Activity-based substrate profiling of N-terminal methyltransferase 1 (NTMT1)

by

Kaimin Jia

B.Eng., China University of Petroleum (East China), 2013

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Chemistry
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2019

Abstract

N-terminal methyltransferase 1 (NTMT1) is a new protein methyltransferase that catalyzes the N-terminal methylation of proteins with a specific N-terminal sequence motif after methionine removal. Protein N-terminal methylation plays multiple cellular roles in DNA-protein interaction, chromatin segregation, and DNA damage repair. Dysregulation of NTMT1 has been implicated in several cancers and developmental disease. Thus, it is vital to identify NTMT1 targets to reveal its involvement in various disease pathology. Here, an activity-based substrate profiling strategy was developed to uncover the potential substrates of NTMT1. For proof of concept, Hey-SAM, a SAM surrogate containing a terminal alkynyl group, was able to alkylate the previously confirmed NTMT1 substrate, regulator of chromosome condensation (RCC1), with a high efficiency in the presence of wild type NTMT1. The alkylated RCC1 was subsequently labelled with a biotin group *via* Cu(I) catalyzed click chemistry, which allowed its purification by streptavidin beads. This strategy was applied to profile NTMT1 substrates from human cell lysates. Human embryonic kidney 293FT (HEK293FT) cell line was edited to generate an NTMT1 knockout (KO) variant using clustered regularly interspaced short palindromic repeats (CRISPR)-cas9 system, which could produce hypomethylated cell lysates specifically for NTMT1. Potential NTMT1 substrates were treated with Hey-SAM, labeled with a biotin group, enriched by streptavidin beads, separated on SDS-PAGE, digested by trypsin, and identified by LC-MS. Statistical and motif analysis revealed 70 putative NTMT1 targets. Among them, three proteins, RCC1, protein SET, and 60S ribosomal protein L23a (RL23A) were reported to be NTMT1 substrates previously. Validation of selected targets confirmed that polybromo-1 (PB1, peptide analog), SH3 and PX domain-containing protein 2B (SPD2B, peptide analog), 40S ribosomal protein S14 (RS14), and Obg-like ATPase 1 (OLA1) were NTMT1 substrates *in vitro*.

Methylation of OLA1 was further investigated *in vivo* using normal and NTMT1 KO HEK293FT cells. OLA1 was reported to play multiple roles in several cellular processes including centrosome amplification, antioxidant response, heat shock response and cell growth. Overexpressed OLA1 in normal cells is dimethylated and this modification is abolished in KO cells. These results confirm that NTMT1 is responsible for catalyzing the N-terminal methylation of OLA1 *in vivo*. Moreover, our preliminary study revealed that methylation of OLA1 does not alter its subcellular location and catalytic activity as an ATPase *in vitro*. Further study is in need to elucidate the cellular function of this important modification on OLA1.

Activity-based substrate profiling of N-terminal methyltransferase 1 (NTMT1)

by

Kaimin Jia

B.Eng., China University of Petroleum (East China), 2013

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Chemistry
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2019

Approved by:

Major Professor
Dr. Ping Li

Copyright

Abstract

N-terminal methyltransferase 1 (NTMT1) is a new protein methyltransferase that catalyzes the N-terminal methylation of proteins with a specific N-terminal sequence motif after methionine removal. Protein N-terminal methylation plays multiple cellular roles in DNA-protein interaction, chromatin segregation, and DNA damage repair. Dysregulation of NTMT1 has been implicated in several cancers and developmental disease. Thus, it is vital to identify NTMT1 targets to reveal its involvement in various disease pathology. Here, an activity-based substrate profiling strategy was developed to uncover the potential substrates of NTMT1. For proof of concept, Hey-SAM, a SAM surrogate containing a terminal alkynyl group, was able to alkylate the previously confirmed NTMT1 substrate, regulator of chromosome condensation (RCC1), with a high efficiency in the presence of wild type NTMT1. The alkylated RCC1 was subsequently labelled with a biotin group *via* Cu(I) catalyzed click chemistry, which allowed its purification by streptavidin beads. This strategy was applied to profile NTMT1 substrates from human cell lysates. Human embryonic kidney 293FT (HEK293FT) cell line was edited to generate an NTMT1 knockout (KO) variant using clustered regularly interspaced short palindromic repeats (CRISPR)-cas9 system, which could produce hypomethylated cell lysates specifically for NTMT1. Potential NTMT1 substrates were treated with Hey-SAM, labeled with a biotin group, enriched by streptavidin beads, separated on SDS-PAGE, digested by trypsin, and identified by LC-MS. Statistical and motif analysis revealed 70 putative NTMT1 targets. Among them, three proteins, RCC1, protein SET, and 60S ribosomal protein L23a (RL23A) were reported to be NTMT1 substrates previously. Validation of selected targets confirmed that polybromo-1 (PB1, peptide analog), SH3 and PX domain-containing protein 2B (SPD2B, peptide analog), 40S ribosomal protein S14 (RS14), and Obg-like ATPase 1 (OLA1) were NTMT1 substrates *in vitro*.

Methylation of OLA1 was further investigated *in vivo* using normal and NTMT1 KO HEK293FT cells. OLA1 was reported to play multiple roles in several cellular processes including centrosome amplification, antioxidant response, heat shock response and cell growth. Overexpressed OLA1 in normal cells is dimethylated and this modification is abolished in KO cells. These results confirm that NTMT1 is responsible for catalyzing the N-terminal methylation of OLA1 *in vivo*. Moreover, our preliminary study revealed that methylation of OLA1 does not alter its subcellular location and catalytic activity as an ATPase *in vitro*. Further study is in need to elucidate the cellular function of this important modification on OLA1.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ABC	ammonium bicarbonate
AD	activation domain
Adox	adenosine-2',3'-dialdehyde
AKG/2OGF	α-ketoglutaric acid
BARD1	BRCA1-associated RING domain protein
BCA	bicinchoninic acid assay
BD	binding domain
BLAST	Basic Local Alignment Search Tool
BRCA1	breast cancer type 1 susceptibility protein
CCAN	constitutive centromere associated protein network
cDNA	complementary DNA
CENP-A	histone H3-like centromeric protein A
CENP-B	centromere protein B
CHIP	carboxy terminus of HSP70-binding protein
CHOP	CCAAT/enhancer-binding protein-homologous protein
co-IP	co-immunoprecipitation
CuAAC	Copper-Catalyzed Azide-Alkyne Cycloaddition
DAPI	4',6-diamidino-2-phenylindole
DBA	diazo biotin azide
DDB2	DNA damage binding protein 2
DDX60L	probable ATP-dependent RNA helicase DDX60-like
DMEM	Dulbecco's Modification of Eagle's Medium
DOC45	DNA damage-regulated overexpressed in cancer 45
E. coli	Escherichia coli
eEF1A	elongation factor 1-alpha 1
ELISA	enzyme-linked immunosorbent assay
FAD	flavin adenine dinucleotide
FBS	fetal bovine serum
GSH	glutathione

HEK	human embryonic kidney
Hey-SAM	(E)-hex-2-en-5-ynyl-SAM
H2A1B	histone H2A type 1-B/E
HRMS	high-resolution mass spectrometry
HSP70	heat shock proteins 70
ICAM1	intercellular adhesion molecule 1
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
KO	knockout
LC-MS	liquid chromatography-mass spectrometry
MCF7	michigan Cancer Foundation-7
MDA-MB-231	M.D. Anderson Metastasis Breast cancer 231
MDCK	Madin-Darby Canine Kidney
MEFs	mouse embryonic fibroblast cells
METTL13	human methyltransferase-like protein 13
METTL13-C	C-terminal domain of METTL13
METTL13-N	N-terminal domain of METTL13
mRNA	messenger RNA
MS	mass spectrometry
MYL9	myosin regulatory light chain 9
NTMT1/NRMT1	protein N-terminal RCC1 methyltransferase 1
NTMT2/NRMT2	protein N-terminal RCC1 methyltransferase 2
OAT	ornithine aminotransferase
OGT1	O-linked β-N-acetylglucosamine transferase
OLA1	Obg-like ATPase 1
ORFs	open reading frames
P/S	penicillin-streptomycin
p21	protein cyclin-dependent kinase inhibitor 1
PAM	protospacer adjacent motif
PARP3	poly (ADP-ribose) polymerase 3
PB1	protein polybromo-1

PCA	principle component analysis
PCR	polymerase chain reaction
PFA	paraformaldehyde
PHB	prohibitin
qRT-PCR	quantitative real-time polymerase chain reaction
RCC1	regulator of chromosome condensation
ROS	reactive oxygen species
RS14	40S ribosomal protein S14
SA	streptavidin
SAH	S-adenosyl homocysteine
SAHH	SAH hydrolase
SAM	S-adenosyl methionine
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
sgRNA	single-guide RNA
SILAC	stable isotope labeling with amino acids in cell culture
SpCas9	Streptococcus pyogenes Cas9
SPD2B	SH3 and PX domain-containing protein 2B
TBTA	tris(benzyltriazolylmethyl)amine
TCEP	tris(2-carboxyethyl)phosphine hydrochloride
TUNEL	terminal deoxynucleotidyl transferase dUTP nick end labeling
UPLC	ultra-performance liquid chromatography
WB	western blot
wt	wild type

Acknowledgements

Upon the completion of this thesis, I would like to take this opportunity to express my sincere gratitude to my major advisor Dr. Ping Li for his support and encouragement throughout my Ph.D. study. I feel very fortunate to have worked in his lab and I would like to thank him for his guidance from early on designing a meaningful experiment, thinking critically and writing a logical paper. With his help, I got a systematic training on the field of molecular biology, protein purification and characterization, enzymology, proteomics, and mammalian cell culture. Especially, I want to thank for his effort in purchasing the high-resolution mass spectrometer which gave me a chance to finish my project.

Next, I would like to thank my committee members: Dr. Duy Hua, Dr. Brian Geisbrecht, Dr. Yulan Xiong and chair Dr. Bradley Olson for their advice on my research and their time on reading my dissertation. I took two classes including organic spectroscopy and synthetic organic chemistry from Dr. Hua and I was impressed with his profound knowledge and professional dedication. Dr. Geisbrecht gave me a lot of advice on isothermal titration calorimetry and protein crystallography. Specially, I would like to express my heartfelt gratitude to Dr. Xiong for her advice on human cell culture and generosity of letting me use their laminar flow hood, fluorescence microscope, and Amersham Imager 600.

I would also like to thank the other organic chemistry and biochemistry faculty members, Dr. Mark Hollingsworth, Dr. Stefan Bossmann, Dr. Ryan Rafferty, Dr. Santosh Aryal, Dr. Michal Zolkiewski, Dr. Anna Zolkiewska, and Dr. Ramaswamy Krishnamoorthi for expanding my knowledge on chemistry and biochemistry.

In addition, I want to thank my colleagues in Dr. Li's lab and Dr. Xiong's lab: Dr. Ruben Shrestha, Dr. Gaochao Huang, Dr. Bingbing Wu, Dr. Xuejie Chen, Mr. Ruikai Cao, Dr. Wei

Zhang, Dr. Chao Chen, Dr. Naren Li, Dr. Guanxing Chen, Mr. Qinfang Liu, Ms. Shifan Chen, and Mr. Liang Hu. Special thanks to Dr. Shrestha who taught me how to purify proteins, Dr. Huang who synthesized Hey-SAM, Dr. Wu who taught me how to synthesize organic compounds and Mr. Liu who helped perform the confocal imaging. I am also grateful to the REU (Research Experiences for Undergraduates) students in Dr. Li's lab: Ms. Chelsea Swartchick, Ms. Julissa Burgos and Ms. Kassidy Deutschmann.

I would also like to acknowledge Department of Chemistry for providing me with graduate teaching assistantship, and Johnson cancer Center, Dane G. Hansen Foundation and National Institute of Health for funding my research. My sincere appreciation also goes to Mr. Michael Hinton, Dr. Louis Wojcinski, Dr. Leila Maurmann, Dr. Simon Sham, Dr. Abhijeet Sinha, Dr. Tingting Liu, Mr. Tobe Eggers, Mr. Ron Jackson, Mr. Jim Hodgson, Ms. Mary Dooley, Ms. Kim Ross and Mr. Bart Bath for all their great help during my study at K-State University.

I would like to extend my thanks to all my friends, classmates and teachers in US and in China. They made my life and my study in US enjoyable, encouraging and meaningful.

Last but not the least, I would like to express my deep gratefulness to my family. My mom is a farmer, and her diligence and perseverance motivate me working hard. My dad always told me to be a responsible, modest and prudent person. My boyfriend, Song Wang, never stopped encouraging and supporting me through the past six years. My elder brother and my sister in law gave me a lot of support for my graduate school application and my Ph.D. study. All their support and deep love bring me up to this stage.

Dedication

To my parents for their deep love!

Preface

The major aim of my doctoral research is to understand the function of protein N-terminal methylation by uncovering the substrates of NTMT1. In chapter 2, an activity-based substrate profiling method has been applied to identify the NTMT1 targets. In chapter 3, we demonstrated the N-terminal methylation of protein OLA1 *in vitro* and *in vivo*. The work contained in Chapter 2 and 3 is in preparation for publication.

In addition, I have participated in several other projects including the purification of α-trichosanthes and E2, and the study of PHA synthases. Some of these results have been published and were not included in this dissertation. During my Ph.D. study, I published the following papers as a co-author:

Jia, K.; Cao, R.; Hua, D. H.; Li, P. Study of class I and class III polyhydroxyalkanoate (PHA) synthases with substrates containing a modified side chain. *Biomacromolecules* **2016,** *17* (4), 1477-1485.

Chen, C.; Shrestha, R.; **Jia, K.**; Gao, P. F.; Geisbrecht, B. V.; Bossmann, S. H.; Shi, J.; Li, P. Characterization of dye-decolorizing peroxidase (DyP) from *Thermomonospora curvata* reveals unique catalytic properties of A-type DyPs. *J. Biol. Chem.* **2015,** *290* (38), 23447-23463.

Huang, G.; Shrestha, R.; **Jia, K.**; Geisbrecht, B. V.; Li, P. Enantioselective synthesis of dilignol model compounds and their stereodiscrimination study with a dye-decolorizing peroxidase. *Org. Lett.* **2017,** *19* (7), 1820-1823.

Madera, R.; Gong, W.; Wang, L.; Burakova, Y.; Lleellish, K.; Galliher-Beckley, A.; Nietfeld, J.; Henningson, J.; **Jia, K.**; Li, P., Bai, J.; Schlup, J.; McVey, S.; Tu, C.; Shi, J. Pigs immunized with a novel E2 subunit vaccine are protected from subgenotype heterologous classical swine fever virus challenge. *BMC Vet. Res.* **2016,** *12* (1), 197-206.

Chapter 1 - Introduction

1.1 Protein N-terminal methylation and protein N-terminal methyltransferases

Protein methylations especially histone methylations have been a focal point for the past two decades due to their crucial roles in epigenetics[1-2]. Protein arginine and lysine methylations are the two most well-studied protein methylations[3-4]. A third type, the protein N-terminal methylation, remained unexplored until the discovery of N-terminal RCC1 methyltransferase 1 (NRMT1, NTMT1) in 2010[5-6], though the first case of N-terminal methylation was reported in 1976[7-8]. For all three types of protein methylation, S-adenosyl methionine (SAM) acts as a methyl donor and is converted into S-adenosyl homocysteine (SAH) after the methyl group is transferred to either the side chains of arginine and lysine residues or the N-terminal amino group of the protein. Currently, three human protein N-terminal methyltransferases have been discovered, which include RCC1 N-terminal methyltransferases 1 and 2 (NRMT1/2, NTMT1/2) and human methyltransferase-like protein 13 (METTL13)[5, 9-10]. Their similarities and differences will be reviewed below.

1.1.1 Methylation reactions catalyzed by N-terminal methyltransferases

As shown in Fig. 1.1, the initial methionine of a protein substrate is usually cleaved by an aminopeptidase before it undergoes mono-, di-, and trimethylations[5]. Multiple methylation catalyzed by protein N-terminal methyltransferases could proceed with a distributive or processive mechanism. While the former mechanism involves the release of intermediates from the enzyme prior to the formation of the final product, the latter one does not[11]. NTMT1 was reported to be a distributive trimethyltransferase, because the mono- and dimethylated products were found to be the most abundant at a low enzyme to substrate ratio (1:50)[12].

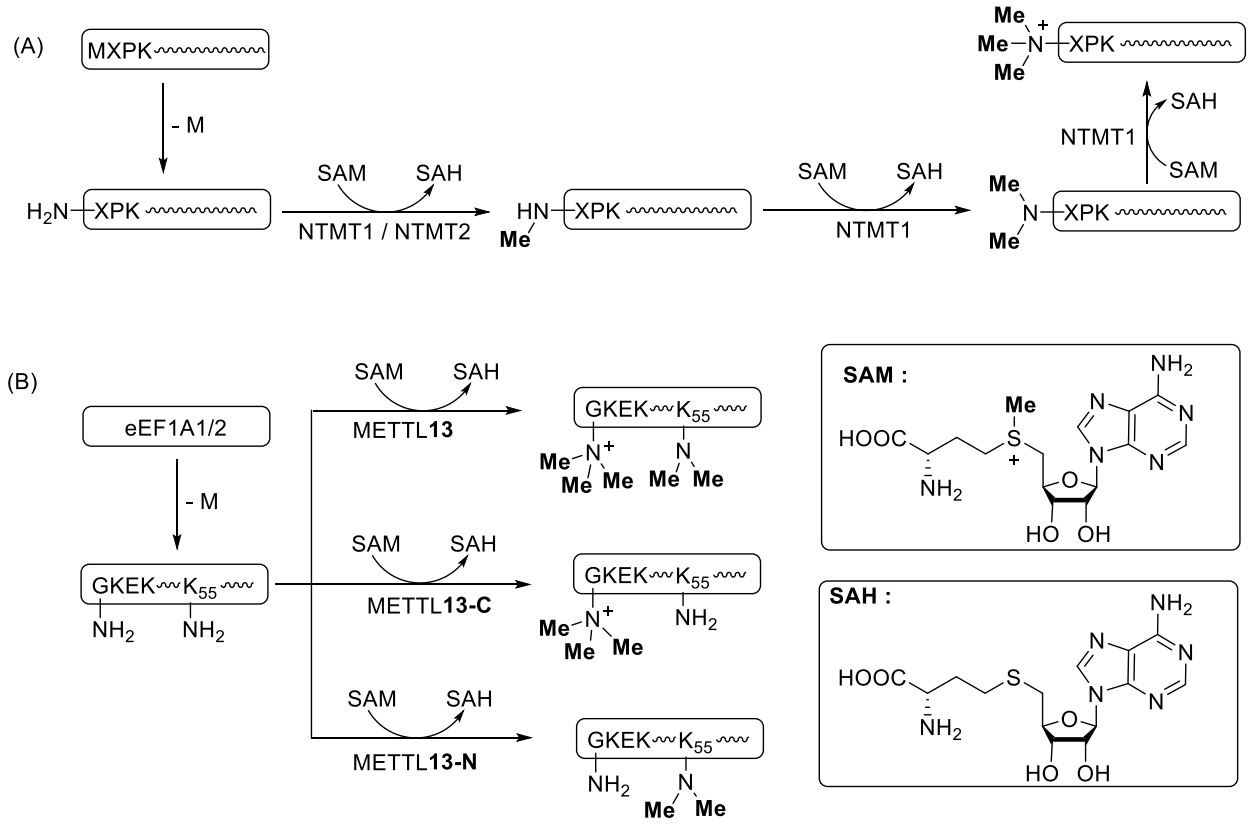


Figure 1.1 Protein methylation catalyzed by NTMT1, NTMT2, and METTL13.

(A) Protein N-terminal methylation catalyzed by NTMT1 and NTMT2 (when X = P, protein can only be dimethylated)[5]; (B) Protein eEF1A1/2 methylation catalyzed by METTL13 (METTL13-C: C-terminal domain; METTL13-N: N-terminal domain)[10].

NTMT2, an isoform of NTMT1, was initially considered as a monomethyltransferase[9]. But later study revealed that it could also catalyze the di- and trimethylations of few peptide substrates with very low catalytic efficiency[13]. METTL13 was reported as a trimethyltransferase but it differs from NTMT1 in that it specifically catalyzes the methylation of elongation factor 1-alpha 1 (eEF1A, eEF1A1/eEF1A2) using a processive mechanism[10]. Jakobsson et al. reported that the major product of METTL13 catalyzed methylation was trimethylated eEF1A and only negligible mono- and dimethylated proteins were detected at various enzyme to substrate ratios (from 0.01-

10), indicating METTL13 undergoes a processive mechanism for methylations[10]. While NTMT1/2 can only methylate the N-terminus of their targets (Fig. 1.1A), METTL13 is a dual enzyme (Fig. 1.1B) which can trimethylate the N-terminus and dimethylate Lys 55 of eEF1A using its C-terminal and N-terminal domains, respectively.

1.1.2 Substrate specificities of N-terminal methyltransferases

Substrates of the well-characterized N-terminal methyltransferases from different species are listed in Table 1. Nearly all the identified substrates of human NTMT1 and yeast YBR261C/Tae1 share an N-terminal Met-Xaa-Pro-Lys/Arg motif, such as regulator of chromosome condensation (RCC1)[5], retinoblastoma 1[5], protein SET[5], DNA damage binding protein 2 (DDB2)[14], poly (ADP-ribose) polymerase 3 (PARP3)[15], histone H3-like centromeric protein A (CENP-A)[16] and centromere protein B (CENP-B)[17]. One exception is the myosin regulatory light chain 9 (MYL9)[18], which contains an N-terminal Met-Ser-Ser-Lys motif. In addition, peptide-based methylation assay reveals that peptides with expanded N-terminal motifs may also serve as NTMT1 substrates, suggesting that NTMT1 may have very broad range of substrates (over 300 proteins)[19]. Though METTL13 is a dual methyltransferase, peptide array-based methylation assay demonstrates that it is a very specific enzyme towards its substrates eEF1A1/2[10]. The "functional homolog of METTL13-C" from yeast, Efm7, is also a dual methyltransferase and exhibits unique substrate specificity, however, it differs from METTL13 in three aspects[10]. First, Efm7 is a dual methyltransferase, but it targets the N-terminus of Gly 2 and the side chain of the adjacent Lys 3 residue[20]. Second, Efm7 cannot methylate the N-terminal peptide (first ten residues without methionine) of eEF1A1, which suggests that the tertiary structure of eEF1A may be important for substrate recognition[20]. And finally, it cannot methylate human eEF1A protein, indicating it may be a species specific enzyme[20]. Taken together, NTMT1/2 recognize an N-terminal motif and may

have very broad range of substrates, while Efm7 and METTL13 are more substrate specific enzymes.

Table 1.1 Substrates of the well-characterized N-terminal methyltransferases.

The first five protein N-terminal residues are listed in the table.

Protein N-terminal methyltransferases	Species	Substrates
YBR261C/Tae1[6]	*S. cerevisiae*	Rpl12ab (MPPKF)[6]; Rps25a/Rps25b (MPPKQ)[6]
NTMT1 (NRMT1)[5]	*Homo sapiens*	RCC1 (MSPKR)[5]; Retinoblastoma 1 (MPPKT)[5]; SET (MAPKR)[5]; DDB2 (MAPKK)[14]; PARP3 (MAPKP)[15]; CENP-A (MGPRR)[16]; CENP-B (MGPKR)[17]; MYL9 (MSSKR)[18];
NTMT2 (NRMT2)[9]	*Homo sapiens*	RCC1 (MSPKR)[9]
dNTMT1[21]	*Drosophila melanogaster*	H2B (MPPKT)[21]
Efm7 (YLR285W)[20]	*S. cerevisiae*	eEF1A (MGKEK)[20]
METTL13[10]	*Homo sapiens*	eEF1A1 (MGKEK)[10]; eEF1A2 (MGKEK)[10]

1.1.3 Catalytic mechanisms of N-terminal methyltransferases

Dong et al. and Wu et al. solved the crystal structures of the ternary complexes containing NTMT1, substrate peptides and SAH independently (PDB: 5E1B, 5CVD)[22-23]. Cheng et al. solved the crystal structures of NTMT2 in the presence of SAM (PDB: 5UBB), and the ternary complex containing NTMT2, substrate peptide and SAH (PDB: 6DUB)[13]. Crystal structure of METTL13 C-terminal domain (PDB: 5WCJ)[10] was solved by Jakobsson et al. The overall structures of all three N-terminal methyltransferases exhibit a typical Rossmann-like fold (Fig. 1.2). NTMT1 and

NTMT2 contain a seven β strand sheets surrounded by five α-helixes[13]. METTL13 contains a seven β strand sheets flanked by six α-helixes[10].

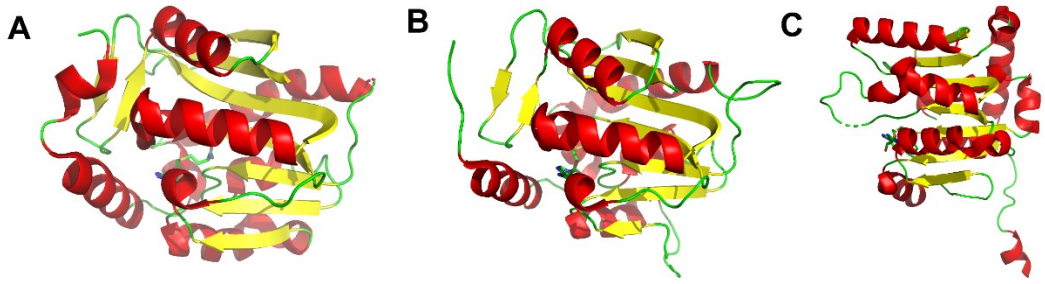


Figure 1.2 Overall structures of NTMT1, NTMT2, and METTL13-C.

(A) Crystal structure of NTMT1 with SAH (PDB: 2EX4). (B) Crystal structure of NTMT2 with SAM (PDB: 5UBB). (C) Crystal structure of METTL13 C-terminal domain with SAH (PDB: 5WCJ). Figures were prepared with PyMOL.

The ternary complex (PDB: 5E1B) of NTMT1, RCC1 peptide (SPKRIA) and SAH reveals that NTMT1 consists of a substrate binding domain and a Rossmann-like fold domain[23]. SAH is anchored inside by the electrostatic interaction between its carboxylate group and arginine 74 residue, the stacking interaction between its sugar ring and tryptophan 20 residue, and the hydrogen bonding between its adenine group and glutamine 120 residue and the main chain of leucine 119 residue (Fig. 1.3)[23]. In addition, the crystal structure uncovers why a substrate harboring an N-terminal Xaa-Pro-Lys motif is preferred for NTMT1. First, the dimethylated glycine locates in a narrow pocket which may also be able to hold other small residues (e.g., proline, serine, alanine)[23]. Second, proline 3 is surrounded by a hydrophobic environment formed by residues leucine 31, isoleucine 37, tryptophan 136 and isoleucine 214 (Fig. 1.3)[23]. Moreover, the stacking interaction between proline 3 and tryptophan 136 residue (Fig. 1.3) makes proline the optimum residue in the third position, which explains why proline 3 is conserved among all the

identified substrates[23]. Third, a positively charged residue in position 4 is crucial for substrate recognition due to the hydrogen bonding between lysine 4 and aspartate 177 and 180 residues (Fig. 1.3)[23]. In addition, Dong et al. pointed out that a positively charged residue in position 5 is preferred because it is in the entrance of a negatively charged channel[22].

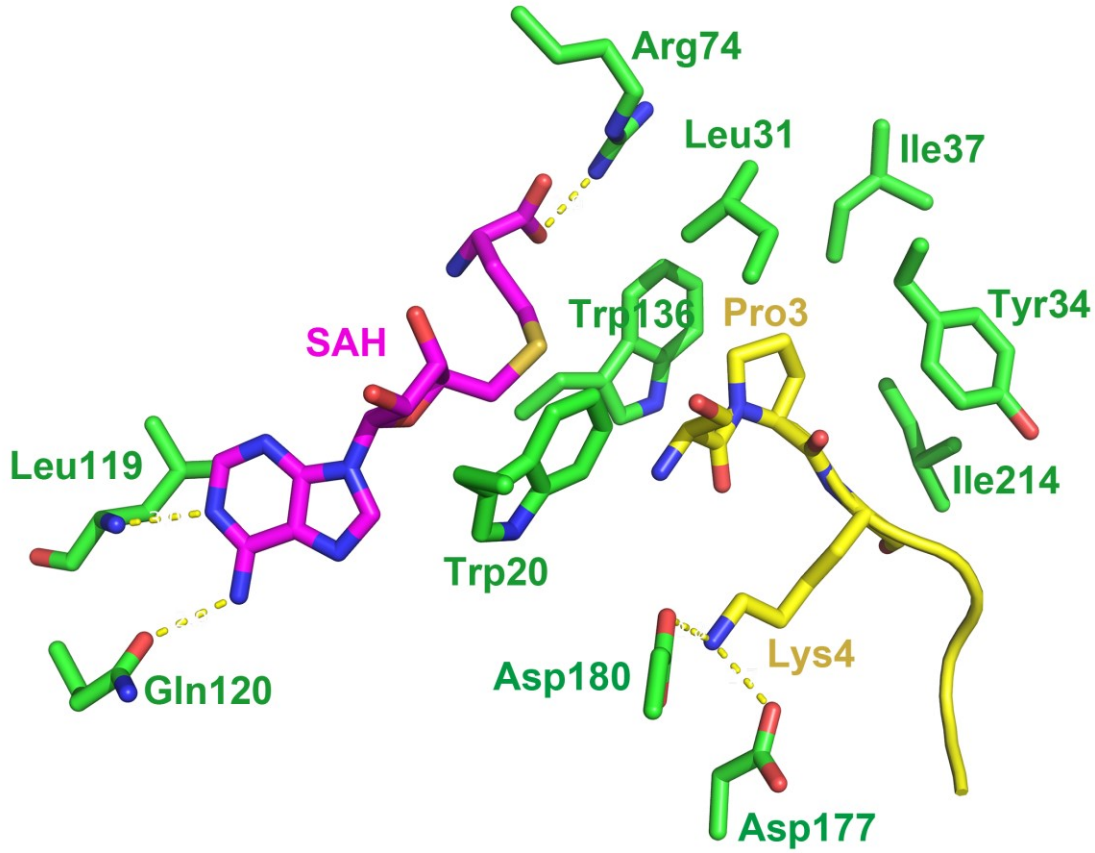


Figure 1.3 Substrate binding domain of NTMT1.

Ternary complex of NTMT1, SAH and RCC1 peptide SPKRIA (PDB: 5E1B). SAH is shown in magenta. NTMT1 residues are shown in green. Peptide SPKRIA is shown in yellow. Figure was prepared with PyMOL.

An S_N2 reaction mechanism for methylation has been proposed according to the crystal structures[22-23]. Briefly, N-terminal α amino group of the substrate is deprotonated by water

mediated hydrogen bonding network consisting of NTMT1 residues aspartic acid 167, aspartic acid 180 and histidine 140[22]. After that, the deprotonated nucleophilic nitrogen can attack the active methyl group of SAM to give the methylated product and SAH.

```
NRMT1              1 M-------------------------------------------------      1
                     |
NRMT2              1 MAHRGAHFAFRSRWQKTDDELCRHSMSFILHKAIRNDFFQSYLYLLEKIP     50

NRMT1              2 -------TSEVIEDEKQFYSKAKTYWKQIPPTVDGMLGGYGHISSIDINS     44
                           ||:||..|.|||::||.:::::|.|.:||:|.:..:||.||.:
NRMT2             51 LVKLYALTSQVINGEMQFYARAKLFYQEVPATEEGMMGNFIELSSPDIQA    100

NRMT1             45 SRKFLQRFLREGPNKTGTSCALDCGAGIGRITKRLLLPLFREVDMVDITE     94
                     |:|||::|:. ||.:.||.||||||:||||::|.:|||:|..|::||:.|
NRMT2            101 SQKFLRKFVG-GPGRAGTDCALDCGSGIGRVSKHVLLPVFNSVELVDMME    149

NRMT1             95 DFLVQAKTYLGEEGKRVRNYFCCGLQDFTPEPDSYDVIWIQWVIGHLTDQ    144
                     .||::|:.||..:|.:|.:|.|..||:|||....|||||||||.|||||:
NRMT2            150 SFLLEAQNYLQVKGDKVESYHCYSLQEFTPPFRRYDVIWIQWVSGHLTDK    199

NRMT1            145 HLAEFLRRCKGSLRPNGIIVIKDNMAQEGVILDDVDSSVCRDLDVVRRII    194
                     .|..||.||:..|:.||||::|||:|:||.|||..||||.||:|::|.:|
NRMT2            200 DLLAFLSRCRDGLKENGIIILKDNVAREGCILDLSDSSVTRDMDILRSLI    249

NRMT1            195 CSAGLSLLAEERQENLPDEIYHVYSFALR-----    223
                     ..:||.:|.:|:|:..|::...|:.|||.
NRMT2            250 RKSGLVVLGQEKQDGFPEQCIPVWMFALHSDRHS    283
```

Figure 1.4 Sequence alignment of NTMT1 and NTMT2.

NTMT1 and NTMT2 are aligned using online EMBOSS Stretcher (EMBL-EBI) which gives 40% identity and 58% similarity. Identical residues are indicated as a vertical line (|); chemically similar residues are indicated as double dots (:); chemically different residues are indicated as one dot (.); missing residues are indicated as a dash line (-). Conserved residues in substrates binding domain and catalytic domain are highlighted in yellow. Asparagine 89 (NTMT2) and its aligned glycine residue (NTMT1) are colored in red.

The key residues in substrate binding domain and catalytic domain are conserved in NTMT1 and NTMT2 (Fig. 1.4), however, unlike NTMT1, NTMT2 is primarily a monomethyltransferase[9]. Based on the ternary structure of NTMT2, SAH and RCC1 peptide SPKRIA, Dong et al. proposed that asparagine 89 might be involved in determining NTMT2 as a monomethyltransferase[13].

Peptide-based activity assay confirmed that NTMT2 N89G mutant was able to dimethylate RCC1 peptide[13]. However, the trimethylated RCC1 peptide was not observed, indicating that additional residues in substrate binding domain may also participate in determining the catalytic specificities of NTMT1/2[13].

1.1.4 Tissue and subcellular locations of N-terminal methyltransferases

NTMT1 and NTMT2 were reported to locate mainly in cell nucleus based on fluorescence signals given by overexpressed GFP tagged proteins in HeLa cells[9]. This is consistent with the subcellular distribution of their substrates (e.g., DDB2, RCC1). One exception is MYL9 which primarily is in the cytosol[18]. According to human protein atlas database, the major location of METTL13 is cell cytosol[24].

It was reported that the transcription level of NTMT1/2 messenger RNA (mRNA) quantified by quantitative real-time polymerase chain reaction (qRT-PCR) of mouse tissue extracts was low, suggesting a low protein expression level of NTMT1/2[9]. The maximum mRNA expression level of NTMT1 was detected in brain tissue, and muscle tissue contained the highest mRNA level of NTMT2[9].

1.2 Functions of protein N-terminal methylation

Protein N-terminal methylation has long been believed to regulate protein degradation because the introduction of methyl groups may block the protein N-terminus and protect it from digestion by aminopeptidases[25]. The N-terminal methylation of protein RCC1 was discovered in 2007, which revealed one novel function of protein N-terminal methylation in regulating protein-DNA interaction. Chen et al. reported that Madin-Darby Canine Kidney (MDCK) cells transfected with methylation deficient RCC1 (K4Q mutant) exhibited a more diffused distribution pattern of RCC1 compared to the ones transfected with wild type (wt) RCC1. Transfection of methylation deficient

RCC1 ultimately caused the abnormal mitosis[5]. It was proposed that trimethylation of RCC1 which introduced a permanent positive charge on its N-terminus would promote its binding with the negative charged phosphate backbone of DNA. Conversely, absence of trimethylation would weaken the protein-DNA interaction and result in the diffused distribution of RCC1[5].

Similar to RCC1, trimethylation of CENP-B N-terminus was found to enhance its binding with DNA[17]. Sathyan et al. reported that the N-terminal methylation of CENP-A played important roles in regulating centromere function and maintaining normal chromosomal segregation[16]. The methylation deficient CENP-A mutant caused the reduced centromere localization of CENP-T and CENP-I which are important components of constitutive centromere associated protein network (CCAN) and promoted the formation of multipolar spindles[16].

In addition, protein N-terminal methylation was reported to participate in DNA damage repair. DDB2 exhibited enhanced cell proliferation ability upon UV light irradiation compared to its methylation deficient mutant (K4Q). The methylation deficient DDB2 displayed less nucleus localization[14]. Cai et al. proposed that the N-terminal methylation of DDB2 might affect protein-protein interactions[14], however, the molecular mechanism for its involvement in DNA damage repair remains unknown. So far, no readers of methylated DDB2 were reported.

Moreover, the N-terminal methylation of MYL9 exhibited enhanced transcriptional activation of protein intercellular adhesion molecule 1 (ICAM1) compared to its methylation deficient mutant (K4Q)[18]. The molecular mechanism is unclear, but methylation of MYL9 was proposed to enhance its interaction with ICAM1 promoter[18].

Functions of eEF1A N-terminal methylation were not well understood. Knockout (KO) of METTL13 in HAP-1 cells affected the translation rate of several specific codons[10]. For example, the translation of alanine codons accelerated, while the translation rate of lysine and histidine

codons decreased[10]. So far, eEF1A is the only identified substrate of METLL13 and it would be possible that the N-terminal methylation of eEF1A participates in influencing the codon translational rates. However, to verify this hypothesis, it would be a good start point to study the effects of methylation deficient eEF1A mutant on codon translation.

Taken together, protein N-terminal methylation plays multiple roles in maintaining cellular processes such as regulation of protein-DNA interaction, maintaining normal centromere function and chromosomal segregation, participating in DNA damage repair, and affecting amino acid codon translational rates.

1.3 Phenotypes related to dysregulation of N-terminal methyltransferase 1

Overexpression of NTMT1 in M. D. Anderson Metastasis Breast cancer 231 (MDA-MB-231) cell line resulted in the decreased cell growth rate[26]. Knockdown of NTMT1 promoted the growth of Michigan Cancer Foundation-7 (MCF7) cell line and enhanced the metastatic potential of SKBR-3 and MDA-MB-231 cell lines[26]. All these results suggest that NTMT1 functions as a tumor suppressor in human breast cancer cell lines[26]. Conversely, Shields et al. reported that NTMT1 knockout in colon cancer HCT116 cell line dramatically reduced the cell proliferation rate, indicating that NTMT1 might act as a tumor oncogene in colorectal cancer[27]. NTMT1 has very broad substrates, which may have different expression levels and play distinct roles in different tumors and thus caused this contradiction. NTMT1 mutations in human cancer tumors can be found from the website of Catalogue of Somatic Mutations in Cancer (COSMIC, https://cancer.sanger.ac.uk/cosmic). Most of the mutation sites reported are not key residues for substrate binding or catalytic activity. Shields et al. reported that two mutants of NTMT1 (N209I in endometrial cancer and P211A in lung cancer) resulted in a much lower catalytic activity compared to wt NTMT1[27].

In addition, Bonsignore et al. generated NTMT1 KO mice (*Ntmt1*$^{-/-}$) which exhibited abnormal phenotypes correlated to double strand DNA damage[28]. Around 40% of NTMT1 KO homozygous mice died in one month after birth and the survived ones showed decreased body size, kyphosis, smaller livers and female infertility[28]. Taken together, dysregulation of N-terminal methyltransferases is related to the carcinogenesis promotion and the cause of developmental diseases.

1.4 Current chemical and biological tools for methylation study

As protein N-terminal methylation plays multiple functions in organism growth and development, it will be of great importance to study their upstream regulators (writers) and downstream regulators (readers and erasers). Writers of methylation are the enzymes that catalyze this modification and are also called "methyltransferases". Readers of methylation are the proteins which can recognize and bind to the methylation sites. Erasers of methylation are the enzymes responsible for removing this modification and are also named as "demethylases". In addition, NTMT1 was predicted to have multiple substrates and only a few of them have been identified. Thus, it is essential to uncover more of its substrates to elucidate the functions of this important modification. Here, the chemical and biological methods for profiling the upstream and downstream regulators are reviewed and organized into four modules: methylation "writers", substrates, "readers" and "erasers". These approaches may be further applied in the study of N-terminal methylation.

1.4.1 Methylation writers

The discovery of NTMT1 as an N-terminal methyltransferase was three years later after the discovery of RCC1 N-terminal methylation[5, 29]. NTMT1 was previously named as methyltransferase-like 11A (METTL11A), indicating that it may function as a methyltransferase,

but its catalytic reaction was unknown. The important roles of RCC1 N-terminal methylation in mitosis triggered the discovery of the corresponding methyltransferase, NTMT1. Tooley et al. developed an activity-based methyltransferase profiling method in which HeLa nuclear extracts were separated into several fractions by hydroxyapatite ion exchange chromatography and each fraction was examined for methylation activity[5]. Each fraction contained hundreds of proteins which greatly narrowed down the number of proteins in the candidate pool. Antibody immunoblotting-based activity assay and enzyme-linked immunosorbent assay (ELISA) were applied as a rapid approach to examine the methylation activity of each fraction and finally, the methylation activity was observed in two fractions which were further analyzed by mass spectrometry[5]. Several key elements should be considered for the activity-based profiling approach: (a) Various subcellular fractions (e.g., nuclear, membrane, cytoplasmic) should be firstly screened for methylation activity to determine the proper fraction for further analysis. Tooley et al. only focused on nuclear fraction which eliminated unnecessary work and simplified the analysis[5]. (b) Multiple protein separation methods such as ion exchange and size exclusion chromatography should be applied to limit the number of proteins in one fraction. (c) A sensitive, simple and fast methylation assay should be developed to avoid false negative results and tedious work. Several sensitive and rapid methylation assays have been developed including fluorescence-based SAH hydrolase (SAHH) coupled assay[12], enzyme coupled ultrasensitive luminescence assay[30], mass spectrometry (MS)-based methylation assay[12], antibody-based immunoblotting[5] and ELISA[5]. (d) And finally, high-resolution mass spectrometry (HRMS) coupled with either an ultra-performance liquid chromatography (UPLC) or a nanoflow liquid chromatography is required for the identification of proteins in low quantities as some endogenous enzymes may have very low expression levels but high catalytic activities.

Moreover, in prokaryotes (e.g., *Escherichia coli* (*E. coli*)) or single-celled eukaryotes (e.g., yeast) which contain less amount of proteins compared to human or mammals and thus have been studied as model organisms for a long time, the activity-based methyltransferase profiling strategy is further simplified. For example, Kimura et al. described a reverse-genetics approach in *E. coli* for profiling RNA methyltransferase which catalyzed the N^6-methylation of N^6-threonylcarbamoyladenosine (t^6A) in transfer RNA^{Thr} ($tRNA^{Thr}$)[31]. The *E. coli* genomic-deletion strains which covered ~50% of *E. coli* open reading frames (ORFs)[31] were analyzed by liquid chromatography-mass spectrometry (LC-MS) to screen the existence of N^6-methylated t6A (m6t6A) in $tRNA^{Thr}$. This modification was only not detected in OCR36 strain, which narrowed down the number of candidate enzymes to 23[31]. The number of candidate proteins was further decreased to 10 following bioinformatic analysis, and finally, gene *yaeB* was identified for coding the responsible methyltransferase[31]. Jackman et al. reported a genomic approach to profile the responsible methyltransferase for m1G9 in tRNA from yeast *Saccharomyces cerevisiae*[32]. In this approach, a yeast strain array constructed by Martzen et al. was used for screening the methylation activity[33]. The array contained 6144 yeast strains and each strain was able to express a specific glutathione S-transferase (GST)-tagged yeast protein[33]. To simplify, 96 strains were grown in a pool and 6144 strains were thus divided into 64 pools[33]. GST-tagged proteins were purified from each pool and their methylation activity was examined[33]. After one-round screen, the number of candidate proteins were efficiently narrowed down to 96[33]. A second-round screen assayed 20 subpools and a protein coded by *YOL093w* was finally confirmed to be the relevant methyltransferase[32]. Meanwhile, three human homolog proteins were also identified[32]. The reverse-genetics and genomic approach is simply because the separation of cellular proteins is not required as either gene KO strains or a strain expressing an individual protein is accessible.

However, one limitation of this approach is that lots of proteins derived from advanced organisms do not have their homologs in either *E. coli* or yeast.

In addition, METTL13 was identified as an N-terminal eEF1A methyltransferase by co-immunoprecipitation (co-IP) with eEF1A N-terminal peptide[10]. Based on the prediction that "writers" may bind to their substrates tighter than their products, trimethylated eEF1A peptide was used as a control for MS-based quantitative pull-down assay[10]. Peptides or proteins used for co-immunoprecipitation are often called "baits" and usually contain a report group which could be an affinity tag or a functional group that is ready to incorporate an affinity tag. Affinity tags could be a biotin which can be pulled down by streptavidin beads due to their strong interaction ($K_d \approx 10^{-14}$ M)[34] or epitope tags (e.g., Flag-tag, Myc-tag) which can be purified by immunoaffinity resins. Jakobsson et al. utilized biotin labeled peptide as a bait for profiling eEF1A methyltransferase[10]. Two key points should be considered for this strategy: (a) One or more controls must be carried out at the same time to eliminate false positive results. Parallel biological experiment repeats, and statistical analysis are usually required for quantitative MS-based protein identification. (b) The bioinformatic analysis may assist to find of the correct enzyme because lots of the proteins identified are not methyltransferase and some may even not be biologically relevant. Jakobsson et al. reported that 157 proteins were enriched by the bait and 174 proteins were enriched by reference peptide[10]. Without bioinformatic analysis, it would be challenging to examine the methylation activity of more than 100 proteins through trial and error.

Currently, bioinformatics has been a powerful tool in the prediction of putative methyltransferases. As more and more methyltransferases are identified and lots of their crystal structures are solved, it becomes easier to predict novel methyltransferases based on the conserved SAM binding motifs or by full sequence alignment with Basic Local Alignment Search Tool

(BLAST)[35]. Though we cannot directly predict a methyltransferase for a specific target through bioinformatic analysis, but it is a powerful tool to provide useful information for further study.

In summary, four methods including activity-based methyltransferase profiling from human cell lysates, reverse-genomic approach, co-immunoprecipitation with substrate peptides as baits and bioinformatics have been reviewed. Each method has its own strength and weakness, so it would always be wise to think of their limitations and applications to make the best choice. Moreover, a sensitive and reliable methylation activity assay or a methylation detection approach is required for all four strategies.

1.4.2 Methylation substrates

A widely used approach for substrate profiling is domain or motif searching since the substrates usually share some common characteristics for a specific enzyme. For example, most of the N-terminal methylated eukaryotic proteins contain an N-terminal Met-Xaa-Pro-Lys/Arg motif[5]. Later, several proteins including DDB2, CENP-A, and CENP-B which harbor the same motif were predicted and successfully confirmed as NTMT1 substrates[14-16]. Moreover, Tooley el. al used anti-me_3-SPK antibody which specifically recognizes trimethylated Ser-Pro-Lys motif to pull down potential NTMT1 targets[5]. Six proteins harboring the N-terminal motif including RCC1 and SET were pulled down[5]. Substrate profiling *via* immunoprecipitation was also applied in identifying arginine methylated proteins[36]. For example, Boisvert et al. identified 200 new proteins as putative substrates for protein arginine methyltransferases through immunoprecipitation with specific antibodies against dimethylated arginine in RG-rich sequences[36].

In addition, peptide array has become a very popular tool for profiling methylation substrates as it can be rapidly synthesized and applied for high-throughput screening. The rapid synthesis of peptide array owed to the invention of solid phase peptide synthesis by Robert Bruce Merrifield,

who won Nobel Prize in chemistry for this innovation. Frank et al. reported an approach to synthesizing multiple peptides on cellular paper discs simultaneously which made peptide array attainable[37-38]. Peptide array was widely applied in the study of protein arginine and protein lysine methylation. Najbauer et al. reported that synthesized peptides with Gly-Arg-Rich (GAR) motif were able to be methylated by protein arginine methyltransferases (PRMTs) and inhibited the methylation of endogenous hypomethylated substrates, which indicated that substrate proteins could be replaced by short peptides for methylation study to some extent[39]. Later, Wooderchak et al. applied peptide array in the study of PRMT1 substrate specificity and concluded that PRMT1 substrates may be not limited to the ones which harbor the dogmatic "RGG" motif[40]. Similarly, Dhayalan et al. applied peptide array in the study of protein lysine methyltransferase SET7/9 and revealed 91 new peptides and 9 non-histone proteins as putative substrates[41]. Recently, peptide array-based methylation assay has been applied in the study of protein N-terminal methylation, which expands the N-terminal recognition motif of NTMT1 substrates and more than 300 proteins have been predicted as putative substrates[19]. Although peptides have been widely used for protein methylation study, it has to be noted that peptide-based substrate profiling may generate false positive or false negative results. For example, yeast Efm7 catalyzed the N-terminal methylation of protein eEF1A but was unable to methylate eEF1A N-terminal peptide[20], suggesting that the protein 3D conformation of eEF1A may be involved in substrate recognition for yeast Efm7. On the contrary, methylation on peptides level does not mean the corresponding proteins are true substrates because the very flexible peptides may be buried inside of the 3D structure of proteins in which methylation sites are blocked.

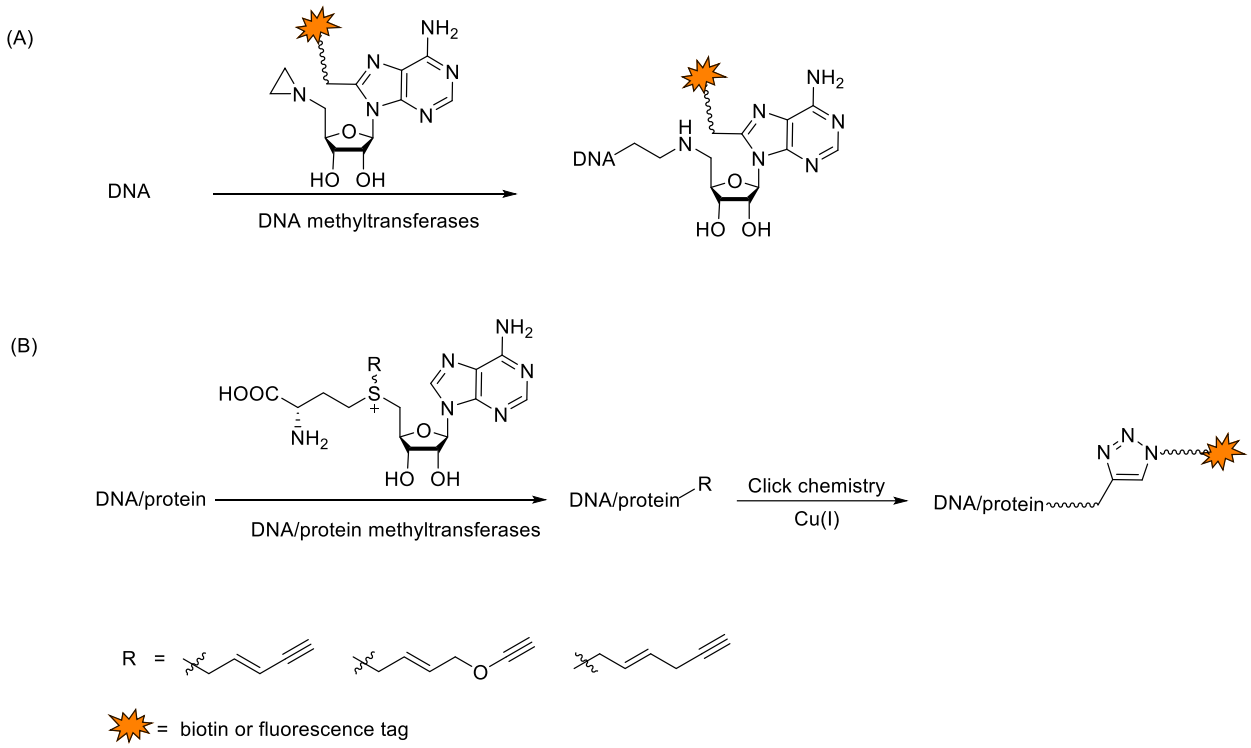


Figure 1.5 Labeling DNA/Protein with SAM analogs.

(A) Labeling DNA with aziridine cofactors[42-43]. (B) Labeling DNA/protein with sulfonium-alkyl SAM analogs[44-48].

Moreover, labeling substrates with a reporter group (e.g., fluorescence tag, affinity tag) *via* chemical approaches has been investigated for two decades. The idea is to introduce a reporter group to the substrates through methyltransferase catalyzed group transfer reactions. To achieve this, two types of SAM analogs (Fig. 1.5) were designed and synthesized. The first type contains an aziridine group and a reporter group[42-43], which can directly label the substrate proteins by one step methylation reaction. Another type is the sulfonium-alkyl SAMs which will not introduce a bulky report group to substrates directly, but instead, the substrate proteins are alkylated with an alkynyl or azide group[44-48]. Through one step Copper-Catalyzed Azide-Alkyne Cycloaddition (CuAAC), substrates can be efficiently labeled with a reporter group. CuAAC was also named as click chemistry and was concurrently discovered by the group of K. Barry Sharpless and Morten

Meldal in 2002 which provides an efficient approach for bond formation between an azide and an alkyne[49-50]. SAM analogs listed in Fig. 1.5 have been applied to label substrates of DNA methyltransferases, protein arginine methyltransferases, and protein lysine methyltransferases[42-48]. Note that wild type methyltransferases usually cannot utilize the bulky sulfonium-alkyl SAMs and protein engineering is required[44-48]. In addition, endogenous substrates should keep in a hypomethylated state for proper enzymatic labeling. A methyltransferase inhibitor, adenosine-2',3'-dialdehyde (Adox), was usually added to the cell culture medium to induce hypomethylated proteins[44]. However, Adox is a broad methyltransferases inhibitor which may affect normal cellular functions and cell growth. Moreover, hypomethylated substrates of irrelevant methyltransferases may also be labeled to some extent which would result in false positive results. Clustered regularly interspaced short palindromic repeats (CRISPR)-cas9 based gene KO system provides a new approach for inducing specific hypomethylated proteins.

Finally, Jakobsson et al. described an HRMS-based substrate profiling approach for METTL13-N[10]. In this method, proteins were labelled using stable isotope labeling with amino acids in cell culture (SILAC) for better quantification, and the corresponding enzyme METTL13 was knocked out[10]. Proteins obtained from normal and METTL13 KO cell lysates which contained the methylation sites were identified, quantified and compared[10]. As a result, Lys55 of eEF1A was identified as a new target site for METTL13-N and METTL13 was confirmed to be a dual methyltransferase[10].

Taken together, four methods including immunoprecipitation, peptide array, activity-based substrates labeling, and HRMS-based substrate profiling are reviewed. Immunoprecipitation is rapid, but it relies on high-quality antibodies. Peptide array is a high-throughput method but may not reflect the real situation of proteins. Activity-based substrate profiling provides a way to

directly label and enrich endogenous proteins, but it is time-consuming to design probes, synthesize probes and engineer proteins. HRMS-based method is straightforward, but it highly relies on expensive equipment and comprehensive MS data analysis.

1.4.3 Methylation readers

To fully understand the functions of protein methylation, it is important to identify the "reader" proteins that determine the function of the methylation. The study of DNA and protein (especially histone) methylation readers uncovered their importance in epigenetic regulations. For example, Harris et al. reported that the binding of transcriptional anti-silencing factor (SUVH1 and SUVH3) to methylated DNA in *Arabidopsis thaliana* recruits DnaJ protein to enhance the expression of proximal genes[51]. Wysocka et al. reported that bromodomain and plant homeodomain (PHD) finger transcription factor (BPTF) specifically recognizes histone K4me3 and recruits the associated ATPase, SNF2L, to facilitate the transcription of chromatin[52]. Protein N-terminal methylation was reported to regulate protein-DNA. For example, trimethylated RCC1 enhanced its interaction with chromosome DNA and maintained normal mitosis[5].

Several motifs of histone lysine methylation readers have been confirmed including chromodomain, tudor domain, WD40-repeat domain, PHD domain, and acidic domain[52-53]. Wu et al. reported that histone K4me2/3 readers, NPM1, pp32, TaF1β, NCL and UBF, all harbored an acidic domain which mainly consisted of aspartic and glutamic acid residues[53]. Further, the acidic domain was confirmed to be necessary for their binding to histone K4me2[53]. Moreover, 655 proteins containing the acidic domain were predicted as putative histone $K4me_2$ readers and protein MTERD2, Polr3gl and VHL were confirmed to bind to methylated histone K4 using pull-down assay[53]. Taken together, histone methylation readers share common domains and it is a rapid approach to predict putative methylation readers by domain searching.

In addition to domain-based prediction, the most widely used approach to profile methylation reader is by co-IP. For example, protein SUVH1 and SUVH3 were identified as DNA methylation readers through co-immunoprecipitation with methylated double strand DNA oligonucleotides consisting a biotin affinity tag[51]. BPTF was found as a binding partner of histone H3K4me3 through co-IP with trimethylated histone H3 peptide[52].

As some of the interactions between reader proteins and the methylation sites are weak and transient, it would be difficult to capture those readers by direct co-immunoprecipitation. A photo-affinity based protein pull-down method which enables covalently crosslinking has been applied in the profiling of histone methylation readers. Several photo-affinity probes containing a methylation marker, a photo reactive group and a report group have been designed and examples are shown in Fig. 1.6. Yang et al. demonstrated that probe 2 with a diazirine group (Fig. 1.6.A) is more efficient than probe 1 (Fig. 1.6.A) in labeling a known H3K4me3 reader, spindlin1[54]. They explained that the bulky benzophenone group would block the binding site and disrupt the interaction between readers and H3K4me3[54]. The mechanisms for radical generation *via* UV radiation are shown in Fig. 1.6.B. Both benzophenone and diazirine generate radicals under relatively longer UV wavelength (~ 350nm), which are good for *in vitro* and *in vivo* protein labeling. The photo excitation of benzophenone is reversible which gives good labeling yield by repeated excitation under UV radiation[55]. Unlike benzophenone, diazirine will release N_2, which makes its photo excitation irreversible[55]. Compared to benzophenone, diazirine is chemically unstable as it isomerizes under light or by heating[55]. However, the small size of diazirine group makes it widely used in peptide-based photo affinity probes.

Recently, Zhai et al. reported a self-assembled multivalent photo-affinity probe (Fig. 1.6.C) which exhibited high efficiency and specificity[56]. Methylated peptide and photo reactive group for

covalent crosslinking were immobilized on the surface of gold nanoparticles (AuNPs) which enabled purification of binding partners through one centrifugation step. Nonporous structure of AuNPs eliminated the non-specific bindings[56]. In addition, its multivalent property improved the crosslinking efficiency[56].

Although peptide-based photo affinity probes are easily synthesized and exhibit high labeling efficiency, they are often applied in cell lysates *in vitro*. Chin et al. reported a method to incorporate unnatural amino acid (UAA), p-benzoyl-L-phenylalanine (Fig. 1.6.D), to proteins synthesized in *E. coli in vivo* by selection of an aminoacyl-tRNA synthetase and tRNA which recognized one of the stop codons, TAG[57]. Incorporation of p-benzoyl-L-phenylalanine to proteins makes *in vivo* photo crosslinking possible.

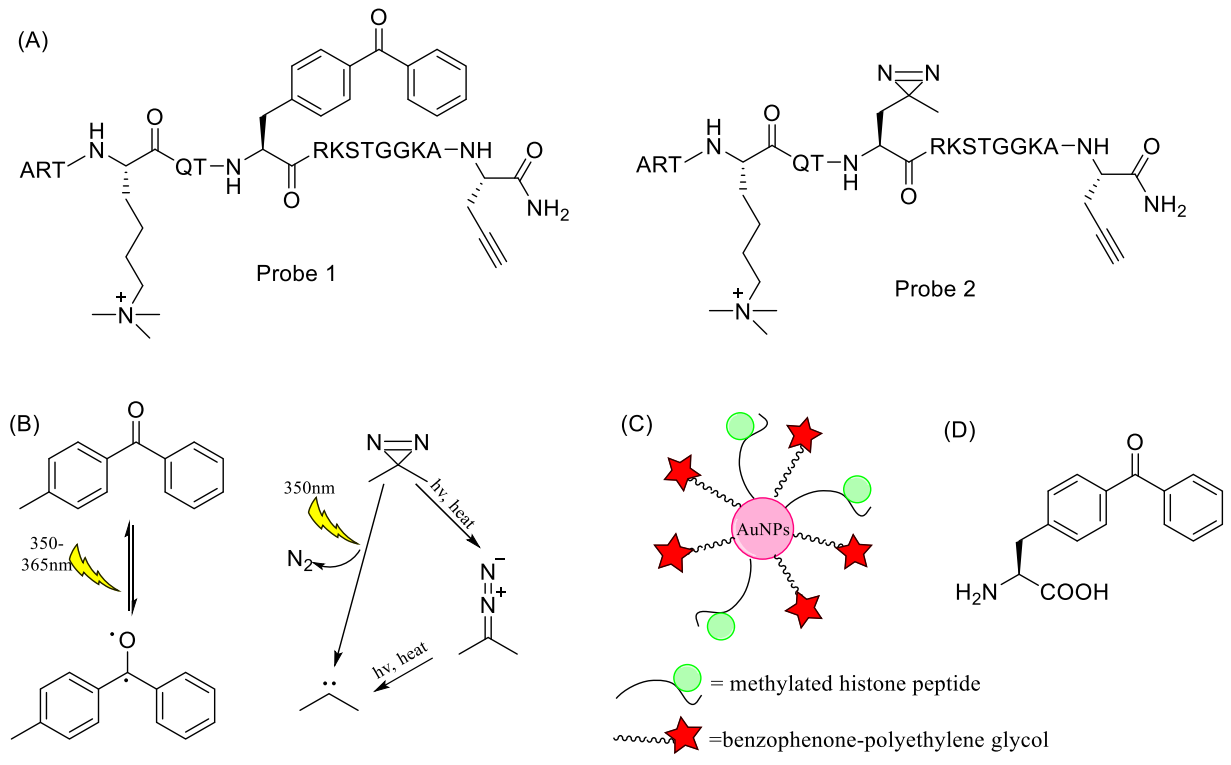


Figure 1.6 Photo active probes.

(A) Photo affinity probes for histone H3K4me3 (probe1: benzophenone containing probe, probe2: diazirine containing probe)[54]. (B) Mechanisms for radical generation of benzophenone or diazirine containing probes[55]. (C) Self-assembled multivalent probe for histone H3K4me3[56]. (D) Photo active amino acid (p-benzoyl-L-phenylalanine)[55].

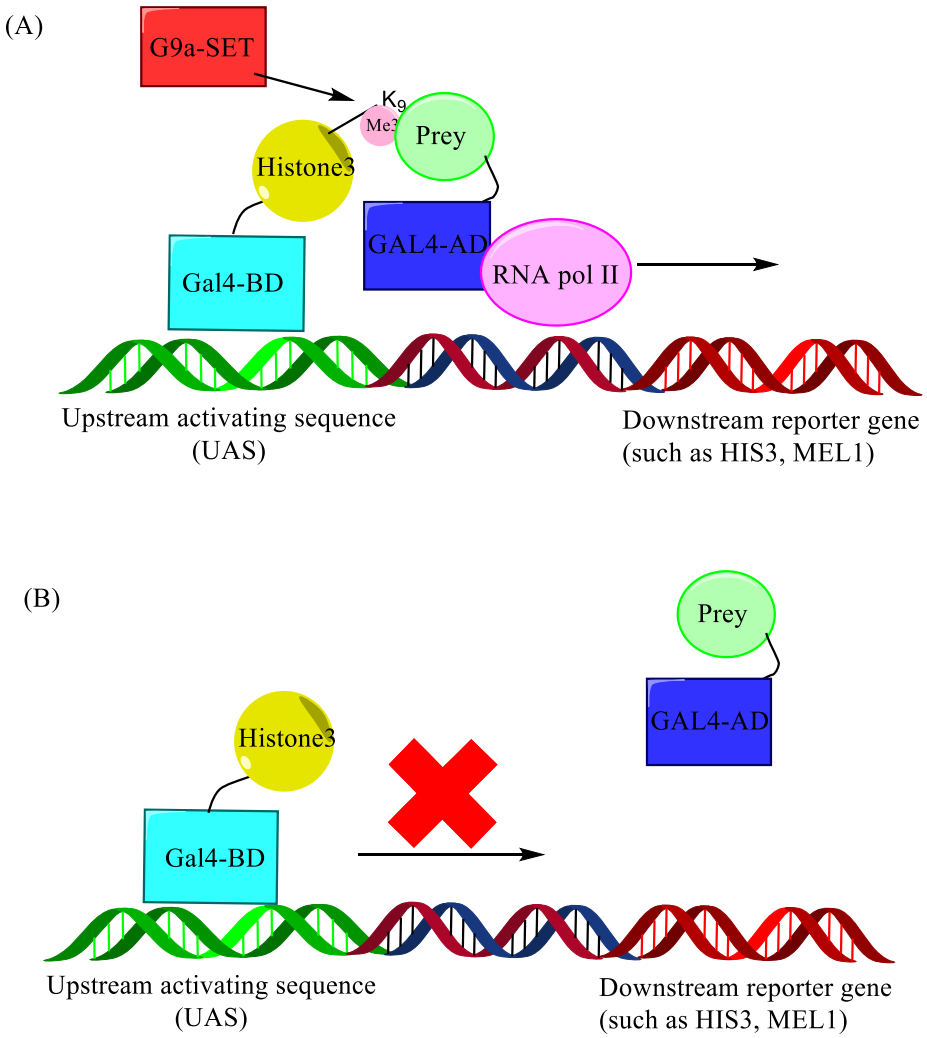


Figure 1.7 Yeast three-hybrid system for screening of histone 3 K9me3 reader.

(A) Histone3 N-terminal peptide fused to GAL4-binding domain (BD) (bait) recognizes and binds to the promoter of the reporter gene. H3K9 is methylated by SET domain of G9a and interacts with the prey protein which is fused to GAL4-activation domain (AD). The reconstitution of GAL4-BD and GAL4-AD activates the transcription of the downstream reporter gene. (B) The SET domain of G9a is not expressed in the system and the non-methylated H3K9 serves as a negative control which will not be recognized by the methylation reader and thus cannot reconstitute GAL4-BD and GAL4-AD. As a result, transcription of the downstream reporter gene is not activated[58].

The yeast two-hybrid (Y2H) system is a rapid and high-throughput approach for studying protein-protein interactions *in vivo*. It does not rely on expensive or sensitive equipment and does not need protein purifications or enrichments. In contrast, it is a genetic-based method for screening protein-protein interaction and its principle is shown in Fig. 1.7[58]. Fields et al. firstly reported this novel genetic system for studying protein-protein interaction in 1989[59]. The basic premise is that the activation domain (AD) and binding domain (BD) for most transcriptional factors (such as GAL4) are modular[60]. The AD activates gene transcription either through covalently fused to BD or stay proximity to BD by interacting with an intermediate protein[59-61]. The Y2H is a high-throughput method as it can screen protein-protein interaction with a gene library containing all the ORFs of an organism. Wieczorek et al. reported a yeast three-hybrid (Y3H) system for the screening of histone lysine methylation readers[58]. It is called three-hybrid as a third protein, histone lysine methyltransferase G9a (SET domain) which can modify the H3K9 in *vivo*, is introduced to the system together with the bait and prey proteins[58]. Four previously confirmed H3K9me3 readers and 20 new putative readers were identified through Y3H[58]. It has to be noted that Y2/3H system with gene library screen often gives false positive and false negative results[62], however, this rapid approach paves a cost-effective way for the rapid identification of protein methylation readers *in vivo*.

Considering the important roles of protein N-terminal methylation in cellular processes and little is known for their downstream signaling pathways, it is necessary to identify the N-terminal methylation readers. As a start point, immunoprecipitation, photo-affinity crosslinking and yeast three-hybrid system provide rapid and effective ways to screen new methylation readers.

1.4.4 Methylation erasers

So far, two types of histone demethylases have been identified (Fig. 1.8). The first histone demethylase, nuclear polyamine oxidase (nPAO, LSD1, KDM1A, Fig. 1.8A) was discovered in 2004, which is a flavin adenine dinucleotide (FAD) dependent monoamine oxidase and demethylates mono- and dimethylated lysine[63]. The second type of demethylases (Fig. 1.8B) are JmjC domain-containing histone demethylases. The first confirmed JmjC domain-containing histone demethylase is JHDM1 (KDM2) which is Fe(II) and α-ketoglutaric acid (AKG, 2OG) dependent hydroxylase and can demethylate mono-, di-, and trimethylated lysine[63].

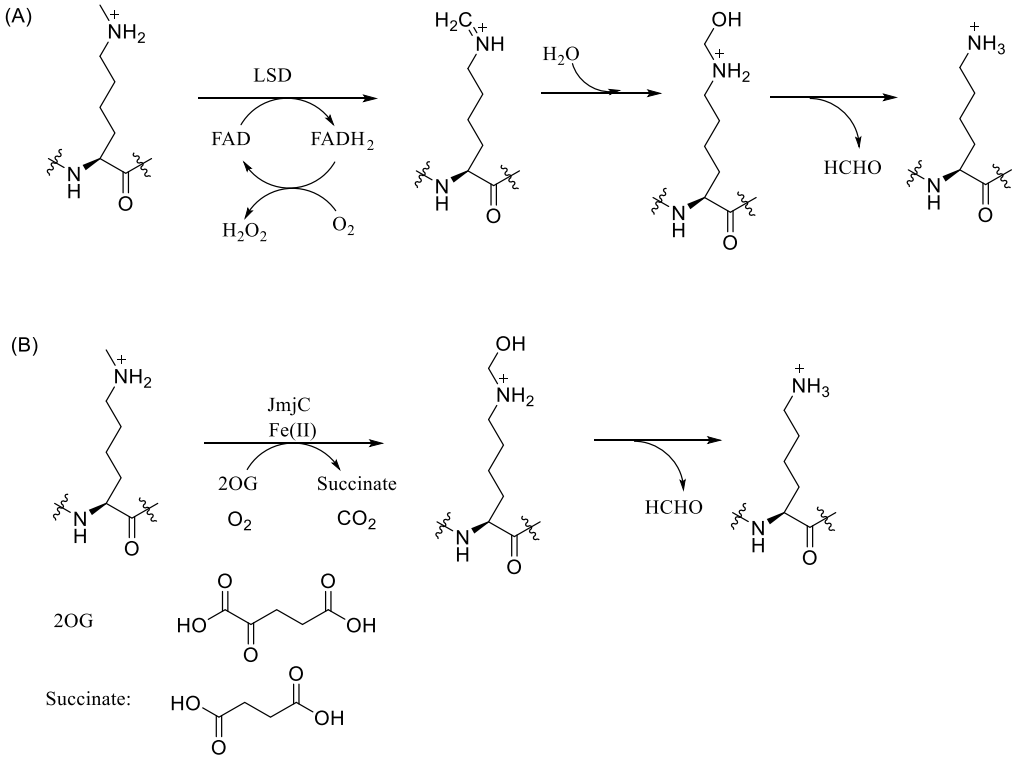


Figure 1.8 Two types of histone lysine demethylases.

(A) FAD dependent histone demethylase[63]; (B) Fe(II) and α-ketoglutaric acid dependent histone demethylase[63].

Like protein methyltransferases, protein demethylases are enzymes and can be identified directly through activity-based assay. For example, Tsukada et al. developed an *in vitro* activity assay and successfully revealed JmjC domain-containing histone demethylase 1 (JHDM1) as a new histone demethylase[64]. Methyl-histone was labeled with ^{3}H through histone methyltransferase-catalyzed methylation using ^{3}H labeled SAM. Formaldehyde, the product of demethylation reaction, was converted into 3,5-diacethyl-1,4-dihydrolutidine for easier extraction and counting[64]. Ammonium iron(II) sulfate and α-ketoglutarate were supplemented in the demethylation buffer since they served as cofactors for previously identified DNA demethylase AlkB[64-65]. JHDM1 was identified as a demethylase by examining the demethylation activity from various HeLa cell lysate fractions[64].

Because protein demethylases are special methylation readers, the profiling methods for protein methylation readers can be applied to identify methylation erasers. Until now, none of the protein N-terminal demethylase was reported. Given that protein N-terminal methylation plays important functions in regulating mitosis and DNA damage repair, it is worthwhile to examine the existence of such erasers.

1.5 Purpose of this study

Protein N-terminal methylation was discovered four decades ago, but little was known about its function. The study of RCC1 N-terminal methylation and the identification of NTMT1 as the corresponding enzyme uncovered its important function in regulating of DNA-protein interaction[5, 29]. Later, the identification of DDB2 as a new substrate for NTMT1 reveals a new

function of N-terminal methylation in regulating DNA damage repair[14]. NTMT1 was up- and down-regulated in colon and breast cancers, respectively[26, 28]. Dysregulation of NTMT1 was associated with developmental diseases and promotion of mammary oncogenesis[26, 28]. So far, the discovered N-terminal methylated proteins share a conserved N-terminal motif consisting of Met-Xaa-Pro-Lys/Arg. Peptide-array based methylation assay indicates that protein with an expanded motif may also serve as NTMT1 substrates[19]. Here, we hypothesize that methylation of various NTMT1 substrates plays important roles in different types of cancers. However, only several proteins including RCC1[5], Retinoblastoma 1[5], protein SET[5], DDB2[14], PARP3[15], CENP-A[16], CENP-B and MYL9 are confirmed as NTMT1 substrates. It would be important and necessary to develop a method to uncover more of NTMT1 substrates. In this case, we developed an activity-based approach to identify NTMT1 substrates from human cell lysates. We believe that some interesting and important substrates would be identified by this approach and it would provide important information for elucidating the specific pathogenic mechanisms in NTMT1-associated tumorigenesis.

1.6 References

1. Rice, J. C.; Allis, C. D. Histone methylation versus histone acetylation: new insights into epigenetic regulation. *Curr. Opin. Cell Biol.* **2001,** *13* (3), 263-273.

2. Allis, C. D.; Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **2016,** *17* (8), 487-500.

3. Wesche, J.; Kühn, S.; Kessler, B. M.; Salton, M.; Wolf, A. Protein arginine methylation: a prominent modification and its demethylation. *Cell. Mol. Life Sci.* **2017,** *74* (18), 3305-3315.

4. Luo, M. Chemical and Biochemical Perspectives of Protein Lysine Methylation. *Chem. Rev.* **2018,** *118* (14), 6656-6705.

5. Tooley, C. E. S.; Petkowski, J. J.; Muratore-Schroeder, T. L.; Balsbaugh, J. L.; Shabanowitz, J.; Sabat, M.; Minor, W.; Hunt, D. F.; Macara, I. G. NRMT is an α-N-methyltransferase that methylates RCC1 and retinoblastoma protein. *Nature* **2010,** *466* (7310), 1125-1128.

6. Webb, K. J.; Lipson, R. S.; Al-Hadid, Q.; Whitelegge, J. P.; Clarke, S. G. Identification of protein N-terminal methyltransferases in yeast and humans. *Biochemistry* **2010,** *49* (25), 5225-5235.

7. Brosius, J.; Chen, R. The primary structure of protein L16 located at the peptidyltransferase center of Escherichia coli ribosomes. *FEBS Lett.* **1976,** *68* (1), 105-109.

8. Wittmann-Liebold, B.; Pannenbecker, R. Primary structure of protein L33 from the large subunit of the Escherichia coli ribosome. *FEBS Lett.* **1976,** *68* (1), 115-118.

9. Petkowski, J. J.; Bonsignore, L. A.; Tooley, J. G.; Wilkey, D. W.; Merchant, M. L.; Macara, I. G.; Tooley, C. E. S. NRMT2 is an N-terminal monomethylase that primes for its homologue NRMT1. *Biochem. J.* **2013,** *456* (3), 453-462.

10. Jakobsson, M. E.; Małecki, J. M.; Halabelian, L.; Nilges, B. S.; Pinto, R.; Kudithipudi, S.; Munk, S.; Davydova, E.; Zuhairi, F. R.; Arrowsmith, C. H. The dual methyltransferase METTL13 targets N terminus and Lys55 of eEF1A and modulates codon-specific translation rates. *Nat. Commun.* **2018,** *9* (1), 3411-3425.

11. Patwardhan, P.; Miller, W. T. Processive phosphorylation: mechanism and biological importance. *Cell. Signal.* **2007,** *19* (11), 2218-2226.

12. Richardson, S. L.; Mao, Y.; Zhang, G.; Hanjra, P.; Peterson, D. L.; Huang, R. Kinetic mechanism of protein N-terminal methyltransferase 1. *J. Biol. Chem.* **2015,** *290* (18), 11601-11610.

13. Dong, C.; Dong, G.; Li, L.; Zhu, L.; Tempel, W.; Liu, Y.; Huang, R.; Min, J. An asparagine/glycine switch governs product specificity of human N-terminal methyltransferase NTMT2. *Commun. Biol.* **2018,** *1* (1), 183-191.

14. Cai, Q.; Fu, L.; Wang, Z.; Gan, N.; Dai, X.; Wang, Y. α-N-methylation of damaged DNA-binding protein 2 (DDB2) and its function in nucleotide excision repair. *J. Biol. Chem.* **2014**, *289* (23), 16046-16056.

15. Dai, X.; Rulten, S. L.; You, C.; Caldecott, K. W.; Wang, Y. Identification and functional characterizations of N-terminal α-N-methylation and phosphorylation of serine 461 in human poly (ADP-ribose) polymerase 3. *J. Proteome Res.* **2015,** *14* (6), 2575-2582.

16. Sathyan, K. M.; Fachinetti, D.; Foltz, D. R. α-amino trimethylation of CENP-A by NRMT is required for full recruitment of the centromere. *Nat. Commun.* **2017,** *8*, 14678-14692.

17. Dai, X.; Otake, K.; You, C.; Cai, Q.; Wang, Z.; Masumoto, H.; Wang, Y. Identification of novel α-N-methylation of CENP-B that regulates its binding to the centromeric DNA. *J. Proteome Res.* **2013,** *12* (9), 4167-4175.

18. Nevitt, C.; Tooley, J. G.; Tooley, C. E. S. N-terminal acetylation and methylation differentially affect the function of MYL9. *Biochem. J.* **2018,** *475* (20), 3201-3219.

19. Petkowski, J. J.; Schaner Tooley, C. E.; Anderson, L. C.; Shumilin, I. A.; Balsbaugh, J. L.; Shabanowitz, J.; Hunt, D. F.; Minor, W.; Macara, I. G. Substrate specificity of mammalian N-terminal α-amino methyltransferase NRMT. *Biochemistry* **2012,** *51* (30), 5942-5950.

20. Hamey, J. J.; Winter, D. L.; Yagoub, D.; Overall, C. M.; Hart-Smith, G.; Wilkins, M. R. Novel N-terminal and lysine methyltransferases that target translation elongation factor 1A in yeast and human. *Mol. Cell. Proteomics* **2016,** *15* (1), 164-176.

21. Villar-Garea, A.; Forne, I.; Vetter, I.; Kremmer, E.; Thomae, A.; Imhof, A. Developmental regulation of N-terminal H2B methylation in Drosophila melanogaster. *Nucleic Acids Res.* **2011,** *40* (4), 1536-1549.

22. Wu, R.; Yue, Y.; Zheng, X.; Li, H. Molecular basis for histone N-terminal methylation by NRMT1. *Genes Dev.* **2015,** *29* (22), 2337-2342.

23. Dong, C.; Mao, Y.; Tempel, W.; Qin, S.; Li, L.; Loppnau, P.; Huang, R.; Min, J. Structural basis for substrate recognition by the human N-terminal methyltransferase 1. *Genes Dev.* **2015,** *29* (22), 2343-2348.

24. https://www.proteinatlas.org/ENSG00000010165-METTL13/cell.

