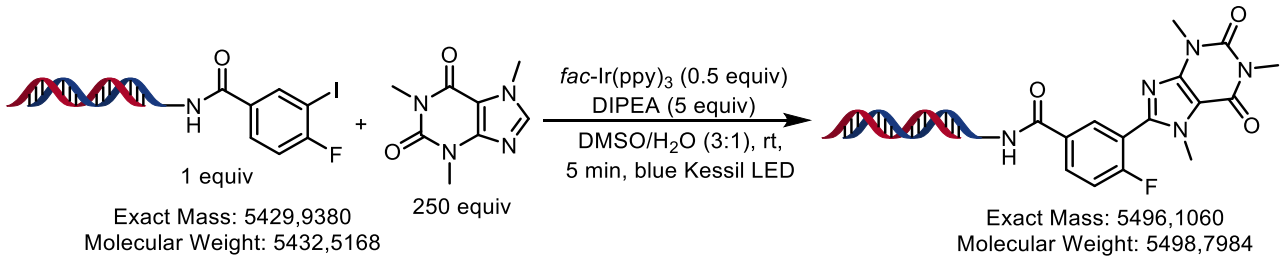
Product **4ab**, 93% yield



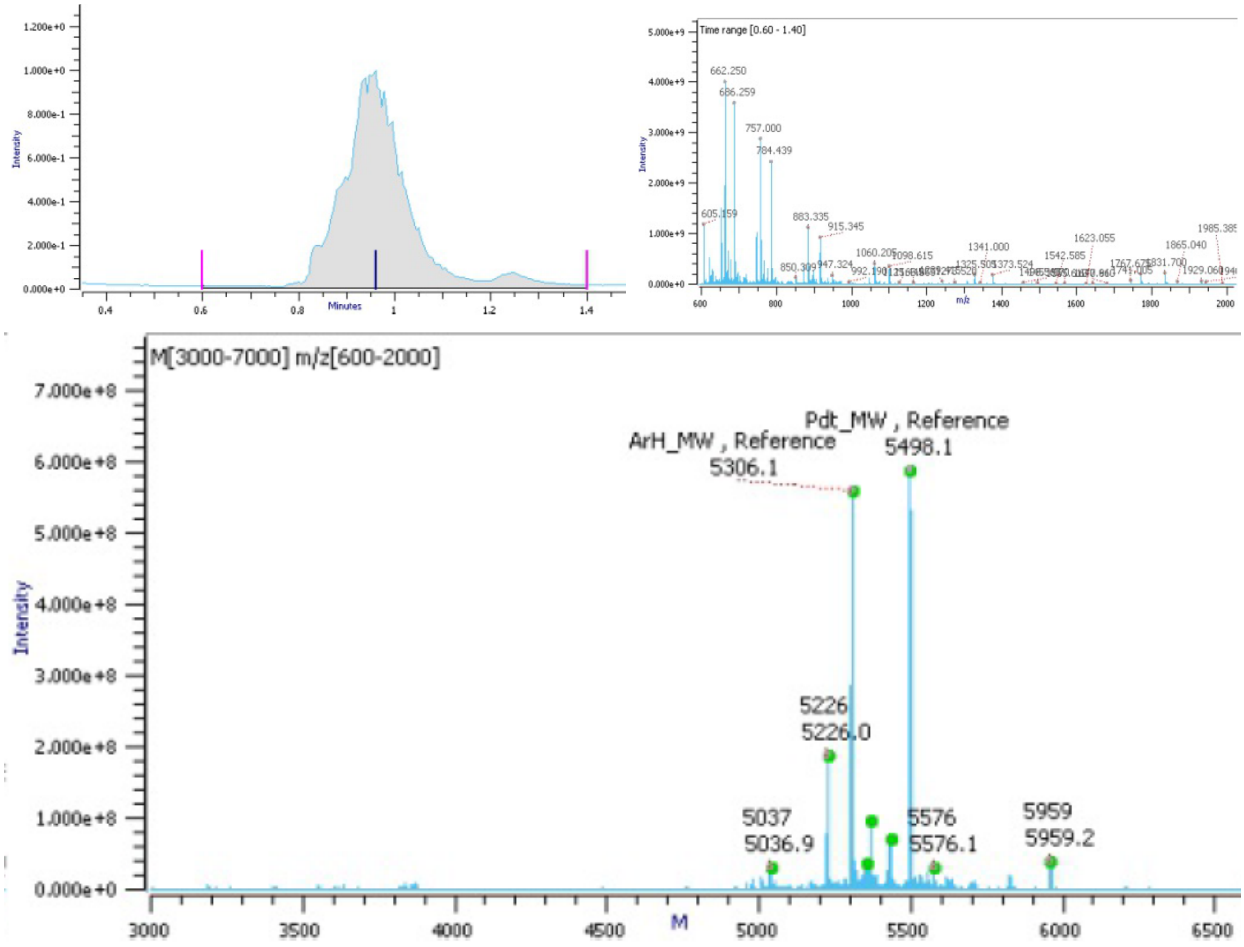
ELTM_433, TW-N78441-10-433.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



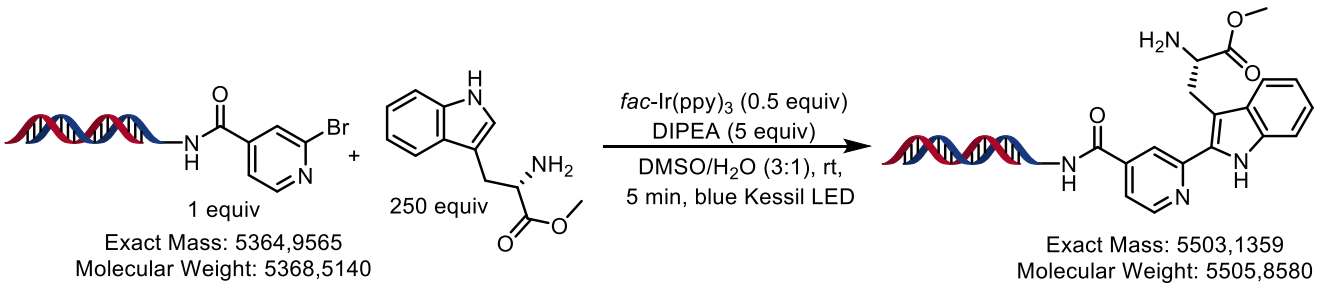
Product **4ac**, 43% yield



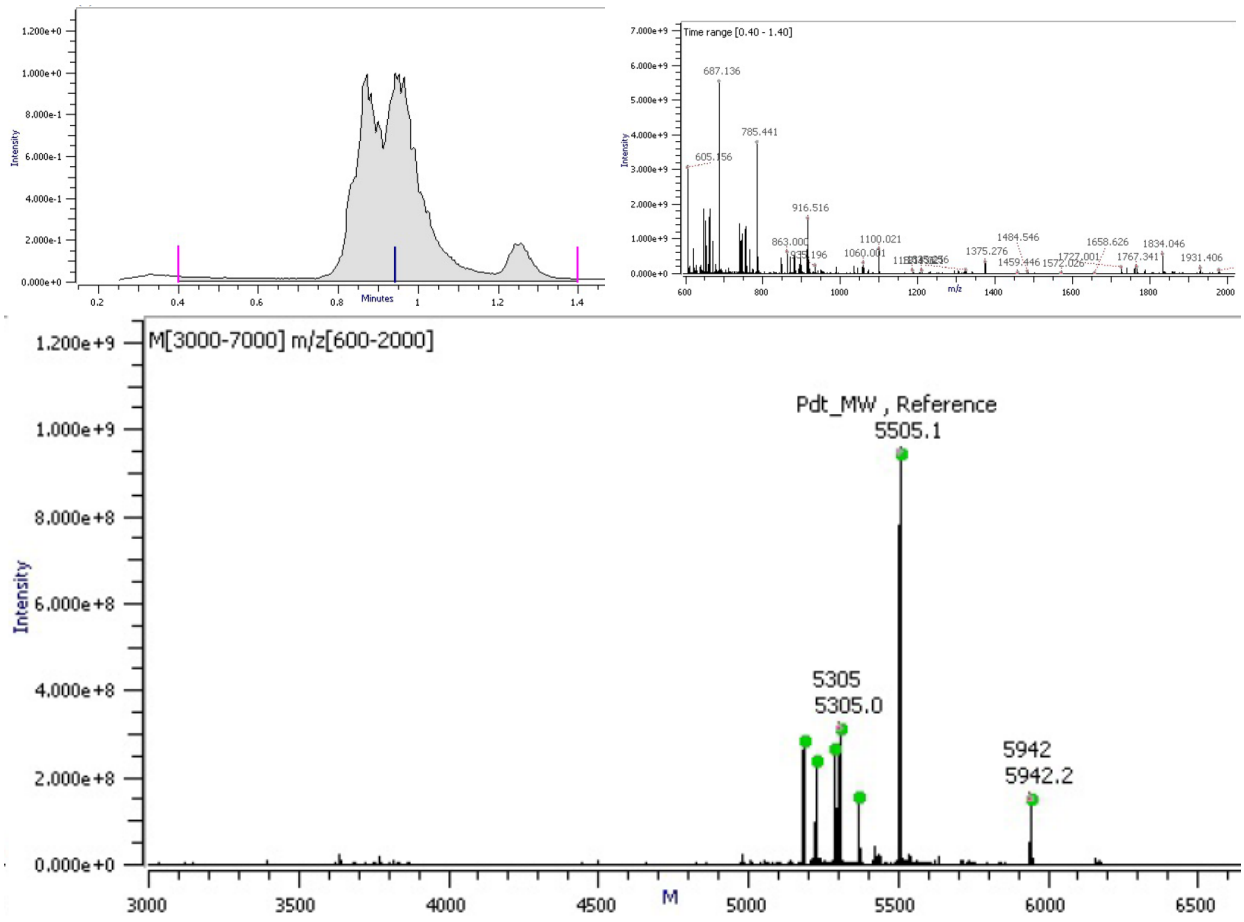
ELTM_388, TW-N78441-9-388.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



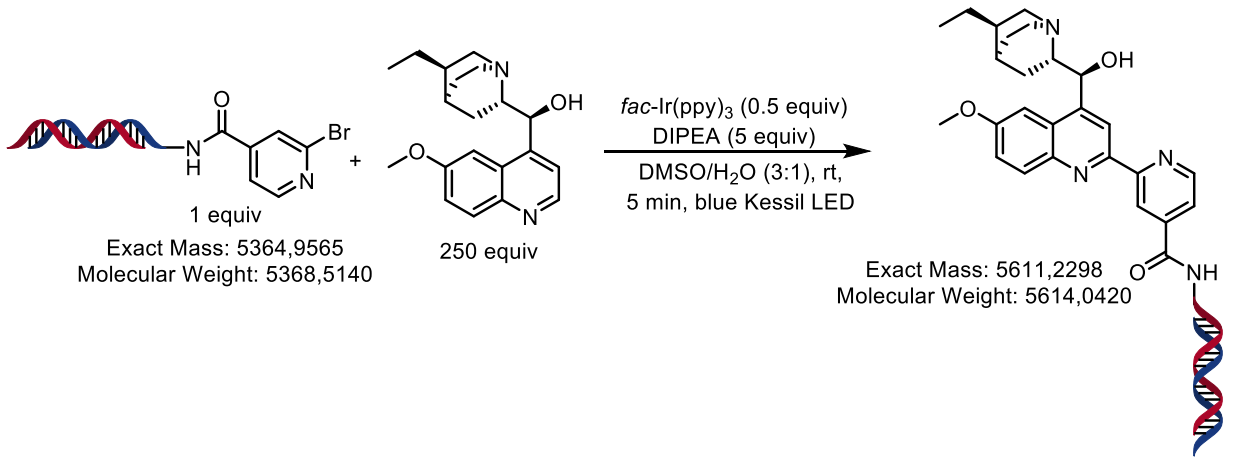
Product **6ad**, 54% yield



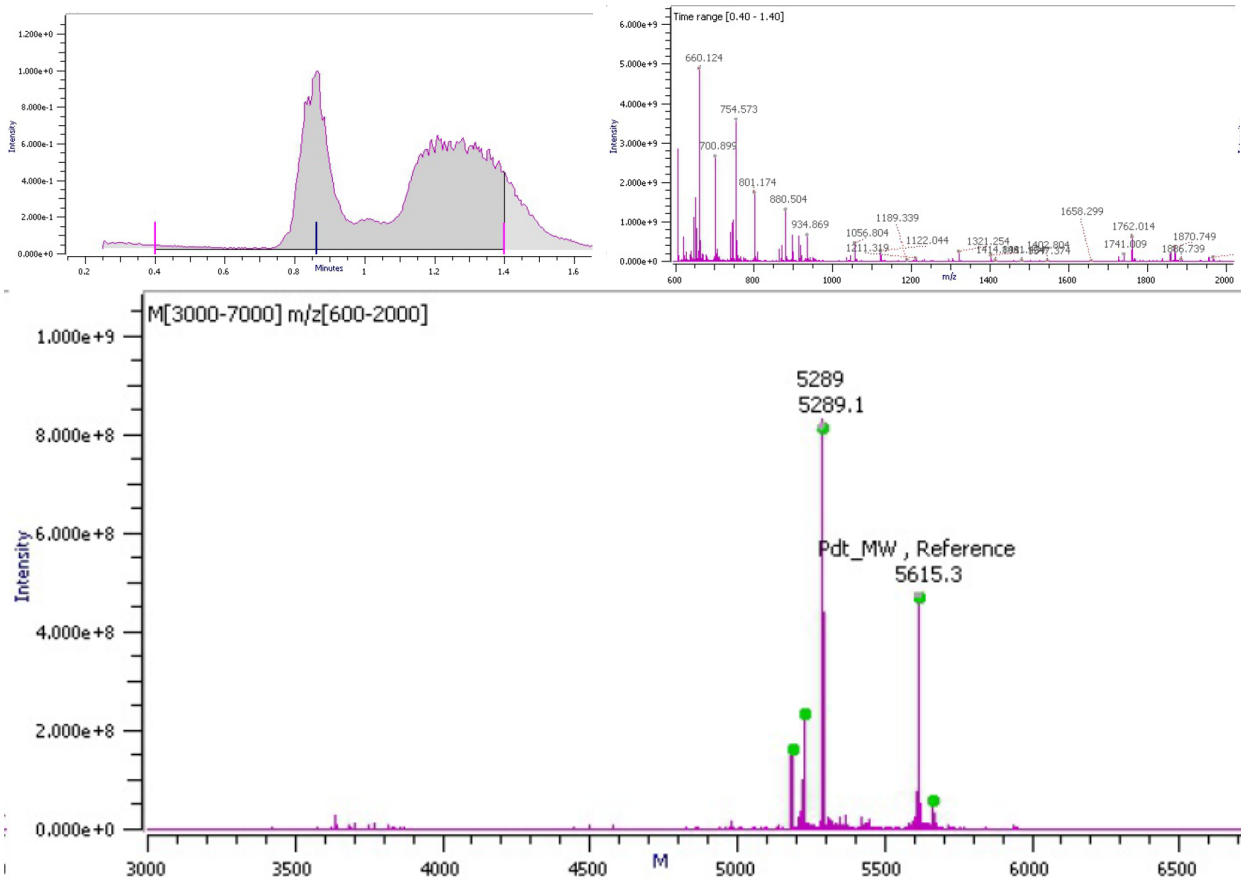
ELTM_480, TW-N78441-10-480.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



Product **6ae**, 37% yield

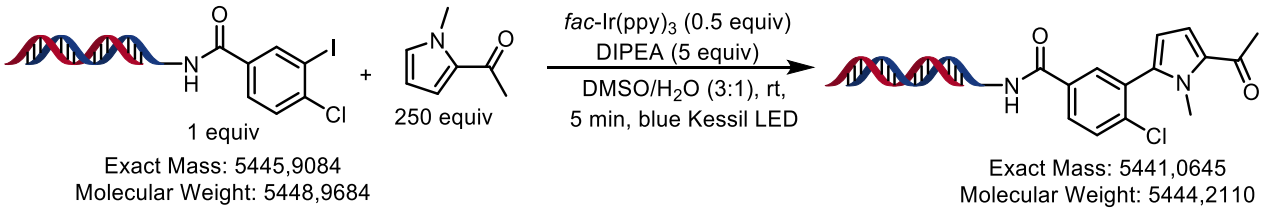


ELTM_464, TW-N78441-10-464.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



Variation of headpiece

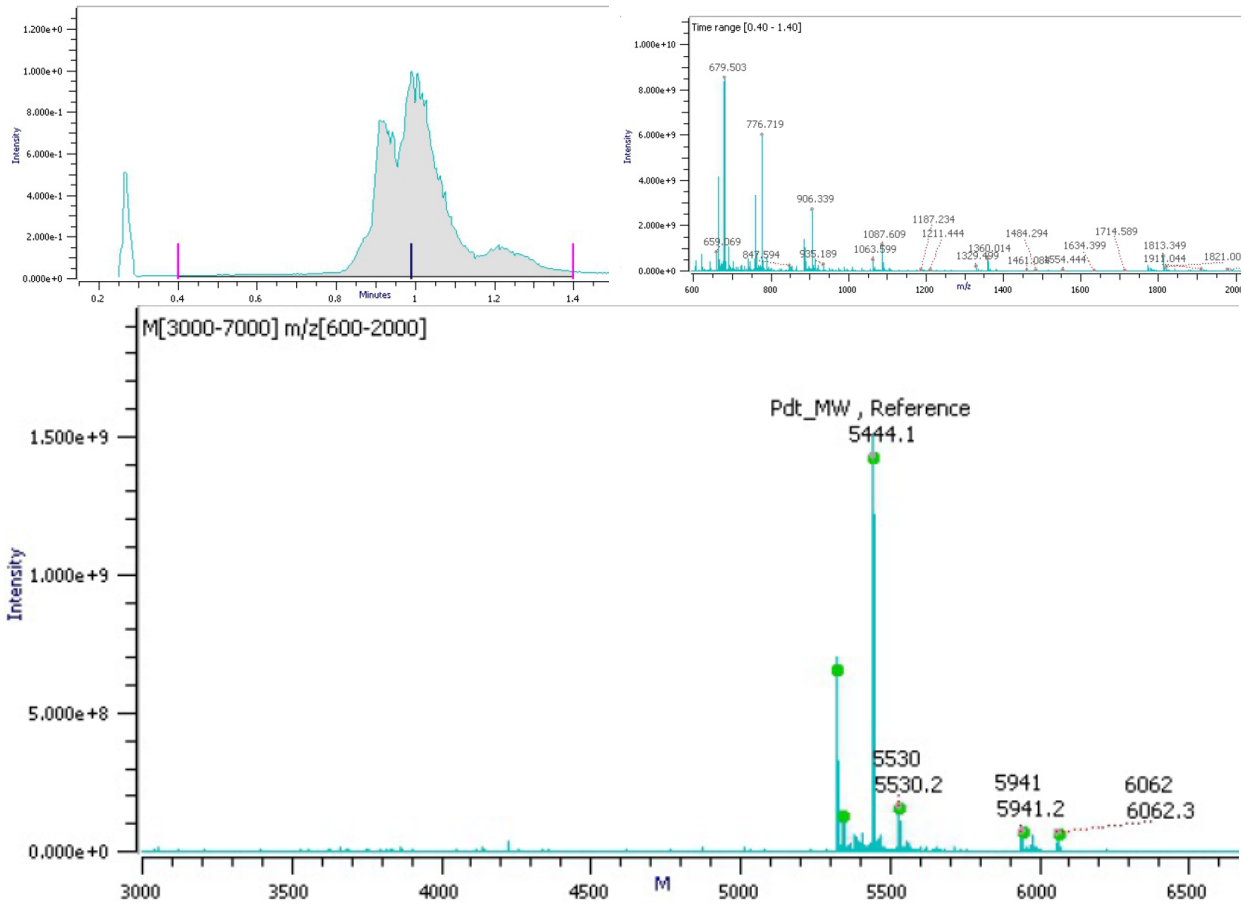
Product **5g**, 67% yield



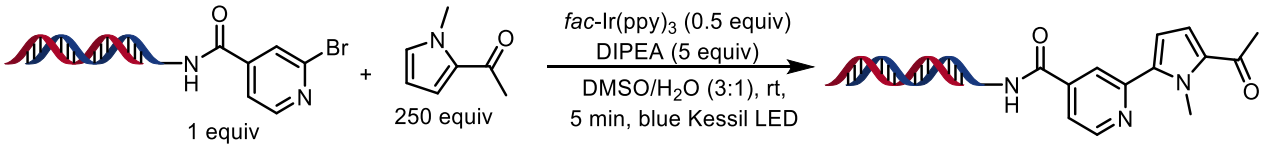
Exact Mass: 5445,9084
Molecular Weight: 5448,9684

