Top-down, middle-down and bottom-up proteomics analysis of medicinal cannabis

D. Vincent[1], S. Rochfort[1,2], G. Spangenberg[1,2]

[1] Agriculture Victoria Research, AgriBio, Centre for AgriBioscience, Bundoora, Victoria 3083, Australia
[2] School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia

ABSTRACT

Cannabis sativa has been dubbed “the plant of the thousand and one molecules” owing to its propensity to produce a plethora of phytochemicals with myriad biological activities as well as fibrous components. Out of the 500 compounds that have been described, more than 90 are phytocannabinoids accumulating within the trichomes of mature buds. Cannabis is a controlled substance and until recently was illegal in many jurisdictions. The recent revised legislation on medicinal cannabis has triggered a surge of medical and clinical research studies evaluating the effect of major cannabis components on human health.

The state of Victoria in Australia was the first jurisdiction to legalise access to medicinal cannabis under the Medicinal Cannabis Act in 2016. In this context, Agriculture Victoria Research (AVR) have controlled access to medicinal cannabis material grown to full maturity in the state-of-the-art Victorian government medicinal cannabis cultivation facility. A comprehensive systems biology approach including genome sequencing, transcriptomics, proteomics and metabolomics was undertaken by AVR to fully characterise the various medicinal cannabis cultivars developed in house.

We have developed three complementary proteomics strategies (**Fig. 1**), namely bottom-up proteomics (BUP) through the use of the most commonly used protease trypsin[1], middle-down proteomics (MDP) by exploiting alternative orthogonal proteases[2], and top-down proteomics (TDP) which required innovative data mining methods and revealed numerous proteoforms[3].

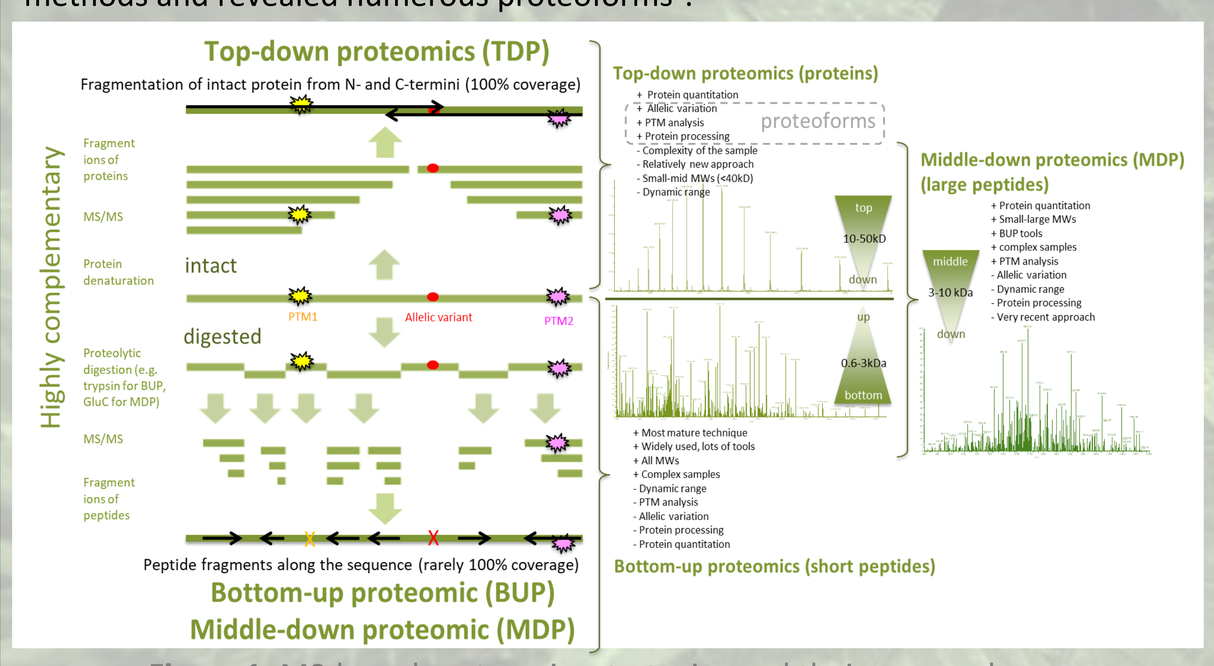


Figure 1: MS-based proteomics strategies and their pros and cons.

