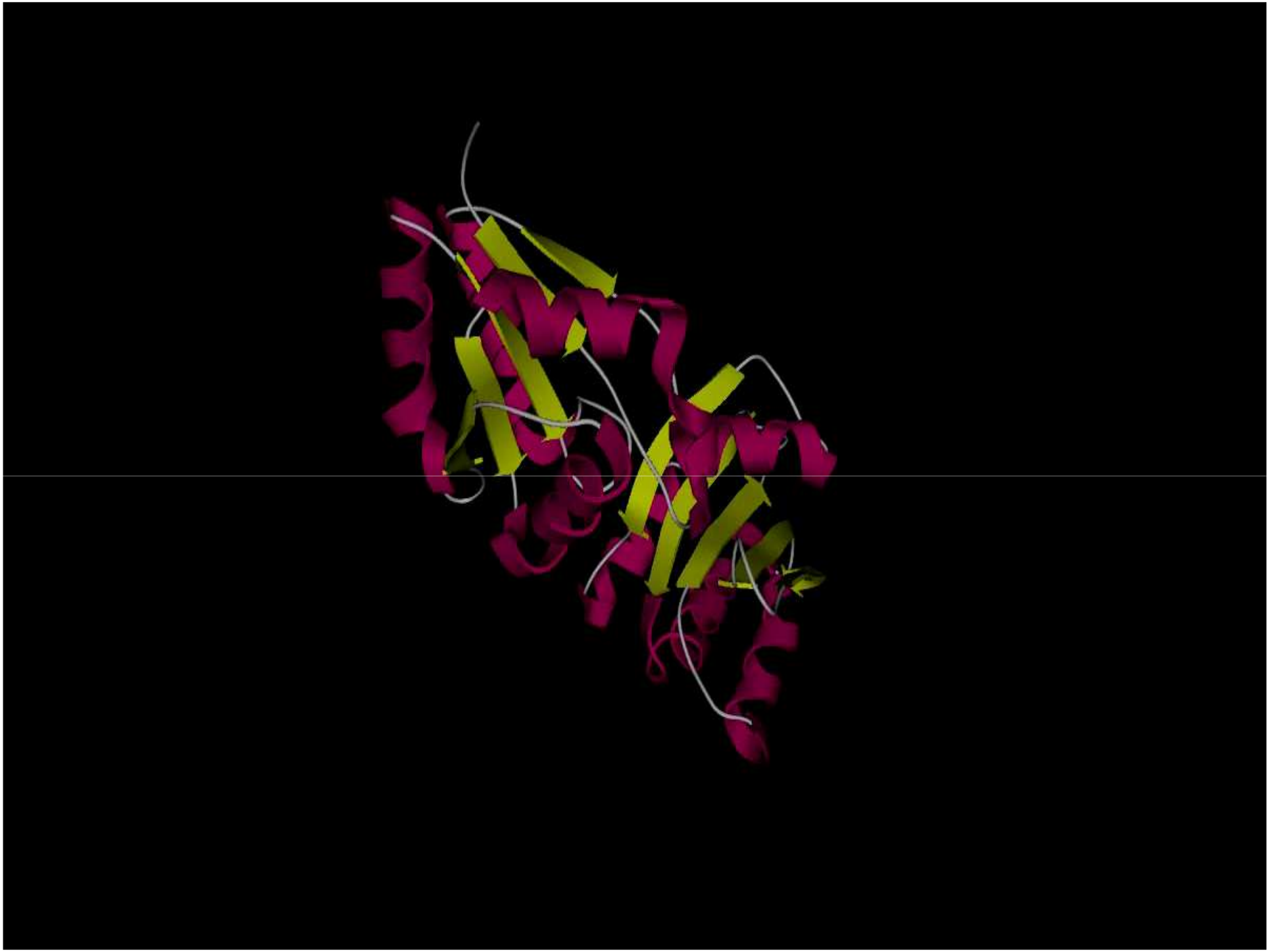
PROTEIN READING

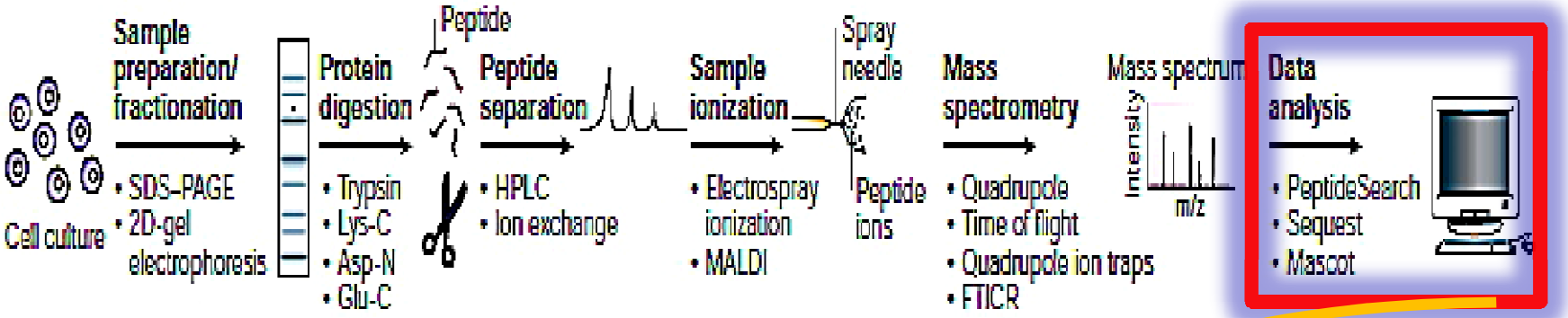
Identifikacija vrsta plijesni proteomskim pristupom

by Mario Cindrić
“Ruđer Bošković Institute, Zagreb, Croatia

Email: mcindric@irb.hr



STANDARD PROTEOMICS WORKFLOW



MATCHING CONCEPT

SEVERAL MATCHING ALGORITHMS:

MASCOT (Perkins, Pappin et al. 1999), Protein Prospector (Clauser, Baker et al. 1999), Peptide Search (Mann and Wilm 1994), SEQUEST (Eng 1994) and OMSSA (Geer, Markey et al. 2004).

