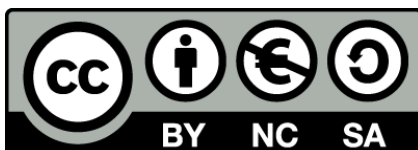




UNIVERSITAT DE
BARCELONA

**Avaluació dels efectes dels factors ambientals
en el metaboloma de la *Daphnia magna* mitjançant
mètodes cromatogràfics i quimiomètrics**

Elba Garreta Lara



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- NoComercial – CompartirIgual 4.0. Espanya de Creative Commons**.

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - NoComercial – CompartirIgual 4.0. España de Creative Commons**.

This doctoral thesis is licensed under the **Creative Commons Attribution-NonCommercial-ShareAlike 4.0. Spain License**.



UNIVERSITAT DE
BARCELONA

**Avaluació dels efectes dels factors ambientals en
el metaboloma de la *Daphnia magna* mitjançant
mètodes cromatogràfics i quimiomètrics**

Elba Garreta Lara



UNIVERSITAT DE
BARCELONA



Avaluació dels efectes dels factors ambientals en el
metaboloma de la *Daphnia magna* mitjançant
mètodes cromatogràfics i quimiomètrics

Elba Garreta Lara



UNIVERSITAT DE
BARCELONA



FACULTAT DE QUÍMICA
DEPARTAMENT D'ENGINYERIA QUÍMICA I QUÍMICA ANALÍTICA

Programa de Doctorat:
Química Analítica i Medi Ambient

Avaluació dels efectes dels factors ambientals en el metaboloma de la *Daphnia magna* mitjançant mètodes cromatogràfics i quimiomètrics

Memòria presentada per

Elba Garreta Lara

per optar al grau de Doctora per la Universitat de Barcelona

Directors:

Dr. Romà Tauler Ferré

Dra. Sílvia Lacorte Bruguera

Departament de Química Ambiental

Institut de Diagnosi Ambiental i Estudis de l'Aigua (IDAEA)

Consejo Superior de Investigaciones Científicas (CSIC)

Tutora:

Dra. M^a Teresa Galcerán Huguet

Departament d'Enginyeria Química i Química Analítica

Universitat de Barcelona (UB)

En **Dr. Romà Tauler Ferré**, professor d'Investigació del Departament de Química Ambiental de l'Institut de Diagnosi Ambiental i Estudis de l'Aigua adscrit al Consejo Superior de Investigaciones Científicas, i na **Dra. Sílvia Lacorte Bruguera**, professora del mateix departament.

FAN CONSTAR,

Que la present memòria titulada: "**Avaluació dels efectes dels factors ambientals en el metaboloma de la *Daphnia magna* mitjançant mètodes cromatogràfics i quimiomètrics**", ha estat realitzada sota la nostra direcció per na Sra. **Elba Garreta Lara** i que tots els resultats són fruit de les experiències realitzades per la citada doctoranda en el Departament de Química Ambiental de l'Institut de Diagnosi Ambiental i Estudis de l'Aigua adscrit al Consejo Superior de Investigaciones Científicas.

I per tal de que així consti expedeixen el present certificat.

Barcelona, Setembre de 2019.

Dr. Romà Tauler Ferré

Dra. Sílvia Lacorte Bruguera

Als meus,

AGRAÏMENTS

Ara que ja s'apropa el final d'aquesta etapa és moment d'agrair a totes aquelles persones amb qui he compartit aquests més de cinc anys de camí. Tots ells han aportat el seu petit granet de sorra, i estic convençuda que sense ells no hagués estat el mateix i per això us vull donar les gràcies de tot cor.

En primer lloc, m'agradaria agrair als meus dos directors per confiar en mi i haver-me donat l'oportunitat de formar part d'aquest bonic projecte CHEMAGEB, el Dr. Romà Tauler i la Dra. Sílvia Lacorte. Gràcies pel vostre suport, la vostra dedicació incondicional i per compartir els vostres coneixements al llarg d'aquests anys. També voldria agrair a la Dra. Maria Teresa Galcerán l'haver acceptat ser la meva tutora durant la Tesi.

Aquesta Tesi tampoc hagués estat possible sense l'ajuda del Dr. Carlos Barata i el Dr. Bruno Campos. Gràcies per donar-me a conèixer el món de les *daphnies*, pel vostre esforç, la vostra dedicació al llarg d'aquest temps i sobretot per les vostres aportacions en la part més biològica d'aquesta Tesi.

Gràcies a tots els companys del CSIC passats, presents i als que hem deixat el relleu (Alejandro, Aline, Amrita, Carma, Clàudia, Cristian, Gabino, Hadi, Helena, Igor, Joaquim, Marc, Marta, Mireia, Miriam, Rita, Stefan, Xin, Yahya) per tots els moments dins i fóra del laboratori, pels dinars, cafès, congressos, excursions, etc. Per ser una font de coneixement i per compartir-los. També agrair d'una manera especial als PhD Chemageb: Eva, Núria, Xisco, Elena, Meri i Víctor, sense vosaltres no estaria escrivint aquestes paraules. Mil gràcies per tot, m'emporto uns amics per tota la vida amb qui espero seguir compartint molts moments.

I would also like to acknowledge Prof. Craig Wheelock and Dr. Toni Checa for the opportunity to be a part of the group in the Karolinska Institutet during my PhD internship in Stockholm. Thank you so much for your support, guidance and your hospitality. També vull agrair a la Cristina Gómez, la Cristina López i el Luciano per tots els moments viscuts durant els mesos que vam compartir, pels dinars de tupper, pels cafès de la màquina de l'Starbucks, pels concerts, pels sopars, per les cerveses, pels

sushis, pel super viatge exprés a Göteborg, per les bogeries, pels consells i per fer de l'estada una experiència que sempre portaré amb mi. P.D. Benvinguts a l'aerobuuus! :)

També vull donar les gràcies a totes les persones que han estat indirectament involucrades en aquesta Tesi, companyes i amigues de l'hoquei passades i presents, companyes i companys d'Anapharm i amigues les quals han patit els nervis, els dies i les setmanes d'estrès i totes les meves queixes i els mals de caps relacionats amb aquesta Tesi. A totes i tots vosaltres gràcies per acompanyar-me en aquest camí i donar-me els ànims necessaris per passar els dies grisos i seguir treballant amb ganes.

Per últim, gràcies a la família, àvies, tietes i cosins gràcies per creure en mi i fer-me sentir que sóc capaç del millor. I en especial als pares i al Nil, gràcies per sempre donar-me tot, pel vostre suport incondicional, per inculcar-me les ganes de seguir treballant i tirar endavant, pels ànims en els moment més complicats i per ajudar-me a créixer i ser millor persona. Gràcies perquè sense vosaltres no seria la persona que sóc.

ÍNDEX

RESUM.....	V
ABREVIACIONS I ACRÒNIMS.....	VII
OBJECTIUS.....	XI
ESTRUCTURA DE LA TESI	XIII
<u>CAPÍTOL 1: INTRODUCCIÓ.....</u>	15
1.1 ÒMICA AMBIENTAL.....	17
1.1.1 METABOLÒMICA	21
1.2 ORGANISME MODEL: <i>DAPHNIA MAGNA</i>.....	24
1.2.1 LA <i>DAPHNIA MAGNA</i> COM A MODEL DE TOXICITAT	25
1.2.2 ESTUDIS DE METABOLÒMICA UTILITZANT LA <i>DAPHNIA MAGNA</i>	26
1.3 ESTRESSANTS AMBIENTALS	36
1.3.1 ESTRÈS ABIÒTIC	36
1.3.2 CONTAMINANTS AMBIENTALS	38
1.4 ESTRATÈGIES I PLATAFORMES ANALÍTIQUES BASADES EN L'ESPECTROMETRIA DE MASSES	42
1.4.1 ANÀLISI DIRIGIDA	44
1.4.2 ANÀLISI NO DIRIGIDA.....	46
1.5 ANÀLISI DE LES DADES.....	50
1.5.1 PREPROCESSAMENT DE LES DADES.....	53
1.5.2 MÈTODES QUIMIOMÈTRICS DE TRACTAMENT DE DADES.....	62
1.6 DESCOBRIMENT DELS MARCADORS BIOLÒGICS.....	80

<u>CAPÍTOL 2: APLICACIÓ DE LA CROMATOGRÀFIA DE GASOS EN L'ESTUDI DEL METABOLOMA DE LA <i>D. MAGNA</i> EXPOSADA A FACTORS ABIÒTICS</u>	83
2.1 INTRODUCCIÓ	85
2.2 ARTICLE CIENTÍFIC I: METABOLIC PROFILING OF <i>DAPHNIA MAGNA</i> EXPOSED TO ENVIRONMENTAL STRESSORS BY GC-MS AND CHEMOMETRIC TOOLS	91
2.3 ARTICLE CIENTÍFIC II: COMBINED EFFECTS OF SALINITY, TEMPERATURE AND HYPOXIA ON <i>DAPHNIA MAGNA</i> METABOLISM	119
2.4 ARTICLE CIENTÍFIC III: CHEMOMETRIC ANALYSIS OF COMPREHENSIVE TWO DIMENSIONAL GAS CHROMATOGRAPHY–MASS SPECTROMETRY METABOLOMICS DATA	137
2.5 ARTICLE CIENTÍFIC IV: CHEMOMETRICS COMPARISON OF GAS CHROMATOGRAPHY WITH MASS SPECTROMETRY AND COMPREHENSIVE TWO-DIMENSIONAL GAS CHROMATOGRAPHY WITH TIME-OF-FLIGHT MASS SPECTROMETRY <i>DAPHNIA MAGNA</i> METABOLIC PROFILES EXPOSED TO SALINITY...	155
2.6 DISCUSSIÓ DELS RESULTATS	182
2.6.1 OPTIMITZACIÓ DE LES CONDICIONS ANALÍTiques.....	183
2.6.2 ANÀLISI DE LES MOSTRES	190
2.6.3 PRETRACTAMENT I TRACTAMENT DE LES DADES.....	192
2.6.4 RELLEVÀNCIA BIOLÒGICA.....	196
<u>CAPÍTOL 3: ESTUDIS DE METABOLÒMICA EN <i>D. MAGNA</i> EXPOSADA A CONTAMINANTS AMBIENTALS MITJANÇANT LC-MS</u>	201
3.1 INTRODUCCIÓ	203
3.2 ARTICLE CIENTÍFIC V: EFFECTS OF PSYCHIATRIC DRUGS ON <i>DAPHNIA MAGNA</i> OXYLIPIN PROFILES	207
3.3 ARTICLE CIENTÍFIC VI: NOVEL COMBINED APPROACH FOR THE TARGET/UNTARGET CHARACTERIZATION OF <i>DAPHNIA MAGNA</i> METABOLOME	223
3.4 DISCUSSIÓ DELS RESULTATS	263
3.4.1 ESTRATÈGIA DIRIGIDA	263
3.4.2 ESTRATÈGIA DIRIGIDA I NO DIRIGIDA.....	266
CONCLUSIONS	273
REFERÈNCIES	279

RESUM

Durant el segle XX i principis del XXI, el clima i els ecosistemes han patit canvis apreciables com a conseqüència directa de l'activitat humana. Els impactes naturals i antropogènics juntament amb els efectes del canvi climàtic han esdevingut un problema urgent per resoldre. Aquestes alteracions químiques i físiques de l'entorn i dels ecosistemes aquàtics produeixen canvis en els sistemes metabòlics dels organismes vius. En el camp de la metabolòmica ambiental, l'estudi i l'anàlisi d'aquests canvis proporciona informació significativa sobre les respostes biològiques a aquestes alteracions ambientals. En aquest context, els enfocaments basats en l'espectrometria de masses (MS) han estat clau en el ràpid desenvolupament de les tecnologies metabolòmiques, ja que han proporcionat un nou enfocament per investigar i avaluar els efectes dels productes químics i altres factors abiòtics en la complexitat molecular dels sistemes biològics. En aquest context, aquesta Tesi s'ha centrat en el desenvolupament de mètodes analítics de metabolòmica dirigida i no dirigida basats en tècniques de separació acoblades a la MS, com per exemple la cromatografia de gasos i la cromatografia de líquids (GC-MS i LC-MS, respectivament), i en l'aplicació dels mètodes quimiomètrics per tal de determinar, identificar i avaluar els canvis produïts en el metaboloma de la *D. magna* a causa de l'exposició a un estressant ambiental.

Per una banda, s'han desenvolupat i optimitzat els mètodes analítics de GC-MS i cromatografia de gasos bidimensional acoblada a l'espectrometria de masses (GCxGC-MS) i de tractament de les dades multivariants per a estudis de metabolòmica no dirigida. En concret, s'han optimitzat les condicions experimentals d'extracció i de derivatització i s'han avaluat diferents tipus de columnes cromatogràfiques per la determinació dels metabòlits presents en la *D. magna*. Per altra banda, s'han presentat dues aproximacions per la compressió dels grans conjunts de dades que es generen per tal d'extreure la informació biològica rellevant i facilitar-ne la seva interpretació. D'aquesta manera, s'ha demostrat la gran eficàcia i la robustesa del mètode quimiomètric de resolució multivariant de corbes per mínims quadrats alternats (MCR-ALS) per la resolució dels perfils d'elució i dels espectres de masses dels metabòlits

presentes en les mostres de metabolòmica de *D. magna* analitzades mitjançant GC-MS i GCxGC-MS. A més, s'han avaluat els efectes de les exposicions a diferents factors abiòtics com la salinitat, la temperatura i la hipòxia en el metaboloma de la *D. magna*. L'aplicació del disseny experimental per a l'estudi dels factors d'exposició ha permès avaluar les possibles interaccions entre els diferents tractaments i llurs combinacions. Els resultats han demostrat la gran influència de la salinitat en el perfil metabolòmic de la *D. magna*, la qual està relacionada amb el metabolisme dels aminoàcids, de les proteïnes, dels sucres, dels àcids grassos i amb el metabolisme energètic, com el cicle de Krebs.

Finalment, s'han presentat els resultats de l'avaluació de les alteracions de diversos contaminants ambientals mitjançant les estratègies dirigida i no dirigida per a l'estudi del metaboloma de la *D. magna*. L'anàlisi dirigida ha permès la quantificació absoluta dels nivells de les oxilipines i l'avaluació dels efectes de la fluoxetina, el diazepam i la carbamazepina en el metaboloma de la *D. magna*. Els resultats han demostrat una influència de les exposicions als productes farmacèutics en els nivells de diversos derivats de l'àcid araquidònic, l'àcid linoleic, l'àcid α -linoleic i àcid eicosapentaenoic. Per altra banda, s'ha realitzat una caracterització exhaustiva dels efectes del bisfenol A en els perfils metabolòmics de la *D. magna* mitjançant la combinació de les estratègies dirigida i no dirigida. S'ha realitzat una interpretació detallada i quantitativa dels metabòlits afectats per l'exposició a bisfenol A mitjançant l'estratègia dirigida. Mentre que, en l'anàlisi no dirigida s'ha aplicat la compressió de les dades mitjançant la selecció de les regions d'interès (ROIs) i la resolució mitjançant MCR-ALS a causa de la complexitat de les dades obtingudes per cromatografia de líquids acoblada a l'espectrometria de masses d'alta resolució (LC-HRMS). La interpretació biològica de les alteracions produïdes a causa de l'exposició al bisfenol A estan relacionades amb el metabolisme del carboni i amb una disminució dels nivells de glutamat, el qual juga un paper central en el metabolisme dels aminoàcids.

ABREVIACIONS I ACRÒNIMS

α -LA	àcid α -linoleic, <i>linoleic acid</i>
1D	unidimensional, <i>one-dimensional</i>
2D	bidimensional, <i>two-dimensional</i>
AA	àcid araquidònic, <i>arachidonic acid</i>
ADN	àcid desoxiribonucleic, <i>deoxyribonucleic acid</i>
ALS	mínims quadrats alternats, <i>alternating least squares</i>
ANOVA	anàlisi de la variància, <i>analysis of variance</i>
ASCA	ANOVA - anàlisi simultània de components, <i>ANOVA-Simultaneous Component Analysis</i>
AsLS	mínims quadrats asimètrics, <i>asymetric least squares</i>
BPA	bisfenol A, <i>bisphenol A</i>
CAWG	grup de treball d'anàlisi química, <i>chemical analysis working group</i>
CE-MS	electroforesis capil·lar acoblada a espectrometria de masses, <i>capillary electrophoresis coupled to mass spectrometry</i>
CEC	contaminants ambientals emergents, <i>contaminants of emerging concern</i>
COW	deformació optimitzada de correlació, <i>correlation optimized warping</i>)
COX	<i>cyclooxygenase</i>
CV	validació creuada, <i>cross validation</i>
CYP	<i>cytochrome 450</i>
DIMS	espectrometria de masses en infusió directa, <i>direct infusion mass spectrometry</i>
DoE	disseny experimental, <i>design of experiment</i>
DOM	matèria orgànica dissolta, <i>dissolved organic matter</i>
EPA	àcid eicosapentaenoic, <i>eicosapentaenoic acid</i>
FN	falsos negatius, <i>false negative</i>
fnnls	mínims quadrats ràpids no negatius, <i>fast non-negative least squares</i>
FP	falsos positius, <i>false positive</i>
FT-ICR-MS	ressonància ió ciclotró de transformada de Fourier acoblada a espectrometria de masses, <i>Fourier-transform ion cyclotron resonance coupled to mass spectrometry</i>
GC	cromatografia de gasos, <i>gas chromatography</i>

GC-MS	cromatografia de gasos acoblada a espectrometria de masses, <i>gas chromatography coupled to mass spectrometry</i>
GCxGC	cromatografia de gasos bidimensional, <i>two-dimensional gas chromatography</i>
GCxGC-MS	cromatografia de gasos bidimensional acoblada a espectrometria de masses, <i>two-dimensional gas chromatography coupled to mass spectrometry</i>
GMD	<i>Golm Metabolome Database</i>
HILIC	<i>cromatografia d'interacció hidrofílica, hydrophilic interaction chromatography</i>
HMDB	<i>Human Metabolome Database</i>
HPLC	<i>cromatografia de líquids d'alta eficàcia, high performance liquid chromatography</i>
HRMS	espectrometria de masses d'alta resolució, <i>high resolution mass spectrometry</i>
KEGG	enciclopèdia de Kyoto de gens i genomes, <i>Kyoto encyclopedia of genes and genomes</i>
LA	àcid linoleic, <i>linoleic acid</i>
LC-MS	cromatografia de líquids, <i>liquid chromatography</i>
<i>lof</i>	falta d'ajust, <i>lack of fit</i>
<i>log P</i>	coeficient de partició, <i>partition coefficient</i>
LOOCV	validació creuada deixant-ne un fora, <i>leave-one-out cross validation</i>
LOX	<i>lipoxigenase</i>
LV	<i>variable latent, latent variable</i>
MCC	coeficient de correlació de Matthews, <i>Matthew's correlation coefficient</i>
MCR	resolució multivariant de corbes, <i>multivariate curve resolution</i>
MCR-ALS	resolució multivariant de corbes per mínims quadrats alternats, <i>multivariate curve resolution by alternating least squares</i>
MeOX	metoximació, <i>methoximation</i>
MF	factor de coincidència, <i>match factor</i>
MOA	mode d'acció, <i>mode of action</i>
MRM	<i>multiple reaction monitoring</i>
MS	espectrometria de masses, <i>mass spectrometry</i>
MS/MS	espectrometria de masses en tàndem, <i>tandem mass spectrometry</i>
MSI	espectroscòpia d'imatges de MS, <i>mass spectrometry imaging</i>

MSTFA	N-metil-N-(trimetilsilil) trifluoroacetamida, <i>N-methyl-N-(trimethylsilyl) trifluoroacetamide</i>
NIST	institut nacional d'estàndards i tecnologia, <i>national institute of science and technology</i>
npls	mínims quadrats no negatius, <i>non negative least squares</i>
OECD	organització per la cooperació econòmica i desenvolupament, <i>organisation for economic co-operation and development</i>
OPEs	esters d'organofosfats, <i>organophosphate ester</i>
OPLS-DA	anàlisi discriminant per mínims quadrats parcials ortogonals, <i>orthogonal partial least squares by discriminant analysis</i>
OSC	correcció ortogonal del senyal, <i>orthogonal signal correction</i>
PAHs	compostos aromàtics policíclics, <i>polycyclic-aromatic hydrocarbons</i>
PC	component principal, <i>principal component</i>
PCA	anàlisi de components principals, <i>principal component analysis</i>
PCBs	bifenils policlorats, <i>polychlorinated biphenyls</i>
PFOS	àcid perfluorooctanosulfònic, <i>perfluorooctanesulfonic acid</i>
PG	prostaglandines, <i>prostaglandins</i>
PLS	mínims quadrats parcials, <i>partial least squares</i>
PLS-DA	anàlisi discriminant per mínims quadrats parcials, <i>partial least squares by discriminant analysis</i>
ppb	partícules per bilió, <i>parts per billion</i>
PUFAs	àcids grassos poliinsaturats, <i>polyunsaturated fatty acids</i>
QC	control de qualitat, <i>quality control</i>
QDs	punts quàntics, <i>quantic dots</i>
QqQ	triple quadrupol, <i>triple quadrupole</i>
R ²	percentatge de la variància explicada, <i>percentage of explained variance</i>
RI	índex de retenció, <i>retention indices</i>
RMF	factor de coincidència inversa, <i>reverse match factor</i>
RMN	ressonància magnètica nuclear, <i>nuclear magnetic resonance</i>
RMSECV	error quadràtic mig de validació creuada, <i>root mean square error of cross-validation</i>
ROI	regions d'interès, <i>regions of interest</i>
RPLC	cromatografia de fase invertida, <i>reversed-phase liquid chromatography</i>

SCA	anàlisi simultània de components, <i>simultaneous component analysis</i>
SIM	<i>selected ion monitoring</i>
SIMPLISMA	anàlisi iterativa d'auto modelatge senzilla d'utilitzar, <i>SIMPL</i> e-to-use <i>iterative self-modeling analysis</i>
SNR	relació senyall soroll, <i>signal to noise ratio</i>
SR	quocients de selectivitat, <i>selectivity ratio</i>
SSRI	inhibidor selectiu del receptor de la serotonina, <i>selective serotonin reuptak inhibitor</i>
SVD	descomposició de valors singulars, <i>singular value decomposition</i>
TBDMS	tert-butildimetilsilil, <i>tert-butyldimethylsilyl</i>
TBOEP	tris(2-butoxietil) fosfat, <i>tris(2-butoxyethyl) phosphate</i>
TCEP	tris(2-cloroetil) fosfat, <i>tris(2-carboxyethyl) phosphine</i>
TIC	cromatogrames d'ions totals, <i>total ion current chromatogram</i>
TMCS	trimetilclorosilà, <i>trimethylchlorosilane</i>
TMS	trimetilsilil, <i>trimethylsilyl</i>
TN	vertaders negatius, <i>true negative</i>
TOF-MS	temps de vol acoblat a espectrometria de masses, <i>time-of-flight mass spectrometry</i>
TP	vertaders positius, <i>true positive</i>
TPhP	trifenil fosfat, <i>triphenyl phosphate</i>
VIP	variables importants en projecció, <i>variable importance in projection</i>
VOCs	compostos orgànics volàtils, <i>volatile organic compounds</i>

NOTACIÓ

En aquesta secció es descriu la notació matemàtica utilitzada en aquesta Tesi. Aquesta és l'acceptada per la comunitat científica.

Les lletres minúscules cursives (per exemple, *x*) indiquen escalars. Les lletres minúscules en negreta (per exemple, ***x***) indiquen vectors. Les lletres majúscules en negreta (per exemple, ***X***) indiquen matrius. Les lletres negreta i subratllades (per exemple, ***X***) indiquen cubs. La transposició d'una matriu s'indica amb una "T" com a superíndex (per exemple, ***X***^T).

OBJECTIUS

En els darrers anys, els impactes del canvi climàtic i de l'alliberament i acumulació de productes químics en els ecosistemes ha esdevingut un problema per als organismes vius i els sistemes ecològics. Aquestes pressions ambientals afecten els sistemes biològics a nivell metabòlic. Els avenços recents en les tècniques òmiques, i principalment les basades amb l'espectrometria de masses permeten identificar i quantificar els canvis en l'expressió dels metabòlits d'un sistema biològic concret. Tot i això, a causa de la mida i la complexitat de la informació generada per aquestes plataformes analítiques és necessari un tractament de les dades mitjançant eines quimiomètriques.

Aquesta Tesi ha estat realitzada en el marc del projecte europeu CHEMAGEB ja finalitzat (*CHEMometric and High-throughput Omics Analytical Methods for Assessment of Global Change Effects on Environmental and Biological Systems, EU Advanced Grant Nr. 320737*). L'objectiu d'aquest projecte ha estat el desenvolupament d'un nou enfocament interdisciplinari de recerca en tres grans àrees, la quimiometria, les tecnologies òmiques i la toxicologia ambiental. Així, a partir d'aquestes eines s'han avaluat els efectes de la contaminació química i d'altres pertorbacions relacionades amb el canvi global sobre diferents organismes biològics model i així proporcionar noves eines per a l'avaluació del risc ambiental. En aquest context, aquesta Tesi s'ha centrat en l'avaluació dels efectes de diversos estressants ambientals, factors abiòtics i contaminants ambientals (salinitat, temperatura, oxigen dissolt en aigua, bisfenol A, fluoxetina, diazepam i carbamazepina) en el metaboloma de la *Daphnia magna* a partir del desenvolupament i l'aplicació de mètodes analítics de metabolòmica dirigida i no dirigida que utilitzen l'espectrometria de masses i les eines quimiomètriques de tractament de les dades.

OBJECTIUS ANALÍTICS

- Desenvolupament i optimització de les condicions d'extracció i derivatització per la determinació dels metabòlits presents en la *D. magna* mitjançant la cromatografia de gasos acoblada a l'espectrometria de masses (GC-MS).

- Desenvolupament i optimització de les condicions experimentals de la cromatografia de gasos unidimensional i bidimensional acoblada a l'espectrometria de masses (GC-MS i GCxGC-MS) per a estudis de metabolòmica no dirigida de la *D. magna*.
- Optimització de les condicions experimentals de la cromatografia de líquids acoblada a l'espectrometria de masses (LC-MS) per a l'estudi del metaboloma de la *D. magna* mitjançant les estratègies dirigides i no dirigides.

OBJECTIUS QUIMIOMÈTRICS

- Desenvolupament i avaluació de diferents metodologies d'anàlisi de les dades metabolòmiques no dirigides obtingudes mitjançant GC-MS i GCxGC-MS, basades especialment en el mètode de la resolució multivariant de corbes per mínims quadrats alternats (MCR-ALS).
- Aplicació de la metodologia ROIMCR mitjançant la selecció de les regions d'interès (ROIs) per la compressió i la resolució mitjançant MCR-ALS de les dades de metabolòmica obtingudes mitjançant l'estratègia no dirigida en LC-HRMS.

OBJECTIUS TOXICOLÒGICS

- Avaluació dels efectes dels diferents estressants ambientals, factors abiòtics i contaminants ambientals (salinitat, temperatura, oxigen dissolt en aigua, bisfenol A, fluoxetina, diazepam i carbamazepina) en el metabolisme de la *D. magna* a partir d'estudis de metabolòmica dirigida i no dirigida, i mitjançant mètodes quimiomètrics.

ESTRUCTURA DE LA TESI

La present memòria està estructurada en quatre capítols els quals es descriuen a continuació.

En el primer capítol s'introdueix de manera general la òmica ambiental, més en concret els estudis de metabolòmica, les estratègies i plataformes analítiques i els mètodes quimiomètrics que s'han emprat en els diferents articles científics que es presenten en aquesta Tesi. Es descriu l'organisme model estudiat i es realitza un recopilatori dels estudis metabolòmics previs on s'ha emprat la *D. magna*. Es detallen les estratègies dirigides i no dirigides emprades en estudis de metabolòmica amb espectrometria de masses i s'expliquen els mètodes quimiomètrics emprats pel pretractament i tractament de les dades.

En el segon capítol es presenten els resultats del desenvolupament i l'optimització de l'estratègia no dirigida dels perfils metabòlics mitjançant GC-MS i GCxGC-MS en l'estudi del metaboloma de la *D. magna*. Per la GC-MS es detalla l'optimització de les condicions analítiques d'extracció i derivatització per a l'anàlisi de les mostres de *D. magna*. Per la GCxGC-MS s'avaluen diferents tipus de columnes cromatogràfiques per la determinació i la identificació dels metabòlits presents en la *D. magna*. També es descriuen les diferents estratègies de pretractament i tractament de les dades de GC-MS i de GCxGC-MS en l'anàlisi de les dades de metabolòmica no dirigides. Finalment, s'avaluen els efectes de les exposicions a diferents factors abiòtics (salinitat, temperatura i hipòxia) en el metaboloma de la *D. magna*.

En el tercer capítol s'avaluen els efectes dels contaminants ambientals, com el bisfenol A, la fluoxetina, el diazepam i la carbamazepina. A més, també es mostren els resultats de l'optimització de les condicions experimentals de les estratègies dirigides i no dirigides per l'estudi del metaboloma de la *D. magna* mitjançant la LC-MS. L'estratègia dirigida permet la identificació i la quantificació de les oxilipines i els eicosanoids derivats dels mediadors lipídics dels àcids grassos poliinsaturats (PUFAs) de la *D. magna*. Per altra banda, la combinació de les dues estratègies permet una caracterització més completa del metaboloma de la *D. magna*. L'estratègia dirigida

permet l'optimització de les condicions analítiques i l'anàlisi quantitativa en termes absoluts dels metabòlits predeterminats. En contrast, l'estratègia no dirigida, mitjançant l'aplicació de la metodologia ROIMCR en l'anàlisi de les dades LC-HRMS *full scan*, permet avaluar d'una manera més completa el perfil metabòlic de la *D. magna*. I realitzar una estimació quantitativa relativa dels canvis de concentració dels metabòlits detectats en diferents mostres. Finalment, un cop coneguts o identificats els diferents metabòlits de la *D. magna*, es determinen els efectes que tenen els diferents contaminants ambientals.

En la darrera part, es recullen les conclusions generals més importants d'aquesta Tesi.

CAPÍTOL 1



INTRODUCCIÓ

1.1 ÒMICA AMBIENTAL

Amb els inicis del Projecte del Genoma Humà, la ciència va entrar en l'era de la biologia postgènica, amb un creixement exponencial de les tècniques anomenades òmiques. Aquesta nova perspectiva dels sistemes biològics, en quant a la manera d'explorar la funció dels gens, va representar un nou i potent enfocament multidisciplinari que va permetre, amb una major capacitat predictiva, la identificació de marcadors biològics. Les ciències òmiques han estat impulsades sobretot arran de la seva àmplia gama d'aplicacions en àmbits tant diversos com les ciències mèdiques, la farmacologia, la fisiologia i les ciències ambientals. La revolució d'aquestes ciències ha permès obtenir dades dels canvis moleculars que tenen lloc en les cèl·lules, en els teixits i en l'organisme sencer quan es sotmeten a un estrès o estímul extern.

Entre les ciències òmiques s'hi inclouen: la genòmica per a l'estudi de les variacions en la seqüència de l'ADN, la transcriptòmica per a la caracterització de l'expressió gènica de tot el genoma mitjançant la mesura dels ARN missatgers (ARNm), la proteòmica per a la mesura dels nivells i de l'expressió de les proteïnes en les cèl·lules i els teixits i la metabolòmica per a l'avaluació global de les concentracions de metabòlits d'un sistema biològic. Actualment hi ha unes 200 disciplines que contenen la paraula òmica. Alguns dels exemples més destacats són la lipidòmica per a l'estudi a gran escala de vies i xarxes de lípids cel·lulars en sistemes biològics, la fluxòmica que descriu els diferents enfocaments que intenten determinar les taxes de les reaccions metabòliques dins d'una entitat biològica i la foodòmica que estudia els dominis de l'alimentació i la nutrició a través de l'aplicació i la integració de les tecnologies òmiques (Lindon et al. 2007).

La utilització d'aquestes ciències permeten obtenir un perfil des de cents fins a mils de gens, productes de gens i metabòlits que es generen simultàniament, i permeten generar, per primera vegada, una impressió global de com els organismes i les cèl·lules responen a un estrès (Snape et al. 2004). La informació que s'obté a través de les diferents ciències òmiques constitueixen la anomenada cascada òmica, tal com es mostra en la **Figura 1.1**. Aquest conjunt de plataformes proporcionen informació des

dels gens (genoma) fins als metabòlits (metaboloma) en resposta als possibles estímuls, com per exemple les malalties o les alteracions ambientals. La cascada òmica s'integra sobre els individus fins arribar als ecosistemes, alhora que inclou les poblacions i les comunitats d'organismes (Watanabe and Iguchi 2006).

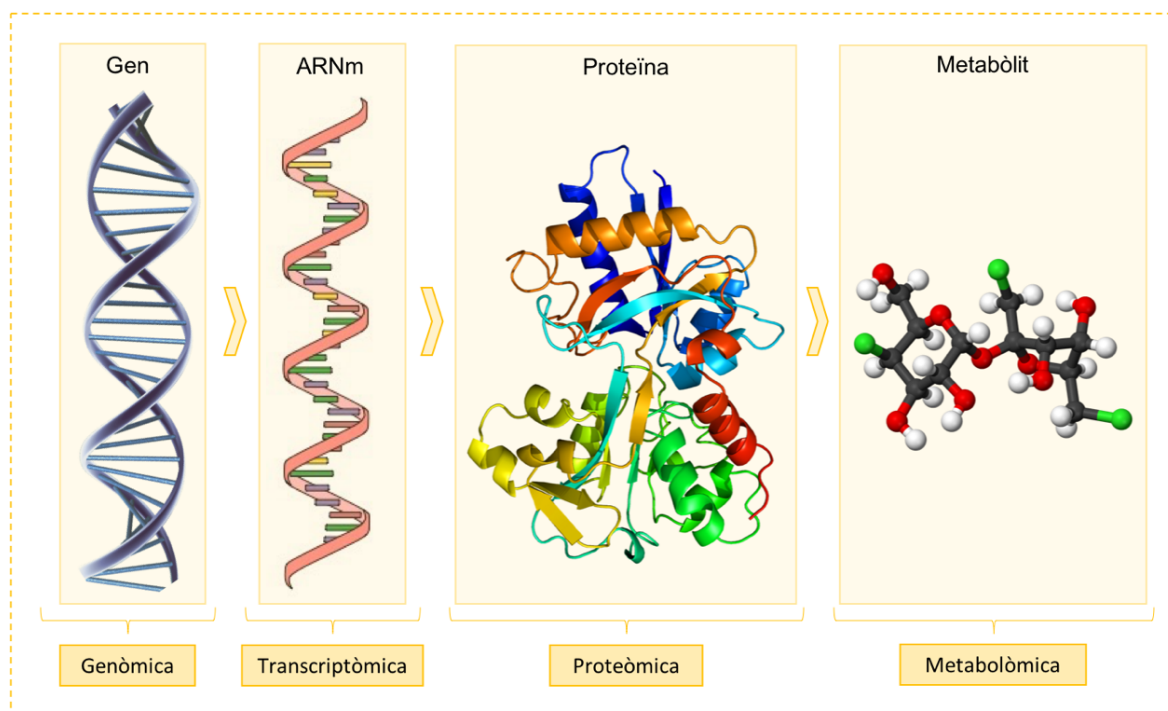


Figura 1.1| Principals ciències que constitueixen la cascada òmica: la genòmica, la transcriptòmica, la proteòmica i la metabolòmica.

En els últims 20 anys, la incorporació de la transcriptòmica, la proteòmica i la metabolòmica en estudis d'ecotoxicologia han permès obtenir informació valuosa sobre els efectes i els impactes de la contaminació en els sistemes biològics. Un dels avantatges de les ciències òmiques és que permeten una avaluació simultània de la resposta genètica, proteica o molecular d'un organisme i, això, en facilita la comprensió de l'estat fisiològic de l'organisme. També permeten una ràpida identificació dels efectes, mitjançant la modificació de l'expressió i l'activitat gènica, proteica o metabòlica. En concret, la òmica ambiental és l'aplicació de les ciències òmiques dirigida a la monitorització i l'avaluació del risc ambiental, per tal de caracteritzar de les interaccions entre els organismes i el medi ambient. Aquestes interaccions poden ser estudiades des dels individus fins a les poblacions, les quals es poden relacionar amb

els camps tradicionals de l'ecofisiologia, on s'estudia l'acció dels efectes biòtics i abiòtics en els organismes, i l'ecologia (Bundy et al. 2009).

L'ús de la òmica ambiental proporciona informació molecular molt detallada, la qual ajuda a identificar els efectes toxicològics alhora que defineix els mecanismes i modes d'acció dels contaminants en els sistemes biològics. Aquest ràpid desenvolupament dels estudis òmics ha esdevingut un avanç en el camp de l'ecotoxicologia, ja que proporcionen informació sobre respostes biològiques associades a malalties, exposicions tòxiques, hormones i estressants ambientals. D'aquesta manera, ha permès millorar-ne el coneixement i la comprensió dels potencials riscos dels contaminants ambientals en la salut dels ecosistemes (Simmons et al. 2015).

En comparació amb les tècniques tradicionals, com els mètodes fisiològics, citològics, histològics i bioquímics, aquestes noves tècniques, anomenades d'alt rendiment, proporcionen una informació molecular molt detallada la qual és més adequada per a la identificació i quantificació dels marcadors moleculars sensibles a una alteració exògena (Van Aggelen et al. 2010). Tot i això, les dades obtingudes mitjançant les ciències òmiques no s'han de considerar de manera independent, sinó que són complementaries de les tècniques tradicionals que estudien els processos biològics.

Les principals vies d'estudi de les ciències òmiques ambientals estan enfocades a: (i) respondre qüestions relacionades amb els modes d'acció de qualsevol estrès ambiental, ja sigui de caire físic o químic; (ii) descobrir com els contaminants químics estan associats a determinades respostes, tant en els individus com en les poblacions; (iii) monitoritzar els efectes produïts en els organismes que habiten en ecosistemes contaminats i (iv) identificar els biomarcadors que s'expressen en resposta a l'exposició d'un estressant. A la pràctica, però, aquests camps se sobreposen. D'aquesta manera, es divideixen els estudis entre els que es realitzen en el laboratori, principalment relacionats amb proves de toxicitat, i els que es basen en estudis de camp, principalment relacionats amb la monitorització ambiental.

Les ciències òmiques poden facilitar una investigació més exhaustiva que permeti millorar el coneixement sobre com els estressants ambientals causen toxicitat en organismes ecològicament rellevants i, a gran escala, efectes adversos a l'ecosistema (Villeneuve et al. 2011). En aquesta direcció, les dades òmiques s'han utilitzat en estudis de contaminació per tal de determinar els biomarcadors de la exposició mitjançant l'expressió gènica (Roling et al. 2007). Falciani et al. (2008) va utilitzar la transcriptòmica i la genòmica per determinar el lloc d'origen d'un peix (*Platichthys flesus*) basant-se en els gens sensibles a l'estrès ambiental. És a dir, que mitjançant la expressió gènica és possible predir un subconjunt de gens amb la expressió alterada com a conseqüència de l'efecte dels estressants ambientals. La influència de l'entorn local sobre el transcriptoma o el metaboloma d'un organisme es pot utilitzar en la monitorització ambiental i, així, caracteritzar els efectes dels estressants antropogènics, com per exemple la contaminació.

Per altra banda, la utilització de les tècniques òmiques en els assajos biològics de proteïnes no es va introduir fins al 2000, quan Shepard et al. (2000) va publicar el primer article científic relacionat amb l'aplicació de la proteòmica en la ecotoxicologia. Igual que en la transcriptòmica, els avanços instrumentals van generar un ràpid creixement en el camp ambiental d'aquesta ciència. Per exemple, Gündel et al. (2012) va descriure canvis en el proteoma i va demostrar la sensibilitat del mètode quan va exposar embrions de peixos zebra a sis concentracions diferents de fenantrè (hidrocarbur aromàtic policíclic), a una concentració letal baixa (LC₂₀). Entre aquestes tècniques, existeix un debat obert sobre si la proteòmica pot ser més informativa que la transcriptòmica, ja que les proteïnes són comparativament menys canviabls que els gens, tenint en compte que representen les molècules funcionals de la cèl·lula. Tot i això, es poden trobar estudis que demostren que la utilització d'ambdós mètodes permeten obtenir informació complementaria (Kanerva et al. 2014). Així, es posa de manifest que un únic enfocament òmic no és suficient per poder caracteritzar la complexitat dels ecosistemes.

1.1.1 METABOLÒMICA

El genoma, la transcriptòmica i el proteoma es poden veure com a mitjà en el flux de l'expressió gènica. Però és el metaboloma el que representa el nivell òmic final en un sistema biològic (**Figura 1.1**) A causa de l'estreta associació entre el metaboloma i l'estat fisiològic de l'organisme, és en aquest punt on es reflecteixen els canvis del fenotip i la funció bioquímica (Ryan and Robards 2006). Aplicada a la recerca en química ambiental, la metabolòmica ambiental té un gran potencial per tal de facilitar una millor comprensió de les rutes metabòliques implicades en els sistemes biològics. D'aquesta manera, permet predir els factors de risc potencials per a la vida salvatge i, així, prendre mesures per tal de minimitzar els efectes de les exposicions a compostos tòxics i estressants ambientals.

En els seus inicis, la utilització dels termes metabonòmica i metabolòmica va portar una mica de controvèrsia. El terme metabonòmica va ser formalitzat per Nicholson et al. (1999) com *la mesura quantitativa de respostes metabòliques dinàmiques multiparamètriques dels sistemes vius a estímuls fisiopatològics o modificacions genètiques*. Per altra banda, Fiehn (2002) va introduir el terme metabolòmic, on va definir aquest terme com a *l'anàlisi global i quantitatiu dels metabòlits en un sistema*. Dit d'una altra manera, mentre que el terme metabonòmica se centra en els canvis en els nivells de determinats metabòlits relacionant-los amb la influència dels estímuls externs, la metabolòmica, en canvi, és la identificació dels metabòlits en un sistema biològic. A posteriori, i tenint en compte els dos conceptes anteriorment descrits, Wishart (2009) va definir la metabolòmica com a *l'estudi sistemàtic d'un gran nombre de metabòlits de baix pes molecular els quals són generats per processos cel·lulars en unes condicions específiques*.

En l'actualitat, la comunitat científica utilitza ambdós termes indistintament, ja que, de fet, les dues definicions se superposen. Per això, es podria definir la metabolòmica com la recopilació de manera qualitativa i quantitativa de les molècules de baix pes molecular (metabòlits) presents en les cèl·lules, els quals participen en les reaccions metabòliques i es requereixen pel manteniment, el creixement i el funcionament normal de la cèl·lula. En general, tot i que també s'inclouen espècies

inorgàniques, s'hi integren espècies orgàniques, com per exemple els aminoàcids, els àcids grassos, els sucres, els nucleòtids, els nucleòsids i els lípids.

La primera publicació que fa referència a la metabolòmica data del 2000, on Fiehn et al. (2000) va descriure la utilització del perfil metabòlic com a una nova eina per a la comparació de les funcions del gens. En aquest estudi, es van analitzar diferents extractes de fulla d'*Arabidopsis thaliana* a través de la cromatografia de gasos acoblada a espectrometria de masses (GC-MS). Un dels avantatges de la metabolòmica, respecte de les altres ciències òmiques, és el fet que generalment els metabòlits són idèntics d'espècie a espècie. A més, els costos de les anàlisis són menors que en altres tècniques òmiques. I, al contrari dels gens i les proteïnes, els metabòlits estan ben definits i mantenen una relació estreta amb l'estat fisiològic de l'organisme.

Una vegada els metabòlits d'interès s'han identificat, els canvis poden ser fàcilment avaluats entre espècies. La metabolòmica analitza les variacions en les concentracions dels metabòlits, els quals són els precursors i els productes de l'activitat enzimàtica, i pretén relacionar-les amb una funció i/o regulació biològica. Per altra banda, la gran diversitat d'estructures moleculars entre metabòlits genera una dificultat addicional a l'hora de determinar estructures metabòliques, en comparació a la poca diversitat entre àcids nucleics i proteïnes, els quals es caracteritzen mitjançant la transcriptòmica i la proteòmica.

Estudis de metabolòmica ambiental utilitzant organismes aquàtics

Una gran part dels estudis de metabolòmica ambiental es centren en les espècies model per tal d'aconseguir un major potencial d'identificació dels efectes. Al 2008, Viant (2008) va publicar una revisió on posava de manifest la utilització d'organismes model com a organismes representatius dels ecosistemes per tal d'expressar les condicions del medi ambient. Entre aquests organismes es troben microorganismes (*Saccharomyces cerevisiae*), invertebrats aquàtics i terrestres (*Daphnia magna* i *Caenorhabditis elegans*), plantes (*Arabidopsis thaliana*), amfibis (*Xenopus tropicalis*) i peixos (*Danio rerio*).

Els estudis on s'usen organismes aquàtics es consideren d'especial rellevància, ja que molts d'aquests sistemes biològics estan exposats a una gran quantitat de contaminants distribuïts en els entorns aquàtics (Størseth and Hammer 2013). Principalment, aquests estudis s'han centrat en organismes d'aigua dolça, com per exemple els peixos. Aquests organismes han estat àmpliament utilitzats en estudis de toxicologia aquàtica per avaluar la toxicitat d'un contaminant i així comprendre millor els seus mecanismes d'acció (Heath 1995). L'aplicació de la metabolòmica en peixos ha estat descrita en varis estudis, els quals demostren els efectes produïts pels contaminants ambientals en aquests organismes aquàtics, com per exemple els pesticides, els productes farmacèutics i els disruptors endocrins. En uns d'aquests primers estudis, Viant et al. (2006) va descriure els canvis en el metaboloma d'embrions del peix *Oryzias latipes* els quals van ser exposats al plaguicida *dinoseb* (6-sec-butil-2,4-dinitrofenol). Com a resposta d'aquesta exposició, va poder observar canvis significatius en llur metabolisme, els quals estaven relacionats amb una reducció del creixement i alteració dels ritmes cardíacs.

Tradicionalment, els musclos també han estat emprats com a models de toxicitat i d'indicadors de la salut dels ecosistemes aquàtics per tal d'avaluar-ne l'impacte dels contaminants (Porte et al. 2006) Aquests organismes creixen adherits i no se separen del lloc on es desenvolupen i, com a conseqüència, acumulen els contaminants presents en aquest indret. Els musclos són una bona eina per a portar a terme estudis de metabolòmica en condicions de camp. D'aquesta manera, es poden trasplantar a diversos llocs per tal d'avaluar les condicions ambientals o exposar-los a uns tòxics concrets. Entre altres exemples, es van caracteritzar les respostes metabòliques del musclo *Mytilus galloprovincialis* en ésser exposat a dues concentracions del disruptor endocrí bisfenol A (BPA). En aquest estudi, es va determinar que ambdues exposicions induïen l'estrès osmòtic i la disrupció en el metabolisme energètic de l'organisme (Ji et al. 2014). Per altra banda, Bonnefille et al. (2018) també va investigar l'impacte produït pel compost farmacèutic diclofenac en el metabolisme d'aquests organismes on va observar una disminució de la tirosina i un augment del triptòfan a causa de l'exposició.

Diversos autors han focalitzat els seus estudis en l'exploració dels canvis en el metaboloma per tal d'avaluar els efectes nocius d'una gran varietat de contaminants i factors ambientals pel medi ambient i els seus ecosistemes. Conseqüentment, han estudiat la toxicitat dels contaminants en diversos organismes aquàtics, fet que els ha permès entendre llurs modes d'acció. D'aquesta manera, la metabolòmica ambiental ha estat capaç de determinar els efectes tòxics d'un gran nombre de contaminants presents en el medi aquàtic, com per exemple dels metalls pesats (Nagato et al. 2013; Zhang et al. 2015), dels pesticides (Gómez-Canela et al. 2017; Zhang and Zhao 2017), dels compostos farmacèutics (Bonnefille, et al. 2018; Gómez-Canela et al. 2016), dels disruptors endocrins (Ji, et al. 2014; Nagato et al. 2016; Ortiz-Villanueva et al. 2017b) i d'altres xenobiòtics (Benskin et al. 2014).

1.2 ORGANISME MODEL: *DAPHNIA MAGNA*

En aquesta Tesi, la *Daphnia magna* és l'organisme escollit per avaluar els canvis en el seu metaboloma quan es sotmet a diferents estressants ambientals. Segons la classificació taxonòmica del Registre Mundial de les Espècies Marines (*World Register of Marine Species*) la *D. magna* és un crustaci planctònic de l'ordre dels Cladòcers *Daphniidae* (Boxshall 2013). En recerca ambiental, és un organisme àmpliament utilitzat donada la seva reduïda variabilitat genètica i llur curt temps de reproducció necessari per a generar clons idèntics (Ebert 1991). Segons la Organització per la Cooperació Econòmica i Desenvolupament (OECD) (OECD 2012) es considera un organisme model per realitzar estudis de toxicitat.

La *D. magna* és un organisme petit, aproximadament d'una mida inferior als 3 mm, el qual és present en llacs, estanys i rierols d'aigua dolça (De Gelas and De Meester 2005). Es tracta d'un organisme fonamental en l'ecosistema pel control de les floracions d'algues, alhora que també és important com a base de l'alimentació, tant pels invertebrats com pels vertebrats (Ebert 2005). Llur principal font d'alimentació són les algues, així com les partícules en suspensió presents en les aigües on viuen. La presència de la *D. magna* és un bon indicador de la qualitat ecològica dels sistemes aquàtics (Hebert 1978).

El cicle de vida d'aquest organisme aquàtic es caracteritza per una reproducció asexual, així es reproduïx cíclicament a través de partenogènesis, amb la generació d'un nou individu a partir d'un òvul sense fecundar (Zaffagnini 1987). La població està compresa exclusivament per femelles i la producció masculina està sota control mediambiental. Les femelles produeixen ous que contenen informació genètica idèntica a la de la mare, i les quals es desenvoluparan en clons idèntics al seu progenitor (veure **Figura 1.2**). A causa d'aquesta reproducció clonal, la *D. magna* es considera una excel·lent eina pels estudis genètics, el quals ajuden a millorar la comprensió de la seva ecologia evolutiva. En condicions de vida difícils, com per exemple durant l'escassetat d'aliments o les onades de calor, la reproducció de les *daphnies* es modifica, i la reproducció es converteix en sexual. Cal esmentar que en el laboratori, les condicions de vida de la *D. magna* estan controlades, i que per tant, només la reproducció clonal està afavorida.

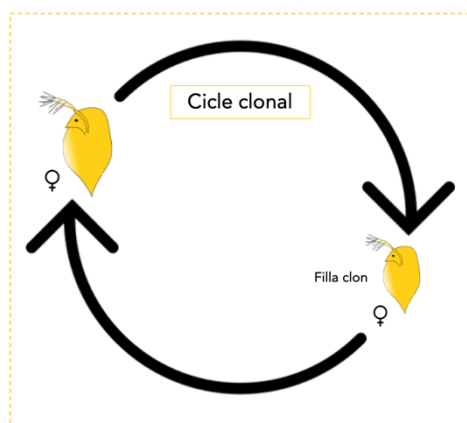


Figura 1.2] Esquema de la reproducció clonal de femella de la *D. magna* adulta. (Imatge adaptada de Arndt (2013)).

1.2.1 LA *DAPHNIA MAGNA* COM A MODEL DE TOXICITAT

Als anys 70 es va publicar el primer article on es demostrava la gran utilitat de la *D. magna* per a estudis de toxicologia aquàtica (Adema 1978). Actualment, aquests organismes es coneixen com a organismes *sentinelles* molt sensibles a una gran varietat d'estressants ambientals. Entre aquests estímuls, tant abiòtics com biòtics, s'inclouen substàncies químiques, hormones sintètiques, acidesa, salinitat, hipòxia, radiació, depredadors, paràsits, etc. (Colbourne et al. 1997). Entre altres exemples, la *D. magna*

s'ha utilitzat per avaluar els efectes tòxics dels metalls pesats (Smolders et al. 2005) o els efectes de diversos fàrmacs (Flaherty and Dodson 2005), on s'han estudiat les respostes de toxicitat i de reproducció a nivell d'organisme i de poblacions amb exposicions als diferents contaminants.

D'aquesta manera, la *D. magna* és un dels organismes més utilitzats en proves de toxicitat, tant aguda com crònica, amb les quals s'han obtingut resultats que han permès avaluar els riscos ambientals. Existeixen diversos mètodes on s'utilitza la *D. magna* per tal de realitzar proves de toxicitat reglamentaries i, més concretament, la OECD recomana aquesta espècie en les seves directrius sobre assaigs químics (OECD 2012). La directiva 79/831/CEE regula les proves de toxicitat aguda en peixos o en *D. magna*, per tal de controlar els efectes ecotoxicològics dels productes químics industrials.

1.2.2 ESTUDIS DE METABOLÒMICA UTILITZANT LA *DAPHNIA MAGNA*

En la última dècada, a part del seu ús en els models de toxicologia, la *D. magna* també ha estat utilitzada per realitzar estudis ecotoxicogenòmics (Kim et al. 2015). En un dels primers exemples publicats, Watanabe et al. (2007) va trobar uns patrons característics de l'expressió gènica a través de la utilització de microarrays d'ADN en exposar la *D. magna* a diferents productes químics tòxics, com el sulfat de coure, el peròxid d'hidrogen, el pentaclorofenol i la β -naftoflavona. Tot i el gran avantatge que representa l'aplicació de les tècniques òmiques en l'avaluació de riscos ecològics, els estudis on es descriuen els perfils metabolòmics de la *D. magna* segueix sent escàs.

Amb l'objectiu de conèixer els estudis de metabolòmica que s'han portat a terme fins al moment, s'ha realitzat una recerca bibliogràfica dels treballs on s'examina la resposta de la *D. magna* a diversos estressants ambientals. En la **Taula 1.1** es mostra el recopilatori d'aquests estudis de metabolòmica ordenats de manera cronològica i segons les tècniques analítiques emprades, on s'indiquen els estressants utilitzats per a l'avaluació de la resposta en la *D. magna* i es descriuen les conclusions principals que s'obtenen amb els resultats publicats en els treballs.

Taula 1.1| Estudis de metabolòmica que examinen la resposta de la *D. magna* a estressants ambientals, ordenats segons les tècniques analítiques emprades.

Autors	Tècniques analítiques	Estressants	Conclusions principals
Taylor et al. (2009)	FT-ICR	Coure.	<ul style="list-style-type: none"> ✓ Reducció dels nivells d'aminoàcids així com marcadors de l'estrès oxidatiu com el glutatió i l'àcid oftàlmic.
Taylor et al. (2010)	FT-ICR	Cadmi, fenvalerat, 2,4-dinitrofenol i propanolol.	<ul style="list-style-type: none"> ✓ Els extractes de l'organisme sencer són més predictius de la toxicitat que la hemolimfa. ✓ Disminució del metabolisme dels àcids grassos i de la síntesi dels eicosanoids.
Taylor et al. (2018)	DIMS	Cadmi, 2,4-dinitrofenol i propanolol i les seves combinacions binàries i terciàries.	<ul style="list-style-type: none"> ✓ Diversos biomarcadors metabòlics poden predir l'aptitud reproductiva de la <i>D. magna</i>. ✓ Els tres contaminants disminueixen significativament la reproducció de la <i>D. magna</i>.
Poynton et al. (2011)	FT-ICR i RMN	Cadmi.	<ul style="list-style-type: none"> ✓ Disrupció dels nivells d'aminoàcids i d'àcids grassos en la hemolimfa.
Vandenbrouck et al. (2010)	RMN i GC-MS	Pirè i fluorantè.	<ul style="list-style-type: none"> ✓ Pertorbacions en el metabolisme dels aminosucre. ✓ Les coexposicions del pirè i el fluorantè produeixen efectes superiors a les exposicions individuals.

Autors	Tècniques analítiques	Estressants	Conclusions principals
Nagato, et al. (2013)	RMN	Arsènic, coure i liti.	<ul style="list-style-type: none"> ✓ L'exposició al coure i al liti contribueixen a la variació dels nivells d'aminoàcids. ✓ Disrupció del metabolisme energètic durant l'exposició a arsènic.
Wagner et al. (2015)	RMN	Variacions en la qualitat i la quantitat de la dieta.	<ul style="list-style-type: none"> ✓ Augment dels aminoàcids en condicions d'alimentació baixes. ✓ Disminució dels aminoàcids amb alts nivells de salinitat.
Scanlan et al. (2015)	RMN	Retardants de flama (pentaBDE, octaBDE, FM550, BZ54, BEH-TEBP, TPhP, BEHP).	<ul style="list-style-type: none"> ✓ El retardant de flama FM550 provoca una resposta tòxica que afecta als aminoàcids. ✓ Augment dels àcids grassos no saturats (HUFA), que estan relacionats amb el creixement de la <i>D. magna</i>.
Li et al. (2015)	RMN	Nanopartícules de plata (AgNPs).	<ul style="list-style-type: none"> ✓ Canvis en l'arginina, la lisina, la glucosa i el 3-hidroxitbutirat que afecten el metabolisme energètic i anaeròbic i produeixen estrès oxidatiu.
Kovacevic et al. (2016)	RMN	Triclosan, ibuprofè i carbamazepina.	<ul style="list-style-type: none"> ✓ Disminueix el nivell d'aminoàcids i augment de la glucosa amb la carbamazepina. ✓ Augment dels aminoàcids i reducció de la glucosa amb el triclosan. ✓ Canvis en els nivells d'aminoàcids i de glucosa a causa de l'ibuprofè.
Nagato, et al. (2016)	RMN	Pesticides organofosfats (diazinon i malatió) i BPA.	<ul style="list-style-type: none"> ✓ Aturada de l'activitat metabòlica, augment dels aminoàcids i disminució de la glucosa i el lactat a causa de les exposicions elevades als organofosfats i al BPA.

Autors	Tècniques analítiques	Estressants	Conclusions principals
Wagner et al. (2017)	RMN	Atrazina, propanolol i PFOS.	<ul style="list-style-type: none"> ✓ Els perfils metabòlics entre nounats i adults difereixen amb la mateixa concentració a l'atrazina i el PFOS. En canvi, les respostes són similars per al propanolol. ✓ La variació en els perfils pot ser deguda a les diferències fisiològiques dels diferents estadis de vida de la <i>D. magna</i>.
Kariuki et al. (2017)	RMN	PFOS.	<ul style="list-style-type: none"> ✓ L'exposició a PFOS mostra una variació en les rutes del metabolisme energètic, així com en la ruta de degradació de proteïnes.
Jeong et al. (2018)	RMN	Propanolol.	<ul style="list-style-type: none"> ✓ Les exposicions van mostrar una disminució del moviment, que significa un impacte en la regulació del metabolisme. ✓ Impacte en el metabolisme energètic amb disminució de glucosa i aminoàcids.
Falanga et al. (2018)	RMN	QD funcionalitzats amb el pèptid inolicidina.	<ul style="list-style-type: none"> ✓ El perfil metabòlic es va veure alterat durant l'exposició subletal dels QDs. ✓ Canvis en els metabòlits relacionats amb l'emmagatzematge energètic (disminució dels triglicèrids), després d'una exposició de 21 dies.
Wagner et al. (2018)	RMN	Propanolol, carbamazepina, PFOS i les seves combinacions binàries i terciària.	<ul style="list-style-type: none"> ✓ L'exposició al propanolol va produir un augment de l'alanina. ✓ La combinació de carbamazepina i PFOS va donar lloc a un augment dels aminoàcids i un descens dels nivells de glucosa. ✓ L'exposició de la combinació terciària va produir un augment dels aminoàcids, amb un augment significatiu de la valina i la isoleucina, i una disminució de la glucosa.

