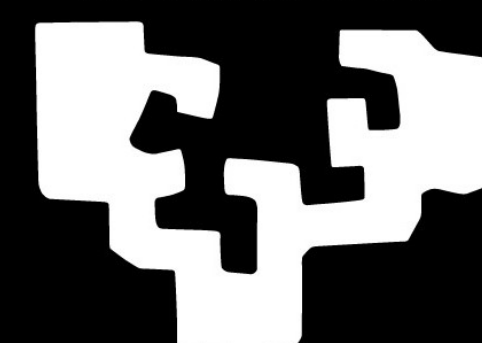


PAnalyzer: A software tool to tackle the protein inference problem

Gorka Prieto¹ <gorka.prieto@ehu.es>, Kerman Aloria², Asier Fullaondo³ and Jesus Mari Arizmendi⁴

¹Department of Electronics and Telecommunications, ²Proteomics Core Facility-SGIKER, ³Department of Genetics, Physical Anthropology and Animal Physiology, ⁴Department of Biochemistry and Molecular Biology. University of the Basque Country, UPV/EHU

eman ta zabal zazu



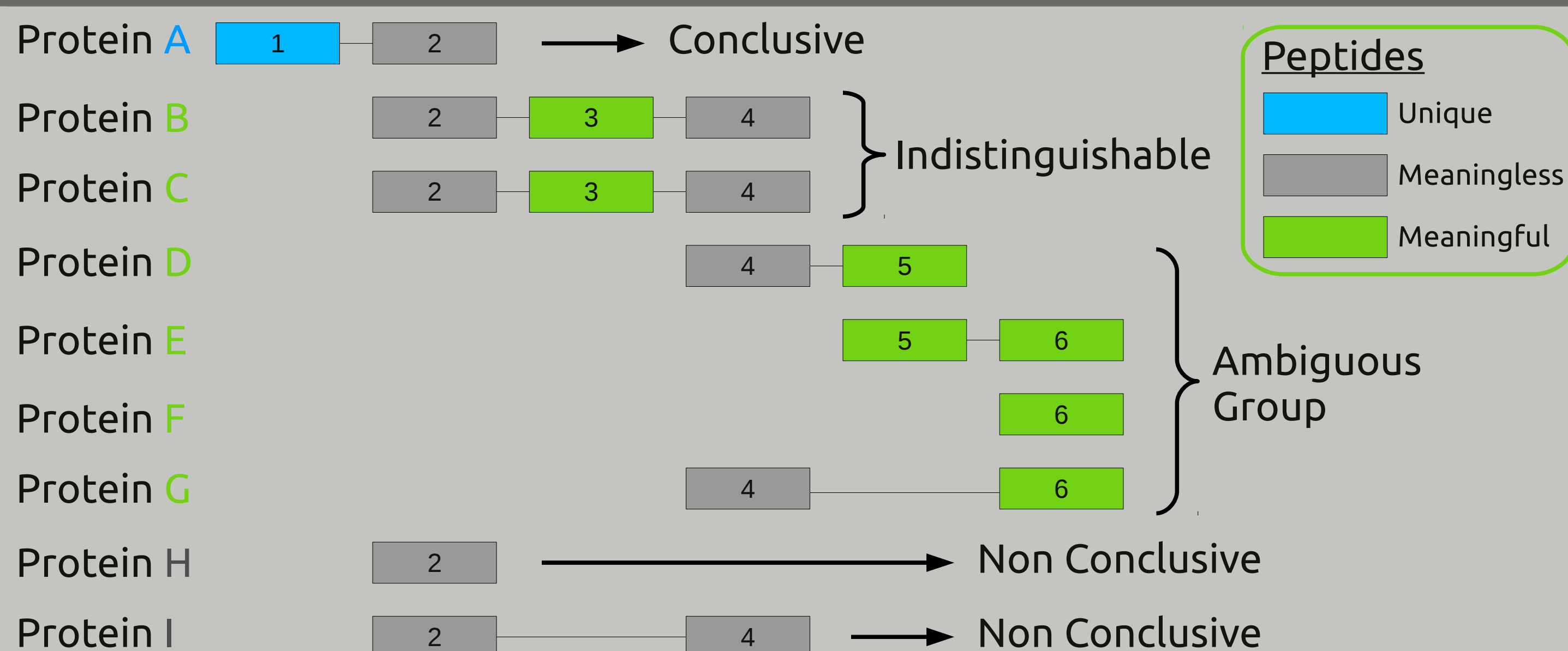
Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