METHODS

1) Protein extraction and digestion

Protein extraction was optimised[1]; the best method was a TCA/acetone precipitation followed by resuspension in a Guanidine-HCl/DTT buffer. Extracted proteins were analysed in their intact reduced state by TDP[3] and further digested using three orthogonal proteases (trypsin/LysC [T], GluC [G], chymotrypsin [C]) on their own or in combination[2]. Peptide digests were analysed by BUP[1] and MDP[2].

2) LC-MS/MS analysis of intact proteins and digested peptides

Intact proteins were separated by RP-UPLC using a C8 column for 120 min and peptide digests were separated by RP-nLC using a C18 column for 60 min prior to MS/MS analyses using a LTQ-orbitrap mass analyser with SID, CID, HCD and ETD fragmentation modes for intact proteins and CID fragmentation for peptide digests.

3) Data mining

Raw MS/MS files were processed using Genedata Expressionist as well as Proteome Discoverer and Mascot algorithm with a *C. sativa* fasta sequence database from Uniprot and NCBI. Statistical analyses (PCA, PLS, HCA, heat map) and PTM assignment were performed using Genedata Expressionist (**Fig. 2**).

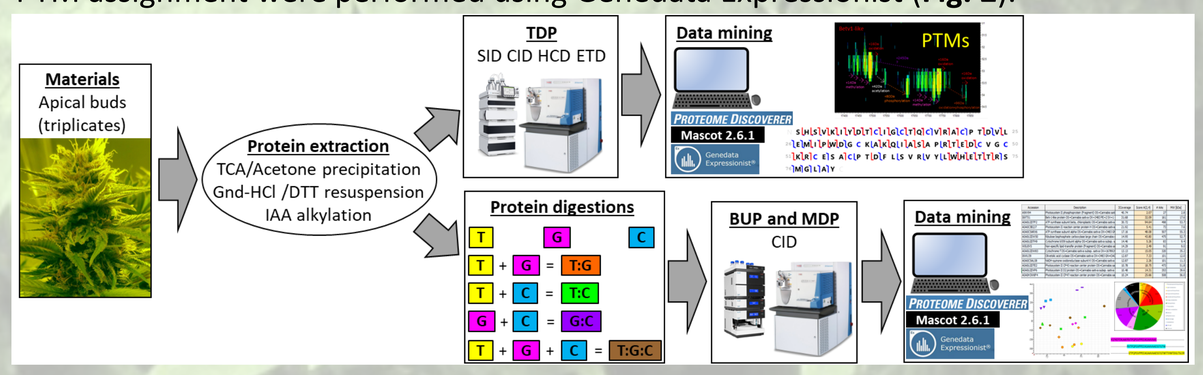
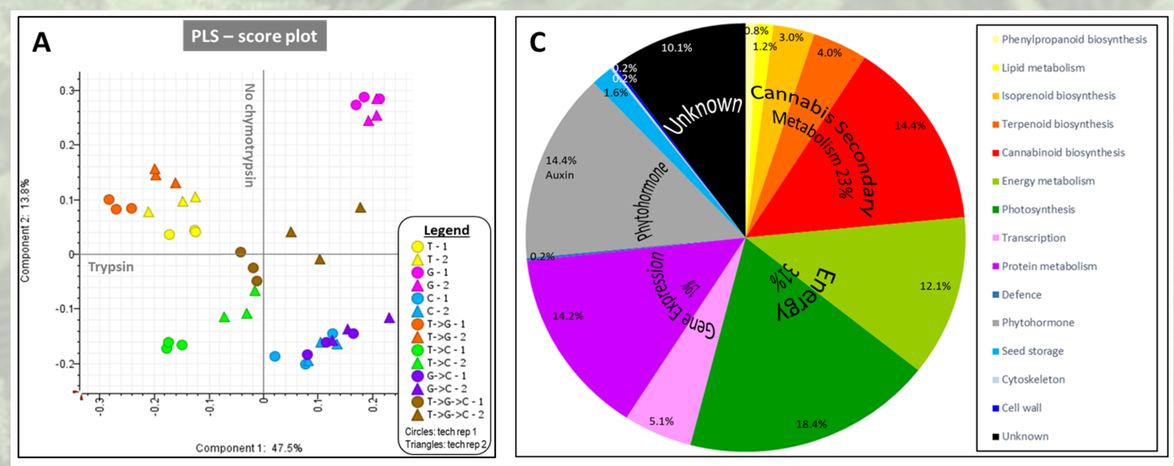


Figure 2: Experimental design.

