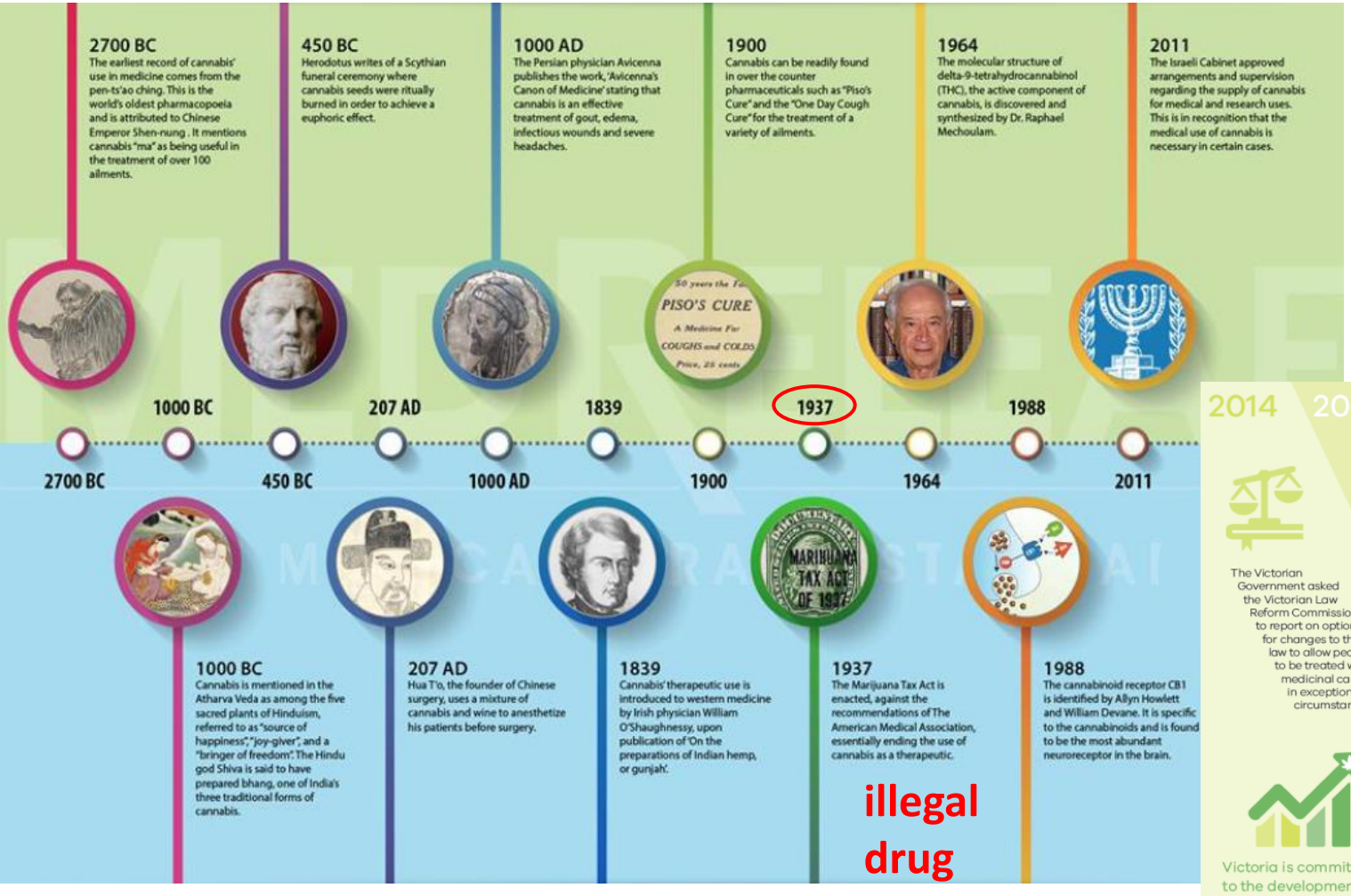
Proteomics tools for medicinal cannabis

Dr Delphine Vincent

13/12/2019

Introduction

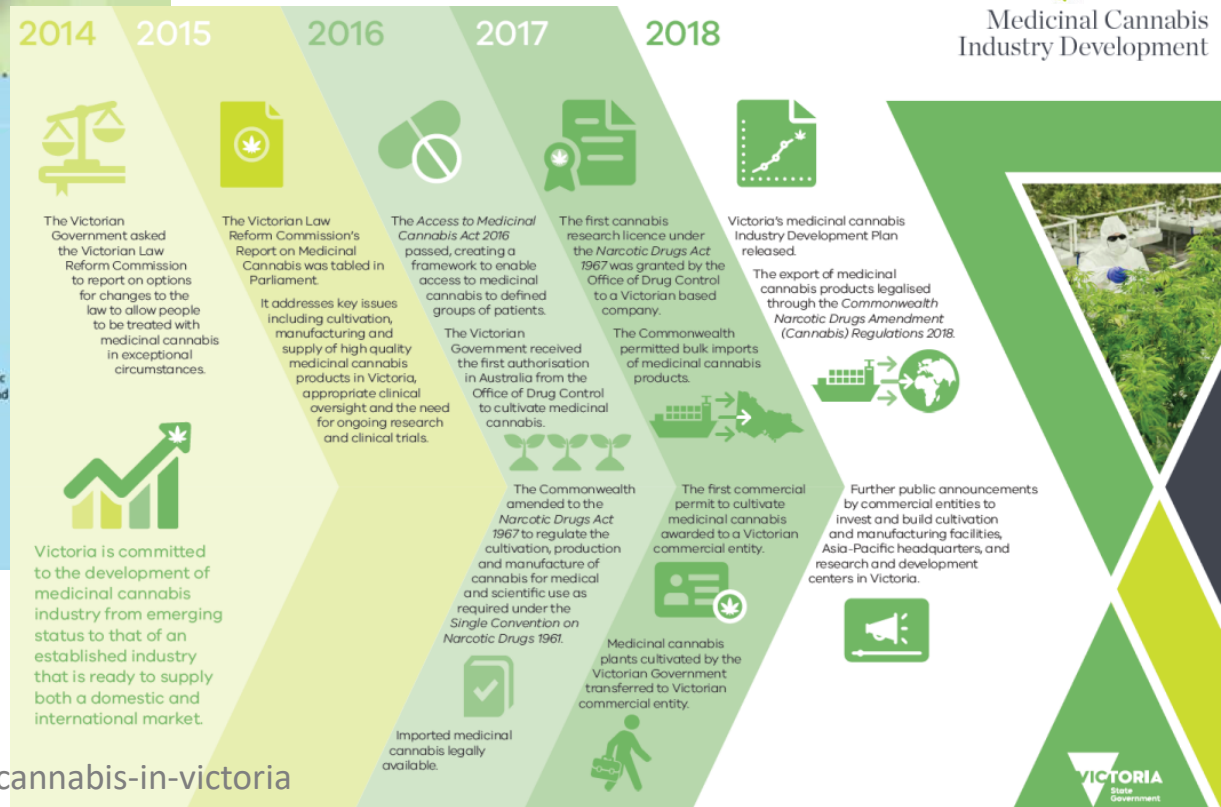
History of cannabis and legislation



Hand et al, 2016

Australia

Dec 2015: Andrews Government (VIC) introduces the Access to Medicinal Cannabis Bill 2015 (cultivation, manufacture and distribution).
First Australian government to legalise medicinal cannabis.
Priority: Children with severe epilepsy.
(SOURCE: www.parliament.vic.gov.au)



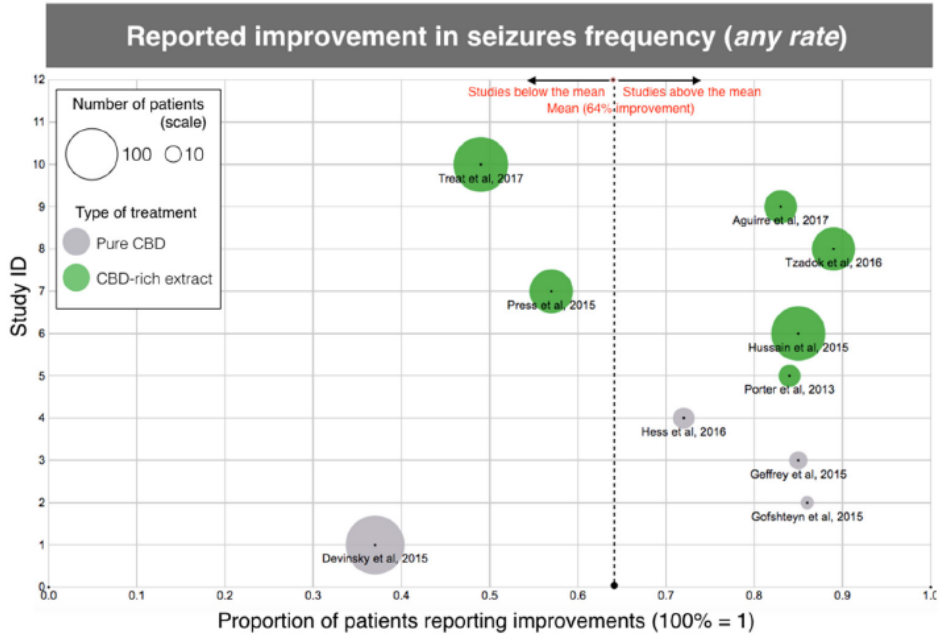
https://djpr.vic.gov.au/medicinal-cannabis-in-victoria

Medicinal cannabis and high CBD content for epilepsy

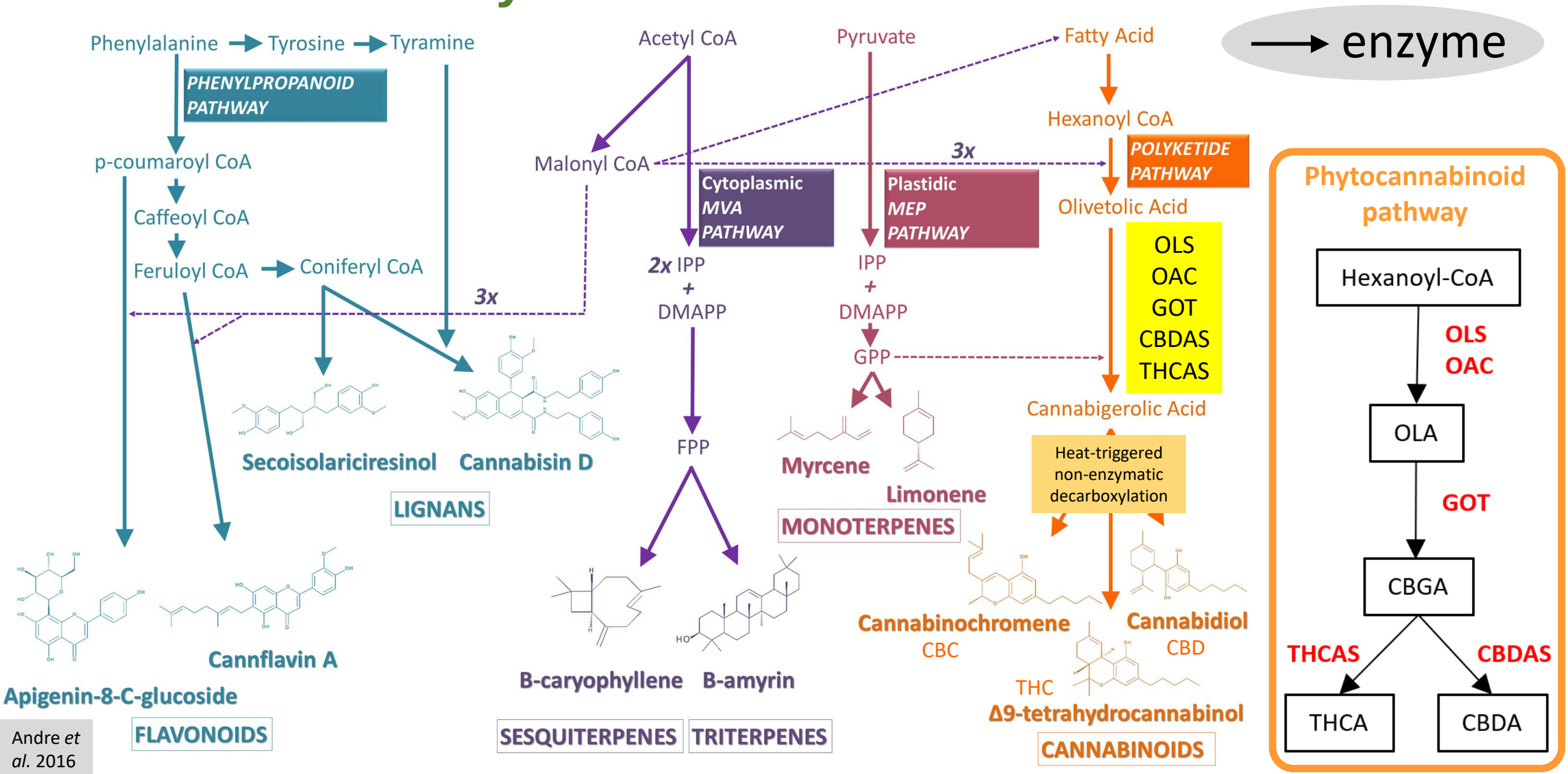
MEDICAL CONDITION	POTENTIAL MEDICAL BENEFITS
Cancer patient undergoing chemotherapy treatment.	Aids in pain management and enhances appetite.
Glaucoma caused by poor blood supply to the optic nerve fibres.	Decreases pressure inside the eye.
Epileptic seizures.	Controls seizures by binding to the brain cells responsible for controlling excitability and regulating relaxation.
Alzheimer diseases.	Slows the formation of amyloid plaques by blocking the enzyme in the brain that makes them.
Painful symptoms of multiple sclerosis.	Binds to receptors in the nerves and muscles to relieve pain.
Treatment for Hepatitis C infection (negative side effects).	Helps lessens treatment side effects such as nausea, muscle aches, loss of appetite, and depression.
Inflammatory bowel diseases like Crohn's disease.	Interacts with cells in the body that play an important role in gut function and immune responses.
Parkinson's disease.	Significantly reduces pain and tremors and improves sleep for Parkinson's disease patients.
Concussion or other traumatic injury.	Lessens the bruising of the brain and helps with healing mechanisms after a traumatic injury.

EverBlu report 2017

Treatment with cannabidiol (CBD)-based products significantly reduces epileptic seizure frequency.
(Pamplona et al, 2018)



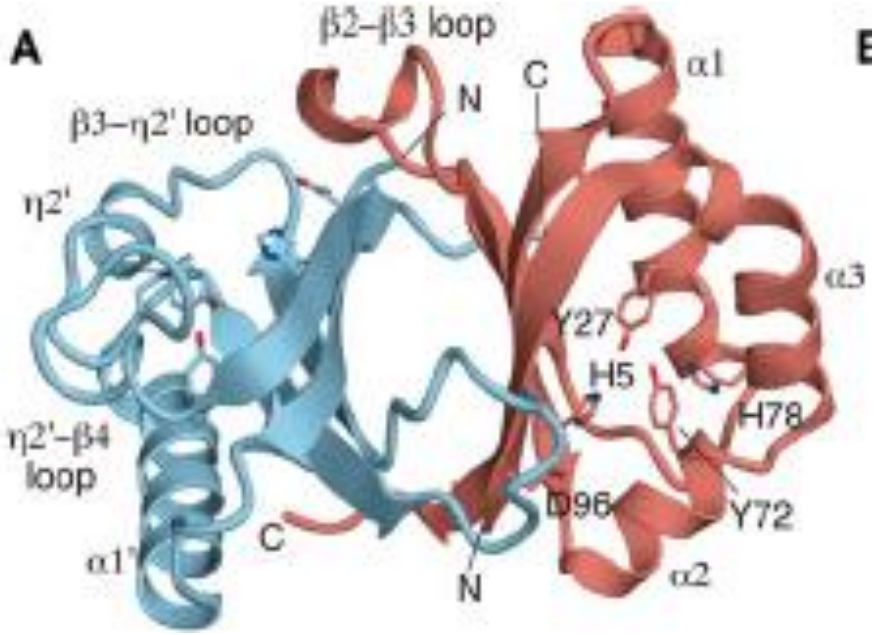
Cannabis secondary metabolism



Andre *et al.* 2016

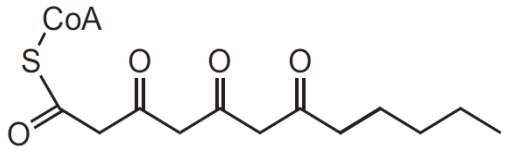
Focus on olivetolic acid cyclase (OAC UniProtKB I6WU39)

Mature sequence 1-101 AA (12kDa), homodimer.
Cytoplasmic (not secreted)
Ligands: 3 x Magnesium

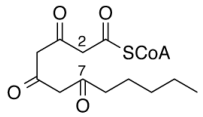


Yang et al, 2016

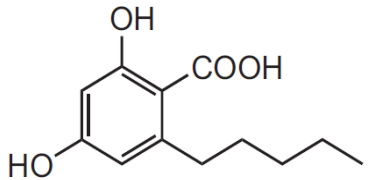
3,5,7-trioxododecanoyl-CoA



Olivetolic Acid Cyclase (OAC)

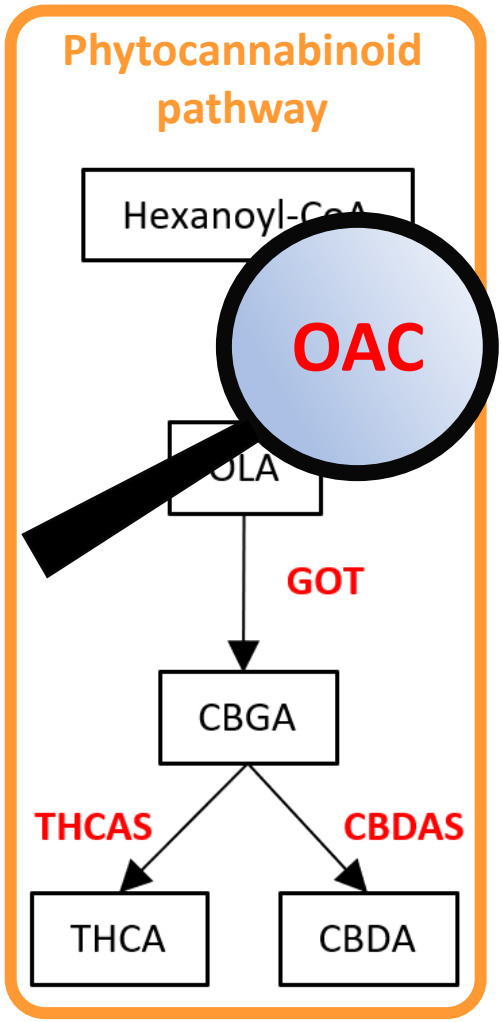


cyclization



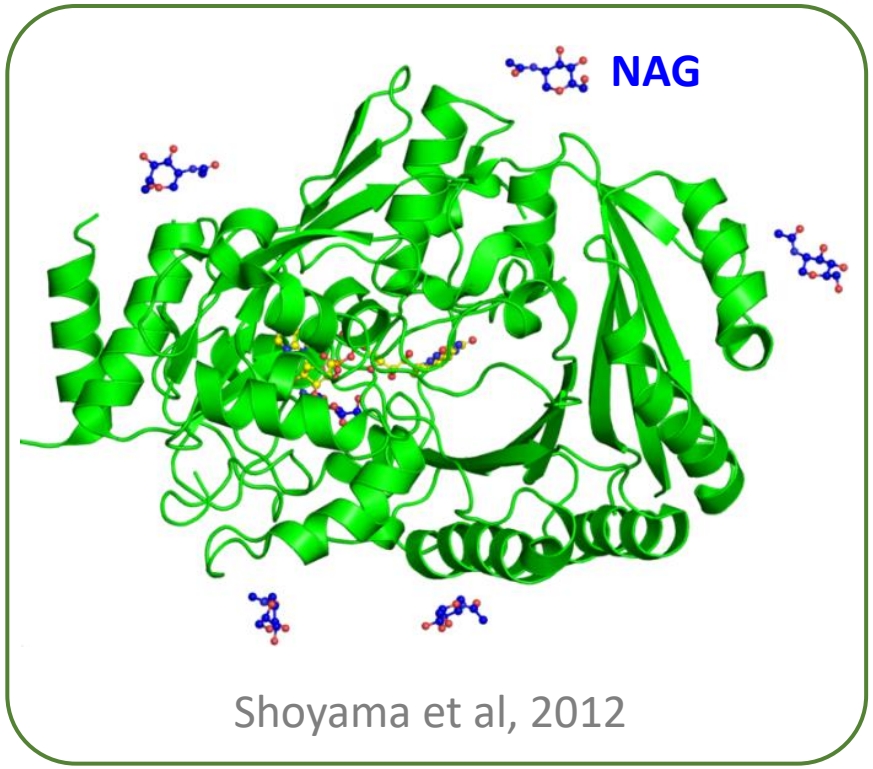
Olivetolic acid

Gagne et al, 2012
Yang et al, 2016

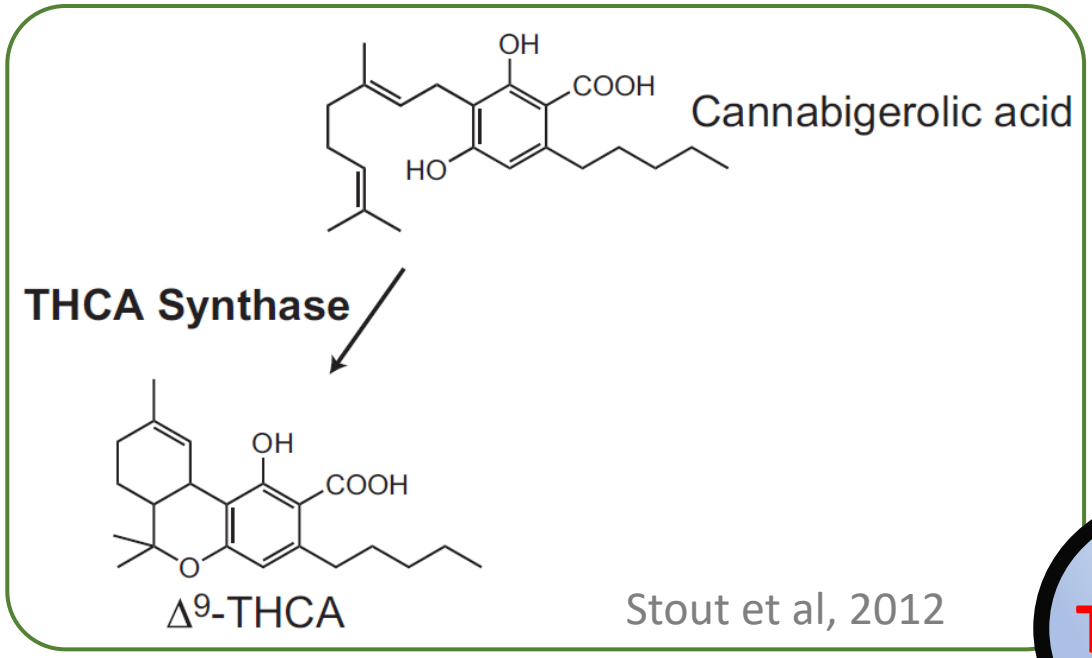


Focus on THCA synthase (THCAS UniProtKB Q8GTB6)

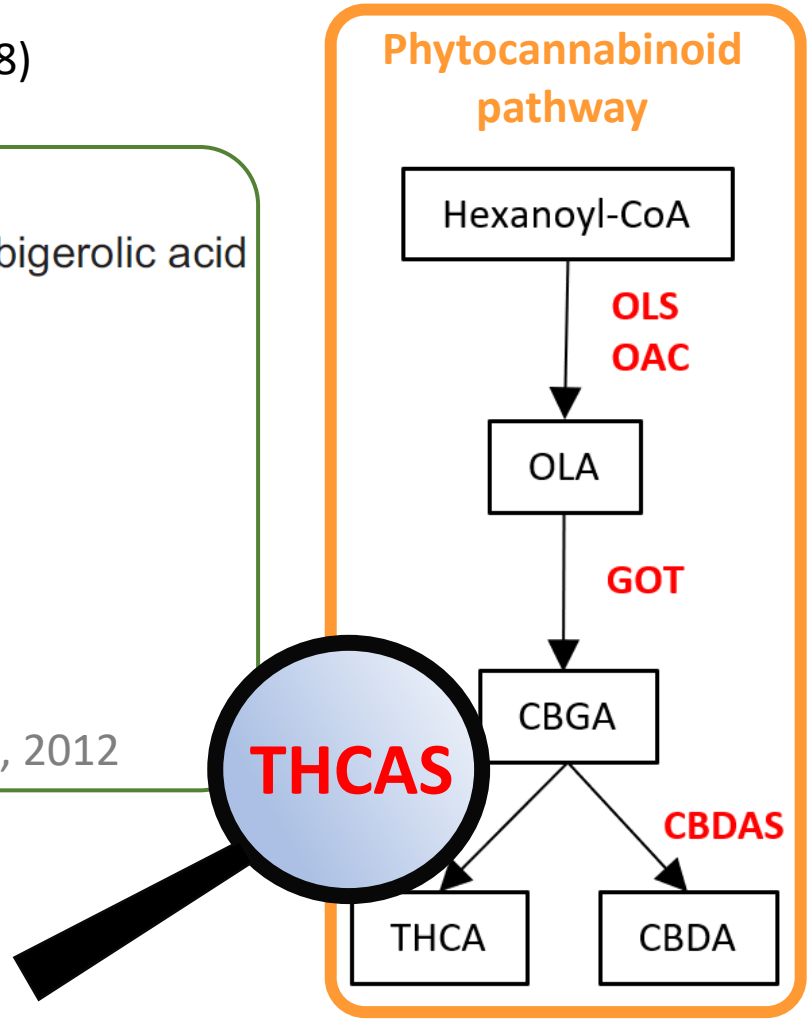
Extracellular (secreted): Signal Peptide (SP) 1-28 AA; mature sequence 29-545 AA (62kDa)
Flavoprotein Cofactor Flavin-adenosine dinucleotide (FAD), disulfide bond (C37-C99)
Ligands: 7(+1) x N-acetyl-D-glucosamine (NAG or GlcNAc, mass=221) (attached to asparagine N residues)
Position of NAG on AA sequence: N2, N65, N89, N168, N297, N305, N329, N467, N499
Some glycosylated N are not needed for enzymatic activity (N89, N499) (Zirpel et al, 2018)



Shoyama et al, 2012

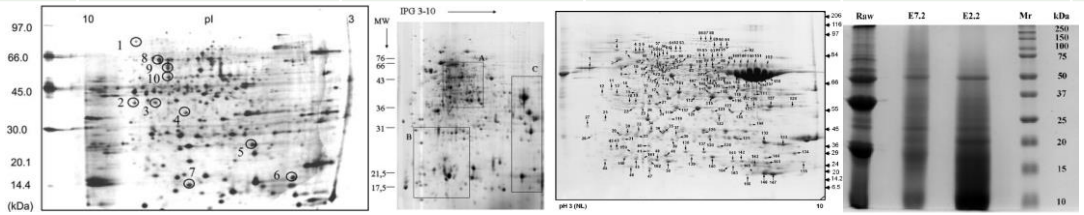


Stout et al, 2012



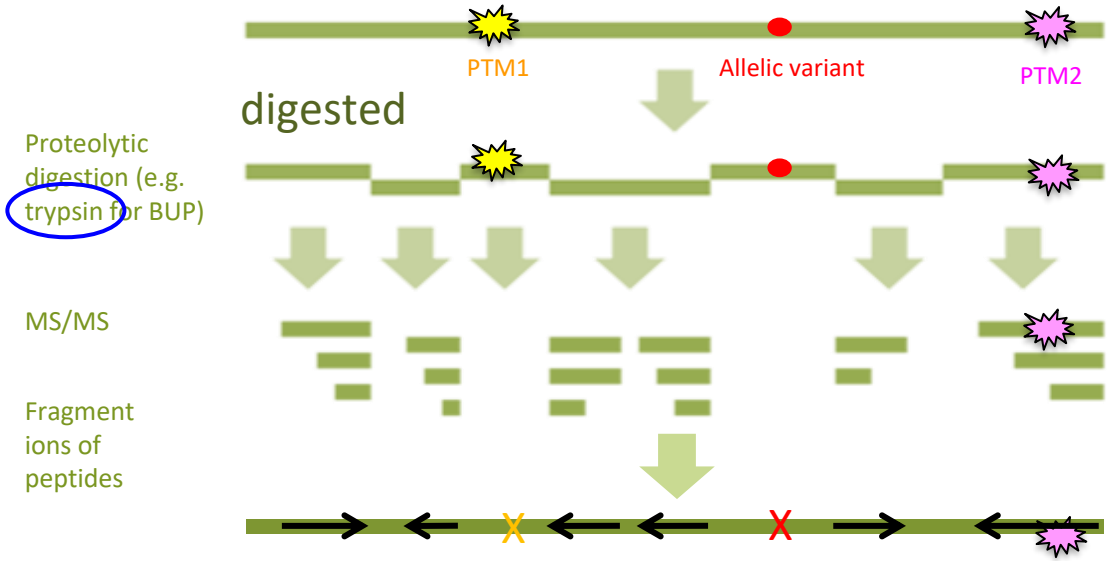
Cannabis proteomics up to 2018

LITERATURE	Raharjo et al, 2004	Bona et al, 2007	Park et al, 2012	Aiello et al, 2016	Behr et al, 2018	SUMMARY
Tissues	leaves, flowers, trichomes	roots	hempseed flour	hempseed flour	hypocotyls	1/5 study on flowers/trichomes (where phytocannabinoids accumulate)
Powder homogenisation		Tris/sucrose/ thiourea/DTT	Sucrose/Na-phosphate	Tris/NaCl/ CHAPS		3/5, different buffers
Protein precipitation	10%TCA/2ME/ acetone		10%TCA/H2O	20%TCA/ H2O	20%/DTT/ acetone	4/5, always TCA, 2/5 with acetone
Phase partition		Phenol/NH4 acetate				1/5
Protein resuspension	Urea/CHAPS/ DTT	Urea/CHAPS/ Triton/DTT	Urea/thiourea /CHAPS/Tris/ DTE	Tris/NaCl/ CHAPS	Urea/thiourea /CHAPS/Tris	4/5 urea-based solutions, different composition
Protein separation	2-DE	2-DE	2-DE	1-DE	1-DE, 2-D DIGE	5/5 gel-based separation
Protein digestion	trypsin	trypsin	trypsin	trypsin	trypsin	5/5 trypsin
Protein identification	PMF	nLC-MS/MS	nLC-MS/MS	nLC-MS/MS	PMF	2/5 PMF, 3/5 nLC-MS/MS



Materials - Methods

Complementary proteomics strategies



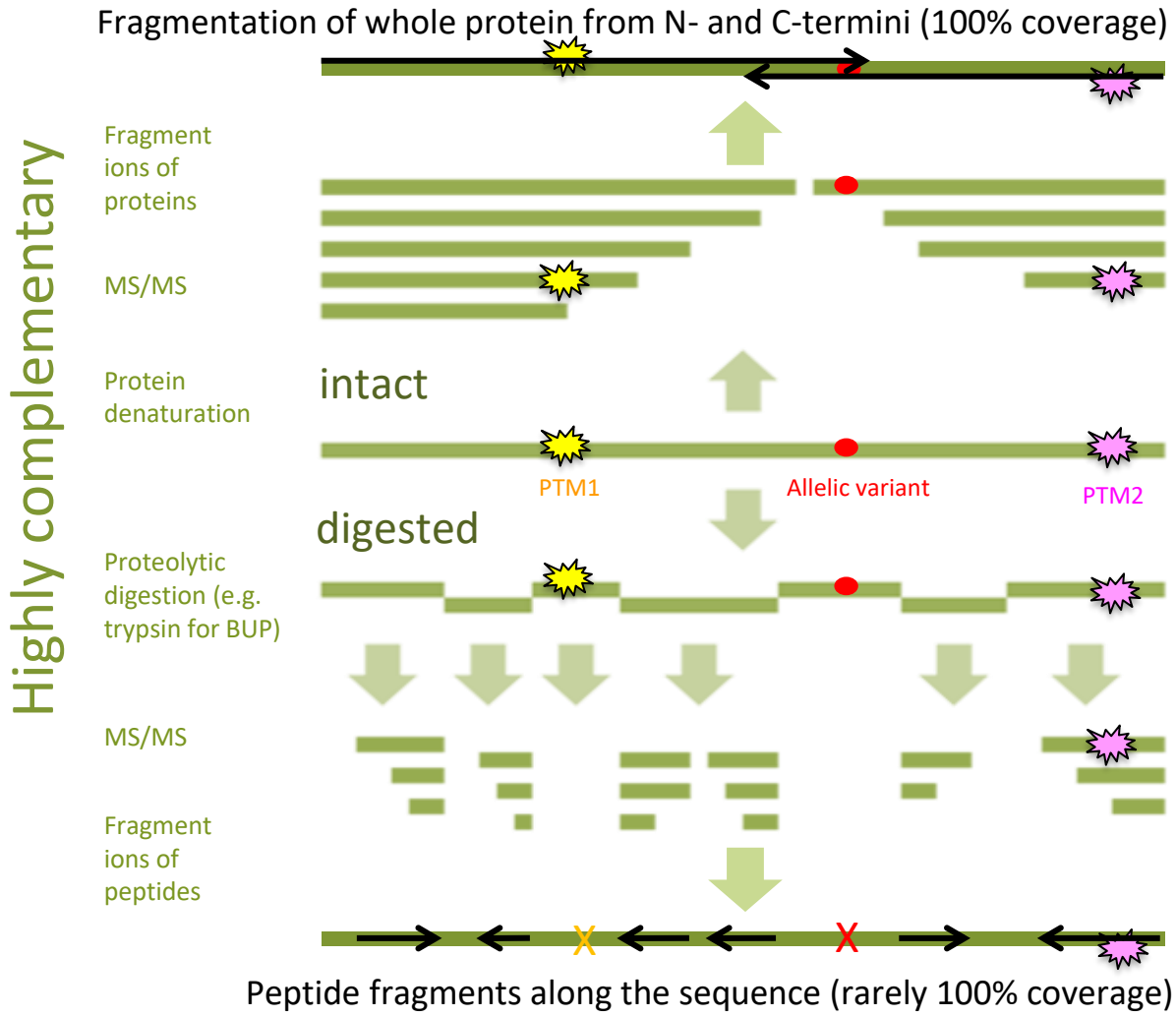
Peptide fragments along the sequence (rarely 100% coverage)