Exact Mass: 5441,0645
Molecular Weight: 5444,2110

ELTM_570, TW-N78441-10-570.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



Product **6g**, 54% yield

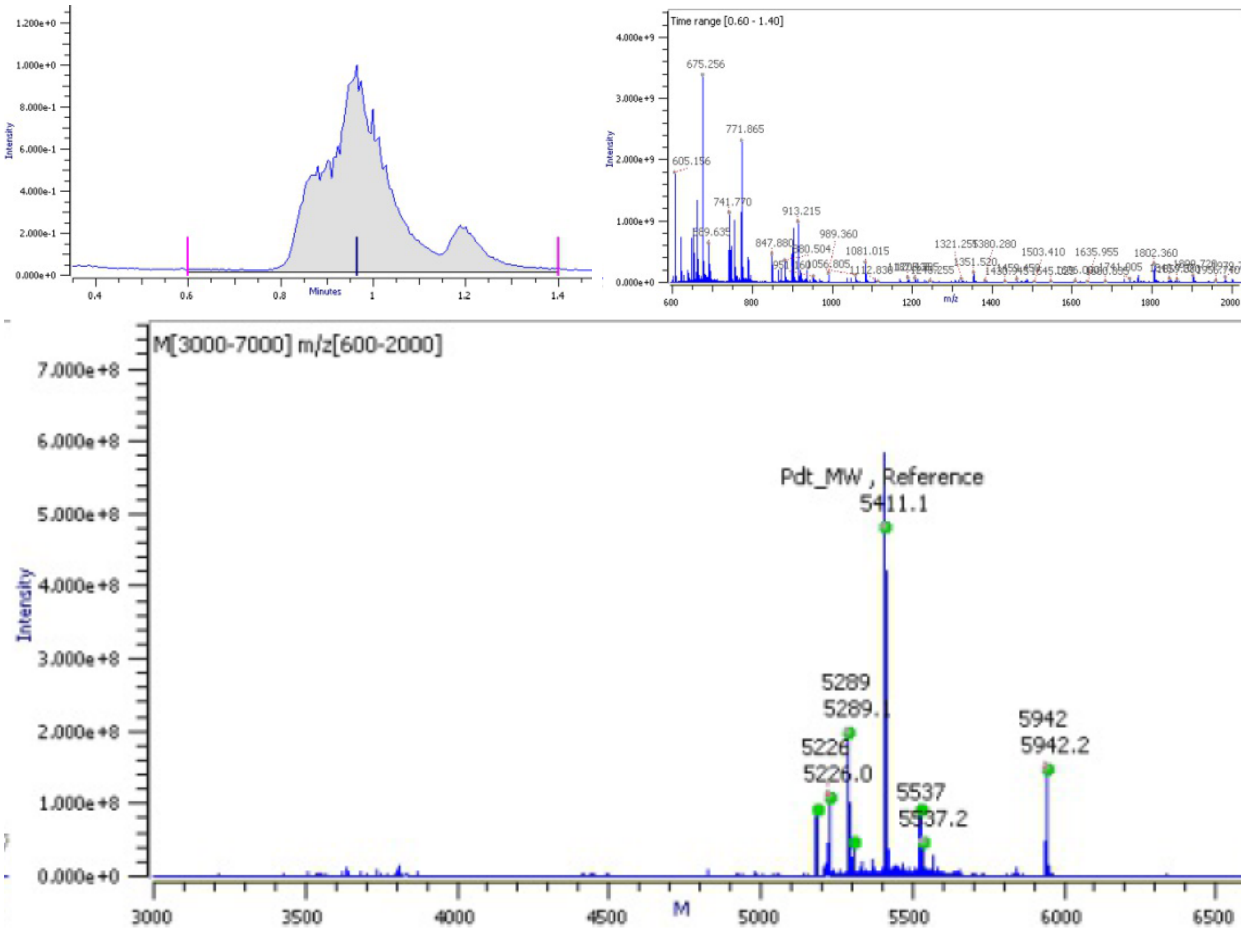


1 equiv
Exact Mass: 5364,9565
Molecular Weight: 5368,5140

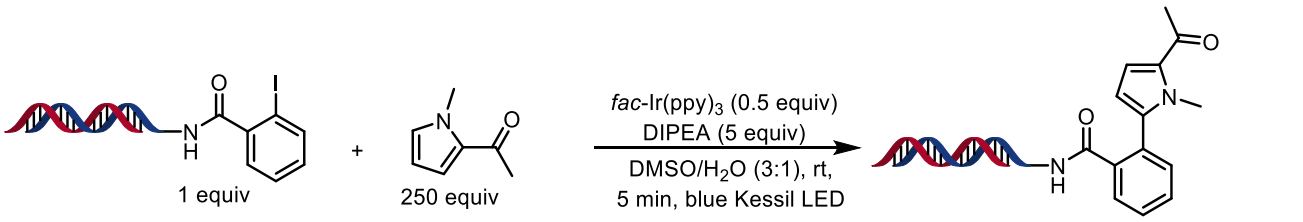
250 equiv

Exact Mass: 5408,0988
Molecular Weight: 5410,7570

ELTM_420, TW-N78441-9-420.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



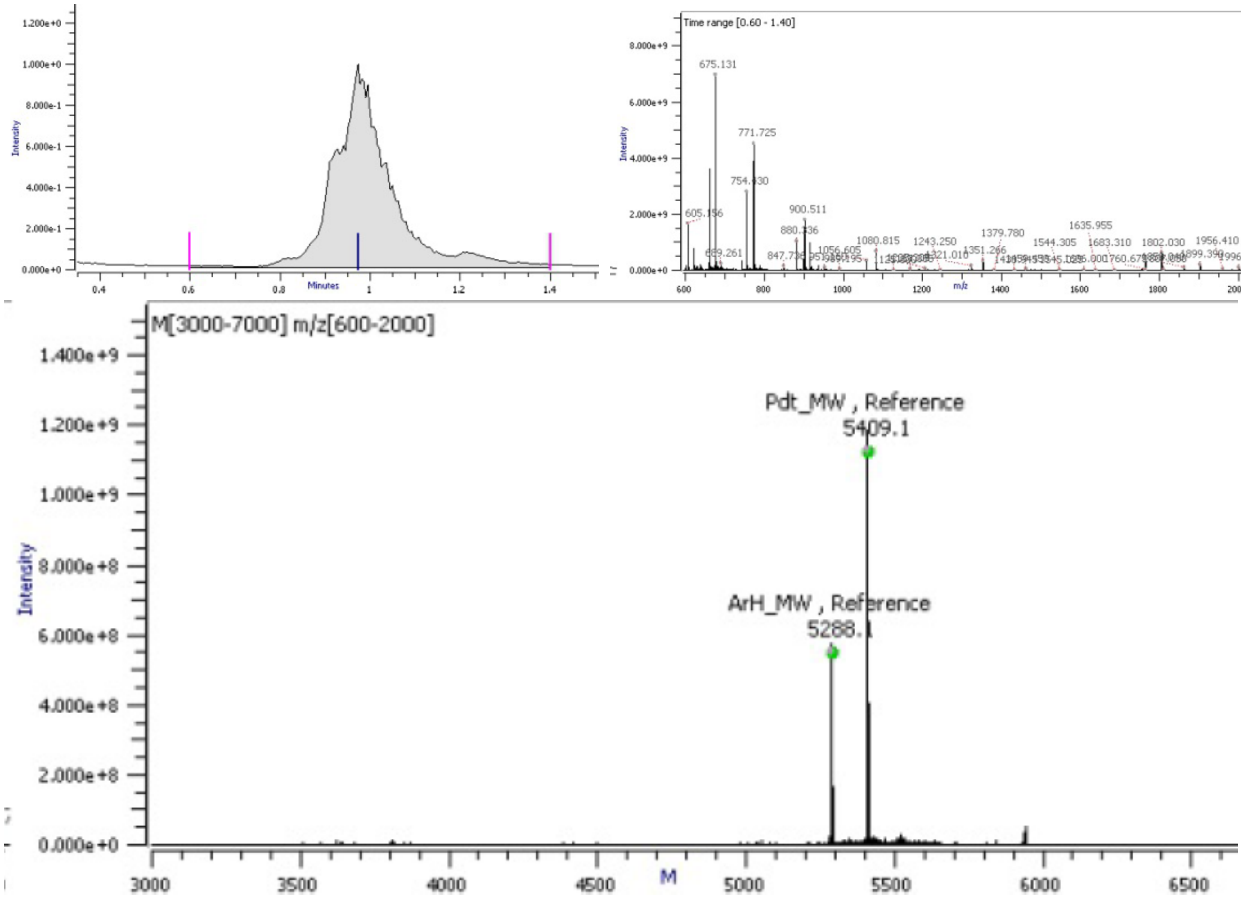
Product **7g**, 71% yield



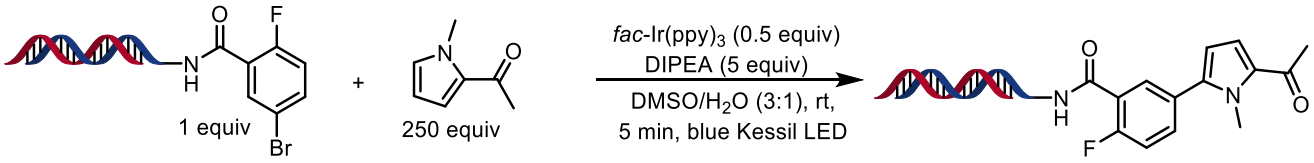
Exact Mass: 5411,9474
Molecular Weight: 5414,5264

Exact Mass: 5407,1035
Molecular Weight: 5409,7690

ELTM_426, TW-N78441-9-426.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



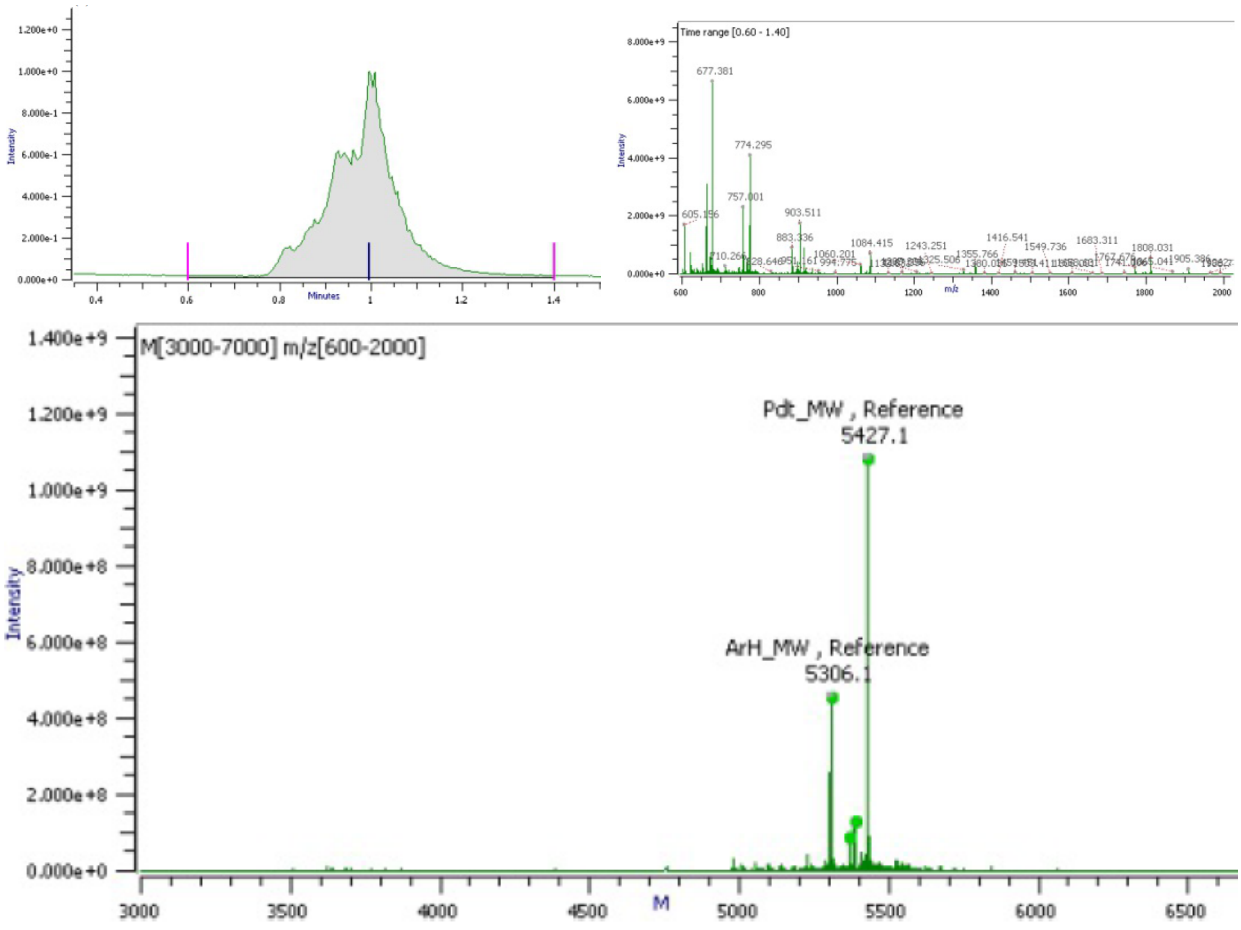
Product **8g**, 61% yield



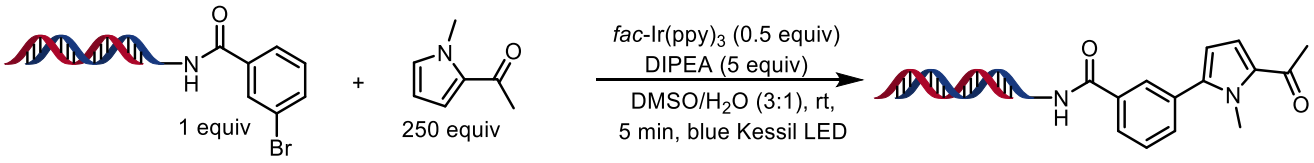
Exact Mass: 5505,0202
Molecular Weight: 5508,6714

Exact Mass: 5425,0941
Molecular Weight: 5427,7594

ELTM_326, TW-N78441-9-326.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



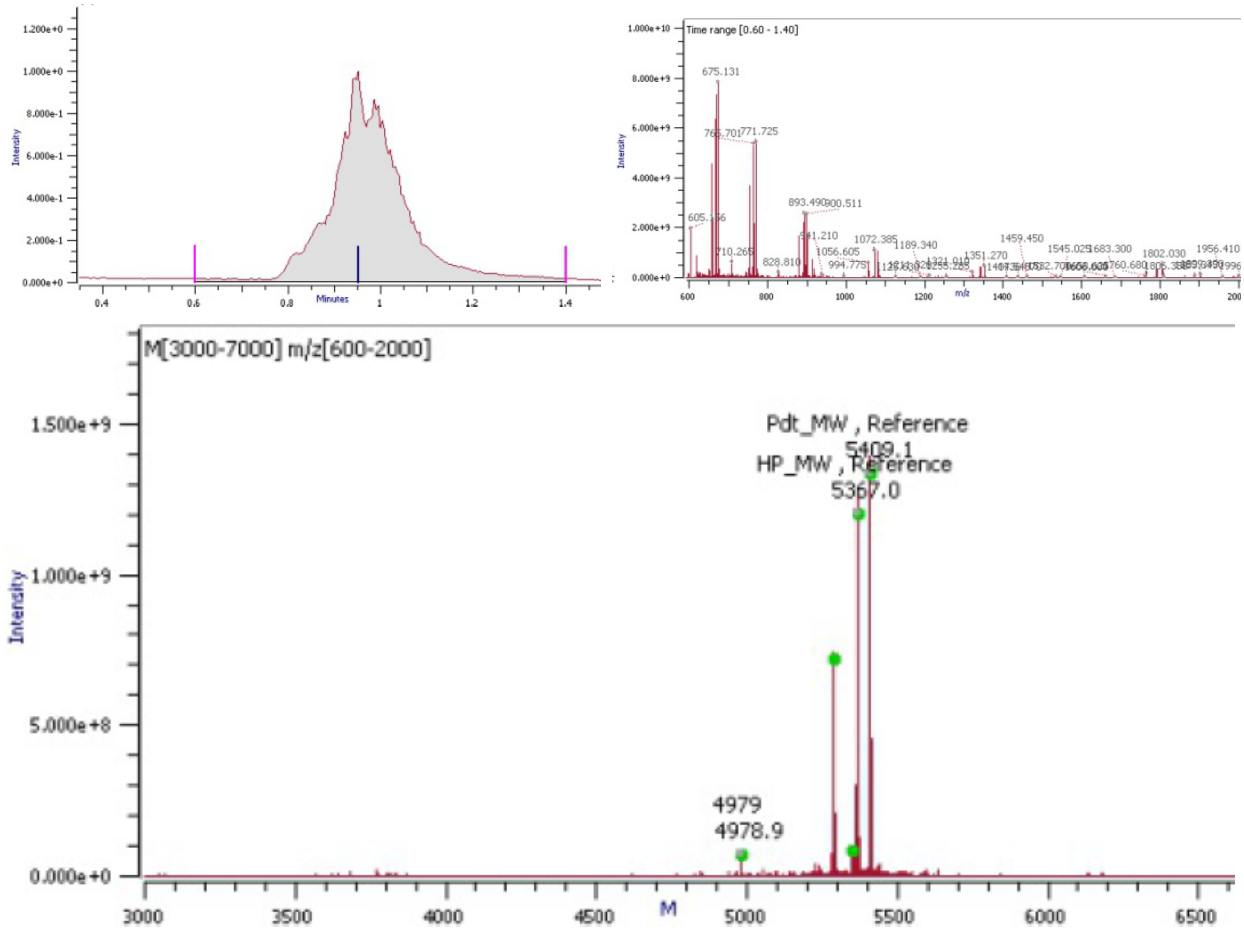
Product **9g**, 41% yield



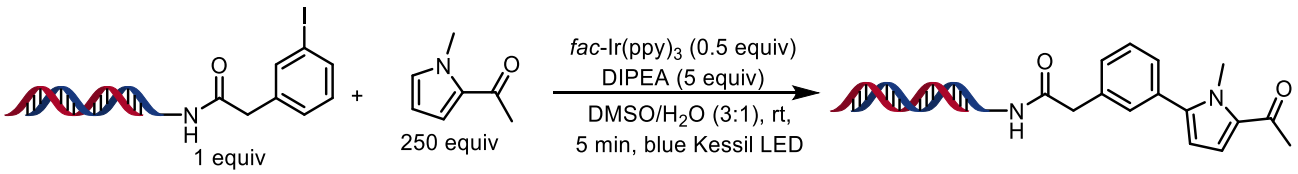
Exact Mass: 5363,9613
Molecular Weight: 5367,5260

Exact Mass: 5407,1035
Molecular Weight: 5409,7690

ELTM_323, TW-N78441-9-323.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



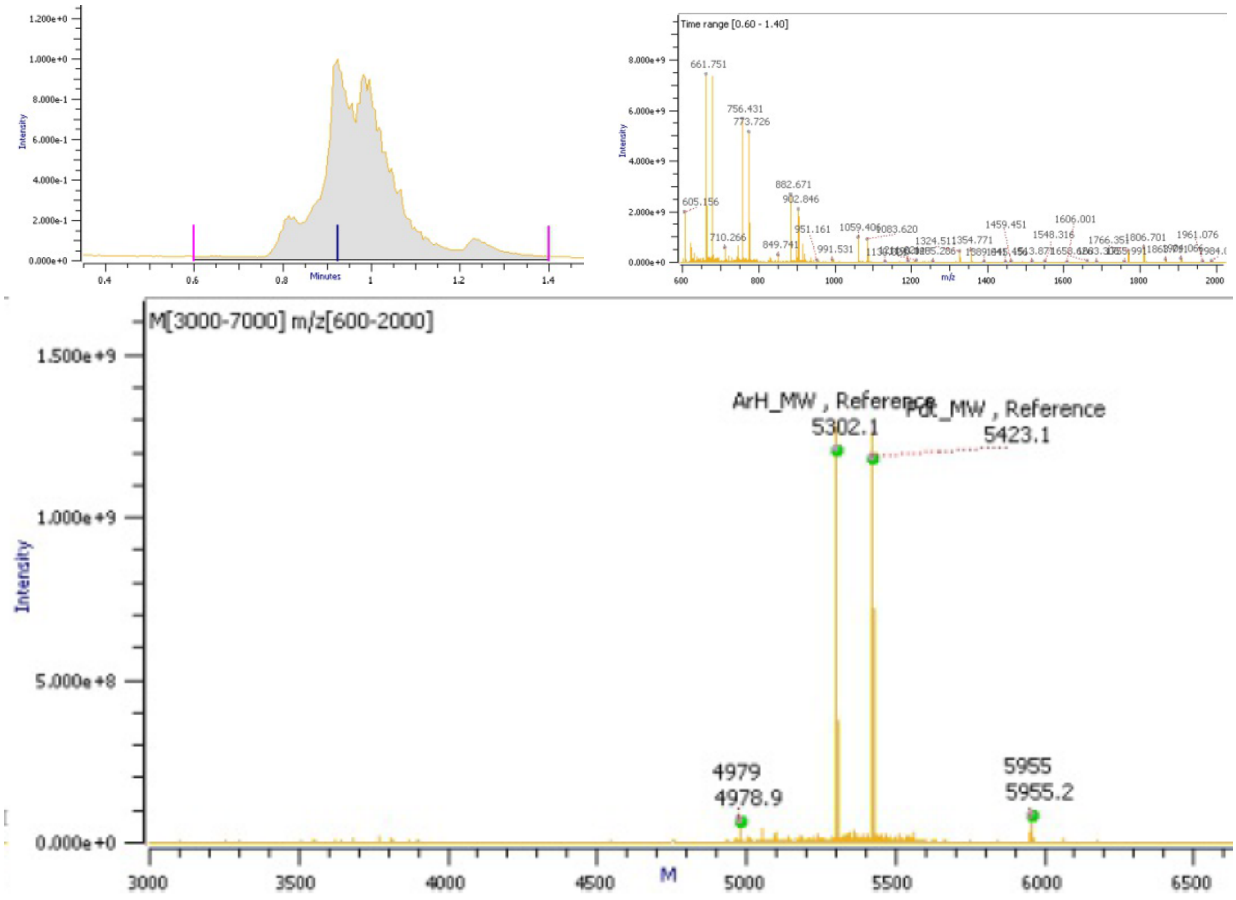
Product **10g**, 51% yield



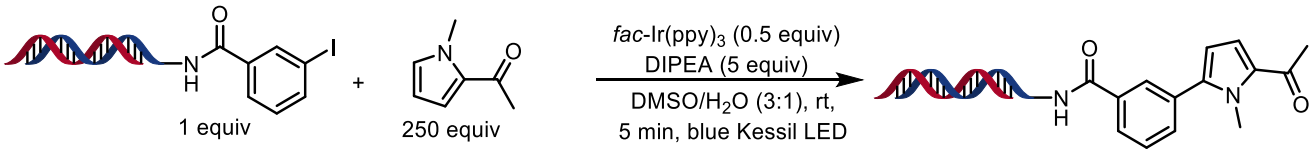
Exact Mass: 5549,0314
Molecular Weight: 5551,7084