Introduction

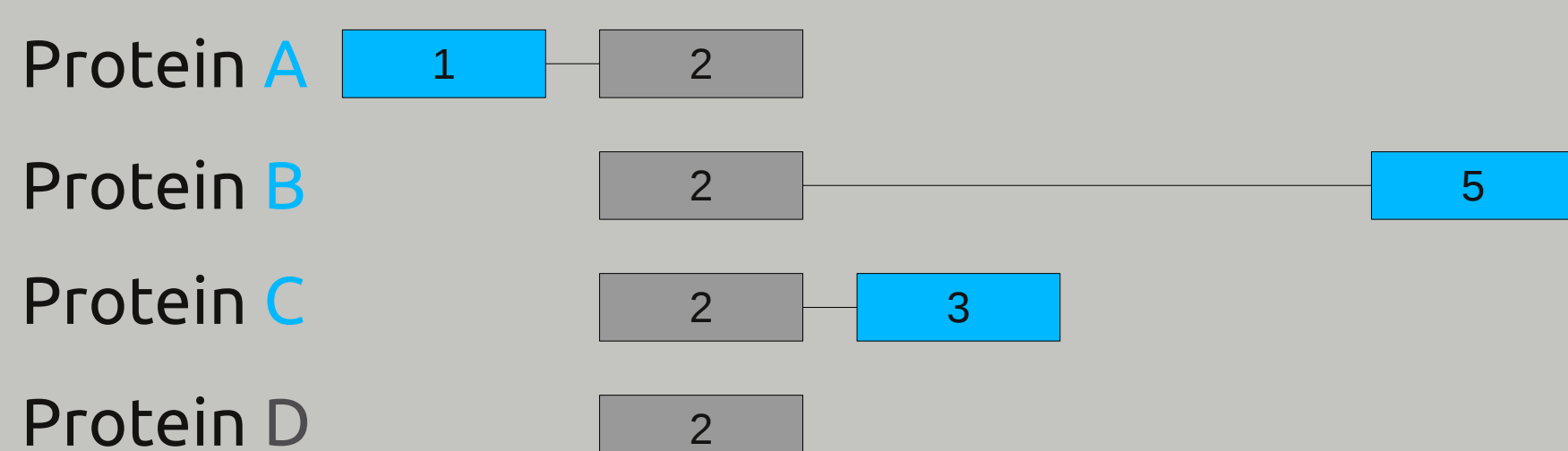
- Protein identification by **shotgun proteomics**
 - It is a peptide-centric approach
 - Same peptide sequence can belong to multiple proteins → **ambiguities**
 - Specially problematic with protein isoforms and protein families
 - Protein identifications with no real evidence are largely reported
- PAnalyzer**
 - Classifies protein identification results according to Nesvizhskii and Aebersold proposal (1)
 - Supports peptide filtering according to score thresholds
 - Supports multiple run integration
 - Multiplatform (.NET/Mono) and free software (GPL)