25. Tooley, J. G.; Schaner Tooley, C. E. New roles for old modifications: Emerging roles of N‐terminal post‐translational modifications in development and disease. *Protein Sci.* **2014,** *23* (12), 1641-1649.

26. Bonsignore, L. A.; Butler, J. S.; Klinge, C. M.; Tooley, C. E. S. Loss of the N-terminal methyltransferase NRMT1 increases sensitivity to DNA damage and promotes mammary oncogenesis. *Oncotarget* **2015,** *6* (14), 12248-12263.

27. Shields, K. M.; Tooley, J. G.; Petkowski, J. J.; Wilkey, D. W.; Garbett, N. C.; Merchant, M. L.; Cheng, A.; Schaner Tooley, C. E. Select human cancer mutants of NRMT1 alter its catalytic activity and decrease N‐terminal trimethylation. *Protein Sci.* **2017,** *26* (8), 1639-1652.

28. Bonsignore, L. A.; Tooley, J. G.; Van Hoose, P. M.; Wang, E.; Cheng, A.; Cole, M. P.; Tooley, C. E. S. NRMT1 knockout mice exhibit phenotypes associated with impaired DNA repair and premature aging. *Mech. Ageing Dev.* **2015,** *146*, 42-52.

29. Chen, T.; Muratore, T. L.; Schaner-Tooley, C. E.; Shabanowitz, J.; Hunt, D. F.; Macara, I. G. N-terminal α-methylation of RCC1 is necessary for stable chromatin association and normal mitosis. *Nat. Cell Biol.* **2007,** *9* (5), 596-603.

30. Ibáñez, G.; McBean, J. L.; Astudillo, Y. M.; Luo, M. An enzyme-coupled ultrasensitive luminescence assay for protein methyltransferases. *Anal. Biochem.* **2010,** *401* (2), 203-210.

31. Kimura, S.; Miyauchi, K.; Ikeuchi, Y.; Thiaville, P. C.; Crécy-Lagard, V. d.; Suzuki, T. Discovery of the β-barrel–type RNA methyltransferase responsible for N 6-methylation of N 6-threonylcarbamoyladenosine in tRNAs. *Nucleic Acids Res.* **2014,** *42* (14), 9350-9365.

32. Jackman, J. E.; Montange, R. K.; Malik, H. S.; Phizicky, E. M. Identification of the yeast gene encoding the tRNA m1G methyltransferase responsible for modification at position 9. *Rna* **2003,** *9* (5), 574-585.

33. Martzen, M. R.; McCraith, S. M.; Spinelli, S. L.; Torres, F. M.; Fields, S.; Grayhack, E. J.; Phizicky, E. M. A biochemical genomics approach for identifying genes by the activity of their products. *Science* **1999,** *286* (5442), 1153-1155.

34. Green, N. M. Avidin. *Adv. Protein Chem.***1975,** 29, 85-133.

35. Katz, J. E.; Dlakić, M.; Clarke, S. Automated identification of putative methyltransferases from genomic open reading frames. *Mol. Cell. Proteomics* **2003,** *2* (8), 525-540.

36. Boisvert, F.-M.; Côté, J.; Boulanger, M.-C.; Richard, S. A proteomic analysis of arginine-methylated protein complexes. *Mol. Cell. Proteomics* **2003,** *2* (12), 1319-1330.

37. Frank, R. The SPOT-synthesis technique: synthetic peptide arrays on membrane supports—principles and applications. *J. Immunol. Methods* **2002,** *267* (1), 13-26.

38. Frank, R.; Döring, R. Simultaneous multiple peptide synthesis under continuous flow conditions on cellulose paper discs as segmental solid supports. *Tetrahedron* **1988,** *44* (19), 6031-6040.

39. Najbauer, J.; Johnson, B. A.; Young, A. L.; Aswad, D. Peptides with sequences similar to glycine, arginine-rich motifs in proteins interacting with RNA are efficiently recognized by methyltransferase (s) modifying arginine in numerous proteins. *J. Biol. Chem.* **1993,** *268* (14), 10501-10509.

40. Wooderchak, W. L.; Zang, T.; Zhou, Z. S.; Acuna, M.; Tahara, S. M.; Hevel, J. M. Substrate profiling of PRMT1 reveals amino acid sequences that extend beyond the "RGG" paradigm. *Biochemistry* **2008,** *47* (36), 9456-9466.

41. Dhayalan, A.; Kudithipudi, S.; Rathert, P.; Jeltsch, A. Specificity analysis-based identification of new methylation targets of the SET7/9 protein lysine methyltransferase. *Chem. Biol.* **2011,** *18* (1), 111-120.

42. Pljevaljcic, G.; Pignot, M.; Weinhold, E. Design of a new fluorescent cofactor for DNA methyltransferases and sequence-specific labeling of DNA. *J. Am. Chem. Soc.* **2003,** *125* (12), 3486-3492.

43. Klimašauskas, S.; Weinhold, E. A new tool for biotechnology: AdoMet-dependent methyltransferases. *Trends Biotechnol.* **2007,** *25* (3), 99-104.

44. Guo, H.; Wang, R.; Zheng, W.; Chen, Y.; Blum, G.; Deng, H.; Luo, M. Profiling substrates of protein arginine N-methyltransferase 3 with S-adenosyl-L-methionine analogues. *ACS Chem. Biol.* **2013,** *9* (2), 476-484.

45. Wang, R.; Zheng, W.; Yu, H.; Deng, H.; Luo, M. Labeling substrates of protein arginine methyltransferase with engineered enzymes and matched S-adenosyl-L-methionine analogues. *J. Am. Chem. Soc.* **2011,** *133* (20), 7648-7651.

46. Islam, K.; Chen, Y.; Wu, H.; Bothwell, I. R.; Blum, G. J.; Zeng, H.; Dong, A.; Zheng, W.; Min, J.; Deng, H. Defining efficient enzyme–cofactor pairs for bioorthogonal profiling of protein methylation. *P. Natl. Acad. Sci. USA* **2013,** *110* (42), 16778-16783.

47. Blum, G.; Islam, K.; Luo, M. Using azido analogue of S-Adenosyl-L-methionine for Bioorthogonal Profiling of Protein Methylation (BPPM). *Curr. Protoc. Chem. Biol.* **2013,** *5*, 45-66.

48. Dalhoff, C.; Lukinavičius, G.; Klimašauskas, S.; Weinhold, E. Synthesis of S-adenosyl-L-methionine analogs and their use for sequence-specific transalkylation of DNA by methyltransferases. *Nat. Protoc.* **2006,** *1* (4), 1879-1886.

49. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A stepwise huisgen cycloaddition process: copper (I)‐catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem., Int. Ed.* **2002,** *41* (14), 2596-2599.

50. Tornøe, C. W.; Christensen, C.; Meldal, M. Peptidotriazoles on solid phase:[1, 2, 3]-triazoles by regiospecific copper (I)-catalyzed 1, 3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **2002,** *67* (9), 3057-3064.

51. Harris, C. J.; Scheibe, M.; Wongpalee, S. P.; Liu, W.; Cornett, E. M.; Vaughan, R. M.; Li, X.; Chen, W.; Xue, Y.; Zhong, Z. A DNA methylation reader complex that enhances gene transcription. *Science* **2018,** *362* (6419), 1182-1186.

52. Wysocka, J.; Swigut, T.; Xiao, H.; Milne, T. A.; Kwon, S. Y.; Landry, J.; Kauer, M.; Tackett, A. J.; Chait, B. T.; Badenhorst, P. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **2006,** *442* (7098), 86-90.

53. Wu, M.; Wei, W.; Chen, J.; Cong, R.; Shi, T.; Li, J.; Wong, J.; Du, J. X. Acidic domains differentially read histone H3 lysine 4 methylation status and are widely present in chromatin-associated proteins. *Sci. China Life Sci.* **2017,** *60* (2), 138-151.

54. Yang, T.; Liu, Z.; Li, X. D. Developing diazirine-based chemical probes to identify histone modification 'readers' and 'erasers'. *Chem. Sci.* **2015,** *6* (2), 1011-1017.

55. Murale, D. P.; Hong, S. C.; Haque, M. M.; Lee, J.-S. Photo-affinity labeling (PAL) in chemical proteomics: a handy tool to investigate protein-protein interactions (PPIs). *Proteome Sci.* **2016,** *15* (1), 14-47.

56. Zhai, G.; Dong, H.; Guo, Z.; Feng, W.; Jin, J.; Zhang, T.; Chen, C.; Chen, P.; Tian, S.; Bai, X. An Efficient Approach for Selective Enrichment of Histone Modification Readers Using Self-Assembled Multivalent Photoaffinity Peptide Probes. *Anal. Chem.* **2018,** *90* (19), 11385-11392.

57. Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. *P. Natl. Acad. Sci. USA* **2002,** *99* (17), 11020-11024.

58. Rawłuszko-Wieczorek, A. A.; Knodel, F.; Tamas, R.; Dhayalan, A.; Jeltsch, A. Identification of protein lysine methylation readers with a yeast three-hybrid approach. *Epigenet. Chromatin* **2018,** *11* (1), 4-15.

59. Fields, S.; Song, O. A novel genetic system to detect protein–protein interactions. *Nature* **1989,** *340* (6230), 245-246.

60. Frankel, A. D.; Kim, P. S. Modular structure of transcription factors: implications for gene regulation. *Cell* **1991,** *65* (5), 717-719.

61. Liu, F.; Green, M. R. A specific member of the ATF transcription factor family can mediate transcription activation by the adenovirus E1a protein. *Cell* **1990,** *61* (7), 1217-1224.

62. Brückner, A.; Polge, C.; Lentze, N.; Auerbach, D.; Schlattner, U. Yeast two-hybrid, a powerful tool for systems biology. *Int. J. Mol. Sci.* **2009,** *10* (6), 2763-2788.

63. Shi, Y. G.; Tsukada, Y. The discovery of histone demethylases. *CSH Perspect Biol.* **2013,** *5* (9), a017947.

64. Tsukada, Y.-i.; Fang, J.; Erdjument-Bromage, H.; Warren, M. E.; Borchers, C. H.; Tempst, P.; Zhang, Y. Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **2006,** *439* (7078), 811-816.

65. Falnes, P. Ø.; Johansen, R. F.; Seeberg, E. AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. *Nature* **2002,** *419* (6903), 178-182.

Chapter 2 - Activity-based substrate profiling of NTMT1

2.1 Introduction

Methylation is a common modification for small bioactive molecules and biomacromolecules among all kinds of organisms. In woody poplar plants, methylation of caffeoyl coenzyme A by caffeoyl coenzyme A O-methyltransferase (CCoAOMT) is essential in lignin biosynthesis[1]. In human body, many essential endogenous chemicals including trace amines (e.g., 3-methoxytyramine, N-methyltyramine, N-methylphenethylamine)[2] and hormones (e.g., adrenaline) are methylated by certain methyltransferases. Methylation of biomacromolecules (e.g., DNA, histone) by relevant methyltransferases (e.g., DNA methyltransferases, histone lysine methyltransferases, histone arginine methyltransferases) have been deeply studied due to their important functions in epigenetics. For example, trimethylation of histone 3 lysine 4 is associated with gene activation, while dimethylation of histone 3 lysine 9 is related to the silence of gene expression[3]. In last decade, three human protein N-terminal methyltransferases (NTMT1, NTMT2, and METTL13) were discovered, which greatly promoted the study of protein N-terminal methylation.

Protein N-terminal methylation has been demonstrated to regulate DNA-protein interaction and DNA damage repair[4-5]. N-terminal trimethylation of RCC1 and CENP-B enhances their binding with chromosome DNA[4, 6]. N-terminal methylation of DDB2 promotes its nucleus localization[5]. Dysregulation of NTMT1 has been implicated in severe diseases such as cancer. NTMT1 functions as a tumor suppressor in breast cancer but additional study indicates that it may act as an oncogene in colorectal cancer[7-8].

So far, the identified NTMT1 substrates including RCC1, DDB2, PARP3, CENP-A, CENP-B, SET and retinoblastoma 1, contain an N-terminal Met-Xaa-Pro-Lys/Arg motif[5-6, 9-11]. Peptide-

based methylation assay indicated that a protein with an expanded motif (Met-Xaa-Xaa-Lys/Arg) may also serve as a substrate for NTMT1[12]. However, peptides only contain very short amino acid sequences and usually do not have a tertiary structure, and therefore, methylation assay on peptide and protein level may not be consistent to each other. Considering the important roles of N-terminal methylation in cellular processes and human diseases, it would be necessary to develop a substrate profiling method for NTMT1.

Activity-based substrate profiling assay has been successfully applied to the study of DNA and protein methylation. It is a powerful tool to label endogenous proteins *via* enzyme catalyzed reactions in physiological condition. Specifically, several sulfonium-alkyl SAM analogs have been reported for substrate profiling of DNA and protein methyltransferases[13-15]. The sulfonium-alkyl SAMs contain either an alkynyl or an azide group to incorporate an affinity tag (e.g., biotin) to the substrates utilizing the efficient click chemistry[13-15]. In most situations, the wild type protein methyltransferases are not able to transfer the bulky alkyl groups to their targets and thus the time-consuming protein engineering is required[13-15]. Here, we found that one of the reported SAM analogs[13-15], Hey-SAM, could be taken by wt NTMT1 as cofactor. Moreover, NTMT1 KO cell line was generated to produce hypomethylated NTMT1 substrates, which avoided the use of Adox, an inhibitor for SAM dependent methyltransferases, and overcame the drawbacks of adding Adox to cell culture.

2.2 Results and discussion

2.2.1 Plasmid construction and purification of NTMT1

To express protein NTMT1 in *E. coli*, plasmid pET28-MHL-NTMT1 containing an N-terminal His_6-tag with a TEV digestion site was constructed (Fig. 2.1A & Fig.2.1B). Protein His_6-NTMT1 (Fig. 2.1C, 27.7kD) was overexpressed in *E. coli* and purified by Ni-NTA affinity chromatography.

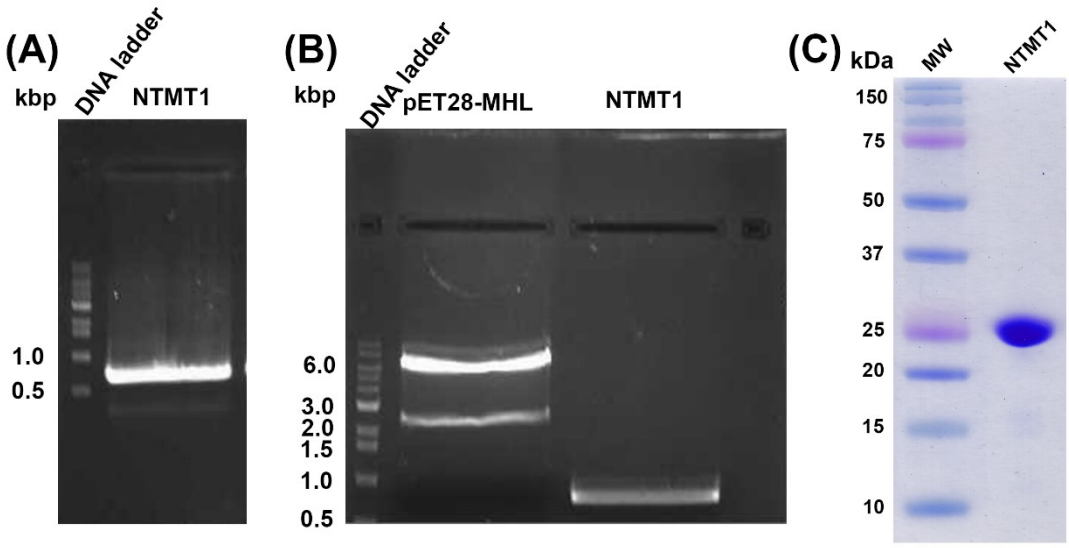


Figure 2.1 Plasmid construction and purification of NTMT1.

(A) PCR products of NTMT1; (B) Digestion of pET28-MHL and PCR products with NdeI and HindIII; (C) SDS-PAGE of purified NTMT1.

2.2.2 Synthesis of Hey-SAM as a SAM surrogate

To perform the activity-based substrate profiling, the natural cofactor SAM was modified to contain a terminal alkynyl for introducing a reporter group by click chemistry. Many SAM analogs have been reported for profiling substrates of PKMTs and PRMTs[13-15]. Most of the surrogates cannot be directly utilized by wt PRMTs or PKMTs due to their bulky sizes but can serve as cofactors for certain mutants[13-15]. NTMT1 crystal structures (Fig. 2.2A) indicate that SAH is deeply buried in a big pocket which may be able to accommodate a bulky SAM surrogate. Among all the reported SAM analogs, *(E)*-hex-2-en-5-ynyl-SAM (Hey-SAM, Fig. 2.2.B), is especially useful, which features a sulfonium-*β-sp2* carbon in addition to a terminal alkyne in its structure[15]. An S_N2 reaction mechanism has been proposed as a common mechanism for SAM dependent methylation[16]. For sulfonium-alkyl SAM analog, the substitution of one hydrogen from the

reactive methyl with an alkyl group destabilizes the transition state and slows down the reaction, which can be rescued by placing a double or triple bond to the adjacent carbon[17]. The driving force for the accelerated rate is the electronic conjugation between the π-orbital and the newly formed p-orbital which stabilizes the transition state[17].

Hey-SAM was synthesized by Dr. Gaochao Huang in our lab using a modified synthetic route, which will be published somewhere else. Through the new synthetic route (Fig. 2.2B), SAH was converted into Hey-SAM almost quantitatively. Hey-SAM was characterized and confirmed by ^{1}H NMR (Appendix A), COSY (Appendix A), ^{13}C NMR (Appendix A), HPLC (Fig. 2.2C) and LC-MS (Fig. 2.2D).

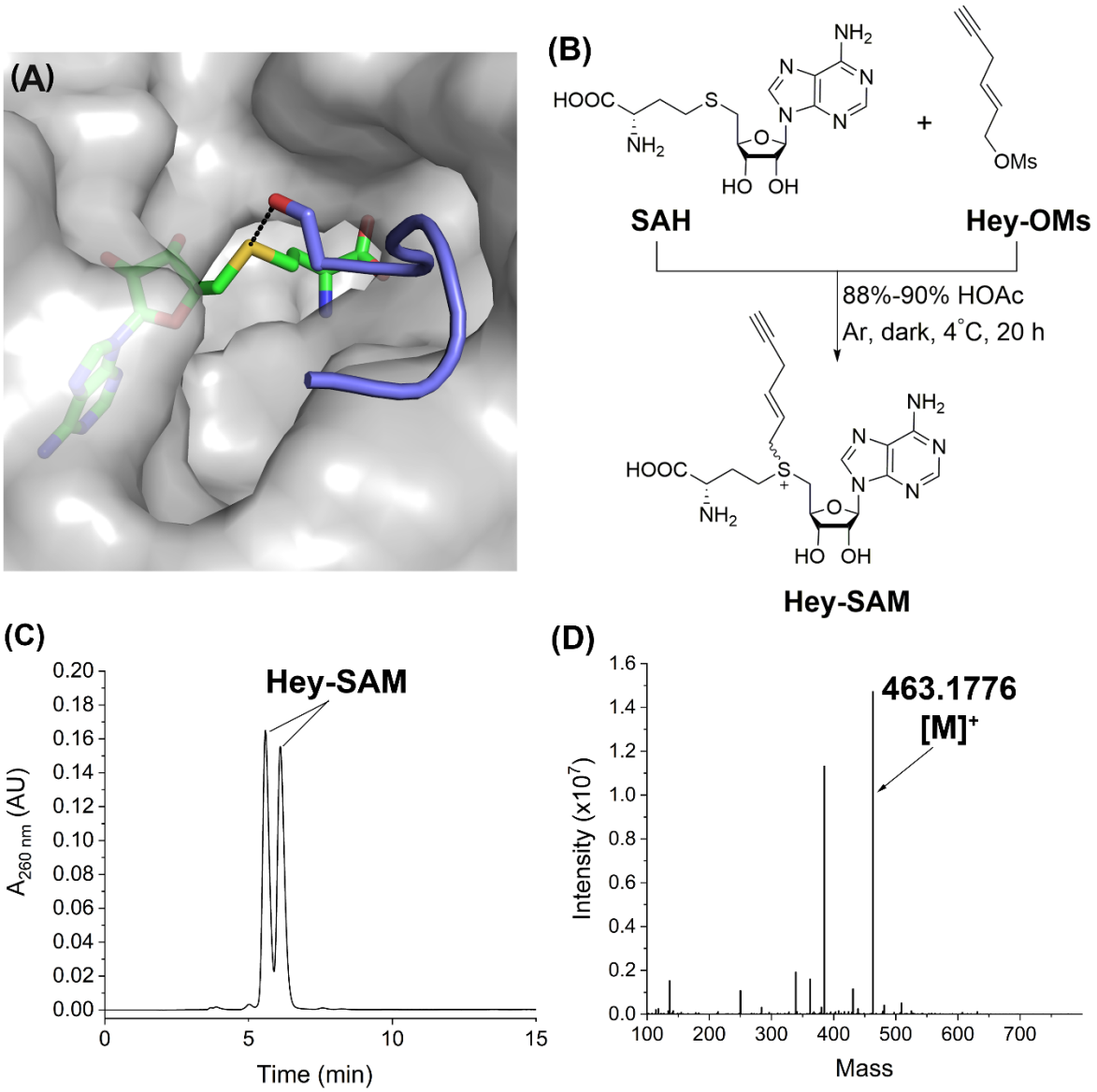


Figure 2.2 Synthesis of Hey-SAM as a SAM surrogate.

(A) Surface model of substrate binding with NTMT1 (PDB: 5E1B). The sulfur atom of SAH is shown in yellow. Peptide SPKRIA is shown in blue. Protein surface is represented in grey. (B) Scheme for Hey-SAM synthesis. (C) HPLC analysis of Hey-SAM (elution condition: 0.01% TFA in water). (D) LC-MS analysis of Hey-SAM (calculated $[M]^+$: 463.1758; detected: 463.1776).

2.2.3 Alkylation of RCC1 peptide with Hey-SAM and wt NTMT1

To investigate whether wt NTMT1 would utilize Hey-SAM as its cofactor, a peptide (N-terminal analog of RCC1, Fig. 2.3A) containing the first ten N-terminal residues (excluding the initial methionine) of RCC1 was synthesized and used as the substrate. As a control, RCC1 peptide was mixed with the natural cofactor SAM with or without wt NTMT1, and mass spectrometry analysis showed that RCC1 peptide was di- and trimethylated only in the presence of wt NTMT1 (Fig. 2.3B & Fig. 2.3C). For the trimethylated product, the loss of a formaldehyde (~ 30 Dalton mass shift) was detected which was also observed in other reported peptides[18-19]. The alkylation reaction between RCC1 peptide and Hey-SAM catalyzed by wt NTMT1 was confirmed by mass spectrometry (Fig. 2.3E). In contrast to SAM, with which di- and trimethylated products were observed, only monoalkylated peptide was formed with Hey-SAM and this may be due to the bulky size of the Hey group which prevents multiple alkylation. The monoalkylation was confirmed to be enzyme-catalyzed as no alkylated products were detected in the absence of wt NTMT1 (Fig. 2.3D).

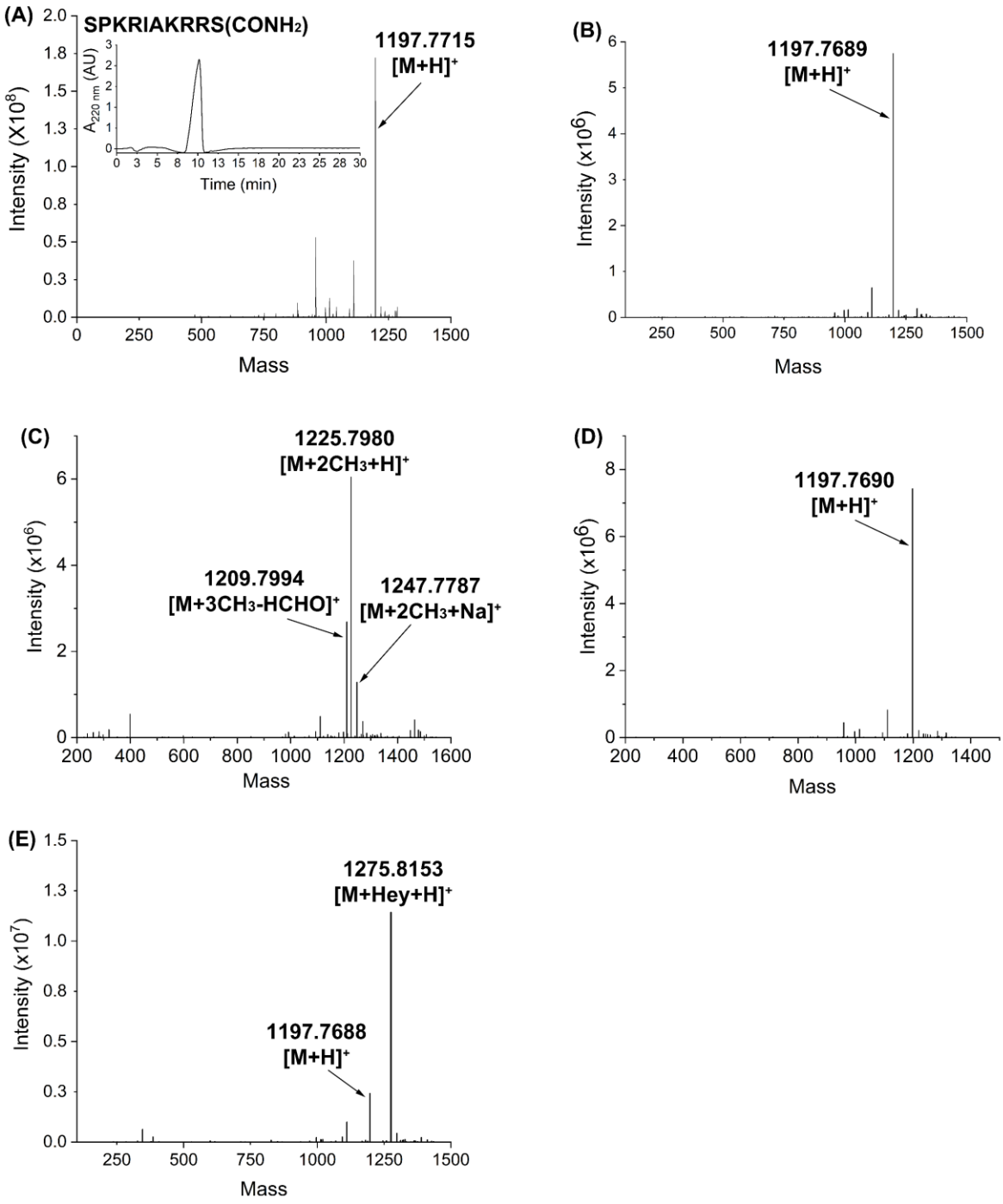

(A)
SPKRIAKRRS(CONH2)
1197.7715
[M+H]+
Intensity (X10^8)
Mass
A220 nm (AU)
Time (min)
(B)
1197.7689
[M+H]+
Intensity (x10^6)
Mass
(C)
1225.7980
[M+2CH3+H]+
1209.7994
[M+3CH3-HCHO]+
1247.7787
[M+2CH3+Na]+
Intensity (x10^6)
Mass
(D)
1197.7690
[M+H]+
Intensity (x10^6)
Mass
(E)
1275.8153
[M+Hey+H]+
1197.7688
[M+H]+
Intensity (x10^7)
Mass

Figure 2.3 MS analysis of RCC1 peptide methylation and alkylation by wt hNTMT1 using different cofactors.

All reactions were carried out with 20 μM RCC1 peptide and 200 μM cofactors for 2h at 37 ℃. (A) Mass spectrum and HPLC analysis (inset) of the synthesized RCC1 peptide (calculated $[M + H]^+$: 1197.7656; detected: 1197.7715). (B) Mass spectrum of RCC1 peptide mixed with SAM in the absence of wt NTMT1 (calculated $[M + H]^+$: 1197.7656; detected: 1197.7689). (C) Mass spectrum of RCC1 peptide methylation in the presence of SAM and wt NTMT1 (calculated $[M + 2CH_3 + H]^+$: 1225.7969, $[M + 2CH_3 + Na]^+$: 1247.7789, $[M + 3CH_3 - HCHO]^+$: 1209.8020; detected: 1225.7980, 1247.7787, and 1209.7994). (D) Mass spectrum of RCC1 peptide mixed with Hey-SAM in the absence of wt NTMT1 (calculated $[M + H]^+$: 1197.7656; detected: 1197.7690). (E) Mass spectrum of RCC1 peptide alkylation in the presence of Hey-SAM and wt NTMT1 (calculated $[M + Hey + H]^+$: 1275.8126; detected: 1275.8153).

2.2.4 Alkylation of RCC1 protein with Hey-SAM and wt NTMT1

Because RCC1 peptide can be alkylated with Hey-SAM and wt NTMT1, next we examined whether RCC1 protein would be alkylated as well. To do this, RCC1 protein with a C-terminal His_6-tag was expressed in *E. coli* and purified by Ni-NTA affinity chromatography (Fig. 2.4B). As the initial methionine removal is a common modification in *E. coli*[4, 20-21], RCC1 purified from *E. coli* was subject to intact protein mass spectrometry to verify whether its initial methionine was removed. Here, we found three forms of RCC1 purified from *E. coli* Bl21 (DE3) strain (Fig. 2.4A): full length (RCC1), initial methionine removed (RCC1 - M), and methionine and serine removed (RCC1 - MS). The full length RCC1 and RCC1 - MS were not the correct substrates of NTMT1[4]. To enhance the yield of the correct substrate, RCC1-M, the protein was expressed in a mutant *E. coli* BL21 (DE3) strain, C41 (DE3) (Fig. 2.4C), which is often used to express toxic proteins. The majority proteins purified from C41 (DE3) were full length and methionine removed RCC1. Thus, we decided to choose C41 (DE3) strain to express the RCC1. When RCC1 was incubated with SAM and wt NTMT1, mono- and dimethylated RCC1 products were detected by intact protein mass (Fig. 2.4E). Similar to the result of RCC1 peptide, only monoalkylated RCC1 (Fig. 2.4G)

was observed after incubating with Hey-SAM and wt NTMT1. RCC1 methylation and alkylation were enzyme-catalyzed because no products were observed in the absence of wt NTMT1 (Fig. 2.4D & Fig. 2.4F). In addition, the modification site of RCC1 was confirmed to be the N-terminus using in-gel trypsin digestion and subsequent analysis with UPLC-HRMS (Fig. 2.4H). All these results demonstrated that Hey-SAM can be taken by wt NTMT1 as a cofactor.

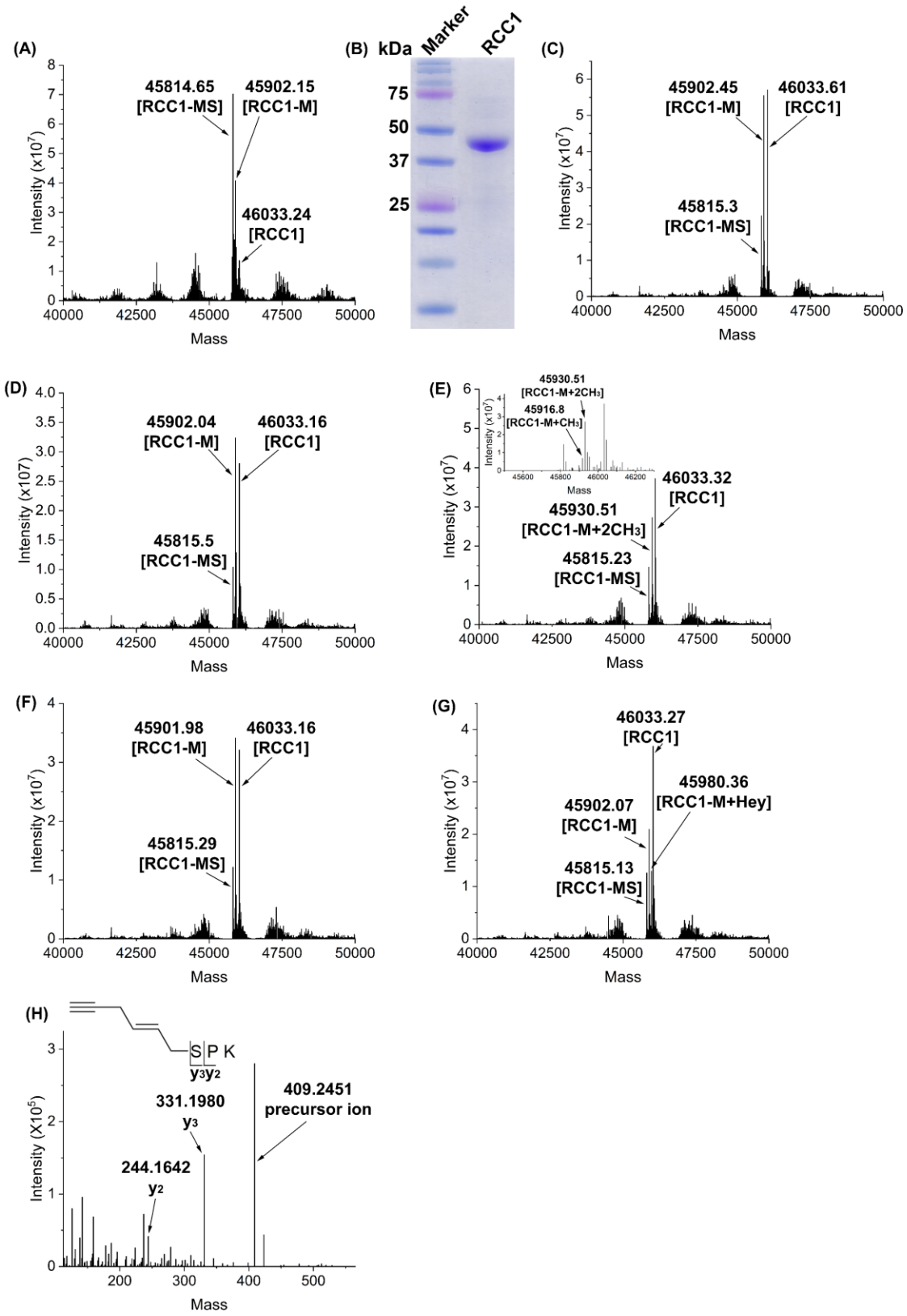

(A)
45814.65
[RCC1-MS]
45902.15
[RCC1-M]
46033.24
[RCC1]
Intensity (x10^7)
Mass
(B) kDa
Marker
RCC1
75
50
37
25
(C)
45902.45
[RCC1-M]
46033.61
[RCC1]
45815.3
[RCC1-MS]
Intensity (x10^7)
Mass
(D)
45902.04
[RCC1-M]
46033.16
[RCC1]
45815.5
[RCC1-MS]
Intensity (x107)
Mass
(E)
45930.51
[RCC1-M+2CH3]
45916.8
[RCC1-M+CH3]
46033.32
[RCC1]
45930.51
[RCC1-M+2CH3]
45815.23
[RCC1-MS]
Intensity (x10^7)
Mass
(F)
45901.98
[RCC1-M]
46033.16
[RCC1]
45815.29
[RCC1-MS]
Intensity (x10^7)
Mass
(G)
46033.27
[RCC1]
45980.36
[RCC1-M+Hey]
45902.07
[RCC1-M]
45815.13
[RCC1-MS]
Intensity (x10^7)
Mass
(H)
S P K
y3 y2
331.1980
y3
409.2451
precursor ion
244.1642
y2
Intensity (X10^5)
Mass

Figure 2.4 MS analysis of RCC1 protein methylation and alkylation by wt hNTMT1 using different cofactors.

All reactions were carried out with 20 μM RCC1 protein and 200 μM cofactors at 37℃. The molecular weight (MW) of full length RCC1-His_6 is 46034.14 D; the MW of methionine removed RCC1-His_6 (RCC1 - M) is 45902.95 D; the MW of methionine and serine removed RCC1-His_6 (RCC1 - MS) is 45815.87 D. (A) Intact protein mass of RCC1 purified from *E. coli* BL21(DE3) strain. (B) SDS-PAGE of RCC1 protein. (C) Intact protein mass of RCC1 purified from *E. coli* C41(DE3) strain. (D) Intact protein mass of RCC1 mixed with SAM. (E) Intact protein mass of RCC1 mixed with SAM and wt NTMT1. (F) Intact protein mass of RCC1 mixed with Hey-SAM. (G) Intact protein mass of RCC1 mixed with Hey-SAM and wt NTMT1. (H) MS^E spectrum of mapping the modification site on RCC1 protein.

2.2.5 Thermodynamic and kinetic analysis of wt NTMT1 with Hey-SAM and RCC1 peptide

To compare the binding affinities of SAM/Hey-SAM with wt hNTMT1, isothermal titration calorimetry (ITC) experiment was performed. ITC was commonly used to determine the thermodynamic parameters of protein-protein interaction and protein-small molecule interaction in solution. Petkowski et al. and Dong et al. have applied ITC to study the binding affinities of NTMT1 with peptide substrates[12, 22]. The titration curves are shown in Fig 2.5 and the thermodynamic parameters including dissociation constants (K_d), binding stoichiometries (N), enthalpy changes (ΔH) and entropy changes (ΔS) are summarized in Table 2.1. SAM and Hey-SAM displayed similar dissociation constants (Kd) and stoichiometries (N), indicating that both have similar binding modes with wt NTMT1. The stoichiometries of NTMT1 with the cofactors or peptide substrate are less than 1 (~ 0.4 - 0.6), suggesting that only half of the enzyme's active sites were occupied with its substrates. NTMT1 purified from *E. coli* will bind with some endogenous SAM/SAH[22], which may contribute to the small N value.

Table 2.1 Thermodynamic parameters of wt NTMT1 with cofactors and RCC1 peptide.

	K_d (μM)	N	ΔH (kcal/mol)	ΔS (cal/(mol·°C))
SAM	1.43 ± 0.325	0.393 ± 0.022	-17.40 ± 0.775	-31.55 ± 3.05
Hey-SAM	1.29 ± 0.165	0.382 ± 0.039	-13.71 ± 1.59	-19 ± 5.1
RCC1 peptide	1.22 ± 0.019	0.63 ± 0.072	-10.95 ± 0.27	-9.65 ± 0.855

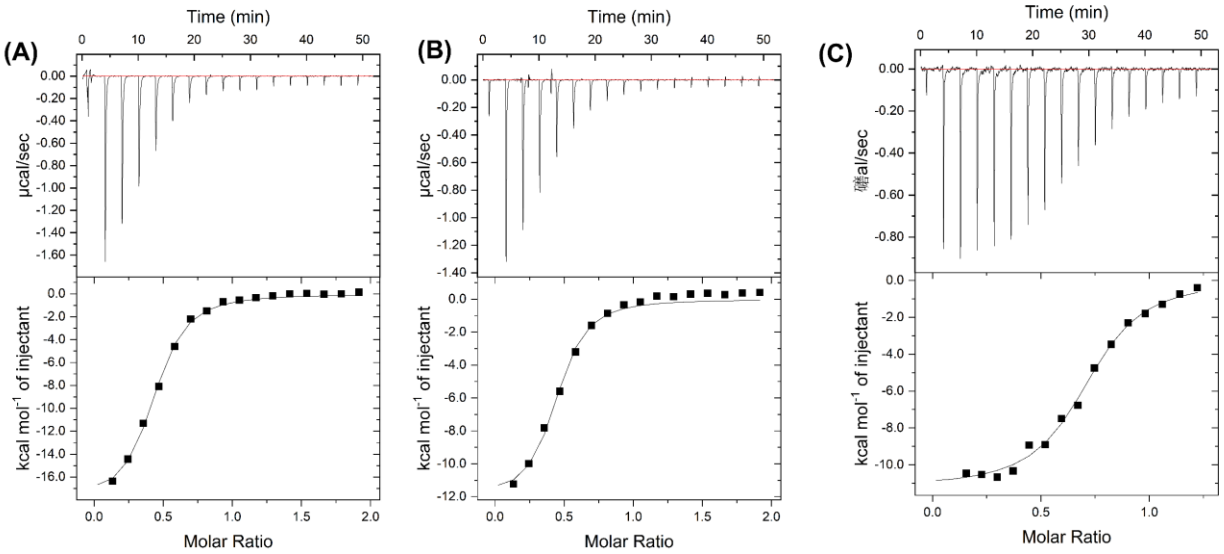


Figure 2.5 Affinity measurements of wt NTMT1 with cofactors and RCC1 peptide by ITC.

(A) Titration curve for SAM. (B) Titration curve for Hey-SAM. (C) Titration curve for RCC1 peptide.

To evaluate and compare the reaction rates of cofactor SAM and Hey-SAM, their kinetic parameters were studied with RCC1 peptide and wt hNTMT1. To measure the reaction rate, SAH release (retention time: 5.2 min, Fig 2.6A & Fig 2.6B) was monitored and quantified by HPLC with adenosine (retention time: 6.8 min, Fig 2.6A & Fig 2.6B) as an external standard. The initial reaction rates and concentrations of substrates were fitted to Michaelis-Menten curve (Fig. 2.6C & Fig. 2.6D). The turnover numbers (k_{cat}) and Michaelis constants (K_M) are shown in Table 2.2.

Table 2.2 Kinetic parameters of wt NTMT1 with RCC1 peptide and cofactors.

	K_M (μM)	k_{cat} (min^{-1})
SAM	0.38 ± 0.029	0.45 ± 0.008
Hey-SAM	1.40 ± 0.085	0.4 ± 0.006
RCC1 peptide	1.96 ± 0.236 /SAM	0.58 ± 0.022 /SAM
	2.09 ± 0.176 /Hey-SAM	0.52 ± 0.014 /Hey-SAM

The k_{cat} values of SAM and Hey-SAM were determined to be similar when NTMT1 was saturated with the RCC1 peptide. However, the K_M of Hey-SAM was ~ 4-times higher than that of SAM, indicating that NTMT1 displays a higher affinity with its native cofactor SAM than with Hey-SAM. Consequently, the catalytic efficiency (k_{cat}/K_M) of SAM (1.97×10^4 $M^{-1}s^{-1}$) was ~ 4-times higher than that of Hey-SAM (4.76×10^3 $M^{-1}s^{-1}$). The K_M and k_{cat} values of RCC1 peptide in the presence of saturated cofactors were very close, suggesting that the cofactors have marginal effects on the binding affinity and turnover of the wt NTMT1 with the peptide substrate.

Taken together, thermodynamic and kinetic studies suggest that Hey-SAM is a good SAM surrogate for wt NTMT1 and can be applied to label substrates with a terminal alkynyl which can undergo Cu(I) catalyzed click chemistry and allow for further labeling with a reporter group such as biotin.

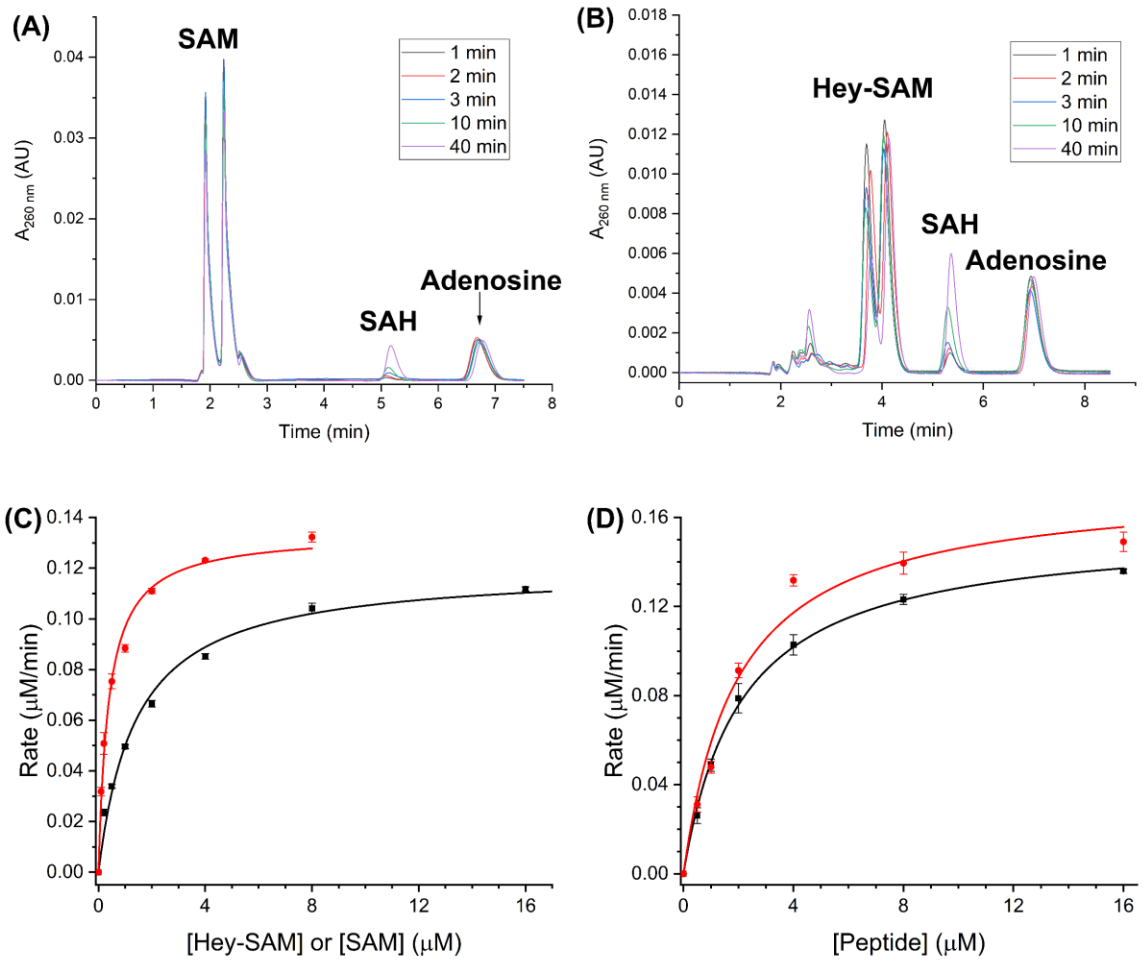


Figure 2.6 Kinetic study of wt NTMT1 with RCC1 peptide and cofactors.

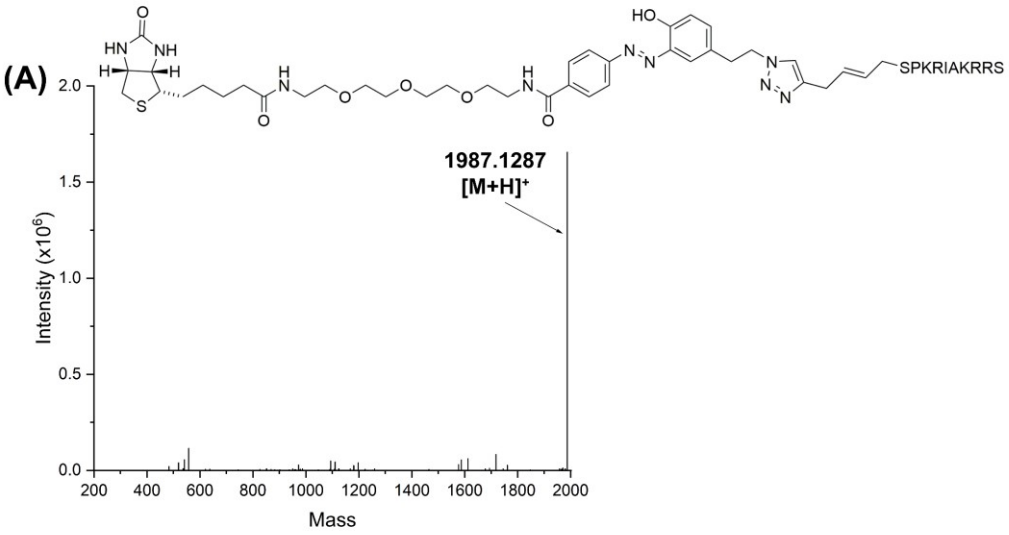
(A) Monitoring SAH release by HPLC for the reaction between 20 μM SAM and 4 μM RCC1 peptide in the presence of 0.3 μM wt NTMT1. (B) Monitoring SAH release by HPLC for the reaction between 20 μM Hey-SAM and 8 μM RCC1 peptide in the presence of 0.3 μM wt NTMT1. (C) Determination of kinetic parameters of cofactor SAM (red) and Hey-SAM (black) in the presence of RCC1 peptide and 0.3 μM wt NTMT1. (D) Determination of kinetic parameters of RCC1 peptide with SAM (red) or Hey-SAM (black) in the presence of 0.3 μM wt NTMT1.

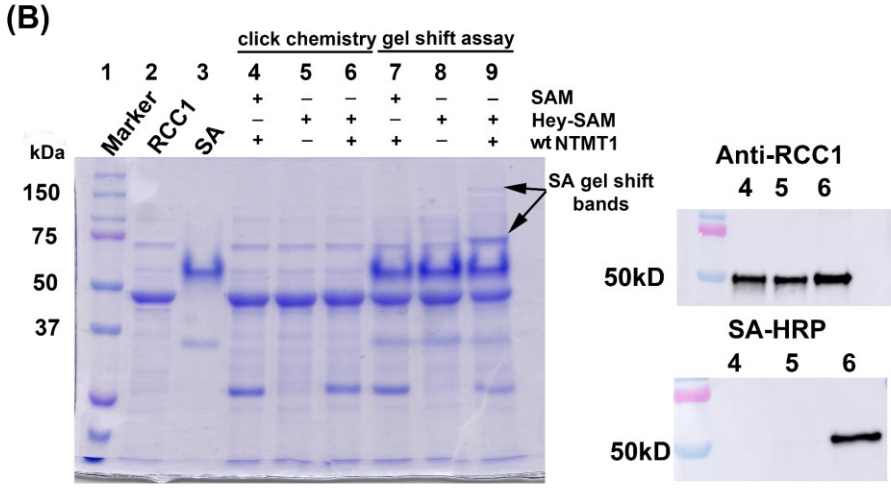
2.2.6 Labeling of RCC1 protein with biotin *via* click chemistry

To determine whether Cu(I) catalyzed click chemistry would work, RCC1 peptide was incubated with Hey-SAM and wt NTMT1 for 2h followed by addition of diazo biotin azide (DBA) in the presence of $CuSO_4$, tris(benzyltriazolylmethyl)amine (TBTA) and tris(2-carboxyethyl)phosphine hydrochloride (TCEP)[23-24]. As expected, LC-MS analysis (Fig. 2.7A) showed that RCC1 was successfully labeled with a biotin group.

To examine whether RCC1 protein would be labeled with a biotin, RCC1 protein was alkylated with Hey-SAM and wt NTMT1, reacted with DBA by Cu(I) catalyzed click chemistry and checked by western blot (WB) with streptavidin-HRP (Fig. 2.7B). Negative controls were performed either in the absence of NTMT1 or using SAM to substitute Hey-SAM (Fig. 2.7B). The attachment of biotin was only detected for the alkylated RCC1 with Hey-SAM and NTMT1 (Fig. 2.7B). Negative controls contained similar loading amount of RCC1, but no biotin signal was detected with streptavidin (SA)-HRP (Fig. 2.7B). Moreover, to confirm that biotin was attached on the N-terminus of RCC1, the SA-based gel shift assay followed by in-gel digestion and LC-MS analysis were performed (Fig. 2.7B). As previously demonstrated, only partial of recombinant RCC1 protein purified from *E. coli* was the active substrate for NTMT1 and gel shift assay was used to enrich the biotin labeled RCC1. To perform gel shift assay, properly alkylated and biotin labeled protein samples were denatured and mixed with native SA before loading to SDS-PAGE. As shown in Fig 2.7B, lane 3 represents protein SA which exists as a mixture of dimer (26.4 kDa) and tetramer (52.8 kDa). Two new bands were specifically formed in lane 9 (Fig 2.7B) and were speculated as RCC1 + 2SA (72 kDa) and 2RCC1 + 4SA (144 kDa) according to their molecular weights on SDS-PAGE. The newly formed bands were excised, digested with trypsin and analyzed

by LC-MS. The biotin labeled N-terminal SPK peptide was detected with a retention time of 24.8 min and confirmed by fragmentation analysis (Fig. 2.7C).


(A)
SPKRIAKRRS
1987.1287
[M+H]+
Intensity (x10^6)
Mass


(B)
click chemistry gel shift assay
1 2 3 4 5 6 7 8 9
Marker
RCC1
SA
SAM
Hey-SAM
wt NTMT1
kDa
150
75
50
37
SA gel shift bands
Anti-RCC1
4 5 6
50kD
SA-HRP
4 5 6
50kD

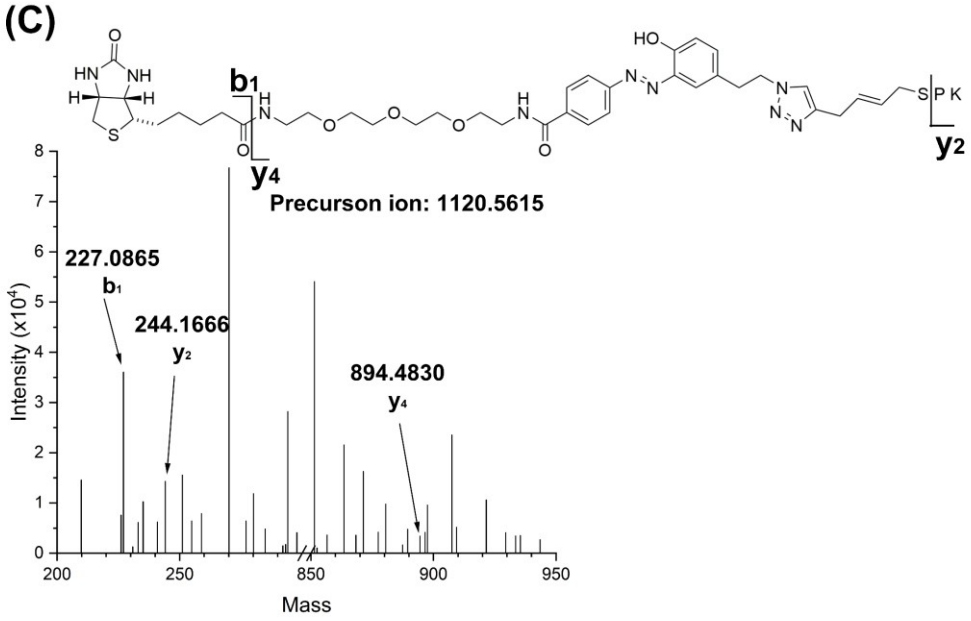

(C)
b1
y4
y2
Precurson ion: 1120.5615
227.0865
b1
244.1666
y2
894.4830
y4
Intensity (x10^4)
Mass

Figure 2.7 RCC1 peptide and protein were labeled with a biotin tag by Hey-SAM and click chemistry.

(A) Mass spectrum of biotin labelled RCC1 peptide (calculated $[M + H]^+$: 1987.1288, detected: 1987.1287). (B) Streptavidin gel shift assay. Lane 2 represents purified RCC1. lane 3 represents purified NTMT1. Lanes 4-6 represent RCC1 sequentially labeled with cofactors and DBA by click chemistry. Lanes 7-8 represent gel shift assay with non-denatured SA. Lanes 4-6 were detected by western blot with rabbit anti-hRCC1 antibody and streptavidin-HRP, respectively. (C) MS^E spectrum of mapping the biotin tag to N-terminal serine (calculated precursor ion $[M + H]^+$: 1120.5608; detected: 1120.5615).

2.2.7 Labeling overexpressed RCC1 protein with biotin by Hey-SAM and click chemistry in *E. coli* lysates

As a proof of concept, the activity-based substrate profiling method was applied to a complex mixture, the *E. coli* cell lysates. Similar to the above procedure for purified RCC1 protein, the *E. coli* lysates containing overexpressed RCC1 were employed as substrates for NTMT1, methylated or alkylated with SAM or Hey-SAM, and reacted with DBA *via* Cu(I) catalyzed click chemistry. As expected, only the properly alkylated and biotin labeled RCC1 (~ 46 kDa) was successfully enriched and purified by streptavidin beads (Lane 9 in Fig. 2.8A & Fig. 2.8B). According to the results shown in SDS-PAGE (Fig 2.8A) and western blot (Fig. 2.8B), control samples either treated with SAM or in the absence of NTMT1 failed to enrich RCC1 though the loading volumes were the same. The purified RCC1 was further quantified by WB (Fig. 2.8C), in which 12 ng, 16 ng, 24 ng and 36 ng of purified RCC1 proteins were loaded as reference. Considering that the amount of active RCC1 substrate purified from *E. coli* varied between 30% and 50% of the total protein, quantification of the purified RCC1 by western blot gave a normalized recovery yield of 8%-13%, indicating that the activity-based substrate profiling approach using Hey-SAM and click chemistry is feasible.

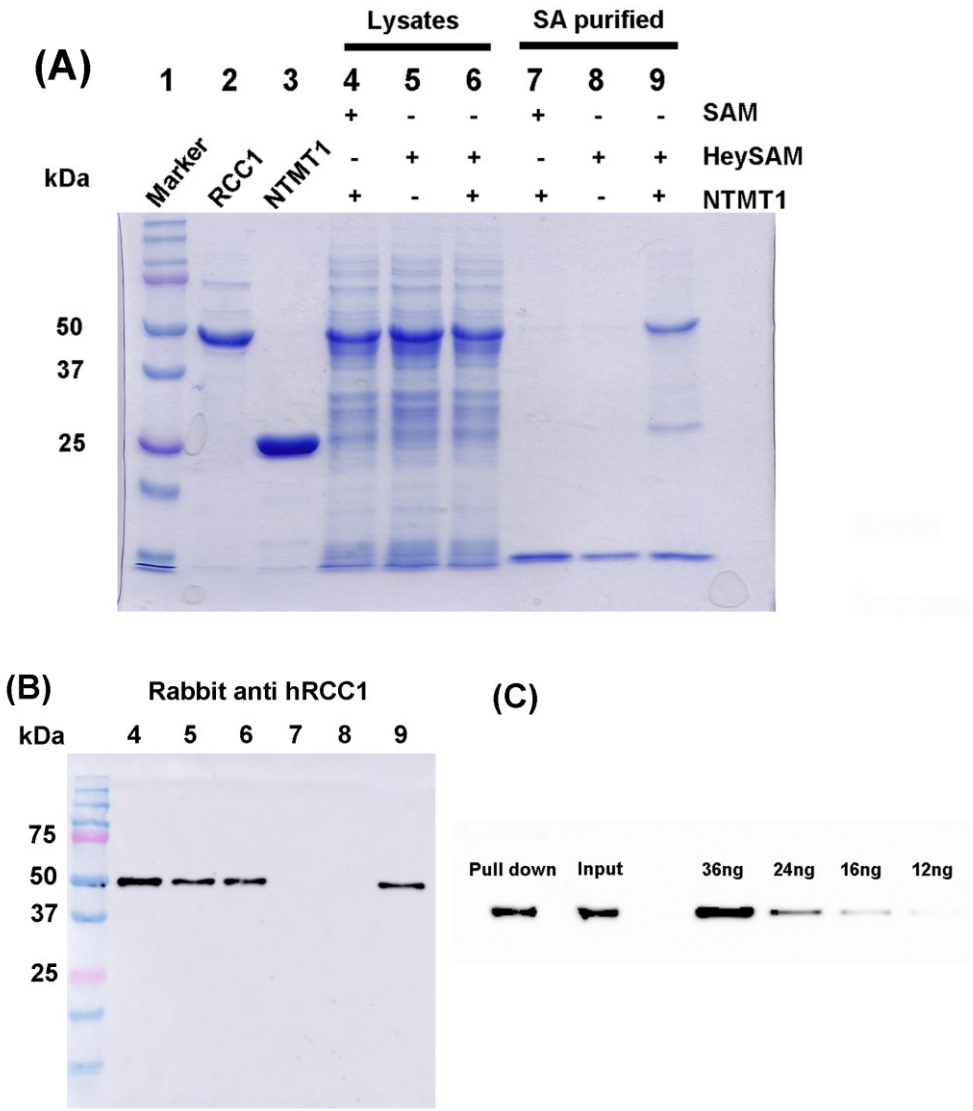


Figure 2.8 Labeling of RCC1 in *E. coli* lysates with a biotin tag *via* activity-based substrate profiling strategy.

(A) Overexpressed RCC1 protein in *E. coli* lysates was biotin labeled and enriched by streptavidin beads. (B) Proteins enriched and purified by streptavidin beads were detected by WB using rabbit anti-RCC1 antibody. (C) Quantification of purified RCC1 by WB (Input: before incubating with streptavidin beads; Pull down: proteins enriched and purified using streptavidin beads).

2.2.8 Knockout of NTMT1 in HEK293FT

Inducing of methyl-accepting proteins in a hypomethylated state is one important premise for the activity-based substrate profiling strategy. Previous protocols rely on methyltransferase inhibitors to generate hypomethylated proteins[25]. One of the most commonly used inhibitors, Adox, were broadly applied to induce hypomethylated proteins in human cell lines[14, 24-26]. Adox, a cell permeable inhibitor of SAH hydrolase, results in the accumulation of intracellular SAH level and therefore inhibits the catalytic activity of methyltransferases. Though Adox was used as a broad inhibitor for all methyltransferases, several studies demonstrated that DNA methyltransferases were less sensitive to its treatment compared to protein methyltransferases[27-28]. The typical concentrations of Adox supplemented in culture medium were 15-20 μM[29-30] because a higher concentration would cause cell apoptosis[31]. In addition, only a small fraction of protein methyltransferases in cells would be inhibited by Adox because Adox only works on the newly translated proteins after its treatment[25]. Though Adox exhibits less inhibition effects for DNA methylation, it can inhibit histone lysine methyltransferases[32] and therefore may affect the regulation of epigenetics.

Considering that Adox is a very broad but not a specific protein methylation inhibitor and many of its adverse effects, we decided to specifically induce NTMT1 substrates to a hypomethylated state by generating NTMT1 gene knockout cell line. Ran et al. reported a protocol for genome editing using CRISPR-Cas9 system[33]. The principle for CRISPR-Cas9 based genome editing is that Cas9 nuclease targets the site of genome DNA which precedes a protospacer adjacent motif (PAM) by a single-guide RNA (sgRNA) consisting of 20 nucleotides complementary with the target DNA sequence[33]. Specifically, the PAM sequence for *Streptococcus pyogenes* Cas9 (SpCas9) is NGG[33]. The cleaved double strands DNA by cas9 can be repaired *via* "nonhomologous

end joining (NHEJ)" or "homology-directed repair (HDR)" in human cells, resulting in nucleotides deletion or addition which eventually may introduce a stop codon or alter the primary sequences of the target protein[33].

As human embryonic kidney (HEK) 293FT cells are easily cultured and their transfection efficiency is usually high, we decided to knock out NTMT1 in HEK293FT cells. To do this, primers containing the 20 nucleotides complementary to the target DNA sequences were synthesized, annealed and inserted into the digested vector PX459 which contains the coding sequences of SpCas9 and puromycin selection marker. Newly constructed plasmids containing the guide sequences were transfected into HEK293FT cells. After puromycin selection, most of the cells died and the survived ones were diluted and seeded into 96-well plates for single cell screening. After two weeks, the cells derived from one single cell were transferred into 24-well plates and protein NTMT1 expression were checked by WB (Fig. 2.9A) using a sensitive rabbit anti-NTMT1 antibody.

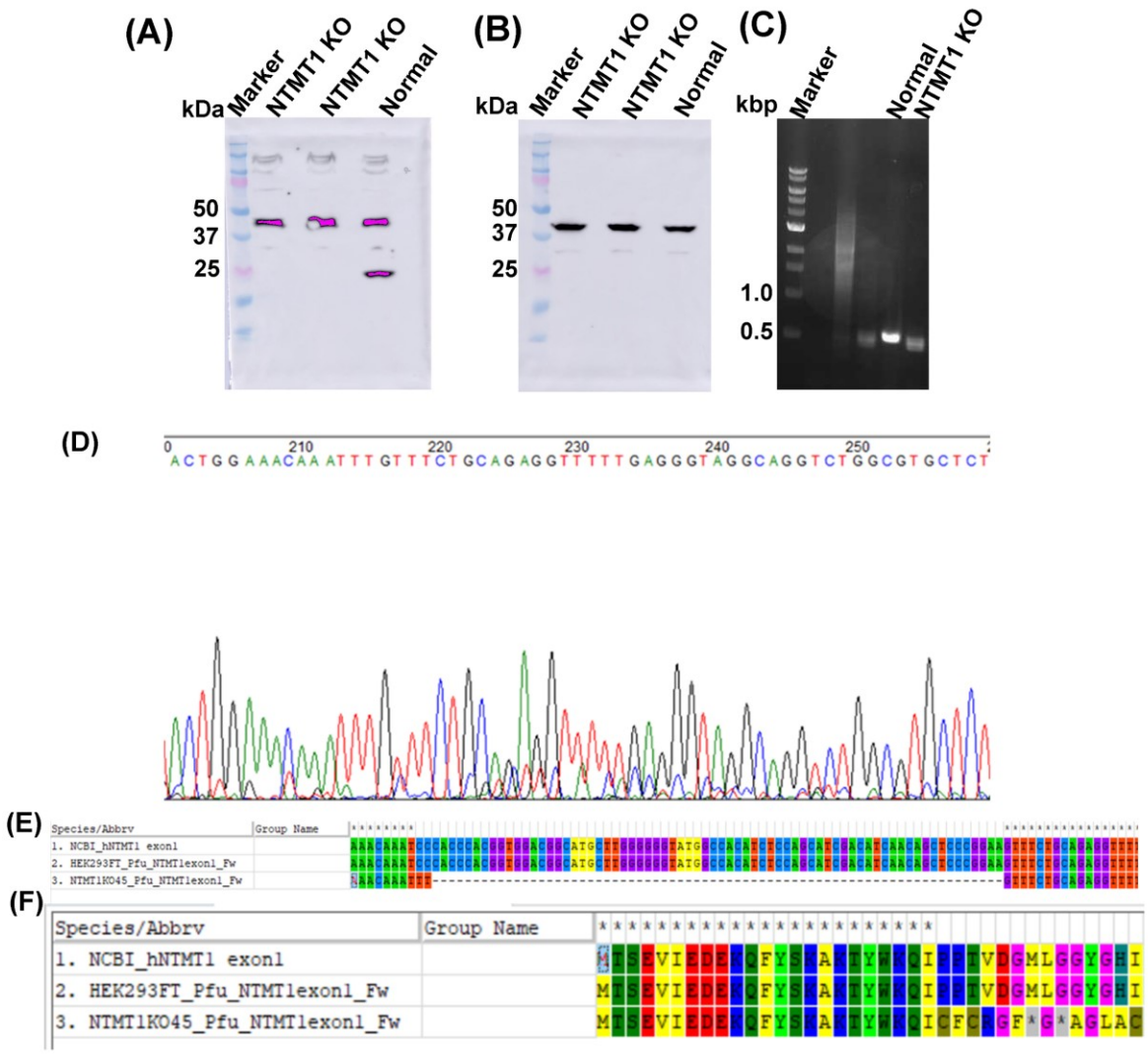


Figure 2.9 Construction of NTMT1 KO cell line.

(A) Detection of NTMT1 expression by WB using anti-NTMT1 antibody in normal HEK293FT and NTMT1 KO cells. (B) Detection the expression of actin as a loading control by WB using anti-actin antibody in normal HEK293FT and NTMT1 KO cells. (C) DNA gel for amplification of NTMT1 exon1 from normal HEK293FT and NTMT1 gene KO cells. (D) DNA sequencing chromatogram of NTMT1 exon1 amplified from NTMT1 gene KO cells. (E) Alignment of NTMT1 exon1 from NCBI, normal HEK293FT and NTMT1 KO cells (KO45 represents HEK293FT NTMT1 gene knockout cell #45). (F) Alignment of translated NTMT1 amino acid sequences from NCBI, normal HEK293FT and NTMT1 gene KO cells.

The rabbit anti-NTMT1 antibody is sensitive to detect endogenous NTMT1 (~ 25 kDa, Fig. 2.9A) from normal HEK293FT cells though a nonspecific band (~ 45 kDa) was also detected. Using actin as a protein loading control (Fig. 2.9B), the absence of NTMT1 shown in WB (Fig. 2.9B) indicated that NTMT1 was successfully knocked out. Further, the gene locus from normal and NTMT1 gene KO HEK293FT cells were extracted with Tissue DNA Kit, amplified by PCR and sequenced. Both DNA electrophoresis (Fig. 2.9C) and sequencing (Fig. 2.9D) results confirmed the deletion of ~ 70 nucleotides (Fig. 2.9C, 2.9D & 2.9E) and insertion of a stop codon (Fig. 2.9F).

2.2.9 Substrate profiling of NTMT1 in NTMT1 KO cell line

Approach for the activity-based NTMT1 substrate profile assay is summarized in Fig. 2.10A. Lysates of NTMT1 KO cells were incubated with either SAM or Hey-SAM in the presence of purified wt His_6-NTMT1. After click chemistry, the biotin conjugated proteins were enriched with streptavidin beads. Further, potential NTMT1 targets were cleaved with sodium dithionite as the DBA probe contains a cleavable azobenzene linker[14]. The purified proteins were separated on SDS-PAGE and visualized by Coomassie staining. The electrophoresis was only run for a short time (20 min) to narrow down the gel area for trypsin digestion. The pull-down samples treated with Hey-SAM contained significantly more proteins than the control samples which were treated with SAM (Fig. 2.10B). As a positive control, RCC1, one of the identified NTMT1 substrates, was successfully detected by WB for the samples properly labeled with Hey-SAM and click chemistry. Only negligible amount of RCC1 was detected in the control samples which were treated with click inactive cofactor SAM, validating the applicability of our protocol for NTMT1 substrate profiling. The lanes (Fig 2.10B) were excised, digested with trypsin and analyzed by LC-MS. Raw data were processed by Progenesis QI for proteomics.

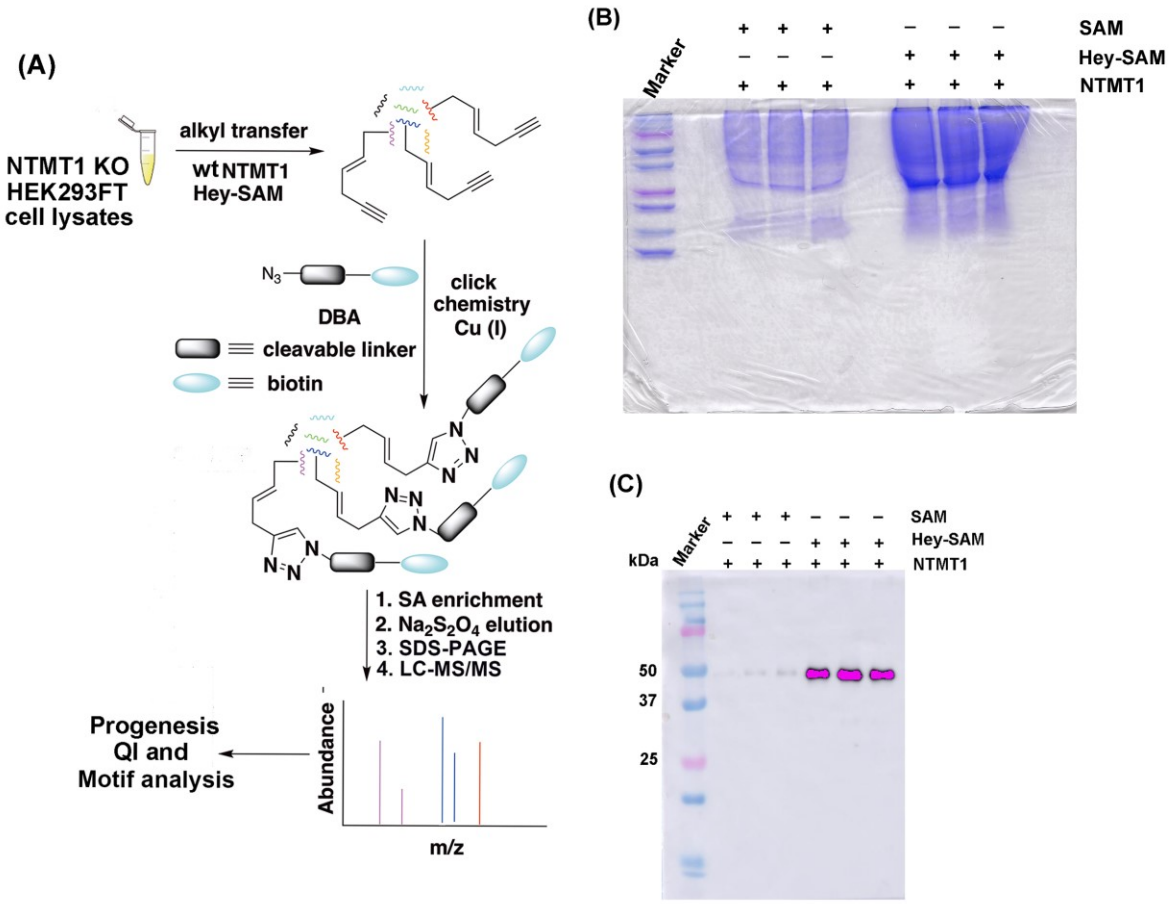


Figure 2.10 Activity-based substrate profiling of NTMT1.

(A) Schematic description of activity-based NTMT1 substrate profiling strategy. (B) Visualization of putative substrates of NTMT1 from NTMT1 KO cells by SDS-PAGE with Coomassie staining. (C) Detection of RCC1 from putative NTMT1 targets by WB using rabbit anti-RCC1 antibody.

2.2.10 Statistical and motif analysis of LC-MS raw data

Mass spectrometry is a powerful and high-throughput analytical technique for proteomic study. MS is highly sensitivity and provides abundant information, thus it would be difficult to analyze the data by hands. Hence, several software has been developed to assist protein identification and quantification. Here, the raw data were analyzed by software Progenesis QI for proteomics and

our object was to identify those proteins that showed abundance differences between different samples (SAM treated, and Hey-SAM treated), and therefore uncover the putative targets of NTMT1. To minimize human errors, three replicates were performed for control sample (treated with SAM) and pull-down sample (treated with Hey-SAM), respectively. In Progenesis, the abundances of peptide ions were calculated and the analysis of variance (Anova value or p value) was applied to determine whether the abundance differences between groups were significant or not. A small Anova value indicates that the difference between groups is significant. Based on Anova value (≤ 0.05) and max fold change (pull down/control ≥ 2), 1189 protein were identified. The principle component analysis (PCA, Fig 2.11A) graph showed that the three replicates in each condition were close to each other.

In addition, motif analysis was applied to assist the identifying of biological meaningful proteins. NTMT1 was previously reported to methylate proteins with an N-terminal sequence of Xaa (Ala/Pro/Ser)-Pro-Lys after methionine removal[9]. Later studies revealed that Xaa can be replaced with any residues except for hydrophobic Leu/Ile/Trp and negatively charged Asp/Glu residues[12]. This is in consistent with the crystal structures of NTMT1, in which the substrate binding pocket is negatively charged and can only provide limited space to accommodate residues smaller than Tyr[22]. Moreover, proline on the third position was predicted to be critical based on crystal structure analysis, however, *in vitro* peptide methylation assay indicated that proline could be substituted with asparagine, glycine, alanine, methionine or serine to some extent[12]. The fourth residue was only limited to positively charged lysine and arginine as it was demonstrated to play key roles in the formation of hydrogen bonding with the substrate binding channel[22]. Thus, the substrate N-terminal motif can be expanded to Xaa (Ala/Gly/Met/Pro/Ser/Thr/Tyr/Arg/Phe)-Xaa (Ala/Gly/Met/Asn/Pro/Ser/Thr)-Lys/Arg. Based on this expanded N-terminal sequence, 733

proteins were identified using a website motif search tool (https://www.genome.jp/tools/motif/MOTIF2.html).

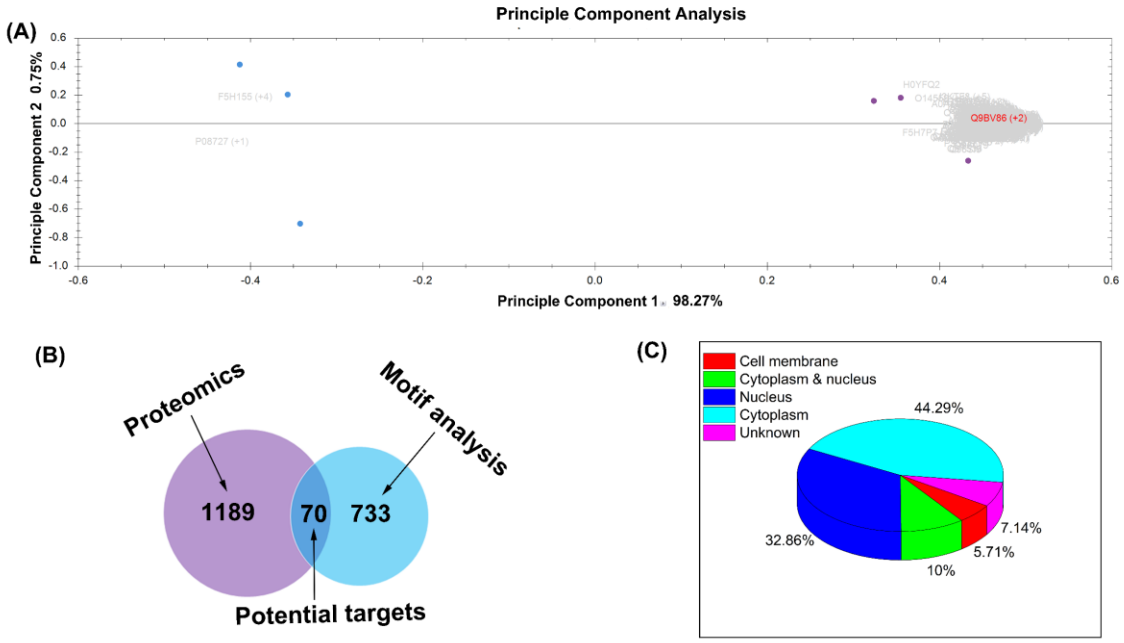


Figure 2.11 Statistical and motif analysis of LC-MS raw data.

(A) Principle component analysis (PCA) (Anova value ≤ 0.05 & max fold change ≥ 2). Blue dots represent the three replicates for samples treated with SAM. Purple dots represent the three replicates for samples treated with Hey-SAM. (B) Venn diagram of putative NTMT1 targets (70 human proteins) from proteomic (1189 human proteins) and motif analysis (733 human proteins). (C) Subcellular distributions of putative NTMT1 targets.

Seventy human proteins (Fig 2.11B, Table 2.3 & Appendix B) were predicted as putative NTMT1 targets according to the statistical and motif analysis. Among them, three of the previous confirmed NTMT1 substrates (highlighted in yellow in Table 2.3) including RCC1, SET and RL23A are included[9]. Subcellular location analysis showed that ~ 60% of the proteins are in cytoplasm and ~ 40% of the proteins are in cell nucleus.

Table 2.3 List of potential substrates of NTMT1 for target validation.

The previously identified NTMT1 substrates were highlighted with yellow color; the validated targets either at peptide level or at protein level are highlighted with green color. Note: protein subcellular distributions are assigned based on UniProt database and some of the locations are predicted by similarity.