Exact Mass: 5421,1192
Molecular Weight: 5423,7960

ELTM_344, TW-N78441-9-344.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



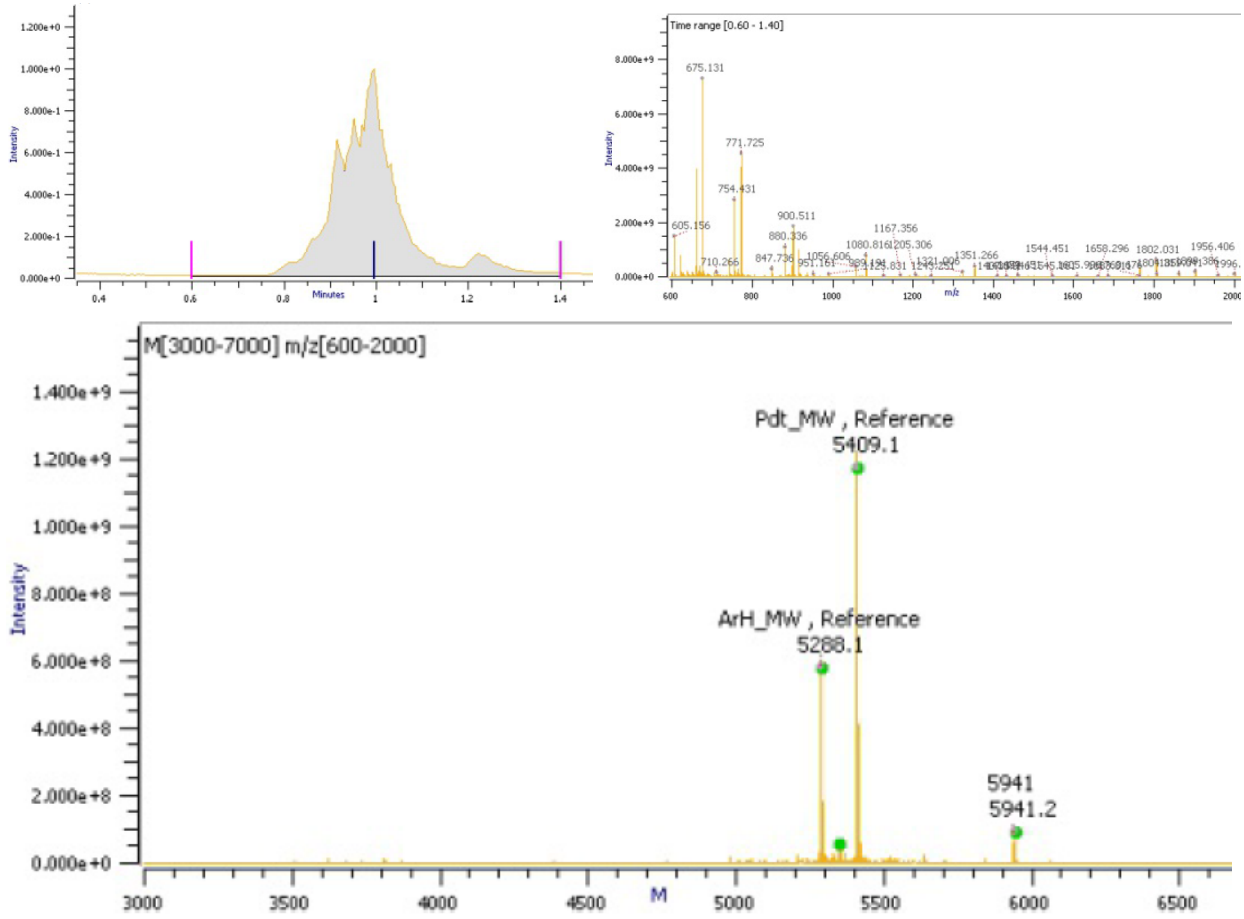
Product **11g**, 68% yield



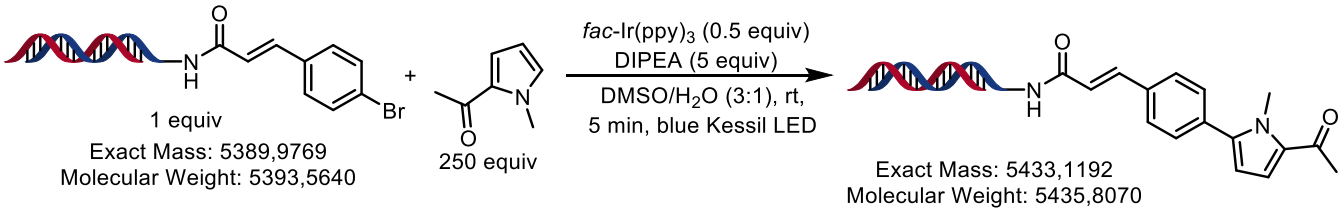
Exact Mass: 5411,9474
Molecular Weight: 5414,5264

Exact Mass: 5407,1035
Molecular Weight: 5409,7690

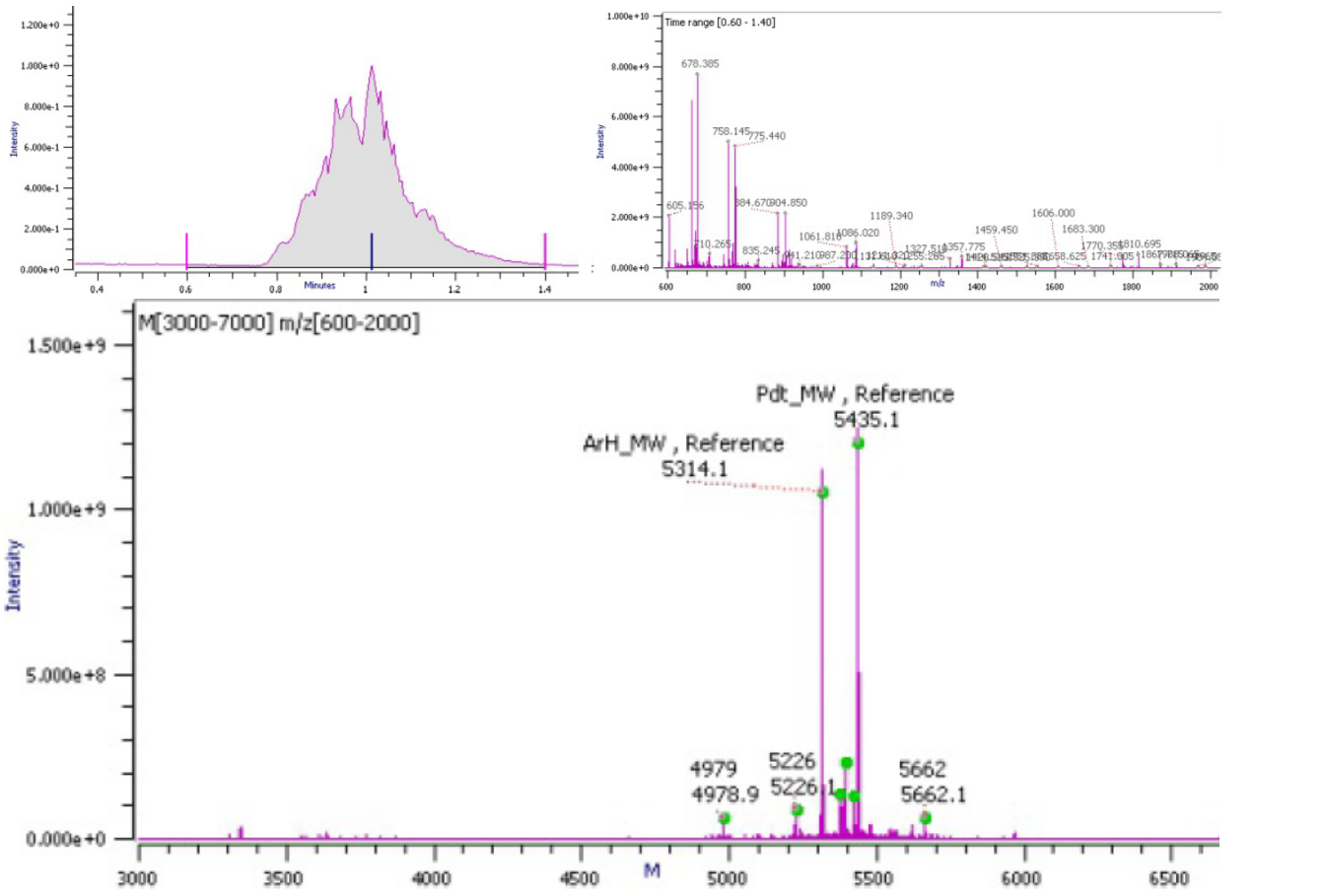
ELTM_301, TW-N78441-9-301.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



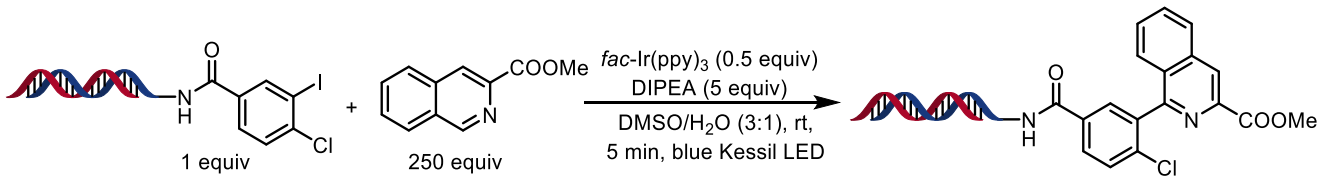
Product **12g**, 47% yield



ELTM_286, TW-N78441-9-286.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



Product **5a**, 45% yield

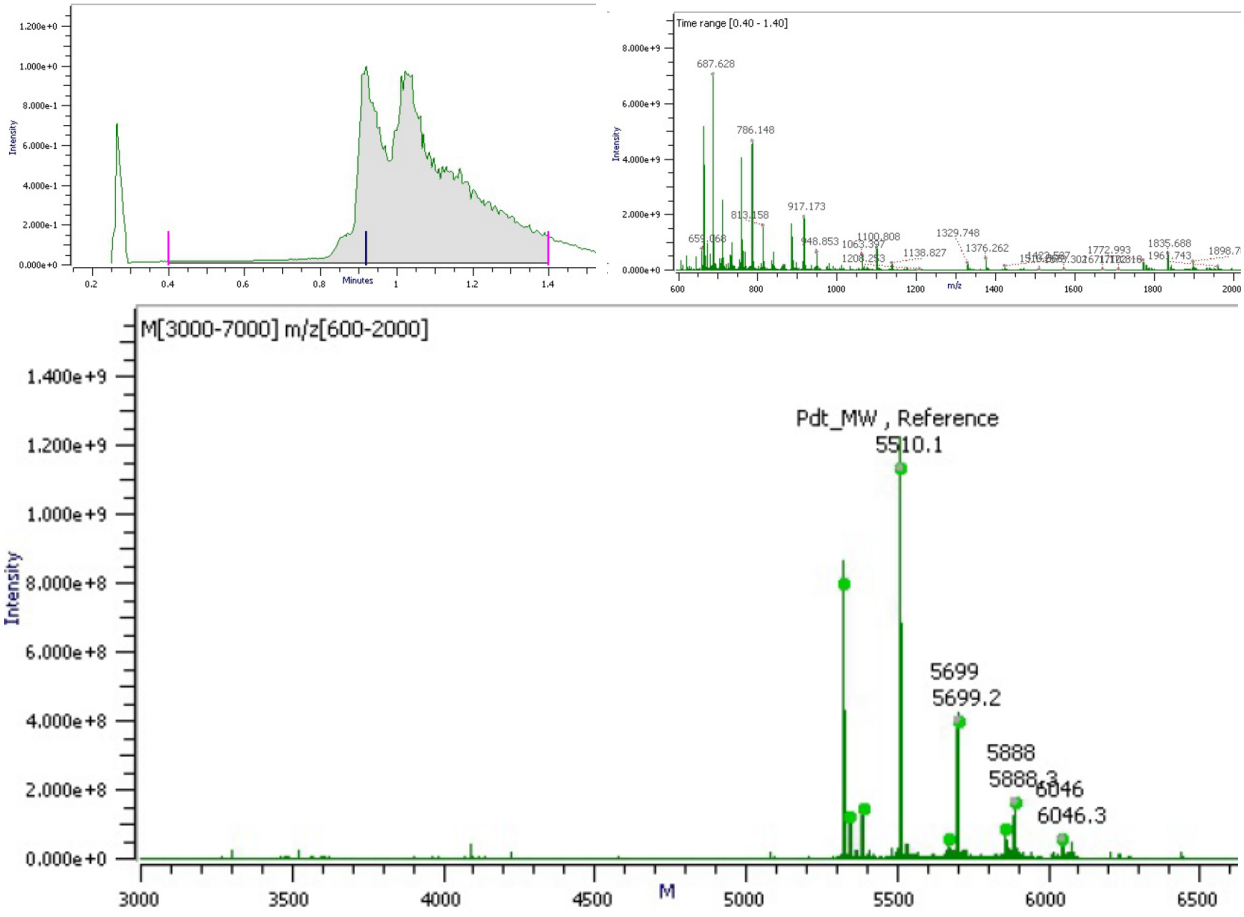


1 equiv
Exact Mass: 5445,9084
Molecular Weight: 5448,9684

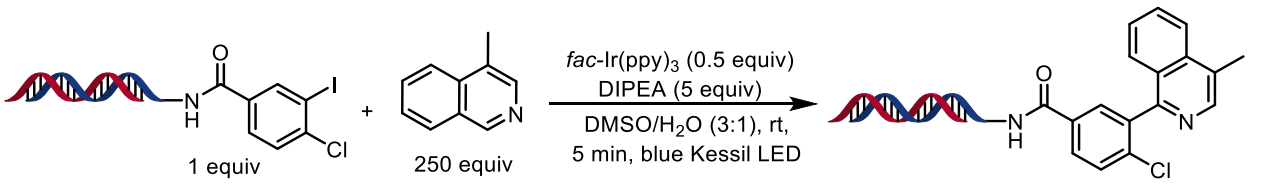
250 equiv

Exact Mass: 5505.0594
Molecular Weight: 5508.2540

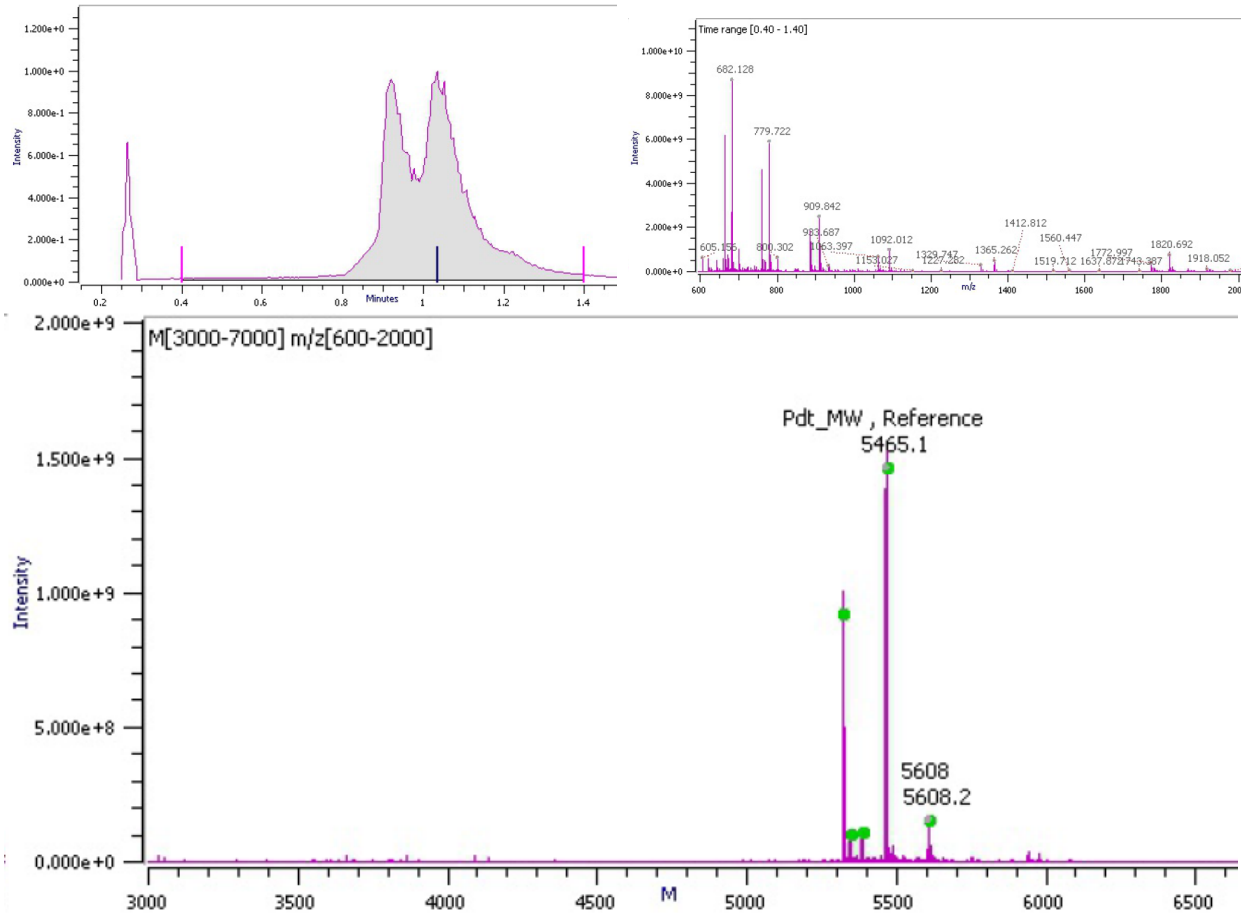
ELTM_569, TW-N78441-10-569.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



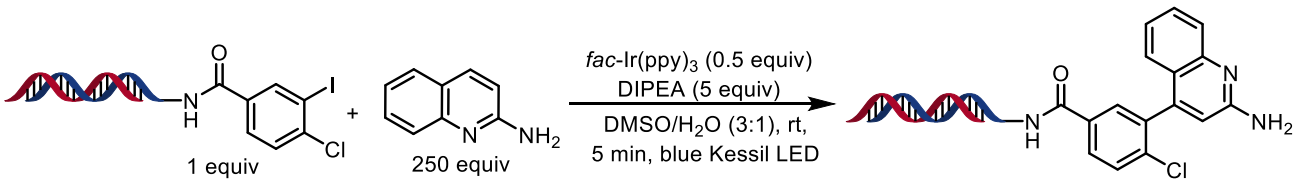
Product **5c**, 63% yield



ELTM_529, TW-N78441-10-529.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



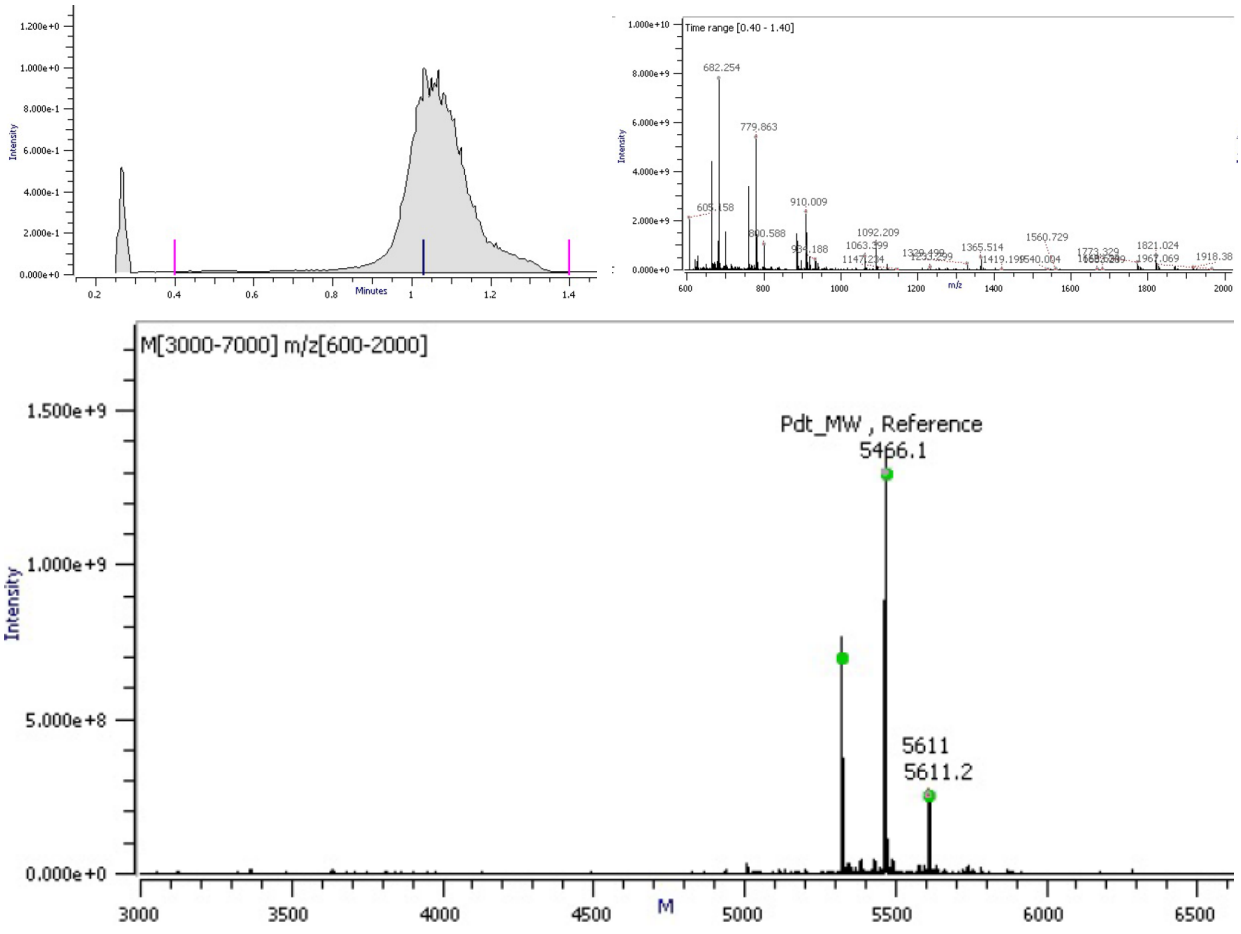
Product **5d**, 68% yield



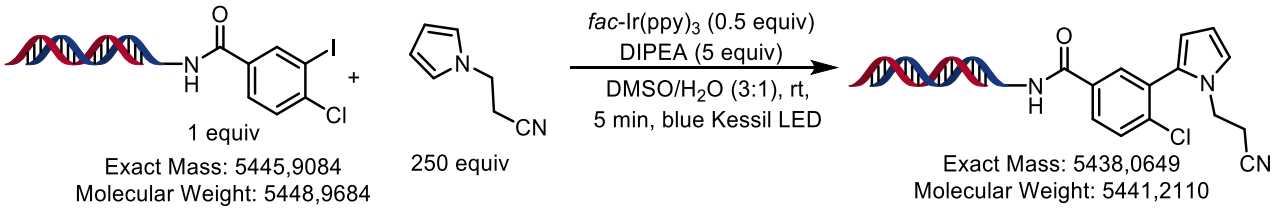
Exact Mass: 5445,9084
Molecular Weight: 5448,9684