Protein classification

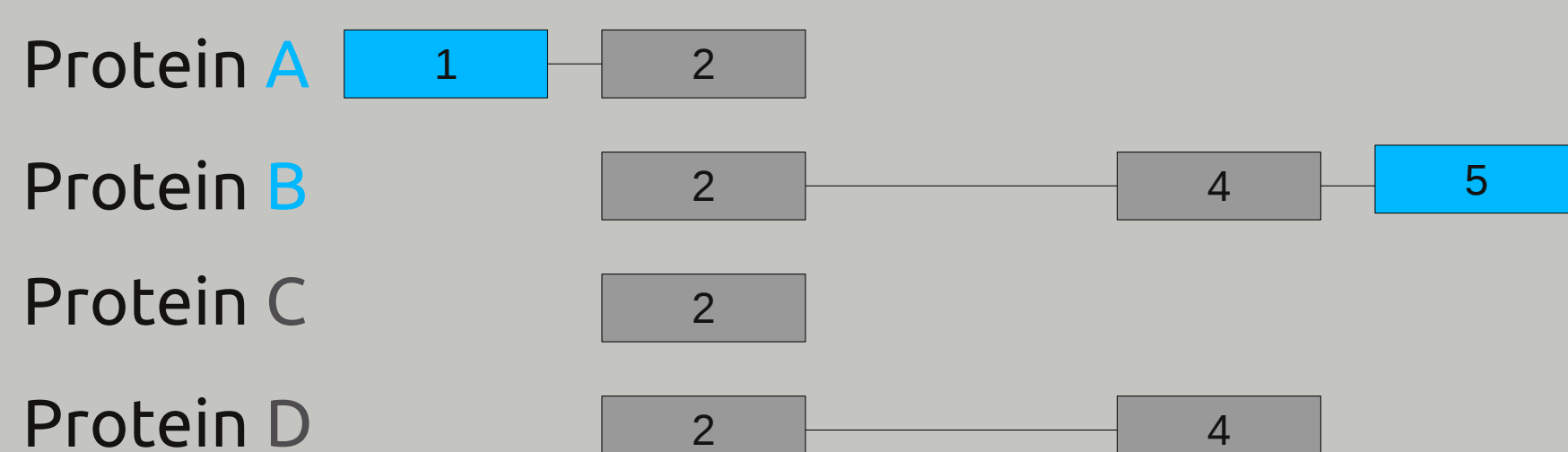


Multiple runs

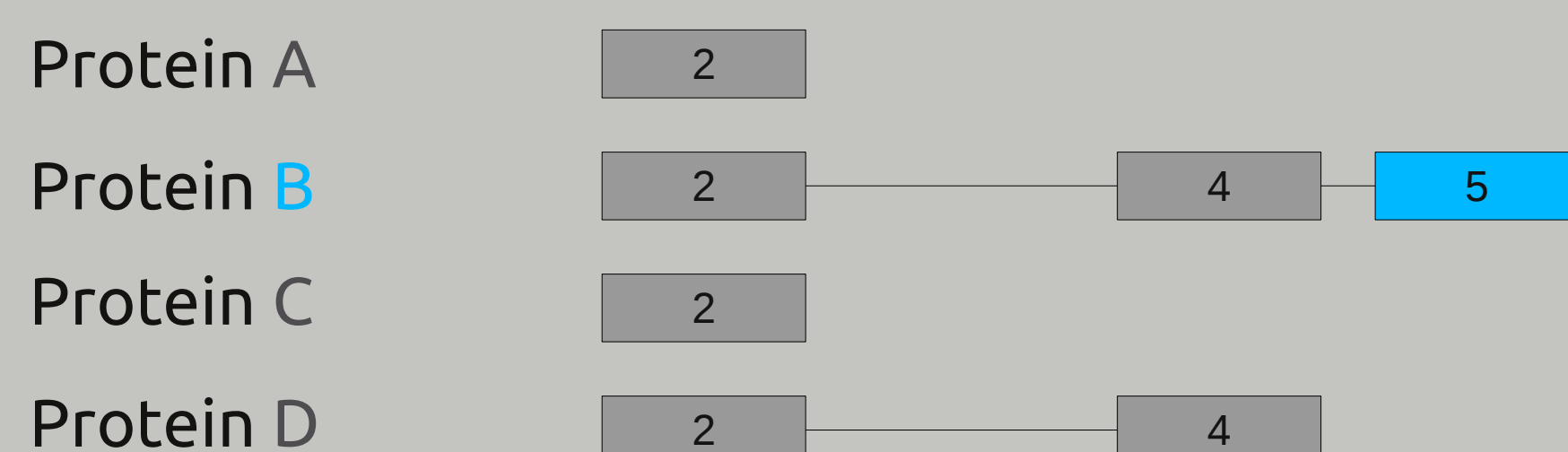
RUN #1



RUN #2



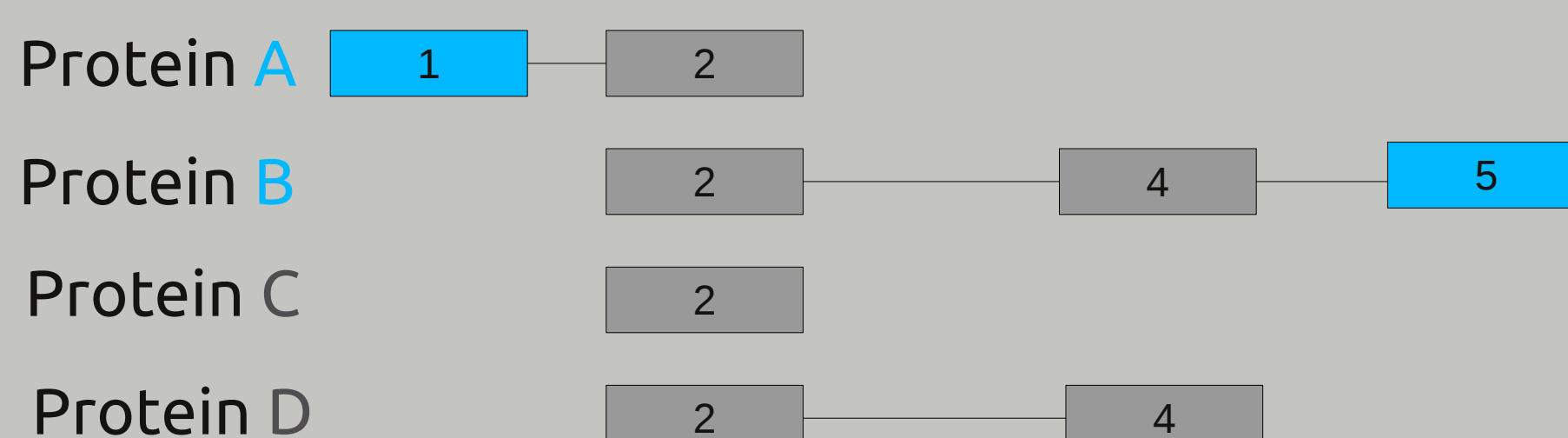
RUN #3



SINGLE RUN ANALYSIS

- RUN #1**
 - Peptide 4 missing
 - A, B, C conclusive; D non conclusive
- RUN #2**
 - Peptide 3 missing
 - A, B conclusive; C, D non conclusive
- RUN #3**
 - Peptides 1 and 3 missing
 - B conclusive; A, C, D non conclusive