Autors	Tècniques analítiques	Estressants	Conclusions principals
Kovacevic et al. (2018a)	RMN	EE2, carbamazepina, imidacloprid i PFOS	<ul style="list-style-type: none"> ✓ L'augment dels aminoàcids produït per l'exposició a EE2, es va veure reduïda amb la presència de la matèria orgànica dissolta. ✓ L'exposició a carbamazepina i a imadacloprid en presència de la matèria orgànica dissolta no afecta al metabolisme. ✓ L'exposició a PFOS va donar lloc a una resposta metabòlica més significativa amb la presència de la matèria orgànica dissolta.
Kovacevic et al. (2018b)	RMN	Esters d'organofosfats: TCEP, TBOEP i TPhP.	<ul style="list-style-type: none"> ✓ La matèria orgànica dissolta genera canvis lleugers en l'alanina i la serina. ✓ L'exposició a TCEP no afecta el metabolisme. ✓ L'exposició a TBOEP altera el metabolisme, amb un increment dels aminoàcids i una disminució de glucosa. ✓ L'exposició a TPhP provoca canvis significatius en vies cel·lulars essencials implicades en la generació d'energia.

En el primer treball que es descriu en la **Taula 1.1**, Taylor, et al. (2009) van utilitzar la tècnica d'infusió directa per ressonància ió ciclotró de transformada de Fourier acoblada a l'espectrometria de masses (*Fourier-transform ion cyclotron resonance mass spectrometry*, FT-ICR-MS) per tal de determinar el gradient de toxicitat del coure a través de l'efecte produït en el metabolisme de la *D. magna*. En aquest treball es va observar una disminució dels aminoàcids i es va apreciar una disminució dels nivells de glutatió i de l'àcid oftàlmic com a resposta a l'estrès oxidatiu produït per l'exposició al coure. En un estudi posterior, Taylor, et al. (2010) va utilitzar la mateixa plataforma analítica per estudiar la toxicitat del cadmi, el 2,4-dinitrofenol, el propanolol i el fenvalerat, on es va determinar que el propanolol indueix un increment dels àcids grassos i dels eicosanoids, mentre que el fenvalerat provoca un augment del nombre de metabòlits implicats en la producció d'aminosucres. A més, també es van avaluar les diferents respostes segons si s'analitzava l'organisme sencer o únicament la hemolimfa de la *D. magna*. Així, es va observar que l'anàlisi de l'organisme sencer és més predictiu de la toxicitat que analitzar únicament l'hemolimfa de l'organisme. Més recentment, el mateix autor va estudiar el perfil metabòlic de la *D. magna* en exposar-se a tres dels contaminants estudiats en l'estudi descrit anteriorment (cadmi, 2,4-dinitrofenol i propanolol) i a les seves combinacions binàries i terciàries mitjançant DIMS. D'aquesta manera, va determinar l'efecte de la toxicitat d'aquests contaminants en la disminució significativa de la reproducció de la *D. magna* després d'una exposició de 21 dies (Taylor, et al. 2018).

Per altra banda, Poynton, et al. (2011) va utilitzar la *D. magna* per avaluar els canvis produïts en el metaboloma després de 24 hores d'exposició al cadmi, on es va observar que els aminoàcids i els àcids grassos van disminuir a l'hora que va determinar una reducció del transport de lípids. En aquest estudi, les mostres van ser analitzades mitjançant la FT-ICR-MS per la seva gran sensibilitat en detectar un nombre elevat de metabòlits en la hemolimfa de la *D. magna*. A més, es van emprar els espectres obtinguts mitjançant la ressonància magnètica nuclear (RMN) per a la identificació i confirmació d'aquests metabòlits.

La combinació de les tècniques de RMN i de GC-MS va ser emprada per Vandembrouck, et al. (2010) per tal de realitzar estudis de metabolòmica en *D. magna*. En aquest estudi es van investigar els efectes de la toxicitat del pirè i el fluorantè, de manera individual i en barreges binàries per ambdues plataformes analítiques, on es van detectar un major nombre de metabòlits amb la GC-MS que amb la tècnica de RMN, aproximadament un total de 86 i 25 metabòlits, respectivament. Els resultats d'aquest estudi van determinar una resposta similar dels dos contaminants, on es van observar canvis importants en el metabolisme dels aminosucres.

La utilització de la RMN com a plataforma analítica en estudis de metabolòmica permet una determinació ràpida, quantitativa i que requereix poca preparació de la mostra per a l'avaluació eficient del metaboloma dels organismes (Clarke and Haselden 2008). Així, Nagato, et al. (2013) va emprar aquesta tècnica per determinar l'efecte de l'arsènic, el coure i el liti en perfils metabòlics de la *D. magna*. Es van observar disruptions en la producció energètica de la *D. magna* a causa de l'exposició al liti i al coure, alhora que ambdós contaminants van afectar de la mateixa manera el metabolisme dels aminoàcids, on van provocar una disminució de l'alanina, el glutamat i la metionina, mentre que l'exposició a l'arsènic va provocar canvis en les concentracions de l'alanina i la lisina.

La RMN va permetre investigar els canvis produïts en la *D. magna* en comparar diferents tipus de dietes alimentàries amb altres estressants com la salinitat i la infecció bacteriana, també anomenat estrès nutricional (Wagner, et al. 2015). En aquest treball es va observar que la *D. magna* creix en condicions alimentàries amb un nivell elevat d'aminoàcids en resposta a la gran necessitat de crear energia. Això contrasta amb un augment de la salinitat, que indueix una resposta contrària, amb accentuades disminucions en diversos aminoàcids, ja que molts organismes compensen l'estrès osmòtic amb un augment en l'alliberació d'aminoàcids per tal de crear un equilibri osmòtic. Per altra banda, Scanlan, et al. (2015) va emprar la combinació de la transcriptòmica i la metabolòmica mitjançant RMN per estudiar les diferents respostes, tant genètiques com metabòliques, de la *D. magna* exposada als retardants de flama bromats. En concret, es va determinar que el pentaPBDE afectava a la transcripció i la

traducció del mRNA, mentre que el FM50 impedia l'absorció dels nutrients en la *D. magna*.

Els perfils metabòlics de la *D. magna* també s'han utilitzat per a la determinació de les nanopartícules alliberades en el medi aquàtic. A causa dels seus efectes nocius, la seva presència és una de les principals preocupacions mediambientals actuals. En aquest context, Li, et al. (2015) van examinar la toxicitat de les nanopartícules de plata (AgNPs), on es van observar canvis significatius en el metabolisme energètic amb un augment dels nivells de lactat que afecten al metabolisme anaeròbic de la *D. magna*. Així mateix, la presència de compostos farmacèutics en el medi ambient també és una preocupació emergent pel seu impacte al medi ambient. Kovacevic, et al. (2016) van estudiar l'efecte del triclosan, l'ibuprofè i la carbamazepina en el metaboloma de la *D. magna*, on es van determinar efectes contraris en els nivells de la glucosa i en el metabolisme dels aminoàcids en resposta al triclosan i a la carbamazepina, és a dir que afecten de manera oposada. Mentre que l'ibuprofè va provocar canvis en els nivells d'aminoàcids.

Els pesticides organofosfats, específicament el diazinon i el malatió, a concentracions d'exposició intermèdies, van mostrar una disminució dels nivells d'aminoàcids i un increment dels nivells de la glucosa i del lactat, suggerint així una mobilització de les reserves energètiques com a causa de l'estrès (Nagato, et al. 2016). En canvi, a concentracions altes dels organofosfats s'observa un increment dels nivells d'aminoàcids i una disminució dels nivells de glucosa i de lactat, que és el mateix patró de resposta que s'obté amb l'exposició al bisfenol A. Això suggereix una resposta generalitzada a l'estrès la qual dona lloc a una alteració dels nivells energètics de la *D. magna*.

La RMN també va permetre avaluar la influència de les respostes metabòliques dels diferents estadis de vida de la *D. magna* en exposar-se a diferents contaminants amb diferents modes d'acció (Wagner, et al. 2017). D'aquesta manera, els nounats i els adults mostren un perfil diferent de metabòlits quan són exposats a l'atrazina o a l'àcid perfluorooctanosulfònic (PFOS). En canvi, quan són exposades al propanolol les respostes metabòliques són semblants en els diferents estadis de vida de la *D. magna*.

Aquests resultats indiquen que el perfil metabòlic es veu alterat com a conseqüència de l'exposició del contaminant, i que aquests canvis no sempre són consistents al llarg dels diferents estadis de vida. A més, Kariuki, et al. (2017) va estudiar l'efecte de la exposició del PFOS, on va observar una disrupció del metabolisme energètic i un increment de la degradació de les proteïnes.

Més recentment, Jeong, et al. (2018) va caracteritzar el mode d'acció del propranolol al llarg del temps per tal de determinar el perfil metabòlic de la *D. magna*. Els resultats mostren una reducció dels nivells de glucosa i dels aminoàcids, a més d'un increment dels nivells de acetat després de l'exposició al propranolol, que poden ser efecte de la disminució en la generació d'energia per part de la *D. magna*. Per altra banda, es van analitzar els diferents perfils metabòlics de la *D. magna* a través de tres generacions per tal d'avaluar els efectes dels punts quàntics (*Quantum dots: QDs*)¹ funcionalitzats amb el pèptid antimicrobià indolicidin, específicament indolicidin, QDs i QDs amb indolicidin (Falanga, et al. 2018). També, Wagner, et al. (2018) va estudiar l'exposició individual i les combinacions binàries i terciàries del propranolol, la carbamazepina i el PFOS durant 48 hores. Es va observar un increment de l'alanina durant l'exposició a concentracions subletals a propranolol. Mentre que, la combinació de la carbamazepina i el PFOS va provocar un augment dels aminoàcids i una disminució de la glucosa, igual que amb la combinació terciària, la qual va donar lloc als mateixos canvis en els aminoàcids i la glucosa, alhora que també van augmentar significativament els nivells de valina i de isoleucina.

Més recentment, Kovacevic et al. va portar a terme dos estudis on va investigar com el metaboloma de la *D. magna* canvia quan la matèria orgànica dissolta (DOM) s'inclou a l'exposició aguda de varis contaminants orgànics amb diferents hidrofobicitats (Kovacevic, et al. 2018a; Kovacevic, et al. 2018b). Entre els contaminants més comuns detectats en el medi aquàtic es troben els esters d'organofosfats (OPEs), els quals modifiquen la biodisponibilitat de la matèria orgànica dissolta (Pang et al. 2013). En el primer d'aquests estudis, es van estudiar els efectes del 17 α -etinilgestradiol (EE2), la carbamazepina, l'imidacloprid i el PFOS, l'augment dels aminoàcids produït

¹ QD (*Quantum Dot*): *Partícula de material semiconductor*.

per l'exposició a l'EE2, es va veure reduïda amb la presència de la DOM (Kovacevic, et al. 2018a). El metaboloma de la *D. magna* es va veure inalterat durant l'exposició a la carbamazepina i a l'imadacloprid, en presència i absència de la DOM, mentre que l'exposició a PFOS amb presència de la DOM va donar lloc a una resposta metabòlica més significativa que en absència de la DOM. En conseqüència, es posa de manifest que la *D. magna* queda alterada quan és exposada a diversos tipus de compostos orgànics.

En un estudi posterior, on es va estudiar l'efecte de tres OPEs individuals: tris(2-cloroetil) fosfat (TCEP), tris(2-butoxietil) fosfat (TBOEP) i trifenil fosfat (TPhP). L'exposició a TCEP no va modificar el metabolisme de la *D. magna* significativament. En canvi, l'exposició a TBOEP va provocar un increment dels aminoàcids i una disminució de la glucosa de la *D. magna*. Aquesta disminució suggereix un esgotament de les reserves d'energia immediates en forma d'hidrats de carboni i l'augment dels aminoàcids estan relacionats en la degradació de les proteïnes per obtenir energia. A més, el metabolisme de la glicina, la serina i la treonina es va veure afectat en exposar la *D. magna* a TPhP. Aquestes vies metabòliques són importants per a la gluconeogènesi, una via cel·lular essencial implicada en la generació d'energia. En l'estudi també es té en compte si la presència de la DOM afecta la toxicitat d'aquests contaminants orgànics. Només en el cas de l'exposició a TPhP la resposta metabòlica és diferent amb presència o no de la DOM (Kovacevic, et al. 2018b).

Tots aquests estudis proporcionen les bases necessàries per a l'acceptació de la metabòlica en estudis ambientals, alhora que també proporcionen informació sobre la capacitat de resposta de la *D. magna* a diferents estressants ambientals. En el **capítol 2** d'aquesta Tesi s'han estudiat els efectes dels estressants abiòtics (salinitat, temperatura i hipòxia) en el metabolisme de la *D. magna*. Mentre que en el **capítol 3** s'han avaluat els efectes de contaminants ambientals com el bisfenol A (BPA), així com les respostes als productes farmacèutics com el diazepam, la fluoxetina i la carbamazepina.

1.3 ESTRESSANTS AMBIENTALS

Durant el segle XX i principis del segle XXI, el clima i els ecosistemes de la Terra van patir canvis importants com a resultat directe de l'activitat humana i del creixement de la seva població. En els darrers 150 anys, la concentració atmosfèrica de gas CO₂ ha augmentat considerablement a causa de l'activitat humana i s'ha posat en evidència que l'emissió de gasos d'efecte hivernacle provoca un escalfament de l'atmosfera i produeix canvis climàtics. A més, cada any, milers de nous productes químics sintetitzats s'alliberen i s'acumulen al medi, els quals provoquen efectes desconeguts sobre els organismes vius i els sistemes ecològics. En el camp de la metabolòmica ambiental l'objectiu principal és estudiar com afecten els diferents estressants en els éssers vius que habiten en els ecosistemes. En general, es poden classificar en factors abiòtics i contaminants ambientals.

1.3.1 ESTRÈS ABIÒTIC

Els estressants abiòtics són els factors tant físics com químics que determinen les característiques d'un biòtop². Estan relacionats amb les alteracions produïdes per fenòmens naturals, com són la sequera, les inundacions, les temperatures elevades, l'increment de la salinitat en les zones d'aigua dolça o la limitació dels nutrients. L'impacte ecològic dels estressants relacionats amb el canvi climàtic són un fet reconegut. Entre els efectes més generalitzats en les poblacions que habiten en el medi aquàtic es troben l'augment de les temperatures de l'aigua, el canvi en el contingut de la salinitat i la reducció dels nivells d'oxigen dissolt (Schindler 2001). Tanmateix, aquests estressants abiòtics també afecten la taxa d'absorció dels contaminants i la sensibilitat dels organismes, fet que provoca canvis en les taxes del metabolisme i en l'alimentació. Per exemple, l'augment de la toxicitat dels contaminants com per exemple els organofosfats i els metalls pesants presents en les aigües està relacionada amb temperatures elevades (Ficke et al. 2007).

² Biòtop: part abiòtica (no viva) de l'ecosistema. En són exemples un llac, un riu, una platja, el cim d'una muntanya o una vall.

Els factors abiòtics estudiats en aquesta Tesi són els següents: la salinitat, la temperatura i l'oxigen dissolt en l'aigua, els quals s'han seleccionat com a representats dels efectes del canvi climàtic. Tal com es descriu al llarg del **capítol 2**, en primer lloc es va portar a terme un estudi dels efectes individuals en el metabolisme de la *D. magna* i, en un segon lloc, es van avaluar els efectes de les diferents combinacions dels factors, binàries i terciària.

La salinitat

Els nivells de salinitat dels ecosistemes d'aigua dolça poden patir importants alteracions com a conseqüència del canvi climàtic global, com ara l'acció de l'aigua de la pluja i el vent, l'evaporació i l'augment del nivell del mar (Schallenberg et al. 2003). Es tracta d'un factor abiòtic determinant en els ecosistemes aquàtics. L'aigua dolça conté 0.5 g L⁻¹ de substàncies dissoltes i l'aigua marina en conté aproximadament 35 g L⁻¹ de mitjana. L'efecte de l'estrès en els organismes d'aigua dolça genera canvis fisiològics per tal d'adaptar-se per sobreviure i poder reproduir-se en aquestes condicions desfavorables.

La temperatura

Una gran part dels organismes d'aigua dolça són ectotèrms, és a dir que no poden regular la seva temperatura corporal (Moyle and Cech 2004). D'aquesta manera, la seva temperatura depèn del medi que els envolta i les fluctuacions de la temperatura de l'aigua alteren les seves funcions fisiològiques com el creixement, el metabolisme, l'alimentació i la reproducció (Paaijmans et al. 2013). Tanmateix, el rang de tolerància a l'augment de la temperatura és específic per a cada espècie.

Oxigen dissolt en l'aigua

Les concentracions d'oxigen dissolt acceptables per a la majoria d'organismes aquàtics es troben al voltant de 5 mg L⁻¹ (Stickney 2000). Les concentracions inferiors a 2-3 mg L⁻¹ es consideren nivells d'hipòxia (Ficke, et al. 2007). Els canvis en els nivells d'oxigen dissolt en l'aigua dependran, entre altres factors, de la temperatura de l'ambient i de la demanda biològica d'oxigen. Com a efecte del canvi climàtic, els

organismes aquàtics es veuran exposats a temperatures elevades que produeix una disminució del subministrament d'oxigen que no podrà assolir-ne la demanda. Per aquest motiu, és possible que els nivells d'oxigen dissolt disminueixin en el medi ambient donada la elevada probabilitat de l'augment de les temperatures (Ficke, et al. 2007).

1.3.2 CONTAMINANTS AMBIENTALS

La presència de contaminants en el medi ambient és un tema de gran rellevància social. Diversos tòxics com els disruptors endocrins, els metalls pesants, els fàrmacs i els compostos orgànics estan àmpliament distribuïts en el medi ambient i poden afectar als éssers vius. Per exemple, es coneix que els disruptors endocrins afecten la reproducció dels organismes, ja que provoquen una reducció del nombre de nounats i la producció de la *Daphnia* masculina (Comber et al. 1993; McLachlan 2001; Tatarazako et al. 2003). Tanmateix, els metalls pesats alteren el desenvolupament de la taxa de creixement de la població i afecten la reproducció (Altshuler et al. 2011). A més, molts compostos orgànics i els productes farmacèutics tenen un efecte perjudicial en la reproducció, en la muda i en el desenvolupament de la *D. magna* (Heckmann et al. 2008a).

Bisfenol A

El bisfenol A és un compost orgànic amb dos grups funcionals fenol que es coneix habitualment com a BPA. El BPA té una gran quantitat d'usos industrials, tot i que el 95 % s'utilitza com a monòmer del policarbonat i les resines epoxi, convertint-lo en un dels productes químics més produïts del món (Vandenberg et al. 2009). La seva ubiqüitat ha provocat una contaminació generalitzada en els sistemes aquàtics (Salgueiro-González et al. 2016). En els rius, s'ha demostrat que el BPA es degrada completament després de 8 dies, tot i que els seus productes intermedis de degradació poden perdurar passades dues setmanes (Suzuki et al. 2004).

Per als humans, la principal ruta d'exposició a aquest contaminant és la dieta, específicament aliments envasats en plàstic que produeixen la migració del BPA als

aliments a concentracions de partícules per bilió (ppb) (Biles et al. 1997). El BPA té un mode d'acció (MOA) similar al de l'estradiol, el qual es coneix que s'uneix als receptors d'estrògens i interfereix en el sistema endocrí humà (Jordão et al. 2016). Encara que la toxicitat del BPA ha estat molt estudiada en els mamífers, es desconeix l'abast de la seva toxicitat en els invertebrats aquàtics. Els estudis realitzats amb la *D. magna* es centren principalment en el seu paper com a disruptor endocrí, en termes de mortalitat i reproducció (Brennan et al. 2006; Mansilha et al. 2013), però pocs estudis examinen l'estrès en el metabolisme. Per aquest motiu, es va creure molt interessant seleccionar aquest contaminant per tal d'estudiar els efectes en el metaboloma de la *D. magna*, tal com es detalla en el **capítol 3** d'aquesta Tesi.

Productes farmacèutics

Els productes farmacèutics són un ventall ampli i divers de compostos orgànics utilitzats en la prevenció i el tractament de les malalties dels éssers humans i dels animals. En l'últim segle, com a conseqüència dels avenços mèdics, s'han desenvolupat un nombre molt elevat de nous medicaments i nous tractaments, els quals han provocat un increment exponencial del seu consum i, alhora, el seu conseqüent alliberament al medi ambient. A causa de la seva extensa aplicació, juntament amb la seva persistència i acumulació, els productes farmacèutics tenen una elevada presència en els ecosistemes, amb concentracions al voltant de ng L^{-1} o $\mu\text{g L}^{-1}$ (Calisto and Esteves 2009). Tot i que les concentracions en què estan presents en el medi són inferiors a les previstes per tal de produir danys en els ésser humans, la seva activitat biològica específica en el sistema neuro-endocrí els converteix en un contaminant ambiental important per investigar i avaluar els seus efectes en els organismes aquàtics (Fong and Ford 2014; Ven et al. 2006).

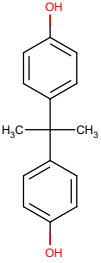
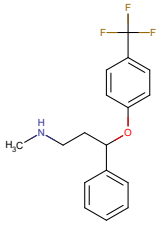
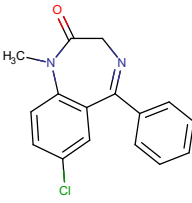
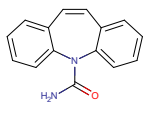
Així mateix, entre els productes farmacèutics més utilitzats en el segle XXI, cal esmentar els antidepressius, els fàrmacs ansiolítics i els neuropàtics, que actuen com a inhibidors selectius del receptor de la serotonina (*selective serotonin reuptak inhibitors*, SSRIs). És a dir, que actuen bloquejant la serotonina en les sinapsis nervioses de la *D. magna* (Gunnarsson et al. 2008). En concret, entre els compostos estudiats s'inclouen la fluoxetina que és un antidepressiu SSRI, el diazepam que se subministra com a

ansiolític i la carbamazepina que és un antiepilèptic. La fluoxetina és un ingredient actiu del Prozac, un medicament àmpliament utilitzat per tractar la depressió clínica. El diazepam, en un principi actiu comercialitzat com a Valium, s'utilitza com a efecte calmant per tractar l'ansietat, intervé en el son, afecta la relaxació muscular, l'activitat anticonvulsiva, la coordinació motora i la memòria. La carbamazepina és un medicament utilitzat en el tractament de l'epilèpsia i el trastorn bipolar, que actua disminuint la transmissió sinàptica. Usualment, els tres compostos són detectats en les aigües residuals i superficials, ja que no s'eliminen en les plantes de tractament (Fent et al. 2006).

En estudis previs es va demostrar que aquests fàrmacs neuroactius augmenten la reproducció de la *D. magna* a concentracions ambientals ($<0.1 \mu\text{g L}^{-1}$) (Rivetti et al. 2016). També, s'han realitzat treballs més exhaustius amb la fluoxetina, on també es va observar un augment de la reproducció i una disminució dels nivells de carbohidrats (Campos et al. 2013; Campos et al. 2012; Jordão, et al. 2016). En aquest context, es va voler explorar un enfocament diferent dels efectes d'aquests fàrmacs en la *D. magna* amb l'avaluació dels nivells d'oxilipines, sobretot les que pertanyen a la ruta de l'àcid araquidònic que estan relacionades amb la reproducció.

Així, en el **capítol 3** d'aquesta Tesi, s'ha estudiat com afecten aquests tres fàrmacs: la fluoxetina, el diazepam i la carbamazepina als perfils de les oxilipines de la *D. magna*. En la **Taula 1.2** es mostra l'estructura d'aquests contaminants i se'n descriuen les propietats i els seus modes d'acció. Entre les propietats esmentades es troben la solubilitat en aigua i el coeficient de partició ($\log P$), el qual està relacionat amb la capacitat d'absorció del compost o el seu potencial de d'acumulació en el medi ambient. Pel que fa referència als fàrmacs descrits en la **Taula 1.2**, els que presenten un valor petit de $\log P$ indiquen una possible mobilitat del compost donada la seva bona solubilitat i biodegradació, és a dir que s'espera una baixa bioacumulació dels compostos. Mentre que un valor gran de $\log P$ indica una probable bioacumulació donada la seva escassa mobilitat que afavoreix la toxicitat d'aquests compostos.

Taula 1.2] Estructura, propietats i mode d'acció dels contaminants ambientals descrits en el DrugBank (www.drugbank.ca).

Contaminant	Estructura	Propietats*	Mode d'acció
Bisfenol A	 <chem>Cc1ccc(O)cc1C(C)c2ccc(O)cc2</chem> $C_{15}H_{16}O_2$	Solubilitat en aigua: $0.0865 \text{ mg mL}^{-1}$ <hr/> Log P: 3.81	Interfereix en el sistema endocrí humà, s'uneix específicament a un element resposta de l'estrogen i activa els gens que controlen la resposta dels estrògens.
Fluoxetina	 <chem>CNCC(Cc1ccccc1)Oc2ccc(C(F)(F)F)cc2</chem> $C_{17}H_{18}F_3NO$	*Solubilitat en aigua: $0.0017 \text{ mg mL}^{-1}$ <hr/> Log P: 4.09	Inhibidor del receptor de la serotonina, bloqueja la recaptació de la serotonina a la bomba de recaptació de serotonina de la membrana neuronal.
Diazepam	 <chem>CN1C=NC2=C1C(=O)C=C(Cl)C2c3ccccc3</chem> $C_{16}H_{13}ClN_2O$	Solubilitat en aigua: $0.0122 \text{ mg mL}^{-1}$ <hr/> Log P: 2.63	S'uneixen als receptors de benzodiazepina que intervenen en el son, afecta la relaxació muscular, l'activitat anticonvulsiva, la coordinació motora i la memòria.
Carbamazepina	 <chem>NC(=O)N1C=CC2=C1C=CC=C2</chem> $C_{15}H_{12}N_2O$	Solubilitat en aigua: 0.152 mg mL^{-1} <hr/> Log P: 2.1	Bloqueja els canals de sodi depenent de l'ús, així com també bloqueja la transmissió sinàptica que controla les convulsions.

*Propietat predictiva proporcionada pel DrugBank mitjançant el software ALOGPS (Tetko et al. 2005).

1.4 ESTRATÈGIES I PLATAFORMES ANALÍTIQUES BASADES EN L'ESPECTROMETRIA DE MASSES

La majoria dels estudis de metabolòmica ambiental es desenvolupen mitjançant tècniques instrumentals avançades, entre les quals trobem l'espectroscòpia de RMN (Puig-Castellví et al. 2015) i l'espectrometria de masses (MS) acoblada a tècniques de separació o mitjançant la injecció directa (*direct infusion mass spectrometry - DIMS*). DIMS) (Howlett et al. 2017). L'aplicació creixent de la RMN i de les plataformes basades en MS en aquesta àrea de la òmica ambiental han millorat en gran mesura la sensibilitat, la selectivitat i el rendiment analític. L'espectroscòpia de RMN és una tècnica quantitativa que ofereix una alta reproductibilitat. Mentre que l'MS ha esdevingut la tècnica més generalitzada gràcies a la possibilitat de realitzar d'una manera ràpida, sensible i selectiva tant els anàlisis qualitius com quantitius dels diferents metabòlits presents en la mostra. Aquests avenços en les tècniques analítiques han esdevingut crucials i han generat un increment exponencial dels estudis enfocats a investigar els efectes biològics dels estressants ambientals a través dels metabòlits. D'aquesta manera, la metabolòmica basada en MS s'ha convertit en una eina molt valuosa ja que proporciona una gran capacitat d'exploració la qual permet obtenir una informació extensa sobre els processos bioquímics i la caracterització d'alteracions de diverses rutes metabòliques (Clendinen et al. 2017).

Per altra banda, l'acoblament de l'MS amb les tècniques de separació és essencial a causa de la complexitat de les mostres biològiques. Entre aquestes tècniques s'hi inclouen la cromatografia de líquids (LC-MS) (Gómez-Canela et al. 2018; Navarro-Reig et al. 2015) i la GC-MS (Tang et al. 2013). A més, encara que no d'una manera tant ampla, també poden estar acoblades a l'MS tècniques com l'electroforesis capil·lar (CE-MS) (Ortiz-Villanueva et al. 2015) i la FT-ICR-MS (Brown et al. 2005; Ghaste et al. 2016), les quals també s'han emprat com a eines per a l'anàlisi de mostres biològiques complexes. Cada una d'aquestes plataformes basades en MS presenta diferents avantatges a la vegada que la combinació de varies d'elles permet una caracterització més completa del metaboloma (Rochat 2016).

Segons la finalitat de l'estudi de metabolòmica que es vol realitzar existeixen dues classes d'estratègies: l'anàlisi dirigida (*target*) i l'anàlisi no dirigida (*non-target* o *untarget*) (Clendinen, et al. 2017). Ambdues es focalitzen en comparar diferents grups biològics que permetin identificar els metabòlits que s'alteren en unes condicions específiques. Tal com es pot veure en la **Figura 1.3**, l'objectiu de l'anàlisi determina quina estratègia és la més adient per aquell estudi. L'anàlisi dirigida es centra en la determinació quantitativa dels metabòlits presents en les mostres biològiques. Per això, un dels primers passos està relacionat amb la preselecció dels metabòlits a analitzar. És en l'optimització del mètode analític i en la preparació de la mostra on es determinen quins són els metabòlits que s'avaluen i, en conseqüència, les rutes metabòliques estudiades. Per altra banda, en l'anàlisi no dirigida es pretén obtenir una visió global del perfil metabòlic del sistema biològic i de les rutes afectades com a conseqüència de l'exposició a la qual està sotmesa l'organisme.

Una vegada s'ha definit l'objectiu, cal realitzar un disseny experimental per tal de donar resposta al problema o hipòtesi que es vol resoldre a partir d'unes proves preliminars on s'han avaluat els efectes toxicològics dels compostos químics que es volen estudiar. Tal com es mostra en la **Figura 1.3**, els aspectes més rellevants a definir que inclou el disseny experimental són (i) els factors experimentals, el temps de l'exposició o la dosi, (ii) el tipus d'experiment en relació a l'organisme que s'utilitza per a realitzar l'estudi, i (iii) el tipus de comparacions que es volen determinar, per exemple si es tracta d'una comparativa entre mostres control i mostres exposades, entre diferents dosis d'exposició o si es tracta d'estudis amb una exposició temporal.

Un altre aspecte important és minimitzar les fonts de variabilitat no relacionades amb els processos biològics d'interès. És per això que cal determinar un nombre mínim de replicats biològics. Normalment es requereix d'un mínim de tres replicats biològics, però per tal de generar resultats consistents és recomanable obtenir-ne un mínim de cinc. A més, en els estudis de metabolòmica no dirigida, és aconsellable l'anàlisi de blancs i l'anàlisi de mostres control (*quality control samples*, QC) al llarg de l'experiment per tal de determinar les variacions instrumentals i experimentals que es puguin produir durant el procés analític (Zhou et al. 2012).

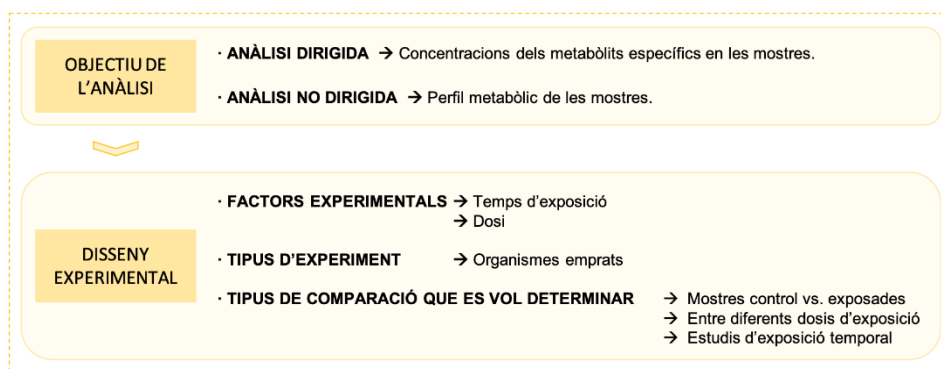


Figura 1.3| Esquema dels aspectes a definir per escollir l'estratègia en els estudis de metabolòmica, segons l'objectiu i el disseny experimental de l'anàlisi.

1.4.1 ANÀLISI DIRIGIDA

La finalitat de realitzar l'anàlisi dirigida és mesurar, detectar i quantificar un grup definit de metabòlits prèviament caracteritzats els quals poden pertànyer a una o varies rutes bioquímiques (Griffiths et al. 2010). Es poden establir les interaccions i les relacions entre aquests metabòlits i interpretar-les en un context biològic. En aquesta estratègia, els mètodes basats en MS s'optimitzen per detectar amb precisió un conjunt predeterminat de compostos en detriment d'una cobertura general del metaboloma.

En la **Figura 1.4** es mostra un exemple dels diferents processos o etapes de treball dels estudis de metabolòmica dirigida. En primer lloc s'analitzen els estàndards dels metabòlits escollits, que s'utilitzen pel desenvolupament i l'optimització del mètode analític amb la corresponent generació de la seva corba de calibratge. Després, s'analitzen les mostres biològiques i es quantifiquen els metabòlits escollits. Per tant, les dades recollides correspondran de forma específica a un grup concret de metabòlits i d'aquelles rutes metabòliques implicades.

Aquesta estratègia dirigida permet la quantificació dels metabòlits presents en la mostra mitjançant l'espectrometria de masses en tàndem (MS/MS) amb l'adquisició per MRM (*Multiple Reaction Monitoring*) dels senyals específics dels metabòlits seleccionats. Així, es poden determinar amb precisió les abundàncies i les concentracions dels metabòlits analitzats encara que es trobin a concentracions baixes. La utilització d'un calibratge intern mitjançant un compost de referència (afegit abans de l'extracció de la mostra) permet minimitzar la variabilitat associada a l'instrument, als

efectes de matriu i a les interferències de la mostra. Un desavantatge important de l'estratègia dirigida és que cal tenir un coneixement previ dels metabòlits d'interès i una disponibilitat com estàndards. Per això, aquest tipus d'estratègia no permet detectar els metabòlits desconeguts o no seleccionats prèviament, i per tant els canvis significatius o que tinguin una gran rellevància en d'un sistema biològic passen inadvertits.

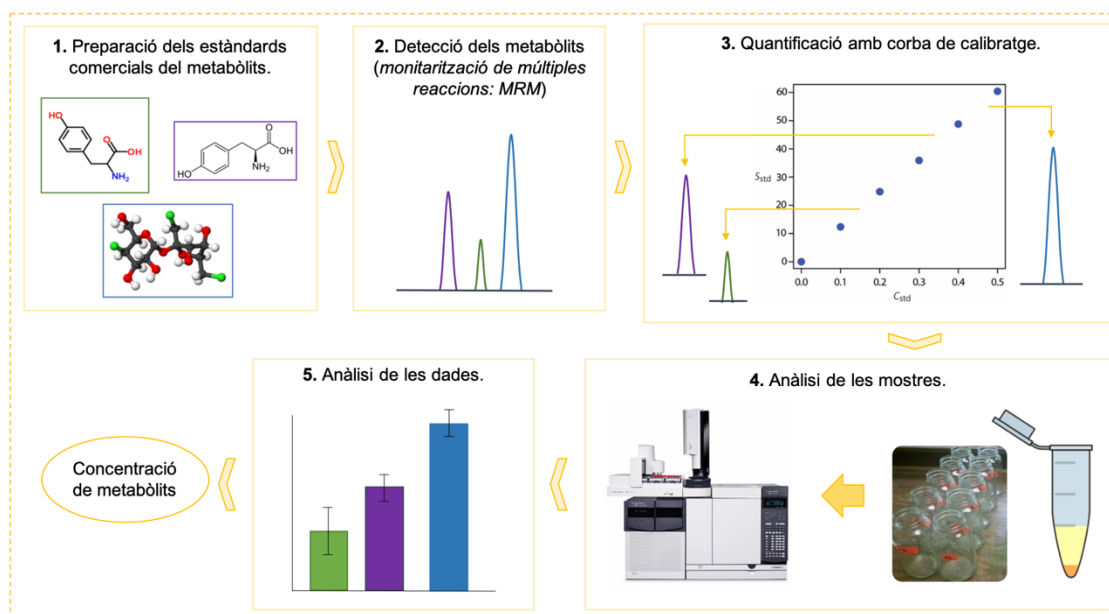


Figura 1.4 | Diagrama de processos per realitzar estudis de metabolòmica basats en tècniques de MS per a estratègies dirigides.

Les dues tècniques basades en MS més utilitzades en els estudis de metabolòmica són la LC-MS i la GC-MS (Begou et al. 2017b). En els seus inicis, la GC-MS va ser àmpliament emprada en el camp de la metabolòmica de plantes, ja que aquesta tècnica és ideal per a l'anàlisi de molècules volàtils. Tot i això, també es pot utilitzar per a la separació de compostos semi-volàtils o no volàtils després de realitzar un procés previ de derivatització dels metabòlits. D'aquesta manera, la GC-MS/MS permet detectar una àmplia gama de compostos endògens, com aminoàcids, àcids orgànics i sucres en diferents tipus de matrius biològiques. Per exemple, Tsugawa et al. (2014) va desenvolupar un mètode de GC-MS/MS per a la determinació de 110 metabòlits, on s'inclouen aminoàcids, àcids orgànics, amines, sucres, nucleòtids, etc., en plasma de ratolí i en llevat (*Saccharomyces cerevisiae*).

En els darrers anys, la LC-MS ha crescut en popularitat ja que permet l'anàlisi de compostos polars i semi-polars sense la necessitat de realitzar el procés de derivatització previ dels metabòlits. Per altra banda, també és possible la combinació de les dues tècniques analítiques per a realitzar estudis de metabolòmica dirigida i determinar un rang més ampli de classes de metabòlits. Per exemple, Kato et al. (2012) va desenvolupar un mètode per a la determinació de metabòlits hidròfils en llevat, capaç de determinar i quantificar un total de 60 metabòlits (sucres, aminoàcids i àcids orgànics) mitjançant la GC acoblada a un quadrupol simple (GC-Q-MS) amb el mètode de detecció SIM (*Selected Ion Monitoring*). A més, es va poder detectar 48 metabòlits (cofactors, fosfats de sucres i àcids orgànics) amb l'anàlisi mitjanant la LC acoblada a un triple quadrupol (LC-QqQ-MS).

Tot i això, no és possible una cobertura completa de totes les classes de metabòlits presents en les mostres biològiques en una única separació. És per això que s'han desenvolupat nombrosos mètodes i protocols que agrupen anàlisis de diverses classes de metabòlits específics, des de compostos hidròfils com els aminoàcids, àcids orgànics i sucres, fins a compostos liòfils com lípids, eicosanoids i oxilipines (Lu et al. 2008). Per exemple, Courant et al. (2018) van avaluar els efectes del diclofenac en musclos (*Mytilus galloprovincialis*) mitjançant un equip d'alta resolució (*high resolution mass spectrometry*, HRMS) *Exactive* LC-MS en mode MS/MS on es van poder determinar i quantificar els nivells de les prostaglandines. També, Gómez-Canela, et al. (2018) van optimitzar un mètode LC-MS/MS per a la determinació de neurotransmissors i de metabòlits relacionats en peixos zebra (*Danio rerio*), el qual va permetre la separació de 38 compostos, 27 dels quals van ser quantificats en els animals exposats a diferents concentracions de plaguicides organofosfats.

1.4.2 ANÀLISI NO DIRIGIDA

La metabolòmica no dirigida també es pot descriure com l'anàlisi global dels perfils metabòlics (*metabolic profiling*) de la mostra analitzada, on es pretén caracteritzar tots els metabòlits presents en una mostra biològica (Patti et al. 2012). En la **Figura 1.5**, es mostra el diagrama de processos que segueixen els estudis de

metabolòmica no dirigida, on s'inclouen la preparació de la mostra, l'anàlisi instrumental, l'anàlisi de les dades i la interpretació de les dades (Dunn and Ellis 2005). Més específicament, els estudis de metabolòmica ambiental es comprenen de les següents etapes:

- 1) Recollida, preparació i tractament de la mostra.
- 2) Elecció de la plataforma analítica a utilitzar per a l'anàlisi.
- 3) Preprocessament de les dades i aplicació de les tècniques estadístiques multivariants que permeten detectar les diferències entre els perfils metabòlics dels diferents grups d'estudi.
- 4) Identificació dels metabòlits.
- 5) Interpretació biològica de les rutes metabòliques afectades per l'exposició.

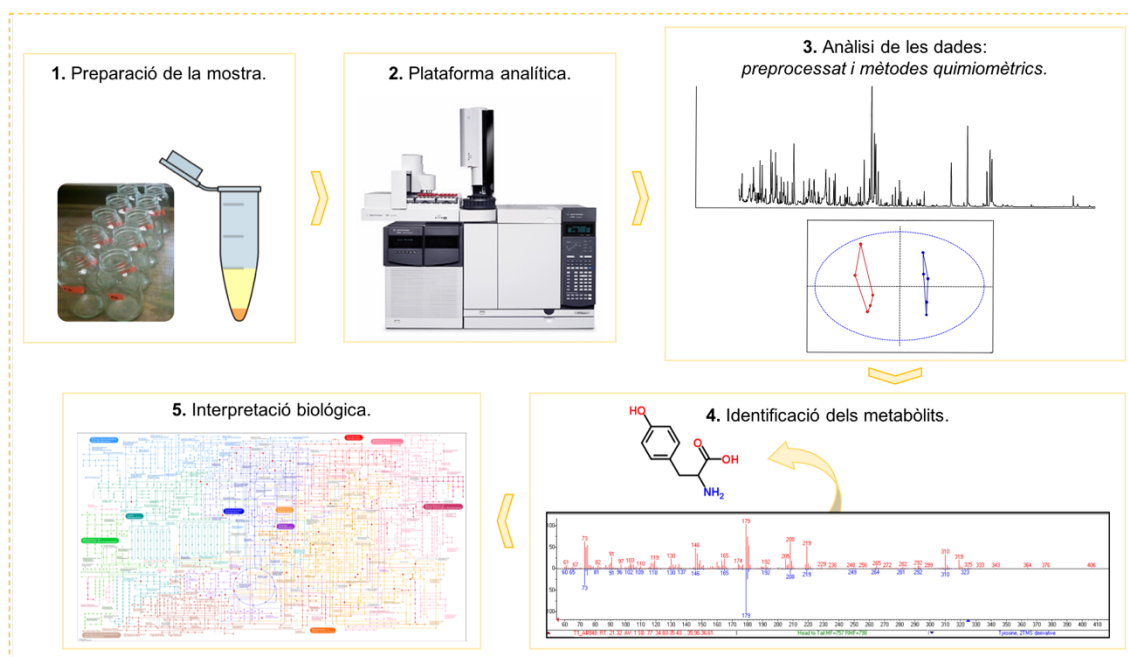


Figura 1.5] Diagrama de processos per estudis de metabolòmica basats en tècniques de MS per a estratègies no dirigides.

Entre els aspectes que caracteritzen l'estratègia no dirigida es troba l'etapa de l'anàlisi de les dades, la qual inclou el preprocessament i el processament o tractament de les dades. Un dels objectius dels mètodes de preprocessament és eliminar les diferències de magnituds entre els metabòlits presents en la mostra, considerant-los igual d'importants en el sistema biològic, tant si es troben a concentracions baixes o

altes. Aquests mètodes de preprocessament de les dades s'expliquen amb més detall en la **Secció 1.5.1** d'aquesta Tesi. Per altra banda, el tractament de les dades permet l'avaluació del conjunt de dades generades per aquests estudis. Els mètodes quimiomètrics que s'han emprat en els estudis de metabolòmica no dirigida d'aquesta Tesi estan descrits en la **Secció 1.5.2**.

La combinació de tècniques de separació acoblades a MS d'alta resolució ha estat l'ingredient essencial en l'èxit de la metabolòmica no dirigida, ja que aquesta combinació facilita l'anàlisi de les múltiples i complexes famílies de metabòlits presents en les mostres biològiques. Entre les tècniques basades en MS, la cromatografia de gasos amb impacte electrònic acoblada a l'espectrometria de masses (GC-EI-MS) és una eina molt utilitzada en metabolòmica no dirigida. La GC convencional (en una dimensió, 1D) proporciona una bona resolució capaç de determinar un gran nombre de metabòlits presents en les mostres biològiques. Tot i això, la tècnica de cromatografia de gasos bidimensional (2D) acoblada a l'espectrometria de masses (GCxGC-MS) constitueix una eina molt valuosa a causa de la seva gran resolució cromatogràfica la qual permet resoldre un ventall molt ampli de metabòlits presents en les mostres ambientals. L'augment multiplicatiu de la capacitat de pic i la millora significativa de la sensibilitat en la detecció permeten una bona separació dels compostos de baix pes molecular en mostres biològiques complexes (Wachsmuth et al. 2013).

Ambdues tècniques, GC-MS i GCxGC-MS, són capaces de determinar centenars de metabòlits diferents en una única anàlisi. La gran disponibilitat de biblioteques espectrals per a GC-EI-MS combinades amb les dades dels índex de retenció fan possible la identificació dels metabòlits desconeguts. Entre altres exemples, Yang et al. (2019) va desenvolupar un mètode GCxGC-TOFMS per a la determinació de les respostes metabòliques de les ostres (*Pinctada fucata martensii*), on es van identificar un total de 84 metabòlits, que juntament amb la combinació de l'anàlisi multivariant va permetre l'avaluació i diferenciació dels perfils metabòlics segons el tipus d'alimentació de les ostres. A més, la GCxGC-TOFMS amb la combinació de l'anàlisi multivariant també va permetre determinar els canvis metabòlics produïts en 45 metabòlits per l'efecte de la sulfametazina en peixos (*Oryzias melastigma*) a les tres concentracions

estudiades (0.01 mg L^{-1} , 0.1 mg L^{-1} i 1 mg L^{-1}) (Liu et al. 2018). Altres autors proposen combinar les dues tècniques per tal d'obtenir un perfil metabòlic més extens. Per exemple, la combinació de la LC-MS i la GC-MS va permetre a Jenkins et al. (2017) identificar un total de 96 metabòlits amb característiques diferents presents en els microorganismes del sòl. La utilització de la LC-MS va permetre identificar 85 metabòlits, on s'inclouïen aminoàcids, derivats dels aminoàcids, àcids mono- i dicarboxílics, nucleobases, nucleòsids i diverses classes de sucres, mentre que la GC-MS va ampliar amb 11 metabòlits més els compostos identificats.

Tot i això, en la gran majoria d'estudis de metabolòmica no dirigida, la tècnica més utilitzada és la LC-MS, sobretot amb els analitzadors dels espectròmetres de masses d'alta resolució com el temps de vol (*Time of flight*, TOF) o l'Orbitrap. Entre altres exemples, Zendong et al. (2015) van posar de manifest el gran potencial de l'HRMS Orbitrap *Exactive* respecte de la LC-QqQ-MS per a la detecció de les toxines d'algues en musclos (*Mytilus galloprovincialis*), concretament per a la determinació de toxines emergents o fins ara no caracteritzades. A més, Sotto et al. (2017) van avaluar els efectes de diferents antibiòtics (claritromicin, florfenicol, sulfametazina i la seva barreja) a una concentració de 0.1 mg L^{-1} en peixos zebra mitjançant una anàlisi no dirigida amb LC acoblada a un analitzador quadrupolar i un temps de vol (QTOF-MS), on es van determinar un total de 810, 919, 303 i 203 metabòlits significatius per a l'exposició als antibiòtics anomenats anteriorment. En conjunt, aquests treballs han permès evidenciar la gran aplicabilitat de la metabolòmica ambiental per avaluar els efectes en els éssers vius, fins i tot a concentracions baixes.

La dificultat principal de l'estratègia no dirigida recau en l'anàlisi simultània d'un ampli ventall de compostos amb propietats i concentracions molt diferents. Per això, és necessària la utilització de bases de dades per identificar els metabòlits i les rutes metabòliques implicades. Entre altres exemples, la *Human Metabolome Database* (HMDB) (Wishart et al. 2013), la *METLIN Database* (METLIN) (Guijas et al. 2018) o la base de dades *Golm Metabolome Database* (GMD) (Hummel et al. 2013) proporcionen informació per realitzar aquesta identificació i determinar les principals vies

metabòliques implicades, així com interpretar la importància biològica dels metabòlits en l'organisme d'estudi.

D'aquesta manera, l'aplicació de les dues estratègies descrites permeten ampliar el coneixement dels compostos desconeguts. Per exemple, Chen et al. (2013) van proposar un estudi on es va combinar ambdues estratègies per tal de cobrir un ventall més ampli de metabòlits que permetés la detecció de més de 500 metabòlits en sèrum humà. El procediment que van proposar s'inicia amb l'anàlisi no dirigida de les mostres de sèrum mitjançant LC-QTOF-MS, i acaba amb la implementació del mètode dirigit utilitzant LC-QqQ, on s'empren les transicions MS/MS seleccionades a partir de l'anàlisi prèvia no dirigida per a la identificació dels metabòlits. Més recentment, Gil-Solsona et al. (2017) van utilitzar un mètode de LC-QTOF-MS per a la determinació dels biomarcadors sensibles a la desnutrició dels peixos de granja. En primer lloc, es van caracteritzar els perfils metabòlics dels dos grups experimentals (peixos alimentats i peixos en dejú), i en segon lloc, a través dels seus espectres MS/MS i de les bases de dades de metabolòmica disponibles es van identificar fins a 45 metabòlits més significatius característics de l'efecte de la desnutrició.

1.5 ANÀLISI DE LES DADES

En els estudis de metabolòmica es generen un gran conjunt de dades complexes, sobretot quan s'apliquen estratègies no dirigides. Els nous avenços instrumentals, tant en sensibilitat com en precisió instrumental, han incrementat de manera considerable la quantitat d'informació disponible. Normalment, els estudis no dirigits requereixen l'ús d'eines estadístiques multivariants capaces de gestionar aquesta gran quantitat de dades generades en forma de variables (metabòlits) i de mostres. En metabolòmica, l'anàlisi multivariant de dades permet entre altres coses reduir la mida dels conjunts de dades generades, i trobar les tendències i les característiques de les mostres que permeten distingir entre mostres control i mostres tractades. La bioinformàtica i els mètodes quimiomètrics han permès facilitar l'anàlisi de la informació generada i interpretar els conjunts de dades obtingudes. Existeixen diferents mètodes

quimiomètrics de tractament de les dades entre els quals s'inclouen l'anàlisi de components principals (PCA), i l'anàlisi de mínims quadrats parcials (PLS-DA).

La informació que es pot obtenir amb l'aplicació dels mètodes quimiomètrics dependrà de com s'han estructurat les dades experimentals. Aquestes dades es poden organitzar de diverses maneres, el més comú és organitzar-les en una matriu de dades en dues direccions (*two-way data sets*), una direcció que conté la informació relacionada amb els cromatogrames, i l'altra direcció amb la informació dels espectres de masses. Les dades obtingudes mitjançant les tècniques LC-MS i GC-MS corresponen a aquest tipus de dades experimentals. En cas de tenir dades en tres direccions, com en el cas de les dades obtingudes mitjançant la GCxGC-MS, les dades es poden agrupar en un cub de dades (*three-way data sets*).

En la **Figura 1.6A** es mostra com s'organitzen aquests diferents tipus de dades. En primer lloc la matriu de dades (**D**), que correspon a una mostra, està formada pels espectres de masses en la direcció de les files de la matriu, i dels cromatogrames en la direcció de les columnes de la matriu. Generalment, en els experiments de metabolòmica s'analitzen un conjunt de mostres, de manera que disposarem de diferents matrius de dades, tal com es mostra en la **Figura 1.6B**. Així, aquestes diferents matrius de dades s'organitzen en forma de cub o en una matriu augmentada (**D_{aug}**) (veure **Figura 1.6C**). En aquesta matriu augmentada, **D_{aug}**, les diverses matrius individuals de dades es poden agrupar en la direcció de les columnes (*column-wise*) o de les files (*row-wise*). La disposició més comuna per als estudis de metabolòmica per cromatografia acoblada a la MS és augmentar les diverses matrius de dades corresponents a diferents mostres en la direcció de les columnes, com es mostra en la **Figura 1.6C**. D'aquesta manera, les matrius de dades de totes les mostres tenen en comú les seves columnes (valors *m/z*), però poden variar en el seu nombre de files (temps de retenció). Aquesta augmentació de les dades és possible perquè totes les matrius individuals de dades (mostres) estan analitzades amb la mateixa tècnica analítica, en aquest cas d'espectrometria de masses.

Per altra banda, la **Figura 1.7** mostra un cub de dades corresponent a les dades obtingudes mitjançant GCxGC-MS. En aquest cub de dades cada mostra està descrita

mitjançant tres direccions: dues de les quals proporcionen la informació cromatogràfica en les dues columnes i una tercera direcció descriu els espectres de masses. Les tres direccions del cub de dades (\mathbf{X}) estan definides per l'eix x corresponent als espectres de masses, i l'eix y i el z que corresponen als temps de retenció en les dues columnes cromatogràfiques (**Figura 1.7A**). Aquest conjunt de dades (\mathbf{X}) conté les diferents modulacions GCxGC. Aquestes modulacions corresponen a la separació cromatogràfica sencera en la segona columna de cada porció de mostra extreta de la primera columna, cadascuna de les quals permeten construir la matriu de dades. Les diferents modulacions conformen l'estructura en tres direccions en forma de cub, a partir de la qual es pot construir també una matriu augmentada (\mathbf{X}_{aug}) en la direcció de les columnes (*column-wise*). Això s'aconsegueix disposant les matrius de cada modulació una sota de l'altra amb els valors de m/z en comú, tal com es mostra en la **Figura 1.7A**. En el cas que es disposi de diverses mostres, aquestes s'agrupen en forma de diversos cubs, com es pot veure en la **Figura 1.7B**. Aquests cubs poden reorganitzar-se en una nova matriu superaugmentada que conté tota la informació de totes les mostres i que recull tota la informació de les diferents mostres.

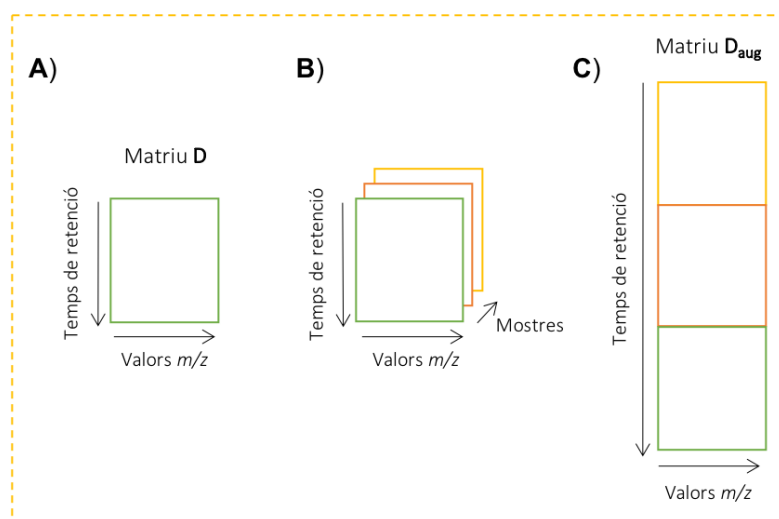


Figura 1.6 **A)** Representació d'una matriu \mathbf{D} de dades bidimensional. **B)** Representació de diverses mostres analitzades d'un estudi de metabolòmica. **C)** Representació d'una matriu augmentada, \mathbf{D}_{aug} , en la direcció de les columnes (*column-wise*).

Per poder aplicar moltes de les eines d'anàlisi multivariant i de l'àlgebra lineal és necessari que les dades estiguin organitzades en forma de matriu de dades. Per això,

aquests conjunts de dades s'augmenten en la direcció de les columnes amb els valors de m/z en comú i així, es genera una nova matriu de dades superaugmentada on es podrà portar a terme la seva anàlisi simultània. En aquesta Tesi s'utilitzen mètodes basats en models matemàtics bilineals, com per exemple el mètode MCR-ALS (*Multivariate Curve Resolution – Alternating Least Squares*) (De Juan and Tauler 2001).

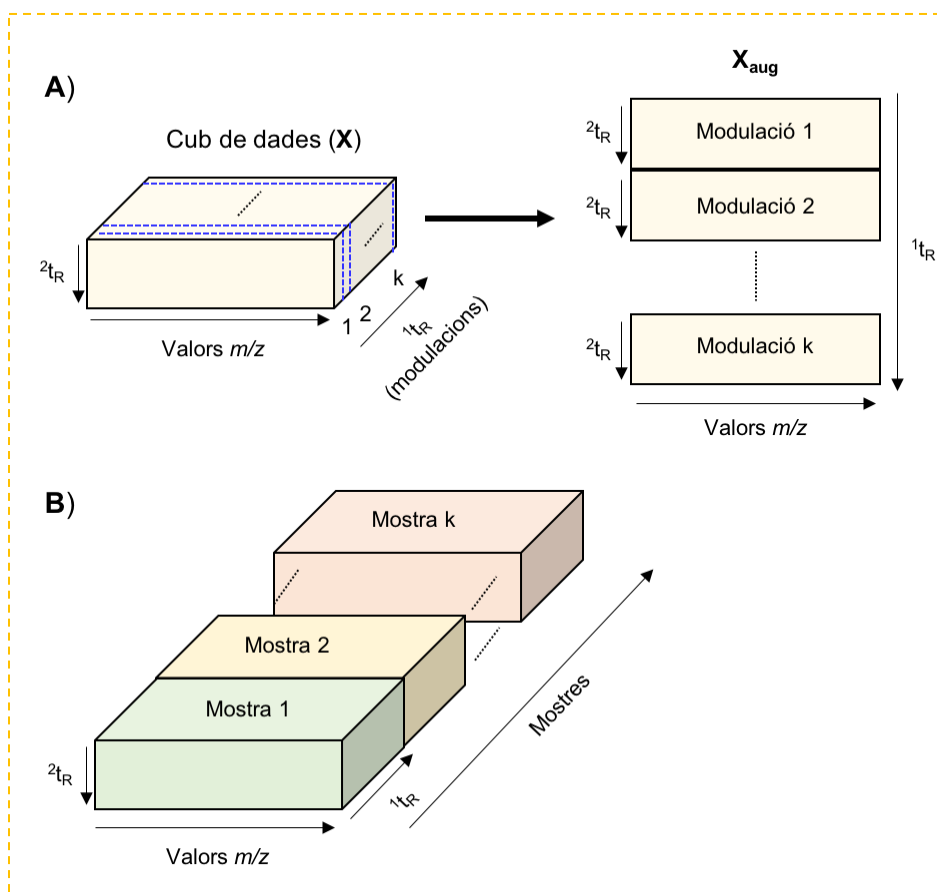


Figura 1.7| **A)** Representació d'un cub de dades tridimensional, on es mostra el desplegament del cub de dades (\mathbf{X}) amb les diferents modulacions donant lloc a una matriu de dades augmentada, \mathbf{X}_{aug} , en la direcció de les columnes (valors m/z). **B)** Representació d'un hiper-cub de dades de diverses mostres en tres dimensions.

1.5.1 PREPROCESSAMENT DE LES DADES

Els mètodes de preprocessament de les dades són força importants per gestionar la informació generada experimentalment d'una manera fàcil i eficient, i millorar la seva posterior anàlisi i interpretació. Aquests tractaments preliminars de les dades permeten reduir la mida de les dades i eliminar o reduir la variació d'aquestes que pot afectar la seva adequada interpretació, com per exemple possibles artefactes

experimentals o instrumentals i/o les variacions sistemàtiques no relacionades amb les mostres d'anàlisi (Zeaiter and Rutledge 2009).

En la **Figura 1.8**, es descriu esquemàticament el flux de treball que s'utilitza generalment en el processament de les dades de metabolòmica, el qual consta de quatre parts: l'adquisició de les dades, el seu preprocessament, l'aplicació dels mètodes quimiomètrics i la interpretació biològica. Com es pot observar, el preprocessament de les dades correspon a un pas intermedi entre l'adquisició de dades mitjançant les plataformes analítiques emprades, i la posterior aplicació dels mètodes quimiomètrics, que permeten la consegüent interpretació biològica de les dades (**Figura 1.8**).