Bottom-up proteomic (BUP)

(adapted from Wikipedia)

Complementary proteomics strategies

Top-down proteomics (TDP)



Bottom-up proteomic (BUP)

(adapted from Wikipedia)

Complementary proteomics strategies

Highly complementary

Top-down proteomics (TDP)

Fragmentation of whole protein from N- and C-termini (100% coverage)

Fragment ions of proteins

MS/MS

Protein denaturation

intact

PTM1 Allelic variant PTM2

digested

Proteolytic digestion (e.g. trypsin for BUP)

MS/MS

Fragment ions of peptides

low specificity/high efficiency protease = small peptides (0.5-3kDa)

X X

Peptide fragments along the sequence (rarely 100% coverage)

Bottom-up proteomic (BUP)

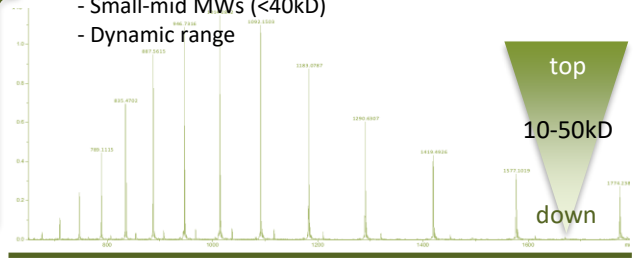
(adapted from Wikipedia)

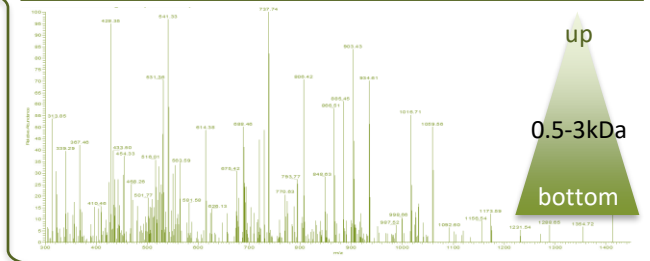
(Vincent et al; 2016, 2018, 2019; Raynes et al. 2018)

Top-down proteomics (proteins)

+ Protein quantitation
+ Allelic variation
+ PTM analysis
+ Protein processing
- Complexity of the sample
- Relatively new approach
- Small-mid MWs (<40kD)
- Dynamic range

proteoforms





+ Most mature technique
+ Widely used, lots of tools
+ All MWs
+ Complex samples
- Dynamic range
- PTM analysis
- Allelic variation
- Protein processing
- Protein quantitation

Bottom-up proteomics (peptides)

(Vincent et al; 2009; 2011; 2012; 2015, 2019)

MDP: bringing the best of both BUP and TDP worlds

Top-down proteomics (TDP)

Fragmentation of whole protein from N- and C-termini (100% coverage)

Highly complementary

Fragment ions of proteins

MS/MS

Protein denaturation

intact

PTM1 Allelic variant PTM2

digested

Proteolytic digestion (e.g. trypsin for BUP, **GluC for MDP**)

high specificity/low efficiency protease = large peptides (3-10kDa)

MS/MS

Fragment ions of peptides

Peptide fragments along the sequence (closer to 100% coverage)

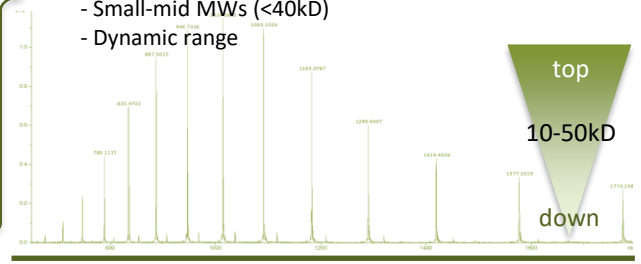
Bottom-up proteomic (BUP)

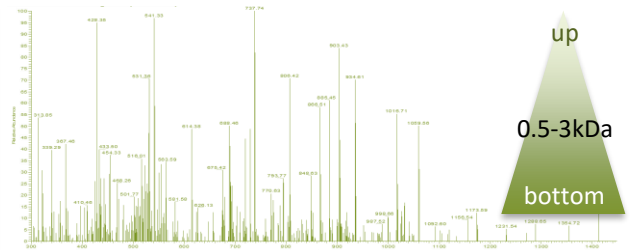
Middle-down proteomics (MDP)

(adapted from Wikipedia)

Top-down proteomics (proteins)

+ Protein quantitation
+ Allelic variation
+ PTM analysis
+ Protein processing
- Complexity of the sample
- Relatively new approach
- Small-mid MWs (<40kD)
- Dynamic range

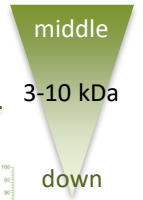




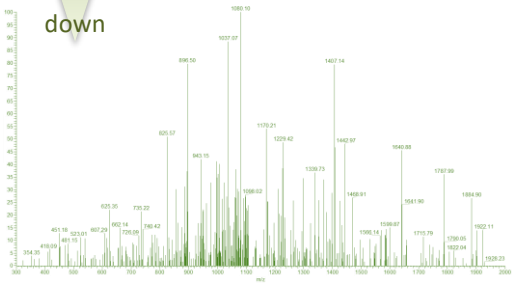
Bottom-up proteomics (peptides)

+ Most mature technique
+ Widely used, lots of tools
+ All MWs
+ Complex samples
- Dynamic range
- PTM analysis
- Allelic variation
- Protein processing
- Protein quantitation

Middle-down proteomics (MDP) (large peptides)



+ Protein quantitation
+ Small-large MWs
+ BUP tools
+ complex samples
+ PTM analysis
- Allelic variation
- Dynamic range
- Protein processing
- Very recent approach

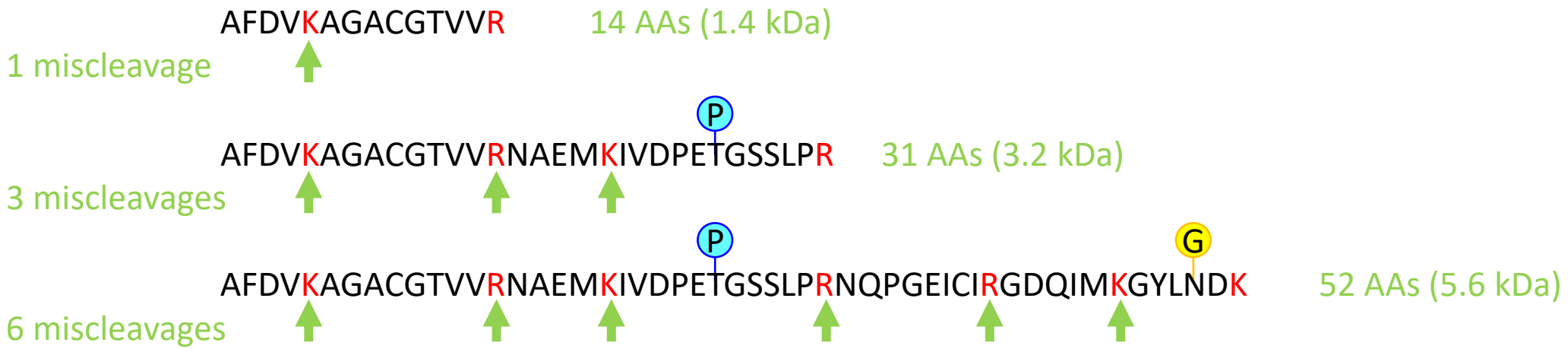


Proteases don’t always get it right! Missed cleavage…

Miscleavages have been reported as a result of the protease skipping a seemingly cleavable residue.

Example: tryptic digest of 4-coumarate:CoA ligase (4CL) (trypsin targets R and K residues)

…TEAGPVLTMSLAFAEKAFDVKAGACGTVVRNAEMKIVDPETGSSLPRNQPGEICIRGDQIMKGYLNDKESTKNTIDKEGWLHTGDIGFVDDDD…

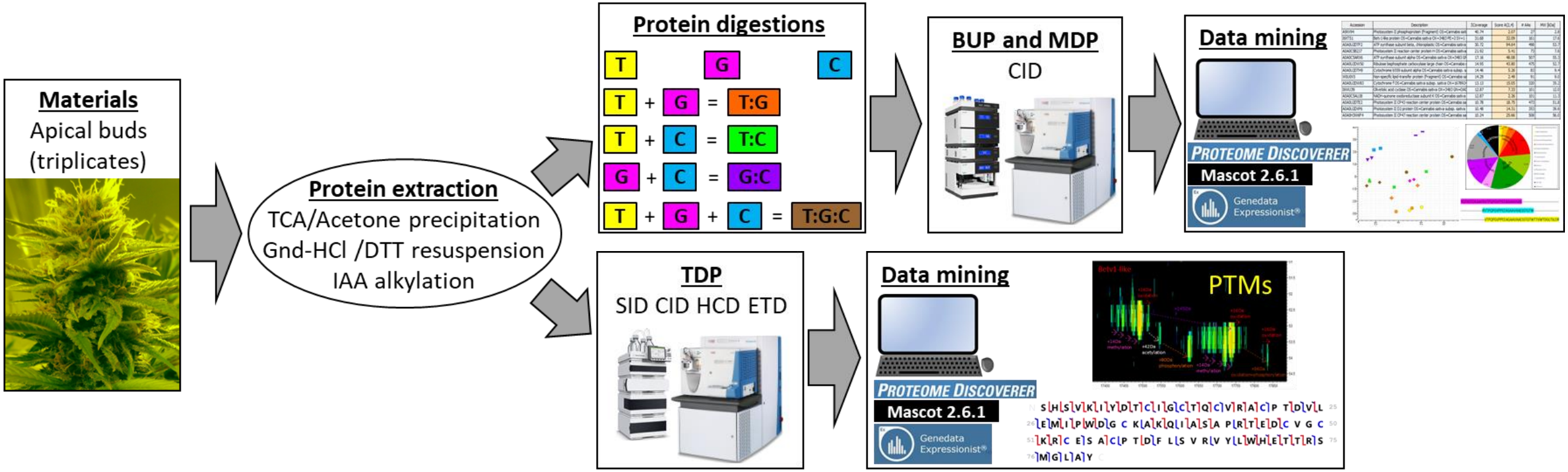


- Miscleavages produce longer peptides.
- Indispensable to discover PTMs.
- Very advantageous for MDP, provided your mass analyser can see them!
- Miscleavages + specific proteases that target rare AAs produce even longer peptides.
- However, a typical BUP study searches for max 2 missed cleavages. Discussed here…

Proof-of-concept studies

- *Overall experimental design*

1st experiment: optimisation of protein extraction from mature buds for BUP (trypsin).
2nd experiment: optimisation of protein digestion for protein identification using BUP and MDP.
3rd experiment: optimisation of intact protein analysis and sequencing using TDP.



Results

Optimisation of protein extraction

Materials
Apical buds
(triplicates)

Protein extraction
TCA/Acetone precipitation
Gnd-HCl /DTT resuspension
IAA alkylation

Protein digestions

T G C

T + G = T:G

T + C = T:C

G + C = G:C

T + G + C = T:G:C

BUP and MDP
CID

TDP
SID CID HCD ETD

Data mining

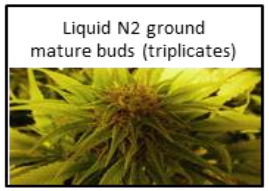
PROTEOME DISCOVERER
Mascot 2.6.1
Genedata Expressionist®

Data mining

PROTEOME DISCOVERER
Mascot 2.6.1
Genedata Expressionist®

1. Bottom-up proteomics (BUP)

Experimental design



Extraction Solutions	
TCA/DTT/A solution: 10% TCA 10mM DTT acetone	TCA/DTT/EtOH solution: 10% TCA 10mM DTT EtOH
Urea buffer: 6M urea, 10mM DTT, 10mM Tris-HCl pH 8.0, 75mM NaCl, 0.05% SDS.	Gnd-HCl buffer: 6M Gdn-HCl, 10mM DTT, 5.37 mM sodium citrate tribasic 2H2O, 0.1 M Bis-Tris.

Extraction number	TCA/DTT/A precipitation	TCA/DTT/E precipitation	Urea buffer resuspension	Gnd-HCl buffer resuspension	Extraction Name
Extraction 1	no	no	yes	no	AB1
Extraction 2	no	no	no	yes	AB2
Extraction 3	yes	no	yes	no	AB3
Extraction 4	yes	no	no	Yes	AB4
Extraction 5	no	yes	yes	no	AB5
Extraction 6	no	yes	No	yes	AB6

18 apical bud extracts

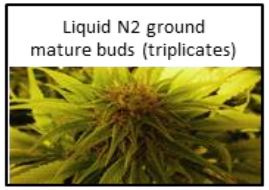
Extraction number	Urea buffer resuspension	Gnd-HCl buffer resuspension	Extraction Name
Extraction 1	yes	no	T1
Extraction 2	no	yes	T2

6 trichome extracts

(Vincent et al., 2019 Molecules + patent)

1. Bottom-up proteomics (BUP)

Experimental design



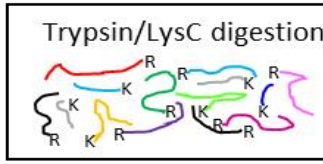
Extraction Solutions	
TCA/DTT/A solution: 10% TCA 10mM DTT acetone	TCA/DTT/EtOH solution: 10% TCA 10mM DTT EtOH
Urea buffer: 6M urea, 10mM DTT, 10mM Tris-HCl pH 8.0, 75mM NaCl, 0.05% SDS.	Gnd-HCl buffer: 6M Gdn-HCl, 10mM DTT, 5.37 mM sodium citrate tribasic 2H2O, 0.1 M Bis-Tris.

Extraction number	TCA/DTT/A precipitation	TCA/DTT/E precipitation	Urea buffer resuspension	Gnd-HCl buffer resuspension	Extraction Name
Extraction 1	no	no	yes	no	AB1
Extraction 2	no	no	no	yes	AB2
Extraction 3	yes	no	yes	no	AB3
Extraction 4	yes	no	no	Yes	AB4
Extraction 5	no	yes	yes	no	AB5
Extraction 6	no	yes	No	yes	AB6

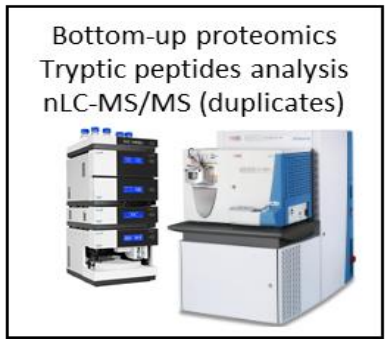
18 apical bud extracts

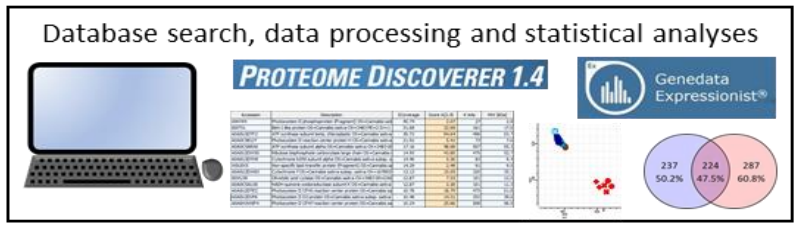
Extraction number	Urea buffer resuspension	Gnd-HCl buffer resuspension	Extraction Name
Extraction 1	yes	no	T1
Extraction 2	no	yes	T2

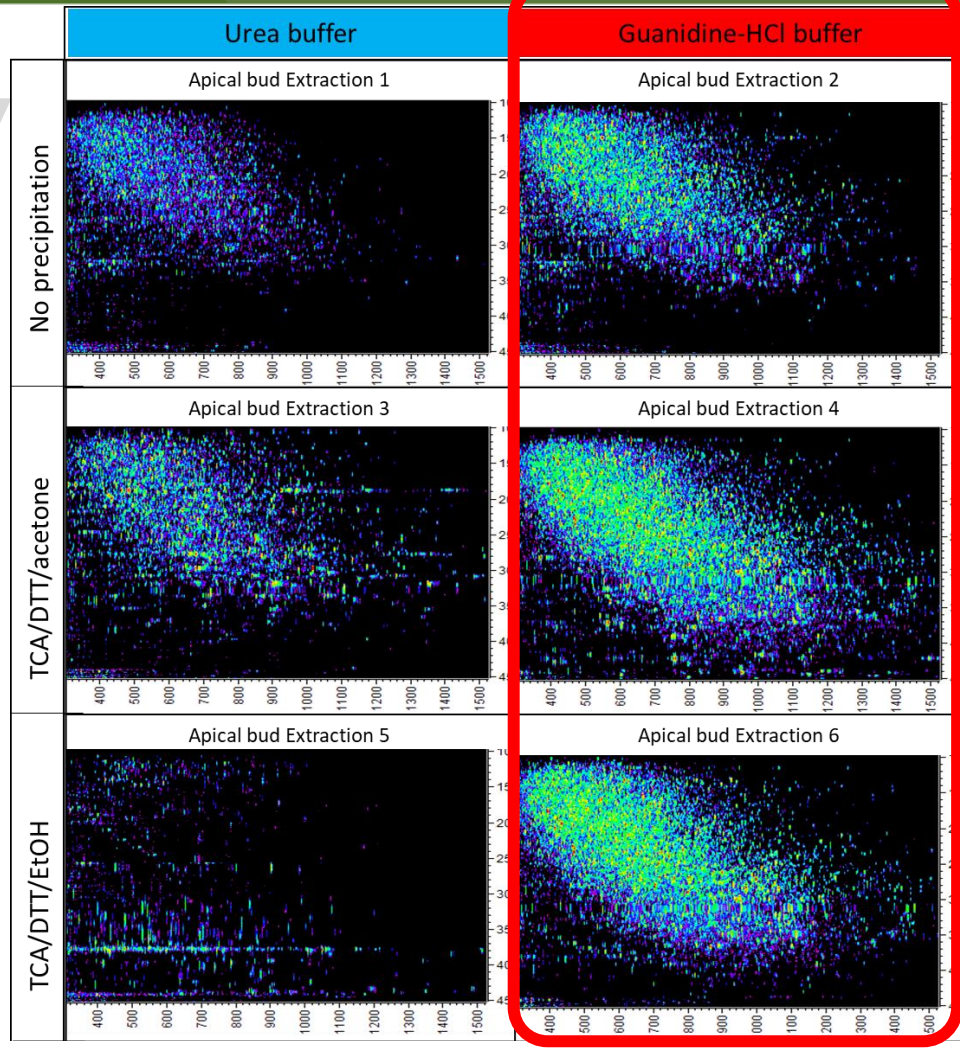
6 trichome extracts



Trypsin (K,R) + LysC (K)
Medium specificity ("ms")
High efficiency ("he")

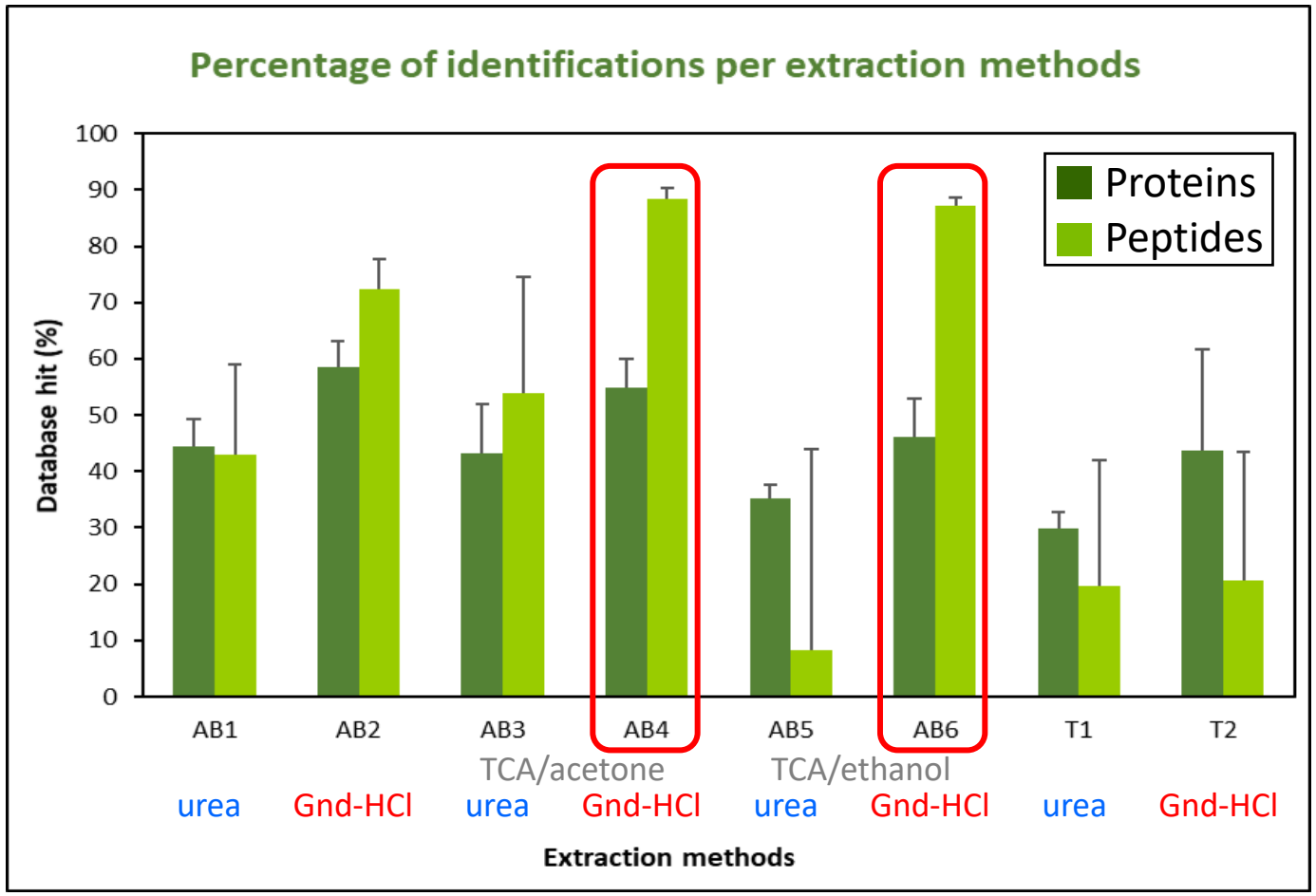






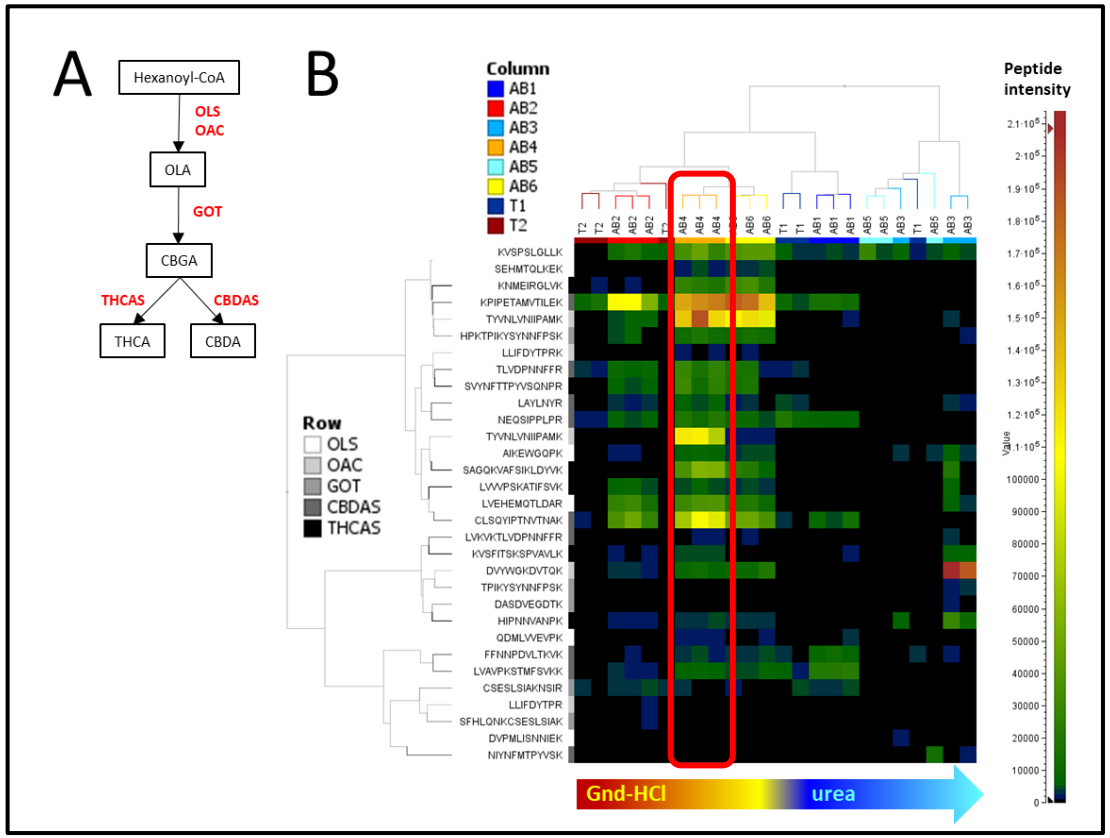
1. Bottom-up proteomics (BUP)

160 proteins identified from 5,675 peptides using BUP.



Methods AB4 and AB6 (TCA/A or TCA/E precipitation, Gnd-HCl resuspension) are the best protein extraction methods for mature buds (more identities, more reproducible).

Methods AB4 and AB6 are the only methods that recover enzymes involved in the phytocannabinoid biosynthesis.



Method chosen: AB4 (TCA/A → Gnd-HCl)

1. Bottom-up proteomics (BUP)

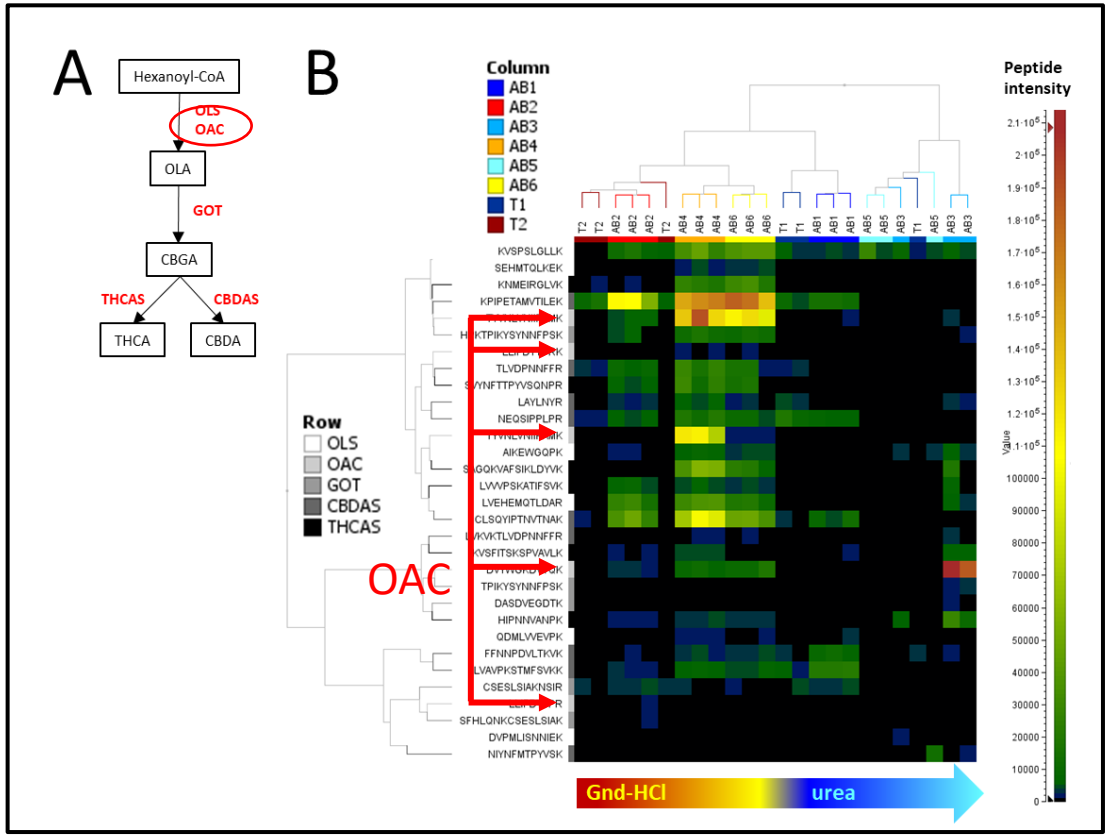
Focus on:

- OAC (olivetolic acid cyclase) (101 AAs, 12kDa)
- THCAS (tetrahydrocannabinolic acid synthase) (516 AAs, 59kDa)

OAC 34% coverage, no PTM (2 miscleavages)

>sp|I6WU39|OLIAC_CANSA Olivetolic acid cyclase OS=Cannabis sativa

MAVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK



1. Bottom-up proteomics (BUP)

Focus on:

- OAC (olivetolic acid cyclase) (101 AAs, 12kDa)
- THCAS (tetrahydrocannabinolic acid synthase) (516 AAs, 59kDa)

OAC 34% coverage, no PTM (2 miscleavages)

>sp|I6WU39|OLIAC_CANSA Olivetolic acid cyclase OS=Cannabis sativa
MAVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK

THCAS 12% coverage, no PTM (2 miscleavages)

>sp|Q8GTB6|29-545|THCAS_CANSA Tetrahydrocannabinolic acid synthase OS=Cannabis sativa
NPRENFLKCFSKHIPNNVANPKLVYTQHDQLYMSILNSTIQNLRFISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHDAEGMSYISQVPFVV

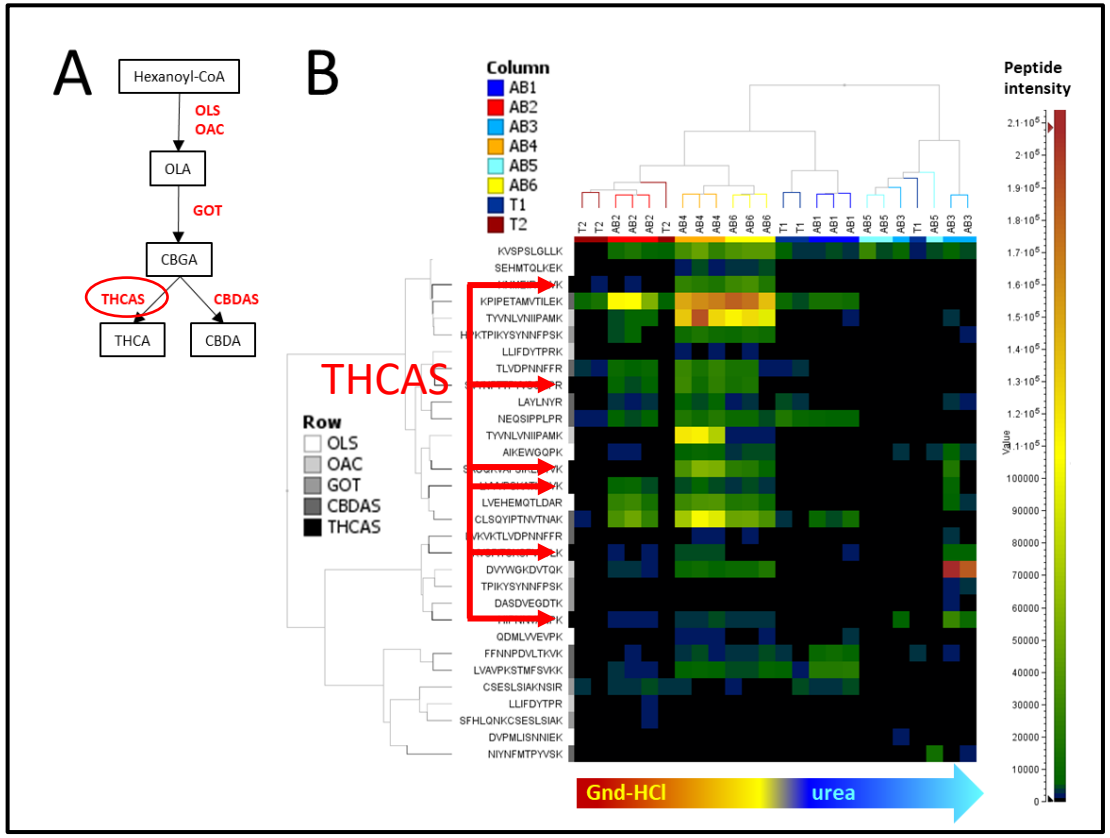
VDLRNMHSIKIDVHSQTAWVEAGATLGEVYYWINEKNENLSFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMG

EDLFWAIRGGGGENFGIIAAWKIKLVAVPSKSTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLVLMTHFITKNITDNHGKNKTTVHGYFSSIFHGGVD

SLVDLMNKSFPELGIKKTDCKEFSWIDTTIFYSGVVNFNTANFKKEILLDRSAGKKTAFSIKLDYVKKPIPETAMVKILEKLYEEDVGAGMYVLYPYGGIME

EISESAIPFPHRAGIMYELWYTASWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNHASPNNYTQARIWGEKYFGKNFNRLVKV

KTKVDPNNFFRNEQSIPPLPPHHH



Optimisation of protein digestion

Materials
Apical buds (triplicates)

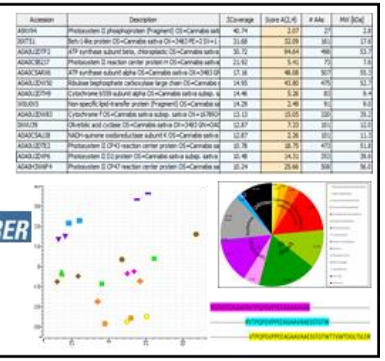
Protein extraction
TCA/Acetone precipitation
Gnd-HCl /DTT resuspension
IAA alkylation

Protein digestions

T	G	C
T + G = T:G		
T + C = T:C		
G + C = G:C		
T + G + C = T:G:C		

BUP and MDP
CID

Data mining



PROTEOME DISCOVERER
Mascot 2.6.1
Genedata Expressionist®

TDP
SID CID HCD ETD

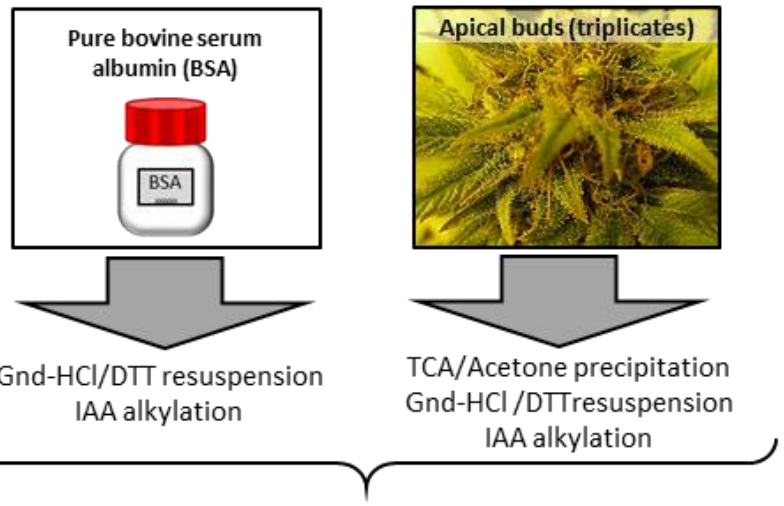
Data mining

PROTEOME DISCOVERER
Mascot 2.6.1
Genedata Expressionist®

PTMs

2. Middle-down proteomics (MDP)

Experimental design



Protein digestion using 3 proteases

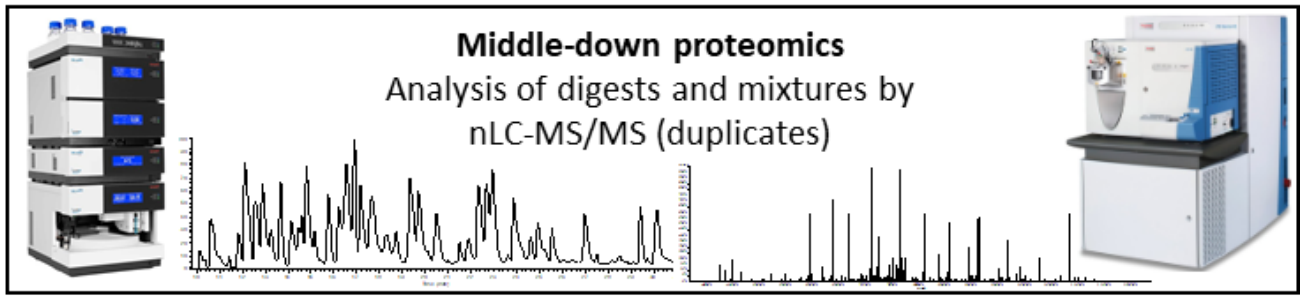
Enzyme code	ms\he	hs\le	ls\me	AA targeted
T	yes	no	no	R, K
G	no	yes	no	E, D
C	no	no	yes	Y, F, W, L
T->G	yes	yes	no	R, K, E, D
T->C	yes	no	yes	R, K, Y, F, W, L
G->C	no	yes	yes	E, D, Y, F, W, L
T->G->C	yes	yes	yes	R, K, E, D, Y, F, W, L

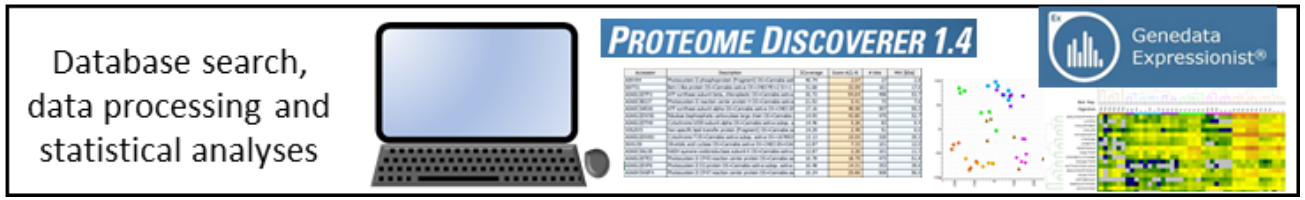
Mixture of digests (1:1 v/v or 1:1:1 v/v/v)

T + G = T:G T + C = T:C G + C = G:C

T + G + C = T:G:C

Peptide digests

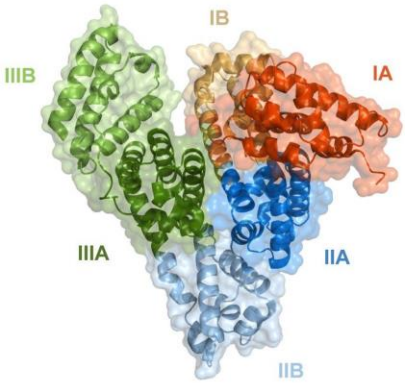




1. Digestion tests on BSA
2. Application to cannabis samples

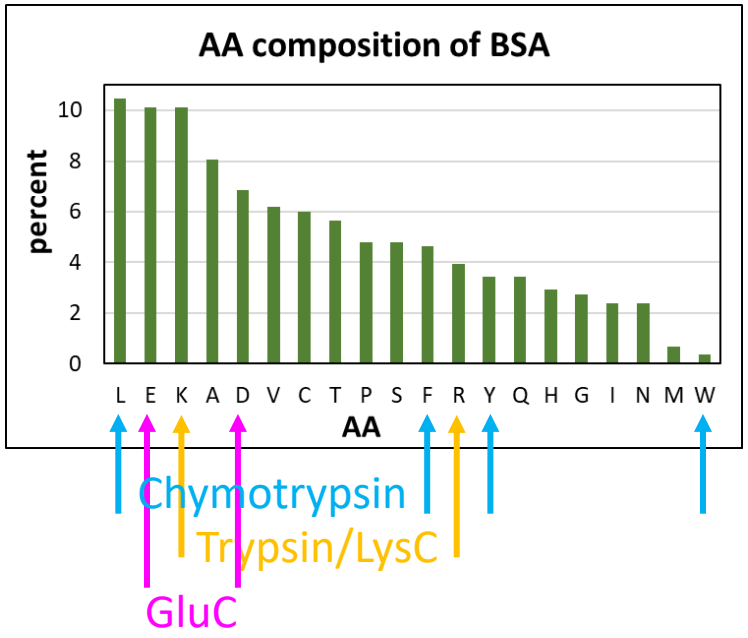
2. Middle-down proteomics (MDP)

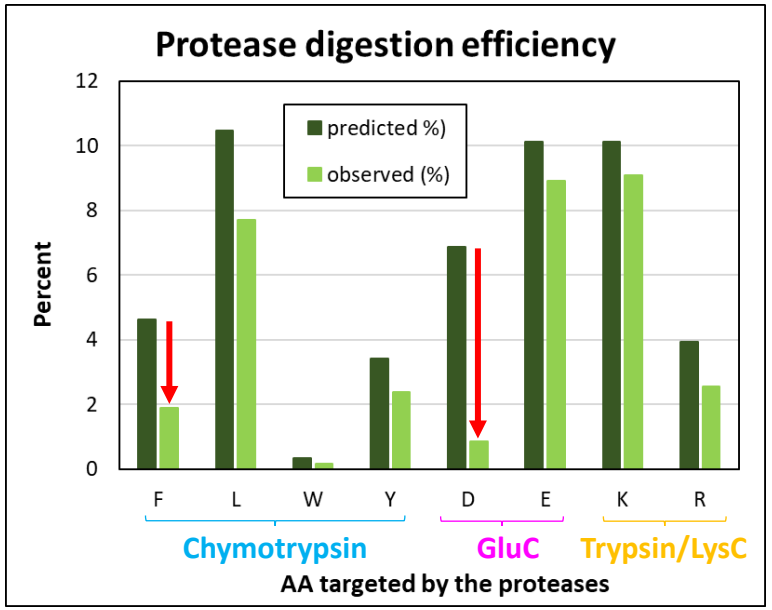
1. Benchmarking using bovine serum albumin (BSA)



BSA structure
(Majorek et al. 2012)

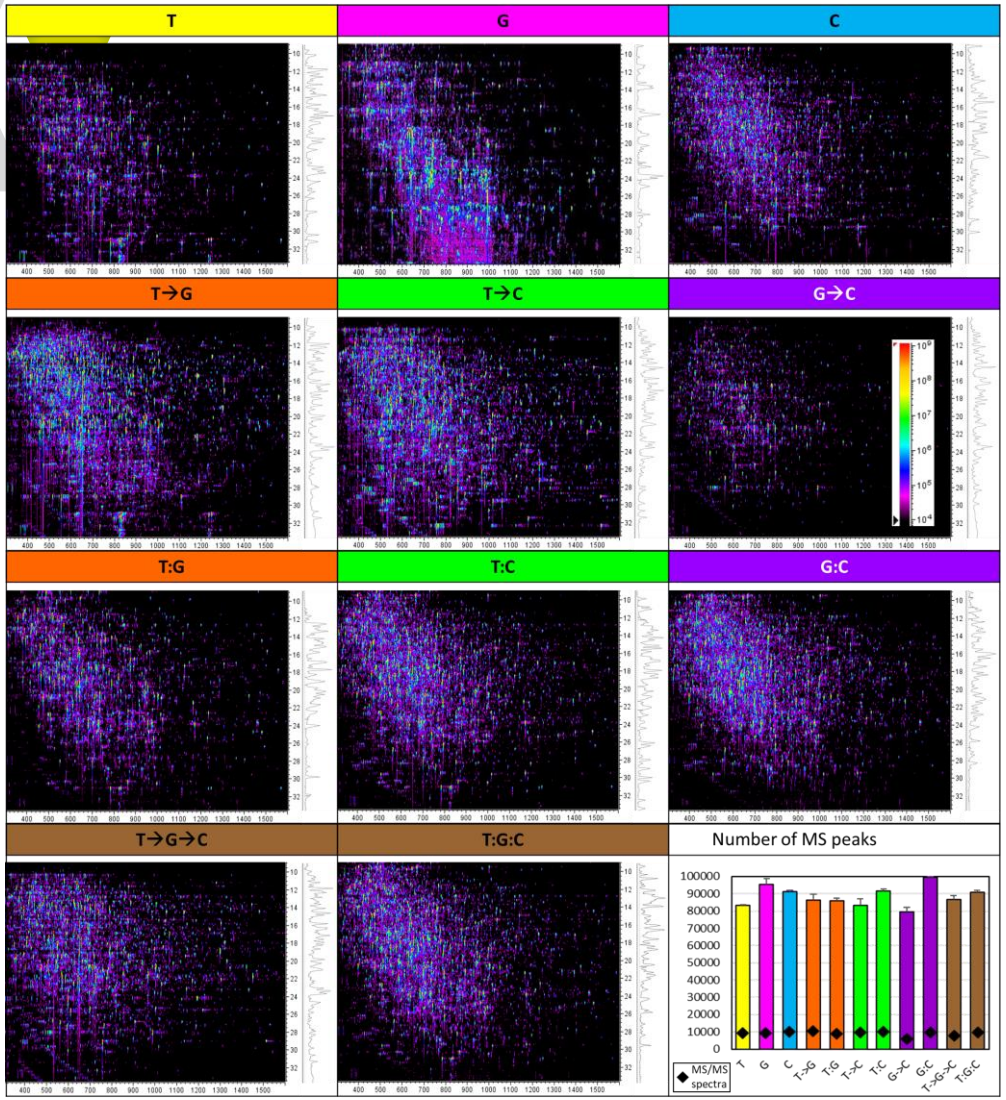
Mature form:
583 AAs, 66kDa





Digestion efficiency varies from one protease to another.
F and D residues often miscleaved.

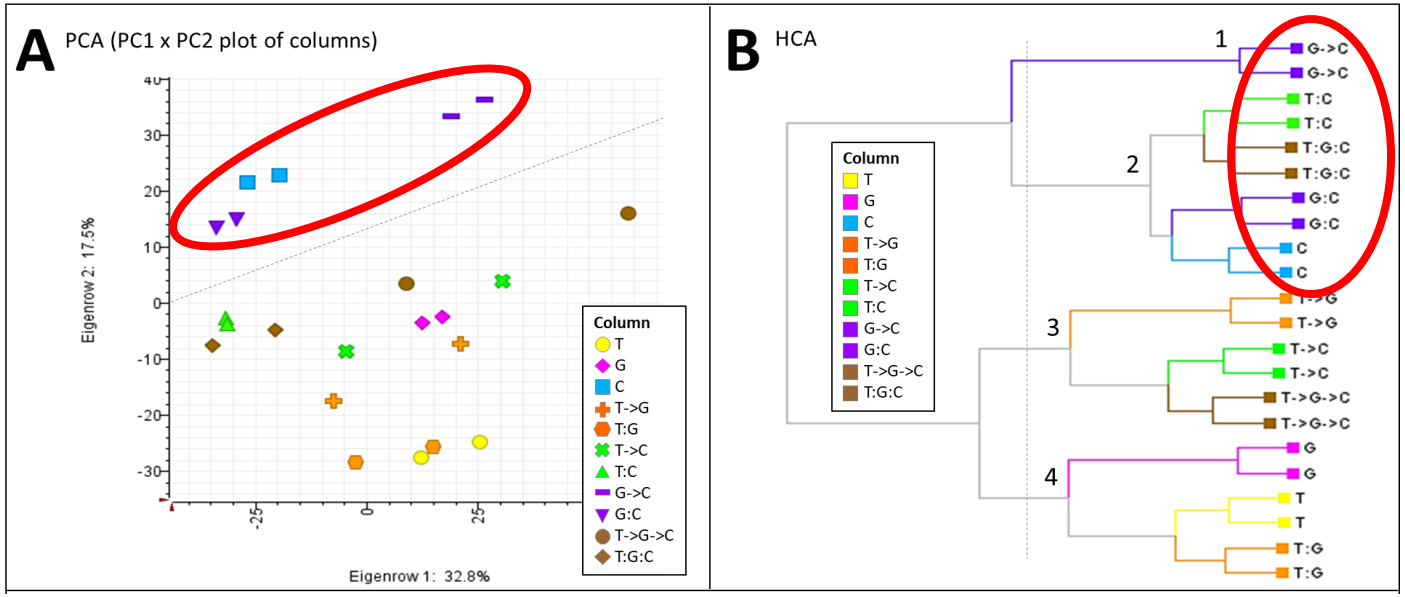
BSA digest LC-MS patterns



BSA peptides elute from 9-39 min (9-39% ACN) along 300-1600 *m/z*.

2. Middle-down proteomics (MDP)

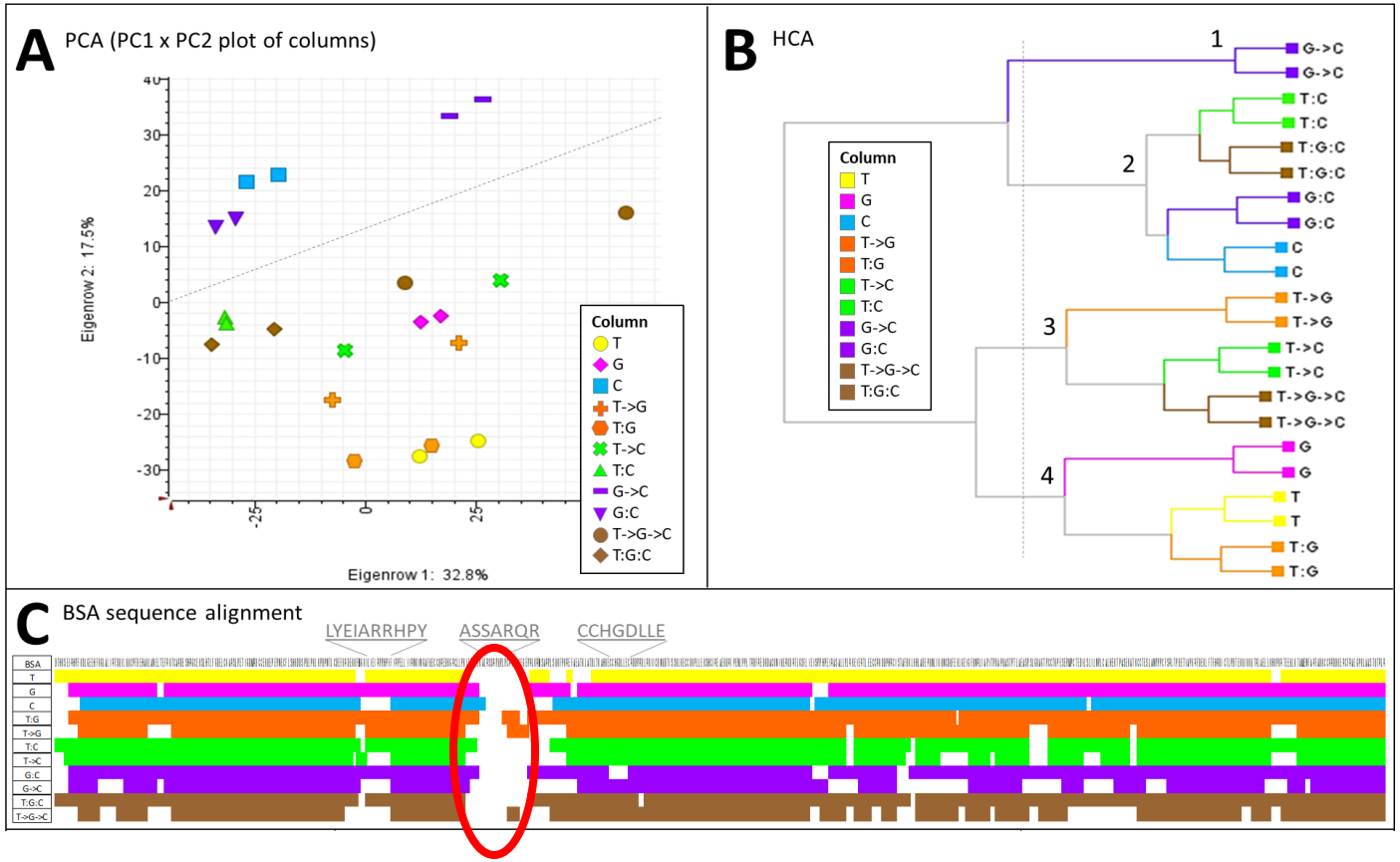
1. Benchmarking using bovine serum albumin (BSA)



- Reproducible analyses.
- Digestions with C (C, GC, TC) more orthogonal than with T and/or G.

2. Middle-down proteomics (MDP)

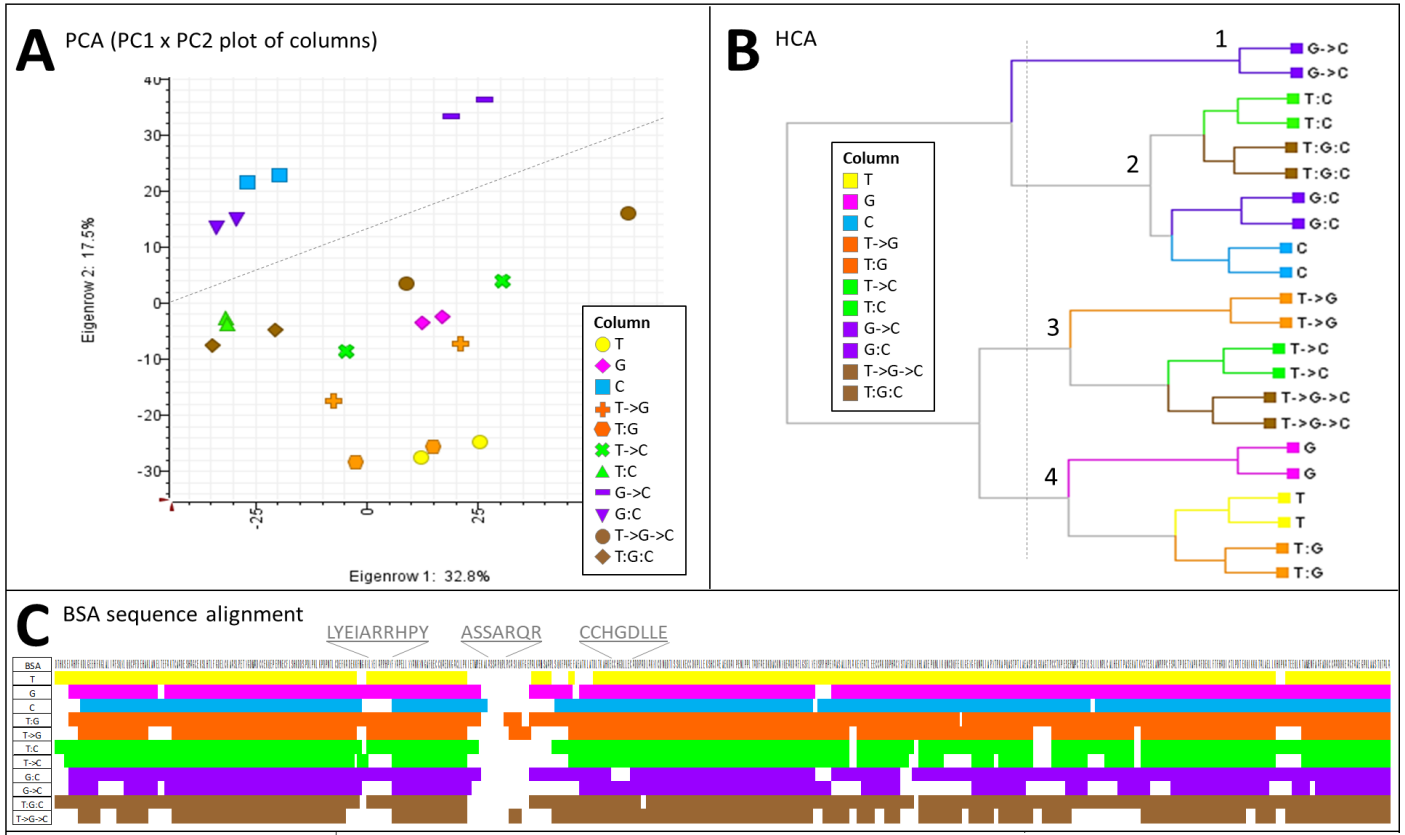
1. Benchmarking using bovine serum albumin (BSA)



- Reproducible analyses.
- Digestions with C (C, GC, TC) more orthogonal than with T and/or G.
- BSA sequence areas resisting digestion.

2. Middle-down proteomics (MDP)

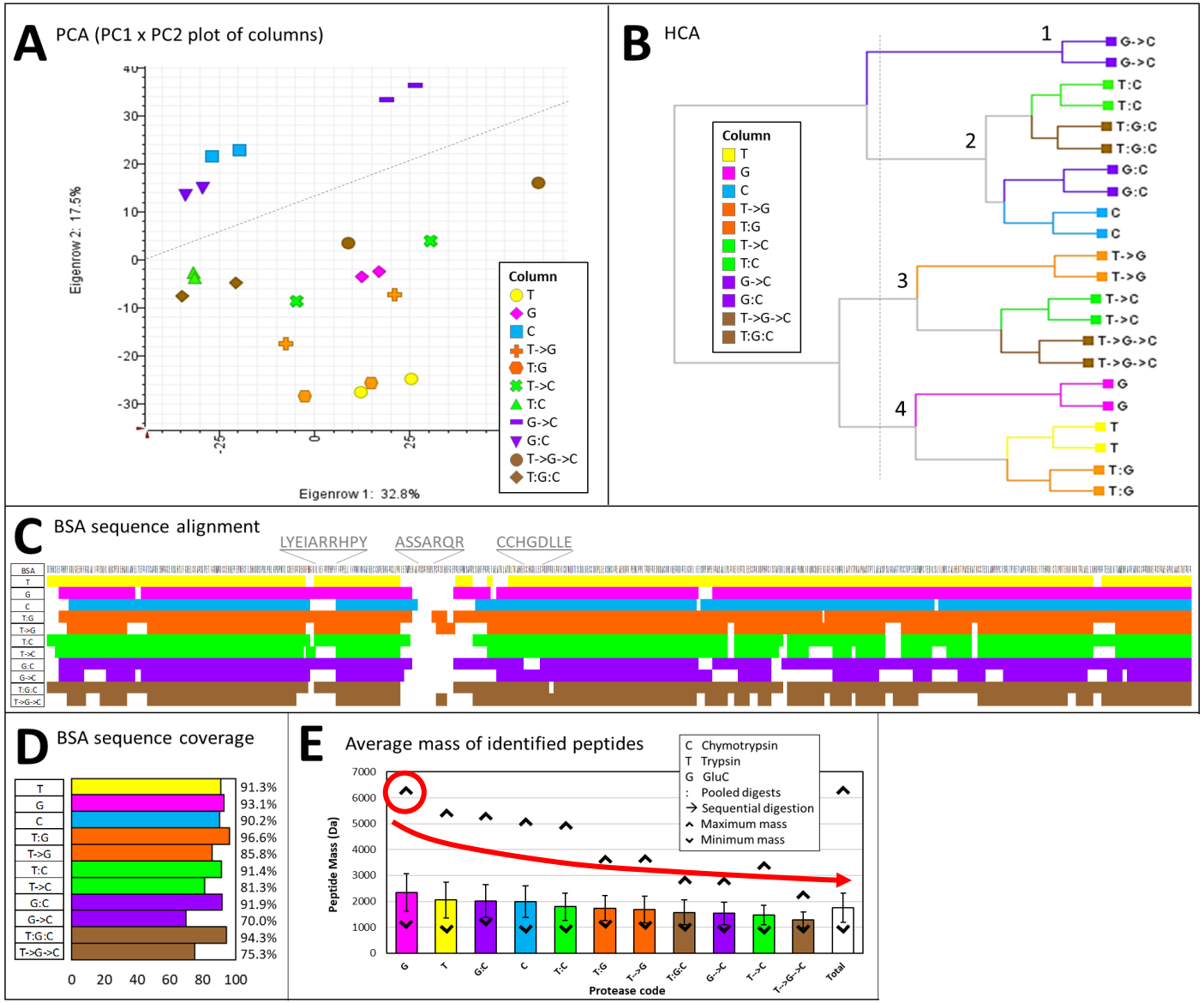
1. Benchmarking using bovine serum albumin (BSA)



- Reproducible analyses.
- Digestions with C (C, GC, TC) more orthogonal than with T and/or G.
- BSA sequence areas resisting digestion.
- Excellent coverage (>90%) using single digestions
- Best coverage using T:G (96.6%).

2. Middle-down proteomics (MDP)

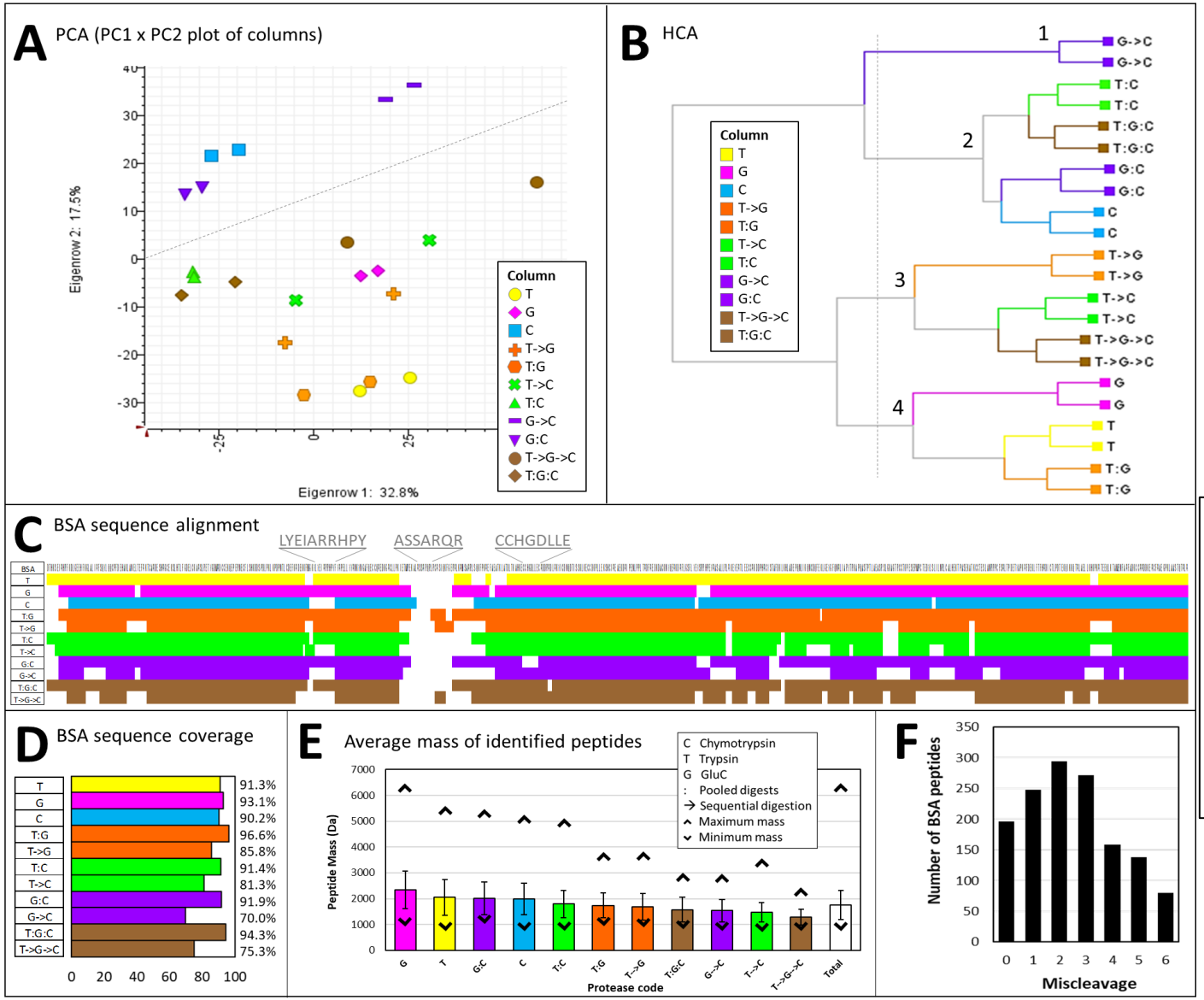
1. Benchmarking using bovine serum albumin (BSA)



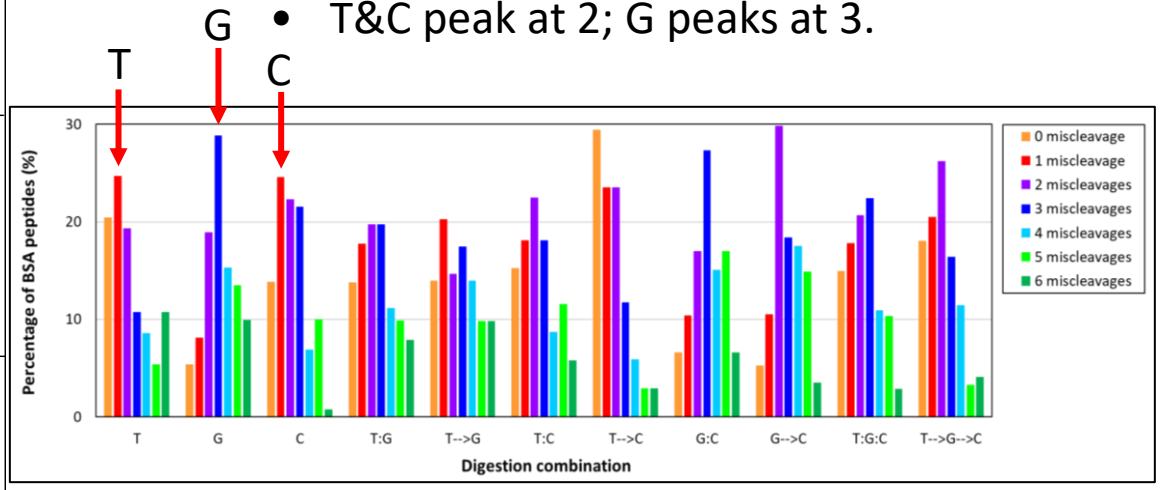
- Reproducible analyses.
- Digestions with C (C, GC, TC) more orthogonal than with T and/or G.
- BSA sequence areas resisting digestion.
- Excellent coverage (>90%) using single digestions
- Best coverage using T:G (96.6%).
- Peptide size decrease when proteases of ls/he are combined
- G produced largest BSA peptides (>6KDa).

