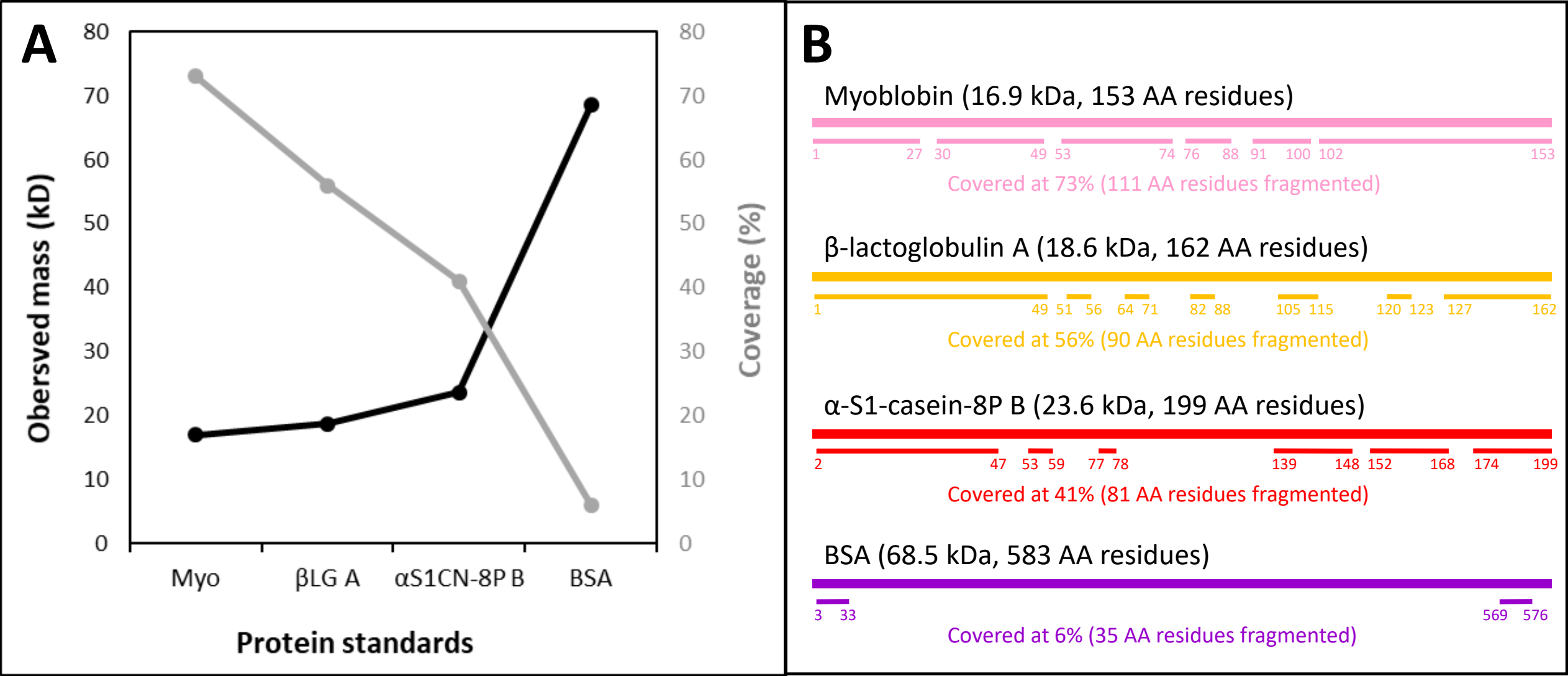


Vincent et al. Supplementary Figure S1.

A. Relationship between the molecular weight (MW) of the protein standards analysed and their sequencing results by top-down proteomics. B. Schematised sequence coverage for each of the four standards; length is not to scale (see Figures 3D and 4B for fragmentation patterns).



Vincent et al. Supplementary Figure S2

Mascot search results of protein standards MS/MS peak lists using (A) a homemade database and (B) Swissprot database.