What is the main source of proteomics data?

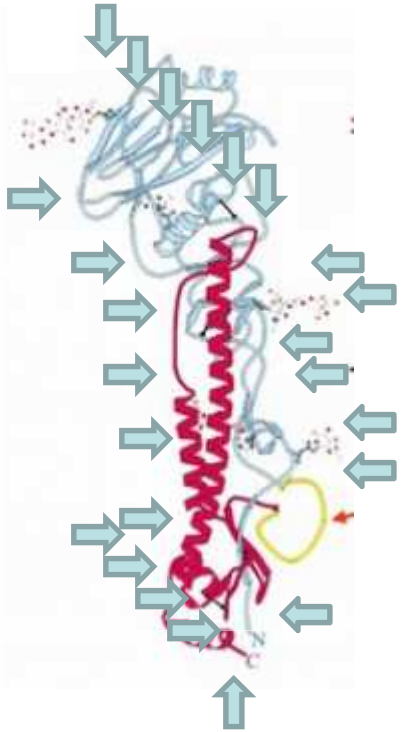
RefSeq

The database incorporates data from over 2400 organisms (6500 in 2013) and includes over one million proteins representing significant taxonomic diversity spanning prokaryotes, eukaryotes and viruses. Nucleotide and protein sequences are explicitly linked, and the sequences are linked to other resources including the NCBI Map Viewer and Gene. Sequences are annotated to include coding regions, conserved domains, variation, references, names, database cross-references, and other features

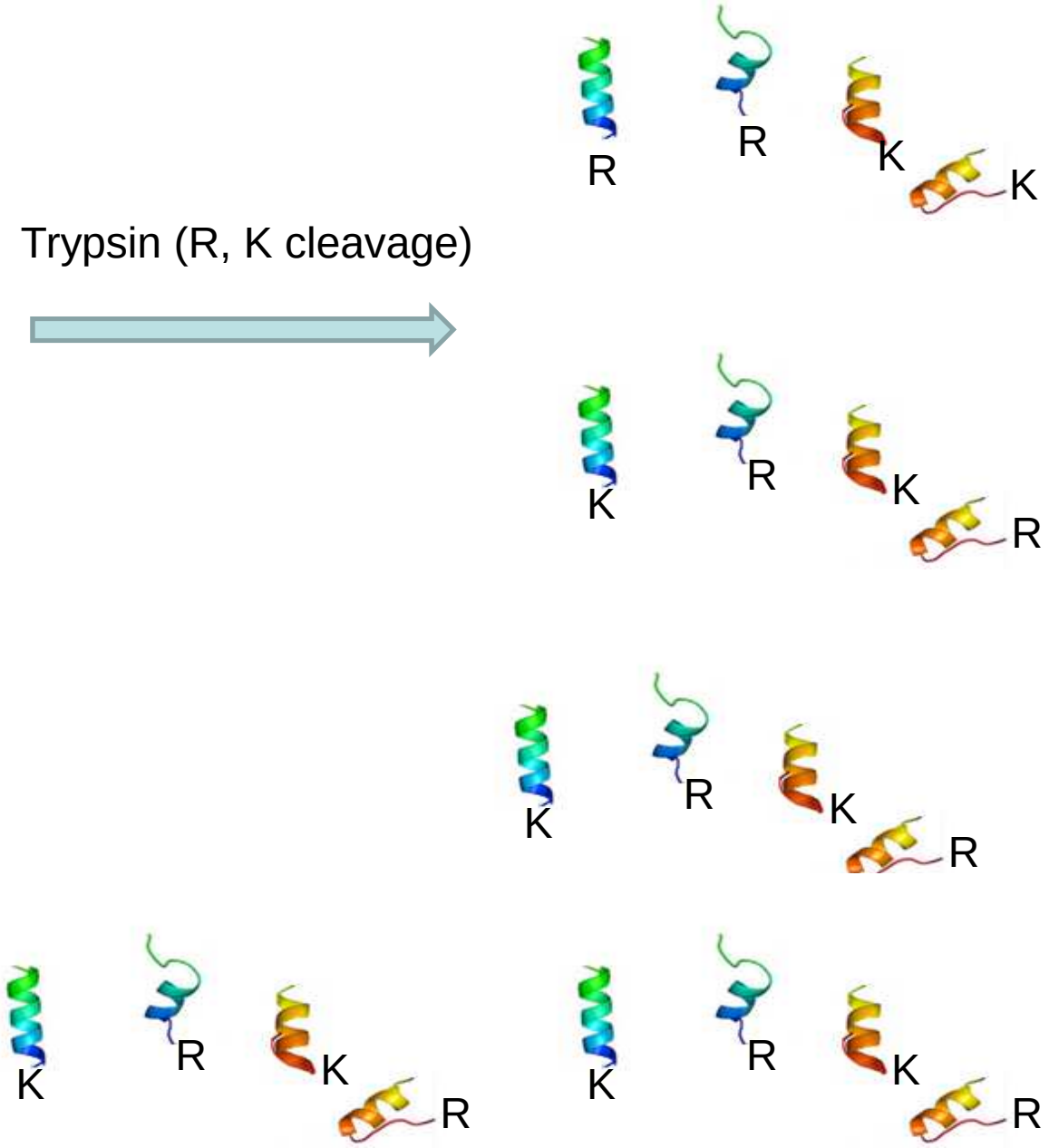
using a combined approach of collaboration and other input from the scientific community, automated annotation, propagation from GenBank and curation by NCBI staff.