2. Middle-down proteomics (MDP)

1. Benchmarking using bovine serum albumin (BSA)



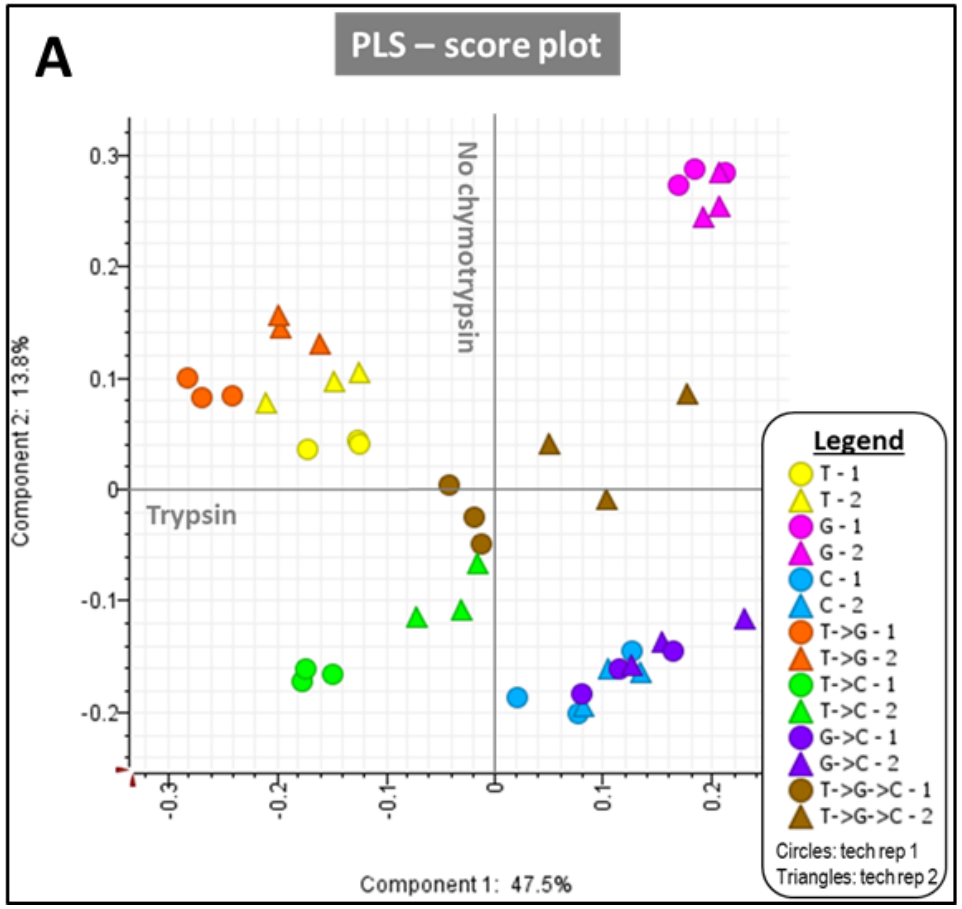
- Reproducible analyses.
- Digestions with C (C, GC, TC) more orthogonal than with T and/or G.
- BSA sequence areas resisting digestion.
- Excellent coverage (>90%) using single digestions
- Best coverage using T:G (96.6%).
- Peptide size decrease when proteases of ls/he are combined
- G produced largest BSA peptides (>6KDa).
- Up to 6 miscleavages observed.
- T&C peak at 2; G peaks at 3.



Digestions were successful on BSA therefore they could be tested on cannabis proteins.

2. Middle-down proteomics (MDP)

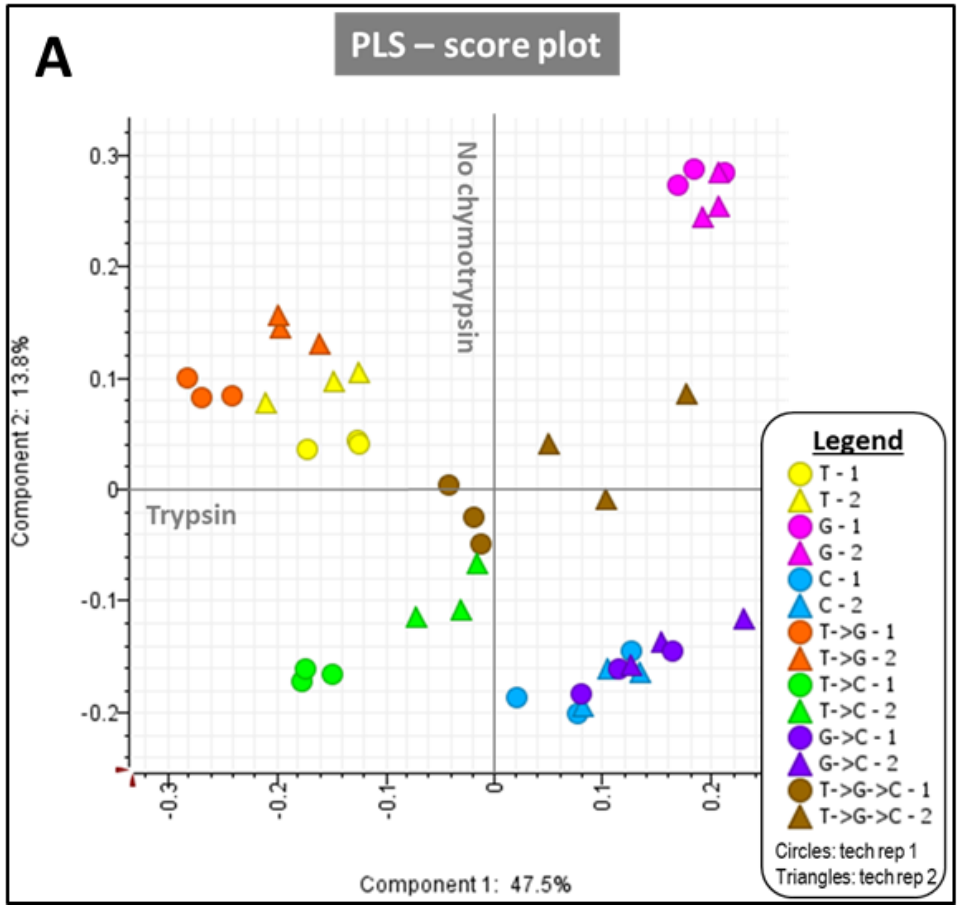
2. Cannabis results

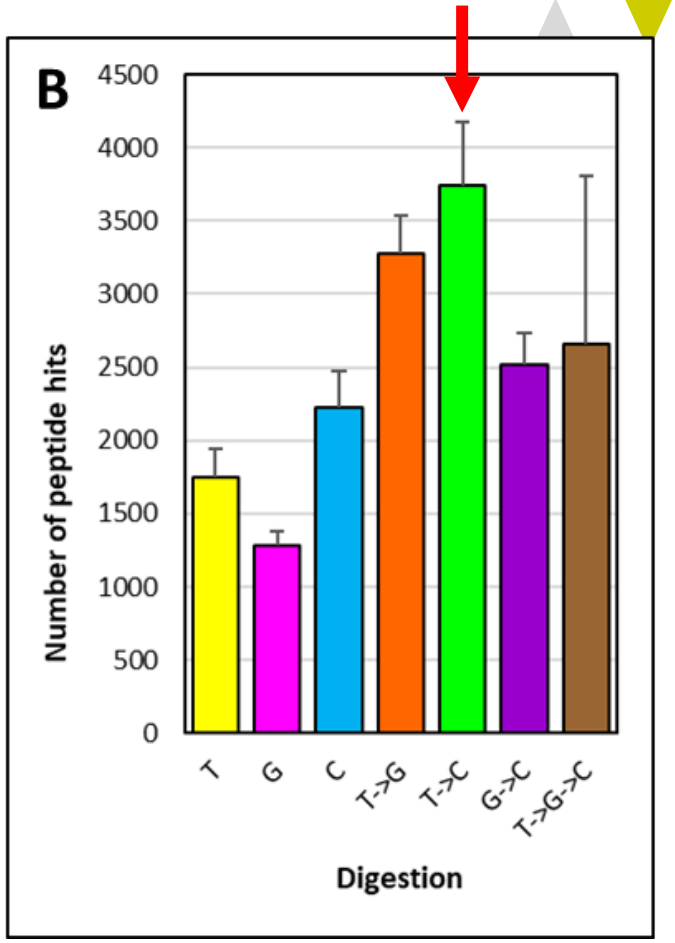


- Reproducible analyses.

2. Middle-down proteomics (MDP)

2. Cannabis results

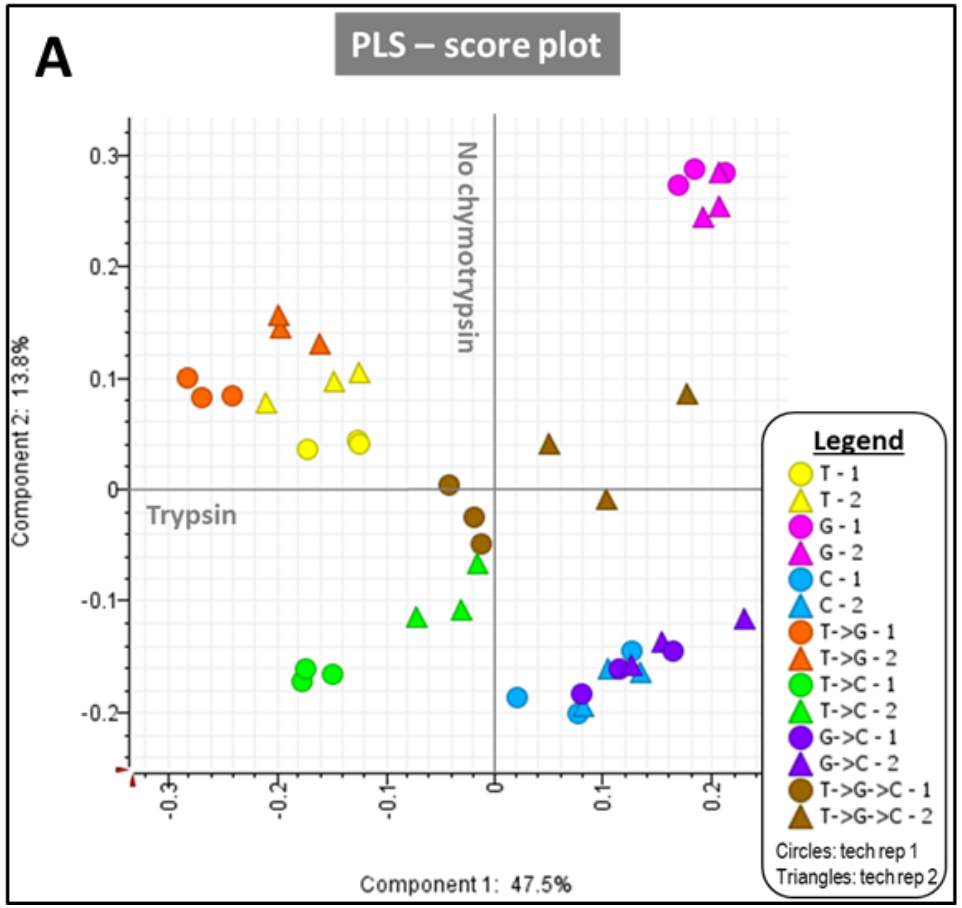


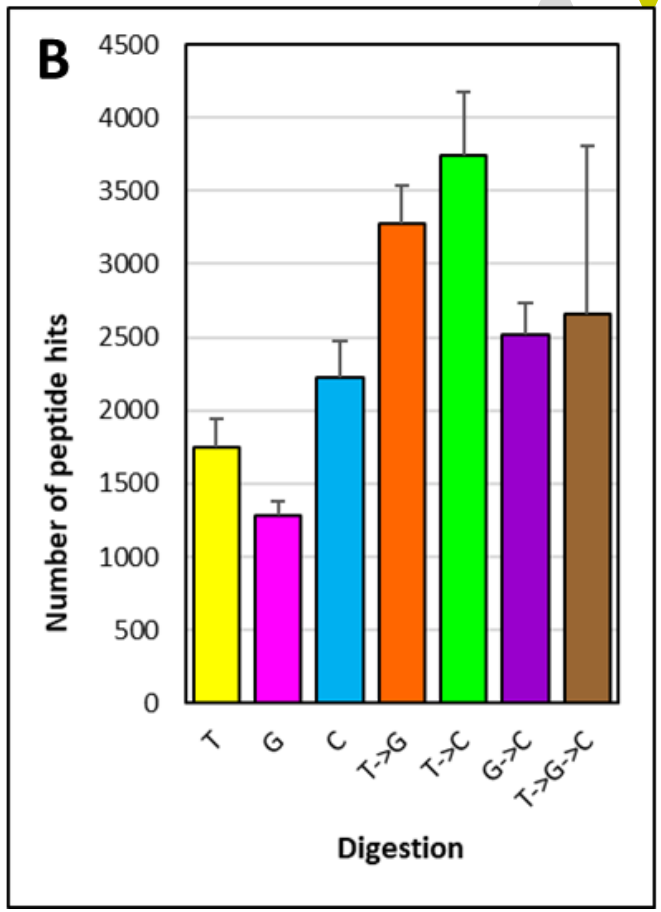


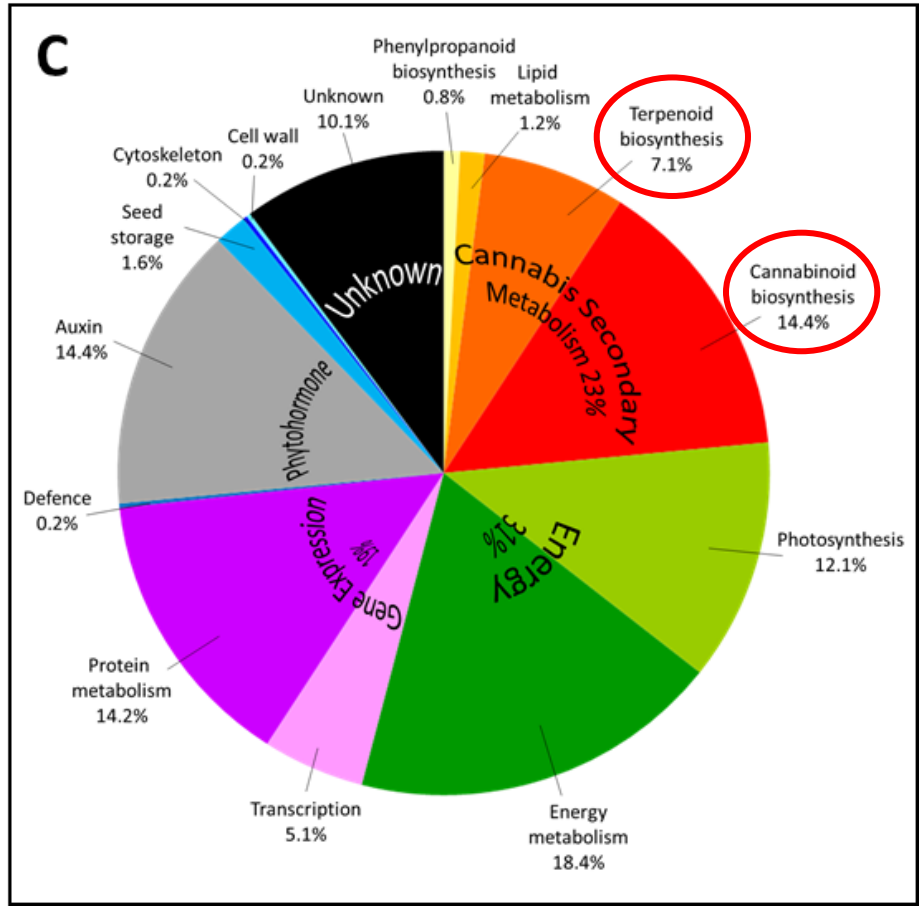
- Reproducible analyses.
- Most peptides identified with TC.

2. Middle-down proteomics (MDP)

2. Cannabis results



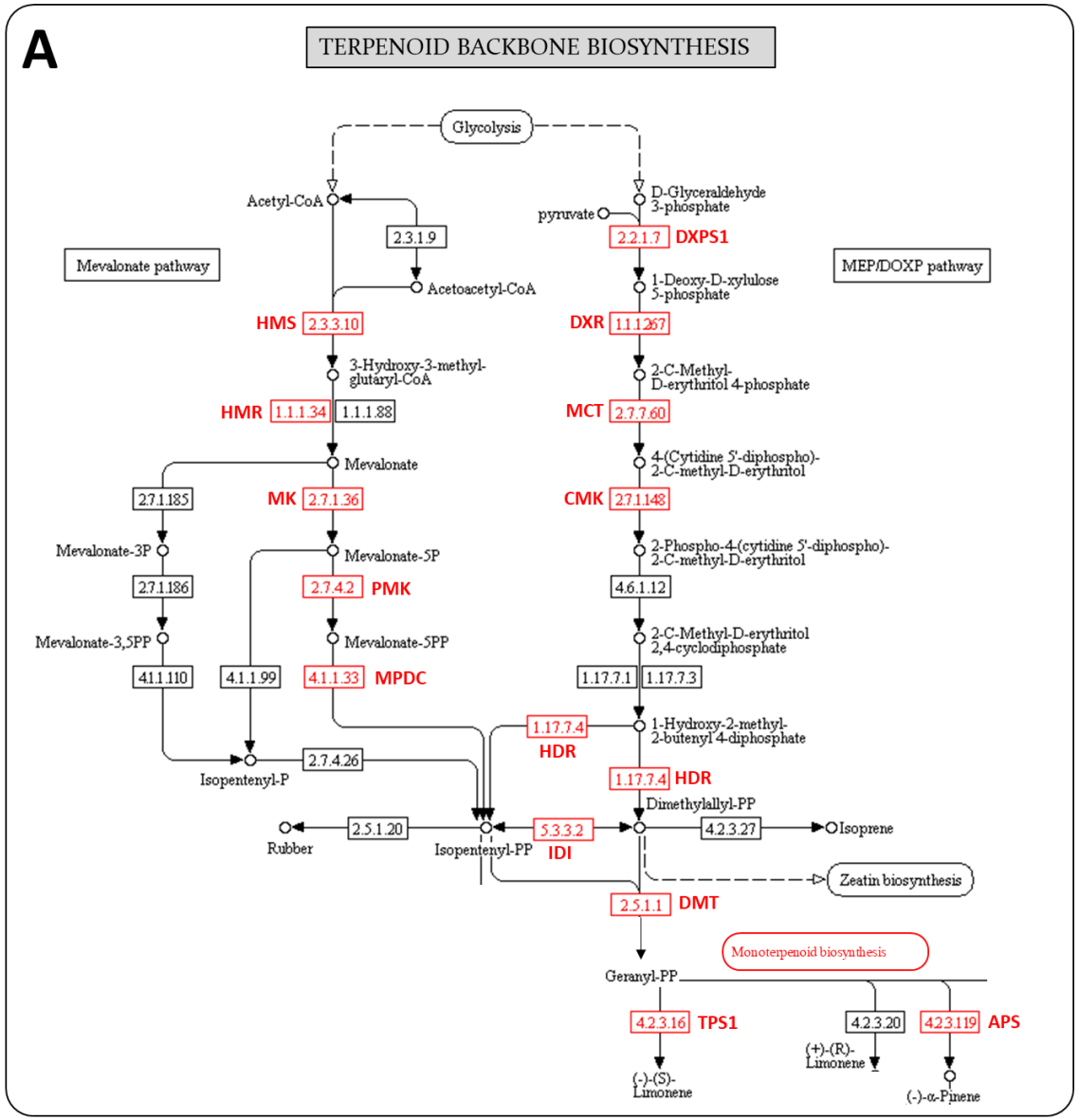




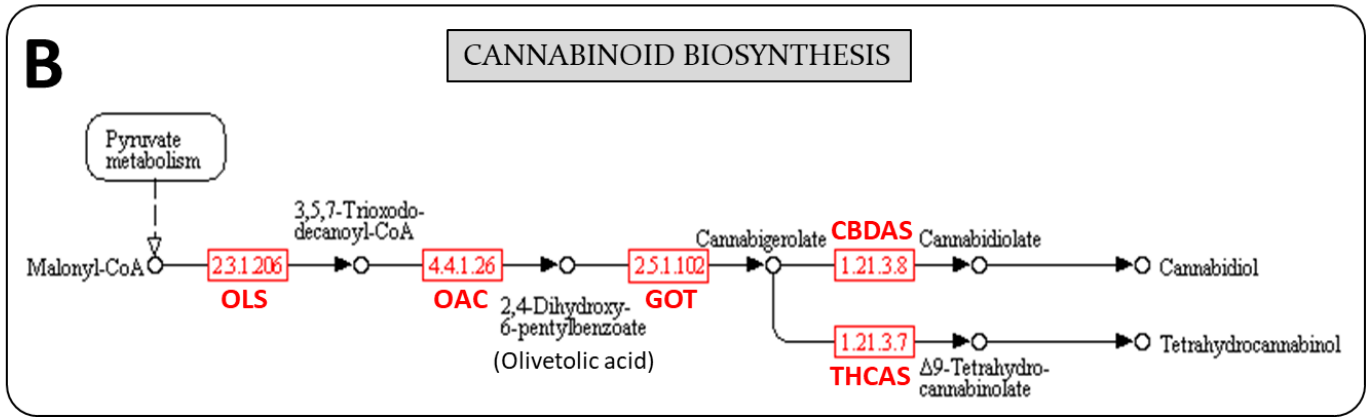
- Reproducible analyses.
- Most peptides identified with TC.
- 27,123 unique peptides identified, 494 accessions and 229 annotations (30% more than BUP).
- Secondary metabolism 23%, energy 31% (photosynthesis 12%), gene expression 19% (protein 14%).

2. Middle-down proteomics (MDP)

2. Cannabis results

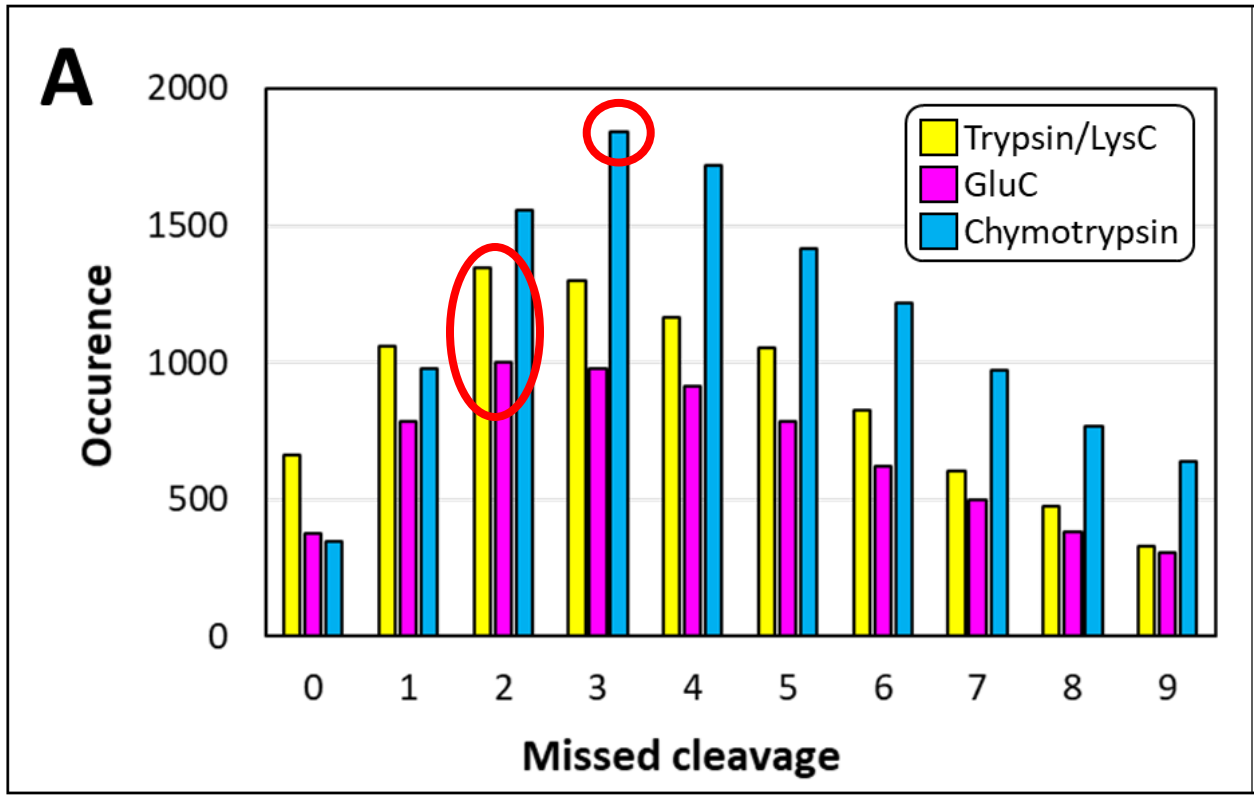


In red, cannabis enzymes identified in this study.



2. Middle-down proteomics (MDP)

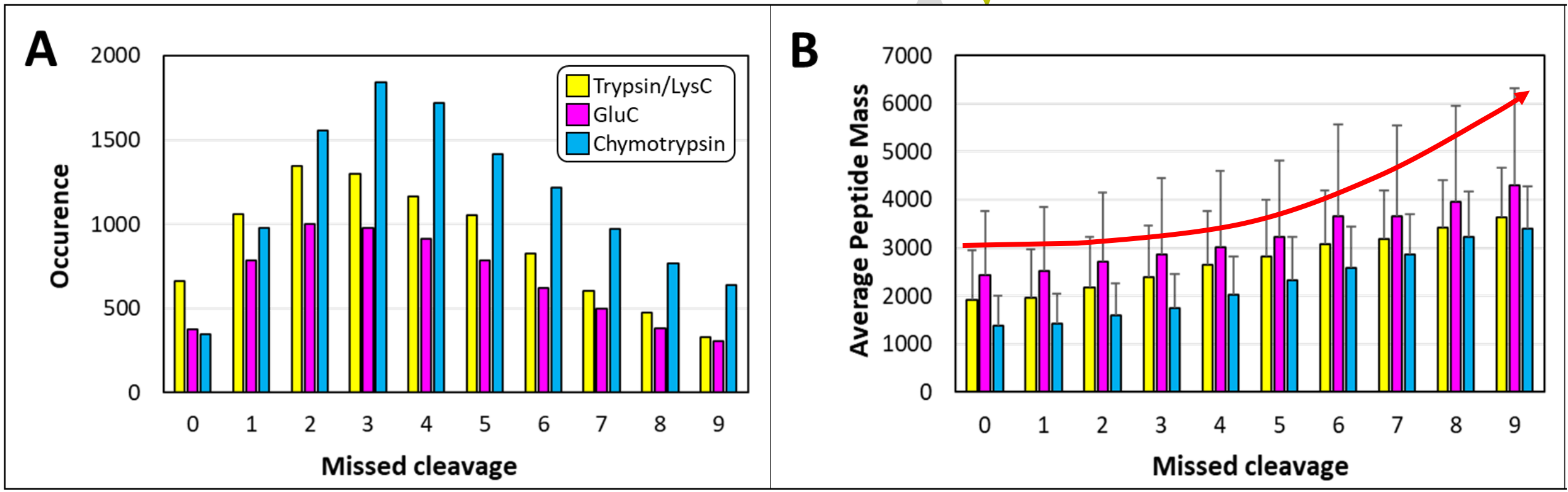
2. Cannabis results



- Up to 9 miscleavages.
- Miscleavage frequency is protease-dependent.

2. Middle-down proteomics (MDP)

2. Cannabis results

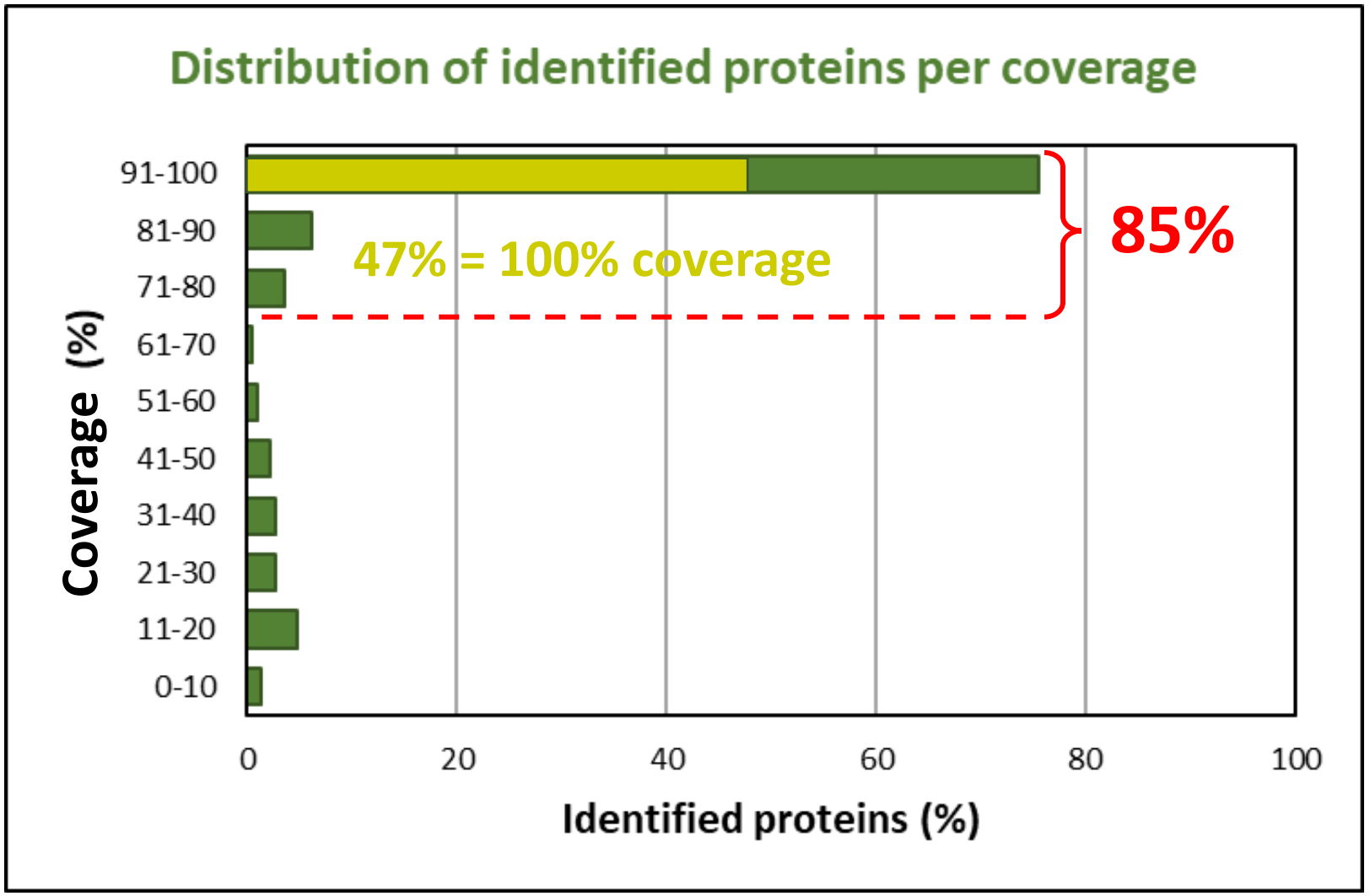


- Up to 9 miscleavages.
- Miscleavage frequency is protease-dependent.
- Peptide size increase with number of miscleavages, irrespective of the protease used.