No existeix un procediment únic estàndard de pretractament de les dades sinó que existeixen un ventall ampli de mètodes segons el tipus d'informació que es vol obtenir (Karaman 2017; Vettukattil 2015; Zeaiter and Rutledge 2009). El pretractament de les dades és crític ja que pot influir en els resultats i la seva interpretació. Per això, és important que els mètodes emprats siguin robustos, per tal d'evitar introduir variacions en les dades que les transformin d'una manera errònia. En les tècniques basades en l'espectrometria de masses els mètodes més comuns de pretractament de les dades són l'alineament dels pics cromatogràfics, la correcció de la línia base, la normalització i l'escalat de les dades. A continuació, es descriuen els mètodes de pretractament utilitzats en els treballs d'aquesta Tesi.

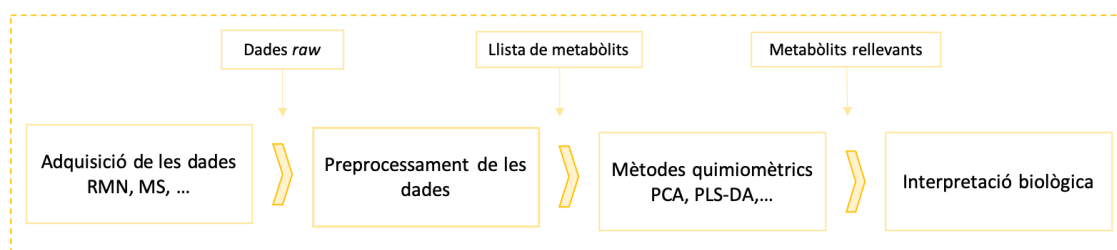


Figura 1.8 Flux de treball on es descriuen esquemàticament els passos a seguir en l'anàlisi de les dades, des de la seva adquisició fins a la interpretació biològica d'aquestes.

Alineament dels pics

L'alineació dels pics cromatogràfics pretén eliminar les diferències presents entre les mostres analitzades, les quals són produïdes per fluctuacions naturals i

ambientals de les mostres, i per la tècnica analítica emprada, la qual pot influir en les mesures instrumentals. Un dels efectes d'aquestes variacions pot ser el desplaçament en els temps de retenció cromatogràfics, produint un desajust a l'hora d'identificar els pics cromatogràfics en els diferents cromatogrames dels conjunt de mostres analitzades. Aquestes diferències entre cromatogrames impedeixen la correcta identificació dels compostos a partir del seu temps de retenció. Per això, amb l'alineació dels cromatogrames s'aconsegueix una millora en la fiabilitat i la robustesa de la interpretació dels resultats.

Existeixen varies tècniques per realitzar aquesta alineació (Yi et al. 2016), on es minimitzen els canvis de les dades inicials per tal de fer-los aplicables als sistemes cromatogràfics. Entre aquests procediments, s'inclou la deformació optimitzada de correlació (COW, *correlation optimized warping*) (Nielsen et al. 1998), la qual és aplicable als cromatogrames d'ions totals (*Total Ion Chromatogram*, TIC) i és útil per afrontar el problema dels canvis dels temps de retenció cromatogràfics al llarg de les mostres analitzades (Christin et al. 2008). L'objectiu d'aquest mètode és alinear un vector de dades en relació a un vector de referència. El vector referència és una mostra escollida com a representativa del conjunt. Quan el nombre de temps de retenció entre el vector a alinear i el vector de referència són diferents, es realitza també una interpolació lineal del primer vector per tal de crear un nou vector de la mateixa magnitud (Zeaiter and Rutledge 2009). Així, l'alineament es produeix en la direcció del temps de retenció dels cromatogrames. En aquesta Tesi, aquest preprocessament s'ha aplicat als cromatogrames TIC abans de l'anàlisi PLS-DA (veure **Secció 1.5.2**).

Correcció de la línia base

Les distorsions de la línia base poden afectar l'aplicació dels mètodes quimiomètrics i la quantificació del metabòlits. Les intensitats del senyal es calculen en referència a la línia base. Per això, una correcció de la línia base inadequada pot interferir de manera negativa en l'anàlisi de les dades. Aquests mètodes s'apliquen per reduir la influència de la variació instrumental no desitjada, la qual pot venir donada per canvis en la resposta del detector o per sorolls electrònics dels instruments analítics. Encara que existeixen diferents maneres de corregir aquesta distorsió, entre els

mètodes més comuns s'inclou el mètode de mínims quadrats asimètrics (AsLS, *Asymmetric Least Squares*) (Peng et al. 2010), el qual s'utilitza per suavitzar la línia base i corregir el soroll de fons. El mètode AsLS és una eina molt efectiva per a la correcció de la línia base tant dels espectres com dels cromatogrames, ja que en millora les seves característiques. En aquesta Tesi, aquest preprocessament s'ha aplicat als cromatogrames TIC abans de l'anàlisi PLS-DA (veure **Secció 1.5.2**).

Normalització i escalat de les dades

L'objectiu principal de la normalització i de l'escalat de les dades és la reducció de la variació sistemàtica entre mostres i les diferències d'escala entre les variables mesurades. Aquests preprocessaments són adients per facilitar la interpretació dels resultats quan les variables tenen escales amb ordres de magnitud diferents.

La normalització de les mostres és necessària per reduir les variacions sistemàtiques degudes a diferències experimentals i a derives instrumentals. Aquest pretractament s'aplica per tal d'ajustar les diferències existents entre les mostres analitzades, i es realitza en la direcció de les files (mostres). Per tal de minimitzar les variacions internes entre mostres durant el procediment experimental s'utilitzen diferents patrons d'extracció o *surrogates*, afegits abans de l'extracció de la mostra, i patrons interns, afegits després de l'extracció. Per altra banda, el pretractament d'escalat en les dades de metabolòmica és útil quan es compararen variables amb magnituds diferents, ja que l'escalat permet que totes les variables tinguin el mateix pes i la mateixa importància relativa dins de l'anàlisi (Zeaiter and Rutledge 2009). D'aquesta manera, és possible comparar diferents metabòlits que es troben a nivells de concentració molt diferents i eliminar la variació que existeix entre elles. En aquesta Tesi, s'han emprat dos mètodes de pretractament de les variables, el centrat (*mean-center*) i l'autoescalat.

El mètode centrat de les dades és un procediment on a cada valor original de les variables es resta el valor de la seva mitjana. Això resulta en una nova matriu de dades on cada variable té una mitjana igual a 0, de manera que totes les dades tenen la mitjana de les dades com a zero o origen (Zeaiter and Rutledge 2009). Aquest

preprocessament és especialment útil per a l'eliminació de desplaçaments sistemàtics o *offsets*. En aquesta Tesi, aquest procediment s'aplica en els cromatogrames TIC abans de fer la seva anàlisi PLS-DA, de manera que permet mostrar millor les semblances o les diferències entre les mostres analitzades (Karaman 2017).

En el pretractament d'autoescalat s'aplica simultàniament el centrat i l'escalat de les variables. D'aquesta manera, a cada valor original de les variables se li resta el valor de la mitjana dels valors d'aquesta variable, i es divideix per la seva desviació estàndard. Així, totes les variables tenen una mitjana igual a 0 i una desviació estàndard igual a 1 (Zeater and Rutledge 2009). El resultat d'aquest preprocessament és semblant a l'escalat de les dades, on per una banda s'igualava l'escala de les variables i, per una altra banda, s'eliminen els efectes constants o *offsets*. Sobretot s'aplica aquest pretractament en mètodes d'anàlisi de dades que es basen en l'avaluació de la variància observada per variables diferents mesurades entre diferents mostres.

Compressió de les dades en la direcció cromatogràfica: finestres de temps i transformades d'ones

Les estratègies d'anàlisi no-dirigides generen grans conjunts de dades multivariants que requereix d'una gran capacitat d'emmagatzematge. Moltes vegades la reducció de la mida del conjunt de dades originals és un pas necessari, sense que això suposi una pèrdua d'informació. En aquesta Tesi, per tal de tractar les dades obtingudes mitjançant les tècniques de GC-MS i GCxGC-MS, s'ha realitzat la seva compressió prèvia en la direcció cromatogràfica (dels temps d'elució) a partir dels procediments basats en les finestres de temps (*time windowing*) i en les transformades d'ones (*wavelet transform*).

Per emprar l'estratègia de les finestres de temps s'han de dividir els cromatogrames de GC-MS en diferents regions de temps, tal com es mostra en la **Figura 1.9**. Les diferents finestres cromatogràfiques considerades poden tenir temps de retenció diferents. Quan s'analitzen diverses mostres simultàniament, la matriu augmentada de les dades GC-MS, D_{aug} , amb la informació de totes les mostres, es divideix en submatrius que contenen les diferents finestres cromatogràfiques. Així, cada

una d'aquestes submatrius es troba augmentada ($DF1_{aug}$) amb un nombre de files igual a la suma del nombre de temps de retenció considerats en cada finestra de temps, a la vegada considerada en cadascuna de les mostres de l'estudi, tal com es descriu en la **Figura 1.10**. Les dades comprimides de totes les mostres s'organitzen en una nova matriu augmentada per a cada finestra de temps, $DF1_{aug}$, que inclou a totes les mostres, per a la posterior aplicació del mètode MCR-ALS (veure la **equació 1.5** en la **Secció 1.5.2**). Així, a partir de MCR-ALS s'obté la matriu comprimida dels perfils d'elució (C_{comp}) i la matriu amb els espectres de masses (S^T).

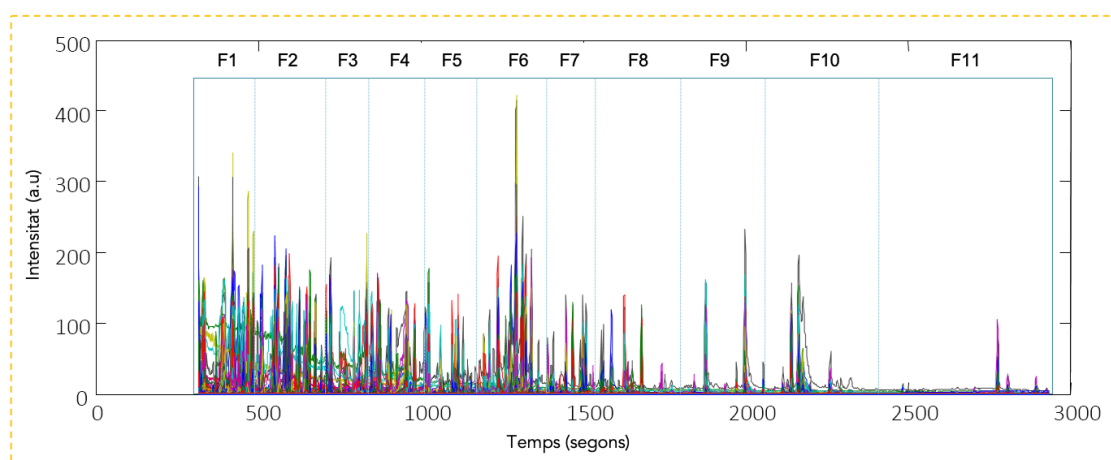


Figura 1.9 | Representació gràfica de la divisió d'un cromatograma de GC-MS en 11 finestres de temps (F).

Una altra alternativa per a la compressió de les dades és la utilització de les transformades d'ones (*wavelet transform*). Aquest procediment és una transformació que expressa la senyal original en termes de funcions bàsiques diferents a les de les dades originals sense perdre informació rellevant. Aquesta transformació d'ones s'aplica independentment a cada cromatograma de la matriu de dades GCxGC-MS (X_{aug}). La compressió es realitza a diferents nivells. Tal com es representa en la **Figura 1.11**, a cada nivell s'obtenen dos conjunts de coeficients, un que permet la descripció aproximada del senyal i l'altre que conté els detalls del senyal, anomenats amb les lletres A i D, respectivament. La transformada d'ones comprimeix el senyal 2^n vegades, on n és el nivell de compressió (Mallat 2009; Meyer 1992; Sayood 2005). Per representar el senyal d'una manera reduïda s'usa el coeficient amb la descripció aproximada, ja que conté la mateixa informació que X_{aug} però comprimida i filtrada amb menys soroll.

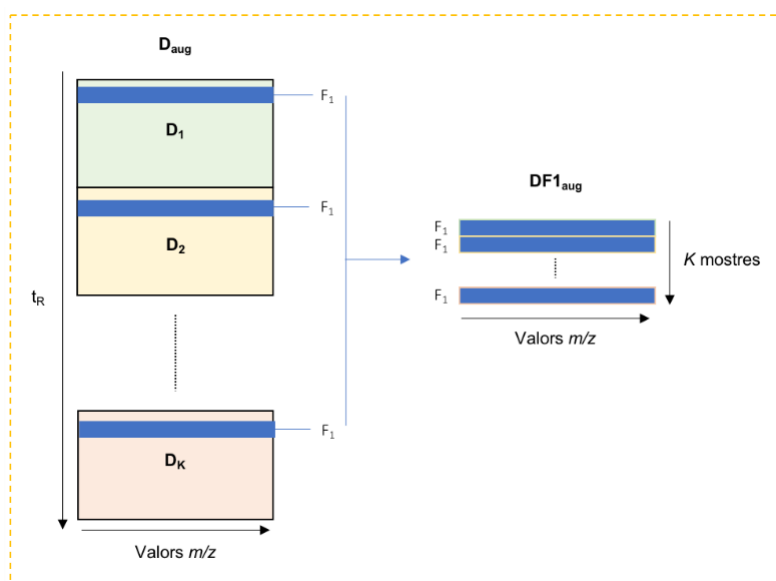


Figura 1.10 Representació gràfica del procediment de compressió per finestres de temps, on es mostra la formació d'una matriu augmentada ($DF1_{aug}$) que correspon a una finestra de temps determinada (F_1).

En els estudis de metabolòmica descrits en aquesta Tesi, la compressió de les dades GCxGC-MS es porta a terme de manera individual per a cada una de les matrius augmentades de cada mostra (X_{aug}). Aquesta matriu comprimida conté la mateixa informació que X_{aug} però necessita menys espai d'emmagatzematge. Les dades comprimides de totes les mostres s'organitzen en una matriu superaugmentada, X_{compr} , i s'analitzen amb el mètode MCR-ALS (veure posteriorment la **equació 1.6** en la **Secció 1.5.2**). Així, s'obté una matriu comprimida amb la informació dels perfils d'elució (C_{compr}) i una matriu amb els espectres de masses (S^T). Un cop obtinguda la matriu dels espectres de masses de tots els components de la mostra, és possible recuperar els perfils d'elució descomprimits a tots els temps de retenció originals ($C_{superaug}$) mitjançant una regressió lineal no negativa per mínims quadrats (*non-negativity constrained least squares*) (Bro and De Jong 1997).

En els estudis de metabolòmica descrits en aquesta Tesi, la compressió de les dades GCxGC-MS es porta a terme de manera individual per a cada una de les matrius augmentades de cada mostra (X_{aug}). Aquesta matriu comprimida conté la mateixa informació que X_{aug} però necessita menys espai d'emmagatzematge. Les dades comprimides de totes les mostres s'organitzen en una matriu superaugmentada, X_{compr} ,

i s'analitzen amb el mètode MCR-ALS (veure posteriorment la **equació 1.6** en la **Secció 1.5.2**). Així, s'obté una matriu comprimida amb la informació dels perfils d'elució (C_{comp}) i una matriu amb els espectres de masses (S^T). Un cop obtinguda la matriu dels espectres de masses de tots els components de la mostra, és possible recuperar els perfils d'elució descomprimits a tots els temps de retenció originals (C_{superaug}) mitjançant una regressió lineal no negativa per mínims quadrats (*non-negativity constrained least squares*) (Bro and De Jong 1997).

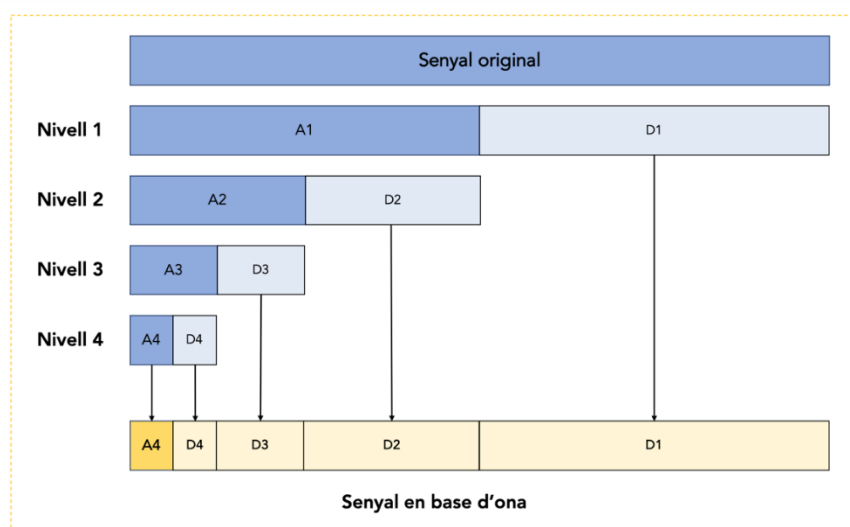


Figura 1.11 | Esquema del procediment de transformació d'ones de nivell 4. Amb la lletra A es representen els coeficients que contenen la informació aproximada del senyal i amb una D els coeficients que contenen els detalls.

Compressió de les dades en la direcció espectral: regions d'interès (ROIs) i procediment ROIMCR

Aquesta estratègia de compressió s'ha emprat especialment en l'anàlisi de les dades obtingudes mitjançant LC-MS en el **capítol 3** d'aquesta Tesi. L'aplicació d'aquest procediment permet la filtració i la reducció de la mida de les dades sense perdre resolució espectral respecte de les dades originals. A diferència dels procediments de compressió en la direcció cromatogràfica descrits anteriorment, el procediment de les ROIs s'aplica en la direcció espectral de les dades. Es basa en la compressió de les dades MS a partir de la recerca de les anomenades regions d'interès (*Regions of interest*, ROIs) en les dades LC-MS (Gorrochategui et al. 2016; Stolt et al. 2006). Es consideren ROIs aquelles regions espectrals que contenen les dades m/z més

rellevants, amb una intensitat significativament més elevada que un valor llinar seleccionat a partir de la relació senyal/soroll de les dades (*Signal to noise ratio threshold*, SNR_{Thr}) i que, segons la exactitud instrumental de l'equip MS emprat, es trobin compresos dins d'un rang d'error dels valors de m/z (m/z error, μ). Per tal de definir correctament els pics cromatogràfics, les ROIs han de tenir un nombre mínim de valors consecutius en la direcció del temps (p_{min}). Amb aquests tres paràmetres s'aconsegueix eliminar els valors que només estan relacionats amb el soroll de fons o amb contribucions al senyal molt poc importants (Gorrochategui, et al. 2016). Aquest procediment es troba descrit en detall a les publicacions (Dalmau et al. 2018; Gorrochategui et al. 2015b; Gorrochategui et al. 2019).

Les ROIs es cerquen espectre per espectre per a cada temps de retenció al llarg del cromatograma i es combinen per obtenir un determinat nombre final de ROIs. Dins d'una mateixa ROI, en un mateix interval d'error, els valors de m/z es calculen a partir de les mitjanes de tots els valor m/z de la sèrie de dades per aquella ROI (veure **Figura 1.12**). Com que el nombre de ROIs pot variar entre les mostres, l'anàlisi simultània d'aquestes ha de considerar tots els valors obtinguts en les diferents mostres. Finalment, es genera una matriu de dades LC-MS per a cada mostra amb els espectres MS a les files (per cada temps de retenció cromatogràfic) i els cromatogrames a les columnes (per cada valor m/z de les ROIs seleccionats).

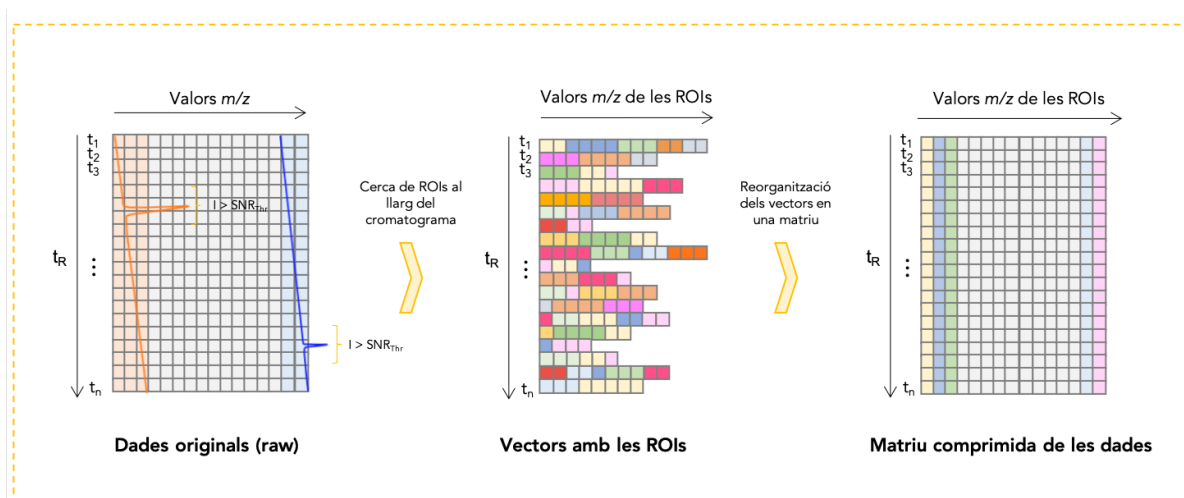


Figura 1.12 | Representació gràfica de la compressió ROI d'una matriu de dades.

L'estratègia d'anàlisi MCR-ALS quan es fa servir les ROIs dona lloc al procediment anomenat ROIMCR. En els estudis de metabolòmica s'analitzen diverses mostres simultàniament, per això és necessària la construcció d'una matriu augmentada que contingui la informació de les ROIs per a les matrius de les dades ROIs obtingudes per a cada mostra. Així, s'unifiquen tots els valors de les ROIs obtingudes i es consideren tant els valors de m/z ROIs comuns com els que no són comuns en totes les mostres. Primer, s'agrupen les ROIs amb una diferència de valors m/z inferior a l'error de massa tolerat. Llavors, aquelles ROIs que no tenen una intensitat significativa en una mostra determinada, amb valors baixos pròxims al nivell del soroll s'eliminen. La matriu de dades augmentada en la direcció de les columnes de les mostres comprimides mitjançant totes les ROIs, D_{aug} , es descompon utilitzant el model bilineal MCR descrit posteriorment en la **equació 1.5** en la **Secció 1.5.2**.

1.5.2 MÈTODES QUIMIOMÈTRICS DE TRACTAMENT DE DADES

Una característica de les dades de metabolòmica és que el nombre de mostres és significativament inferior al nombre de variables mesurades. Els mètodes quimiomètrics apliquen eines matemàtiques i estratègies analítiques multivariants per tal de maximitzar la informació present en les dades obtingudes. Les diferents estratègies d'anàlisi de dades emprades pretenen avaluar les diferències entre grups de mostres, ja sigui d'una manera univariant, és a dir, paràmetre per paràmetre, com per exemple a través d'un anàlisi t -test o de l'anàlisi de la variància (ANOVA) (**Figura 1.13A**), o mitjançant tècniques d'anàlisi multivariants on s'analitzen simultàniament les relacions entre múltiples metabòlits i mostres, com per exemple l'anàlisi de components principals, PCA, o l'anàlisi per mínims quadrats parcials, PLS (**Figura 1.13B**).

La limitació principal de l'aplicació dels mètodes d'anàlisi de dades univariants en metabolòmica recau en la dificultat d'investigar les correlacions entre els metabòlits (Hendriks et al. 2011). Un dels exemples d'aquest tipus d'enfocament univariant seria l'estudi de metabolòmica dirigida realitzat per Altmaier et al. (2008), on s'investiguen els canvis en els nivells induïts per medicaments dels metabòlits individuals mitjançant ANOVA two-way o bidireccional. En aquesta Tesi, s'han utilitzat el t -test (Kim 2015) i

l'ANOVA (Kim 2014) per avaluar les diferències entre dos o més grups de mostres i determinar si la mitjana d'una variable, en el nostre cas la concentració relativa d'un metabòlit, conté una diferència significativa entre grups diferents de mostres.

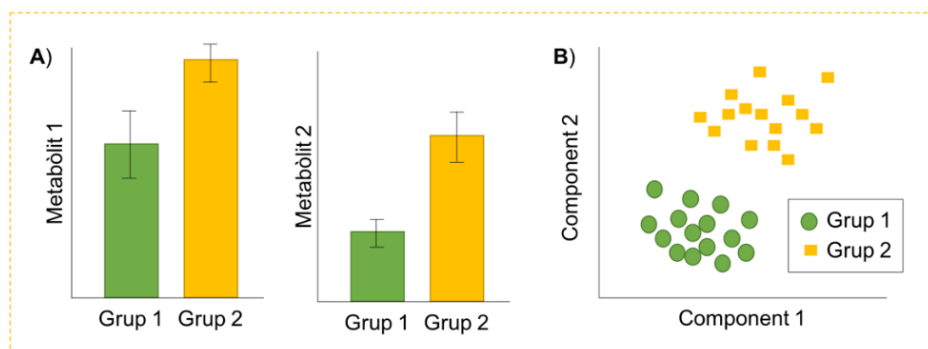


Figura 1.13| Diferents tipus d'enfocaments analítics en els estudis de metabolòmica. **A)** Anàlisi univariant: diferències dels nivells de concentració de diferents metabòlits entre dos grups de mostres. **B)** Anàlisi multivariant: diferències entre grups de mostres a partir de les correlacions entre múltiples metabòlits. (Imatge adaptada de Bartel et al. (2013)).

Les plataformes analítiques emprades en la majoria d'estudis de metabolòmica, especialment en estratègies no-dirigides, proporcionen respostes i dades multivariants. Per aquest motiu, els mètodes quimiomètrics esdevenen una eina essencial en el processament, l'exploració i la classificació del conjunt de dades generades per aquests estudis. Els mètodes quimiomètrics d'anàlisi i d'exploració multivariant de dades permeten la determinació de les variables més importants que causen la variació experimental observada i les seves relacions, i facilita la diferenciació entre els tipus de mostres estudiades (controls i exposades). Tal com es mostra en la **Figura 1.14**, en quimiometria, hi ha tres categories bàsiques d'anàlisi de dades (Trygg et al. 2006):

- a) Anàlisi multivariant exploratòria de dades: permet una visió global de les dades per tal de determinar tendències, patrons o grups (*clusters*) (**Figura 1.14A**).
- b) Anàlisi classificatòria (*cluster analysis*) i anàlisi discriminant: classifiquen les mostres per categories o classes (**Figura 1.14B**).
- c) Anàlisi de regressió i models de predicció: estableixen una relació quantitativa entre blocs de dades (**Figura 1.14C**).

En aquesta Tesi s'han aplicat diferents mètodes quimiomètrics exploratoris, classificatoris i discriminants en els estudis de metabolòmica de la *D. magna* fonamentalment a partir de dades GC-MS. En concret, s'ha utilitzat l'anàlisi de components principals (*Principal Component Analysis*, PCA), l'anàlisi de la variància - anàlisi simultània de components (*ANOVA-Simultaneous Component Analysis*, ASCA), l'anàlisi de resolució multivariant de corbes per mínims quadrats alternats (*Multivariate Curve Resolution-Alternating Least Squares*, MCR-ALS), l'anàlisi discriminant per mínims quadrats parcials (*Partial Least Squares-Discriminant Analysis*, PLS-DA) i l'anàlisi discriminant per mínims quadrats parcials ortogonals (*Orthogonal Partial Least Squares-Discriminant Analysis*, OPLS-DA), els quals es descriuen a continuació.

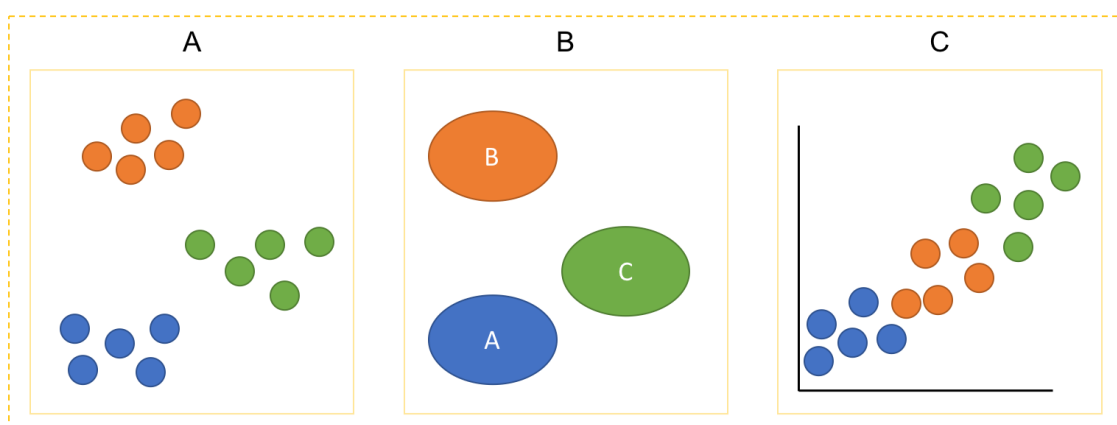


Figura 1.14 | Visió general de les categories bàsiques dels anàlisis de quimiometria: **A**) anàlisi exploratòria; **B**) anàlisi classificatòria i anàlisi discriminant; **C**) anàlisi de regressió (Trygg, et al. 2006).

Anàlisi de components principals: PCA

L'anàlisi de components principals (PCA) és una eina matemàtica àmpliament utilitzada per a la descomposició ortogonal no supervisada de dades (Esbensen and Geladi 2010; Wold et al. 1987). És un mètode multivariant d'anàlisi de dades dissenyat per extreure i visualitzar la variació sistemàtica de la informació continguda en un conjunt de dades. L'anàlisi PCA ajuda a entendre i a interpretar els factors de variació presents en grans conjunts de dades complexes, ja que redueix la dimensió del conjunt de dades i facilita la classificació dels grups de mostres basant-se en les variables observades.

Aquesta eina funciona condensant tota la informació en un nombre reduït de components, els anomenats components principals (PCs). En aquest mètode, a partir de la combinació lineal de les variables originals, es dedueix un nou conjunt de variables ortogonals no correlacionades (components principals) les quals expliquen per a cada una d'elles la màxima variància de les dades originals. Aquests nous components principals són ortogonals, així que la informació que contenen cadascun d'ells no està present en els altres components. A més, els components es troben ordenats segons la quantitat de variància explicada, de manera que el primer component explica més variància que el segon i així successivament. El mètode PCA dóna lloc a uns nous eixos de representació de les dades, on es descriu la màxima variància explicada. El nou conjunt d'eixos ortogonals de coordenades (*loadings*) sobre els que es poden projectar les dades originals (*scores*) proporciona una millor interpretació dels fenòmens que produeixen la variància observada. La matriu original de dades \mathbf{X} es descompon en el producte de dues matrius ortogonals \mathbf{T} i \mathbf{P}^T (vegeu **equació 1.1**):

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad \text{Equació 1.1}$$

On \mathbf{X} és la matriu de dades originals, \mathbf{T} representa a la matriu de *scores*, \mathbf{P}^T defineix els *loadings* i \mathbf{E} correspon a la matriu dels residuals. Els *scores* descriuen la informació sobre el tipus de mostra i els *loadings* corresponen al mapa de les variables. L'objectiu del PCA és maximitzar la variància explicada en les dades amb un nombre mínim de components. Les matrius de factors \mathbf{T} i \mathbf{P}^T representen un resum concís de les dades originals que poden ajudar a interpretar les fonts de variància de les dades. La visualització dels *scores* \mathbf{T} s'anomena gràfic dels *scores* o *scores plot*. Aquesta representació dóna una visió general de tot el conjunt de dades analitzades i de les relacions entre aquestes. A partir del *scores plot* es poden observar els agrupament de les mostres, les tendències i els valors anòmals (*outliers*). En el gràfic dels *loadings* \mathbf{P} o *loadings plot* es descriu la influència de les variables experimentals de la matriu \mathbf{X} sobre cada un dels components descrits pel model PCA.

Els resultats obtinguts mitjançant PCA permeten descriure la complexitat del sistema estudiat i donen la possibilitat d'interpretar les variacions descrites pels components a partir dels gràfics de *scores* i de *loadings*. Les direccions dels gràfics de

scores i dels gràfics dels *loadings* són les mateixes, d'aquesta manera es pot realitzar una interpretació conjunta dels dos gràfics i relacionar els tipus de mostres analitzades amb els metabòlits més rellevants (Esbensen and Geladi 2010). Tot i això, abans d'analitzar un conjunt de dades mitjançant el mètode PCA, si és necessari es poden aplicar els mètodes de preprocessament descrits en la secció anterior (**Secció 1.5.1**).

Anàlisi de variància – Anàlisi simultània de components: ASCA

L'anàlisi simultània de components d'ANOVA o ASCA és útil com a eina d'exploració per interpretar i avaluar la significació estadística dels canvis que tenen lloc en un conjunt de dades multivariants obtingudes a partir d'un disseny experimental (Jansen et al. 2005; Smilde et al. 2005). En aquesta Tesi, s'utilitza l'ASCA per a determinar les possibles relacions que existeixen entre els diferents factors estudiats i els seus efectes en el metaboloma de la *D. magna*. En concret, es va emprar el mètode ASCA per descriure les relacions dels factors de salinitat, temperatura i hipòxia, i avaluar si els efectes d'aquests factors eren significatius.

Les contribucions a la variació dels diferents factors a estudiar es descomponen en diferents matrius utilitzant els models lineals d'ANOVA. L'ASCA combina la capacitat de l'ANOVA per separar les fonts de variació amb els avantatges del PCA per explicar la màxima variància comuna entre les diferents variables. Així, en un primer pas s'utilitza l'ANOVA per descompondre la matriu original de dades en diverses matrius additives que caracteritzen cada un dels factors del disseny experimental, les quals contenen els efectes per separat dels factors escollits i una matriu interacció, a més de la variància natural i l'error experimental, que correspon a la variància no explicada per aquests factors. Les matrius dels factors resultants són analitzades pel mètode de l'anàlisi simultània de components (SCA) (Hoefsloot et al. 2010; Jansen, et al. 2005; Smilde, et al. 2005). Aquesta anàlisi s'aplica individualment a cada matriu dels factors experimentals. S'utilitza un model bilineal de descomposició de les dades que permet aproximar la informació continguda en la matriu original de dades d'una manera molt semblant al PCA. Cal tenir en compte, que l'ASCA s'ha d'aplicar en un disseny balancejat.

Per tal de provar estadísticament els resultats i d'avaluar els efectes dels factors estudiats i de les seves interaccions, s'aplica un test de permutació. Aquest test permet determinar la significació estadística dels efectes de tots els factors i de les seves interaccions. La hipòtesi nul·la (H_0) del test de permutació es basa en què el factor considerat no produeix cap efecte (Hoefsloot, et al. 2010). Així, el test s'aplica de manera que es permuten aleatòriament totes les files de les matrius de dades originals de tots els factors (normalment, 10000 permutacions). D'aquesta manera, s'obté una estimació de la variància aleatòria que després pot utilitzar-se per determinar si un factor concret és significatiu. La significació estadística s'avalua mitjançant el valor p (p -value), el qual es calcula dividint el nombre de casos en els que l'error de les matrius permutades és més gran a l'original, dividida pel nombre total de permutacions portades a terme. Es considera que el factor té un efecte significatiu quan el valor de p és menor que el nivell de significació prefixat (per exemple, $p < 0.05$) i es rebutja la H_0 (Hoefsloot, et al. 2010).

Resolució multivariant de corbes per mínims quadrats alternats: MCR-ALS

El mètode de resolució multivariant de corbes (*Multivariate Curve Resolution*, MCR) és un mètode quimiomètric que permet resoldre els constituents d'una mescla a partir de dades multivariants sense necessitat de la postulació prèvia de la seva composició. Permet resoldre problemes relacionats amb la identificació i la determinació analítica de compostos en presència d'interferències desconegudes en mostres complexes (De Juan et al. 2014; De Juan and Tauler 2006; Jaumot et al. 2005; Tauler 1995; Tauler and Barceló 1993; Tauler et al. 1993; Tauler et al. 1995).

Aquest mètode ha estat utilitzat en l'anàlisi de dades químiques de camps molt diversos, com per exemple en l'anàlisi de dades analítiques espectromètriques (Bedia et al. 2017), cromatogràfiques (Marín-García et al. 2018), ambientals (Malik et al. 2016) i d'imatges espectrals (Piqueras et al. 2018), i també més recentment en l'anàlisi de dades metabolòmiques (Navarro-Reig et al. 2017a) i lipidòmiques (Gorrochategui et al. 2015a) a partir de mètodes cromatogràfics amb detecció per MS. En aquesta Tesi aquest mètode s'ha aplicat especialment a conjunts de dades de GC-MS, GCxGC-MS i LC-MS. Quan el mètode MCR-ALS s'aplica a matrius de dades cromatogràfiques de

metabolòmica (**D**), es vol descriure la variació observada en les dades a partir d'un model bilineal amb un nombre reduït de components, definits en el cas d'aquestes tècniques analítiques com a perfils d'elució i espectres de masses. Així, tal com es mostra en l'**equació 1.2**, el mètode MCR es basa en un model de descomposició bilineal similar al del mètode PCA.

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad \text{Equació 1.2}$$

On **C** correspon la matriu de concentracions, la qual descriu els perfils d'elució dels N ($n=1, \dots, N$) *components* (metabòlits) presents en les mostres analitzades, **S^T** és la matriu d'espectres de masses associats als N components resolts (metabòlits) i **E** és la part de la variància que no explica el model, és a dir, l'error de les dades (veure **Figura 1.15**). El nombre de files en les matrius **D** i **C** coincidirà, així com el nombre de columnes de **D** i **S^T**, que serà el mateix. El nombre de components escollit en el model, que corresponen al nombre de columnes en **C** i el nombre de files en **S^T**, és el mínim necessari per tal de descriure la variància en la matriu de dades **D**.

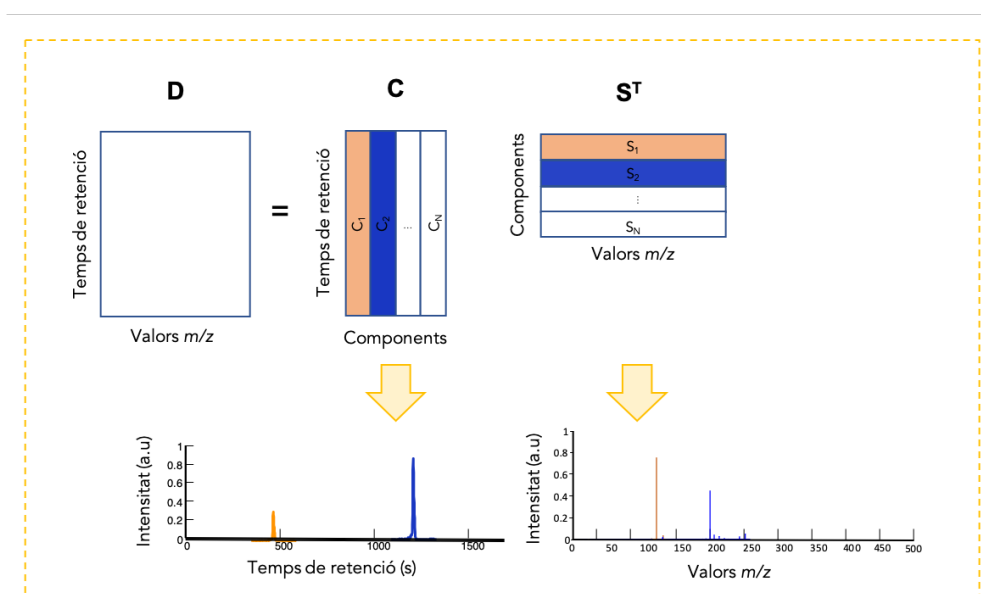


Figura 1.15] Representació del model bilineal de MCR-ALS per a una matriu de dades metabolòmiques.

En el mètode MCR-ALS, l'**equació 1.2** es resol iterativament a través d'un algoritme de mínims quadrats alternats (*Alternating Least Squares, ALS*), el qual calcula

les concentracions (\mathbf{C}) i els espectres purs (\mathbf{S}^T) dels components de les mostres analitzades minimitzant les diferències entre els valors de les dades experimentals i els valors calculats segons el model bilineal. Aquesta optimització es realitza per un nombre proposat de components estimats inicialment, per exemple a partir d'una descomposició de valors singulars, SVD (Golub and Reinsch 1970) i unes estimacions inicials de \mathbf{C} o \mathbf{S}^T . Aquestes estimacions inicials de \mathbf{C} i \mathbf{S}^T es poden obtenir per exemple a partir de les columnes o de les files de la matriu original de dades, \mathbf{D} , que siguin més diferents entre si i que no continguin només soroll experimental. Una manera de realitzar aquesta estimació és a través del mètode SIMPLISMA (*SIMPLE-to-use Iterative Self-Modeling Analysis*) (Windig 2005; Windig and Guilment 1991).

Una vegada ja s'han realitzat aquestes estimacions inicials del nombre de components i de \mathbf{C} o \mathbf{S}^T s'inicia un procediment d'optimització iteratiu per mínims quadrats alternats (*Alternating, Least Squares, ALS*) (De Juan, et al. 2014). Durant aquesta optimització per ALS es poden aplicar diverses restriccions (*constraints*) per modelar les formes dels perfils \mathbf{C} i \mathbf{S}^T , i donar sentit físic i interpretable a les solucions obtingudes del model bilineal, el més semblants possibles en els perfils d'elució i espectrals dels constituents analitzats. La convergència s'aconsegueix quan en dos cicles iteratius consecutius, les diferències relatives en les desviacions estàndards dels residus entre els valors de dades calculats experimentals i l'ALS són inferiors a un valor prèviament seleccionat, generalment el valor triat serà el 0.1%.

En el context de les dades cromatogràfiques de metabolòmica en l'optimització per ALS, normalment s'aplicaran les restriccions de no-negativitat (en les matrius \mathbf{C} i \mathbf{S}^T) i de normalització dels espectres de masses (\mathbf{S}^T) (De Juan, et al. 2014; Tauler 1995). La restricció de no-negativitat evita la presència de valors negatius tant en els perfils de concentració (\mathbf{C}) com en els perfils dels espectres de masses (\mathbf{S}^T), ja que en les dades de cromatografia, tant les concentracions com els espectres de masses dels compostos químics han de ser valors positius o zeros. Existeixen diferents maneres d'aplicar aquesta restricció, mitjançant la substitució directa dels valors negatius per zero durant les iteracions, per penalització de la funció a optimitzar o, per altra banda, amb l'ús d'algoritmes rigorosos, de mínims quadrats no negatius (*non negative least squares*,

npls) (Lawson and Hanson 1995) o de mínims quadrats ràpids no negatius (*fast non-negative least squares, fnpls*) (Bro and De Jong 1997). La restricció de normalització s'aplica per eliminar l'ambigüitat d'escala en els perfils resolts i per facilitar les iteracions durant l'optimització de l'algorisme ALS. Per aplicar aquesta restricció als perfils dels espectres (\mathbf{S}^T) es divideix tots els valors de l'espectre d'un component determinat per la seva àrea o alçada. Així, es forcen els perfils de tots els components a tenir la mateixa àrea o alçada igual a 1.

Per a les dades de GCxGC-MS, també es pot aplicar la restricció de la trilinealitat (MCR-ALS trilineal) en els perfils d'elució, a més de les de no-negativitat i normalització dels espectres de masses. En aquest model, els perfils d'elució i els seus temps de retenció d'un determinant component han d'ésser els mateixos en les successives modulacions, i només poden diferir en la seva proporció relativa (àrees i intensitats dels pics). Aquesta restricció només es pot aplicar a dades que compleixin aquesta condició d'invariabilitat dels perfils d'elució ja que si no es compleix l'ajust que tindrà el model a les dades estudiades serà pitjor i les solucions obtingudes no seran correctes.

Per tal d'avaluar la qualitat del model MCR-ALS s'utilitzen dos paràmetres: el percentatge de falta d'ajust (*lack of fit, % lof*) i el percentatge de la variància explicada (% R^2), descrits en les **equacions 1.3** i **1.4** respectivament.

$$\% \text{ lof} = 100 \times \sqrt{\frac{\sum_{i,j} e_{i,j}^2}{\sum_{i,j} d_{i,j}^2}} \quad e_{i,j} = d_{i,j} - \hat{d}_{i,j} \quad \text{Equació 1.3}$$

$$\% R^2 = \left(\frac{\sum_{i,j} d_{i,j}^2 - \sum_{i,j} e_{i,j}^2}{\sum_{i,j} d_{i,j}^2} \right) \times 100 \quad \text{Equació 1.4}$$

On $d_{i,j}$ és el valor de la matriu de dades original de la mostra i i la columna j , $e_{i,j}$ correspon al residual relacionat, que s'obté de la diferència entre el valor $d_{i,j}$ de les dades originals i el valor $\hat{d}_{i,j}$ obtingut a través del model MCR-ALS.

El valor d'aquests paràmetres de qualitat està relacionat amb la qualitat de les dades originals i amb la relació senyal soroll de les mesures. L'objectiu del model és que aquest expliqui el màxim de variància possible (R^2), tenint en compte l'error experimental. Quan s'apliquen diferents models amb un nombre creixent de

components, s'observa un augment de la variància explicada (R^2 més elevat) i un millor ajust del model (*lof* més petit). Pel que fa a la variància explicada, si el valor de R^2 millora respecte del model anterior, significa que el nou model descriu d'una millor manera el sistema. En canvi, si el valor de R^2 s'estabilitza o creix molt poc indica que l'addició de components al model no és necessària perquè ja no millora el model de manera significativa. En relació a l'ajust del model, si el valor de *lof* disminueix significativament, indica que el model proposat incorpora informació per descriure el sistema d'una millor manera. Per contra, si el % de *lof* pràcticament no canvia significativament, indica que no és necessària l'addició de nous components al model. Tot i l'avaluació d'aquests paràmetres, en la selecció del model final també és necessari avaluar la versemblança dels perfils \mathbf{C} i \mathbf{S}^T obtinguts. Quan l'addició de nous components comporta la recuperació de perfils d'elució o espectres de masses inversemblants, l'addició d'aquests components addicionals no es recomanable.

MCR-ALS aplicat a diverses mostres

Normalment, en els estudis de metabolòmica s'investiguen varies mostres del mateix sistema biològic alhora. Per tal de poder distingir els diferents perfils metabòlics en les diferents mostres és interessant analitzar totes les mostres simultàniament. D'aquesta manera, es poden veure millor els canvis que es produeixen en els perfils metabòlics de les diverses mostres estudiades. La informació que proporciona l'anàlisi d'una matriu de dades d'una única mostra sovint no és generalment suficient per a poder resoldre un sistema complex concret. Per això és necessari l'anàlisi simultània de diverses matrius de dades de mostres diferents del mateix sistema, preferiblement obtingudes sota les condicions d'un disseny experimental estadístic.

Per tal d'analitzar totes les mostres simultàniament, aquestes s'han d'organitzar d'una determinada manera. En aquesta Tesi, s'han organitzat de manera que generen una nova matriu de dades augmentada en la direcció de les columnes (*column-wise*) (\mathbf{D}_{aug}). Tal com s'ha descrit anteriorment en la **Figura 1.6C**, les diferents matrius individuals de dades s'agrupen fent la seva augmentació en la direcció dels valors m/z analitzats (en referència a les columnes de les matrius de les mostres). Aquesta manera

d'organitzar les matrius de dades és útil quan les diferents mostres s'han analitzat utilitzant la mateixa tècnica analítica, per exemple GC-MS o LC-MS.

Per altra banda, quan la tècnica utilitzada és la GCxGC-MS per a cada una de les mostres analitzades, les dades s'organitzen en una matriu ja augmentada per columnes que inclou les diferents modulacions a la segona columna, \mathbf{X}_{aug} , tal com s'ha comentat anteriorment en la **Figura 1.7**. Les dades GCxGC-MS es poden organitzar en forma de matrius superaugmentades ($\mathbf{X}_{\text{superaug}}$) les quals es generen a partir de les matrius augmentades individuals corresponents a cada una de les mostres simultàniament analitzades. En aquest cas, també les diferents matrius de dades es troben apilades en la direcció de les columnes (valors de m/z analitzats). El model de MCR-ALS permet l'anàlisi conjunta de diverses matrius de dades simultàniament. En l'**equació 1.5**, es mostra el model MCR per a dades GC-MS o LC-MS. La matriu augmentada \mathbf{D}_{aug} està formada per la concatenació de diverses matrius experimentals que corresponen a totes les mostres (K). Aquesta \mathbf{D}_{aug} es descompon en el producte de la matriu augmentada dels perfils d'elució (\mathbf{C}_{aug}) i la dels espectres (\mathbf{S}^T), deixant una matriu residual, \mathbf{E}_{aug} , que conté la variància no explicada del model MCR.

El mètode MCR-ALS també permet l'anàlisi simultània de mostres analitzades amb GCxGC-MS, tal com s'expressa en l'**equació 1.6**. Aquí, la matriu superaugmentada $\mathbf{X}_{\text{superaug}}$ està formada per totes les modulacions (M) de totes les mostres (K) analitzades. A partir de la descomposició bilineal de la $\mathbf{X}_{\text{superaug}}$, s'obté la matriu superaugmentada dels perfils d'elució en les dues dimensions cromatogràfiques ($\mathbf{C}_{\text{superaug}}$) i la matriu els espectres de masses purs dels components resolts (\mathbf{S}^T).

En els models de MCR descrits en les **equacions 1.5 i 1.6**, els perfils d'elució d'un mateix component resolt en les diferents matrius \mathbf{C}_k i $\mathbf{C}_{1,M}$ poden tenir els temps de retenció desplaçats i tenir canviada la seva forma. Això és un inconvenient en l'aplicació de mètodes d'anàlisi de les dades multidireccionals amb models multilineals (Bro 1997) però no ho és en l'aplicació del MCR-ALS bilineal, tal com s'ha emprat en aquesta Tesi. En aquesta darrera aproximació, no existeixen problemes en relació a la dificultat d'alineació ni canvis de forma dels pics cromatogràfics entre les diferents mostres provinents de diferents anàlisis cromatogràfiques, tant en les dades GC-MS o

LC-MS (\mathbf{D}_k) com en les dades GCxGC-MS ($\mathbf{X}_{k,M}$). Aquest fet és un gran avantatge de la modelització bilineal del mètode MCR-ALS, ja que permet evitar l'aplicació de procediments d'alineament complexos i de modelització dels pics cromatogràfics. Per contra, els espectres de masses dels components resolts comuns a les diferents mostres que es troben en la matriu espectral \mathbf{S}^T han d'ésser els mateixos per a totes les matrius originals individuals. Això, significa que les condicions experimentals d'anàlisi de les mostres \mathbf{D}_k , control i tractades, comparteixen compostos amb el mateix espectre MS.

$$\mathbf{D}_{\text{aug}} = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \mathbf{D}_3 \\ \vdots \\ \mathbf{D}_K \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \mathbf{C}_3 \\ \vdots \\ \mathbf{C}_K \end{bmatrix} \mathbf{S}^T + \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \\ \mathbf{E}_3 \\ \vdots \\ \mathbf{E}_K \end{bmatrix} = \mathbf{C}_{\text{aug}} \mathbf{S}^T + \mathbf{E}_{\text{aug}} \quad \text{Equació 1.5}$$

$$\mathbf{X}_{\text{superaug}} = \begin{bmatrix} \mathbf{X}_{1,1} \\ \mathbf{X}_{1,2} \\ \mathbf{X}_{1,3} \\ \vdots \\ \mathbf{X}_{K,M} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{1,1} \\ \mathbf{C}_{1,2} \\ \mathbf{C}_{1,3} \\ \vdots \\ \mathbf{C}_{K,M} \end{bmatrix} \mathbf{S}^T + \begin{bmatrix} \mathbf{E}_{1,1} \\ \mathbf{E}_{1,2} \\ \mathbf{E}_{1,3} \\ \vdots \\ \mathbf{E}_{K,M} \end{bmatrix} = \mathbf{C}_{\text{superaug}} \mathbf{S}^T + \mathbf{E}_{\text{superaug}} \quad \text{Equació 1.6}$$

En les dades de metabolòmica analitzades mitjançant GC-MS o LC-MS, cada matriu \mathbf{D}_k conté l'anàlisi d'una mostra. En la **Figura 1.16**, les diferents matrius de dades (\mathbf{D}_k) corresponen a diferents mostres en diferents condicions experimentals d'un mateix estudi metabolòmic, per exemple les mostres control i les diferents mostres sotmeses a un tractament determinat o a unes condicions ambientals d'estrès. La matriu augmentada \mathbf{C}_{aug} contindrà la informació dels perfils d'elució i de concentració resolts en les diferents mostres analitzades (mostres control i de les mostres exposades). Els perfils de concentració (\mathbf{C}_k) de la seva matriu augmentada \mathbf{C}_{aug} apareixen seguint el mateix ordre que en les dades originals individuals (\mathbf{D}_k), abans d'organitzar-se per generar la matriu augmentada \mathbf{D}_{aug} .

Tal com s'observa en la **Figura 1.17**, per a les dades analitzades per GCxGC-MS, cada matriu individual ($\mathbf{D}_{1,M}$) conté la informació d'una sola modulació en la segona dimensió cromatogràfica. Així, cada mostra està representada per una matriu augmentada, $\mathbf{D}_{1,\text{aug}}$, formada per un determinat numero de matrius corresponent al nombre de modulacions (M), $\mathbf{D}_{1,M}$. És a dir, aquesta matriu augmentada d'una mostra

contindrà tantes matrius com el nombre de modulacions que es generin en la segona dimensió cromatogràfica, tal com es descriu en la **Figura 1.17**.

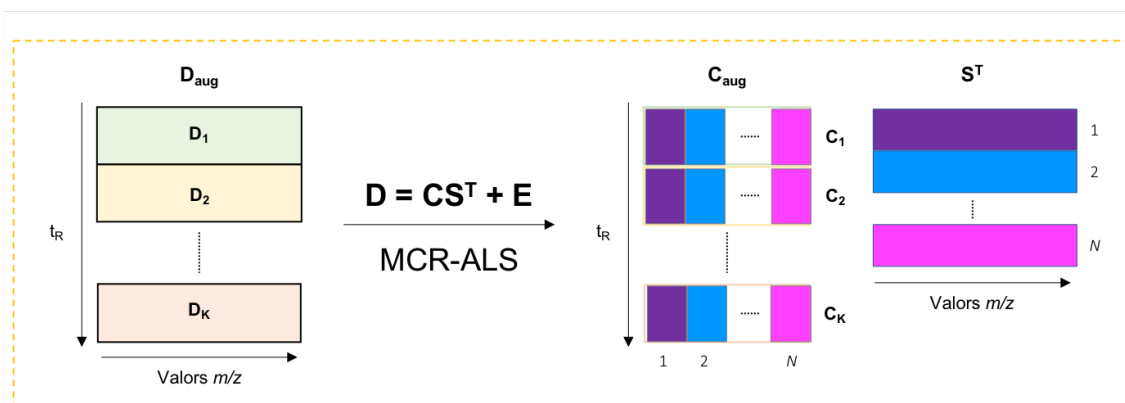


Figura 1.16 Representació gràfica de la resolució pel mètode MCR-ALS per una matriu augmentada de dades, D_{aug} , de dades de GC-MS o LC-MS que conté K mostres.

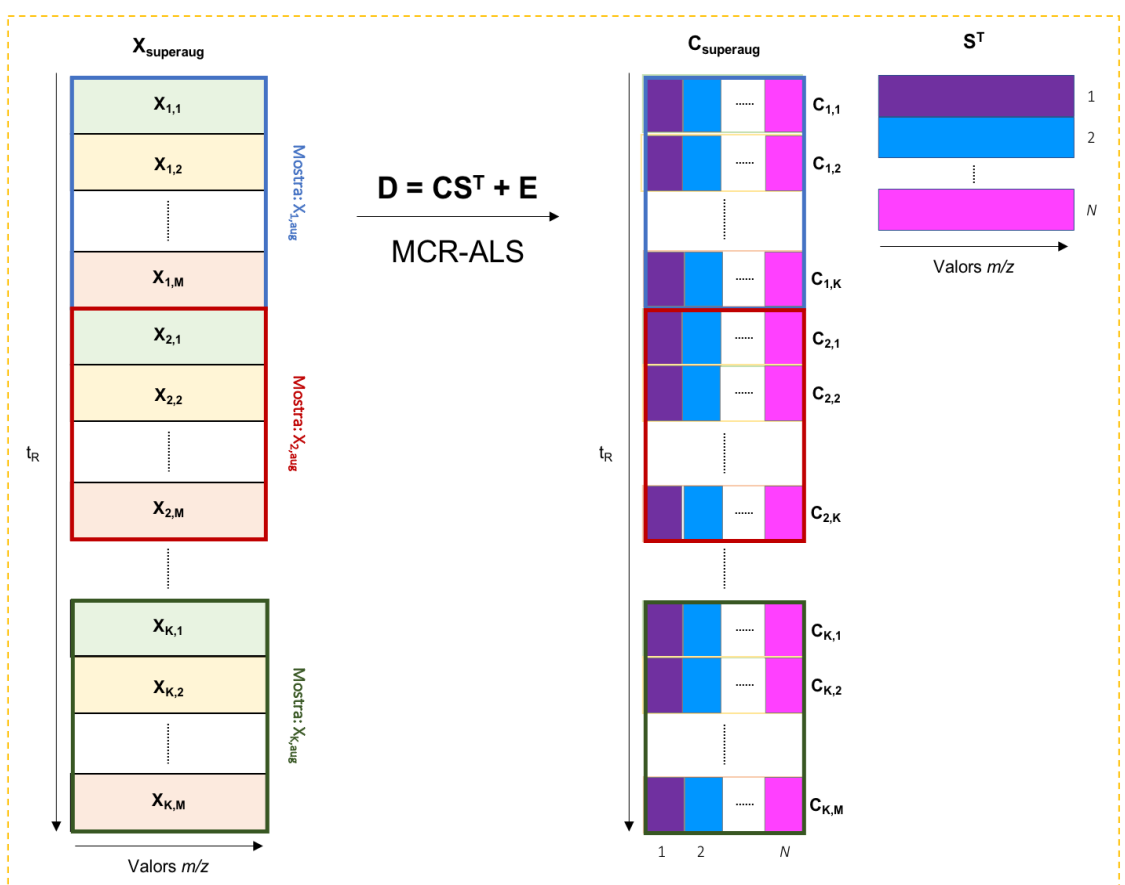


Figura 1.17 Representació gràfica de la resolució MCR-ALS d'una matriu de dades GCxGC-MS superaugmentada, $X_{superaug}$, la qual conté K mostres en matrius augmentades de M modulacions ($X_{K,M}$).

Les diferents matrius D_K es refereixen a l'anàlisi cromatogràfica en una dimensió, per GC-MS o LC-MS, o en dues dimensions per GCxGC-MS, de diferents mostres d'un mateix sistema metabòlic sota diferents condicions experimentals (mostres controls i mostres exposades). Les matrius C_{aug} o $C_{superaug}$ contenen la informació dels perfils d'elució dels diferents components resolts en les diferents mostres analitzades. D'aquesta manera, a partir de les àrees dels perfils d'elució dels diferents components de les mostres es pot fer una quantificació relativa dels constituents (metabòlits) associats als components resolts obtinguts. És possible obtenir aleshores la informació sobre els canvis que presenten aquests components quan ells es comparen en les mostres controls i en les mostres exposades. D'altra banda, els espectres de masses dels components resolts que conté la matriu S^T , permetrà obtenir informació qualitativa i identificar els metabòlits presents en les mostres analitzades.