UniProt ID	Gene name	Unique peptides	Confidence score	Anova (p)	Max fold change	Subcellular location
P04181	OAT	13	115.58	8.47E-05	2.88	Cytoplasm
Q5H9U9	DDX60L	5	92.45	5.94E-05	4.12	Unknown
P62750	RPL23A	5	70.68	4.02E-05	4.77	Cytoplasm
P18754	RCC1	4	56.32	5.01E-05	3.46	Cytoplasm, nucleus
P35232	PHB	1	43.24	0.001095	3.05	Cytoplasm
P62263	RPS14	2	34.10	0.000113	4.20	Cytoplasm, nucleus
P04908	HIST1H2AB	1	29.50	0.00172	2.57	Nucleus
P62805	HIST1H4A	1	26.25	0.000414	2.72	Nucleus
Q9NTK5	OLA1	1	16.12	0.000527	3.57	Cytoplasm, nucleus
Q86U86	PBRM1	1	15.47	4.33E-06	7.29	Nucleus
Q01105	SET	2	11.44	8.35E-05	4.14	Cytoplasm
A1X283	SH3PXD2B	1	5.37	0.000273	2.94	Cytoplasm

2.2.11 Validation of target proteins *in vitro*

NTMT1 substrates RCC1, SET and RL23A were successfully identified by this activity-based substrate profiling approach, suggesting that the assay worked as intended. Thus, we decided to select some other proteins (Table 2.3) which either have important biological functions or have very low or high confidence scores for *in vitro* validation. The open reading frames of protein

histone H2A type 1-B/E (H2A1B), histone H4, ornithine aminotransferase (OAT), prohibitin (PHB), obg-like ATPase 1 (OLA1) and 40S ribosomal protein S14 (RS14) were amplified and inserted into *E. coli* expression vectors (pET30a or pET28b) containing a C-terminal His_6 tag. Among those proteins, only OLA1 and RS14 (Fig 2.12A, inset) were successfully purified using Ni-NTA resin as the other proteins were either not well expressed or existed in inclusion bodies. OLA1 was reported to play important roles in multiple cellular processes[34-38] and therefore we investigated its methylation *in vitro* and in human cells which were described in Chapter 3. Here, the intact protein mass showed that the initial methionine of purified RS14 was removed (Fig. 2.12A) and monomethylated protein (Fig. 2.12C) was detected after incubating RS14 with SAM and 3 μM NTMT1 for 2h at 37℃. Methylation of RS14 was enzyme-catalyzed because no methylated products were detected in the absence of NTMT1 (Fig. 2.12B). The major product of RS14 methylation was the monomethylated protein, indicating that the methylation reaction rate was slow.

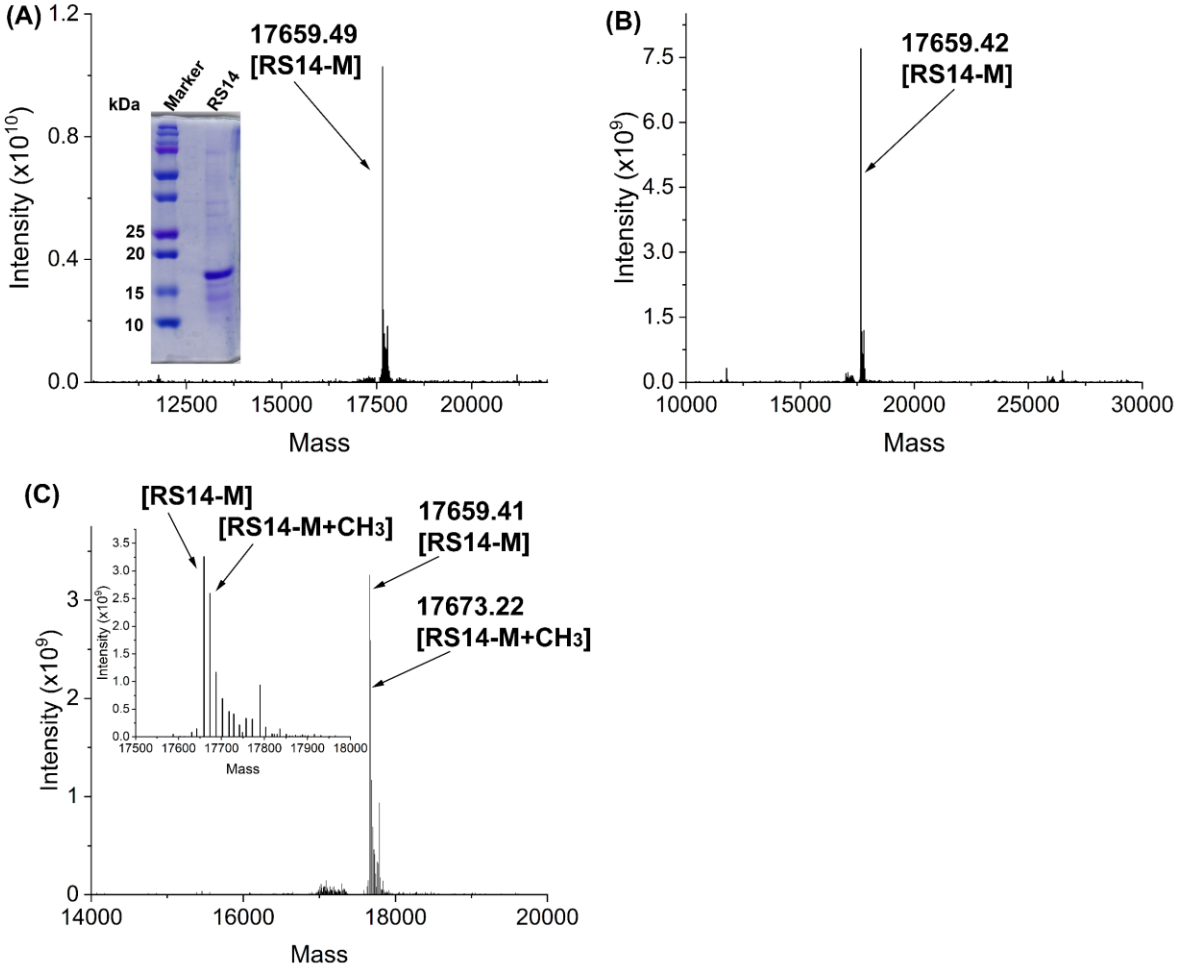


Figure 2.12 MS analysis of protein RS14 methylation with SAM and wt NTMT1.

(A) Intact protein mass and SDS-PAGE (inset) of the purified RS14 protein (calculated MW for initial methionine removed RS14 [RS14 - M]: 17661.2; detected MW: 17659.49). (B) Intact protein mass of 20 μM RS14 protein mixed with 200 μM SAM in the absence of NTMT1 (detected MW: 17659.42). (C) Intact protein mass of RS14 protein mixed with 200 μM SAM in the presence of 3 μM NTMT1 (calculated [RS14 - M + CH_3]: 17675.2, detected MW: 17673.22).

Moreover, many identified proteins (e.g., probable ATP-dependent RNA helicase DDX60-like (DDX60L), protein polybromo-1 (PB1), SH3 and PX domain-containing protein 2B (SPD2B)) have relatively large molecular masses, thus it could be challenging to express and purify them in

E. coli. To simplify, we synthesized their N-terminal peptide analogs as instead. Meanwhile, the N-terminal peptide of protein H2A1B was also synthesized because histone H2B in *Drosophila melanogaster* was reported as a substrate for dNTMT[39] and we wanted to verify whether human histone proteins would be methylated by human NTMT1. Peptides containing the first ten residues (without methionine) of PB1 (Fig. 2.13A), SPD2B (Fig. 2.14A), DDX60L (Fig 2.15A) and H2A1B (Fig. 2.16A) were synthesized and characterized by HPLC and LC-MS. MS-based methylation assay confirmed that PB1 N-terminal peptide with a GSK motif was trimethylated in the presence of SAM and wt NTMT1 (Fig 2.13B & 2.13C). SPD2B N-terminal peptide with a PPR motif was dimethylated in the presence of SAM and wt NTMT1 (Fig. 2.14B & 2.14C). To our surprise, DDX60L N-terminal peptide harboring the same motif as PB1 was not methylated in the presence of SAM and 3 μM NTMT1 (Fig. 2.15B & 2.15C). Though DDX60L and PB1 share the same GSK motif, their fifth residues are different. DDX60L contains a negatively charged aspartate residue in position 5 while PB1 contains a positively charged arginine residue. According to the crystal structure of NTMT1, aspartate 5 is not preferred as it locates in the entrance of a negatively charged channel[16]. In addition, H2A1B N-terminal peptide was not methylated in the presence of 200 μM SAM and 3 μM NTMT1 (Fig. 2.16B & 2.16C). Human H2A1B contains an SGR motif which is quite different from the PPK motif of H2B in *Drosophila melanogaster*, indicating that SGR was not a good substrate motif for human NTMT1.

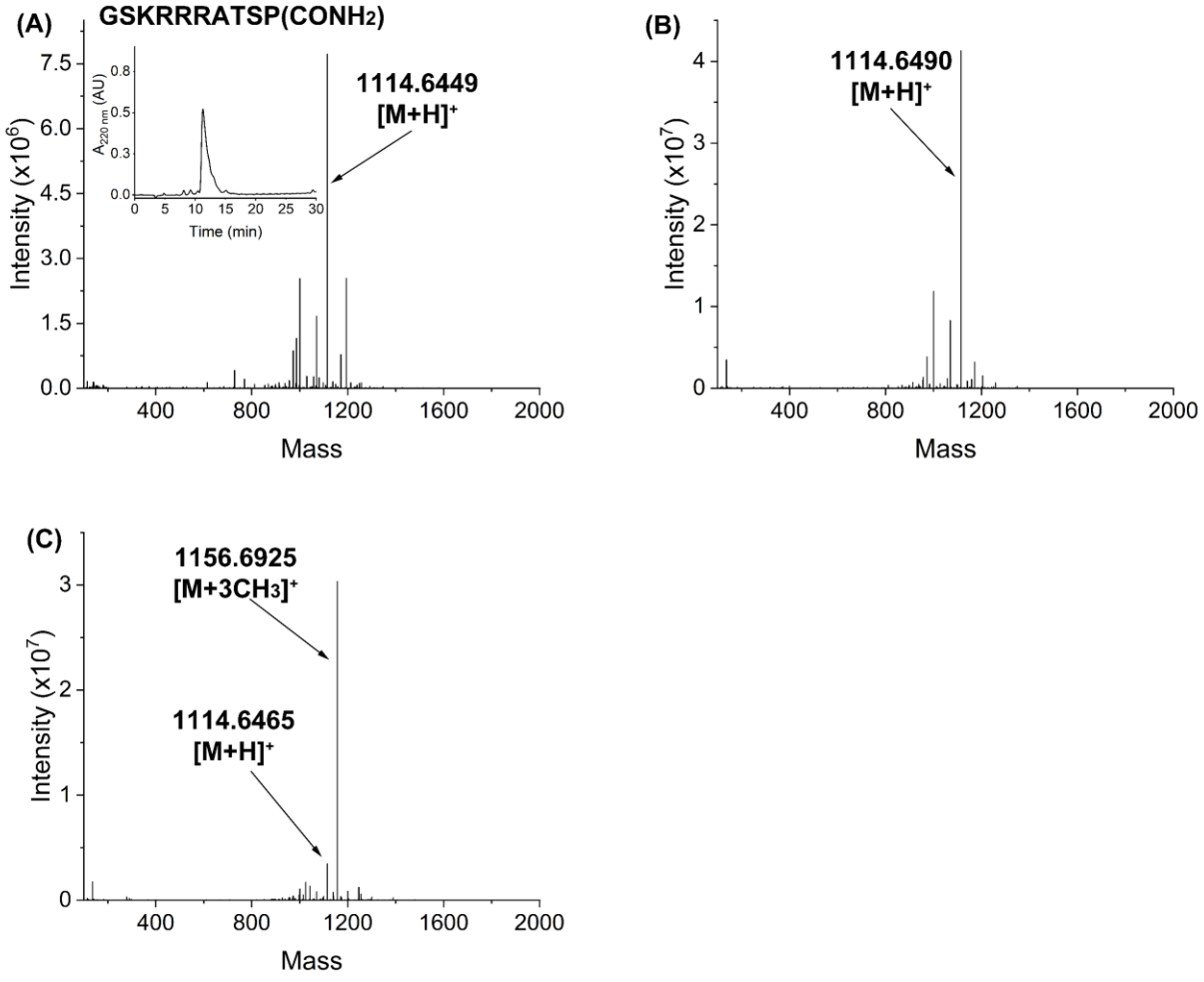


Figure 2.13 MS analysis of PB1 peptide methylation with SAM and wt NTMT1.

(A) Mass spectrum and HPLC analysis (inset) of the synthesized PB1 peptide (calculated $[M + H]^+$: 1114.6557; detected: 1114.6449). (B) Mass spectrum of 20 μM PB1 peptide mixed with 200 μM SAM in the absence of NTMT1 (calculated $[M + H]^+$: 1114.6557; detected: 1114.6490). (C) Mass spectrum of PB1 peptide mixed with 200 μM SAM in the presence of 3 μM NTMT1 (calculated $[M + 3CH_3]^+$: 1156.7027, detected: 1156.6925).

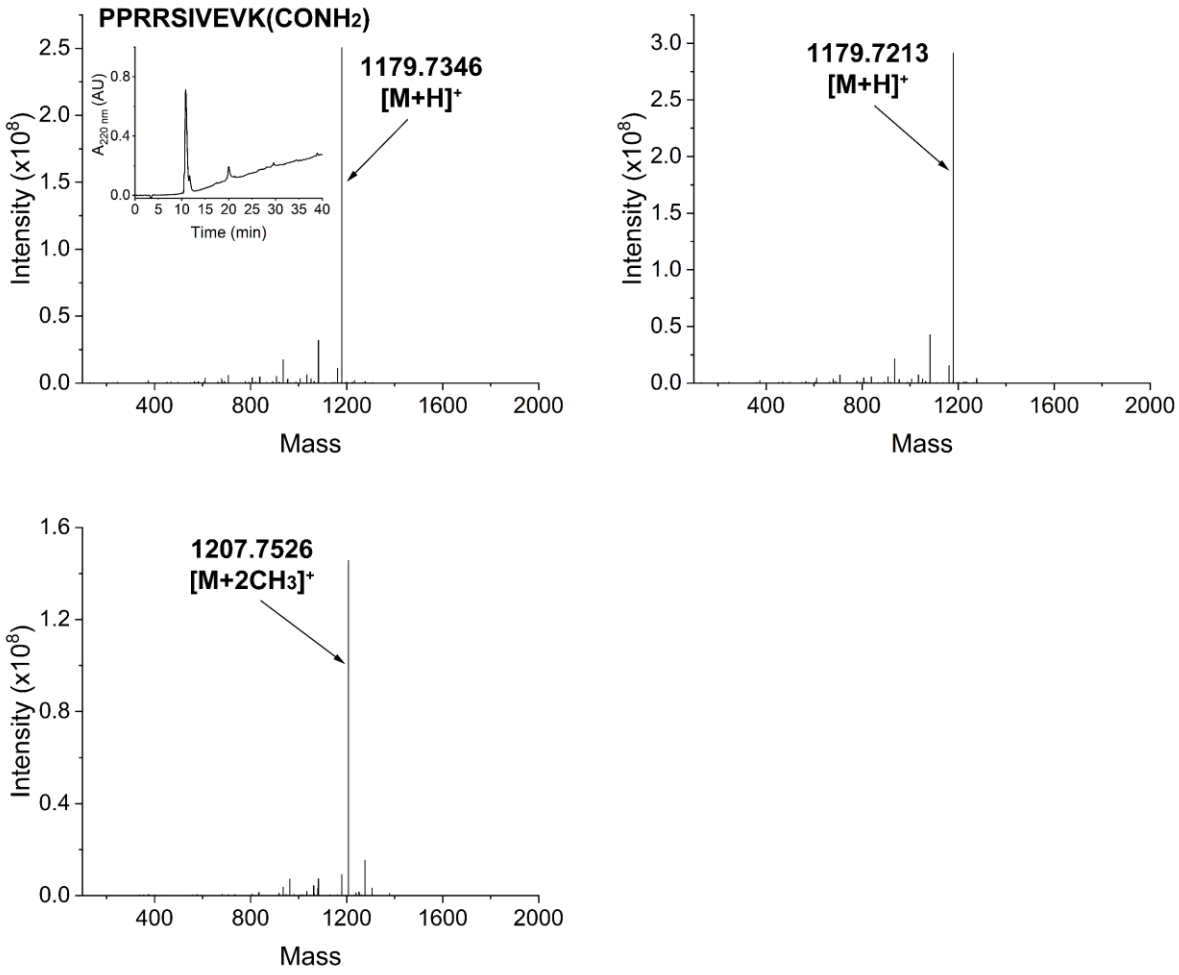


Figure 2.14 MS analysis of SPD2B peptide methylation with SAM and wt NTMT1.

(A) Mass spectrum and HPLC analysis (inset) of the synthesized SPD2B peptide (calculated $[M + H]^+$: 1179.7326; detected: 1179.7346). (B) Mass spectrum of 20 μM SPD2B peptide mixed with 200 μM SAM in the absence of NTMT1 (calculated $[M + H]^+$: 1179.7326; detected: 1179.7213). (C) Mass spectrum of SPD2B peptide mixed with 200 μM SAM in the presence of 3 μM NTMT1 (calculated $[M + 2CH_3]^+$: 1207.7639, detected: 1207.7526).

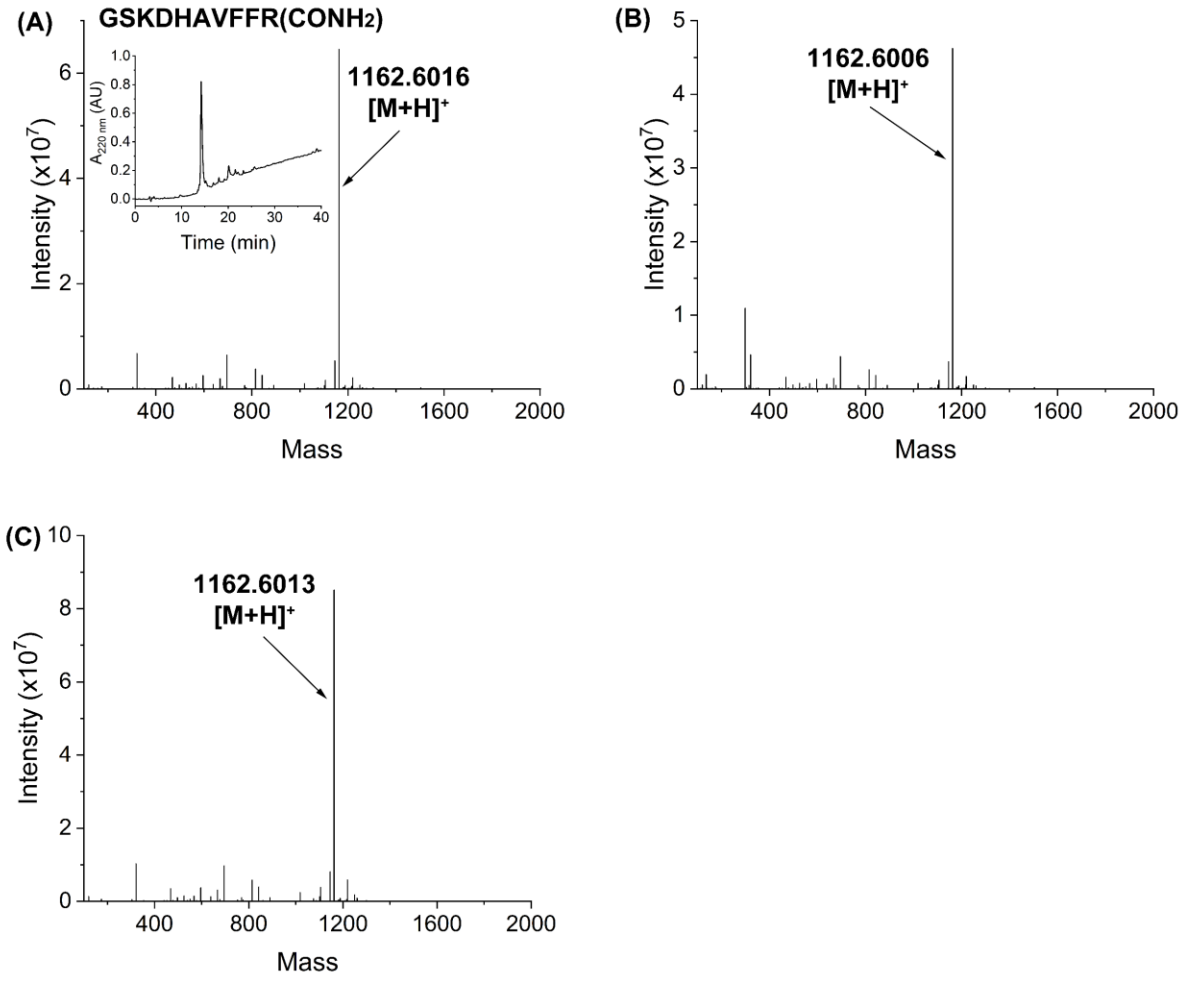


Figure 2.15 MS analysis of DDX60L peptide methylation with SAM and wt NTMT1.

(A) Mass spectrum and HPLC analysis (inset) of the synthesized DDX60L peptide (calculated $[M + H]^+$: 1162.6122; detected: 1162.6016). (B) Mass spectrum of 20 μM DDX60L peptide mixed with 200 μM SAM in the absence of NTMT1 (calculated $[M + H]^+$: 1162.6122; detected: 1162.6006). (C) Mass spectrum of DDX60L peptide mixed with 200 μM SAM in the presence of 3 μM NTMT1 (calculated $[M + 3CH_3]^+$: 1204.6591, detected: 1162.6013).

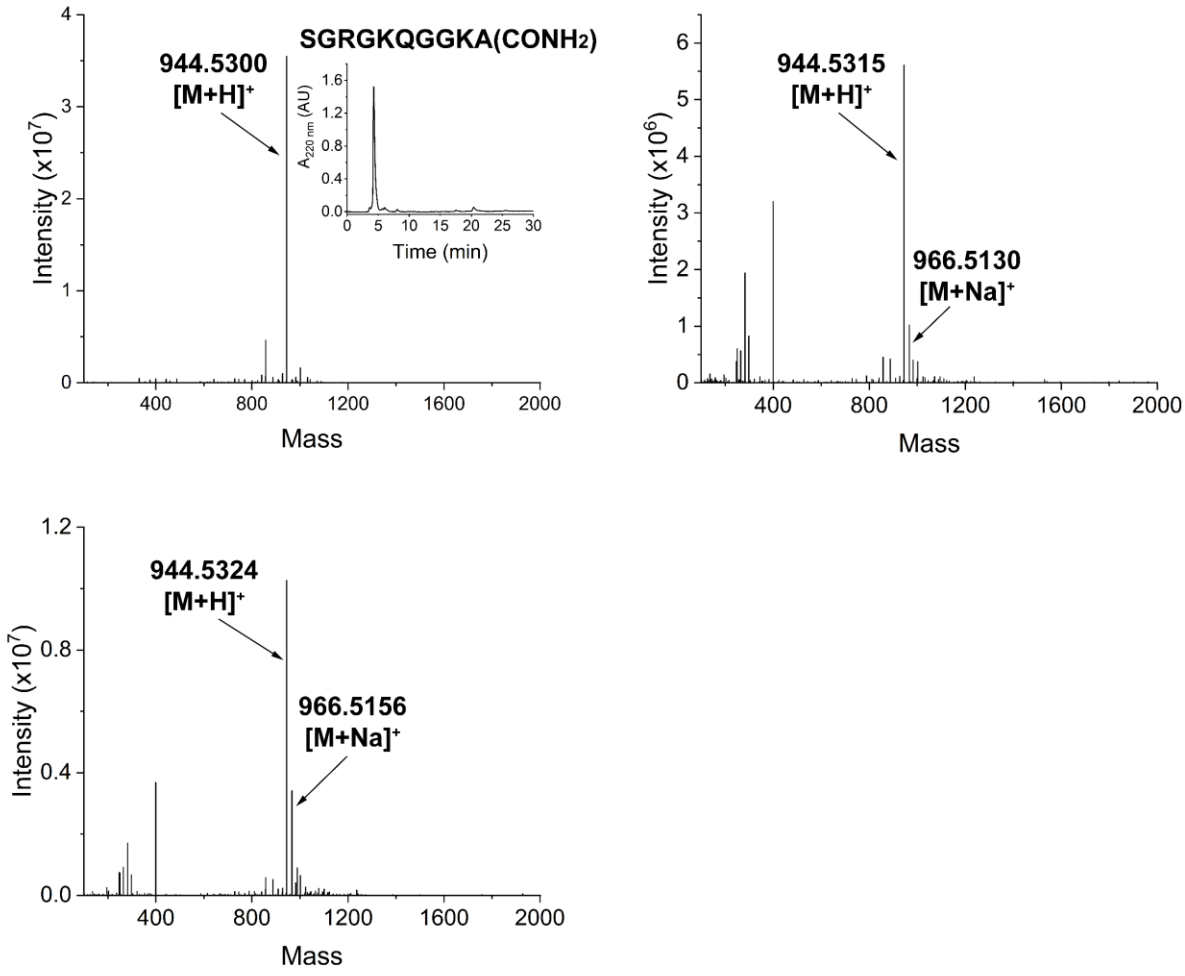


Figure 2.16 MS analysis of H2A1B peptide methylation with SAM and wt NTMT1.

(A) Mass spectrum and HPLC analysis (inset) of the synthesized H2A1B peptide (calculated $[M + H]^+$: 944.5390; detected: 944.5300). (B) Mass spectrum of 20 μM H2A1B peptide mixed with 200 μM SAM in the absence of NTMT1 (calculated $[M + H]^+$: 944.5390; detected: 944.5315). (C) Mass spectrum of H2A1B peptide mixed with 200 μM SAM in the presence of 3 μM NTMT1 (calculated $[M + 3CH_3]^+$: 986.5859, detected: 944.5324).

Taken together, validation of NTMT1 putative targets revealed protein RS14 as NTMT1 targets *in vitro*. In addition, the N-terminal peptides of PB1 and SPD2B was tri- and dimethylated with NTMT1 respectively, suggesting that their corresponding proteins are likely to be methylated

by NTMT1. Moreover, the importance of the fifth residue adjacent to Xaa-Xaa-Lys/Arg motif may be previously underestimated as DDX60L peptide containing a GSKD motif cannot be methylated while PB1 peptide with a GSKR motif can be trimethylated.

To summarize, target validation demonstrated that protein RS14 and the N-terminal peptide analogs of PB1 and SPD2B were NTMT1 substrates *in vitro*. In addition, lots of the candidate substrates have important biological functions. For example, the bromodomain containing protein PB1 functions as a "reader" for histone lysine acetylation and it plays important roles in maintaining the chromosomal stability[40-41]. Thus, it would be meaningful to investigate its methylation at protein level in future. Moreover, OLA1 was reported to play important roles in multiple cellular processes including regulation of centrosome, heat shock response and cell growth[34-38], and therefore we decided to investigate its methylation *in vitro* and *in vivo* which were described in Chapter 3.

2.3 Materials and methods

2.3.1 General information

Unless otherwise specified, HPLC was performed with a Breeze 2 system consisting of a 1525 pump and 2998 photodiode array detector (Waters). High-resolution mass spectroscopy was acquired with an Acquity H-class or M-class UPLC coupled with XEVO G2-XS QTof (Waters). To determine the intact protein mass, proteins were separated by either a BEH C4 column (1.7 μm, 300 Å, 2.1 mm × 50 mm, for H class UPLC, Waters) or an M-class BEH C4 (1.7 μm, 300 Å, 300 μm × 150 mm, Waters) with an M-class BEH C4 trap column (5 μm, 300 Å, 300 μm × 50 mm, Waters). For small molecules, synthesized peptides, peptides mapping, protein identification and quantification, either a BEH C18 (1.7 μm, 130 Å, 2.1 mm × 50 mm, for H class UPLC, Waters) or an M-class HSS T3 column (1.8 μm, 100 Å, 300 μm × 100 mm, Waters) with an M-class

symmetry C18 trap column (5 μm, 100Å, 300 μm × 50 mm, Waters). Mobile phase A and B consisted of 0.1% formic acid in water and acetonitrile, respectively. HPLC samples were centrifuged to remove particles before injection. MS samples were filtered using either 0.2 μm PTFE syringe filters (Fisher Scientific) or ultrafree PTFE membranes (Millipore). NMR spectrums were obtained on a 400 MHz (Varian) or 600 MHz (Bruker) spectrometer.

2.3.2 Plasmids construction

Human NTMT1 gene was synthesized by Genscript and amplified by polymerase chain reaction (PCR) to introduce restriction sites NdeI and HindIII. Human RCC1 gene was amplified from pST50Tr-hRCC1 (generous gift from Dr. Tan's lab)[42] to introduce NdeI and XhoI restriction sites. The open reading frames of protein RS14, PHB, OAT, H2A1B and H4 were amplified from the complementary DNA (cDNA) originated from HEK293FT cell line. The amplified PCR products and vectors were digested with the corresponding restriction endonucleases. NTMT1 gene was inserted into pET28-MHL (Addgene) to generate pET28-MHL-NTMT1 containing an N-terminal His_6 tag. Genes of RCC1, PHB, OAT, H2A1B and H4 were inserted into vector pET30a containing a C-terminal His_6 tag. RS14 gene was inserted into pET28b vector to generate pET28b-RS16 containing a C-terminal His_6 tag.

To generate NTMT1 KO cell line using CRISPR-Cas9 system, guide sequences were designed according to an online CRISPR Design Tool (http://tools.genome-engineering.org) established by Dr. Zhang group[33]. The synthesized oligos were annealed to give sticky BbsI ends and inserted into the BbsI digested vector PX459 (generous gift from Dr. Xiong lab) which contains single guide RNA scaffold, cas9 coding sequence and puromycin selection marker.

All the insertions were verified by DNA sequencing before use. The sequences of all used primers are listed in Table 2.4.

Table 2.4 Primer sequences for plasmids construction.

The restriction sites are bolded and underlined.

Plasmids	Primer sequences (restriction sites bolded and underlined)	
pET28-MHL-NTMT1	NdeI_Fw	TAGCG**CATATG**ACGAGCGAGGTGATAG
	HindIII_Rv	TGACG**AAGCTT**CATCTCAGGGCAAAGC
pET30a-RCC1	NdeI_Fw	TCGTTC**CATATG**TCACCCAAGCGCATAGC
	XhoI_Rv	AAACTA**CTCGAG**GCTCTGTTCTTTGTCCTTGAC
pET28b-RS14	NcoI_Fw	TTTTTA**CCATGG**CACCTCGAAAGGGGAAGG
	HindIII_Rv	TTTTTT**AAGCTT**CAGACGGCGACCACGGCGACC C
pET30a-PHB	NdeI_Fw	AACCCA**CATATG**GCTGCCAAAGTGTTTGAGTC
	XhoI_Rv	AAAATT**CTCGAG**CTGGGGCAGCTGGAGGAGCAC
pET30a-OAT	NdeI_Fw	CCGAAA**CATATG**TTTTCCAAACTAGCACATTTGC
	HindIII_Rv	CCGAAT**AAGCTT**GAAAGACAAGATGGTCTTGTTAAT
pET30a-H2A1B	NdeI_Fw	TTTTAA**CATATG**TCTGGTCGCGGCAAACAAGGC
	HindIII_Rv	AAATCG**AAGCTT**CTTTCCCTTGGCCTTATGATGGCTC
pET30a-H4	NdeI_Fw	TTTTAA**CATATG**TCCGGCAGAGGAAAGGGCGGAAAAG
	HindIII_Rv	AAAATT**AAGCTT**GCCTCCGAAGCCGTACAGGGTGCGC
PX459-$NTMT1_{sgRNA1}$	BbsI_Fw	**CACC**GCTGGAAACAAATCCCACCCA
	BbsI_Rv	**AAAC**TGGGTGGGATTTGTTTCCAGC
PX459-$NTMT1_{sgRNA2}$	BbsI_Fw	**CACC**GCATGCCGTCCACCGTGGGT
	BbsI_Rv	**AAAC**ACCCACGGTGGACGGCATGC
Seq-$NTMT1_{exon1}$	NdeI_Fw	AAAACC**CATATG**CTCTGCACTCAGCAGGATGTGTGAC
	XhoI_Rv	AAAATT**CTCGAG**ATCTAGAAGTTACCCCACTGTGCC

2.3.3 Cell culture conditions

Human embryonic kidney 293FT cells (generous gift from Dr. Yulan Xiong) and the NTMT1 gene KO cells were cultured in Dulbecco's Modification of Eagle's Medium (DMEM, with 4.5

g/L glucose, L-glutamine & sodium pyruvate, Corning cellgro) supplemented with 10% fetal bovine serum (heat inactivated FBS, Gibco) and 1% penicillin-streptomycin (P/S, Gibco) in a humidified atmosphere containing 5% CO_2 at 37 °C.

2.3.4 Expression and purification of recombinant proteins

Plasmids pET28-MHL-NTMT1, pET28b-RS14, pET30a-PHB, pET30a-OAT, pET30a-H2A1B and pET30a-H4 were transformed into *E. coli* BL21(DE3) competent cells (Lucigen). Plasmid pET30a-RCC1 was transformed into *E. coli* C41 (DE3) competent cells (Lucigen). Bacterial were grown in 4L of Luria-Bertani (LB) medium in the presence of 50 μg/ml kanamycin. Isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.4 mM as the final concentration) was added into the culture to induce the expression of recombinant proteins when $OD_{600\ nm}$ reached to 0.6 - 0.8. Bacterial were then cultured for additional 18 h at 18 ℃ and harvested by centrifugation at 5,000 × g for 20 min at 4 ℃. Cell pellets were collected and frozen at -80 ℃ for further use.

The ingredients of buffers used in this study are summarized in Table 2.5. Protein purification was performed at 4 ℃. Briefly, bacterial pellets were suspended and lysed in buffer A by sonication for 15 min (5-second on and 3-second off) at 40% amplitude. After centrifugation at 18,000 × g for 45 min at 4 ℃, the cell debris was removed, and the supernatant were incubated with Ni-NTA resin (Qiagen) which had been previously equilibrated with buffer A for 30 min. After washing with 30 column volumes (CV) of buffer B, recombinant proteins containing the His_6 tag were eluted with 5 CV of buffer C. The eluted fractions containing the target protein were combined and exchanged to buffer D using either a HiPrep™ 26/10 desalting column (GE Healthcare) or Amicon Ultra-15 Centrifugal Rilter Unites (30 kDa or 10 kDa, Millipore). Proteins were further exchanged to 10 mM ammonium bicarbonate (ABC) for determining their intact protein masses. Proteins were further concentrated to 10-50 mg/ml, flash frozen with liquid N_2 and

stored in aliquots at -80 ℃ for further use. Protein purity was examined by 12% tris-tricine or tris-glycine SDS-PAGE and their concentrations were determined by bicinchoninic acid assays (BCA, ThermoFisher) using a BSA calibration curve.

Table 2.5 Ingredients of buffers and click chemistry cocktail.

Buffer	Composition
A	25 mM HEPES (pH 7.4), 300 mM NaCl and 5 mM Imidazole
B	25 mM HEPES (pH 7.4), 300 mM NaCl and 30 mM Imidazole
C	25 mM HEPES (pH 7.4), 300 mM NaCl and 250 mM Imidazole
D	25 mM HEPES (pH 7.4), 150 mM NaCl and 50 mM KCl
E	25 mM Tris (pH 7.5) and 50 mM KCl
F	25 mM HEPES (pH 7.4), 150 mM NaCl and 1% SDS
G	25 mM HEPES (pH 7.4) and 300 mM NaCl
H	1% SDS in 1 × PBS (pH 7.4)
I	1 × PBS (pH 7.4)
J	4 M urea in 1× PBS (pH 7.4)
K	25 mM HEPES (pH 7.4), 50 mM KCl, 10% glycerol with 1× protease inhibitor (Pierce)
Click chemistry cocktail	one volume of 5 mM DBA and ten volumes of premixed solution consisting of 1 mM TBTA, 10 mM $CuSO_4$, and 10 mM TCEP (pH was adjusted to 7.0 using 1 M NaOH)

2.3.5 Peptide synthesis

RCC1 peptide [SPKRIAKRRS($CONH_2$)], PB1 peptide [GSKRRRATSP($CONH_2$)], SPD2B peptide [PPRRSIVEVK($CONH_2$)], DDX60L peptide [GSKDHAVFFR($CONH_2$)], and histone H2A1B peptide [SGRGKQGGKA($CONH_2$)] were synthesized on Rink amide resins (Peptide International) using standard Fmoc chemistry[43]. The synthesized peptides were characterized by

HPLC and mass spectrometry. RCC1 peptide and SPD2B peptide were further purified through semi-preparative HPLC. The conditions of analytical and semi-preparative HPLC are summarized in table 2.6.

Table 2.6 HPLC conditions for peptide purification and characterization.

		Semi-preparative HPLC	Analytical HPLC
Column		Luna C18(2), 5 μm, 100 Å, 10 mm × 250 mm (Phenomenex)	XBridge BEH Shield RP 18, 5 μm, 130 Å, 4.6 mm × 250 mm (Waters)
Flow Rate		3.0 mL/min	1.0 mL/min
Mobile Phases		**A:** 0.1% TFA in H_2O; **B:** 0.1% TFA in MeOH	
Gradient (**B**)	RCC1 peptide (retention time)	0-5 min: 5%; 5-30 min: 5%-30% (23.7 min)	0-40 min: 5%-95% (10.2 min)
	PB1 peptide (retention time)	N/A	0-30 min: 1%-15%; (11.2 min)
	SPD2B peptide (retention time)	0-40 min: 10%-40% (18.3 min)	0-40 min: 10%-90% (10.8 min)
	DDX60L peptide (retention time)	N/A	0-40 min: 10%-90% (14.2 min)
	H2A1B peptide (retention time)	N/A	0-30 min: 1%-15%; (4.3 min)

2.3.6 MS-based methylation and alkylation assay for RCC1 peptide and protein

Methylation and alkylation of RCC1 at peptide and protein levels by wt NTMT1 and cofactors were investigated using mass spectrometry. Specifically, 20 μM RCC1 peptide dissolved in 10 mM ABC or RCC1 protein dissolved in buffer D, 200 μM SAM or Hey-SAM, and 3 μM wt NTMT1 were mixed and incubated at 37 ˚C for 2 h. Negative controls were performed in the

absence of wt NTMT1. The peptide reaction mixtures were quenched with equal volume of water containing 0.1% formic acid and then analyzed by LC-MS. The protein reaction mixtures were not quenched and directly analyzed by LC-MS, which was performed on a M-class BEH C4 column eluted at a flow rate of 0.4 mL/min. The gradient for intact protein mass analysis was: 0-1 min, 5% B; 1-3.5 min, 5%-100% B; and 3.5-4.7 min, 100% B. Elution from 2–4.7 min was injected into the mass spectrometer, which was operated in a MS mode with a positive polarity. Raw data of intact protein mass were processed by BiopharmaLynx (Waters).

2.3.7 HPLC-based steady state kinetic study

The steady state kinetics were performed by measuring the SAH release using HPLC. The concentrations of SAH formed were quantified by adenosine which served as an external standard. Experiments were performed by varying concentrations of one substrate in the presence of the other one at saturation. Specifically, concentrations of SAM and Hey-SAM varied from 0.1 to 8 μM and from 0.2 to 16 μM, respectively, in the presence of 40 μM RCC1 peptide. Concentration of RCC1 peptide varied from 0.5 to 16 μM in the presence of 20 μM SAM or Hey-SAM.

Experiments were carried out in the following way. In a final volume of 400 μL, 0.3 μM wt NTMT1 was added to the mixtures of cofactor and RCC1 peptide in buffer E at 37 ˚C. Aliquots of 40 μL were withdrawn at selected time intervals (1 to 90 min) and the reactions were quenched with 40 μL of 0.4% TFA solution containing 4 μM adenosine. The reaction mixtures were centrifuged at 18,000 × g for 2 min and 40 μL of the supernatant were injected into an XBridge Shield RP 18 column (3.5 μm, 130 Å, 4.6 mm × 150 mm, Waters) using an autosampler (Agilent). The mobile phase contained an isocratic gradient of 1% methanol in water supplemented with 0.01% TFA and the flow rate was 1 ml/min. Measurements were performed in triplicate. Concentrations of SAH were determined by area ratios of the peaks corresponding to SAH and

adenosine under 260 nm. The data obtained were fitted into the Michaelis-Menten equation ($v = \frac{V_{max}[S]}{K_M+[S]}$) using SigmaPlot.

2.3.8 Isothermal titration calorimetry

ITC measurements were carried out with MicroCal iTC200 (GE Healthcare) at 25 ˚C. Purified wt NTMT1 was diluted to a final concentration of 50 μM using buffer D. Cofactors or RCC1 peptide were dissolved and diluted to 300-500 μM using the same buffer. For titration, the first syringe injection volume was 0.5 μl and the volume for the following 18 injections was 2 μl. The interval between each injection was 180 s and the reference power was 7 μcal/s. Measurements were performed in duplicate.

2.3.9 Verification of RCC1 modification site by in-gel trypsin digestion and peptide mapping

To identify the modification site of RCC1 protein by Hey-SAM and wt NTMT1, the alkylation reaction mixture was separated by SDS-PAGE. In-gel trypsin digestion for LC-MS analysis was performed following the standard protocol[44]. Briefly, the target gel band was excised, cut into 1.0 mm^3 pieces, reduced with 100 mM dithiothreitol (DTT), and then alkylated with 55 mM iodoacetamide. After dehydration with acetonitrile, protein digestion was implemented with sequencing-grade trypsin (Promega) at 20 ng/μL in 10 mM ABC containing 10% (vol/vol) acetonitrile at 37 ˚C for 18 h. After extraction with 100 μl of 5% fomic acid, 100 μl of 100% acetonitrile/10% fomic acid (1:1, vol/vol), and 100 μl of pure acetonitrile, respectively, the digestion products were further concentrated by SpeedVac.

For LC−MS analysis, peptides were re-dissolved in water containing 5% formic acid and separated on an BEH C18 column. The flow rate for mobile phase was 0.2 mL/min and the gradient of mobile phase B increased from 0 to 50% in 50 min. The mass spectrometer was operated in an

MS^E mode with a ramp collision energy ranging from 20 to 45 V. Raw data were analyzed by BiopharmaLynx (Waters). The parameters for peptide mapping included peptide mass tolerance of 10 ppm, MS^E mass tolerance of 30 ppm, no missed cleavages allowed, fixed modification of cysteine carbamidomethylation and variable modification of *N*-terminal chain by Hey.

2.3.10 Detection of biotin labeled RCC1 protein

In a final volume of 200 µl, 20 µM purified RCC1 was incubated with 50 µM Hey-SAM and 3 µM wt NTMT1 in buffer D at 4 °C for overnight. Negative controls were carried out either using SAM to substitute Hey-SAM or in the absence of wt NTMT1. Cold methanol (1 ml) were added to the mixtures to precipitate proteins. After centrifuging at 14,100 × g for 2 min and removing the supernatant, protein pellets were washed with 1 mL cold methanol for three times, air-dried for 15 min, and re-dissolved in 200 µL buffer F. A total volume of 22 µL click chemistry cocktail (Table 2.5) were added into each sample and the click reaction was performed in the dark for 2 h at room temperature[15]. Proteins were then precipitated with 1 mL cold methanol and kept overnight at −80 °C. After washing with 1 mL cold methanol, the protein pellets were re-dissolved in 100 µL buffer F. Presence of RCC1 and biotin were verified by Western blot (WB) as described below in Section 2.4.12 using rabbit anti-RCC1 antibody (Abcam) and streptavidin-HRP (SA-HRP, Abcam), respectively.

The biotin labeled RCC1 were separated from non-labeled RCC1 using gel shift assay with streptavidin. Briefly, 10 µL of the above re-dissolved protein samples after click chemistry were mixed with 10 µL of 2 × Laemmli buffer and denatured at 100 °C for 5 min. After cooling down to room temperature, the samples were incubated with 15 µL of 1mg/ml streptavidin solution (NEB) at room temperature for 30 min[45] and separated by SDS-PAGE. Bands corresponding to

the RCC1-streptavidin complexes were excised and analyzed following the procedures for in-gel trypsin digestion and peptide mapping described in Section 2.4.9.

2.3.11 Pull-down of overexpressed RCC1 from *E. coli* lysates *via* activity-based substrate profiling strategy

E. coli C41(DE3) cells transformed with pET30a-RCC1 were cultured and harvested as described above in Section 2.4.4. Bacterial pellets were re-suspended in buffer G, sonicated for 15 min (5-second on and 3-second off) at 40% amplitude and centrifuged at 18,000 × g for 45 min. The supernatant (1 ml) were withdrawn and incubated with 50 μM Hey-SAM and 3 μM wt NTMT1 overnight at 4 ˚C. Negative controls were carried out either using SAM to substitute Hey-SAM or in the absence of wt NTMT1. Proteins were precipitated out by adding 4 mL cold methanol, air-dried for 15 min, and re-dissolved in 4 mL buffer F. A total volume of 440 μL click chemistry cocktail (Table 2.5) were added to each sample and click chemistry was performed in dark for 2 h at room temperature[15]. Proteins were precipitated out by adding 16 mL cold methanol, re-dissolved in 1 mL buffer H and incubated with 60 μL streptavidin beads (prewashed with buffer I, GE Healthcare) for 2 h at room temperature with gentle shaking. Samples were centrifuged at 4,000 × g for 2 min and the supernatant containing unbound proteins were pipetted out. The streptavidin beads were then washed sequentially with buffer I, buffer J, and 100 mM ABC, each for three times. Finally, the biotin labelled proteins were cleaved and eluted from the beads using 60 μL of 25 mM sodium dithionite in 20 mM ABC. The eluted proteins were mixed with 2 × Laemmli buffer and analyzed by SDS-PAGE. Presence of RCC1 was detected and quantified by western blot as described below in Section 2.4.12. Standard curve of signal volumes versus protein amounts were generated using a dilution series of purified RCC1 from *E. coli* with known concentrations.

2.3.12 Western blot

Equal volumes of cell lysates or protein samples were separated in a 12% tris-glycine SDS-PAGE and then transferred to a supported nitrocellulose membrane (Bio-Rad). The membranes were incubated with rabbit anti-RCC1 antibody (1:1000, Abcam), rabbit anti-NTMT1 antibody (1:1000, Abcam), or streptavidin-HRP (1:3000, Abcam) overnight at 4 ˚C. After washing with TBS-T (0.1% Tween 20 in 1 × TBS) buffer, membranes for RCC1 or NTMT1 detection were incubated with goat anti-rabbit antibody (1:5000, Abcam) by gentle shaking for 1 h at room temperature. Upon similar washing, bands were detected using SuperSignal West Pico PLUS chemiluminescent substrate (Thermo Fisher Scientific) following manufacturer's protocol. The bands were visualized using Amersham Imager 600 (GE Healthcare).

2.3.13 NTMT1 knockout using CRISPR-Cas9 system

Procedures for knockout of NTMT1 gene in HEK293FT cell line were similar to the published nature protocal[33]. Briefly, 1 μg of plasmids PX459-NTMT1$_{sgRNA1}$ and 1 μg of PX459-NTMT1$_{sgRNA2}$ were mixed and transfected to HEK293FT seeded in a 6-well plate using transfection reagent LipoD293 (SignaGen Laboratories) and Opti-MEM® I Reduced Serum Medium (ThermoFisher) when cell confluency reached to 50%-70%. After 36h, transfected cells were treated with 3 μg/ml puromycin for 1 week. Cells were then dissociated with TrypLE™ Express Enzyme (ThermoFisher), counted and serially diluted in DMEM supplemented with 20% FBS to a final cell density of 50-100 cells per 15 ml medium to reduce the possibility of having multiple cells in one well. The diluted cells were plated to two 96-well plates and cultured for 2 weeks. After passaging to 24-well pate, cells were screened by western blot using anti-NTMT1 antibody (Abcam). For those cells without the expression of NTMT1, their genome DNAs were

extracted with Tissue DNA Kit (Omega), and the exon1 of NTMT1 were amplified and sequenced. After confirmed by DNA sequencing, the NTMT1 KO cell lines were frozen in liquid N_2.

2.3.14 Activity-based NTMT1 substrate pull-down and proteomic identification of NTMT1 targets in NTMT1 KO cell line

The culture condition for NTMT1 KO cells was described in Section 2.4.3. At 70%–90% confluence stage, cells (4 of 175 cm^2 flasks) were harvested and lysed in 9 mL cold buffer K by sonication for 8 min (5-second on and 3-second off) at 40% amplitude. Cell lysates were centrifuged at 15,000 × g for 15 min at 4 ˚C to remove cell debris. The supernatant was incubated with either 50 μM Hey-SAM or SAM (negative control) in the presence of 3 μM wt NTMT1 for 4 h at 37 ˚C. Four-time volumes of cold methanol was added into each sample and kept at −80 ˚C overnight to precipitate proteins. The samples were centrifuged at 7,197 × g for 30 min and the supernatant was removed. The protein pellets were air-dried for 15 min and re-dissolved in 4.0 mL buffer F. Click chemistry and subsequent treatment were performed in the same way as described above in Section 2.4.11. As a positive control, the presence of RCC1 was detected by western blot as described above in Section 2.4.12.

The lanes of Hey-SAM and SAM (negative control) treated were excised and digested with trypsin following the procedure described in section 2.4.9. For LC-MS analysis, protein digestion products were re-dissolved in water containing 1% formic acid, filtered with Ultrafree PTFE membrane (Millipore) and injected to a M class UPLC-XEVO G2-XS QTof (Waters) equipment. Peptides were firstly cleaned by a trapping column (ACQUITY UPLC M-Class Symmetry C18 Trap Column, 100 Å, 5 μm, 300 μm × 50 mm, Waters) and separated on an analytical column (ACQUITY UPLC M-Class HSS T3 Column, 1.8 μm, 300 μm × 100 mm, Waters) with a flow rate of 8 ul/min. Mobile phase A and B consisted of 0.1% formic acid in water and acetonitrile,

respectively. The gradient for separation is 1 % to 40 % of mobile phase B for 45 min. The mass spectrometer was operated in positive polarity, sensitivity mode, and an MS^E mode with a ramp collision energy ranging from 20 to 45 V. Raw data were analyzed by Progenesis QI for Proteomics (Waters). The statistical and motif analysis were described in Section 2.2.11.

2.3.15 Targets validation *via* MS-based methylation assay

MS-based methylation assays of RS14 protein, PB1 peptide, SPD2B peptide, DDX60L peptide and histone H2A1B peptide were performed similarly to the previous described procedure in Section 2.2.6 except that the MS data were acquired on a M class UPLC coupled with XEVO G2-XS QTof (Waters). Raw data of peptides were processed by MassLynx (Waters) and raw data of intact protein mass were analyzed by BiopharmaLynx (Waters).

2.4 References

1. Zhong, R.; Morrison, W. H.; Himmelsbach, D. S.; Poole, F. L.; Ye, Z.-H. Essential role of caffeoyl coenzyme A O-methyltransferase in lignin biosynthesis in woody poplar plants. *Plant Physiol.* **2000,** *124* (2), 563-578.

2. Khan, M. Z.; Nawaz, W. The emerging roles of human trace amines and human trace amine-associated receptors (hTAARs) in central nervous system. *Biomed. Pharmacother.* **2016,** *83*, 439-449.

3. Greer, E. L.; Shi, Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* **2012,** *13* (5), 343-357.

4. Chen, T.; Muratore, T. L.; Schaner-Tooley, C. E.; Shabanowitz, J.; Hunt, D. F.; Macara, I. G. N-terminal α-methylation of RCC1 is necessary for stable chromatin association and normal mitosis. *Nat. Cell Biol.* **2007,** *9* (5), 596-603.

5. Cai, Q.; Fu, L.; Wang, Z.; Gan, N.; Dai, X.; Wang, Y. α-N-methylation of damaged DNA-binding protein 2 (DDB2) and its function in nucleotide excision repair. *J. Biol. Chem.* **2014**, *289* (23), 16046-16056.

6. Dai, X.; Otake, K.; You, C.; Cai, Q.; Wang, Z.; Masumoto, H.; Wang, Y. Identification of novel α-N-methylation of CENP-B that regulates its binding to the centromeric DNA. *J. Proteome Res.* **2013,** *12* (9), 4167-4175.

7. Bonsignore, L. A.; Butler, J. S.; Klinge, C. M.; Tooley, C. E. S. Loss of the N-terminal methyltransferase NRMT1 increases sensitivity to DNA damage and promotes mammary oncogenesis. *Oncotarget* **2015,** *6* (14), 12248-12263.

8. Shields, K. M.; Tooley, J. G.; Petkowski, J. J.; Wilkey, D. W.; Garbett, N. C.; Merchant, M. L.; Cheng, A.; Schaner Tooley, C. E. Select human cancer mutants of NRMT1 alter its catalytic activity and decrease N‐terminal trimethylation. *Protein Sci.* **2017,** *26* (8), 1639-1652.

9. Tooley, C. E. S.; Petkowski, J. J.; Muratore-Schroeder, T. L.; Balsbaugh, J. L.; Shabanowitz, J.; Sabat, M.; Minor, W.; Hunt, D. F.; Macara, I. G. NRMT is an α-N-methyltransferase that methylates RCC1 and retinoblastoma protein. *Nature* **2010,** *466* (7310), 1125-1128.

10. Dai, X.; Rulten, S. L.; You, C.; Caldecott, K. W.; Wang, Y. Identification and functional characterizations of N-terminal α-N-methylation and phosphorylation of serine 461 in human poly (ADP-ribose) polymerase 3. *J. Proteome Res.* **2015,** *14* (6), 2575-2582.

11. Sathyan, K. M.; Fachinetti, D.; Foltz, D. R. α-amino trimethylation of CENP-A by NRMT is required for full recruitment of the centromere. *Nat. Commun.* **2017,** *8*, 14678-14692.

12. Petkowski, J. J.; Schaner Tooley, C. E.; Anderson, L. C.; Shumilin, I. A.; Balsbaugh, J. L.; Shabanowitz, J.; Hunt, D. F.; Minor, W.; Macara, I. G. Substrate specificity of mammalian N-terminal α-amino methyltransferase NRMT. *Biochemistry* **2012,** *51* (30), 5942-5950.

13. Wang, R.; Zheng, W.; Yu, H.; Deng, H.; Luo, M. Labeling substrates of protein arginine methyltransferase with engineered enzymes and matched S-adenosyl-L-methionine analogues. *J. Am. Chem. Soc.* **2011,** *133* (20), 7648-7651.

14. Guo, H.; Wang, R.; Zheng, W.; Chen, Y.; Blum, G.; Deng, H.; Luo, M. Profiling substrates of protein arginine N-methyltransferase 3 with S-adenosyl-L-methionine analogues. *ACS Chem. Biol.* **2013,** *9* (2), 476-484.

15. Islam, K.; Chen, Y.; Wu, H.; Bothwell, I. R.; Blum, G. J.; Zeng, H.; Dong, A.; Zheng, W.; Min, J.; Deng, H. Defining efficient enzyme–cofactor pairs for bioorthogonal profiling of protein methylation. *P. Natl. Acad. Sci. USA* **2013,** *110* (42), 16778-16783.

16. Wu, R.; Yue, Y.; Zheng, X.; Li, H. Molecular basis for histone N-terminal methylation by NRMT1. *Genes Dev.* **2015,** *29* (22), 2337-2342.

17. Klimašauskas, S.; Weinhold, E. A new tool for biotechnology: AdoMet-dependent methyltransferases. *Trends Biotechnol.* **2007,** *25* (3), 99-104.

18. Richardson, S. L.; Mao, Y.; Zhang, G.; Hanjra, P.; Peterson, D. L.; Huang, R. Kinetic mechanism of protein N-terminal methyltransferase 1. *J. Biol. Chem.* **2015,** *290* (18), 11601-11610.

19. Pu, D.; Cassady, C. J. Negative ion dissociation of peptides containing hydroxyl side chains. *Rapid Commun. Mass Spectrom.* **2008,** *22* (2), 91-100.

20. Wingfield, P. T. N‐Terminal Methionine Processing. *Curr. Protoc Protein Sci.* **2017,** *88* (1), 6.14. 1-6.14. 3.

21. Sherman, F.; Stewart, J. W.; Tsunasawa, S. Methionine or not methionine at the beginning of a protein. *Bioessays* **1985,** *3* (1), 27-31.

22. Dong, C.; Mao, Y.; Tempel, W.; Qin, S.; Li, L.; Loppnau, P.; Huang, R.; Min, J. Structural basis for substrate recognition by the human N-terminal methyltransferase 1. *Genes Dev.* **2015,** *29* (22), 2343-2348.

23. Islam, K.; Zheng, W.; Yu, H.; Deng, H.; Luo, M. Expanding cofactor repertoire of protein lysine methyltransferase for substrate labeling. *ACS Chem. Biol.* **2011,** *6* (7), 679-684.

24. Bothwell, I. R.; Islam, K.; Chen, Y.; Zheng, W.; Blum, G.; Deng, H.; Luo, M. Se-adenosyl-L-selenomethionine cofactor analogue as a reporter of protein methylation. *J. Am. Chem. Soc.* **2012,** *134* (36), 14905-14912.

25. Chen, D.-H.; Wu, K.-T.; Hung, C.-J.; Hsieh, M.; Li, C. Effects of adenosine dialdehyde treatment on in vitro and in vivo stable protein methylation in HeLa cells. *J. Biochem.* **2004,** *136* (3), 371-376.

26. Blum, G.; Islam, K.; Luo, M. Using azido analogue of S-Adenosyl-L-methionine for Bioorthogonal Profiling of Protein Methylation (BPPM). *Curr. Protoc. Chem. Biol.* **2013,** *5*, 45-66.

27. Liteplo, R. G. DNA (cytosine) methylation in murine and human tumor cell lines treated with S-adenosylhomocysteine hydrolase inhibitors. *Cancer Lett.* **1988,** *39* (3), 319-327.

28. Esse, R.; Rocha, M. S.; Barroso, M.; Florindo, C.; Teerlink, T.; Kok, R. M.; Smulders, Y. M.; Rivera, I.; Leandro, P.; Koolwijk, P. Protein arginine methylation is more prone to inhibition by S-adenosylhomocysteine than DNA methylation in vascular endothelial cells. *PLoS One* **2013,** *8* (2), e55483.

29. Chang, H.-H.; Hu, H.-H.; Lee, Y.-J.; Wei, H.-M.; Fan-June, M.-C.; Hsu, T.-C.; Tsay, G. J.; Li, C. Proteomic analyses and identification of arginine methylated proteins differentially recognized by autosera from anti-Sm positive SLE patients. *J. Biomed. Sci.* **2013,** *20* (1), 27-36.

30. Blum, G.; Islam, K.; Luo, M. Bioorthogonal Profiling of Protein Methylation (BPPM) Using an Azido Analog of S‐Adenosyl‐L‐Methionine. *Curr. Protoc. Chem. Biol.* **2013**, 45-66.

31. Schwerk, C.; Schulze-Osthoff, K. Methyltransferase inhibition induces p53-dependent apoptosis and a novel form of cell death. *Oncogene* **2005,** *24* (47), 7002-7011.

32. Heit, R.; Rattner, J. B.; Chan, G. K.; Hendzel, M. J. G2 histone methylation is required for the proper segregation of chromosomes. *J. Cell Sci.* **2009,** *122* (16), 2957-2968.

33. Ran, F. A.; Hsu, P. D.; Wright, J.; Agarwala, V.; Scott, D. A.; Zhang, F. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **2013,** *8* (11), 2281-2308.

34. Matsuzawa, A.; Kanno, S.-i.; Nakayama, M.; Mochiduki, H.; Wei, L.; Shimaoka, T.; Furukawa, Y.; Kato, K.; Shibata, S.; Yasui, A. The BRCA1/BARD1-interacting protein OLA1 functions in centrosome regulation. *Mol. Cell* **2014,** *53* (1), 101-114.

35. Balasingam, N.; Brandon, H.; Ross, J. A.; Wieden, H.-J.; Thakor, N. Cellular roles of the human Obg-like ATPase 1 (hOLA1) and its bacterial homologue, YchF. *Biochem. Cell Biol.* **2019,** 1-11.

36. Mao, R.; Rubio, V.; Chen, H.; Bai, L.; Mansour, O.; Shi, Z.-Z. OLA1 protects cells in heat shock by stabilizing HSP70. *Cell Death & Dis.* **2013,** *4* (2), e491.

37. Ding, Z.; Liu, Y.; Rubio, V.; He, J.; Minze, L. J.; Shi, Z.-Z. OLA1, a translational regulator of p21, maintains optimal cell proliferation necessary for developmental progression. *Mol. Cell. Biol.* **2016,** *36* (20), 2568-2582.

38. Jeyabal, P. V.; Rubio, V.; Chen, H.; Zhang, J.; Shi, Z.-Z. Regulation of cell-matrix adhesion by OLA1, the Obg-like ATPase 1. *Biochem. Biophys. Res. Commun.* **2014,** *444* (4), 568-574.

39. Villar-Garea, A.; Forne, I.; Vetter, I.; Kremmer, E.; Thomae, A.; Imhof, A. Developmental regulation of N-terminal H2B methylation in Drosophila melanogaster. *Nucleic Acids Res.* **2011,** *40* (4), 1536-1549.

40. Chandrasekaran, R.; Thompson, M. Polybromo-1-bromodomains bind histone H3 at specific acetyl-lysine positions. *Biochem. Biophys. Res. Commun.* **2007,** *355* (3), 661-666.

41. Varela, I.; Tarpey, P.; Raine, K.; Huang, D.; Ong, C. K.; Stephens, P.; Davies, H.; Jones, D.; Lin, M.-L.; Teague, J. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* **2011,** *469* (7331), 539-542.

42. Makde, R. D.; Tan, S. Strategies for crystallizing a chromatin protein in complex with the nucleosome core particle. *Anal. Biochem.* **2013,** *442* (2), 138-145.

43. Coin, I.; Beyermann, M.; Bienert, M. Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. *Nat. Protoc.* **2007,** *2* (12), 3247-3256.

44. Shevchenko, A.; Tomas, H.; Havli, J.; Olsen, J. V.; Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006,** *1* (6), 2856-2860.

45. Sorenson, A. E.; Askin, S. P.; Schaeffer, P. M. In-gel detection of biotin–protein conjugates with a green fluorescent streptavidin probe. *Anal. Methods* **2015,** *7* (5), 2087-2092.

Chapter 3 - Investigation of hOLA1 methylation *in vitro* and in human cells

3.1 Introduction

3.1.1 Obg-like ATPase 1 (OLA1)

Nucleotide binding proteins (e.g., kinases, ATPases, GTPases) are one of the most important families of proteins which play vital roles in multiple cellular processes including cation transport, protein translation, signal transduction, mRNA transcription, microRNA biogenesis, cell motility, DNA replication, mitosis and cytoskeleton organization[1-8]. Obg-like ATPase 1 (OLA1) belongs to the family of P-loop GTPases, however, preferentially binds and hydrolyzes ATP over GTP *in vitro*[9]. Koller-Eichhorn et al. solved the crystal structure of hOLA1 (PDB: 2OHF) and explained that the hydrogen bonding between the carbonyl group of Leu 231 and the exocyclic amino group in ATP was responsible for its preferred binding with ATP[9].

OLA1 was also named as DNA damage-regulated overexpressed in cancer 45 (DOC45) because it was strongly downregulated by DNA damage-inducing chemicals (e.g., doxorubicin, etoposide) and UV radiation[10]. Sun et al. reported that OLA1 was overexpression in several different human cancers (e.g., colorectal, rectum, stomach, lung, ovary, uterus) and oncogenic mutant of H-Ras or TC21 significantly increased the expression of OLA1[10]. Recent studies indicated that OLA1 have multiple roles in several cellular processes including centrosome amplification, antioxidant response, heat shock response and cell growth[10-15].

3.1.2 OLA1 regulates centrosome amplification

OLA1 was reported to directly interact with breast cancer type 1 susceptibility protein (BRCA1), BRCA1-associated RING domain protein (BARD1) and γ-tubulin (a centrosome

marker)[11, 16]. Knockdown of OLA1 resulted in abnormal amplification of centrosome in breast cancer cell lines (e.g., Hs578T) which was rescued by overexpression of wt OLA1 but not the OLA1 E168Q mutant[11]. OLA1 E168Q mutant failed to bind to BRCA1, and therefore, OLA1 may regulate centrosome amplification by binding with BRCA1 protein[11]. Moreover, Yoshino et al. reported that overexpression of wt OLA1 in breast cancer cell line Hs578T caused abnormal centrosome amplification while the overexpression of OLA1 E168Q mutant did not result in extra amplification of centrosome[12]. These results suggest that either downregulation or upregulation of OLA1/BRCA1 complex would affect the centrosome amplification and keeping a regular expression level of OLA1 is necessary for maintaining the normal numbers of centrosome.

3.1.3 OLA1 functions as an antioxidant suppressor for oxidative stress

Oxidative stress refers to the oxygen toxicity by oxidants (e.g., hydrogen peroxide)[17]. Zhang et al. reported that HeLa cells with OLA1 knockdown by siRNAs exhibited enhanced resistance to oxidative agents (e.g., hydrogen peroxide, tert-butyl hydroperoxide), while the overexpression of OLA1 in HeLa cells increased cell cytotoxicity induced by tert-butyl hydroperoxide[13]. Moreover, lower amount of reactive oxygen species (ROS) were observed in OLA1-knockdown HeLa cells compared to normal HeLa cells after the treatment of tert-butyl hydroperoxide[13]. In addition, the concentration of glutathione (GSH), a vital chemical in protecting cells from oxidative toxicity, remained a high level in OLA1-knockdown HeLa cells but dramatically decreased in normal HeLa cells after treatment with tert-butyl hydroperoxide[13]. All these results demonstrate that OLA1 regulates cellular oxidative stress by acting as an antioxidant suppressor[13].

3.1.4 OLA1 regulates the stability of heat shock protein 70 (HSP70)

Heat shock protein 70 (HSP70) belongs to the family of heat shock proteins and functions as a molecular chaperone to assist the folding of newly synthesized proteins, unfolding and refolding

of damaged or aggregated proteins[18]. Mao et al. reported that knockdown or knockout of OLA1 gene in cultured cells (e.g., MDA-MB-231, mouse embryonic fibroblast cells (MEFs)) enhanced the cells' sensitivity towards heat shock treatment (42 ℃ and 45 ℃) compared to normal cells[14]. Moreover, overexpression of OLA1 in HEK293T cells increased cell survival after heat shock treatment[14]. These results indicate that OLA1 act as a regulator for heat shock response. Further, the expression level of HSP70 was dramatically decreased in either OLA1 KO MEFs or OLA1 knockdown human cell lines (e.g., MDA-MB-231, HEK293T), demonstrating that OLA1 regulated the expression of HSP70[14]. Overexpression of OLA1 in either MDA-MB-231 or HEK293T cells caused an increased expression level of HSP70[14]. Further study indicated that downregulation of OLA1 did not have a significant influence on the transcription of HSP70[14]. To investigate the regulation mechanism, co-IP of OLA1 and C-terminal domain of HSP70 was performed and OLA1 was found to bind with HSP70[14]. Further study revealed that the interaction of OLA1 with HSP70 blocked the binding site for CHIP (carboxy terminus of HSP70-binding protein), impeded the CHIP induced ubiquitination of HSP70 and therefore stabilized HSP70[14, 19]. These results suggest that OLA1 regulates the stability of HSP70 *via* a post translational modification mechanism[14].

3.1.5 OLA1 regulates cell growth

Overexpression of OLA1 is found in many types of cancers, thus OLA1 has been proposed to influence cell growth[10]. Sun et al. reported that OLA1 knockdown in two cultured human colon cancer cell lines (RKO and HT29) had a negative effect on cellular proliferation[10]. In contrast to this *in vitro* study, xenograft experiments performed by Chen et al. showed that OLA1-knockdown MDA-MB231 cells grew into larger and heavier tumors compared to normal cells when they were inoculated into nude mice[15]. Terminal deoxynucleotidyl transferase dUTP nick end labeling

(TUNEL) assay indicated that OLA1 knockdown cells had a lower apoptotic rate. Moreover, the downregulation of CCAAT/enhancer-binding protein-homologous protein (CHOP) which plays critical role in regulating apoptosis[20-21], was observed in OLA1-knockdown tumors[15]. All these results suggest that the accelerative tumor growth for OLA1 knockdown breast cancer cells results from a net decreased cell death with a normal cell growth but a low apoptosis[15].

3.1.6 Phenotypes of OLA1 knockout mice

To study the physiological functions of OLA1, Ding et al. generated OLA1 knockout mice[22]. Though the heterozygous mice ($Ola1^{+/-}$) looked similar to the normal mice, 80% the homozygous mice ($Ola^{-/-}$) died within 24 h of birth[22]. The survived homozygous mice had smaller body size, decreased body weight and delayed cell proliferation compared to control mice[22]. The upregulation of protein cyclin-dependent kinase inhibitor 1 (p21) in OLA1 knockout MEFs was observed and p21 knockout in $Ola1^{-/-}$ mice ($p21^{-/-}/Ola1^{-/-}$) partially rescued the growth defects[22]. Protein p21 is known as an inhibitor for cyclin-dependent kinases mediated protein phosphorylation and may arrest the cell cycle[23-24]. These results indicate that OLA1 may be involved in regulating p21 expression. The accumulation of p21 in $Ola1^{-/-}$ mice inhibited the cell proliferation, which partially explained why the $Ola1^{-/-}$ mice exhibited the delayed progression of development[22].

3.1.7 Purpose of this study

The activity-based substrate profiling of NTMT1 (Chapter 2) revealed OLA1 as one of the putative targets (Table 2.3). In addition, OLA1 contains an N-terminal Met-Pro-Pro-Lys-Lys motif, which was predicted to be the best one for NTMT1 substrates[25]. Previous studies demonstrated that NTMT1 was mainly localized in cell nucleus[26-27]. Many of the identified substrates (e.g., DDB2, RCC1) have the same subcellular localization as NTMT1 in the nucleus[28-29]. Human OLA1 was reported to localize in both cytoplasm and nucleus but with a major

distribution in cytoplasm[10]. Moreover, treatment with leptomycin B, a nuclear export inhibitor, significantly enhanced the nucleus distribution of endogenous OLA1 in MCF-7 cells[10]. These results indicate that OLA1 exhibited a unique nucleocytoplasmic subcellular localization[10]. Taken together, we hypothesized that OLA1 would be a candidate substrate for NTMT1 and therefore we investigated its methylation *in vitro* and *in vivo*.

3.2 Results and discussion

3.2.1 Methylation of OLA1 peptide by NTMT1

To investigate whether OLA1 would be methylated by wt NTMT1, we firstly synthesized its peptide equivalent (OLA1 peptide, Fig. 3.1A) which contains the first 10 N-terminal residues without the initial methionine and used it as the substrate for methylation reaction. The reaction products, methylated peptide and SAH, were analyzed by MS and HPLC, respectively. MS analysis showed that OLA1 peptide methylation was enzyme-catalyzed (Fig. 3.1B & 3.1C) and OLA1 peptide was fully converted into dimethylated product (Fig. 3.1C) after incubating with SAM and 1 μM NTMT1 for 30 min at 37 ℃. SAH release was monitored by HPLC (Fig. 3.1D) and quantified using adenosine as an external reference. The initial rate for the reaction between 100 μM SAM and 40 μM OLA1 peptide with 1 μM NTMT1 was 2.17 μM/min (Fig. 1E) which is higher than the value for RCC1 peptide (0.58 min) at the same condition. This result is in accordance with the previous reported ITC results which indicated that Pro-Pro-Lys was a better motif than Ser-Pro-Lys[25].

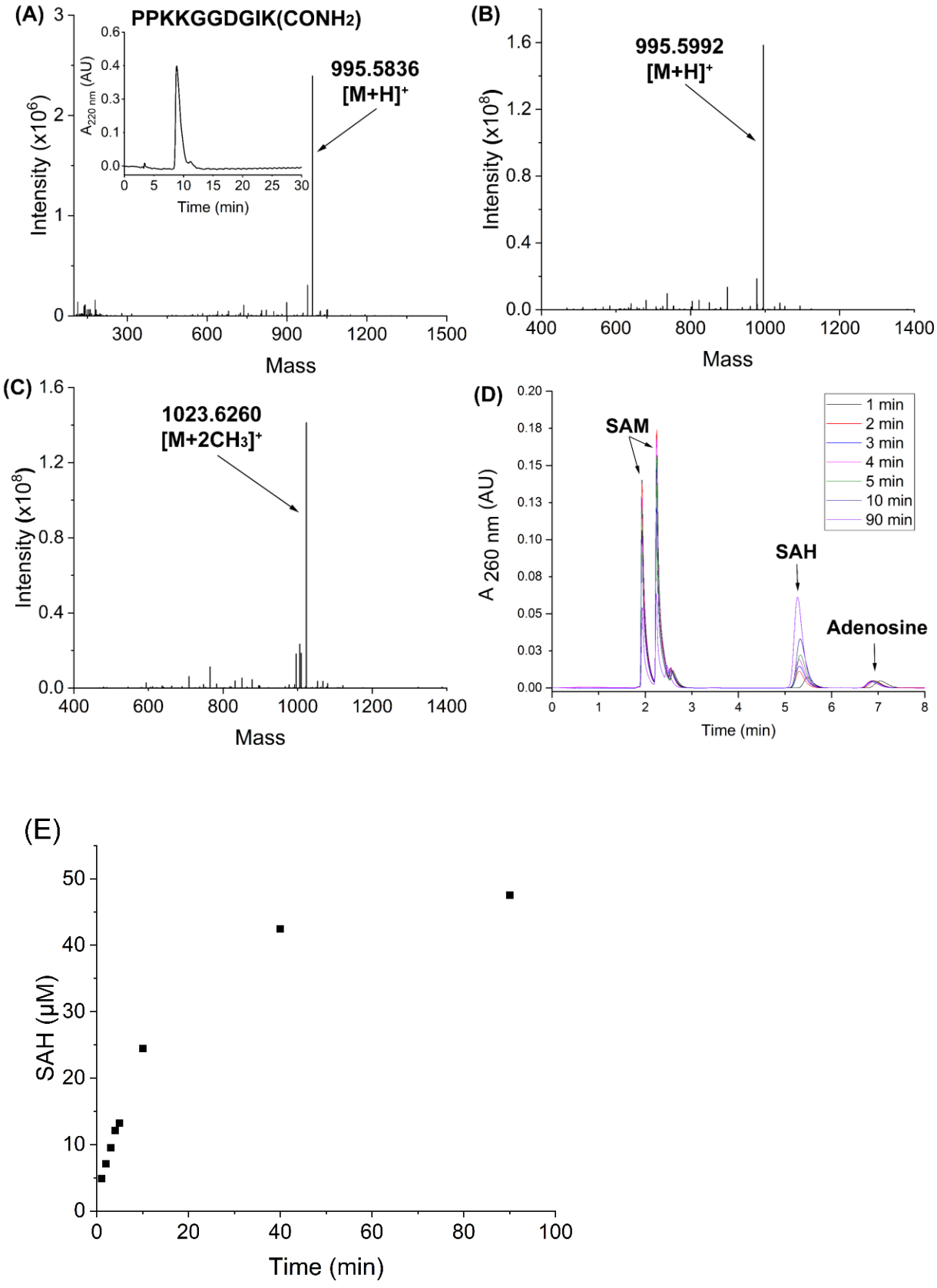

(A)
PPKKGGDGIK(CONH2)
995.5836
[M+H]+
Intensity (x10^6)
Mass
A220 nm (AU)
Time (min)
(B)
995.5992
[M+H]+
Intensity (x10^8)
Mass
(C)
1023.6260
[M+2CH3]+
Intensity (x10^8)
Mass
(D)
SAM
SAH
Adenosine
1 min
2 min
3 min
4 min
5 min
10 min
90 min
A 260 nm (AU)
Time (min)
(E)
SAH (μM)
Time (min)

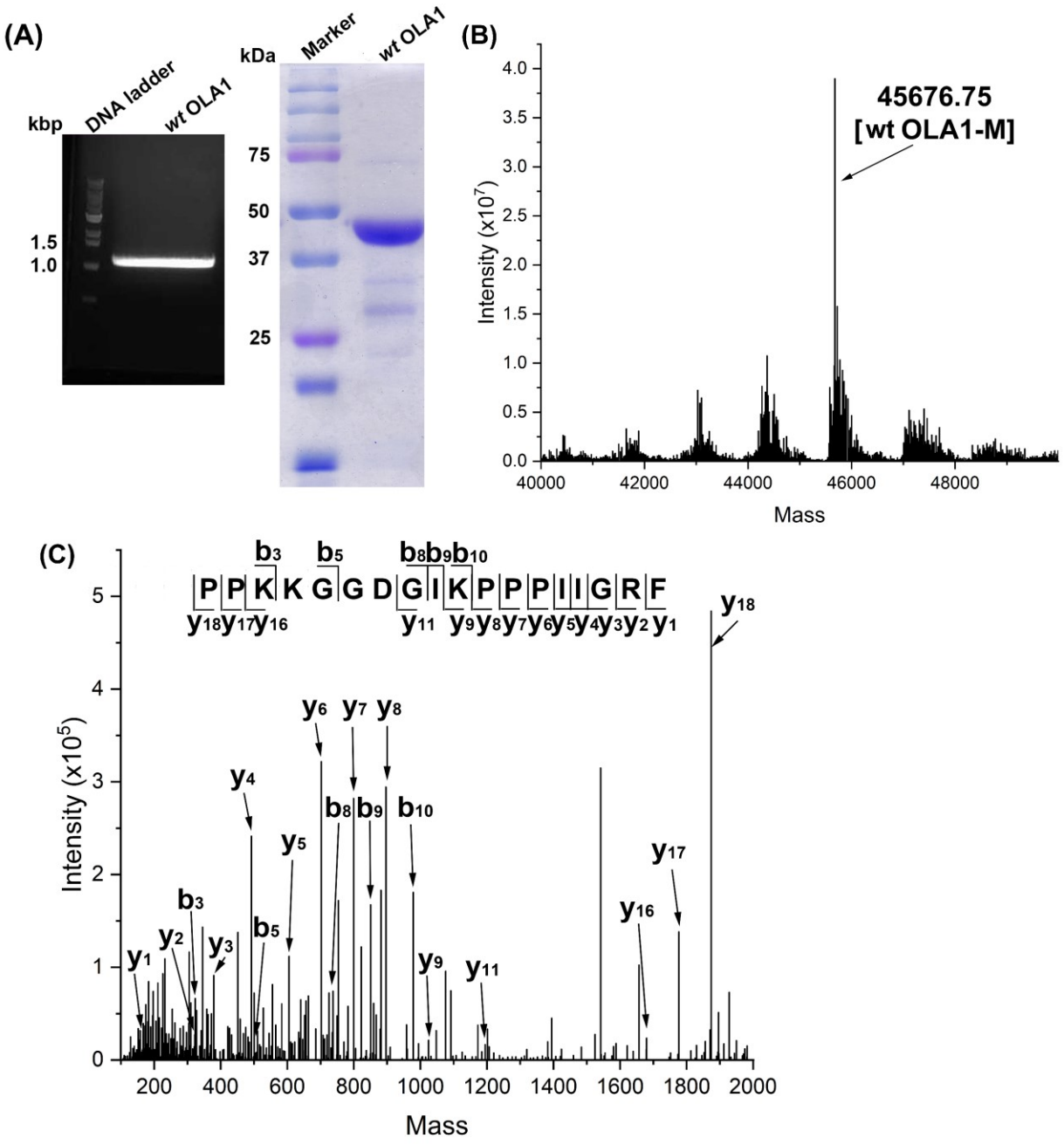
Figure 3.1 MS and HPLC analysis of OLA1 peptide methylation.

All reactions were carried out at 37 ℃. (A) Mass spectrum and HPLC purity (inset) of the synthesized 10-residue OLA1 peptide (calculated $[M + H]^+$: 995.6002, detected: 995.5836). (B) Mass spectrum of the reaction between 20 μM OLA1 peptide and 200 μM SAM in the absence of NTMT1 (calculated $[M + H]^+$: 995.6002, detected: 995.5992). (C) Mass spectrum of the reaction between 20 μM OLA1 peptide and 200 μM of SAM with 1 μM NTMT1 (calculated $[M + 2CH_3]^+$: 1023.6315, detected: 1023.6260). (D) HPLC monitoring of SAH release for the reaction between 40 μM of OLA1 peptide and 100 μM of SAM with 1 μM NTMT1 at different time intervals. (E) Initial rate of SAH release for the reaction described in (D).