A

MATRIX SCIENCE

MASCOT Search Results

User :

E-mail :

Search title : 2019-05-15_std_Xtract_mgf-export_1

MS data file : 2019-05-15_std_Xtract_mgf-export.mgf

Database : TDS_milk-protein-variants-sequences 20190501 (59 sequences; 10,517 residues)

Timestamp : 31 May 2019 at 05:22:07 GMT

Re-search

☒ All ☐ Non-significant ☐ Unassigned

[\[help\]](#)

Export

As

XML

Not what you expected? Try [the peptide summary](#).

▼Search parameters

Type of search : MS/MS Ion Search

Enzyme : NoCleave

Fixed modifications : [Carbamidomethyl \(C\)](#)

Variable modifications : [Oxidation \(M\)](#), [Acetyl \(Protein N-term\)](#), [Phospho \(ST\)](#)

Mass values : Monoisotopic

Protein mass : Unrestricted

Peptide mass tolerance : ± 1 %

Fragment mass tolerance : ± 50 ppm

Max missed cleavages : 9

Instrument type : ETD-CID-HCD

Number of queries : 371

►Score distribution

►Modification statistics

►Legend

Protein Family Summary

Format

Significance threshold p< 0.05

Max. number of families AUTO

[\[help\]](#)

Display non-sig. matches ☐

Dendrograms cut at 0

Preferred taxonomy All entries

►Sensitivity and FDR (reversed protein sequences)

Proteins (4)

Report Builder

Unassigned (265)

Protein hits (4 proteins)

►Columns: Standard (12 out of 16)

►Filters: (none)

Export as CSV

?Family	M	DB	Accession	Score	Mass	Matches	Match(sig)	Sequences	Seq(sig)	emPAI	Description
1	1	TDS_milk-protein-variants-sequences	myoglobin (P68082)	3782	16941	97	97	1	1	2.94	myoglobin (P68082)
2	1	TDS_milk-protein-variants-sequences	aS1CN B (P02662)	123	22960	6	6	1	1	1.16	aS1CN B (P02662)
3	1	TDS_milk-protein-variants-sequences	bLG E (P02754)	21	18531	1	1	1	1	0.17	bLG E (P02754)
4	1	TDS_milk-protein-variants-sequences	bLG F (P02754)	17	18472	1	1	1	1	0.17	bLG F (P02754)

B

MATRIX SCIENCE

MASCOT Search Results

User :

E-mail :

Search title : 2019-05-15_std_Xtract_mgf-export_2_swissprot

MS data file : 2019-05-15_std_Xtract_mgf-export.mgf

Database : SwissProt 2019_02 (559,228 sequences; 200,905,869 residues)

Timestamp : 3 Jun 2019 at 05:45:50 GMT

Re-search

☒ All ☐ Non-significant ☐ Unassigned

[\[help\]](#)

Export

As

XML

Not what you expected? Try [the peptide summary](#).

▼Search parameters

Type of search : MS/MS Ion Search

Enzyme : NoCleave

Fixed modifications : [Carbamidomethyl \(C\)](#)

Variable modifications : [Oxidation \(M\)](#), [Phospho \(ST\)](#)

Mass values : Monoisotopic

Protein mass : Unrestricted

Peptide mass tolerance : ± 1 %

Fragment mass tolerance : ± 50 ppm

Max missed cleavages : 9

Instrument type : ETD-CID-HCD

Number of queries : 371

►Score distribution

►Modification statistics

►Legend

Protein Family Summary

Format

Significance threshold p< 0.05

Max. number of families AUTO

[\[help\]](#)

Display non-sig. matches ☐

Dendrograms cut at 0

Preferred taxonomy All entries

►Sensitivity and FDR (reversed protein sequences)

Proteins (1)

Report Builder

Unassigned (329)

Protein hits (1 proteins)