Anàlisi discriminant per mínims quadrats parcials: PLS-DA

L'anàlisi discriminant per mínims quadrats parcials (*Partial Least Squares – Discriminant Analysis*, PLS-DA) s'aplica a problemes de regressió, on la resposta (y) està formada per variables categòriques que representen grups (Barker and Rayens 2003; Kalivodová et al. 2015). Es tracta d'un enfocament d'anàlisi supervisada de les dades, que s'ha utilitzat àmpliament en la classificació de dades metabolòmiques. El PLS-DA és una extensió de la regressió de mínims quadrats parcials (PLS) definida per un vector variable y de variables binàries que descriuen la pertinença a unes determinades categories de les mostres de les quals es té la seva resposta analítica en la matriu X . Més concretament, la matriu X conté la informació relativa a les concentracions relatives dels metabòlits en cada una de les mostres analitzades, alhora que el vector y representa la informació referent a les categories o classes de les mostres analitzades. En el model PLS-DA es realitza una estimació eficient de les variables latents (*latent variables*, LVs) de la matriu de dades X , que es troben correlacionades de manera òptima amb els canvis observats en el vector de classes y . En aquest model es realitza una regressió inversa, en la qual el model calculat relaciona les LVs de X amb y (variables a predir) segons el vector regressió b d'acord amb l'**equació 1.7**:

$$\mathbf{y} = \mathbf{bX}$$

|Equació 1.7

On \mathbf{y} és el vector resposta de les variables dependents, \mathbf{b} és el vector que conté els coeficients de regressió calculats durant la construcció del model de calibratge i \mathbf{X} és la matriu que conté la informació de les variables predictores. D'aquesta manera, es construeix un model que maximitza la covariància entre la \mathbf{X} i la \mathbf{y} amb el mínim nombre de LVs ortogonals. Aquest nombre de LVs és generalment molt inferior al nombre de variables originals. Per a cada LV s'obté un vector de coeficients de pes (*weights*, \mathbf{w}) que determina quines variables de la matriu \mathbf{X} combinen millor per formar la matriu de scores, \mathbf{T} , en la predicció de les classes en el vector \mathbf{y} (Kalivodová, et al. 2015).

Per tal de determinar el nombre adequat de LVs per predir i explicar suficientment bé la variable \mathbf{y} , en aquesta Tesi s'ha utilitzat el mètode de la validació creuada (*cross validation*, CV) pel calibratge intern del model PLS. Aquest mètode es basa en dividir el conjunt de dades originals en un nombre reduït de subgrups, sobre els quals s'aplica el model de PLS en tots els subgrups excepte en un. Aquest grup exclòs s'usa per validar l'anàlisi realitzat. Així, el model de predicció es genera amb les dades on s'aplica el model PLS. En aquest model predictiu es comparen les dades originals o mesurades amb les dades predites d'una manera iterativa. A partir de la comparació entre els valors predits i els valors mesurats de les dades excloses es calcula l'error de CV (RMSECV), tal i com es descriu en l'**equació 1.8**:

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

|Equació 1.8

On \hat{y}_i és el valor predit i y_i és el valor mesurat. En aquesta Tesi, s'ha utilitzat la validació creuada (CV) de manera aleatòria i el mètode *leave-one-out cross validation* (LOOCV). En la CV aleatòria, la separació entre el conjunt de dades on s'aplica el model PLS, i el subgrup exclòs es realitza de manera aleatòria en cada iteració, i en grups de mida variable. En canvi, en el LOOCV per a cada iteració, una mostra és el subgrup exclòs, i la resta constitueixen el subgrup que genera el model predictiu. Segons el nombre de mostres del conjunt de dades a analitzar, s'utilitza un mètode o un altre. Per

exemple, quan el conjunt de dades conté menys de 20 mostres, el mètode recomanat és el LOOCV.

Per tal de validar el model PLS-DA es calcula el coeficient de correlació de Matthews (*Matthew's correlation coefficient*, MCC) (Matthews 1975). El MCC és un coeficient de correlació que mesura els encerts entre les classificacions binàries observades i previstes, tal com es descriu en l'**equació 1.9**, on els valors retornats estan entre +1 i -1. El coeficient de +1 indica una predicció perfecta, un valor pròxim a 0 representa una predicció semblant a el què seria una predicció aleatòria, i un valors de -1 indica un desacord total entre la predicció i la observació.

$$MCC = \frac{TP \times TN - FR \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad \text{Equació 1.9}$$

On TP correspon al nombre de vertaders positius (*true positive*), TN el nombre de vertaders negatius (*true negative*), FP el nombre de falsos positius (*false positive*) i FN el nombre de falsos negatius (*false negative*), on es fa referència a una classificació correcte o incorrecte com a membres del grup estudiat.

Anàlisi discriminant per mínims quadrats parcials ortogonals: OPLS-DA

L'anàlisi discriminant per mínims quadrats parcials ortogonals (OPLS-DA) és una extensió del mètode de regressió PLS amb la integració d'un filtre preliminar de correcció ortogonal del senyal (*Orthogonal signal correction*, OSC) que va ser desenvolupat per Wold et al. (1998) (Trygg and Wold 2002). Aquest model es va introduir com una millora del mètode PLS-DA descrit anteriorment, el qual permet discriminar dos o més grups de mostres a partir de dades multivariants (Westerhuis et al. 2010). A diferència del PLS-DA, l'OPLS-DA utilitza la informació de la matriu resposta **Y** per descompondre la matriu **X** en dos blocs de variació estructurada, un amb la variació correlacionada i l'altre bloc amb la informació ortogonal a **Y**, respectivament (Bylesjö et al. 2006). Així, aquest model descompon la matriu **X** en tres parts diferents tal com es descriu en l'**equació 1.10**, on **T_p** correspon a la matriu scores predictiva per a **X**, **P_p** correspon la matriu *loadings* predictiva per a **X**, **T_o** correspon a la matriu scores

ortogonal de \mathbf{Y} , \mathbf{P}_o correspon la matriu *loadings* dels components ortogonals de \mathbf{Y} i \mathbf{E} representa la matriu dels residuals de \mathbf{X} (Trygg and Wold 2002).

$$\mathbf{X} = \mathbf{T}_p \mathbf{P}_p^T + \mathbf{T}_o \mathbf{P}_o^T + \mathbf{E} \quad \text{Equació 1.10}$$

La classificació de mostres amb el mètode OPLS-DA es realitza en dos passos. En primer lloc, la variació ortogonal de \mathbf{Y} s'elimina de la matriu de dades \mathbf{X} , tal com es mostra a l'**equació 1.11**. En segon lloc, la matriu predictiva, anomenada \mathbf{Y}_{hat} es calcula utilitzant la \mathbf{X}_p actualitzada i els components predictius del model OPLS-DA, que es calculen a partir del conjunt de dades on s'aplica el model PLS-DA. L'avantatge de OPLS-DA en comparació al PLS-DA recau en que només cal un sol component per a predir les classes, mentre que els altres components descriuen la variació ortogonal al primer component predictiu, en cas de tenir dues classes (Westerhuis, et al. 2010). Aquest únic component predictiu del model OPLS-DA en realitat descriu l'efecte del tractament estudiat en el model en comparació a la mitjana de les dues classes estudiades.

$$\mathbf{X}_p = \mathbf{X} - \mathbf{T}_o \mathbf{P}_o^T \quad \text{Equació 1.11}$$

Selecció de variables

Els models PLS-DA i OPLS-DA són capaços d'obtenir informació sobre quines són les variables, metabòlits, de la matriu \mathbf{X} més rellevants per a una millor discriminació entre les mostres. La selecció de variables permet la reducció de la complexitat de les dades i la millora de la seva interpretació biològica, ja que es selecciona un conjunt, el més petit possible, de variables que estiguin correlacionades amb la variables resposta. D'aquesta manera, es poden determinar quins són els metabòlits que presenten diferències significatives entre les diferents categories de mostres (normalment, mostres control i mostres exposades), especialment útil en estudis de metabolòmica no dirigida.

És possible obtenir aquesta informació a partir de les variables que descriuen el pes dels *loadings*, de les variables importants en projecció (*variable importance in projection*, VIP) i dels quocients de selectivitat (*selectivity ratio*, SR). En aquesta Tesi,

s'ha realitzat aquesta selecció de variables mitjançant la selecció de les variables importants en la projecció (VIPs) i dels quocients de selectivitat (SR).

El mètode de selecció de VIPs dóna una mesura de la influència individual de cada una de les variables de \mathbf{X} en la construcció del model PLS-DA i OPLS-DA (Wold et al. 1993). Els VIPs es calculen a partir de la suma ponderada dels quadrats dels coeficients de pes (*weights*, \mathbf{w}), els quals tenen en compte la quantitat de variància de cada variable latent obtinguda. Els VIPs donen una mesura útil per seleccionar quines són les variables que més contribueixen a l'explicació de la variància del vector \mathbf{y} . Per cada model de PLS-DA i OPLS-DA d'un conjunt de dades concretes sempre hi haurà un vector de VIPs que contindrà la informació de la importància de les variables \mathbf{X} en la predicció de les variables \mathbf{y} . Tenint en compte que la mitjana dels valors dels VIP al quadrat ha d'ésser igual a 1, normalment s'escollirà aquest valor o de més gran de 1 com a criteri per a la selecció de les variables més importants (Chong and Jun 2005). Aquesta selecció del valor "més gran que la unitat" es recomana de manera inicial i orientativa. En la **Figura 1.18** es mostra un exemple del gràfic d'scores VIPs on els metabòlits amb els valors superiors a la unitat serien escollits com a metabòlits rellevants per a la discriminació entre classes i avaluats com a potencials biomarcadors.

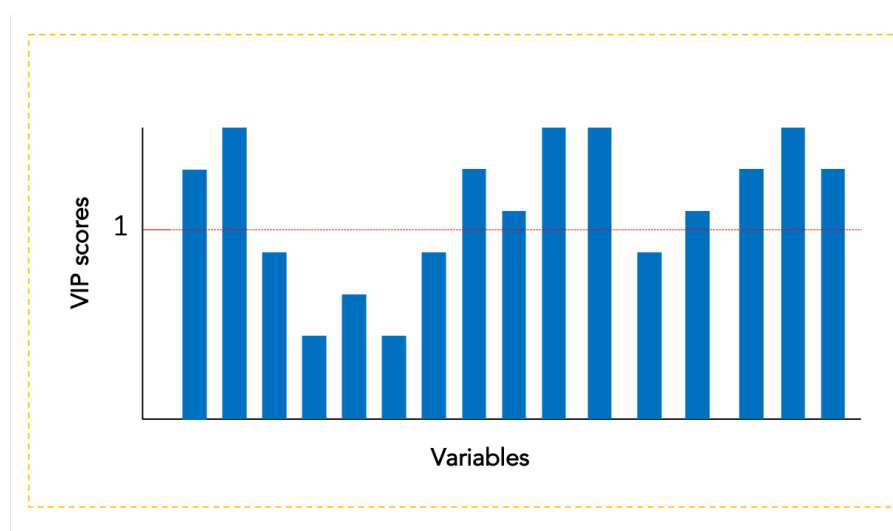


Figura 1.18 | Exemple de gràfic d'scores VIPs de l'anàlisi OPLS-DA.

Un altre mètode per a la selecció de variables emprat en aquesta Tesi, és el quocient de selectivitat (*Selectivity Ratio*, SR) (Rajalahti et al. 2009a; Rajalahti et al.

2009b). Aquest paràmetre és una eina de visualització de les variables X que contribueixen millor a l'explicació de la variància de la variable resposta y . Kvalheim and Karstang (1989) van desenvolupar un procediment anomenat vector projectat (*target projection*, TP) el qual permet una simplificació en la interpretació de les variables latents en els models de regressió. Aquest procediment genera un únic component predictiu mitjançant la projecció de la descomposició de les variables latents sobre la variable resposta (Kvalheim and Karstang 1989). El quocient SR de cada variable es defineix com la relació entre la variància explicada i la seva variància residual (no explicada) en el vector projectat.

1.6 DESCOBRIMENT DELS MARCADORS BIOLÒGICS

Els mètodes estadístics i les eines quimomètriques són útils per a comparar les concentracions dels metabòlits dels diferents grups de mostres, les control i les exposades. Aquests canvis de concentració permeten seleccionar aquells metabòlits més rellevants (estadísticament significatius), explicar les seves variacions en els diferents grups de mostres i deduir-ne una explicació biològica. Un biomarcador és una mesura interna de les alteracions a nivell molecular o cel·lular que poden aparèixer durant o després un procés d'estrès les quals indiquen l'estat biològic d'un organisme (Boudah et al. 2013).

Les ciències òmiques han permès obrir un nou ventall de possibilitats en el descobriment d'aquests biomarcadors (Bennett and Waters 2000). Concretament, a través de la informació que proporciona la MS, la qual permet determinar les assignacions i identificacions dels metabòlits, mitjançant la combinació de la massa exacta, de la seva distribució isotòpica i dels possibles patrons de fragmentació, a més de la informació cromatogràfica disponible (Patti, et al. 2012).

En relació als estudis de metabolòmica, ja a l'any 2007 es va arribar a un consens promogut pel grup de treball d'anàlisi química (*Chemical Analysis Working Group*, CAWG) el qual va proposar la iniciativa de normalització dels estudis de metabolòmica (*Metabolomics Standards Initiative*). Aquestes normes estan publicades en el treball de

Sumner et al. (2007), on es distingeixen en 4 nivells diferents el rigor de la identificació de tots els metabòlits detectats:

- i) Compostos identificats: es necessiten dues dades independents per validar la identificació dels metabòlits, per exemple, el temps de retenció o l'índex de retenció i l'espectre de masses.
- ii) Compostos amb anotació qualitativa: es consideren identificacions *putatives* o qualitatives on els autors documenten la informació espectral per validar les identificacions. Es tracta de valors reportats per altres estudis o de dades de les biblioteques disponibles *online*. Tot i això, no tots els metabòlits detectats en les mostres biològiques poden ser identificats mitjançant les bases de dades disponibles.
- iii) Classes de compostos qualitius caracteritzats: la manca de disponibilitat d'estàndards de metabòlits comercials impedeixen una identificació definitiva del metabòlit.
- iv) Compostos desconeguts: la identificació dels quals no és possible per manca d'informació espectral per tal de portar a terme una identificació inequívoca d'aquells metabòlits.

Una vegada els metabòlits s'han identificat de manera correcta, el pas següent recau en la exploració biològica d'aquests metabòlits i de les rutes metabòliques involucrades les quals permeten descriure la seva funció biològica. En aquest punt, la utilització de les bases de dades disponibles és fonamental per entendre la relació entre els metabòlits i les rutes metabòliques a les quals estan relacionats. Entre altres exemples, la base de dades *Kyoto Encyclopedia of Genes and Genomes* (KEGG) (Kanehisa et al. 2016) és de les més comunes i més àmpliament utilitzades per realitzar aquesta interpretació biològica. Aquesta base de dades és un recurs bioinformàtic que permet entendre les diferents funcions dels sistemes biològics a nivell molecular. Dins d'aquesta base de dades existeixen varies aplicacions, entre elles en la secció de rutes metabòliques (*KEGG pathway*) l'eina *KEGG mapper* (http://www.genome.jp/kegg/tool/map_pathway1.html) permet situar els metabòlits sobre els mapes de rutes metabòliques de l'organisme estudiat. D'aquesta manera, es

pot aprofundir en els processos bioquímics associats als possibles efectes sobre les mostres estudiades. En la **Figura 1.19** es mostra un exemple de la representació gràfica de les rutes metabòliques que ofereix la base de dades KEGG, on es poden visualitzar les rutes metabòliques i on cada punt correspon als metabòlits involucrats. Tot i això, no tots els metabòlits identificats tenen un codi KEGG. Així, aquells metabòlits que no el tinguin no es podran representar en les xarxes o mapes que proporciona aquesta base de dades.

En els últims anys, la plataforma MetaboAnalyst (Xia et al. 2015) ha incorporat dotze mòduls organitzats en 4 grups per analitzar diversos conjunts de dades i analitzar-les amb diferents finalitats: (i) anàlisi estadística exploratòria, (ii) anàlisi de l'enriquiment de metabòlits, (iii) integració de les dades i dels sistemes biològics i (iv) processament de les dades. Aquesta eina permet avaluar d'una manera *online* els diferents metabòlits, la qual proporciona una interpretació exhaustiva de les dades metabòliques. Entre altres opcions, inclou una anàlisi estadística de les dades, una anàlisi de les rutes metabòliques en diferents organismes model i una anàlisi dels biomarcadors. Es tracta d'una eina molt completa per a la interpretació de les dades de metabolòmica obtingudes mitjançant qualsevol de les tècniques emprades GC-MS, LC-MS i RMN.

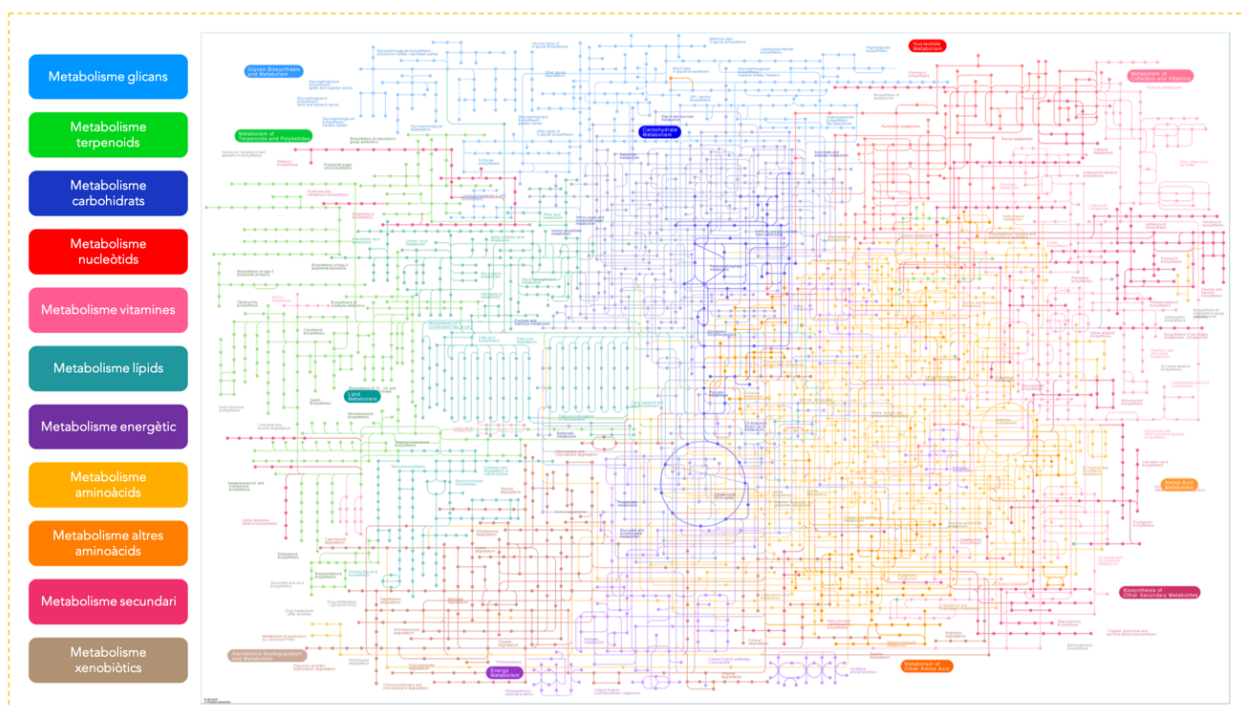


Figura 1.19 Mapa cartogràfic KEGG per a l'anàlisi de les rutes metabòliques.

CAPÍTOL 2



APLICACIÓ DE LA CROMATOGRÀFIA DE
GASOS EN L'ESTUDI DEL METABOLOMA DE
LA *D. MAGNA* EXPOSADA A FACTORS
ABIÒTICS

2.1 INTRODUCCIÓ

La cromatografia de gasos acoblada a l'espectrometria de masses (GC-MS) ha estat àmpliament utilitzada en estudis de metabolòmica per tal de determinar una sèrie de metabòlits i biomolècules que s'expressen en organismes sotmesos a un estrès ambiental (Jones et al. 2012; Mishra et al. 2017). Aquesta tècnica analítica proporciona una elevada sensibilitat i reproductibilitat, alhora que permet una identificació dels metabòlits a través de l'àmplia disponibilitat de bases de dades com les del *National Institute of Science and Technology* (NIST) (Simón-Manso et al. 2013), la *Golm Metabolome Database* (Kopka et al. 2005) i la *Fiehn Library* (Kind et al. 2009).

Tot i el seu extens ús, la GC-MS només permet analitzar compostos volàtils. Per analitzar metabòlits, que són compostos no volàtils i polars, és necessari un procés de derivatització química per tal de transformar-los en derivats volàtils (Orata 2012). La derivatització implica la reacció de tots els metabòlits extrets en la mostra biològica amb l'agent derivatitzant. Així, aquest pas comprèn un conjunt de reaccions paral·leles dels diferents metabòlits de diferents estructures i característiques químiques amb un mateix agent derivatitzant (Kanani et al. 2008). El procediment de derivatització requereix la optimització de:

- La universalitat d'un únic agent derivatitzant per a tots els metabòlits presents en la mostra biològica.
- La quantitat d'agent derivatitzant que cal afegir en la mostra: per tal de portar a terme una completa derivatització de tots els metabòlits presents en la mostra l'agent derivatitzant ha d'estar en excés.
- La durada de la derivatització: per tal que les reaccions de derivatització s'hagin completat en el moment d'anàlisi a través de la GC-MS.
- L'estabilitat de la derivatització: ja que les reaccions de derivatització no són permanents al llarg del temps.

Finalment, tenint en compte la gran varietat de condicions que poden condicionar la derivatització dels metabòlits (Kanani and Klapa 2007), en aquesta Tesi

es va decidir portar a terme una optimització de les condicions de derivatització per tal de maximitzar la detecció dels metabòlits de la *D. magna*. Entre altres, els agents sililants més destacats per a la derivatització dels grups hidroxils i les amines primàries es troben el trimetilsilil (TMS) i el *tert*-butildimetilsilil (TBDMS), a més de la metoximació (MeOX) per a la derivatització dels grups carbonils. La utilització dels agents derivatitzants més comuns permet portar a terme la identificació dels metabòlits derivatitzats a partir de les bases de dades espectrals (Halket et al. 2005).

En la cromatografia de gasos l'adequada selecció de la fase estacionària està relacionada amb la detecció dels metabòlits derivatitzats. En el camp de la metabolòmica generalment s'utilitza una columna convencional de fase estacionària no polar amb un 5% difenil i 95% dimetilpolisiloxà de dimensions de 30 m de llargada x 0.25 mm de diàmetre intern x 0.25 µm de grossor de partícula (Jones, et al. 2012; Mishra, et al. 2017).

Per altra banda, la cromatografia de gasos en dues dimensions (GCxGC) es pot definir com una tècnica instrumental formada per dues columnes cromatogràfiques de fases estacionàries amb característiques químiques independents, capaces de proporcionar una resposta de manera bidimensional. La finalitat principal de l'ús d'aquesta tècnica recau en la capacitat d'identificar un gran nombre de compostos. Tal com es mostra en la **Figura 2.1**, es tracta d'una tècnica multidimensional en la qual s'aplica el poder de resolució de dues columnes diferents amb característiques de separació ortogonals, que permet resoldre molts pics diferents en mostres complexes. El modulador (**Figura 2.1**) actua com a nexa d'unió entre les dues dimensions, el qual permet la connexió entre ambdues columnes en sèrie. D'aquesta manera, la mostra es separa en una primera columna i, en segon lloc, mitjançant el modulador, la mostra es focalitza en fraccions estretes, les quals són injectades a intervals regulars i curts a la segona columna, de dimensions més curtes que la primera (veure **Figura 2.1**). L'efecte d'enfocament comporta el benefici addicional de crear pics estrets en la segona dimensió, de manera que augmenten les alçades de pic i la resolució, que a la vegada milloren la sensibilitat de detecció. Existeixen varis tipus de moduladors, entre els més habituals trobem els neumàtics o el tèrmics. Per una banda, els neumàtics focalitzen

mitjançant línies de transferència i diferències de pressió, mentre que els tèrmics focalitzen les fraccions treballant per sobre o per sota de la temperatura ambient. Entre els moduladors tèrmics trobem els moduladors criogènics que utilitzen el nitrogen gas per atrapar i focalitzar les fraccions per injectar-les a la segona dimensió. Aquest tipus de modulador és l'utilitzat en els treballs de GCxGC-TOFMS d'aquesta Tesi (**articles científics III i IV**).

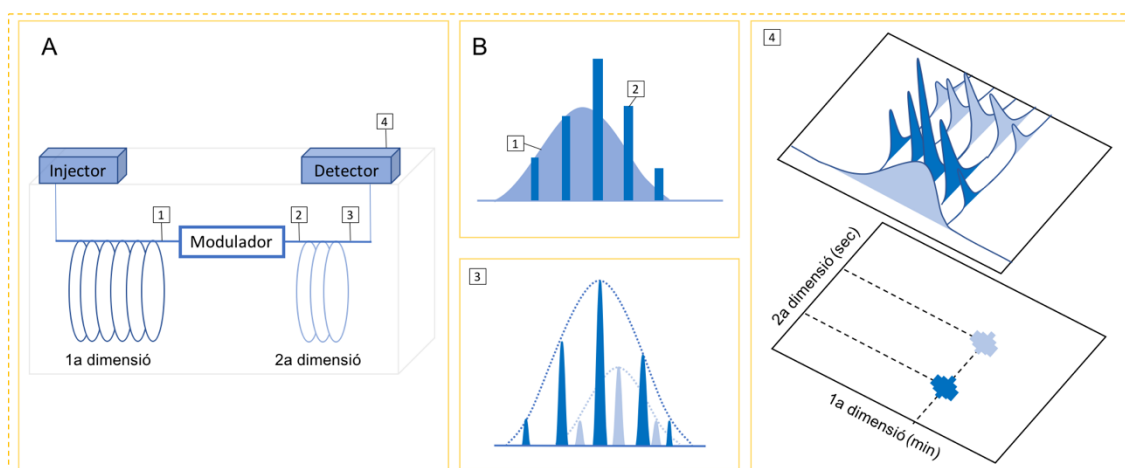


Figura 2.1| Configuració de la GCxGC (A) i el principi de modulació (B). Els pics que elueixen des de la primera columna (1) són focalitzats pel modulador (2) i re-injectats en la segona columna per una separació més exhaustiva (3). El senyal registrat pel detector (4) es divideix segons el període de modulació; la combinació de cromatogrames en les dues dimensions permet una reconstrucció dels gràfics de contorn en dues dimensions (imatge adaptada de (Vendevre et al. 2007)).

La cromatografia GCxGC permet una major eficiència en la separació dels metabòlits en comparació amb la cromatografia unidimensional. Una de les principals combinacions d'aquesta plataforma analítica és el seu acoblament a detectors d'adquisició ràpida, com per exemple el temps de vol acoblat a l'espectrometria de masses (*time-of-flight mass spectrometry*, TOFMS). Aquesta combinació, GCxGC-TOFMS, és un dels mètodes més utilitzats en estudis de metabolòmica, ja que permet determinar un perfil metabòlic de mostres biològiques complexes d'una manera precisa donada la seva elevada velocitat d'enregistrament del senyal. A més, permet una elevada capacitat de separació, una millora de la resolució i una millor identificació i quantificació dels metabòlits analitzats (Koek et al. 2011b), així com la desconvolució

de pics superposats, la quantificació i la identificació dels metabòlits a través de les biblioteques espectrals de masses (Almstetter et al. 2012).

En la cromatografia GCxGC, s'utilitzen dues columnes de diferent polaritat i, al ser possible, de diferent diàmetre en cada posició per tal de maximitzar la capacitat de pic (Koek et al. 2008). Així, en aquesta Tesi s'ha estudiat la influència de diferents configuracions de columnes analítiques en la detecció dels metabòlits mitjançant GCxGC-TOFMS (**article científic III**). Tot i això, en la majoria dels estudis de metabolòmica s'utilitza una columna convencional de fase estacionària no polar amb un 5% difenil i 95% dimetilpolisiloxà en la primera dimensió (1D) de 30 m de llargada x 0.25 mm de diàmetre intern x 0.25 µm d'espessor de fase, la qual permet una separació dels compostos basada en la seva volatilitat. Com a resultat, cada petita fracció que passa a la segona dimensió estarà formada per compostos de volatilitats semblants. En la segona dimensió (2D), la columna utilitzada és de polaritat intermèdia amb un 50% difenil i 50% dimetilpolisiloxà i de dimensions 1-2 m de llargada x 0.1 mm de diàmetre intern x 0.1 µm d'espessor de fase. El principal avantatge d'utilitzar una columna no polar en la 1D és que permet transferir els mètodes que s'han desenvolupat en la GC monodimensional. Els dos processos cromatogràfics, en la 1D i en la 2D, són independents, és a dir que la separació és ortogonal i, en aquestes condicions, la capacitat de pic de la separació de la 2D és màxima. En aquesta dimensió, l'anàlisi és molt ràpida, pel que no hi ha pràcticament dependència de la volatilitat dels compostos en la seva separació, sinó que la seva retenció estarà relacionada amb interaccions específiques dels compostos amb la fase estacionària.

En la bibliografia trobem nombroses aplicacions de la utilització de la tècnica GCxGC (especialment GCxGC-TOFMS), per exemple en el camp de l'anàlisi de residus químics de la utilització de pesticides presents en aliments (Zrostílková et al. 2003) o bifenils policlorats (PCBs) (Focant et al. 2004). També, s'han publicat articles relacionats en la caracterització d'hidrocarburs, com per exemple compostos orgànics volàtils (VOCs), compostos aromàtics policíclics (PAHs) o retardants de flama halogenats presents en els sòls (Escobar-Arnanz et al. 2018). Els estudis en el camp de la metabolòmica són més escassos, tot i que s'ha utilitzat la GCxGC-TOFMS sobretot en

metabolòmica de plantes, per exemple per determinar els perfils metabolòmics característics de diverses espècies (Hope et al. 2005; Pierce et al. 2006), o per avaluar la resposta metabòlica de plantes exposades a contaminants ambientals emergents (CECs) (Hurtado et al. 2017).

En aquesta Tesi es proposa una metodologia basada en GCxGC-TOFMS per estudiar el metaboloma de la *D. magna*. Aquesta tècnica analítica millora la selectivitat i la sensibilitat a causa de l'efecte de focalització de la mostra que proporciona el modulador en comparació a la GC-MS. D'aquesta manera permet ampliar al màxim el nombre de metabòlits que es poden detectar.

En aquest capítol s'inclouen les següents publicacions:

- **Article científic I:** *Metabolic profiling of Daphnia magna exposed to environmental stressors by GC-MS and chemometric tools*. E. Garreta-Lara, B. Campos, C. Barata, S. Lacorte, R. Tauler. *Metabolomics* 12 (2016), 86.

En aquest article es van optimitzar les condicions per detectar metabòlits en mostres de *D. magna* mitjançant la GC-MS. Es van posar a punt les condicions de derivatització, així com l'anàlisi de dades quimiomètriques mitjançant PCA i MCR-ALS. A més, mitjançant l'aplicació del mètode PLS-DA es van identificar els metabòlits que presentaven canvis deguts als factors estressants estudiats (salinitat, temperatura i hipòxia) i relacionar-los amb les rutes metabòliques afectades durant l'exposició individual de cada factor.

- **Article científic II:** *Combined effects of salinity, temperature and hypoxia on Daphnia magna metabolism*. E. Garreta-Lara, B. Campos, C. Barata, S. Lacorte, R. Tauler. *Science of the Total Environment* 610-611 (2018), 602-612.

En aquest article es va estudiar d'una manera més exhaustiva els efectes dels tres factors abiòtics (salinitat, temperatura i hipòxia) mitjançant un disseny experimental, i com afecten en el metaboloma de la *D. magna*. Les dades s'analitzen mitjançant MCR-ALS, i s'utilitzen les eines quimiomètriques PCA, ASCA i OPLS-DA per a la interpretació

dels metabòlits afectats pels diferents factors estudiats, amb l'objectiu de determinar el factor o la combinació de factors que tenen més influència dins del model.

- **Article científic III:** *Chemometric analysis of comprehensive two dimensional gas chromatography–mass spectrometry metabolomics data.* Y. Izadmanesh, E. Garreta-Lara, J. B. Ghasemi, S. Lacorte, V. Matamoros, R. Tauler. *Journal of Chromatography A*, 1488 (2017), 113-125.

En aquest article s'ha desenvolupat una metodologia per a l'anàlisi per GCxGC-TOFMS en mostres de *D. magna*. S'han avaluat diferents configuracions de columnes per tal d'optimitzar la separació dels analits i facilitar la identificació del metaboloma de la *D. magna*. Degut a la gran quantitat d'informació que contenen les mostres, ha estat necessària el desenvolupament d'un procediment per a la compressió prèvia de les dades a través de *wavelets* per poder analitzar-les a través del mètode MCR-ALS.

- **Article científic IV:** *Chemometrics comparison of gas chromatography with mass spectrometry and comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry Daphnia magna metabolic profiles exposed to salinity.* H. Parastar, E. Garreta-Lara, B. Campos, C. Barata, S. Lacorte, R. Tauler. *Journal of Separation Science* 41 (2018), 2368-2379.

En aquest article es va realitzar un estudi comparatiu entre la GC-MS en 1D i en 2D (GCxGC-TOFMS) a partir de l'anàlisi dels perfils metabòlics de la *D. magna* exposada a un augment de la salinitat. Es va portar a terme una anàlisi de les dades quimiomètrica basada en la compressió de les dades mitjançant les finestres de temps i les *wavelets*, i l'ús del mètode MCR-ALS per tal de resoldre i identificar els metabòlits de la *D. magna*. Els metabòlits més significatius per a la diferenciació entre les mostres controls i les exposades a la salinitat van ser determinats mitjançant l'anàlisi PLS-DA, on es van utilitzar els valors de la projecció de les variables importants (VIP) i la relació de selectivitat (SR).

2.2 ARTICLE CIENTÍFIC I

Metabolic profiling of Daphnia magna exposed to environmental stressors by GC-MS and chemometric tools.

E. Garreta-Lara, B. Campos, C. Barata, S. Lacorte, R. Tauler.

Metabolomics 12 (2016), 86.



Metabolic profiling of *Daphnia magna* exposed to environmental stressors by GC–MS and chemometric tools

Elba Garreta-Lara¹ · Bruno Campos¹ · Carlos Barata¹ · Silvia Lacorte¹ · Romà Tauler¹

Received: 18 November 2015 / Accepted: 2 March 2016 / Published online: 16 March 2016
© Springer Science+Business Media New York 2016

Abstract

Introduction Climate change is a major concern for the scientific community, demanding novel information about the effects of environmental stressors on living organisms. Metabolic profiling is required for achieving the most extensive possible range of compounds and their concentration changes on stressed conditions.

Objectives Individuals of the crustacean species *Daphnia magna* were exposed to three different abiotic factors linked to global climate change: high salinity, high temperature levels and hypoxia. Advanced chemometric tools were used to characterize the metabolites affected by the exposure.

Method An exploratory analysis of gas chromatography–mass spectrometry (GC–MS) data was performed to discriminate between control and exposed *daphnid* samples. Due to the complexity of these GC–MS data sets, a comprehensive untargeted analysis of the full scan data was performed using multivariate curve resolution–alternating least squares (MCR–ALS) method. This approach enabled to resolve most of the metabolite signals from interference peaks caused by derivatization reactions. Metabolites with significant changes in their peak areas were tentatively identified and the involved metabolic pathways explored.

Results *D. magna* metabolic biomarkers are proposed for the considered physical factors. Metabolites related with

energy metabolic pathways including some amino acids, carbohydrates, organic acids and nucleosides were identified as potential biomarkers of the investigated treatments. **Conclusions** The proposed untargeted GC–MS metabolomics strategy and multivariate data analysis tools were useful to investigate *D. magna* metabolome under environmental stressed conditions.

Keywords Metabolic profiling · Gas chromatography–mass spectrometry · *Daphnia magna* · Environmental stressors · Chemometric tools

1 Introduction

Metabolomics is concerned with the study of naturally occurring, low molecular weight organic metabolites within a cell or tissue (Griffiths 2008; Harrigan and Goodacre 2003; Lindon et al. 2007). An important application of metabolomics is the study of the interactions of living organisms with their environment, which is defined as environmental metabolomics (Bundy et al. 2009; Miller 2007). A definition proposed for this field is “the application of metabolomics to the investigation of both free-living organisms obtained directly from the natural environment and of organisms reared under laboratory conditions” (Morrison et al. 2007). Metabolomics presents several advantages for the study of organism–environment interactions and for assessing the health status of organisms in the field (Bundy et al. 2009; Krastanov 2010). Metabolomic measurements report on the actual functional status of biological organism levels, (i.e. cell, tissue) which, in principle, can be mechanistically anchored to effects occurring at lower (i.e. genomic) or higher (phenotypic) levels of biological organization (Fiehn 2001). As such, at present, the number of metabolomic

Electronic supplementary material The online version of this article (doi:10.1007/s11306-016-1021-x) contains supplementary material, which is available to authorized users.

✉ Romà Tauler
roma.auler@idaea.csic.es

¹ Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, 08034 Barcelona, Catalonia, Spain

studies applied to environmental problems are growing considerably trying to improve the understanding of organism responses to abiotic stressors (Arbona et al. 2013; Sardans et al. 2011). Metabolomic studies are based on simultaneous measurement of multiple metabolites, using inherently parallel analytical techniques such as nuclear magnetic resonance spectroscopy or mass spectrometry, followed by appropriate statistical analysis (Lindon and Nicholson 2008a, b; Dunn et al. 2011). The application of metabolomic techniques, however, is still limited, due to the inherent problems of separation and identification of unique metabolites and hence metabolic changes (Viant 2008; Bundy et al. 2009; Samuelsson and Larsson 2008; Miller 2007).

To evaluate the effects of stressors on the metabolome, gas chromatography coupled to mass spectrometry (GC–MS) in electron ionization (EI) is a widely used analytical technique as it provides high sensitivity and reproducibility, and permits compound identification using spectral libraries (Steinhauser and Kopka 2007; Vandebrouck et al. 2010b). The major limitation of GC–MS is the restriction to semi-volatile analytes, and hence the requirement to convert many polar metabolites to volatile derivatives through chemical derivatization processes. Oximation and trimethylsilylation (TMS) reactions of organic compounds are the classical and most widely used derivatization procedures for metabolome analysis by GC–MS (Villas-Boas et al. 2011; Kanani et al. 2008; Yi et al. 2014; Gullberg et al. 2004; De Souza 2013; Little 1999). Derivatization, however, increases the amount of adduct-derivatives and hence the complexity of metabolomics data. Understanding this complex big-data generated from GC–MS is still currently a bottleneck (Vandebrouck et al. 2010b; Dunn et al. 2013; Wehrens et al. 2014).

In order to maximize the extraction of information, a pipeline for GC–MS metabolomic data analysis was designed and principal component analysis (PCA), partial least squares-discriminant analysis (PLS-DA) and multivariate curve resolution-alternating least squares (MCR-ALS) (De Juan et al. 2014; van Stokkum et al. 2009; Tauler 1995) were proposed for the resolution of complex chemical mixtures. This last chemometric tool is a powerful data analysis tool that solves the mixture analysis and interferences co-elution problems in complex natural samples as well as problems derived from GC–MS systems, such as baseline drift, spectral background, noise contributions or low S/N ratio values (Parastar et al. 2012). MCR-ALS has been also used in previous works related with ecological and environmental problems (Navarro-Reig et al. 2015; Malik and Tauler 2015; Alier et al. 2010; Terrado et al. 2009).

Daphnia magna, a freshwater crustacean, is used extensively as an aquatic test species (Ikenaka et al. 2006;

Martins et al. 2007; Barata and Baird 2000a) being the object of standardized testing guidelines from the Organization for Economic Co-operation and Development (OECD 2012). Acute and chronic tests of *D. magna* are among the most frequently performed studies in aquatic toxicology because these animals are relatively easy to culture, have a short lifecycle, and can be maintained at high population densities in relatively small volumes and thus are cost-effective (Barata et al. 2005; De Meester and Vanoverbeke 1999). *Daphnia magna* is also a keystone species in freshwater food webs, thus it is also an ecological model species to study effects of global climate change. Water temperature, salinity and oxygen levels of freshwater bodies are likely to be affected by climate change (IPCC 2014; Adrian et al. 2009; Coutant 1990).

Here an untargeted GC–MS metabolomics study of *D. magna* exposed to mild stress provoked by high salinity, increased temperature and low dissolved oxygen is presented.

2 Materials and methods

2.1 Chemicals and standard solutions

Pure metabolites: pimelic acid, malic acid, L-phenylalanine, L-tyrosine, putrescine, dopamine hydrochloride, L-histidine, L-methionine, D-maltose monohydrate, myo-inositol, D-ribose, D-galactose and cytidine used as test standards were purchased from Sigma-Aldrich (St. Louis, MO, USA). D-glucose (U-13C6, 99 %), used as the internal standard (IS), was supplied by Cambridge Isotope Laboratories, Inc. (Andover, MA, USA). Pyridine (anhydrous, 99.8 %), chlorotrimethylsilane (TMCS), methoxyamine hydrochloride (98 %) (MeOX) and *N*-methyl-*N*-trimethylsilyl trifluoroacetamide (>98.5 %) (MSTFA), used as derivatizing agents, and the saturated Alkane standard mixture for the performance test of GC systems from C₇ to C₃₀ were also obtained from Sigma-Aldrich (St. Louis, MO, USA). Hexane, methanol and chloroform were analytical reagent grade, and sodium chloride (NaCl) salt was supplied by Merck (Darmstadt, Germany).

2.2 Sample preparation

2.2.1 *Daphnia magna* growth conditions

A single laboratory clone F, which has been the subject of many investigations (Barata and Baird 2000b) was used. Bulk cultures of 10 animals L⁻¹ were maintained in ASTM hard synthetic water (ASTM, 1995) as described by Barata and Baird (2000b). Individual or bulk cultures were fed daily with *Chorella vulgaris* Beijerinck (5 × 10⁵ cells

mL^{-1} , corresponding to $1.8 \mu\text{g C mL}^{-1}$) (Barata and Baird 1998). The culture medium was changed every other day, and neonates were removed within 24 h. Photoperiod was set to 14 h light: 10 h dark cycle and temperature at $20 \pm 1 \text{ }^\circ\text{C}$.

2.2.2 Experimental design

Four days old *D. magna* juveniles were used for experiment. Experimental animals were obtained from bulk cultures of 100 individuals reared in 10 L of ASTM as described above ("*Daphnia magna* growth conditions"). Randomly collected animals were exposed to high salinity, elevated temperature and hypoxia treatments in groups of ten individuals in 1.2 L jars filled with 1 L of ASTM hard water. A control treatment of individuals exposed to ASTM hard water maintained at $20 \text{ }^\circ\text{C}$ under saturating oxygen conditions ($\approx 9 \text{ mg L}^{-1}$) without salt (0 g L^{-1} NaCl) was also included. Each treatment was replicated five times. Salinity exposed samples were prepared with 5 g L^{-1} NaCl in ASTM hard water, temperature samples were prepared in a water bath at $25 \text{ }^\circ\text{C}$ and hypoxia samples at $\leq 2 \text{ mg O}_2 \text{ L}^{-1}$ were prepared by bubbling liquid nitrogen into the ASTM hard water thus displacing oxygen from it. The oxygen levels were controlled by an oxygen sensor, approximately the levels had to be $\leq 2 \text{ mg L}^{-1}$ in every jar sample. After 24 h of exposition *daphnids* were collected from each jar, pooled in an eppendorf, the water was removed and then samples were frozen in liquid N_2 and stored at -80°C before metabolite extraction.

2.2.3 Extraction and derivatization of metabolites

In the first part of this work, GC–MS method and derivatization step were optimized to obtain high sensitivity and to maximize the number of detected compounds. Several standard mixtures (i.e. organic acids, sugars, nucleotides, nucleosides and amino acids) were tested to optimize the experimental conditions and the derivatization step in order to ensure a comprehensive and reliable metabolic profiling. An aqueous standard solution ($1000 \mu\text{g mL}^{-1}$) of each standard was prepared in methanol and stored in the freezer at $-20 \text{ }^\circ\text{C}$ until their use. Working standard solutions of $50 \mu\text{g mL}^{-1}$ were obtained by diluting the stock solutions with water. Diluted standard solutions of $5 \mu\text{g mL}^{-1}$ were used to optimize experimental conditions and also several derivatization conditions were tested. Metabolite extraction was performed with methanol, water and chloroform and the chemical derivatization was carried out with two agents, methoxyamine (MeOX) in pyridine and *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (MSTFA) with 1 % of trimethylchlorosilane (TMCS) as a catalyst of the

reaction, which are very commonly used in metabolomics studies by GC–MS analysis.

Polar metabolites of the whole organism were extracted with 400 μL methanol, sonicated for 15 min, and then 200 μL water and 400 μL chloroform were added before centrifugation at $10,000\times g$ during 15 min, in order to separate the aqueous and the lipid phase. The aqueous phase of every sample was transferred to a new eppendorf and 10 μL of the Internal Standard (IS) [D-glucose (U-13C6, 99 %)] were added at a concentration of $50 \mu\text{g mL}^{-1}$. Afterwards, the extract was evaporated to dryness with a Speedvac (Thermo Scientific) at $40 \text{ }^\circ\text{C}$ during 3 h. When samples were completely dry they were stored at $-80 \text{ }^\circ\text{C}$ until analysis.

Prior the analysis by GC–MS, a derivatization method was performed to create thermally stable derivatives suitable for the analysis (Orata 2012). The initial conditions were adjusted from (Vandenbrouck et al. 2010a) and adapted to the needs of this study, in order to detect the maximum number of metabolites. The first step was methoximation with methoxyamine hydrochloride (MeOX) in pyridine, which was used to stabilize carbonyl radicals of metabolites and to stop charge transfer. Then, trimethylsilylation was carried out with *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (MSTFA) with 1 % of trimethylchlorosilane (TMCS) as catalyst. Trimethylsilylation was applied to replace active hydrogens on a wide range of polar compounds with a trimethylsilyl [$-\text{Si}(\text{CH}_3)_3$] group. This derivatization method permits to detect the silylated molecules in the gas chromatography analysis because it increases the volatility of the metabolites. Different MeOX concentrations (20, 30 and 40 mg mL^{-1}), MeOX volumes (30, 65 and $100 \mu\text{L}$), MeOX temperatures ($20, 30$ and $40 \text{ }^\circ\text{C}$), MeOX times (90, 525 and 960 min), MSTFA volumes (30, 65 and $100 \mu\text{L}$), MSTFA temperatures ($20, 30$ and $40 \text{ }^\circ\text{C}$) and MSTFA times (30, 60 and 90 min) were tested to improve the derivatization reaction and to determine the most suitable conditions for the analysis of *D. magna* metabolites (Pacchiarotta et al. 2010; Kanani et al. 2008; Azizan et al. 2012; Kanani and Klapa 2007; Strehmel et al. 2008).

After the extraction of polar metabolites, the final protocol of the derivatization conditions was implemented: to the dry residue obtained after extraction from *daphnids*, 65 μL of $20 \mu\text{g mL}^{-1}$ MeOX in pyridine were added. After mixing for 1 min, the mixture was incubated for 90 min at $30 \text{ }^\circ\text{C}$. Thereafter, 30 μL of MSTFA (1 % TMCS) were added, vortex mixed for a minute and then incubated for another 30 min at room temperature. Prior to injection, *daphnid* derivatized extracts were finally filtrated through a $0.22 \mu\text{m}$ filters (Ultrafree[®]-MC, Millipore) and, then transferred to a chromatographic vial and 5 μL of hexane

were added to obtain a final volume of 100 μL in order to be injected into the GC system.

2.3 GC–MS analysis

A Trace GC–MS Thermo Fisher system operated in EI mode at 70 eV and equipped with a ZB-5MS column (Phenomenex, 30 m \times 0.25 mm ID \times 0.25 μm) (Vandenbrouck et al. 2010a) was used. The oven temperature program was set at 70 $^{\circ}\text{C}$ and then increased at 10 $^{\circ}\text{C min}^{-1}$ to 250 $^{\circ}\text{C}$, and then to 310 $^{\circ}\text{C}$ at 5 $^{\circ}\text{C min}^{-1}$ and held for 5 min, the delay time was 5.3 min and the time of analysis was 49 min. Other operating conditions were as follows: carrier gas was He, with a flow rate of 0.6 mL min^{-1} ; source temperature, 200 $^{\circ}\text{C}$; interface temperature, 280 $^{\circ}\text{C}$; split ratio of 1:20; detector voltage, 440 V; 2 μL were injected in the split mode injection. The m/z values were recorded in full scan mode in the range of m/z 60–650 amu. The mass spectrometer was interfaced to a computer workstation running Xcalibur software (version 2.2, Thermo Scientific) for data acquisition and processing. For each detected peak, a linear retention index (RI) (Strehmel et al. 2008) was calculated using GC RI standards (hydrocarbons from C_7 to C_{30} were used as internal standards). Therefore, an *n*-alkane Standard mixture was prepared dissolved in hexane at a concentration of 10 $\mu\text{g mL}^{-1}$, and was injected at the same programme conditions as the samples in the GC–MS instrument.

2.4 Data analysis

Original full scan GC–MS data sets “.raw“ from Xcalibur software (version 2.2, Thermo Scientific) were converted to “.cdf“ data files by File Converter tool of Xcalibur Software to be further processed with MATLAB (The Mathworks, Inc., Natick, MA, USA). Every data matrix corresponds to one of the samples analyzed by GC–MS (control and exposed), with 2581 rows (retention times, 49 min run) and the same number of rows equal to 590 columns (m/z values, from 60 to 650 with one unit of mass resolution). Data values in every data matrix were normalized with the peak area value of the Internal Standard (D-glucose U-13C6) to correct for the instrumental intensity drifts among injections, and to scale data internally. A total of 21 individual data matrices [1 blank, 5 controls and 15 exposed (3 treatments \times 5 exposed samples)] with dimensions of 2581 \times 590 (retention times, m/z values) were obtained and these resulting pre-processed data matrices were analyzed by MCR-ALS.

Additionally, for initial exploration of GC–MS data, a new data set was arranged consisting in a single data matrix where every individual total ion current (TICs) Chromatogram of every control and exposed sample were

arranged in the rows of this matrix. Three different single data matrices of TICs were obtained, one for each of the three investigated factors (salinity, temperature and hypoxia), with dimensions of 10 \times 2581 [number of samples (5 controls and 5 exposed) \times retention times], which were then submitted to PCA and PLS-DA using PLS Toolbox 7.3.1 (Eigenvector Research Inc., Wenatchee, WA, USA).

2.4.1 Initial exploration of GC–MS data

PCA (Wold et al. 1987; Esbensen and Geladi 2010) is an unsupervised technique that assumes no prior knowledge of class information and permits the unbiased comparison of the samples. This allows an initial exploration of the GC–MS TIC data comparing the behavior between control and exposed samples according to salinity, temperature and oxygen levels changes, finding patterns and detecting possible outliers in the samples. Prior to the chemometric analysis, some pretreatment methods were applied using PLS Toolbox 7.3.1 (Eigenvector Research Inc., Wenatchee, WA, USA). The first step was to align the chromatograms to compensate for possible retention time shifts between chromatographic runs using the correlation optimized warping method (Tomasi et al. 2004; Nielsen et al. 1998). After alignment, automatic weighted least squares from PLS Toolbox 7.3.1 (Eigenvector Research Inc., Wenatchee, WA, USA) was applied for automatically removing baseline offsets from TIC data matrices and variables were mean-centered also using PLS Toolbox 7.3.1 (Eigenvector Research Inc., Wenatchee, WA, USA).

Partial least squares discriminant analysis (PLS-DA) (Kalivodová et al. 2015; Wold et al. 2001) is a supervised linear regression technique in which the class for each group of samples is assigned prior to the construction of the sample scores plot, and the maximum separation of the classes (control and exposed samples) is achieved. Variable importance in projection (VIP) scores (Wold et al. 2001) were calculated to recognize the chromatographic regions of interest and to reveal the most influent variables (peak retention times related with metabolite concentration changes) discriminating controls and exposed samples.

2.4.2 MCR-ALS analysis

MCR-ALS method allows for the resolution of the pure elution and mass spectra profiles of all sample constituents in one or multiple samples analyzed by GC–MS, in a similar way to the one previously shown for LC-MS analysis (Farrés et al. 2014) or for other hyphenated chromatographic techniques (Ortiz-Villanueva et al. 2015). From the resolved elution profiles of the constituents, their peak heights or areas can be easily evaluated and compared

between runs. Due to the high fragmentation of molecules observed in GC–MS and to the presence of a large number of derivatized subproducts, the proposed MCR-ALS analysis is especially helpful because it resolves and distinguishes the relevant metabolites profiles of sample constituents from the irrelevant contributions of the derivatizing signal, and therefore, it allows for the discrimination of the relevant *D. magna* elution profiles (metabolites) from the large number of interfering peaks (undesired derivatized compounds).

MCR analysis can be mathematically expressed with the bilinear model shown in the Eq. 1:

$$D = CS^T + E \tag{1}$$

where C contains the chromatographic profiles of all the resolved components, S^T their resolved (pure) mass spectra, and E is the matrix expressing the error or the variance not explained by the model, related to background and other unknown noise contributions. Every full scan GC–MS gives a data matrix, D, which has the mass spectra at all retention times in its rows, and the chromatograms at all m/z values in its columns.

As shown previously, the solution to Eq. 1 for C and S^T is ambiguous if no additional information is available, in other words, this solution contains rotational and scale freedom. This problem is usually referred as the factor analysis ambiguity problem (Tauler 2001). This problem can affect the resolution of the component profiles. Thus, exploratory information should be used as constraints during the resolution process and it can also be used to build good initial estimates of concentration and spectra profiles. Constraints such as non-negativity and others are frequently used in MCR studies (De Juan and Tauler 2007). In the case of GC–MS data, due to the very high spectral

selectivity, possible ambiguities in the resolution of elution and spectral profiles are very much diminished or totally eliminated in most of the cases.

Apart from the derivatizing agent blank sample, five control and five exposed *D. magna* samples were obtained for every investigated treatment: salinity, temperature and hypoxia. This gives a total number of 21 samples analyzed by GC–MS [1 blank, 5 controls and 15 exposed (3 treatments × 5 exposed samples)], giving 21 individual data matrices, each of them of dimensions 2581 retention times (between 5.3 min and 49 min) and 590 m/z values (between 60 and 650 m/z values). These individual GC–MS full scan large sized data matrices were subdivided into eleven different chromatographic time sub-windows to reduce computer storage and execution time requirements. The size limits of these time subwindows were initially obtained considering a sufficient number of retention times (approximately 200 retention times) to cover complete peak clusters and to avoid peak halving. See Table 1 for the details of these eleven time windows and their retention times.

In order to analyze all these data sets, eleven new column-wise augmented data matrices D_{aug} , corresponding to every time window containing the information of one blank derivatizing agent sample, five replicate control samples and five replicate abiotic exposed *daphnids* samples, were built for every type of treatment (salinity, temperature and hypoxia). A total number of thirty three (3 treatments × 11 time windows) column-wise augmented data matrices were obtained by arranging the corresponding data matrices one on top of each other, containing in all cases the same number of columns (590 m/z values). In this way, all the measured mass spectra were linked across all data (with common m/z values). GC–MS data matrices of a blank sample containing only the derivatizing agent were included in all analyses to

Table 1 MCR-ALS results in the analysis of the augmented data matrices for the three treatments (salinity, temperature and hypoxia)

Time window (Tw)	Retention time (min)	Dimensions of the Tw	Salinity			Temperature			Hypoxia		
			Estimated components	Resolved peaks	R ² (%)	Estimated components	Resolved peaks	R ² (%)	Estimated components	Resolved peaks	R ² (%)
1	5.3–8.7	202 × 591	12	8	95.2	13	9	95.9	18	10	96.1
2	8.7–10.0	80 × 591	9	6	96.2	12	8	98	9	6	97.8
3	10.0–12.2	129 × 591	12	11	96.4	13	11	97.4	13	7	98.6
4	12.2–17.0	291 × 591	13	10	94.8	15	12	96.4	17	11	98.2
5	17.0–20.0	181 × 591	8	6	95.7	13	8	97.8	13	9	98.6
6	20.0–22.6	156 × 591	11	9	98.6	11	9	98.7	11	9	98.9
7	22.6–25.2	157 × 591	12	9	98.4	10	7	97.7	13	5	99.6
8	25.2–29.5	260 × 591	10	7	97.9	10	5	97.8	12	3	99.5
9	29.5–34.5	301 × 591	7	4	98.7	7	3	99.2	9	4	98.5
10	34.5–38.7	251 × 591	9	6	99.1	7	3	98.8	9	5	98.7
11	38.7–49.0	581 × 591	8	5	99.1	5	2	98.1	8	1	98.8

determine the influence of derivatization and to describe the components influenced by it. In Fig. 1 these column-wise matrix arrangements are displayed.

Every column-wise augmented data matrix, \mathbf{D}_{aug} , corresponding to a time subwindow and containing 11 submatrices (one blank derivatizant, \mathbf{D}_b ; five control ($\mathbf{D}_1 \dots \mathbf{D}_5$) and the five exposed samples ($\mathbf{D}_6 \dots \mathbf{D}_{10}$) according to a specific treatment) can be decomposed according to the bilinear model Eq. 2 (see also this matrix decomposition in Fig. 1):

$$\mathbf{D}_{\text{aug}} = \begin{bmatrix} \mathbf{D}_b \\ \mathbf{D}_1 \\ \mathbf{D}_2 \\ \vdots \\ \mathbf{D}_{10} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_b \\ \mathbf{C}_1 \\ \mathbf{C}_2 \\ \vdots \\ \mathbf{C}_{10} \end{bmatrix} \mathbf{S}^T + \begin{bmatrix} \mathbf{E}_b \\ \mathbf{E}_1 \\ \mathbf{E}_2 \\ \vdots \\ \mathbf{E}_{10} \end{bmatrix} = \mathbf{C}_{\text{aug}} \mathbf{S}^T + \mathbf{E}_{\text{aug}} \quad (2)$$

MCR-ALS applied to this augmented data matrix gave the resolved augmented elution profiles of the resolved components (\mathbf{C}_{aug}) for every sample and a matrix of the pure mass spectra of the resolved components (\mathbf{S}^T) for all samples.

