Non-target proteomics study of wastewater by ROIMCR procedure using polymeric passive samplers

Carlos Pérez López [a] , Damià Barceló [a] , Antoni Ginebreda [a] , Joaquin Abian [b], Montserrat Carrascal [b] and Romà Tauler [a]

[a] *Institute of Environmental Assessment and Water Studies (IDAEA-CSIC), Department of Environmental Chemistry, Jordi Girona 18-26, 08034 Barcelona, Spain*
[b] *Proteomics Laboratory CSIC/UAB, Institute of Biomedical Research of Barcelona, Spanish National Research Council (IIBB-CSIC/IDIBAPS), Rosellón 161, E-08036 Barcelona, Spain*

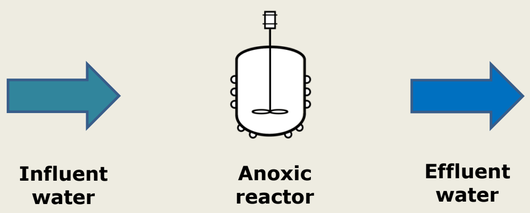
INTRODUCTION

Wastewater provides a large amount of information about the inhabitants of a territory, from consumption of illegal drugs to genetic biomarkers studies[1]. Currently, the latter is rising in interest, because of COVID-19 pandemia and fields like proteomics or metabolomics are becoming an important part of environmental, toxicological and epidemiological studies. Although, proteomics provides a large amount of environmental information, it is a new research area where identification and quantification technologies are being developed to perform a proper interpretation of environmental hazards. In this study, the ROIMCR procedure[2,3], a methodology developed recently for metabolomics studies, is proposed for non-target analysis in environmental proteomic studies.

METHODOLOGY

Experimental procedure

Polymer probes exposed to 3 different sites of the Gavà-Viladecans (Barcelona, Spain) wastewater treatment plant were analyzed:



Three devices were placed per sampling sites during 11 days

↓

Polymer surfaces were cut and purified
+
Trypsin digestion

Agilent 1200 HPLC system

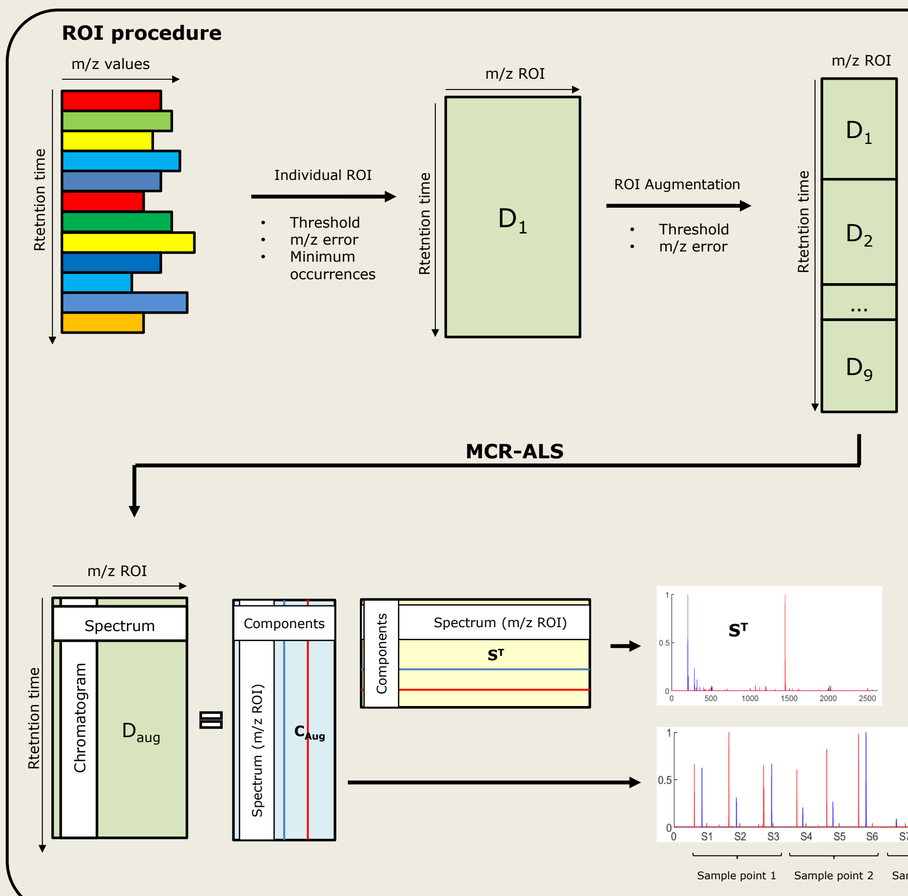
https://www.labx.com

Orbitrap XL

https://www.labx.com

Data analysis in MATLAB environment

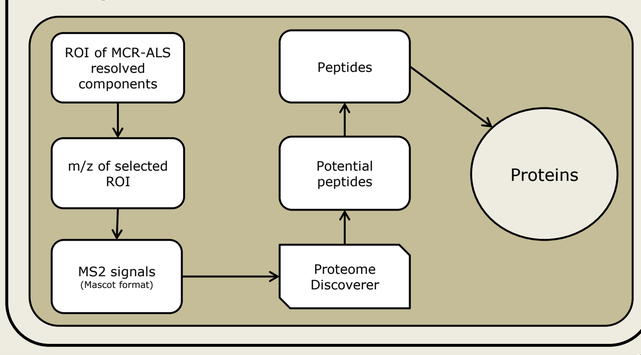
ROI procedure



Peptide identification

Result from ROIMCR analysis were employed to protein identification step whit the following characteristics.

- **Software**: Proteome Discoverer
- **Methodology**: Bottom-up peptide identification approach
- **Limitations**: Only those signals fragmented by previous authors have MS2 signals, necessary for protein identification
- **Procedure**: Only those matching signals between ROIMCR and previous results from MS2 fragmented signals were considered to be studied by Proteome Discoverer in Mascot format
- **Drawback**: There are signals from ROIMCR result that were not identified due to lack of MS2 signals.



RESULTS

ROIMCR

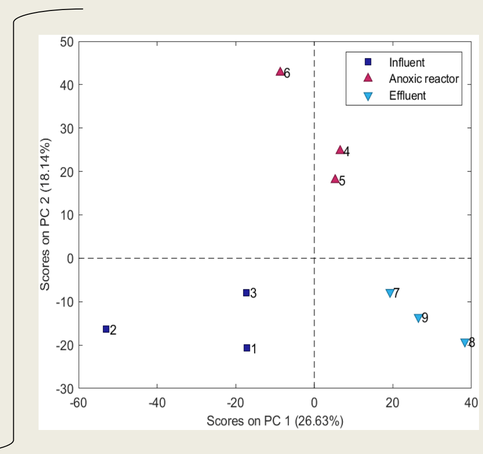
Individual regions of interest (ROI) were searched from full scan raw LC-HRMS data as a first step of ROI concatenation in column-wise augmented data matrices. Similar intensity threshold parameter were employed at the same sample WTP site (10^5, $5{\cdot}10^4$ and 10^4 respectively) and the same mass tolerance and minimum occurrences (with little changes) were used for every sample ROI calculation (0,001 a.u. and 10-12 occurrences respectively). Finally, between 492 and 883 ROI were obtained.

Point	Threshold	Mass Tolerance	ROI
Site 1	100000	0.005	1297
Site 2	50000	0.005	1325
Site 3	10000	0.005	1124
Total	**50000**	**0.005**	**2855**

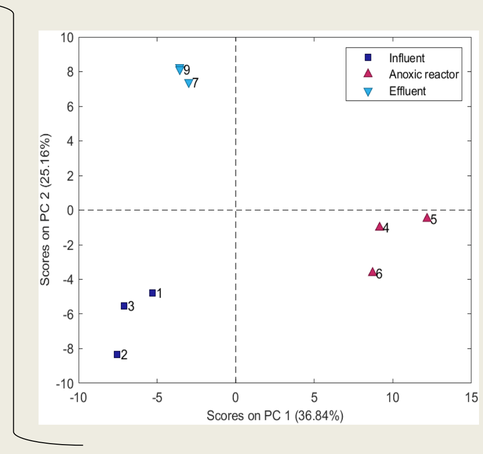
From ROI augmented matrix, a total number of 162 MCR components were resolved explaining 95.15% of variance.