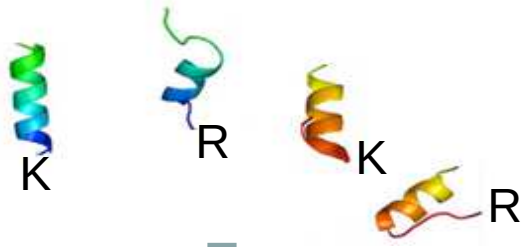
Pruitt, Nucleic Acids Research, 2005, Vol. 33

EXAMPLE OF TRYPSIN CLEAVAGE?

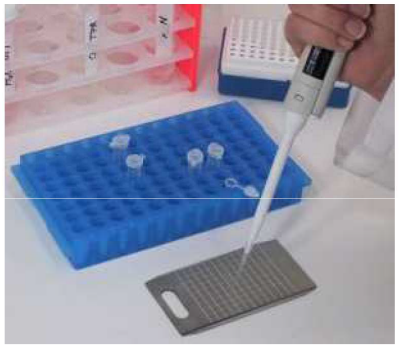


Trypsin (R, K cleavage)

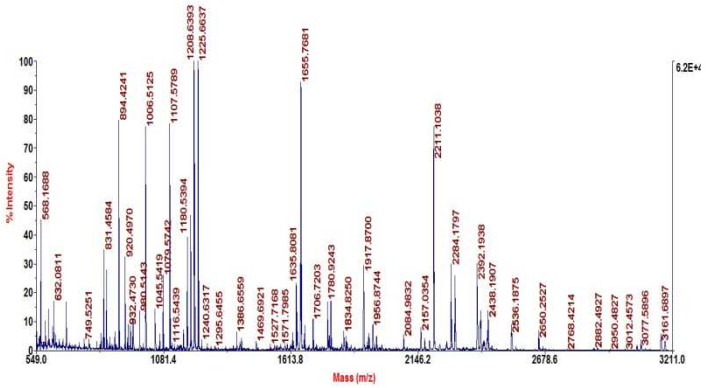




Loading Sample/Matrix on MALDI Target



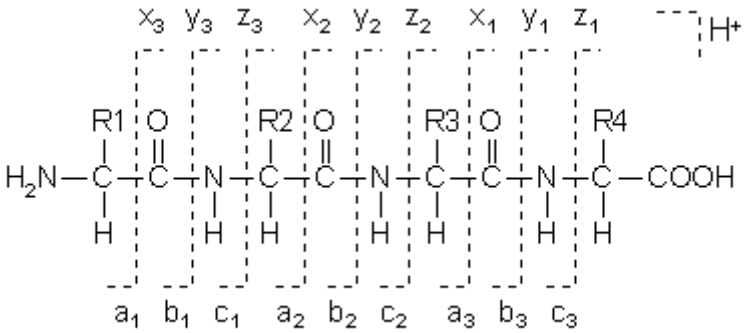
Common way of loading sample & matrix on MALDI target