3.2.2 Methylation of OLA1 protein *in vitro* by NTMT1

To examine whether NTMT1 could catalyze OLA1 methylation at the protein level, recombinant wt OLA1 (Fig. 3.2 A) with a His_6 tag on its C-terminus was expressed and purified from *E. coli.* Removal of the initial methionine residue is a common modification in *E. coli*[30-31]. However, the reported catalytic activity of *E. coli* methionine aminopeptidase 1 (MAP1) towards substrate peptides containing a proline at the third position was very low[32]. As described in chapter 2, the initial methionine of RCC1 can be partially or completely removed when expressed in *E. coli*, and thus we decided to check the intact protein mass of OLA1 before performing the methylation assay. Here, the MS results indicated that the initial methionine of wt hOLA1 expressed in *E. coli* was completely cleaved (Fig. 3.2B). Protein in-gel chymotrypsin digestion and subsequent LC-MS analysis confirmed the initial methionine removal of OLA1 (Fig. 3.2C, retention time: 20 min). Purified OLA1 was directly used as a substrate for methylation assay. Dimethylated OLA1 protein was only detected in the sample containing both SAM and NTMT1, indicating that OLA1 methylation is enzyme-catalyzed (Fig. 3.2D & 3.2E). Note that the dimethylation of proline would introduce a positive charge for OLA1, which would make the deconvoluted intact mass processed by BiopharmaLynx software (Waters) one Dalton lower than

the theoretical mass. In-gel chymotrypsin digestion of methylated OLA1 and subsequent analysis by LC-MS verified this protein N-terminal methylation (Fig. 3.2F). To further confirm that the mass shift was caused by NTMT1 and SAM, the SAH release was monitored and quantified by HPLC (Fig. 3.2D & 3.2E inset). In consistent with MS results, 31.4 μM of SAH were formed for 20 μM of wt OLA1, which indicated that most of wt OLA1 had been dimethylated (Fig. 3.2E inset).


(A)
kbp
DNA ladder
wt OLA1
1.5
1.0
kDa
Marker
wt OLA1
75
50
37
25
(B)
45676.75
[wt OLA1-M]
Intensity (x10^7)
Mass
(C)
P P K K G G D G I K P P P I I G R F
y18 y17 y16 y11 y9 y8 y7 y6 y5 y4 y3 y2 y1
b3 b5 b8 b9 b10
Intensity (x10^5)
Mass

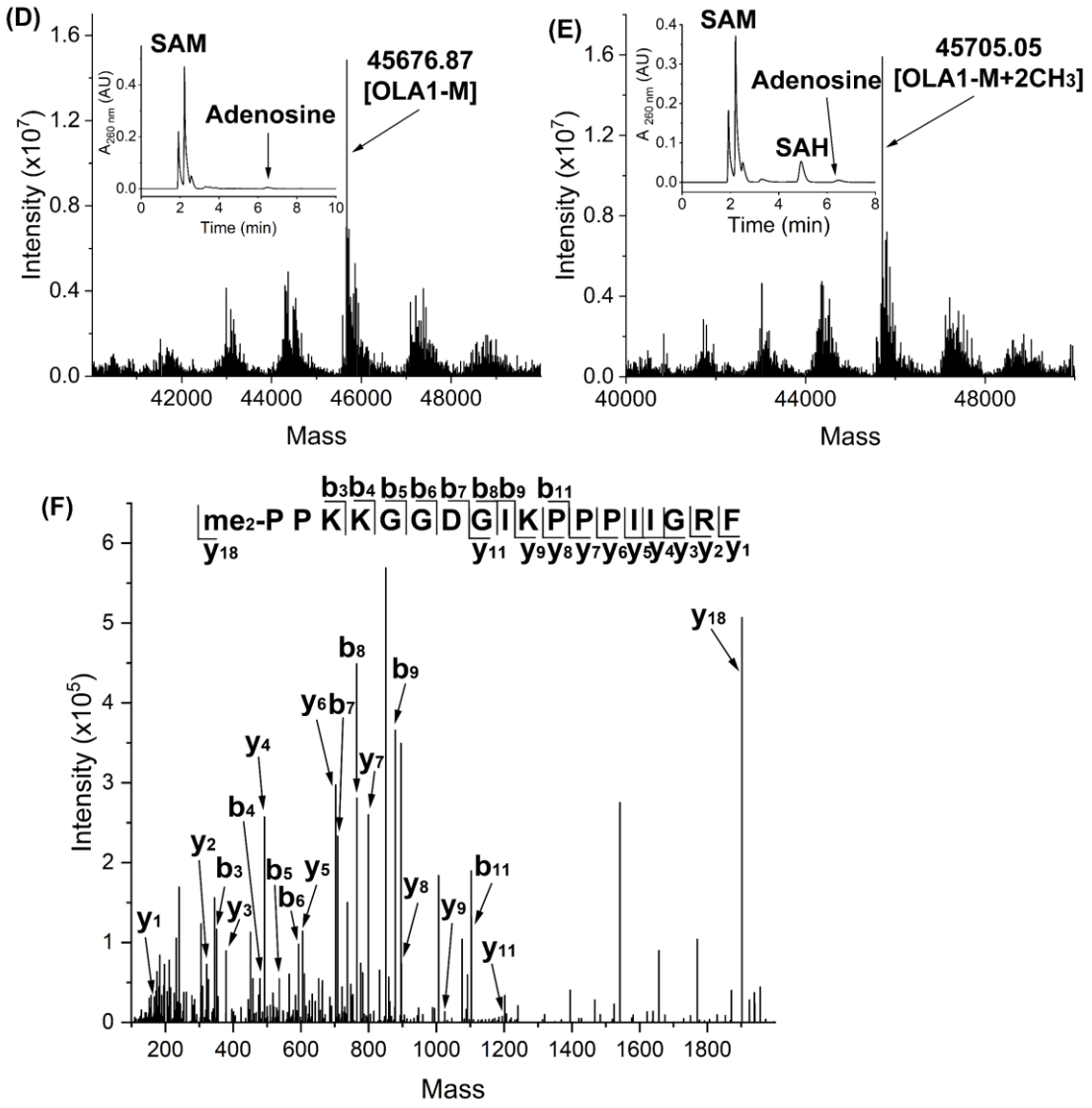


Figure 3.2 MS and HPLC analysis of OLA1 protein methylation.

All reactions were carried out at 37 ℃. (A) Plasmid construction (left panel: amplification of OLA1 gene from human cDNA) and protein purification (wt OLA1: 45 kDa). (B) Intact protein mass of wt hOLA1 (calculated MW for deletion of initial methionine [wt hOLA1 - M]: 45677.5 Da, detected MW: 45676.75 Da). (C) MS^E spectrum of mapping the wt OLA1 N-terminal peptide (calculated precursor ion: 1874.1011, detected: 1874.0760). (D) Intact protein mass of wt hOLA1 after incubating with SAM (detected MW: 45676.87 Da) and HPLC monitoring of SAH release (inset). (E) Intact protein mass of wt hOLA1 after incubation with SAM and NTMT1 (calculated [wt hOLA1 – M + $2CH_3$]: 45706.56Da, detected MW: 45705.05 Da) and HPLC monitoring of SAH release (inset). (F) MS^E spectrum of mapping the dimethylated wt OLA1 N-terminal peptide (calculated precursor ion: 1902.1324, detected: 1902.1045).

3.2.3 The importance of Pro-Pro-Lys motif for OLA1 N-terminal methylation

Considering the importance of lysine residue in the Xaa-Pro-Lys motif and its glutamine mutant dramatically affected the methylation of RCC1[28] and DDB2[29], we decided to examine its importance for OLA1. The peptide containing the lysine to glutamine mutation (OLA1 K4Q peptide) was synthesized and characterized by HPLC (Fig. 2.3A inset) and MS (Fig. 2.3A). MS analysis of methylation assay showed that NTMT1 was not able to catalyze the methylation of OLA1 K4Q peptide (Fig. 2.3B & 2.3C). Compared to PPK peptide, the SAH release for the mixture of 100 μM SAM, 40 μM PPQ peptide and 1 μM NTMT1 was negligible (Fig. 2.3D) and the initial reaction rate was only 0.017 μM/min (Fig. 2.3E). These results indicate that lysine residue in the Pro-Pro-Lys motif is vital for OLA1 N-terminal methylation.

Next, we examined whether Lys 4 is important for OLA1 methylation at the protein level. The K4Q mutant (Fig. 2.3F, inset) was generated, expressed and purified from *E. coli*. Unlike wt OLA1, intact protein mass showed that the majority of K4Q mutant purified from *E. coli* lacked the initial methionine together with the proline residue (Fig. 2.3F). The N-terminal Pro-Pro-Lys motif was destroyed and the newly formed Pro-Gln-Lys N-terminus was not methylated according to the protein intact mass result (Fig. 2.3G). In consistent with intact protein mass result, SAH release was not detected by HPLC (Fig. 2.3G, inset). Here we conclude that the Pro-Pro-Lys motif is important for the N-terminal methylation of OLA1.

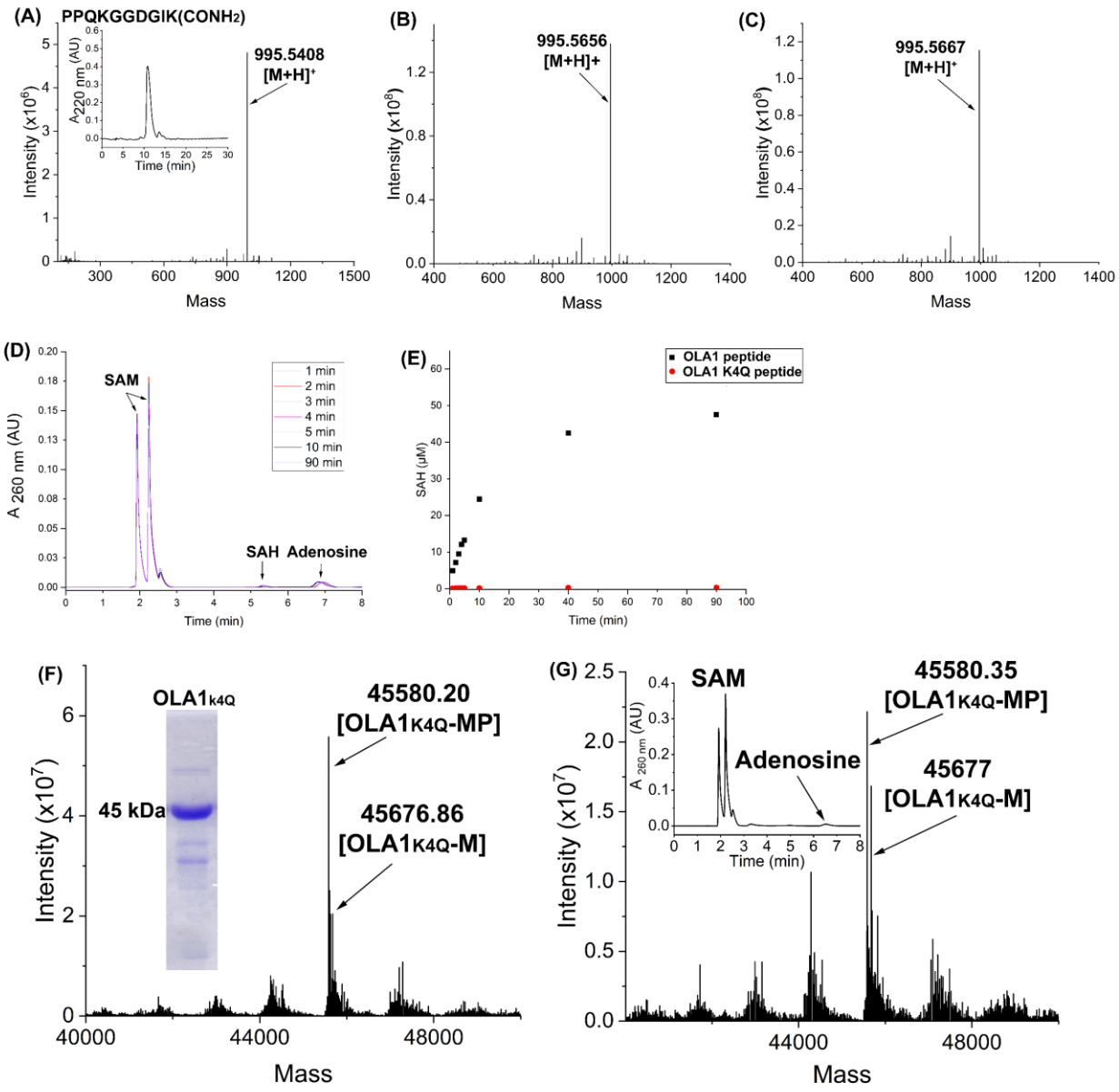

(A)
PPQKGGDGIK(CONH2)
995.5408
[M+H]+
Intensity (x10^6)
Mass
A220 nm (AU)
Time (min)
(B)
995.5656
[M+H]+
Intensity (x10^8)
Mass
(C)
995.5667
[M+H]+
Intensity (x10^8)
Mass
(D)
SAM
SAH
Adenosine
1 min
2 min
3 min
4 min
5 min
10 min
90 min
A 260 nm (AU)
Time (min)
(E)
OLA1 peptide
OLA1 K4Q peptide
SAH (μM)
Time (min)
(F)
OLA1k4Q
45 kDa
45580.20
[OLA1K4Q-MP]
45676.86
[OLA1K4Q-M]
Intensity (x10^7)
Mass
(G)
SAM
Adenosine
45580.35
[OLA1K4Q-MP]
45677
[OLA1K4Q-M]
A 260 nm (AU)
Time (min)
Intensity (x10^7)
Mass

Figure 3.3 MS and HPLC analysis of the methylation reaction using OLA1 K4Q mutant peptide and protein as substrates.

All reactions were carried out at 37 ℃. (A) MS and HPLC (inset) analysis of the synthesized 10-residue OLA1 K4Q peptide (calculated $[M + H]^+$: 995.5638, detected: 995.5408). (B) Mass spectrum of the reaction between 20 μM OLA1 K4Q peptide and 100 μM SAM (calculated $[M + H]^+$: 995.5638, detected : 995.5656). (C) Mass spectrum of the reaction between 20 μM OLA1 K4Q peptide and 100 μM SAM in the presence of 1 μM NTMT1 (calculated $[M + H]^+$: 995.5638, detected: 995.5667). (D) HPLC monitoring of SAH release for the reaction between 40 μM of OLA1 K4Q peptide and 100 μM of SAM in the presence of 1 μM NTMT1 at different time intervals. (E) Comparison of the initial rates for SAH release between OLA1 and OLA1 K4Q peptides. (F) Intact protein mass of OLA1 K4Q mutant (calculated MW for deletion of initial methionine and proline [$OLA1_{K4Q}$ - MP]: 45580.34 Da, detected MW: 45580.2 Da; calculated MW for deletion of initial methionine [$OLA1_{K4Q}$ - M]: 45677.45 Da, detected MW: 45676.86 Da). (G) Intact protein mass of methylation assay using OLA1 K4Q mutant as substrate (detected MW: 45580.35 Da & 45677 Da) and SAH release monitored by HPLC (inset).

3.2.4 Investigation of OLA1 methylation in human cells

As OLA1 can be methylated by NTMT1 at peptide and protein level *in vitro*, we continued to ask whether it could be methylated *in vivo*. To test this, the open reading frame of OLA1 was inserted into the modified pEGFP-n1 vector which contains a C-terminal Flag tag. The newly constructed plasmid was transfected to HEK293FT cells and the Flag-tagged protein was purified using anti-Flag M2 magnetic beads (Fig. 3.4D). Protein band for OLA1-EGFP-Flag (~ 73 kDa) was excised, in-gel digested by chymotrypsin, and subsequently analyzed by LC-MS. MS raw data were analyzed and mapped to the primary sequence of protein OLA1-EGFP-Flag by BiopharmaLynx and numerous peptides derived from this protein were identified with a sequence coverage of 78.4%. Among them, the N-terminal peptide PPKKGGDGIKPPPIIGRF was not detected, but instead, its dimethylated form was heavily found (Fig. 3.4A, retention time: 20.1 min). The LC-MS results indicated that the first methionine was removed in human cells and nearly

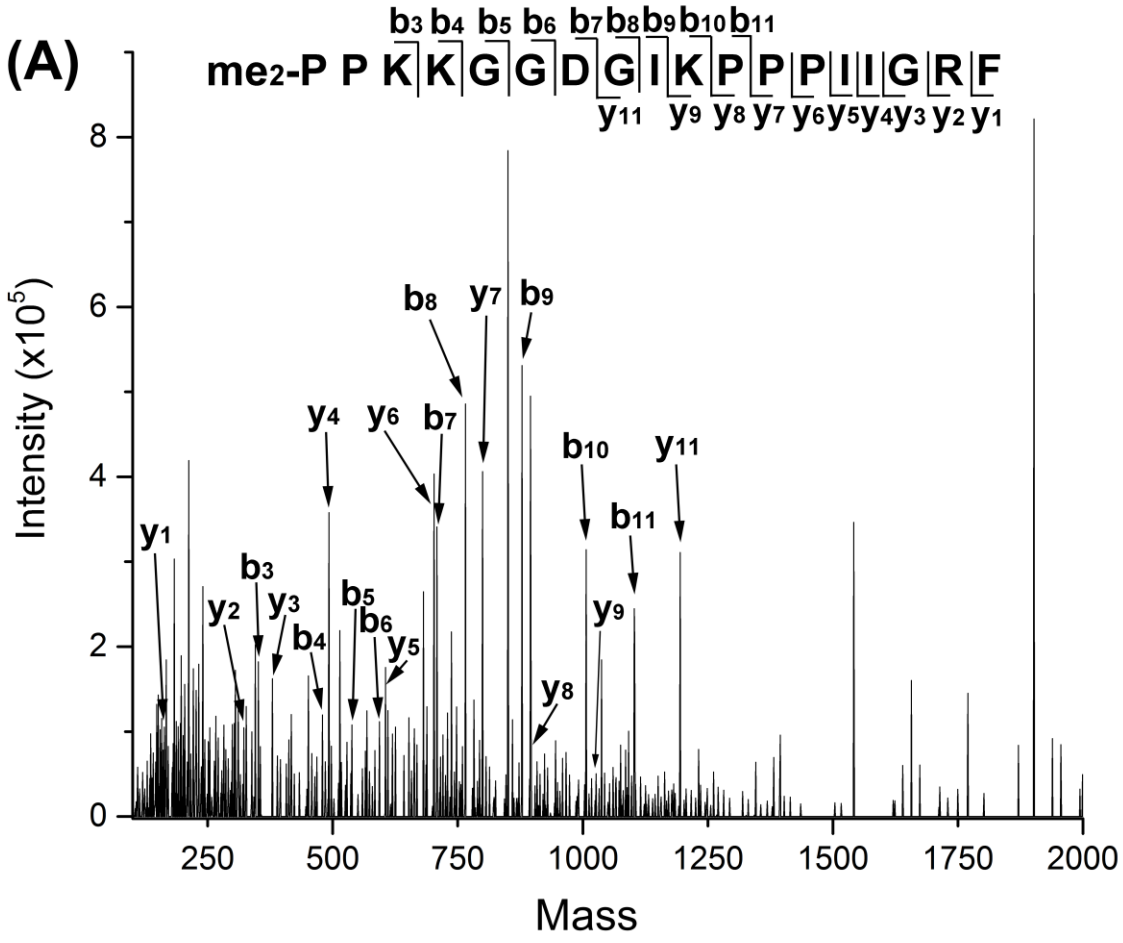
all the OLA1 detected were dimethylated. Peptide fragmentations were analyzed manually, and many b-ions and y-ions were identified, confirming that OLA1 is N-terminal methylated in human cells (Fig. 3.4A).

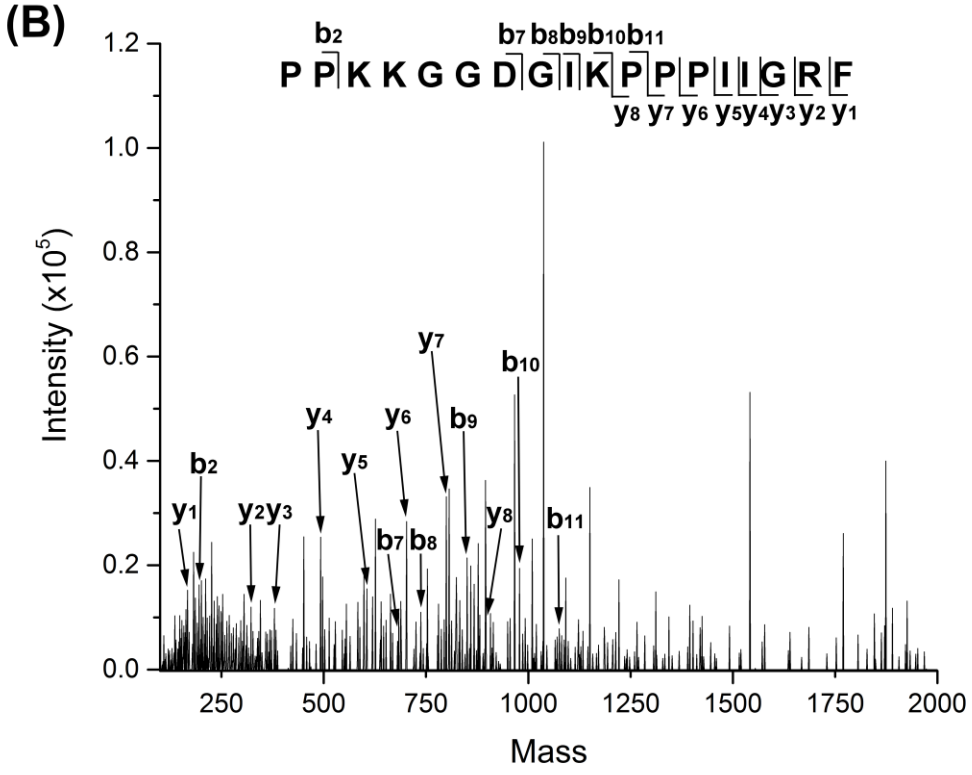
Next, we determined whether this modification was catalyzed by NTMT1 *in vivo*. Previous studies applied *si*RNA knockdown technique to reduce the expression level of NTMT1, and LC-MS/MS was used to quantify the declined methylation level of its substrates (e.g., DDB2, PARP3)[29, 33]. To avoid this semiquantitative method, HEK293FT NTMT1 gene KO cell line was generated using CRISPR-Cas9 (Chapter 2)[34]. Plasmid pEGFP-n1-OLA1-EGFP-Flag was transfected to this NTMT1 KO cell line and the Flag-tagged protein was purified and analyzed as previously described. On the contrary to wt HEK293FT cell line, dimethylated N-terminal peptide PPKKGGDGIKPPPIIGRF was not detected by LC-MS. Instead, the unmodified N-terminal peptide was found (Fig. 3.4B, retention time: 20.0 min). Peptide fragmentation confirmed that OLA1 was not methylated in NTMT1 KO cells (Fig. 3.4B). These results demonstrate that OLA1 is dimethylated in human cell line and NTMT is required for catalyzing this modification.

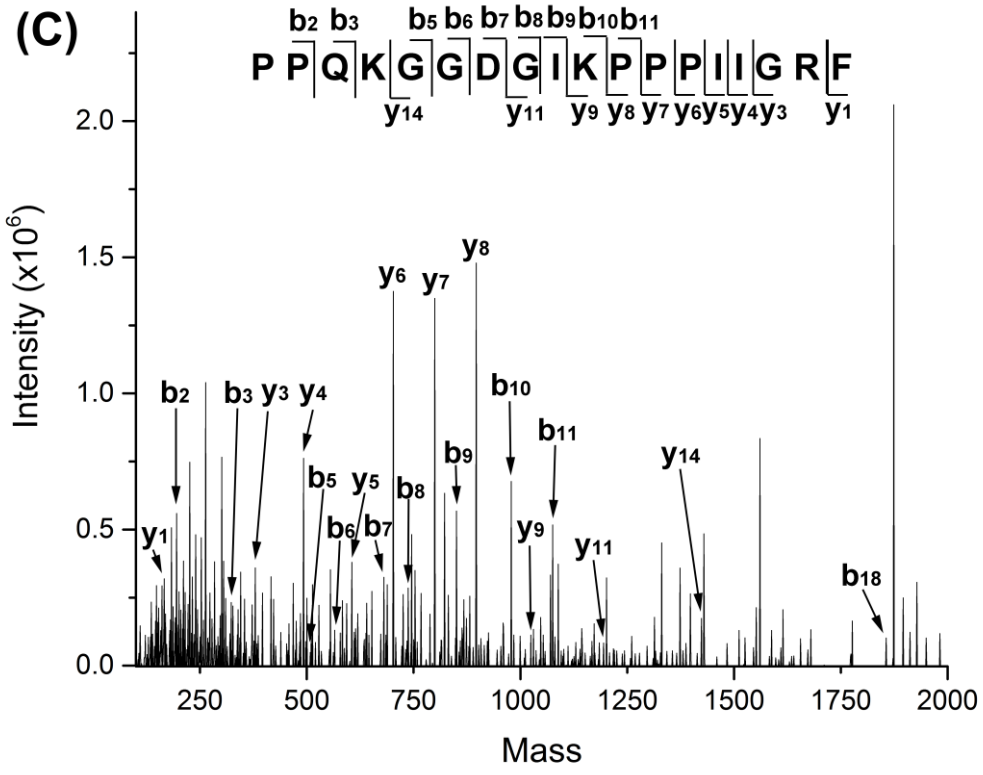
Moreover, to verify whether Pro-Pro-Lys motif is important for its *in vivo* methylation, plasmid pEGFP-n1-OLA1$_{K4Q}$-EGFP-Flag was constructed and transfected to unedited HEK293FT cell line. Flag-tagged protein was treated the same way as previously described and only the unmethylated N-terminal peptide of OLA1$_{K4Q}$ mutant (Fig. 3.4C, retention time: 21.8 min) was identified by LC-MS analysis. Our results indicate that Pro-Pro-Lys especially the lysine residue is critical for OLA1 methylation *in vitro* and *in vivo.* The K4Q mutant would be a good control for studying the functions of OLA1 N-terminal methylation.

To verify whether NTMT1 would methylate endogenous OLA1, we wanted to generate an antibody which specifically recognizes the methylated OLA1. Here, we synthesized the N-methyl-

L-proline and stachydrine (characterized by ^{1}H and ^{13}C NMR shown in Appendix A), which can be used to synthesize the N-terminal OLA1 peptides with methylated prolines as antigens for antibody production in future.




(B)
b2
b7 b8 b9 b10 b11
P P K K G G D G I K P P P I I G R F
y8 y7 y6 y5 y4 y3 y2 y1
Intensity (x10^5)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
y7
b10
y4
y6
b9
b2
y5
y1
y2 y3
y8
b11
b7
b8
250
500
750
1000
1250
1500
1750
2000
Mass


(C)
b2 b3
b5 b6 b7 b8 b9 b10 b11
P P Q K G G D G I K P P P I I G R F
y14
y11
y9 y8 y7 y6 y5 y4 y3
y1
Intensity (x10^6)
2.0
1.5
1.0
0.5
0.0
y8
y6
y7
b10
b2
b3
y3
y4
b11
b9
y14
b5
y5
b8
b6
b7
y9
y11
b18
250
500
750
1000
1250
1500
1750
2000
Mass

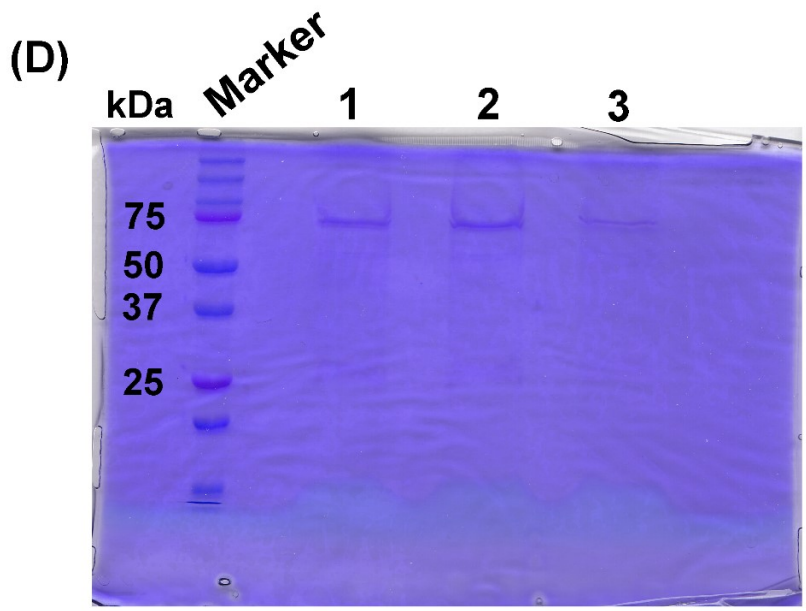


Figure 3.4 LC-MS analysis of OLA1 N-terminal methylation in HEK293FT cells.

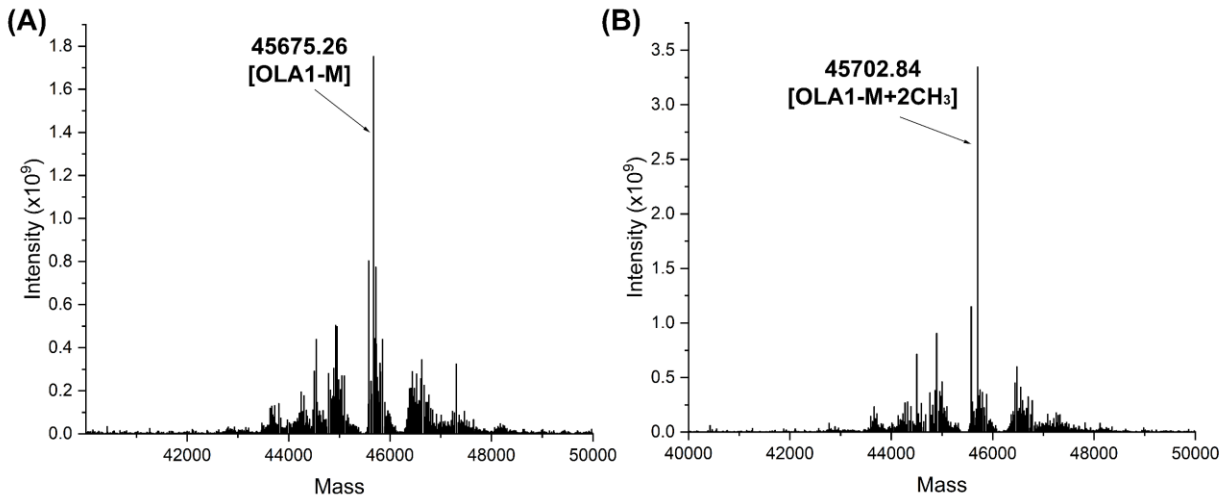
(A) MS^E spectrum of mapping the N-terminal peptide of wt OLA-EGFP-Flag expressed in unedited HEK293FT cell line (calculated precursor ion: 1902.1324, detected: 1902.1068). (B) MS^E spectrum of mapping the N-terminal peptide of wt OLA-EGFP-Flag expressed in NTMT1 KO HEK293FT cell line (calculated precursor ion: 1874.1011, detected: 1874.0739). (C) MS^E spectrum of mapping the N-terminal peptide of $OLA1_{K4Q}$-EGFP-Flag expressed in unedited HEK293FT cell line (calculated precursor ion: 1874.0647, detected: 1874.0394). (D) SDS-PAGE of FLAG tagged proteins purified using anti-Flag M2 magnetic beads (1: wt OLA-EGFP-Flag expressed in unedited HEK293FT cell line; 2: $OLA1_{K4Q}$-EGFP-Flag expressed in unedited HEK293FT cell line; 3: wt OLA1-EGFP-Flag expressed in NTMT1 KO HEK293FT cell line).

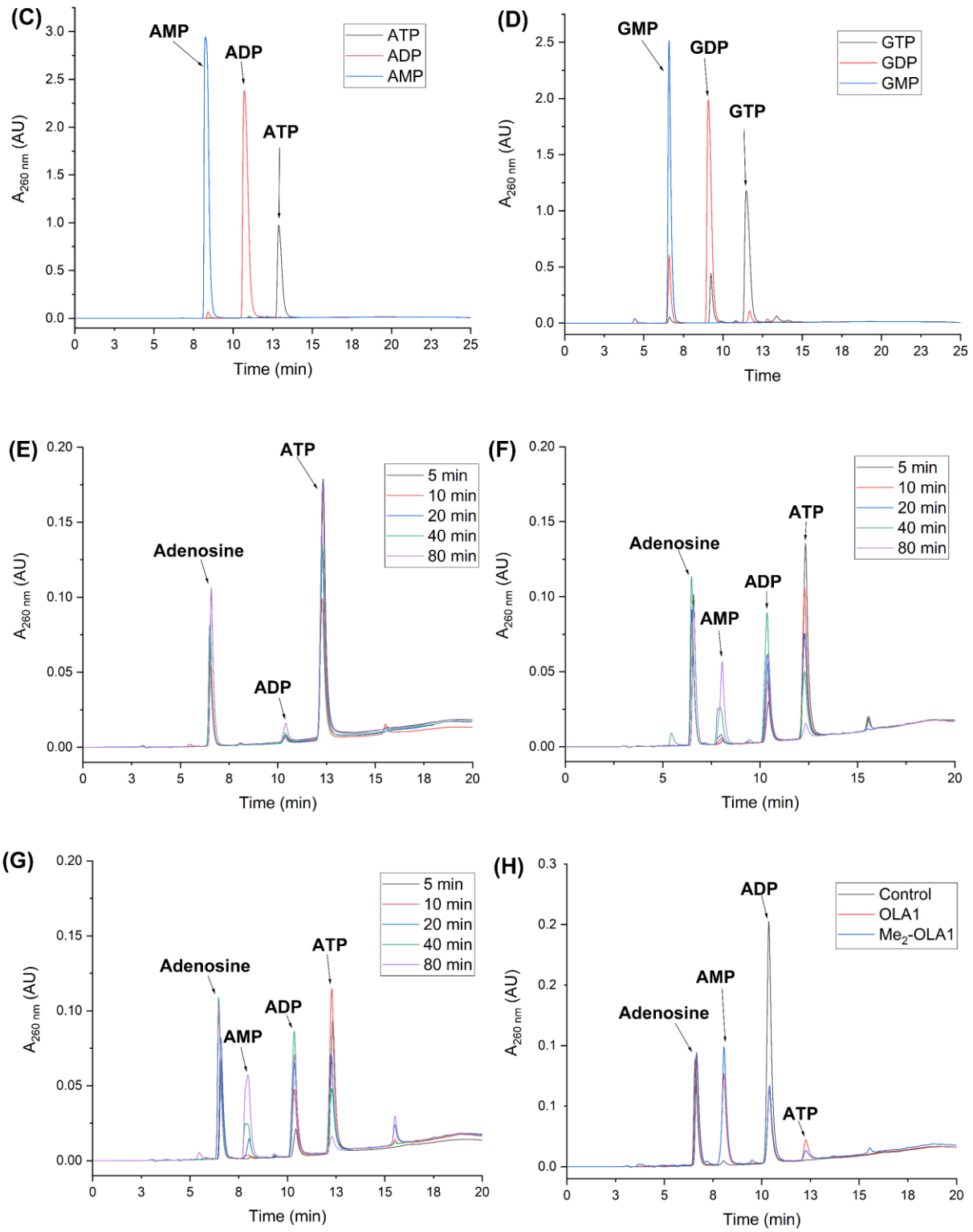
3.2.5 N-terminal methylation of OLA1 does not alter its ATPase and GTPase activity *in vitro*

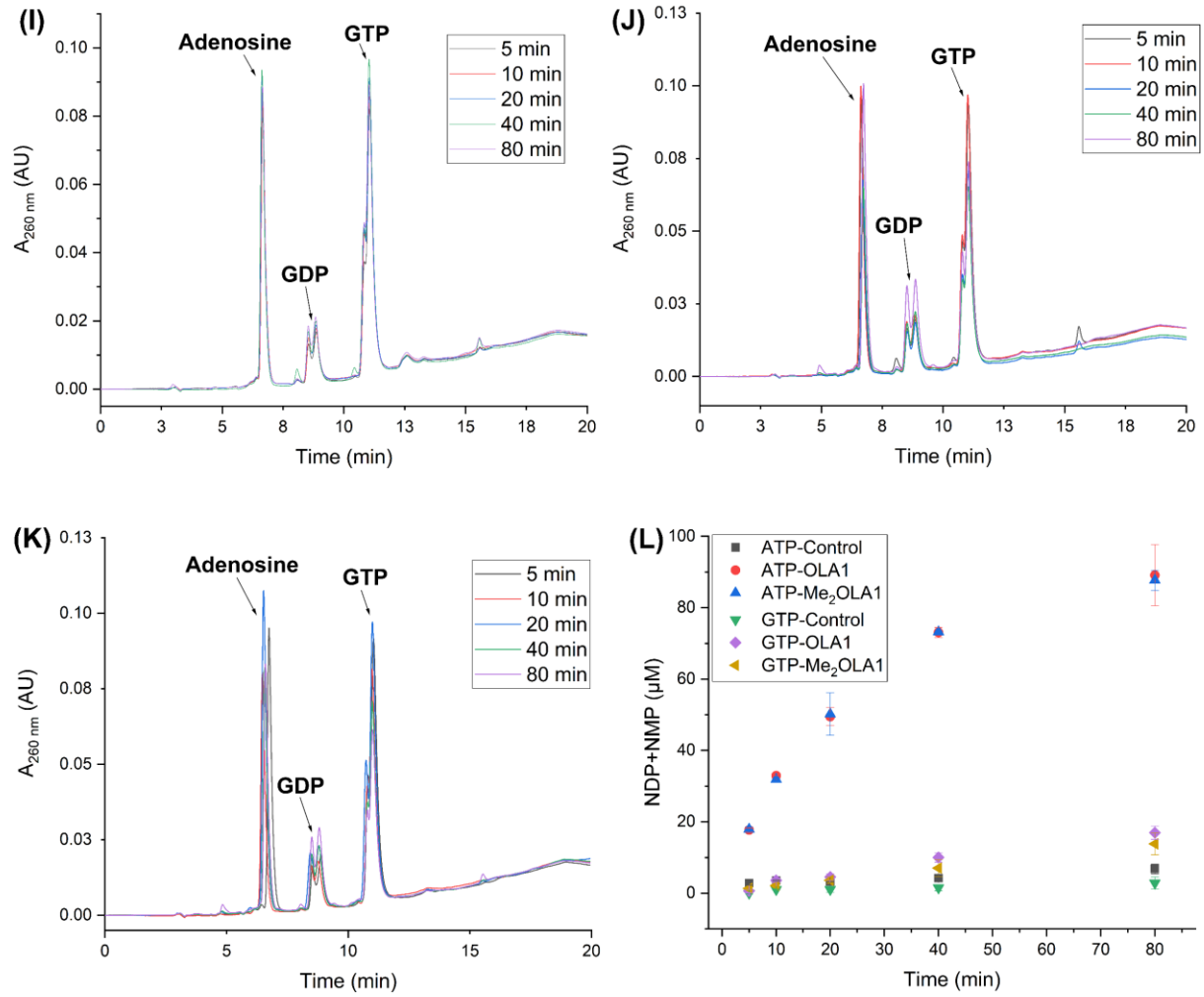
OLA1 belongs to the family of GTPases but was reported to function as an ATPase[9-10]. Here we determined whether the N-terminal methylation of OLA1 would affect or alter its catalytic activity. Dimethylated OLA1 was obtained by incubating OLA1 with SAM in the presence of NTMT1. As a control, OLA1 was incubated with SAM in the absence of NTMT1. SAM and SAH

were removed by gel filtration. Dimethylated OLA1 was further confirmed by intact protein mass (Fig. 3.5A & 3.5B).

The hydrolysis of ATP and GTP were measured by HPLC, and adenosine was added as an external standard (retention time: 6.6 min). The retention times for standards (ATP, ADP, AMP, GTP, GDP, GMP) were shown in Fig. 3.5C & 3.5D. The hydrolysis of ATP in the absence of OLA1 was slow (0.05 μM/min, Fig. 3.5E, 3.5L & 3.5M). As expected, both OLA1 and dimethylated OLA1 were able to catalyze the hydrolysis of ATP with high activities, however, the reaction rates were undistinguishable from each other (1.51 μM/min for OLA1, 1.52 μM/min for Me_2-OLA1, Fig. 3.5F, 3.5G, 3.5L & 3.5M). We observed two products, ADP and AMP, for ATP hydrolysis. ADP was previously reported as the only product[9]. To confirm that OLA1 can hydrolyze ADP as well, we performed an ADP hydrolysis assay. Our results showed that both OLA1 and methylated OLA1 can catalyze the hydrolysis of ADP (Fig. 3.5H).




(C)
AMP
ADP
ATP
ATP
ADP
AMP
A260 nm (AU)
Time (min)
(D)
GMP
GDP
GTP
GTP
GDP
GMP
A260 nm (AU)
Time
(E)
ATP
Adenosine
ADP
5 min
10 min
20 min
40 min
80 min
A260 nm (AU)
Time (min)
(F)
Adenosine
AMP
ADP
ATP
5 min
10 min
20 min
40 min
80 min
A260 nm (AU)
Time (min)
(G)
Adenosine
AMP
ADP
ATP
5 min
10 min
20 min
40 min
80 min
A260 nm (AU)
Time (min)
(H)
Adenosine
AMP
ADP
ATP
Control
OLA1
Me2-OLA1
A260 nm (AU)
Time (min)


(I)
Adenosine
GTP
GDP
5 min
10 min
20 min
40 min
80 min
A260 nm (AU)
Time (min)
(J)
Adenosine
GTP
GDP
5 min
10 min
20 min
40 min
80 min
A260 nm (AU)
Time (min)
(K)
Adenosine
GTP
GDP
5 min
10 min
20 min
40 min
80 min
A260 nm (AU)
Time (min)
(L)
ATP-Control
ATP-OLA1
ATP-Me2OLA1
GTP-Control
GTP-OLA1
GTP-Me2OLA1
NDP+NMP (μM)
Time (min)

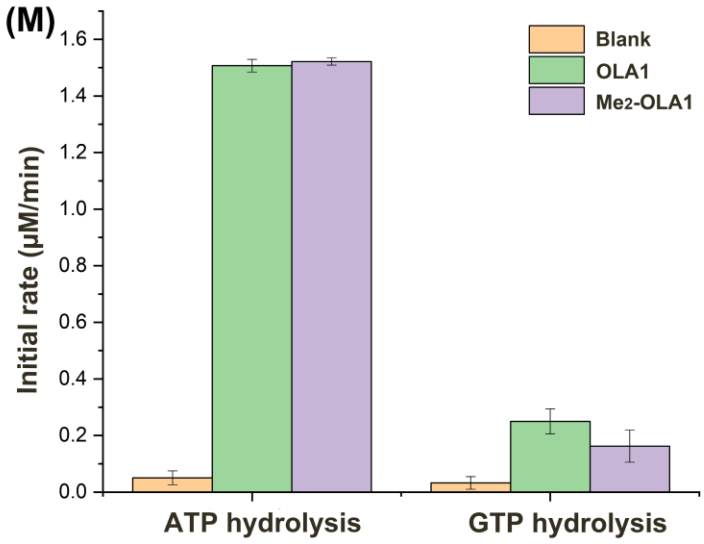

(M)
Blank
OLA1
Me2-OLA1
Initial rate (μM/min)
ATP hydrolysis
GTP hydrolysis

Figure 3.5 HPLC-based ATPase and GTPase assay for OLA1 and dimethylated OLA1.

All the reactions were carried out at 37 ℃. (A) Intact protein mass of OLA1 after incubating with SAM. (B) Intact protein mass of dimethylated OLA1. (C) HPLC analysis of ATP (retention time: 12.9 min), ADP (retention time: 10.7 min) and AMP (retention time: 8.2 min) standards. (D) HPLC analysis of GTP (11.5 min), GDP (retention time: 9.1 min) and GMP (retention time: 6.6 min) standards. (E) HPLC monitoring of ATP (125 μM) hydrolysis in the absence of OLA1 at different time intervals. (F) HPLC monitoring of ATP (125 μM) hydrolysis in the presence of 5 μM OLA1 at different time intervals. (G) HPLC monitoring of ATP (125 μM) hydrolysis in the presence of 5 μM dimethylated OLA1 at different time intervals. (H) HPLC monitoring of ADP (125 μM) hydrolysis with OLA1, dimethylated OLA1 or without OLA1. The reactions were carried out for 1 h. (I) HPLC monitoring of GTP (125 μM) hydrolysis in the absence of OLA1 at different time intervals. (J) HPLC monitoring of GTP (125 μM) hydrolysis in the presence of 5 μM OLA1 at different time intervals. (K) HPLC monitoring of GTP (125 μM) hydrolysis in the presence of 5 μM dimethylated OLA1 at different time intervals. (L) Comparison of the products formation for ATP/GTP hydrolysis with OLA1, dimethylated OLA1 or without OLA1. (M) Comparison of the initial reaction rates for ATP/GTP hydrolysis with OLA1, dimethylated OLA1 or without OLA1.

Next, we examined the GTPase activity of OLA1 (Fig. 3.5I, 3.5J & 3.5K). Note that the GTP often contains contaminative GDP, and our batch contained ~18% of GDP (quantified by HPLC). Thus, the GDP background was subtracted for calculating the initial reaction rates (Fig. 3.5L &3.5M). OLA1 was able to hydrolyze GTP (0.25 μM/min) but the reaction rate was much lower compared to ATP (1.51 μM/min). These results are in consistent with previous reports by Koller-Eichhorn et al. and Sun et al. but are on the contrary to the one published by Chen et al.[1, 10, 15] Chen et al. reported that the GTPase activity of OLA1 was ~ 7-fold higher than ATPase activity by measuring the release of [^{32}P] Pi[15]. Both the radiolabeled assay and HPLC-based assay were performed *in vitro*, and our HPLC results clearly showed that the ATP hydrolysis rate was much faster than the rate of GTP hydrolysis.

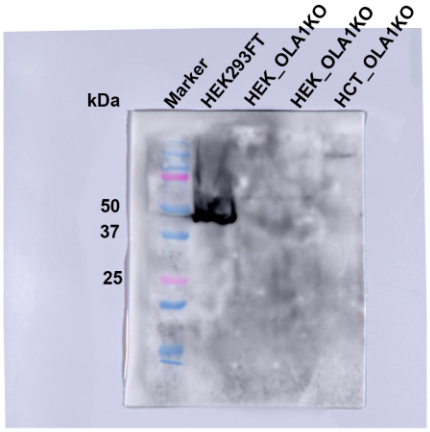
Taken together, both OLA1 and its dimethylated form exhibit a higher activity for ATP hydrolysis than GTP hydrolysis. Moreover, OLA1 can further hydrolyze ADP to form AMP. However, the N-terminal methylation of OLA1 has no influence on its ATPase or GTPase activity *in vitro*. The flexible N-terminus of OLA1 was not seen in the crystal structure (PDB: 2OHF) and may be not involved in the ATP/GTP binding site, which explains why its methylation has no effect on its catalytic activity.

3.2.6 Knockout of OLA1 in HEK293FT and HCT116

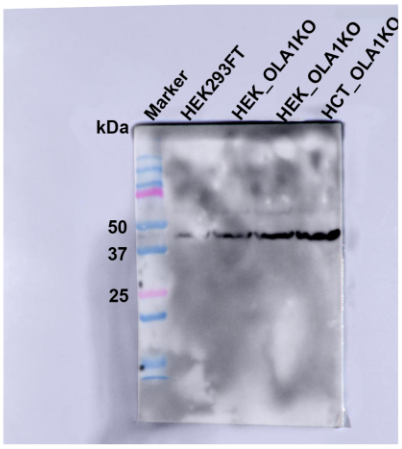
To eliminate the influences of endogenous OLA1 for the study of N-terminal methylation in human cell lines, OLA1 was knocked out in HEK293FT cell line and colon cancer cell line HCT116 using CRISPR-cas9 system. OLA1 gene KO cells were confirmed by western blot (Fig. 3.6A & 3.6B) and DNA sequencing (Fig. 3.6C & 3.6D).

Till now, the functions of OLA1 N-terminal methylation remain unknown. Previous study indicated that OLA1 regulates centrosome amplification and cell growth[11-12, 22]. To test whether its N-terminal methylation would participate in such regulations, plasmids pEGFP-n1-OLA1-EGFP-Flag and pEGFP-n1-OLA1$_{K4Q}$-EGFP-Flag (production of methylation deficient OLA1) were transfected into OLA1 KO HCT116 cell line. However, the transfection efficiency was very low, and one of our future plans is to generate stable cell lines expression OLA1-EGFP or OLA1$_{K4Q}$-EGFP.

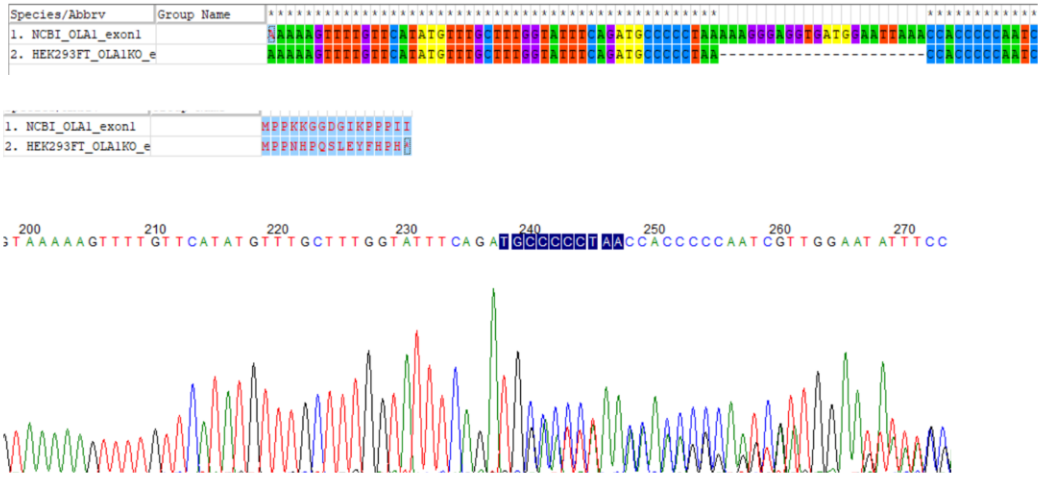
(A)


Marker
HEK293FT
HEK_OLA1KO
HEK_OLA1KO
HCT_OLA1KO
kDa
50
37
25

(B)


Marker
HEK293FT
HEK_OLA1KO
HEK_OLA1KO
HCT_OLA1KO
kDa
50
37
25

(C)


Species/Abbrv
Group Name
1. NCBI_OLA1_exon1
2. HEK293FT_OLA1KO_e
1. NCBI_OLA1_exon1
2. HEK293FT_OLA1KO_e
200
210
220
230
240
250
260
270

(D)

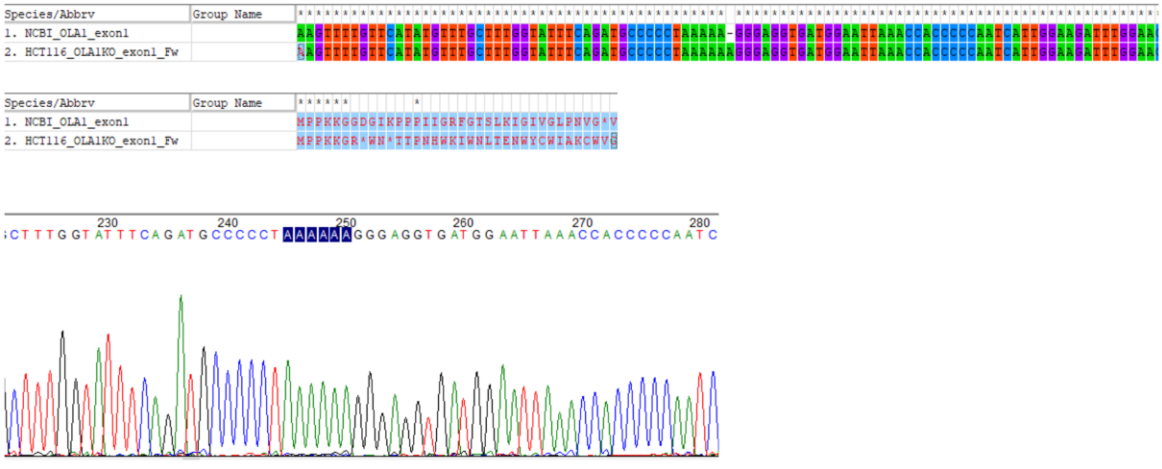

Species/Abbrv
Group Name
1. NCBI_OLA1_exon1
2. HCT116_OLA1KO_exon1_Fw
Species/Abbrv
Group Name
1. NCBI_OLA1_exon1
2. HCT116_OLA1KO_exon1_Fw
230
240
250
260
270
280

Figure 3.6 OLA1 gene knockout using CRISPR-Cas9 system.

(A) Western blot analysis of OLA1 KO cell lines (HEK293FT & HCT116) using rabbit anti-OLA1 antibody. (B) Western blot analysis of actin as a loading control for OLA1 KO cell lines (HEK293FT & HCT116). (C) DNA sequencing result of OLA1 KO HEK293FT cell line. (D) DNA sequencing result of OLA1 KO HCT116 cell line.

3.2.7 Methylation of overexpressed OLA1-EGFP-Flag does not alter its subcellular location

OLA1 was previously reported to exhibit a nucleocytoplasmic distribution and may shuttle between cytoplasm and nucleus[10]. Cai et al. reported that N-terminal trimethylation of DDB2 would promote its nuclear localization while the methylation deficient mutant displayed a reduced distribution in cell nuclear fraction[29]. Here, we wanted to test whether the N-terminal methylation would affect the subcellular distribution of OLA1. Plasmids pEGFP-n1-OLA1-EGFP-Flag, pEGFP-n1-OLA1$_{K4Q}$-EGFP-Flag and pEGFP-n1 empty vector were transfected into OLA1 KO HEK293FT cells and the green fluorescence signal was visualized by confocal microscopy. The dimethylated OLA1 and the unmodified OLA1 K4Q mutant expressed in HEK293FT OLA1 KO cells were confirmed by protein in-gel digestion and LC-MS analysis (See Appendix C). Majority of OLA1-EGFP and OLA1$_{K4Q}$-EGFP were localized in cytoplasm. EGFP, as a control, was distributed in both nucleus and cytoplasm. Taken together, the subcellular location of overexpressed OLA1 was not affected by its N-terminal methylation. This result suggests that the N-terminal methylation of OLA1 may function differently from the N-terminal methylation of DDB2.

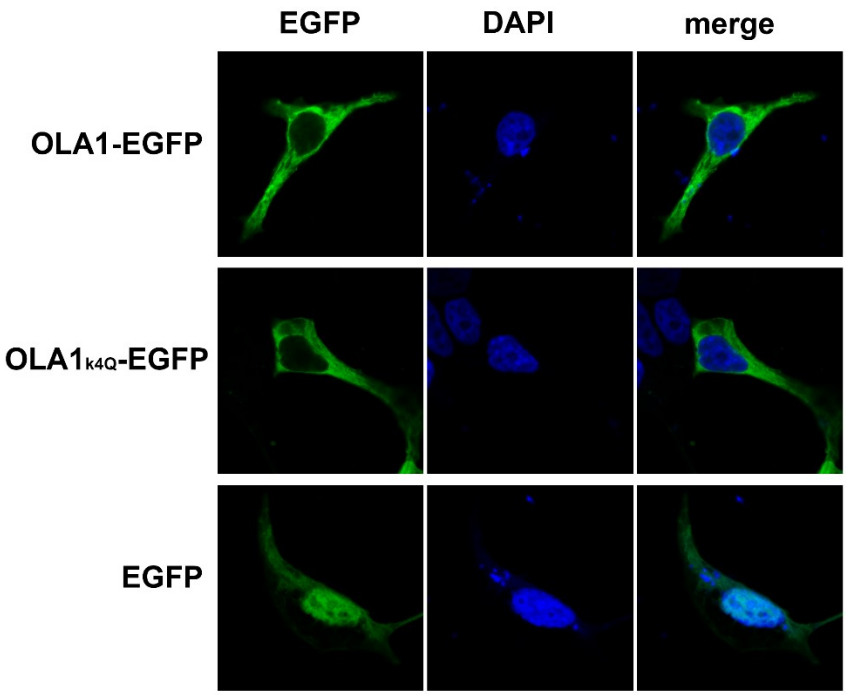


Figure 3.7 Comparison of the subcellular location of OLA1 and K4Q mutant using confocal microscopy.

Plasmids pEGFP-n1-OLA1-EGFP-Flag, pEGFP-n1-OLA1$_{K4Q}$-EGFP-Flag and pEGFP-n1 empty vector were transfected into OLA1 KO HEK293FT cells. OLA1 and EGFP proteins were imaged by green fluorescence. Cell nucleus was stained with DAPI.

3.3 Materials and methods

3.3.1 Cell culture conditions

Human embryonic kidney (HEK) 293FT cells (generous gift from Dr. Yulan Xiong) were cultured in DMEM (with 4.5 g/L glucose, L-glutamine & sodium pyruvate, Corning cellgro) supplemented with 10% FBS (heat inactivated FBS, Gibco) and 1% P/S (Gibco) in a humidified atmosphere containing 5% CO_2 at 37 ˚C. Human colorectal carcinoma cell line HCT116 (generous gift from Dr. Ryan Rafferty) were cultured in McCoy's 5a Modified Medium (Corning cellgro) supplemented with 10% FBS and 1% P/S in a humidified atmosphere containing 5% CO_2 at 37 ˚C.

3.3.2 Plasmids construction

Plasmid pET28-MHL-NTMT1 was constructed as previously described. The open reading frame of human OLA1 was amplified from HEK293FT cDNA to introduce NdeI and XhoI restriction sites, and subcloned into *E. coli* expression vector pET30a (+) (Novagen) to generate plasmid pET30a-OLA1-His_6. Plasmid pET30a-$OLA1_{K4Q}$-His_6 was constructed similarly, except that the forward primer contained a K4Q mutation site. Mammalian expression vector pEGFP-n1 was modified to introduce a FLAG epitope tag (DYKDDDDK) at the C terminus of EGFP. To express wt OLA1 and K4Q mutant in mammalian cells, their coding sequences were cloned and inserted into the modified pEGFP-n1-Flag vector with XhoI and BanHI restriction sites. Plasmids for OLA1 knockout using the CRISPR-Cas9 system were constructed similarly to the one described in chapter 2[34]. Briefly, the guide sequences were designed based on an online CRISPR Design Tool (http://tools.genome-engineering.org) established by Dr. Zhang group[34]. The designed oligos were annealed to give sticky BbsI ends and inserted into vctor PX459 (generous gift from Dr. Yulan Xiong) linearized by BbsI. All the constructed plasmids were confirmed by DNA sequencing before use. All primers are listed in Table 3.1.

Table 3.1 Primer sequences for plasmids construction.

Plasmid	Primer sequence (restriction site bolded and underlined)	
pET30a-OLA1-His_6	NdeI_Fw	TTGATA**CATATG**CCCCCTAAAAAGGGAGGTGATG
	XhoI_Rv	AAACAC**CTCGAG**TTTCTTCTTCGGTTGTTGAG
pET30a-$OLA1_{K4Q}$-His_6	NdeI_Fw	TTGATA**CATATG**CCCCCTCAAAAGGGAGGTGATG
	XhoI_Rv	AAACAC**CTCGAG**TTTCTTCTTCGGTTGTTGAG
pEGFP-n1-EGFP-Flag	BamHI_Fw	AAAATT**GGATCC**ACCGGTCGCCACCATGG
	NotI_Rv	AA**GCGGCCGC**CTACTTATCGTCATCGTCCTTGTAATCC TTGTACAG CTCGTCCATGCC
pEGFP-n1-OLA1-EGFP-Flag	XhoI_Fw	TTGATA**CTCGAG**GCCACCATGCCCCCTAAAAAGGGAG GTGATG
	BamHI_Rv	AAACAC**GGATCC**ACTTTCTTCTTCGGTTGTTGAG
pEGFP-n1-$OLA1_{K4Q}$-EGFP-Flag	XhoI_Fw	ATA**CTCGAG**GCCACCATGCCCCCTCAAAAGGGAGGTG ATG
	BamHI_Rv	AAACAC**GGATCC**ACTTTCTTCTTCGGTTGTTGAG
PX459-$OLA1_{sgRNA1}$	BbsI_Fw	**CACC**GCAGATGCCCCCTAAAAAGGG
	BbsI_Rv	**AAAC**CCCTTTTTAGGGGGCATCTGC
PX459-$OLA1_{sgRNA2}$	BbsI_Fw	**CACC**GATTTCAGATGCCCCCTAAAA
	BbsI_Rv	**AAAC**TTTTAGGGGGCATCTGAAATC
PX459-$OLA1_{sgRNA3}$	BbsI_Fw	**CACC**GCCATCACCTCCCTTTTTAGG
	BbsI_Rv	**AAAC**CCTAAAAAGGGAGGTGATGGC
PX459-$OLA1_{sgRNA4}$	BbsI_Fw	**CACC**GTCTCTTACCTTGCTGGTTTG
	BbsI_Rv	**AAAC**CAAACCAGCAAGGTAAGAGAC
Seq-$OLA1_{exon1}$	Forward	ACCGTGCCCGGCCCAACAAATATTGATTTTAG
	Reverse	TGTAGGTACCACACATAGCCAACCAGCACTG
Seq-$OLA1_{exon2}$	Forward	TCCTCATGGATCTTATTTGAGCATTG
	Reverse	GTAGACAGTGGTTTACCATGTTGGC

3.3.3 Expression and purification of recombinant proteins

His_6-NTMT1 was expressed and purified as previously described (Chapter 2). Protein OLA1-His_6 and its K4Q mutant were expressed in *E. coli.* BL21 (DE3) (Lucigen). Bacteria were grown in 2 L of LB medium supplemented with 50 μg/mL kanamycin. When $OD_{600\ nm}$ reached to 0.6 - 0.8, IPTG was added to a final concentration of 0.4 mM to induce protein expression. Bacteria were grown at 18 ˚C for an additional 18 h and then harvested by centrifugation at 5,000 × g for 20 min at 4˚C. Bacterial pellets were collected and stored at −80 ˚C until further use.

Composition and pH values of buffers used in this study are summarized in Table 3.2. Proteins were purified following a modified procedure in a published literature[9]. Cell pellets were re-suspended in buffer A and lysed by sonication for 15 min (5-second on and 3-second off) at 40% amplitude. The cell debris was removed by centrifugation at 18,000 × g for 45 min. The supernatant was decanted and incubated with Ni-NTA resin (Qiagen) that had been pre-equilibrated with buffer A for 30 min. The resin was washed with 30 column volumes of buffer B and then eluted with 5 column volumes of buffer C. Fractions containing the targeted protein were combined and protein was exchanged into buffer D using a HiPrep^{TM} 26/10 desalting column (GE Healthcare). The purified proteins were concentrated to 10−50 mg/ml and stored in aliquots at −80 ˚C until further use. Protein purity was assessed by 12% tris-glycine SDS-PAGE and their concentrations were determined by bicinchoninic acid assays (BCA, ThermoFisher) using a BSA calibration curve.

Table 3.2 Ingredients of buffers.

Buffer	Composition
A	50 mM Tris (pH7.9), 700 mM NaCl, 5 mM $MgCl_2$ and 10% glycerol
B	50 mM Tris (pH 7.9), 100 mM NaCl, 5 mM $MgCl_2$, 5% glycerol and 40 mM Imidazole
C	50 mM Tris (pH 7.9), 100 mM NaCl, 5 mM $MgCl_2$, 5% glycerol and 250 mM Imidazole
D	50 mM Tris (pH7.9), 100 mM NaCl
E	25 mM Tris (pH 7.5), 50 mM KCl
F	1 × PBS (pH 7.4), 0.2% Triton X-100
G	500 mM Tris (pH7.9), 10% SDS, 20 mM TCEP
H	100 mM Tris (pH 7.9), 100 mM iodoacetamide,
I	100 mM Tris (pH 7.9), 10 mM $CaCl_2$
J	50 mM Tris (pH 7.5), 200 mM NaCl, 5 mM $MgCl_2$

3.3.4 Peptide synthesis

Peptide analogs of N-terminal wt OLA1 [PPKKGGDGIK($CONH_2$)] and OLA1 K4Q mutant [PPQKGGDGIK($CONH_2$)] were synthesized on Rink amide resins (Peptides International) using standard Fmoc chemistry as previously described[35]. The synthesized peptides were purified by HPLC and their purities and identities were confirmed by HPLC and LC-MS, respectively. Conditions of semi-preparative and analytical HPLC are summarized in table 3.3.

Table 3.3 HPLC conditions for peptide purification and characterization.

		Semi-preparative HPLC	Analytical HPLC
Column		Luna C18(2), 5 μm, 100 Å, 10 × 250 mm (Phenomenex)	XBridge BEH Shield RP 18, 5 μm, 130 Å, 4.6 × 250 mm (Waters)
Flow Rate		3.0 mL/min	1.0 mL/min
Mobile Phases		**A:** 0.1% TFA in H_2O; **B:** 0.1% TFA in MeOH	
Gradient (**B**)	wt OLA1 peptide (retention time)	0–25 min: 5-40%; 25–30 min: 40-90% (11.9 min)	0-30 min: 5–15%; 30-35 min: 15-70% (9.0 min)
	OLA1 K4Q peptide (retention time)	0–30 min: 5-15%; 30-50 min: 15-50% (37.0 min)	0-30 min: 5–15%; 30-35 min: 15-70% (10.8 min)

3.3.5 HPLC and MS-based *in vitro* methylation assay

Methylation of OLA1 at peptide and protein levels by wt NTMT1 and SAM was investigated using LC-MS and HPLC. For LC-MS analysis, 20 μM OLA1 peptides or proteins in 10 mM ammonium bicarbonate (ABC), 200 μM SAM, and 1 μM wt NTMT1 were incubated at 37 ˚C for 30 min. Negative controls were performed in the absence of wt NTMT1. The peptide reaction mixtures were quenched with equal volume of water containing 0.1% formic acid and then analyzed by LC-MS. The protein reaction mixtures were not quenched and directly analyzed by intact protein mass, which was performed on a C4 analytical column eluted at a flow rate of 0.4 ml/min. Mobile phases A and B consisted of 0.1% formic acid in water and in acetonitrile, respectively. The gradient is: 0–1 min, 5% B; 1–3.5 min, 5–100% B; and 3.5–4.7 min, 100% B.

Elution from 2–4.7 min was injected into the mass spectrometer, which was operated in a MS positive mode. Raw data was processed by MassLynx (Waters) and BiopharmaLynx (Waters).

SAH release was monitored by HPLC and adenosine was added as an external standard to quantify SAH. Experiments were carried out in the following way. In a final volume of 400 μL, 1 μM wt NTMT1 was added to the mixture of 100 μM SAM and 40 μM OLA1 peptides at 37 ˚C in buffer E. Aliquots of 40 μL were withdrawn at selected time intervals (1 to 90 min) and quenched with 40 μL of 0.4% TFA solution containing 4 μM adenosine. After centrifugation at 18,000 *g* for 2 min, 50 μL of the supernatant was injected into an XBridge Shield RP 18 column (3.5 μm, 130 Å, 4.6 mm × 150 mm, Waters). The column was eluted at 1 mL/min with an isocratic gradient of 1% methanol in water supplemented with 0.01% TFA. Concentrations of SAH were determined by $A_{260\ nm}$ ratios of the peak areas corresponding to SAH and adenosine.

3.3.6 Mass spectrometric analysis of Flag tagged OLA1 proteins purified from human cells

HEK293FT and HEK293FT NTMT1 KO cells were grown in 175 cm^2 flasks at 37℃ in DMEM medium supplemented with 10% FBS and 100 U/ml P/S in a humidified atmosphere containing 5% CO_2. At 70-90% confluence stage, cells were transfected with 20 μg plasmid pEGFP-n1-OLA1-EGFP-Flag or pEGFP-n1-OLA1$_{K4Q}$-EGFP-Flag using transfection reagent Lipo D293 (SignaGen Laboratories). After 36 h, cells were harvested, suspended in 3 ml of 1× PBS (pH 7.4) buffer and lysed by sonication for 4 min (5-second on and 3-second off) at 40% amplitude. The cell debris was removed by centrifugation at 15,000 × g for 15 min and the supernatant was incubated with 30 μl Flag M2 magnetic beads (Sigma) for 4 h at 4 ℃. After washing with buffer F for three times, Flag tagged proteins were eluted with 60 μl of 200 mM glycine (pH 2.5). The pH of the eluted samples was adjusted to ~7 using 1 M Tris (pH 7.9). Proteins

were reduced by 6 μl buffer G for 10 min at 37 ℃ and alkylated with buffer H for 15 min at 37℃[36]. Protein samples were concentrated to ~ 10 μl by speed-vac, mixed with 10 μl of 2 × Laemmli sample (Bio-Rad) buffer and heated at 100 ℃ for 5 min before loading to SDS-PAGE. After staining and destaining, bands of OLA1-EGFP-Flag or $OLA1_{K4Q}$-EGFP-Flag (~73kD) were excised, cut into small pieces and washed with 100 mM ammonium bicarbonate (ABC)/acetonitrile (1:1, vol/vol) at 37 ℃ to completely destain the gel. After shrinking and dehydrating with acetonitrile, gel pieces were incubated with chymotrypsin (Promega, 200 ng diluted in buffer I) for 18 h at 25 ℃. Digestion products were extracted with 100 μl of 5% formic acid, 100 μl of 100% acetonitrile/10% formic acid (1:1, vol/vol), and 100 μl of pure acetonitrile, respectively. Samples were concentrated by speed-vac, filtered with Ultrafree PTFE membrane (Millipore) and injected to an M class UPLC-XEVO G2-XS QTof (Waters) equipment. Protein digestion products were firstly cleaned using a trapping column (ACQUITY UPLC M-Class Symmetry C18 Trap Column, 100 Å, 5 μm, 300 μm × 50 mm, Waters) and separated on an analytical column (ACQUITY UPLC M-Class HSS T3 Column, 1.8 μm, 300 μm × 100 mm, Waters) with a flow rate of 8 ul/min. Mobile phase A and B consisted of 0.1% formic acid in water and acetonitrile, respectively. The gradient for separation is 1 % to 40 % of mobile phase B for 40min. The mass spectrometer was operated in positive polarity, sensitivity mode, and MS^E mode with a ramp collision energy ranging from 20 to 45 V. All data were mapped to the primary structure of target proteins by BiopharmaLynx (Waters) and peptide fragmentations were analyzed manually.

3.3.7 HPLC-based ATPase and GTPase assay

The dimethylated OLA1 was produced by incubating OLA1 (20 μM) purified from *E. coli* with SAM (200 μM) and NTMT1 (3 μM) at 37 ℃ for 45 min in buffer E. Proteins were desalted

into buffer J using a HiPrep^{TM} 26/10 desalting column (GE Healthcare). Dimethylated proteins were concentrated and confirmed by intact protein mass. Protein concentrations were determined by bicinchoninic acid assays (BCA, ThermoFisher) using a BSA calibration curve.

The ATP/GTP hydrolysis was monitored by HPLC and adenosine was added as an external standard to quantify the formation of products ADP/AMP/GDP. Experiments were carried out in the following way. In a final volume of 400 µL, 5 µM OLA1 or dimethylated OLA1 were incubated with 125 µM ATP/GTP at 37 ˚C in buffer J. Aliquots of 60 µL were withdrawn at selected time intervals (5 to 80 min) and quenched with 60 µL of 0.2% TFA solution containing 50 µM adenosine. After centrifugation at 18,000 × g for 2 min, 50 µL of the supernatant were injected into an XBridge Shield RP 18 column (3.5 µm, 130 Å, 4.6 × 250 mm, Waters). Mobile phase A consisted of 10 mM potassium phosphate buffer (pH 6.5) supplemented with 5 mM tetrabutylammonium bromide (Chem-Impex) and mobile B consisted of 10 mM potassium phosphate buffer (pH 6.5)/ methanol (3/2 vol/vol) supplemented with 5 mM tetrabutylammonium bromide[9]. The gradient for separation is 20 % to 100 % of mobile phase B for 15 min with a flow rate of 1ml/min. Concentrations of SAH were determined by $A_{260\ nm}$ ratios of the peak areas corresponding to ADP/AMP/GDP and adenosine. Measurements were performed in duplicate.

3.3.8 OLA1 knockout using CRISPR-Cas9 system

Procedures for OLA1 gene KO in HEK293FT cell line and HCT116 cell line were similar to the one described in Chapter 2[34]. Briefly, 0.5 µg of plasmids PX459-OLA1$_{sgRNA1}$ and 0.5 µg of PX459-OLA1$_{sgRNA2}$, or 0.5 µg of plasmids PX459-OLA1$_{sgRNA3}$ and 0.5 µg of PX459-OLA1$_{sgRNA4}$ were mixed and transfected into HEK293FT seeded in a 6-well plate using transfection reagent LipoD293 (SignaGen Laboratories) and Opti-MEM® I Reduced Serum Medium (ThermoFisher) when cell confluency reached to 50%-70%. For HCT116 cell line, plasmids were transfected with

lipofectamine 2000 (ThermoFisher). After 36h, cells were treated with 3 μg/ml of puromycin for 3 days. Cells were then dissociated with TrypLE™ Express Enzyme (ThermoFisher), counted and serially diluted in DMEM medium (for HEK293FT) or McCoy's 5a Modified Medium (for HCT116) supplemented with 20% FBS to a final cell density of 50-100 cells per 15 ml medium. The diluted cells were plated to two 96-well plates and cultured for 2 weeks. After passaging to 24-well pate, cells were screened by western blot using anti-OLA1 antibody (Bethyl). After confirmed by DNA sequencing, the OLA1 KO cell lines were frozen in liquid N_2.

3.3.9 Western blot

Equal volumes of cell lysates were separated in a 12% tris-glycine SDS-PAGE and then transferred to a supported nitrocellulose membrane (Bio-Rad). The membranes were incubated with rabbit anti-OLA1 antibody (1:1000) overnight at 4 ˚C. After washing with TBS-T buffer, membranes for were incubated with goat anti-rabbit antibody (1:5000) by gentle shaking for 1 h at room temperature. Upon similar washing, bands were detected using SuperSignal West Pico PLUS chemiluminescent substrate (Thermo Fisher Scientific) following manufacturer's protocol. The bands were visualized using Amersham Imager 600 (GE Healthcare).

3.3.10 Subcellular location of overexpressed OLA1 observed by confocal microscopy

HEK293FT OLA1 KO cells were seeded on glass coverslips coated with poly-L-ornithine hydrobromide (Sigma) and transfected with plasmid pEGFP-n1-OLA1-EGFP-Flag, pEGFP-n1-$OLA1_{K4Q}$-EGFP-Flag or empty vector pEGFP-n1 using transfection reagent Lipo D293 (SignaGen Laboratories). After 36 h, cells were fixed with 4% paraformaldehyde (PFA) in 1 × PBS for 20min. Cell nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI). All confocal images were acquired on a Carl Zeiss 700 microscope.

3.3.11 Synthesis of N-methyl-L-proline and stachydrine

Procedure for the synthesis of N-methyl-L-proline was similar to the published literature[37]. Briefly, 37% formaldehyde (1.6 ml, 19.7 mmol) was added to the methanol solution containing L-proline (2g, 17.4 mmol). 10% palladium on carbon (Pd/C, 500mg) was added to the mixture and the resulting slurry was stirred in a hydrogen atmosphere (27 psi) overnight. The Pd/C catalyst was removed by filtering through a Celite pad. The Celite pad was washed with methanol and the filtrates were concentrated under reduced pressure. The residues were re-dissolved in methanol and concentrated a second time to give a solid, which was recrystallized with methanol-diethyl ether to provide fine needles (2.14 g, 95% yield). ^{1}H NMR (600 MHz, D_2O) δ 3.85 (m, 1H), 3.69 (m, 1H), 3.11 (m, 1H), 2.88 (s, 3H), 2.53 -2.38 (m, 1H), 2.11 (m, 1H), 2.08 - 2.00 (m, 1H), 1.99 - 1.90 (m, 1H). ^{13}C NMR (151 MHz, D2O) δ 173.65, 70.49, 56.19, 40.56, 28.69, 22.72.

Stachydrine was synthesized with a modified procedure of the published literature[38]. Briefly, in a 300 ml round bottom flask, L-proline (2 g, 17.4 mmol) was dissolved in 45 ml NaOH solution (2 M). Dimethyl sulfate (4.12 ml, 43.4 mmol) was added to the solution drop by drop. The mixture was stirred at 45 ℃ for 4 h. The reaction was quenched by adding 20 ml of NaOH solution (12M) and refluxed at 100 ℃ for 2 h. After cooling down, the pH was adjusted to ~2 using concentrated HCl. Water was removed under reduced pressure and the residues were extracted with ethanol. After removing ethanol by reduced pressure, the residues were dissolved in 25 ml water and incubated with 20 g of IRA 402 hydroxide resin (Amberlite). The flow through were concentrated under reduced pressure and the residues were taken up by ethanol. Removing of ethanol under reduced pressure provided a white solid (2.43 g, 78% yield). ^{1}H NMR (600 MHz, D_2O) δ 4.30 (m, 1H), 3.71 (m, 1H), 3.56 (m, 1H), 3.30 (s, 3H), 3.10 (s, 3H), 2.52 (m, 1H), 2.44 - 2.28 (m, 1H), 2.27 - 2.02 (m, 2H). ^{13}C NMR (151 MHz, D_2O) δ 169.18, 74.31, 67.70, 52.11, 46.05, 24.55, 18.47.

3.4 References

1. Balasingam, N.; Brandon, H.; Ross, J. A.; Wieden, H.-J.; Thakor, N. Cellular roles of the human Obg-like ATPase 1 (hOLA1) and its bacterial homologue, YchF. *Biochem. Cell Biol.* **2019,** 1-11.

2. Xiao, Y.; Wang, Y. Global discovery of protein kinases and other nucleotide‐binding proteins by mass spectrometry. *Mass Spectrom Rev.* **2016,** *35* (5), 601-619.

3. Stahl, W. L. (Na++ K+)‐ATPase: Function, structure, and conformations. *Ann. Neurol.* **1984,** *16* (S1), S121-S127.

4. Milon, P.; Tischenko, E.; Tomšic, J.; Caserta, E.; Folkers, G.; La Teana, A.; Rodnina, M. V.; Pon, C. L.; Boelens, R.; Gualerzi, C. O. The nucleotide-binding site of bacterial translation initiation factor 2 (IF2) as a metabolic sensor. *P. Natl. Acad. Sci. USA* **2006,** *103* (38), 13962-13967.

5. Spiegel, A. M. Signal transduction by guanine nucleotide binding proteins. *Mol. Cell. Endocrinol.* **1987,** *49* (1), 1-16.

6. Barwe, S. P.; Anilkumar, G.; Moon, S. Y.; Zheng, Y.; Whitelegge, J. P.; Rajasekaran, S. A.; Rajasekaran, A. K. Novel role for Na, K-ATPase in phosphatidylinositol 3-kinase signaling and suppression of cell motility. *Mol. Biol. Cell* **2005,** *16* (3), 1082-1094.