MCR-ALS analysis requires an initial estimation of the number of components that can be obtained looking at the sizes of the singular values (Golub et al. 2000) of the investigated data matrix. Larger singular values are associated to systematic changes and lower ones to noise and minor contributions. This initial estimation can be afterwards revised according to the explained variances and to the interpretability of the recovered profiles. The number of pre-selected components should describe a sufficient amount of data variance and to include minor contributions such as background and derivatizing agents effects. Only components with reliable chromatographic peak shapes and meaningful metabolite mass spectra are finally

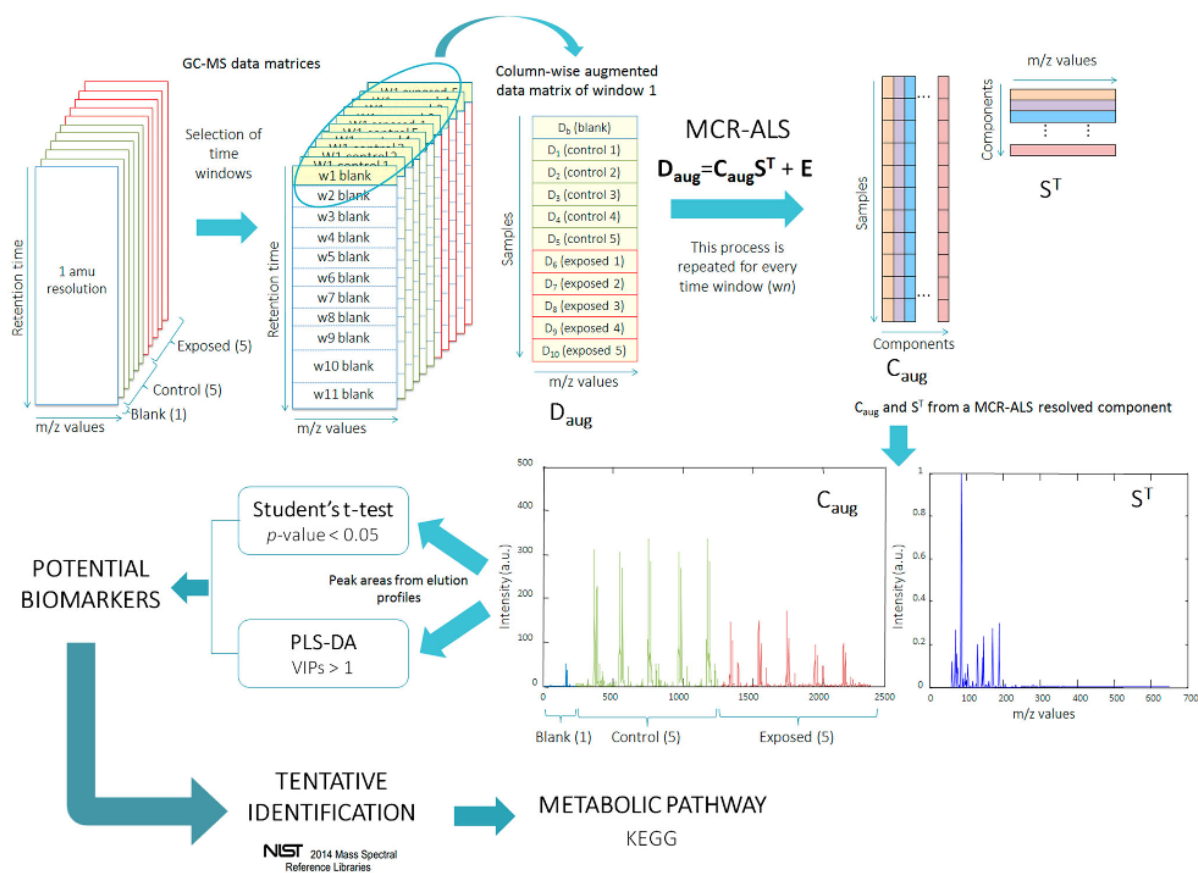


Fig. 1 GC-MS data sets for blank (1), control (5) and exposed (5) samples. Selection of time and column-wise matrix augmentation for the first time window (w1) \mathbf{D}_{aug} to be analyzed by MCR-ALS. Processing steps involve resolution of elution profiles (\mathbf{C}_{aug}) of the components present in this time window in the eleven samples [blank

(1), control (5) and exposed (5)] and of their pure mass spectra (\mathbf{S}^T). Statistical evaluation of peak area changes between control and exposed samples via Student's *t* test and PLS-DA. Tentative identification of potential biomarkers via NIST mass spectra matching and proposal of possible metabolic pathways involved

considered. MCR-ALS requires also the use of constraints to give meaningful solutions (for instance positive elution and spectra profiles) (Tauler and Barceló 1993; Tauler 1995; De Juan et al. 2010). In this work, constraints of non-negativity to chromatographic and mass spectra profiles and spectra normalization of the pure mass spectra profiles (spectra equal height) were applied. The quality of MCR-ALS models was measured evaluating the lack of fit, which is the difference among the input data D_{aug} and the data reproduced by MCR-ALS ($C_{\text{aug}}S^T$), and the percent of explained data variance (R^2).

2.4.3 Biomarkers detection and NIST identification

Resolved elution and spectra profiles can be used then to investigate potential biomarkers of the exposed effects (see Fig. 1). Two different procedures were applied to detect discriminant metabolites. First, a Student's *t* test was tested within the data matrix C_{aug} , resolved components showing significant concentration differences between groups (*p* value lower than 0.05) were finally considered and identified by their corresponding MS spectrum in S^T . Secondly, resolved elution profiles were autoscaled to give equal relevance to all metabolites whatever is their total concentration and, focusing more in their relative concentration changes in control and exposed samples. Then, PLS-DA was applied to MCR-ALS autoscaled resolved peak areas for selecting candidate metabolites. Variable importance in projection (VIP) scores (Eriksson et al. 2006) were chosen as selection criteria. Peak areas or elution profiles showing higher VIP scores than 1 were considered discriminant metabolites between control and exposed samples. Metabolites from selected peak areas or elution profiles were identified comparing RIs (from MCR-ALS resolved elution profiles) and mass fragmentation patterns (from MCR-ALS resolved mass spectra profiles) with the standard mass spectral database of the National Institute of Standards and Technology (USA) (NIST 2014) (www.nist.gov/srd/nist1a.htm) and of the Golm metabolome database (GMD) of derivatized compounds (Kopka et al. 2005; Schauer et al. 2005). Two different values for the matching factors of experimental and theoretical mass spectra can be obtained by the NIST *Mass Spectral Search* 2.2 program distributed with the NIST 2014 library, a match factor for the unknown and the library spectrum (direct match), and a match factor for the unknown and the library spectrum ignoring any peaks in the unknown that are not in the library spectrum, match factor (MF) and reverse match factor (RMF), respectively. For each mass spectrum, 100 hits were retrieved. These matching factors reported are between 0 (no match) and 1000 (perfect match). As a general guide, a value of 900 or greater was considered to be a very good matching; between 800 and

900, a good match; between 700 and 800, a fair match; and less than 600 a poor or very poor match. In the calculation of MF, the experimental spectrum is used as a template, whereas for RMF, the template is the library spectrum. To increase the reliability of the identification, we have included the internal linear RI markers in the evaluation of the library hits, by injecting a saturated Alkane standard mixture from C_7 to C_{30} . An advantage of using this MCR-ALS strategy is that it allows for a better comparison of the MS spectra of the resolved peaks with those from the NIST 2014 library, because of the simultaneous use of peak elution retention times and of the presence of the GC–MS fragment ions in the resolved mass spectrum assigned to these peaks.

Finally, as shown in the workflow of Fig. 1, metabolites tentatively identified as potential biomarkers of salinity, temperature and hypoxia treatments were further characterized by Kyoto encyclopaedia of genes and genomes (KEGG) database (www.genome.jp/kegg/ligand.html) (Kanehisa and Goto 2000) to investigate the possible metabolic pathways involved.

2.5 Software

Data analyses were performed using MATLAB 2013a (The Mathworks Inc. Natick, MA, USA) and PLS Toolbox 7.3.1 (Eigenvector Research Inc., Wenatchee, WA, USA) was used for PCA, PLS-DA and variables importance in projection (VIP) scores calculations. MCR-ALS toolbox (Jaumot et al. 2015) were used in MATLAB 2012b (The Mathworks Inc. Natick, MA, USA) for resolution of metabolite profiles from full MS scan augmented data matrices. The tentative identification of the metabolites was carried out by NIST *Mass Spectral Search* program (version 2.2) (<http://chemdata.nist.gov/mass-spc/ms-search/>) distributed with NIST 2014 *Mass Spectral Library*.

3 Results and discussion

3.1 PCA and PLS-DA exploratory and discriminatory analysis of GC–MS TIC of *D. magna* control and exposed samples

Preliminary analysis of GC–MS TIC data matrices (see Fig. 2) was performed for the three treatments (salinity, Fig. 2a, d, temperature, Fig. 2b, e, and hypoxia, Fig. 2c, f) using PCA and PLS-DA, respectively. In Fig. 2a, b, PC1-PC2 scores plot separates control (red) from exposed (blue) samples. In Fig. 2c (hypoxia treatment) PCA did not distinguish between control and exposed *D. magna* samples.

PLS-DA analysis was applied to the same three GC–MS TIC data matrices and VIP scores were calculated for every factor: salinity (Fig. 2d), temperature (Fig. 2e) and hypoxia

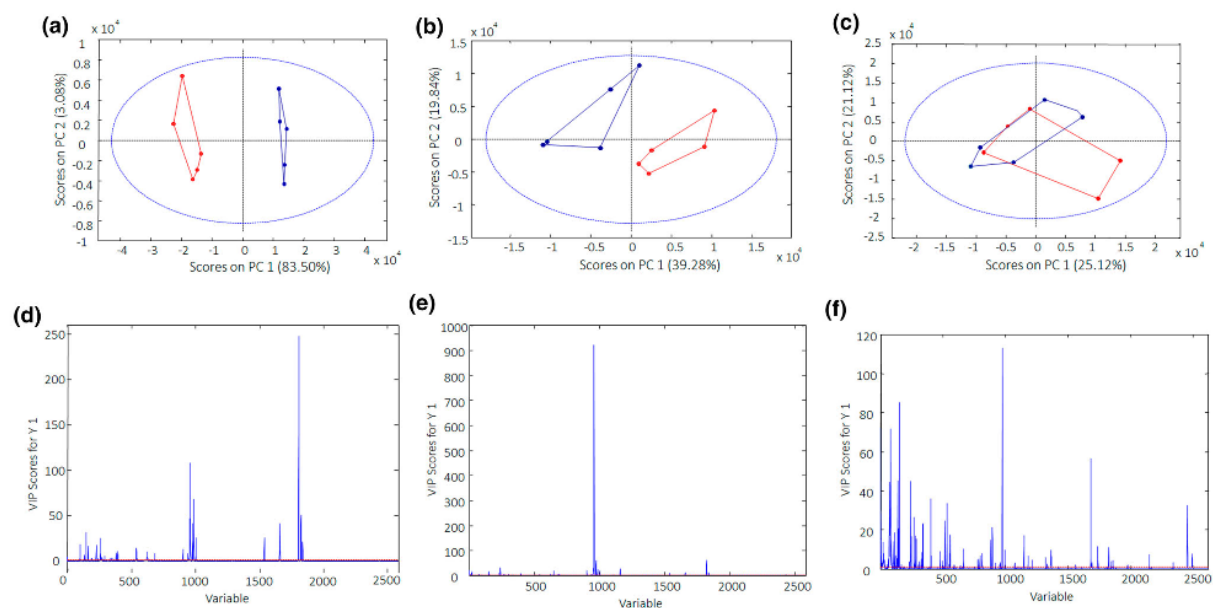


Fig. 2 PCA scores plots of the GC–MS TICs of *D. magna* extracts for the salinity (**a** 5 samples), temperature (**b** 5 samples) and hypoxia (**c** 5 samples) treatments. *Red* are control samples (control conditions) and *blue* are exposed samples (salinity, temperature or hypoxia). PLS-

DA VIP (variables importance in projection) scores plots for GC–MS TIC retention time variables across salinity (**d**), temperature (**e**), and hypoxia (**f**) treatments (Color figure online)

(Fig. 2f). In the case of salinity a single latent variable was selected explaining 83.5 and 98.31 % of the **X**- and **y**-variances respectively. In the case of temperature, two latent variables were selected explaining 56.19 and 97.19 % of the **X**- and **y**-variances respectively, and in the case of hypoxia, three latent variables were selected to explain 46.24 and 96.18 % of the **X**- and **y**-variances respectively. In all cases, the number of latent variables used for the PLS-DA model were much lower than the number of samples used in the analyses (10 samples). According to these graphs, for salinity and temperature, a small group of retention times were selected to explain most of the variance across control and exposed samples, whereas many retention times were needed to explain the hypoxia treatment.

These two exploratory analysis (PCA and PLS-DA) showed that exposures to salinity and temperature produced larger distinctive effects on the metabolite profile of *D. magna* samples. However, in the case of low oxygen, effects were not so pronounced and discrimination of samples was difficult. However, although GC–MS retention times with larger VIP scores could already be preliminary investigated, their very large number and the presence of co-eluting chromatographic peak prevented determining precisely metabolite profiles of exposed *D. magna* samples changed from this preliminary analysis. Further analysis was then attempted using a deeper MCR-ALS analysis of the set of individual full scan GC–MS data matrices for all analyzed samples (control and exposed).

3.2 MCR-ALS analysis of GC–MS full scan data matrices of *D. magna* control and exposed samples

MCR-ALS investigation of full scan data matrices from GC–MS analysis of control and exposed *D. magna* samples allowed the simultaneous resolution of the elution and mass spectra profiles of a large number of components. Number of finally resolved peaks, number of resolved peaks in each time window augmented data matrix for each treatment, and explained variances of the MCR-ALS model are summarized in Table 1. Dimensions of the 11 time window augmented data matrices for the three treatments and their dimensions are also given in Table 1.

For all treatments, explained variance (R^2) percentages were higher than 94.8 %, and, in all cases, the number of estimated components was higher than the number of resolved peaks with reliable chromatographic shapes. However, this additional number of estimated components was useful to explain other possible experimental data variance sources such as background and derivatization signals. Examples of results obtained in the MCR-ALS analysis of the GC–MS full scan data are given in Fig. 3 where the representation of resolved concentration and mass spectra profiles for different components are given.

In Fig. 3a, results from the application of MCR-ALS analysis on time window 5 for salinity treatment are shown. In this particular case, optimal results were obtained using

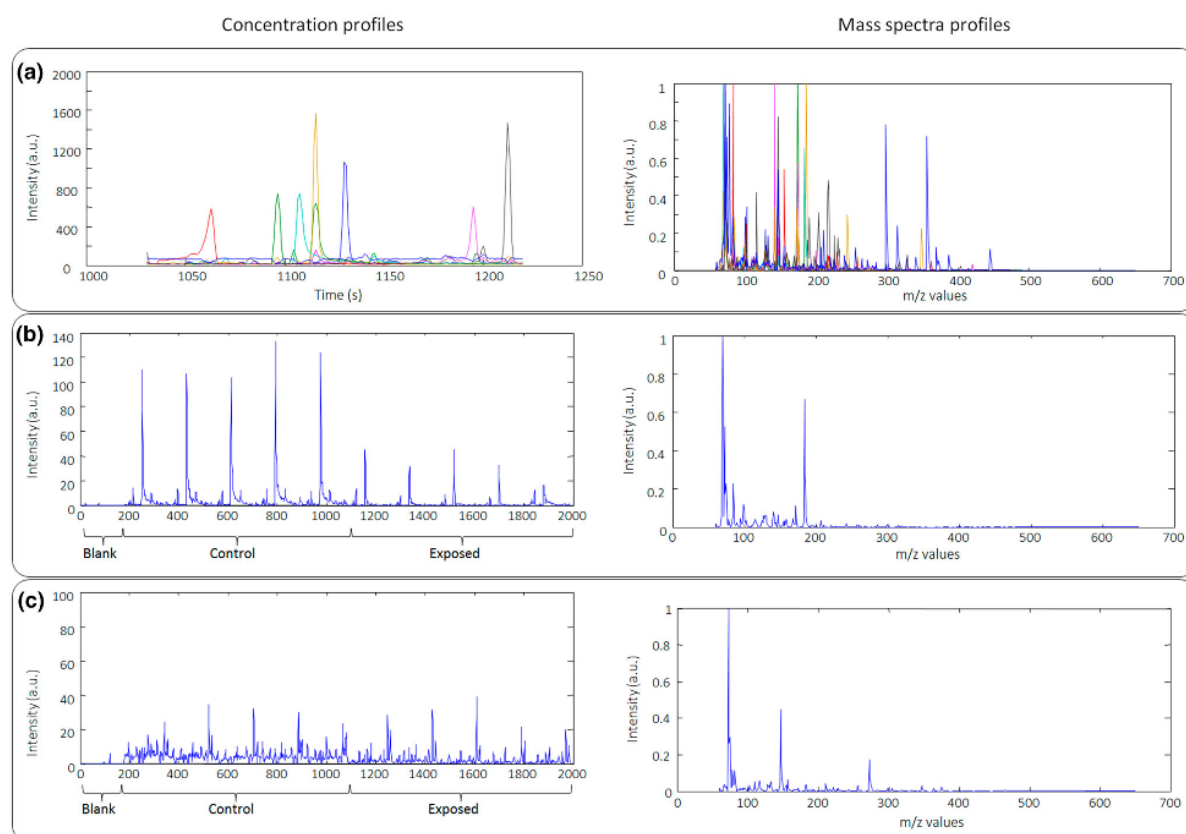


Fig. 3 Examples of elution profiles (*left*) and of corresponding mass spectra profiles (*right*) resolved by MCR-ALS analysis: **a** applied to one single control sample in the 5th time window for the salinity treatment, showing in *different colors* the 8 component resolved elution profiles in this time window (*left*), and the corresponding mass spectra profiles (*right*); **b** applied to the salinity treatment column-

wise augmented data matrix showing the resolution of one of the components in the 11 simultaneously analyzed matrices, in the order 1st blank sample, 2nd–6th control samples and 7th–11th exposed samples. And **c** applied to the same salinity treatment column-wise augmented data matrix than in **b** but showing the resolution of the derivatizing species too (Color figure online)

eight components (in different colors), some of them with highly overlapped (embedded peaks) elution profiles (on the left) and with multiple MS spectra signals (on the right). On the left of Fig. 3b, the 11 elution profiles of one of the resolved components in the different samples (1 blank, 5 control and 5 exposed samples) are given. As it is clearly shown, control samples showed considerably higher peaks than exposed samples. On the right of Fig. 3b the mass spectrum of this resolved component is given. Mass spectrum of this component gives the fragmentation pattern of a metabolite assigned to Arginine [–NH₃] (2TMS) derivative according to NIST 2014 *MS Spectral Search* program (www.nist.gov/srd/nist1a.htm). In Fig. 3c concentration and mass spectra profiles of one component related with the derivatizing agent, MSTFA ($m/z = 73$) is shown. Additional profiles related to derivatized sub-products are resolved giving peak shapes with an irregular chromatographic shape. All components of this type were finally discarded for their further metabolite identification.

Only those MCR-ALS resolved components showing reasonable chromatographic peak shapes (like the one in Fig. 3b) were considered for their identification and linked to *D. magna* metabolome.

3.3 Tentative metabolite identification

Two strategies were implemented to investigate component profiles changing with the treatments. Prior to the application of these analyses, peak areas of the same MCR-ALS component profiles in every sample and for all components and treatments were arranged in a data table for every treatment. PLS-DA was applied to these three peak area tables of MCR-ALS resolved components in control and exposed samples for each treatment (10 salinity samples \times 81 peak areas, 10 temperature samples \times 77 peak areas, and 10 hypoxia samples \times 70 peak areas). In the case of salinity a single latent variable was selected explaining 67.53 and 96.45 % of

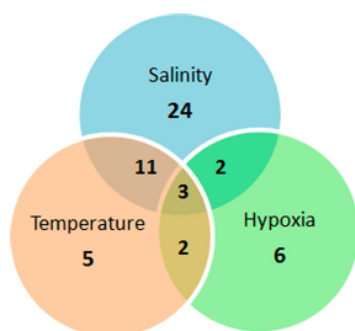


Fig. 4 Venn diagram of the metabolites detected in every treatment. Salinity factor is represented by 24, temperature by 5 and hypoxia by 6 unique metabolites, respectively. Salinity and temperature have 11 metabolites in common, whereas temperature and hypoxia, and salinity and hypoxia have only 2 metabolites coincident for each

the **X**- and **y**-variances respectively. In the case of temperature, four latent variables were selected explaining 82.28 and 98.10 % of the **X**- and **y**-variances respectively, and in the case of hypoxia, four latent variables were selected to explain 73.07 and 98.33 % of the **X**- and **y**-variances respectively. In all cases, the number of latent variables used for the PLS-DA model were much lower than the number of samples used in the analyses.

PLS-DA VIP scores provided the most important variables (resolved peaks) (see Online Resource 1) that help to discriminate between control and exposed samples, with a threshold above 1. For salinity, temperature and hypoxia 50 of the 81 resolved peaks, 27 of the 77 resolved peaks and 24 of the 70 resolved peaks, respectively (Online Resource 2), were determined as potential biomarkers (metabolites with significant concentration changes). Peak areas of MCR-ALS resolved profiles of control and exposed samples were compared using Student's *t* test, which confirmed that 66, 26 and 16 peaks were changing significantly ($p < 0.05$) for salinity, temperature and hypoxia, respectively (Online Resource 2). Finally, only the peaks chosen by both approaches were finally selected to discard false positives. Numbers of selected biomarkers determined for each treatment agreed well with sample discrimination already observed in PCA. As with PCA, there was a good separation of samples of the control vs exposed samples for salinity and temperature treatments and a poor separation for hypoxia treatment.

Tentative identification of metabolites for salinity, temperature and hypoxia treatments using resolved mass spectra and NIST library spectra matching are shown in Online Resource 3, 4 and 5, respectively. Most of the metabolites were identified with high RMF value in their evaluation in NIST library hits. Metabolite identifications for salinity, temperature and hypoxia treatments did not take into account that the derivatization with MSTFA produced different products from the same metabolite. For

instance, Lysine (3TMS) and Lysine (4TMS) (Online Resource 3) were found as biomarkers for salinity treatment, both corresponding to the same metabolite (Lysine). Finally, a total number of 74 potential metabolites whose concentrations (peak areas) changed with the treatment were identified, 40, 21 and 13 metabolites for salinity, temperature and hypoxia treatments, respectively.

For the salinity treatment (Online Resource 3) most of the 50 metabolites detected as potential biomarkers were identified with RMF values higher than 715, which can be considered a relatively fair matching factor. For most metabolites the difference between the theoretical and the experimental RI, was lower than 148, except for the resolved peak 52. This peak 52 was identified as Galactose, and it had a difference between experimental and theoretical RI of 205. However, this (Galactose) peak had a very good RMF value (927) showing a very good mass spectral coincidence with the theoretical mass values. For the salinity treatment, most of the identified metabolites were amino acids, carbohydrates, organic acids and nucleosides. Most of these metabolites were down-regulated, except 6 metabolites peak areas, 14, 55, 65, 73, 75 and 76 which were up-regulated.

For the temperature treatment (Online Resource 4), the identification of the 25 metabolites was reached with RMF values higher than 747, considered as fair matches, and with differences between theoretical and experimental RI lower than 52. The theoretical RI value of the resolved peak 76 could not be calculated because of the *n*-alkane standard only permitted to adjust up to a value of RI of 3000, which corresponds to the alkane C₃₀. Carbohydrates, amino acids and organic acids were the most commonly detected metabolites for changes in the temperature. Only two metabolites were up-regulated: 1,3-diaminopropane and putrescine.

Twelve of the 13 profiles with peak areas changing with the hypoxia treatment were identified (Online Resource 5). This identification of metabolites were reached with a difference of theoretical and experimental RI lower than 34, and with a RMF value higher than 826, considered as good matches. Carbohydrates, amino acids and organic acids were the most commonly detected metabolites. All biomarkers detected were down-regulated, except resolved peak number 17 which was up-regulated.

3.4 Metabolic pathways involved

After identifying potential biomarkers of salinity, temperature and hypoxia treatments, these metabolites were investigated in KEGG database to detect the possible affected metabolic pathways. In Fig. 4, the Venn diagram shows the metabolites identified, commons and unique, in each treatment. Salinity treatment had 24 unique metabolites identified, temperature 5, and hypoxia 6 unique metabolites. By comparing common metabolites for salinity and temperature, 11 metabolites were

Table 2 Most disrupted KEGG metabolic pathways following exposures of *D. magna* to salinity, high temperature and hypoxia. Unique metabolites from Online Resources 3, 4 and 5 were included in the metabolic KEGG pathway analysis. Total number of unique metabolites in KEGG pathway database, N number of altered metabolites, % is percentage of pathway coverage (N/total)

	Total	Salinity			Temperature			Hypoxia		
		N	%	Metabolites	N	%	Metabolites	N	%	Metabolites
ko00970 Aminoacyl-tRNA biosynthesis	23	14	60.9	GLU, LYS, MET, GLY, ARG, TRP, ALA, SER, TYR, PRO, HIS, LEU, VAL, THR, ASP	4	17.4	LYS, PRO, ASP, SER	3	13.0	LYS, SER, TYR
ko00290 Valine, leucine and isoleucine biosynthesis	18	3	16.7	LEU, VAL, THR						
ko00061 Fatty acid biosynthesis	10	1	10.0	C16						
ko00220 Arginine biosynthesis	24	4	16.7	GLU, ARG, ORN, FMR	2	8.3	FMR, ASP			
ko00280 Valine, leucine and isoleucine degradation	18	2	11.1	LEU, VAL						
ko00020 Citrate cycle (TCA cycle)	23	2	8.7	FMR, MAL	2	8.7	FMR, MAL	3	13.0	FMR, MAL, CIT
ko00260 Glycine, serine and threonine metabolism	44	5	11.4	GLY, SER, TRP, THR, DAP	4	9.1	SER, ASP, DAP, GLE	1	2.3	SER
ko00052 Galactose metabolism	45	3	6.7	GLC, GOL, GAL	3	6.7	GLC, GAL, MYO			
ko00480 Glutathione metabolism	40	4	10.0	GLU, GLY, ORN, PYROGLU	1	2.5	PUT	1	2.5	PYROGLU
ko00250 Alanine, aspartate and glutamate metabolism	30	3	10.0	GLU, ALA, FMR	2	6.7	FMR, ASP	2	6.7	FMR, CIT
ko00500 Starch and sucrose metabolism	59	4	6.8	GLC, XYL, MALT, TRE	2	3.4	GLC, XYL			
ko00620 Pyruvate metabolism	28	2	7.1	FMR, MAL	2	7.1	FMR, MAL	2	7.1	FMR, MAL

ALA alanine, ARG arginine, ASP aspartate, C16 hexadecanoic acid, CIT citrate, DAP 1,3-diaminopropane, FMR fumarate, GAL galactose, GLC glucose, GLE glycerate, GLU glutamate, GLY glycine, GOL glycerol, HIS histidine, LEU leucine, LYS lysine, MAL malate, MALT maltose, MET methionine, MYO myo-inositol, ORN ornithine, PRO proline, PUT putrescine, PYROGLU pyroglutamic acid, SER serine, THR threonine, TRE trehalose, TRP tryptophan, TYR tyrosine, VAL valine, XYL xylose

affected in both treatments. Temperature and hypoxia and, salinity and hypoxia had 2 metabolites in common in each pair of treatments.

In Table 2 specific enriched KEGG pathways affected by two or more metabolites were depicted. Metabolic pathways were sorted considering the percentage coverage of the pathway (the ratio between the altered metabolites and the total unique metabolites in the KEGG pathway database). The top three KEGG pathways represented in Table 2 coincided with three of the top five pathways described in *D. magna* exposed to Cd (Poynton et al. 2011). In this previous study, metabolomic changes were determined using different tissues and analytical methods: hemolymph samples and using an FT-ICR mass spectrometer, respectively. This may indicate that many of the obtained altered metabolites were unspecific

and probably related with stress. The three tested treatments decreased the relative amount of two to three metabolites from the pyruvate metabolism and Cytrate or Krebs cycle, which are considered critical energy metabolic pathways under hypoxia or normoxia. Indeed individuals of *D. magna*, exposure to high salinity levels supress energy metabolic rates and oxygen demand (Arnér and Koivisto 1993). Increasing temperature raised up metabolic demands for oxygen in *Daphnia*, which causes hypoxia (Paul et al. 2004a; Paul et al. 2004b). *D. magna* is well adapted to hypoxia levels or anoxia, increasing its hemoglobin content and under severe conditions switching to anaerobic metabolism (Paul et al. 1998). This means that the three studied environmental stressors had in common several features related to hypoxia and as a result shared several metabolites and metabolomic pathways.

Specific metabolomic effects of salinity included increased levels of glycerol and trehalose, which are related to cell osmoregulation (Diamant et al. 2001).

4 Concluding remarks

Detection and identification of *Daphnia magna* metabolites whose concentrations suffered changes during the exposition to the three abiotic factors studied (salinity, temperature and hypoxia) was achieved. Changes on metabolite GC–MS peak areas of controls and exposed samples were statistically assessed for their discrimination. 74 metabolites were identified as possible biomarkers of the studied effects. For salinity treatment 40 metabolites were identified, for temperature and hypoxia treatments 21 and 13 metabolites were identified, respectively. The three treatments shared effects in several metabolites related with pyruvate metabolism and Citrate or Krebs cycle, which are linked to energy metabolic pathways under hypoxia. Apart from common metabolites altered, salinity treatment had impact in glycerol and trehalose, which are related to cell osmoregulation.

The methodology proposed in this work confirmed the usefulness of the combination of GC–MS and multivariate data analysis tools for metabolomics studies of *D. magna*, exposed to mild stress of salinity, temperature and hypoxia. MCR–ALS analysis is a powerful tool to resolve directly the chromatographic signals (elution and spectra profiles) of most of the constituents (metabolites) present in complex biological samples, and differentiate them from the large number of GC–MS interfering derivatizing signals.

Acknowledgments The research leading to these results received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement No. 320737.

Compliance with ethical standards

Conflict of interest Elba Garreta-Lara, Bruno Campos, Carlos Barata, Silvia Lacorte and Romà Tauler declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Adrian, R., O'Reilly, C. M., Zagarese, H., Baines, S. B., Hessen, D. O., Keller, W., et al. (2009). Lakes as sentinels of climate change. *Limnology and Oceanography*, *54*(6 part 2), 2283–2297.
- Alier, M., Felipe, M., Hernández, I., & Tauler, R. (2010). Trilinearity and component interaction constraints in the multivariate curve resolution investigation of NO and O₃ pollution in Barcelona.

- Analytical and Bioanalytical Chemistry*, *399*(6), 2015–2029. doi:10.1007/s00216-010-4458-1.
- Arbona, V., Manzi, M., Ollas, C., & Gómez-Cadenas, A. (2013). Metabolomics as a Tool to Investigate Abiotic Stress Tolerance in Plants. *International Journal of Molecular Sciences*, *14*(3), 4885.
- Arnér, M., & Koivisto, S. (1993). Effects of salinity on metabolism and life history characteristics of *Daphnia magna*. *Hydrobiologia*, *259*(2), 69–77. doi:10.1007/BF00008373.
- Azizan, K. A., Baharum, S. N., Resson, H. W., & Noor, N. M. (2012). GC–MS analysis and PLS–DA validation of the trimethyl silyl-derivatization techniques. *American Journal of Applied Sciences*, *9*(7), 1124–1136.
- Barata, C., & Baird, D. J. (1998). Phenotypic plasticity and constancy of life-history traits in laboratory clones of *Daphnia magna* straus: Effects of neonatal length. *Functional Ecology*, *12*(3), 442–452. doi:10.1046/j.1365-2435.1998.00201.x.
- Barata, C., & Baird, D. J. (2000a). Determining the ecotoxicological mode of action of chemicals from measurements made on individuals: Results from instar-based tests with *Daphnia magna* Straus. *Aquatic Toxicology*, *48*(2–3), 195–209. doi:10.1016/S0166-445X(99)00038-7.
- Barata, C., & Baird, D. J. (2000b). Determining the ecotoxicological mode of action of toxicants from measurements on individuals: results from short duration chronic tests with *Daphnia magna* Straus. *Aquatic Toxicology*, *48*, 195–209.
- Barata, C., Carlos Navarro, J., Varo, I., Carmen Riva, M., Arun, S., & Porte, C. (2005). Changes in antioxidant enzyme activities, fatty acid composition and lipid peroxidation in *Daphnia magna* during the aging process. *Comparative Biochemistry and Physiology B: Biochemistry and Molecular Biology*, *140*(1), 81–90. doi:10.1016/j.cbpc.2004.09.025.
- Bundy, J. G., Davey, M. P., & Viant, M. R. (2009). Environmental metabolomics: A critical review and future perspectives. *Metabolomics*, *5*(1), 3–21. doi:10.1007/s11306-008-0152-0.
- Coutant, C. C. (1990). Temperature-oxygen habitat for freshwater and coastal striped bass in a changing climate. *Transactions of the American Fisheries Society*, *119*(2), 240–253. doi:10.1577/1548-8659.
- De Juan, A., Jaumot, J., & Tauler, R. (2014). Multivariate curve resolution (MCR). Solving the mixture analysis problem. *Analytical Methods*, *6*(14), 4964. doi:10.1039/c4ay00571f.
- De Juan, A., Rutan, S. C., & Tauler, R. (2010). Two-way data analysis: multivariate curve resolution—iterative resolution methods. *Comprehensive Chemometrics*, *2*, 325–344.
- De Juan, A., & Tauler, R. (2007). Factor analysis of hyphenated chromatographic data: Exploration, resolution and quantification of multicomponent systems. *Journal of Chromatography A*, *1158*(1–2), 184–195. doi:10.1016/j.chroma.2007.05.045.
- De Meester, L., & Vanoverbeke, J. (1999). An uncoupling of male and sexual egg production leads to reduced inbreeding in the cyclical parthenogen *Daphnia*. *Proceedings of the Royal Society B: Biological Sciences*, *266*(1437), 2471–2477.
- De Souza, D. (2013). Detection of polar metabolites through the use of gas chromatography–mass spectrometry. In U. Roessner & D. A. Dias (Eds.), *Metabolomics tools for natural product discovery: Methods in molecular biology* (Vol. 1055, pp. 29–37). Totowa: Humana Press.
- Diamant, S., Eliahu, N., Rosenthal, D., & Goloubinoff, P. (2001). Chemical chaperones regulate molecular chaperones in vitro and in cells under combined salt and heat stresses. *Journal of Biological Chemistry*, *276*(43), 39586–39591. doi:10.1074/jbc.M103081200.
- Dunn, W. B., Broadhurst, D. I., Atherton, H. J., Goodacre, R., & Griffin, J. L. (2011). Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear

- magnetic resonance spectroscopy. *Chemical Society Reviews*, 40(1), 387–426.
- Dunn, W., Erban, A., Weber, R. J. M., Creek, D. J., Brown, M., Breitling, R., et al. (2013). Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics*, 9(Suppl.1), 44–66. doi:10.1007/s11306-012-0434-4.
- Eriksson, L., Johansson, E., Kettaneh-Wold, N., Trygg, J., Wikström, C., & Wold, S. (2006). *Multi- and megavariable data analysis: Part I: Basic principles and applications*. Umeå: Umetrics.
- Esbensen, K. H., & Geladi, P. (2010). Principal component analysis: concept, geometrical interpretation, mathematical background, algorithms, history, practice. *Comprehensive Chemometrics*, 2, 211–226.
- Farrés, M., Piña, B., & Tauler, R. (2014). Chemometric evaluation of *Saccharomyces cerevisiae* metabolic profiles using LC-MS. *Metabolomics*, 11(1), 210–224. doi:10.1007/s11306-014-0689-z.
- Fiehn, O. (2001). Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comparative and Functional Genomics*, 2(3), 155–168. doi:10.1002/cfg.82.
- Golub, G., Sølna, K., & Van Dooren, P. (2000). Computing the SVD of a general matrix product/quotient. *SIAM Journal on Matrix Analysis and Applications*, 22(1), 1–19. doi:10.1137/S0895479897325578.
- Griffiths, W. J. (2008). *Metabolomics, metabonomics, and metabolite profiling*. Cambridge: Royal Society of Chemistry.
- Gullberg, J., Jonsson, P., Nordstrom, A., Sjoström, M., & Moritz, T. (2004). Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of *Arabidopsis thaliana* samples in metabolomic studies with gas chromatography/mass spectrometry. *Analytical Biochemistry*, 331(2), 283–295. doi:10.1016/j.ab.2004.04.037.
- Harrigan, G. G., & Goodacre, R. (2003). *Metabolic profiling: its role in biomarker discovery and gene function analysis*. Boston: Kluwer Academic.
- Ikenaka, Y., Eun, H., Ishizaka, M., & Miyabara, Y. (2006). Metabolism of pyrene by aquatic crustacean *Daphnia magna*. *Aquatic Toxicology*, 80(2), 158–165. doi:10.1016/j.aquatox.2006.08.005.
- IPCC. (2014). *Climate change 2013: The physical science basis: Working group I contribution to the fifth assessment report of the Intergovernmental Panel on Climate Change*. New York: Cambridge University Press.
- Jaumot, J., de Juan, A., & Tauler, R. (2015). MCR-ALS GUI 2.0: New features and applications. *Chemometrics and Intelligent Laboratory Systems*, 140, 1–12. doi:10.1016/j.chemolab.2014.10.003.
- Kalivodová, A., Hron, K., Filzmoser, P., Najdekr, L., Janečková, H., & Adam, T. (2015). PLS-DA for compositional data with application to metabolomics. *Journal of Chemometrics*, 29(1), 21–28. doi:10.1002/cem.2657.
- Kanani, H., Chrysanthopoulos, P. K., & Klapa, M. I. (2008). Standardizing GC-MS metabolomics. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 871(2), 191–201. doi:10.1016/j.jchromb.2008.04.049.
- Kanani, H., & Klapa, M. I. (2007). Data correction strategy for metabolomics analysis using gas chromatography-mass spectrometry. *Metabolic Engineering*, 9(1), 39–51. doi:10.1016/j.ymben.2006.08.001.
- Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1), 27–30. doi:10.1093/nar/28.1.27.
- Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., et al. (2005). GMD@CSB.DB: The Golm Metabolome Database. *Bioinformatics*, 21(8), 1635–1638. doi:10.1093/bioinformatics/bti236.
- Krastanov, A. (2010). Metabolomics—the state of art. *Biotechnology and Biotechnological Equipment*, 24(1), 1537–1543. doi:10.2478/V10133-010-0001-y.
- Lindon, J. C., & Nicholson, J. K. (2008a). Analytical technologies for metabonomics and metabolomics, and multi-omic information recovery. *TrAC Trends in Analytical Chemistry*, 27(3), 194–204. doi:10.1016/j.trac.2007.08.009.
- Lindon, J. C., & Nicholson, J. K. (2008b). Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. *Annual Review of Analytical Chemistry*, 1(1), 45–69. doi:10.1146/annurev.anchem.1.031207.113026.
- Lindon, J. C., Nicholson, J. K., & Holmes, E. (2007). *The handbook of metabonomics and metabolomics*. Oxford: Elsevier.
- Little, J. L. (1999). Artifacts in trimethylsilyl derivatization reactions and ways to avoid them. *Journal of Chromatography A*, 844(1–2), 1–22. doi:10.1016/S0021-9673(99)00267-8.
- Malik, A., & Tauler, R. (2015). Exploring the interaction between O₃ and NO_x pollution patterns in the atmosphere of Barcelona, Spain using the MCR-ALS method. *Science of the Total Environment*, 517, 151–161. doi:10.1016/j.scitotenv.2015.01.105.
- Martins, J., Oliva Teles, L., & Vasconcelos, V. (2007). Assays with *Daphnia magna* and Danio rerio as alert systems in aquatic toxicology. *Environment International*, 33(3), 414–425. doi:10.1016/j.envint.2006.12.006.
- Miller, M. G. (2007). Environmental metabolomics: A SWOT analysis (strengths, weaknesses, opportunities, and threats). *Journal of Proteome Research*, 6(2), 540–545. doi:10.1021/pr060623x.
- Morrison, N., Bearden, D., Bundy, J. G., Collette, T., Currie, F., Davey, M. P., et al. (2007). Standard reporting requirements for biological samples in metabolomics experiments: Environmental context. *Metabolomics*, 3(3), 203–210. doi:10.1007/s11306-007-0067-1.
- Navarro-Reig, M., Jaumot, J., García-Reiriz, A., & Tauler, R. (2015). Evaluation of changes induced in rice metabolome by Cd and Cu exposure using LC-MS with XCMS and MCR-ALS data analysis strategies. *Analytical and Bioanalytical Chemistry*, 407(29), 8835–8847. doi:10.1007/s00216-015-9042-2.
- Nielsen, N.-P. V., Carstensen, J. M., & Smedsgaard, J. (1998). Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *Journal of Chromatography A*, 805(1–2), 17–35. doi:10.1016/S0021-9673(98)00021-1.
- OECD. (2012). OECD Guidelines for the testing of chemicals. In OECD (Ed.), *Daphnia magna Reproduction Test (211)*. Paris: OECD.
- Orata, F. (2012). *Derivatization reactions and reagents for gas chromatography analysis*. Rijeka: INTECH Open Access Publisher.
- Ortiz-Villanueva, E., Jaumot, J., Benavente, F., Piña, B., Sanz-Nebot, V., & Tauler, R. (2015). Combination of CE-MS and advanced chemometric methods for high-throughput metabolic profiling. *Electrophoresis*, 36(18), 2324–2335. doi:10.1002/elps.201500027.
- Pacchiarotta, T., Nevedomskaya, E., Carrasco-Pancorbo, A., Deelder, A. M., & Mayboroda, O. A. (2010). Evaluation of GC-APCI/MS and GC-FID as a complementary platform. *Journal of Biomolecular Techniques*, 21(4), 205–213.
- Parastar, H., Jalali-Heravi, M., Sereshti, H., & Mani-Varnosfaderani, A. (2012). Chromatographic fingerprint analysis of secondary metabolites in citrus fruits peels using gas chromatography-mass spectrometry combined with advanced chemometric methods.

- Journal of Chromatography A*, 1251, 176–187. doi:10.1016/j.chroma.2012.06.011.
- Paul, R. J., Colmorgen, M., Pirow, R., Chen, Y. H., & Tsai, M. C. (1998). Systemic and metabolic responses in *Daphnia magna* to anoxia. *Comparative Biochemistry and Physiology A: Molecular and Integrative Physiology*, 120(3), 519–530. doi:10.1016/S1095-6433(98)10062-4.
- Paul, R. J., Lamkemeyer, T., Maurer, J., Pinkhaus, O., Pirow, R., Seidl, M., et al. (2004a). Thermal acclimation in the microcrustacean *Daphnia*: A survey of behavioural, physiological and biochemical mechanisms. *Journal of Thermal Biology*, 29(7–8 SPEC. ISS), 655–662. doi:10.1016/j.jtherbio.2004.08.035.
- Paul, R. J., Zeis, B., Lamkemeyer, T., Seidl, M., & Pirow, R. (2004b). Control of oxygen transport in the microcrustacean *Daphnia*: Regulation of haemoglobin expression as central mechanism of adaptation to different oxygen and temperature conditions. *Acta Physiologica Scandinavica*, 182(3), 259–275. doi:10.1111/j.1365-201X.2004.01362.x.
- Poynton, H. C., Taylor, N. S., Hicks, J., Colson, K., Chan, S., Clark, C., et al. (2011). Metabolomics of microliter hemolymph samples enables an improved understanding of the combined metabolic and transcriptional responses of *Daphnia magna* to cadmium. *Environmental Science and Technology*, 45(8), 3710–3717.
- Samuelsson, L. M., & Larsson, D. G. J. (2008). Contributions from metabolomics to fish research. *Molecular BioSystems*, 4(10), 974–979. doi:10.1039/b804196b.
- Sardans, J., Peñuelas, J., & Rivas-Ubach, A. (2011). Ecological metabolomics: overview of current developments and future challenges. *Chemoecology*, 21(4), 191–225. doi:10.1007/s00049-011-0083-5.
- Schauer, N., Steinhauser, D., Strelkov, S., Schomburg, D., Allison, G., Moritz, T., et al. (2005). GC–MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Letters*, 579(6), 1332–1337. doi:10.1016/j.febslet.2005.01.029.
- Steinhauser, D., & Kopka, J. (2007). Methods, applications and concepts of metabolite profiling: primary metabolism. *EXS*, 97, 171–194.
- Strehmel, N., Hummel, J., Erban, A., Strassburg, K., & Kopka, J. (2008). Retention index thresholds for compound matching in GC–MS metabolite profiling. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 871(2), 182–190. doi:10.1016/j.jchromb.2008.04.042.
- Tauler, R. (1995). Multivariate curve resolution applied to second order data. *Chemometrics and Intelligent Laboratory Systems*, 30(1), 133–146. doi:10.1016/0169-7439(95)00047-X.
- Tauler, R. (2001). Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution. *Journal of Chemometrics*, 15(8), 627–646. doi:10.1002/cem.654.
- Tauler, R., & Barceló, D. (1993). Multivariate curve resolution applied to liquid chromatography-diode array detection. *Trends in Analytical Chemistry*, 12(8), 319–327. doi:10.1016/0165-9936(93)88015-W.
- Terrado, M., Barceló, D., & Tauler, R. (2009). Quality assessment of the multivariate curve resolution alternating least squares method for the investigation of environmental pollution patterns in surface water. *Environmental Science and Technology*, 43(14), 5321–5326. doi:10.1021/es803333s.
- Tomasi, G., Van Den Berg, F., & Andersson, C. (2004). Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics*, 18(5), 231–241. doi:10.1002/cem.859.
- van Stokkum, I. H. M., Mullen, K. M., & Mihaleva, V. V. (2009). Global analysis of multiple gas chromatography–mass spectrometry (GC/MS) data sets: A method for resolution of co-eluting components with comparison to MCR-ALS. *Chemometrics and Intelligent Laboratory Systems*, 95(2), 150–163. doi:10.1016/j.chemolab.2008.10.004.
- Vandenbrouck, T., Jones, O. A., Dom, N., Griffin, J. L., & De Coen, W. (2010). Mixtures of similarly acting compounds in *Daphnia magna*: From gene to metabolite and beyond. *Environment International*, 36(3), 254–268. doi:10.1016/j.envint.2009.12.006.
- Viant, M. R. (2008). Recent developments in environmental metabolomics. *Molecular BioSystems*, 4(10), 980–986. doi:10.1039/b805354e.
- Villas-Boas, S. G., Smart, K. F., Sivakumaran, S., & Lane, G. A. (2011). Alkylation or Silylation for Analysis of Amino and Non-Amino Organic Acids by GC–MS? *Metabolites*, 1(1), 3–20. doi:10.3390/metabo1010003.
- Wehrens, R., Weingart, G., & Mattivi, F. (2014). MetaMS: An open-source pipeline for GC–MS-based untargeted metabolomics. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 966, 109–116. doi:10.1016/j.jchromb.2014.02.051.
- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1–3), 37–52. doi:10.1016/0169-7439(87)80084-9.
- Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109–130. doi:10.1016/S0169-7439(01)00155-1.
- Yi, L., Shi, S., Yi, Z., He, R., Lu, H., & Liang, Y. (2014). MeOx-TMS derivatization for GC–MS metabolic profiling of urine and application in the discrimination between normal C57BL/6 J and type 2 diabetic KK-Ay mice. *Analytical Methods*, 6(12), 4380. doi:10.1039/c3ay41522h.

Online Resources

Metabolic profiling of *Daphnia magna* exposed to environmental stressors by GC-MS and chemometric tools

Elba Garreta-Lara, Bruno Campos, Carlos Barata, Silvia Lacorte, Romà Tauler*

Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, 08034, Catalonia, Spain.

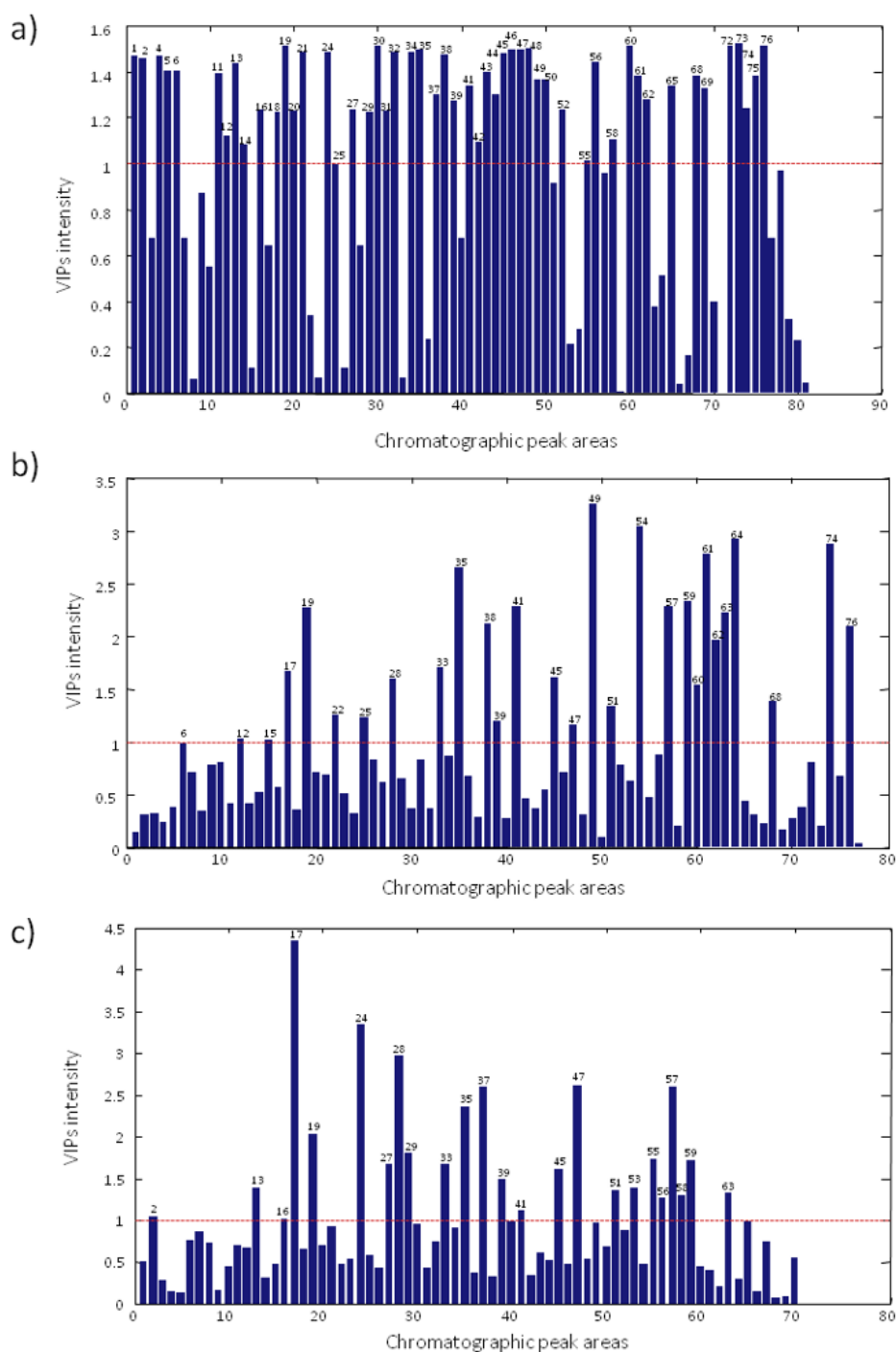
*Corresponding author: Romà Tauler

Postal address: Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, 08034, Catalonia, Spain.

Telephone: +34934006140

Fax: +34932045904

E-mail: roma.tauler@idaea.csic.es



Online Resource 1. Plot of PLS-DA Variables Importance in Projection (VIP) scores of the autoscaled MCR-ALS peak areas. **(a)** Controls vs salinity exposed samples; **(b)** controls vs temperature exposed samples; and **(c)** controls vs hypoxia exposed samples. Peak areas with a threshold value (horizontal red line) above 1 are considered as potential biomarkers of each treatment.

Online Resources

Metabolic profiling of *Daphnia magna* exposed to environmental stressors by GC-MS and chemometric tools

Elba Garreta-Lara, Bruno Campos, Carlos Barata, Silvia Lacorte, Romà Tauler*

Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, 08034, Catalonia, Spain.

*Corresponding author: Romà Tauler

Postal address: Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, 08034, Catalonia, Spain.

Telephone: +34934006140

Fax: +34932045904

E-mail: roma.tauler@idaea.csic.es

Online Resource 2. Potential biomarkers detected for the three treatments (salinity, temperature and hypoxia) with VIP scores above 1, or with a significance p value < 0.05 in the Student's t-test in the comparison of mean peak areas between control and exposed samples.

	Salinity	Temperature	Hypoxia
PLS-DA (VIPs > 1)	1, 2, 4, 5, 6, 11, 12, 13, 14, 16, 18, 19, 20, 21, 24, 27, 29, 30, 31, 32, 34, 35, 37, 38, 39, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 55, 56, 58, 60, 61, 62, 65, 68, 69, 72, 73, 74, 75, 76	12, 15, 17, 19, 22, 25, 28, 33, 35, 38, 39, 41, 45, 47, 49, 51, 54, 57, 59, 60, 61, 62, 63, 64, 68, 74, 76	2, 13, 16, 17, 19, 24, 27, 28, 29, 33, 35, 37, 39, 41, 45, 47, 51, 53, 55, 56, 57, 58, 59, 63
Student t-test ($p < 0.05$)	1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 22, 24, 25, 27, 28, 29, 30, 31, 32, 34, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 55, 56, 57, 58, 60, 61, 62, 63, 64, 65, 68, 69, 70, 72, 73, 74, 75, 76, 77, 78	6, 15, 17, 19, 22, 25, 28, 33, 35, 38, 39, 41, 45, 47, 49, 54, 57, 59, 60, 61, 62, 63, 64, 68, 74, 76	13, 17, 19, 24, 28, 29, 35, 37, 39, 41, 45, 47, 51, 56, 57, 63
Common resolved peaks	1, 2, 4, 5, 6, 11, 12, 13, 14, 16, 18, 19, 20, 21, 24, 27, 29, 30, 31, 32, 34, 35, 37, 38, 39, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 55, 56, 58, 60, 61, 62, 65, 68, 69, 72, 73, 74, 75, 76 (50 out of 66)	15, 17, 19, 22, 25, 28, 33, 35, 38, 39, 41, 45, 47, 49, 54, 57, 59, 60, 61, 62, 63, 64, 68, 74, 76 (26 out of 27)	13, 17, 19, 24, 28, 29, 35, 37, 39, 41, 45, 47, 51, 56, 57, 63 (16 out of 24)

Numbers in the table refer to MCR-ALS resolved peak areas in every treatment.

Online Resources

Metabolic profiling of *Daphnia magna* exposed to environmental stressors by GC-MS and chemometric tools

Elba Garreta-Lara, Bruno Campos, Carlos Barata, Silvia Lacorte, Romà Tauler*

Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, 08034, Catalonia, Spain.

*Corresponding author: Romà Tauler

Postal address: Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, 08034, Catalonia, Spain.

Telephone: +34934006140

Fax: +34932045904

E-mail: roma.tauler@idaea.csic.es

Online Resource 3. Tentative identification of all peaks resolved by MCR-ALS to be considered as potential biomarkers of the salinity effect on *D. magna* metabolome.

Peak number	tR (min)	RI theoretical	RI experimental	^a (RI theo - RI exp)	Metabolite	Derivatization	KEGG-ID	RMF	% Change	Trend
1	8.57	1217	1224	-7	L-Valine	(2TMS)	C00183	821	154.2	Down
2	7.87	1173	1167	6	L-Leucine	(TMS)	C00123	802	206.8	Down
4	7.4	1141	1105	36	Glycine	(2TMS)	C00037	865	135.3	Down
5	5.62	1022	874	148	Alanine	(TMS)	C00041	811	98.8	Down
6	7.07	1112	1094	18	L-Valine	(TMS)	C00183	715	98.9	Down
11	9.2	1256	1252	4	L-Serine	(3TMS)	C00065	830	118.2	Down
12	9.83	1296	1305	-9	L-Proline	(2TMS)	C00148	841	42.6	Down
13	9.77	1292	1290	2	L-Threonine	(2TMS)	C00188	848	108.0	Down
14	9.4	1269	1262	7	Glycerol	(3TMS)	C00116	842	42.5	Up
16	10.47	1431	1335	96	Uracil	(2TMS)	C00106	884	60.7	Down
18	10.7	1444	1359	85	Fumaric acid	(2TMS)	C00122	932	38.1	Down
19	11.83	1506	1416	90	L-Methionine	(TMS)	C00073	871	446.3	Down
20	11.28	1475	1367	108	L-Threonine	(3TMS)	C00188	897	151.6	Down
21	11	1460	None	None	None	None	None	None	369.1	Down
24	12.05	1517	None	None	None	None	None	None	343.8	Down
27	15.93	1709	1654	55	L-Glutamic acid	(3TMS)	C00025	950	60.7	Down
29	13.92	1611	1522	89	L-5-Oxoproline	(2TMS)	C01879	840	38.1	Down
30	17.1	1766	1669	97	Xylose	(MeOX)(4TMS)	C00181	903	446.3	Down
31	13.25	1577	1492	85	Malic acid	(3TMS)	C00149	961	151.6	Down
32	15.13	1670	1592	78	L-Proline	(3TMS)	C00148	945	369.1	Down
34	15.87	1706	1614	92	L-Glutamic acid	(3TMS)	C00025	943	343.8	Down
35	16.37	1730	1645	85	1,3-Diaminopropane	(4TMS)	C00986	896	118.1	Down
37	17.72	1796	1742	54	L-Lysine	(3TMS)	C00047	870	449.6	Down
38	18.45	1831	1742	89	L-Arginine [-NH3]	(2TMS)	C00062	957	266.2	Down
39	19.92	1903	1792	111	DL-Ornithine	(4TMS)	C00077	823	201.3	Down
41	20.22	1918	None	None	None	None	None	None	57.8	Down
42	21.9	1899	1901	-2	Glucose	(MeOX)(5TMS)	C00031	812	29.0	Down
43	21.3	1972	1845	127	Tyrosine	(2TMS)	C00082	799	138.9	Down
44	20.65	1939	1849	90	Lysine	(3TMS)	C00047	891	102.1	Down
45	22.07	1990	1920	70	Lysine	(4TMS)	C00047	859	1265.5	Down
46	22.35	1975	1952	23	Tyrosine	(3TMS)	C00082	864	534.7	Down
47	21.93	1997	1942	55	Histidine	(3TMS)	C00135	853	2082.3	Down
48	21.58	1986	None	None	None	None	None	None	198.7	Down
49	20.65	1939	1849	90	Lysine	(3TMS)	C00047	891	157.5	Down
50	21.55	1984	None	None	None	None	None	None	66.8	Down
52	24.12	2117	1912	205	Galactose	(MeOX)(5TMS)	C00124	927	54.5	Down
55	24.47	2135	2050	85	Palmitic acid	(TMS)	C00249	913	34.9	Up
56	24.53	2139	None	None	None	None	None	None	367.7	Down
58	25.52	2192	None	None	None	None	None	None	34.8	Down
60	27.12	2281	2215	66	Tryptophan	(2TMS)	C00078	886	1172.3	Down
61	26.45	2244	2185	59	Tryptophan	(TMS)	C00078	870	452.7	Down

62	25.95	2216	None	None	None	None	None	None	103.2	Down
65	27.53	2304	2219	85	Linoleic acid	(TMS)	C01595	935	162.53	Up
68	31.32	2529	2499	30	Uridine	(3TMS)	C00299	905	221.0	Down
69	33.32	2657	2596	61	Inosine	(4TMS)	C00294	923	259.6	Down
72	36.05	2841	2772	69	Guanosine	(4TMS)	C00387	923	376.6	Down
73	35.8	2823	2743	80	Trehalose	(8TMS)	C01083	854	654.2	Up
74	36.28	2857	None	None	None	None	None	None	909.6	Down
75	35.82	2824	2743	81	Trehalose	(8TMS)	C01083	850	610.0	Up
76	35.65	2813	2707	106	Maltose	(8TMS)	C00208	925	1057.6	Up

None: Unknown identified metabolite.

^a: the negative sign does not have any numerical significance.

Online Resources

Metabolic profiling of *Daphnia magna* exposed to environmental stressors by GC-MS and chemometric tools

Elba Garreta-Lara, Bruno Campos, Carlos Barata, Silvia Lacorte, Romà Tauler*

Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, 08034, Catalonia, Spain.

*Corresponding author: Romà Tauler

Postal address: Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, 08034, Catalonia, Spain.

Telephone: +34934006140

Fax: +34932045904

E-mail: roma.tauler@idaea.csic.es

Online Resource 4. Tentative identification of all peaks resolved by MCR-ALS to be considered as potential biomarkers of the temperature effect on *D. magna* metabolome.