Shotgun proteomics study should apply more than 2 missed cleavages!

2. Middle-down proteomics (MDP)

2. Cannabis results



- Coverage rate spans 6-100% in a MW-dependent manner.
- Most proteins (85%) identified with coverage > 70%.
- Almost half (47%) fully covered (100% coverage).
- PTMs discovered

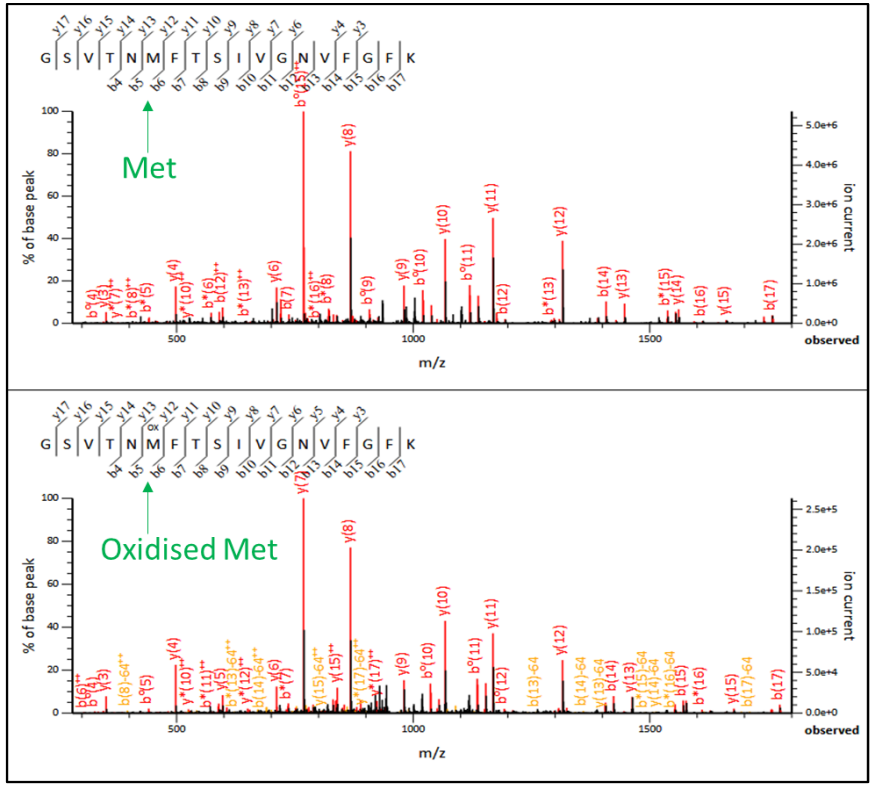
AGRICULTURE VICTORIA

2. Middle-down proteomics (MDP)

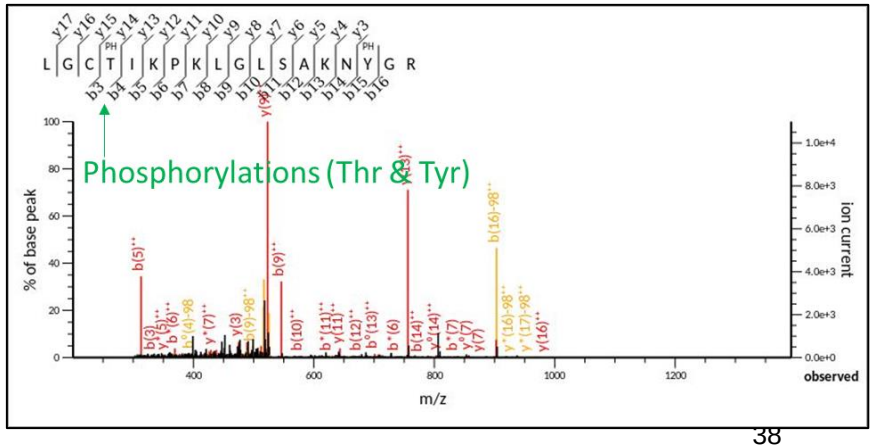
2. Cannabis results

- PTMs discovered

	Fixed modifications	Dynamic modifications			
Proteases	Carbamidomethylation	Acetylation	Phosphorylation	Oxidation	*Total*
Trypsin/LysC	1362	296	6213	2927	*10798*
Chymotrypsin	1483	238	7683	3520	*12924*
GluC	1396	149	4820	2789	*9154*
Total	*4241*	*683*	*18716*	*9236*	*32876*



Multiproteases needed to achieve 100% coverage and identify PTMs.



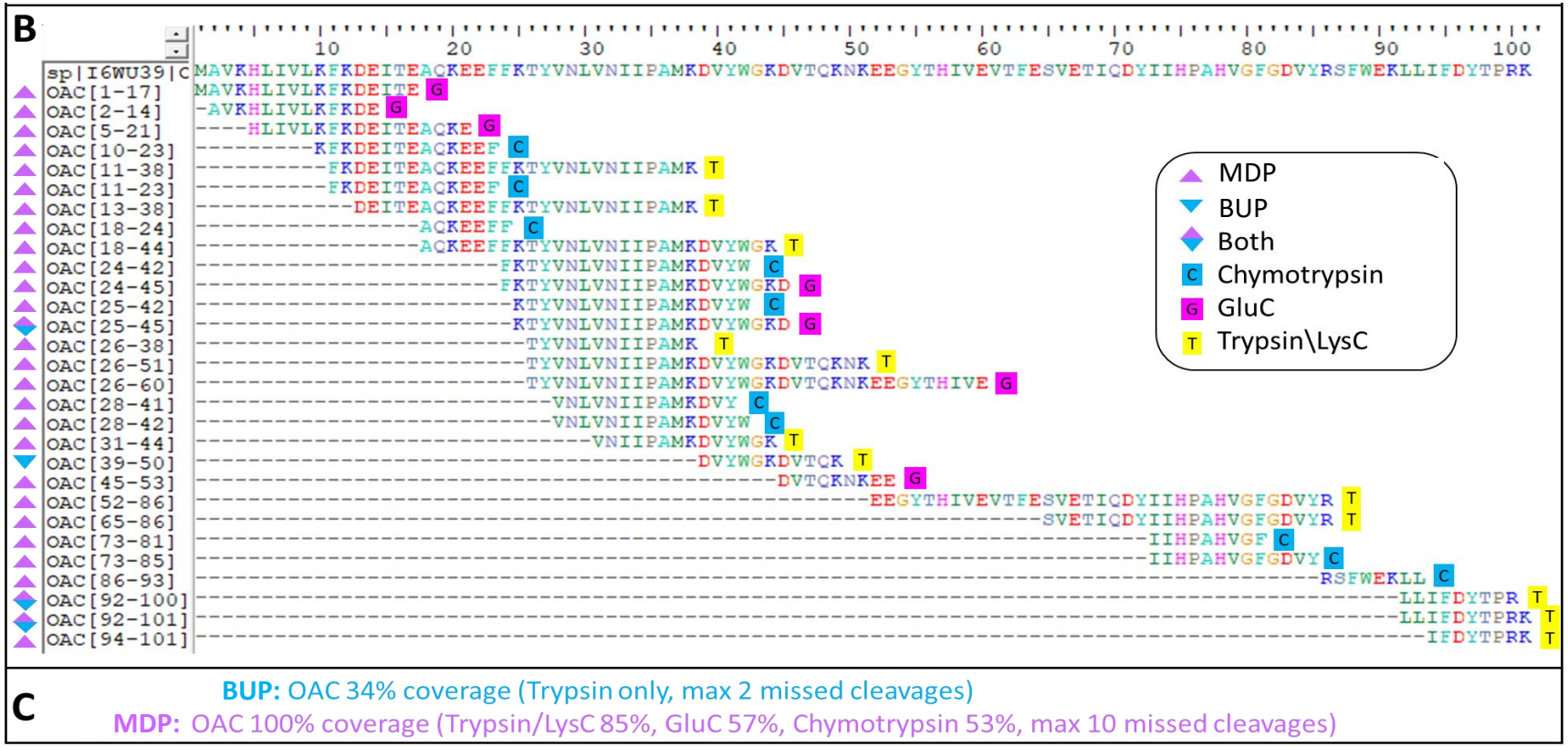
Focus on:

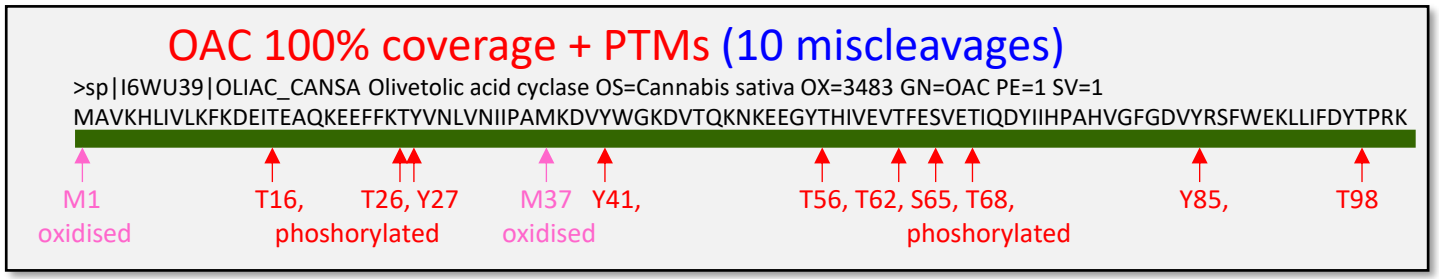
- OAC (olivetolic acid cyclase) (101 AAs, 12kDa)
- THCAS (tetrahydrocannabinolic acid synthase) (516 AAs, 59kDa)

2. Middle-down proteomics (MDP)

2. Cannabis results

Focus on: OAC (100% sequence coverage) (34% BUP)





2. Middle-down proteomics

2. Cannabis results

Focus on: THCAS

Peptide Sequence	Modifications	MH+ [Da]	# Missed Cleavages
IDVHSQTAW		1056.52	1
IDVHSQTAWVE		1284.63	2
SMGEDLFWAIR		1324.64	4
ILKENPNLCAY	C9(Carbamidomethyl)	1334.68	2
SMGEDLFWAIR	M2(Oxidation)	1340.63	4
KLYEEDVGAGMY		1374.63	5
GGGGENFGIIAAWK		1376.70	3
WGVLFGFGPGLTVE		1478.77	3
KPIPETAMVKILEK	K1(Methyl); M8(Oxidation)	1626.96	3
SKHIPNNVANPKLVY		1693.94	2
ILEKLYEEDVGAGMY		1729.84	6
GMSYISQVPFVVVDLR		1809.96	3
ENPNLCAYEAPSLDAR	C6(Carbamidomethyl)	1819.83	4
GLAADNIIDAHLVNVDGK		1834.98	3
GGIMEEISESAIPFPHR		1869.93	3
IDVHSQTAWVEAGATLGE		1883.92	3
AGGTVLRLAKDLAENNK	N-Term(Acetyl); T4(Phospho); K10(Methyl)	1906.00	4
NRLVKVKTKVDPNNF	N-Term(Acetyl); K5(Acetyl); T8(Phospho)	1935.99	4
YFGKNFNRLVKVKTK	N-Term(Acetyl); K4(Methyl); T14(Phospho)	1978.07	7
NYGLAADNIIDAHLVNVDGK		2112.08	4
LVSAAQTILPDSDGAIDGHLRE		2278.18	4
LGKEAATKAIKEWGQPKSK	T7(Phospho); K11(Methyl); S18(Phospho); K19(Acetyl)	2286.14	7
VKKPIPETAMVKILEKLY	T8(Phospho); K12(Methyl); K16(Acetyl); Y18(Phospho)	2316.19	5
TTIFYSGVVNFNTANFKKE	T1(Phospho); T13(Phospho); K18(Methyl)	2354.07	6
IKLVAVPSKSTIFSVKKNME	S14(Phospho); K16(Methyl); K17(Acetyl); M19(Oxidation)	2371.29	5
LNYRDLDLGKTNHASPNNY	K10(Methyl); N18(NAG)	2440.20	5
VLYPYGGIMEEISESAIPFPHR		2505.25	4
KIKLVAVPSKSTIFSVKKNME	S9(Phospho); S15(Phospho); K18(Methyl); M20(Oxidation)	2537.30	6
ITHLVFCTTSGVDMPGADYQLTK	C7(Carbamidomethyl)	2554.26	4
WIAHPGGPAILDQVESKLALKTE	K17(Methyl); K21(Acetyl); T22(Phospho)	2609.35	5
LDYVKKPIPETAMVKILEKLY	Y3(Phospho); T11(Phospho); K19(Acetyl); Y21(Phospho)	2773.34	7
IDVHSQTAWVEAGATLGEVYYWINEK		2979.47	8
AEGPATIMAIGTATPANCVLQSEYPDYYFR	C18(Carbamidomethyl)	3306.57	6
AFKPLGISDWNSLFWIAHPGGPAILDQVESK		3393.79	7
GGGGENFGIIAAWKIKLVAVPSKSTIFSVKK	K16(Methyl); T25(Phospho); S28(Phospho); K30(Acetyl)	3418.82	8
DKDLVLMTHFITKNITDNHGKNKTTVHGY	K23(Acetyl); Y29(Phospho)	3462.71	8
CTTSGVDMPGADYQLTKLLGLRPSVKRLMMY	C1(Carbamidomethyl); Y13(Phospho); K17(Methyl); K26(Methyl); Y31(Phospho)	3689.74	6
YDKDLVLMTHFITKNITDNHGKNKTTVHGY	Y1(Phospho); K3(Methyl); K22(Methyl); Y30(Phospho)	3691.71	9
LYEEDVGAGMYVLYPYGGIMEEISESAIPFPHR		3732.77	9
SLNEAFKPLGISDWNSLFWIAHPGGPAILDQVESK		3836.95	8
KHINWVRSVYNFTTPYVSQNPRLAYLNYR	K1(Methyl); S8(Phospho); S18(Phospho); Y28(Phospho)	3853.81	9
QLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAKD	T3(Phospho); C22(Carbamidomethyl)	3929.05	7
ITAVTFRGPNDTHLDSLVGQALFGDGSAALIVGSDPIPE		3951.05	6
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTR	T5(Phospho); K7(Methyl); C26(Carbamidomethyl); T36(Phospho)	4260.25	4
GQPKSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVK	T8(Phospho); C13(Carbamidomethyl); T28(Phospho); K38(Acetyl)	4345.18	7
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHD	K7(Methyl); S15(Phospho); C26(Carbamidomethyl); S27(Phospho); K28(Acetyl)	4755.41	5
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHDAE	S15(Phospho); C26(Carbamidomethyl); K28(Methyl); K29(Acetyl)	4875.54	6

2. Middle-down proteomics

2. Cannabis results

Focus on: THCAS

- peptide masses 1-5kDa

Peptide Sequence	Modifications	MH+ [Da]	# Missed Cleavages
IDVHSQTAW		1056.52	1
IDVHSQTAWVE		1284.63	2
SMGEDLFWAIR		1324.64	4
ILKENPNLCAY	C9(Carbamidomethyl)	1334.68	2
SMGEDLFWAIR	M2(Oxidation)	1340.63	4
KLYEEDVGAGMY		1374.63	5
GGGGENFGIIAAWK		1376.70	3
WGVLFGFGPGLTVE		1478.77	3
KPIPETAMVKILEK	K1(Methyl); M8(Oxidation)	1626.96	3
SKHIPNNVANPKLVY		1693.94	2
ILEKLYEEDVGAGMY		1729.84	6
GMSYISQVPFVVVDLR		1809.96	3
ENPNLCAYEAPSLDAR	C6(Carbamidomethyl)	1819.83	4
GLAADNIIDAHLVNVDGK		1834.98	3
GGIMEEISESAIPFPHR		1869.93	3
IDVHSQTAWVEAGATLGE		1883.92	3
AGGTVLRLAKDLAENNK	N-Term(Acetyl); T4(Phospho); K10(Methyl)	1906.00	4
NRLVKVKTKVDPNNF	N-Term(Acetyl); K5(Acetyl); T8(Phospho)	1935.99	4
YFGKNFNRLVKVKTK	N-Term(Acetyl); K4(Methyl); T14(Phospho)	1978.07	7
NYGLAADNIIDAHLVNVDGK		2112.08	4
LVSAAQTILPDSDGAIDGHLRE		2278.18	4
LGKEAATKAIKEWGQPKSK	T7(Phospho); K11(Methyl); S18(Phospho); K19(Acetyl)	2286.14	7
VKKPIPETAMVKILEKLY	T8(Phospho); K12(Methyl); K16(Acetyl); Y18(Phospho)	2316.19	5
TTIFYSGVVNFNTANFKKE	T1(Phospho); T13(Phospho); K18(Methyl)	2354.07	6
IKLVAVPSKSTIFSVKKNME	S14(Phospho); K16(Methyl); K17(Acetyl); M19(Oxidation)	2371.29	5
LNYRDLDLGKTNHASPNNY	K10(Methyl); N18(NAG)	2440.20	5
VLYPYGGIMEEISESAIPFPHR		2505.25	4
KIKLVAVPSKSTIFSVKKNME	S9(Phospho); S15(Phospho); K18(Methyl); M20(Oxidation)	2537.30	6
ITHLVFCTTSGVDMPGADYQLTK	C7(Carbamidomethyl)	2554.26	4
WIAHPGGPAILDQVESKLALKTE	K17(Methyl); K21(Acetyl); T22(Phospho)	2609.35	5
LDYVKKPIPETAMVKILEKLY	Y3(Phospho); T11(Phospho); K19(Acetyl); Y21(Phospho)	2773.34	7
IDVHSQTAWVEAGATLGEVYYWINEK		2979.47	8
AEGPATIMAIGTATPANCVLQSEYPDYYFR	C18(Carbamidomethyl)	3306.57	6
AFKPLGISDWNSLFWIAHPGGPAILDQVESK		3393.79	7
GGGGENFGIIAAWKIKLVAVPSKSTIFSVKK	K16(Methyl); T25(Phospho); S28(Phospho); K30(Acetyl)	3418.82	8
DKDLVLMTHFITKNITDNHGKNKTTVHGY	K23(Acetyl); Y29(Phospho)	3462.71	8
CTTSGVDMPGADYQLTKLLGLRPSVKRLMMY	C1(Carbamidomethyl); Y13(Phospho); K17(Methyl); K26(Methyl); Y31(Phospho)	3689.74	6
YDKDLVLMTHFITKNITDNHGKNKTTVHGY	Y1(Phospho); K3(Methyl); K22(Methyl); Y30(Phospho)	3691.71	9
LYEEDVGAGMYVLYPYGGIMEEISESAIPFPHR		3732.77	9
SLNEAFKPLGISDWNSLFWIAHPGGPAILDQVESK		3836.95	8
KHINWVRSVYNFTTPYVSQNPRLAYLNYR	K1(Methyl); S8(Phospho); S18(Phospho); Y28(Phospho)	3853.81	9
QLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAKD	T3(Phospho); C22(Carbamidomethyl)	3929.05	7
ITAVTFRGPNDTHLDSLVGQALFGDGSAALIVGSDPIPE		3951.05	6
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTR	T5(Phospho); K7(Methyl); C26(Carbamidomethyl); T36(Phospho)	4260.25	4
GQPKSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVK	T8(Phospho); C13(Carbamidomethyl); T28(Phospho); K38(Acetyl)	4345.18	7
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHD	K7(Methyl); S15(Phospho); C26(Carbamidomethyl); S27(Phospho); K28(Acetyl)	4755.41	5
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHDAE	S15(Phospho); C26(Carbamidomethyl); K28(Methyl); K29(Acetyl)	4875.54	6

2. Middle-down proteomics

2. Cannabis results

Focus on: THCAS

- peptide masses 1-5kDa
- most peptides contain > 2 miscleavages

Peptide Sequence	Modifications	MH+ [Da]	# Missed Cleavages
IDVHSQTAW		1056.52	1
IDVHSQTAWVE		1284.63	2
SMGEDLFWAIR		1324.64	4
ILKENPNLCAY	C9(Carbamidomethyl)	1334.68	2
SMGEDLFWAIR	M2(Oxidation)	1340.63	4
KLYEEDVGAGMY		1374.63	5
GGGGENFGIIAAWK		1376.70	3
WGVLFGFGPGLTVE		1478.77	3
KPIPETAMVKILEK	K1(Methyl); M8(Oxidation)	1626.96	3
SKHIPNNVANPKLVY		1693.94	2
ILEKLYEEDVGAGMY		1729.84	6
GMSYISQVPFVVVDLR		1809.96	3
ENPNLCAYEAPSLDAR	C6(Carbamidomethyl)	1819.83	4
GLAADNIIDAHLVNVDGK		1834.98	3
GGIMEEISESAIPFPHR		1869.93	3
IDVHSQTAWVEAGATLGE		1883.92	3
AGGTVLRLAKDLAENNK	N-Term(Acetyl); T4(Phospho); K10(Methyl)	1906.00	4
NRLVKVKTKVDPNNF	N-Term(Acetyl); K5(Acetyl); T8(Phospho)	1935.99	4
YFGKNFNRLVKVKTK	N-Term(Acetyl); K4(Methyl); T14(Phospho)	1978.07	7
NYGLAADNIIDAHLVNVDGK		2112.08	4
LVSAAQTILPDSDGAIDGHLRE		2278.18	4
LGKEAATKAIKEWGQPKSK	T7(Phospho); K11(Methyl); S18(Phospho); K19(Acetyl)	2286.14	7
VKKPIPETAMVKILEKLY	T8(Phospho); K12(Methyl); K16(Acetyl); Y18(Phospho)	2316.19	5
TTIFYSGVVNFNTANFKKE	T1(Phospho); T13(Phospho); K18(Methyl)	2354.07	6
IKLVAVPSKSTIFSVKKNME	S14(Phospho); K16(Methyl); K17(Acetyl); M19(Oxidation)	2371.29	5
LNYRDLDLGKTNHASPNNY	K10(Methyl); N18(NAG)	2440.20	5
VLYPYGGIMEEISESAIPFPHR		2505.25	4
KIKLVAVPSKSTIFSVKKNME	S9(Phospho); S15(Phospho); K18(Methyl); M20(Oxidation)	2537.30	6
ITHLVFCTTSGVDMPGADYQLTK	C7(Carbamidomethyl)	2554.26	4
WIAHPGGPAILDQVESKLALKTE	K17(Methyl); K21(Acetyl); T22(Phospho)	2609.35	5
LDYVKKPIPETAMVKILEKLY	Y3(Phospho); T11(Phospho); K19(Acetyl); Y21(Phospho)	2773.34	7
IDVHSQTAWVEAGATLGEVYYWINEK		2979.47	8
AEGPATIMAIGTATPANCVLQSEYPDYYFR	C18(Carbamidomethyl)	3306.57	6
AFKPLGISDWNSLFWIAHPGGPAILDQVESK		3393.79	7
GGGGENFGIIAAWKIKLVAVPSKSTIFSVKK	K16(Methyl); T25(Phospho); S28(Phospho); K30(Acetyl)	3418.82	8
DKDLVLMTHFITKNITDNHGKNKTTVHGY	K23(Acetyl); Y29(Phospho)	3462.71	8
CTTSGVDMPGADYQLTKLLGLRPSVKRLMMY	C1(Carbamidomethyl); Y13(Phospho); K17(Methyl); K26(Methyl); Y31(Phospho)	3689.74	6
YDKDLVLMTHFITKNITDNHGKNKTTVHGY	Y1(Phospho); K3(Methyl); K22(Methyl); Y30(Phospho)	3691.71	9
LYEEDVGAGMYVLYPYGGIMEEISESAIPFPHR		3732.77	9
SLNEAFKPLGISDWNSLFWIAHPGGPAILDQVESK		3836.95	8
KHINWVRSVYNFTTPYVSQNPRLAYLNYR	K1(Methyl); S8(Phospho); S18(Phospho); Y28(Phospho)	3853.81	9
QLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAKD	T3(Phospho); C22(Carbamidomethyl)	3929.05	7
ITAVTFRGPNDTHLDSLVGQALFGDGSAALIVGSDPIPE		3951.05	6
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTR	T5(Phospho); K7(Methyl); C26(Carbamidomethyl); T36(Phospho)	4260.25	4
GQPKSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVK	T8(Phospho); C13(Carbamidomethyl); T28(Phospho); K38(Acetyl)	4345.18	7
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHD	K7(Methyl); S15(Phospho); C26(Carbamidomethyl); S27(Phospho); K28(Acetyl)	4755.41	5
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHDAE	S15(Phospho); C26(Carbamidomethyl); K28(Methyl); K29(Acetyl)	4875.54	6

2. Middle-down proteomics

2. Cannabis results

Focus on: THCAS

- peptide masses 1-5kDa
- most peptides contain > 2 miscleavages
- 72% sequence coverage (12% BUP)

Peptide Sequence	Modifications	MH+ [Da]	# Missed Cleavages
IDVHSQTAW		1056.52	1
IDVHSQTAWVE		1284.63	2
SMGEDLFWAIR		1324.64	4
ILKENPNLCAY	C9(Carbamidomethyl)	1334.68	2
SMGEDLFWAIR	M2(Oxidation)	1340.63	4
KLYEEDVGAGMY		1374.63	5
GGGGENFGIIAAWK		1376.70	3
WGVLFGFGPGLTVE		1478.77	3
KPIPETAMVKILEK	K1(Methyl); M8(Oxidation)	1626.96	3
SKHIPNNVANPKLVY		1693.94	2
ILEKLYEEDVGAGMY		1729.84	6
GMSYISQVPFVVVDLR		1809.96	3
ENPNLCAYEAPSLDAR	C6(Carbamidomethyl)	1819.83	4
GLAADNIIDAHLVNVDGK		1834.98	3
GGIMEEISESAIPFPHR		1869.93	3
IDVHSQTAWVEAGATLGE		1883.92	3
AGGTVLRLAKDLAENNK	N-Term(Acetyl); T4(Phospho); K10(Methyl)	1906.00	4
NRLVKVKTKVDPNNF	N-Term(Acetyl); K5(Acetyl); T8(Phospho)	1935.99	4
YFGKNFNRLVKVKTK	N-Term(Acetyl); K4(Methyl); T14(Phospho)	1978.07	7
NYGLAADNIIDAHLVNVDGK		2112.08	4
LVSAAQTILPDSDGAIDGHLRE		2278.18	4
LGKEAATKAIKEWGQPKSK	T7(Phospho); K11(Methyl); S18(Phospho); K19(Acetyl)	2286.14	7
VKKPIPETAMVKILEKLY	T8(Phospho); K12(Methyl); K16(Acetyl); Y18(Phospho)	2316.19	5
TTIFYSGVVNFNTANFKKE	T1(Phospho); T13(Phospho); K18(Methyl)	2354.07	6
IKLVAVPSKSTIFSVKKNME	S14(Phospho); K16(Methyl); K17(Acetyl); M19(Oxidation)	2371.29	5
	K10(Methyl); N18(NAG)	2440.20	5
		2505.25	4
	S9(Phospho); S15(Phospho); K18(Methyl); M20(Oxidation)	2537.30	6
	C7(Carbamidomethyl)	2554.26	4
	K17(Methyl); K21(Acetyl); T22(Phospho)	2609.35	5
	Y3(Phospho); T11(Phospho); K19(Acetyl); Y21(Phospho)	2773.34	7
		2979.47	8
FR	C18(Carbamidomethyl)	3306.57	6
ESK		3393.79	7
KK	K16(Methyl); T25(Phospho); S28(Phospho); K30(Acetyl)	3418.82	8
	K23(Acetyl); Y29(Phospho)	3462.71	8
MMY	C1(Carbamidomethyl); Y13(Phospho); K17(Methyl); K26(Methyl); Y31(Phospho)	3689.74	6
GY	Y1(Phospho); K3(Methyl); K22(Methyl); Y30(Phospho)	3691.71	9
FPHR		3732.77	9
LDQVESK		3836.95	8
R	K1(Methyl); S8(Phospho); S18(Phospho); Y28(Phospho)	3853.81	9
VLRLAKD	T3(Phospho); C22(Carbamidomethyl)	3929.05	7
ITAVTFRGPNDTHLDSLVGQALFGDGSAALIVGSDPIPE		3951.05	6
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTR	T5(Phospho); K7(Methyl); C26(Carbamidomethyl); T36(Phospho)	4260.25	4
GQPKSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVK	T8(Phospho); C13(Carbamidomethyl); T28(Phospho); K38(Acetyl)	4345.18	7
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHD	K7(Methyl); S15(Phospho); C26(Carbamidomethyl); S27(Phospho); K28(Acetyl)	4755.41	5
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHDAE	S15(Phospho); C26(Carbamidomethyl); K28(Methyl); K29(Acetyl)	4875.54	6

THCAS 72% coverage, many PTMs (10 miscleavages)

>sp|Q8GTB6|29-545|THCAS_CANSA Tetrahydrocannabinolic acid synthase OS=Cannabis sativa

NPRENFLKCFSKHIPNNVANPKLVYTQHDQLYMSILNSTIQNLRFISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHDAEGMSYISQVPFVV

VDLRNMHSIKIDVHSQTAWVEAGATLGEVYYWINEKNENLSFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMG

EDLFWAIRGGGGENFGIIAAWKIKLVAVPSKSTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLVLMTHFITKNITDNHGKNKTTVHGYFSSIFHGGVD

SLVDLMNKSFPELGIKKTDCKEFSWIDTTIFYSGVVNFNTANFKKEILLDRSAGKKTAFSIKLDYVKKPIPETAMVKILEKLYEEDVGAGMYVLYPYGGIME

EISESAIPFPHRAGIMYELWYTASWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNHASPNNYTQARIWGEKYFGKNFNRLVKV

KTKVDPNNFFRNEQSIPPLPPHHH

2. Middle-down proteomics

2. Cannabis results

Focus on: THCAS

- peptide masses 1-5kDa
- most peptides contain > 2 miscleavages
- 72% sequence coverage (12% BUP)
- Most peptides are modified

THCAS 72% coverage, many PTMs (10 miscleavages)

>sp|Q8GTB6|29-545|THCAS_CANSA Tetrahydrocannabinolic acid synthase OS=Cannabis sativa