7. Moore, J. D. The Ran‐GTPase and cell‐cycle control. *Bioessays* **2001,** *23* (1), 77-85.

8. van Nieuw Amerongen, G. P.; van Hinsbergh, V. W. Cytoskeletal effects of Rho-like small guanine nucleotide–binding proteins in the vascular system. *Arter. Thromb. Vasc. Biol.* **2001,** *21* (3), 300-311.

9. Koller-Eichhorn, R.; Marquardt, T.; Gail, R.; Wittinghofer, A.; Kostrewa, D.; Kutay, U.; Kambach, C. Human OLA1 defines an ATPase subfamily in the Obg family of GTP-binding proteins. *J. Biol. Chem.* **2007,** *282* (27), 19928-19937.

10. Sun, H.; Luo, X.; Montalbano, J.; Jin, W.; Shi, J.; Sheikh, M. S.; Huang, Y. DOC45, a Novel DNA Damage–Regulated Nucleocytoplasmic ATPase That Is Overexpressed in Multiple Human Malignancies. *Mol. Cancer Res.* **2010,** *8* (1), 57-66.

11. Matsuzawa, A.; Kanno, S.; Nakayama, M.; Mochiduki, H.; Wei, L.; Shimaoka, T.; Furukawa, Y.; Kato, K.; Shibata, S.; Yasui, A. The BRCA1/BARD1-interacting protein OLA1 functions in centrosome regulation. *Mol. Cell* **2014,** *53* (1), 101-114.

12. Yoshino, Y.; Qi, H.; Fujita, H.; Shirota, M.; Abe, S.; Komiyama, Y.; Shindo, K.; Nakayama, M.; Matsuzawa, A.; Kobayashi, A. BRCA1-interacting protein OLA1 requires interaction with BARD1 to regulate centrosome number. *Mol. Cancer Res.* **2018,** *16* (10), 1499-1511.

13. Zhang, J.; Rubio, V.; Lieberman, M. W.; Shi, Z.-Z. OLA1, an Obg-like ATPase, suppresses antioxidant response via nontranscriptional mechanisms. *P. Natl. Acad. Sci. USA* **2009,** *106* (36), 15356-15361.

14. Mao, R.; Rubio, V.; Chen, H.; Bai, L.; Mansour, O.; Shi, Z.-Z. OLA1 protects cells in heat shock by stabilizing HSP70. *Cell Death Dis.* **2013,** *4* (2), e491.

15. Chen, H.; Song, R.; Wang, G.; Ding, Z.; Yang, C.; Zhang, J.; Zeng, Z.; Rubio, V.; Wang, L.; Zu, N. OLA1 regulates protein synthesis and integrated stress response by inhibiting eIF2 ternary complex formation. *Sci. Rep.* **2015,** *5*, 1324-1340.

16. Fant, X.; Gnadt, N.; Haren, L.; Merdes, A. Stability of the small γ-tubulin complex requires HCA66, a protein of the centrosome and the nucleolus. *J. Cell Sci.* **2009,** *122* (8), 1134-1144.

17. Breitenbach, M.; Eckl, P. Introduction to oxidative stress in biomedical and biological research. *Biomolecules* **2015**, *5* (2), 1169-1177.

18. Mayer, M.; Bukau, B. Hsp70 chaperones: cellular functions and molecular mechanism. *Cell. Mol. Life Sci.* **2005,** *62* (6), 670-684.

19. Qian, S.-B.; McDonough, H.; Boellmann, F.; Cyr, D. M.; Patterson, C. CHIP-mediated stress recovery by sequential ubiquitination of substrates and Hsp70. *Nature* **2006,** *440* (7083), 551-555.

20. Nishitoh, H. CHOP is a multifunctional transcription factor in the ER stress response. *J. Biochem.* **2012,** *151* (3), 217-219.

21. Ohoka, N.; Hattori, T.; Kitagawa, M.; Onozaki, K.; Hayashi, H. Critical and functional regulation of CHOP (C/EBP homologous protein) through the N-terminal portion. *J. Biol. Chem.* **2007,** *282* (49), 35687-35694.

22. Ding, Z.; Liu, Y.; Rubio, V.; He, J.; Minze, L. J.; Shi, Z.-Z. OLA1, a translational regulator of p21, maintains optimal cell proliferation necessary for developmental progression. *Mol. Cell. Biol.* **2016,** *36* (20), 2568-2582.

23. Harper, J. W.; Elledge, S. J.; Keyomarsi, K.; Dynlacht, B.; Tsai, L.-H.; Zhang, P.; Dobrowolski, S.; Bai, C.; Connell-Crowley, L.; Swindell, E. Inhibition of cyclin-dependent kinases by p21. *Mol. Bio. Cell* **1995,** *6* (4), 387-400.

24. Gartel, A. L.; Tyner, A. L. The role of the cyclin-dependent kinase inhibitor p21 in apoptosis 1 supported in part by NIH grant R01 DK56283 (to ALT) for the p21 research and Campus Research Board and Illinois Department of Public Health Penny Severns Breast and Cervical Cancer grants (to ALG). 1. *Mol. Cancer Ther.* **2002,** *1* (8), 639-649.

25. Petkowski, J. J.; Schaner Tooley, C. E.; Anderson, L. C.; Shumilin, I. A.; Balsbaugh, J. L.; Shabanowitz, J.; Hunt, D. F.; Minor, W.; Macara, I. G. Substrate specificity of

mammalian N-terminal α-amino methyltransferase NRMT. *Biochemistry* **2012,** *51* (30), 5942-5950.

26. Petkowski, J. J.; Bonsignore, L. A.; Tooley, J. G.; Wilkey, D. W.; Merchant, M. L.; Macara, I. G.; Tooley, C. E. S. NRMT2 is an N-terminal monomethylase that primes for its homologue NRMT1. *Biochem. J.* **2013,** *456* (3), 453-462.

27. Tooley, C. E. S.; Petkowski, J. J.; Muratore-Schroeder, T. L.; Balsbaugh, J. L.; Shabanowitz, J.; Sabat, M.; Minor, W.; Hunt, D. F.; Macara, I. G. NRMT is an α-N-methyltransferase that methylates RCC1 and retinoblastoma protein. *Nature* **2010,** *466* (7310), 1125-1128.

28. Chen, T.; Muratore, T. L.; Schaner-Tooley, C. E.; Shabanowitz, J.; Hunt, D. F.; Macara, I. G. N-terminal α-methylation of RCC1 is necessary for stable chromatin association and normal mitosis. *Nat. Cell Biol.* **2007,** *9* (5), 596-603.

29. Cai, Q.; Fu, L.; Wang, Z.; Gan, N.; Dai, X.; Wang, Y. α-N-methylation of damaged DNA-binding protein 2 (DDB2) and its function in nucleotide excision repair. *J. Biol. Chem.* **2014**, *289* (23), 16046-16056.

30. Wingfield, P. T. N‐Terminal Methionine Processing. *Curr. Protoc. Protein Sci.* **2017,** *88* (1), 6.14. 1-6.14. 3.

31. Sherman, F.; Stewart, J. W.; Tsunasawa, S. Methionine or not methionine at the beginning of a protein. *Bioessays* **1985,** *3* (1), 27-31.

32. Xiao, Q.; Zhang, F.; Nacev, B. A.; Liu, J. O.; Pei, D. Protein N-terminal processing: substrate specificity of Escherichia coli and human methionine aminopeptidases. *Biochemistry* **2010,** *49* (26), 5588-5599.

33. Dai, X.; Rulten, S. L.; You, C.; Caldecott, K. W.; Wang, Y. Identification and functional characterizations of N-terminal α-N-methylation and phosphorylation of serine 461 in human poly (ADP-ribose) polymerase 3. *J. Proteome Res.* **2015,** *14* (6), 2575-2582.

34. Ran, F. A.; Hsu, P. D.; Wright, J.; Agarwala, V.; Scott, D. A.; Zhang, F. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **2013,** *8* (11), 2281-2308.

35. Coin, I.; Beyermann, M.; Bienert, M. Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. *Nat. Protoc.* **2007,** *2* (12), 3247-3256.

36. Speicher, K.; Kolbas, O.; Harper, S.; Speicher, D. Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies. *J. Biomol. Tech.* **2000,** *11* (2), 74-86.

37. Aurelio, L.; Box, J. S.; Brownlee, R. T.; Hughes, A. B.; Sleebs, M. M. An efficient synthesis of N-methyl amino acids by way of intermediate 5-oxazolidinones. *J. Org. Chem.* **2003,** *68* (7), 2652-2667.

38. HE, L.; WANG, X. Synthesis of stachydrine. *Huaxue Shijie (Chemical World)* **2005,** *5*, 296-298.

Chapter 4 - Conclusions and future perspectives

4.1 Conclusions

Here, an activity-based substrate profiling strategy was performed to uncover the potential targets of NTMT1. For proof of concept, Hey-SAM, as a SAM surrogate containing a terminal alkynyl group, was able to alkylate RCC1 protein with a high efficiency in the presence of wt NTMT1. Cu(I) catalyzed click chemistry was then performed to label the alkylated RCC1 with a biotin group which allowed its purification using streptavidin beads.

The profiling strategy was applied to identify NTMT1 substrates from human cell lysates. NTMT1 was knocked out in HEK293FT cell line by CRISPR-cas9 which resulted in the production of hypomethylated NTMT1 substrates. After alkylation using Hey-SAM and click chemistry, the biotin labelled proteins were enriched by streptavidin beads, separated on SDS-PAGE, digested by trypsin and analyzed by LC-MS. The statistical analysis of identified and quantified proteins together with the protein N-terminal motif analysis identified 70 proteins as putative NTMT1 targets. Among them, three of the previous confirmed NTMT1 substrates including RCC1, SET and RL23A were included. Target validation confirmed that N-terminal peptide of PB1, N-terminal peptide of SPD2B, protein RS14 and protein OLA1 were NTMT1 substrates *in vitro*.

In addition, the *in vivo* investigation of OLA1 N-terminal methylation was carried out in normal and NTMT1 KO cells. Overexpressed OLA1 protein in normal HEK293FT was dimethylated and this modification was absent in NTMT1 KO cells. These results confirm that NTMT1 is the enzyme for catalyzing the N-terminal methylation of OLA1 in human cells. Moreover, our preliminary study reveals that N-terminal methylation of OLA1 does not alter its ATPase activity and subcellular location.

4.2 Merits and limitations

As previously reported, lots of protein lysine and arginine methyltransferases (e.g., PRMT1, PRMT3, EuHMT1, EuHMT2) cannot utilize the bulky SAM analogs (e.g., Hey-SAM)[1-2]. To perform activity-based substrate profiling assay, protein engineering is usually required. One drawback of protein engineering is that the mutants may have different substrate specificity compared to wt enzyme. Here, we reported that a bulky SAM surrogate, Hey-SAM, was able to be taken by wt NTMT1 as a cofactor, which avoided the time-consuming protein engineering. In addition, NTMT1 knockout cells were generated to specifically produce the hypomethylated NTMT1 substrates. Previously reported method for inducing hypomethylated proteins relies on Adox, an inhibitor of SAH hydrolase, which is not specific. One concern for our approach is that NTMT2, the reported monomethyltransferase, might have the potential to monomethylate the endogenous NTMT1 substrates. However, none of its physiological substrates were confirmed so far. Moreover, monomethylated OLA1-EGFP-Flag was not observed in NTMT1 KO cells according to our LC-MS results, which demonstrated that NTMT2 might have very low or no activity towards OLA1 in human cells. These results indicate that it is feasible to generate hypomethylated NTMT1 substrates using NTMT1 KO cells. Another strength of our approach is the application of statistics and motif analysis in substrates identification, which narrows down the number of potential targets and points to the biologically meaningful substrates.

Despite that 70 proteins were identified as putative NTMT1 substrates, not all of the known substrates (e.g., DDB2, PARP3) were included[3-4]. Possible reasons include that some proteins may have very low expression level and some proteins were not properly labeled. Moreover, target validation showed that the identified NTMT1 putative targets contain some false positive proteins (e.g., H2A1B, DDX60L). This may be caused by the nonspecific labeling and high expression

level of some housekeeping proteins. Though motif analysis can eliminate partial of the false positive protein, further improvements (e.g., higher labeling efficiency, lower nonspecific binding) and novel approach are necessary to solve this challenging issue for MS-based protein identification.

4.3 Future perspectives

As protein N-terminal methylation plays important roles in mitosis and DNA damage repair, and dysregulation of NTMT1 promotes the growth and metastasis potential of cancer cells, it would be important to elucidate the function of N-terminal methylation by uncovering substrates of NTMT1. Here, our activity-based substrate profiling approach uncovered several new substrates, however, it has its intrinsic limitations. For future perspectives, it is important to exploit new strategies to identify more of the true substrates.

Jakobsson et al. reported an HRMS-based substrate profiling approach for METTL13-N[5]. In this method, proteins were labelled using SILAC for better quantification, and the corresponding enzyme METTL13 was knocked out[5]. Proteins obtained from normal and METTL13 KO cell lysates which contained the methylation sites were identified, quantified and compared[5]. Application of this method requires two prerequisites. First, substrates should be modified specifically by the target enzyme and the enzyme KO would abolish the modification. Second, a high-resolution mass spectrometry is required for precise protein identification and quantification.

Hu et al. reported a strategy to profile substrates of O-linked β-N-acetylglucosamine transferase (OGT) using electrophilic probes[6]. This strategy provides some hints on eliminating the nonspecific protein labeling. The electrophilic probes containing an allyl halide group (e.g., allyl bromide, allyl chloride) bind to OGT and react with the adjacent cysteine to form a covalent bond[6]. After group transfer reaction catalyzed by OGT, the substrates were covalently attached to

the complex of OGT and the probe[6]. His_6-tagged OGT could be purified by Ni-NTA resin and its substrates would be pulled-down together as well[6]. This strategy is very specific as only the true substrates can form a complex with the target enzyme. A proposed strategy for profiling NTMT1 targets using an electrophilic SAM analog is shown in Fig 4.1. NTMT1 should be engineered to introduce a cysteine residue in the SAM binding pocket to react with the electrophilic SAM analog. However, several limitations should be considered before performing this approach. First, the electrophilic probe is very reactive and may react with some surface residues (e.g., cysteine), which will result in nonspecific labeling. Second, the probe is not stable in basic condition and it would a challenge for enzyme to work in a weak acid solution. Third, the mechanism for methylation reaction is S_N2, and the steric hindrance generated by the NTMT1-probe complex might slow down the reaction. Therefore, it would be important to optimize the probe and engineer NTMT1.

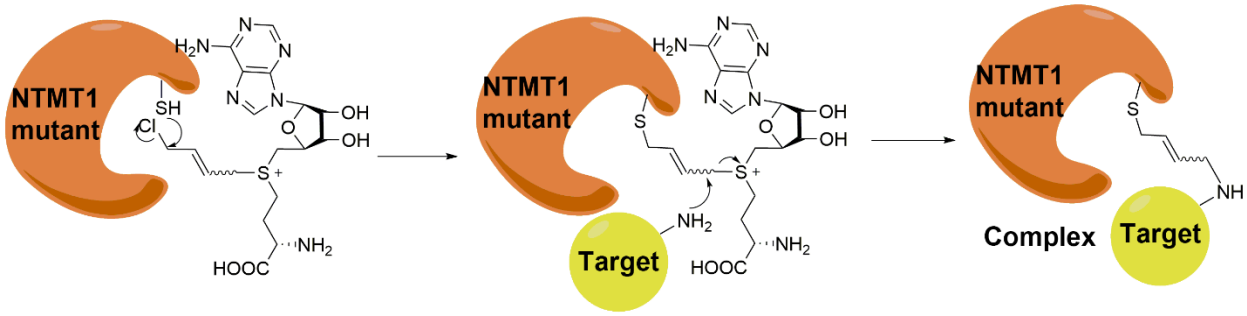


Figure 4.1 Proposed strategy for NTMT1 substrates profiling using an electrophilic probe.

Another future perspective is to elucidate the function of OLA1 N-terminal methylation. Profiling the methylation readers will be a good start point. In chapter 1, several methods for methylation reader profiling including immunoprecipitation, photo-affinity crosslinking and yeast three-hybrid system have been reviewed. It will be necessary to try these methods to uncover the methylation readers. Moreover, it is worthwhile to test the existence of protein N-terminal

demethylase by performing a demethylase assay. All these studies will help us understand this important N-terminal methylation.

4.4 References

1. Wang, R.; Zheng, W.; Yu, H.; Deng, H.; Luo, M. Labeling substrates of protein arginine methyltransferase with engineered enzymes and matched S-adenosyl-L-methionine analogues. *J. Am. Chem. Soc.* **2011,** *133* (20), 7648-7651.

2. Guo, H.; Wang, R.; Zheng, W.; Chen, Y.; Blum, G.; Deng, H.; Luo, M. Profiling substrates of protein arginine N-methyltransferase 3 with S-adenosyl-L-methionine analogues. *ACS Chem. Biol.* **2013,** *9* (2), 476-484.

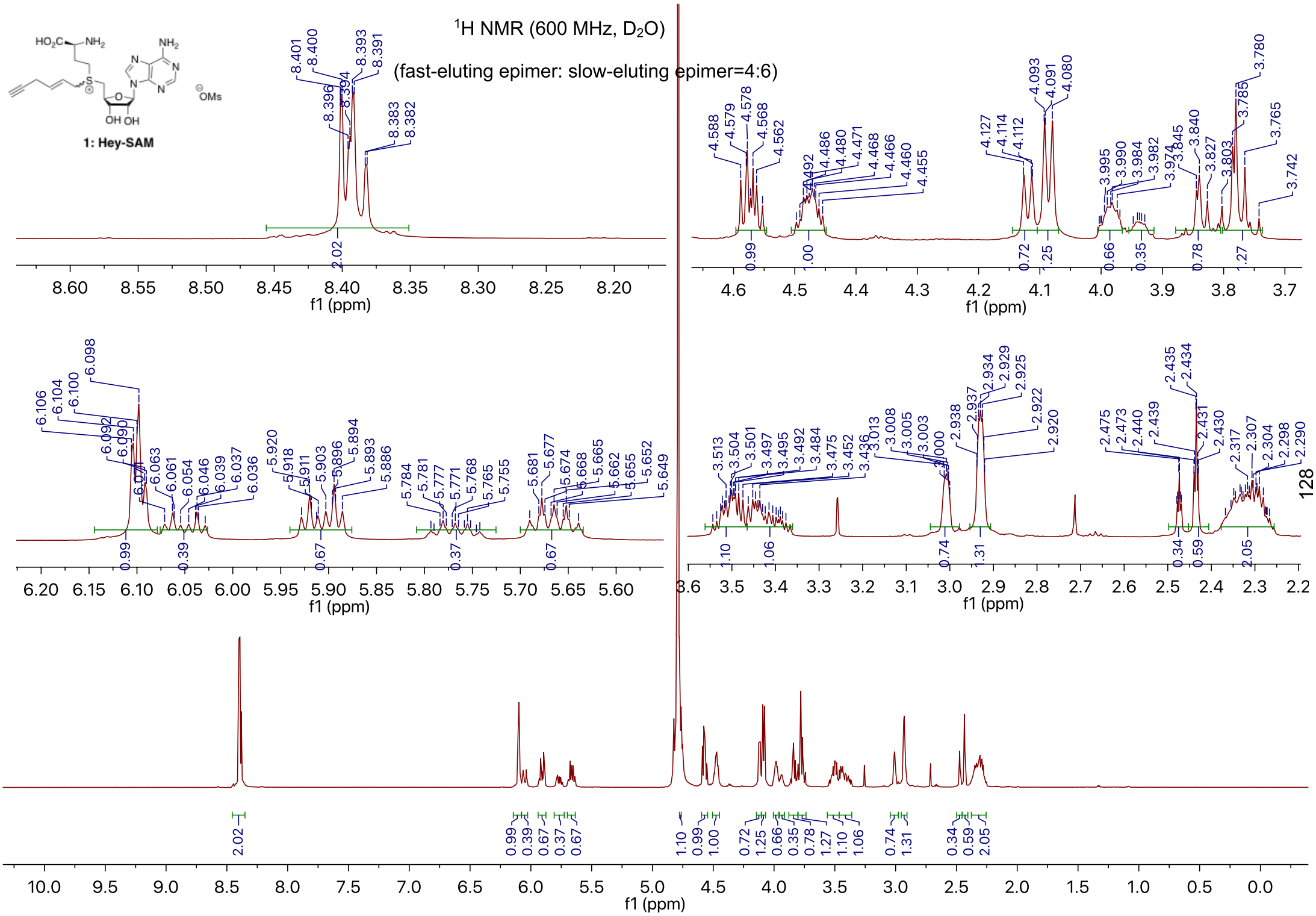
3. Cai, Q.; Fu, L.; Wang, Z.; Gan, N.; Dai, X.; Wang, Y. α-N-methylation of damaged DNA-binding protein 2 (DDB2) and its function in nucleotide excision repair. *J. Biol. Chem.* **2014**, *289* (23), 16046-16056.

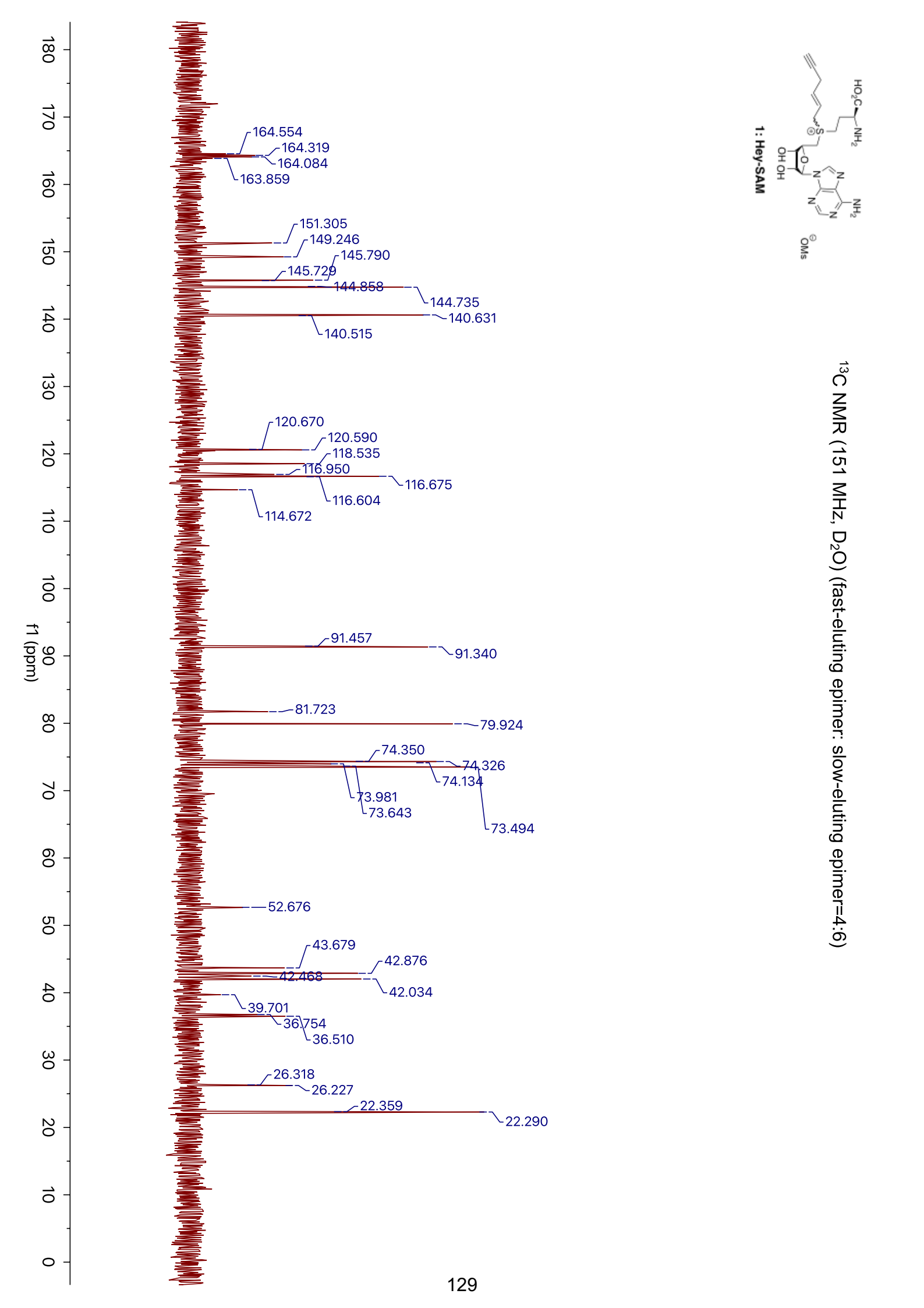
4. Dai, X.; Rulten, S. L.; You, C.; Caldecott, K. W.; Wang, Y. Identification and functional characterizations of N-terminal α-N-methylation and phosphorylation of serine 461 in human poly (ADP-ribose) polymerase 3. *J. Proteome Res.* **2015,** *14* (6), 2575-2582.

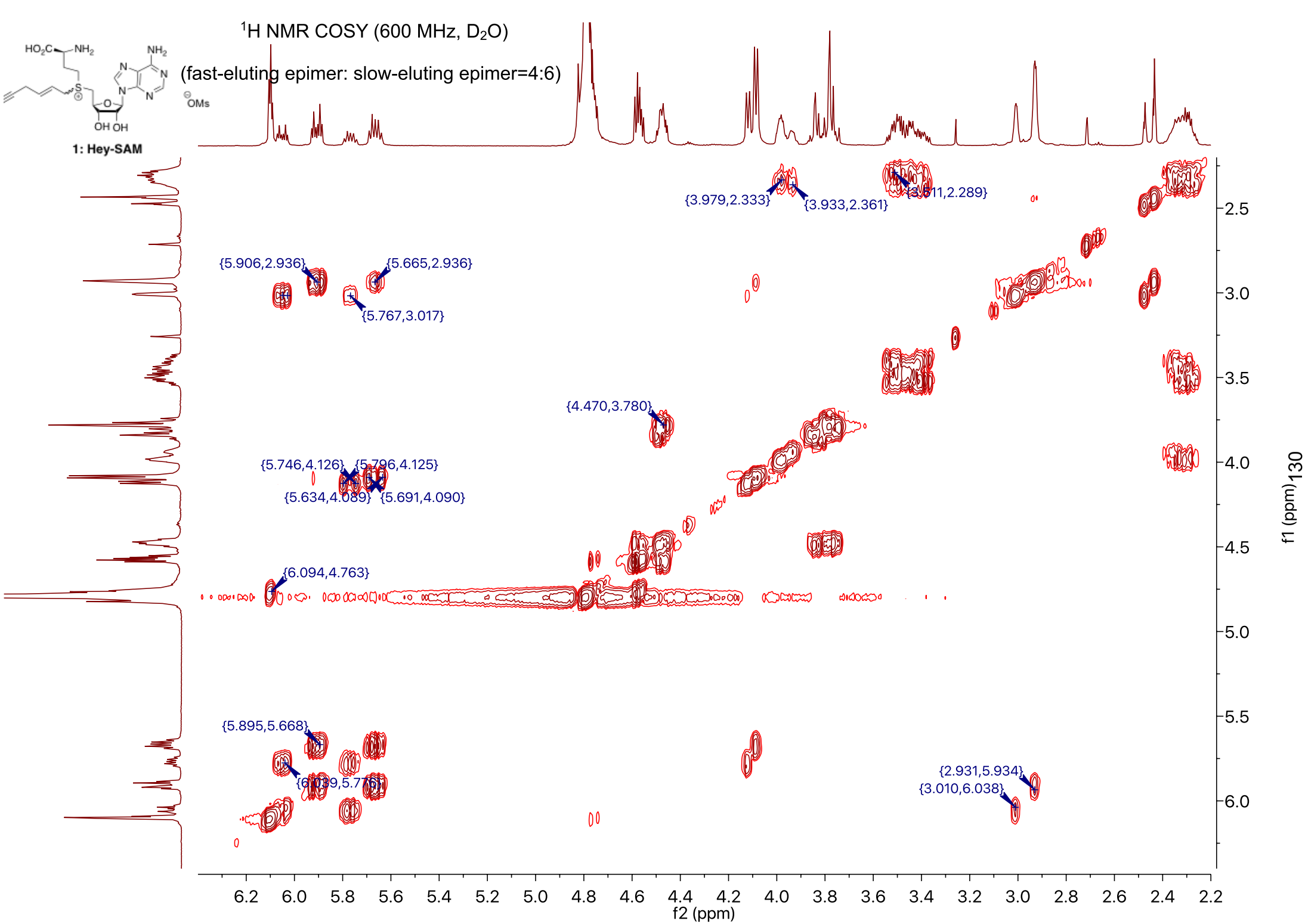
5. Jakobsson, M. E.; Małecki, J. M.; Halabelian, L.; Nilges, B. S.; Pinto, R.; Kudithipudi, S.; Munk, S.; Davydova, E.; Zuhairi, F. R.; Arrowsmith, C. H. The dual methyltransferase METTL13 targets N terminus and Lys55 of eEF1A and modulates codon-specific translation rates. *Nat. Commun.* **2018,** *9* (1), 3411-3425.

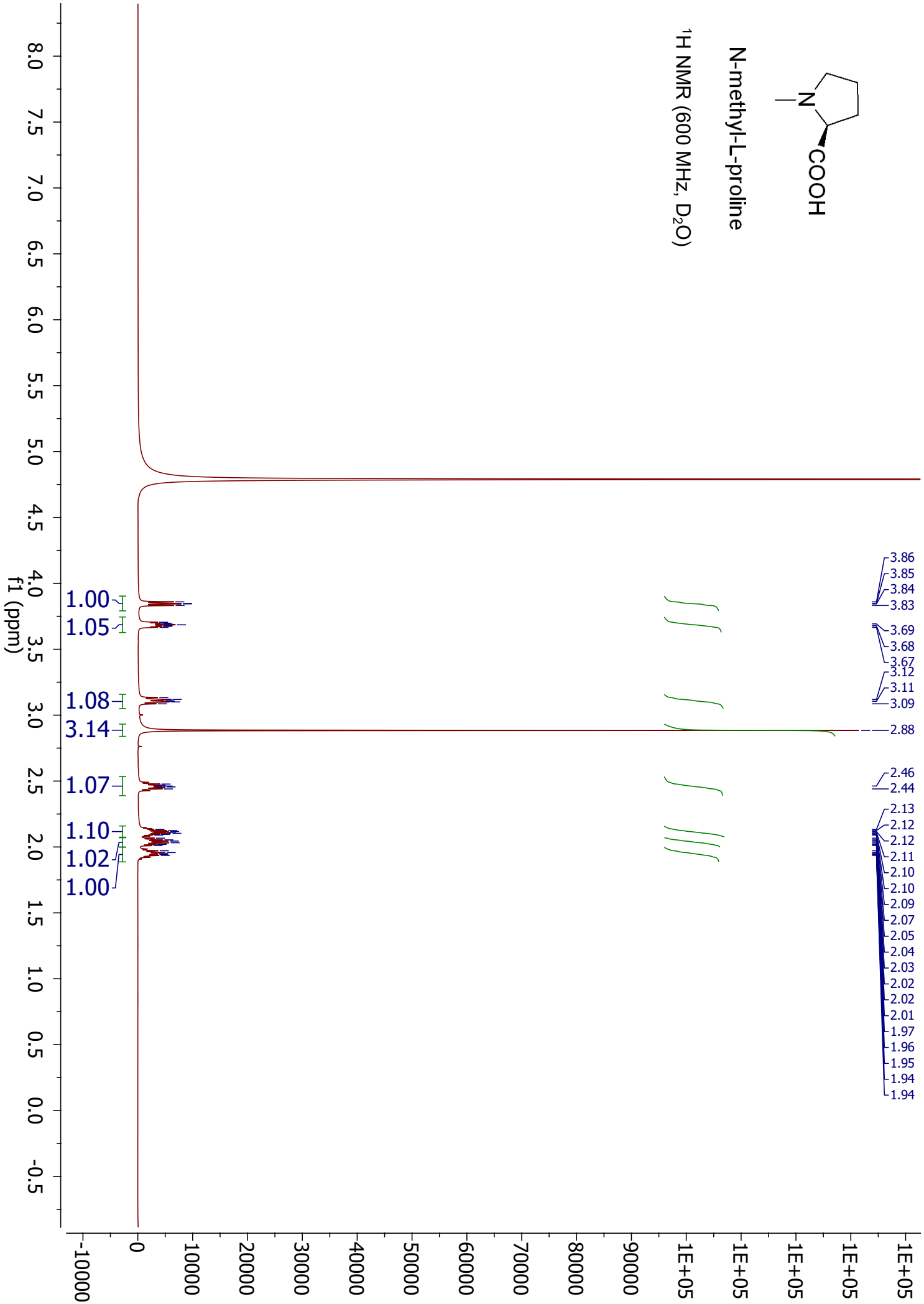
6. Hu, C.-W.; Worth, M.; Fan, D.; Li, B.; Li, H.; Lu, L.; Zhong, X.; Lin, Z.; Wei, L.; Ge, Y. Electrophilic probes for deciphering substrate recognition by O-GlcNAc transferase. *Nat. Chem. Biol.* **2017,** *13* (12), 1267-1273.

Appendix A - ^{1}H NMR and ^{13}C NMR spectrums

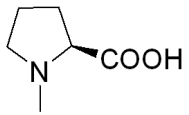

1: Hey-SAM
^{1}H NMR (600 MHz, D_2O)
(fast-eluting epimer: slow-eluting epimer=4:6)
f1 (ppm)


164.554
164.319
164.084
163.859
151.305
149.246
145.790
145.729
144.858
144.735
140.631
140.515
120.670
120.590
118.535
116.950
116.675
116.604
114.672
91.457
91.340
81.723
79.924
74.350
74.326
74.134
73.981
73.643
73.494
52.676
43.679
42.876
42.468
42.034
39.701
36.754
36.510
26.318
26.227
22.359
22.290
f1 (ppm)
1: Hey-SAM
^{13}C NMR (151 MHz, D_2O) (fast-eluting epimer: slow-eluting epimer=4:6)


^{1}H NMR COSY (600 MHz, D_2O)
(fast-eluting epimer: slow-eluting epimer=4:6)
1: Hey-SAM
{3.979,2.333}
{3.933,2.361}
{3.911,2.289}
{5.906,2.936}
{5.665,2.936}
{5.767,3.017}
{4.470,3.780}
{5.746,4.126}
{5.796,4.125}
{5.634,4.089}
{5.691,4.090}
{6.094,4.763}
{5.895,5.668}
{2.931,5.934}
{3.010,6.038}
f2 (ppm)
f1 (ppm)

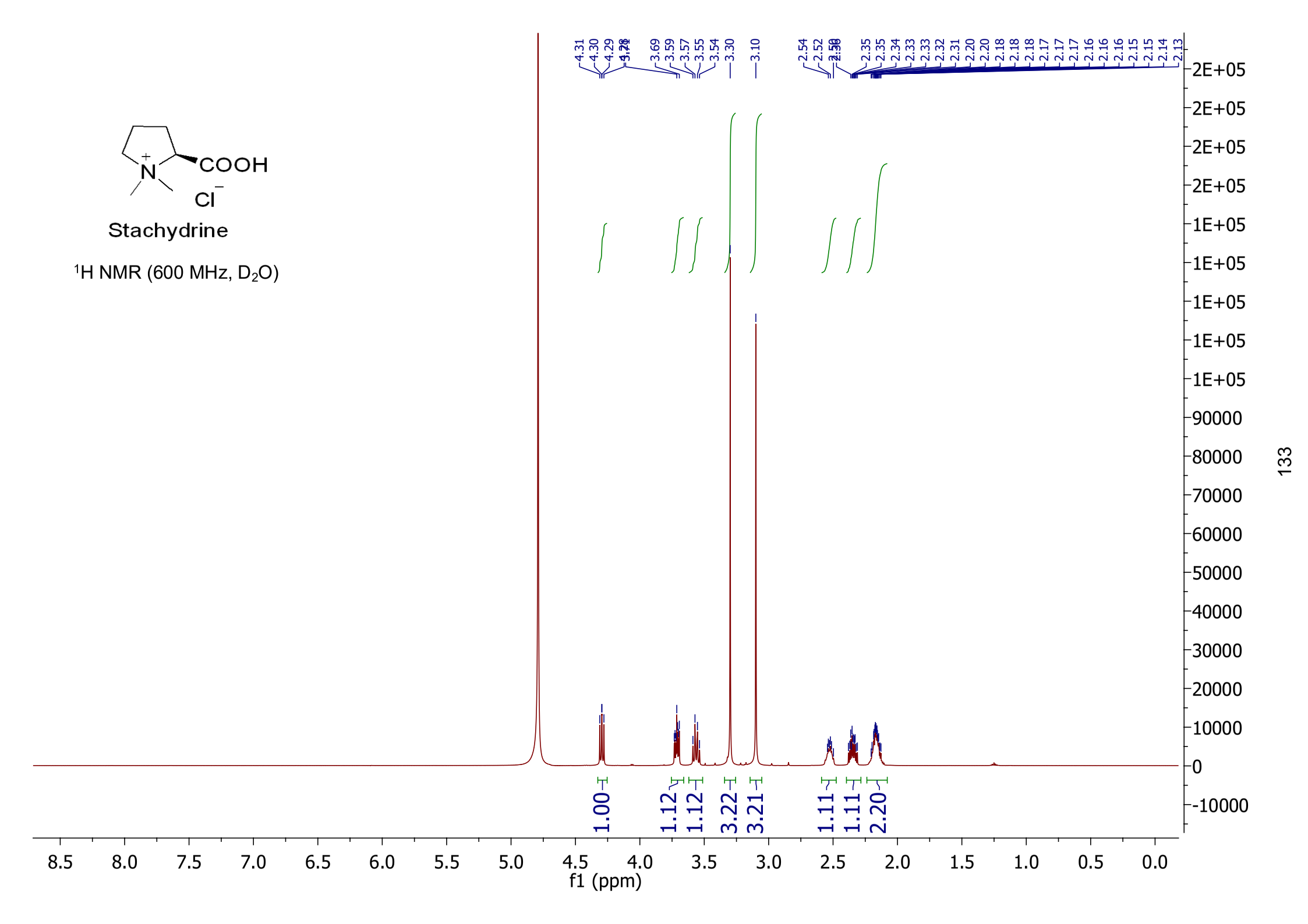

N-methyl-L-proline
^{1}H NMR (600 MHz, D_2O)
COOH
f1 (ppm)
1.00
1.05
1.08
3.14
1.07
1.10
1.02
1.00
3.86
3.85
3.84
3.83
3.69
3.68
3.67
3.12
3.11
3.09
2.88
2.46
2.44
2.13
2.12
2.12
2.11
2.10
2.10
2.09
2.07
2.05
2.04
2.03
2.02
2.02
2.01
1.97
1.96
1.95
1.94
1.94

^{13}C NMR (151 MHz, D_2O) δ 173.65 (s, -1H), 70.49 (s, 9H), 56.19 (s, 10H), 40.56 (s, 8H), 28.69 (s, 7H), 22.72 (s, 5H).



N-methyl-L-proline

^{13}C NMR (151 MHz, D_2O)


COOH
Cl⁻
Stachydrine
¹H NMR (600 MHz, D₂O)
4.31
4.30
4.29
3.69
3.59
3.57
3.55
3.54
3.30
3.10
2.54
2.52
2.35
2.35
2.34
2.33
2.33
2.32
2.31
2.20
2.20
2.18
2.18
2.17
2.17
2.17
2.16
2.16
2.16
2.15
2.15
2.14
2.13
1.00
1.12
1.12
3.22
3.21
1.11
1.11
2.20
f1 (ppm)
133

^{13}C NMR (151 MHz, D_2O)
Stachydrine
N
+
Cl
−
COOH
169.18
74.31
67.70
52.11
46.05
24.55
18.47
f1 (ppm)
210
190
170
150
130
110
90
80
70
60
50
40
30
20
10
0
-5000
0
5000
10000
15000
20000
25000
30000
35000
40000
45000
50000
55000

Appendix B - Candidate substrates of NTMT1

The previously identified NTMT1 substrates were highlighted with yellow color; the validated targets either at peptide level or at protein level are highlighted with green color. Note: protein subcellular distributions are assigned based on UniProt database and some of the locations are predicted by similarity.

UniProt ID	Gene name	Unique peptides	Confidence score	Anova (p)	Max fold change	Subcellular location
P08670	VIM	17	449.57	7.82E-06	2.88	Cytoplasm
P53396	ACLY	34	439.42	9.73E-05	2.95	Cytoplasm
Q9NR22	PRMT8	12	289.60	1.34E-05	6.37	Cell membrane
P54652	HSPA2	4	246.86	0.000118	2.97	Cytoplasm
P22102	GART	11	173.66	0.000116	2.52	Cytoplasm
P13667	PDIA4	10	137.95	5.01E-05	2.40	Cytoplasm
P04181	OAT	13	115.58	8.47E-05	2.88	Cytoplasm
Q5H9U9	DDX60L	5	92.45	5.94E-05	4.12	Unknown
Q05BV3	EML5	4	90.08	7.60E-05	3.16	Cytoplasm
P37802	TAGLN2	5	75.66	0.000104	3.12	Cytoplasm
P62750	RPL23A	5	70.68	4.02E-05	4.77	Cytoplasm
Q6GYQ0	RALGAPA1	4	68.95	5.95E-05	2.36	Cytoplasm, nucleus
Q8N7X0	ADGB	1	64.13	0.002031	2.91	Unknown
P15121	AKR1B1	2	59.92	2.11E-06	3.15	Cytoplasm
P42336	PIK3CA	2	58.43	0.000209	4.44	Cytoplasm
P18754	RCC1	4	56.32	5.01E-05	3.46	Cytoplasm, nucleus
Q9Y6M1	IGF2BP2	2	50.81	6.73E-05	2.85	Cytoplasm, nucleus
P46776	RPL27A	3	45.86	3.58E-05	3.28	Cytoplasm

P61158	ACTR3	4	43.57	0.000311	2.70	Cytoplasm, nucleus
P35232	PHB	1	43.24	0.001095	3.05	Cytoplasm
Q659A1	ICE2	2	42.68	0.000314	3.08	Nucleus
Q6UB35	MTHFD1L	2	42.05	0.00011	2.43	Cytoplasm
O00160	MYO1F	1	38.42	1.53E-05	7.83	Cytoplasm
B4DIP2	ERBIN	2	38.40	0.000237	2.79	Unknown
Q9Y4E1	WASHC2C	2	37.10	0.000257	8.23	Cytoplasm
Q641Q2	WASHC2A	2	37.10	0.000257	8.23	Cytoplasm
Q8IYD8	FANCM	4	36.06	0.000281	3.22	Nucleus
Q9Y5S2	CDC42BPB	3	34.31	2.07E-05	3.44	Cell membrane
P62263	RPS14	2	34.10	0.000113	4.20	Cytoplasm, nucleus
Q16576	RBBP7	1	32.22	0.000883	4.23	Nucleus
Q8NBS9	TXNDC5	2	29.80	0.000341	2.77	Cytoplasm
Q8TF47	ZFP90	1	29.54	0.000505	2.82	Nucleus
P04908	HIST1H2AB	1	29.50	0.00172	2.57	Nucleus
P0C0S8	HIST1H2AG	1	29.50	0.00172	2.57	Nucleus
P20671	HIST1H2AD	1	29.50	0.00172	2.57	Nucleus
Q16777	HIST2H2AC	1	29.50	0.00172	2.57	Nucleus
Q6FI13	HIST2H2AA3	1	29.50	0.00172	2.57	Nucleus
Q7L7L0	HIST3H2A	1	29.50	0.00172	2.57	Nucleus
Q93077	HIST1H2AC	1	29.50	0.00172	2.57	Nucleus
Q96KK5	HIST1H2AH	1	29.50	0.00172	2.57	Nucleus
Q99878	HIST1H2AJ	1	29.50	0.00172	2.57	Nucleus
Q9BTM1	H2AFJ	1	29.50	0.00172	2.57	Nucleus
P16104	H2AFX	1	29.50	0.00172	2.57	Nucleus
Q8IUE6	HIST2H2AB	1	29.50	0.00172	2.57	Nucleus
Q96QV6	HIST1H2AA	1	29.50	0.00172	2.57	Nucleus

Q13017	ARHGAP5	2	26.46	0.000342	2.05	Cytoplasm, cell membrane
Q9Y2Z9	COQ6	1	26.40	7.67E-05	2.74	Cytoplasm
P62805	HIST1H4A	1	26.25	0.000414	2.72	Nucleus
P10606	COX5B	1	26.03	0.000181	4.29	Cytoplasm
Q8IWG1	WDR63	1	23.87	0.000666	3.67	Cytoplasm
Q8NB90	SPATA5	2	23.41	7.21E-05	3.09	Cytoplasm
P46459	NSF	2	21.53	0.000219	2.44	Cytoplasm
Q01955	COL4A3	2	20.50	7.26E-06	8.67	Extracellular matrix
Q2PPJ7	RALGAPA2	2	20.03	0.000238	3.61	Cytoplasm
Q5T5U3	ARHGAP21	2	19.69	1.60E-05	3.01	Cytoplasm
Q9UQB3	CTNND2	1	17.29	0.004354	2.48	Nucleus
Q9NTK5	OLA1	1	16.12	0.000527	3.57	Cytoplasm, nucleus
Q12965	MYO1E	2	15.73	0.000128	2.90	Cytoplasm
Q86U86	PBRM1	1	15.47	4.33E-06	7.29	Nucleus
Q9BQ39	DDX50	1	12.22	5.70E-05	2.94	Nucleus
Q86VW1	SLC22A16 OCT6	1	11.67	1.39E-05	3.90	Membrane
Q6ZRP0	PRR23C	1	11.44	5.58E-06	5.67	Unknown
Q01105	SET	2	11.44	8.35E-05	4.14	Cytoplasm
Q96L34	MARK4	2	11.40	0.000136	2.85	Cytoplasm
Q9Y3F4	STRAP	1	10.70	0.000349	2.99	Cytoplasm, nucleus
Q6NSJ2	PHLDB3	1	9.44	0.000226	2.32	Unknown
Q96GM5	SMARCD1	1	6.01	0.001129	3.81	Nucleus

P42679	MATK	1	5.76	5.39E-05	2.67	Cytoplasm, cell membrane
A1X283	SH3PXD2B	1	5.37	0.000273	2.94	Cytoplasm
P26358	DNMT1	1	4.99	3.55E-05	6.33	Nucleus

Appendix C - MS raw data

Figure 2.2D (TOF transformed)

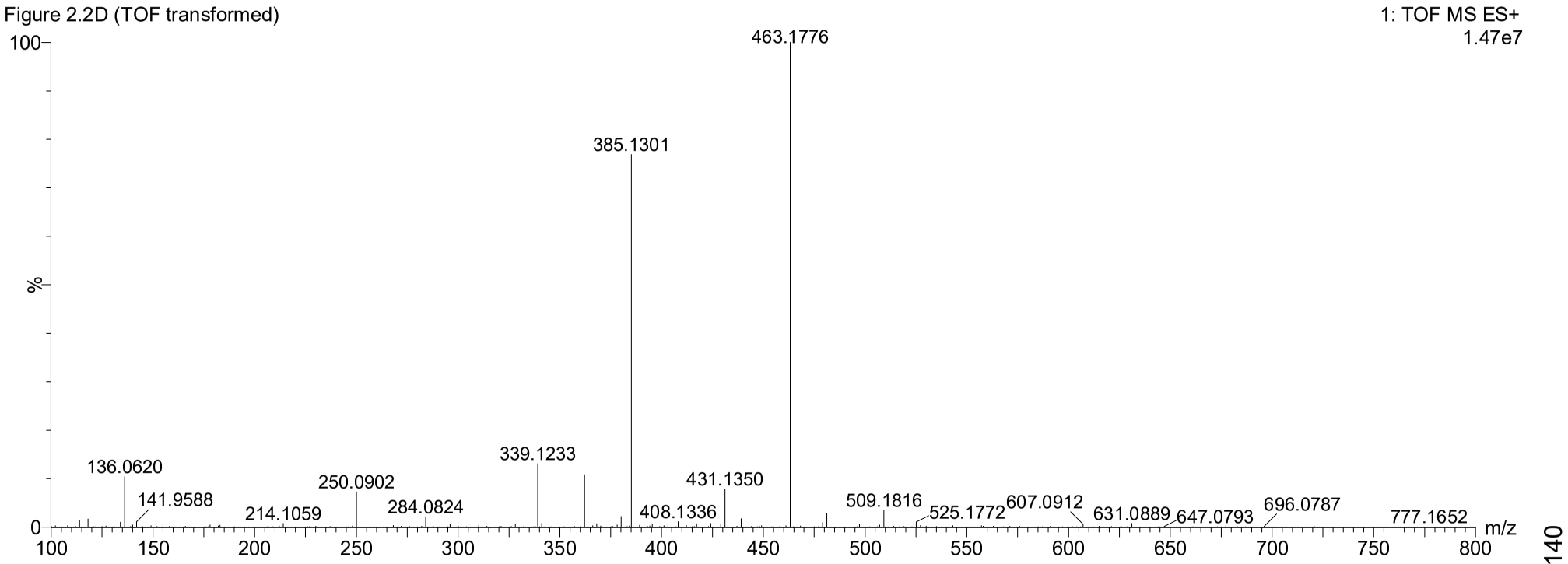
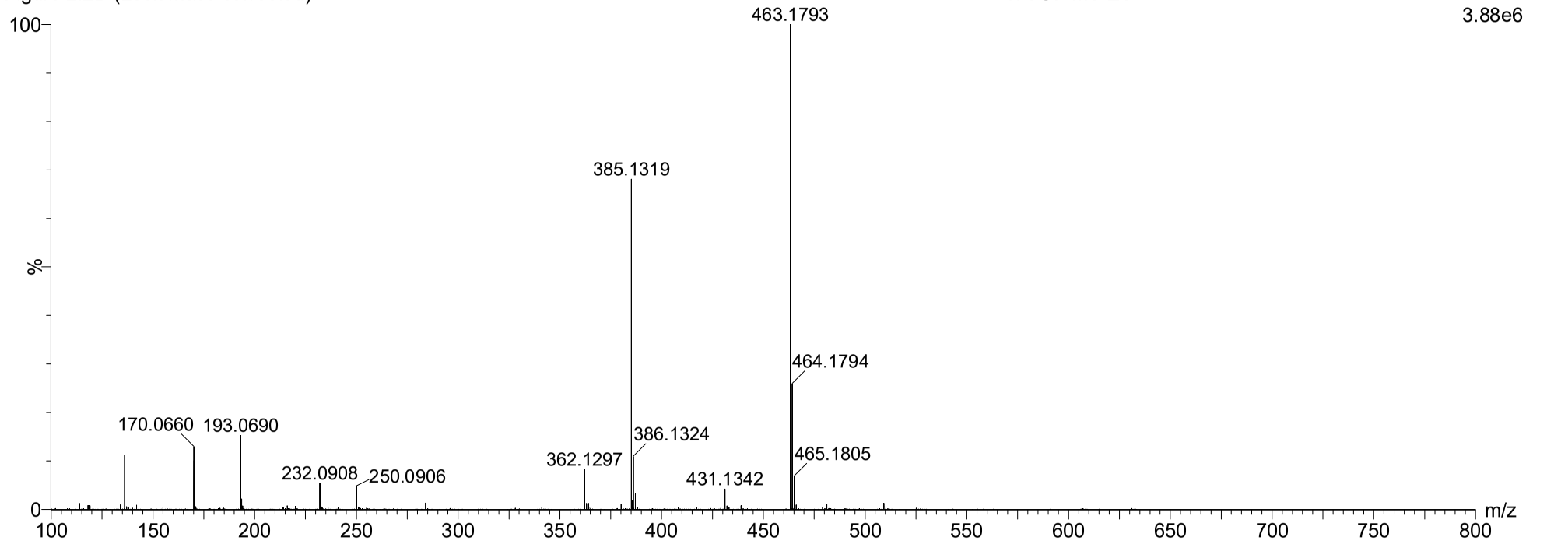


Figure 2.2D (Lock mass corrected)



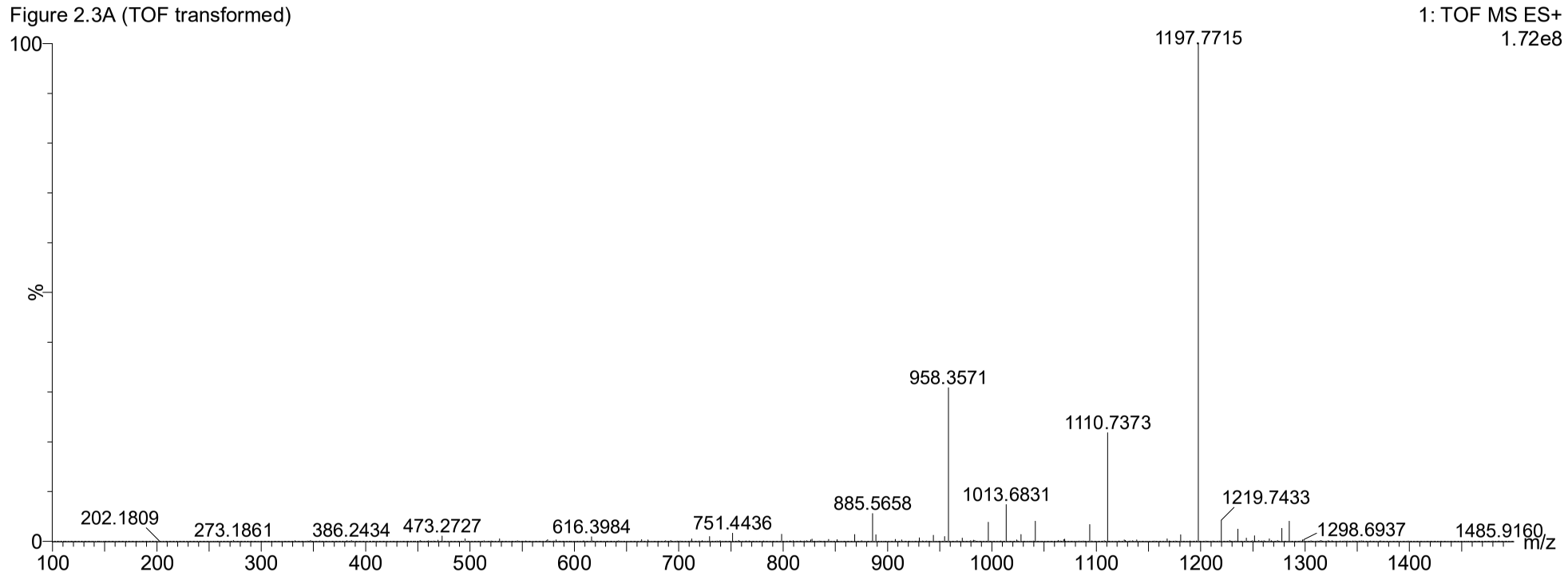


Figure 2.3A (TOF transformed)
1: TOF MS ES+
1.72e8
100
%
0
1197.7715
958.3571
1110.7373
1013.6831
1219.7433
885.5658
202.1809
273.1861
386.2434
473.2727
616.3984
751.4436
1298.6937
1485.9160
m/z
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400

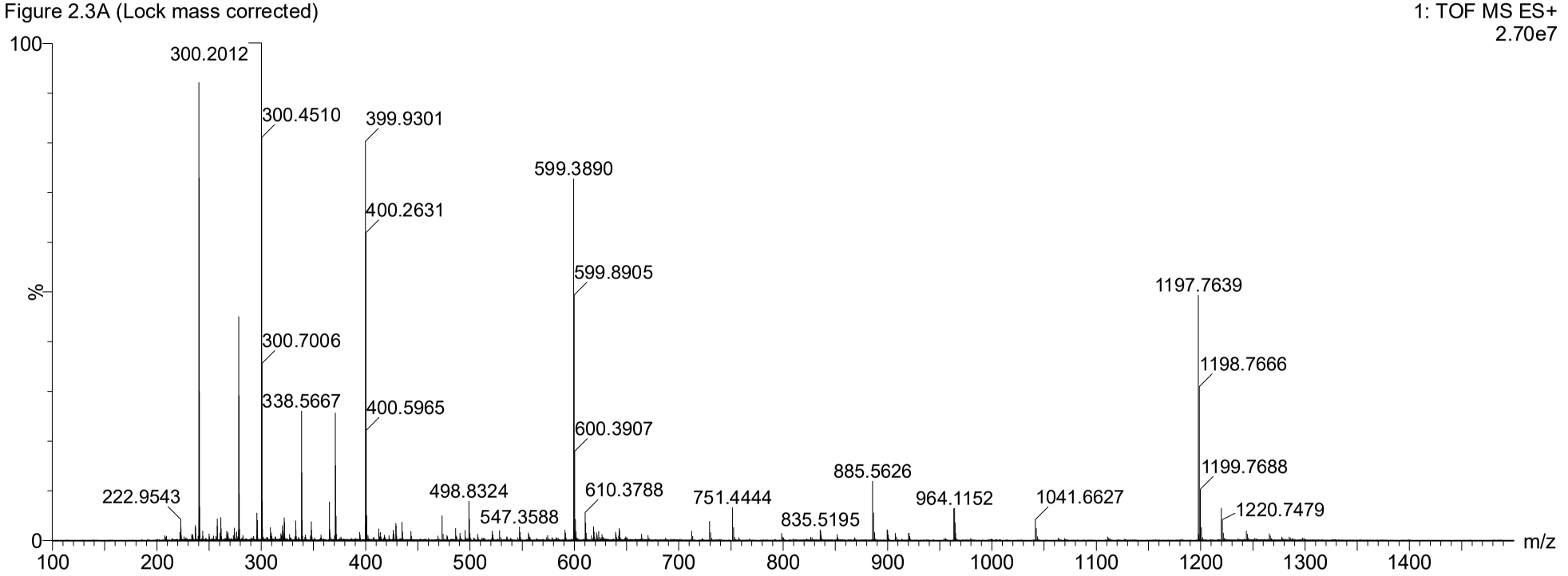


Figure 2.3A (Lock mass corrected)
1: TOF MS ES+
2.70e7
100
%
0
300.2012
300.4510
399.9301
599.3890
400.2631
599.8905
1197.7639
300.7006
1198.7666
338.5667
400.5965
600.3907
1199.7688
885.5626
222.9543
498.8324
610.3788
751.4444
964.1152
1041.6627
547.3588
835.5195
1220.7479
m/z
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400

Figure 2.3B (TOF transformed)

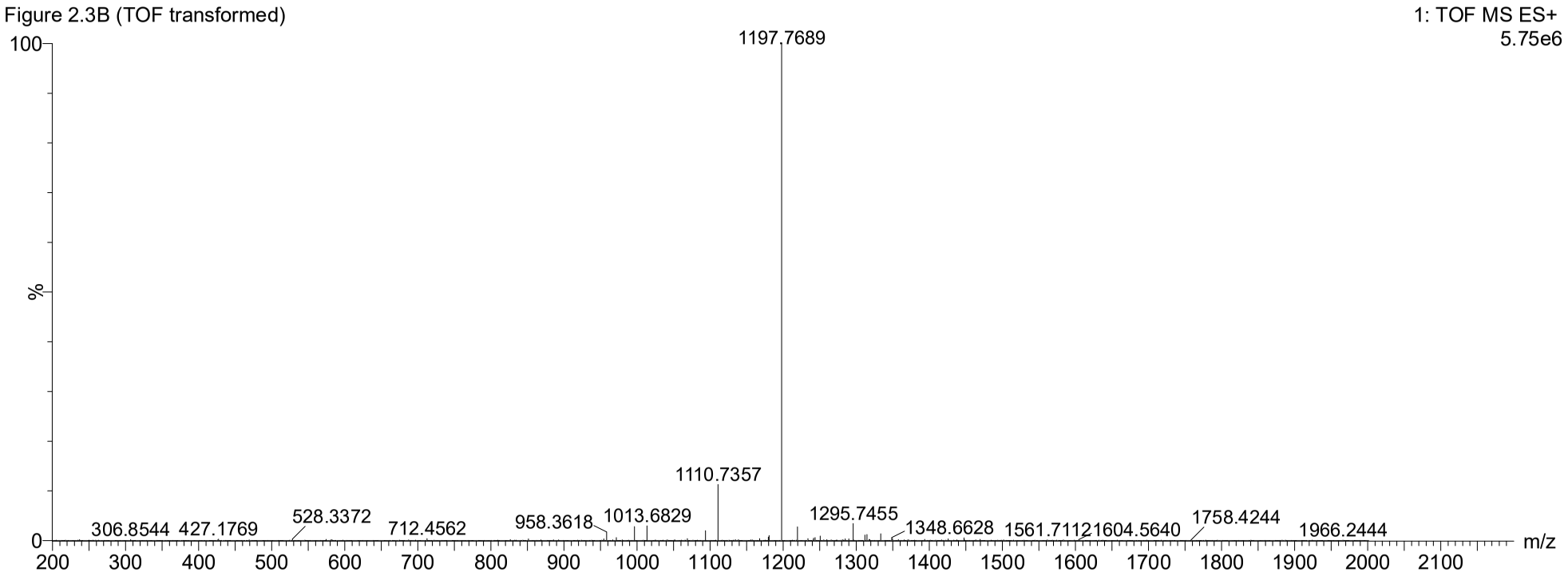


Figure 2.3B (Lock mass corrected)

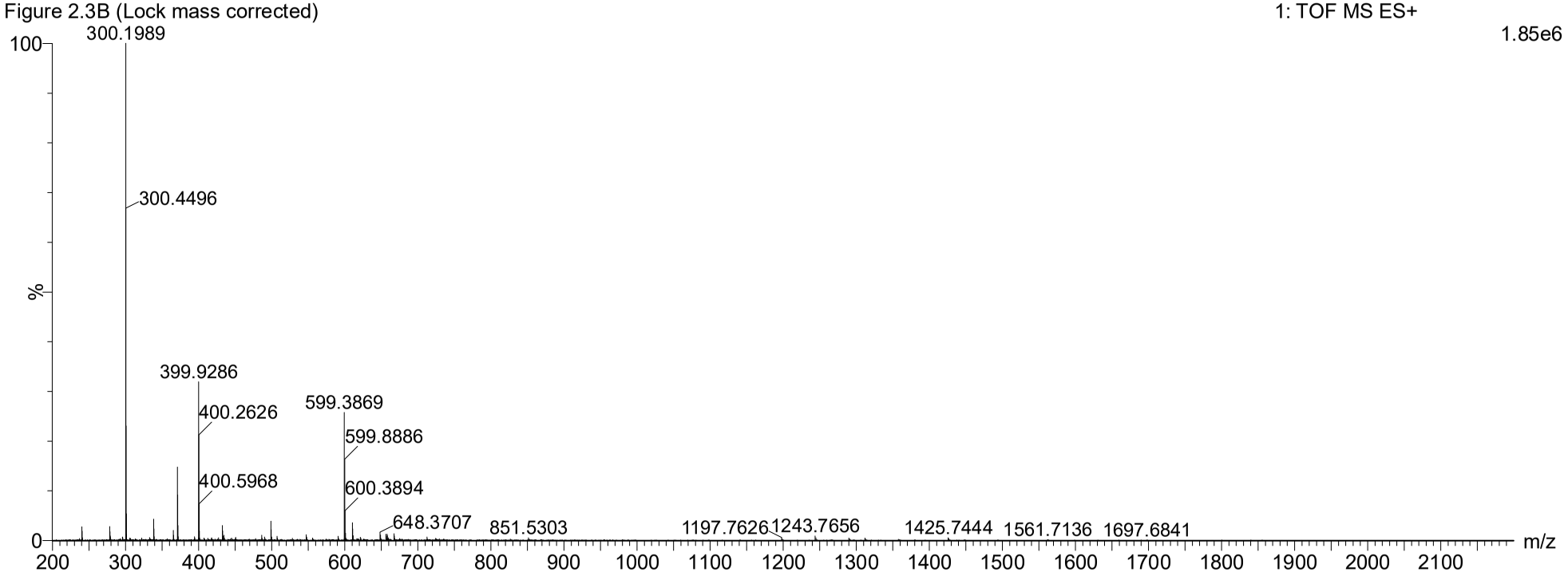


Figure 2.3C (TOF transformed)

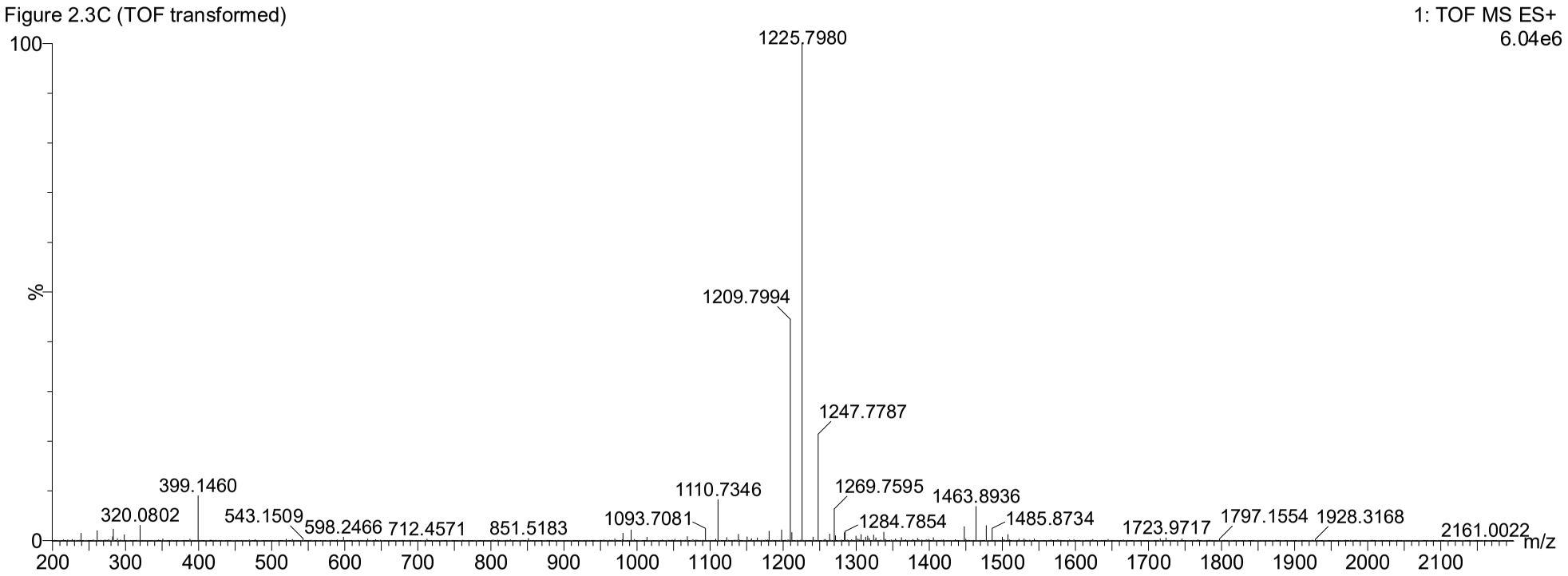


Figure 2.3C (Lock mass corrected)

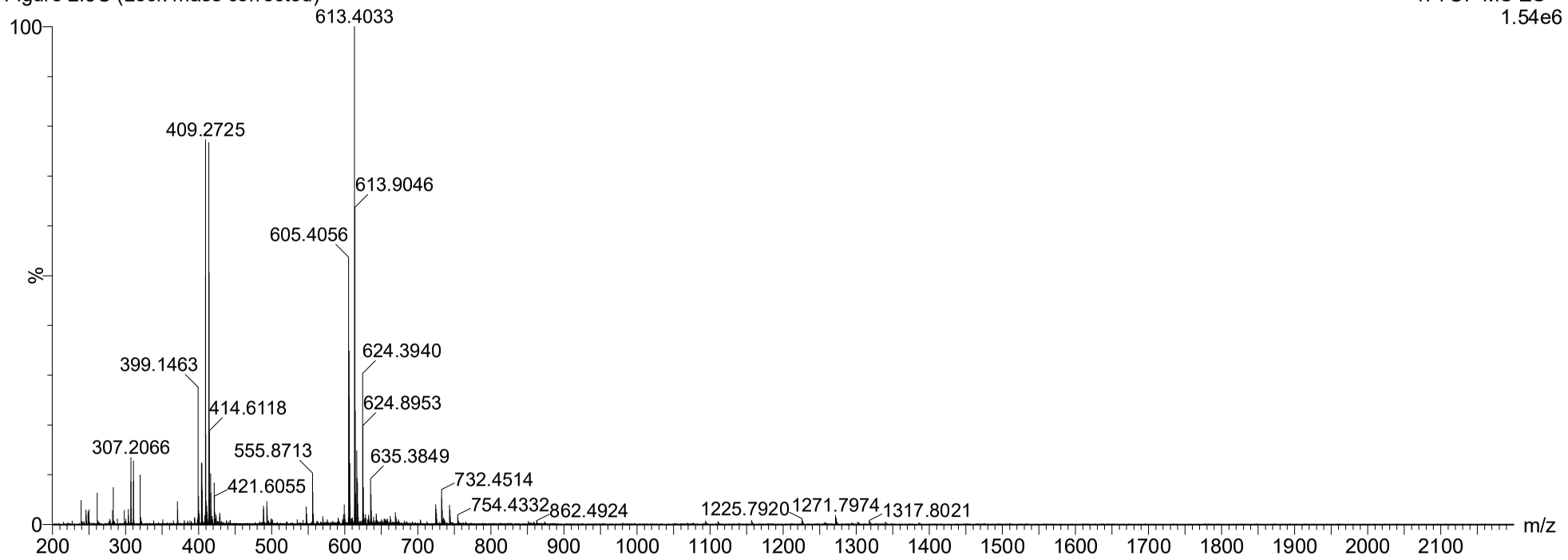


Figure 2.3D (TOF transformed)

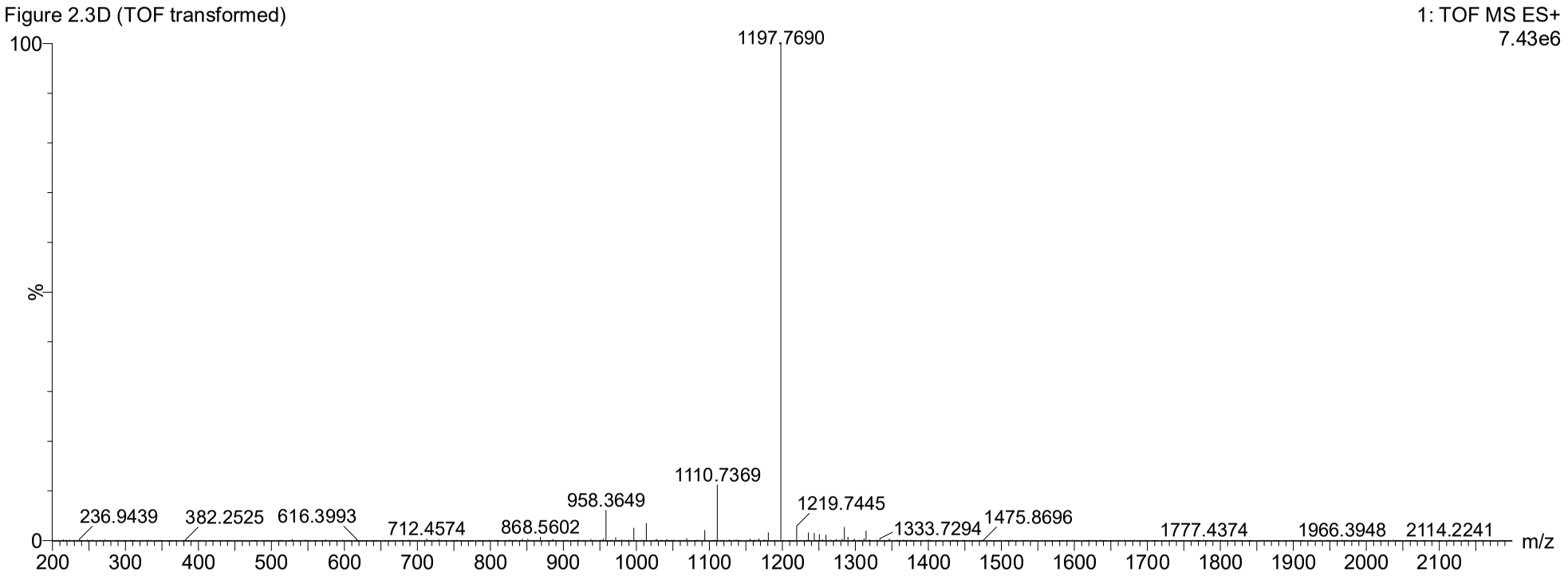


Figure 2.3D (Lock mass corrected)

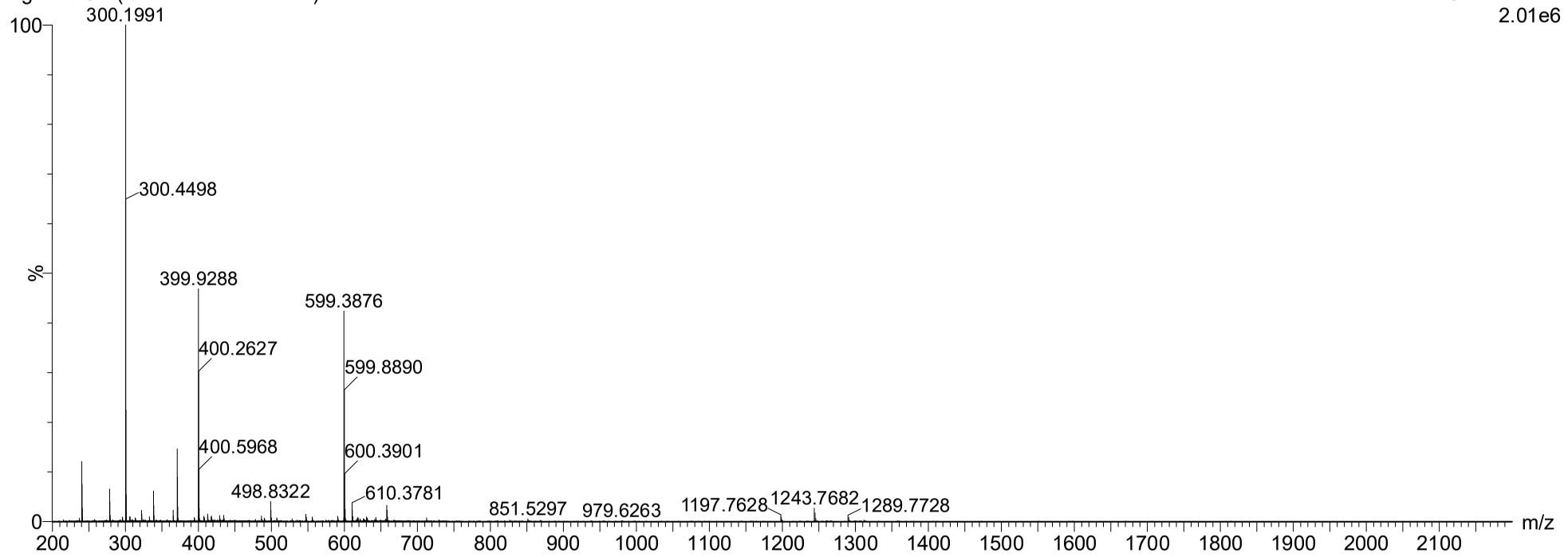


Figure 2.3E (TOF transformed)

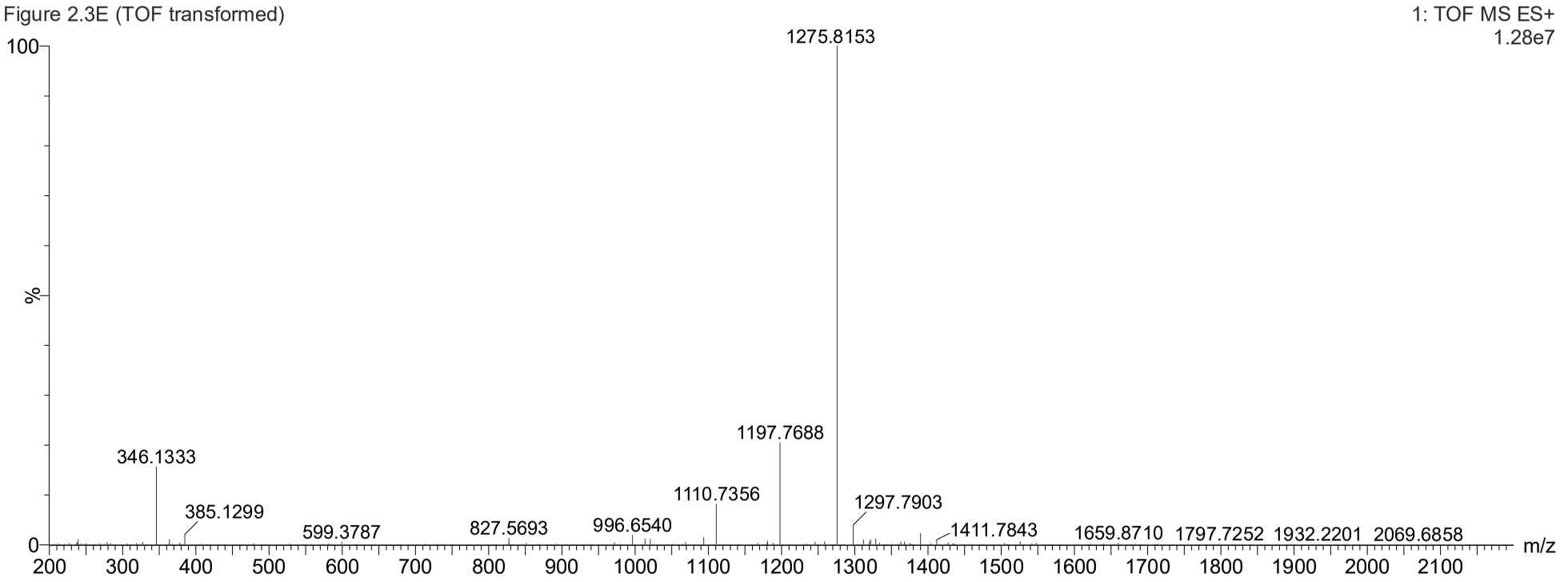


Figure 2.3E (Lock mass corrected)

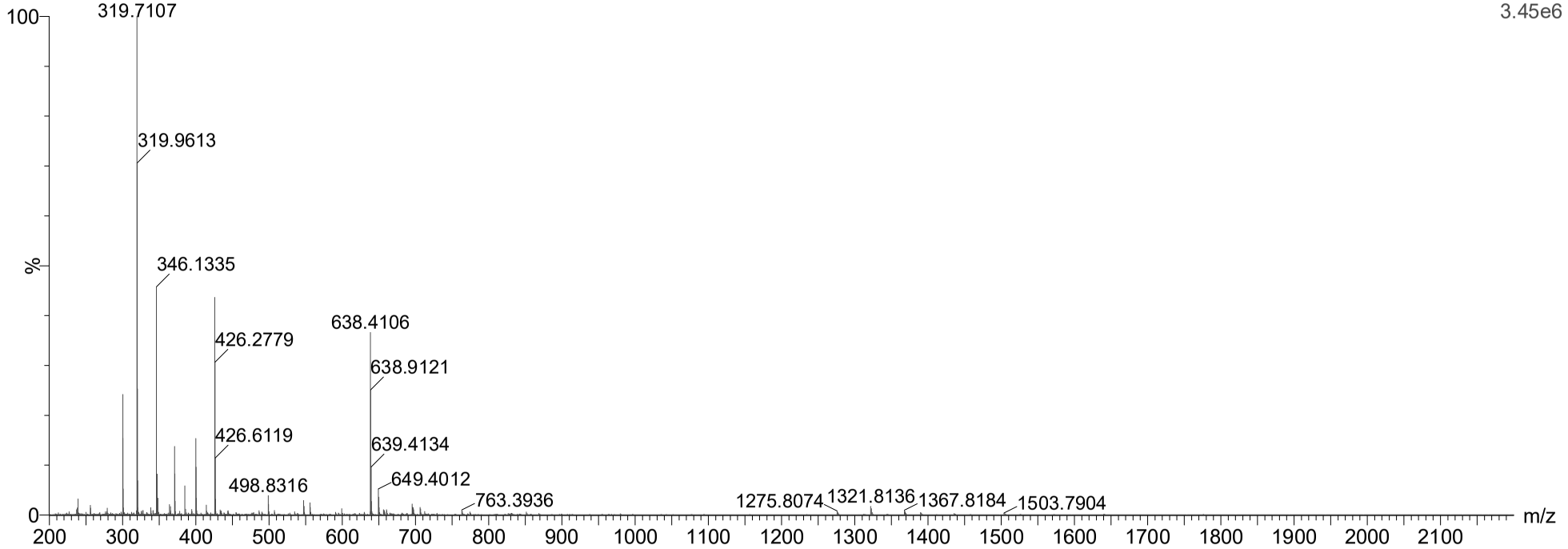


Figure 2.4A (smoothed and background subtracted)