Peak number	tR (min)	RI theoretical	RI experimental	^a (RI theo - RI exp)	Metabolite	Derivatization	KEGG-ID	RMF	% Change	Trend
15	9.4	1269	1281	-12	Phosphoric acid	(3TMS)	C00009	852	23.5	Down
17	9.2	1257	1252	5	L-Serine	(2TMS)	C00065	788	19.5	Down
19	10.47	1331	1343	-12	Uracil	(2TMS)	C00106	888	27.6	Down
22	10.72	1345	1353	-8	Fumaric acid	(2TMS)	C00122	930	21.7	Down
25	12.05	1417	1420	-3	L-Aspartic acid	(2TMS)	C00049	747	16.0	Down
28	10.31	1323	1340	-17	Glyceric acid	(3TMS)	C00258	941	45.2	Down
33	17.1	1666	1669	-3	Xylose	(MeOX)(4TMS)	C00181	901	22.2	Down
35	16.37	1630	1645	-15	1,3-Diaminopropane	(4TMS)	C00986	871	30.5	Up
38	13.27	1478	1492	-14	Malic acid	(3TMS)	C00149	960	35.5	Down
39	15.13	1570	1592	-22	L-Proline	(2TMS)	C00148	927	21.2	Down
41	18.28	1723	1734	-11	Putrescine	(4TMS)	C00134	925	43.3	Up
45	17.72	1696	1702	-6	L-Lysine	(3TMS)	C00047	857	62.0	Down
47	20.22	1817	None	None	None	None	None	None	17.5	Down
49	21.58	1886	1883	3	D-Glucose	(MeOX)(5TMS)	C00031	791	56.9	Down
54	21.57	1885	1876	9	D-Galactose	(MeOX)(5TMS)	C00124	813	46.4	Down
57	21.9	1902	1871	31	Idose	(5TMS)	C06466	813	24.7	Down
59	24.12	2017	1970	47	Glucose	(5TMS)	C00031	791	38.6	Down
60	25.15	2072	2090	-18	Myo-inositol	(6TMS)	C00137	963	31.5	Down
61	24.97	2063	2072	-9	N-acetyl-D-glucosamine	(MeOX)(4TMS)	C00140	937	88.2	Down
62	24.53	2039	2026	13	Octopamine	(4TMS)	C04227	811	42.1	Down
63	24.97	2063	None	None	None	None	None	None	39.5	Down
64	24.12	2017	None	None	None	None	None	None	79.3	Down
68	26.05	2121	2173	-52	D-Galactose	(5TMS)	C00124	868	31.1	Down
74	36.05	2740	2772	-32	Guanosine	(4TMS)	C00387	926	30.7	Down
76	46.38	*None	3549	None	Maltotriose	(MeOX)(11TMS)	C01835	916	47.0	Down

*The alkane standard only permits to calculate up to a RI value of 3000, corresponding to the alkane C₃₀.

None: Unknown identified metabolite.

^a: the negative sign does not have any numerical significance.

Online Resources

Metabolic profiling of *Daphnia magna* exposed to environmental stressors by GC-MS and chemometric tools

Elba Garreta-Lara, Bruno Campos, Carlos Barata, Silvia Lacorte, Romà Tauler*

Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, 08034, Catalonia, Spain.

*Corresponding author: Romà Tauler

Postal address: Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, 08034, Catalonia, Spain.

Telephone: +34934006140

Fax: +34932045904

E-mail: roma.tauler@idaea.csic.es

Online Resource 5. Tentative identification of all peaks resolved by MCR-ALS to be considered as potential biomarkers of the hypoxia effect on *D. magna* metabolome.

Peak number	tR (min)	RI theoretical	RI experimental	^a (RI theo - RI exp)	Metabolite	Derivatization	KEGG-ID	RMF	% Change	Trend
13	9.18	1261	1252	9	L-Serine	(2TMS)	C00065	833	7.8	Down
17	11	1366	None	None	None	None	None	None	11.1	Up
19	10.7	1350	1353	-3	Fumaric acid	(2TMS)	C00122	917	24.3	Down
24	13.88	1515	1522	-7	L-5-Oxoproline	(2TMS)	C01879	915	17.3	Down
28	13.52	1497	1511	-14	L-5-Oxoproline	(TMS)	C01879	874	16.8	Down
29	13.25	1484	1491	-7	Malic acid	(3TMS)	C00149	964	26.6	Down
35	18.85	1757	1775	-18	Glycerol-3-phosphate	(4TMS)	C00093	976	13.6	Down
37	17.12	1673	1654	19	D-Arabinose	(MeOX)(4TMS)	C00259	845	24.6	Down
39	18.88	1759	1749	10	Tetraethylene glycol	(2TMS)	C01380	826	9.4	Down
41	18.87	1758	1780	-22	Glycerol-3-phosphate	(4TMS)	C00093	842	21.0	Down
45	21.3	1879	1845	34	Tyrosine	(2TMS)	C00082	878	15.1	Down
47	20	1814	1827	-13	Citric acid	(4TMS)	C00158	959	31.7	Down
51	20.65	1846	1849	-3	L-Lysine	(3TMS)	C00047	877	8.2	Down
56	24.98	2071	2068	3	N-acetyl-D-glucosamine	(MeOX)(4TMS)	C00140	898	21.3	Down
57	23.37	1984	1997	-13	Panhotenic acid	(3TMS)	C00864	879	26.5	Down
63	33.49	2577	2609	-32	Adenosine	(3TMS)	C00212	921	10.2	Down

None: Unknown identified metabolite.

^a: the negative sign does not have any numerical significance.

2.3 ARTICLE CIENTÍFIC II

Combined effects of salinity, temperature and hypoxia on Daphnia magna metabolism.

E. Garreta-Lara, B. Campos, C. Barata, S. Lacorte, R. Tauler.

Science of the Total Environment 610-611 (2018), 602-612.



Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv



Combined effects of salinity, temperature and hypoxia on *Daphnia magna* metabolism



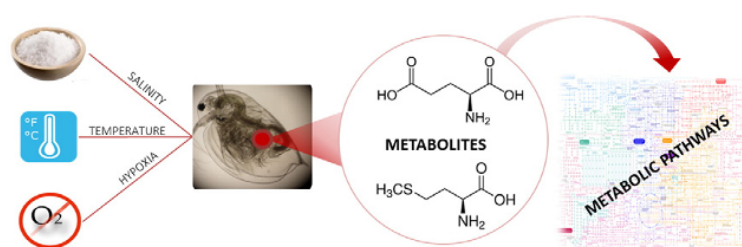
Elba Garreta-Lara, Bruno Campos, Carlos Barata, Silvia Lacorte, Romà Tauler*

Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona 08034, Catalonia, Spain

HIGHLIGHTS

- Study of the effects of salinity, temperature and hypoxia on *D. magna* metabolism
- Evaluation of changes on *D. magna* metabolome by chemometric methods
- Salinity was the most influential factor on *D. magna* metabolome.
- Energy metabolism pathways were altered by salinity.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 4 April 2017
 Received in revised form 18 May 2017
 Accepted 20 May 2017
 Available online 17 August 2017

Editor: D. Barcelo

Keywords:

Experimental design
 Abiotic factors
 GC-MS
 Metabolomics
Daphnia magna
 Chemometrics

ABSTRACT

Metabolic changes of *Daphnia magna* pools due different abiotic factors linked to global climate change (salinity, temperature and hypoxia) were investigated using untargeted GC-MS and advanced chemometric strategies using a three factors two-level full factorial experimental design (DoE). Effects of these three factors and identity of the metabolites whose concentrations changed because of them were investigated. The simultaneous analysis of GC-MS data sets using Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) allowed the resolution of the elution and mass spectra profiles of a large number of *D. magna* metabolites. Changes in peak areas of these metabolites were then analyzed by Principal Component Analysis (PCA), by ANOVA-Simultaneous Component Analysis (ASCA) and by Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA), and the combined effects of the three investigated stressors were assessed. Results confirmed the strong influence of increasing environmental salinity levels on the *D. magna* metabolome. This impact was specially highlighted by changes on the cellular content of carbohydrates, fatty acids, organic acids and amino acid molecules. In contrast, these effects were less significant for the other two factors (temperature and hypoxia) at the moderate stressing experimental conditions investigated in this work when they were not combined with salinity.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The ecological impact of stressors related with climate change is a recognized fact, although there is little information on the effects of some

abiotic factors towards the organism and ecosystems integrity. The general effects of climate change on freshwater systems will likely increase water temperatures, change salinity contents, decrease dissolved oxygen levels, and increase toxicity of pollutants (Ficke et al., 2007). In the last years, the exposure of organisms to natural stressors is receiving attention in risk assessment studies (Bergman Filho et al., 2011).

Apart from the interaction between organisms and environmental systems, it is necessary to understand the interactions between

* Corresponding author.

E-mail address: roma.tauler@idaea.csic.es (R. Tauler).

stressors, and how individual and combinations of factors affect organisms. Statistical design of experiments (DoE) is a common tool to maximize the information obtained. Hence, the main goal of this approach is to generate the most useful information focusing on the investigation of the effects by means of their interaction. Using a full-factorial design every possible combination of factors at the designated levels is performed, resulting L^k possible combinations of k factors at L levels (Hibbert, 2012).

Metabolomics can provide a snapshot of the molecular events triggered by a variation of an environmental condition. The study of the variation in metabolome permits to obtain information regarding the status of the organism. Characterizing the metabolic responses of organisms to climate change stressors will highlight the power of metabolomics to reveal the interaction of biological organisms with their natural environment (Viant, 2008). To fully determine the metabolome of a biological organism, non-targeted analytical approaches allow the determination of a large number of metabolites, collecting as much information as possible (Alonso et al., 2015).

The crustacean *Daphnia magna* is a model organism in ecotoxicology, because of its easily culture, rapid growth and defined constant genetic background with a clonal reproduction. Changes in the metabolome of *D. magna* and related species have been studied across metals, polycyclic aromatic hydrocarbons, pharmaceuticals, pesticides, flame-retardants, perfluorated compounds and nanoparticles using Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS), nuclear magnetic resonance (NMR), GC-Q-TOF/MS and LC-Q-TOF/MS (Kovacevic et al., 2016; Li et al., 2015; Nagato et al., 2013; Nagato et al., 2016; Poynton et al., 2011; Scanlan et al., 2015; Taylor et al., 2016a; Taylor et al., 2016b; Taylor et al., 2009; Taylor et al., 2010; Toyota et al., 2016; Vandembrouck et al., 2010; Wagner et al., 2016; Zhang et al., 2017). In most studies reported above, changes in metabolomic profiles alone or in combination with transcriptomic ones, allowed to identify pollutant-specific biomarkers and hence differentiate modes of toxicity across the studied chemicals. The study of metabolic changes along the life-cycle or across natural stressors, such as low food supply, high salinity, high temperature and low oxygen levels, allowed to identify common and dissimilar metabolic pathways across natural occurring stressors (Garreta-Lara et al., 2016; Jones et al., 2012; Wagner et al., 2015; Zhang et al., 2017). In only one study, effects of mixtures were addressed although interactive effects were not statistically tested (Vandembrouck et al., 2010). Recently, multivariate ASCA methods were applied on the lipidomic profile of *D. magna* exposed to TBT during its reproductive cycle. This allowed testing statistically that TBT and the reproductive cycle affected independently and additively the lipidome of *D. magna* (Malik et al., 2016).

In our previous study, we assessed metabolomics changes of *D. magna* pools exposed acutely to sublethal levels of high salinity and temperature and low dissolved oxygen levels, which produced disturbances in the energy metabolic pathways (Garreta-Lara et al., 2016). However, the combination of these factors was not evaluated. Following this line of research a new study was proposed in order to achieve broader perspectives on the impact of these abiotic factors and about how their combination can affect the metabolome. The present work will evaluate the combined effects of salinity, temperature and hypoxia on *D. magna* metabolome, to anticipate the effects of these physical stressors on their metabolism and to identify the metabolites affected by these environmental stressors.

2. Materials and methods

2.1. Chemicals and standard solutions

D-glucose (U-13C6, 99%), used as the Internal Standard (IS), was supplied by Cambridge Isotope Laboratories, Inc. (Andover, MA, USA).

Pyridine (anhydrous, 99.8%), derivatizing agents: methoxyamine hydrochloride (98%) (MeOX), chlorotrimethylsilane (TMCS) and *N*-methyl-*N*-trimethylsilyl trifluoroacetamide (>98.5%) (MSTFA), and the saturated Alkane standard mixture for the performance test of GC systems from C₇ to C₃₀ were also obtained from Sigma-Aldrich (St. Louis, MO, USA). Hexane, methanol and chloroform were analytical reagent grade, and sodium chloride (NaCl) salt was supplied by Merck (Darmstadt, Germany).

2.2. Sample preparation and experimental design

2.2.1. Experimental animals

All experiments were performed using a well characterized clone F of *D. magna* maintained indefinitely as pure parthenogenetic cultures (Barata and Baird, 1998). Photoperiod was set to 16 h light: 8 h dark cycle and temperature at 20 ± 1 °C. Bulk cultures of 10 animals L⁻¹ were maintained in ASTM International hard water at high food ration levels (5×10^5 cells mL⁻¹ of *Chlorella vulgaris*, respectively), as described in Barata and Baird (1998). The culture medium was changed every other day, and neonates were removed within 24 h.

2.2.2. Design of experiments (DoE)

Four days old *D. magna* juveniles were used. Animals were obtained from bulk cultures of 100 individuals reared in 10 L of ASTM as described above ("Experimental animals"). A two-level full factorial experiment with 3 factors (salinity (0 and 5 g L⁻¹ NaCl), temperature (20 and 25 °C) and hypoxia (9 and 2 mg L⁻¹ O₂)), 5 treatment replicate samples at each factor level and their combination was designed statistically (DOE) (Hibbert, 2012). These levels were selected to not affect the survival of the organisms. All treatment replicate samples implied independent *D. magna* sample pools, independent sample extractions and independent GC-MS analysis. Treatments with all factors at a low-level: no-salt with low levels of temperature and oxygen represent the control samples (C).

In the last 50 years, global average surface temperature increased by 0.10–0.16 °C per decade and, it is projected to rise by 1.8–4.0 °C by 2099 (relative to 1999) (IPCC, 2007). Recent advancing industrialization and urbanization have increased salt concentrations in formerly freshwater habitats. *D. magna* is mainly recognized as a freshwater cladoceran, but there are some strains that grow in brackish waters having salinities up to 12.5 g L⁻¹ (Arnér and Koivisto, 1993). Natural populations of *D. magna* often migrate to deep hypoxic waters during the day to avoid fish predation, and then swim to surface water to graze on microalgae during the night (Cousyn et al., 2001). Thus, *daphnids* are physiologically well adapted to abrupt changes of oxygen levels (Paul et al., 1998; Zeis et al., 2009). This means that the selected temperature, increase of 5 °C, is close to that predicted from global climate change. Also, the tested levels of salinity (5 g L⁻¹) and low oxygen levels are similar to those that field *D. magna* populations may encounter.

Animals were exposed to individual factors: high salinity, 5 g L⁻¹ NaCl (S), high temperature, 25 °C (T), and hypoxia conditions, around 2 mg O₂ L⁻¹ (O), and to their binary combinations at high levels: high salinity with high temperature (ST), high salinity with hypoxia (SO) and high temperature with hypoxia conditions (TO). Finally, a tertiary mixture of the three factors was also tested at their high levels: high salinity with high temperature and hypoxia conditions (exposed samples (STO)). Randomly collected animals were exposed to different treatments according to the experimental design in groups of 10 individuals in 1.2 L glass jars completely filled with ASTM hard water. Based on a two-level three factors full factorial design (2³), a total number of 8 independent treatments were done.

Salinity exposed samples were prepared with 5 g L⁻¹ NaCl in ASTM hard water, high temperature samples were prepared in a water bath at 25 °C and hypoxia samples at ≤2 mg O₂ L⁻¹ were prepared by bubbling liquid nitrogen into the ASTM hard water thus removing oxygen from it. The oxygen levels were controlled at the beginning and at the end of the exposure by an oxygen meter (Model 782, Strathkelvin Instruments, Glasgow) provided with an oxygen electrode (Model 1302, Strathkelvin Instruments, Glasgow). Approximately, the levels had to be ≤2 mg L⁻¹ in every sample. After 24 h of exposure, which represents an acute toxicity condition, *daphnids* were randomly collected from each jar, pooled in an Eppendorf, then the water was removed and samples were flash frozen in liquid N₂ and stored at -80 °C until metabolite extraction.

2.2.3. Extraction and derivatization of metabolites

Polar metabolites of the whole organism were extracted using liquid-solid extractions. *Daphnids* were suspended with 400 µL methanol, containing 10 µL of the Internal Standard (IS, D-glucose-U-13C6, 99%) at a concentration of 50 µg mL⁻¹, then samples were vortexed for 1 min and homogenized by sonication for 15 min. Following homogenization, 200 µL water and 400 µL chloroform were added and then samples were centrifuged at 10,000g during 15 min to separate the aqueous and the lipid phase. The aqueous phase was then transferred to a new Eppendorf and the extracts were evaporated to dryness with a Speedvac (Thermo Scientific) at 40 °C during 3 h approximately. When samples were completely dry they were stored at -80 °C until their GC-MS analysis. The quality control in the GC-MS analysis was done by the use of D-glucose (U-13C6, 99%) as Internal Standard (IS). With the use of this IS, the extraction and the analysis of the samples were better assessed and controlled to assure the quality of the work. All the samples were randomized during the different steps of the experimental method performed.

Before GC-MS analysis, sample extracts were derivatized with methoxyamine (MeOX) in pyridine and *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (MSTFA) with 1% of trimethylchlorosilane (TMCS) as catalyst. The derivatization method used in this study were optimized in our previous study (Garreta-Lara et al., 2016) where different derivatization conditions were tested to determine the most suitable conditions for the analysis of *D. magna* metabolites. Briefly, to the dry residue (sample extracts previously obtained), 65 µL of 20 µg mL⁻¹ MeOX in pyridine were added. After vortexing for 1 min, the mixture was incubated for 90 min at 30 °C. Thereafter, 30 µL of MSTFA (1% TMCS) were added, vortex mixed for a minute and then incubated for another 30 min at room temperature. Prior to injection, *daphnid* derivatized extracts were finally filtered through a 0.22 µm filters (Ultrafree®-MC, Millipore) and, then transferred to a chromatographic vial and reconstituted to 100 µL with hexane and injected into the GC system.

2.3. Instrumental GC-MS analysis

GC-MS analysis was performed as described in the previous work by Garreta-Lara et al. (2016). A Trace GC-MS Thermo Fisher system operated in electron ionization (EI) mode at 70 eV and equipped with a ZB-5MS column (Phenomenex, 30 m length × 0.25 mm ID × 0.25 µm film thickness) was used. The oven temperature program was set at 70 °C, increasing at 10 °C min⁻¹ to 250 °C, then to 310 °C at 5 °C min⁻¹ and held for 5 min, the delay time was 5.5 min and the time of analysis was 49 min. Other operating conditions were as follows: carrier gas was He, with a flow rate of 0.6 mL min⁻¹; source temperature, 200 °C; interface temperature, 280 °C; split ratio of 1:20; detector voltage, 440 V; 2 µL were injected in the split mode injection. The *m/z* values in full scan mode were in the range of *m/z* 60–650 amu. The mass spectrometer was interfaced to a computer workstation running

Xcalibur software (version 2.2, Thermo Scientific) for data acquisition and processing. Samples were run in a randomized order. Linear retention index (RI) was calculated for each detected peak, using an *n*-alkane standard mixture C₇ to C₃₀ as internal standard (IS). The standard mixture was dissolved in hexane at a concentration of 10 µg mL⁻¹ and injected in the GC-MS instrument with the same programme conditions as the *daphnids* samples.

2.4. Chemometric analysis of GC-MS data

The chemometric data analysis workflow used in this study was similar to the one used in our previous work (Garreta-Lara et al., 2016). Raw full scan GC-MS data sets from Xcalibur software (version 2.2, Thermo Scientific) were converted to cdf data files by File Converter tool of Xcalibur Software and further stored and processed under MATLAB computer environment (The Mathworks, Inc., Natick, MA, USA). Every sample analyzed by GC-MS gives one data matrix, with 2606 rows (retention times, from 5.5 to 49 min) and 590 columns (*m/z* values, from 60 to 650, with one unit of mass resolution). Data values in every data matrix were normalized using the peak area value of the Internal Standard (D-glucose U-13C6) to correct for the instrumental intensity drifts and size effects among injections, and to scale data internally.

Full scan GC-MS data matrices were analyzed by the Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) method. MCR-ALS permits the resolution of the pure elution and mass spectra profiles of the metabolite constituents in the different samples simultaneously analyzed by GC-MS. Using this method a better resolution of the elution profiles of *D. magna* metabolites from those of the large number of interfering derivatized compounds was achieved. Also, from these resolved elution profiles, their peak areas in the different chromatographic runs (samples) can be more easily evaluated and compared.

A total number of 41 GC-MS raw data matrices were obtained, one from a blank sample of the derivatizants (only containing the derivatizing agents) and forty from the exposed samples submitted to the experimental design previously described (8 treatments × 5 treatment replicate samples). Every one of these data matrices had 2606 retention times and 590 *m/z* values. Due to the large dimensions of these 41 individual data matrices, they were subdivided into 12 time windows. These windows were not equal in size and the division was made manually depending on the abundance of the peak signals in the chromatogram area. The 41 individual data matrices corresponding to the same time window were then arranged in a single column-wise augmented data matrix for their simultaneous MCR-ALS analysis. Hence, a total number of 12 new column-wise augmented data matrices were obtained that have the information of the 41 samples (1 blank derivatizant and 40 treatment samples following the proposed experimental design) in one of the preselected time windows.

The application of MCR-ALS to these 12 windowed column-wise augmented data matrices gave as a result the elution profiles of the resolved components in the 41 simultaneously analyzed samples at the considered time window, and their corresponding pure mass spectra. Only components with well resolved chromatographic peak shapes were finally considered and their MS spectra analyzed too. Peak areas were examined taking into account the signal of the derivatizing agents. Peaks present also in the blank derivatizant sample were removed from data analysis. Moreover, samples with a percentage of their relative standard deviations above 30% within the same factor treatment were considered outliers and excluded from the analysis too. Finally, relative peak areas of all the other well resolved elution profiles were calculated and further evaluated.

Table 1
 Summary of results obtained in the MCR-ALS analysis of the 12 window column-wise augmented data matrices (see Section 2.4).

Time window (Tw)	Retention time (min)	Dimensions of the Tw	Estimated components	Resolved peaks	R ² (%)
1	5.5–7.1	100 × 590	22	13	99.0
2	7.1–8.8	101 × 590	20	10	98.5
3	8.8–11.1	141 × 590	22	14	98.6
4	11.1–14.3	196 × 590	17	10	98.6
5	14.3–17.3	176 × 590	16	10	98.0
6	17.3–20.9	220 × 590	17	11	98.5
7	20.9–24.3	202 × 590	15	10	98.9
8	24.3–26.9	161 × 590	11	8	99.4
9	26.9–31.9	301 × 590	12	5	99.1
10	31.9–34.7	166 × 590	13	8	99.2
11	34.7–37.4	166 × 590	10	6	99.0
12	37.4–49.0	692 × 590	10	5	98.0

2.4.1. Identification of metabolites and analysis of their peak areas

MCR-ALS resolved components were identified by their corresponding retention indexes (RI) and MS spectra. To implement this tentative

identification, the MCR-ALS resolved mass spectra were compared with the theoretical mass spectra using the NIST Mass Spectral Search 2.2 program distributed with the NIST 2014 library (<http://chemdata.nist.gov/mass-spc/ms-search>) and the Golm metabolome database (GMD) (<http://gmd.mpimp-golm.mpg.de/download>), including derivatized compounds (Kopka et al., 2005). Match Factor (MF) and Reverse Match Factor (RMF) NIST library parameters were used. The rank of the matching factors used in the identification by the NIST library hits were between 0 (no match) and 1000 (perfect match). To increase the reliability of the identification, the internal linear retention index (RI) markers were also included in the evaluation of the library hits, using the RI values obtained by injecting a saturated Alkane standard mixture from C₇ to C₃₀ (Babushok, 2015). After the identification of the metabolites, Principal Component Analysis (PCA), ANOVA-Simultaneous Component Analysis (ASCA) and Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA) were applied to their corresponding peak areas.

Principal Component Analysis (PCA) (Wold et al., 1987) was applied to the autoscaled peak areas of the resolved profiles obtained after MCR-ALS analysis. ANOVA-Simultaneous Component Analysis (ASCA) (Smilde et al., 2005) was used to interpret and evaluate the statistical

Table 2
 Tentative identification of *D. magna* metabolites extracted by GC–MS and MCR-ALS data analysis. RI: retention index, RMF: reverse match factor.

	tR (min)	RI theoretical	RI experimental	(RI theo-RI exp) ^a	Metabolites	Derivatization	KEGG-ID	RMF
1	6.78	1107	1066	41	Lactic acid	(2TMS)	C00186	913
2	6.02	1056	1032	24	Alanine	(TMS)	C00041	825
3	8.22	1203	1200	3	Leucine	(TMS)	C01933	732
4	8.67	1231	1224	7	Valine	(2TMS)	C00183	900
5	7.95	1185	1282	−97	1-Nonanol	(TMS)	C14696	886
6	9.93	1308	1310	−2	Glycine	(3TMS)	C00037	951
7	9.43	1278	1272	6	Isoleucine	(2TMS)	C00123	916
8	10.47	1338	1343	−5	Uracil	(2TMS)	C00106	913
9	10.83	1358	1367	−9	Serine	(3TMS)	C00065	917
10	9.37	1275	1284	−9	Glycerol	(3TMS)	C00116	910
11	10.72	1352	1353	−1	Fumaric acid	(2TMS)	C00122	847
12	9.87	1305	1303	2	Nicotinic acid	(TMS)	C00253	854
13	10.92	1363	1355	8	Nonanoic acid	(TMS)	C01601	809
14	9.82	1302	1292	10	Threonine	(2TMS)	C00188	926
15	9.40	1276	1282	−6	Phosphoric acid	(3TMS)	C00009	900
16	14.12	1528	1538	−10	N-Acetylglytamic acid	(2TMS)	C00624	908
17	11.78	1409	1416	−7	Methionine	(TMS)	C00073	863
18	11.23	1380	1432	−52	Erythrose	(MeOX)(3TMS)	C01796	613
19	13.22	1483	1492	−9	Malic acid	(3TMS)	C00149	901
20	13.50	1497	1511	−14	Pyroglutamic acid	(TMS)	C02237	834
21	12.02	1422	1422	0	Aspartic acid	(2TMS)	C00049	852
22	13.85	1515	1522	−7	L-5-Oxoproline	(2TMS)	C01879	911
23	15.10	1576	1592	−16	Proline	(3TMS)	C00148	934
24	14.42	1542	1559	−17	Phenylalanine	(TMS)	C00079	889
25	16.30	1634	1645	−11	1,3-Diaminopropane	(4TMS)	C00986	909
26	15.88	1613	1630	−17	Glutamic acid	(3TMS)	C00025	917
27	17.05	1670	1667	3	Xylose	(MeOX)(4TMS)	C00181	866
28	18.78	1754	1780	−26	L-Glycerol-3-phosphate	(4TMS)	C00093	960
29	19.93	1810	1827	−17	Citric acid	(4TMS)	C00158	935
30	18.40	1735	1742	−7	Arginine [−NH ₃]	(2TMS)	C00062	914
31	21.23	1875	1887	−12	Tyrosine	(2TMS)	C00082	874
32	23.13	1973	1926	47	Glucose (Z)	(MeOX)(5TMS)	C00031	920
33	21.97	1913	1920	−7	Lysine	(4TMS)	C00047	896
34	21.45	1886	1876	10	Galactose (Z)	(MeOX)(5TMS)	C00124	856
35	21.13	1870	1928	−58	Methylgalactose	(4TMS)	C03619	958
36	21.50	1889	1870	19	Idose	(MeOX)(5TMS)	C06466	812
37	25.08	2076	2090	−14	Myo-Inositol	(6TMS)	C00137	957
38	24.90	2067	2068	−1	N-Acetyl-D-glucosamine	(MeOX)(4TMS)	C00140	856
39	24.43	2042	2050	−8	Palmitic acid	(TMS)	C00249	907
40	31.22	2433	2469	−36	Uridine	(3TMS)	C00299	934
41	27.98	2239	2246	−7	Stearic acid	(TMS)	C01530	904
42	27.47	2210	2218	−8	Oleic acid (Z)	(TMS)	C00712	966
43	34.15	2622	2653	−31	Sucrose	(8TMS)	C00089	877
44	33.23	2560	2562	−2	Inosine	(4TMS)	C00294	949
45	35.62	2721	2744	−23	Maltose	(MeOX)(8TMS)	C00208	965
46	46.13	3480	3549	−69	Maltotriose	(MeOX)(11TMS)	C01835	930
47	37.65	2866	2838	28	Uridine-5'-Monophosphate	(5TMS)	C00105	833

^a The negative sign does not have any numerical significance.

significance of the changes occurred in the peak areas in terms of the applied experimental design (Jansen et al., 2005). In this analysis, peak areas of the metabolites were mean centered. In the ASCA model, the effects of the three factors (salinity, temperature and hypoxia) at two levels (low and high-levels) were considered. Possible interactions among these treatments have been evaluated. In order to establish the significance of the observed differences, a permutation test was applied to the samples in which the null hypothesis (H_0) assumes that there is no effect of the considered factor. In this work, the number of permutation was set to 10,000.

Orthogonal Partial Least Squares–Discriminant Analysis (OPLS-DA) (Bylesjö et al., 2006; Trygg and Wold, 2002) is an extension to the supervised PLS-DA method featuring an integrated Orthogonal signal correction (OSC) filter. This method uses information in the Y matrix to decompose the X matrix into blocks of structured variation correlated to and orthogonal to Y , respectively. In simple terms, OPLS-DA is calculated between the multivariate data and a response variable that only contains class information. OPLS-DA analysis was applied to the autoscaled peak areas obtained by MCR-ALS analysis. Matthew's Correlation Coefficient (MCC) (Matthews, 1975) was calculated to validate the discrimination model. MCC is in essence a correlation coefficient between the observed and predicted binary classifications; returned values are between +1 and -1. A coefficient +1 indicates a perfect prediction, a value of 0 represents no better than a random prediction and a value of -1 indicates a total disagreement between prediction and observation.

Finally, potential biomarkers produced by the factors studied were analyzed by means of two different strategies: one-way ANOVA and the OPLS-DA discrimination analysis by the use of Variable Importance in Projection (VIP) scores and Selectivity Ratio (SR) values. Besides, the identified metabolites were mapped on Kyoto Encyclopedia of Genes and Genomes (KEGG) database (www.genome.jp/kegg/ligand.html) (Kanehisa and Goto, 2000) to investigate the possible metabolic pathways involved in each case.

2.5. Software

Data analyses were performed using MATLAB 2015a (The Mathworks Inc. Natick, MA, USA) and PLS Toolbox 7.3.1 (Eigenvector Research Inc., Wenatchee, WA, USA) was used for PCA, ASCA and OPLS-DA calculations. MCR-ALS toolbox (Jaumot et al., 2015) was used in MATLAB 2012b (The Mathworks Inc. Natick, MA, USA) for resolution of metabolite profiles from full MS scan augmented data matrices.

3. Results and discussion

3.1. Chemometric analysis

A summary of the results obtained by MCR-ALS analysis of the 12 windowed full scan GC–MS augmented data matrices (including the simultaneous analysis of the 41 individual full scan GC–MS data matrices) for the eight experimental design treatments is shown in Table 1 (see Materials and methods section). In these GC–MS full-scan analysis, the derivatization process played an important role in the detection of the metabolites. Thus, discrimination between the elution profiles belonging to the interfering part of the derivatizing agents and the elution profiles of the *D. magna* metabolites was performed successfully. Between 10 and 22 MCR-ALS components were resolved for each windowed column-wise augmented data matrix achieving in all cases a variance of 98% or more. A total number of 185 components were directly resolved by MCR-ALS analysis, 110 of them with well resolved MS spectra and elution profiles (with good peak shapes). This difference between the estimated number of MCR-ALS resolved components and the final number of selected components is due to the presence of the signals

from the derivatizing agents and also from instrumental and solvent/background contributions.

3.2. Assessment of the effects of the three factors (salinity, temperature and hypoxia)

Interpretation of the metabolite concentration changes caused by the three investigated factors (salinity, temperature and hypoxia) required first the identification of these metabolites. As indicated above, 110 MCR-ALS resolved components showed reasonable chromatographic peak shapes. Their corresponding MS spectra were then considered for their possible identification using the NIST library and the Golm metabolome databases. The identification of the metabolites was tentative. Most of the metabolites were identified with a high value of the Reverse Match Factor (RMF) in NIST library hits.

Since derivatization process with MSTFA was not complete, different products from the same metabolite or from the derivatization reagent were detected and the preliminary list of MCR-ALS resolved components was then reduced to 47 metabolites. The finally selected metabolites were mainly organic acids, carbohydrates and amino acids. They were selected as potential descriptors of the changes caused by the stressing factors in metabolic profiles of *D. magna*. These metabolites were identified with high match factor values, RMF values, between 800 and 1000 as shown in Table 2, except for leucine and erythrose, with RMF values of 732 and 613 match factors, respectively. Another feature considered in their identification was the difference between theoretical and experimental retention indices, RI (Table 2). For most of the metabolites this difference was <50, which confirmed the good identification carried out, except for 1-nonanol, erythrose, methylgalactose and maltotriose, where these differences were 97, 52, 58 and 69, respectively. The selected metabolites were best candidates according to the RMF values or the RI of the NIST and the Golm databases.

Some of the *D. magna* samples produced outlying results and they were excluded from subsequent data analysis. Additionally, another few samples were also excluded from the analysis because the reproducibility of their peak areas on treatment replicate samples was above 30%. Thus, a final number of 24 *D. magna* samples, 3 replicate

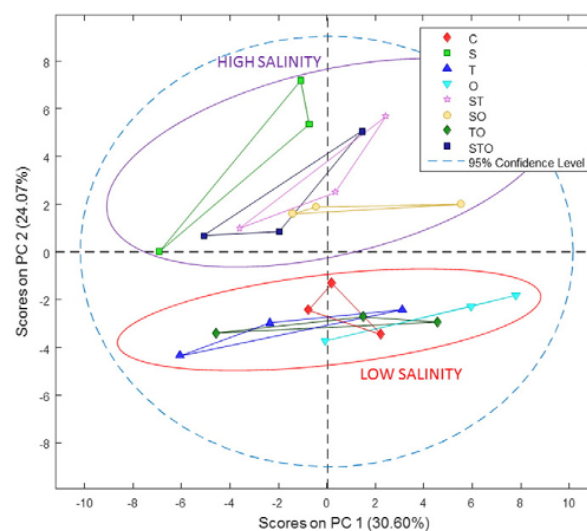


Fig. 1. PCA scores plot of the 24 analyzed samples exposed to 8 different treatments. Different symbols indicate the different sample treatments studied in the experimental design. (For identification of the legend symbols, the reader is referred to Section 2.2.2).

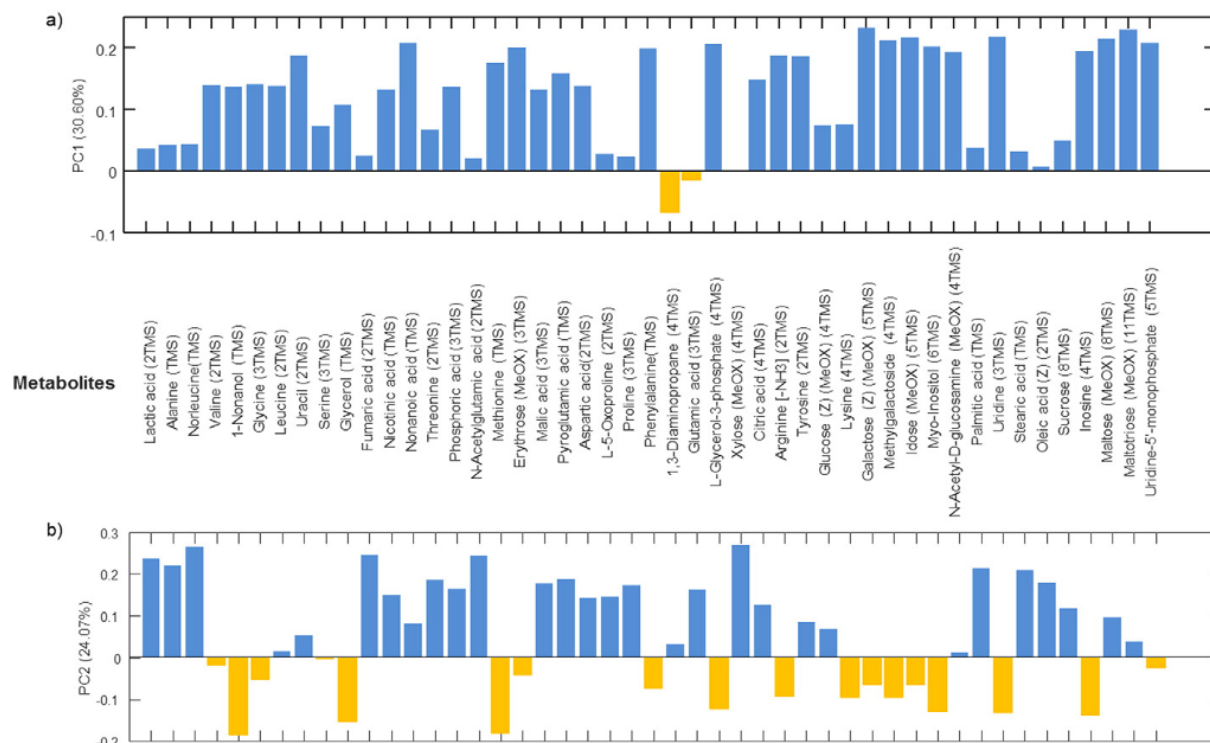


Fig. 2. PCA loading plots of peak areas of the 47 identified metabolites. a) PC1 loadings plot. b) PC2 loadings plot. Positive loadings are displayed in blue and negative loadings in orange.

samples per each treatment, and 47 unique metabolites (Table 2) was considered.

Once the identification of these metabolites and of their peak areas in the different samples were calculated, PCA, OPLS-DA and ASCA analysis were applied to them. Results of PCA applied to the autoscaled peak areas of these 47 identified metabolites for all factor treatments are shown in Fig. 1. Principal component 1 (PC1) captured 30.60% of total data variance, and PC2 captured 24.07% of the data variance. Interpretation of PC1 scores resulted to be ambiguous because of the large differences among samples with the same treatment, most probably due to the natural variation among different organisms of these samples. Samples from different treatments were however distinguished along PC2, with two clusters (marked with high and low salinity circles). In Fig. 1, the group of samples in the

upper side of the plot is related with the higher levels of the salinity factor (S, ST, SO and STO sample treatments). The other group of samples in the lower part of the plot is related with sample treatments with low (no) levels of the salinity factor (C, O, T and TO). Therefore, this salinity factor already allowed the correct separation of the metabolic responses of the investigated samples.

Metabolite PC1 and PC2 loading plots are represented in Fig. 2. In blue, positive loadings and in orange, negative loadings are displayed. Most of the metabolites have positive loadings in PC1 except 1,3-diaminopropane and glutamic acid (negative loadings) and xylose with a close to zero loading in this PC. Metabolites with high positive (blue) loadings in PC2 (Fig. 2b) have higher concentrations in samples with higher salinity (on top of the PCA scores plot in Fig. 1) such as in S, ST, OS and OTS treatments. Metabolites in orange (Fig. 2b) with

Table 3

Potential biomarkers produced by the three factor models: (A) low vs. high salinity treatments, (B) low vs. high temperature treatments and (C) low vs. high hypoxia treatments. Results for one-way ANOVA using p -value < 0.1 and OPLS-DA analysis using VIP scores > 2 and SR threshold value > 2.07 are shown.

	A Salinity	B Temperature	C Hypoxia
One-way ANOVA	LAC, ALA, LEU, NON, GLY, FMR, THR, AGLU, MET, ERY, MAL, ASP, OXO, PRO, PHE, GLU, XYL, G3P, ARG, GAL, MBG, MYO, URD, INO	URA, IDO, PAL, STE, OLE	SER, PHO, TYR, GAL, MBG, IDO, MYO, AGLU, PAL, STE, OLE, URI
OPLS-DA	ALA, LEU, NON, FMR, AGLU, MET, OXO, PRO, G3P	URA, DAP, GLC, IDO, INO, URD	SER, OLE, URI
	SR > 2.07	—	GAL, MBG, IDO, MYO, AGLU
	NON, G3P	—	—

Metabolites abbreviations: AGLU: N-acetyl-D-glucosamine; ALA: Alanine; ARG: Arginine; ASP: Aspartate; DAP: 1,3-Diaminopropane; ERY: Erythrose; FMR: Fumarate; GYL: Glycerol; G3P: Glycerol-3-phosphate; GAL: Galactose; GLC: Glucose; GLU: Glutamate; IDO: Idose; INO: Inosine; LAC: Lactate; LEU: Leucine; MAL: Malate; MBG: Methyl beta-D-Galactoside; MET: Methionine; MYO: Myo-Inositol; NON: 1-Nonanol; OLE: Oleic acid; OXO: L-5-Oxoproline; PAL: Palmitic acid; PHE: Phenylalanine; PHO: Phosphoric acid; PRO: Proline; SER: Serine; STE: Stearic acid; THR: Threonine; TYR: Tyrosine; URA: Uracil; URD: Uridine; XYL: Xylose.

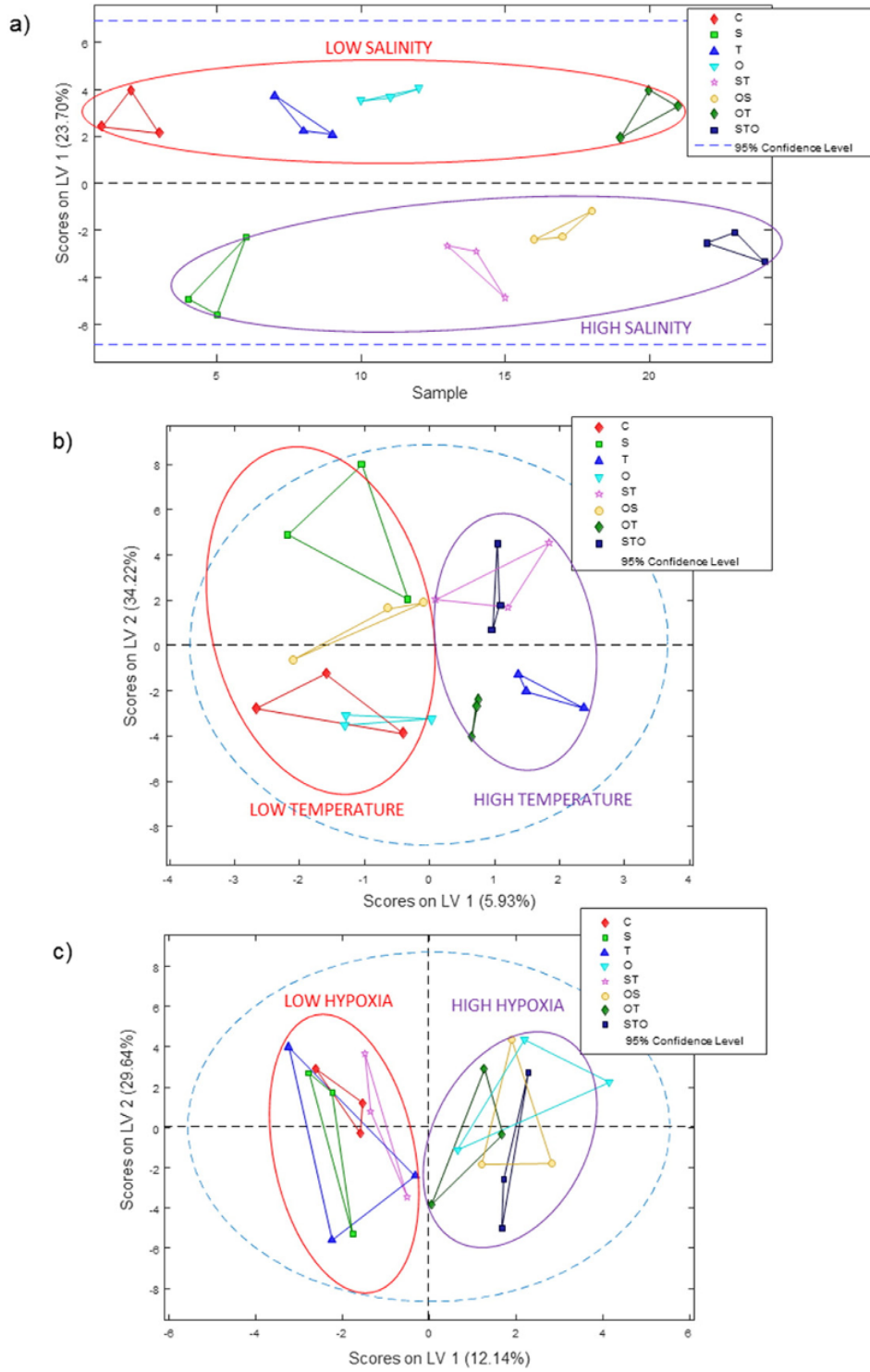


Fig. 3. OPLS-DA score plots for the 24 analyzed samples exposed to 8 different treatments: a) low and high salinity treatments, b) low and high temperature treatments and c) low and high hypoxia treatments. (For identification of the legend symbols, the reader is referred to Section 2.2.2).

negative loadings in PC2 have higher concentrations in samples with the lower salinity (on the bottom of the PCA scores plot in Fig. 1): C, T, O and OT treatments.

As deduced from the comparison of Figs. 1 and 2b, lactic acid, alanine, leucine, fumaric acid, *N*-acetylglutamic acid, and xylose were the metabolites with the highest positive PC2 loadings (Fig. 2b) and therefore their concentrations were higher in samples with high salinity (upper part of Fig. 1). On the other side, 1-nonanol, glycerol, methionine, myo-inositol, uridine and inosine metabolites had the highest negative loadings (Fig. 2b) and therefore they were relatively at higher concentrations in samples with the low salinity (no salt added) (lower part of Fig. 1).

Statistical inference of the global changes in metabolite peak areas of the elution profiles resolved by MCR-ALS was further studied using an ASCA three-factor model with interactions considering two-levels of the three factors. Results showed that the salinity factor (S) (p -value = 0.0015), had a clear statistically significant effect, whereas hypoxia factors (O) (p -value = 0.1162) and specially temperature (T) (p -value = 0.6026) were not statistically significant at the p -value < 0.05. In addition, no interaction between the analyzed factors resulted to be statistically significant (p -value < 0.05). These results confirmed that salinity (S) was the most relevant factor affecting *D. magna* metabolome, and that hypoxia (O) and specially temperature (T) factors did not show strong global effects on the metabolite levels compared to control samples.

Determination of the effects of the three stressing factors (salinity, temperature and hypoxia) on the individual metabolites was further investigated by one-way ANOVA. One-way ANOVA is a good univariate statistical strategy to evaluate the differences in the peak areas of every individual metabolite produced by the different factors studied. In addition, OPLS-DA analysis together with the Variable Importance in Projection (VIP) scores and the Selectivity Ratio (SR) values were then used as a multivariate data analysis tool to reveal which metabolites could have the greatest influence on the discrimination among sample treatments.

Table 3 shows a summary of the results obtained for all detected metabolites when one-way ANOVA and OPLS-DA models were used. Potential biomarkers are indicated separately for the three different factor models studied ((A) salinity, (B) temperature and (C) hypoxia treatments). Results of these strategies show differences in the choice of biomarkers depending on the significance levels or threshold values selected in each case. With one-way ANOVA test, the significance level of 10% (p -value < 0.1) was tested to select the potential biomarkers. The total number of identified biomarkers were 24 for salinity, 5 for temperature, and 12 for hypoxia treatments respectively.

OPLS-DA was applied to discriminate samples according to salinity (S, ST, OS, STO versus C, T, O, OT samples), according to temperature (T, ST, OT, STO, versus C, S, O, OS samples), and according to hypoxia (O, OS, OT, STO versus C, S, T, ST samples) treatments. Different components resulted optimal in every case. Whereas one single latent variable was needed for the salinity treatment (explained class-variance of 89.19%) model, four and three latent variables were needed for the temperature and hypoxia models (71.30% and 79.31%, explained class-variances respectively). Matthew's Correlation Coefficient (MCC, see Materials and methods section) for the salinity OPLS-DA model was equal to 1 indicating a very good classification between the two groups of samples, whereas MCC for temperature and hypoxia models were equal to 0.17 and 0.43, respectively. Correct positive (sensitivity = 1) and negative (specificity = 1) assignments were totally achieved in the case of the salinity treatment, whereas it was not for temperature model (0.67 and 0.50 respectively), nor for hypoxia (0.58 and 0.83 respectively). These results confirm the results obtained previously by ASCA analysis, where it was already observed the lower influence of temperature and hypoxia factor treatments in *D. magna* metabolome at the investigated levels of this study. Therefore, the effects of

temperature and hypoxia treatments would be less clearly interpretable than the salinity effects.

In Fig. 3 the score plots corresponding to the three different discrimination models are shown. A first important conclusion is that a large dispersion across the treatment replicate samples is confirmed, most probably due to the intrinsic natural variation among the *D. magna* organisms in the different analyzed samples. As noted for the salinity model (Fig. 3a), samples with higher salinity levels (in the lower part of the plot) were clearly distinguished along the first latent variable (LV1), from the samples with lower salinity levels (in the upper site of the plot), giving two sample clusters. In Fig. 3b, samples at higher temperatures were grouped in the right side of the plot, whereas samples at lower temperatures were distributed in the left side of the plot, although both classes of samples were not well separated in this case. In Fig. 3c, class samples related with higher hypoxia were distributed in the right side of the plot while in the left side of the plot sample treatments with lower hypoxia were grouped, giving again two separate clusters along LV1.

Table 4
Fold changes, percentages of concentration changes and trends of the identified metabolites in the salinity treatments (S, ST, OS, STO samples) compared to all the other samples (C, T, O, OT). Significant differences are given as ** p < 0.05 and * p < 0.1.

	Metabolites	KEGG-ID	Fold changes	%Changes	Trends
1	Lactic acid	C00186	1.2**	20.7**	Up
2	Alanine	C00041	1.3**	34.3**	Up
3	Leucine	C01933	1.3**	34.5**	Up
4	Valine	C00183	0.9	-6.3	-
5	1-Nonanol	C14696	0.5**	-47.0**	Down
6	Glycine	C00037	0.9	-9.9	-
7	Isoleucine	C00123	0.8	-21.3	-
8	Uracil	C00106	1.0	2.9	-
9	Serine	C00065	1.1	6.0	-
10	Glycerol	C00116	0.8**	-17.1**	Down
11	Fumaric acid	C00122	1.6**	64.6**	Up
12	Nicotinic acid	C00253	1.1	12.8	-
13	Nonanoic acid	C01601	1.0	-4.7	-
14	Threonine	C00188	1.2**	20.9**	Up
15	Phosphoric acid	C00009	1.1	10.3	-
16	<i>N</i> -Acetylglutamic acid	C00624	1.4**	41.2**	Up
17	Methionine	C00073	0.7**	-29.6**	Down
18	Erythrose	C01796	0.8*	-16.1*	Down
19	Malic Acid	C00149	1.2*	16.4*	Up
20	Pyroglutamic acid	C02237	1.1	13.1	-
21	Aspartic acid	C00049	1.2**	21.4**	Up
22	L-5-Oxoproline	C01879	1.4**	36.3**	Up
23	Proline	C00148	1.8**	83.8**	Up
24	Phenylalanine	C00079	0.9**	-13.1**	Down
25	1,3-Diaminopropane	C00986	1.1	9.9	-
26	Glutamic acid	C00025	1.5**	45.6**	Up
27	Xylose	C00181	0.7**	-34.6**	Down
28	L-Glycerol-3-phosphate	C00093	2.1**	107.8**	Up
29	Citric acid	C00158	1.1	9.4	-
30	Arginine [—NH ₃]	C00062	0.8**	-18.6**	Down
31	Tyrosine	C00082	1.0	-2.8	-
32	Glucose (Z)	C00031	1.0	0.8	-
33	Lysine	C00047	0.8	-16.1	-
34	Galactose (Z)	C00124	0.8**	-18.5**	Down
35	Methylgalactose	C03619	0.7**	-30.2**	Down
36	Idose	C06466	0.8	-22.0	-
37	Myo-inositol	C00137	0.8**	-20.3**	Down
38	<i>N</i> -Acetyl-D-glucosamine	C00140	1.0	0.7	-
39	Palmitic acid	C00249	1.1	8.7	-
40	Uridine	C00299	0.8**	-19.6**	Down
41	Stearic acid	C01530	1.1	6.3	-
42	Oleic acid (Z)	C00712	1.0	2.7	-
43	Sucrose	C00089	1.1	12.2	-
44	Inosine	C00294	0.5**	-51.9**	Down
45	Maltose	C00208	1.1	5.6	-
46	Maltotriose	C01835	0.9	-11.7	-
47	Uridine-5'-monophosphate	C00105	0.8	-17.7	-

Metabolites more important in the different OPLS-DA discrimination models using VIP scores > 2 and SR threshold value > 2.07 (Rajalahti et al., 2009) (see **Materials and methods** section) are given in Table 3. For the salinity treatments 9 and 2 metabolites were selected according to VIPs and SR values. For the temperature treatments 6 and 0 metabolites, and for the hypoxia treatments 8 and 0 metabolites. As already described in the literature, variables selected by VIP method may be in some cases false positive candidates, while those selected by the SR method are false negative candidates (Farrés et al., 2015). Thus, the best approach should be performed according to the aim of the variable selection. In this work, we decided not to select the most restrictive conditions among the different strategies, since the aim of this study was to investigate possible changes on the metabolic profile of stressed *D. magna* at the given conditions in a non-restrictive way. Therefore, metabolites obtained using one-way ANOVA with a p -value < 0.1 , which corresponds approximately to a fold change of 1.2 or 0.8 in Table 4 below, were finally selected as potential biomarkers of the effects produced by the three factors models (see Table 3). Thus, a total of 24, 5 and 12 metabolites resulted for the salinity, temperature and hypoxia treatments, respectively.

3.3. Biological interpretation of the changing metabolites

Results from the distinct chemometric strategies performed in this work highlighted the strong effects of salinity in *D. magna* metabolome, whereas the effects of temperature and hypoxia treatments were much less effective changing metabolites levels of *D. magna*. Table 4 provides the fold changes, percentages of concentration changes and trends (up- or down-regulated) of all identified de-regulated metabolites due to salinity treatment. Additional tables corresponding to fold changes, percentages of concentration changes and trends caused by temperature and hypoxia treatments are given in Online Resource 1 and in Online Resource 2, respectively. All these fold changes and percentages of changes were calculated using the mean peak areas of these metabolites

in the samples treated at the two conditions (high and low) of the factors using a Student's t -test.

For the salinity treatment (Table 4), 12 metabolites were up-regulated and 12 metabolites were down-regulated. Four of the 47 metabolites identified had a percentage of change greater than a 50% in the samples with high salt content. L-glycerol-3-phosphate was the metabolite with the highest percentage of change with a 107.8%, followed by proline with a 83.8% of change, fumaric acid with a 64.6% of change and inosine with a -51.9% of change which was down-regulated. Note that up-regulated metabolites by salinity treatment in Table 4 had positive loadings in PC2 PCA plot (blue in Fig. 2b) and down-regulated metabolites had negative loadings in PC2 PCA plot (orange in Fig. 2b). For the temperature treatment (Online Resource 1), only 5 metabolites showed some significant changes and all of them were down-regulated. On the other hand, for the hypoxia treatment (Online Resource 2) 12 metabolites were altered with significant differences, 9 out of them were up-regulated and the rest were down-regulated. In addition, in the simultaneous temperature and hypoxia treatments (both factors at high levels) 4 of these altered metabolites were coincident: idose, palmitic acid, stearic acid and oleic acid molecules.

Metabolites whose concentrations were altered by salinity treatment given in Table 4 explained the differences between groups described by both PCA (Fig. 1) and OPLS-DA (Fig. 3a) analysis. These metabolites were further used to investigate the possible KEGG metabolic pathways affected by the salinity treatments. In Fig. 4 the most disrupted KEGG metabolic pathways and the number of altered metabolites are given graphically. These metabolites and the disrupted KEGG metabolic pathways are also listed in Table 5. Some of these metabolites appear in the top seven pathways described in Fig. 4 and in Table 5, and they are coincident and related with biosynthesis of secondary metabolites, ABC transporters, carbon metabolism, biosynthesis of amino acids and of aminoacyl-tRNA, protein digestion and mineral absorption. Salinity treatments modified the relative amounts of two main metabolites (fumarate and malate) (Table 5) from the Krebs cycle, both involved in the energy metabolism. We

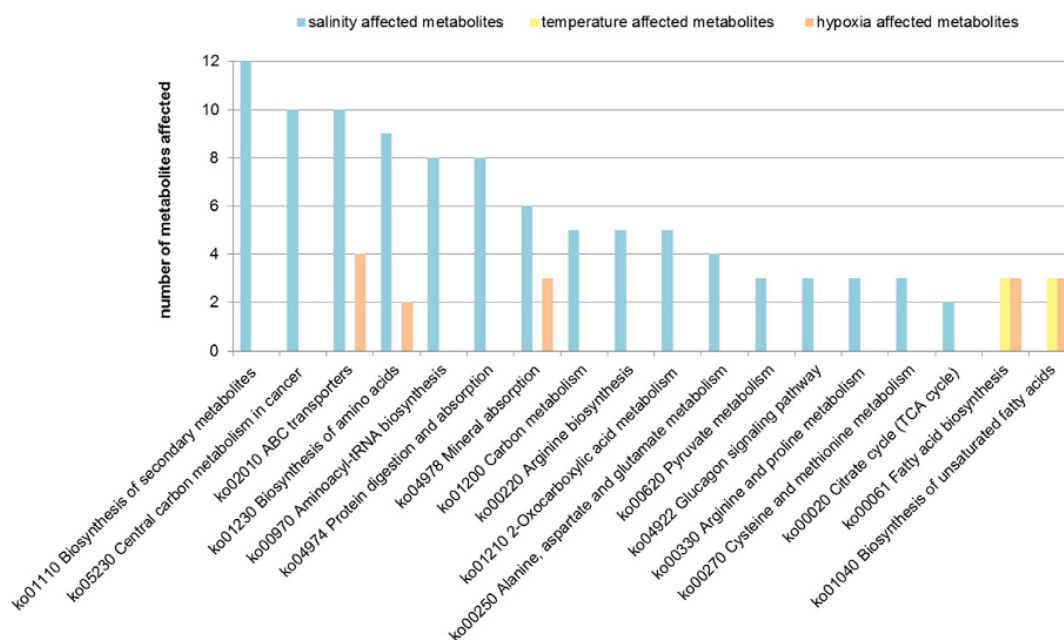


Fig. 4. Number of *D. magna* metabolites in KEGG metabolic pathways most disrupted by salinity, temperature and hypoxia treatments.

Table 5
Most disrupted KEGG metabolic pathways in *D. magna* by salinity, temperature and hypoxia treatments.

Metabolic pathways	Salinity		Temperature		Hypoxia	
	N ^a	Metabolites ^b	N ^a	Metabolites ^b	N ^a	Metabolites ^b
ko01110 Biosynthesis of secondary metabolites	12	GLU, ASP, ARG, MET, PHE, G3P, FMR, PRO, MAL, LAC, THR, AGLU	0		0	
ko05230 Central carbon metabolism in cancer	10	GLU, ALA, ASP, ARG, MET, PHE, FMR, PRO, MAL, LAC	0		0	
ko02010 ABC transporters	10	GLU, ALA, ASP, ARG, PHE, G3P, MYO, PRO, THR, MBG	0		4	PHO, SER, AGLU, MBG
ko01230 Biosynthesis of amino acids	9	GLU, ALA, ASP, ARG, MET, PHE, PRO, THR, MBG	0		0	
ko00970 Aminoacyl-tRNA biosynthesis	8	GLU, ALA, ASP, ARG, MET, PHE, PRO, THR	0		0	
Ko04974 Protein digestion and absorption	8	GLU, ALA, ASP, ARG, MET, PHE, PRO, THR	0		0	
ko04978 Mineral absorption	6	ALA, MET, PHE, GAL, PRO, THR	0		3	PHO, SER, GAL
ko01200 Carbon metabolism	5	GLU, ALA, ASP, FMR, MAL	0		0	
ko00220 Arginine biosynthesis	5	GLU, ASP, ARG, FMR, AGLU	0		0	
ko01210 2-Oxocarboxylic acid metabolism	5	GLU, ASP, MET, PHE, AGLU	0		0	
ko00250 Alanine, aspartate and glutamate metabolism	4	GLU, ALA, ASP, FMR	0		0	
ko00620 Pyruvate metabolism	3	FMR, MAL, LAC	0		0	
ko04922 Glucagon signaling pathway	3	FMR, MAL, LAC	0		0	
ko00330 Arginine and proline metabolism	3	GLU, ARG, PRO	0		0	
Ko00270 Cysteine and methionine metabolism	3	ALA, ASP, MET	0		0	
ko00020 Citrate cycle (TCA cycle)	2	FMR, MAL	0		0	
ko00061 Fatty acid biosynthesis	0		3	PAL, OLE, STE	3	PAL, OLE, STE
ko01040 Biosynthesis of unsaturated fatty acids	0		3	PAL, OLE, STE	3	PAL, OLE, STE

^a Number of altered metabolites.