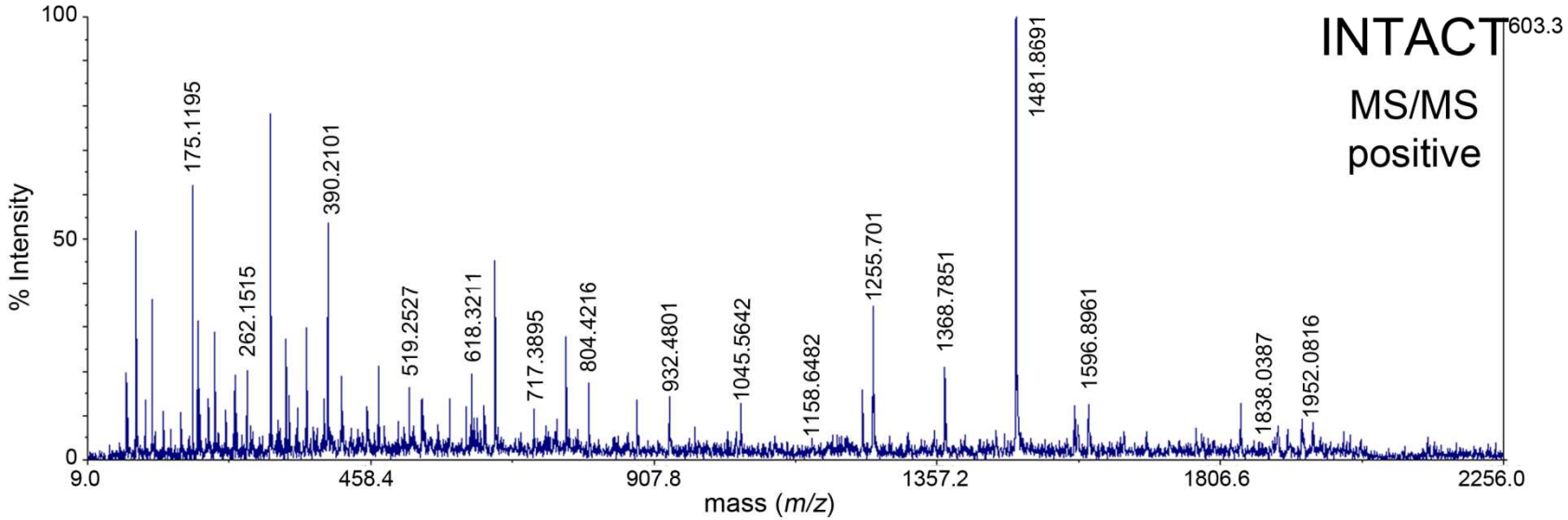


MS analysis

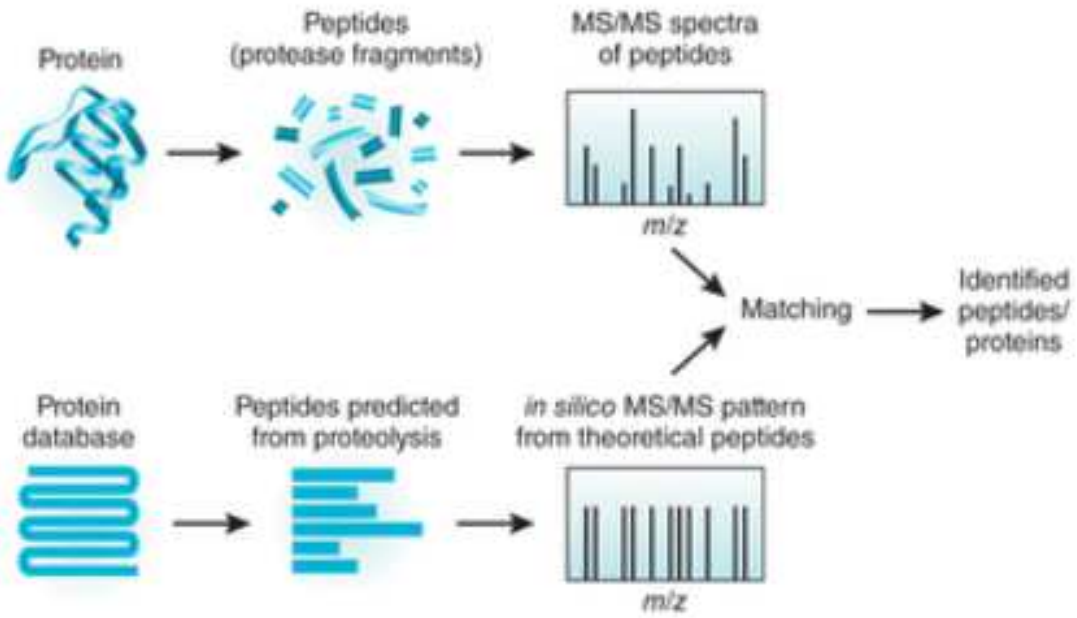
PRECURSOR SELECTION FROM MS DATA?

- a-ions,
- b-ions,
- c-ions,
- x-ions,
- y-ions,
- z-ions …





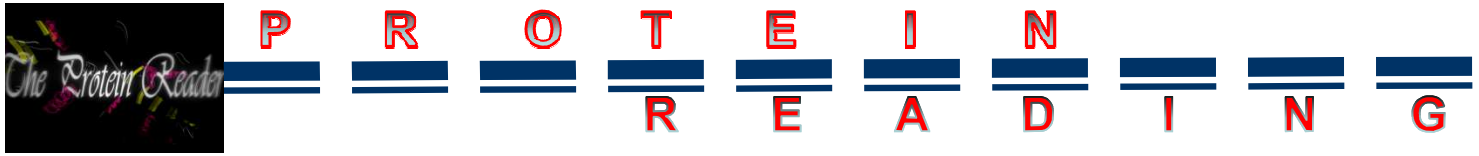
MATCHING IN SILICO AND EXPERIMENTAL DATA?



Duncan, 2010, Nat. Biotech.

These were general aspects of spectra matching concept

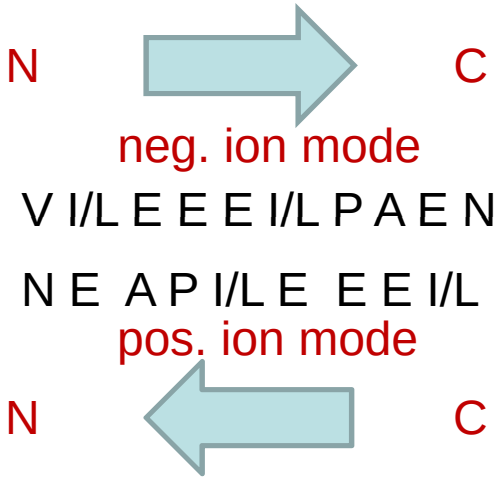
- Is there any other concept?



Protein reading concept:

1. Peptides MS negative ionization
2. *De novo* sequencing in neg. MS/MS
3. *De novo* sequencing in pos. MS/MS
4. Looking for overlapped readings
5. ID of the microorganisms by overlapped readings

e.g.