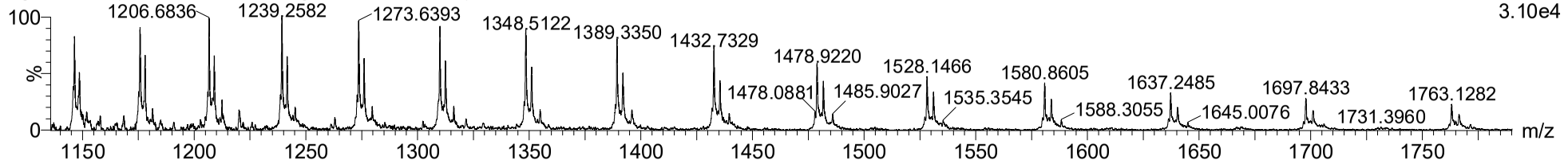


Figure 2.4C (smoothed and background subtracted)

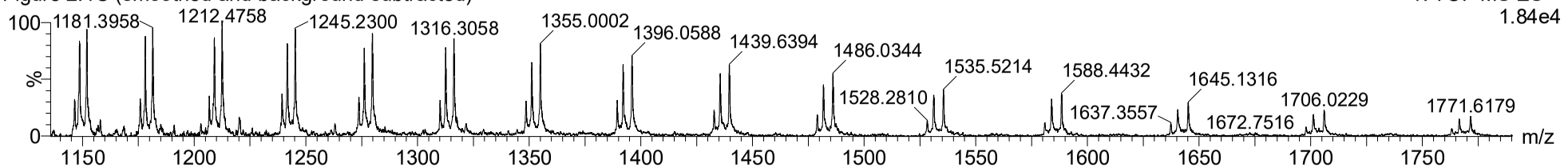


Figure 2.4D (smoothed and background subtracted)

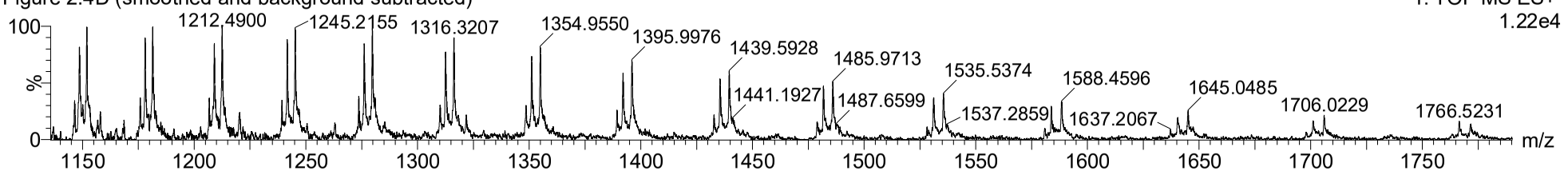


Figure 2.4E (smoothed and background subtracted)

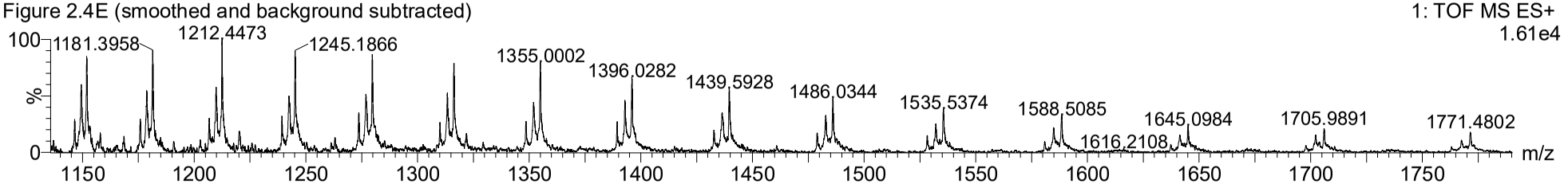


Figure 2.4F (smoothed and background subtracted)

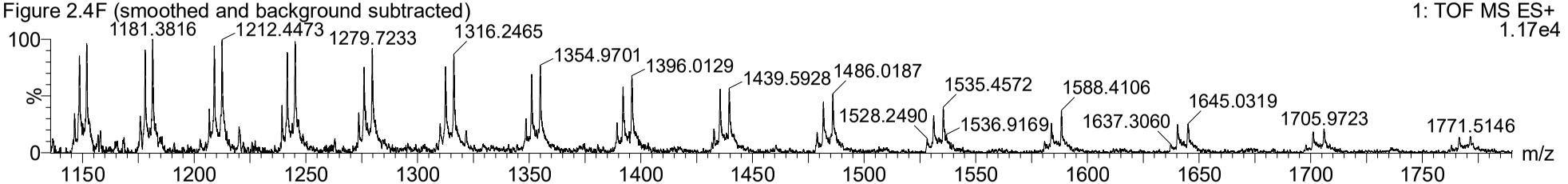


Figure 2.4G (smoothed and background subtracted)

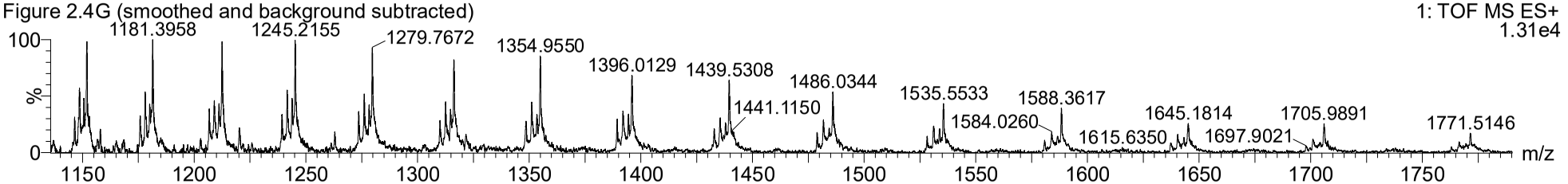


Figure 2.4H (TOF transformed)

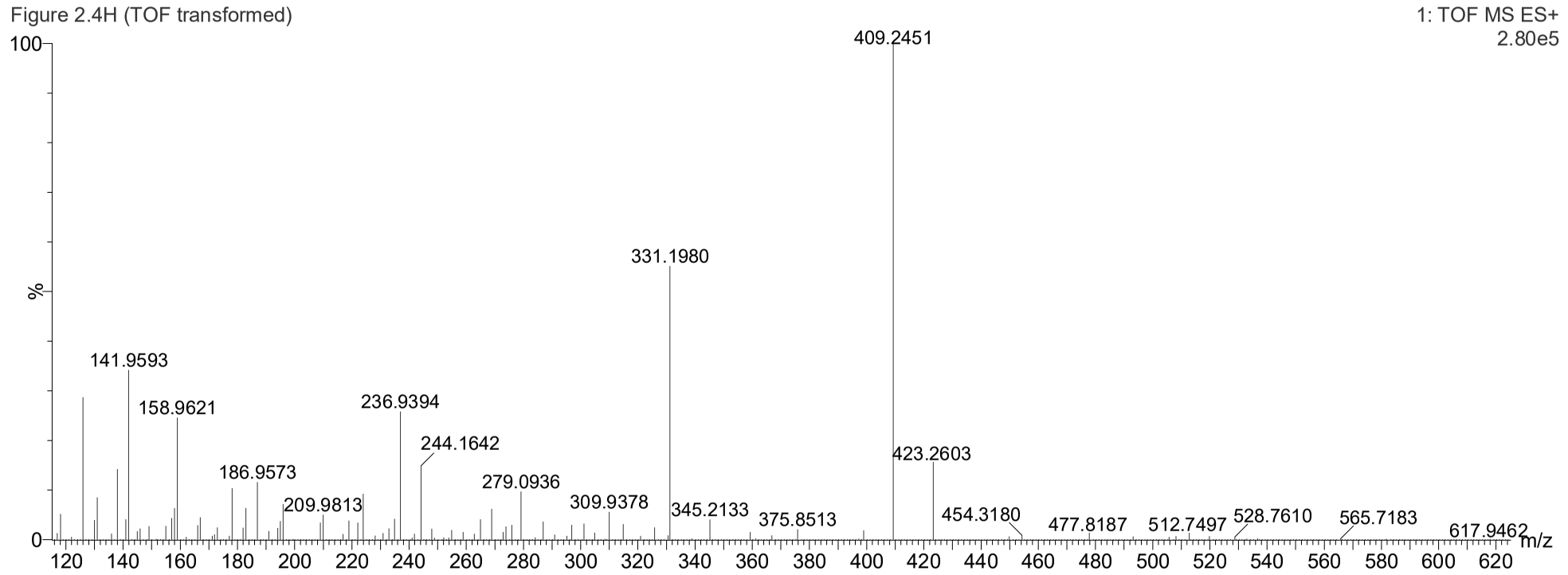


Figure 2.4H (Lock mass corrected)

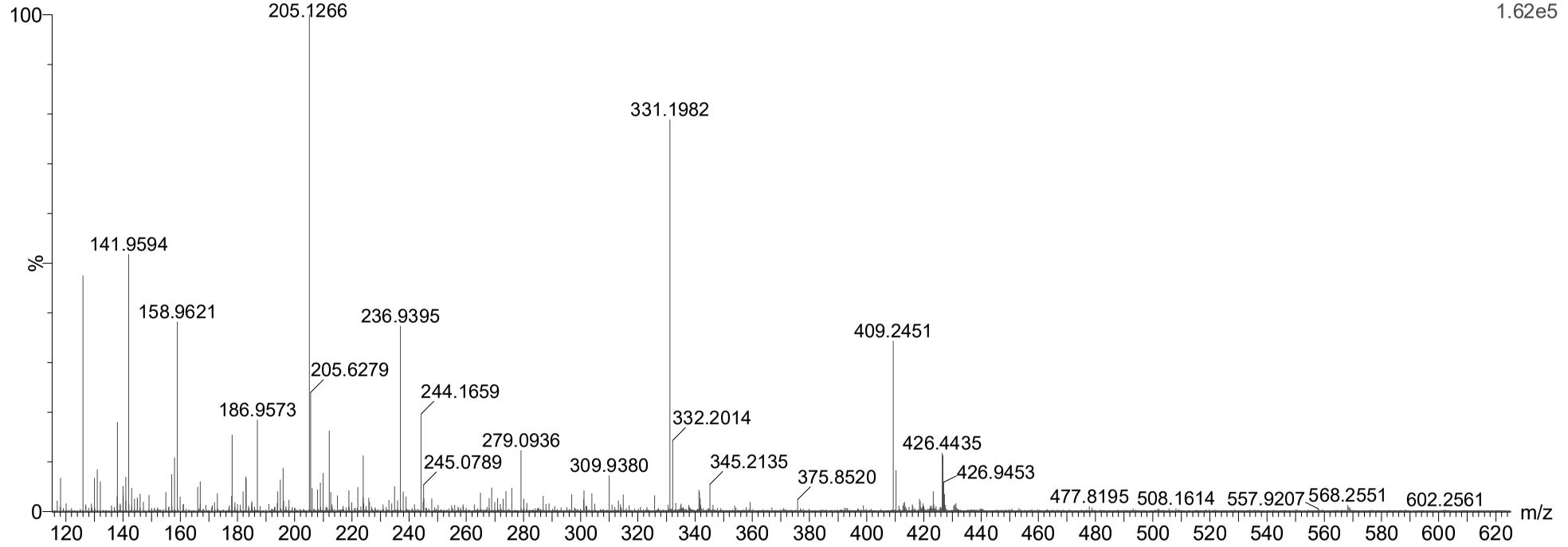


Figure 2.7A (TOF transformed)

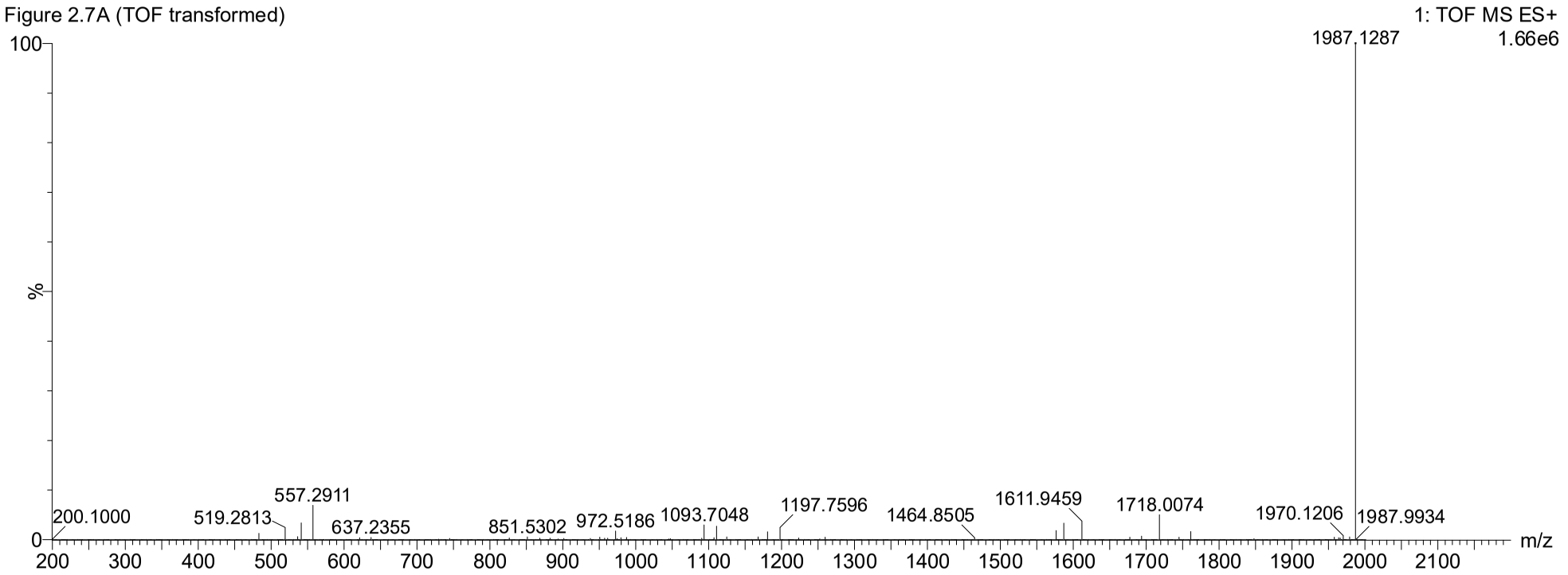


Figure 2.7A (Lock mass corrected)

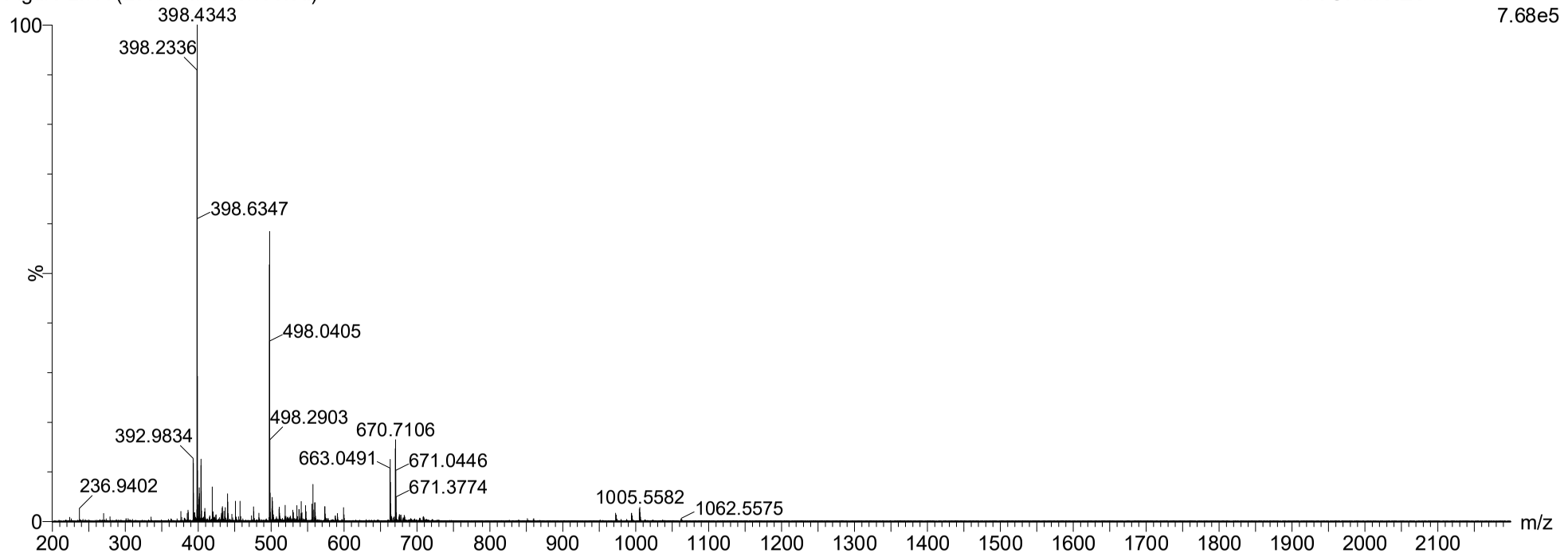


Figure 2.7C (Precursor, TOF transformed)

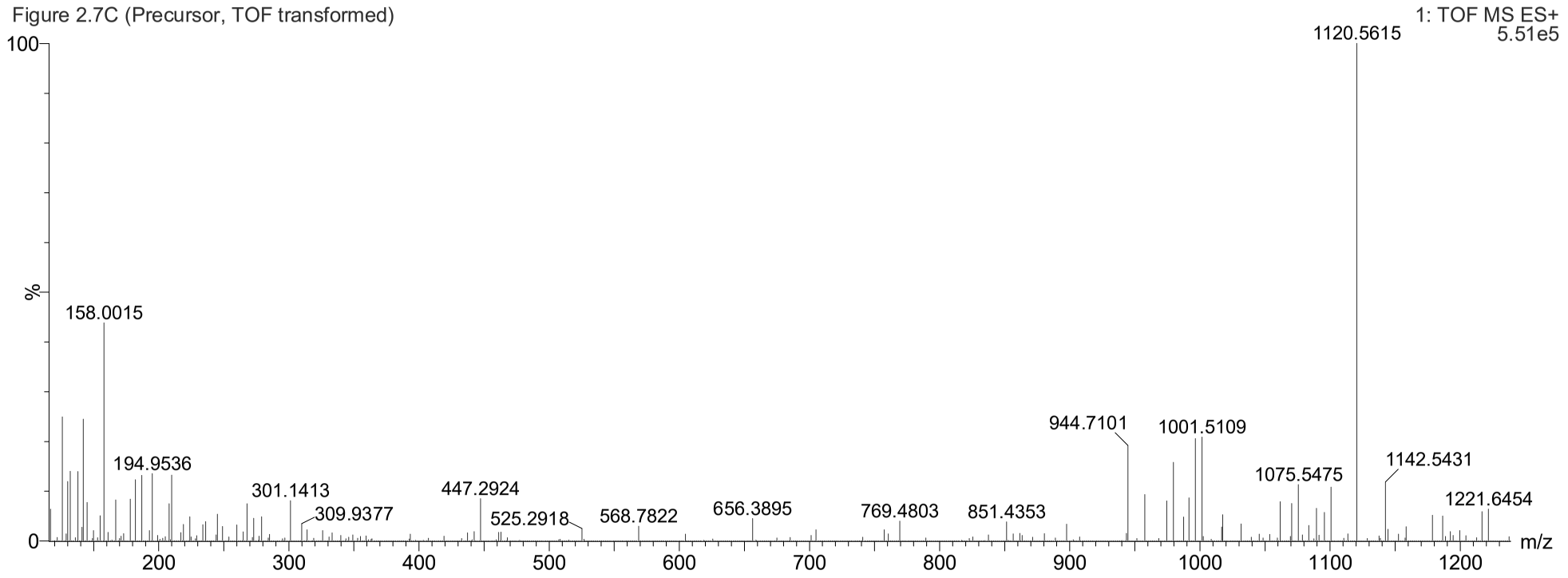


Figure 2.7C (Precursor, Lock mass corrected)

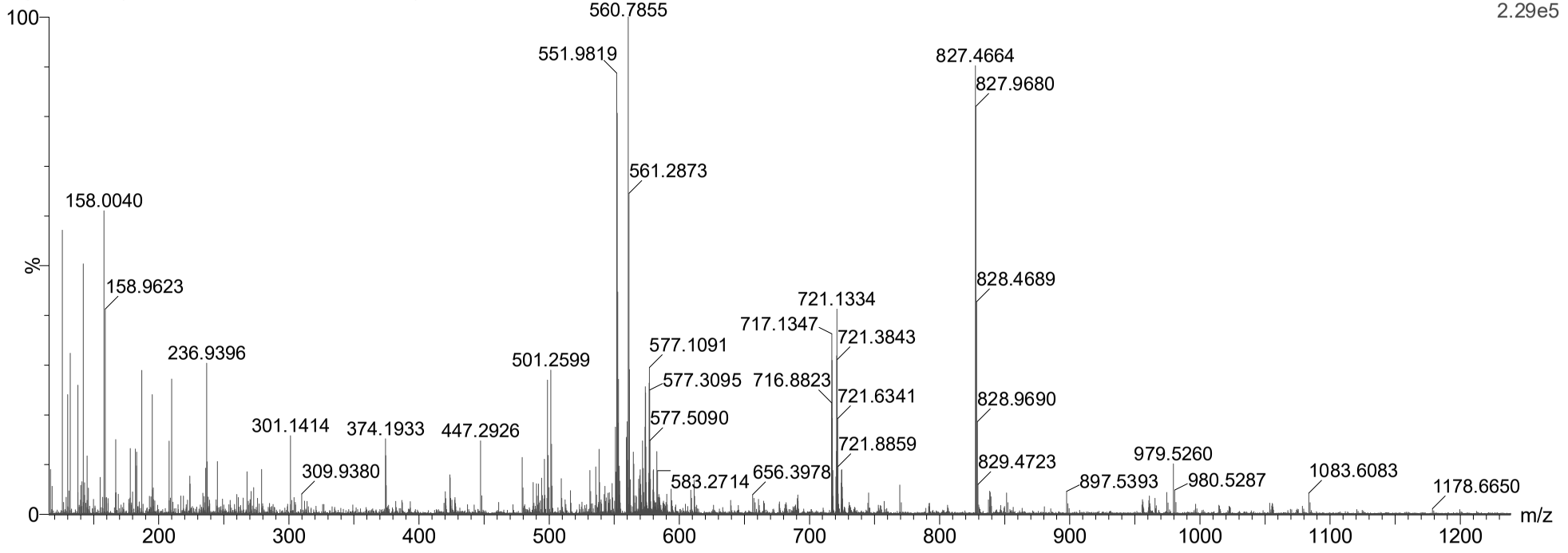


Figure 2.7C (TOF transformed)

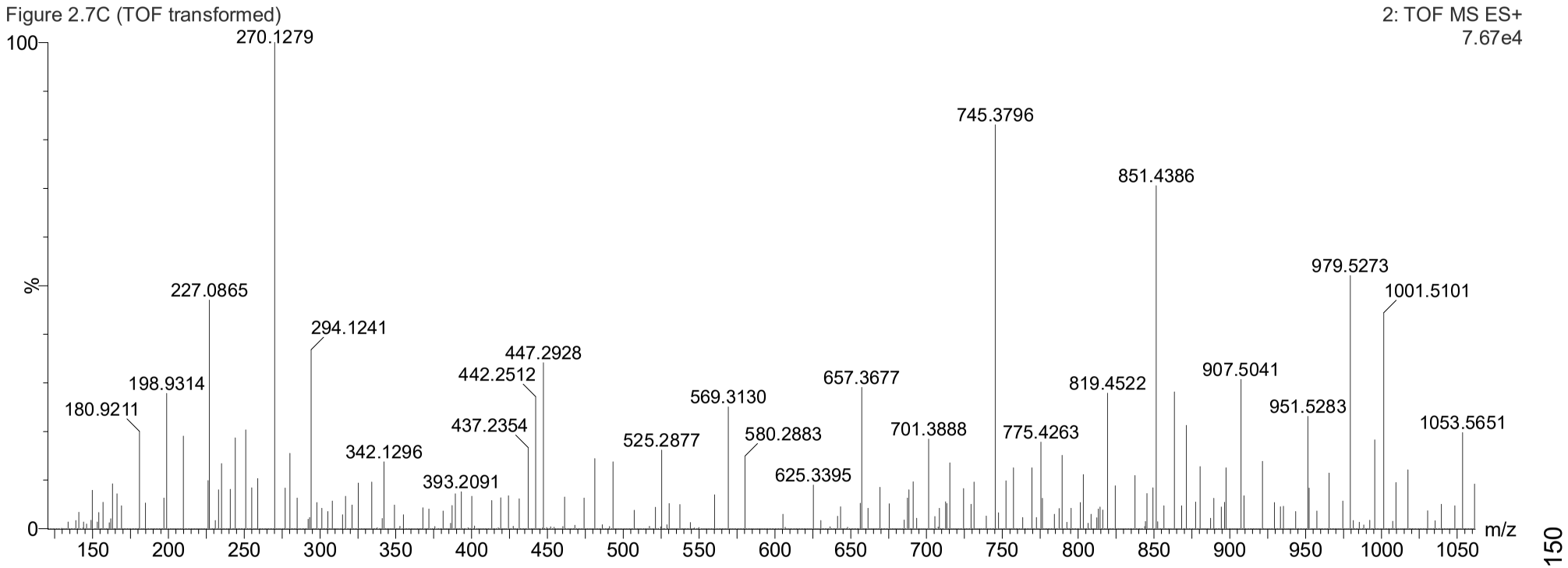


Figure 2.7C (Lock mass corrected)

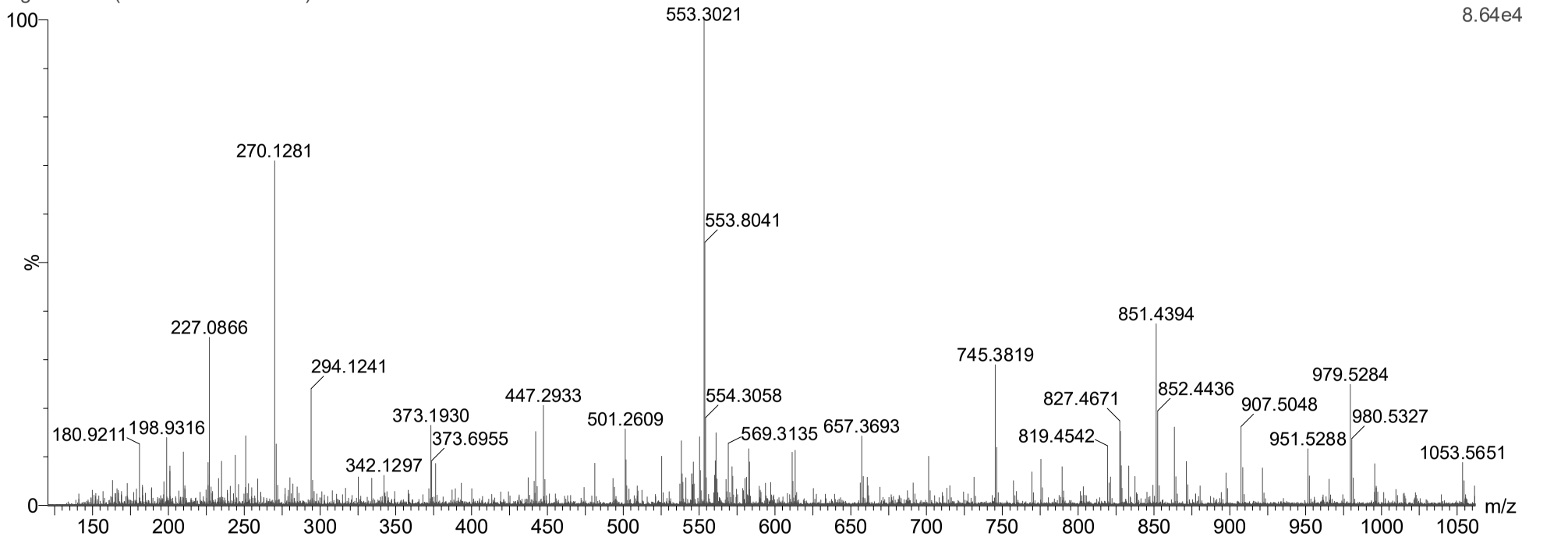


Figure 2.12A (smoothed and background subtracted)

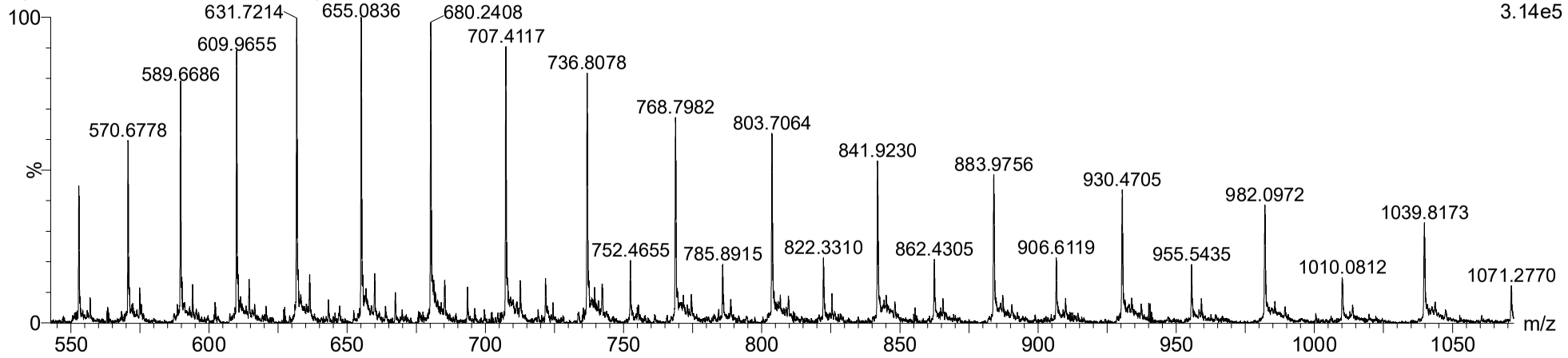


Figure 2.12B (smoothed and background subtracted)

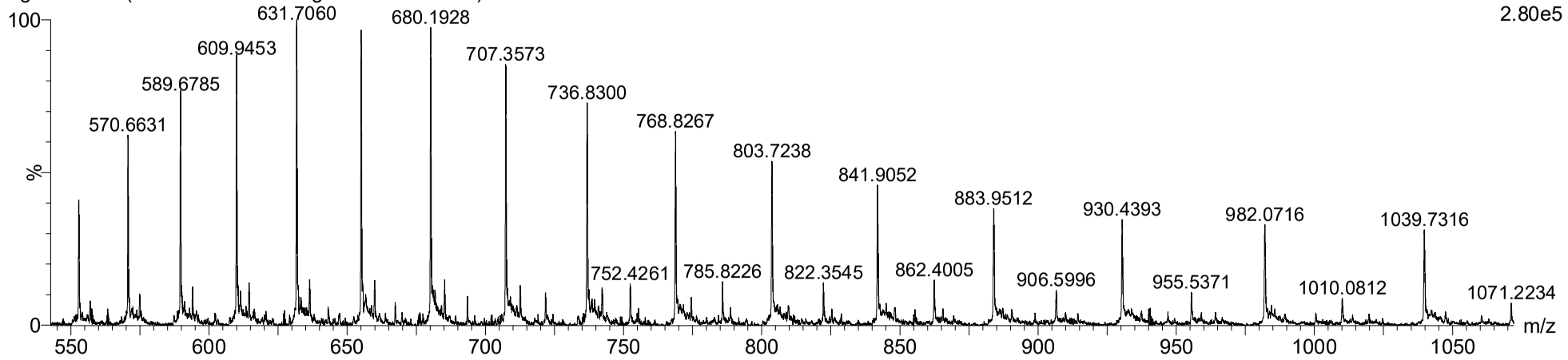


Figure 2.12C (smoothed and background subtracted)

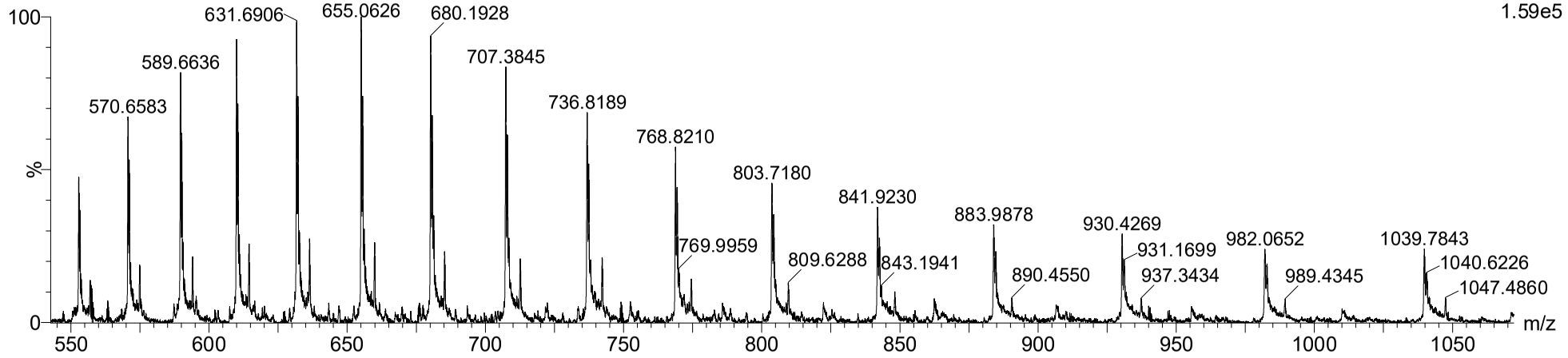


Figure 2.12A (smooth and background subtracted)

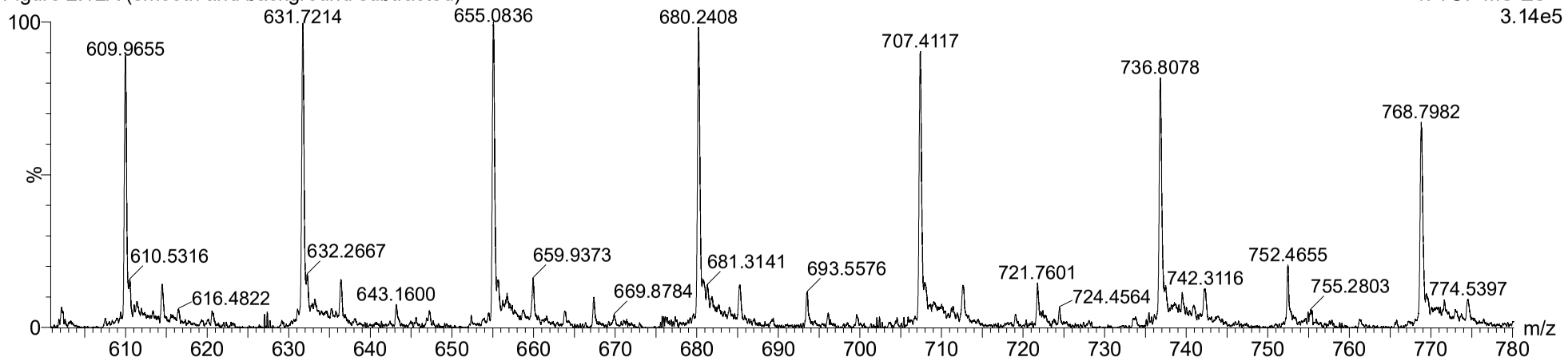


Figure 2.12B (smooth and background subtracted)

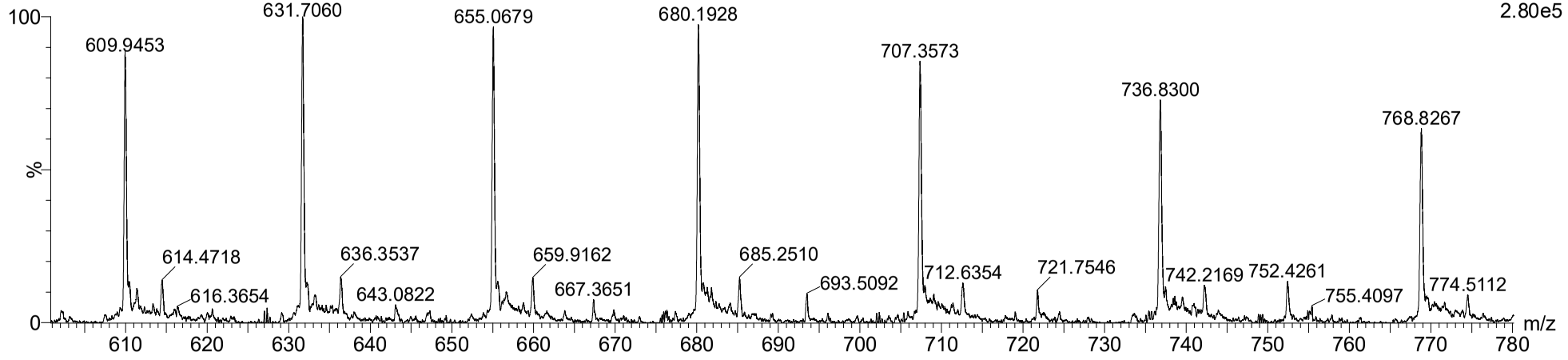


Figure 2.12C (smooth and background subtracted)

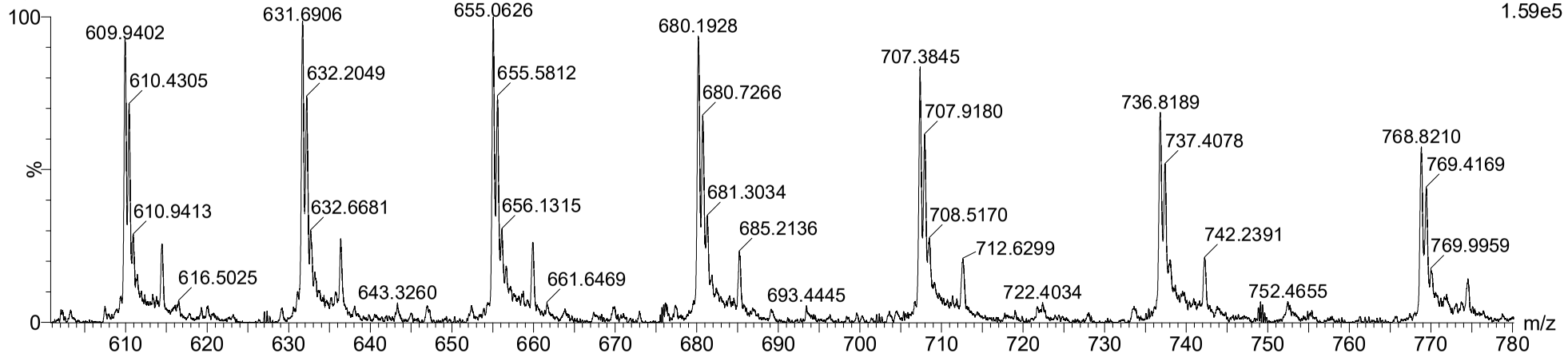


Figure 2.13A (TOF transformed)

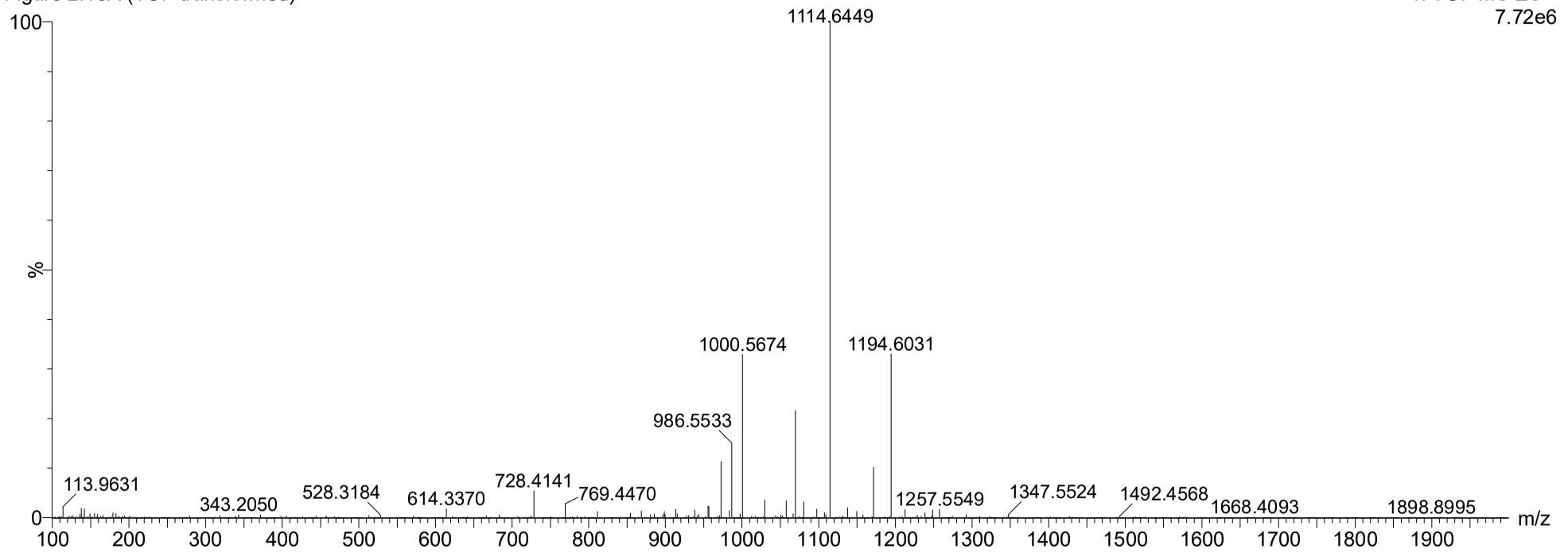


Figure 2.13A (Lock mass corrected)

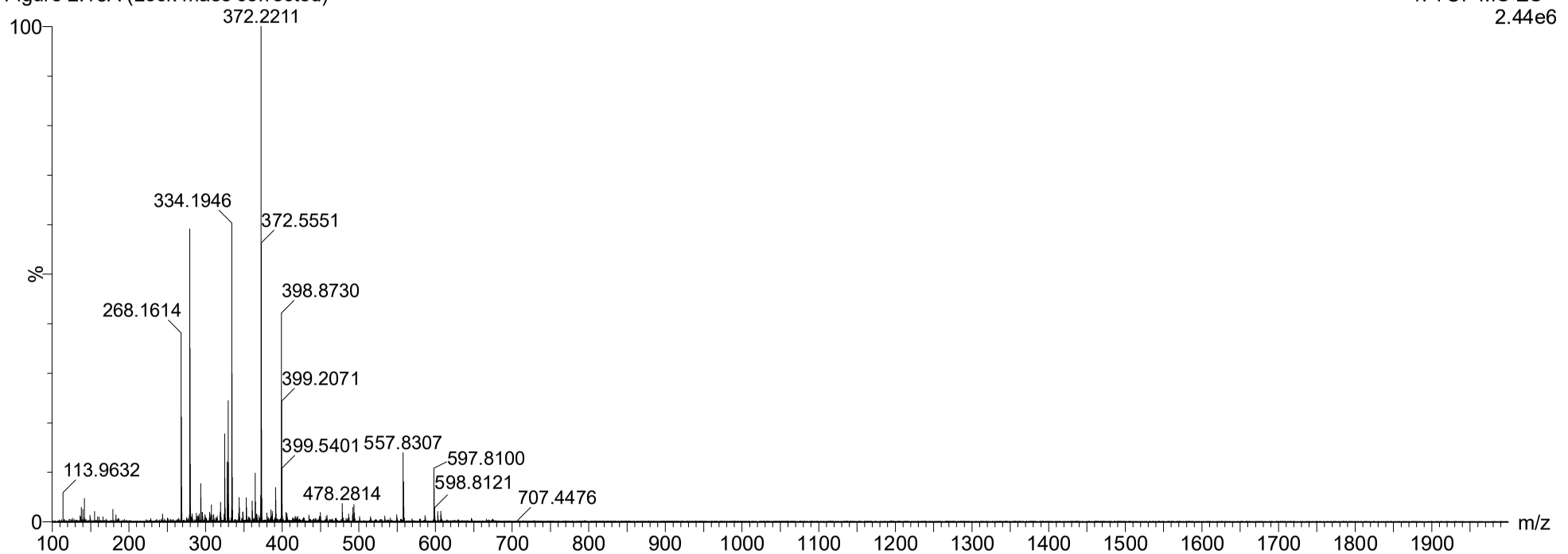


Figure 2.13B (TOF transformed)

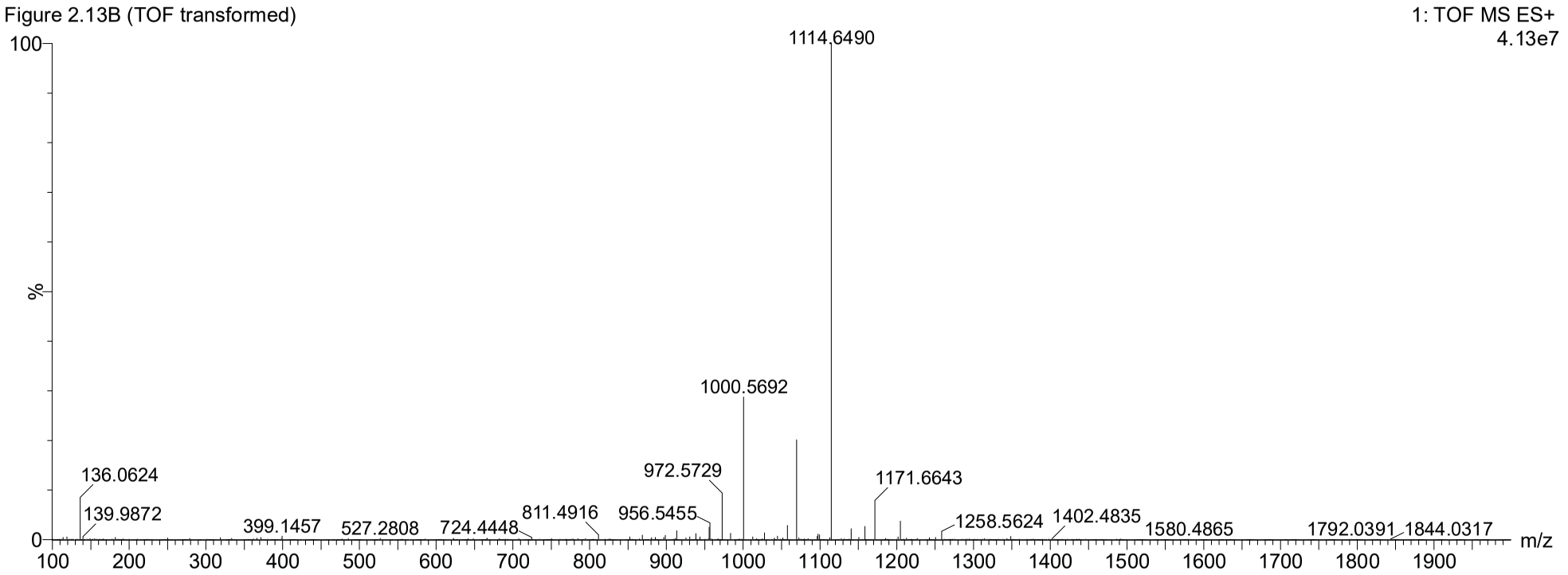


Figure 2.13B (Lock mass corrected)

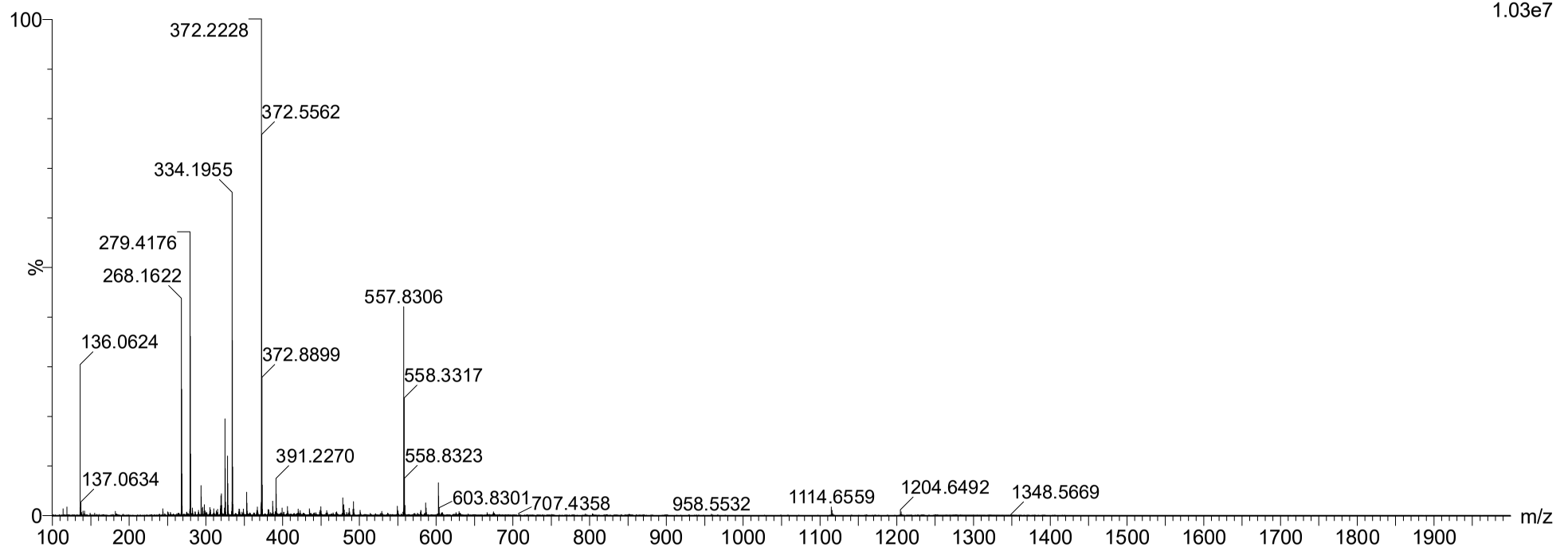


Figure 2.13C (TOF transformed)

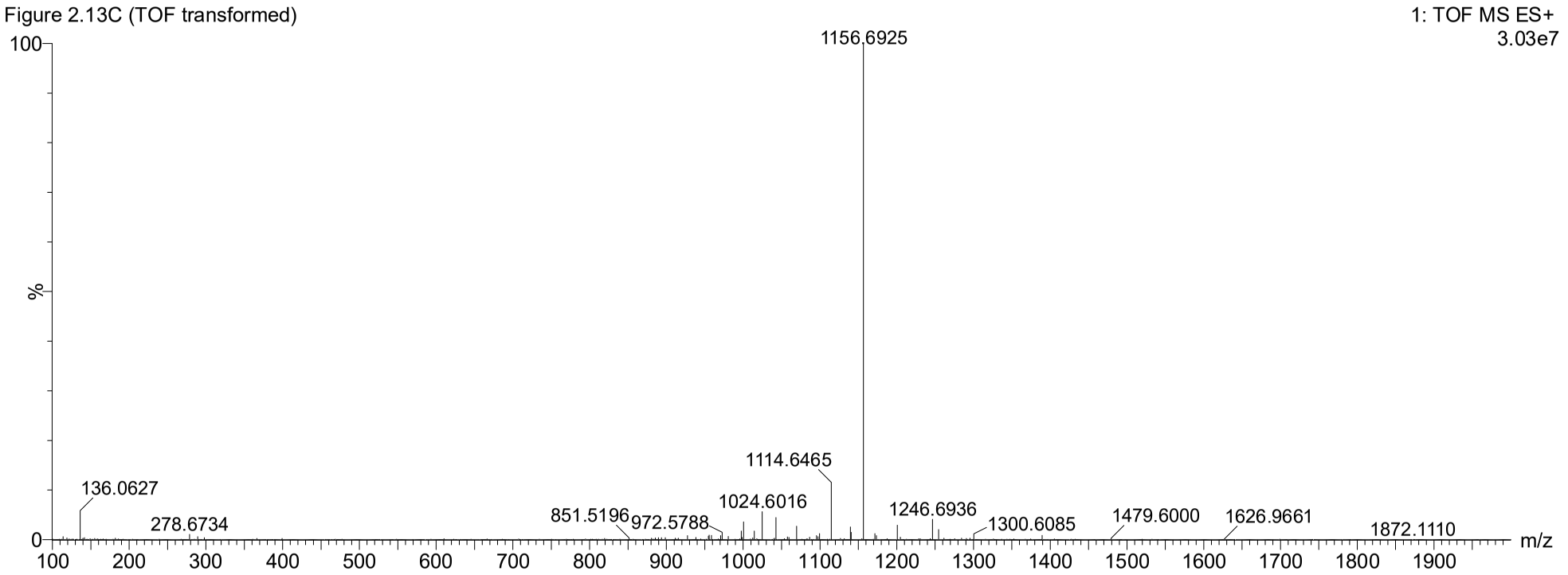


Figure 2.13C (Lock mass corrected)

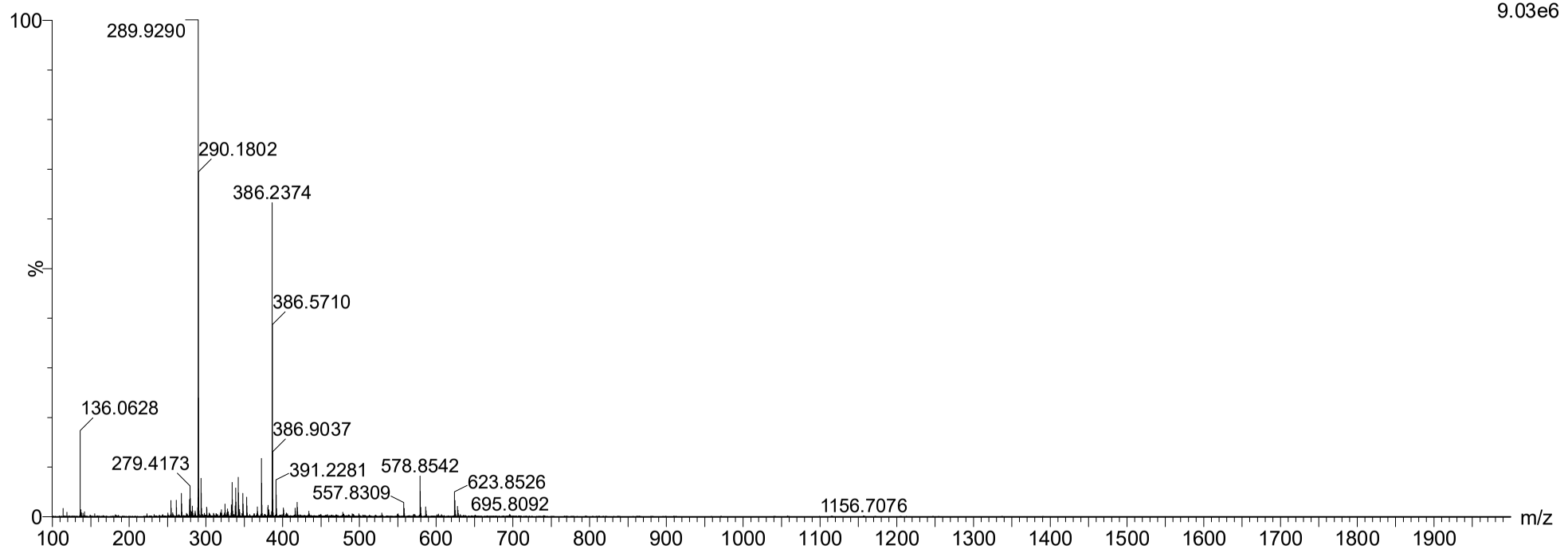


Figure 2.14A (TOF transformed)

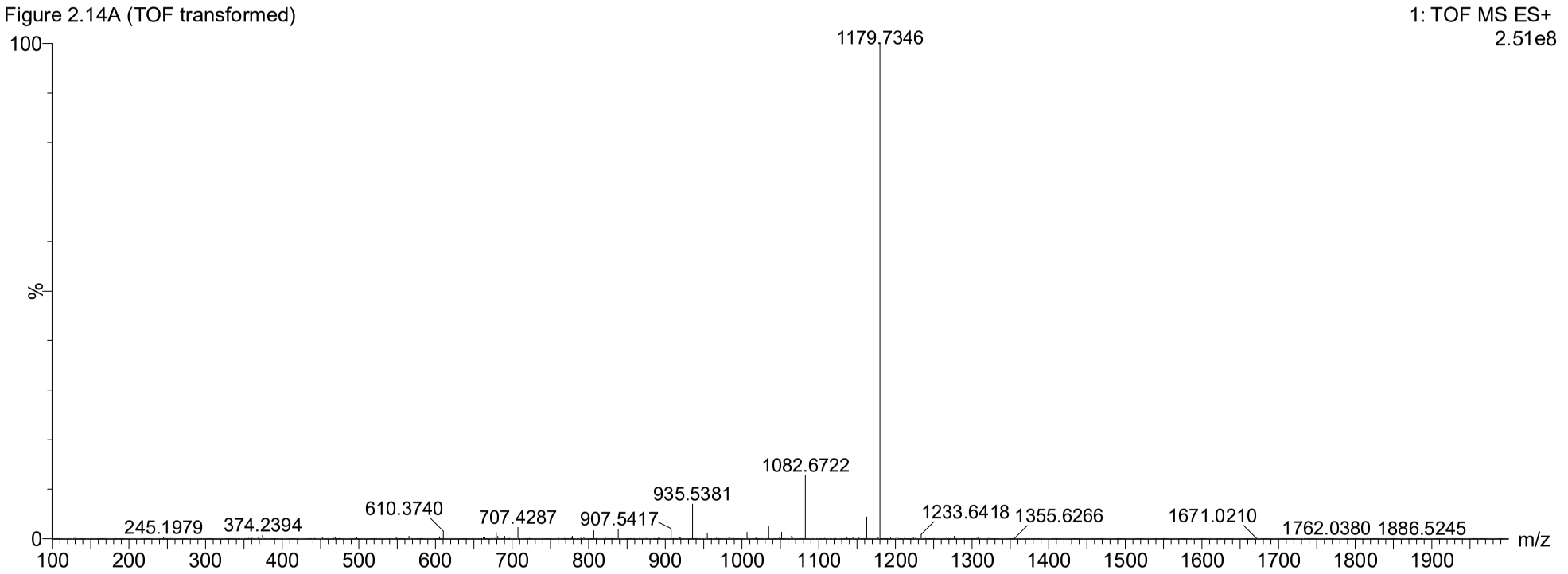


Figure 2.14A (Lock mass corrected)

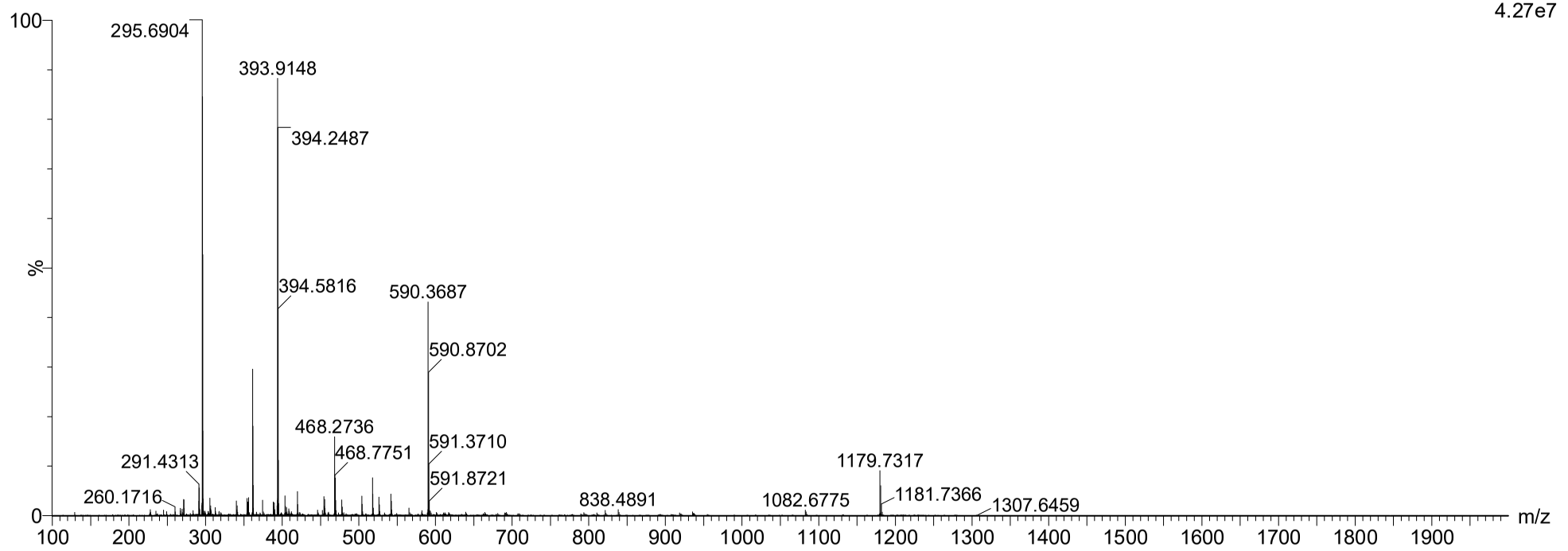


Figure 2.14B (TOF transformed)

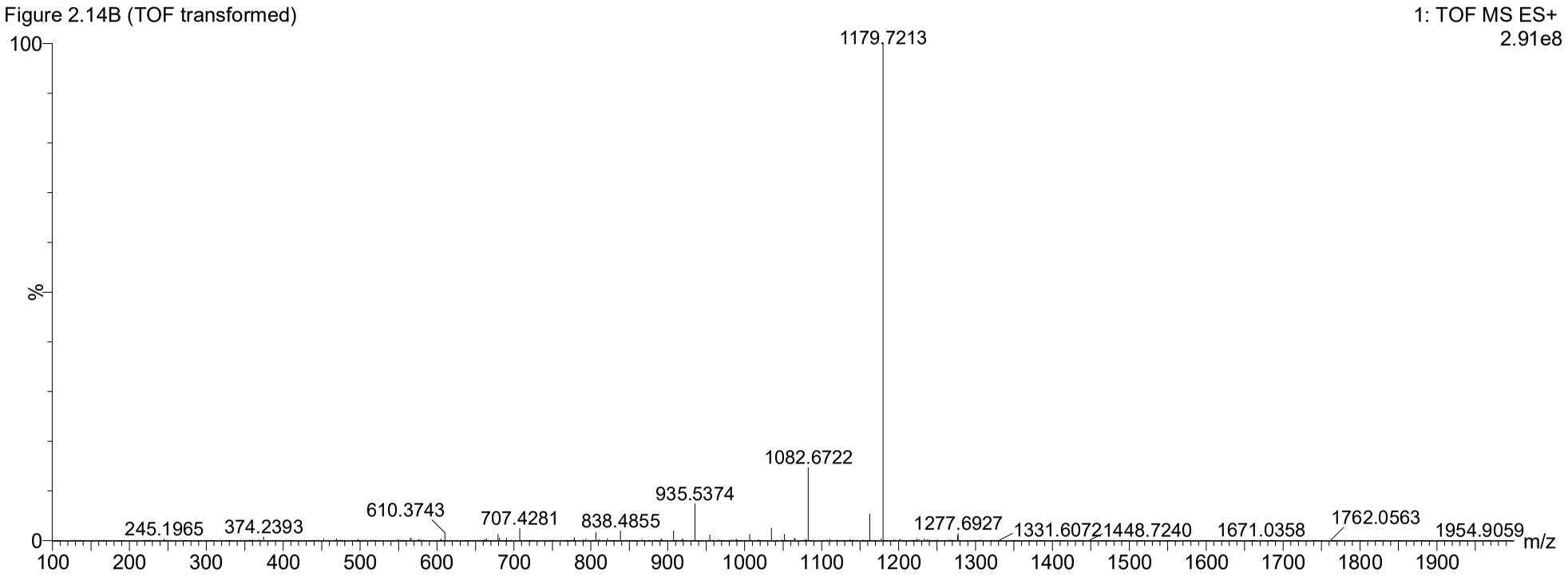


Figure 2.14B (Lock mass corrected)

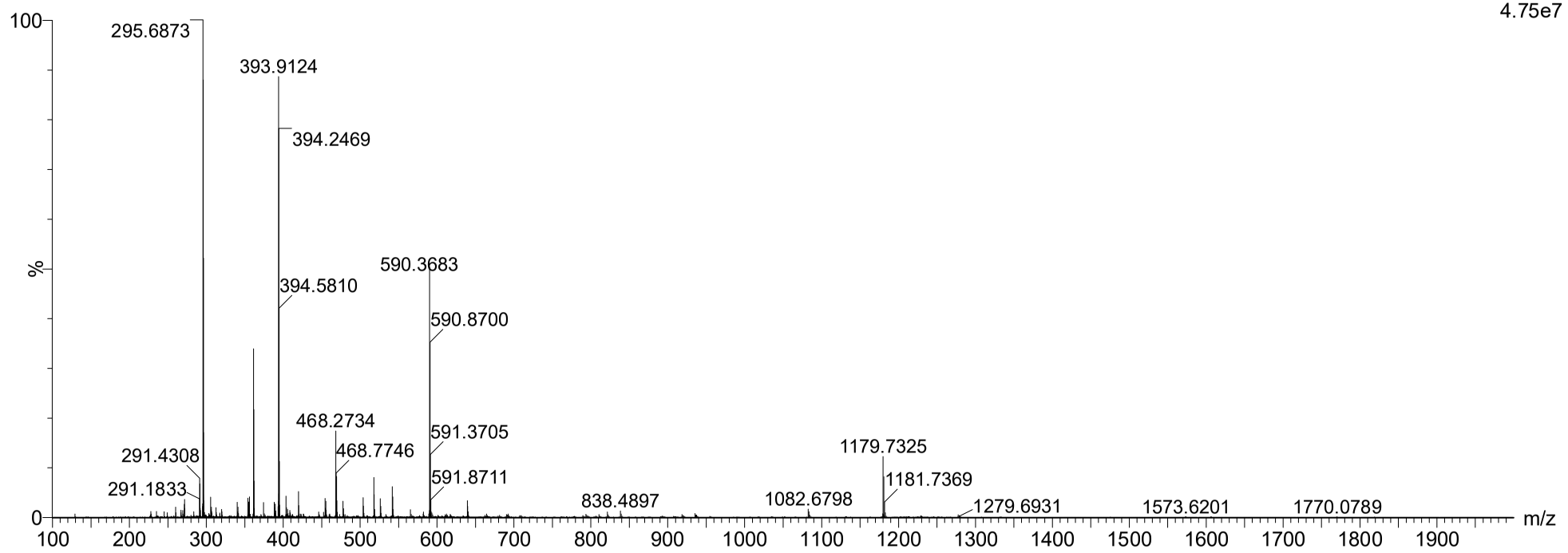


Figure 2.14C (TOF transformed)

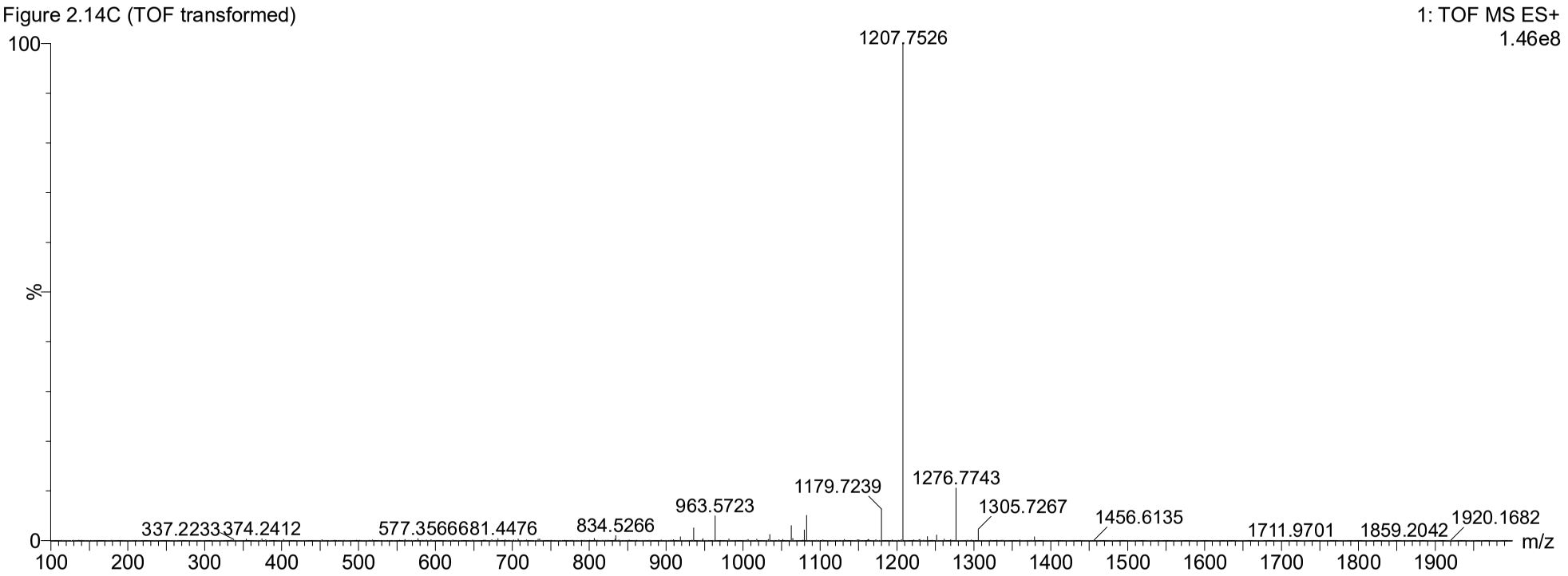


Figure 2.14C (Lock mass corrected)

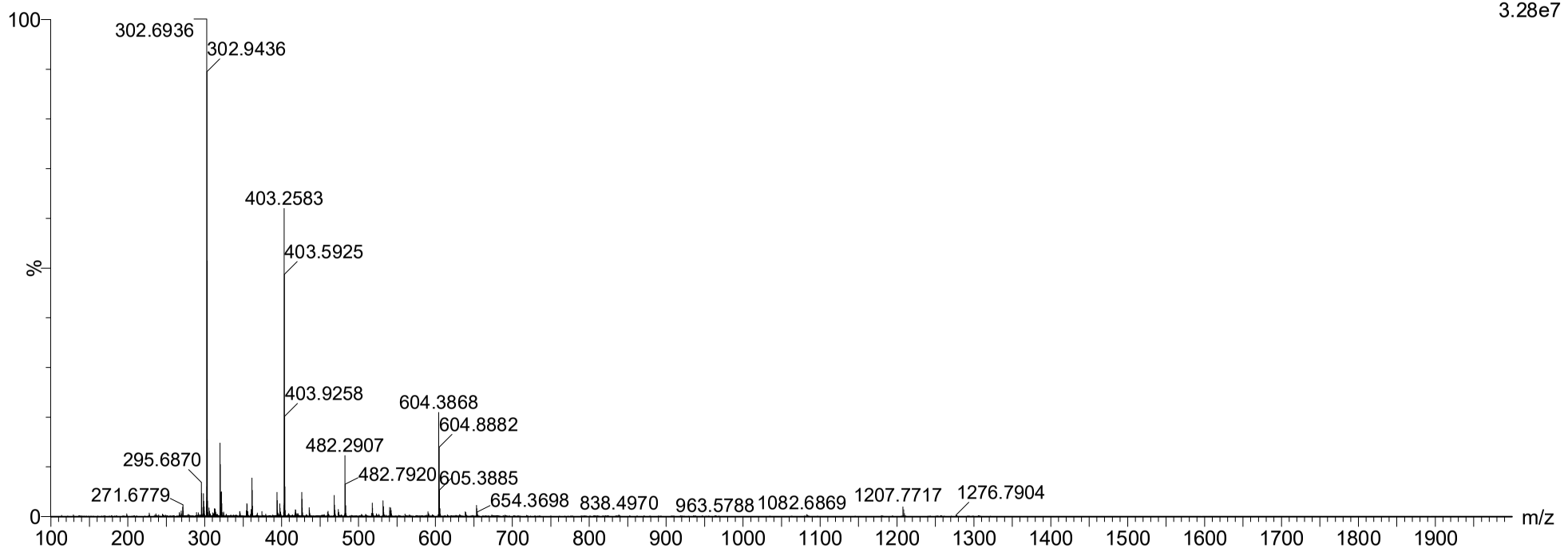


Figure 2.15A (TOF transformed)

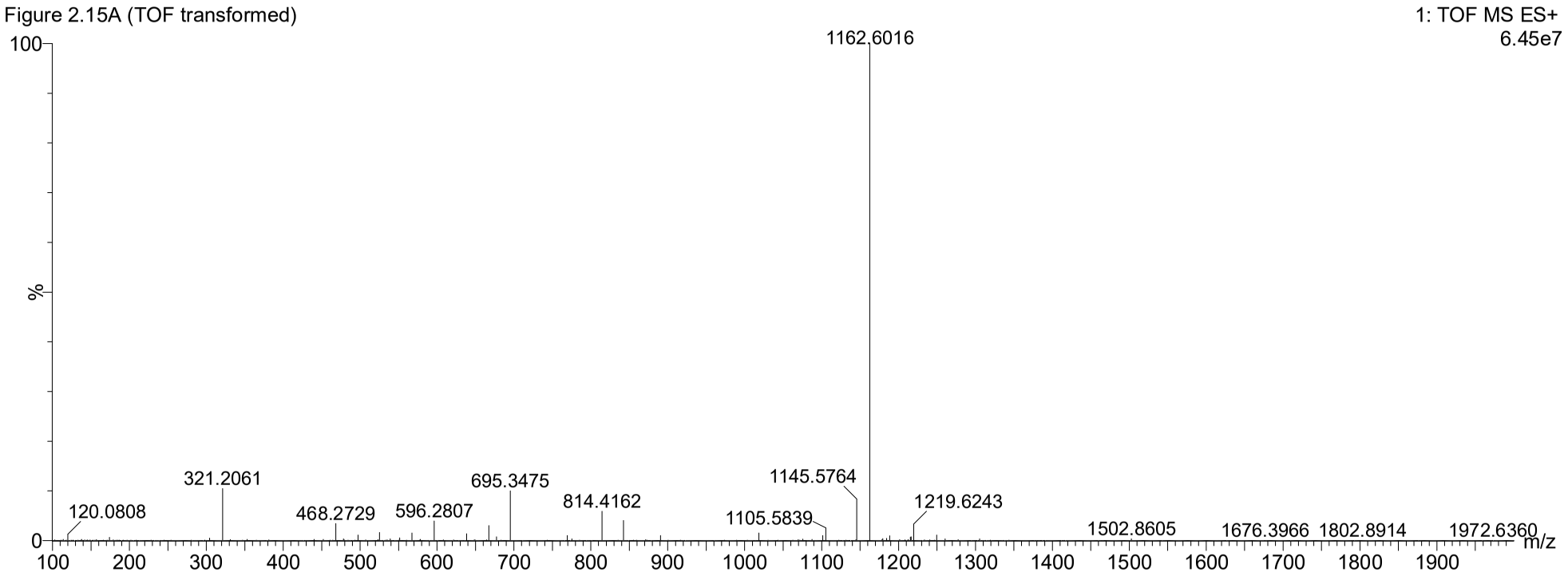


Figure 2.15A (Lock mass corrected)

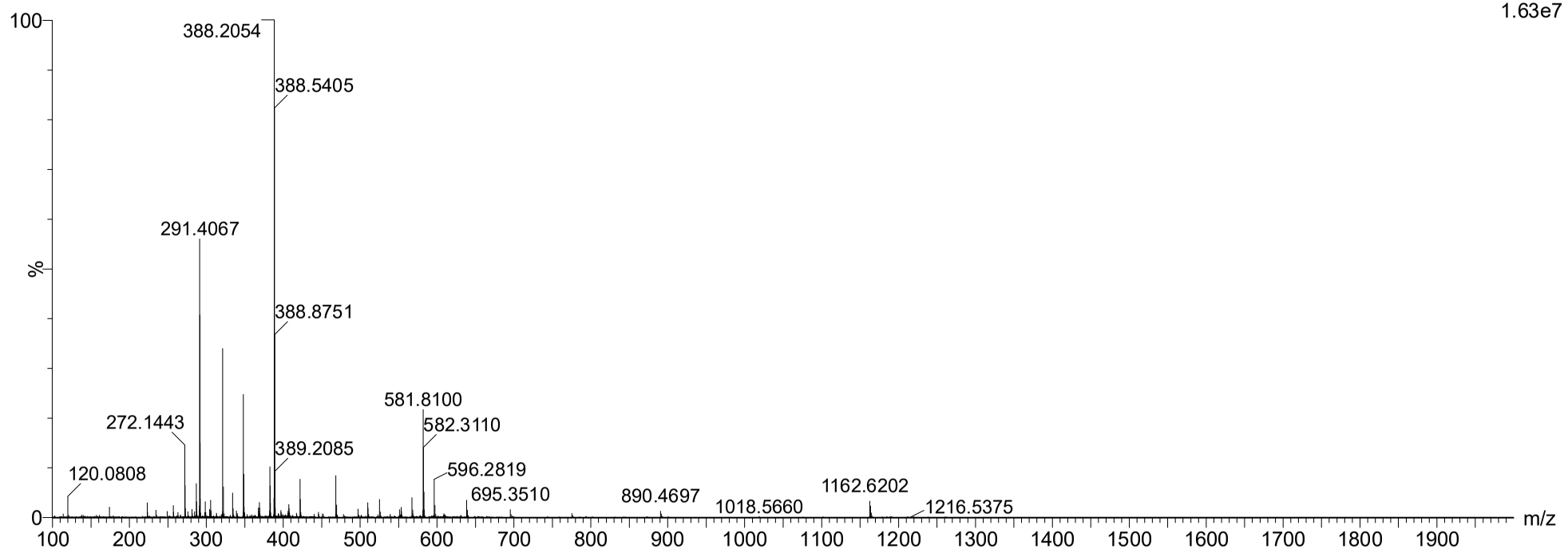
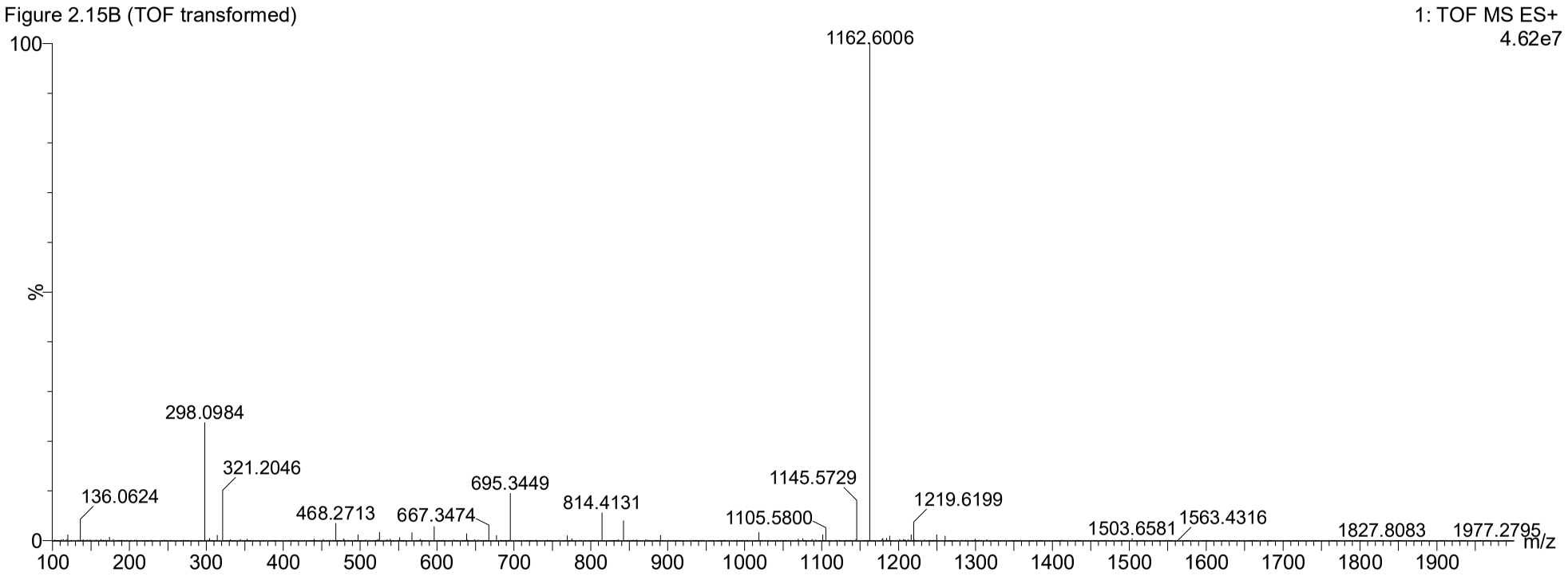