Exact Mass: 5462,0645
Molecular Weight: 5465,2330

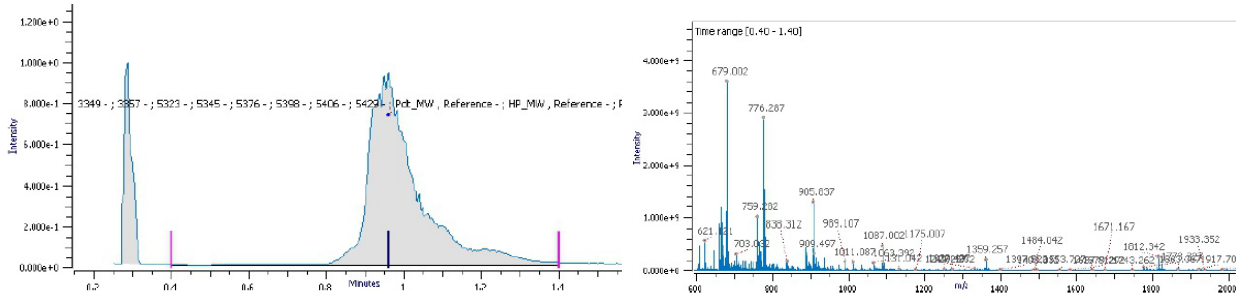
ELTM_532, TW-N78441-10-532.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)

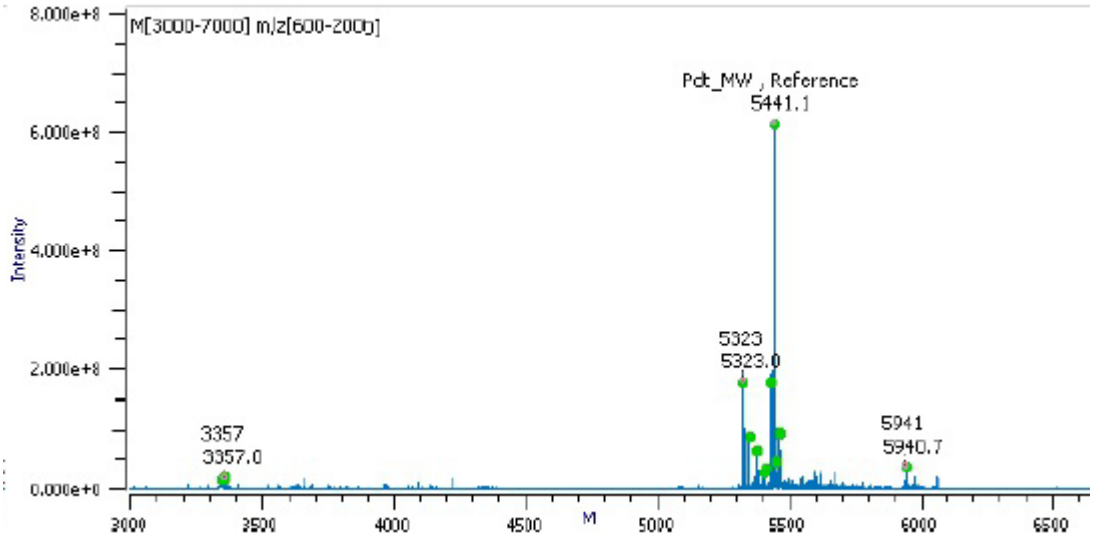


Product **5i**, 43% yield

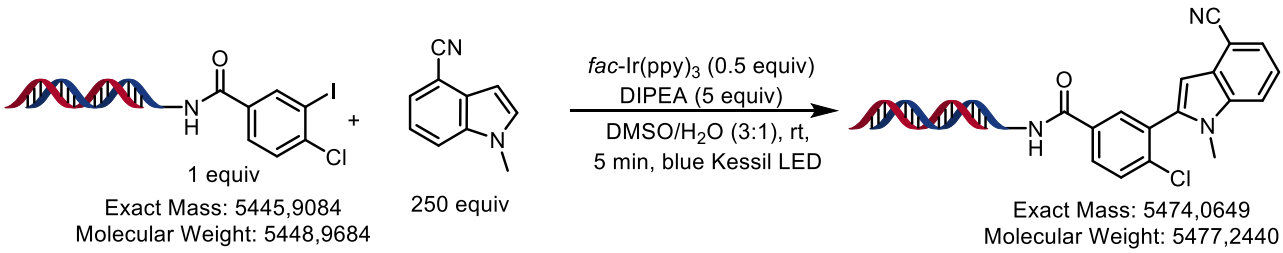


ELTM_650, TW-N78441-14-650.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)

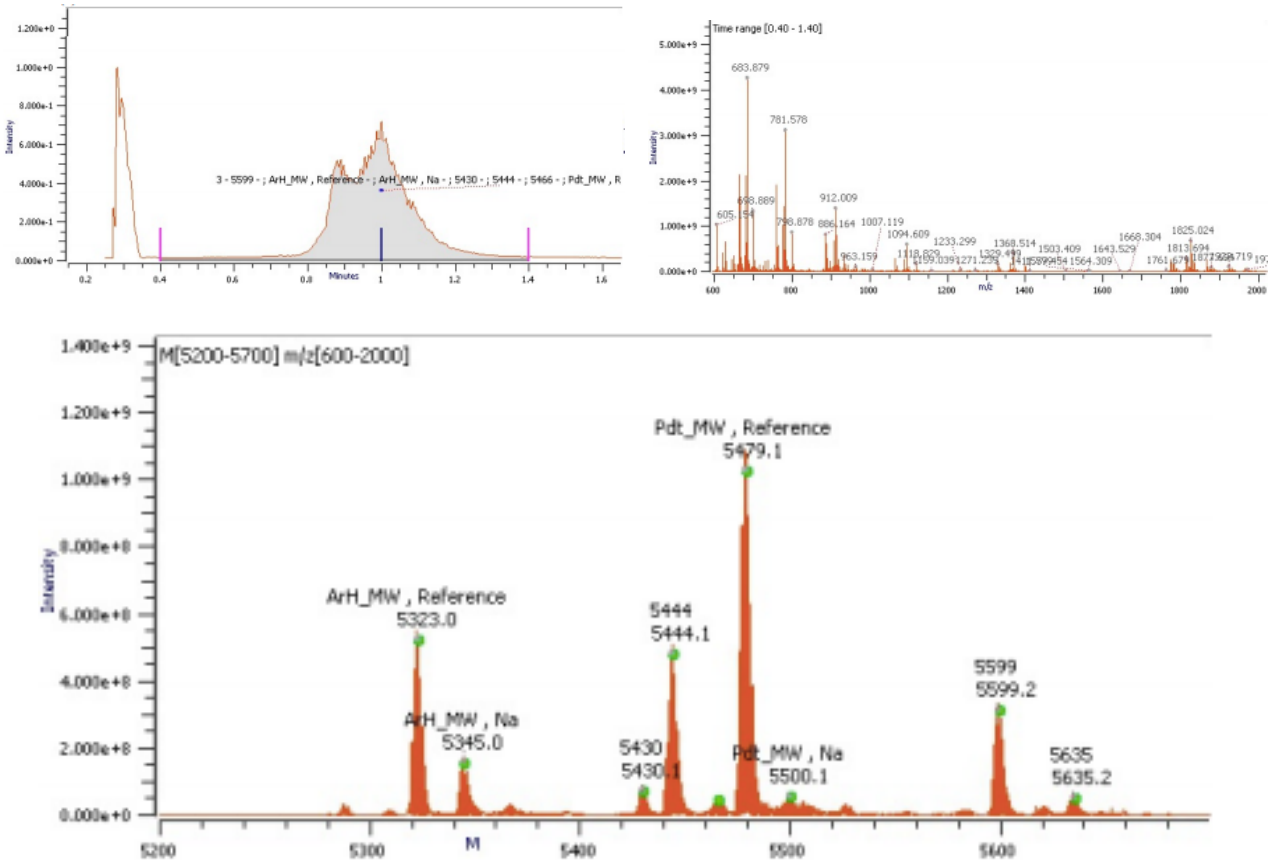




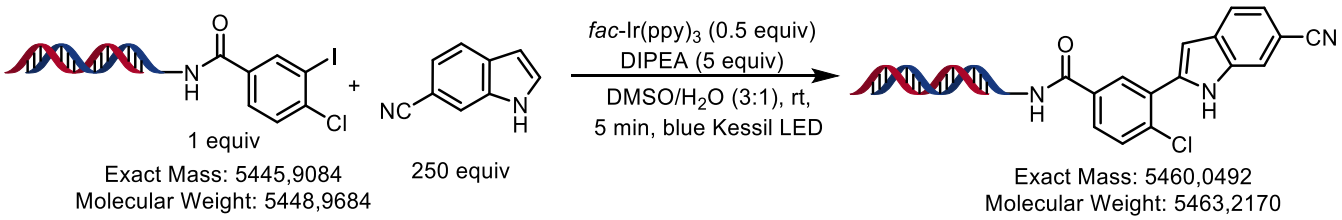
Product **5k**, 40% yield



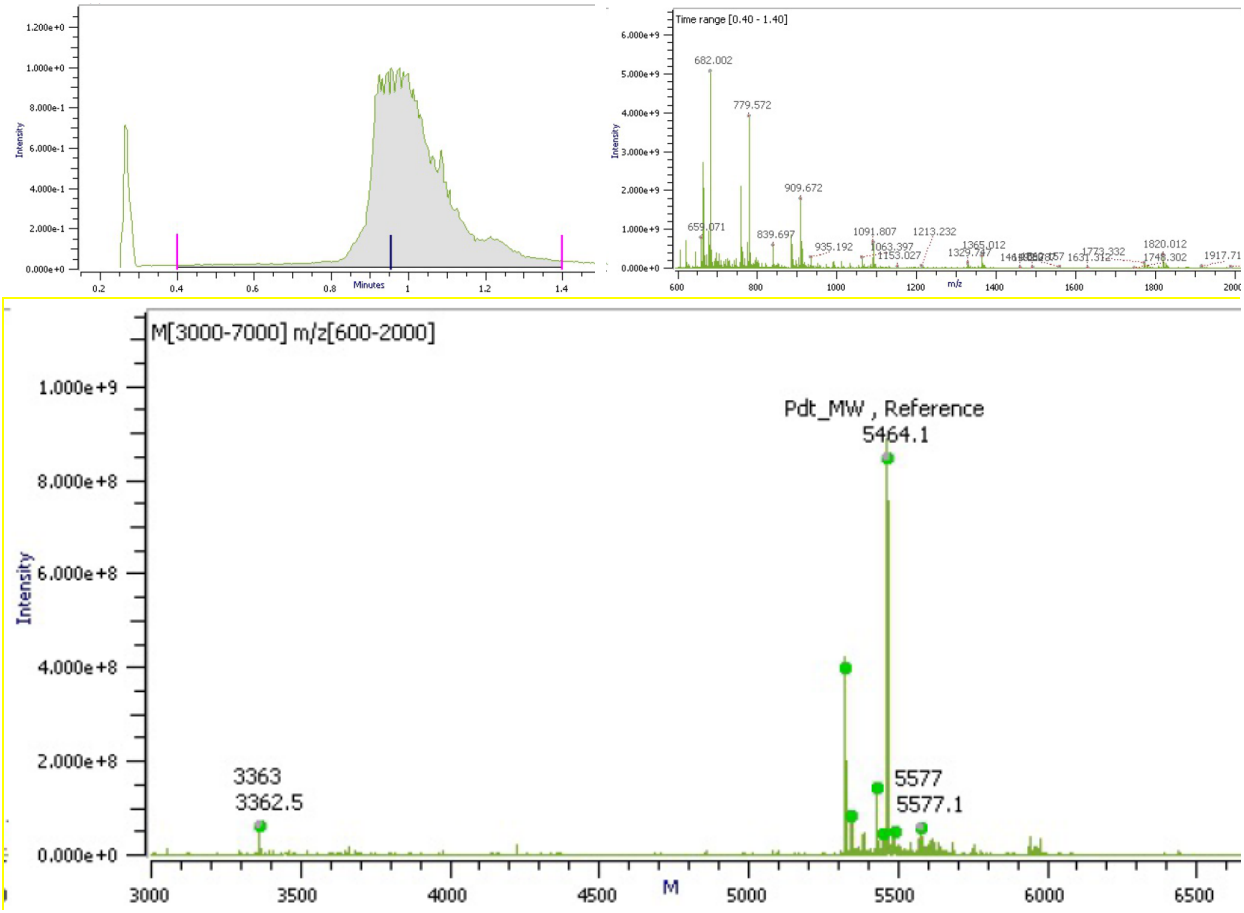
TW-N78441-15-702 - (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



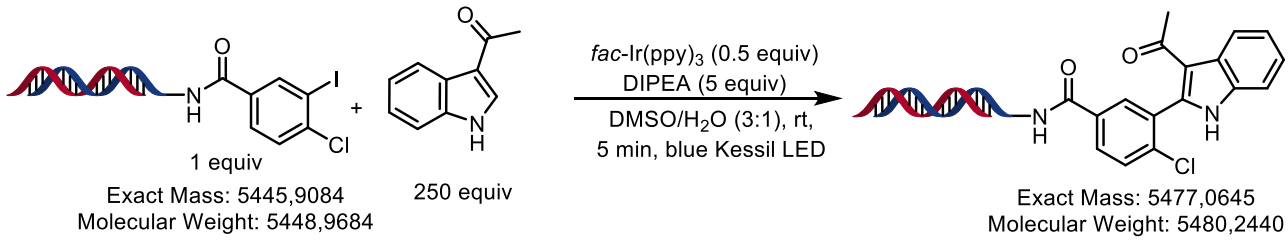
Product **5l**, 60% yield



ELTM_537, TW-N78441-10-537.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



Product **5o**, 77% yield

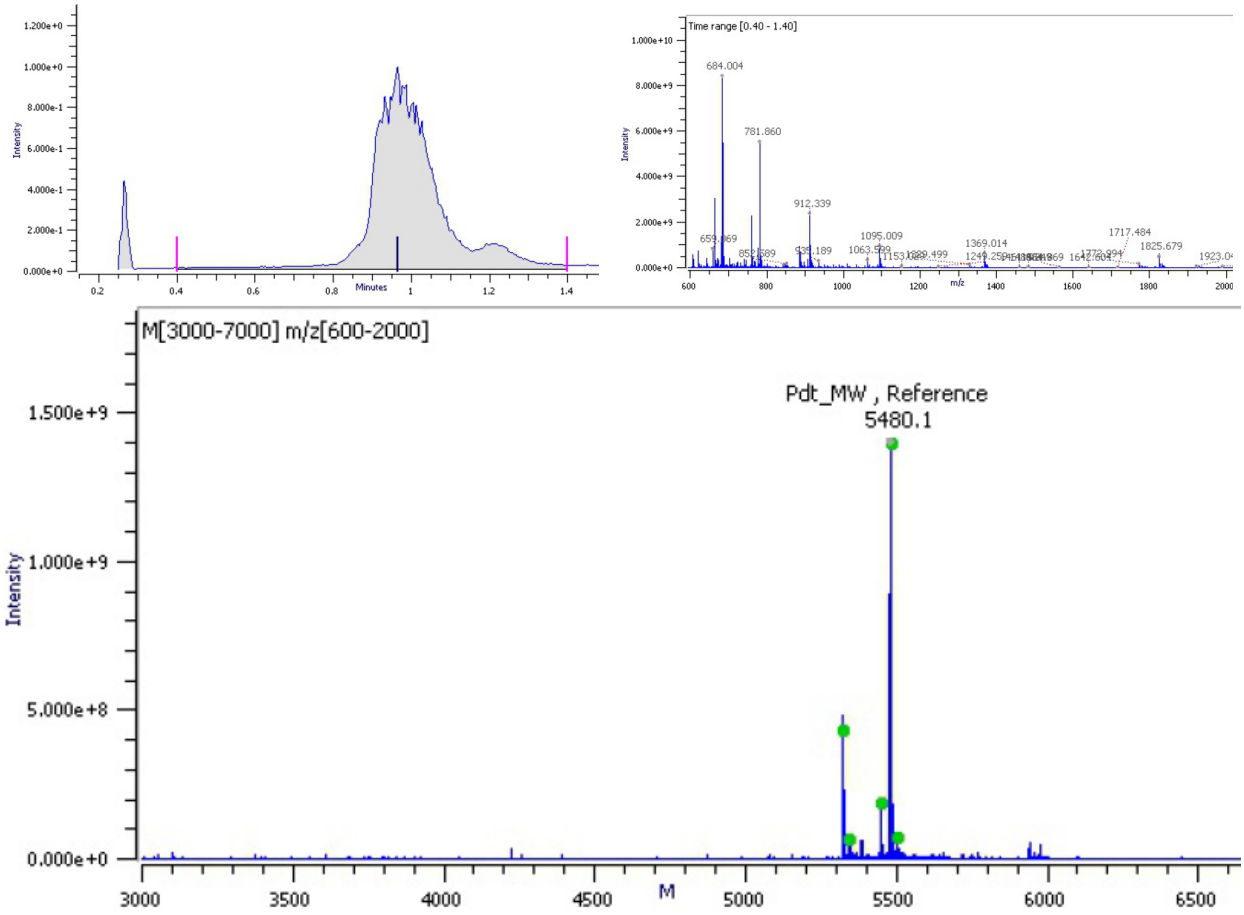


1 equiv
Exact Mass: 5445,9084
Molecular Weight: 5448,9684

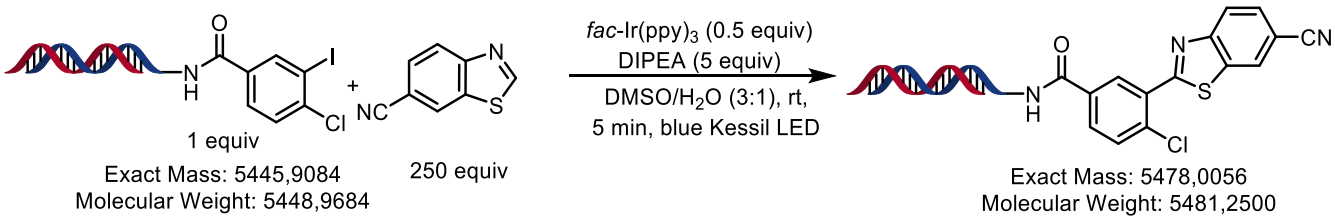
250 equiv

Exact Mass: 5477,0645
Molecular Weight: 5480,2440

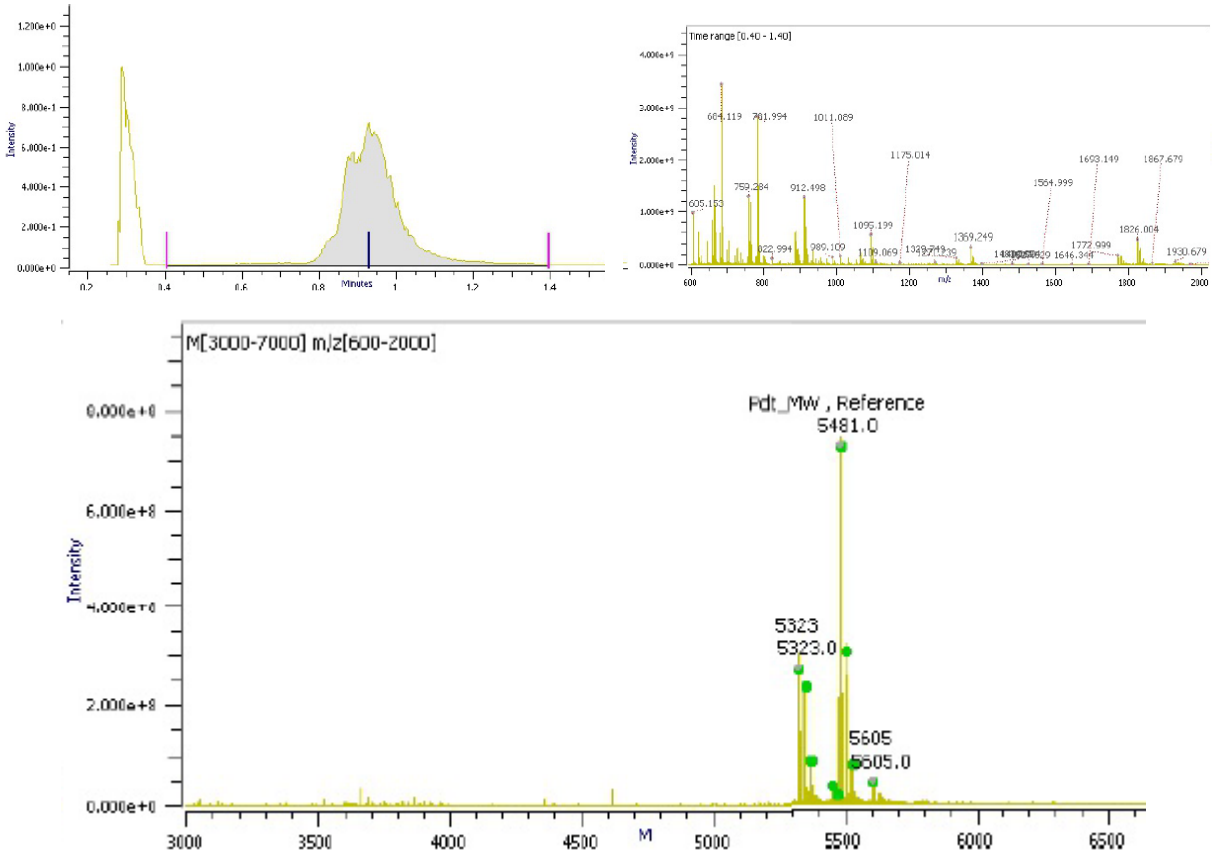
ELTM_539, TW-N78441-10-539.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