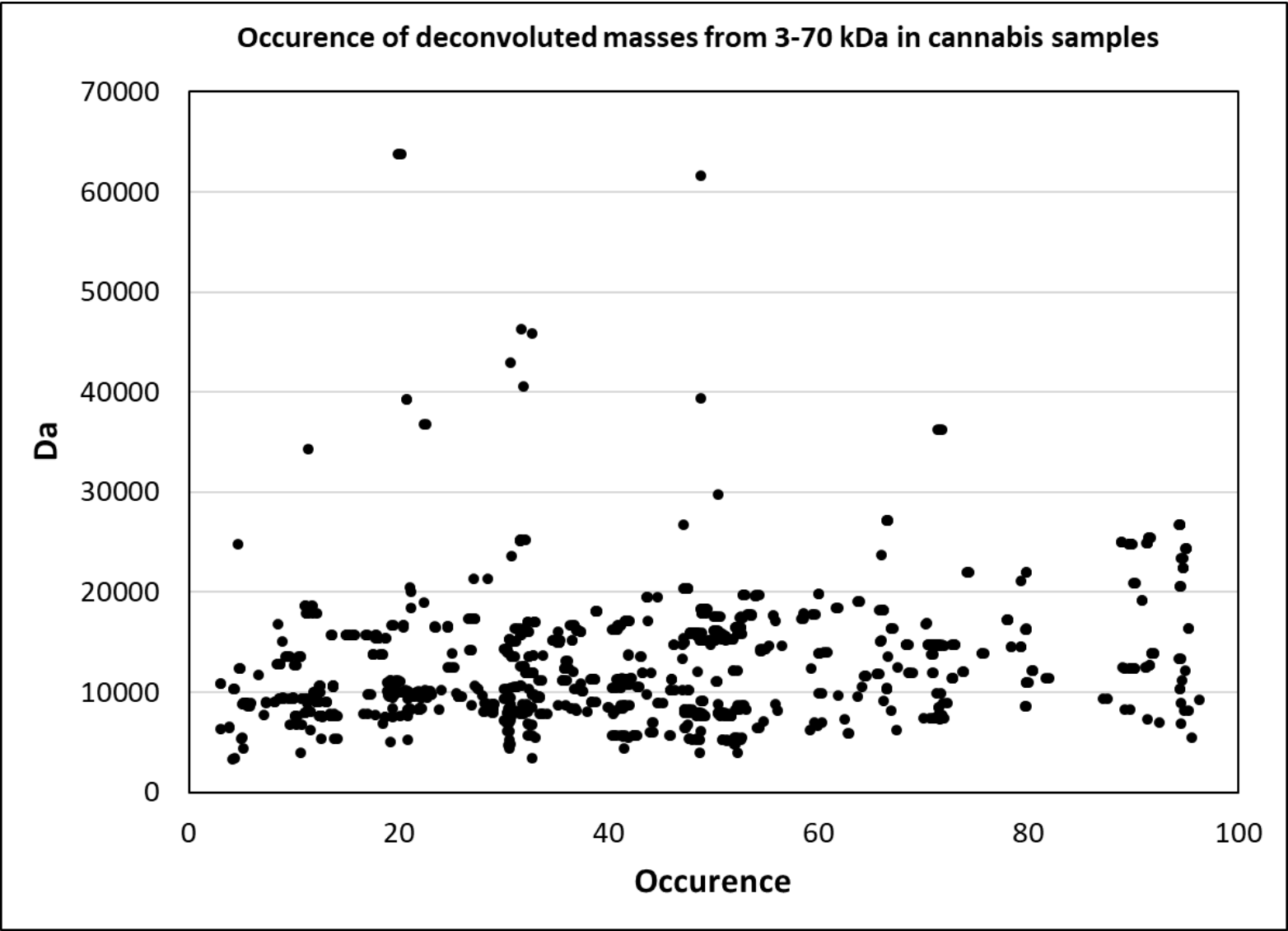
►Columns: Standard (12 out of 16)

►Filters: (none)