NPRENFLKCFSKHIPNNVANPKLVYTQHDQLYMSILNSTIQNLRFISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHDAEGMSYISQVPFVV

VDLRNMHSIKIDVHSQTAWVEAGATLGEVYYWINEKNENLSFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDGKVLDRKSMG

EDLFWAIRGGGGENFGIIAAWKIKLVAVPSKSTIFSVKKNMEIHGLVKLFNKWQNIAYKYDKDLVLMTHFITKNITDNHGKNKTTVHGYFSSIFHGGVD

SLVDLMNKSFPELGIKKTDCKEFSWIDTTIFYSGVVNFNTANFKKEILLDRSAGKKTAFSIKLDYVKKPIPETAMVKILEKLYEEDVGAGMYVLYPYGGIME

EISESAIPFPHRAGIMYELWYTASWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAYLNYRDLDLGKTNHASPNNYTQARIWGEKYFGKNFNRLVKV

KTKVDPNNFFRNEQSIPPLPPHHH

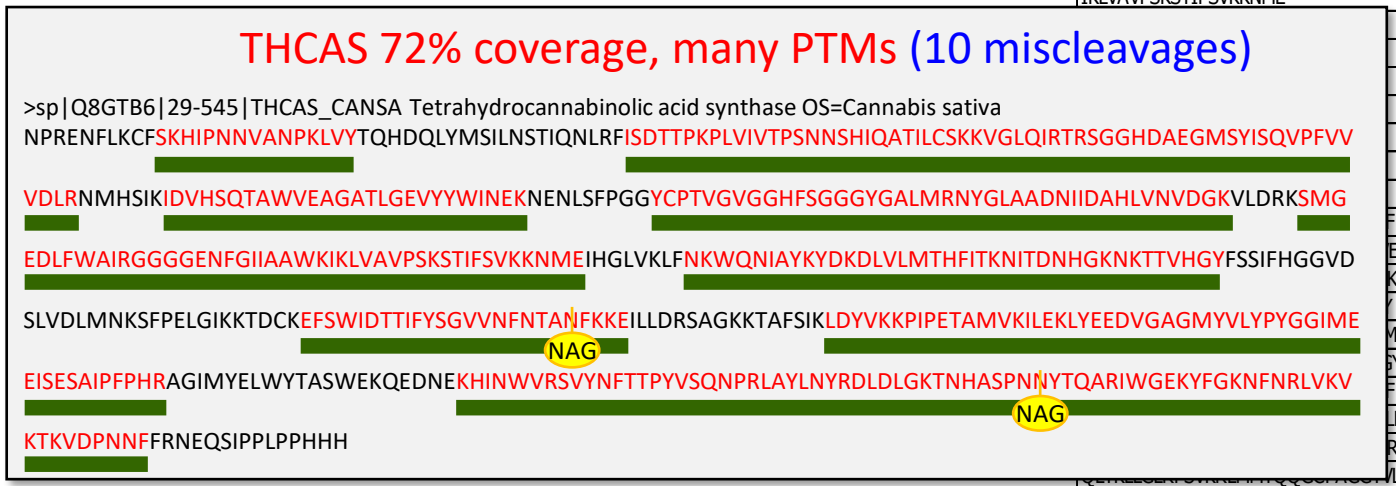
Peptide Sequence	Modifications	MH+ [Da]	# Missed Cleavages
IDVHSQTAW		1056.52	1
IDVHSQTAWVE		1284.63	2
SMGEDLFWAIR		1324.64	4
ILKENPNLCAY	C9(Carbamidomethyl)	1334.68	2
SMGEDLFWAIR	M2(Oxidation)	1340.63	4
KLYEEDVGAGMY		1374.63	5
GGGGENFGIIAAWK		1376.70	3
WGVLFGFGPGLTVE		1478.77	3
KPIPETAMVKILEK	K1(Methyl); M8(Oxidation)	1626.96	3
SKHIPNNVANPKLVY		1693.94	2
ILEKLYEEDVGAGMY		1729.84	6
GMSYISQVPFVVVDLR		1809.96	3
ENPNLCAYEAPSLDAR	C6(Carbamidomethyl)	1819.83	4
GLAADNIIDAHLVNVDGK		1834.98	3
GGIMEEISESAIPFPHR		1869.93	3
IDVHSQTAWVEAGATLGE		1883.92	3
AGGTVLRLAKDLAENNK	N-Term(Acetyl); T4(Phospho); K10(Methyl)	1906.00	4
NRLVKVKTKVDPNNF	N-Term(Acetyl); K5(Acetyl); T8(Phospho)	1935.99	4
YFGKNFNRLVKVKTK	N-Term(Acetyl); K4(Methyl); T14(Phospho)	1978.07	7
NYGLAADNIIDAHLVNVDGK		2112.08	4
LVSAAQTILPDSDGAIDGHLRE		2278.18	4
LGKEAATKAIKEWGQPKSK	T7(Phospho); K11(Methyl); S18(Phospho); K19(Acetyl)	2286.14	7
VKKPIPETAMVKILEKLY	T8(Phospho); K12(Methyl); K16(Acetyl); Y18(Phospho)	2316.19	5
TTIFYSGVVNFNTANFKKE	T1(Phospho); T13(Phospho); K18(Methyl)	2354.07	6
IKLVAVPSKSTIFSVKKNME	S14(Phospho); K16(Methyl); K17(Acetyl); M19(Oxidation)	2371.29	5
	K10(Methyl); N18(NAG)	2440.20	5
		2505.25	4
	S9(Phospho); S15(Phospho); K18(Methyl); M20(Oxidation)	2537.30	6
	C7(Carbamidomethyl)	2554.26	4
	K17(Methyl); K21(Acetyl); T22(Phospho)	2609.35	5
	Y3(Phospho); T11(Phospho); K19(Acetyl); Y21(Phospho)	2773.34	7
		2979.47	8
FR	C18(Carbamidomethyl)	3306.57	6
ESK		3393.79	7
KK	K16(Methyl); T25(Phospho); S28(Phospho); K30(Acetyl)	3418.82	8
	K23(Acetyl); Y29(Phospho)	3462.71	8
MMY	C1(Carbamidomethyl); Y13(Phospho); K17(Methyl); K26(Methyl); Y31(Phospho)	3689.74	6
GY	Y1(Phospho); K3(Methyl); K22(Methyl); Y30(Phospho)	3691.71	9
FPHR		3732.77	9
LDQVESK		3836.95	8
R	K1(Methyl); S8(Phospho); S18(Phospho); Y28(Phospho)	3853.81	9
VLRLAKD	T3(Phospho); C22(Carbamidomethyl)	3929.05	7
ITAVTFRGPNDTHLDSLVGQALFGDGSAALIVGSDPIPE		3951.05	6
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTR	T5(Phospho); K7(Methyl); C26(Carbamidomethyl); T36(Phospho)	4260.25	4
GQPKSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVK	T8(Phospho); C13(Carbamidomethyl); T28(Phospho); K38(Acetyl)	4345.18	7
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHD	K7(Methyl); S15(Phospho); C26(Carbamidomethyl); S27(Phospho); K28(Acetyl)	4755.41	5
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHDAE	S15(Phospho); C26(Carbamidomethyl); K28(Methyl); K29(Acetyl)	4875.54	6

2. Middle-down proteomics

2. Cannabis results

Focus on: THCAS

- peptide masses 1-5kDa
- most peptides contain > 2 miscleavages
- 72% sequence coverage (12% BUP)
- Most peptides are modified
- 2/7 NAG ligands found (N329 and N467)
- Other dynamic PTMs: methylations, phosphorylation, acetylation, oxidation



Peptide Sequence	Modifications	MH+ [Da]	# Missed Cleavages
IDVHSQTAW		1056.52	1
IDVHSQTAWVE		1284.63	2
SMGEDLFWAIR		1324.64	4
ILKENPNLCAY	C9(Carbamidomethyl)	1334.68	2
SMGEDLFWAIR	M2(Oxidation)	1340.63	4
KLYEEDVGAGMY		1374.63	5
GGGGENFGIIAAWK		1376.70	3
WGVLFGFGPGLTVE		1478.77	3
KPIPETAMVKILEK	K1(Methyl); M8(Oxidation)	1626.96	3
SKHIPNNVANPKLVY		1693.94	2
ILEKLYEEDVGAGMY		1729.84	6
GMSYISQVPFVVVDLR		1809.96	3
ENPNLCAYEAPSLDAR	C6(Carbamidomethyl)	1819.83	4
GLAADNIIDAHLVNVDGK		1834.98	3
GGIMEEISESAIPFPHR		1869.93	3
IDVHSQTAWVEAGATLGE		1883.92	3
AGGTVLRLAKDLAENNK	N-Term(Acetyl); T4(Phospho); K10(Methyl)	1906.00	4
NRLVKVKTKVDPNNF	N-Term(Acetyl); K5(Acetyl); T8(Phospho)	1935.99	4
YFGKNFNRLVKVKTK	N-Term(Acetyl); K4(Methyl); T14(Phospho)	1978.07	7
NYGLAADNIIDAHLVNVDGK		2112.08	4
LVSAAQTILPDSDGAIDGHLRE		2278.18	4
LGKEAATKAIKEWGQPKSK	T7(Phospho); K11(Methyl); S18(Phospho); K19(Acetyl)	2286.14	7
VKKPIPETAMVKILEKLY	T8(Phospho); K12(Methyl); K16(Acetyl); Y18(Phospho)	2316.19	5
TTIFYSGVVNFNTANFKKE	T1(Phospho); T13(Phospho); K18(Methyl)	2354.07	6
IKLVAVPSKSTIFSVKKNME	S14(Phospho); K16(Methyl); K17(Acetyl); M19(Oxidation)	2371.29	5
	K10(Methyl); N18(NAG)	2440.20	5
		2505.25	4
	S9(Phospho); S15(Phospho); K18(Methyl); M20(Oxidation)	2537.30	6
	C7(Carbamidomethyl)	2554.26	4
	K17(Methyl); K21(Acetyl); T22(Phospho)	2609.35	5
	Y3(Phospho); T11(Phospho); K19(Acetyl); Y21(Phospho)	2773.34	7
		2979.47	8
FR	C18(Carbamidomethyl)	3306.57	6
ESK		3393.79	7
KK	K16(Methyl); T25(Phospho); S28(Phospho); K30(Acetyl)	3418.82	8
	K23(Acetyl); Y29(Phospho)	3462.71	8
MMY	C1(Carbamidomethyl); Y13(Phospho); K17(Methyl); K26(Methyl); Y31(Phospho)	3689.74	6
GY	Y1(Phospho); K3(Methyl); K22(Methyl); Y30(Phospho)	3691.71	9
FPHR		3732.77	9
LDQVESK		3836.95	8
R	K1(Methyl); S8(Phospho); S18(Phospho); Y28(Phospho)	3853.81	9
VLRLAKD	T3(Phospho); C22(Carbamidomethyl)	3929.05	7
ITAVTFRGPNDTHLDSLVGQALFGDGSAALIVGSDPIPE		3951.05	6
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTR	T5(Phospho); K7(Methyl); C26(Carbamidomethyl); T36(Phospho)	4260.25	4
GQPKSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVK	T8(Phospho); C13(Carbamidomethyl); T28(Phospho); K38(Acetyl)	4345.18	7
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHD	K7(Methyl); S15(Phospho); C26(Carbamidomethyl); S27(Phospho); K28(Acetyl)	4755.41	5
ISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRSGGHDAE	S15(Phospho); C26(Carbamidomethyl); K28(Methyl); K29(Acetyl)	4875.54	6

Optimisation of intact protein analysis

Materials
Apical buds
(triplicates)

Protein extraction
TCA/Acetone precipitation
Gnd-HCl /DTT resuspension
IAA alkylation

Protein digestions

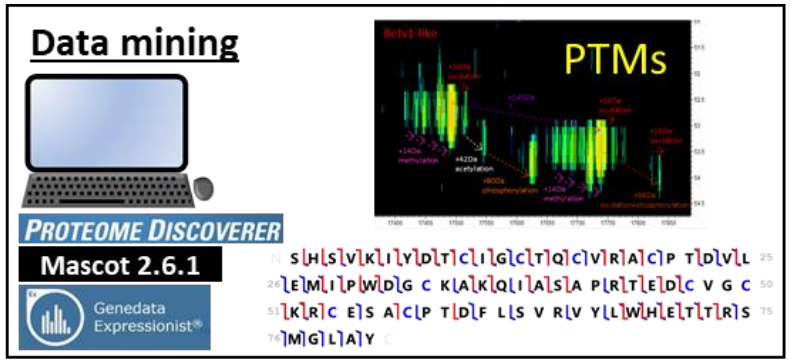
T		G	C
T	+ G =	T:G	
T	+ C =	T:C	
G	+ C =	G:C	
T	+ G + C =	T:G:C	

BUP and MDP
CID

Data mining

TDP
SID CID HCD ETD

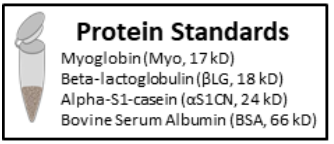
Data mining



3. Top-down proteomics (TDP)

Experimental design

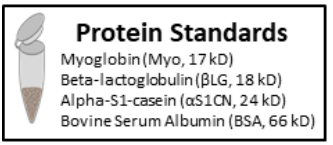
1. Benchmarking MS/MS fragmentation using protein standards



3. Top-down proteomics (TDP)

Experimental design

1. Benchmarking MS/MS fragmentation using protein standards



Protein Standards

Myoglobin (Myo, 17 kD)
Beta-lactoglobulin (βLG, 18 kD)
Alpha-S1-casein (αS1CN, 24 kD)
Bovine Serum Albumin (BSA, 66 kD)

Standard preparation

Steps	Methods/Solutions
1/ Weighing	10 mg
2/ Reconstitution and Reduction	Reconstitution solution (1 mL): 50% ACN 0.1% FA 10mM DTT
3/ Alkylation	20mM IAA (final concentration)
4/ Desalting	SPE C8 gravity column Elution solution (1 mL): 80% ACN 0.1% FA

3. Top-down proteomics (TDP)

Experimental design

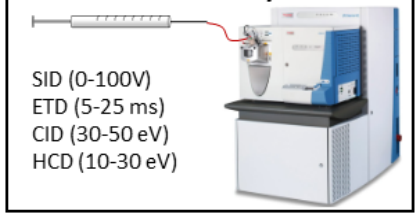
1. Benchmarking MS/MS fragmentation using protein standards

Protein Standards
Myoglobin (Myo, 17 kD)
Beta-lactoglobulin (βLG, 18 kD)
Alpha-S1-casein (αS1CN, 24 kD)
Bovine Serum Albumin (BSA, 66 kD)

Standard preparation

Steps	Methods/Solutions
1/ Weighing	10 mg
2/ Reconstitution and Reduction	Reconstitution solution (1 mL): 50% ACN 0.1% FA 10mM DTT
3/ Alkylation	20mM IAA (final concentration)
4/ Desalting	SPE C8 gravity column Elution solution (1 mL): 80% ACN 0.1% FA

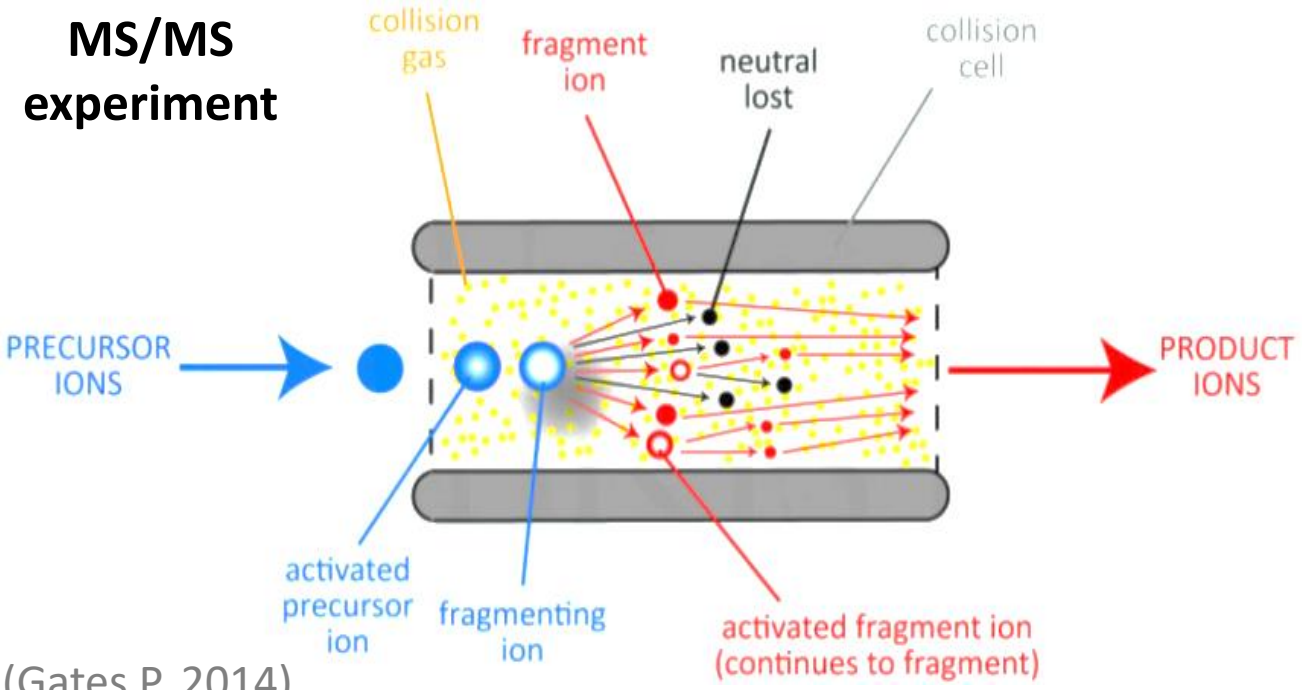
Direct infusion analysis



SID (0-100V)
ETD (5-25 ms)
CID (30-50 eV)
HCD (10-30 eV)

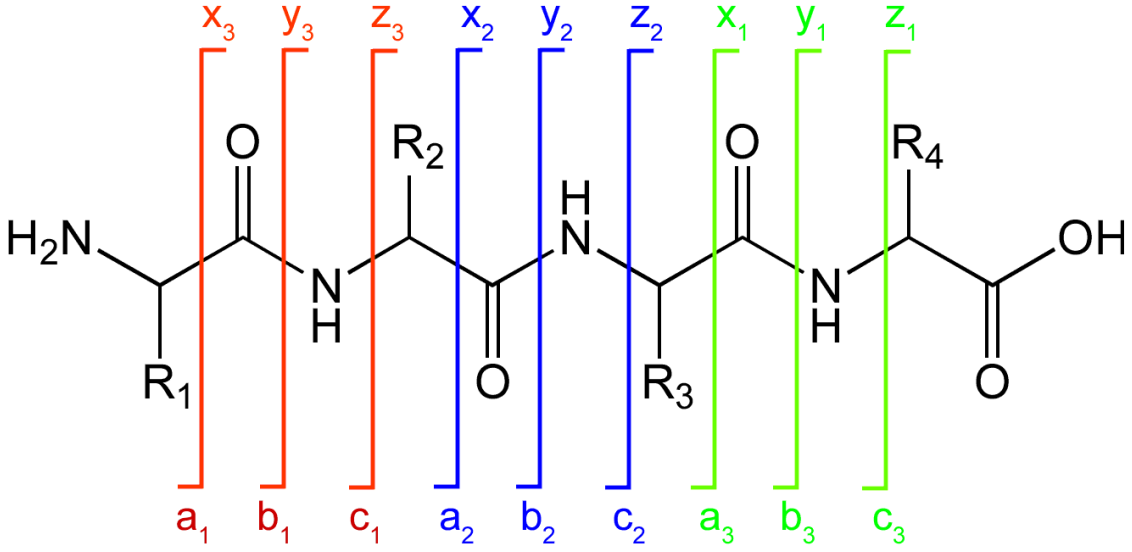
SID, CID, HCD → b-ions and y-ions
ETD → c-ions and z-ions

MS/MS experiment



(Gates P, 2014)

Fragmentation notation



(SOURCE: wikipedia)

3. Top-down proteomics (TDP)

Experimental design

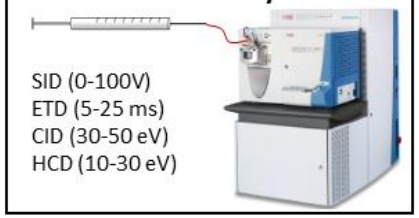
1. Benchmarking MS/MS fragmentation using protein standards

Protein Standards
Myoglobin (Myo, 17 kD)
Beta-lactoglobulin (βLG, 18 kD)
Alpha-S1-casein (αS1CN, 24 kD)
Bovine Serum Albumin (BSA, 66 kD)

Standard preparation

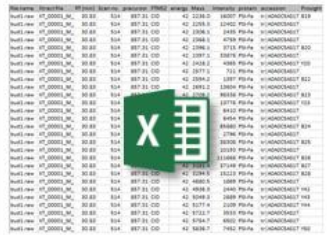
Steps	Methods/Solutions
1/ Weighing	10 mg
2/ Reconstitution and Reduction	Reconstitution solution (1 mL): 50% ACN 0.1% FA 10mM DTT
3/ Alkylation	20mM IAA (final concentration)
4/ Desalting	SPE C8 gravity column Elution solution (1 mL): 80% ACN 0.1% FA

Direct infusion analysis



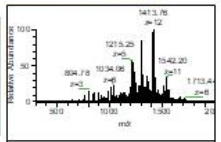
SID (0-100V)
ETD (5-25 ms)
CID (30-50 eV)
HCD (10-30 eV)

Data processing

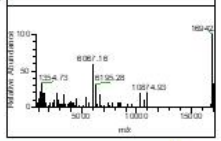


Manual search
SID ETD CID HCD

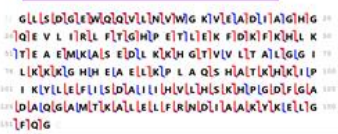
Xcalibur 3.1



Xtract

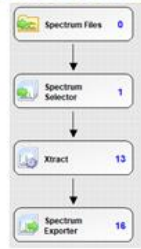


ProSight Lite 1.4



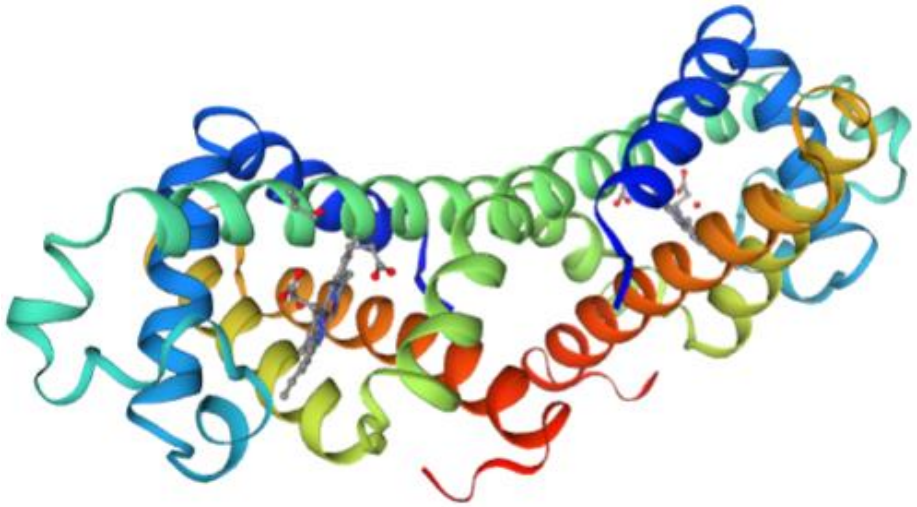
Automatic search
ETD CID HCD

Proteome Discoverer 2.2



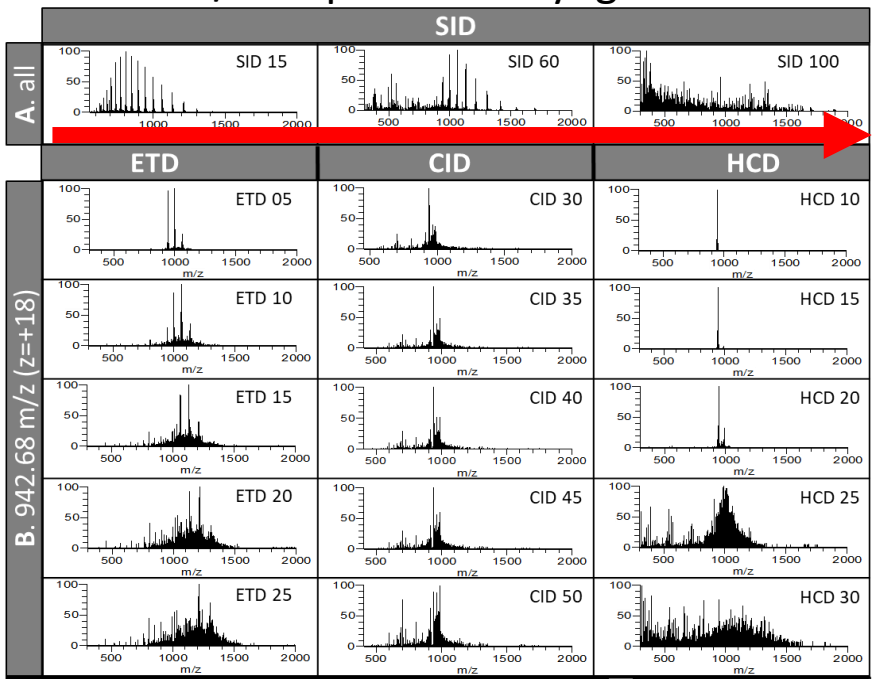
MATRIX SCIENCE Mascot 2.6.1

Example: Myoglobin (P68082 MYG_HORSE, 153 AAs, 17kD)



3. Top-down proteomics (TDP)

MS/MS spectra of myoglobin



more fragments as SID energy increases

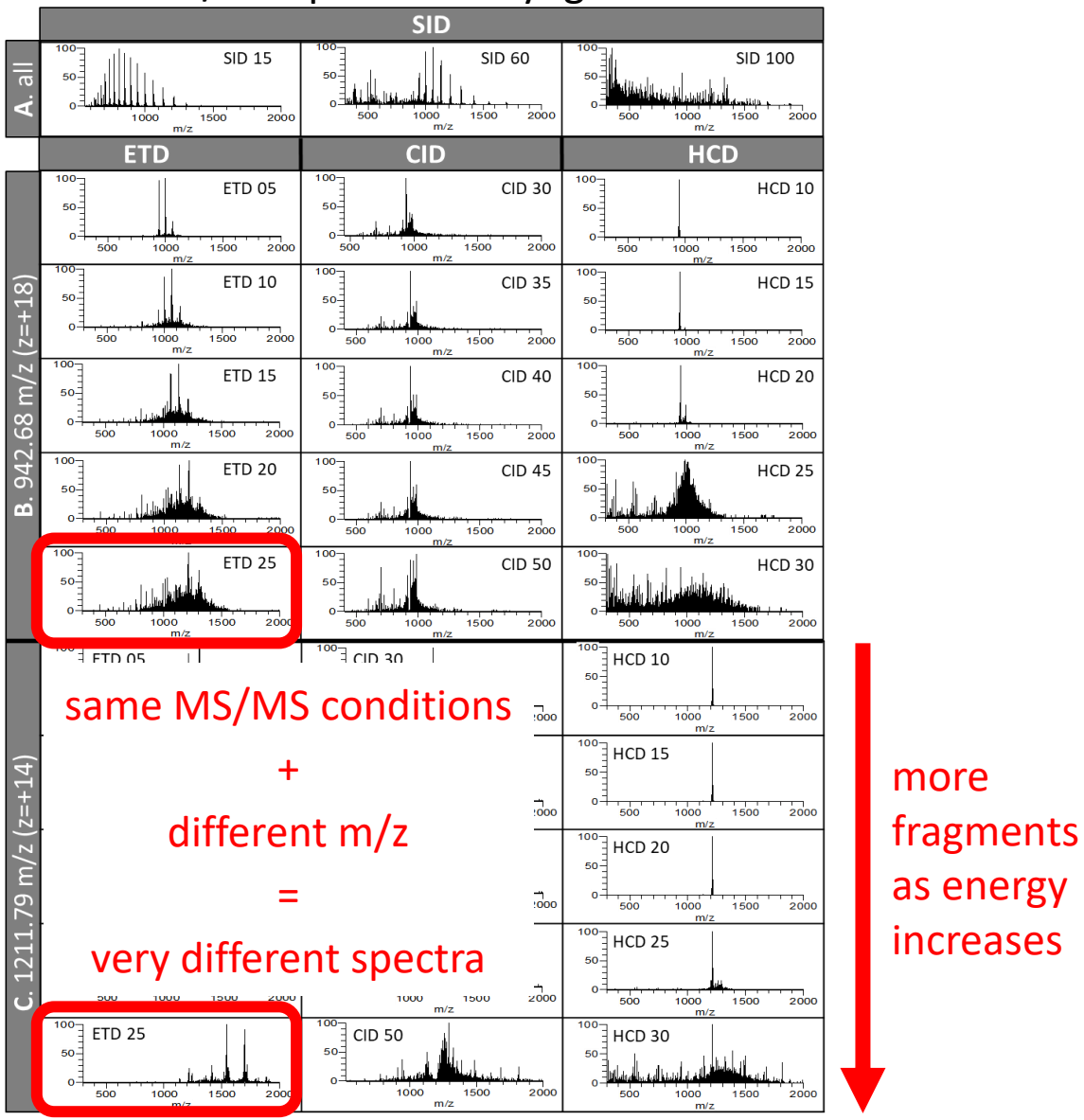
more fragments as ETD, CID, HCD energy increases

- Each MS/MS mode produces different fragmentation patterns in an energy-dependent way.

3. Top-down proteomics (TDP)

- Each MS/MS mode produces different fragmentation patterns in an energy-dependent way.
- Fragmentation efficiency is precursor-dependent.

MS/MS spectra of myoglobin



same MS/MS conditions

+

different m/z

=

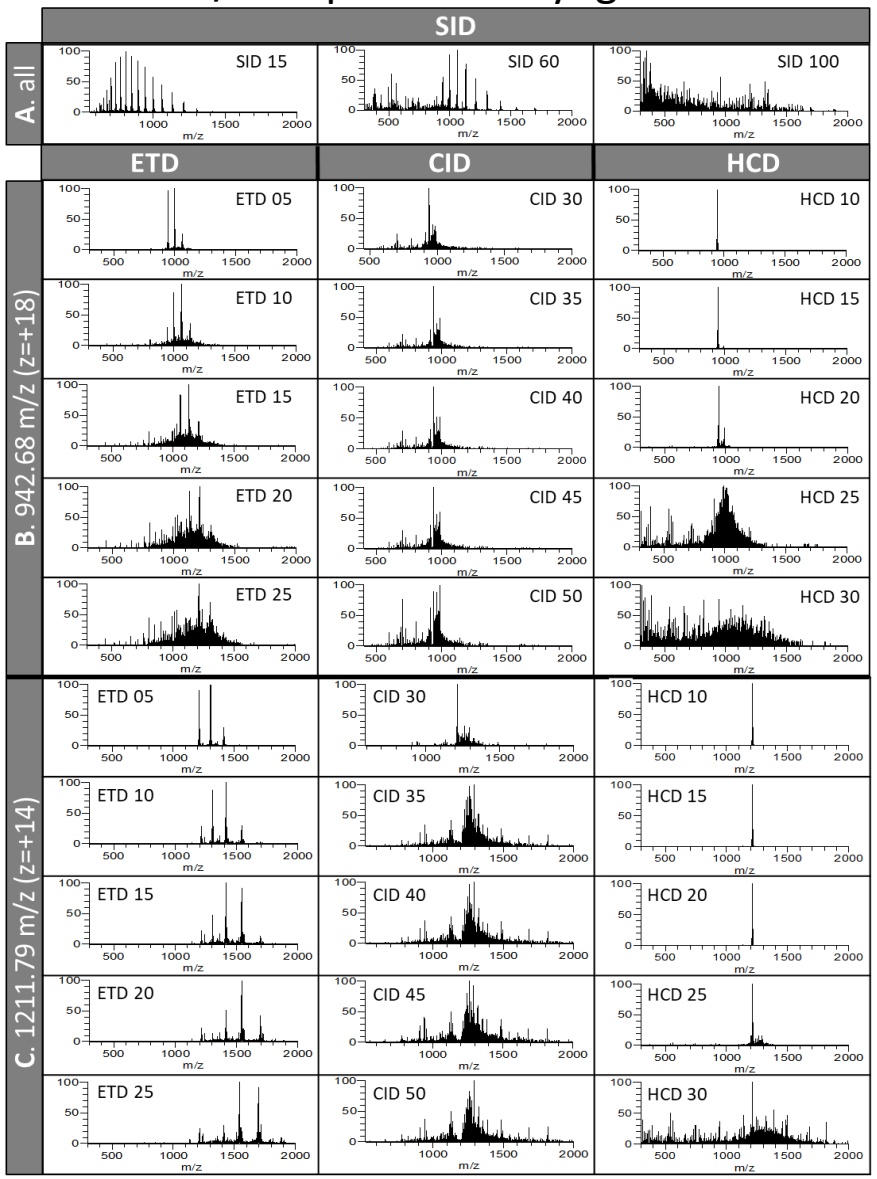
very different spectra

more fragments as energy increases

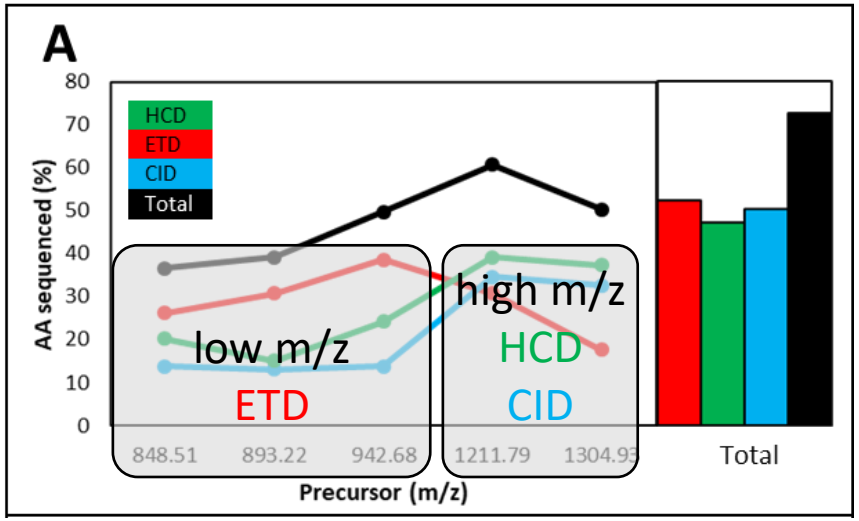
3. Top-down proteomics (TDP)

- Each MS/MS mode produces different fragmentation patterns in an energy-dependent way.
- Fragmentation efficiency is precursor-dependent.