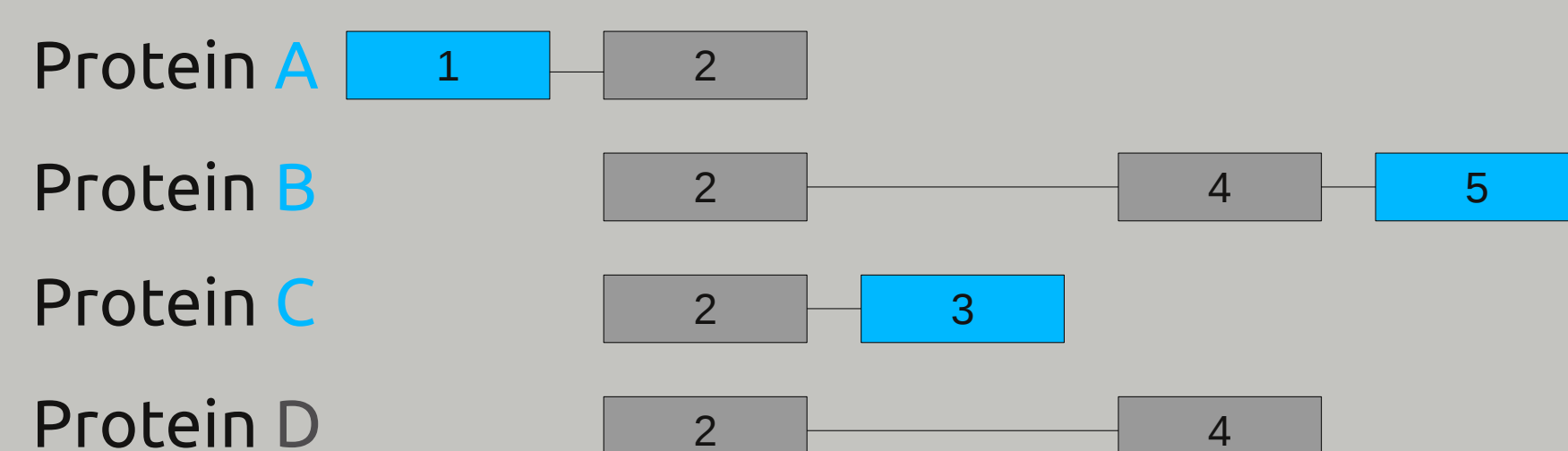
VOTING



- Voting: peptides identified in more than half of the runs
- Peptide 3 is filtered
- A, B conclusive; C, D non conclusive