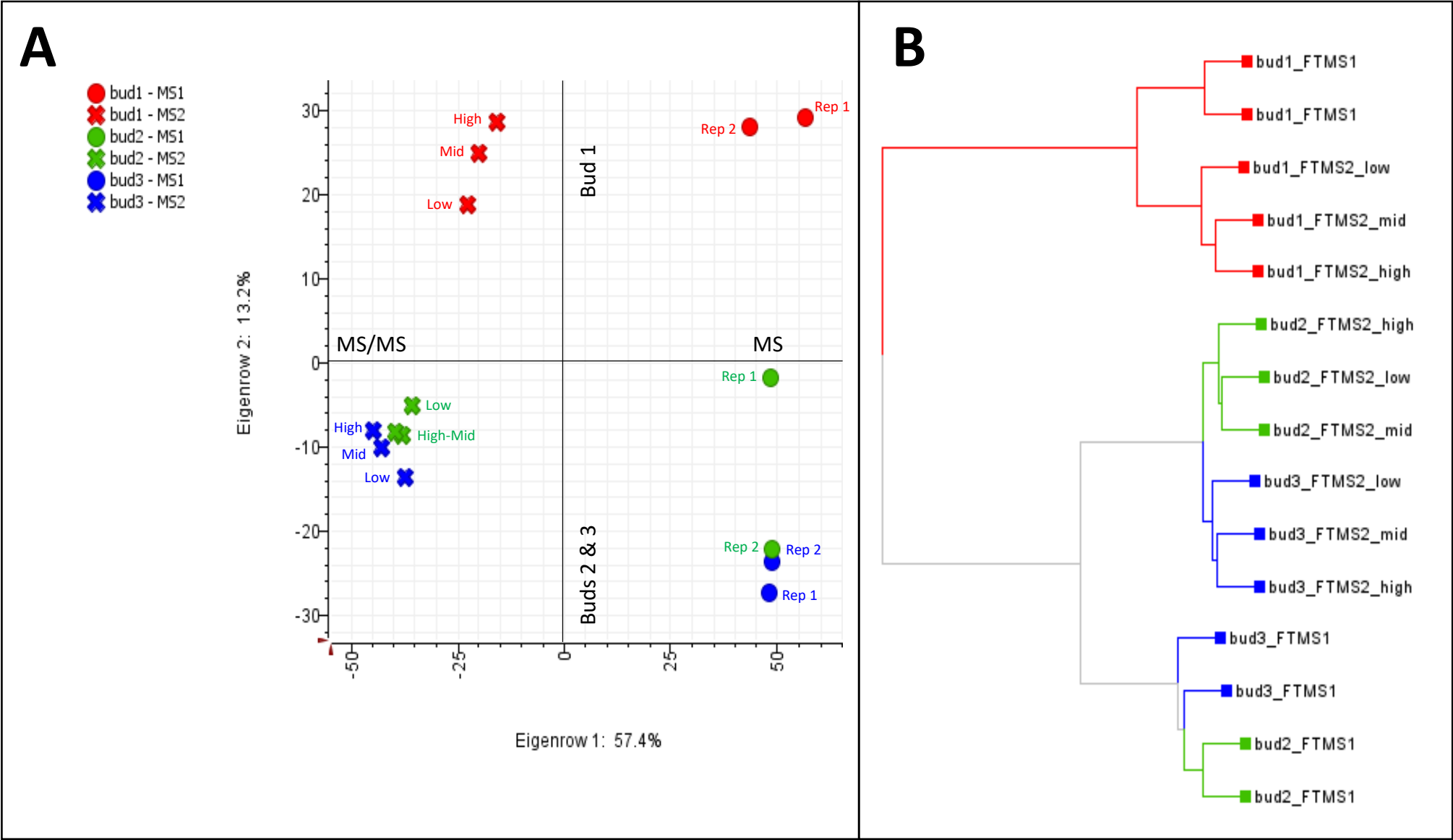
Export as CSV

?Family	M	DB	Accession	Score	Mass	Matches	Match(sig)	Sequences	Seq(sig)	emPAI	Description
1	1	SwissProt	MYG_EQUBU	1456	17072	46	46	2	2	2.91	Myoglobin OS=Equus burchelli OX=9790 GN=MB PE=1 SV=2

Vincent et al. Supplementary Figure S3
Distribution of cannabis proteins according to their accurate masses (3-70 kDa).