RESULTS

1) Shotgun proteomics (BUP and MDP)

BUP and MDP led to the identifications of 160 and 494 accessions, respectively. Digestion with orthogonal proteases produced distinct patterns (**Fig. 3A**). A greater number of identities were obtained with chymotrypsin on its own or in combination with the other proteases (**Fig. 3B**). Bud proteins are involved in energy (31%) and secondary (23%) metabolisms, including phytocannabinoid pathway (14%), as well as gene expression (19%) (**Fig. 3C**). MDP helped achieved not only more identities but also with greater sequence coverage (**Fig. 3D**).



D Coverage (%)	BUP[1]	MDP[2]
0-10 %	20	1
11-20 %	32	5
21-30 %	7	3
31-40 %	5	3
41-50 %	4	2
51-60 %	2	1
61-70 %	0	0
71-80 %	0	3
81-90 %	0	6
91-100 %	30	76

Figure 3: Shotgun (BUP and MDP) proteomics results.

2) Top-down proteomics (TDP)

TDP yielded 46 accessions and 136 proteoforms bearing various PTMs including the excision of N-terminal methionine, N-terminal acetylation, methylation and acetylation of lysine resides, and phosphorylation.

Examples of proteoforms are displayed in **Figure 4**.

Most identified proteins are involved in photosynthesis, translation, and ATP production.

Only one protein belongs to the phytocannabinoid biosynthesis pathway, olivetolic acid cyclase.

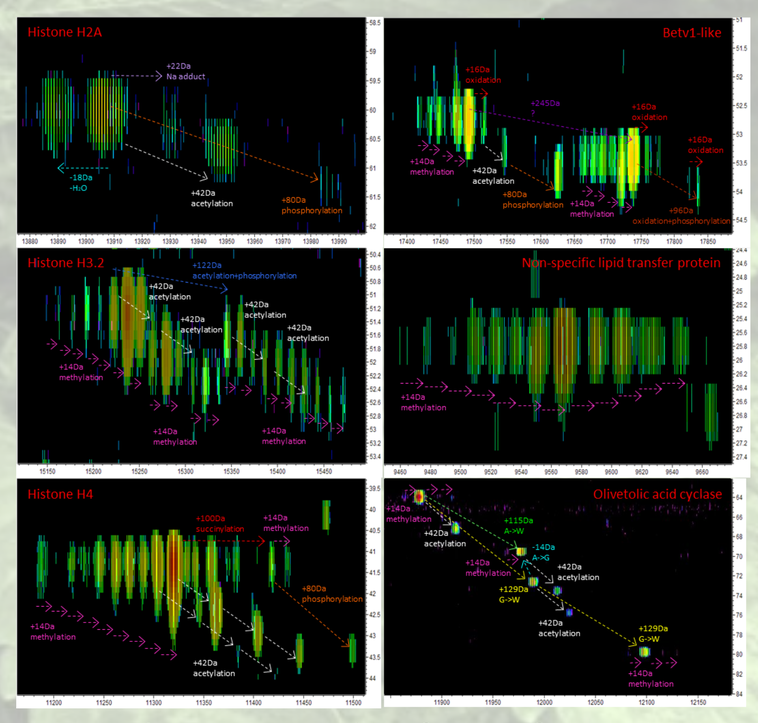


Figure 4: Examples of proteoforms.

CONCLUSIONS

The three methods developed here are highly complementary (**Fig. 5**) and all together offer a more comprehensive view of the cannabis proteome. These techniques will be applied to various medicinal cannabis strains as part of a molecular phenotyping program.

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

Figure 5: Comparison of BUP, MDP and TDP.

CITATIONS

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