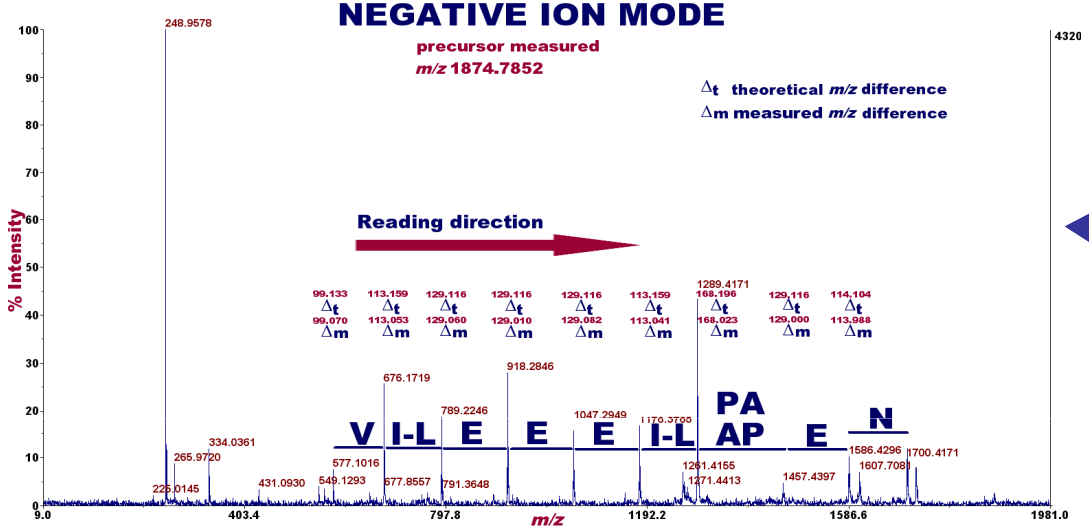
P R O T E I N
R E A D I N G

NEGATIVE ION MODE

precursor measured
m/z 1874.7852

Δ_t theoretical *m/z* difference
Δm measured *m/z* difference

Reading direction

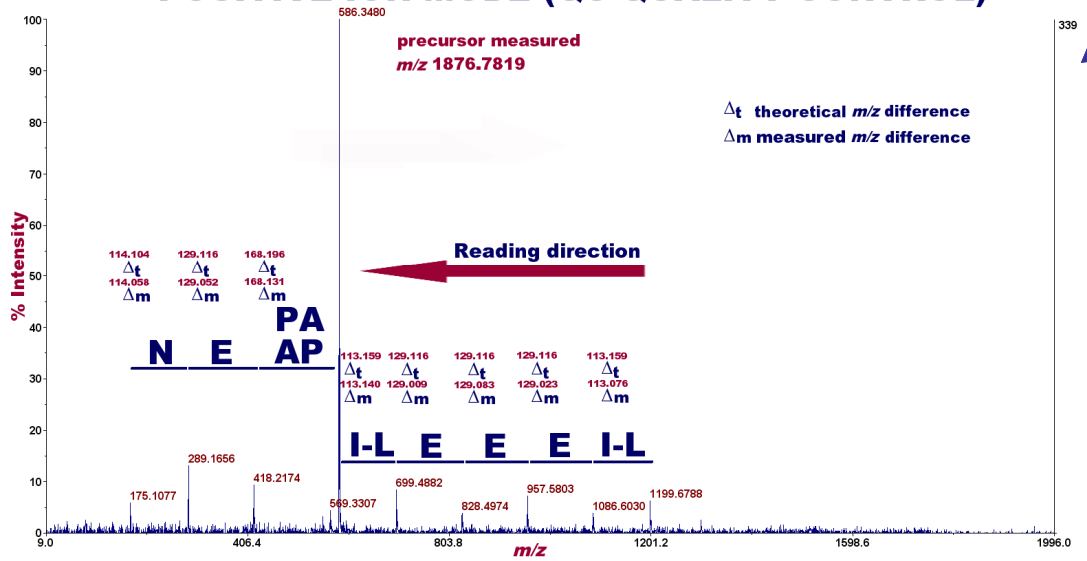


POSITIVE ION MODE (QC-QUALITY CONTROL)