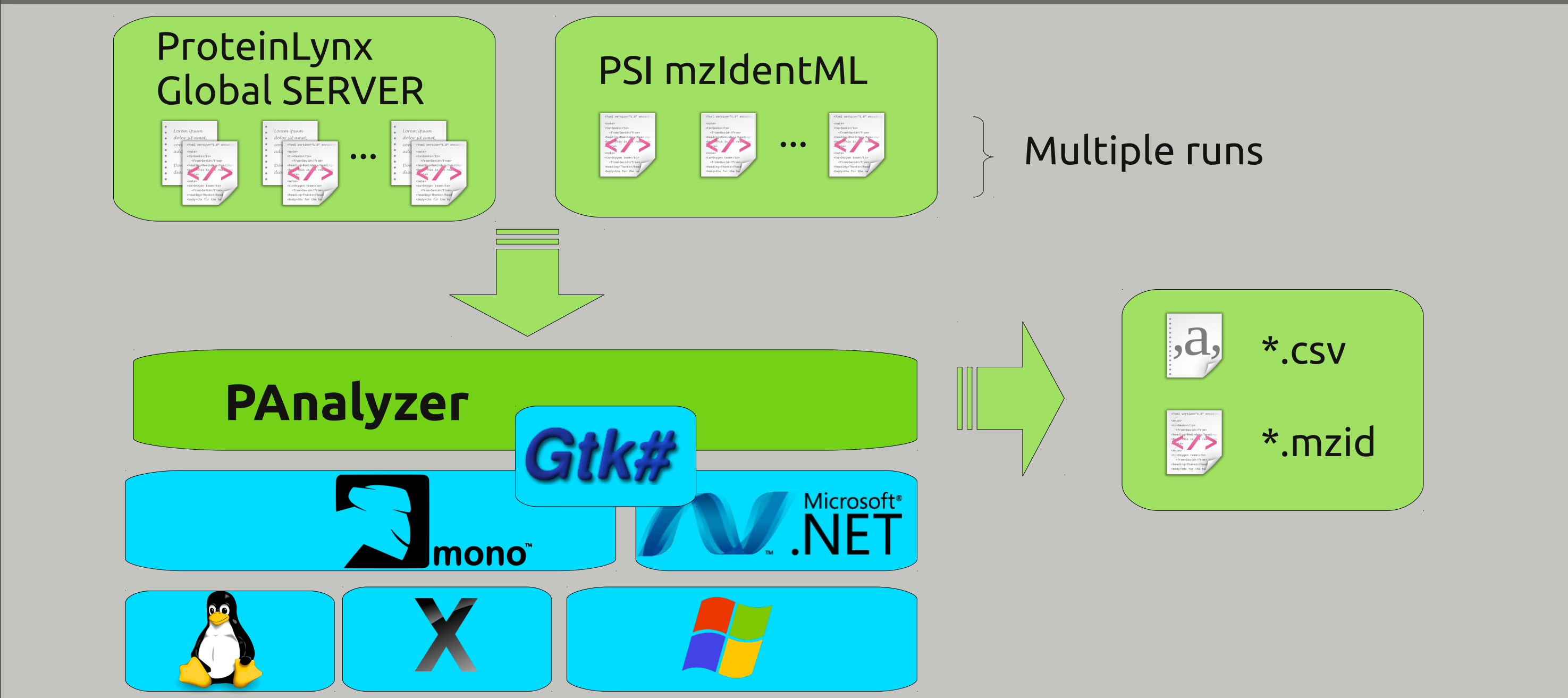
MULTIPLE RUN ANALYSIS

MERGING

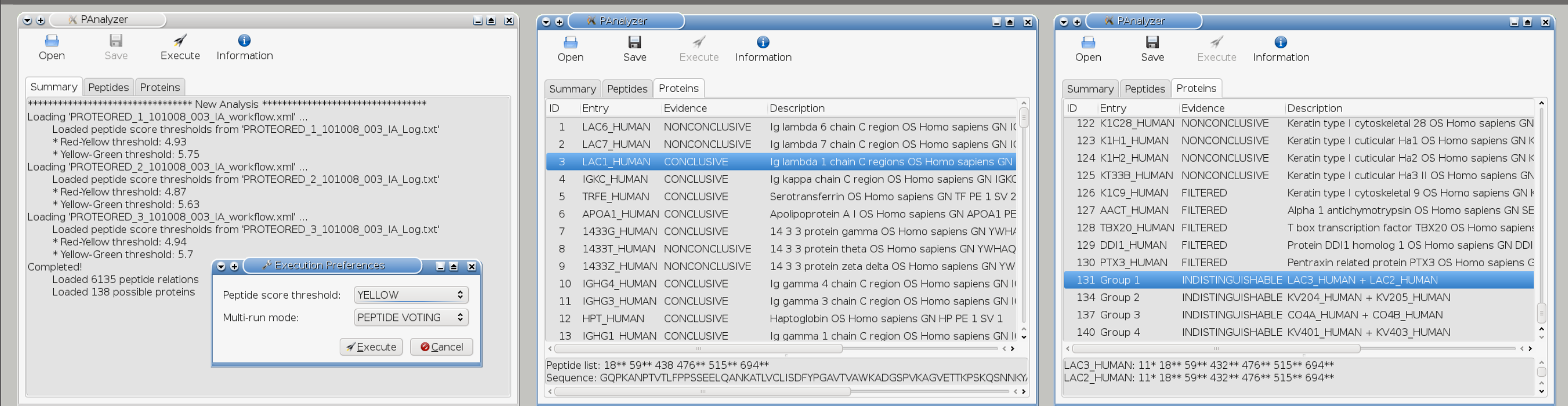


- Merging: all peptides are considered
- Peptide 3 is also included
- A, B, C conclusive; D non conclusive