Figure 2.15B (TOF transformed)



160

Figure 2.15B (Lock mass corrected)

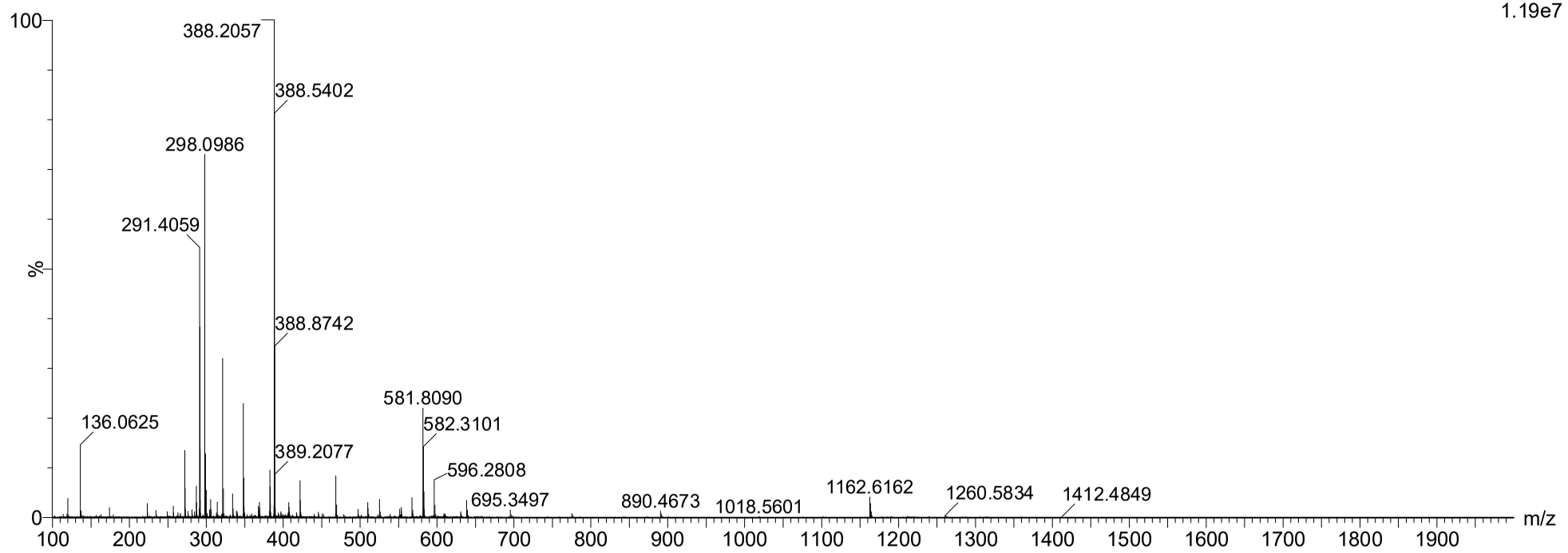


Figure 2.15C (TOF transformed)

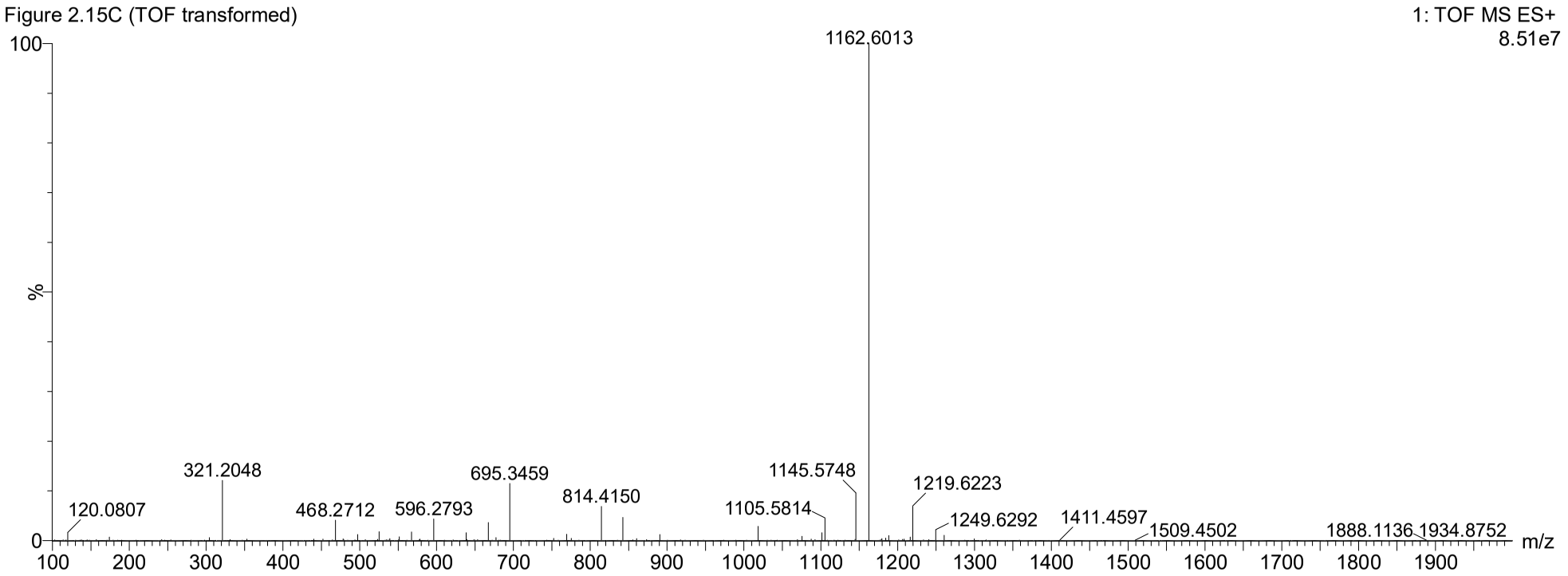


Figure 2.15C (Lock mass corrected)

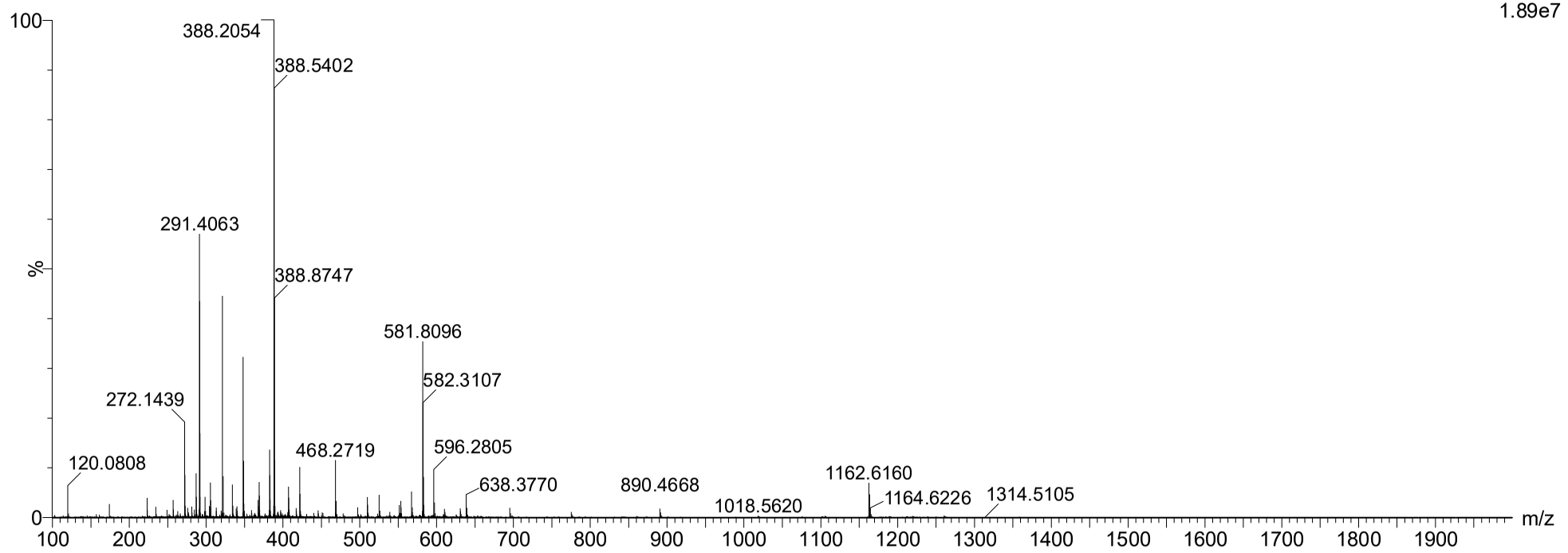


Figure 2.16A (TOF transformed)

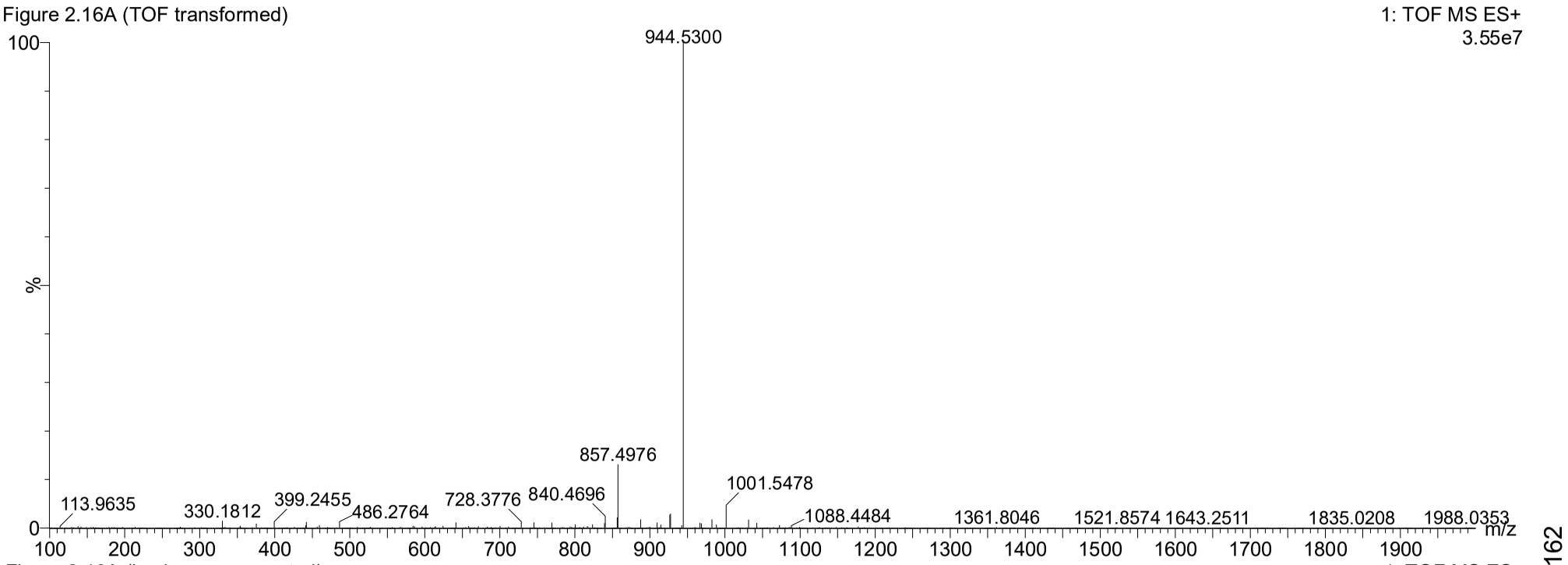


Figure 2.16A (Lock mass corrected)

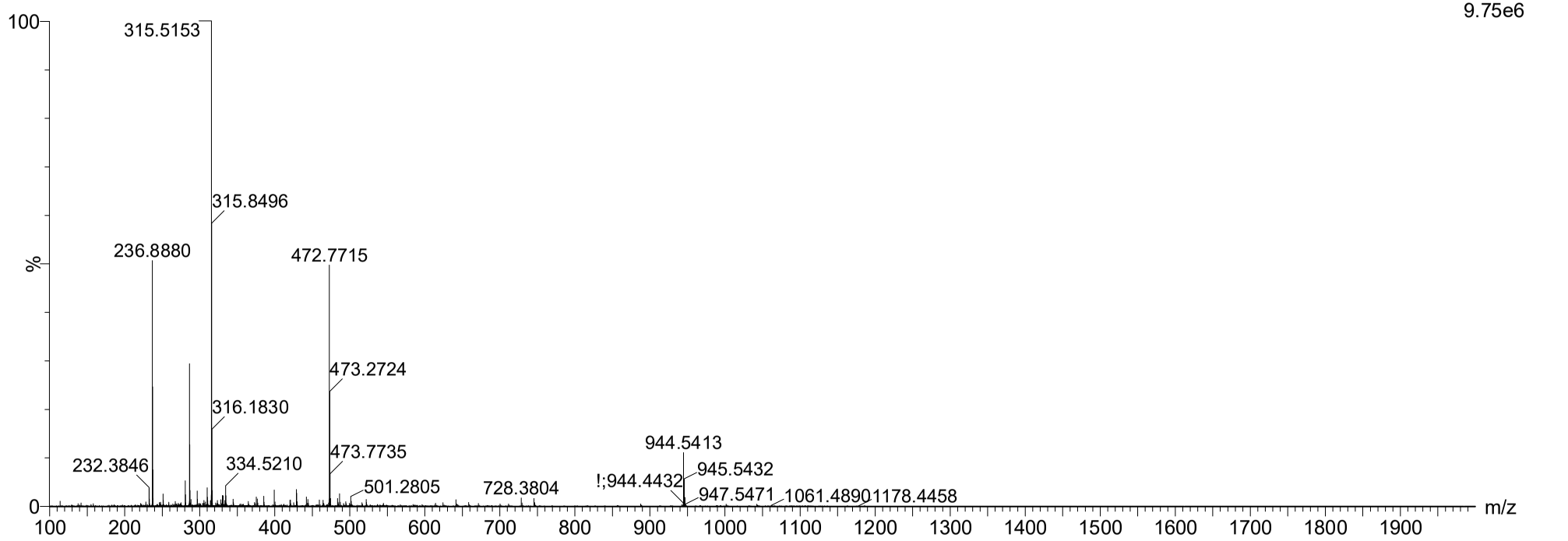


Figure 2.16B (TOF transformed)

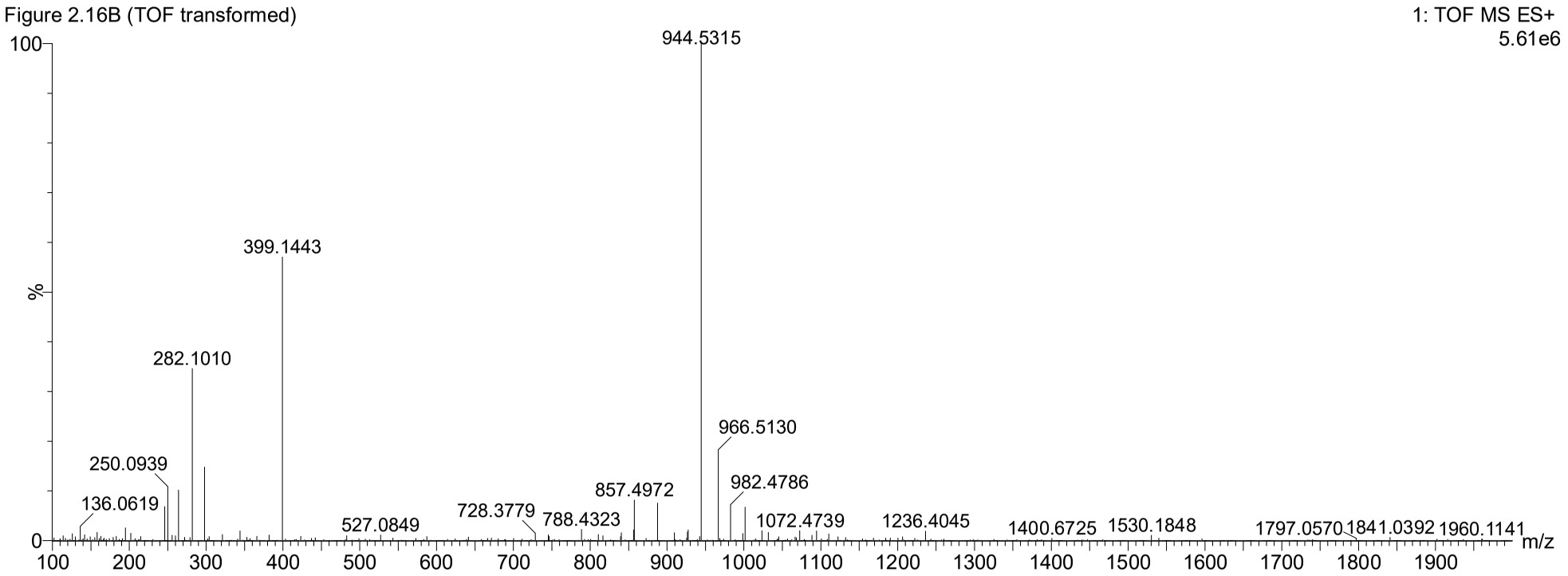


Figure 2.16B (Lock mass corrected)

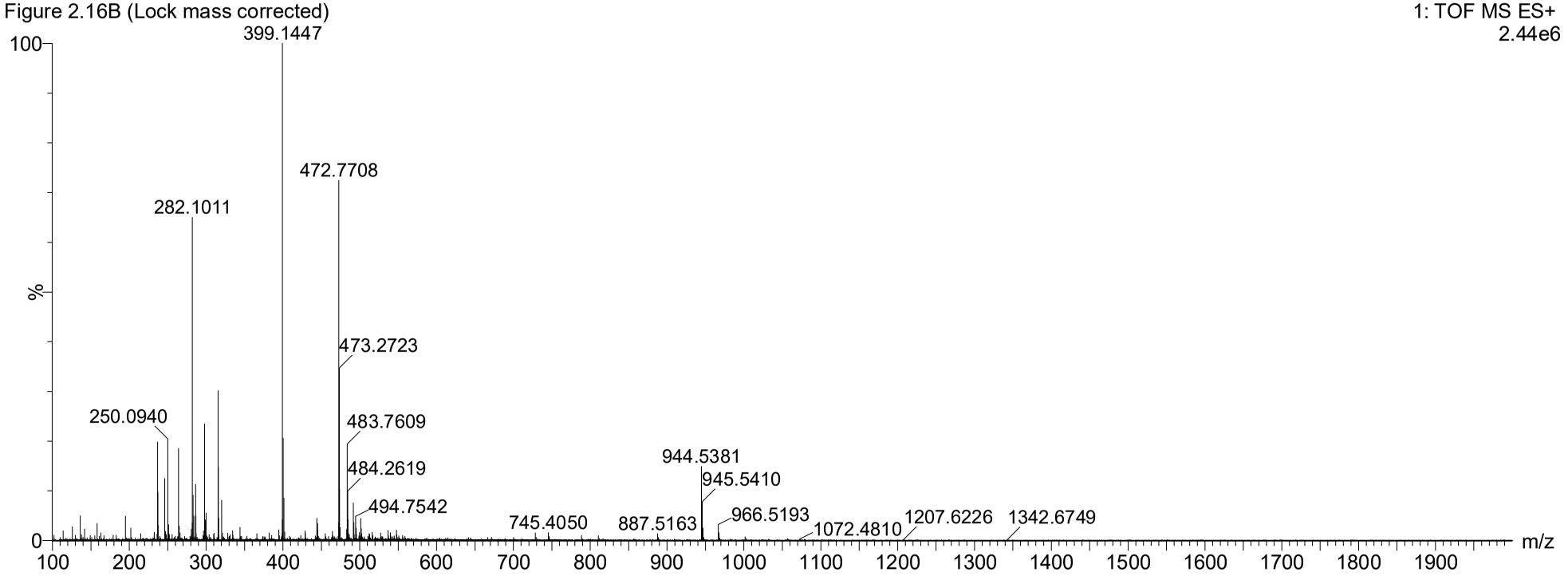


Figure 2.16C (TOF transformed)

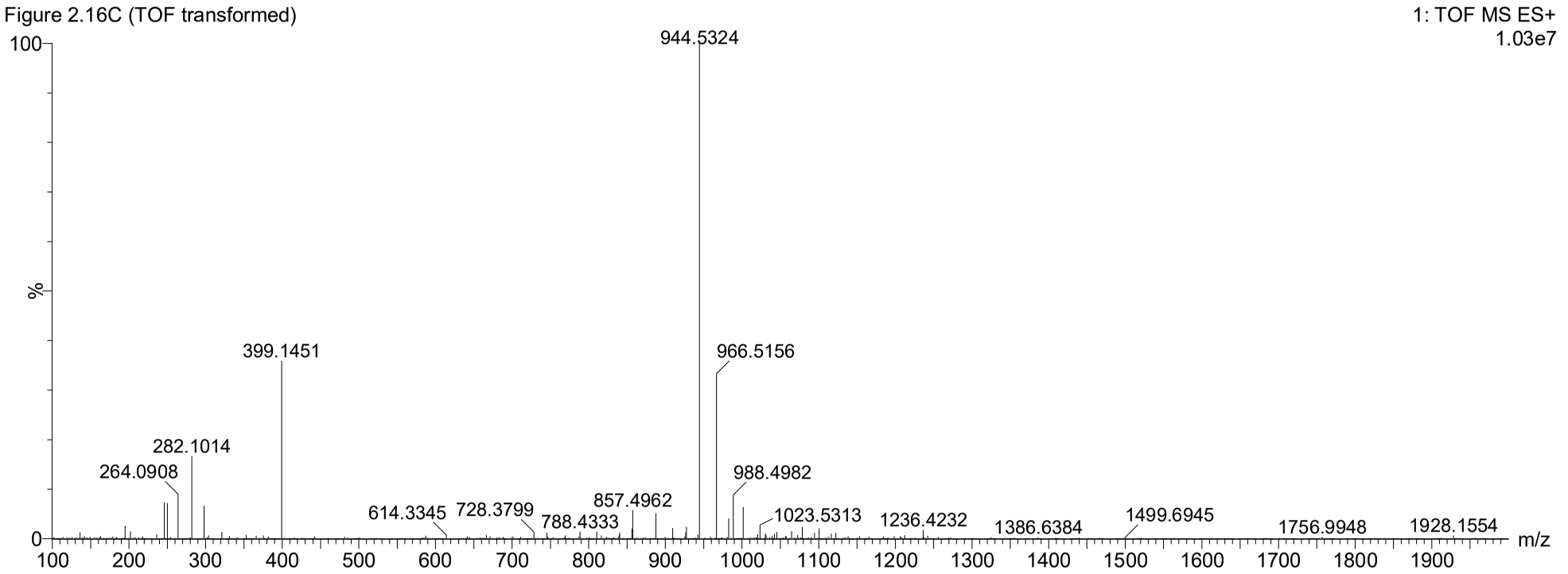


Figure 2.16C (Lock mass corrected)

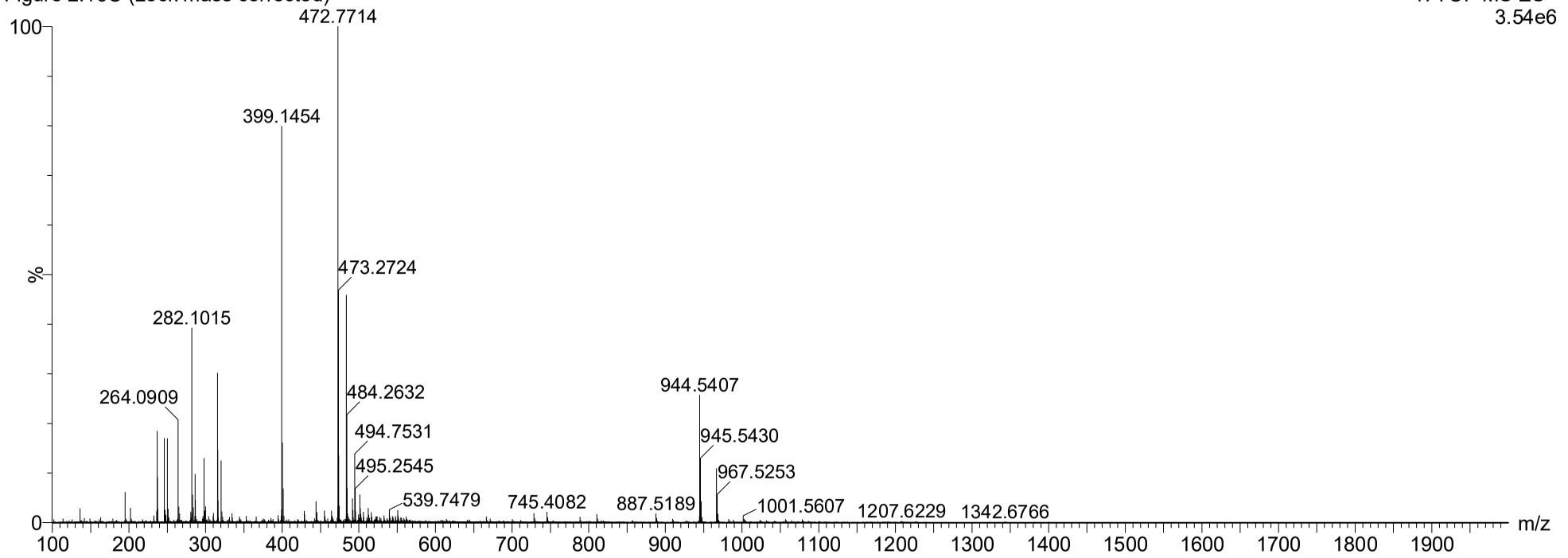


Figure 3.1A (TOF transformed)

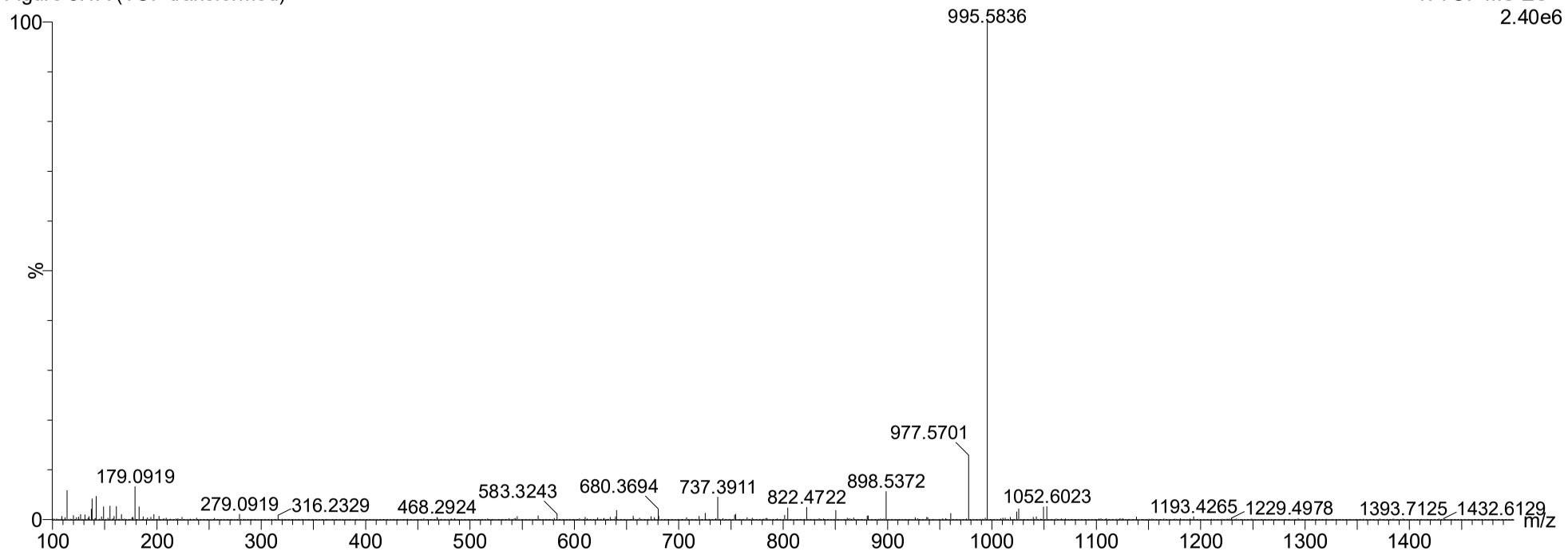


Figure 3.1A (Lock mass corrected)

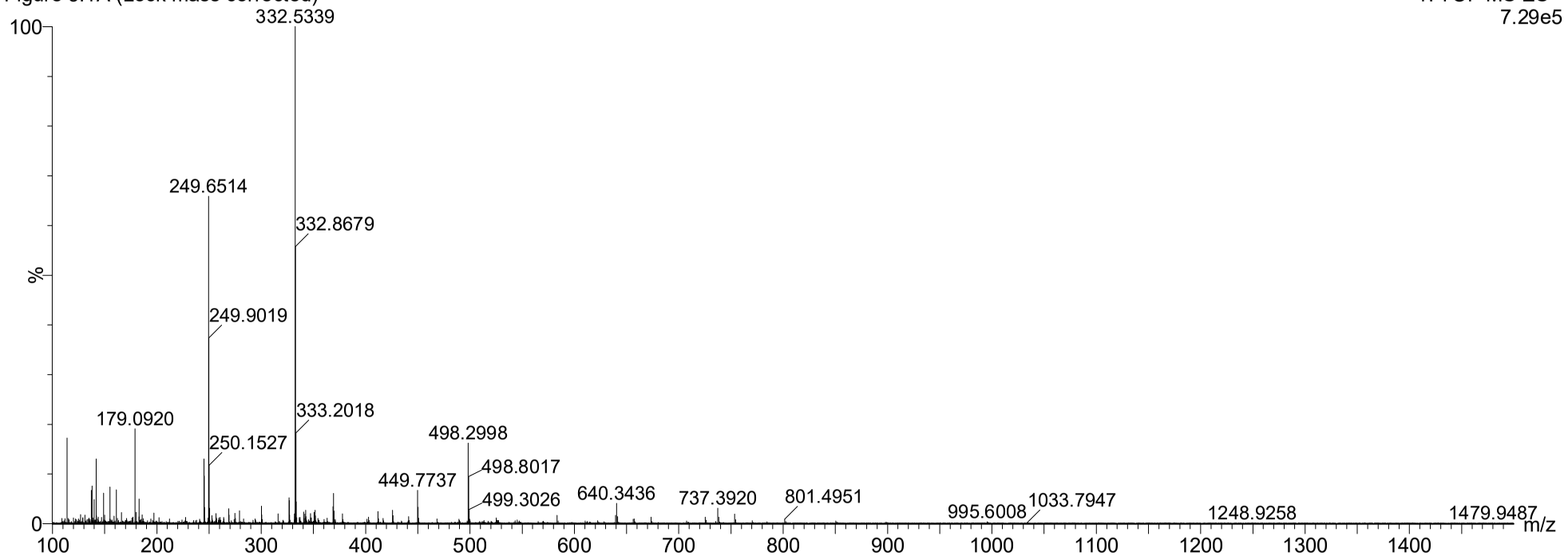


Figure 3.1B (TOF transformed)

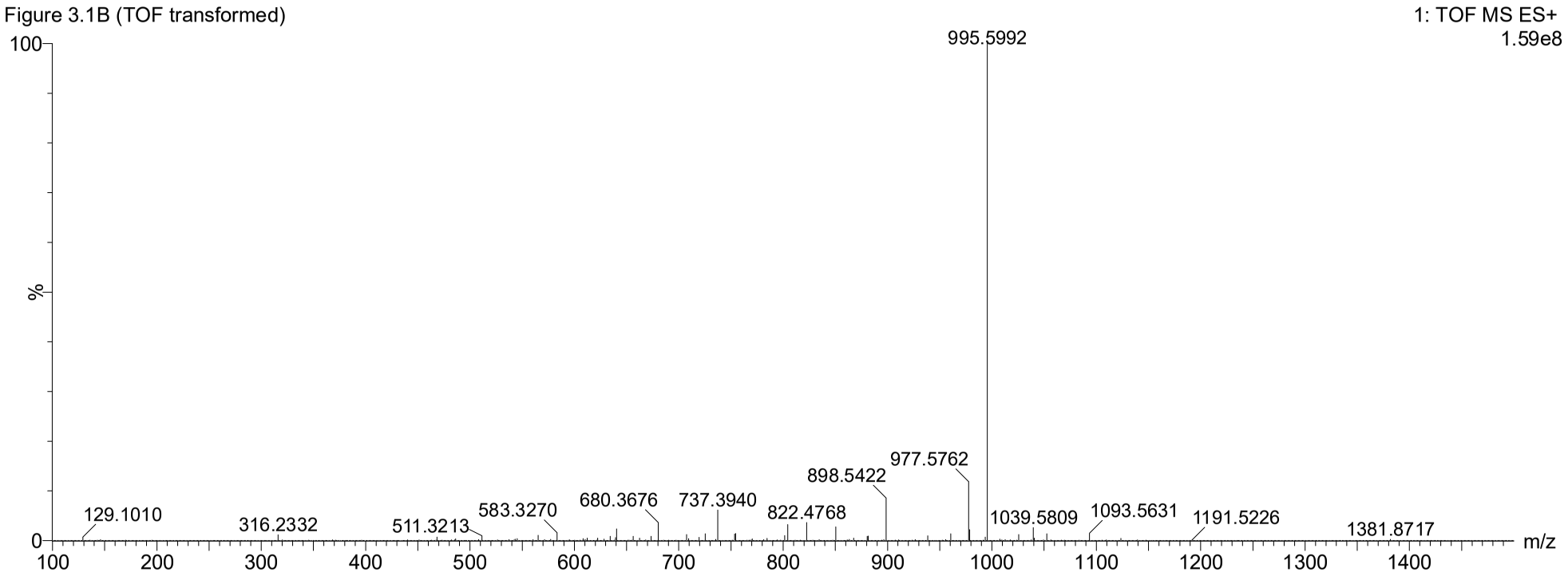


Figure 3.1B (Lock mass corrected)

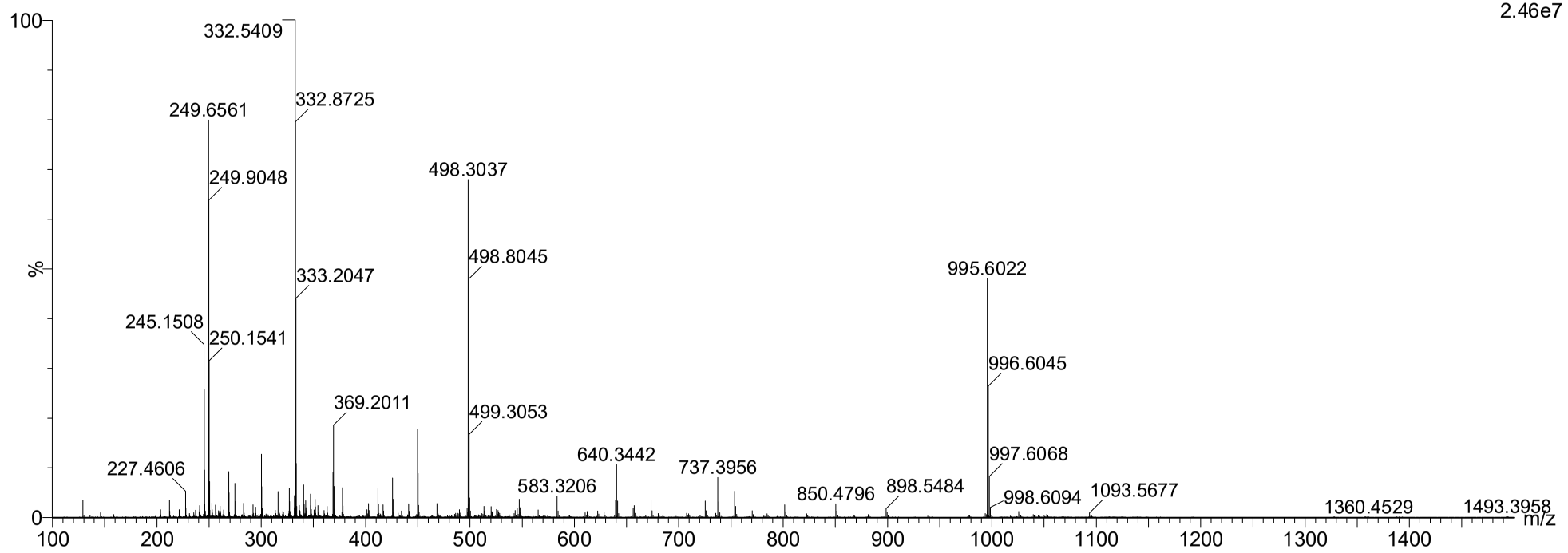


Figure 3.1C (TOF transformed)

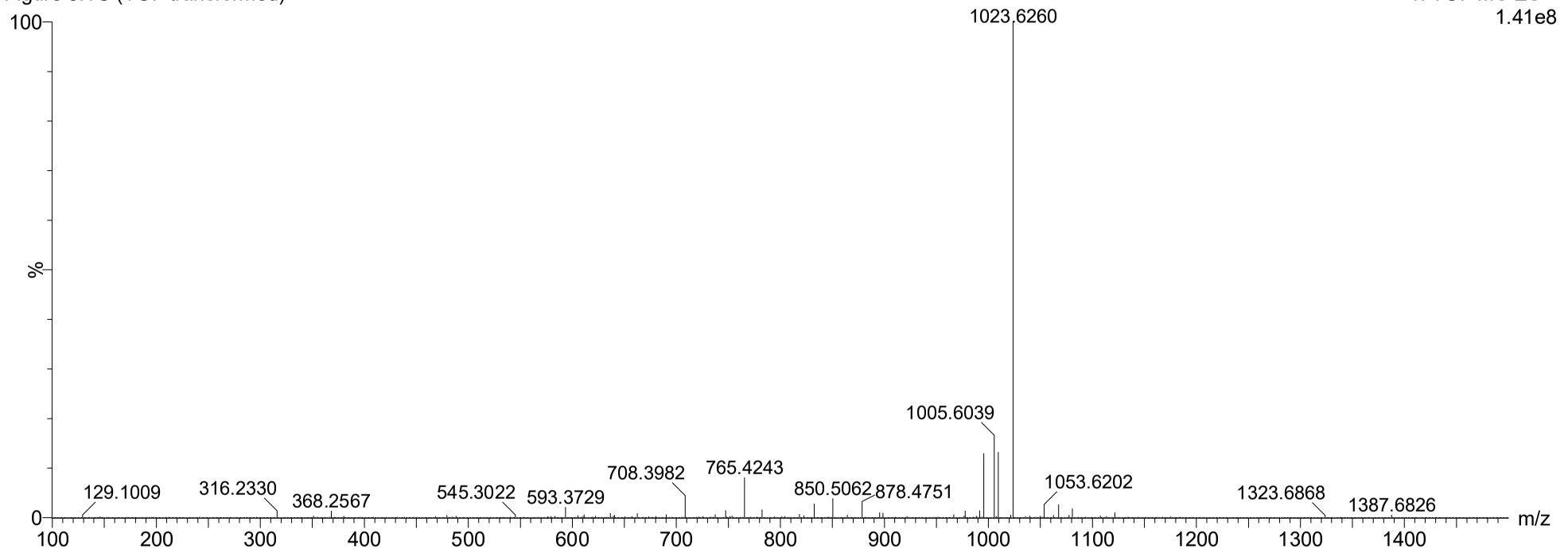


Figure 3.1C (Lock mass corrected)

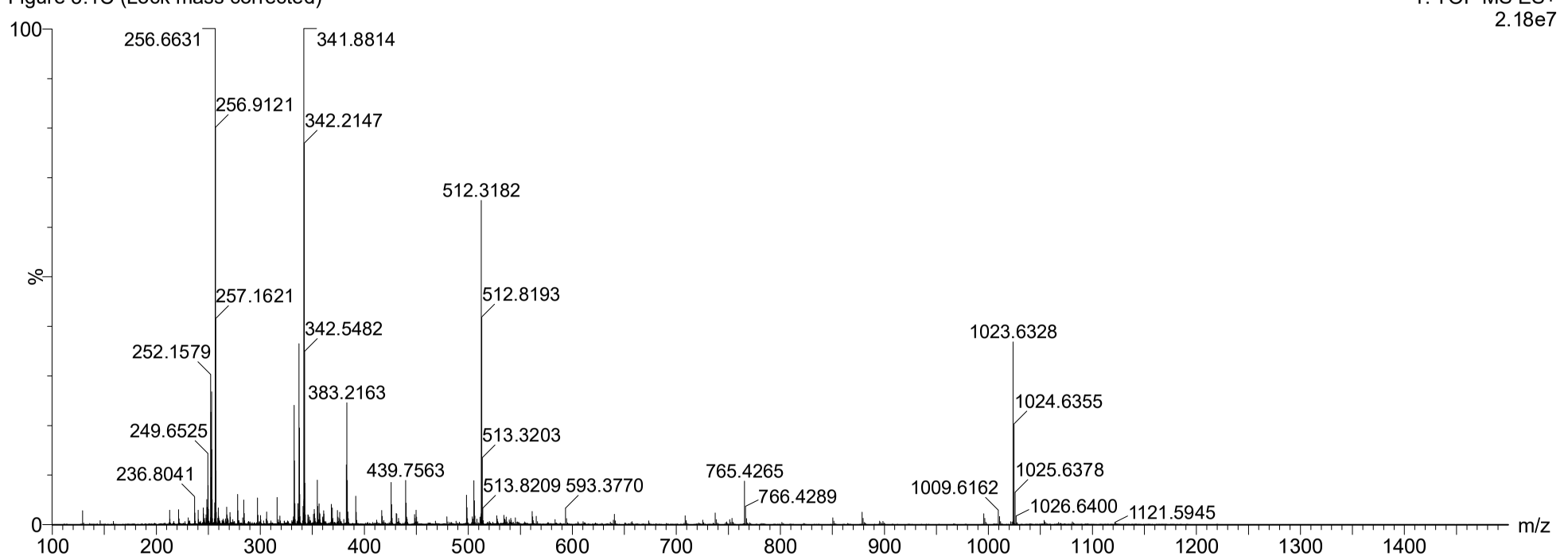


Figure 3.2B (smooth and background subtracted)

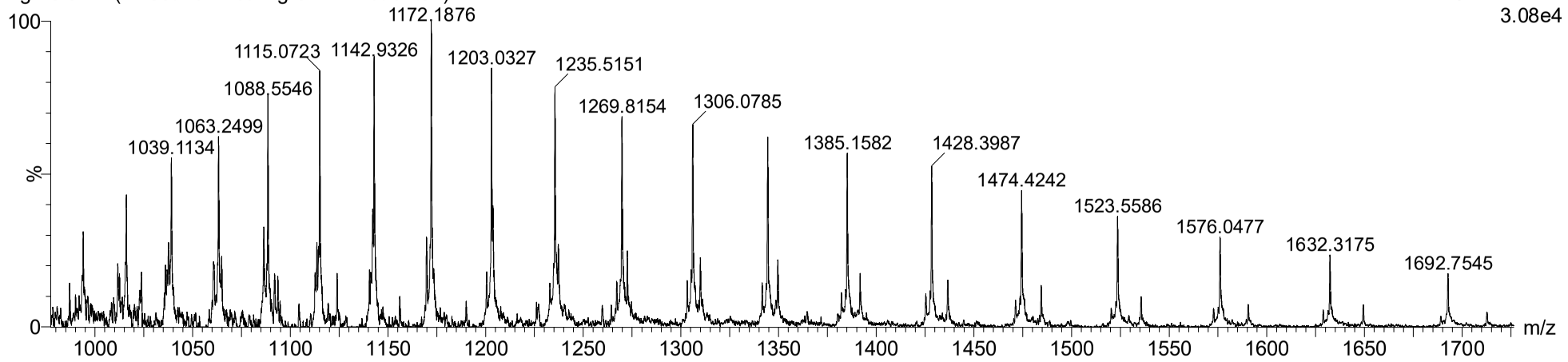


Figure 3.2D (smooth and background subtracted)

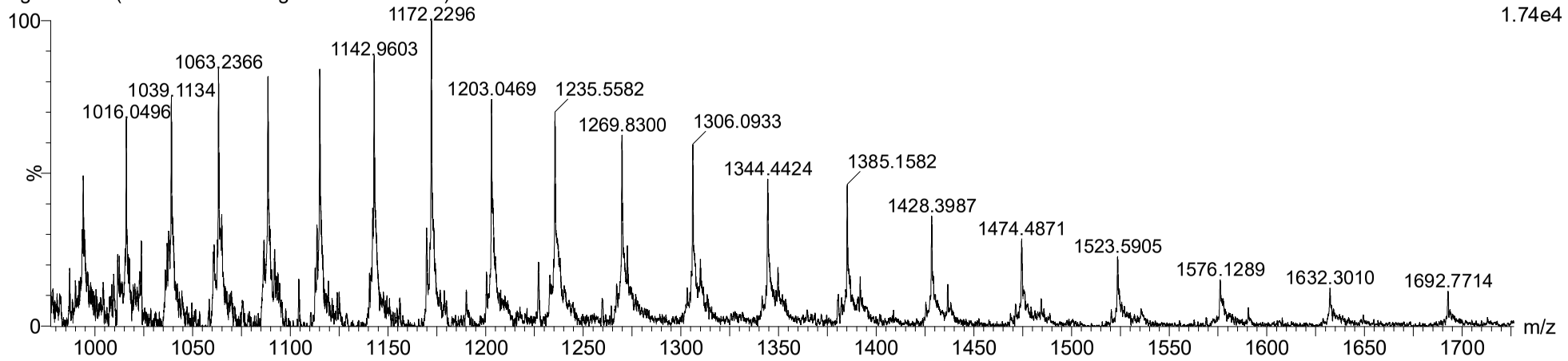


Figure 3.2E (smooth and background subtracted)

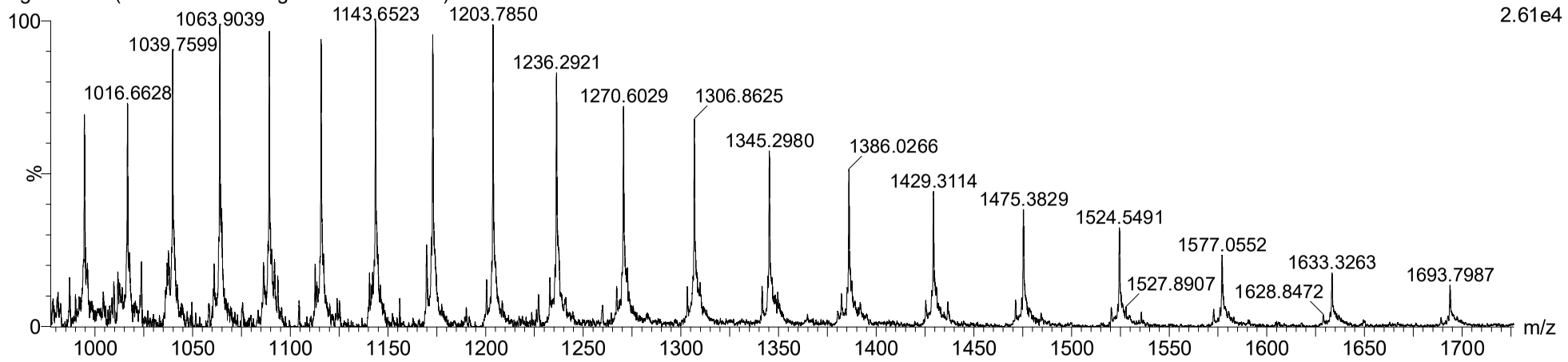


Figure 3.2C (Precursor, TOF transformed)

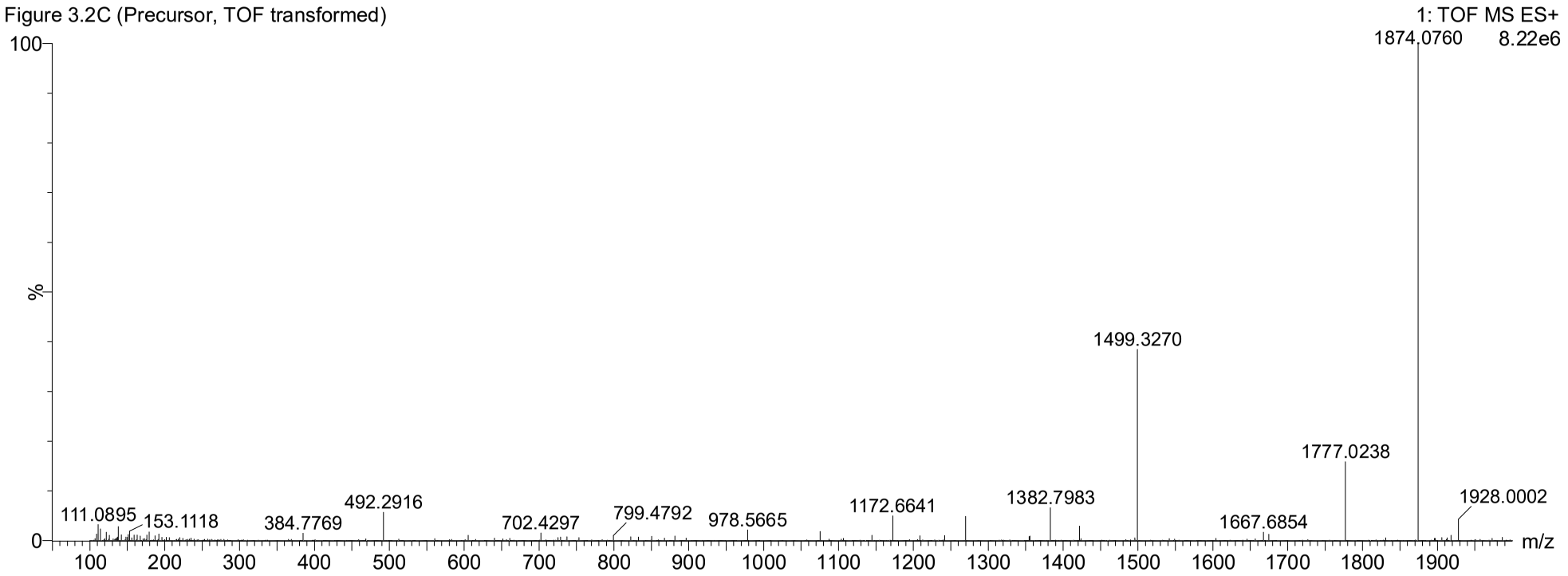


Figure 3.2C (Precursor, Lock mass corrected)

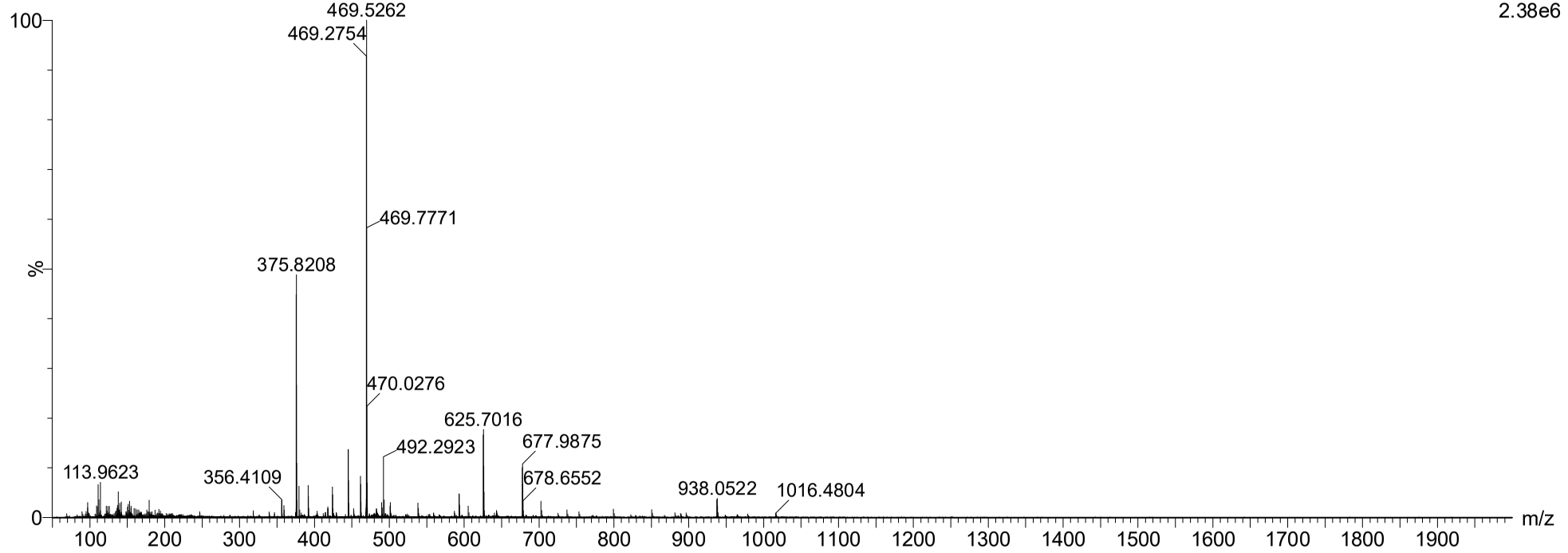


Figure 3.2C (TOF transformed)

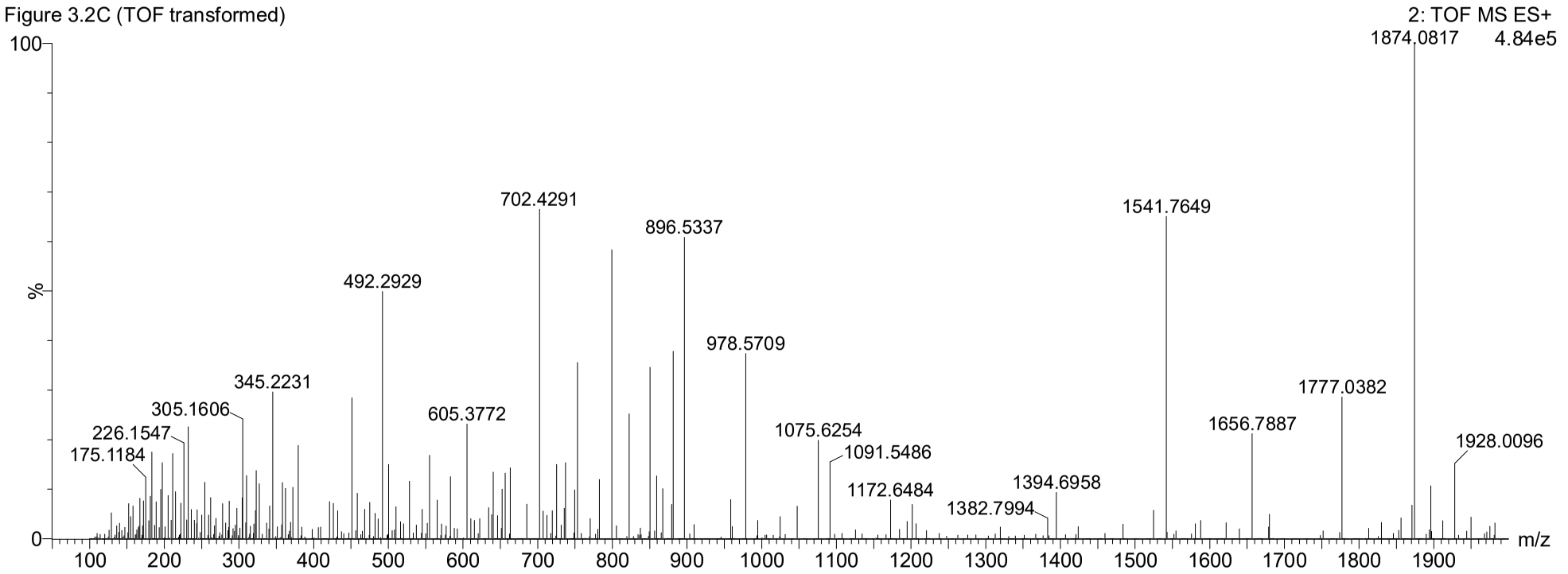


Figure 3.2C (Lock mass corrected)

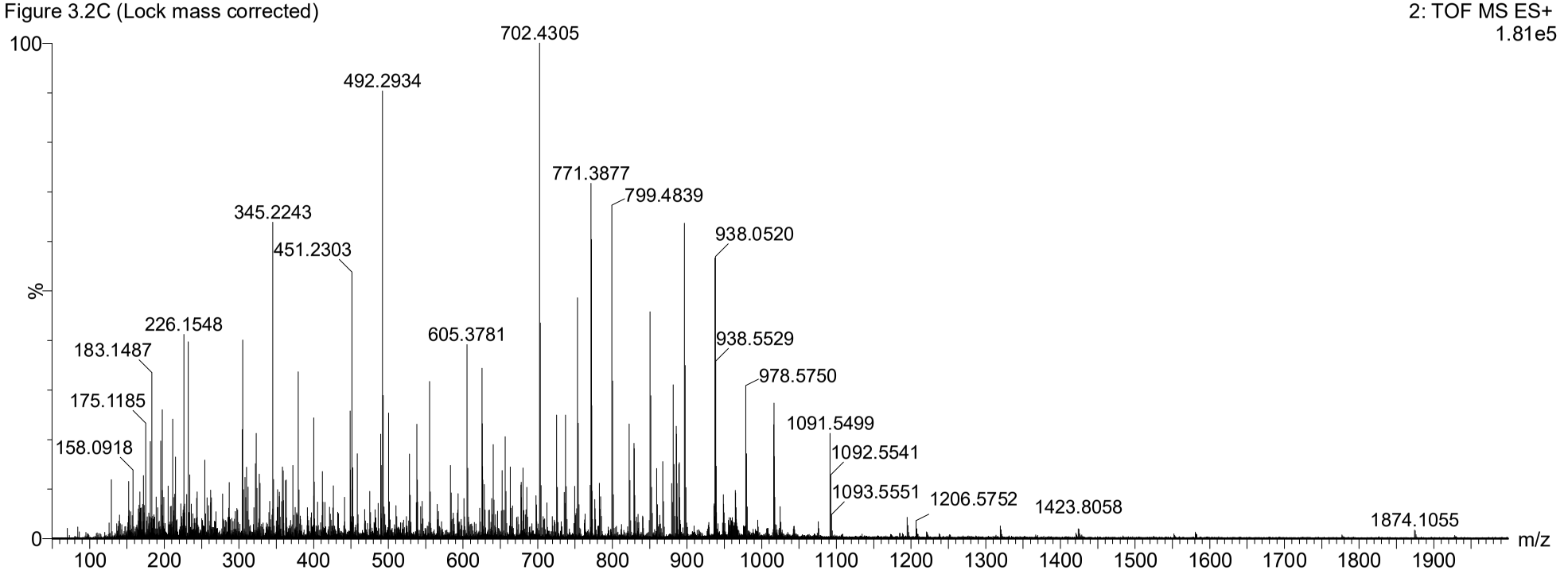


Figure 3.2F (Precursor, TOF transformed)

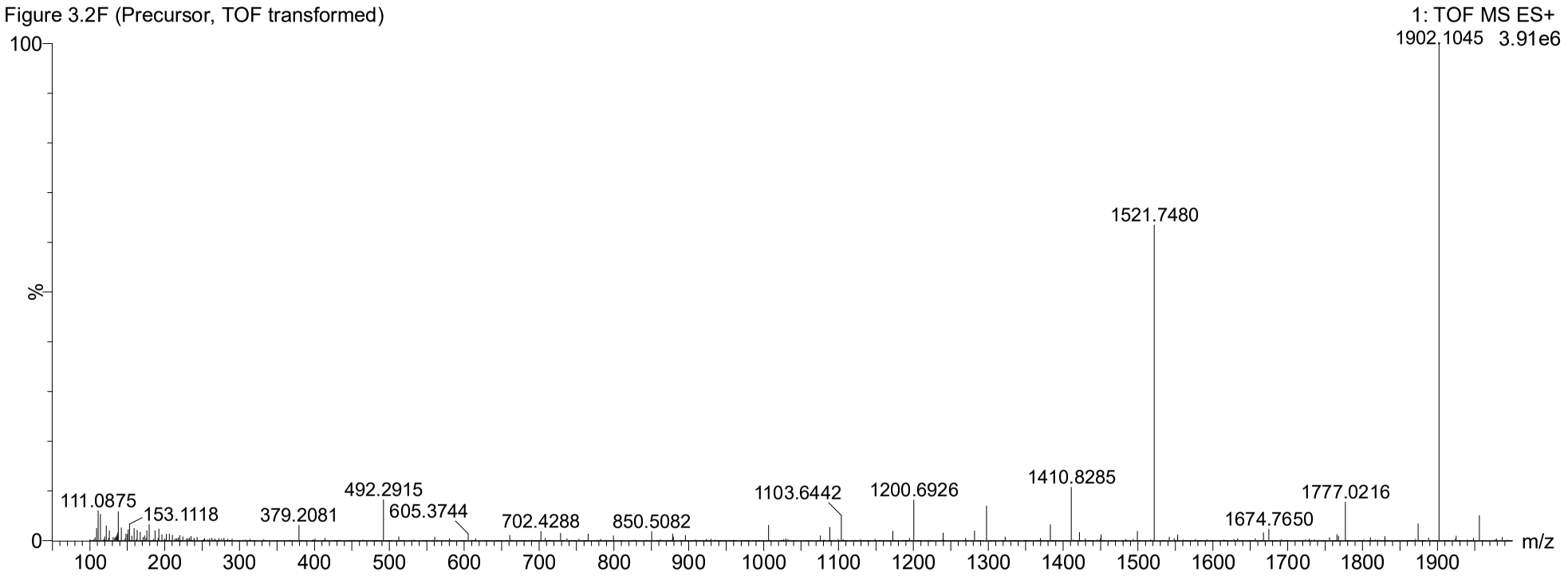


Figure 3.2F (Precursor, Lock mass corrected)

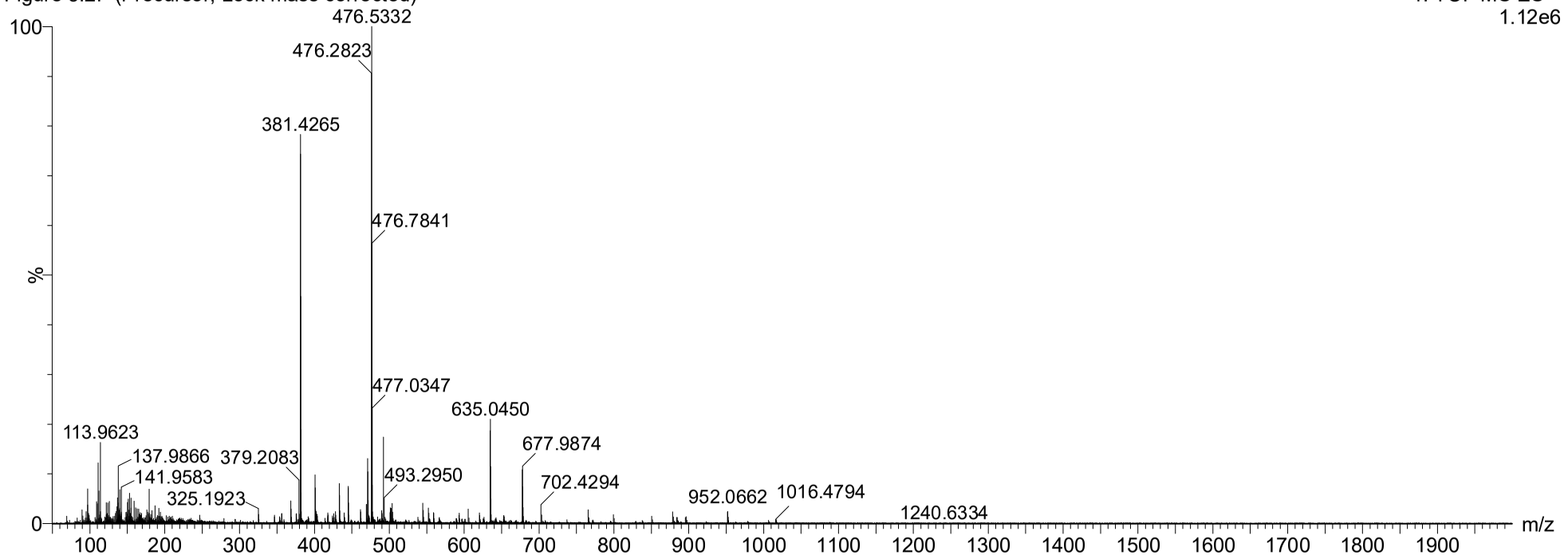


Figure 3.2F (TOF transformed)

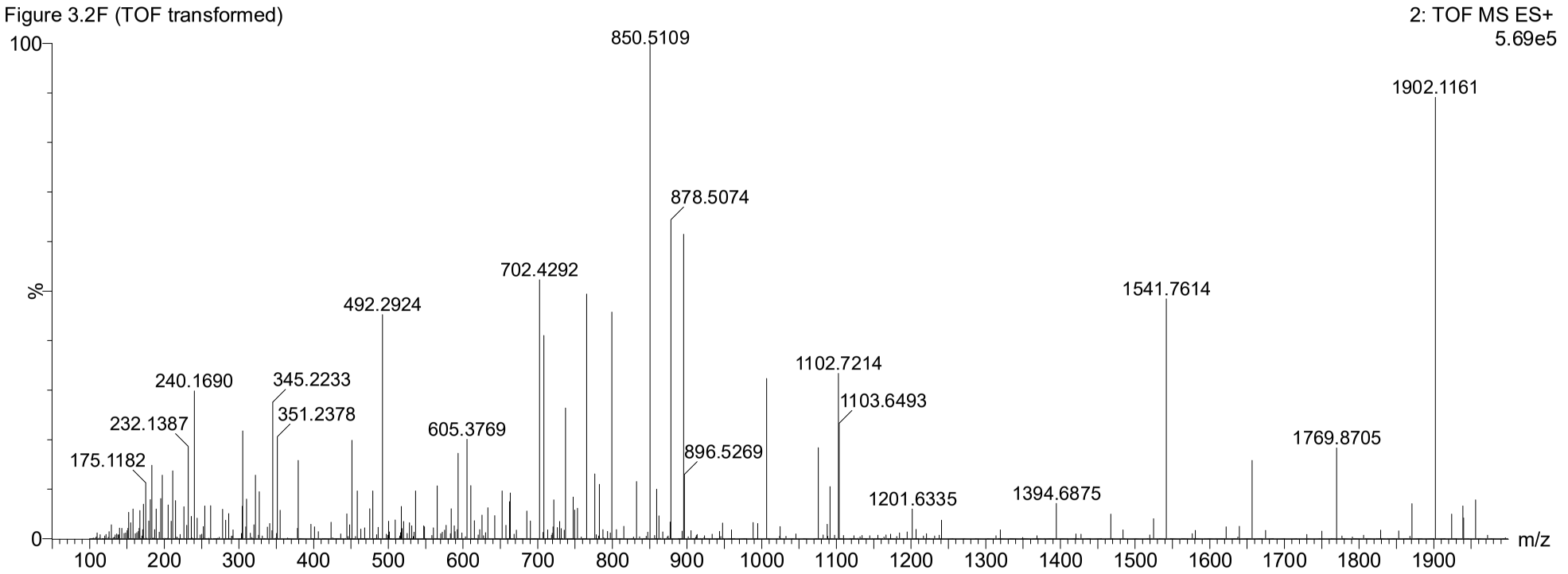


Figure 3.2F (Lock mass corrected)

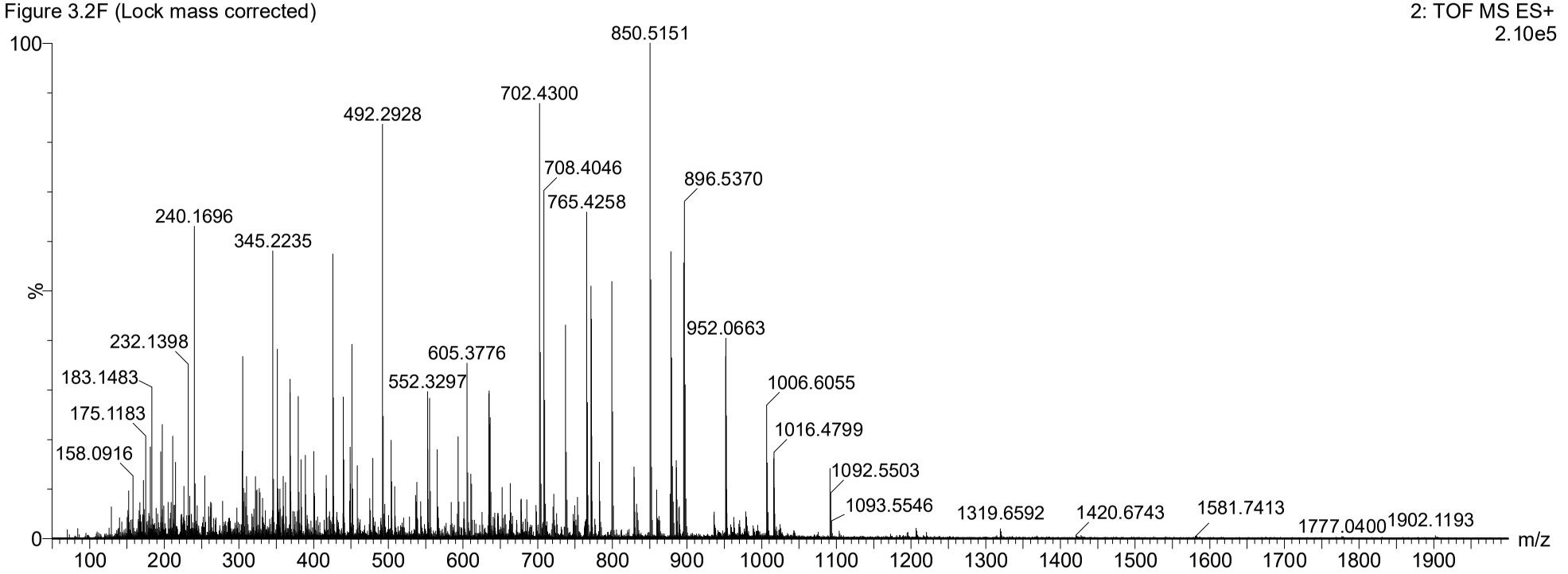


Figure 3.3F (smooth and background subtracted)

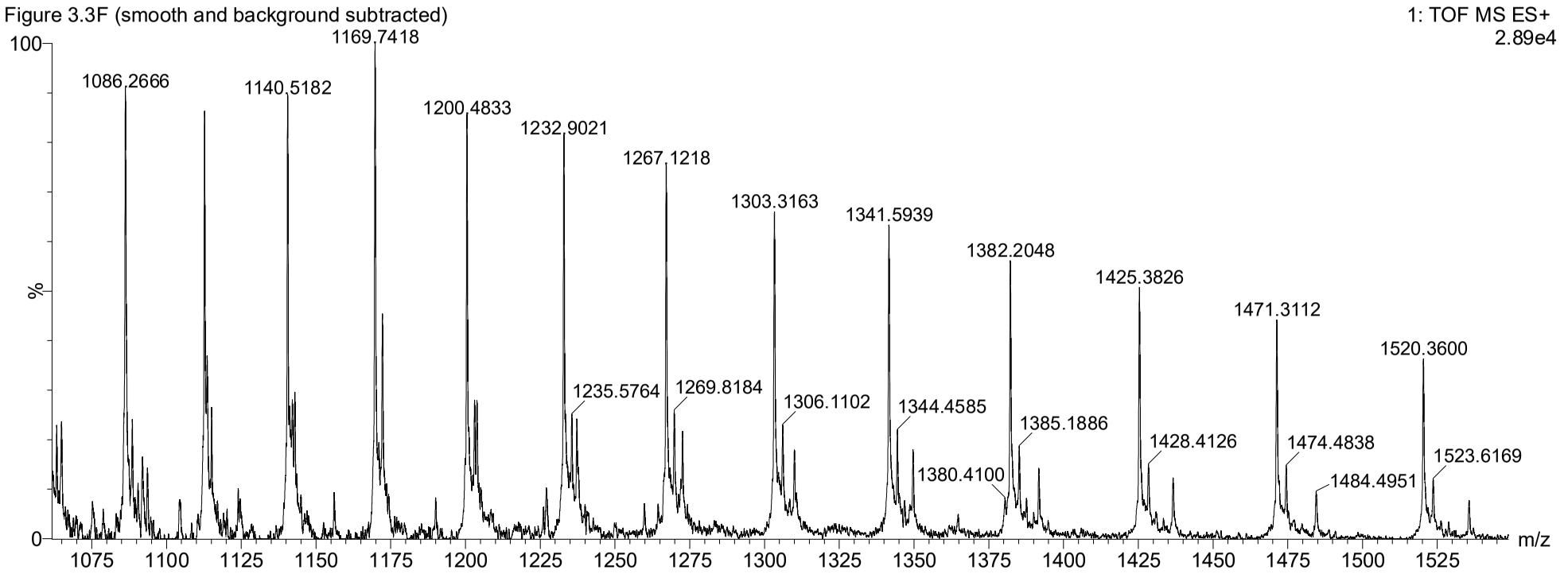


Figure 3.3G (smooth and background subtracted)

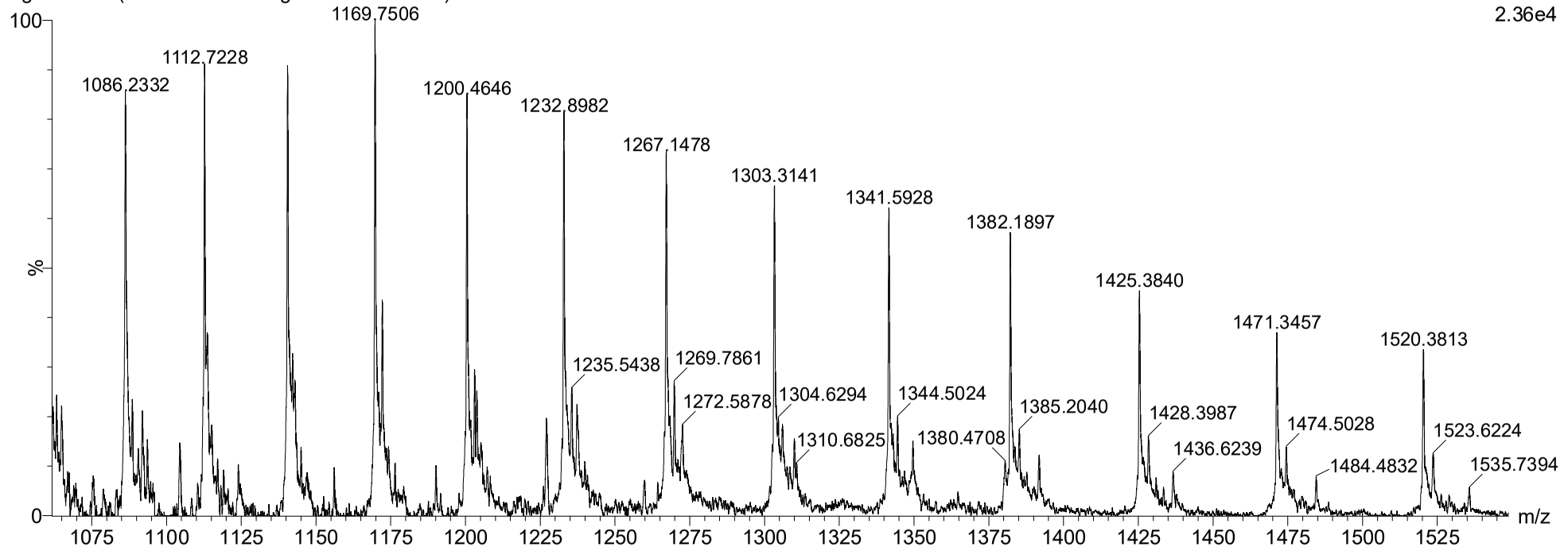


Figure 3.3A (TOF transformed)

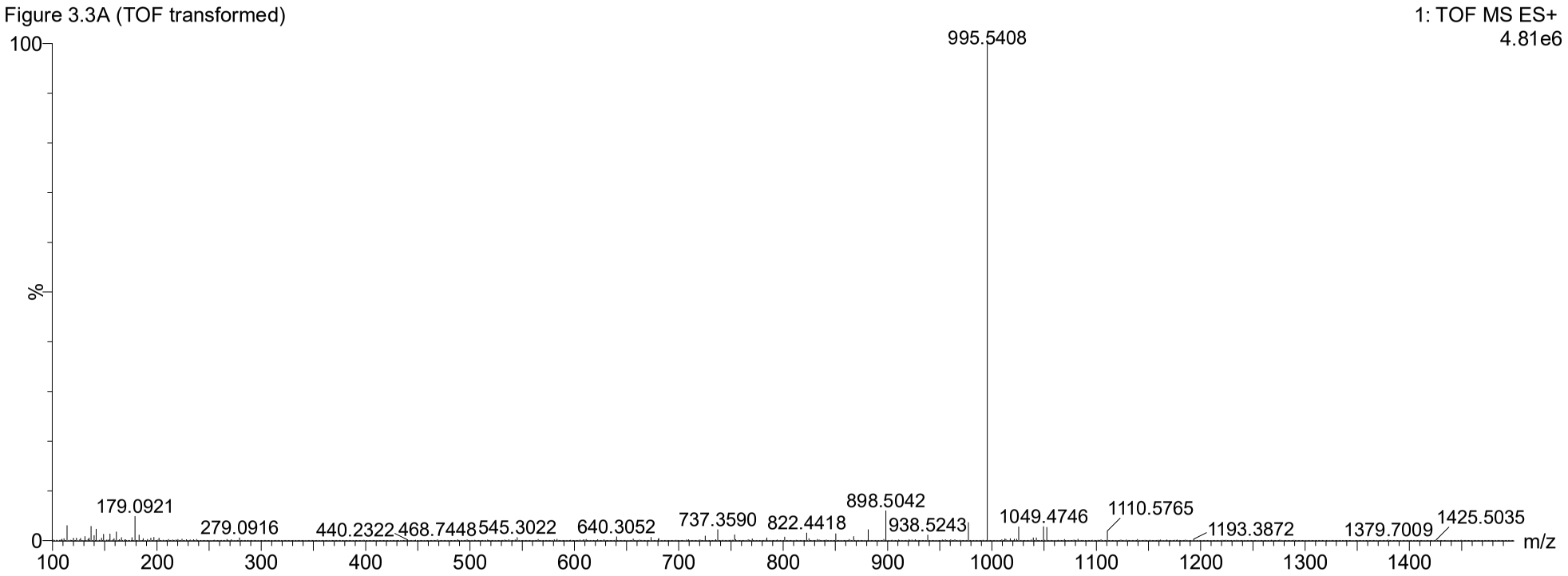


Figure 3.3A (Lock mass corrected)