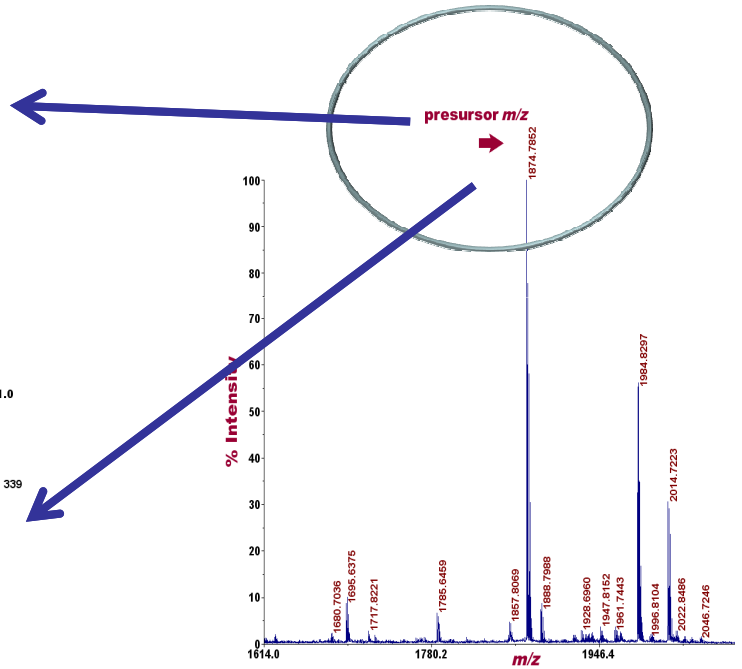
precursor measured
m/z 1876.7819

Δ_t theoretical *m/z* difference
Δm measured *m/z* difference

Reading direction



MS negative ion mode



CAF-/CAF+ or shorter
CAF/CAF

The Protein Reader

P R O T E I N R E A D I N G

Classical MALDI analysis

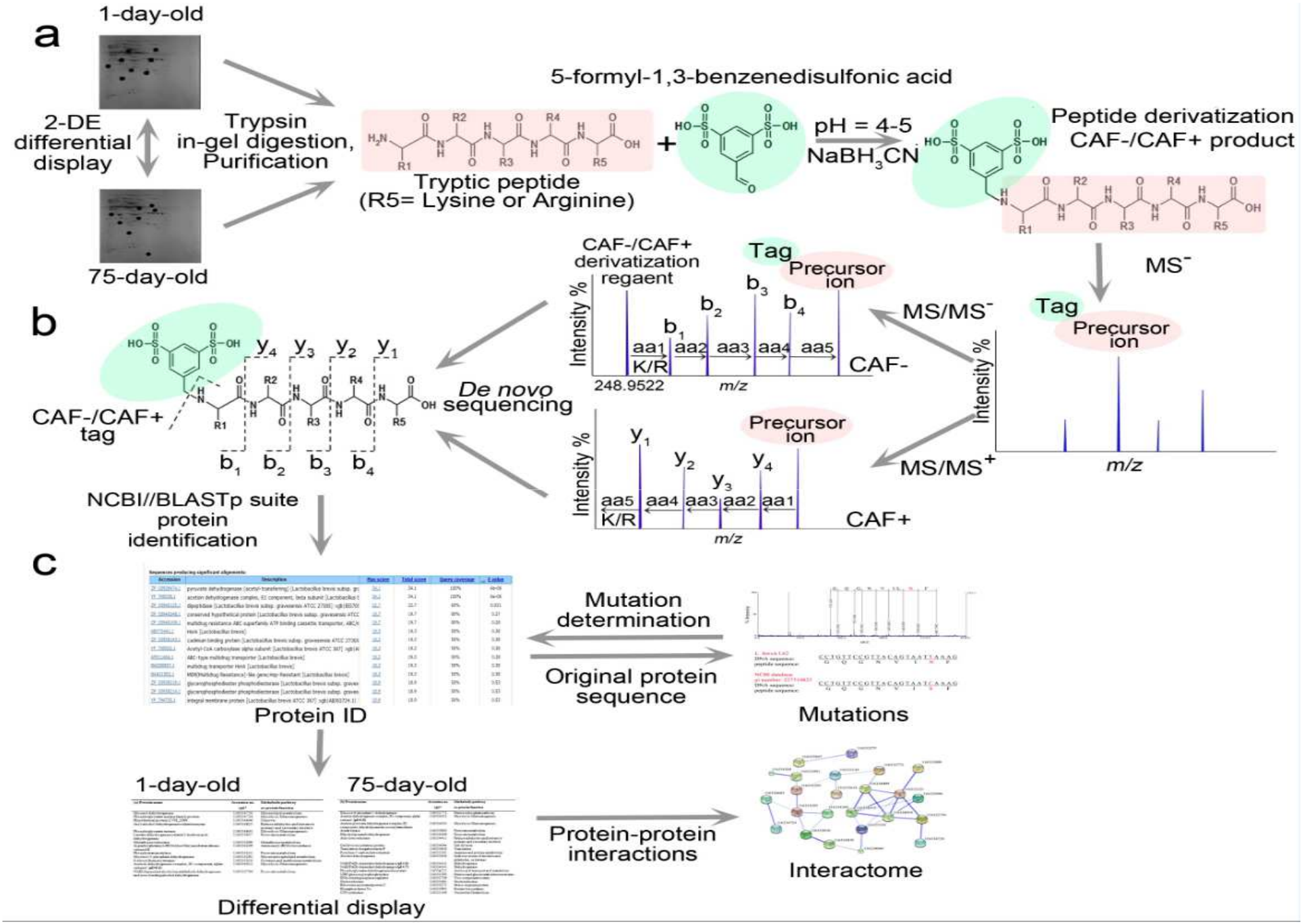
Analysis in positive ion mode of peptides