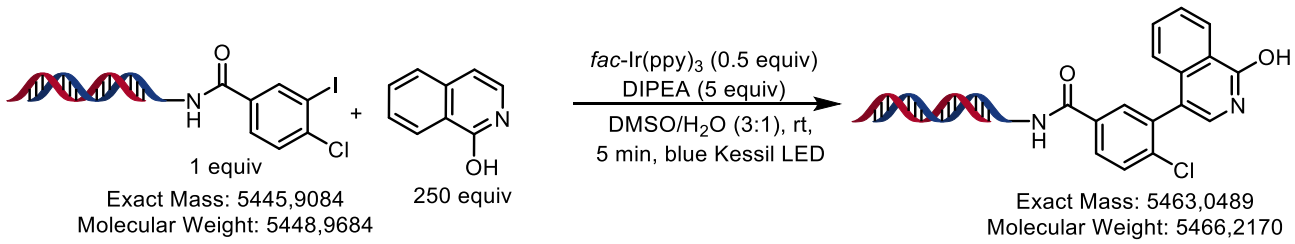
Product **5w**, 40% yield



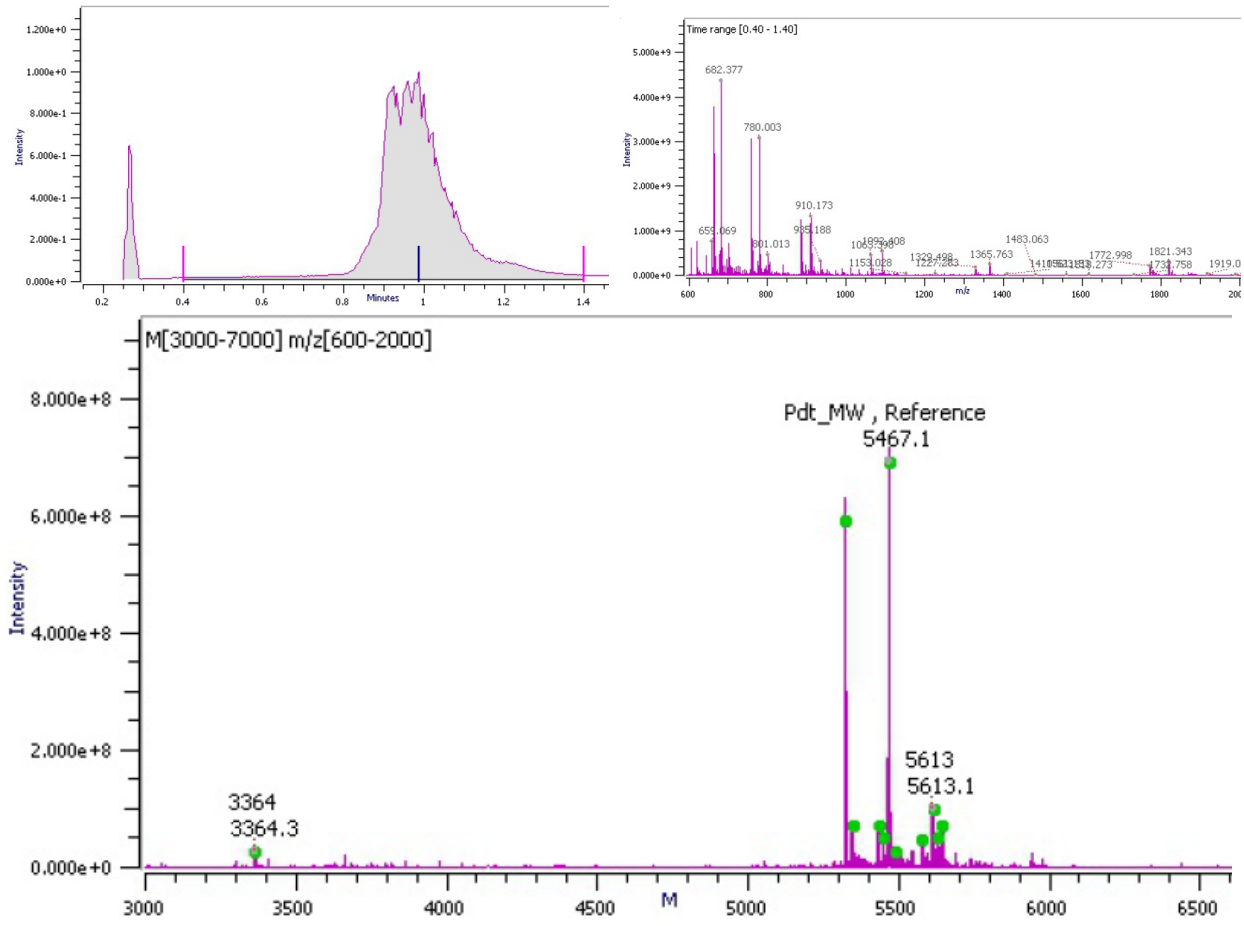
ELTM_712, TW-N78441-15-712.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



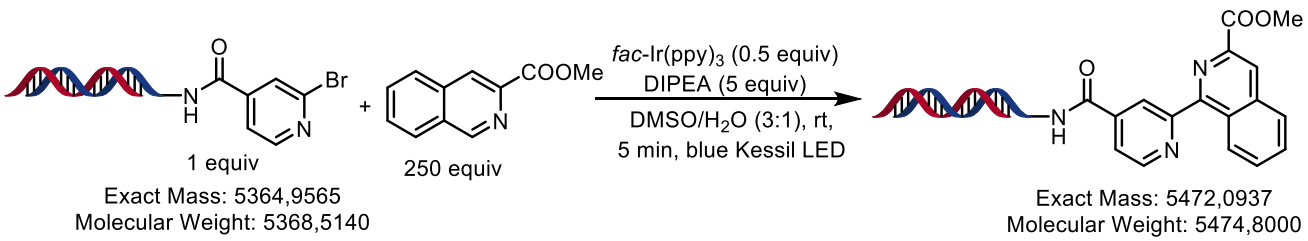
Product **5af**, 45% yield



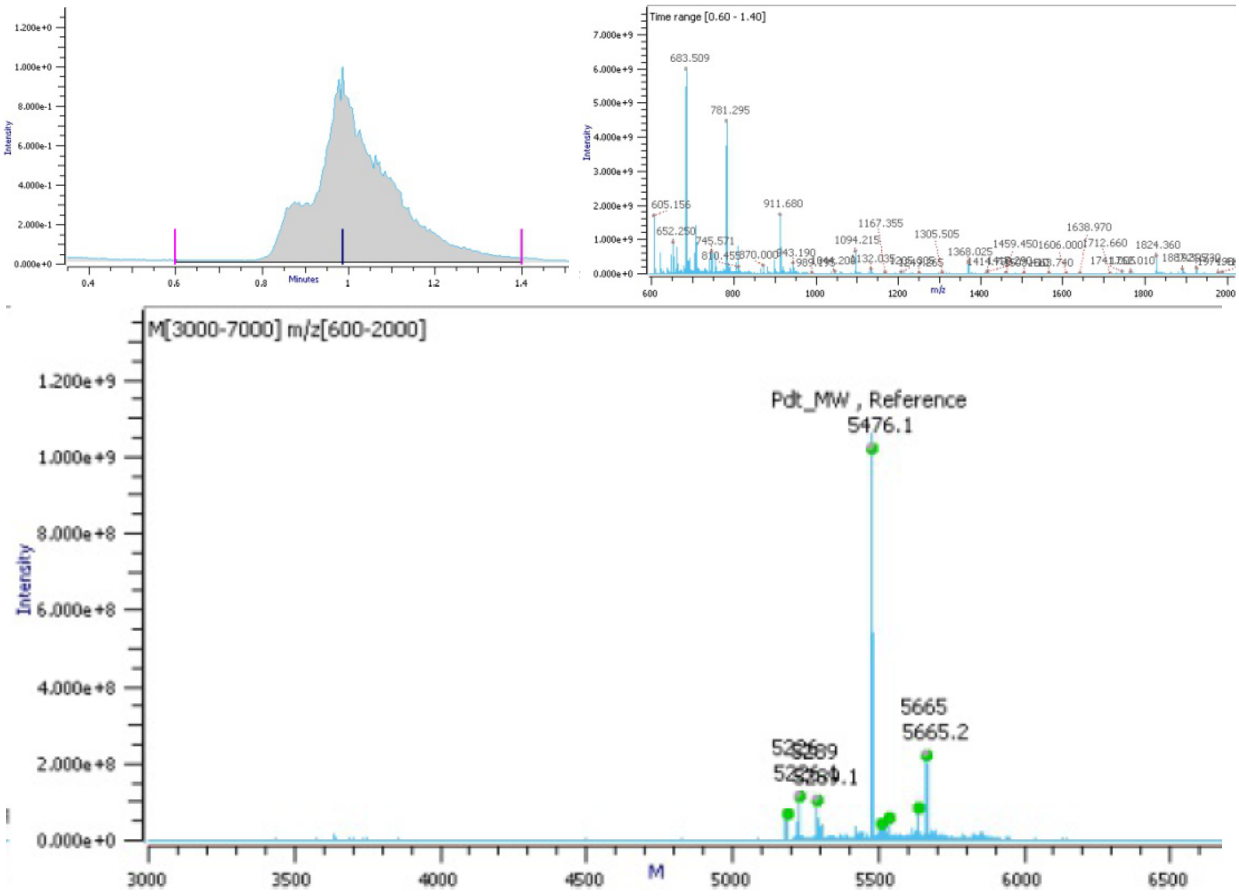
ELTM_542, TW-N78441-10-542.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



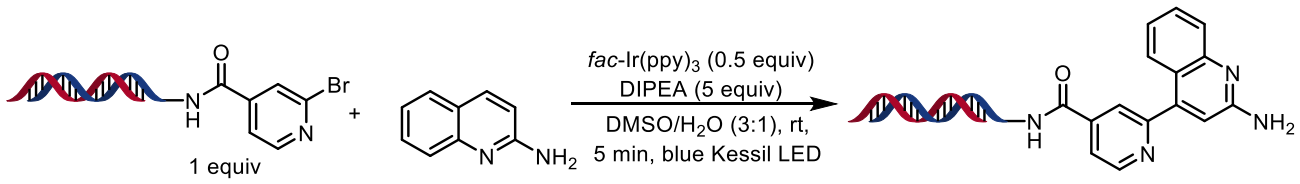
Product **6a**, 81% yield



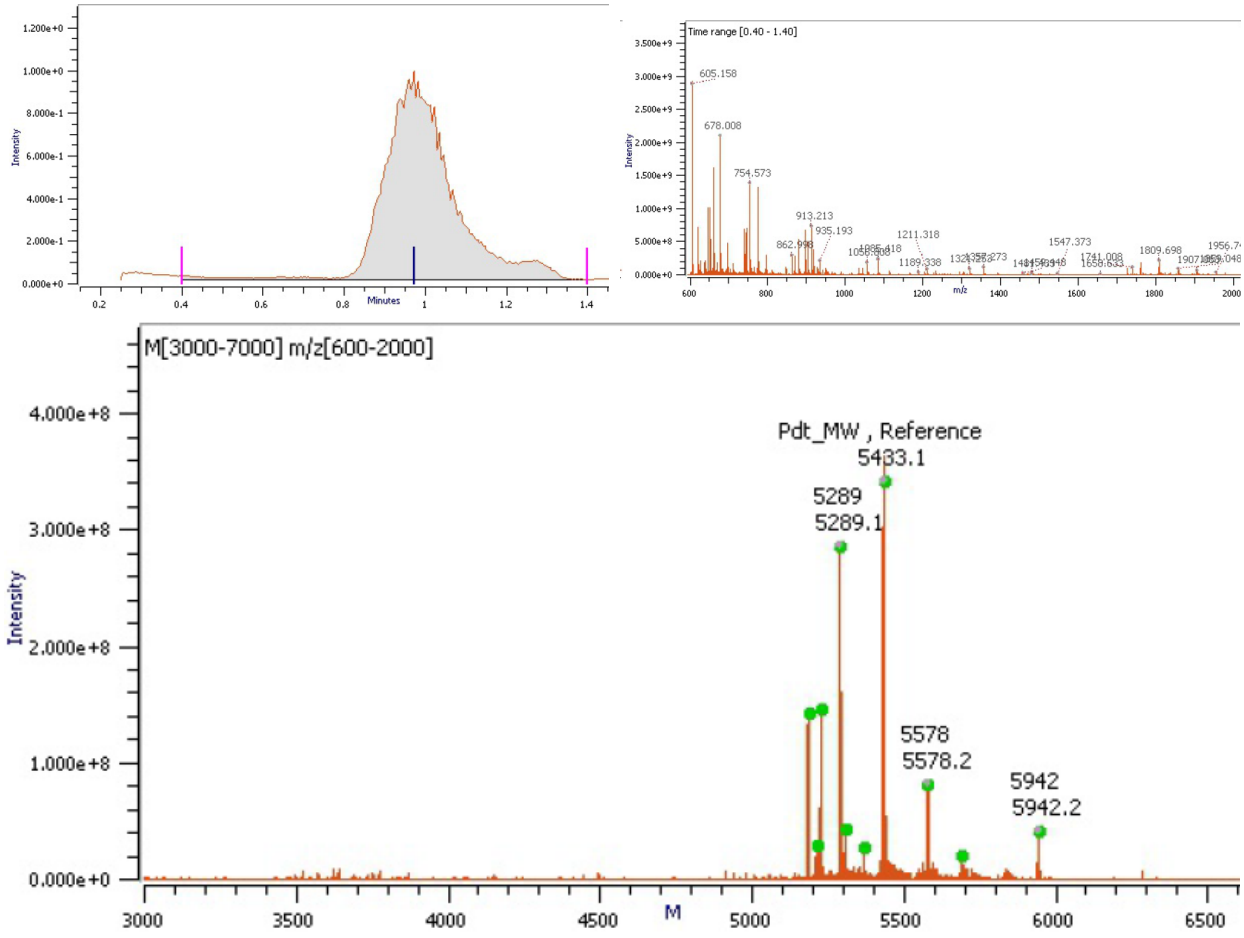
ELTM_419, TW-N78441-9-419.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



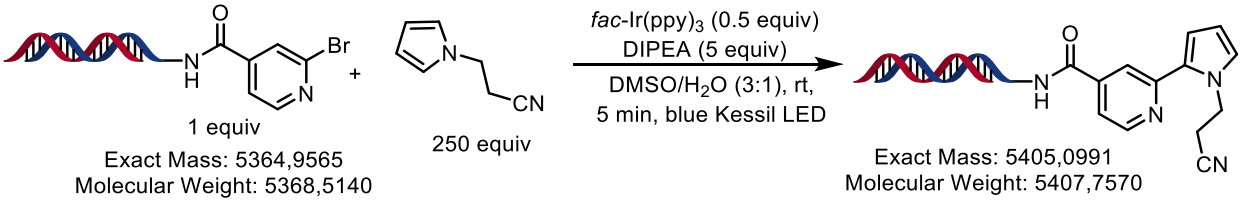
Product **6d**, 40% yield



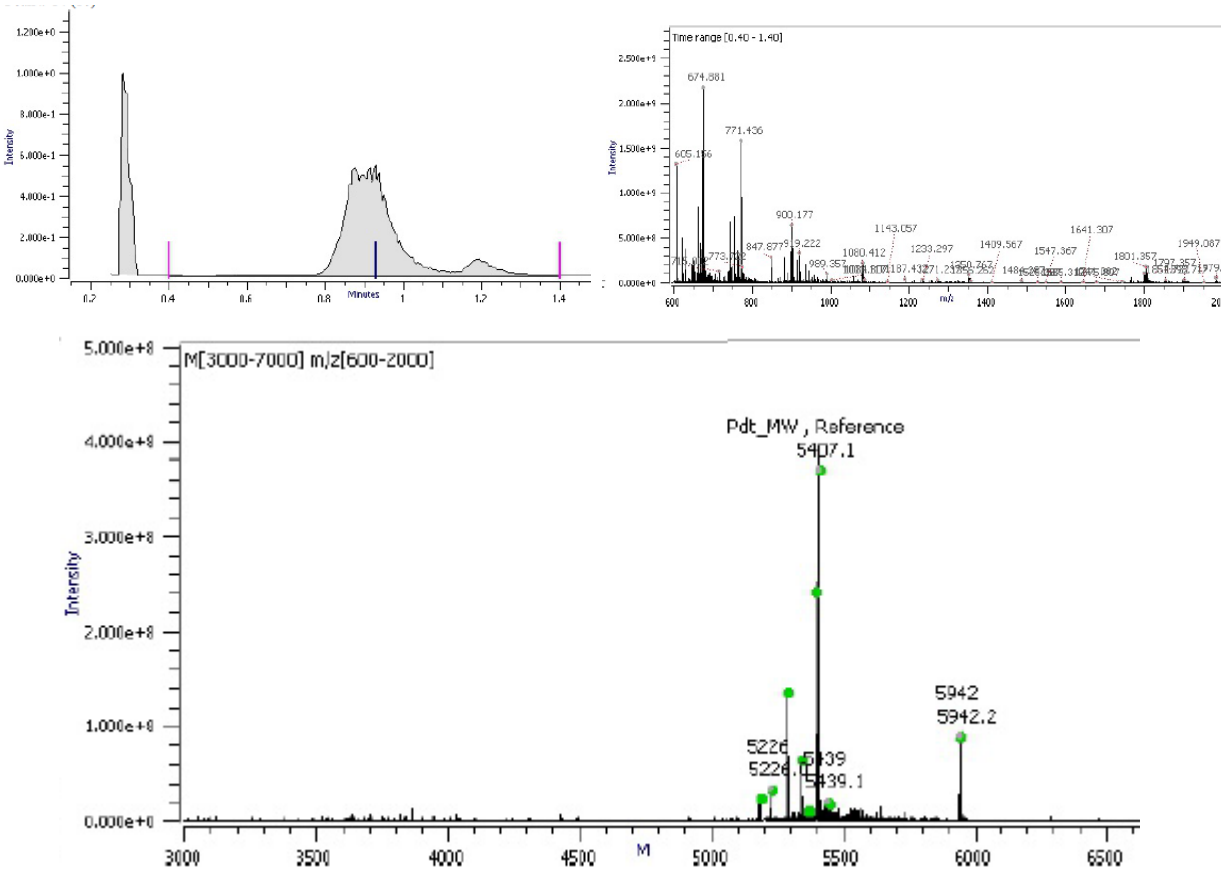
ELTM_469, TW-N78441-10-469.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



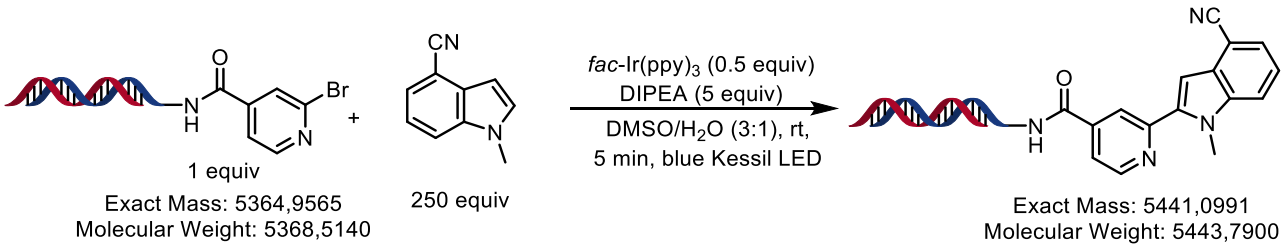
Product **6i**, 46% yield



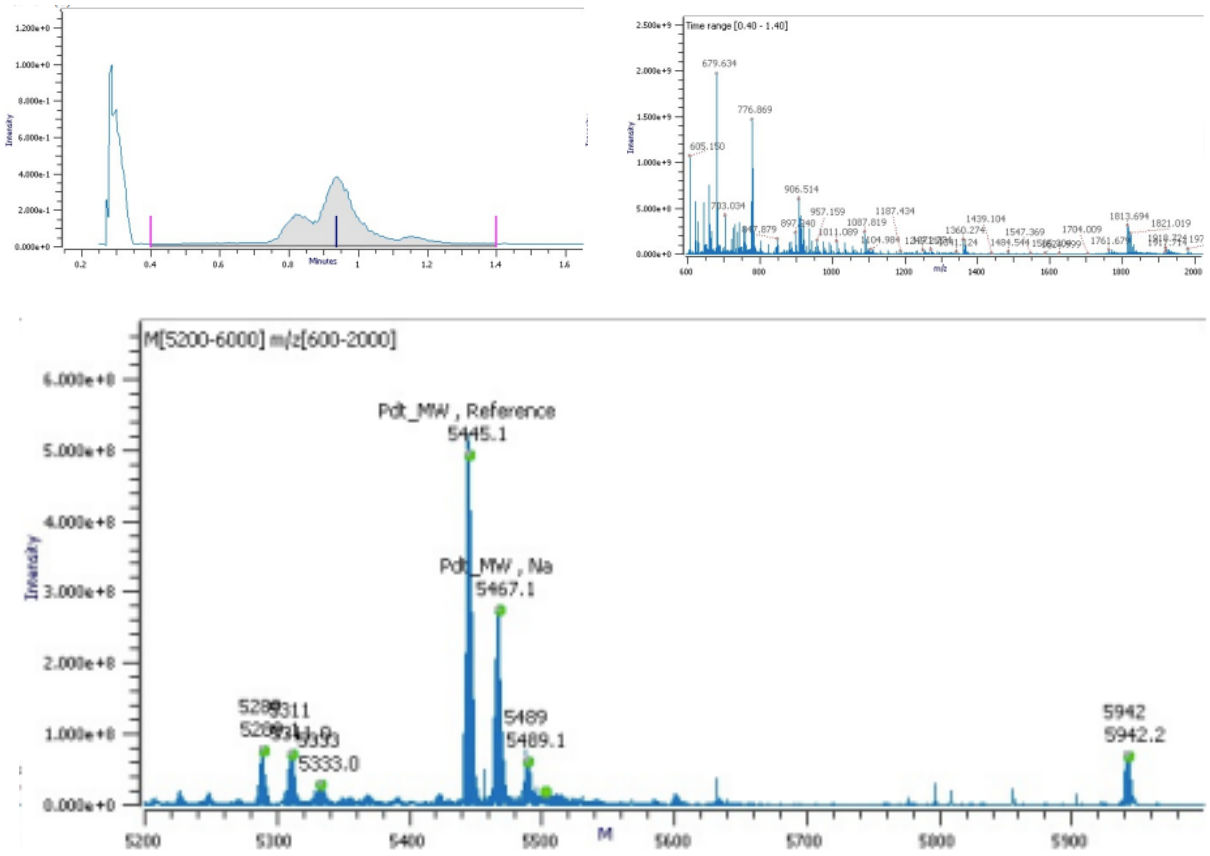
ELTM_649, TW-N78441-14-649.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