Block diagram



Screenshots

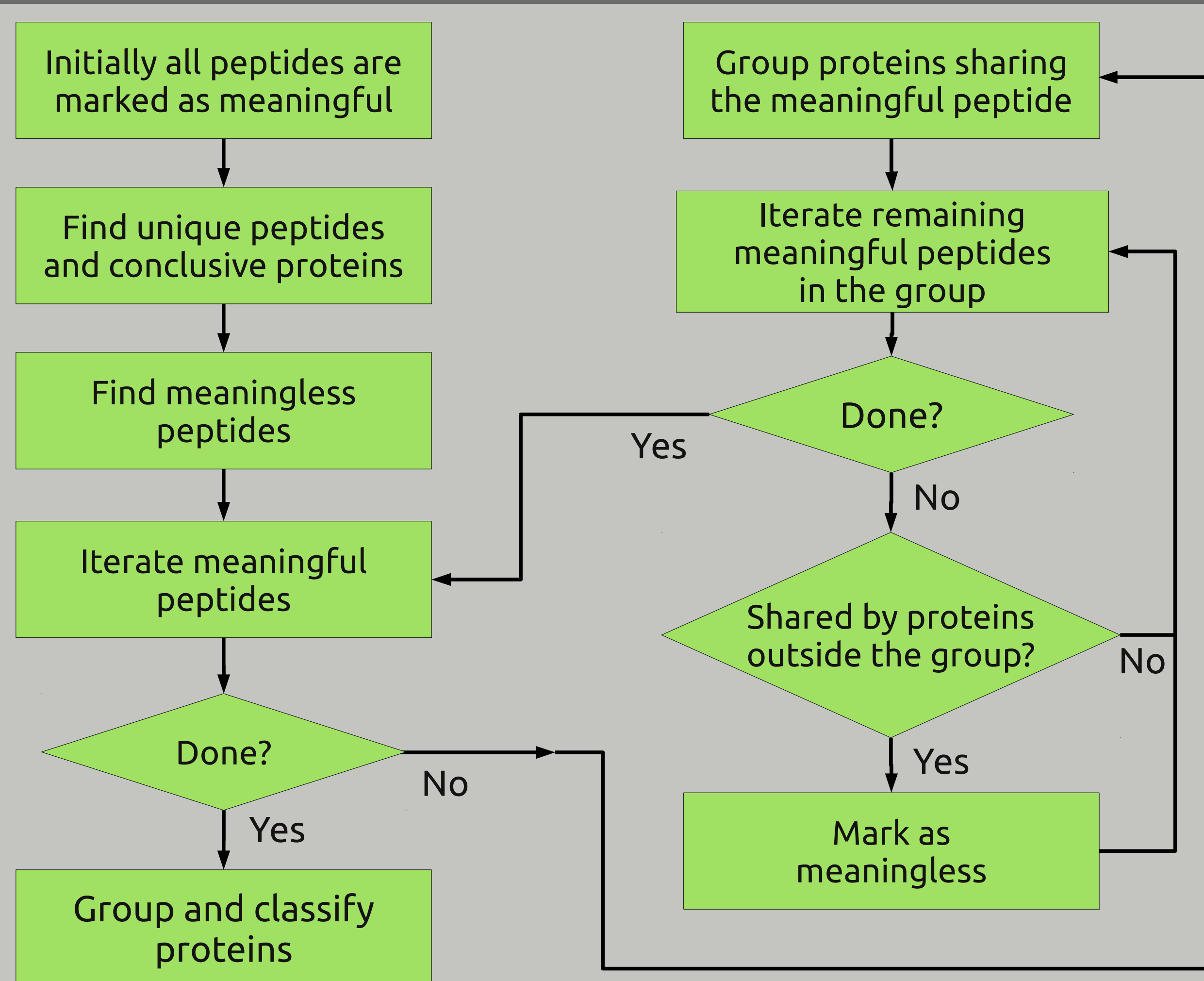


Plasma sample analysis

- MS^E based identification of proteins in a fraction of plasma prepared by ProteoRed-ISCIH working group 2

	SINGLE RUN			MULTIPLE RUN	
	RUN #1	RUN #2	RUN #3	VOTING	MERGING
ProteinLynx Global SERVER	106	108	122	149	
PAnalyzer					
Conclusive	84	93	87	72	127
Indistinguishable	4 (2 groups)	2 (1 groups)	4 (2 groups)	8 (4 groups)	0
Ambiguous group	0	0	0	0	0
Non conclusive	18	13	30	36	21
Filtered	0	0	1	33	1

Flow diagram



Conclusions

- A **software tool** has been developed to cope with protein inference ambiguities and classify proteins as: **conclusive**, **indistinguishable**, **ambiguous group**, and **non conclusive**
- Multirun** analysis is supported using two approaches:
 - Voting**: conservative approach
 - Merging**: less restrictive approach
- File formats supported are **mzIdentML** and ProteinLynx Global SERVER
- PAnalyzer is free software and licensed under the GNU **GPL**
- Cross-platform** interoperability is provided using .NET and GTK#

References and acknowledgements

(1) Nesvizhskii, A.I. and Aebersold, R. (2005). *Mol. Cell. Proteomics*. 4:1419-1440

This work was supported by the Basque Government Department of Industry (Etortek-IE09-2.56 to J.M.A.) Proteomics Core Facility-SGIKER is member of ProteoRed-ISCIH and is supported by UPV/EHU, MCINN, GV/EJ, ERDF and ESF