Principal Component Analysis (PCA) was applied to the heights of the different components after a Standard Normal Variate (SNV) normalization step

PCA



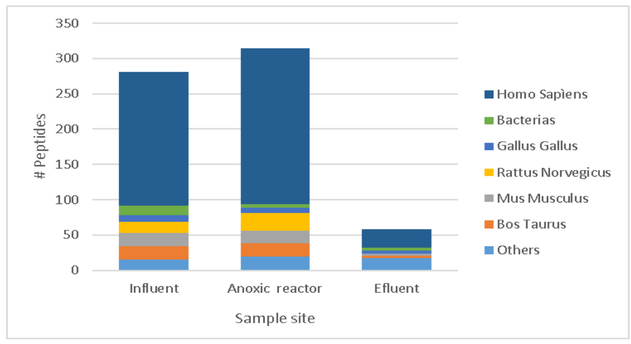
ROI data



MCR-ALS data

Peptide identification

From Proteome Discoverer results, the peptides were grouped by species they belong at each sampling site, obtaining 281, 314 and 58 respectively. This peptides are representatives of a large variety of species e.g human, chicken, rat, mouse or bacterial species among others.



Peptides were grouped taking into account all proteins they could belong to. Proteins presented below are the 10 most abundant in the 3 sampling sites.

Uniprot ID	Protein	Specie	kingdom	# Peptides		
				Influent	Anoxic reactor	Efluent
-	Keratins	Homo sapiens	Eukaryote	145	165	20
-	Keratins	Others	Eukaryote	24	47	15
P09093	Chymotrypsin-like elastase family member 3A	Homo sapiens	Eukaryote	25	22	-
Q99895	Chymotrypsin-C	Homo sapiens	Eukaryote	7	6	-
P00761	Trypsin	Sus scrofa	Eukaryote	2	3	2
P07477	Trypsin-1	Homo sapiens	Eukaryote	3	1	1
P01877	Immunoglobulin heavy constant alpha 2	Homo sapiens	Eukaryote	1	3	-
P21933	ATP synthase subunit beta	Streptococcus downei	Bacteria	-	3	-
P45574	Sulfite reductase, dissimilatory-type subunit alpha	Desulfovibrio vulgaris	Bacteria	3	-	-
P85917	Putative heat shock protein 2.	Pseudotsuga menziesii	Eukaryote	2	1	-

CONCLUSION

- ✓ The results confirm the similarity of the proteomic profiles of the samples from the same WWTP sampling site and the differences between the samples at the different sampling sites of the WWTP.
- ✓ Data compression and analysis performed by ROIMCR allows to reduce the amount of data without losing the own proteomic signal of each sample. ROIMCR improved PCA differentiation among different WWTP sites
- ✓ The results of protein identification coincide with some proteins among the most abundant proteins identified by previous authors[1]. Further work is pursued to study precision and accuracy of ROIMCR in proteomic studies.

ACKNOWLEDGEMENTS

This work has been financially supported by the Generalitat de Catalunya (Consolidated Research Group "2017 SGR 1404 - Water and Soil Quality Unit"), by the Spanish Ministry of Science and Innovation (Project CEX2018-000794-S, PID2019-105732GB) and the IBERAQUA-NET project (RED2018-102737-T)

REFERENCES

1. M. Carrascal, J. Abian, A. Ginebreda, D. Barceló. Discovery of large molecules as new biomarkers in wastewater using environmental proteomics and suitable polymer probes. 2020. Sci. Total Environ. 747, 141145
2. Gorrochategui E, Jaumot J, Tauler R (2019) ROIMCR: a powerful analysis strategy for LC-MS metabolomic datasets, BMC Bioinformatics 17;20(1):256.
3. Jaumot J, de Juan A, Tauler R (2015) MCR-ALS GUI 2.0: New features and applications. Chemom Intell Lab Syst 140:1–12.