^b Metabolites abbreviations: AGLU: N-acetyl-D-glucosamine; ALA: Alanine; ARG: Arginine; ASP: Aspartate; FMR: Fumarate; G3P: Glycerol-3-phosphate; GAL: Galactose; GLU: Glutamate; LAC: Lactate; MAL: Malate; MBG: Methyl beta-D-Galactoside; MET: Methionine; MYO: Myo-Inositol; OLE: Oleic acid; PAL: Palmitic acid; PHE: Phenylalanine; PHO: Phosphoric acid; PRO: Proline; SER: Serine; STE: Stearic acid; THR: Threonine.

conclude therefore that individuals of *D. magna*, exposed to high salinity levels suppress energy metabolic rates and oxygen demands (Arnér and Koivisto, 1993), which affect the Krebs cycle and their involved metabolites.

Possible KEGG metabolic pathways involved in temperature and hypoxia treatments (see Online Resources 1 and 2) were also investigated. Three of the affected metabolites in both treatments (palmitic acid, stearic acid and oleic acid) were involved in the biosynthesis of fatty acids and unsaturated fatty acids metabolic pathways (ko00061 and ko01040, respectively) as shown in Table 5. Besides, hypoxia affected metabolites (Online Resource 2) were also involved in two metabolic pathways (Table 5): ABC transporters and mineral absorption (ko02010 and ko04978, respectively). Temperature and hypoxia treatments enhanced anabolic metabolic pathways, which may be opposed to salinity. For instance, increasing temperatures raised up metabolic and anabolic demands in *D. magna* (i.e. protein turnover rates (Dölling et al., 2016; Schwerin et al., 2009)). Moreover, *D. magna* is well adapted to hypoxia, increasing its hemoglobin content and up-regulating >50 proteins related with glycolytic enzymes, and enzymes involved in the degradation of storage and structural carbohydrates (Paul et al., 1998; Zeis et al., 2009). The previous mentioned metabolic effects would explain the difficult interpretation of these two effects when they are combined in the same experiments with the salinity as in the present work.

According to the risk assessment in the mixture framework approach, noninteractive joint effects of dissimilar acting contaminants or stressors should be additive and predicted in an additive manner (Altenburger et al., 2003), which means that the interaction of those mixtures should not be significant or contributed little to the explained variance (De Coninck et al., 2013). In this work, salinity was the dominant factor affecting most *D. magna* metabolome. In our previous study (Garreta-Lara et al., 2016) it was also found that salinity was the dominant factor. In this previous study, salinity alone already accounted for about 70% of the total de-regulated metabolites, whereas temperature and hypoxia alone accounted each for only 15% of them. Thus, both studies are coincident that salinity had the greatest effects on *D. magna* metabolome. Furthermore, 5 of the 12 de-regulated metabolic KEGG pathways reported in our previous study (Garreta-Lara et al. (2016)) were also found in this work, which were related with amino acid, protein, sugar, fatty

acid and energetic metabolism (ko00970, ko00020, ko00250, ko00620, ko00061). Main differences observed between both studies were that in our previous study (Garreta-Lara et al., 2016) energy metabolic pathways gave somewhat more evidence about metabolites de-regulated by hypoxia and temperature. This was mostly due to the concentration changes in two metabolites: malate and fumarate. However, in the present work this could not be confirmed, since malate and fumarate were only affected by salinity and very little by temperature and hypoxia treatments. Further studies should be required to confirm these results.

4. Conclusions

Alterations in *D. magna* metabolism observed in this work can be used as sensitive indicators of environmental stress. The results of this study confirmed that increasing salinity, and to a much lower extent increasing temperatures and hypoxia, affect the concentrations of different metabolites of *D. magna*, which can be related to respiration, energy stress and cell osmoregulation metabolic pathways. Metabolites whose concentrations changed most were carbohydrates, fatty acids, organic acids and amino acid molecules.

Results obtained in this work have also shown that the joint effect of different environmental stressors on the whole metabolome of an organism can be analyzed using a comprehensive non-targeted GC-MS metabolomics approach combined with the application of appropriate multivariate statistical and chemometric data analysis methods.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.scitotenv.2017.05.190>.

Compliance with ethical standards.

Conflict of interest

The authors declare no financial conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Acknowledgments

The research leading to these results received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013)/ERC Grant Agreement no. 320737.

References

- Alonso, A., Marsal, S., Julià, A., 2015. Analytical methods in untargeted metabolomics: state of the art in 2015. *Front. Bioeng. Biotechnol.* 3, 23.
- Altenburger, R., Nendza, M., Schürmann, G., 2003. Mixture toxicity and its modeling by quantitative structure-activity relationships. *Environ. Toxicol. Chem.* 22, 1900–1915.
- Arnér, M., Koivisto, S., 1993. Effects of salinity on metabolism and life history characteristics of *Daphnia magna*. *Hydrobiologia* 259, 69–77.
- Babushok, V.I., 2015. Chromatographic retention indices in identification of chemical compounds. *TrAC Trends Anal. Chem.* 69, 98–104.
- Barata, C., Baird, D.J., 1998. Phenotypic plasticity and constancy of life-history traits in laboratory clones of *Daphnia magna* Straus: effects of neonatal length. *Funct. Ecol.* 12, 442–452.
- Bergman Filho, T.U., Soares, A.M.V.M., Loureiro, S., 2011. Energy budget in *Daphnia magna* exposed to natural stressors. *Environ. Sci. Pollut. Res.* 18, 655–662.
- Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J.K., Holmes, E., Trygg, J., 2006. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemom.* 20, 341–351.
- Cousyn, C., De Meester, L., Clombourne, J.K., Brendonck, L., Verschuren, D., Volckaert, F., 2001. Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. *Proc. Natl. Acad. Sci.* 98, 6256–6260.
- De Coninck, D.I.M., De Schampelaere, K.A.C., Jansen, M., De Meester, L., Janssen, C.R., 2013. Interactive effects of a bacterial parasite and the insecticide carbaryl to life-history and physiology of two *Daphnia magna* clones differing in carbaryl sensitivity. *Aquat. Toxicol.* 130–131, 149–159.
- Dölling, R., Becker, D., Hawat, S., Koch, M., Schwarzenberger, A., Zeis, B., 2016. Adjustments of serine proteases of *Daphnia pulex* in response to temperature changes. *Comp. Biochem. Physiol. B: Biochem. Mol. Biol.* 194–195, 1–10.
- Farrés, M., Platikanov, S., Tsakovski, S., Tauler, R., 2015. Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *J. Chemom.* 29, 528–536.
- Ficke, A.D., Myrick, C.A., Hansen, L.J., 2007. Potential impacts of global climate change on freshwater fisheries. *Rev. Fish Biol. Fish.* 17, 581–613.
- Garreta-Lara, E., Campos, B., Barata, C., Lacorte, S., Tauler, R., 2016. Metabolic profiling of *Daphnia magna* exposed to environmental stressors by GC-MS and chemometric tools. *Metabolomics* 12, 86.
- Hibbert, D.B., 2012. Experimental design in chromatography: a tutorial review. *J. Chromatogr. B* 910, 2–13.
- IPCC, 2007. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. *Climate Change 2007: Synthesis Report*. Cambridge University Press.
- Jansen, J.J., Hoefsloot, H.C.J., Van Der Greef, J., Timmerman, M.E., Westerhuis, J.A., Smilde, A.K., 2005. ASCA: analysis of multivariate data obtained from an experimental design. *J. Chemom.* 19, 469–481.
- Jaumot, J., de Juan, A., Tauler, R., 2015. MCR-ALS GUI 2.0: new features and applications. *Chemom. Intell. Lab. Syst.* 140, 1–12.
- Jones, O.A.H., Swain, S.C., Svendsen, C., Griffin, J.L., Sturzenbaum, S.R., Spurgeon, D.J., 2012. Potential new method of mixture effects testing using metabolomics and *Caenorhabditis elegans*. *J. Proteome Res.* 11, 1446–1453.
- Kanehisa, M., Goto, S., 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.
- Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., et al., 2005. GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* 21, 1635–1638.
- Kovacevic, V., Simpson, A.J., Simpson, M.J., 2016. 1H NMR-based metabolomics of *Daphnia magna* responses after sub-lethal exposure to triclosan, carbamazepine and ibuprofen. *Comp. Biochem. Physiol. D: Genom. Proteomics* 19, 199–210.
- Li, L., Wu, H., Ji, C., van Gestel, C.A.M., Allen, H.E., Peijnenburg, W.J.G.M., 2015. A metabolomic study on the responses of *Daphnia magna* exposed to silver nitrate and coated silver nanoparticles. *Ecotoxicol. Environ. Saf.* 119, 66–73.
- Malik, A., Jordao, R., Campos, B., Casas, J., Barata, C., Tauler, R., 2016. Exploring the disruptive effects of TBT on lipid homeostasis of *Daphnia magna* using chemometric methods. *Chemom. Intell. Lab. Syst.* 159, 58–68.
- Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *BBA - Protein Struct.* 405, 442–451.
- Nagato, E.G., D'Eon, J.C., Lankadurai, B.P., Poirier, D.G., Reiner, E.J., Simpson, A.J., et al., 2013. 1H NMR-based metabolomics investigation of *Daphnia magna* responses to sub-lethal exposure to arsenic, copper and lithium. *Chemosphere* 93, 331–337.
- Nagato, E.G., Simpson, A.J., Simpson, M.J., 2016. Metabolomics reveals energetic impairments in *Daphnia magna* exposed to diazinon, malathion and bisphenol-a. *Aquat. Toxicol.* 170, 175–186.
- Paul, R.J., Colmorgen, M., Pirow, R., Chen, Y.H., Tsai, M.C., 1998. Systemic and metabolic responses in *Daphnia magna* to anoxia. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 120, 519–530.
- Poynton, H.C., Taylor, N.S., Hicks, J., Colson, K., Chan, S., Clark, C., et al., 2011. Metabolomics of microliter hemolymph samples enables an improved understanding of the combined metabolic and transcriptional responses of *Daphnia magna* to cadmium. *Environ. Sci. Technol.* 45, 3710–3717.
- Rajalahti, T., Arneberg, R., Kroksveen, A.C., Berle, M., Myhr, K.M., Kvalheim, O.M., 2009. Discriminating variable test and selectivity ratio plot: quantitative tools for interpretation and variable (biomarker) selection in complex spectral or chromatographic profiles. *Anal. Chem.* 81, 2581–2590.
- Scanlan, L.D., Loguinov, A.V., Teng, Q., Antczak, P., Dailey, K.P., Nowinski, D.T., et al., 2015. Gene transcription, metabolite and lipid profiling in eco-indicator *Daphnia magna* indicate diverse mechanisms of toxicity by legacy and emerging flame-retardants. *Environ. Sci. Technol.* 49, 7400–7410.
- Schwerin, S., Zeis, B., Lamkemeyer, T., Paul, R.J., Koch, M., Madlung, J., et al., 2009. Acclimatory responses of the *Daphnia pulex* proteome to environmental changes. II. Chronic exposure to different temperatures (10 and 20 °C) mainly affects protein metabolism. *BMC Physiol.* 9.
- Smilde, A.K., Jansen, J.J., Hoefsloot, H.C.J., Lamers, R.-J.A.N., van der Greef, J., Timmerman, M.E., 2005. ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics* 21, 3043–3048.
- Taylor, N.S., Weber, R.J.M., Southam, A.D., Payne, T.G., Hrydziszko, O., Arvanitis, T.N., et al., 2009. A new approach to toxicity testing in *Daphnia magna*: application of high throughput FT-ICR mass spectrometry metabolomics. *Metabolomics* 5, 44–58.
- Taylor, N.S., Weber, R.J.M., White, T.A., Viant, M.R., 2010. Discriminating between different acute chemical toxicities via changes in the daphnid metabolome. *Toxicol. Sci.* 118, 307–317.
- Taylor, N.S., Kirwan, J.A., Johnson, C., Yan, N.D., Viant, M.R., Gunn, J.M., et al., 2016a. Predicting chronic copper and nickel reproductive toxicity to *Daphnia pulex-Pulicaria* from whole-animal metabolic profiles. *Environ. Pollut.* 212, 325–329.
- Taylor, N.S., Kirwan, J.A., Yan, N.D., Viant, M.R., Gunn, J.M., McGeer, J.C., 2016b. Metabolomics confirms that dissolved organic carbon mitigates copper toxicity. *Environ. Toxicol. Chem.* 35, 635–644.
- Toyota, K., Gavin, A., Miyagawa, S., Viant, M.R., Iguchi, T., 2016. Metabolomics reveals an involvement of pantothenate for male production responding to the short-day stimulus in the water flea, *Daphnia pulex*. *Sci. Rep.* 6.
- Trygg, J., Wold, S., 2002. Orthogonal projections to latent structures (O-PLS). *J. Chemom.* 16, 119–128.
- Vandenbrouck, T., Jones, O.A.H., Dom, N., Griffin, J.L., De Coen, W., 2010. Mixtures of similarly acting compounds in *Daphnia magna*: from gene to metabolite and beyond. *Environ. Int.* 36, 254–268.
- Viant, M.R., 2008. Recent developments in environmental metabolomics. *Mol. BioSyst.* 4, 980–986.
- Wagner, N.D., Lankadurai, B.P., Simpson, M.J., Simpson, A.J., Frost, P.C., 2015. Metabolomic differentiation of nutritional stress in an aquatic invertebrate. *Physiol. Biochem. Zool.* 88, 43–52.
- Wagner, N.D., Simpson, A.J., Simpson, M.J., 2016. Metabolomic responses to sublethal contaminant exposure in neonate and adult *Daphnia magna*. *Environ. Toxicol. Chem.*
- Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. *Chemom. Intell. Lab. Syst.* 2, 37–52.
- Zeis, B., Lamkemeyer, T., Paul, R.J., Nunes, F., Schwerin, S., Koch, M., et al., 2009. Acclimatory responses of the *Daphnia pulex* proteome to environmental changes. I. Chronic exposure to hypoxia affects the oxygen transport system and carbohydrate metabolism. *BMC Physiol.* 9.
- Zhang, B., Zhang, H., Du, C., Ng, Q.X., Hu, C., He, Y., et al., 2017. Metabolic responses of the growing *Daphnia similis* to chronic AgNPs exposure as revealed by GC-Q-TOF/MS and LC-Q-TOF/MS. *Water Res.* 114, 135–143.

Online Resources

Combined effects of salinity, temperature and hypoxia on *Daphnia magna* metabolism

Elba Garreta-Lara, Bruno Campos, Carlos Barata, Silvia Lacorte, Romà Tauler*

Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, 08034, Catalonia, Spain.

*Corresponding author: Romà Tauler

Postal address: Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, 08034, Catalonia, Spain.

Telephone: +34934006140

Fax: +34932045904

E-mail: roma.tauler@idaea.csic.es

Online Resource 1. Fold changes, percentages of concentration changes and trends of the identified metabolites in the temperature treatments (T, ST, OT, STO samples) compared to all the other samples. Significant differences are given as $**p < 0.05$ and $*p < 0.1$.

	Metabolites	KEGG-ID	Fold changes	%Changes	Trends
1	Lactic acid	C00186	0.9	-5.3	=
2	Alanine	C00041	1.0	-4.3	=
3	Leucine	C01933	0.9	-8.8	=
4	Valine	C00183	0.9	-9.6	=
5	1-Nonanol	C14696	0.9	-5.2	=
6	Glycine	C00037	1.0	4.8	=
7	Isoleucine	C00123	0.8	-17.8	=
8	Uracil	C00106	0.9*	-12.6*	Down
9	Serine	C00065	1.0	2.2	=
10	Glycerol	C00116	1.0	-2.3	=
11	Fumaric acid	C00122	0.9	-11.2	=
12	Nicotinic acid	C00253	1.0	-2.9	=
13	Nonanoic acid	C01601	1.0	-4.7	=
14	Threonine	C00188	1.0	2.5	=
15	Phosphoric acid	C00009	1.0	1.9	=
16	N-Acetylglutamic acid	C00624	1.0	-1.2	=
17	Methionine	C00073	0.9	-10.3	=
18	Erythrose	C01796	1.0	1.4	=
19	Malic acid	C00149	0.9	-8.7	=
20	Pyroglutamic acid	C02237	0.9	-7.2	=
21	Aspartic acid	C00049	0.9	-7.8	=
22	L-5-Oxoproline	C01879	1.0	0.2	=
23	Proline	C00148	1.0	0.5	=
24	Phenylalanine	C00079	1.0	0.6	=
25	1,3-Diaminopropane	C00986	1.1	14.7	=
26	Glutamic acid	C00025	0.9	-8.5	=
27	Xylose	C00181	1.0	-1.5	=
28	L-Glycerol-3-phosphate	C00093	1.0	-1.9	=
29	Citric acid	C00158	0.9	-8.9	=
30	Arginine [-NH3]	C00062	0.9	-6.1	=
31	Tyrosine	C00082	1.0	-2.3	=
32	Glucose (Z)	C00031	1.0	-3.3	=
33	Lysine	C00047	0.9	-8.3	=
34	Galactose (Z)	C00124	0.9	-6.1	=
35	Methylgalactose	C03619	0.9	-11.4	=
36	Idose	C06466	0.7*	-30.0*	Down
37	Myo-Inositol	C00137	1.0	-4.1	=
38	N-Acetyl-D-glucosamine	C00140	0.9	-14.2	=
39	Palmitic acid	C00249	0.7**	-25.2**	Down
40	Uridine	C00299	0.9	-5.8	=
41	Stearic acid	C01530	0.8**	-23.8**	Down
42	Oleic acid (Z)	C00712	0.8*	-21.6*	Down
43	Sucrose	C00089	0.9	-5.4	=
44	Inosine	C00294	0.9	-8.4	=
45	Maltose	C00208	0.9	-6.7	=
46	Maltotriose	C01835	0.7	-26.1	=
47	Uridine-5'-monophosphate	C00105	0.8	-16.7	=

Online Resources

Combined effects of salinity, temperature and hypoxia on *Daphnia magna* metabolism

Elba Garreta-Lara, Bruno Campos, Carlos Barata, Silvia Lacorte, Romà Tauler*

Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, 08034, Catalonia, Spain.

*Corresponding author: Romà Tauler

Postal address: Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, 08034, Catalonia, Spain.

Telephone: +34934006140

Fax: +34932045904

E-mail: roma.tauler@idaea.csic.es

Online Resource 2. Fold changes, percentages of concentration changes and trends of the identified metabolites in the hypoxia treatments (O, OS, OT, STO samples) compared to all the other samples. Significant differences are given as $**p < 0.05$ and $*p < 0.1$.

	Metabolites	KEGG-ID	Fold changes	%Changes	Trends
1	Lactic acid	C00186	1.1	5.3	=
2	Alanine	C00041	0.9	-6.0	=
3	Leucine	C01933	0.9	-12.7	=
4	Valine	C00183	1.1	12.6	=
5	1-Nonanol	C14696	1.1	8.3	=
6	Glycine	C00037	1.0	1.4	=
7	Isoleucine	C00123	1.2	21.4	=
8	Uracil	C00106	1.0	3.2	=
9	Serine	C00065	1.2**	21.8**	Up
10	Glycerol	C00116	1.0	1.8	=
11	Fumaric acid	C00122	1.0	-0.8	=
12	Nicotinic acid	C00253	1.1	9.9	=
13	Nonanoic acid	C01601	1.0	-3.9	=
14	Threonine	C00188	0.9	-9.8	=
15	Phosphoric acid	C00009	1.3**	26.7**	Up
16	N-Acetylglutamic acid	C00624	1.0	-4.1	=
17	Methionine	C00073	1.1	8.2	=
18	Erythrose	C01796	0.9	-6.0	=
19	Malic acid	C00149	1.0	0.9	=
20	Pyroglutamic acid	C02237	1.0	-2.2	=
21	Aspartic acid	C00049	1.1	6.9	=
22	L-5-Oxoproline	C01879	1.0	3.2	=
23	Proline	C00148	1.2	23.1	=
24	Phenylalanine	C00079	1.1	7.3	=
25	1,3-Diaminopropane	C00986	0.8	-15.2	=
26	Glutamic acid	C00025	1.0	-0.8	=
27	Xylose	C00181	0.9	-7.8	=
28	L-Glycerol-3-phosphate	C00093	1.0	1.6	=
29	Citric acid	C00158	1.1	5.4	=
30	Arginine [-NH3]	C00062	1.0	0.7	=
31	Tyrosine	C00082	1.1*	14.3*	Up
32	Glucose (Z)	C00031	1.0	-0.6	=
33	Lysine	C00047	0.9	-6.1	=
34	Galactose (Z)	C00124	1.2*	16.1*	Up
35	Methylgalactose	C03619	1.6**	60.6**	Up
36	Idose	C06466	1.5**	54.8**	Up
37	Myo-Inositol	C00137	1.2**	22.6**	Up
38	N-Acetyl-D-glucosamine	C00140	1.5**	50.1**	Up
39	Palmitic acid	C00249	0.8*	-20.9*	Down
40	Uridine	C00299	1.1	8.7	=
41	Stearic acid	C01530	0.8*	-21.3*	Down
42	Oleic acid (Z)	C00712	0.7**	-27.0**	Down
43	Sucrose	C00089	1.1	9.3	=
44	Inosine	C00294	0.9	-13.0	=
45	Maltose	C00208	1.1	8.8	=
46	Maltotriose	C01835	1.0	1.6	=
47	Uridine-5'-monophosphate	C00105	1.5**	50.5**	Up

2.4 ARTICLE CIENTÍFIC III

Chemometric analysis of comprehensive two dimensional gas chromatography–mass spectrometry metabolomics data.

Y. Izadmanesh, E. Garreta-Lara, J. B. Ghasemi, S. Lacorte, V. Matamoros, R. Tauler.

Journal of Chromatography A, 1488 (2017), 113-125.



Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Chemometric analysis of comprehensive two dimensional gas chromatography–mass spectrometry metabolomics data

Yahya Izadmanesh^{a,b}, Elba Garreta-Lara^b, Jahan B. Ghasemi^c, Silvia Lacorte^b, Victor Matamoros^b, Roma Tauler^{b,*}^a Faculty of Chemistry, K. N. Toosi University of Technology, Tehran, Iran^b Institute of Environmental Assessment and Water Research, Spanish Council for Scientific Research (CSIC), Jordi Girona 18, Barcelona 08034, Spain^c Faculty of Chemistry, University of Tehran, Tehran, Iran

ARTICLE INFO

Article history:

Received 3 October 2016

Received in revised form

25 December 2016

Accepted 22 January 2017

Available online 23 January 2017

Keywords:

GC×GC-TOFMS

Wavelet compression

Chemometrics

Multivariate curve resolution-alternating

least squares (MCR-ALS)

Metabolomics

ABSTRACT

Comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GC×GC-TOFMS) is a powerful tool for separation and identification of analytes in complex natural samples. In this paper, different chemometric methods were compared for fast non-targeted GC×GC-TOFMS metabolomic profiling of the crustacean species *Daphnia magna* and a general chemometric strategy and workflow is proposed. The strategy proposed in this work combined the compression of GC×GC-TOFMS data matrices in the retention time direction using wavelets and the appropriate column-wise data matrix augmentation arrangement, and its modeling by Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS). Using the proposed strategy, eighty different *D. magna* metabolites were resolved and identified. After calculation of the peak capacities of different columns and peak area changes of these metabolites, the best instrumental configuration and column combination for the GC×GC-TOFMS metabolomic study of *D. magna* are proposed and discussed. The procedure described in this work can be applied as a general method for untargeted GC×GC-TOFMS data processing and metabolomic profiling.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Comprehensive two-dimensional gas chromatography, GC×GC, is a well-established and useful instrumental platform for analyzing biological samples [1–4]. In comparison to one dimensional (one column) chromatographic approaches, using comprehensive two-dimensional gas chromatography–mass spectrometry (GC×GC-TOFMS) allows for higher separation efficiencies, improved resolution, better identification and quantitation of co-eluting analytes [5–7]. However, some difficulties are still present even when using GC×GC-TOFMS: (1) complete separation of all detectable analytes cannot be achieved in complex biological samples, because of existing limitations in experimental and instrumental conditions; (2) baseline/background changes often cause problems in GC×GC-TOFMS analysis due to the presence of two chromatographic columns; and (3), every GC×GC-TOFMS run produces a very large, dense, information-rich three-way array of data, typically of 500–1500 megabytes (MB) per sample, depending on

run time and data collection rate. When multiple samples are simultaneously analyzed and compared, current laboratory computer technologies have still difficulties to manage and process adequately the huge amounts of data produced.

The first two problems can be solved to some extent with application of multivariate chemometric methods such as bilinear multivariate curve resolution-alternating least squares (MCR-ALS) [5–9], and in some particular circumstance by parallel factor analysis (PARAFAC) [10,11], PARAFAC2 [12], and trilinear MCR-ALS [13] methods. While each of these chemometric methods used for analyzing GC×GC-TOFMS data have their own strengths and weaknesses, their analytical goals are generally the same; essentially, combination of analyte resolution, identification, and quantification. However, progressive and random elution time changes in position or shape which cause limitations in the structure of the GC×GC-TOFMS data, and restricts application of some of these chemometric methods [10].

In order to address the third challenge related with the huge size of GC×GC-TOFMS data sets, several data handling strategies such as pixel level [14–16], peak table [17], or peak region basis [18] have been proposed. These strategies propose a systematic way to GC×GC-TOFMS data analysis, but the main challenge (large-size of the data sets) is still present. Data compression methods

* Corresponding author.

E-mail address: roma.auler@idaea.csic.es (R. Tauler).

can be used to reduce data sizes without missing relevant information, so that reasonable low capacity storage is maintained and computational demands are alleviated [19–21]. Data compression algorithms based on wavelet transform have been shown to be very effective methods in chemistry [22] and they have the advantages of simplicity, speed, and multi-resolution capability. Wavelet compressed data are amenable to secondary processing such as multivariate curve resolution modeling methods [23].

Metabolomics was defined in 1990s to describe techniques aimed at measuring metabolites present within a cell, tissue or organism during a genetic alteration or physiological stimulus [24]. Metabolomics strategies have been divided into two distinct approaches, untargeted and targeted metabolomics, each with their own inherent advantages and disadvantages. The objective of untargeted metabolomics is to comprehensively survey as many metabolites as possible and maximize number of detected peaks to construct a global profile. Due to its comprehensive nature, untargeted metabolomics must be coupled to advanced chemometric techniques, to reduce the extensive datasets into a smaller set of manageable signals. Major challenges of this approach lie in the protocols and time required to process data, difficulties in identifying and characterizing unknown small molecules, reliance on the intrinsic analytical coverage of the employed platform, and bias towards detection of high-abundance molecules. By contrast, targeted metabolomics is the measurement of predefined list of chemically characterized and biochemically annotated metabolites.

Daphnia magna, a freshwater crustacean, is used extensively as an aquatic test species [25], being the object of standardized testing guidelines from the Organization for Economic Co-operation and Development (OECD) [26]. Acute and chronic tests of *D. magna* are among the most frequently performed studies in aquatic toxicology because these animals are relatively easy to culture, have a short lifecycle, and can be maintained at high population densities in relatively small volumes and thus are cost-effective [27].

In this paper, a general workflow is proposed for metabolomic profiling of *D. magna* by GC×GC-TOFMS and chemometric analysis. A comparison is made on the performance of different chemometric methods and on the best GC×GC-TOFMS instrumental configurations for general metabolomics work.

2. Experimental

2.1. Chemical and reagents

D-glucose (U-13C6, 99%), used as internal standard (IS), was supplied by Cambridge Isotope Laboratories, Inc. (Andover, MA, USA). Triphenylamine (98%) used also as IS, was obtained from Sigma-Aldrich (St. Louis, MO, USA). Pyridine (anhydrous, 99.8%), chlorotrimethylsilane (TMCS), methoxyamine hydrochloride (98%) (MeOX) and *N*-methyl-*N*-trimethyl-silyl trifluoroacetamide (>98.5%) (MSTFA), used as derivatization agents and the saturated Alkane standard mixture for the performance test of GC systems from C₇ to C₃₀ were also obtained from Sigma-Aldrich (St. Louis, MO, USA). Analytical reagent grade hexane, methanol and chloroform were supplied from Merck (Darmstadt, Germany).

2.2. Sample preparation and metabolite extraction

Metabolites were extracted from the whole organism of *D. magna* using the same method described in the previous work [28]. Polar metabolites of the whole organism were extracted with 400 μ L methanol, vortexed and sonicated for 15 min, and then 200 μ L water and 400 μ L chloroform were added before centrifugation at 10,000 *g* during 15 min, in order to separate the aqueous

and the lipid phase. The aqueous phase of every sample was transferred to a new eppendorf and 10 μ L of D-glucose were added as Internal Standard (IS) at a concentration of 50 μ g mL⁻¹. Afterwards, the extract was evaporated to dryness with a Speedvac (Thermo Scientific) at 40 °C during 3 h. When samples were completely dry they were stored at -80 °C until analysis.

The obtained dried extracts were derivatized as follows. 65 μ L of MeOX (20 μ g μ L⁻¹) in pyridine were added to each sample. After mixing for 1 min, the mixture was incubated for 90 min at 30 °C. Thereafter, 30 μ L of MSTFA (1% TMCS) were added, vortex mixed for a minute and then incubated for another 30 min at room temperature. Prior to injection, 10 μ L of triphenylamine were added as IS in a concentration of 50 μ g mL⁻¹ and, *daphnid* derivatized extracts were finally filtrated through a 0.22 μ m filters (Ultrafree[®]-MC, Millipore). Then, the final polar fraction was transferred to a chromatographic vial and injected into GC×GC-TOFMS instrument. GC×GC-TOFMS analysis was performed within 12 h from derivatization.

2.3. GC×GC-TOFMS

The comprehensive two-dimensional gas chromatography was coupled to time-of-flight mass spectrometry (GC×GC-TOFMS) using a Pegasus 4D (LECO, St. Joseph, MI, USA), system which has a 6890 gas chromatograph (Agilent Technologies, Palo Alto, CA, USA) equipped with a split/splitless injector, a secondary oven to fit the secondary column, and a ZX1 (Zoex, Houston, TX, USA) two-stage thermal modulator. Liquid nitrogen was used to cool nitrogen gas for cold pulses. The column combinations and method parameters used for analysis of *D. magna* samples are shown in Table 1.

The six chromatographic configurations used for analysis of metabolomic samples of *D. magna* are named S1, S2, S3, S4, S5 and S6 in Table 1. Three different column configurations were tested: A, B and C (see Table 1) and the best chromatographic configuration for *D. magna* metabolomic study was investigated based on theoretical peak capacities. The first columns (¹D) used in all column configurations were coated with 5% diphenyl, 95% dimethylpolysiloxane (TRB-5MS (30 m × 0.18 mm ID × 0.18 μ m film thickness) from Teknokroma (Sant Cugat del Vallès, Spain)) in configuration A, and ZB-5MS from Phenomenex (30 m × 0.25 mm ID × 0.25 μ m film thickness) in configuration B and C. The used second columns (²D) were coated with 50% diphenyl, 50% dimethylpolysiloxane, (TRB-50HT (2 m × 0.10 mm ID × 0.10 μ m film thickness)) in configuration A and B, and TRB-50 (2.0 m × 0.25 mm ID × 0.25 μ m film thickness) from Teknokroma in configuration C.

The secondary oven temperature was 5 °C higher than preliminary oven and with a similar ramp. The helium carrier gas flow rate, modulation period, hot pulse duration, and temperature offset of modulator versus primary oven temperature are also given for each column combination in Table 1. The MS transfer line was held at 250 °C, and TOFMS instrument, equipped with an electron ionization ion source, was set to scan mass range of 60–700 *m/z*. The ion source temperature was 200 °C. Detector operation was at 1800 V and applied electron energy was 70 eV. Acquisition rate was 100 spectra per second with unit mass resolution. 2.0 μ L of each sample was injected via an auto-sampler (Agilent 7890) in splitless injection mode.

3. Methods

3.1. GC×GC-TOFMS data structure

Comprehensive two-dimensional gas chromatography (GC×GC) is a technique with two columns of different selec-

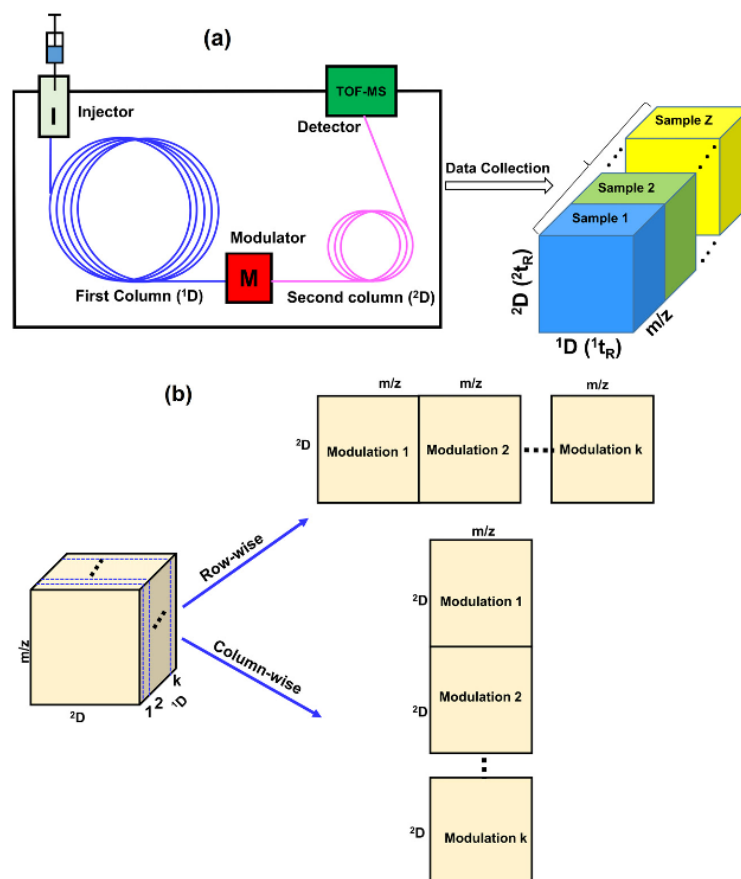


Fig. 1. Schematic representation of (a) GCxGC-TOFMS system, and (b) column-wise and row-wise augmented data arrays of one GCxGC-TOFMS sample.

tivity connected in series by a modulation device [29]. A general scheme of a GCxGC system is presented in Fig. 1 (a). First column or mode (¹D column) is a conventional column, while the second column (²D column) is a shorter and efficient column as those commonly used in fast GC [30]. ¹D column is connected to injector and ²D column is fitted into detector. Both columns can be placed in the same oven or each one separated. At the beginning of ²D, a modulator (M) is placed, which is the main part of the GCxGC instrument and functions in three steps: (i) Collection of continuously narrow fractions of the ¹D effluent, ensuring that separation in the first column is maintained; (ii) focus or refocus the effluent and (iii) quickly transfer the ¹D collected fraction sample and inject it as a narrow pulse to ²D column. The combination of these three steps is called modulation cycle, which is repeated throughout the chromatographic run. Required time to perform a cycle is called modulation period (MP), which is typically 2–10 s and depends on the time required for the compounds to be eluted in ²D. The MP should be as small as possible in order to avoid separation losses in ¹D. Therefore, the entire first column chromatogram is modulated into a series of high-speed short secondary chromatograms of a length equal to MP, which are continuously recorded by the detector. K is the total number of modulations in one GCxGC-TOFMS run.

Every modulation produces a data matrix, X (I, J) where I (1, 2, ..., I) is number of mass spectra scans in each modulation and J is the predefined m/z range to be recorded for every scan. If all modulations (k = 1, 2, ..., K) are considered simultaneously, the whole set

of acquired data of GCxGC-TOFMS analysis of a single sample, gives a three-way data array with dimensions equal to $\times (I, J, K)$. If several samples are analyzed, a four-way data array will be obtained with dimensions equal to (I, J, K, Z), where I, J, and K are same as before, and Z is the number of analyzed samples (Fig. 1 (a)). As shown in Fig. 1 (b), the three-way data array obtained in the analysis of every sample can be converted into a large two way array, X_{aug} , in a column-wise or in a row wise augmentation manner [31,32]. For column-wise augmentation, $X_{aug} (I \times K \times J)$, the different individual data matrices (or slices of three-way array) from every modulation can be put on top of each other, in a way that, the m/z direction is common among all modulations [7]. For row-wise augmentation, $X_{aug} (I \times K \times J)$, the different individual data matrices (slices of three-way array) of modulation are set one besides the other, in a way that, the retention times in ²D column are common among all different modulations. If three-way data arrays of more than one sample are available, for simultaneous analysis, the column or row-wise augmented data matrices of different samples can be super-augmented ($X_{Superaug}$) in the same manner [7].

Fig. 2 shows the proposed chemometrics workflow for analysis of the GCxGC-TOFMS metabolomic data of *D. magna*. In step I, data were first explored for their internal structure. To check whether GCxGC-TOFMS three-way data structure conforms to the trilinear model, rank analysis (by singular value decomposition, SVD, [33]) of the column and row wise augmented data matrices was performed. Chemometric methods based on different model assumptions were compared and evaluated. The tested chemo-

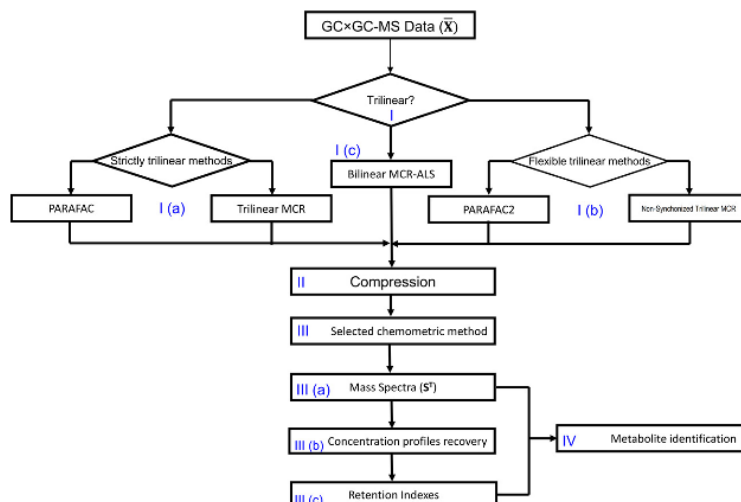


Fig. 2. Workflow of the proposed chemometrics methodology for GC×GC-TOFMS metabolomic profiling of *D. magna*.

metric methods were PARAFAC and MCR-ALS with the trilinear constraint (step I (a)), PARAFAC2 and non-synchronized trilinear MCR-ALS methods which allow for peak shift deviations (step I (b)), and bilinear MCR-ALS method (step I (c)) [10,34,35]. In step II, the column-wise augmented data set was compressed along the elution time mode using wavelets [21] to reduce computer storage and computation time (see below). In step III (a), compressed data were analyzed using the most suitable chemometric approach to estimate the mass spectra of the resolved metabolites. In step III (b), the elution (concentration) profiles of these metabolites in the two columns were recovered (Section 3.5). In step III (c), retention indexes (RI) were calculated. Finally, in step IV, metabolites were identified using their RI and mass spectra in S^T , using NIST data base library. More detailed explanation of all these different steps are given below.

3.2. Data structure and modeling testing of GC×GC-TOFMS data

When GC×GC-TOFMS data conforms with the trilinear [10] model, for each component, the same elution profiles and spectra in the two columns are recovered for all modulation data matrices, only differing in a scale factor [7]. A fast way, for testing fulfillment of the trilinearity for a particular chromatographic run data set, is to perform singular value decomposition (SVD) of the corresponding row- and column-wise augmented data matrices [36] (see Fig. 1(b)). If data are trilinear, then, in absence of noise, both arrays will share the same row- and column- vector spaces and give the same number of large singular values (larger than those due to noise). In this conditions, the chemical rank (mathematical rank in absence of experimental noise [36]) should be the same and equal to the total number of eluted components [32]. In case of fulfillment of the trilinear model, and of absence of rank deficiencies in any of the data modes (i.e. every component gives a distinct profile in every mode), the same 'chemical' rank (mathematical rank in absence of experimental noise) should be obtained. If for instance, SVD analysis shows a higher number of components in row-wise augmented data matrix than column-wise augmented data matrix, then data structure is not in agreement with the trilinear model, probably due to changes in the elution profiles caused by time shifts and shape changes between modulations. This situation has been shown to be present in many cases as shown in several previous

studies of chromatographic data and described in more detail in previous works [6–9,34].

3.3. Bilinear MCR-ALS

MCR-ALS is based on a bilinear model, which decomposes the measured data, collected in a data matrix, into a set of pure component contributions in two factor matrices. For GC×GC-TOFMS data, the bilinear MCR-ALS decomposition model can be described by Eq. (1)

$$\mathbf{X} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (1)$$

Where \mathbf{X} (I, J) is the data matrix collected in a single modulation, I is retention times in 2D column and J is measured m/z values. \mathbf{C} (I, N) has the 2D elution profiles of the N eluted chemical components, and \mathbf{S}^T (N, J) has the pure mass spectra, which can be used for identification of possible metabolites. NIST MS spectra library was used for this purpose [28]. \mathbf{E} (I, J) is the residual matrix having the data variance unexplained by the model [5,8]. \mathbf{C} and \mathbf{S}^T matrices can be estimated by an iterative algorithm based on two constrained alternating linear least-squares steps [37]. This bilinear model for MCR-ALS data decomposition can be easily extended to simultaneous analysis of multiple modulations of the same GC×GC-TOFMS run and to multiple GC×GC-TOFMS runs (Eqs. (2) and (3)) [5–9,38].

$$\mathbf{X}_{aug} = \mathbf{C}_{aug}\mathbf{S}^T + \mathbf{E}_{aug} \quad (2)$$

$$\mathbf{X}_{Superaug} = \mathbf{C}_{Superaug}\mathbf{S}^T + \mathbf{E}_{Superaug} \quad (3)$$

To assess the quality and reliability of MCR-ALS solutions, lack of fit, LOF (Eq. (4)) and percent of explained variance, R^2 (Eq. (5)) values were calculated according to equations defining these two parameters:

$$LOF(\%) = 100 \sqrt{\frac{\sum_{ij} e_{ij}^2}{\sum_{ij} x_{ij}^2}} \quad (4)$$

$$R^2 = 100 \sqrt{\frac{\sum_{ij} x_{ij}^2 - \sum_{ij} e_{ij}^2}{\sum_{ij} x_{ij}^2}} \quad (5)$$

where x_{ij} is an element of the experimental data matrix, \hat{x}_{ij} is the corresponding value calculated by MCR-ALS, and $e_{ij} = x_{ij} - \hat{x}_{ij}$ is the residual value [37].

Table 1
 The ¹D and ²D column combinations and method parameters of the GC×GC–MS determinations.

Conf. set ^a	Column ID	¹ D (mm ID)	² D (mm ID)	¹ D temp °C ^b	² D temp °C ^b	¹ D ramp °C/min ^c	² D ramp °C/min ^c	¹ D time min ^d	² D time min ^d	² D temp °C/min ^c	² D time min ^d	² D modulation offset °C	² D modulation time s
S1	A	TRB-5MS (0.18)	TRB-50HT (0.1)	70, 100, 260, 310	7, 5, 10	7, 5, 10	7, 5, 10	2, 5	2, 5	7, 5, 10	2, 5	30	4
S2	A	TRB-5MS (0.18)	TRB-50HT (0.1)	70, 100, 260, 310	7, 5, 10	7, 5, 10	7, 5, 10	2, 5	2, 5	7, 5, 10	2, 5	15	4
S3	A	TRB-5MS (0.18)	TRB-50HT (0.1)	70, 100, 260, 310	5, 3, 10	5, 3, 10	5, 3, 10	3, 5	3, 5	5, 3, 10	3, 5	30	4
S4	B	ZB-5MS (0.25)	TRB-50HT (0.1)	70, 100, 260, 310	7, 5, 10	7, 5, 10	7, 5, 10	2, 5	2, 5	7, 5, 10	2, 5	30	4
S5	C	ZB-5MS (0.25)	TRB-50 (0.25)	70, 100, 260, 310	7, 5, 10	7, 5, 10	7, 5, 10	2, 5	2, 5	7, 5, 10	2, 5	30	2,5
S6	C	ZB-5MS (0.25)	TRB-50 (0.25)	70, 100, 260, 295	7, 5, 5	7, 5, 5	7, 5, 5	3, 5	3, 5	7, 5, 5	3, 5	15	2,5

All configurations have ²D hot pulse time of 0.5 s.

^a Column sets A, B and C had flow rates of 0.6 mL·min⁻¹, 1 mL·min⁻¹ and 1 mL·min⁻¹, respectively.

^b Temperatures of the ramp temperature program.

^c Temperature increase for minute, in the first, second and third ramp temperature program.

^d Duration of initial and final temperatures of the ramp temperature program.

The number of components in MCR-ALS modeling was first estimated by visual inspection of the number of singular values [33] of the considered data matrix (\mathbf{X} , \mathbf{X}_{aug} or $\mathbf{X}_{superaug}$) which were larger than those already associated to noise (i.e. when the size of singular values are not changing anymore and they are small). Increasing the number of components was considered appropriate, only if a diminution in the lack of fit (Eq. (4)) and an increase in the explained variance (Eq. (5)) after MCR-ALS were observed. Also, visual inspection of the shapes of resolved elution and spectra profiles can help to ascertain the reliability of the selected components and to discern between those with chromatographic and MS signal features and those that are just noise or experimental artifacts. MS raw signal intensities were initially divided by 10^4 to make computations more manageable and to facilitate evaluations, graphical representations and comparison of the results.

3.4. Data compression and MCR-ALS analysis

Due to huge size in one single GC×GC-TOFMS data set, especially in the time direction and for \mathbf{X}_{aug} and $\mathbf{X}_{superaug}$ matrices, a data compression strategy based on the use of wavelets [20,39,40] allowed an easier chemometric analysis and reduced computer storage requirements considerably. Wavelet compression consists of two algorithms: first, there is a compression algorithm that takes the input data matrix, \mathbf{X} , and generates a representation of it, $\hat{\mathbf{X}}$, which requires fewer computer storage (bits), and there is also a reconstruction algorithm that operates on this representation, $\hat{\mathbf{X}}$, and generates reconstructed matrix [21]. Reconstructed matrix contains the same information as \mathbf{X} , but requires much less computer storage.

In this work, the discrete wavelet transform compression/filtering algorithm was used to reduce the size of the huge time mode of GC×GC data without missing relevant information [20]. This wavelet function can be described by Eq. (6) [19].

$$\langle f(t), \Psi_{m,n}(t) \rangle = \int_{-\infty}^{+\infty} f(t) \Psi_{m,n}(t) dt \quad (6)$$

where $f(t)$ is the signal to be compressed, $\Psi_{m,n}(t)$ is the mother wavelet function, m is the scale and n is the shift in time. The used mother wavelet in this study was the simple Haar function (Eq. (7)).

$$\Psi_{m,n}(t) = \begin{cases} 1 & \text{if } 0 < t < 0.5 \\ -1 & \text{if } 0.5 < t < 1 \\ 0 & \text{elsewhere} \end{cases} \quad (7)$$

By changing size of the mother wavelet function, or scaling and translating it, other mother functions can be easily obtained, which provides the multi-resolution property [39].

Multi-resolution property of wavelets which means adapting the signal resolution makes it applicable to stationary and non-stationary signals [41].

Wavelet decomposition and compression was applied independently on every column (m/z) of the \mathbf{X}_{aug} matrix. Compression reduces the size of data 2^n times, which, n is the compression level [20,39,40]. The compressed matrix contains the same information as \mathbf{X}_{aug} , but needs much lower computer storage. For the datasets under study in this paper, it is demonstrated that level-4 wavelet compression is reliable without significant loss of relevant information in the elution time direction. Data compression was performed individually on augmented matrix of each samples (\mathbf{X}_{aug}). The compressed data of all samples were then super-augmented to obtain \mathbf{X}_{compr} for MCR-ALS analysis.

Fig. 3 shows the general procedure followed for simultaneous MCR analysis of wavelet compressed GC×GC-TOFMS data of multiple chromatographic runs and subsequent recovery of elution

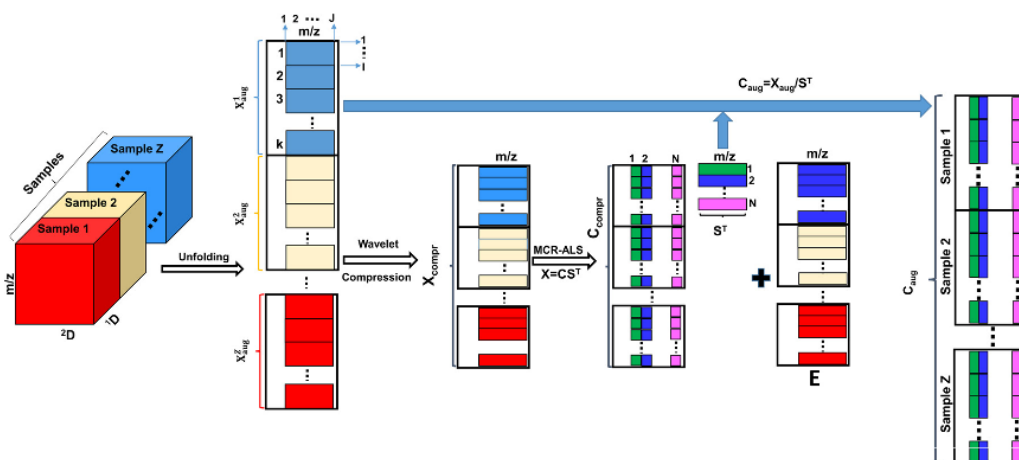


Fig. 3. Schematic representation of the wavelet compression and simultaneous MCR-ALS modeling of several samples. Each data cube represents a three-way data array obtained by the GC×GC-TOFMS analysis of one sample. $X_{aug}^1, X_{aug}^2, \dots, X_{aug}^Z$ are the corresponding matrixed data arrays of the different samples. Non-negative least squares projection of S^T onto X_{aug}^z allows the recovery of the corresponding uncompressed elution profiles for every sample.

profiles in 1^D and 2^D columns using fast non-negative least squares [42] post-processing step. MCR-ALS on X_{compr} gave the compressed elution profiles (C_{compr}) and mass spectral profiles (S^T). Once S^T was properly estimated, it was possible to obtain the uncompressed elution profiles ($C_{superaug}$) using a non-negativity constrained least squares postprocessing step [42]. C_{aug} was then extracted from $C_{superaug}$ taking into account the order used for supraugmentation of the different samples in $D_{superaug}$ matrix (see Section 3.3).

3.5. Elution profiles in first and second columns

C_{aug} is the column-wise augmented concentration matrix of one sample and contains the concatenated 2^D elution profiles. The elution profiles of metabolites in 1^D and 2^D columns can be recovered by refolding the augmented elution profile of every component in C_{aug} . It contains second-column elution profiles for all N components in all modulations. To obtain first and second-column elution profiles of every component in each sample, every column in C_{aug} should be appropriately refolded to give a matrix with (I, K) dimensions, where I is the number of elution times in the second column and K is the number of modulations, i.e. the number of elution times in the first column. The sum along the columns of this refolded data matrix gives an estimation of the corresponding first-column elution profiles. Therefore, for every sample, a matrix of first-column elution profiles of dimensions (N, K) is obtained, N being the number of resolved components.

In case of several samples ($X_{superaug}$), the concentration matrix of every sample (C_{aug}) is extracted from $C_{superaug}$, by keeping in mind the arrangement of samples in 'supraugmentation' step (Fig. 3). Then the procedure for obtaining 1^D and 2^D elution profiles is the same as explained before.

3.6. Metabolites identification

MCR-ALS resolved elution and spectra profiles were assigned to metabolites and identified by comparing the retention indexes (RI) of the peak 1^D profiles on one side and from the mass fragmentation patterns associated to the MCR-ALS resolved mass spectra profiles on the other side. Standard mass spectral database of the National Institute of Standards and Technology (www.nist.gov/srd/nist1a.htm) was used to identify spectra fragmentation patterns. For each mass spectrum, 100 hits were retrieved by the

NIST Mass Spectral Search 2.2 software distributed with the NIST 2014 library. A reverse match factor (RMF) based on the correlation coefficient between the MCR-AS resolved and experimental mass spectra reported by NIST software was used for selection of the best identified compound for MCR-ALS resolved mass spectra. This match factor is reported between 0 (no match) and 1000 (perfect match). As a general guide, a value of 900 or greater was considered to be a very good matching; between 800 and 900, a good match; between 700 and 800, a fair match; and less than 600 a poor or very poor match. To increase the reliability of the identification, NIST internal linear RI markers of first column were also included in the evaluation of the library hits, using the injection of a saturated Alkanes standard mixture (C_7 to C_{30}) [43].

3.7. Calculation of peak capacities

The most suitable chromatographic configuration (column and instrumental setup) was chosen by comparison of the peak capacities of all chromatographic configurations given in Table 1. Peak capacity in the first column was calculated by Eq. (8) [44]:

$$CP_1 = \frac{(t_{R,1} - t_{M,1})}{4\delta} \quad (8)$$

where CP_1 , $t_{m,1}$ and $t_{R,1}$ are respectively peak capacity, column holdup and retention times of the last peak eluting in the first column, and 4δ is the average peak width. In the second column, the separation is approximately isothermal because of the short length of 2^D column, and peak capacity was calculated using Eq. (9) [45].

$$CP_2 = 1 + \frac{\sqrt{N}}{4} \times \ln(t_{R,2}/t_{m,2}) \times 20 \quad (9)$$

where CP_2 , $t_{m,2}$ and $t_{R,2}$ are the peak capacity, column holdup and maximum retention times for the last peak eluting from 2^D column, respectively [44,45]. The peak width and retention times of glucose were used to estimate the number of plates (N) in the second column. Overall peak capacity was obtained by multiplying CP_1 and CP_2 .

3.8. Software

GC×GC-TOFMS data were acquired using ChromaTOF software version 3.3.2 (LECO, St. Joseph, MI, USA). Data from ChromaTOF software were imported and converted into MATLAB in the 60–700 m/z

range by using bioinformatics toolbox (The Mathworks, Inc., Natick, MA, USA). The compression was done by the wavelet toolbox (The Mathworks, Inc., Natick, MA, USA). NIST MS Search version 2.2 (National Institute of Standards Technology, USA) and the Golm metabolome database (GMD) of derivatized compounds [46,47] were used for metabolite identifications. Wavelet toolbox of Matlab was used [48] for data compression. PARAFAC and PARAFAC2 from PLS toolbox (<http://www.eigenvector.com/>) and MCR-ALS toolbox [37] <http://mcrals.wordpress.com/download/mcr-als-toolbox/> were used for chemometrical analysis

4. Results and discussion

4.1. GC×GC-TOFMS experimental profiles and selection of the best chromatographic configuration

Fig. 4 shows the 3D TIC chromatograms of the *D. magna* sample extract analyzed by GC×GC-TOFMS using the different column configurations set in Table 1. The peak shape and peak capacity of metabolites, as well as their distribution along the chromatogram are influenced by the internal diameter, operating conditions and film thickness of the columns [1,44]. According to Fig. 4, the column set A (see Table 1) showed an apparent worse result on the resolution of the peaks and peak shapes of the metabolites than column set B. The change in offset modulations from 30 °C to 15 °C did not bring any significant improvement in the separation conditions of samples analyzed by S1, S2 and S3 configurations (see Fig. 4(a)–(c), respectively). Taking into account these considerations, in the column set B, ¹D column was changed by another with an ID of 0.25 mm and 0.25 μm of film thickness.

In column configuration S4, using the column set B, with a ¹D column with ID 0.25 mm, and a ²D column with ID 0.10 mm (see Table 1), the parameters were modified with the intention of obtaining the highest peak capacity (see Fig. 4(d)). Distribution of the peaks was better than those observed for the previous three configurations. In order to test an additional configuration, column configuration C was also tested, in which, the two columns had the same internal diameter (0.25 mm ID), (see Table 1). This new configuration was intended to determine the effects of increasing diameter in the second column and, to avoid peak overload and obtain optimum flow rates in both columns (see Fig. 4(e) and (f)).

Results showed that column set C was probably the configuration providing better chromatographic results. However, to ensure that this configuration was better than others, chromatographic parameters (peak capacity and resolution) of all identified metabolites were also assessed and compared for the different column configurations. In this regard, selection of the best column configuration was done by comparison of the peak capacities of configurations (Section 3.6).

Fig. 5(a) shows the calculated first and second column peak capacities and total peak capacities of the different investigated chromatographic setups. The peak capacities of S1, S2, S3 and S4 setups were lower than peak capacities of S5 and S6 setups. In S1, S2, S3 and S4 setups, separation of metabolites achieved in ¹D column was partially lost in ²D column because of longer modulation periods. In second column, the column length is short and with longer modulation periods, the probability of remixing of components and therefore overlapping of elution profiles increases. In contrast, in configurations S5 and S6, the separation efficiencies achieved in the first column were maintained in ²D column, because of their shorter modulation periods (2.5 s). The peak capacity of S4 setup was higher than peak capacity of S1, S2 and S3 setups, probably because the stationary film thickness of ¹D and ²D columns in S4 was higher and the metabolites separation was more effective. When comparing S5 and S6 setups, the final temperature in the former was

Table 2

LOF and R² values of chemometric analysis of GC×GC-MS data by different methods. (see Eq. 3 and 4)

Method	LOF (%)	R ² (%)
Bilinear MCR-ALS	6.22	99.61
Non-synchronized trilinear MCR	11.76	98.61
synchronized trilinear MCR	24.98	93.76
PARAFAC	11.50	98.68
PARAFAC2	78.36	38.58

310 °C, which was too high for temperature tolerance of ²D column, hence, S6 setup was finally preferred over S5 setup as the best column combination. It has the highest peak capacity and therefore it allowed for a better distribution of metabolites over the whole available chromatographic space.

Moreover, the best instrumental chromatographic configuration was also investigated by comparison of the peaks areas of metabolites obtained in the analysis of different *D. magna* samples using the mentioned chromatographic configurations (Table 1) and instrumental setups. Fig. 5 (b) compares the peak area of one of the metabolites (lactic acid) when analyzed by mentioned chromatographic setups. According to Fig. 5(b), the highest peak area of lactic acid has been obtained when the sample analyzed by configuration S6. Therefore, configuration S6 was additionally confirmed to be the best GC×GC configuration for metabolomic study of *D. magna*, and configuration S6 is proposed to be used in future GC×GC-TOFMS metabolomics studies.

4.2. Testing GC×GC-TOFMS data structure and selection of the best chemometric approach

To explore the structure of the GC×GC-MS datasets, modulations 94–100 of sample S1 were first selected and analyzed in detail. Comparison of SVD analysis of row- and column-wise augmented data matrices showed some differences on how the size of the singular values decrease with the number of components. Sizes of singular values of column-wise augmented matrix (sharing the same MS spectral vector space) decrease much faster than those of row-wise augmented matrix, especially for the first ones. This is a preliminary indication of possible deviations from the trilinear model.