CAF/CAF MALDI analysis

Analysis in positive and negative ion mode of peptides

Soft derivatization

BENEFITS!!!
N-term → C-term
C-term → N-term



Cindrić, 2013, RCMS, US patent 8647880

PROTEIN READING

SAMPLE PREPARATION WORKFLOW

Pick a sampe into buffer

Blend it well

Harvest and collect supernatant

Protein sample

Add Trypsin beads Incubate

Peptide sample

Add CAF$^-$/CAF$^+$ reagent

Microwave 10 min

Peptide separation and collection

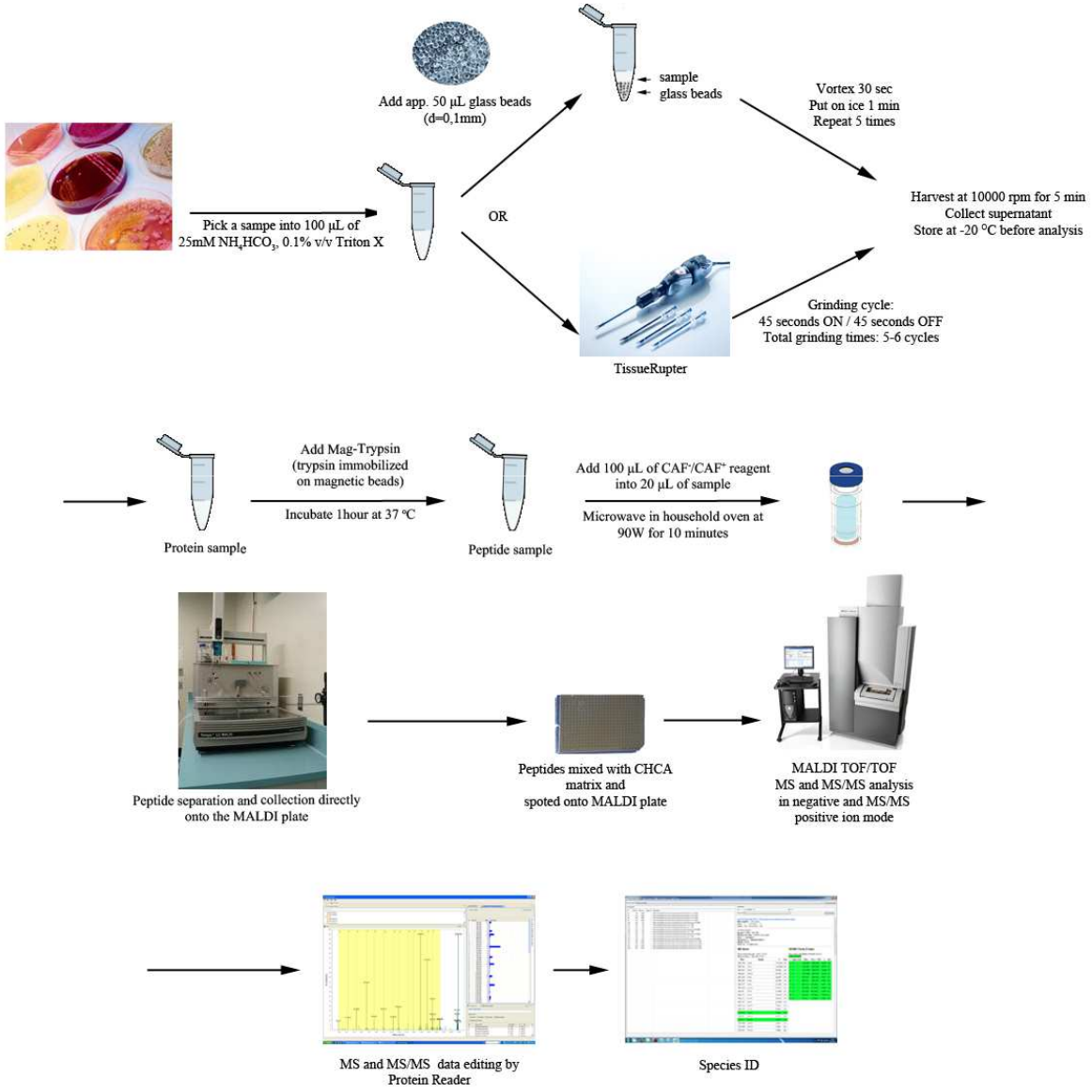
MS and MS/MS analysis in negative and MS/MS positive ion mode

MS and MS/MS data editing by Protein Reader

Species ID

P R O T E I N
R E A D I N G

SAMPLE PREPARATION WORKFLOW



P R O T E I N
R E A D I N G

COMPARISON WITH OTHER MASS SPECTROMETRY BIOTYPIZATION APPROACHES

Bruker's Biotyper results

specie	strain	result	best match	score	secound best match	score	positive identification	
Lactobacillus plantarum	strain 7.1	(+++)(A)	Lactobacillus plantarum	2.438	Lactobacillus plantarum	2.43	+	
Enterococcus faecium	strain 7.2	(++)(A)	Enterococcus faecium	2.224	Enterococcus faecium	2.221	+	
Pediococcus pentosaceus	strain 2.2	(++)(A)	Pediococcus pentosaceus	2.134	Pediococcus pentosaceus	2.067	+	
Lactobacillus casei	strain 16.1	(++)(A)	Lactobacillus paracasei	2.257	Lactobacillus paracasei	2.119	+	False positive match!!!
Lactobacillus fermentum	strain 8.2	(++)(A)	Lactobacillus fermentum	2.05	Lactobacillus fermentum	1.794	+	
Lactobacillus fermentum	strain 10.2	(+)(B)	Lactobacillus fermentum	1.85	Lactobacillus fermentum	1.701	+/-	
Lactobacillus brevis	strain 15.2	(+)(B)	Lactobacillus brevis	1.938	Lactobacillus brevis	1.892	+/-	
Lactobacillus buchneri	strain 12.3	(+)(B)	Lactobacillus buchneri	1.895	Lactobacillus kefiri	1.725	+/-	
Lactobacillus plantarum	strain 16.3	(+)(B)	Lactobacillus plantarum	1.97	Lactobacillus plantarum	1.849	+/-	
Lactobacillus fermentum	strain 9.1	(+)(B)	Lactobacillus fermentum	1.885	Lactobacillus fermentum	1.719	+/-	
Lactoccocus lactis	strain 1.1	(-)(C)	7 not reliable identification	1.3837	not reliable identification	1.355	-	
Lactococcus lactis	strain 6.8	(-)(C)	16 Not reliable identification	1.27216	Not reliable identification	1.234	-	

legend	score
A - positive	>2
B - maybe	1.7 to 2
negative resulti	< 1,7

A - positive identification
B - maybe positive identification
C - not reliable identification