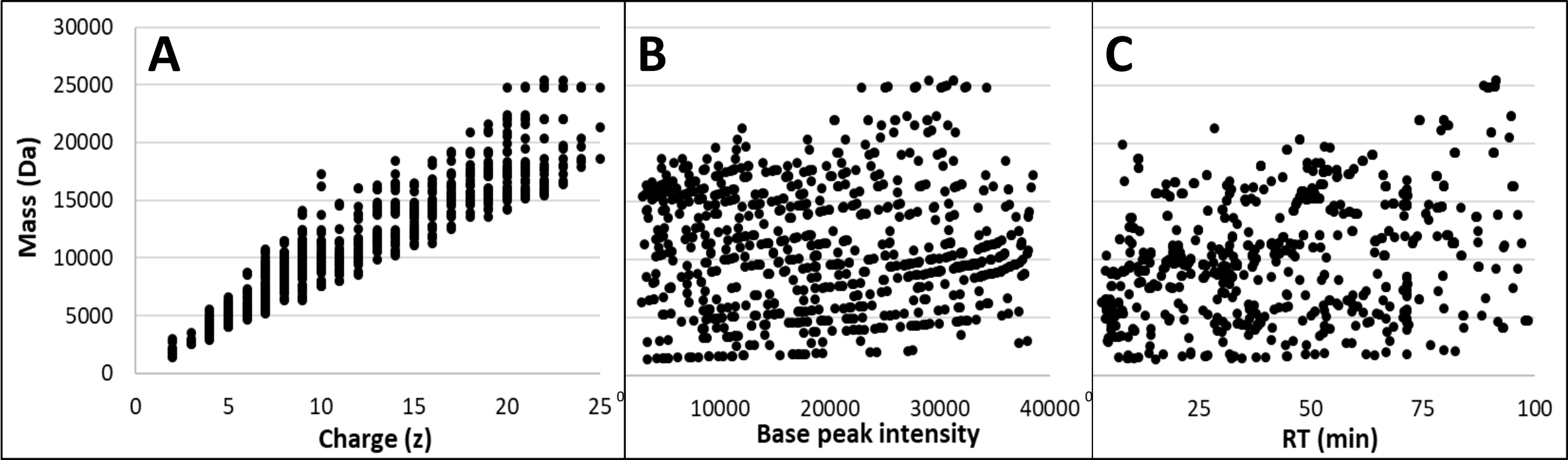


Vincent et al. Supplementary Figure S4
Multivariate statistical analyses using LC-MS data from cannabis protein samples. **(A)** Principal Component Analysis (PCA); **(B)** Hierarchical Clustering Analysis (HCA).



Vincent et al. Supplementary Figure S5

Statistics on parent ions from cannabis proteins analysed by LC-MS/MS. **(A)** Distribution of deconvoluted mass according to their charge state; **(B)** Distribution of deconvoluted masses according to their base peak intensity; **(C)** Distribution of deconvoluted masses according to their elution times.



Top-down sequencing result from Mascot for *C. sativa* Cytochrome b559 subunit alpha (A0A0C5ARS8). **(A)** Protein view; **(B)** Peptide view.

MATRIX SCIENCE MASCOT Search Results

☒ All
 ☐ Non-significant
 ☐ Unassigned
 [\[help\]](#)

 As XML

Not what you expected? Try [the select summary](#).