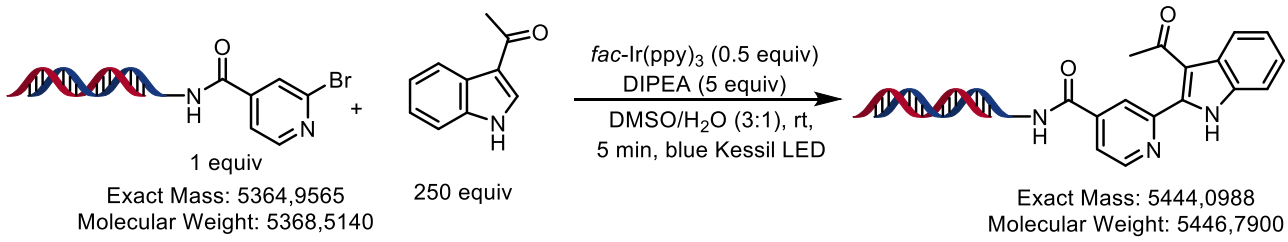
Product **6k**, 70% yield



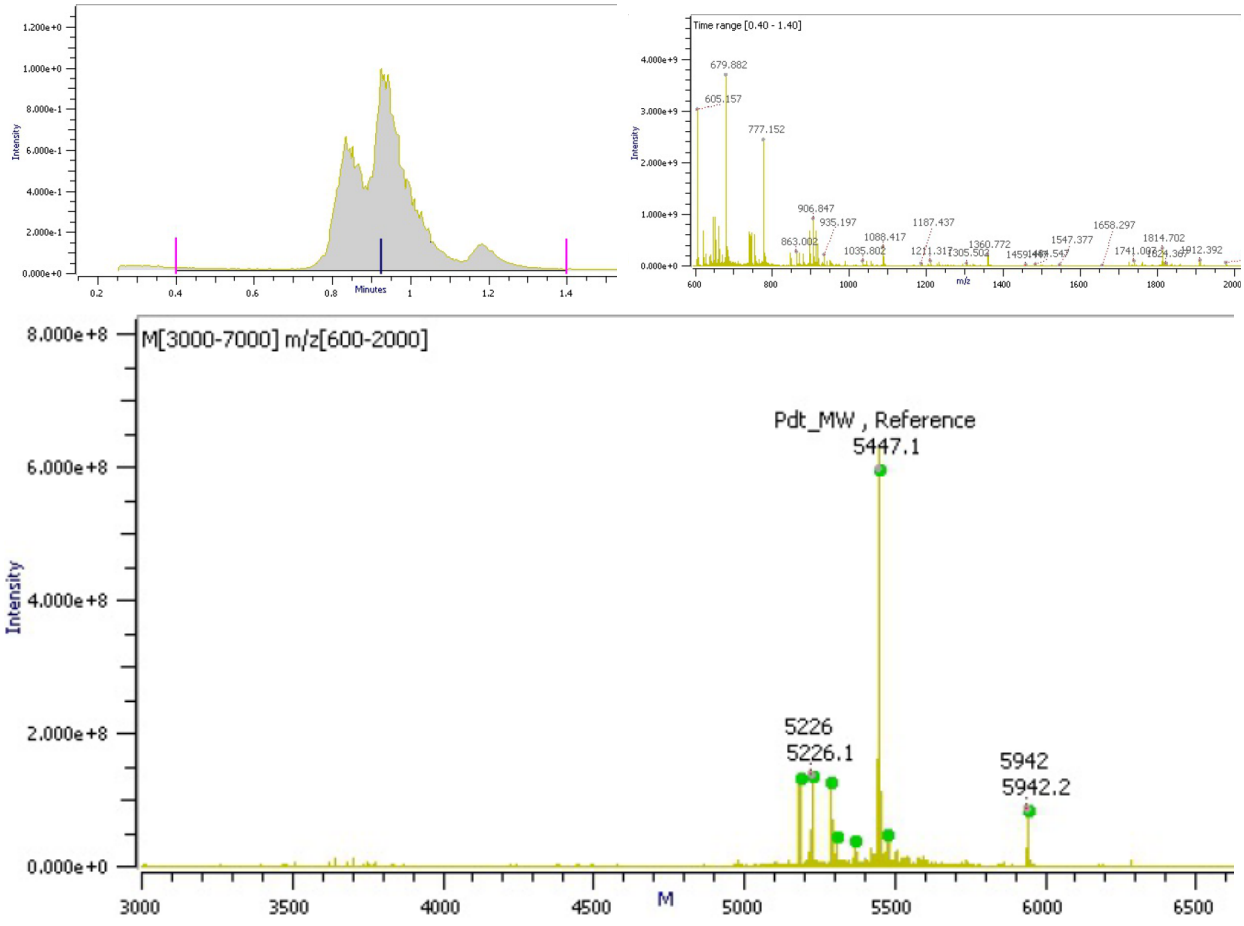
TW-N78441-15-681 - (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



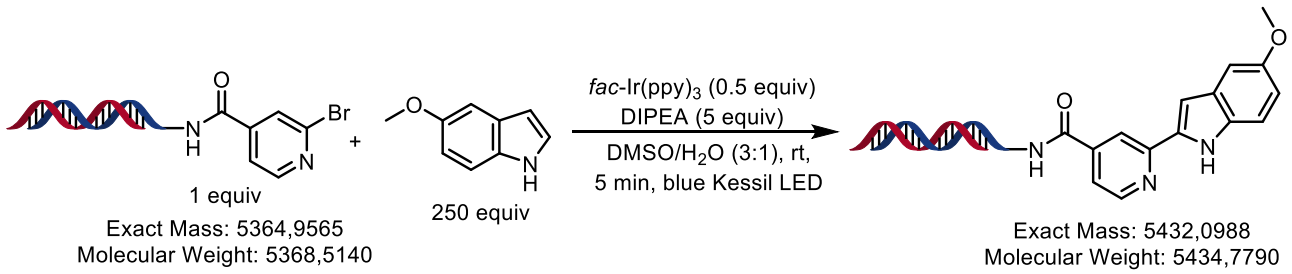
Product **6o**, 67% yield



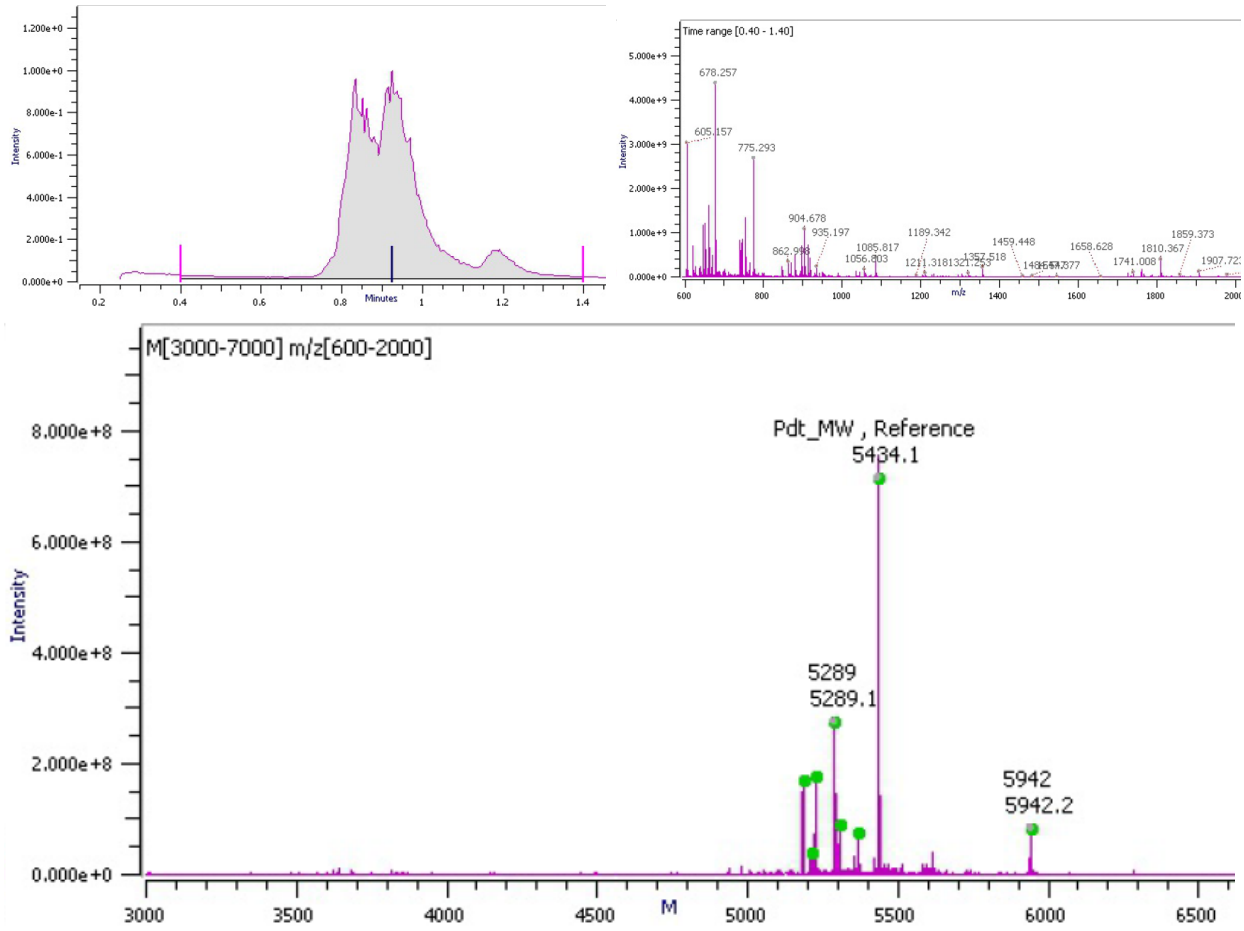
ELTM_476, TW-N78441-10-476.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



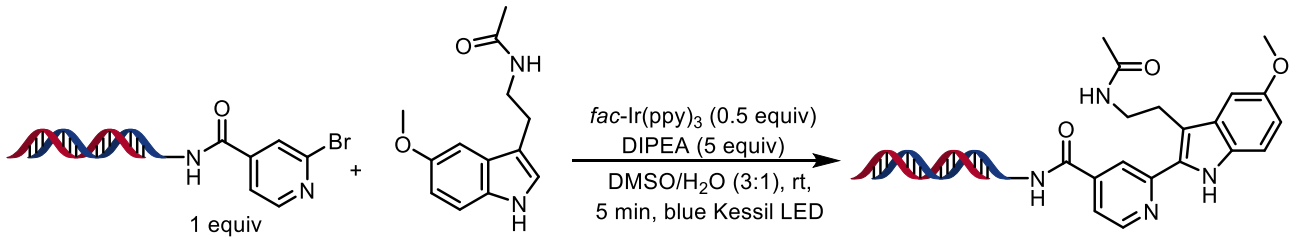
Product **6p**, 60% yield



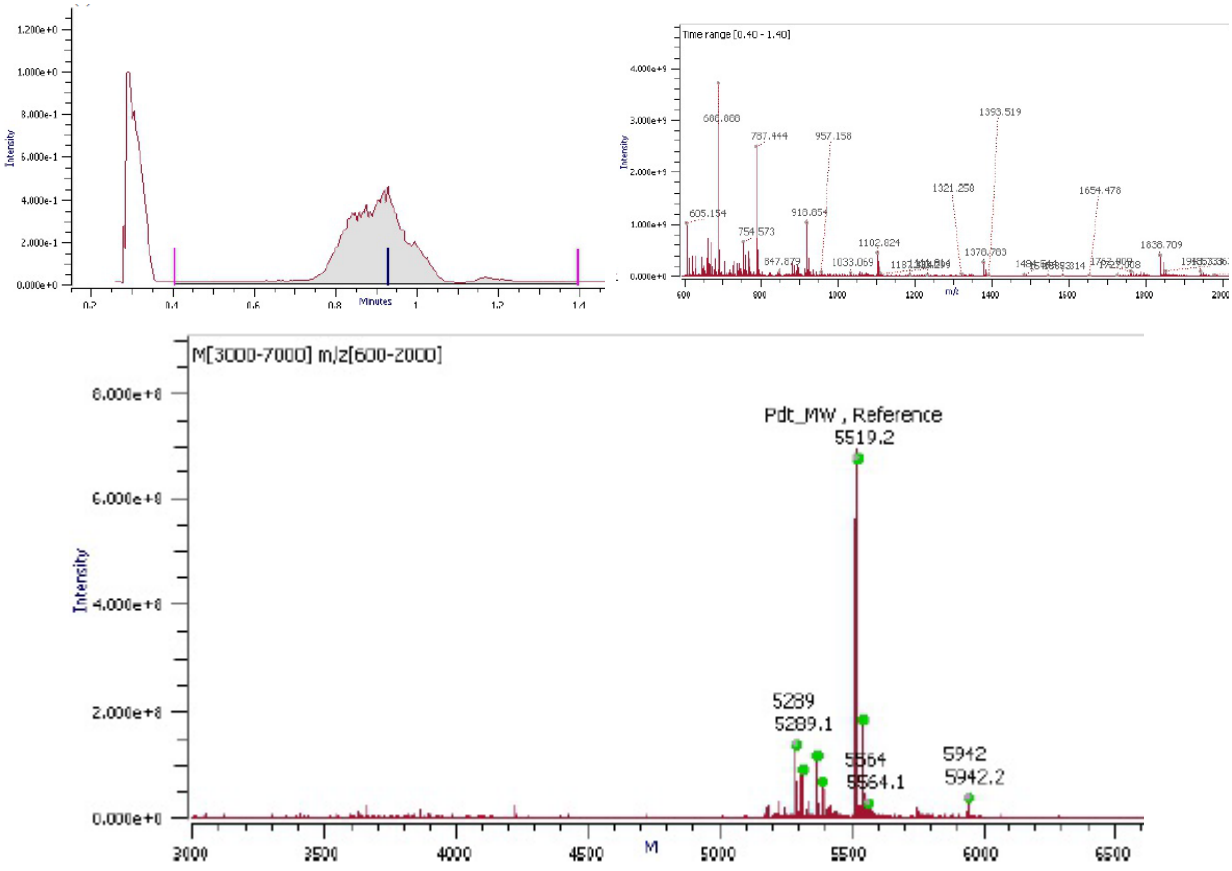
ELTM_477, TW-N78441-10-477.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



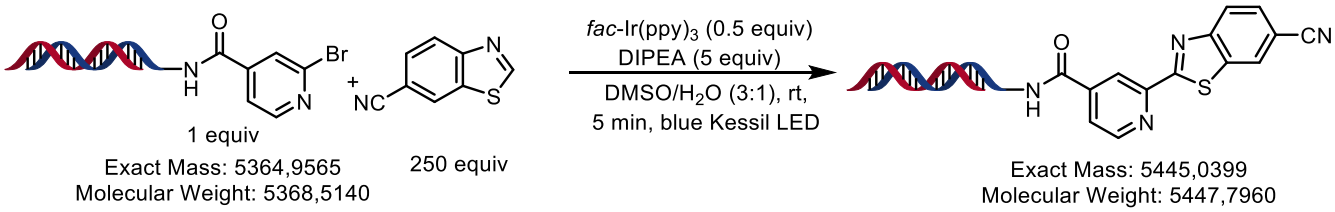
Product **6q**, 50% yield



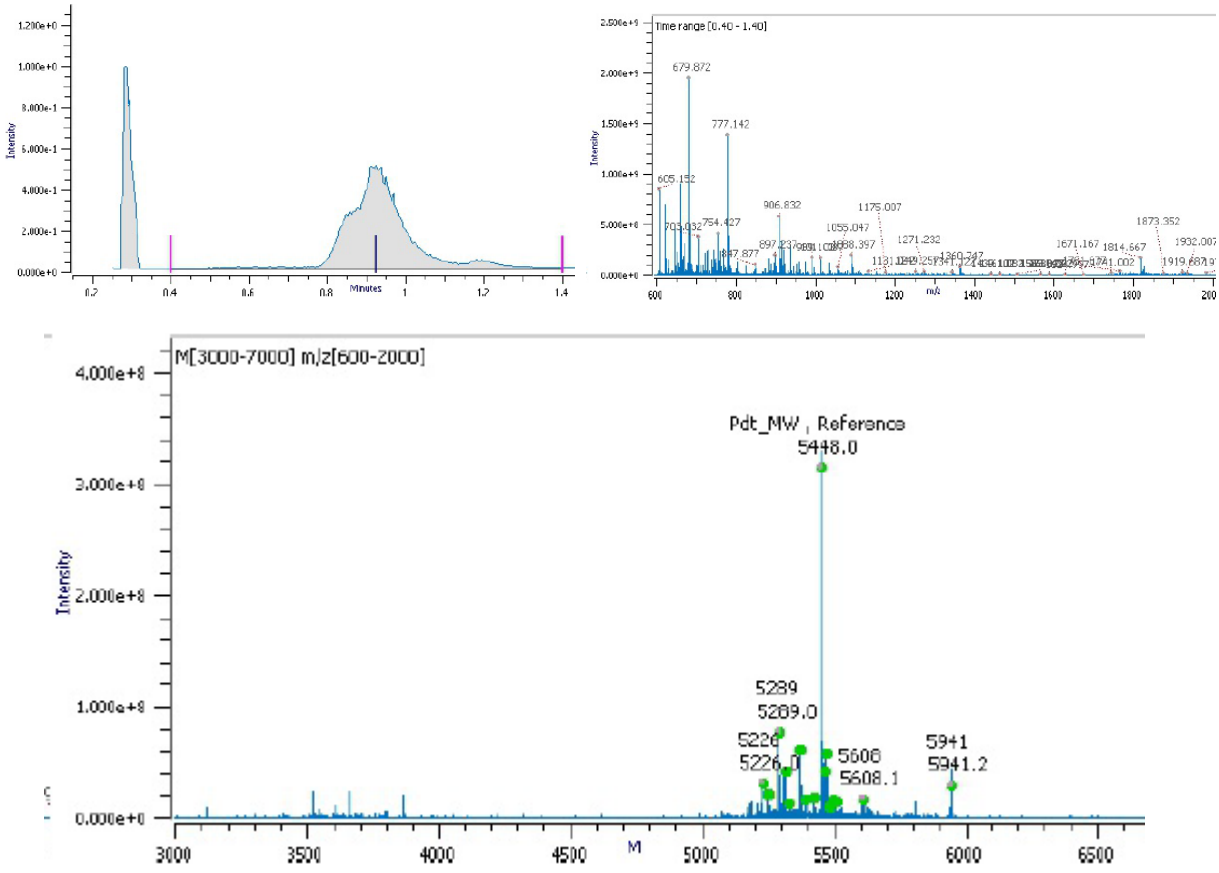
ELTM_672, TW-N78441-15-672.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



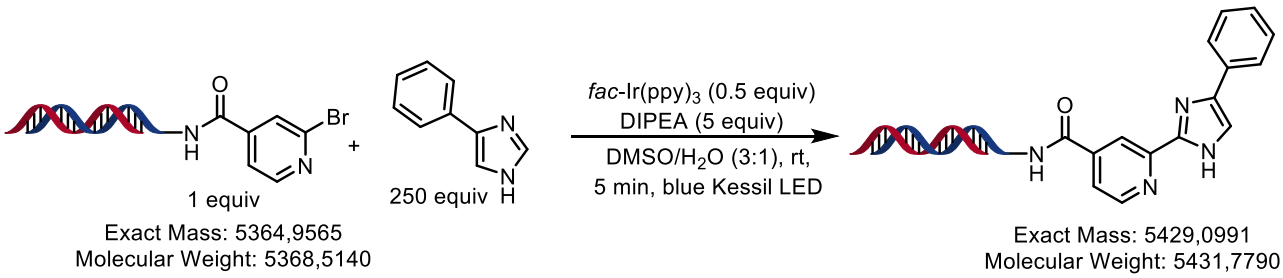
Product **6w**, 54% yield



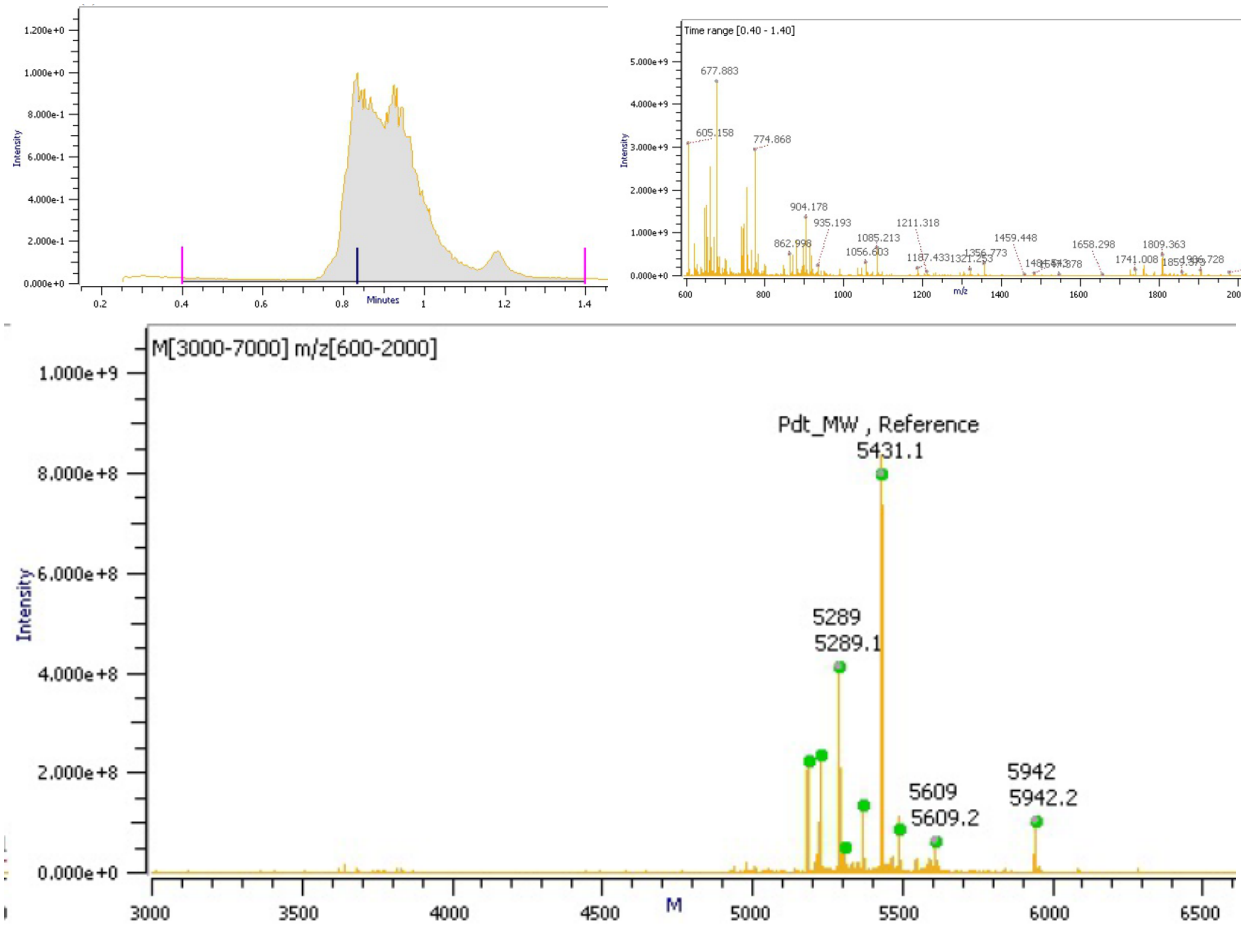
ELTM_637, TW-N78441-14-637.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