THE PROTEIN READER

Final result: 4 out of 12

CAF+ / CAF- results	positive identification				
Lactobacillus plantarum	+	Pediococcus pentosaceus	8 spots	5 precursors	15 read sequences
Enterococcus faecium	+	Lactobacillus brevis	8 spots	5 precursors	20 read sequences
Pediococcus pentosaceus	+	Lactobacillus casei	8 spots	5 precursors	34 read sequences
Lactobacillus casei	+	Lactobacillus plantarum	8 spots	5 precursors	17 read sequences
Lactobacillus fermentum	+	Enterococcus faecium	8 spots	5 precursors	12 read sequences
Lactobacillus fermentum	+	Lactobacillus fermentum	8 spots	5 precursors	14 read sequences
Lactobacillus brevis	+	Lactoccocus lactis	8 spots	5 precursors	36 read sequences
Lactobacillus buchneri	+	Lactobacillus buchneri	8 spots	5 precursors	33 read sequences
Lactobacillus plantarum	+				
Lactobacillus fermentum	+				
Lactoccocus lactis	+				
Lactococcus lactis	+				

Final result: 12 out of 12

Examples of sequences reading

PROTEIN READING

Common clinical strains		Rare clinical strains	
Proteus mirabilis str. 33	*Acitenobacter baumanni* str. 3	*Trichophyton rubrum* CBS 118892	*Bacillus anthracis*
Proteus mirabilis str. 1312R	*Enterococcus faecium* str. E124	*Aspergillus fumigatus* Af293	*Aspergillus flavus*
Staphylococcus aureus strain 59	*Pseudomonas aurigenosa* strains: P1, P3, P104, P91, P33, P123	*Aureobasidium pullulans*	*Candida albicans*
Staphylococcus aureus MRSA	*Escherichia coli* CFT073	*Fusarium dimerum*	*Candida glabrata*
Yersinia enterocolica str. 94	*Helicobacter pylori* P12	*Fusarium delphinoides*	*Cryptococcus neoformans*
Indentified by PR in 3 h		Indentified by PR	

Tablica 1.
Skupni prikaz rezultata analiza ukupnog broja kvasaca i plijesni te plijesni iz *Aspergillus flavus* grupe (=*Aspergillus* sekcija Flavi) te iz roda *Fusarium*:

Broj	Protokolarni broj uzorka Zavoda za mikrobiologiju FBF-a	Oznaka uzorka merkantilnog kukuruza roda 2013. godine	Ukupan broj kvasaca i plijesni u 1 g (Tolerancija)	Broj plijesni u 1 g	Broj plijesni roda *Aspergillus* iz sekcije *Flavi* u 1 g (identificiran *Aspergillus* spp.)	Broj plijesni *Fusarium* spp. u 1 g (identificirano kao *Fusarium* spp.)	Odgovara li pravilniku (NN 26/98, čl. 71)? DA= odgovara NE=ne odgovara
1	**MFBF 11221**	**AO 110**	$1{,}8 \times 10^6$ (5%)	5×10^5	136 (*Aspergillus oryzae)*	$1{,}1 \times 10^5$	NE
2	**MFBF 11222**	**MK 106**	$1{,}3 \times 10^5$ (12%)	$1{,}2 \times 10^5$	Nije nađeno	$1{,}2 \times 10^5$	DA
10	**MFBF 11230**	**TŽ-230**	2×10^5 (6%)	$1{,}7 \times 10^5$	Nije nađeno	$1{,}4 \times 10^5$	DA
11	**MFBF 11231**	**IZ-225**	7×10^5 (1%)	7×10^5	300 (*Aspergillus flavus*)	$6{,}4 \times 10^4$	NE
20	**MFBF 11240**	**V.Š./13-7**	7×10^4 (9%)	6×10^4	Nije nađeno	4×10^4	DA
21	**MFBF 11241**	**V.Š./13-8**	$1{,}6 \times 10^5$ (4%)	7×10^4	182 (*Aspergillus flavus*)	$5{,}4 \times 10^4$	DA
22	**MFBF 11242**	**V.Š./13-9**	$1{,}1 \times 10^4$ (14%)	3×10^3	Nije nađeno	$1{,}3 \times 10^3$	DA
27	**MFBF 11247**	**FT-2/13**	$9{,}4 \times 10^4$ (5%)	$7{,}5 \times 10^3$	Nije nađeno	$6{,}4 \times 10^3$	DA
28	**MFBF 11248**	**DE-2/13**	3×10^4 (11%)	3×10^4	136 (*Aspergillus oryzae*)	$1{,}9 \times 10^4$	DA
29	**MFBF 11249**	**CJ-2/13**	$5{,}2 \times 10^4$ (2%)	$1{,}4 \times 10^4$	500 (*Aspergillus flavus)*	$3{,}2 \times 10^3$	DA
30	**MFBF 11250**	**PM-2/13**	$9{,}4 \times 10^3$ (14%)	$9{,}3 \times 10^3$	364 (*Aspergillus oryzae*)	$5{,}4 \times 10^3$	DA

PROTEIN READING

BENEFITS

- For the first time mass spectrometry can exploit both positive and negative ion mode for protein and species identification
- Sequence reading includes only b-ions (neg. ion mode) and y-ions (pos. ion mode) making sequence reading easy and unambiguous
- Clinical application requires quality control gained by two orthogonal techniques (MS/MS neg. and pos.)
- Sample preparation is finished in 3 h and it can be completely automatized
- Biotypization is relied on genetic sequence not on the “home made” database
- Developed software named PROTEIN READER exploits the benefits accomplished by CAF/CAF chemistry for fast, accurate, robust and reliable protein and species identification