MS/MS spectra of myoglobin



Top-down sequencing of myoglobin

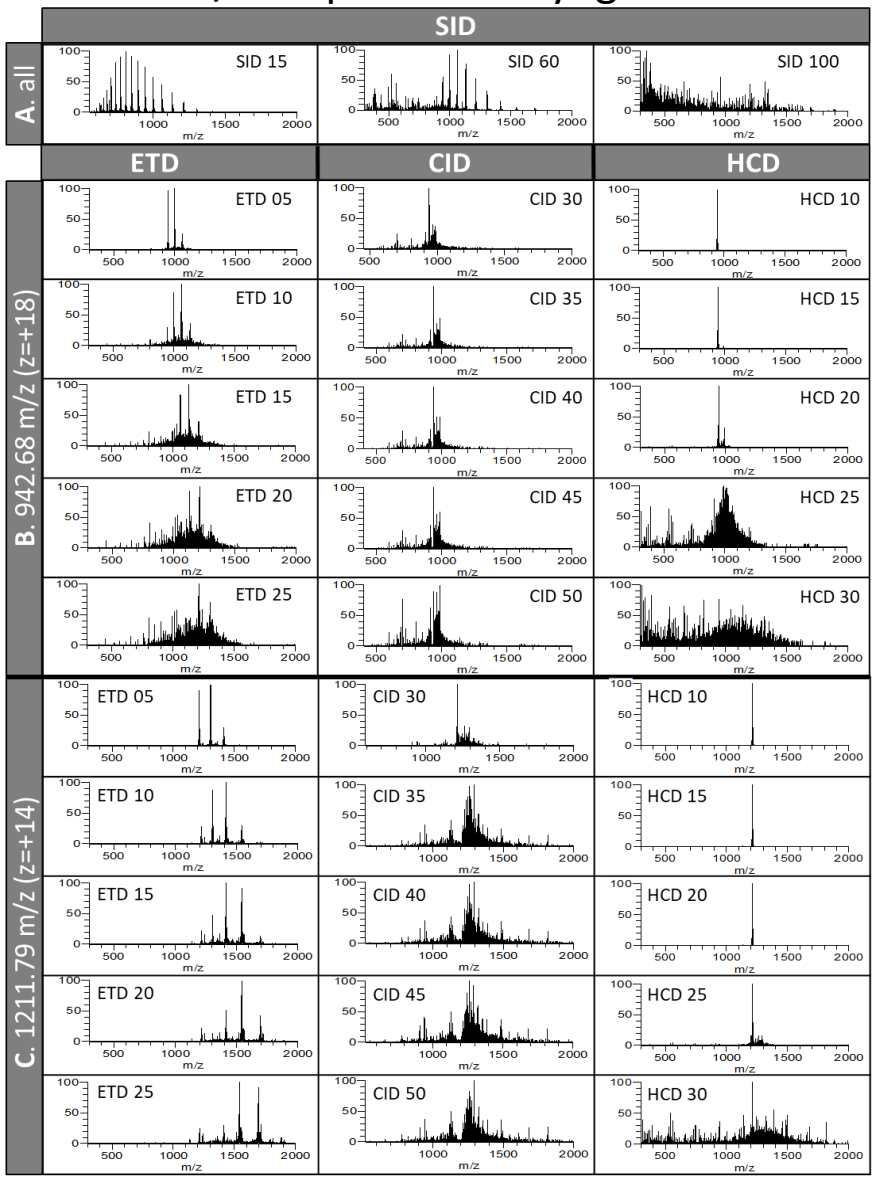


- ETD is complementary to CID and HCD MS/MS modes.

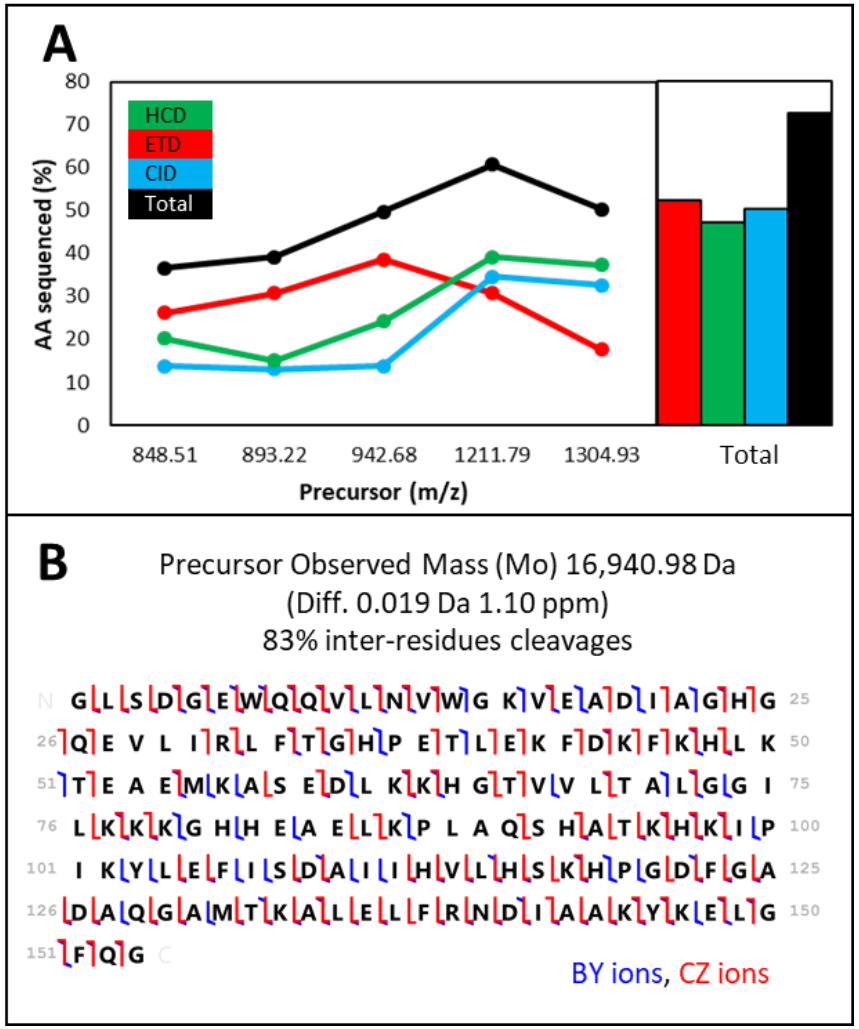
3. Top-down proteomics (TDP)

- Each MS/MS mode produces different fragmentation patterns in an energy-dependent way.
- Fragmentation efficiency is precursor-dependent.

MS/MS spectra of myoglobin



Top-down sequencing of myoglobin

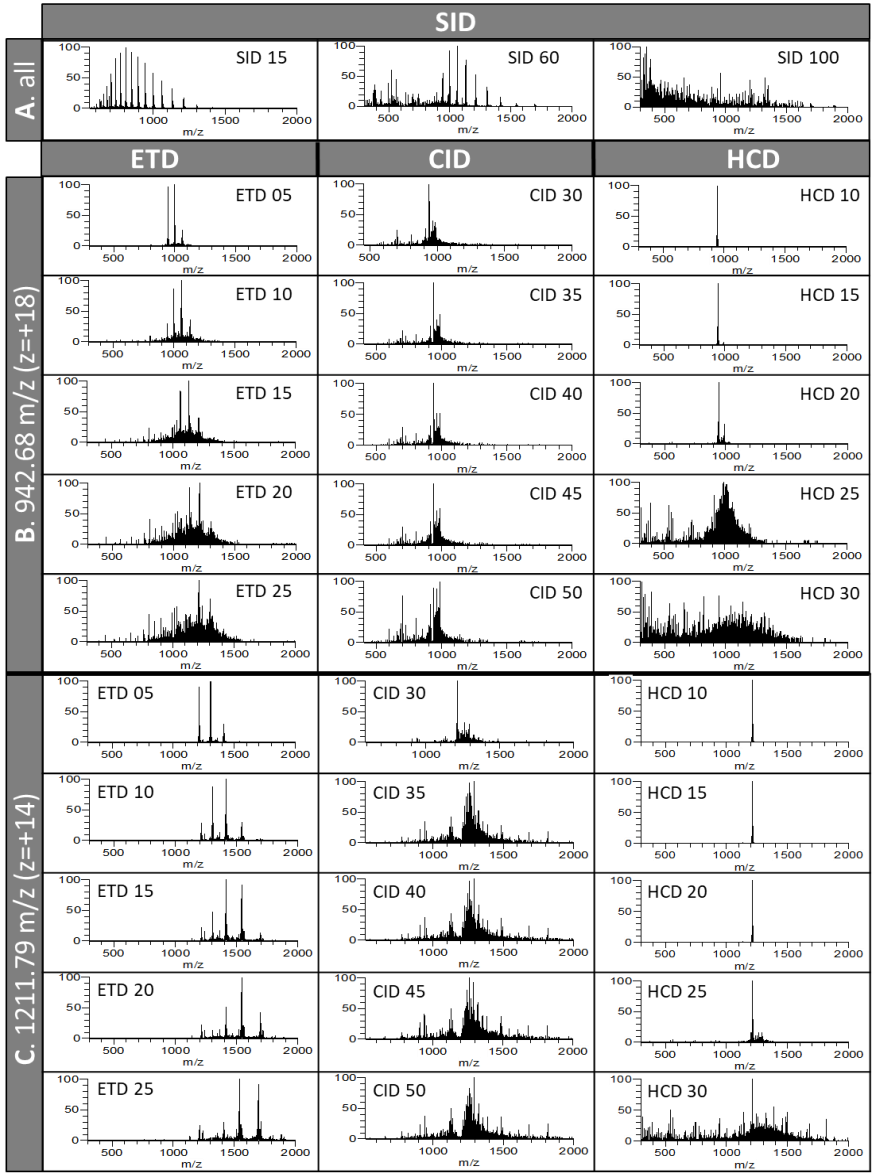


- ETD is complementary to CID and HCD MS/MS modes.
- 83% myoglobin residues fragmented.
- Identified myoglobin fragments give 100% coverage.

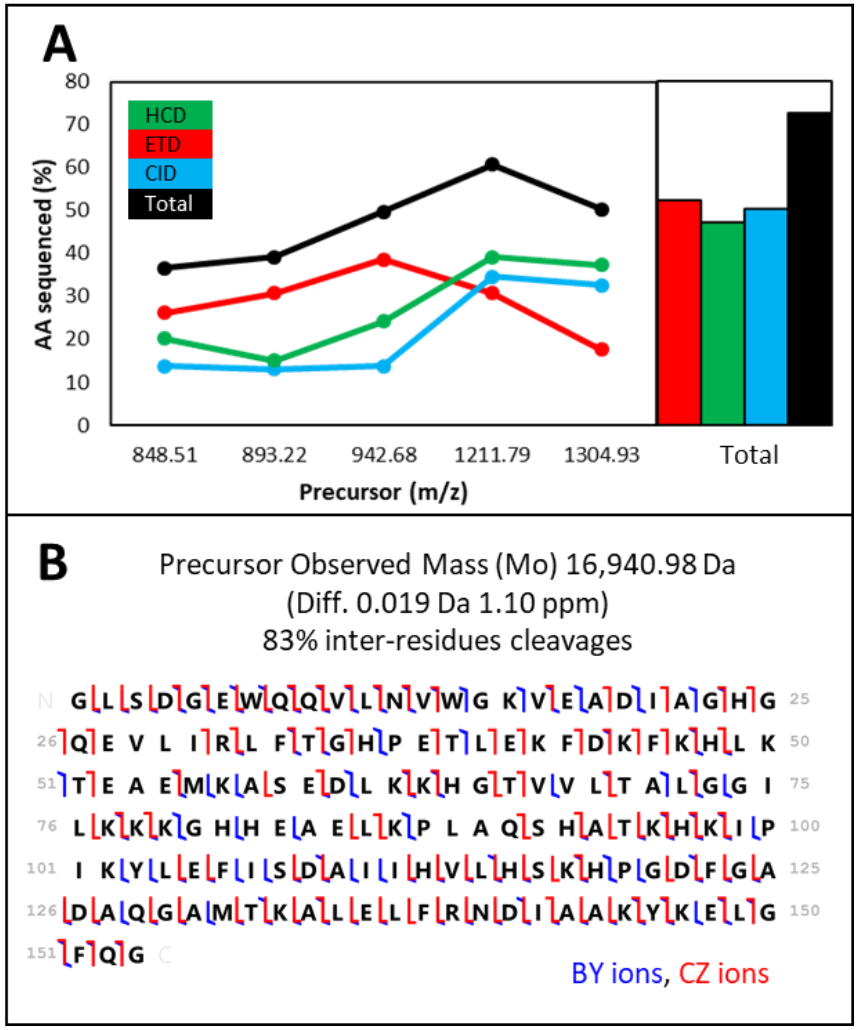
3. Top-down proteomics (TDP)

- Each MS/MS mode produces different fragmentation patterns in an energy-dependent way.
- Fragmentation efficiency is precursor-dependent.

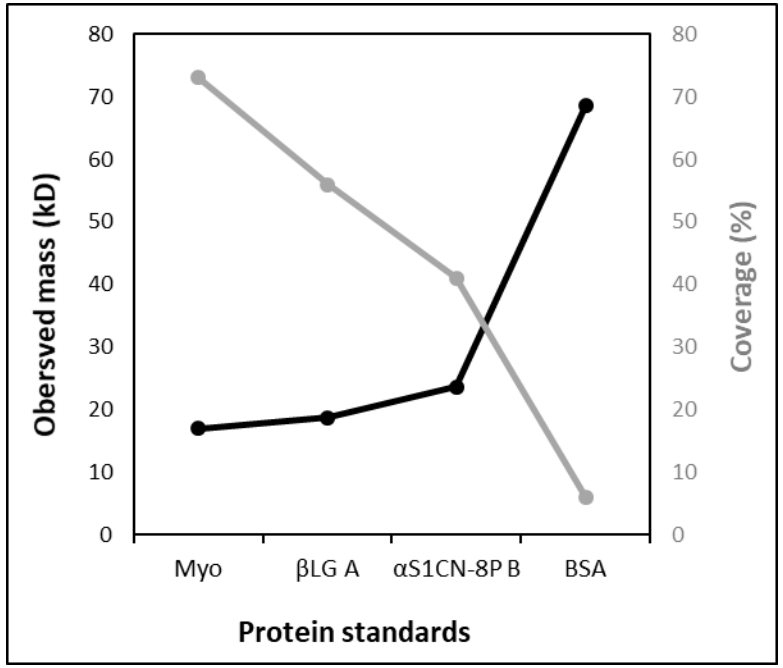
MS/MS spectra of myoglobin



Top-down sequencing of myoglobin



- ETD is complementary to SID, CID and HCD MS/MS modes.
- 83% myoglobin residues fragmented.
- Identified myoglobin fragments give 100% coverage.
- Sequencing success ↓with ↑ MWs.



3. Top-down proteomics (TDP)

Experimental design

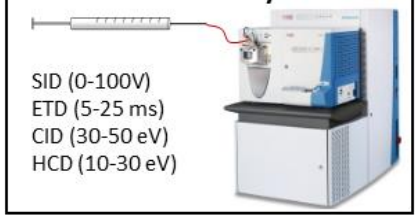
1. Benchmarking MS/MS fragmentation using protein standards

Protein Standards
Myoglobin (Myo, 17 kD)
Beta-lactoglobulin (βLG, 18 kD)
Alpha-S1-casein (αS1CN, 24 kD)
Bovine Serum Albumin (BSA, 66 kD)

Standard preparation

Steps	Methods/Solutions
1/ Weighing	10 mg
2/ Reconstitution and Reduction	Reconstitution solution (1 mL): 50% ACN 0.1% FA 10mM DTT
3/ Alkylation	20mM IAA (final concentration)
4/ Desalting	SPE C8 gravity column Elution solution (1 mL): 80% ACN 0.1% FA

Direct infusion analysis



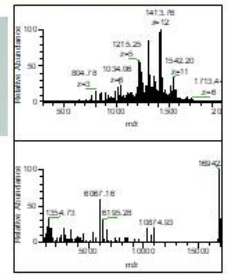
SID (0-100V)
ETD (5-25 ms)
CID (30-50 eV)
HCD (10-30 eV)

Data processing

Manual search
SID ETD CID HCD

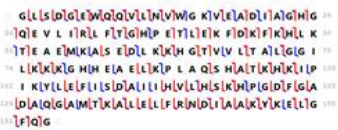
Xcalibur 3.1

Qual Browser



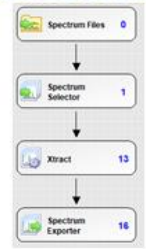
Xtract

ProSight Lite 1.4



Automatic search
ETD CID HCD

Proteome Discoverer 2.2



MATRIX SCIENCE Mascot 2.6.1

2. Apply best MS/MS methods to cannabis extracts

Cannabis mature buds
(triplicates)

Protein extraction

Steps	Methods/Solutions
1/ Grinding	Liquid N2 in mortar & pestle
2/ Precipitation (O/N)	TCA/DTT/A solution: 10% TCA 10mM DTT Acetone
3/ Washing (x2)	DTT/A solution: 10mM DTT Acetone
4/ Drying	Supernatant discarded Pellet air-dried
5/ Resuspension	Gnd-HCl buffer: 6M Gdn-HCl, 10mM DTT, 5.4 mM Na-citrate tribasic $2H_2O$, 0.1 M Bis-Tris.
6/ Protein assay	BCA assay
7/ Dilution	Gnd-HCl buffer added so that all concentrations match
7/ Alkylation	20mM IAA (final concentration)
8/ Desalting	SPE C8 gravity column Elution solution (1 mL): 80% ACN 0.1% FA
9/ Evaporation	SpeedVac Evaporator (90 min)

3. Top-down proteomics (TDP)

Experimental design

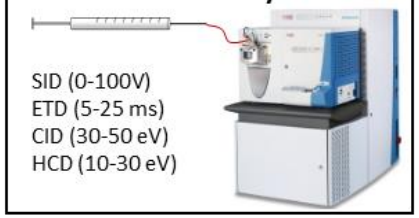
1. Benchmarking MS/MS fragmentation using protein standards

Protein Standards
Myoglobin (Myo, 17 kD)
Beta-lactoglobulin (βLG, 18 kD)
Alpha-S1-casein (αS1CN, 24 kD)
Bovine Serum Albumin (BSA, 66 kD)

Standard preparation

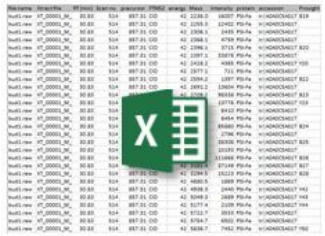
Steps	Methods/Solutions
1/ Weighing	10 mg
2/ Reconstitution and Reduction	Reconstitution solution (1 mL): 50% ACN 0.1% FA 10mM DTT
3/ Alkylation	20mM IAA (final concentration)
4/ Desalting	SPE C8 gravity column Elution solution (1 mL): 80% ACN 0.1% FA

Direct infusion analysis

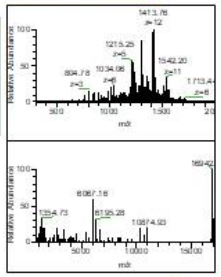


SID (0-100V)
ETD (5-25 ms)
CID (30-50 eV)
HCD (10-30 eV)

Data processing

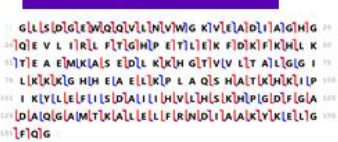


Manual search
SID ETD CID HCD
Xcalibur 3.1

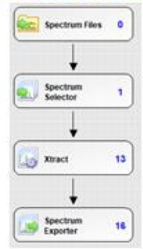


Xtract

ProSight Lite 1.4



Automatic search
ETD CID HCD
Proteome Discoverer 2.2



MATRIX SCIENCE Mascot 2.6.1

2. Apply best MS/MS methods to cannabis extracts

Cannabis mature buds
(triplicates)

Protein extraction

Steps	Methods/Solutions
1/ Grinding	Liquid N2 in mortar & pestle
2/ Precipitation (O/N)	TCA/DTT/A solution: 10% TCA 10mM DTT Acetone
3/ Washing (x2)	DTT/A solution: 10mM DTT Acetone
4/ Drying	Supernatant discarded Pellet air-dried
5/ Resuspension	Gnd-HCl buffer: 6M Gdn-HCl, 10mM DTT, 5.4 mM Na-citrate tribasic 2H_2O, 0.1 M Bis-Tris.
6/ Protein assay	BCA assay
7/ Dilution	Gnd-HCl buffer added so that all concentrations match
7/ Alkylation	20mM IAA (final concentration)
8/ Desalting	SPE C8 gravity column Elution solution (1 mL): 80% ACN 0.1% FA
9/ Evaporation	SpeedVac Evaporator (90 min)

LC-MS and LC-MS/MS analyses

Intact protein analysis
5 injections per bud
UPLC-MS (duplicates)
UPLC-MS/MS (3 methods)

	ETD (ms)	CID (eV)	HCD (eV)
Low	5	35	19
Mid	10	42	23
High	15	50	27

3. Top-down proteomics (TDP)

Experimental design

1. Benchmarking MS/MS fragmentation using protein standards

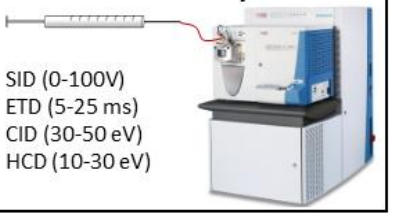
Protein Standards
Myoglobin (Myo, 17 kD)
Beta-lactoglobulin (βLG, 18 kD)
Alpha-S1-casein (αS1CN, 24 kD)
Bovine Serum Albumin (BSA, 66 kD)

Standard preparation

Steps	Methods/Solutions
1/ Weighing	10 mg
2/ Reconstitution and Reduction	Reconstitution solution (1 mL): 50% ACN 0.1% FA 10mM DTT
3/ Alkylation	20mM IAA (final concentration)
4/ Desalting	SPE C8 gravity column Elution solution (1 mL): 80% ACN 0.1% FA

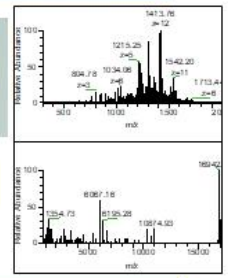
Direct infusion analysis

SID (0-100V)
ETD (5-25 ms)
CID (30-50 eV)
HCD (10-30 eV)

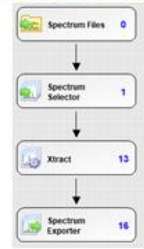


Data processing

Manual search
SID ETD CID HCD
Xcalibur 3.1
Qual Browser
Xtract
ProSight Lite 1.4



Automatic search
ETD CID HCD
Proteome Discoverer 2.2



MATRIX SCIENCE Mascot 2.6.1

2. Apply best MS/MS methods to cannabis extracts

Cannabis mature buds (triplicates)

Protein extraction

Steps	Methods/Solutions
1/ Grinding	Liquid N2 in mortar & pestle
2/ Precipitation (O/N)	TCA/DTT/A solution: 10% TCA 10mM DTT Acetone
3/ Washing (x2)	DTT/A solution: 10mM DTT Acetone
4/ Drying	Supernatant discarded Pellet air-dried
5/ Resuspension	Gnd-HCl buffer: 6M Gdn-HCl, 10mM DTT, 5.4 mM Na-citrate tribasic 2H_2O, 0.1 M Bis-Tris.
6/ Protein assay	BCA assay
7/ Dilution	Gnd-HCl buffer added so that all concentrations match
7/ Alkylation	20mM IAA (final concentration)
8/ Desalting	SPE C8 gravity column Elution solution (1 mL): 80% ACN 0.1% FA
9/ Evaporation	SpeedVac Evaporator (90 min)

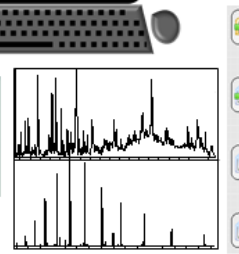
LC-MS and LC-MS/MS analyses

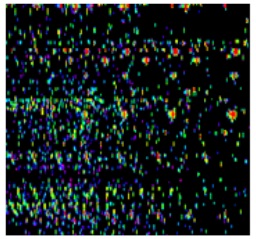
Intact protein analysis
5 injections per bud
UPLC-MS (duplicates)
UPLC-MS/MS (3 methods)

	ETD (ms)	CID (eV)	HCD (eV)
Low	5	35	19
Mid	10	42	23
High	15	50	27

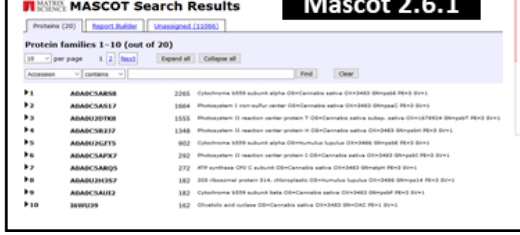
Data processing

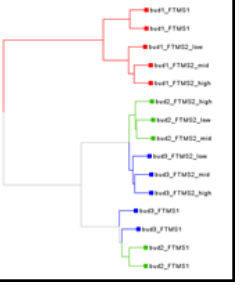
Proteome Discoverer 2.2
Genedata Expressionist®
Qual Browser



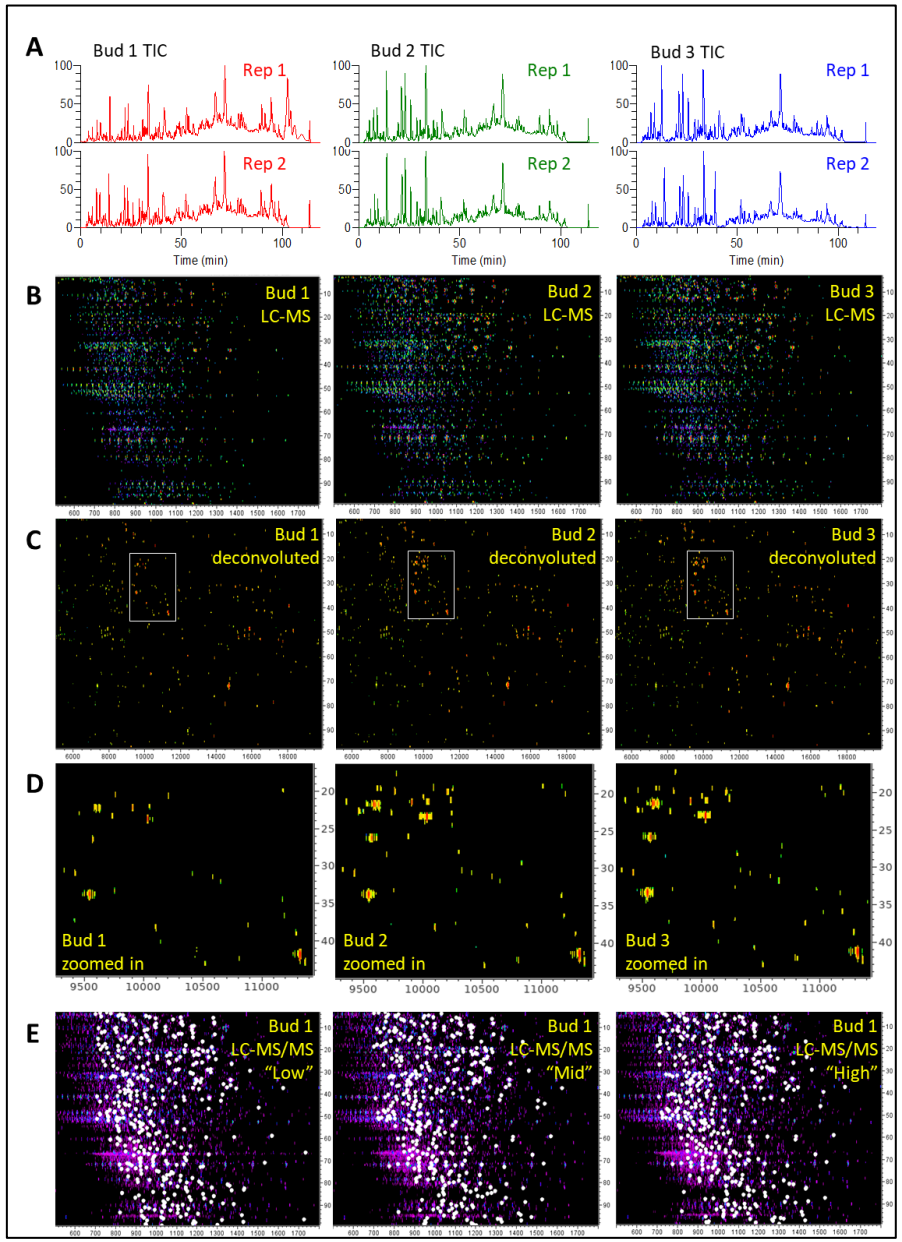


MASCOT Search Results
Mascot 2.6.1





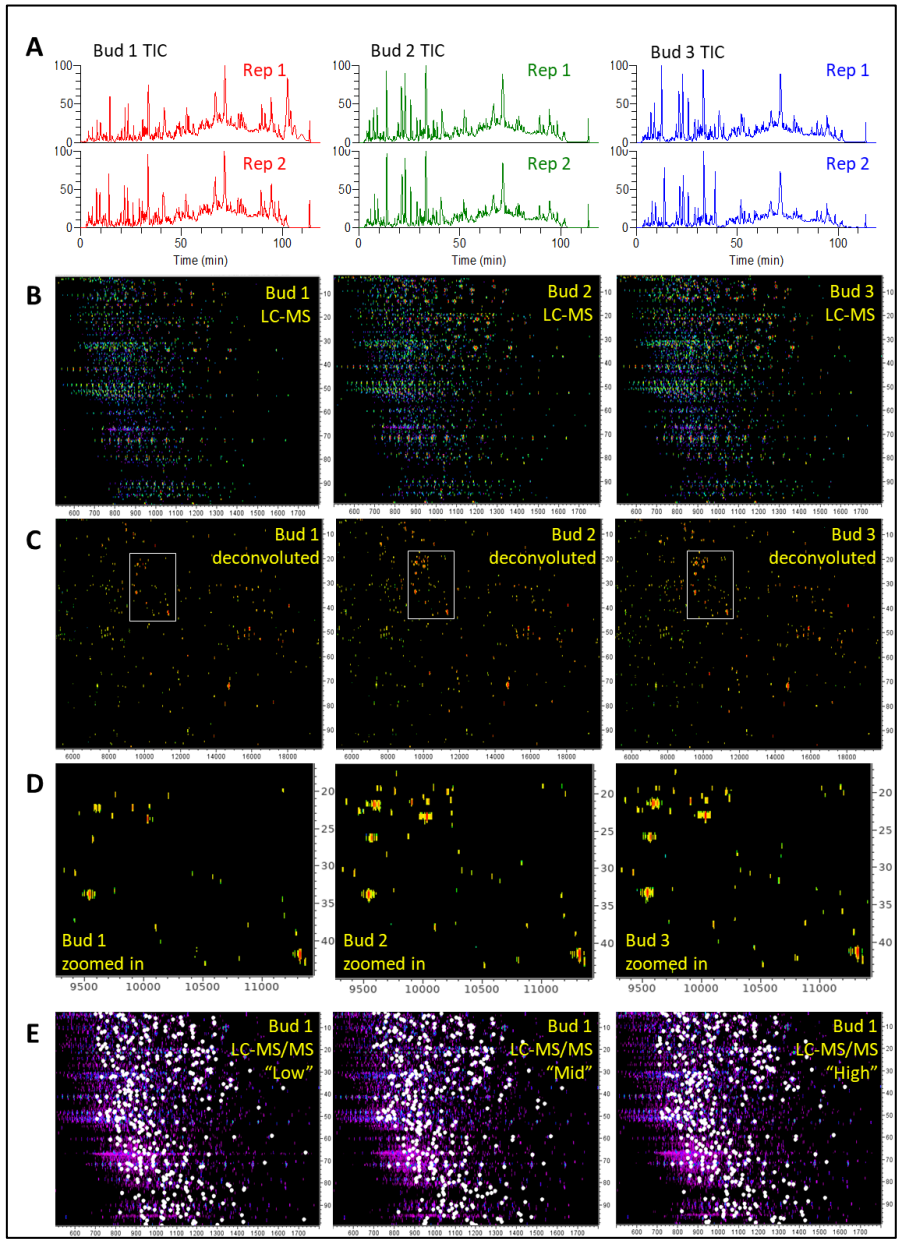
3. Top-down proteomics (TDP)