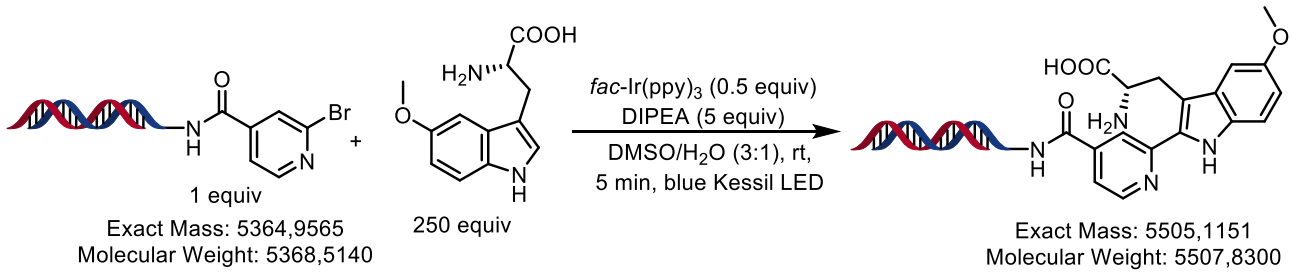
Product **6aa**, 51% yield



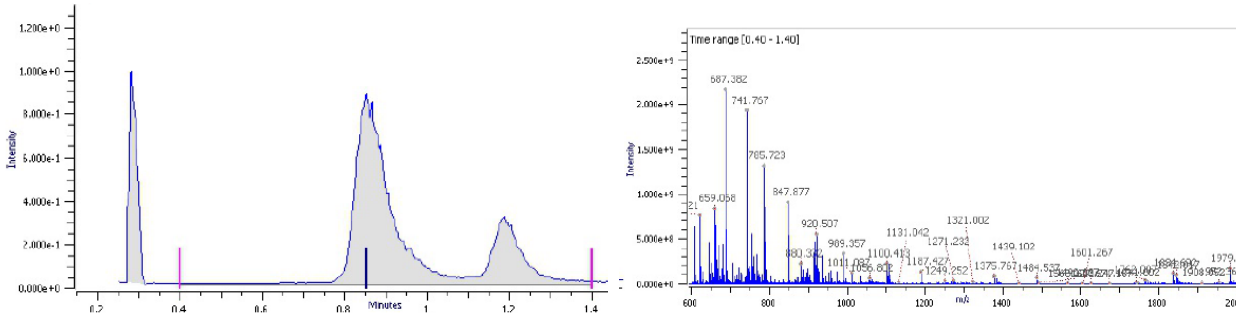
ELTM_483, TW-N78441-10-483.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)

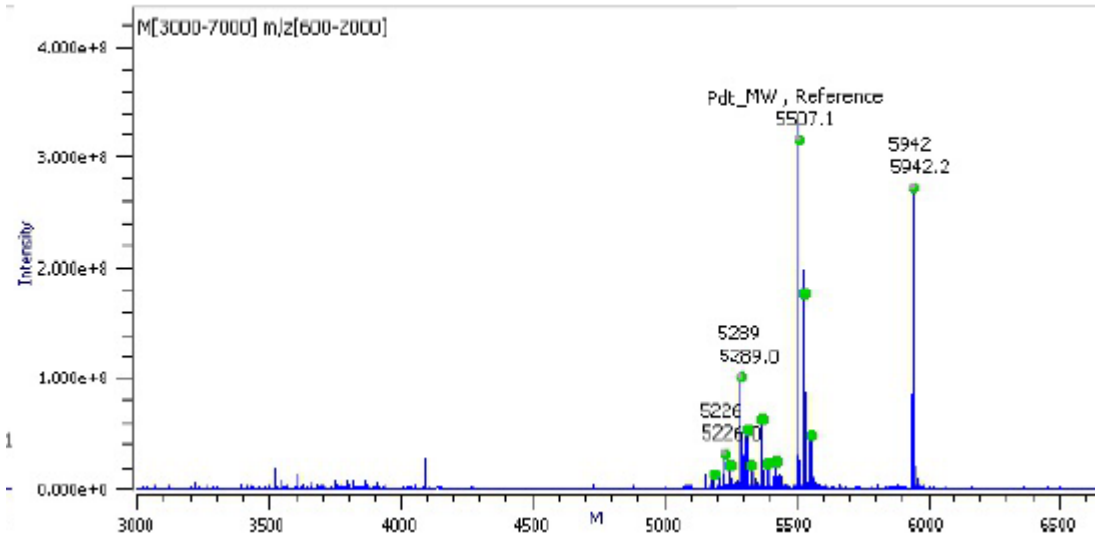


Product **6ag**, 37% yield

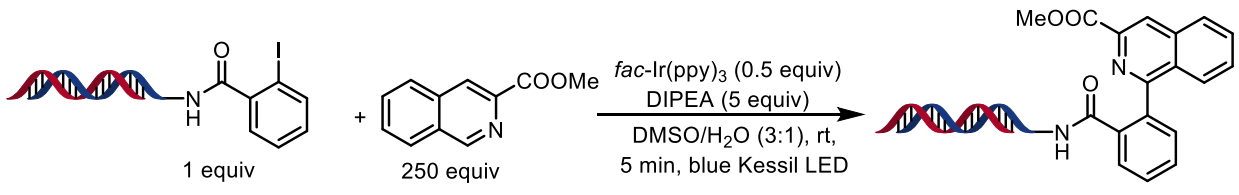


ELTM_643, TW-N78441-14-643.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)





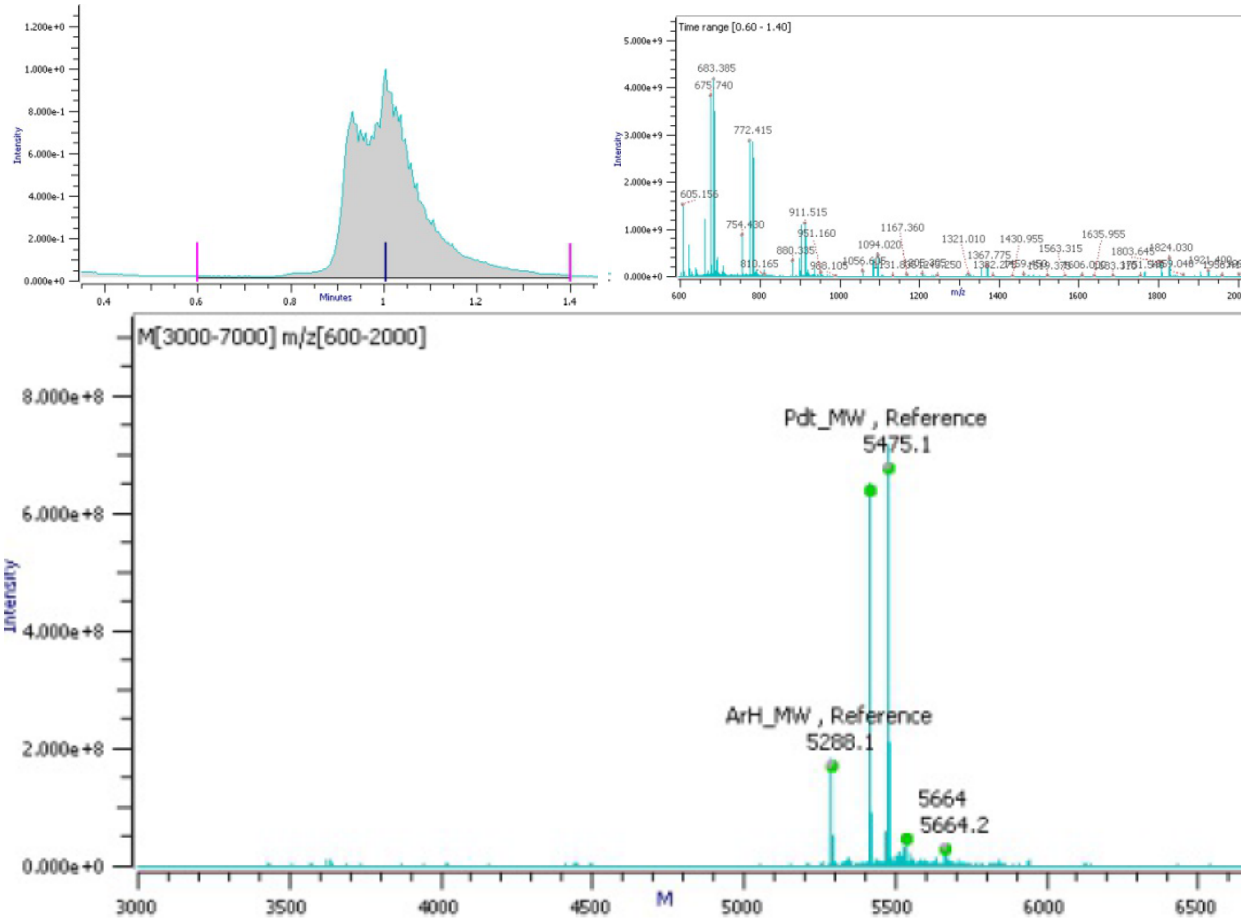
Product **7a**, 46% yield



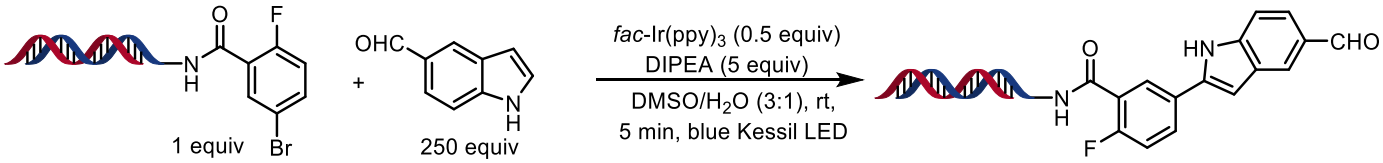
Exact Mass: 5411,9474
Molecular Weight: 5414,5264

Exact Mass: 5471,0984
Molecular Weight: 5473,8120

ELTM_425, TW-N78441-9-425.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



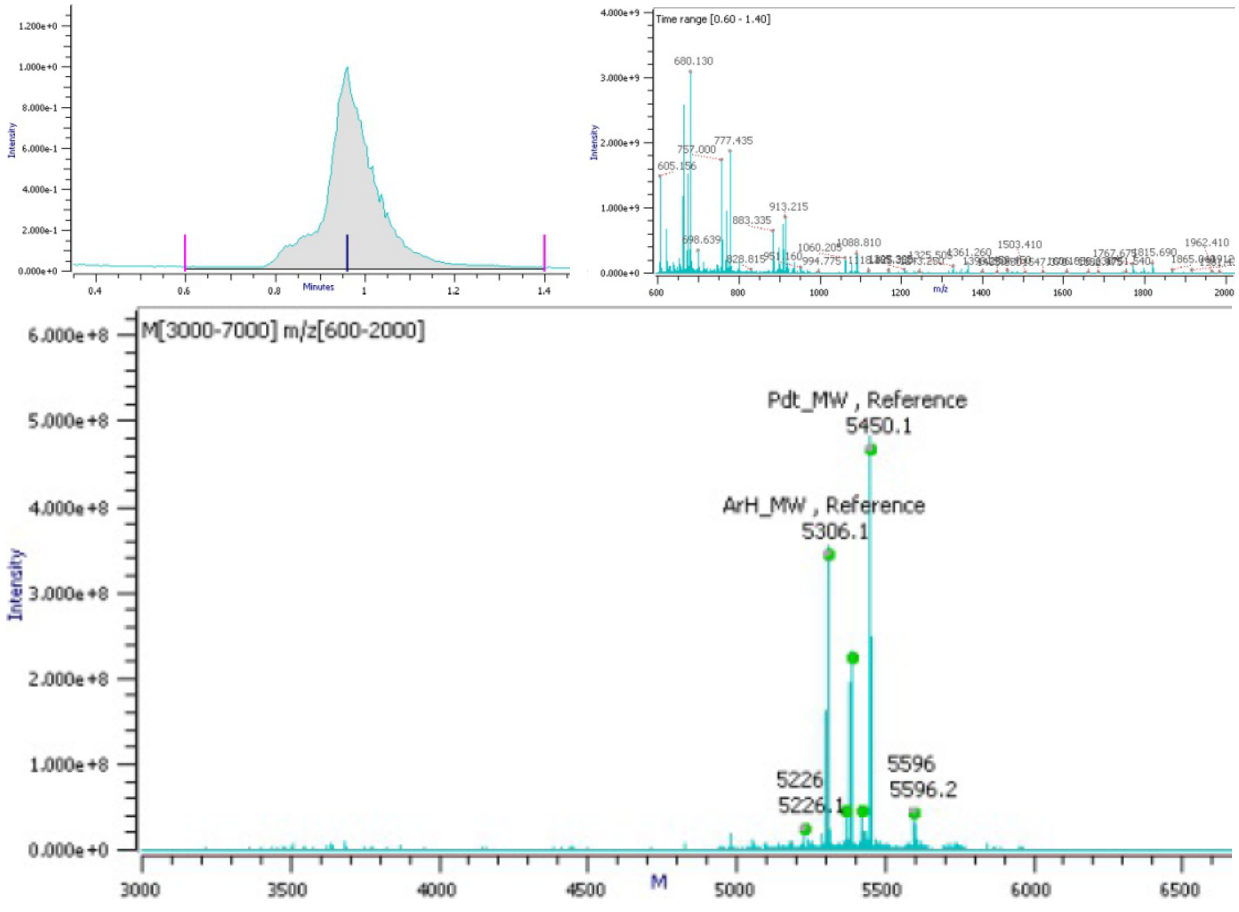
Product **8m**, 39% yield



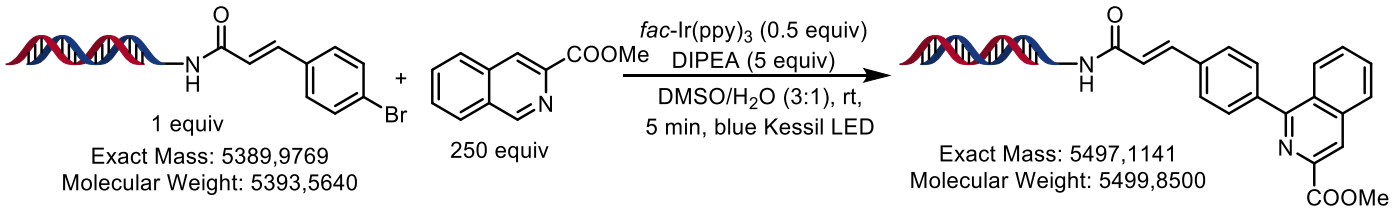
Exact Mass: 5505,0202
Molecular Weight: 5508,6714

Exact Mass: 5447,0784
Molecular Weight: 5449,7654

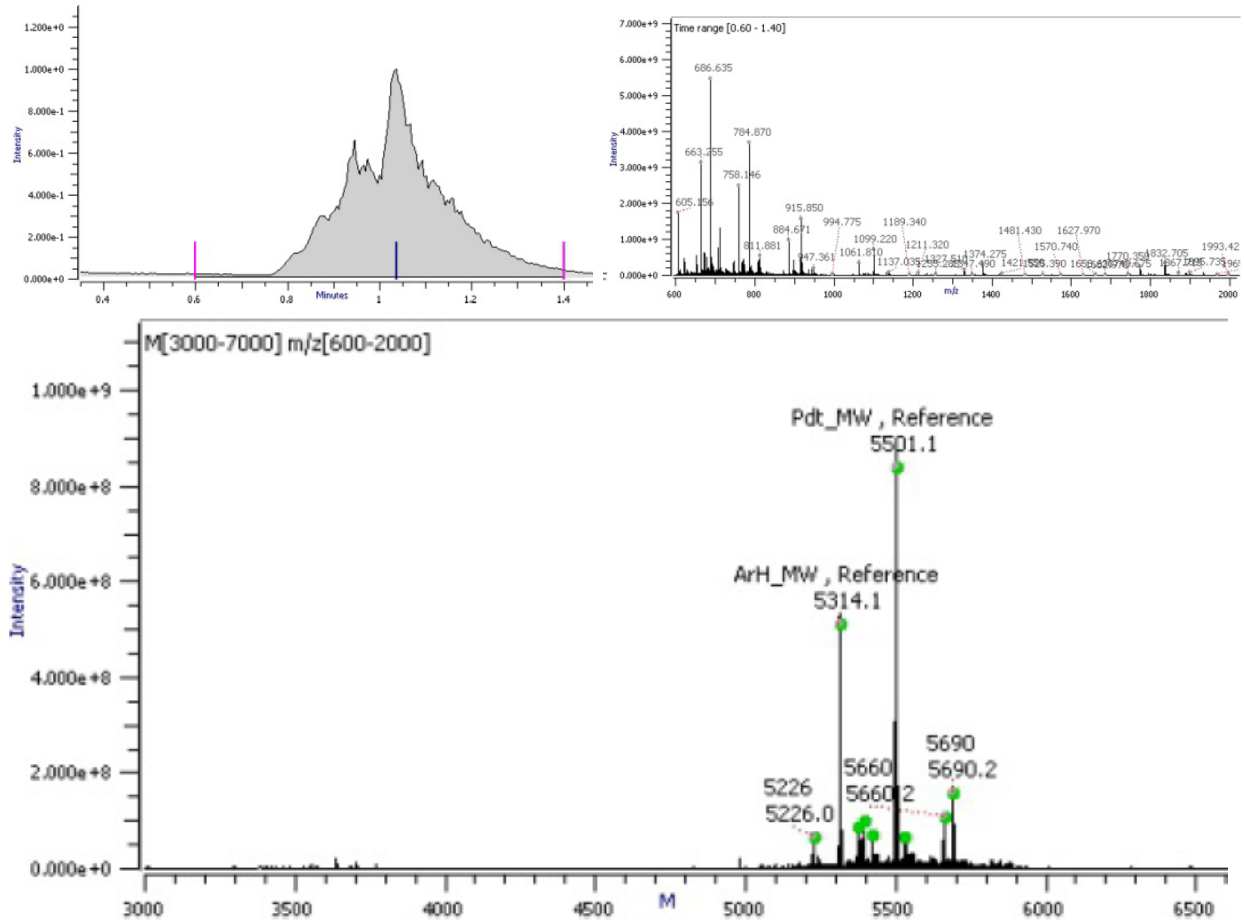
ELTM_327, TW-N78441-9-327.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



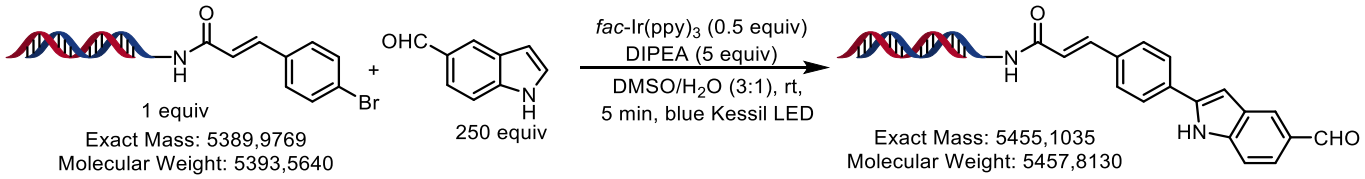
Product **12a**, 50% yield



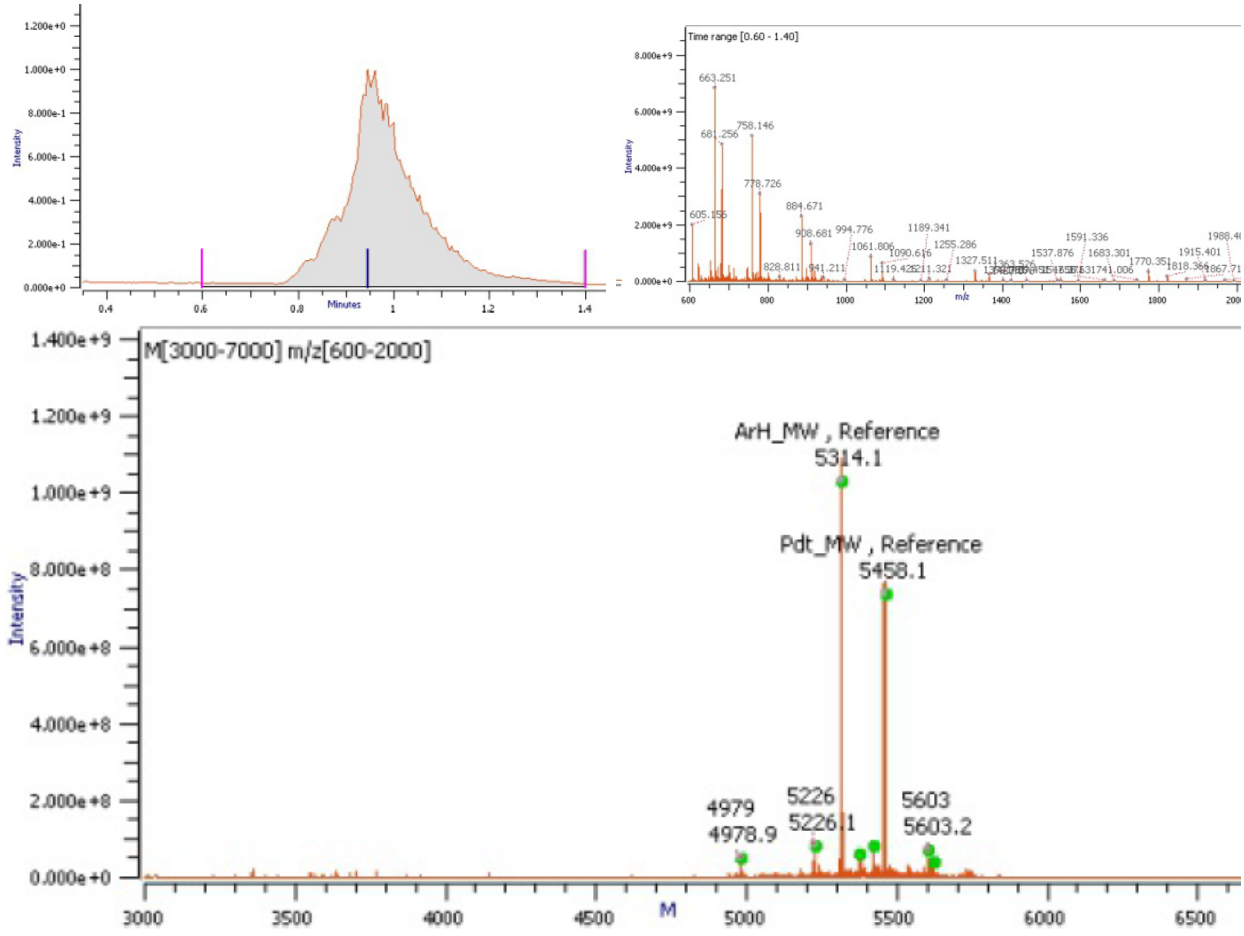
ELTM_328, TW-N78441-9-328.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