▼ *Search parameters*

► Score distribution

▼ *Modification statistics*

► **Legend**

Protein Family Summary

Significance threshold $p <$
 Max. number of families
[\[help\]](#)

Display non-sig. matches ☐
 Dendrograms cut at

Show Percolator scores ☐

Preferred taxonomy

► Sensitivity and FDR (reversed protein sequences)

Proteins (36) Report Builder Unassigned (11008)

Protein families 1–10 (out of 36)

10 per page 1 2 3 4 Next Expand all Collapse all

Accession contains Find

▼1 3::tr|A0A0C5ARS8|A... 1641 Cytochrome b559 subunit alpha OS=Cannabis sativa OX=3483 GN=psbE PE=3 SV=1

		Score	Mass	Matches	Sequences	emPAI
1.1	m3:tr A0A0C5ARS8 A	1641	9367	29 (29)	1 (1)	0.83 Cytochrome b559 subunit alpha OS=Cannabis sativa OX=3483 GN=csbE PE=3 SV=1

▼29 peptide matches (2 non-duplicate, 27 duplicate)

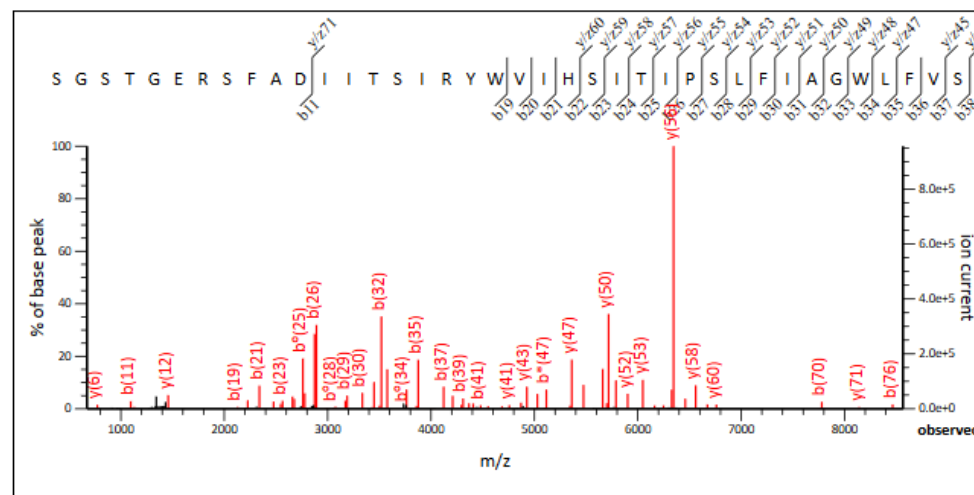
☒ Auto-fit to window

Query Dupes	Observed	Mr (expt)	Mr (calc)	$\frac{m}{z}$	M Score	Expect	Rank	U	Peptide
#3456	26	9237.6657	9236.6584	9236.6465	0.0110	0	1.1e-17	1	U M .SGSTGERSFADIITSIRVWVNSITIPSLFIAGMLFVSTGLAYDVFGSPRPNVEYTESRGGVPLITGRFDSLQELDFSRSF. -
#3453	1	9278.6719	9277.6646	9277.6571	0.0001	0	0.0011	17	U M .SGSTGERSFADIITSIRVWVNSITIPSLFIAGMLFVSTGLAYDVFGSPRPNVEYTESRGGVPLITGRFDSLQELDFSRSF. - + Acetyl (Protein N-term)

MATRIX SCIENCE Mascot Search Results

Peptide View

Match to Query 3456: 9236.658414 from(9237.665690,1+) intensity(1069305.5000) rtinseconds(5783) scans(1537) index(3594)
Title: File: "D:\Delphine\computer\Delphine DEPI_20180403\Med Can\proteomics\2018-08-29_pilot study\3_TDS
2019-05-09_TDS_MC_SpeedVac_LC-MS2\shorter names\MC-bud1_ETD15-CID50_HCD27.raw"; SpectrumID: "10650"; scans: "1537"
Data file 2019-05-09_MC_buds123_expant-mgf 1z.mgf



672.36 to 8565.16

Label all possible matches ☒ Label matches used for scoring ☐

Fixed modifications: Carbamidomethyl (C) (apply to specified residues or termini only)