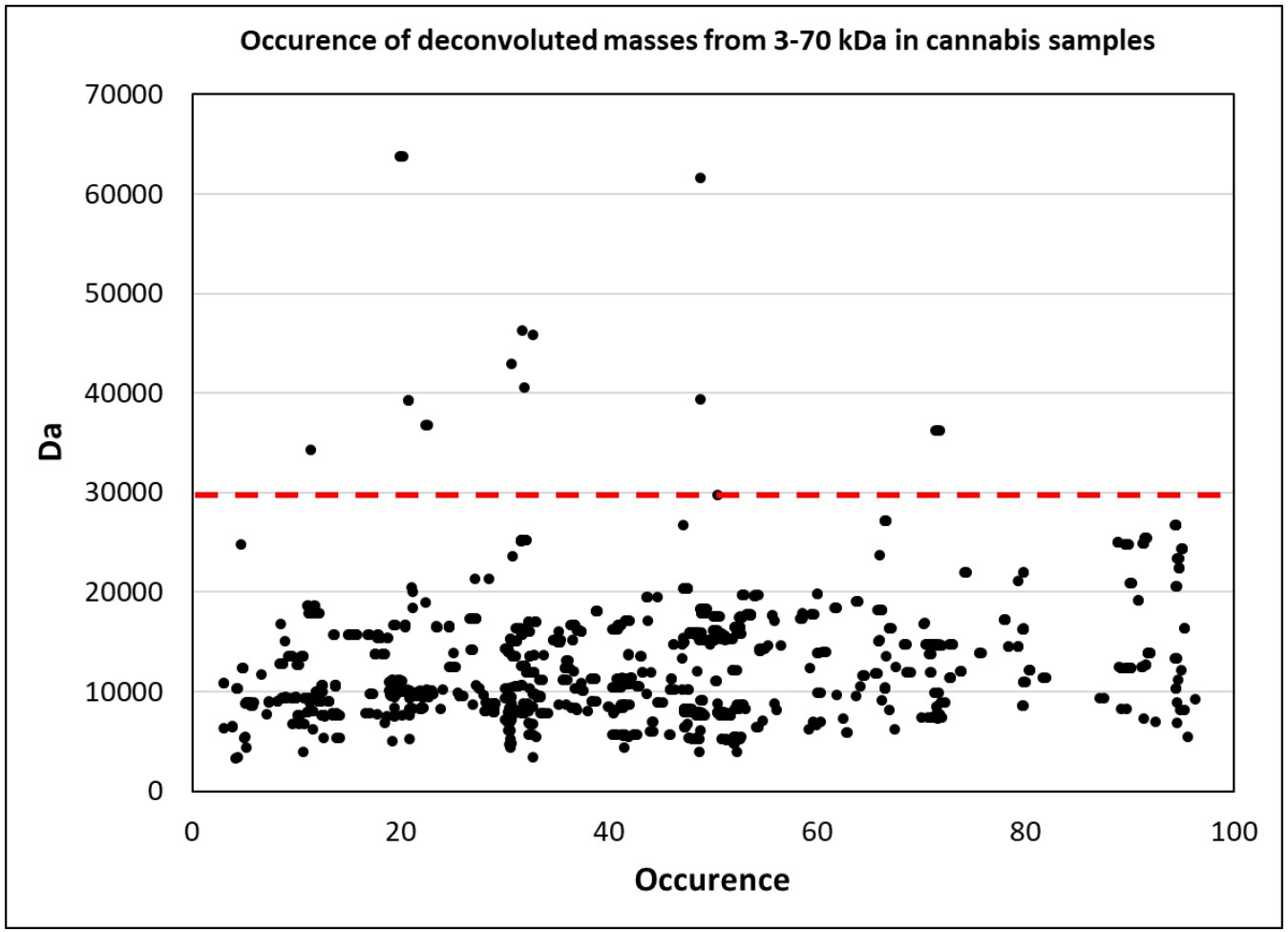


- Excellent reproducibility

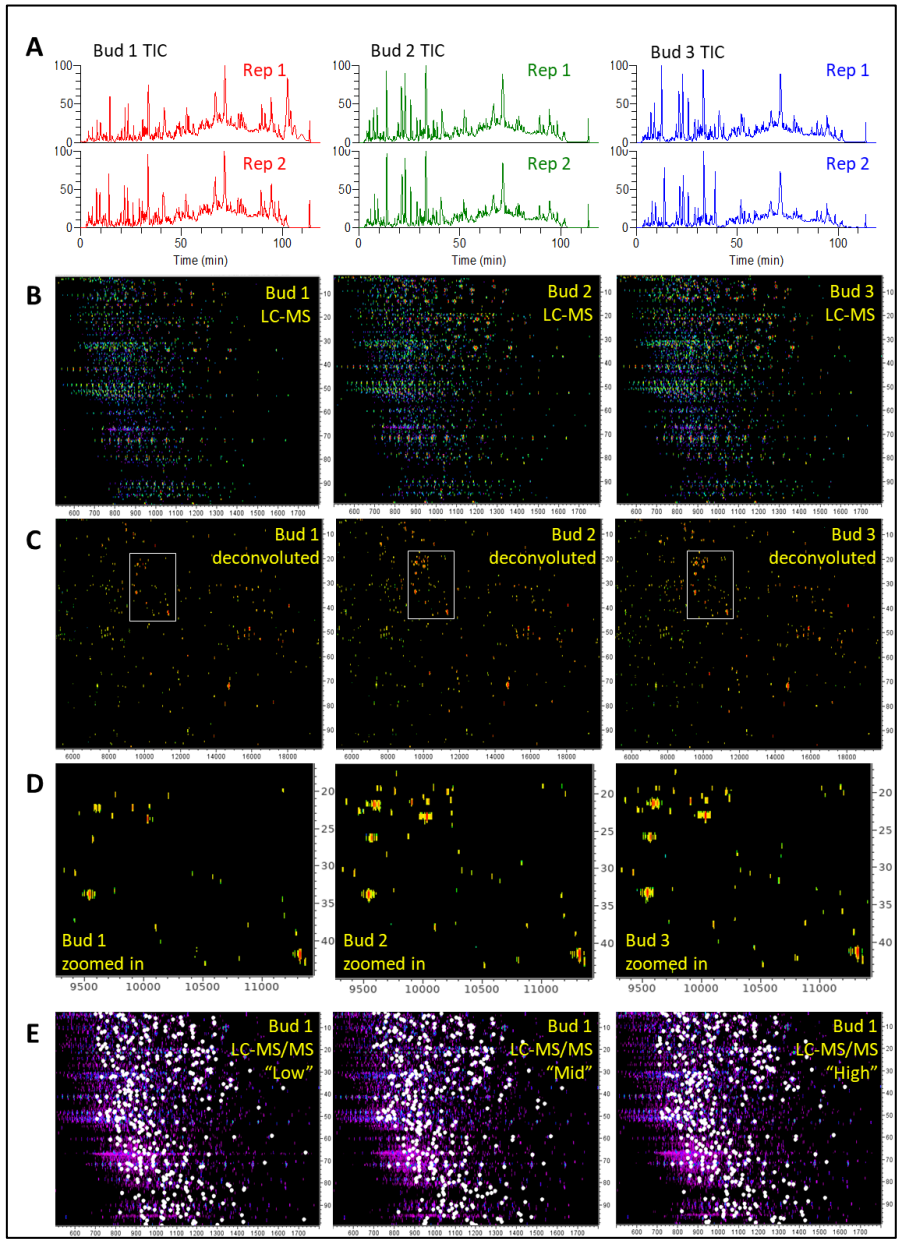
3. Top-down proteomics (TDP)

- Excellent reproducibility
- Most cannabis proteins <30kDa

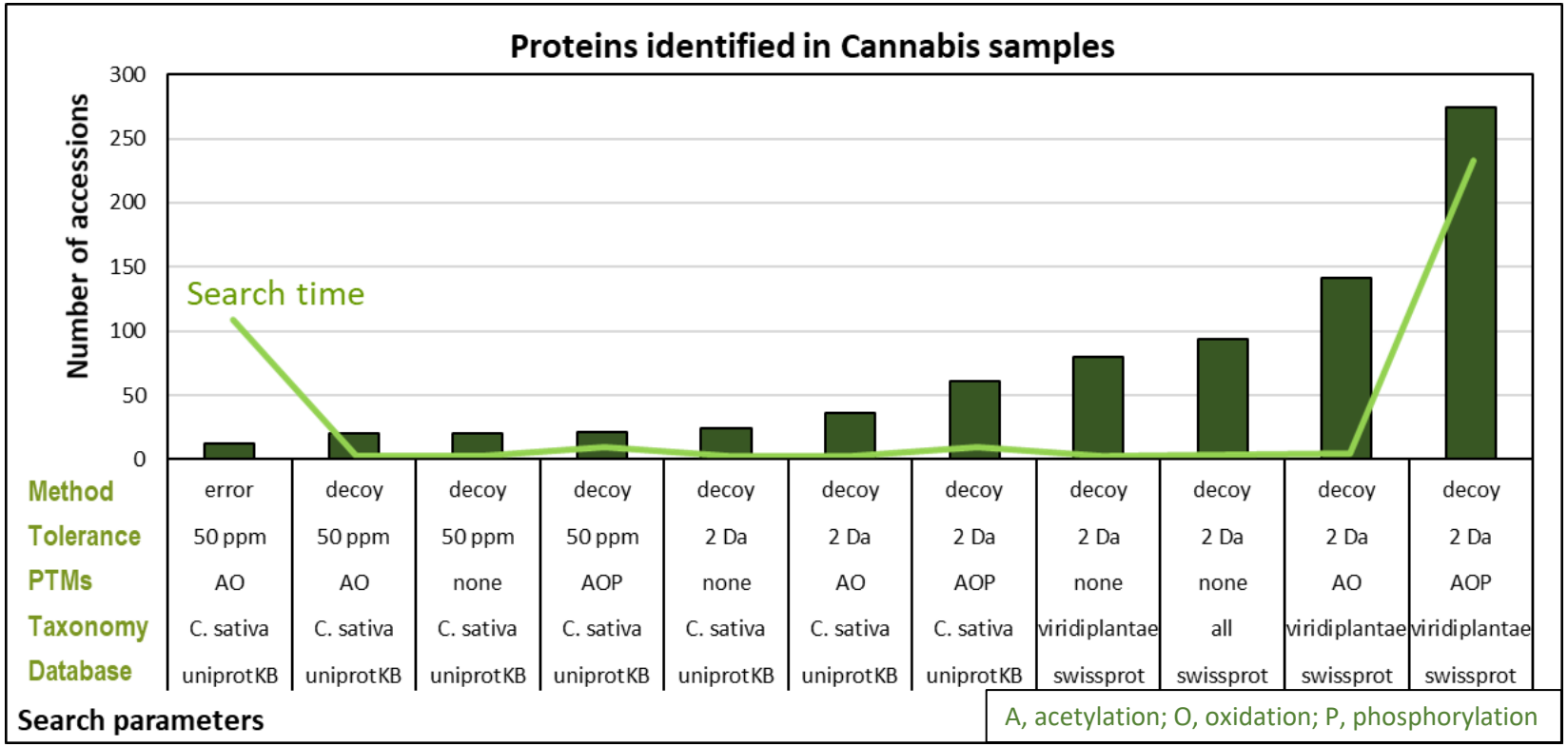




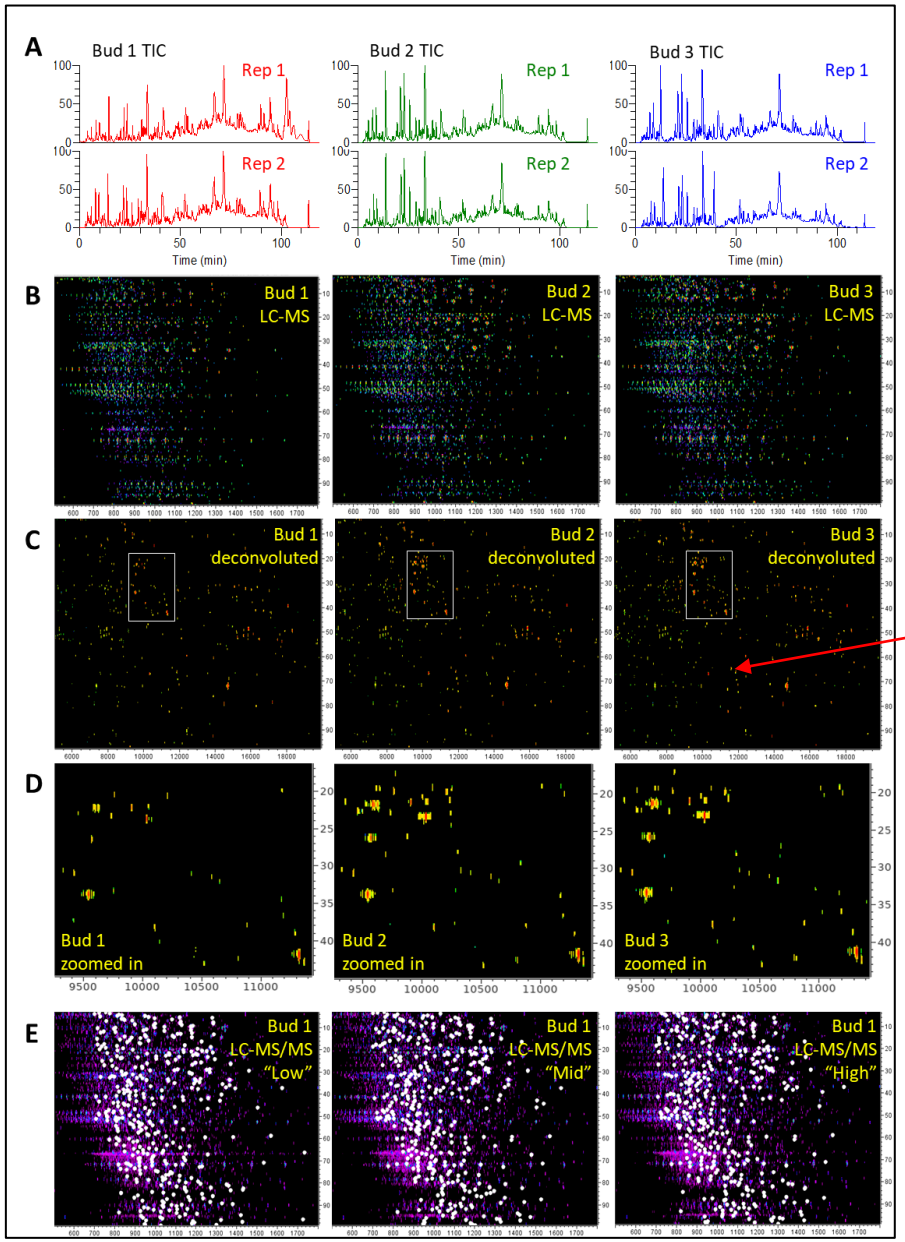
3. Top-down proteomics (TDP)

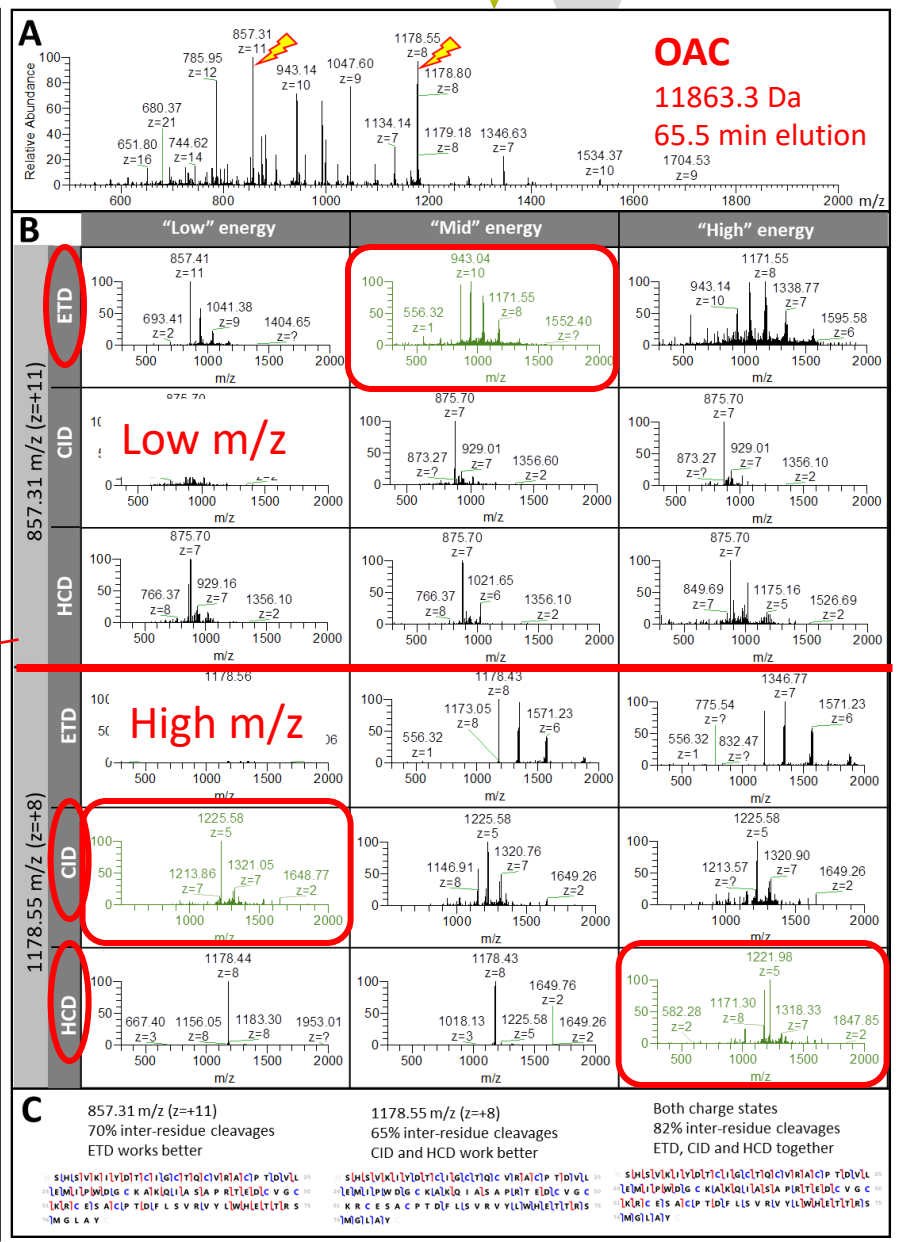


- Excellent reproducibility
- Most cannabis proteins <30kDa
- 11,250 MS/MS spectra
- 213-1863 (2-17%) matched in Mascot
- Search time spanned min to weeks depending on the DB and parameters (PTMs)!



3. Top-down proteomics (TDP)





Focus on OAC:

Elutes at 64-66 min (40%ACN)

Low m/z → ETD

High m/z → HCD, CID

3. Top-down proteomics (TDP)

Focus on OAC:
Elutes at 64-66 min (40%ACN)
Low m/z → ETD
High m/z → HCD, CID

MASCOT Search Results

Protein View: I6WU39|OLIAC_CANSA

Olivetolic acid cyclase OS=Cannabis sativa OX=3483 GN=OAC PE=1 SV=1

Database:	OAC_I6WU39
Score:	270
Monoisotopic mass (M_r):	11994
Calculated pI:	5.77

Sequence similarity is available as an NCBI BLAST search of I6WU39|OLIAC_CANSA against nr.

Search parameters

MS data file:	2019-05-04_MC_buds123_export-mgf_1x.mgf
Enzyme:	NoCleave: cuts C-term side of J unless next residue is ABCDEFGHIJKLMNOPQRSTUVWXYZ.
Variable modifications:	Methyl (K), Acetyl (K), Acetyl (Protein N-term), Acetyl (N-term), Oxidation (M), Phospho (ST), Phospho (Y)

Protein sequence coverage: 99%

Matched peptides shown in ***bold red***.

```
  1 MAVKHLIVLK FKDEITEAQK EEFFKTYVNL VNIIPAMKDV YWGKDVTQKN
 51 KEEGYTHIVE VTFESVETIQ DYIIHPAHVG FGDVYRSFWE KLLIFDYTPR
101 K
```

Unformatted sequence string: 101 residues (for pasting into other applications).

Sort by ◉ residue number ○ increasing mass ○ decreasing mass
Show ◉ matched peptides only ○ predicted peptides also

Query	Start – End	Observed	Mr(expt)	Mr(calc)	Δ	M	Score	Expect	Rank	U	Peptide
6704	2 – 101	11866.3070	11865.2997	11863.1629	0.0180	0	20	0.04	1	U	M.AVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK.-
6795	2 – 101	11910.3060	11909.2987	11905.1735	0.0347	0	54	4.1e-05	1	U	M.AVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK.- + Acetyl (K)
6702	2 – 101	11866.3070	11865.2997	11943.1292	-0.6517	0	48	5.4e-05	1	U	M.AVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK.- + Phospho (ST)
6703	2 – 101	11866.3070	11865.2997	11943.1292	-0.6517	0	45	0.00013	1	U	M.AVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK.- + Phospho (ST)
6725	2 – 101	11869.2875	11868.2802	11943.1292	-0.6267	0	60	4.1e-06	1	U	M.AVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK.- + Phospho (ST)
6729	2 – 101	11869.2935	11868.2862	11943.1292	-0.6267	0	25	0.014	1	U	M.AVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK.- + Phospho (ST)
6732	2 – 101	11869.3080	11868.3007	11943.1292	-0.6265	0	39	0.00055	1	U	M.AVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK.- + Phospho (ST)
6733	2 – 101	11869.3080	11868.3007	11943.1292	-0.6265	0	31	0.0034	1	U	M.AVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK.- + Phospho (ST)
6735	2 – 101	11870.2102	11869.2030	11943.1292	-0.6190	0	38	0.00069	1	U	M.AVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK.- + Phospho (ST)
6723	2 – 101	11869.2875	11868.2802	11957.1449	-0.7432	0	21	0.035	1	U	M.AVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK.- + Methyl (K); Phospho (ST)
6727	2 – 101	11869.2935	11868.2862	11957.1449	-0.7431	0	41	0.00033	1	U	M.AVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK.- + Methyl (K); Phospho (ST)
6728	2 – 101	11869.2935	11868.2862	11957.1449	-0.7431	0	29	0.0058	1	U	M.AVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK.- + Methyl (K); Phospho (ST)
6713	2 – 101	11868.1750	11867.1677	11961.1997	-0.7861	0	26	0.011	1	U	M.AVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK.- + Methyl (K); 2 Acetyl (K)
6831	2 – 101	11995.3102	11994.3029	12001.1711	-0.0572	0	24	0.0061	1	U	M.AVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK.- + 3 Methyl (K); Oxidation (M); Phospho (ST)
6830	2 – 101	11983.3118	11982.3045	12079.1218	-0.8015	0	42	0.0022	1	U	M.AVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK.- + Methyl (K); Acetyl (K); Phospho (ST); Phospho (Y)

OAC 100% coverage (N-Term protein M excised) + PTMs

>sp|I6WU39|OLIAC_CANSA Olivetolic acid cyclase OS=Cannabis sativa OX=3483 GN=OAC PE=1 SV=1
MAVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQKNKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK

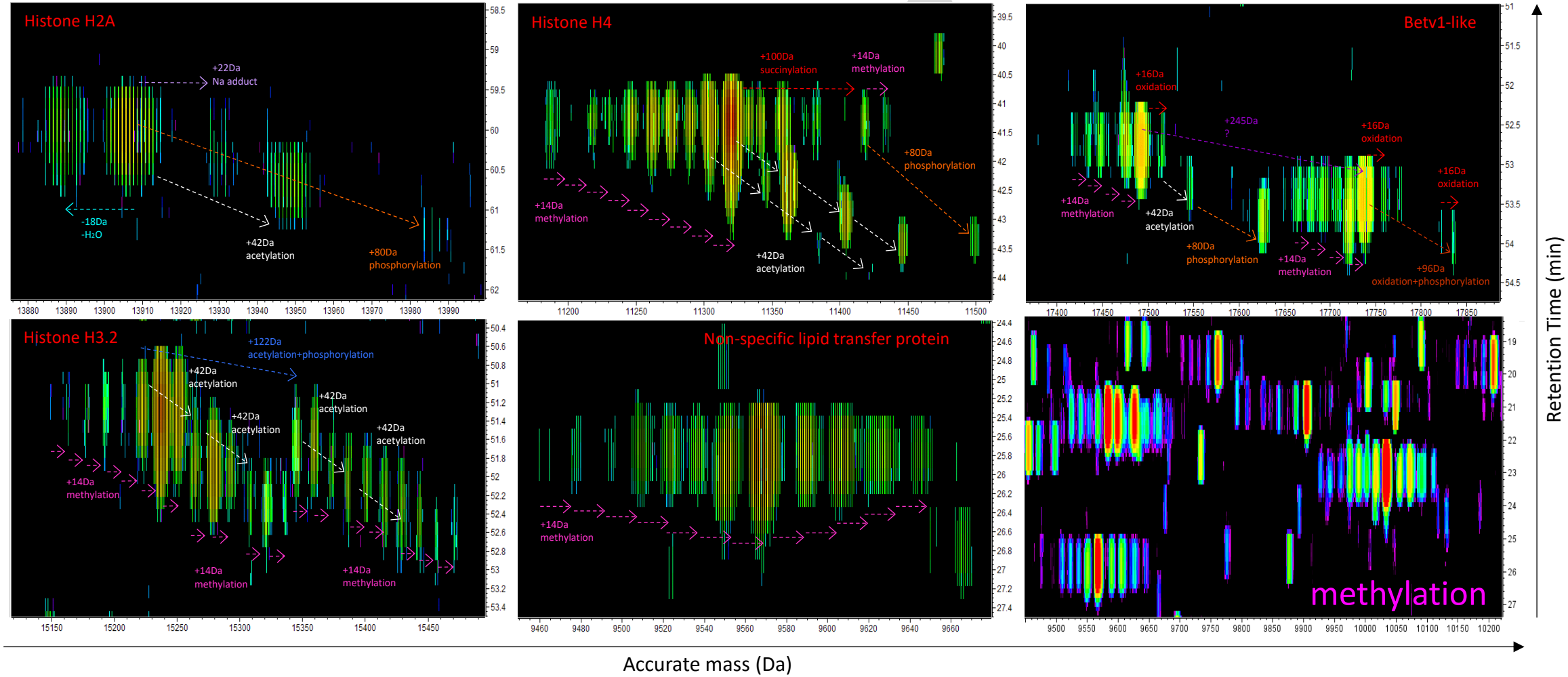
N-term acetyl: K4, K10, K12, K20, K25 methylated; M37 K38, K44 T47 K49,K51 acetylated; T56, T62,S65,T68,Y72 phoshorylated; S87, T98

N-term M excision of OAC and PTMs not referenced in UniprotKB

OAC proteoforms

3. Top-down proteomics (TDP)

Examples of PTMs discovered by TDP.

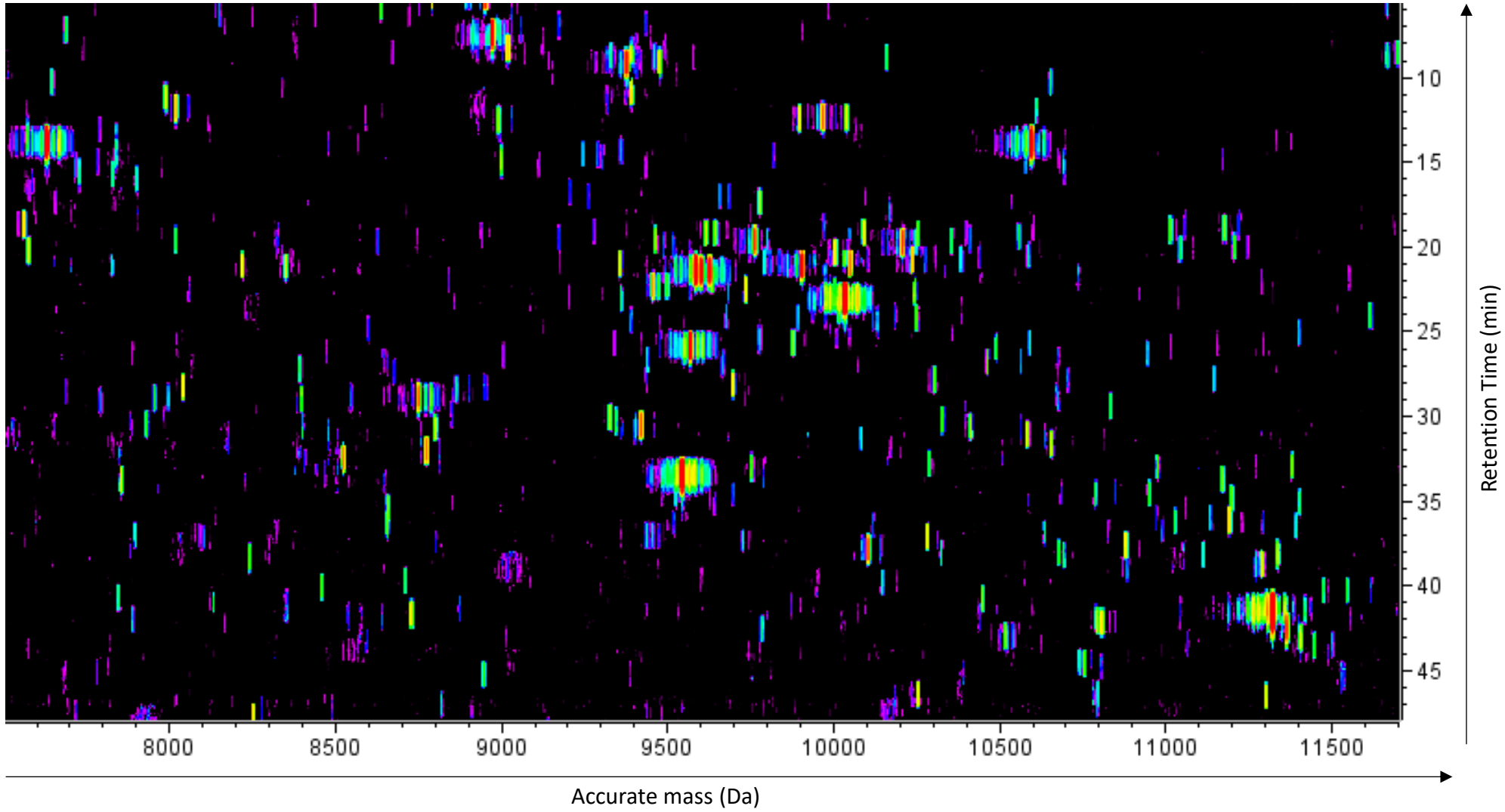


46 cannabis proteins identified, 136 proteoforms with different PTMs
(N-term M excision, N-term acetylation, methylation, acetylation, phosphorylation).

3. Top-down proteomics (TDP)

Protein methylation.

Cannabis proteins heavily methylated.



K methylation of histones determines chromatin structure and affect gene expression.

K methylation of non-histone proteins enables protein-protein interactions.

Conclusions

- **BUP**: Robust qualitative method that gives access to phytocannabinoid enzymes [1].

- **MDP**: Similar to BUP with greater sequence coverage and more proteins identified [2].

- **TDP**: Quantitative method but mostly unknown features. Proteoforms of identified proteins [3].

Highly complementary strategies.

Validates genome annotation (proteogenomics).

Future directions

Apply each of these methods to various cultivars of medicinal cannabis and hemp.

	BUP[1]	MDP[2]	TDP[3]
speed	-	--	++
costs	+	++	--
tools	+	+	-
identities	+	++	--
quantitative	-	-	++
proteoforms	-	+	+++

PUBLICATIONS

1. Vincent, D.; Rochfort, S.; Spangenberg, G. Optimisation of protein extraction from medicinal cannabis mature buds for bottom-up proteomics. Molecules 2019, 24, 659. doi:10.3390/molecules24040659.

2. Vincent, D.; Ezernieks, V.; Rochfort, S.; Spangenberg, G. A multiple protease strategy to optimise shotgun proteomics of medicinal cannabis mature buds. International Journal of Molecular Sciences 2019, 20, 5630; doi:10.3390/ijms20225630.

3. Vincent, D.; Binos, S.; Rochfort, S.; Spangenberg, G. Top-down proteomics of medicinal cannabis. Proteomes 2019, 7, 33. doi:10.3390/proteomes7040033.


Thank you!