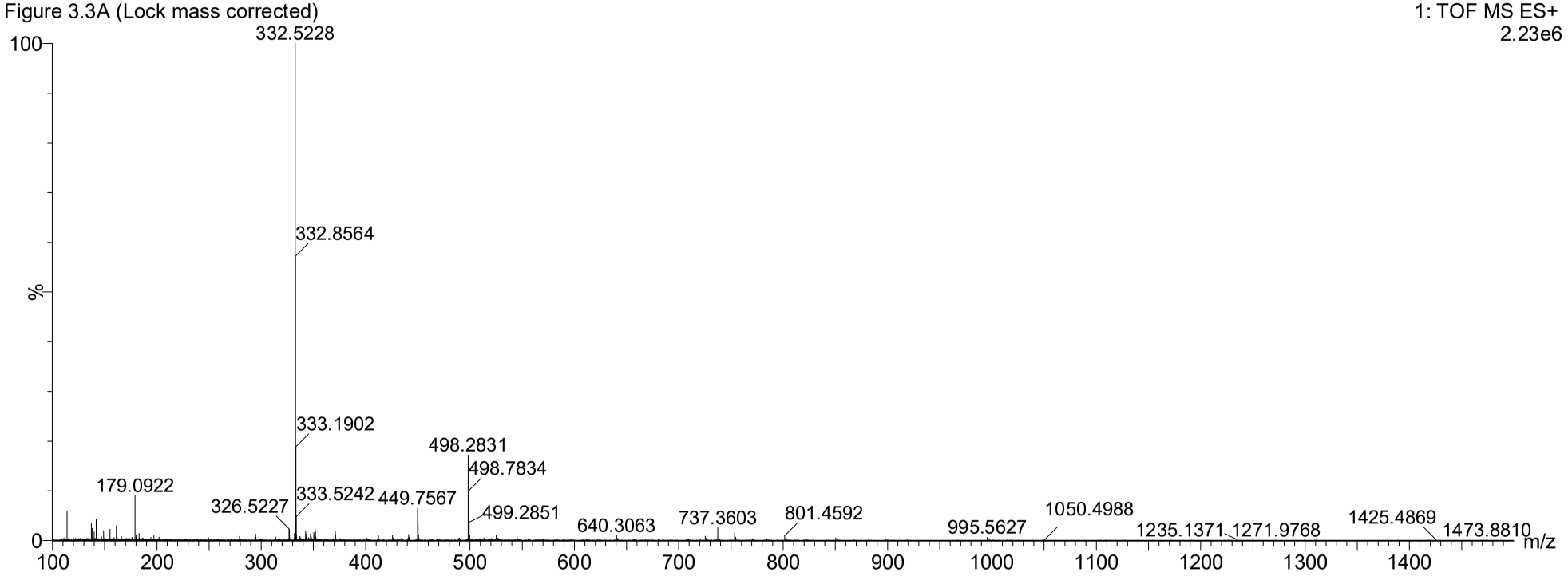


Figure 3.3B (TOF transformed)

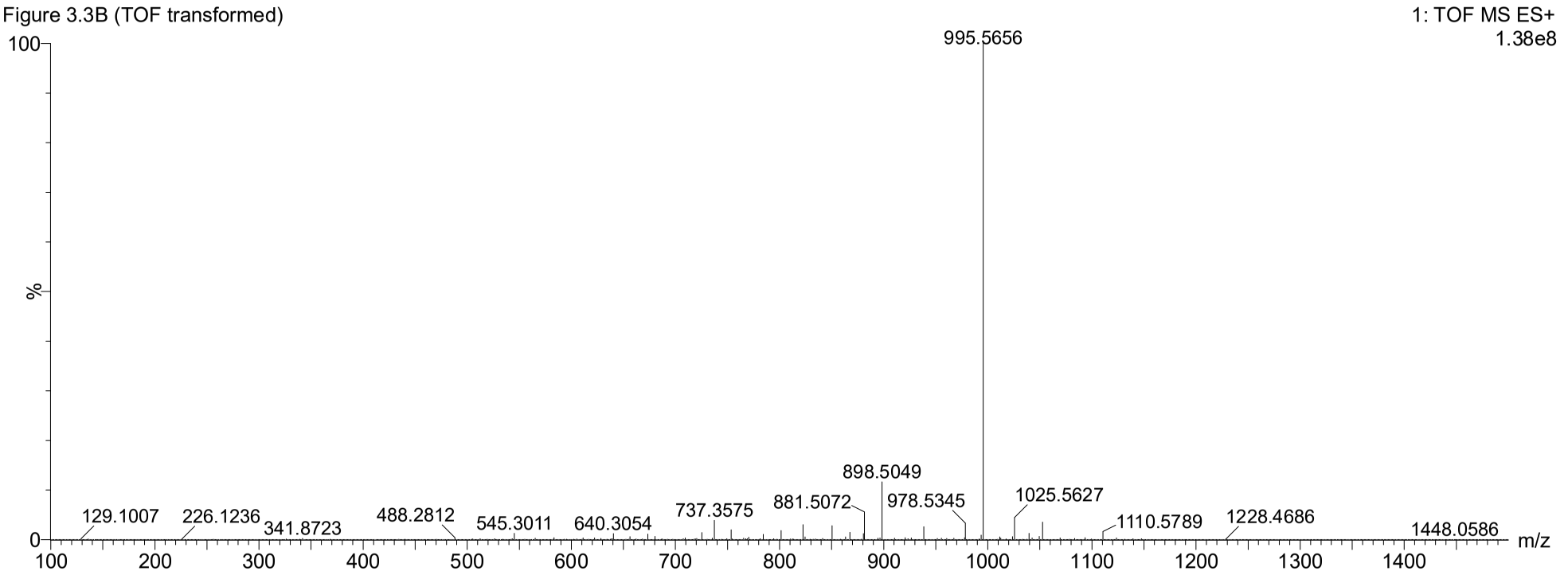


Figure 3.3B (Lock mass corrected)

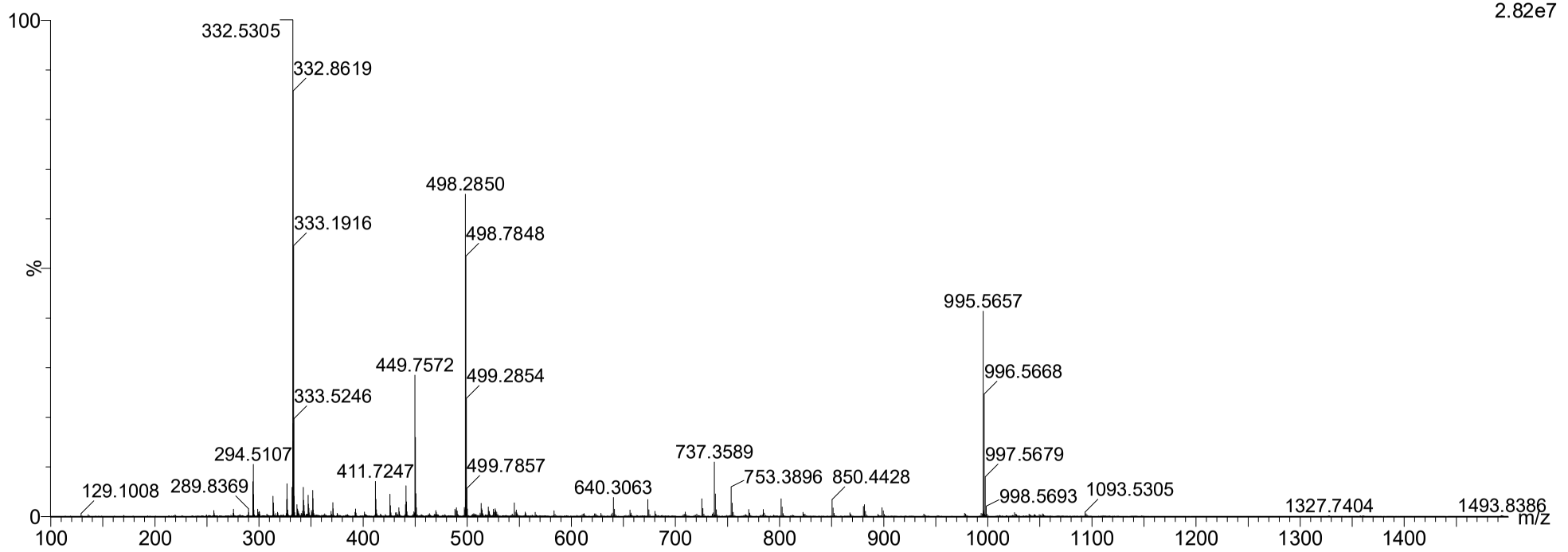


Figure 3.3C (TOF transformed)

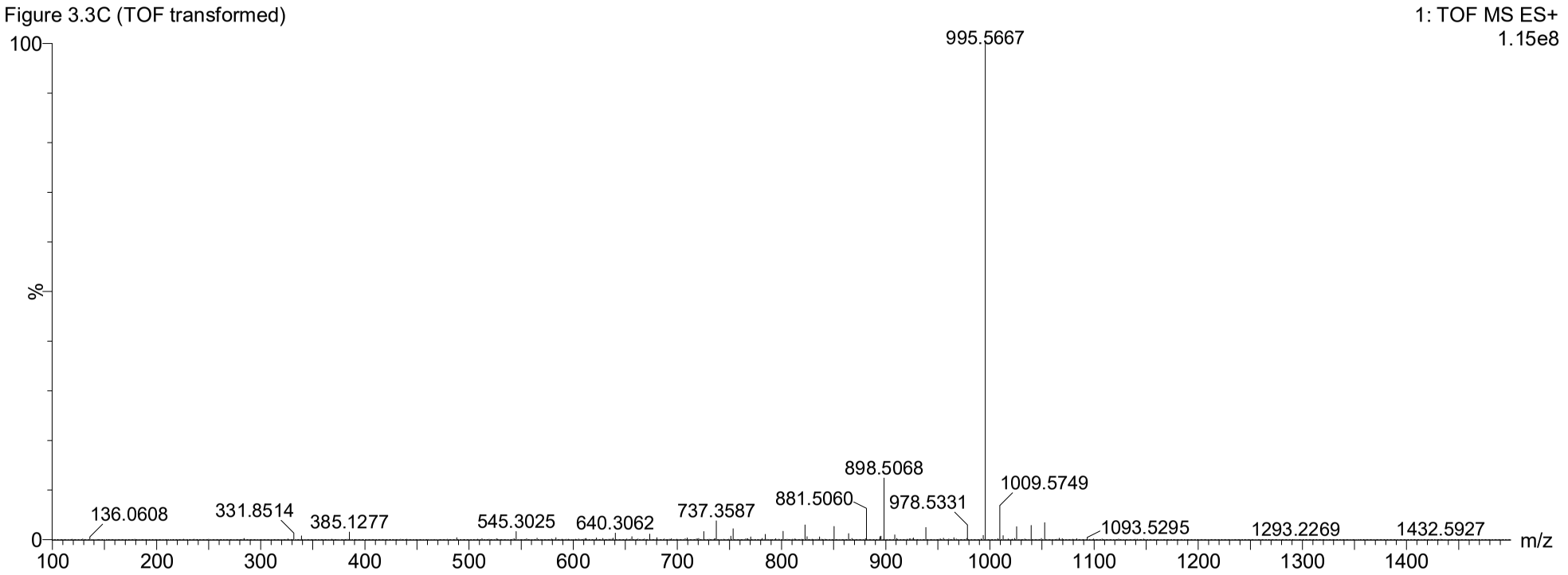


Figure 3.3C (Lock mass corrected)

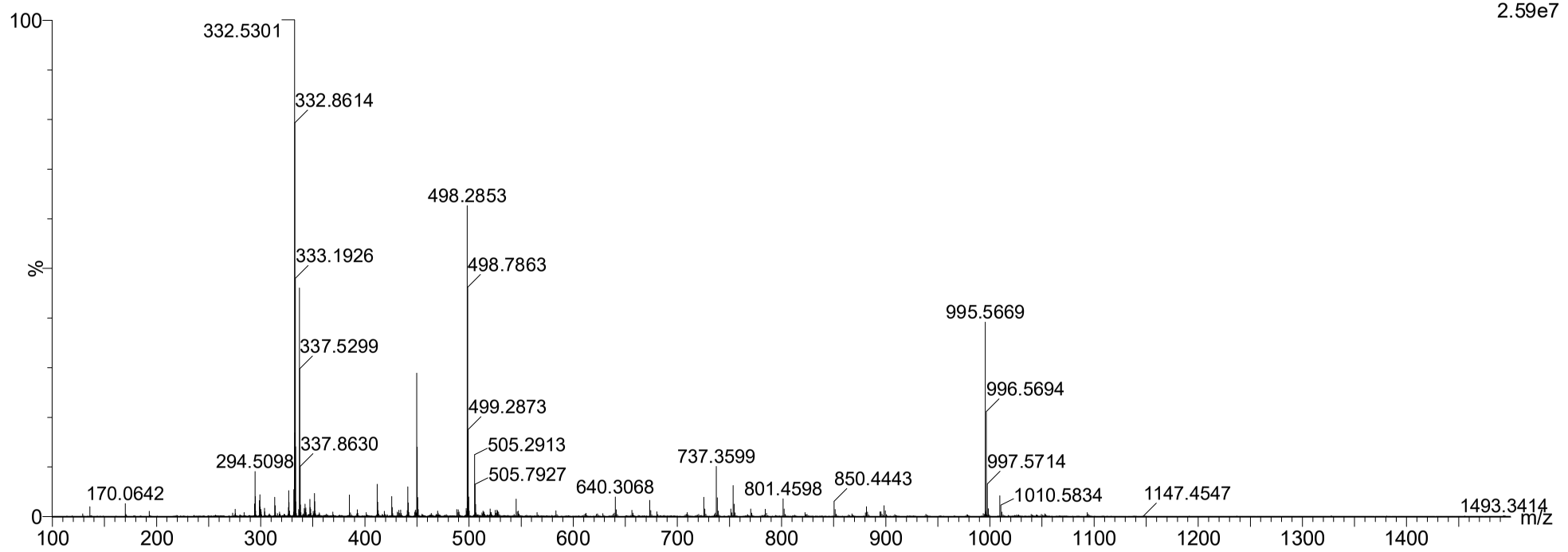


Figure 3.4A (Precursor, TOF transformed)

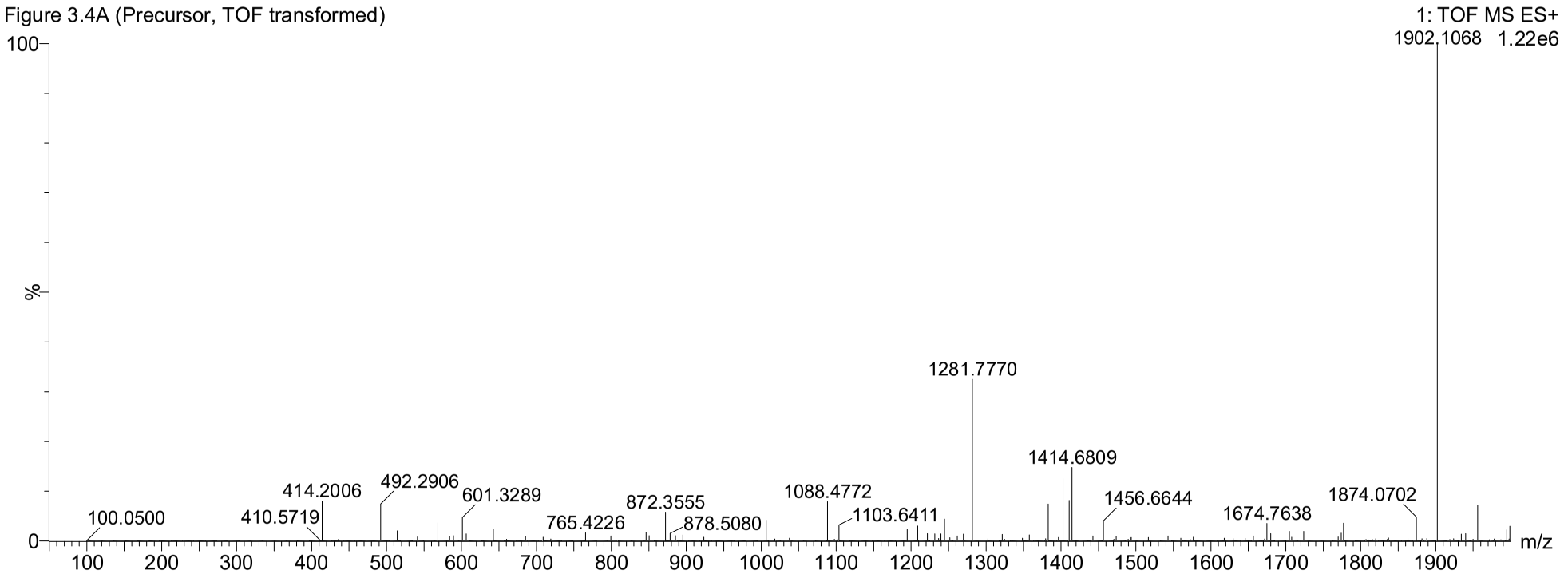


Figure 3.4A (Precursor, Lock mass corrected)

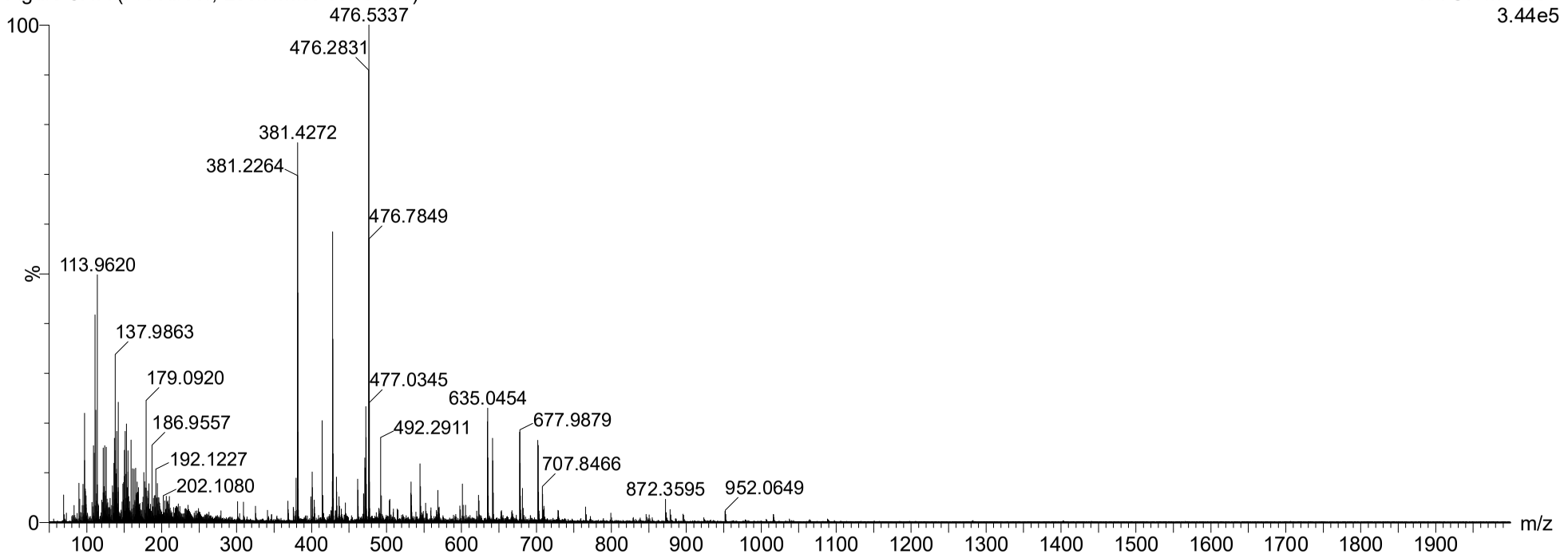


Figure 3.4A (TOF transformed)

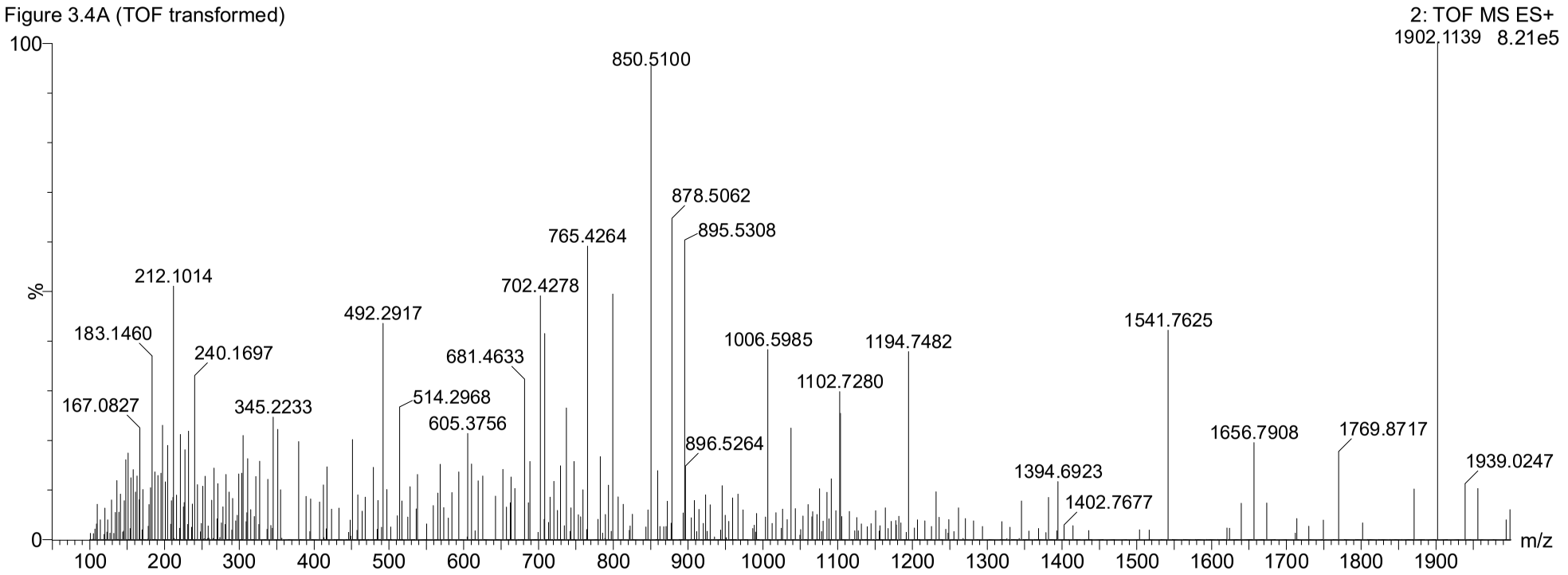


Figure 3.4A (Lock mass corrected)

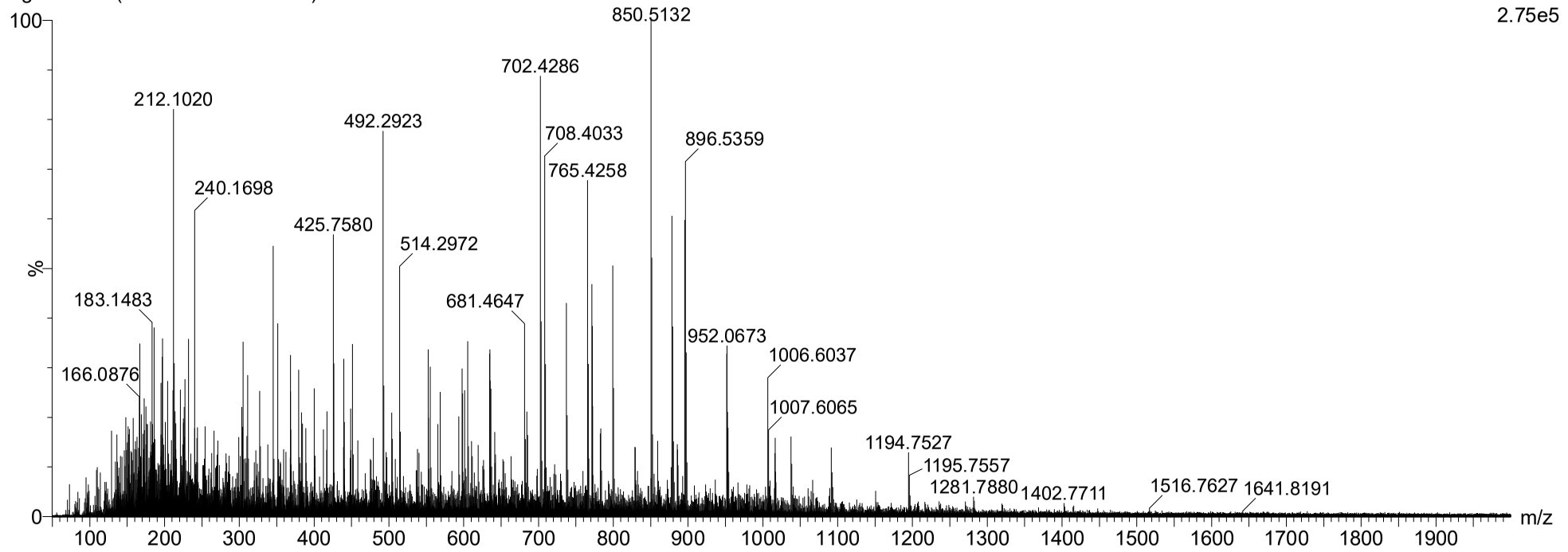


Figure 3.4B (Precursor, TOF transformed)

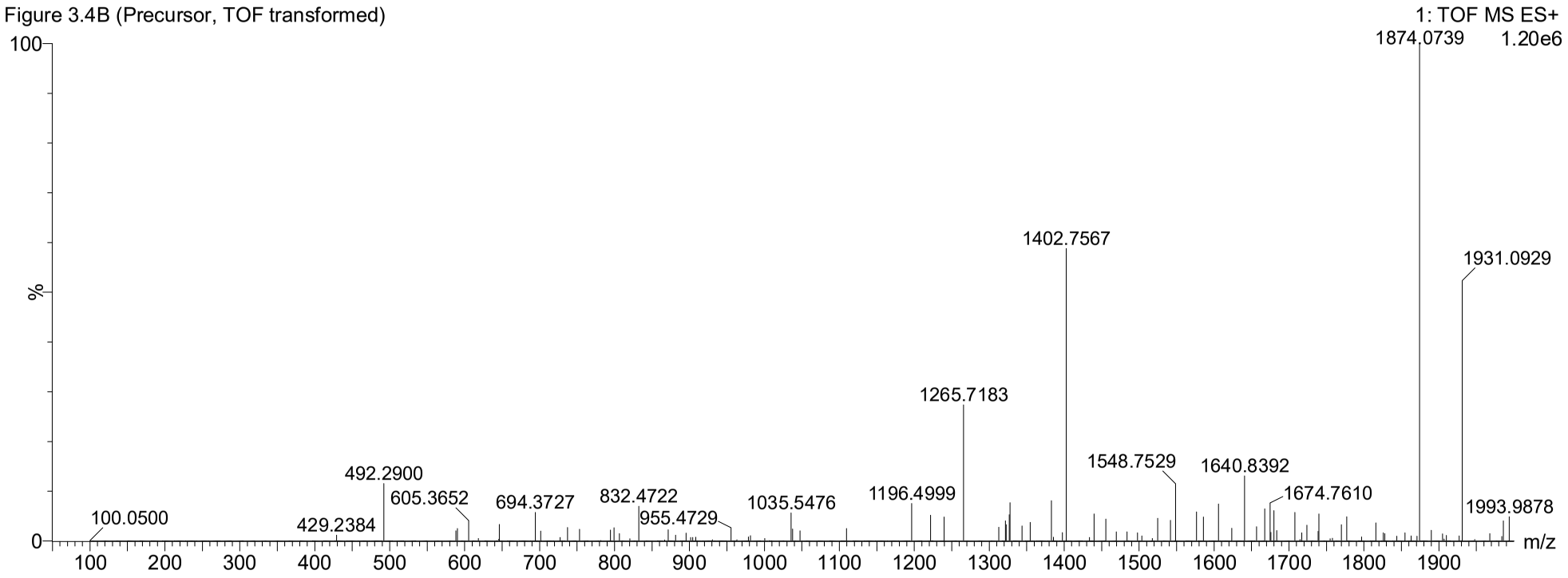
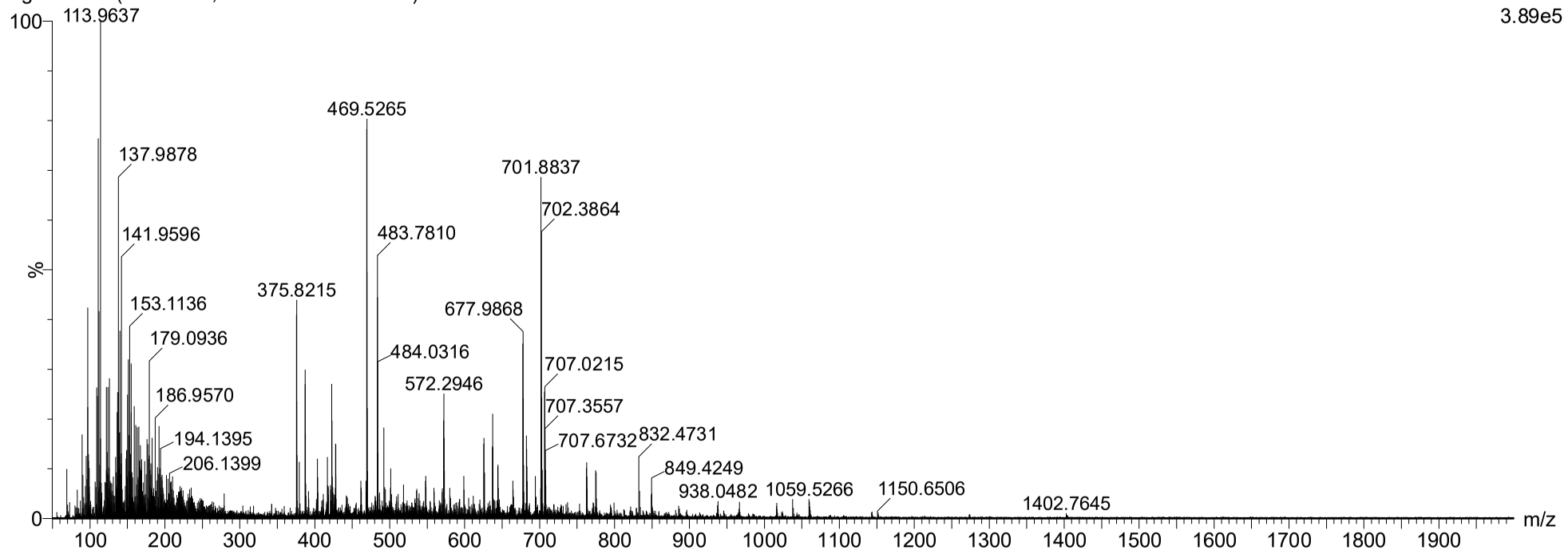


Figure 3.4B (Precursor, Lock mass corrected)



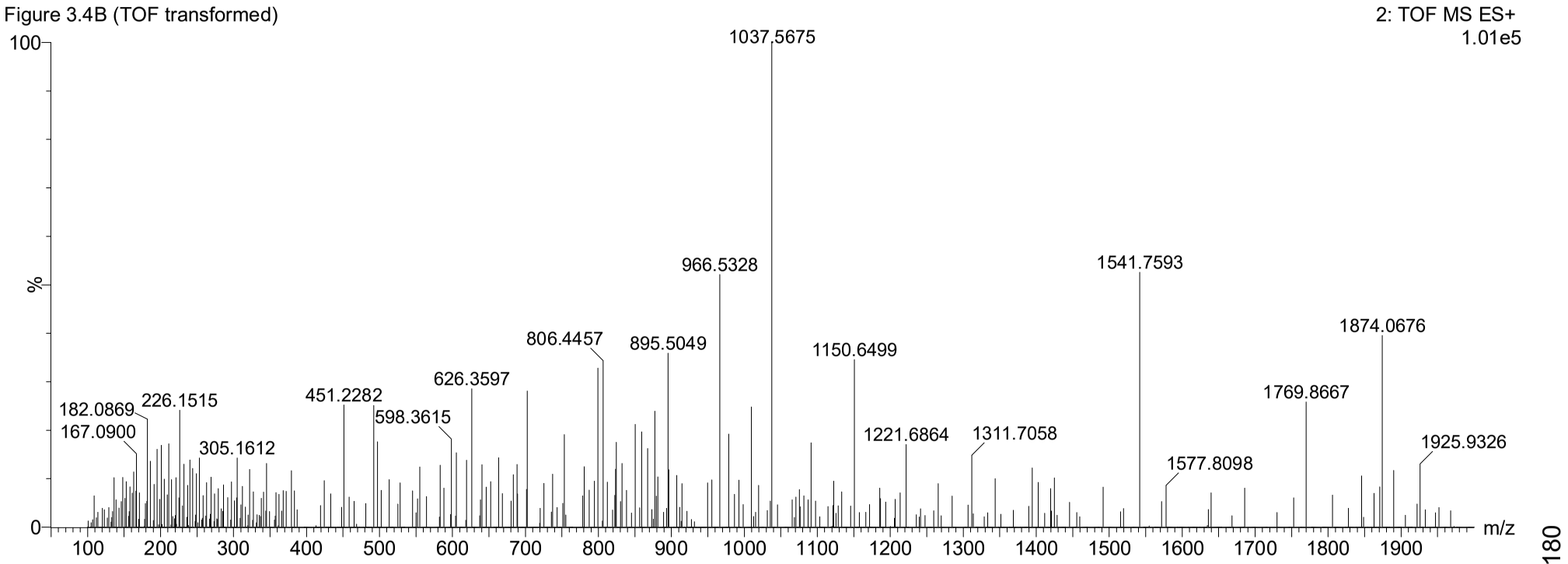


Figure 3.4B (TOF transformed)
2: TOF MS ES+
1.01e5
100
%
0
1037.5675
966.5328
1541.7593
1874.0676
806.4457
895.5049
1150.6499
626.3597
451.2282
182.0869
226.1515
598.3615
167.0900
305.1612
1221.6864
1311.7058
1769.8667
1925.9326
1577.8098
m/z
100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900

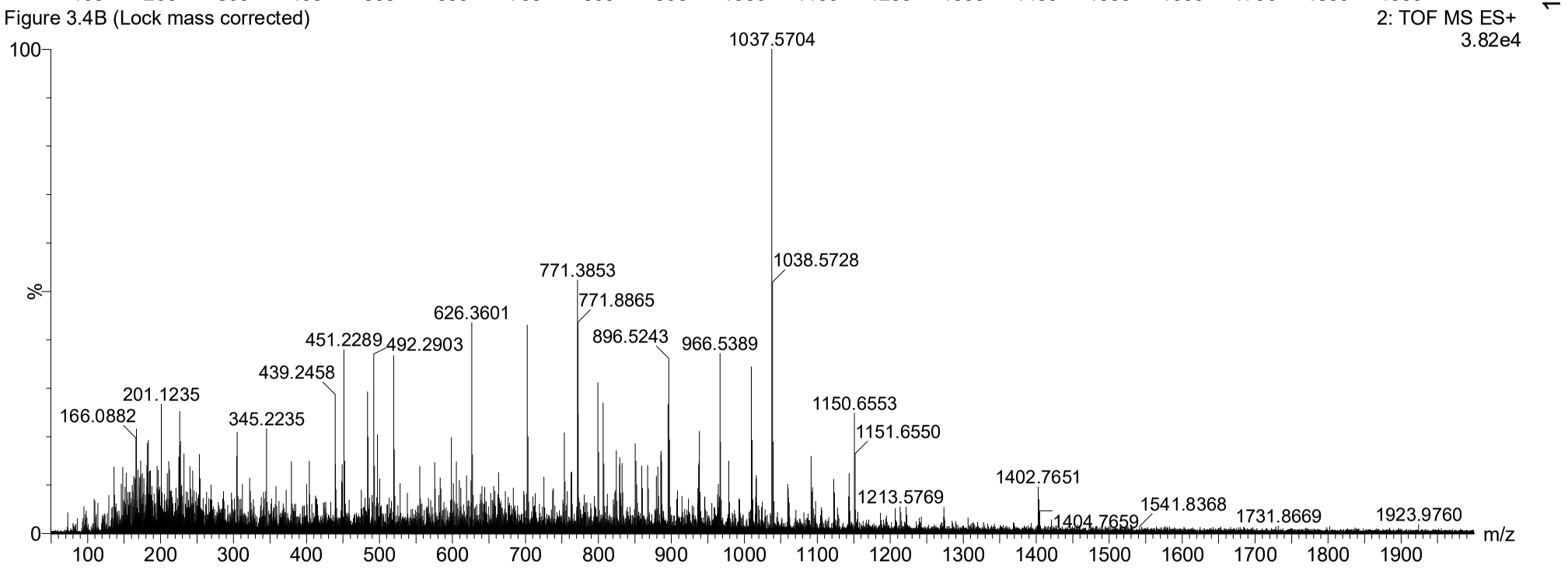


Figure 3.4B (Lock mass corrected)
2: TOF MS ES+
3.82e4
100
%
0
1037.5704
1038.5728
771.3853
771.8865
626.3601
896.5243
966.5389
451.2289
492.2903
439.2458
201.1235
166.0882
345.2235
1150.6553
1151.6550
1402.7651
1213.5769
1541.8368
1731.8669
1923.9760
1404.7659
m/z
100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900

Figure 3.4C (Precursor, TOF transformed)

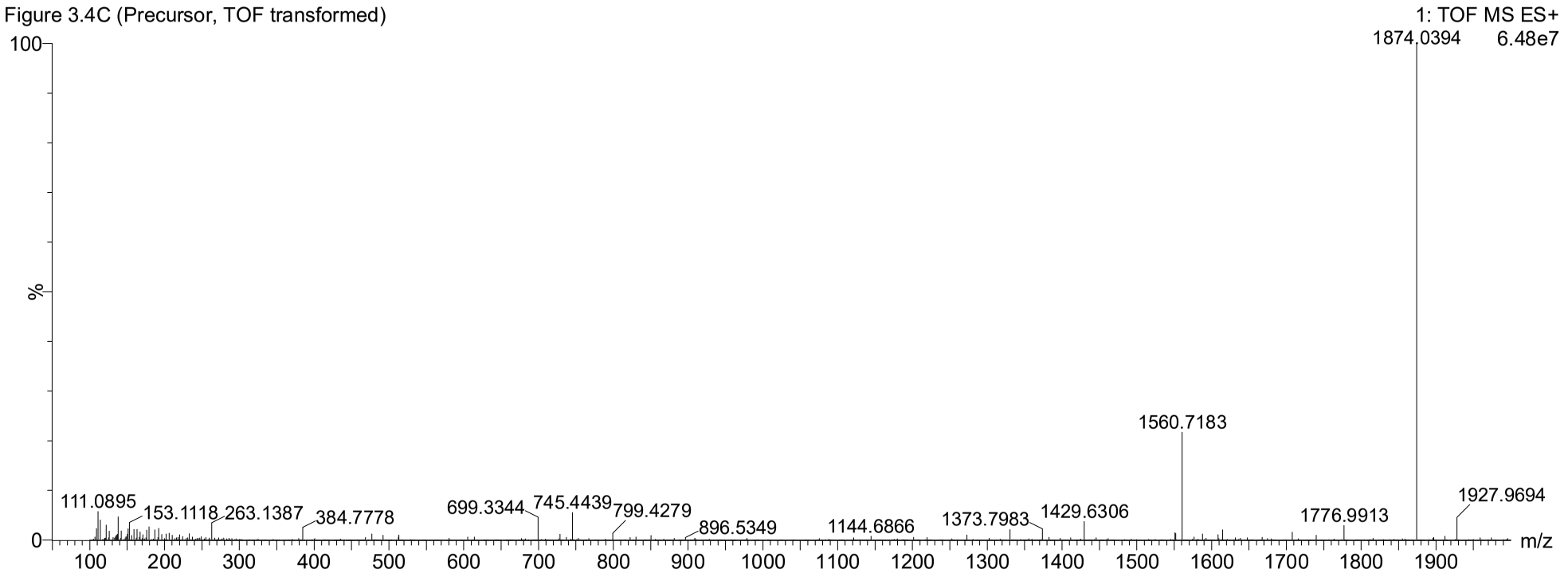


Figure 3.4C (Precursor, Lock mass corrected)

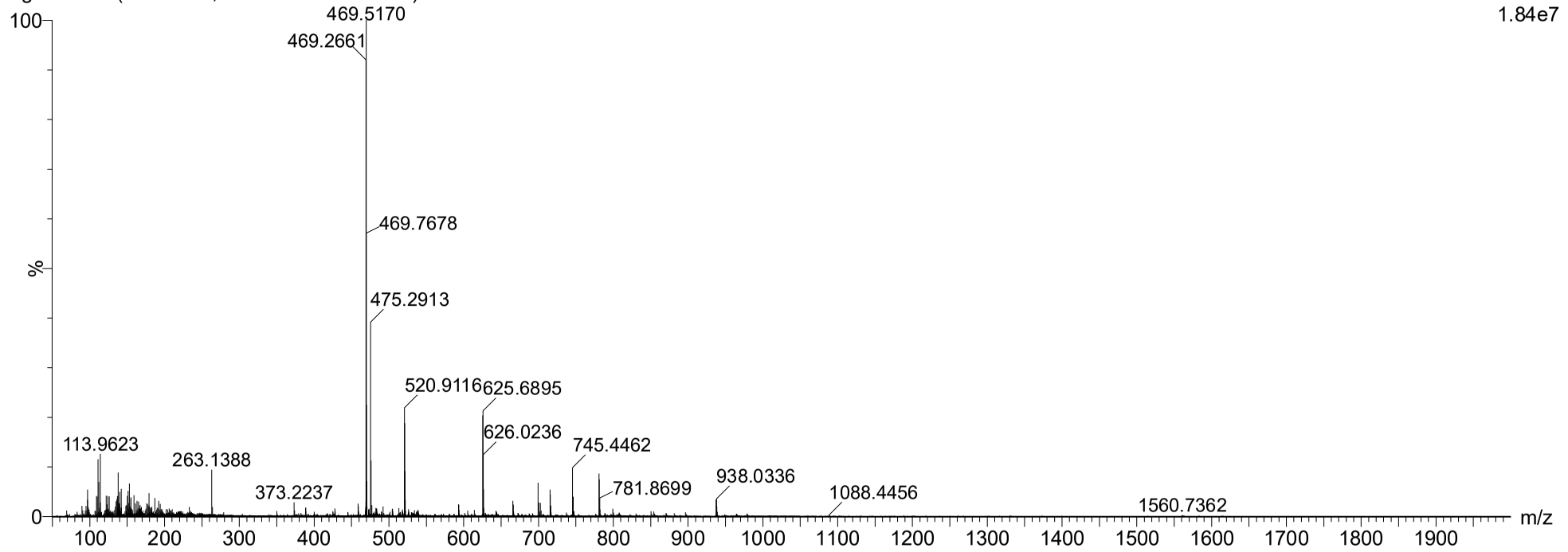


Figure 3.4C (TOF transformed)

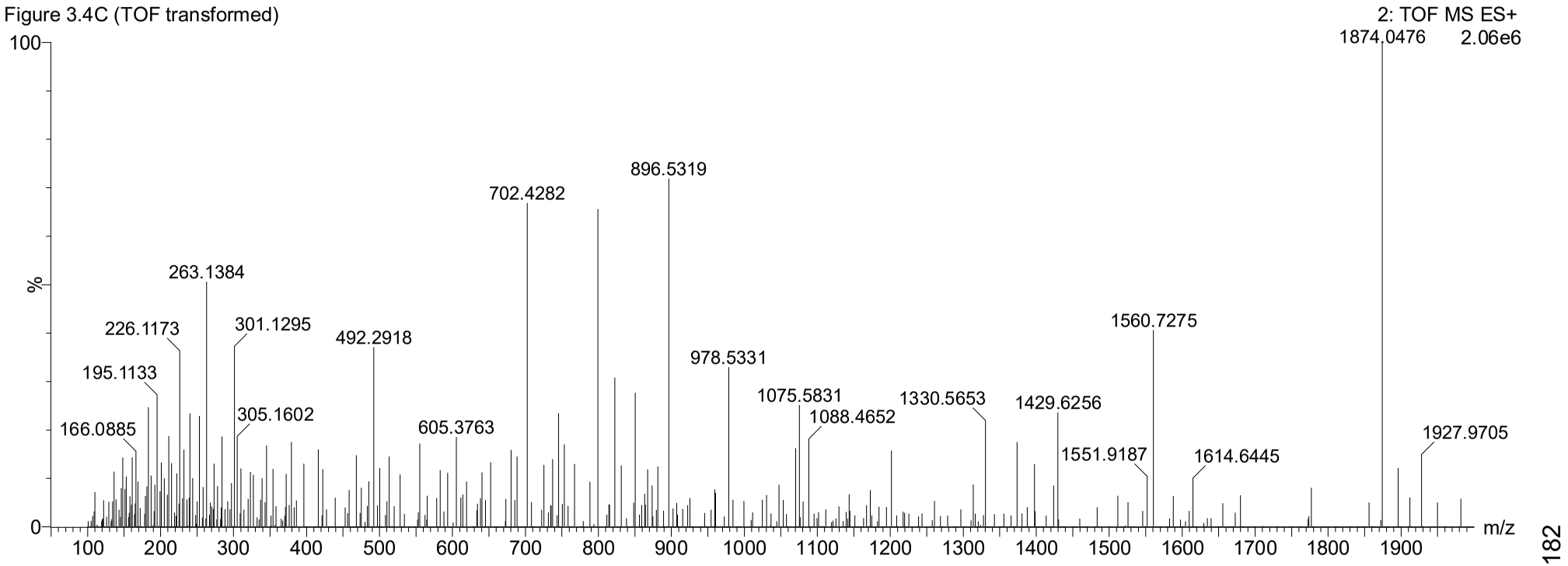


Figure 3.4C (Lock mass corrected)

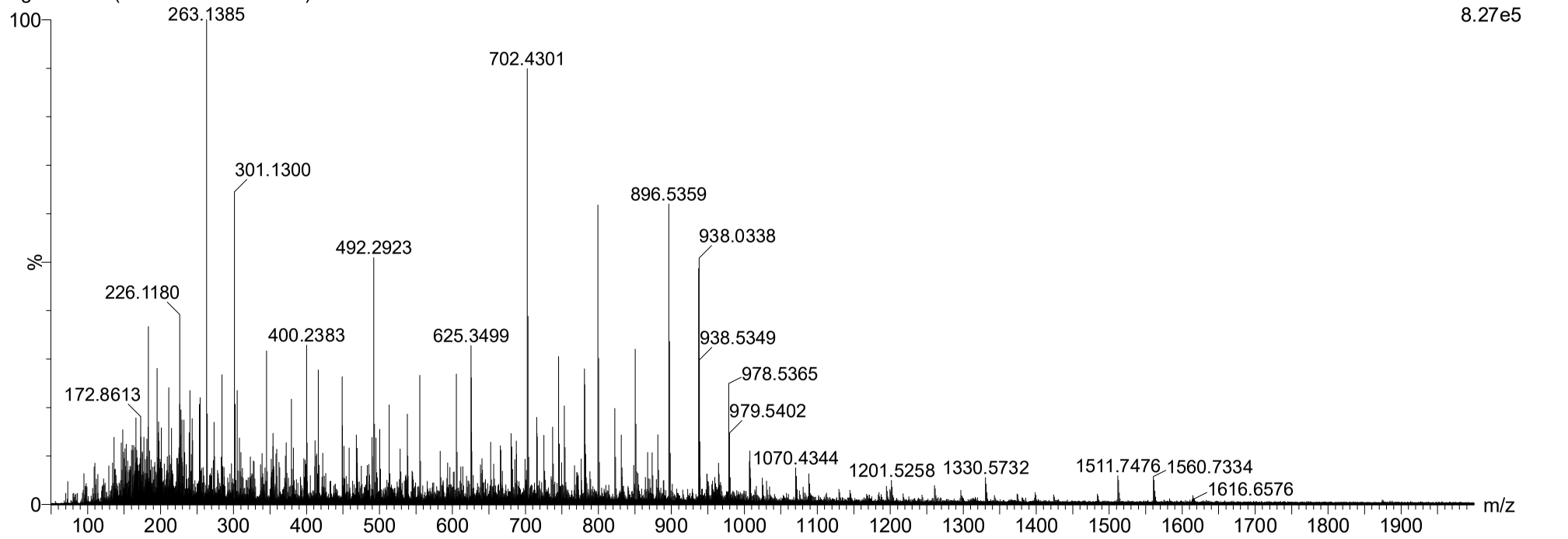


Figure 3.5A (smooth and background subtracted)

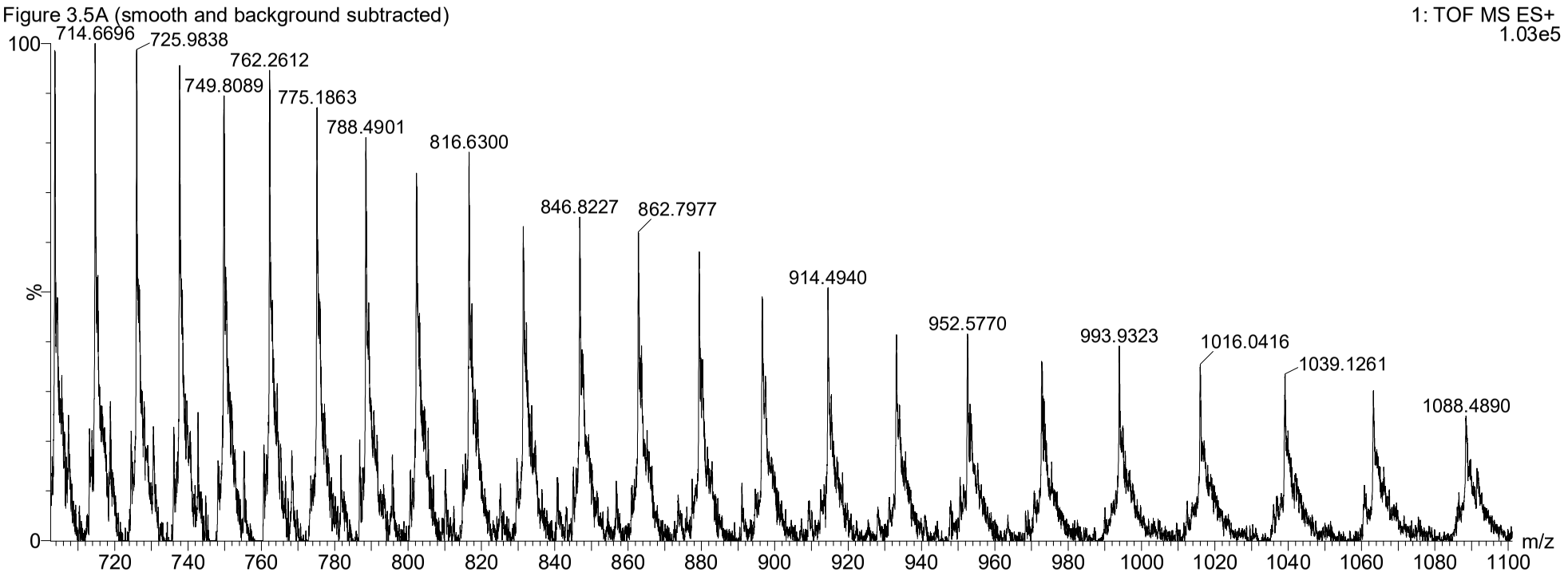
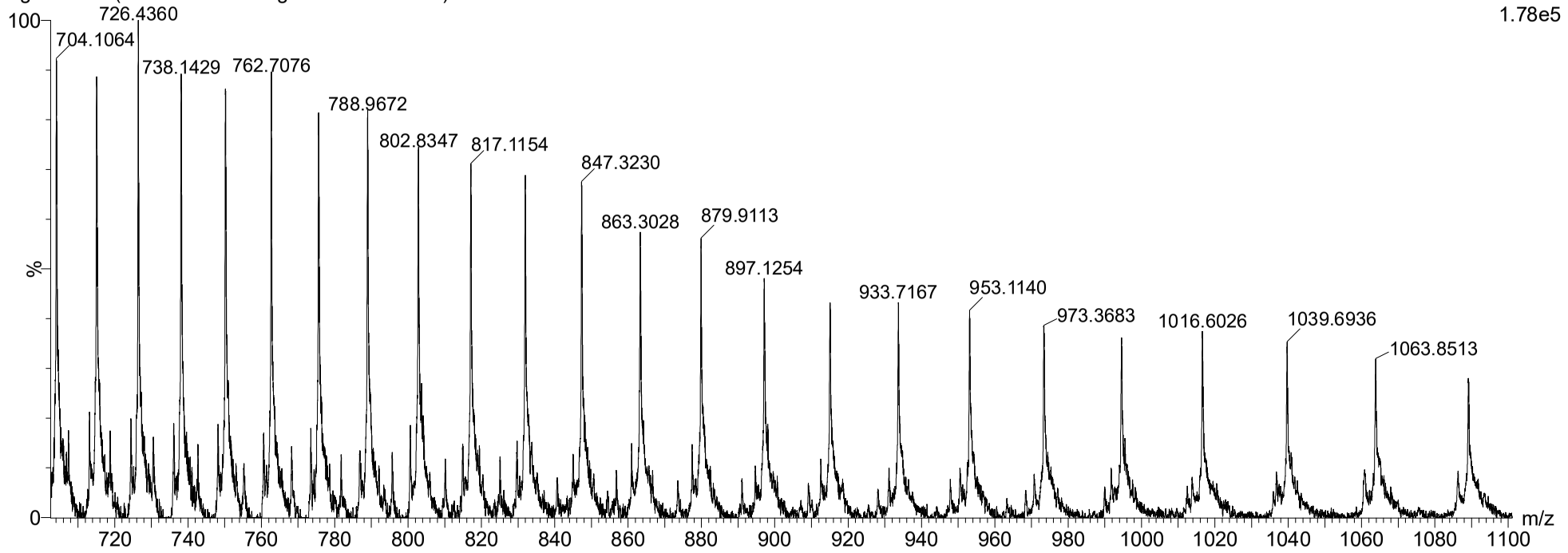
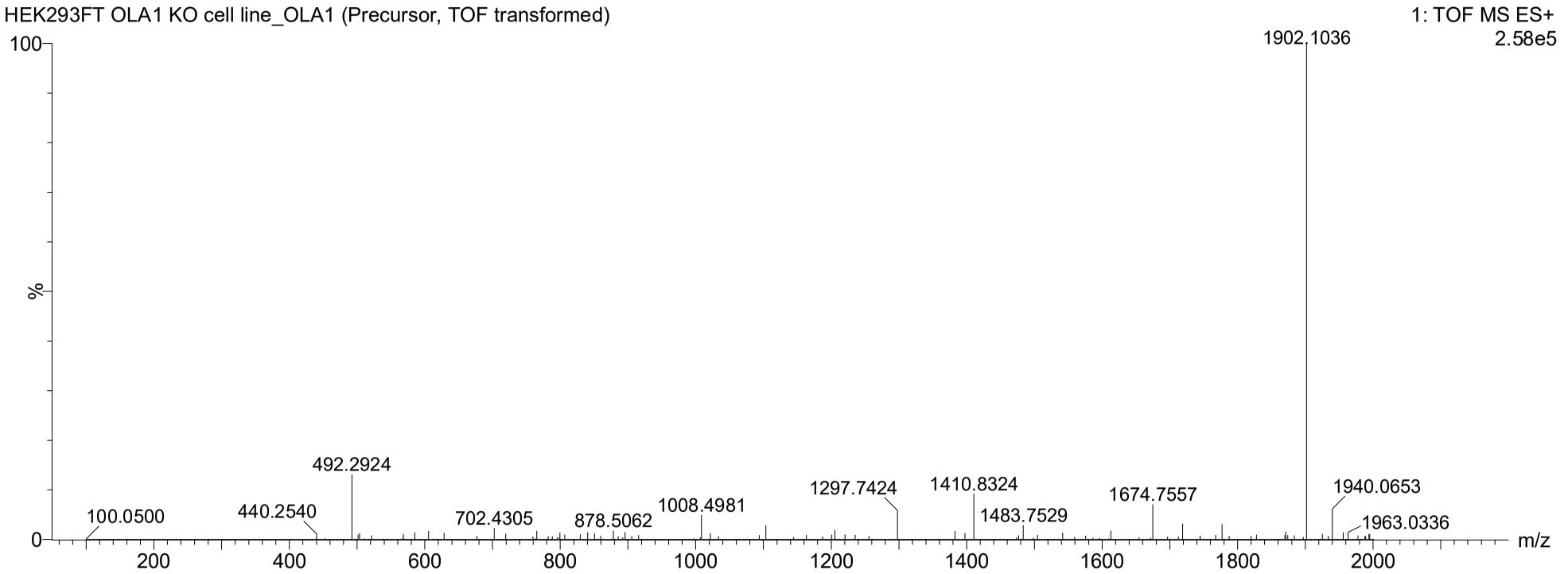
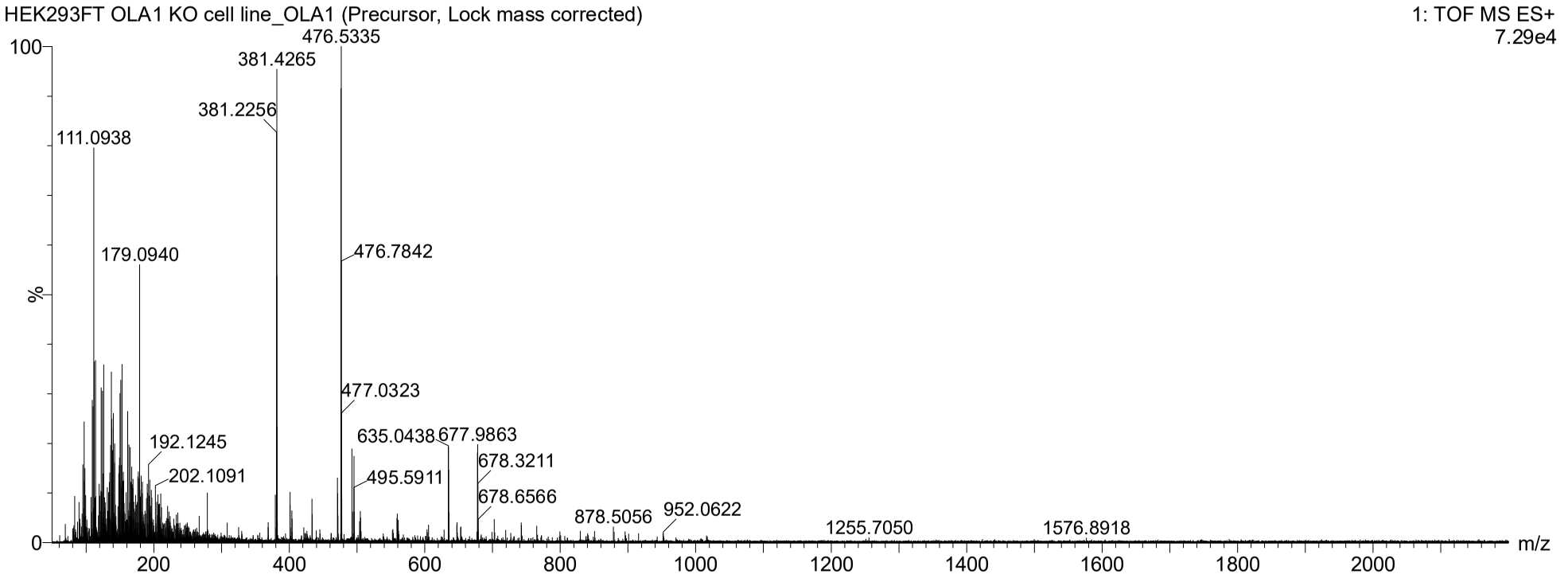
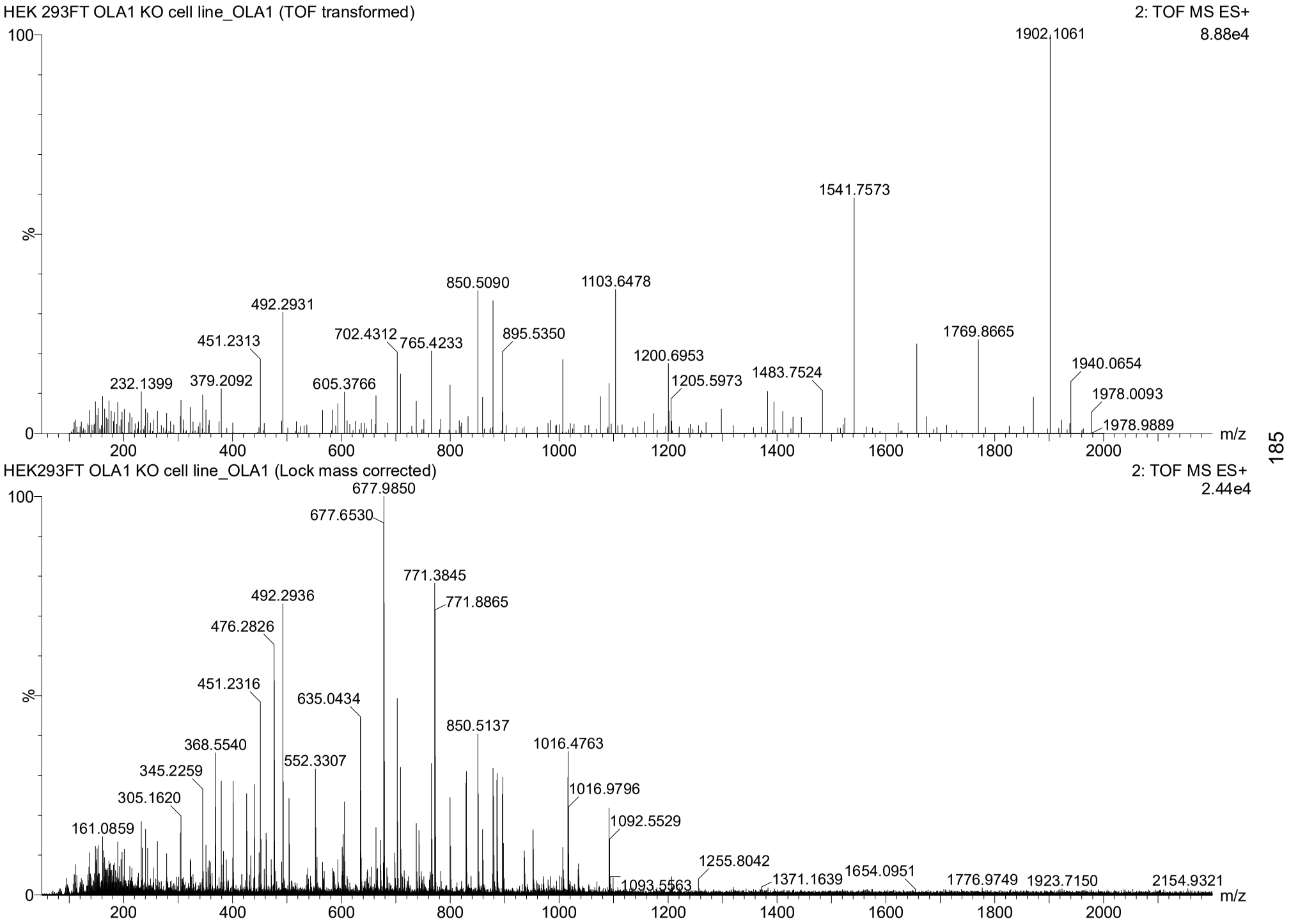


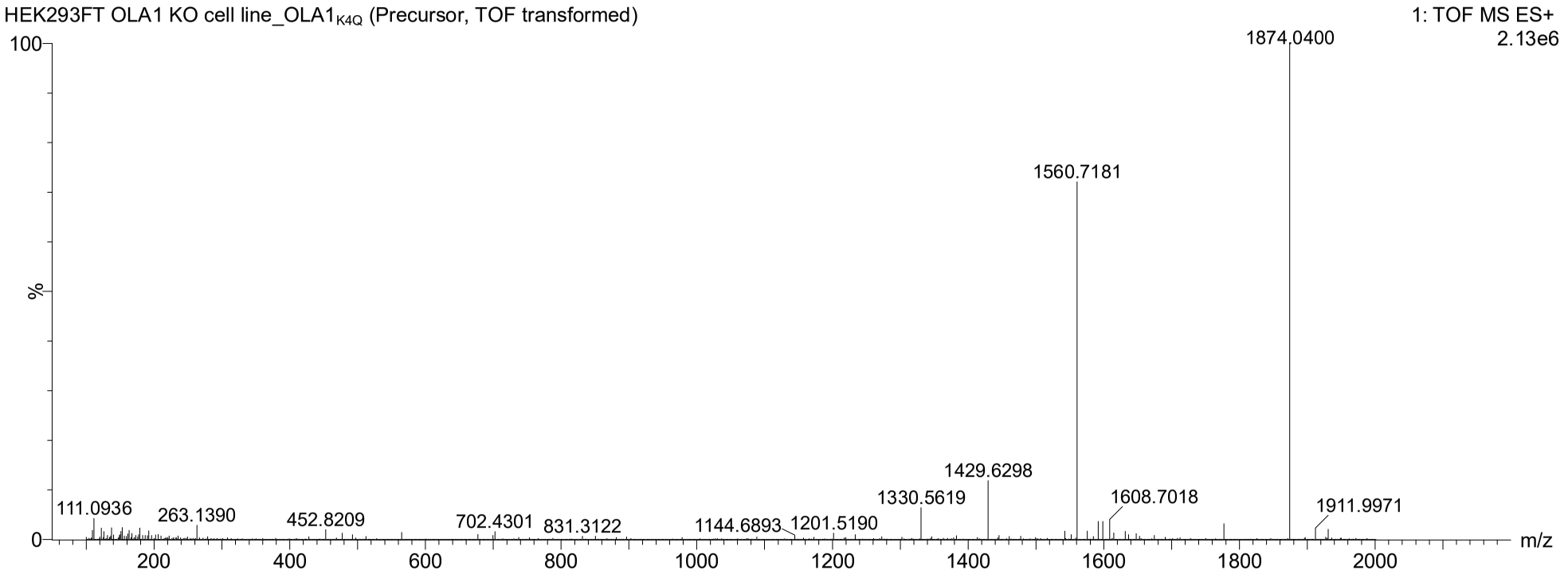
Figure 3.5B (smooth and background subtracted)

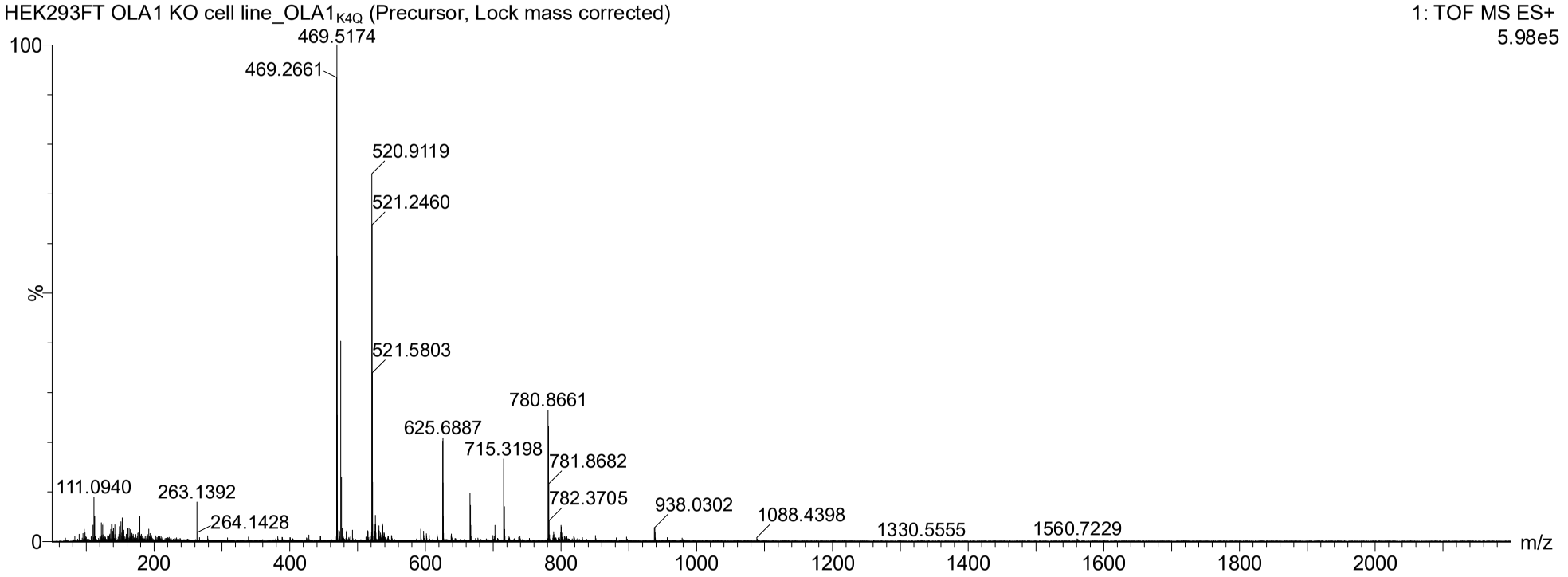


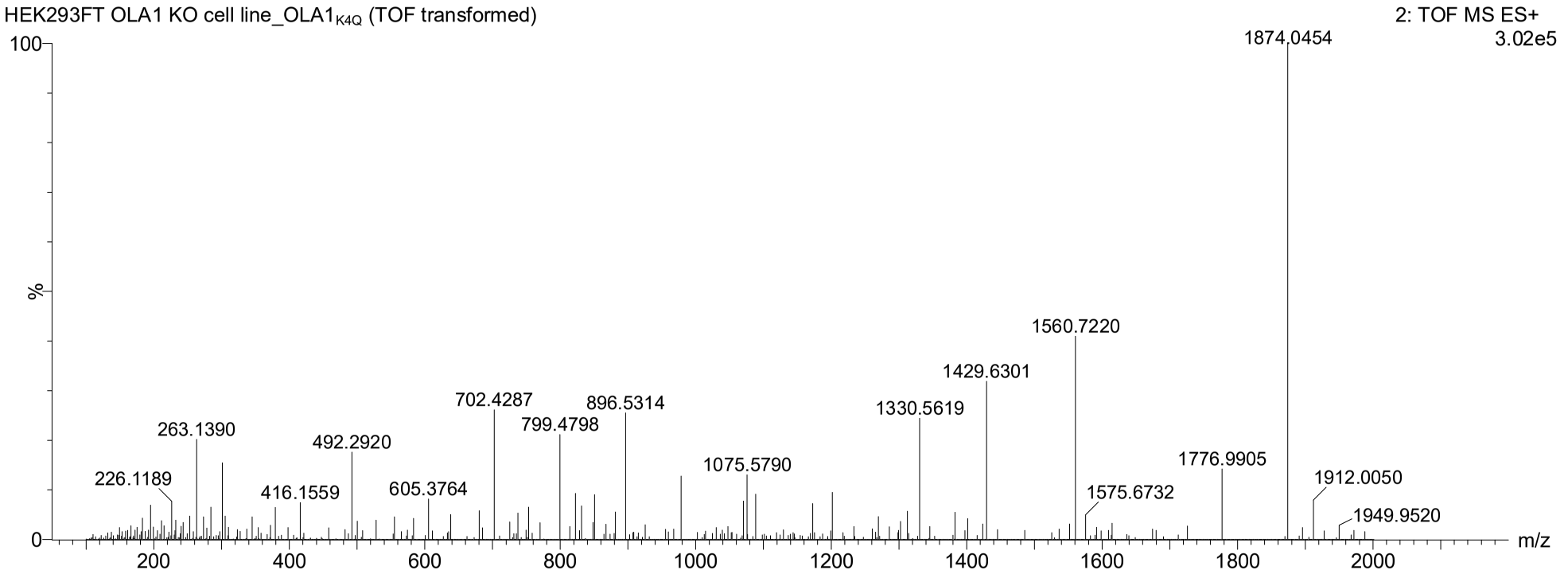

HEK293FT OLA1 KO cell line_OLA1 (Precursor, TOF transformed)
1: TOF MS ES+
2.58e5
100
%
0
1902.1036
492.2924
100.0500
440.2540
702.4305
878.5062
1008.4981
1297.7424
1410.8324
1483.7529
1674.7557
1940.0653
1963.0336
m/z
200
400
600
800
1000
1200
1400
1600
1800
2000

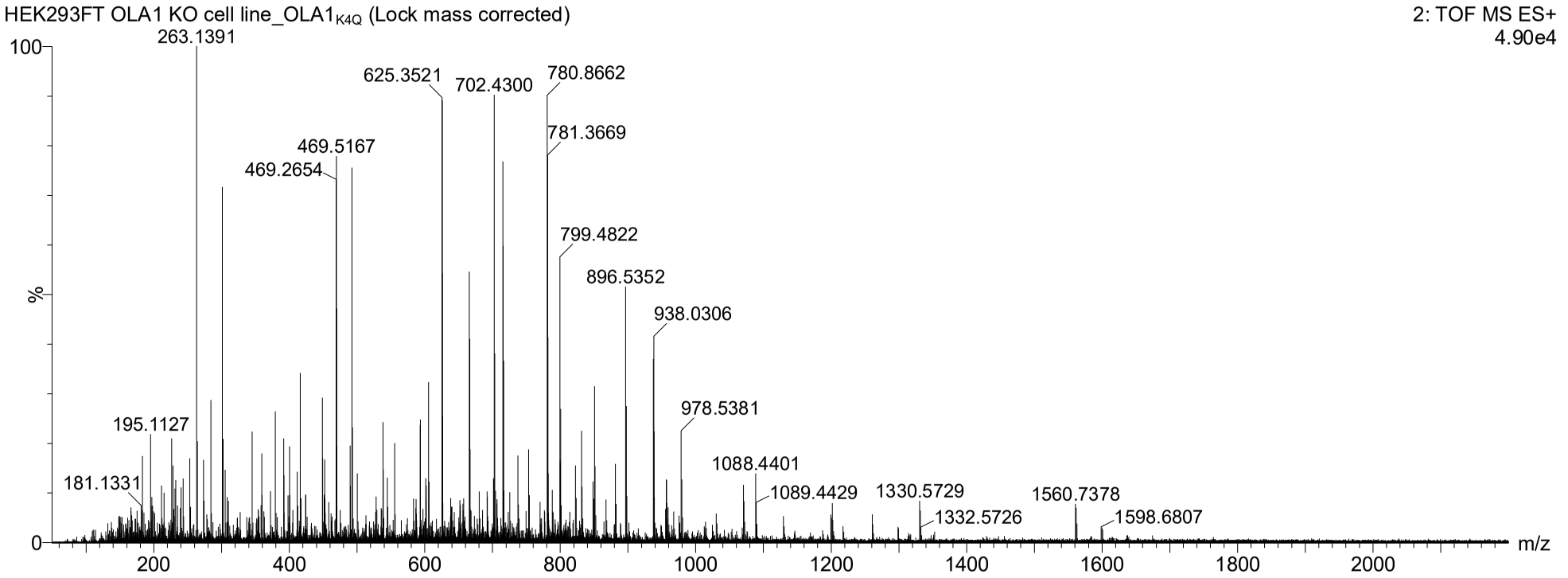

HEK293FT OLA1 KO cell line_OLA1 (Precursor, Lock mass corrected)
1: TOF MS ES+
7.29e4
100
%
0
476.5335
381.4265
381.2256
111.0938
179.0940
476.7842
477.0323
635.0438
677.9863
678.3211
678.6566
192.1245
202.1091
495.5911
878.5056
952.0622
1255.7050
1576.8918
m/z
200
400
600
800
1000
1200
1400
1600
1800
2000


HEK 293FT OLA1 KO cell line_OLA1 (TOF transformed)
2: TOF MS ES+
8.88e4
100
%
0
1902.1061
1541.7573
850.5090
1103.6478
492.2931
451.2313
702.4312
765.4233
895.5350
1769.8665
1200.6953
1940.0654
1483.7524
232.1399
379.2092
605.3766
1205.5973
1978.0093
1978.9889
m/z
200
400
600
800
1000
1200
1400
1600
1800
2000
185
HEK293FT OLA1 KO cell line_OLA1 (Lock mass corrected)
2: TOF MS ES+
2.44e4
100
%
0
677.9850
677.6530
771.3845
771.8865
492.2936
476.2826
451.2316
635.0434
850.5137
1016.4763
368.5540
552.3307
345.2259
1016.9796
305.1620
1092.5529
161.0859
1255.8042
1093.5563
1371.1639
1654.0951
1776.9749
1923.7150
2154.9321
m/z
200
400
600
800
1000
1200
1400
1600
1800
2000


HEK293FT OLA1 KO cell line_OLA1K4Q (Precursor, TOF transformed)
1: TOF MS ES+
2.13e6
100
%
0
1874.0400
1560.7181
1429.6298
1330.5619
1608.7018
1911.9971
111.0936
263.1390
452.8209
702.4301
831.3122
1144.6893
1201.5190
m/z
200
400
600
800
1000
1200
1400
1600
1800
2000


HEK293FT OLA1 KO cell line_OLA1K4Q (Precursor, Lock mass corrected)
1: TOF MS ES+
5.98e5
100
%
0
469.5174
469.2661
520.9119
521.2460
521.5803
780.8661
625.6887
715.3198
781.8682
782.3705
938.0302
1088.4398
1330.5555
1560.7229
111.0940
263.1392
264.1428
m/z
200
400
600
800
1000
1200
1400
1600
1800
2000


HEK293FT OLA1 KO cell line_OLA1K4Q (TOF transformed)
2: TOF MS ES+
3.02e5
100
%
0
1874.0454
1560.7220
1429.6301
1330.5619
702.4287
896.5314
799.4798
263.1390
492.2920
1075.5790
1776.9905
226.1189
416.1559
605.3764
1575.6732
1912.0050
1949.9520
m/z
200
400
600
800
1000
1200
1400
1600
1800
2000


HEK293FT OLA1 KO cell line_OLA1K4Q (Lock mass corrected)
2: TOF MS ES+
4.90e4
100
%
0
263.1391
625.3521
702.4300
780.8662
781.3669
469.5167
469.2654
799.4822
896.5352
938.0306
978.5381
195.1127
1088.4401
181.1331
1089.4429
1330.5729
1560.7378
1332.5726
1598.6807
m/z
200
400
600
800
1000
1200
1400
1600
1800
2000