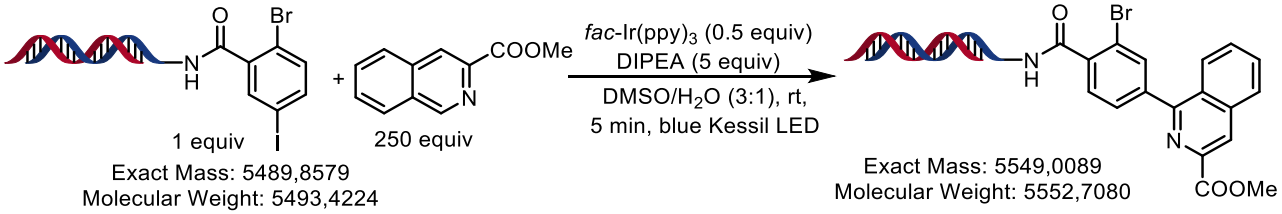
Product **12m**, 40% yield



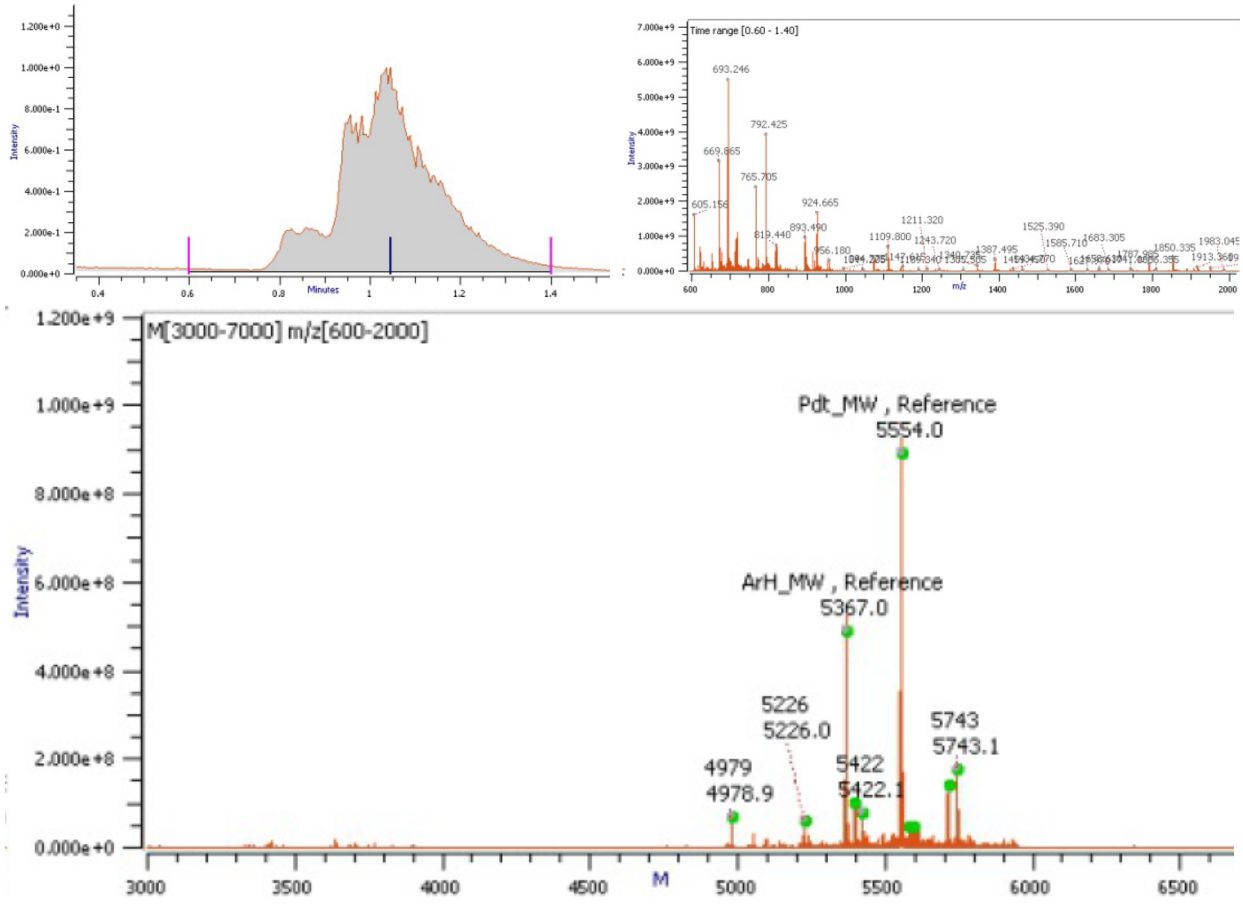
ELTM_330, TW-N78441-9-330.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



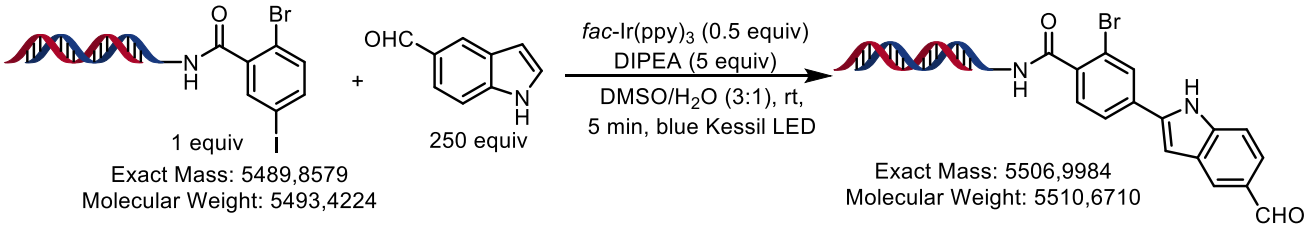
Product **13a**, 44% yield



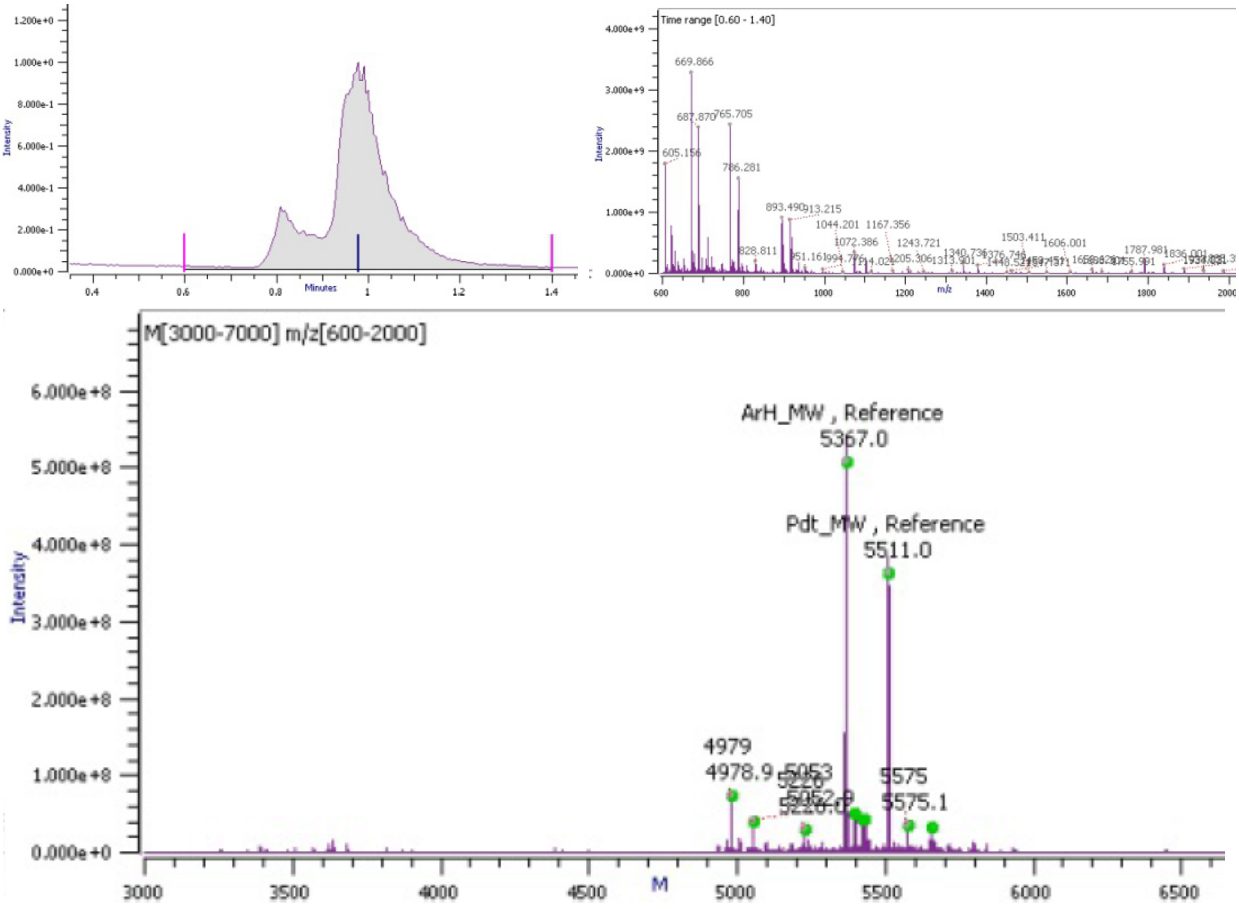
ELTM_240, TW-N78441-9-240.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



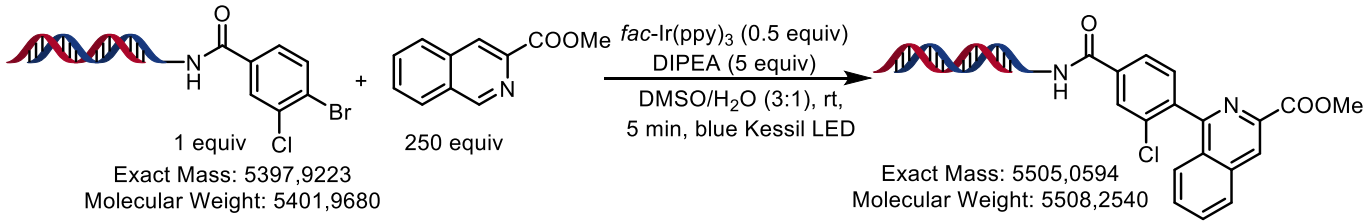
Product **13m**, 31% yield



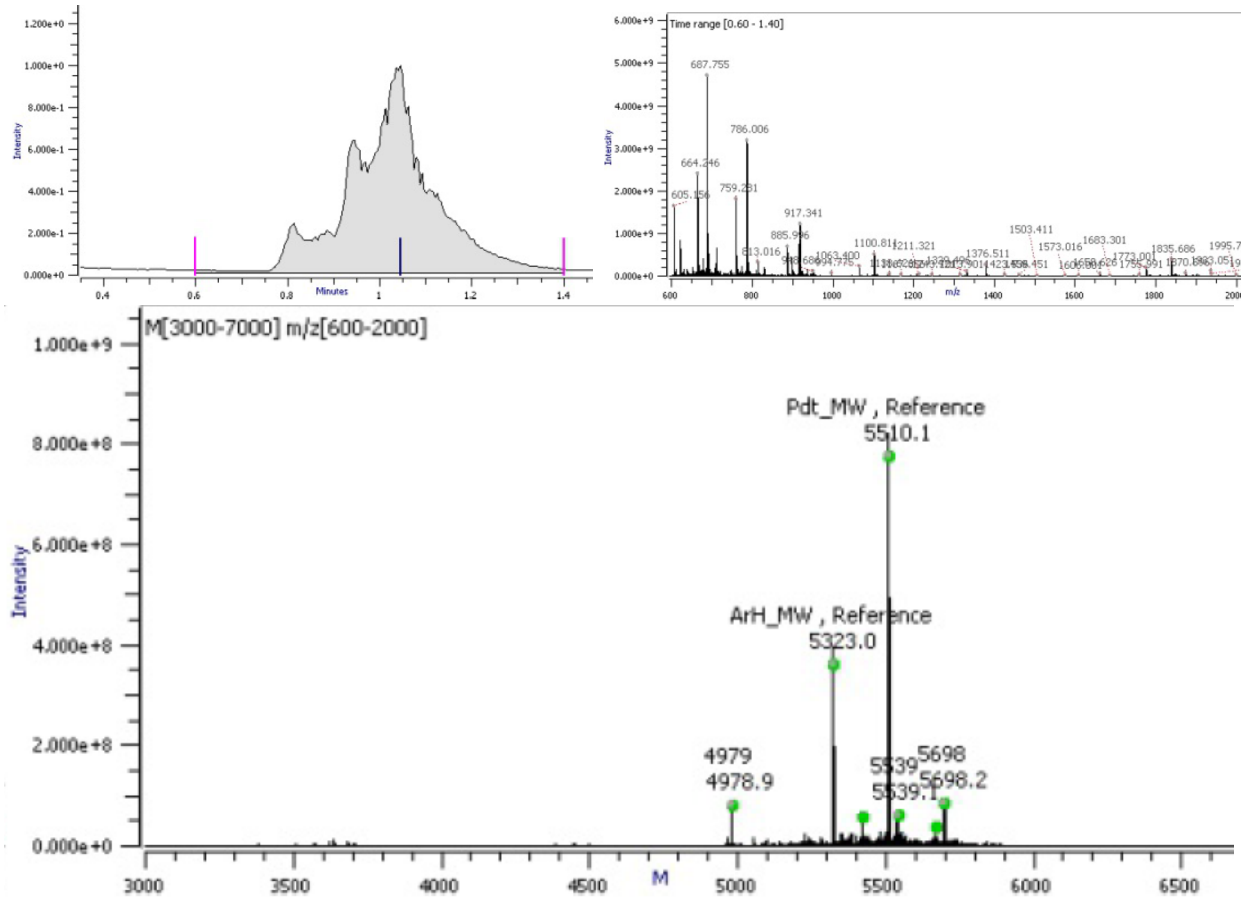
ELTM_242, TW-N78441-9-242.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



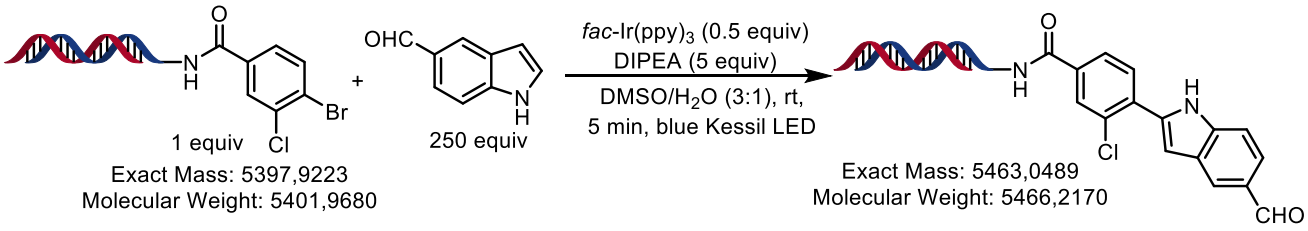
Product **14a**, 56% yield



ELTM_297, TW-N78441-9-297.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



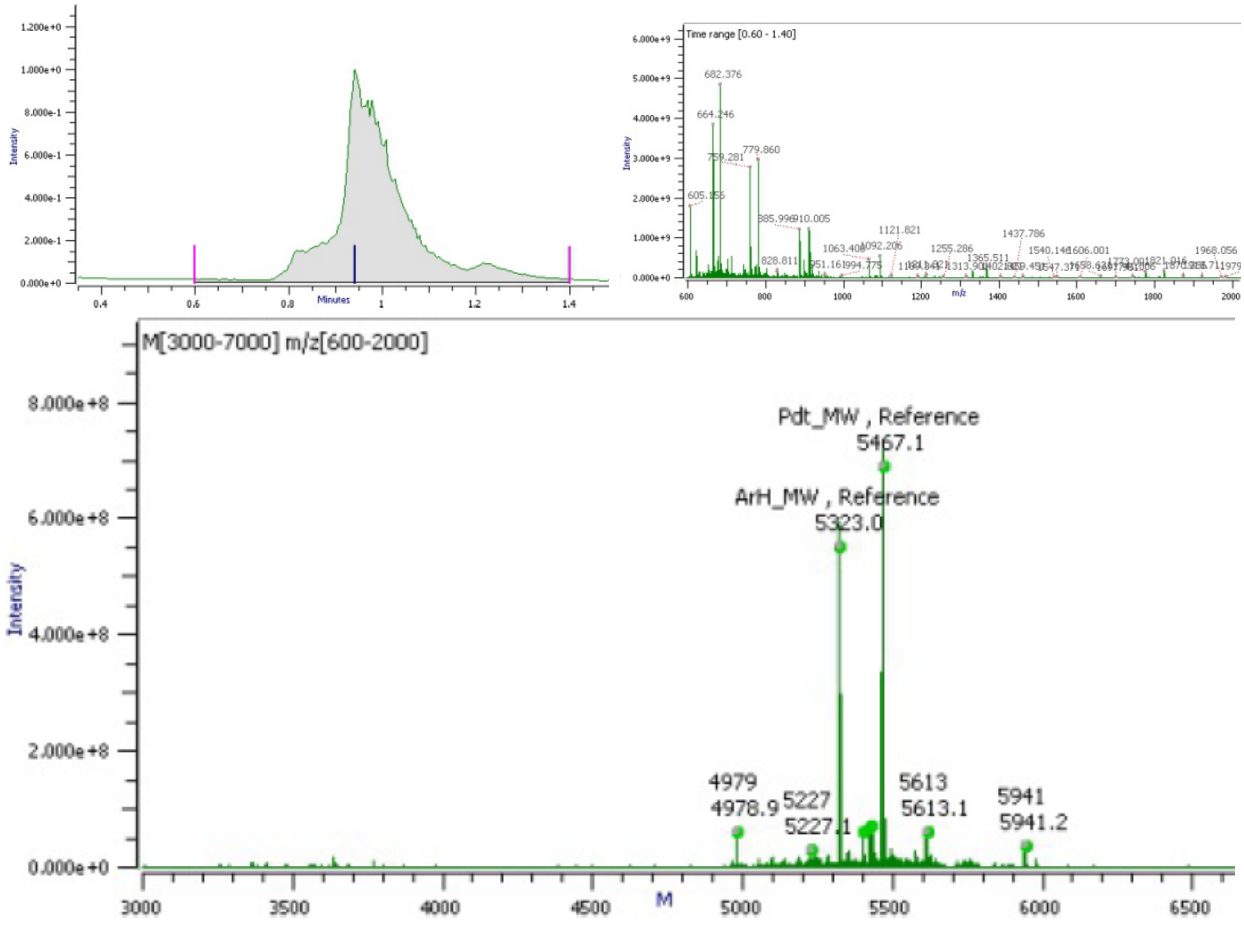
Product **14m**, 44% yield



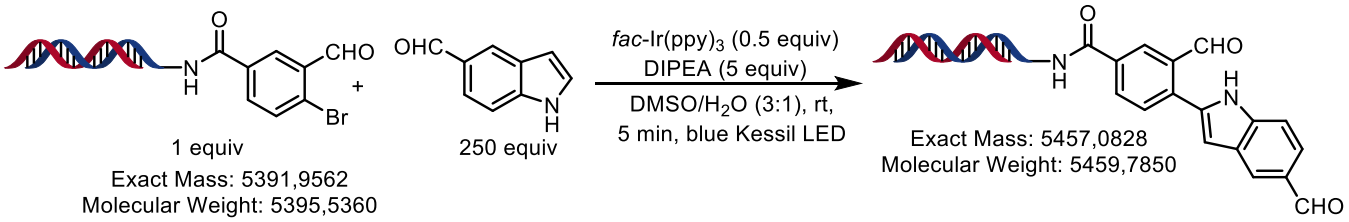
Exact Mass: 5397,9223
Molecular Weight: 5401,9680

Exact Mass: 5463,0489
Molecular Weight: 5466,2170

ELTM_339, TW-N78441-9-339.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)



Product **15m**, 50% yield



ELTM_239, TW-N78441-9-239.raw (Plot Elution Peak - Plot MS1 - Plot Deconvoluted Mass spectrum)