Fig. 6 shows the ²D elution profiles of one of the resolved components in modulations 94–100 by different modeling methods. ²D elution profile resolved by PARAFAC2 (in blue) was physically meaningless (with negative values and more than one peak). In case of bilinear MCR-ALS, ²D elution profiles had small changes in the shape and position in successive modulations. In non-synchronized trilinear MCR-ALS, shapes of elution profiles are forced to be the same in successive modulations, but, due to shape changes of elution profiles from modulation-to-modulation in raw data, the fitting was worse in comparison with bilinear MCR-ALS [49,50]. In synchronized trilinear MCR-ALS, shapes of elution profiles and their retention times are forced to be the same in successive modulations, but, due to within run retention time shifts from modulation to modulation in raw data, the fitting was worse in comparison with non-synchronized trilinear MCR-ALS. The elution profile of PARAFAC is rather acceptable, but because of retention time shifts and trilinearity deviation of the data, the LOF and R² values are worse in comparison with bilinear MCR-ALS. From results in Table 2, rank analysis of row and column wise augmented data matrices and results shown in Fig. 6, bilinear MCR-ALS was concluded to be the best chemometric approach for the investigated data. The remaining models include artificial constraints which are not applicable at all, at least in principle, to chromatographic data. An additional advantage of bilinear MCR-ALS modeling is that data

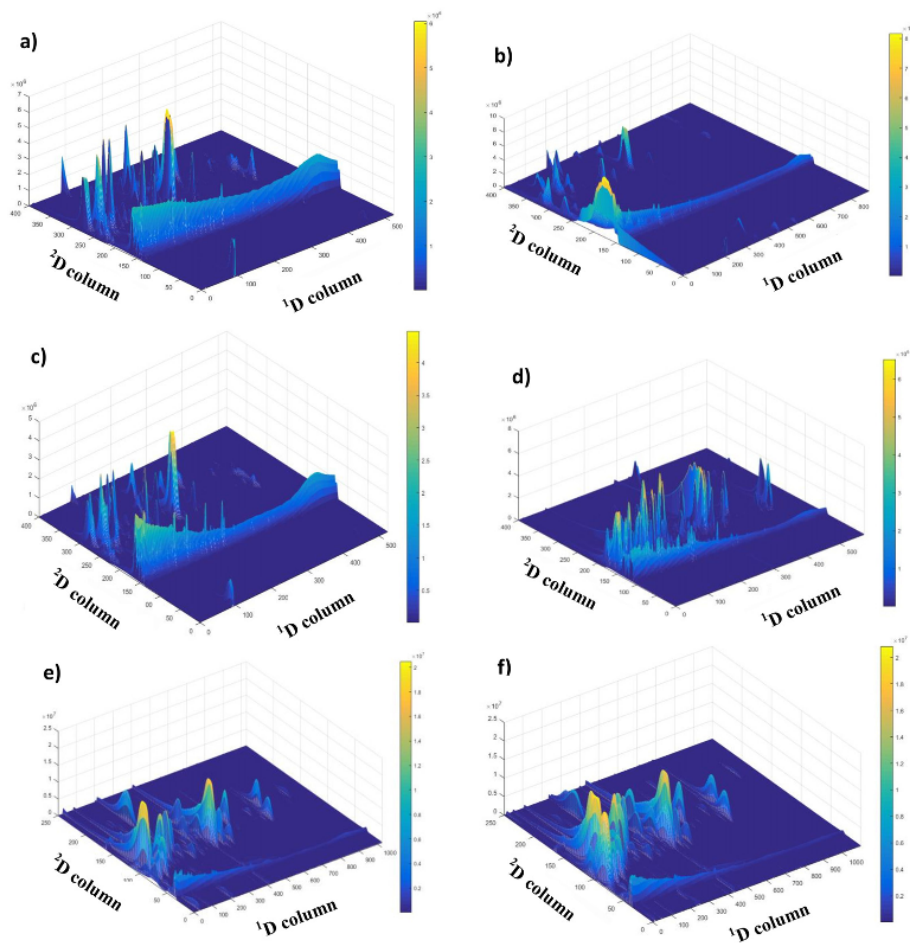


Fig. 4. Three-dimensional total ion chromatograms (TIC) of the *D. magna* sample extracts analyzed by GC×GC-TOFMS using different chromatographic configurations. TIC plot of: (a) S1, (b) S2, (c) S3, (d) S4, (e) S5, and (f) S6.

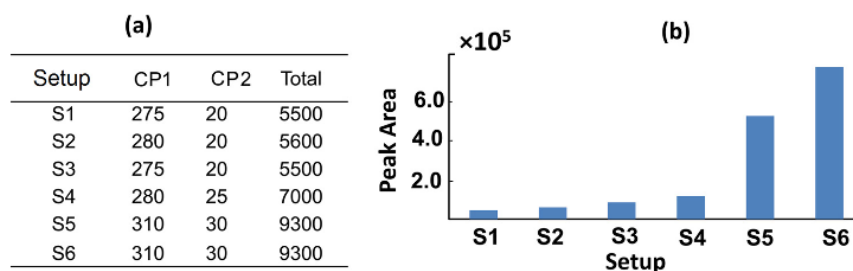


Fig. 5. (a) Calculated peak capacities of the different chromatographic configurations used for *D. Magna* samples analysis, PC1: peak capacity in first column, PC2: peak capacity in second column, Total: total peak capacity. (b) Peak area bar plots of metabolite 1 (lactic acid) in the six *D. Magna* samples analyzed by the different chromatographic configurations. The peak areas are obtained from C_{Superaug} (Section 3.3).

matrices from different samples or modulations could have different number of rows (retention times).

MCR-ALS results in the analysis of one *Daphnia* sample from a single GC×GC-MS chromatographic run (\mathbf{X}_{aug})

Fig. 7 shows MCR-ALS results in the analysis of sample analyzed by configuration S1, using the compressed and uncompressed column-wise augmented data matrices. Fig. 7 (a) shows the contour plot of the total ion chromatogram (TIC) of this sample. Chromato-

graphic signals appearing around scan number at 150 in second column and at all t_R were mostly only due to the derivatization reagents. In GC×GC analysis, derivatization of metabolites produced more than one reactant and hence elution profile for the same metabolite, which made the analysis of the full scan GC×GC-TOFMS data more challenging. The inset of Fig. 7 (a) shows the small part of the chromatogram which was analyzed (modulations 94–100) in more detail. This chromatographic region has a peak

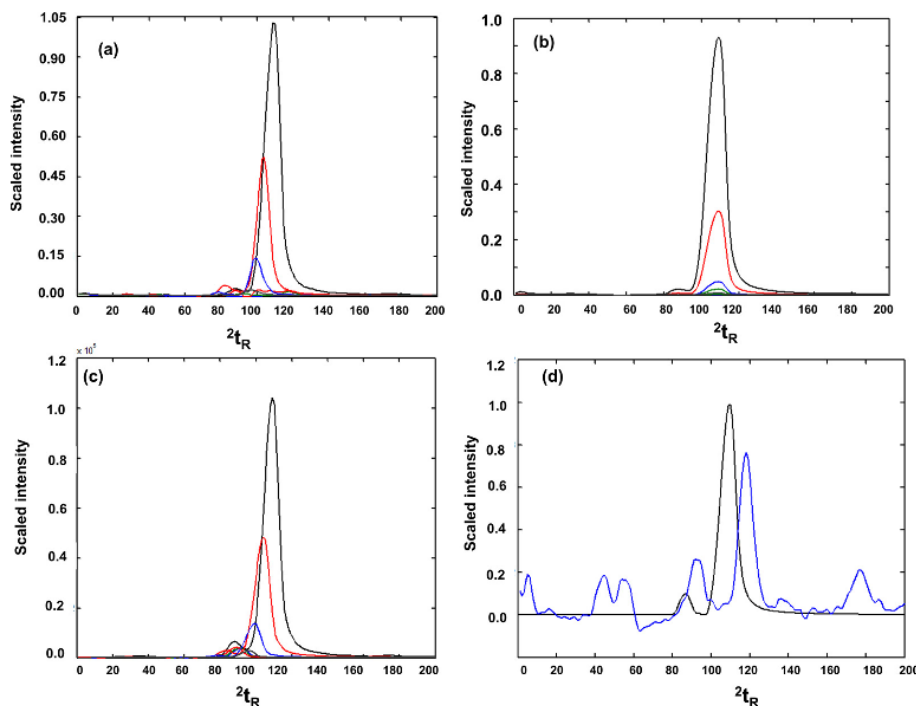


Fig. 6. $2D$ elution profiles of one of the resolved components (ribulose (MeOx) (4TMS)) on modulations 94–100 of sample S1 using different modeling methods: (a) Bilinear MCR-ALS, (b) synchronized trilinear MCR-ALS, (c) non-synchronized trilinear MCR-ALS, (d) PARAFAC (black) and PARAFAC2 (blue). Different colors in (a)–(c) correspond to MCR-ALS resolved $2D$ elution profiles of the same component in successive modulations. For brevity, only the elution profiles of one of the components are shown here. Only m/z 200–400 scan values were considered for all modulations analyzed here. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cluster of potential metabolites that are highly overlapped, and their identification would be rather difficult without using chemometric methods. Raw chromatographic and MS spectral data from this region were arranged into a column-wise augmented data matrix with 2800 rows and 250 m/z values in its columns. This data matrix needed 5.6 megabytes of memory storage. Fig. 7(b) shows the elution profiles of the metabolites resolved directly by MCR-ALS analysis of this single data matrix. The compressed data matrix using level-4 wavelets (see Section 3.2) had a reduced number of 175 rows, and the same number of columns (m/z). It only required computer storage of 0.35 megabytes (reduction in storage of 16 times). In Fig. 7(b) and (c), elution profiles resolved by MCR-ALS of these two augmented data matrices (raw- and compressed) are compared. Since data compression was performed only in the retention time direction, shapes of elution profiles in Fig. 7(b) and (c) were different, but resolved mass spectra profiles were practically the same. Fig. 7(d) shows the MCR-ALS resolved mass spectra obtained using the compressed augmented data matrix, in which the relevant chemical metabolites are already identified in the legend. These chemical compounds were tri methyl silyl (volatilization agent) derivatives including, 2, 2, 2-trifluoryl-*N*-methyl acetamide, trimethyl silanol and dimethyl silyl fluoride which were fragments of the derivatization agent. Other identified compound was pyridine which was used as derivatization media and it was present in all modulations. The two remaining components were the 1-Nonanol, and the MeOx and TMS derivatives of ribulose. Fig. 7(e) shows the final first column ($1D$) elution profiles and Fig. 7(f) shows the final second column ($2D$) elution profiles of the identified components (Section 3.3). Fig. 7(e) confirms that pyridine and other fragments of the derivatization agent were constantly present in all modulations, while the two metabolites of interest were modulated

for a limited number of times (typically within 2–4 modulations in the case understudy) from $1D$ to $2D$ columns. These results demonstrate the reliability of the proposed MCR-ALS analysis and its results.

LOF and R^2 (Eqs. (4) and (5)) values obtained in the MCR-analysis of the uncompressed data were equal to 6.05% and 99.63% respectively. These results were reasonable considering the experimental error level of the data. LOF and R^2 values of MCR-ALS analysis of the same data once were compressed were 2.11% and 99.95% respectively. This improvement is due to the noise filtering of the wavelet decomposition compression procedure, which removes low and high frequency noisy parts from data.

4.3. Elution profiles recovery after simultaneous analysis of two samples

Fig. 8(a) shows $1D$ elution profiles and Fig. 8(b) shows $2D$ elution profiles of modulations 94–100 when samples analyzed by S1 and S2 condition sets were simultaneously modeled by MCR-ALS using the column-wise super-augmented data matrix, $X_{superaug}$ (Fig. 3). In order to check for the differences between elution profiles obtained from compressed and uncompressed data, only scans 201–400 of modulations 94–100 are given in Fig. 8. Original raw chromatographic data were 16 times compressed (level-4 compression), with a change of computer storage from 2.8 megabytes to 0.175 megabytes. Compression was done for each sample column-wise augmented data matrix separately (as in previous Section 4.5) before building the two samples column-wise super-augmented data matrix. Four components were selected to model the systematic data variance observed for this particular GC \times GC-TOFMS data section. LOF (Eq. (4)) and R^2 (Eq. (5)) values were equal to 6.04%

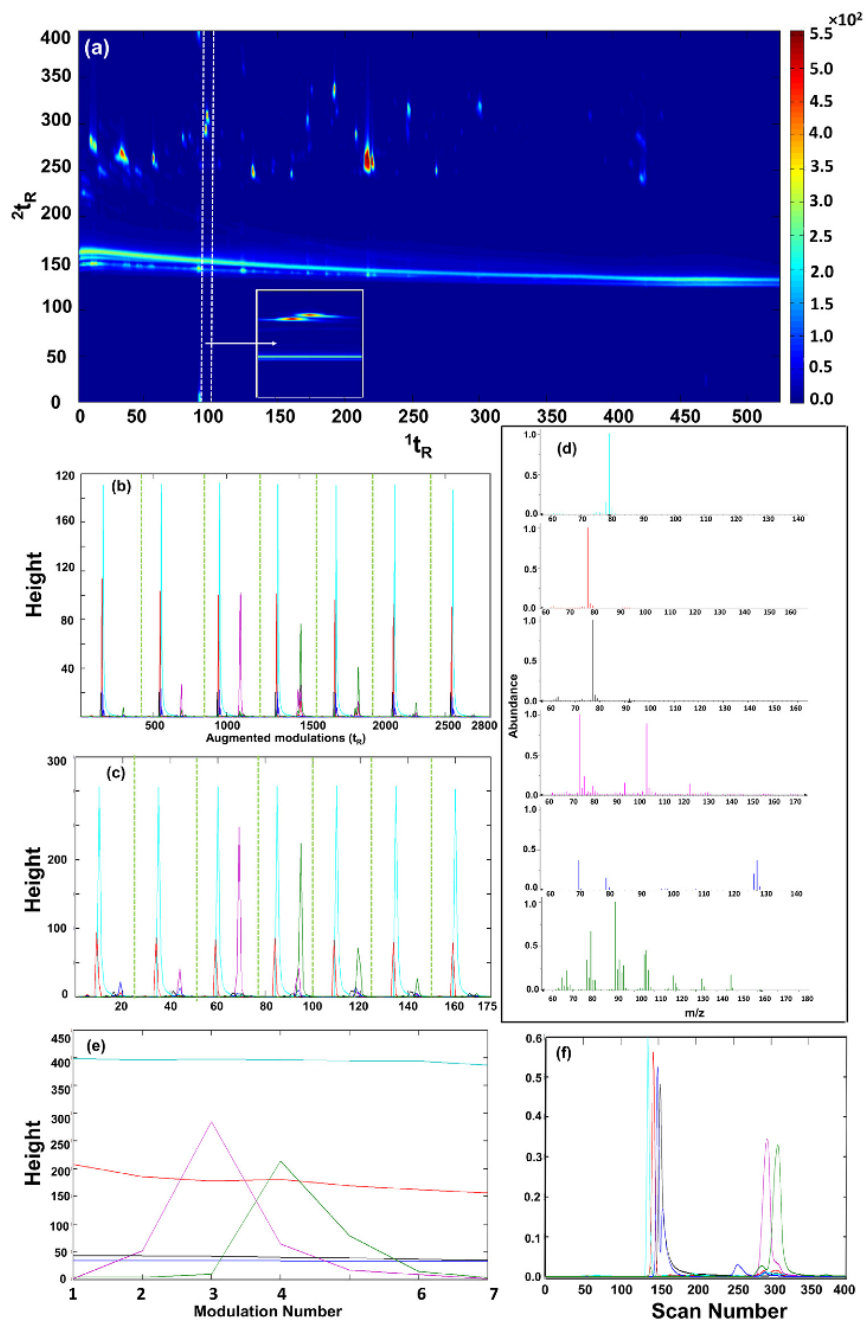


Fig. 7. MCR-ALS analysis of the column-wise augmented GC \times GC-TOFMS data matrix from modulations 94–100 of *Daphnia* sample S1. (a) Contour plot of total ion chromatogram (TIC) of sample S1. The inset figure shows the chromatographic region used for data analysis. (b) MCR-ALS resolved elution profiles from uncompressed data. Numbers in the x-axis are the scan number in every modulation, i.e., 1–400 shows the scan numbers of modulation 94, 401–800 shows the scan numbers of modulation 95 and so on. (c) MCR-ALS resolved elution profiles from compressed data. Numbers in x-axis are the scan number of compressed modulations, in the same order as in Fig. 7 (b). (d) MCR-ALS resolved mass spectra from compressed data. (e) 1D and (f) 2D elution profiles after recovering elution profiles in the two columns using non-negative least squares procedure described in Section 3.2. 2D elution profiles were normalized to one (Frobenius norm). Cyan: pyridine, Red: trimethyl silanol, Black: trimethylsilyl fluoride, Pink: 1-nonanol (TMS), blue: 2,2,2-trifluoro-*N*-methyl-acetamide, green: ribulose (Meox) (4TMS). The reader is recommended to see the web color version of this figure.

and 99.53 for the raw uncompressed data and 6.19% and 99.61% for compressed data. In this case, LOF values for wavelets compressed data were practically equal to uncompressed data. Compared to the previous case, where only one augmented data matrix was ana-

lyzed, the LOF values of uncompressed data are higher, because, between run differences for the two samples were also affecting the fit. Since the two chromatographic runs were wavelet compressed separately along their retention times, differences between sam-

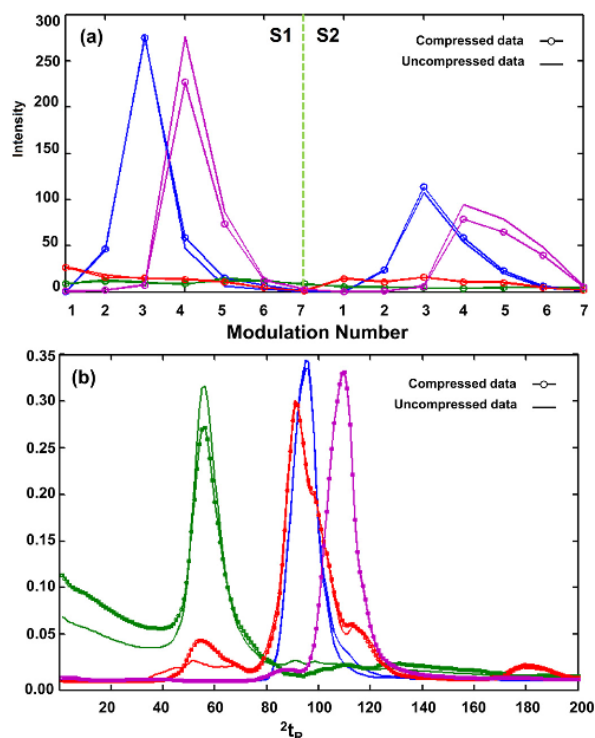


Fig. 8. elution profiles of (a) first column and (b) second column obtained by MCR-ALS simultaneous analysis of the column-wise super-augmented GC×GC-TOFMS (Eq. (3)), X_{Superaug} , compressed and uncompressed data matrices, for modulations 94–100 of two samples, analyzed using setups S1 and S2. Different colors are used for elution profiles of the different components resolved by MCR-ALS, and different symbols are used for elution profiles resolved from compressed and uncompressed chromatograms (see text in the figure).

ples raised and LOF values increased a little bit although wavelet compression reduced data noise for every run separately [51–53]. The two effects were compensating each other, and final fit resulted to be similar in both cases.

As shown in Fig. 8, MCR-ALS elution profiles obtained from raw uncompressed and wavelets compressed data, resulted to be very similar, which proved again the reliability of bilinear MCR-ALS results in the simultaneous analysis of the two samples analyzed by chromatographic setups S1 and S2. This shows that wavelet compression doesn't result in the loss of useful information and it only stores the same information in much lower digital space, because the elution profiles recovered from uncompressed and compressed data are nearly the same. Fig. 8 does show that peak heights/areas/shapes are well preserved after compression with little changes. Accepting these small changes might be a compromise that is necessary to reap the benefits of applying MCR-ALS to such a large dataset. This result indicated that, it could be possible to extend the proposed procedure of wavelet compression to the simultaneous analysis of multiple (more than two) chromatographic GC×GC-TOFMS runs (multiple samples) by MCR-ALS [54,55]. This will be especially helpful when multiple samples are compared and analyzed simultaneously (like in metabolomic studies), which due to the huge data storage requirements in these cases, the size reduction by wavelet procedure proposed here is necessary for fast metabolomics profiling.

Table 3
Computer storage and modeling time requirements of samples datasets analyzed by different chromatographic configurations (in megabytes, Mb) before and after compression.

Sample	Uncompressed (Mb)	Time (min) ^a	Compressed (Mb)	Time (min)
S1	419	480	26	30
S2	419	480	26	30
S3	708	760	44	47
S4	459	550	29	34
S5	641	680	32	42
S6	489	690	30	43
all samples	3135	3700	187	230

^a Time (minutes) required for data modeling by MCR-ALS.

4.4. MCR-ALS simultaneous analysis of multiple GC×GC-TOFMS chromatographic runs (X_{Superaug})

Datasets of *D. magna* samples analyzed by different chromatographic conditions listed in Table 1, were individually submitted to level-4 wavelet compression along the retention time direction, which computer storage reduced 16 times. Then, compressed data were column-wise augmented in a single super-augmented data matrix. Computer storages of every chromatographic run and time required for MCR-ALS modeling of every dataset before and after level-4 wavelet compression are given in Table 3. The computer storage for one single chromatographic run (sample) was around 550 megabytes and for the sum of the six chromatographic runs of uncompressed raw data matrices this resulted to be more than three gigabytes. This was too big for its direct simultaneous analysis by MCR-ALS. Without any compression this would need several hours to complete MCR-ALS modeling. Once data were compressed, much less time was needed (around one hour). After data compression, the computer storage required to store the same six chromatographic runs decreased up to 187 megabytes, which could be easily analyzed by MCR-ALS using current laboratory PC computers with only some few Gb of memory.

150 components were resolved by MCR-ALS analysis of the super-augmented compressed data matrix of six samples. Since the goal was to identify as many metabolites as possible, this large number was initially chosen to model most of the data variance presented in the data. Non-negativity constraints were applied to concentration and spectral profiles, and spectra were also normalized to have equal height. MCR-ALS LOF value was only 2.60% which indicated a very good data fitting, and R^2 was equal to 99.93%. By proposing larger number of components, no additional interpretable information was obtained.

Individual mass spectra of MCR-ALS resolved components were converted one-by-one into .txt files readable by NIST library. Elution profiles of the corresponding components in the two columns (1D and 2D) were also analyzed (Section 3.3). Retention index (RI) values of components were calculated at the maximum height of 1D elution profiles.

Metabolites were identified by the procedure described in section 0. Identified metabolites were distinguished from derivatization related components by comparison of the corresponding recovered chromatographic peaks. The derivatization neutralized the polar ends of the metabolites and byproducts of the derivatization were present during the whole chromatographic run giving irregular peak shapes. All components of this type were finally discarded and not used for identification. On the contrary, identified metabolites had regular chromatographic peak profiles and they were only modulated for a limited number of times from 1D column to 2D column. Only those MCR-ALS resolved components showing reasonable chromatographic peak shapes were finally considered for identification and were assumed to be linked to *D. magna* metabolome. A total number of 80 metabolites were finally

identified. Most of them were amino acids, organic acids and carbohydrates, from unstressed *D. magna* metabolism [28] and, mostly contained the tri methyl silyl masking agents from the derivatization agent. Table M1 in Supplementary material shows metabolites of *D. magna* that were identified, with their retention times, RI and RMF.

5. Conclusions

In this work, a fast and efficient workflow for untargeted metabolomic profiling of GC×GC-TOFMS data obtained in the analysis of crustacean *D. magna* samples is proposed. For this purpose, wavelet compression of chromatographic data sets in time direction is proposed to reduce data storage and allow their further high throughput analysis with currently used laboratory computers. Application of chemometric methods helped to overcome the incomplete highly coeluted separations of the metabolites in *D. magna* samples, background noise correction and multicomponent analysis challenges of GC×GC-TOFMS data sets. GC×GC-TOFMS data did not conform to the trilinear model requirements in general and use of the bilinear MCR-ALS chemometric method was preferred.

The chemometrics workflow proposed in this work was used for metabolomic profiling of crustacean species, *D. magna* and eighty different metabolites were identified. To find the best GC×GC-TOFMS conditions, metabolic extracts of *D. magna* were analyzed with six different chromatographic configurations. The best chromatographic configuration was selected by comparing peak capacities and peak areas of the resolved metabolites in each setup. The best column set included a relatively thick non-polar stationary phase in the first column: ZB-5MS (30 m × 0.25 mm ID × 0.25 μm film thickness), and a thick stationary film of intermediate polarity with the same internal diameter in the second column: TRB-50 (2.0 m × 0.25 mm ID × 0.25 μm film thickness). The combinations of these two columns in first and second chromatographic dimension provided the highest peak capacity among the other column sets studied in this work, and it is proposed for future metabolomics work.

Acknowledgements

The authors would like to thank Dr. B. Campos and Dr. C. Barata who supplied the biological samples for this study, and Dr. R. Chaler and D. Fanjul for GC×GC-TOFMS support. Yahya Izadmanesh acknowledges partial funding of his research stay in Institute of Environmental Assessment and Water Research (IDAEA) of CSIC by the Ministry of Science, Research and Technology (MSRT) of Iran. Romà Tauler and Elba Garreta-Lara acknowledge research funding from MINECO Spain grant Nr. CTQ2015-66254-C2-1-P and to CHEMAGEB project, (EU FP/2007-2013/ERC Grant Agreement n. 320737).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.chroma.2017.01.052>.

References

- [1] M.M. Koek, F.M. van der Kloet, R. Kleemann, T. Kooistra, E.R. Verheij, T. Hankemeier, Semi-automated non-target processing in GC×GC-MS metabolomics analysis: applicability for biomedical studies, *Metabolomics* 7 (2011) 1–14.
- [2] S. Ly-Verdú, T.M. Gröger, J.M. Arteaga-Salas, S. Brandmaier, M. Kahle, S. Neschen, M.H. de Angelis, R. Zimmermann, Combining metabolomic non-targeted GC×GC-ToF-MS analysis and chemometric ASCA-based study of variances to assess dietary influence on type 2 diabetes development in a mouse model, *Anal. Bioanal. Chem.* 407 (2015) 343–354.
- [3] H.D. Bean, J.E. Hill, J.-M.D. Dimandja, Improving the quality of biomarker candidates in untargeted metabolomics via peak table-based alignment of comprehensive two-dimensional gas chromatography–mass spectrometry data, *J. Chromatogr. A* 1394 (2015) 111–117.
- [4] G.J. Patti, R. Tautenhahn, D. Rinehart, K. Cho, L.P. Shriver, M. Manchester, I. Nikolskiy, C.H. Johnson, N.G. Mahieu, G. Siuzdak, A view from above: cloud plots to visualize global metabolomic data, *Anal. Chem.* 85 (2012) 798–804.
- [5] H. Parastar, J.R. Radović, J.M. Bayona, R. Tauler, Solving chromatographic challenges in comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry using multivariate curve resolution–alternating least squares, *Anal. Bioanal. Chem.* 405 (2013) 6235–6249.
- [6] H. Parastar, M. Jalali-Heravi, R. Tauler, Comprehensive two-dimensional gas chromatography (GC×GC) retention time shift correction and modeling using bilinear peak alignment, correlation optimized shifting and multivariate curve resolution, *Chemom. Intell. Lab. Syst.* 117 (2012) 80–91.
- [7] H. Parastar, J.R. Jalali-Heravi, M. Radovic, S. Diez, J.M. Bayona, R. Tauler, Resolution and quantification of complex mixtures of polycyclic aromatic hydrocarbons in heavy fuel oil sample by means of GC×GC-TOFMS combined to multivariate curve resolution, *Anal. Chem.* 83 (2011) 9289–9297.
- [8] H. Parastar, R. Tauler, Multivariate curve resolution of hyphenated and multidimensional chromatographic measurements: a new insight to address current chromatographic challenges, *Anal. Chem.* 86 (2013) 286–297.
- [9] J.R. Radović, K.V. Thomas, H. Parastar, S. Diez, R. Tauler, J.M. Bayona, Chemometrics-assisted effect-directed analysis of crude and refined oil using comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry, *Environ. Sci. Technol.* 48 (2014) 3074–3083.
- [10] D.K. Pinkerton, B.A. Parsons, T.J. Anderson, R.E. Synovec, Trilinearity deviation ratio: a new metric for chemometric analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data, *Anal. Chim. Acta* 871 (2015) 66–76.
- [11] R. Bro, PARAFAC. Tutorial and applications, *Chemom. Intell. Lab. Syst.* 38 (1997) 149–171.
- [12] H.A. Kiers, J.M. Ten Berge, R. Bro, PARAFAC2-Part I. A direct fitting algorithm for the PARAFAC2 model, *J. Chemom.* 13 (1999) 275–294.
- [13] A. De Juan, R. Tauler, Comparison of three-way resolution methods for non-trilinear chemical data sets, *J. Chemom.* 15 (2001) 749–771.
- [14] T. Gröger, M. Schäffer, M. Pütz, B. Ahrens, K. Drew, M. Eschner, R. Zimmermann, Application of two-dimensional gas chromatography combined with pixel-based chemometric processing for the chemical profiling of illicit drug samples, *J. Chromatogr. A* 1200 (2008) 8–16.
- [15] S. Furbo, A.B. Hansen, T. Skov, J.H. Christensen, Pixel-based analysis of comprehensive two-dimensional gas chromatograms (color plots) of petroleum: a tutorial, *Anal. Chem.* 86 (2014) 7160–7170.
- [16] K.M. Pierce, B.A. Parsons, R.E. Synovec, Pixel-level data analysis methods for comprehensive two-dimensional chromatography, *Fundam. Anal. Appl. Multi-way Calibration* 29 (2015) 427.
- [17] M.F. Almstetter, I.J. Appel, K. Dettmer, M.A. Gruber, P.J. Oefner, Comparison of two algorithmic data processing strategies for metabolic fingerprinting by comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry, *J. Chromatogr. A* 1218 (2011) 7031–7038.
- [18] P.M. Harvey, R.A. Shellie, Data reduction in comprehensive two-dimensional gas chromatography for rapid and repeatable automated data analysis, *Anal. Chem.* 84 (2012) 6501–6507.
- [19] J. Trygg, N. Kettaneh-Wold, L. Wallbäck, 2D wavelet analysis and compression of on-line industrial process data, *J. Chemom.* 15 (2001) 299–319.
- [20] I. Daubechies, Ten Lectures on Wavelets, Vol. 61 of CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, Pa USA, 1992.
- [21] K. Sayood, Introduction to data compression, *Newnes* (2012).
- [22] J.C. Cobas, P.G. Tahoces, M. Martin-Pastor, M. Penedo, F.J. Sardina, Wavelet-based ultra-high compression of multidimensional NMR data sets, *J. Magn. Reson.* 168 (2004) 288–295.
- [23] G. Chen, P.B. Harrington, Real-time two-dimensional wavelet compression and its application to real-time modeling of ion mobility data, *Anal. Chim. Acta* 490 (2003) 59–69.
- [24] S.G. Oliver, M.K. Winson, D.B. Kell, F. Baganz, Systematic functional analysis of the yeast genome, *Trends Biotechnol.* 16 (1998) 373–378.
- [25] C. Barata, J.C. Navarro, I. Varo, M.C. Riva, S. Arun, C. Porte, Changes in antioxidant enzyme activities, fatty acid composition and lipid peroxidation in *Daphnia magna* during the aging process, *Comp. Biochem. Physiol. B: Biochem. Mol. Biol.* 140 (2005) 81–90.
- [26] Oecd, OECD guidelines for the testing of chemicals, *Org. Econ.* (1994).
- [27] T. Vandenbrouck, O.A. Jones, N. Dom, J.L. Griffin, W. De Coen, Mixtures of similarly acting compounds in *Daphnia magna*: from gene to metabolite and beyond, *Environ. Int.* 36 (2010) 254–268.
- [28] E. Garreta-Lara, B. Campos, C. Barata, S. Lacorte, R. Tauler, Metabolic profiling of *Daphnia magna* exposed to environmental stressors by GC-MS and chemometric tools, *Metabolomics* 12 (2016) 1–14.
- [29] P. Wojtowicz, J. Zrostlíková, V. Štátná, E. Dostálová, L. Žídková, P. Bruheim, T. Adam, Comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry in human metabolomics, *Gas Chromatography-Biochemicals, Narcotics and Essential Oils*, Rijeka: InTech (2012) 29–50.

- [30] P. Korytár, H.-G. Janssen, E. Matisová, A.T. Udo, Practical fast gas chromatography: methods, instrumentation and applications, *TrAC Trends Anal. Chem.* 21 (2002) 558–572.
- [31] R. Tauler, Multivariate curve resolution applied to second order data, *Chemom. Intell. Lab. Syst.* 30 (1995) 133–146.
- [32] R. Tauler, A. Smilde, B. Kowalski, Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution, *J. Chemom.* 9 (1995) 31–58.
- [33] G.H. Golub, C.F. Van Loan, *Matrix computations*, JHU (2016) Press2012.
- [34] S.A. Bortolato, A.C. Olivieri, Chemometric processing of second-order liquid chromatographic data with UV–vis and fluorescence detection. A comparison of multivariate curve resolution and parallel factor analysis 2, *Anal. Chim. Acta* 842 (2014) 11–19.
- [35] A.K. Smilde, R. Tauler, J. Saurina, R. Bro, Calibration methods for complex second-order data, *Anal. Chim. Acta* 398 (1999) 237–251.
- [36] R. Tauler, I. Marqués, E. Casassas, Multivariate curve resolution applied to three-way trilinear data: study of a spectrofluorimetric acid–base titration of salicylic acid at three excitation wavelengths, *J. Chemom.* 12 (1998) 55–75.
- [37] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: new features and applications, *Chemom. Intell. Lab. Syst.* 140 (2015) 1–12.
- [38] H. Parastar, N. Akvan, Multivariate curve resolution based chromatographic peak alignment combined with parallel factor analysis to exploit second-order advantage in complex chromatographic measurements, *Anal. Chim. Acta* 816 (2014) 18–27.
- [39] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 11 (1989) 674–693.
- [40] Y. Meyer, *Wavelets and Operators*, Cambridge University Press, 1995.
- [41] M.B. Priestley, *Spectral Analysis and Time Series*, 1981.
- [42] R. Bro, S. De Jong, A fast non-negativity-constrained least squares algorithm, *J. Chemom.* 11 (1997) 393–401.
- [43] N. Strehmel, J. Hummel, A. Erban, K. Strassburg, J. Kopka, Retention index thresholds for compound matching in GC–MS metabolite profiling, *J. Chromatogr. B* 871 (2008) 182–190.
- [44] M.M. Koek, B. Muilwijk, L.L. van Stee, T. Hankemeier, Higher mass loadability in comprehensive two-dimensional gas chromatography–mass spectrometry for improved analytical performance in metabolomics analysis, *J. Chromatogr. A* 1186 (2008) 420–429.
- [45] C.F. Poole, *The Essence of Chromatography*, Elsevier, 2013.
- [46] J. Kopka, N. Schauer, S. Krueger, C. Birkemeyer, B. Usadel, E. Bergmüller, P. Dormann, W. Weckwerth, Y. Gibon, M. Stitt, L. Willmitzer, A.R. Fernie, D. Steinhauser, GMD@CSB.DB: the golm metabolome database, *Bioinformatics* 21 (2005) 1635–1638.
- [47] N. Schauer, D. Steinhauser, S. Strelkov, D. Schomburg, G. Allison, T. Moritz, K. Lundgren, U. Roessner-Tunali, M.G. Forbes, L. Willmitzer, A.R. Fernie, J. Kopka, GC–MS libraries for the rapid identification of metabolites in complex biological samples, *FEBS Lett.* 579 (2005) 1332–1337.
- [48] M. Misić, Y. Misić, G. Oppenheim, J.-M. Poggi, *Wavelet Toolbox*, The MathWorks Inc., Natick, MA, 1996.
- [49] R. Tauler, D. Barceló, Multivariate curve resolution applied to liquid chromatography–diode array detection, *TrAC Trends Anal. Chem.* 12 (1993) 319–327.
- [50] R. Tauler, M. Maeder, A. De Juan, Multiset data analysis: extended multivariate curve resolution|NOVA, Univ. Newcastle's Digital Repository (2009).
- [51] U. Dępczynski, K. Jetter, K. Molt, A. Niemöller, The fast wavelet transform on compact intervals as a tool in chemometrics. II. Boundary effects, denoising and compression, *Chemom. Intell. Lab. Syst.* 49 (1999) 151–161.
- [52] C. Mittermayr, S. Nikolov, H. Hutter, M. Grasserbauer, Wavelet denoising of Gaussian peaks: a comparative study, *Chemom. Intell. Lab. Syst.* 34 (1996) 187–202.
- [53] B.K. Alsberg, A.M. Woodward, M.K. Winson, J. Rowland, D.B. Kell, Wavelet denoising of infrared spectra, *Analyst* 122 (1997) 645–652.
- [54] E. Gorroategui, J. Casas, C. Porte, S. Lacorte, R. Tauler, Chemometric strategy for untargeted lipidomics: biomarker detection and identification in stressed human placental cells, *Anal. Chim. Acta* 854 (2015) 20–33.
- [55] F. Puig-Castellví, I. Alfonso, B. Piña, R. Tauler, A quantitative ¹H NMR approach for evaluating the metabolic response of *Saccharomyces cerevisiae* to mild heat stress, *Metabolomics* 11 (2015) 1612–1625.

“Chemometric analysis of comprehensive two dimensional gas chromatography-mass spectrometry metabolomics data”

Yahya Izadmanesh^{1,2}, Elba Garreta-Lara², Jahan B. Ghasemi³, Silvia Lacorte², Victor Matamoros², Roma Tauler^{2,*}

¹ Faculty of Chemistry, K. N. Toosi University of Technology, Tehran, Iran

² Institute of Environmental Assessment and Water Research, Spanish Council for Scientific Research (CSIC), Jordi Girona 18, Barcelona, 08034, Spain

³ Faculty of Chemistry, University of Tehran, Tehran, Iran

*Corresponding author: roma.tauler@idaea.csic.es

Table M1. Identified metabolites according to the retention time, retention index (RI) experimental and theoretical, and relative match factor (RMF).

Metabolites	Retention time (min)	RI (experimental)	Identification	RI (theoretical)	RMF
1	370	1070	Lactic acid (2TMS)	1044	838
2	922	1534	Pyroglutamic acid (2TMS)	1521	895
3	600	1280	Phosphoric acid (3TMS)	1262	891
4	905	1524	Unknown	-	-
5	464	1154	Norleucine (TMS)	1199	845
6	800	1456	Unknown	-	-
7	1510	1930	Mannose (5TMS)	1970	744
8	900	1520	Unknown	-	-
9	1407	1856	Idose (Meox)(5TMS)	1872	735
10	952	1553	Phenylalanine (TMS)	1556	898
11	578	1258	Unknown	-	-
12	450	1142	Valine (TMS)	1082	874
13	702	1387	Fumaric acid (2TMS)	1359	847
14	1007	1589	2-hydroxyglutaric acid (3TMS)	1569	837
15	450	1142	Unknown	-	-
16	482	1169	Proline ethyl ester	1172	907
17	677	1360	Uracil (2TMS)	1335	886
18	770	1436	Aspartic acid (2TMS)	1349	677
19	577	1257	Serine (2TMS)	1269	727
20	1407	1856	3-O-methylglucose (MeOx)(4TMS)	1808	548
21	1407	1856	Allose (MeOX) (5TMS) BP	1885	571
22	350	1050	1,3-Propanediol (2TMS)	1073	651
23	1511	1931	Unknown	-	-
24	1615	2008	Unknown	-	-
25	1338	1807	Unknown	-	-
26	487	1174	Phosphoric acid (2TMS)	1167	792
27	665	1348	Lumichrome (2Meox)	1332	545
28	682	1365	2-(2-Butoxyethoxy)ethoxy (TMS)	1348	780
29	450	1142	Oxalic (2TMS)	1136	777
30	520	1201	Valine (2TMS)	1220	893
31	700	1385	Alanine (3TMS)	1361	938
32	774	1439	Methionine (TMS)	1416	837
33	717	1401	1,9-Nonanediol	1401	841
34	477	1165	Unknown	-	-
35	427	1123	Glycine (2TMS)	1111	859
36	987	1576	Threonic acid (4TMS)	1545	854
37	1005	1587	Proline (2TMS)	1583	903
38	665	1348	Lumichrome (2Meox)	1332	904
39	1075	1632	1,3-Diaminopropane (4TMS)	1645	851

40	908	1525	Unknown	-	-
41	980	1572	Unknown	-	-
42	1512	1932	Pantothenic aci (3TMS)	1987	563
43	672	1355	Glyceric acid (3TMS)	1320	891
44	2287	2548	Unknown	-	-
45	600	1280	Glycerol (3TMS)	1283	912
46	620	1300	Nicotinic acid (TMS)	1301	956
47	562	1243	Benzoic acid (TMS)	1254	833
48	366	1066	Pyruvic acid (MeOx)(TMS)	1037	684
49	1385	1840	Tyrosine (2TMS)	1887	722
50	470	1159	3-hydroxybutanoic acid (2TMS)	1148	859
51	745	1420	Threonine (3TMS)	1383	689
52	1250	1746	Glycerol-2-phosphate (4TMS)	1714	590
53	890	1515	Malic acid (3TMS)	1479	841
54	380	1080	Hexanoic acid (TMS)	993	607
55	712	1397	3-hydroxynorvaline (2TMS)	1360	555
56	2287	2548	Unknown	-	-
57	1440	1880	Pentadecanoic acid methyl ester	1807	577
58	992	1579	Serine (4TMS)	1564	549
59	1424	1869	Unknown	-	-
60	1328	1800	Unknown	-	-
61	1615	2008	Unknown	-	-
62	945	1549	Ditertbutylphenol (TMS)	1541	896
63	1610	2004	Unknown	-	-
64	1140	1674	Ribose (MeOx) (4TMS)	1666	725
65	2295	2554	Unknown	-	-
66	1270	1760	Terephthalic acid (2TMS)	1788	872
67	405	1104	Oxalic acid (2TMS)	1136	686
68	857	1493	Aminomalonic acid (3TMS)	1416	752
69	1410	1859	Unknown	-	-
70	1315	1791	Citric acid (4TMS)	1805	728
71	772	1437	Phosphomycin (2TMS)	1359	604
72	652	1334	Succinic acid (2TMS)	1310	808
73	2240	2510	Unknown	-	-
74	800	1456	1,3-Benzenediol	1372	705
75	1197	1710	Putrescine (4TMS)	1735	613
76	780	1443	Tromethamine (3TMS)	1430	789
77	622	1302	Threonine (2TMS)	1290	883
78	1577	1890	Unknown	-	-
79	1787	2146	Unknown	-	-
80	410	1108	Oxalic acid (2TMS)	1118	602

2.5 ARTICLE CIENTÍFIC IV

Chemometrics comparison of gas chromatography with mass spectrometry and comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry Daphnia magna metabolic profiles exposed to salinity.

H. Parastar, E. Garreta-Lara, B. Campos, C. Barata, S. Lacorte, R. Tauler.

Journal of Separation Science 41 (2018), 2368-2379.


Received: 10 November 2017 | Revised: 17 February 2018 | Accepted: 20 February 2018

DOI: 10.1002/jssc.201701336

RESEARCH ARTICLE

JOURNAL OF
SEPARATION SCIENCE

Chemometrics comparison of gas chromatography with mass spectrometry and comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry *Daphnia magna* metabolic profiles exposed to salinity

Hadi Parastar^{1*}  | Elba Garreta-Lara^{2*} | Bruno Campos² | Carlos Barata² |
Silvia Lacorte² | Roma Tauler²

¹Department of Chemistry, Sharif University of Technology, Tehran, Iran

²Department of Environmental Chemistry, IDAEA-CSIC, Barcelona, Spain

Correspondence

Dr. Hadi Parastar, Department of Chemistry, Sharif University of Technology, P.O. Box 11155-3516, Tehran, Iran.

Email: h.parastar@sharif.edu; h.parastar@gmail.com

*Hadi Parastar and Elba Garreta-Lara contributed equally to this work.

The performances of gas chromatography with mass spectrometry and of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry are examined through the comparison of *Daphnia magna* metabolic profiles. Gas chromatography with mass spectrometry and comprehensive two-dimensional gas chromatography with mass spectrometry were used to compare the concentration changes of metabolites under saline conditions. In this regard, a chemometric strategy based on wavelet compression and multivariate curve resolution–alternating least squares is used to compare the performances of gas chromatography with mass spectrometry and comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry for the untargeted metabolic profiling of *Daphnia magna* in control and salinity-exposed samples. Examination of the results confirmed the outperformance of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry over gas chromatography with mass spectrometry for the detection of metabolites in *D. magna* samples. The peak areas of multivariate curve resolution–alternating least squares resolved elution profiles in every sample analyzed by comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry were arranged in a new data matrix that was then modeled by partial least squares discriminant analysis. The control and salt-exposed *daphnids* samples were discriminated and the most relevant metabolites were estimated using variable importance in projection and selectivity ratio values. Salinity de-regulated 18 metabolites from metabolic pathways involved in protein translation, transmembrane cell transport, carbon metabolism, secondary metabolism, glycolysis, and osmoregulation.

KEYWORDS

chemometrics, comprehensive two-dimensional gas chromatography, *Daphnia magna*, mass spectrometry, metabolomics

Article Related Abbreviations: KEGG, Kyoto Encyclopedia of Genes and Genomes; LV, latent variable; MCC, Matthew's Correlation Coefficient; MCR-ALS, multivariate curve resolution–alternating least squares; MeOX, methoxyamine; MSTFA, N-methyl-N-trimethyl-silyl trifluoroacetamide; NIST, National Institute of Standards and Technology; PLS, partial least

squares; PLS–DA, partial least squares discriminant analysis; RI, retention index; RMF, reverse match factor; SR, selectivity ratio; TIC, total ion chromatogram; TMCS, chlorotrimethylsilane; VIP, variable importance in projection

Conflict of interest: The authors have declared no conflict of interest.

1 | INTRODUCTION

In recent years, metabolomics has increasingly been adopted as a tool for the systemic study of metabolite profiles, their composition, and dynamics when affected by genetic modifications, physiological stimuli, environmental, nutritional, or other factors [1–4]. Metabolites are low-molecular-weight compounds (<1500 Da) formed by numerous biosynthetic and catabolic pathways within a biological system [5–7]. The ultimate ambition in metabolomics is the global detection and quantification of all metabolites.

Daphnia magna (*D. magna*), a freshwater crustacean, is used extensively as an aquatic test species being the object of standardized testing guidelines from the Organization for Economic Co-operation and Development [8]. *D. magna* is also a keystone species in freshwater food webs, thus it is also an ecological model species to study effects of global climate change, such as water temperature, salinity, and oxygen levels of freshwater bodies [8].

To date, multiple analytical platforms have been used in metabolomics for metabolic profiling and/or fingerprinting. NMR spectroscopy [8–10], GC–MS [11–15], and LC–MS [11–13,16] are frequently used techniques in metabolomics among others [17–22]. Compositional analysis by GC–MS is one of the most common approaches in metabolomics. Among hyphenated techniques, GC–MS is highly efficient, sensitive, reproducible, and widely used tool in metabolomics [11,23]. Moreover, it is quantitative and its compound identification capabilities are superior to other separation techniques because GC–MS instruments obtain mass spectra with reproducible fragmentation patterns, which allow for the creation of public databases [24]. Although conventional 1D GC provides adequate resolution for a variety of analytical tasks, it does not resolve the numerous components present in extracts of biological samples (e.g., polar and non-polar metabolites).

Comprehensive 2D GC (GC \times GC) has been reported as a very successful analytical technique for the analysis of volatile and semi-volatile organic compounds in complex samples [25–27]. In the last decade, GC \times GC–TOF–MS has emerged as a powerful technique suitable for the separation of complex mixtures because of its high resolution and high peak capacity [28,29]. However, complete separation of all detectable components cannot be achieved because of the extremely high complexity of biological samples, and because of limitations in experimental and instrumental conditions [29]. Chemometric resolution techniques can be used complementarily to address incomplete separation issues during GC \times GC–TOF–MS analysis of complex mixtures [30]. Different chemometric methods including parallel factor analysis [31], partial least squares (PLS) [32], and multivariate curve resolution-alternating least squares (MCR–ALS) [33,34] have been proposed for the analysis of

chromatographic data [30,35,36]. Among these methods, the MCR–ALS method is one of the more established chemometric techniques that can be used for untargeted analysis and its potential has been already shown for the analysis of different types of chromatographic problems and especially GC \times GC data [35,37–42].

We have recently proposed a general chemometric strategy for fast nontargeted GC \times GC–TOF–MS metabolic profiling of *D. magna* [27]. This strategy was based on GC \times GC–TOF–MS data compression in the retention time direction using wavelet transform and appropriate data matrix arrangement and modeling by MCR–ALS [27]. Since measured mass spectra for the same component had the same m/z ratios in all second column slices during the different GC \times GC–TOF–MS runs, column-wise data augmentation keeping in common these m/z values (in the columns of augmented data matrix) is an optimal strategy for MCR bilinear modeling of GC \times GC–TOF–MS data [27]. The advantage of using this strategy was that both retention time shifts and baseline/background contribution can be properly handled during the resolution process without any need to correct them before chemometric analysis [40].

The objective of the present study was application of this previously developed chemometric strategy (i.e., wavelet compression and MCR–ALS) for GC \times GC–TOF–MS metabolic profiling with the following new aspects: (1) comparison of the performance of GC–MS and GC \times GC–TOF–MS for untargeted metabolic profiling of *D. magna* in control and exposed samples to salinity as a metabolomics study; it is important to note that even though the advantages of GC \times GC–TOF–MS over GC–MS have been extensively discussed in literature, few real world applications of these two analytical platforms for metabolite discovery in complex samples have been shown in detail [43]. (2) Partial least squares discriminant analysis (PLS–DA) [32,44,45] was used as a tool to classify control and salt-exposed *D. magna* samples according to their peak areas obtained by MCR–ALS. MCR–ALS as a multi-set resolution algorithm for analysis of GC–MS and GC \times GC–MS data of multiple samples without need to correct chromatographic artifacts before analysis because of bilinear model assumption of this technique. Also, MCR–ALS model gives pure elution profiles in first and second chromatographic dimensions as well as pure mass spectra of metabolites. In this way, there is possibility of identifying existing components in the system and then calculate their relative concentrations (i.e., peak areas) to prepare a new matrix analyzed by PLS–DA to see the similarities and dissimilarities between control and exposed samples. (3) Discovery of the most important metabolites able to discriminate between control and salt exposed *D. magna* samples, using the variable importance in projection (VIP) and the selectivity ratio (SR) [46] methods. (4) Finding metabolic pathways of the identified metabolites using Kyoto Encyclopedia of Genes and

Genomes (KEGG) and their detailed interpretation from biochemical points of view.

2 | MATERIALS AND METHODS

2.1 | Chemicals and reagents

D-Glucose (U-13C6, 99%), used as the internal standard (IS), was supplied by Cambridge Isotope Laboratories (Andover, MA, USA). Triphenylamine (98%) also used as IS, was obtained from Sigma-Aldrich (St. Louis, MO, USA). Pyridine (anhydrous, 99.8%), chlorotrimethylsilane (TMCS), methoxyamine (MeOX) hydrochloride (98%) and *N*-methyl-*N*-trimethyl-silyl trifluoroacetamide (MSTFA; > 98.5%), used as derivatizing agents and the saturated alkane standard mixture for the performance test of GC systems from C₇ to C₃₀ were also obtained from Sigma-Aldrich (St. Louis, MO, USA). Hexane, methanol, and chloroform were analytical reagent grade, and NaCl was supplied by Merck (Darmstadt, Germany).

2.2 | Experimental animals

All experiments were performed using the characterized single clone F of *D. magna* maintained indefinitely as pure parthenogenetic cultures [8,47]. Photoperiod was set to 16 h light: 8 h dark cycle and temperature at 20 ± 1°C. Bulk cultures of 10 animals/L were maintained in American Society for Testing and Materials (ASTM) International hard water at high food ration levels (5 × 10⁵ cells/mL of *Chlorella vulgaris*, respectively), as described in Barata and Baird [47]. The culture medium was changed every other day and neonates were removed within 24 h.

2.3 | Experimental design to evaluate the effect of salinity

Four days old *D. magna* juveniles were used for the experimental part. Animals were obtained from bulk cultures of 100 individuals reared in 10 L of ASTM as described above in Section 2.2. Control samples were exposed to ASTM hard water without salt (0 g/L NaCl). Salinity exposed samples were prepared with 5 g/L NaCl in ASTM hard water and replicated six times. After 24 h of exposition, *daphnids* were collected from each jar, pooled in an Eppendorf, the water was removed and then samples were frozen in liquid N₂ and stored at -80°C before metabolite extraction.

2.4 | Extraction and derivatization protocols

Daphnids were suspended with 400 µL methanol, and 10 µL of the IS D-glucose (U-13C6, 99%) was added at a concentration of 50 µg/mL and bead beating balls were introduced to the

Eppendorf. Bead beating was sustained for 60 s at 30 pulses/s using the homogenizer settings to break *daphnids* shell. Then, bead beating balls were removed and 200 µL water and 400 µL chloroform were added before centrifugation at 10 000 × g for 15 min to separate the aqueous and the lipid phase. The aqueous phase of every sample was transferred to a new Eppendorf. Afterwards, the extract was evaporated to dryness with a Speedvac (Thermo Scientific, USA) at room temperature for 3 h. When samples were completely dried, they were stored at -80°C until analysis.

Aqueous samples were prepared for GC-MS and GC × GC-TOF-MS analysis by derivatization with MeOX in pyridine and MSTFA with 1% of TMCS as catalyst as previously described [8]. The dried extracts obtained were derivatized as follows. Sixty five microliter of MeOX (20 µg/µL) in pyridine was added to each sample. After mixing for 1 min, the mixture was incubated for 90 min at 30°C. Thereafter, 30 µL of MSTFA (1% TMCS) were added, vortex mixed for 1 min and then incubated for another 30 min at room temperature. Before injection, 10 µL of triphenylamine was added as IS in a concentration of 50 µg/mL and, *daphnid* derivatized extracts were finally filtered through a 0.22 µm filters (Ultrafree®-MC, Millipore). Then, the final polar fraction was transferred to a chromatographic vial and injected into GC-MS and GC × GC-TOF-MS instrument. GC-MS and GC × GC-TOF-MS analysis were performed within 12 h from derivatization. Two internal standards were used during the extraction protocol. D-Glucose (U-13C6, 99%) was used to determine the extraction efficiency and the triphenylamine (IS) to assess the stability of the chromatographic injections.

2.5 | Chromatographic conditions

A Trace GC-MS Thermo Scientific DSO II system operated in electron ionization mode at 70 eV and equipped with a ZB-5MS column (Phenomenex, 30 m length × 0.25 mm id × 0.25 µm film thickness) was used. The conditions of GC-MS were previously optimized by Garreta-Lara et al. [8] and are reported in supporting information.

The GC × GC-TOF-MS using a Pegasus 4D (LECO, St. Joseph, MI, USA), system which has a 6890 gas chromatograph (Agilent Technologies, Palo Alto, CA, USA) equipped with a split/splitless injector, a secondary oven to fit the secondary column, and a ZX1 (Zoex, Houston, TX, USA) two-stage thermal modulator. Liquid nitrogen was used to cool nitrogen gas for cold pulses. The conditions of the GC × GC-TOF-MS analysis were previously performed by Izadmanesh et al., [27] and are reported in supporting information.

2.6 | Chemometric strategies

Figure 1 shows the general framework used in this work as an example for GC × GC-TOF-MS data analysis of *D. magna*

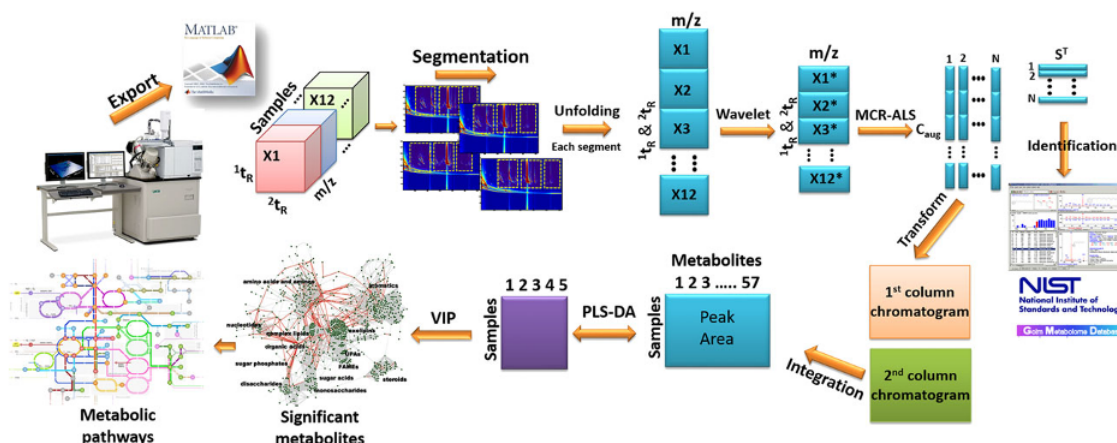


FIGURE 1 General chemometric strategy used in this work for the analysis of GC \times GC-TOF-MS data of *D. magna* samples

metabolites. The same strategy was used for GC-MS data but with different data arrangement. More details about these steps can be found in supporting information. GC-MS and GC \times GC-TOF-MS data of different *D. magna* samples (control and exposed) were transferred to the MATLAB numerical computing environment. For multiple samples, GC-MS data will be arranged in a column-wise augmented data matrix with elution time for different samples as rows and m/z values as columns. Data from GC \times GC-TOF-MS can be also column-wise super-augmented with elution times in first and second-chromatographic columns as rows and m/z values as columns. For multiple samples, this column-wise super-augmented data matrix was extended for different samples but with the same structure [37,39,40,48]. Due to the huge size of the data sets collected in GC-MS and GC \times GC-TOF-MS data sets of 12 *D. magna* samples, a data compression strategy based on the use of wavelets in time direction was used to make more feasible their chemometric analysis and reduce computer storage requirements [27]. Wavelet decomposition and compression were independently applied to every column (m/z) of original data matrix [49,50].

The column-wise augmented data matrix for GC-MS and the column-wise super-augmented data matrixes for four segments (A-D) for GC \times GC-TOF-MS in 12 samples were then analyzed using MCR-ALS.

The peak areas of resolved components were used to build a new data matrix having the changes in the relative concentrations of all resolved metabolites for 12 *D. magna* samples (control and exposed). Then, this new data matrix was used to find those metabolites that better discriminate between the two groups of samples (control and exposed), by means of PLS-DA. Matthew's Correlation Coefficient (MCC) [51] was calculated to validate the discrimination model. Finally, for the interpretation of the PLS-DA model and of the more important metabolic changes produced by salinity changes, VIP scores [52] and SR were used [46].

2.7 | Identification of metabolites

MS spectra profiles of MCR-ALS resolved components, both in the GC-MS and GC \times GC-TOF-MS analysis of the *D. magna* samples were assigned to metabolites and identified by comparing their mass fragmentation patterns with those in the standard mass spectral database of the National Institute of Standards and Technology (NIST; www.nist.gov/srd/nist1a.htm). To increase the reliability of the identification, internal linear retention indexes (RI) markers were also included in the evaluation of the library hits, using the injection of a saturated alkanes standard mixture (C_7 - C_{30}).

Metabolites whose concentrations had suffered statistically significant changes between control and treated samples because of salinity were identified through NIST library, and further mapped on KEGG database (www.genome.jp/kegg/ligand.html) [53] to investigate what were the possible metabolic pathways more affected due to the exposition of *D. magna* to salinity.

2.8 | Software

The GC-MS data sets were acquired using Xcalibur version 2.2 (Thermo Scientific, USA) and the GC \times GC-TOF-MS data sets were acquired using ChromaTOF version 3.3.2 (LECO, St. Joseph, MI, USA), converted to CDF format in the 60-700 m/z range and imported into MATLAB version R2015a (The Mathworks, Natick, MA, USA). NIST MS Search version 2.2 (National Institute of Standards and Technology, USA) and the Golm metabolome database of derivatized compounds were used for metabolite identifications. Wavelet toolbox of MATLAB was used for data compression. MCR-ALS toolbox was used for chemometric analysis [54,55]. PLS-DA modeling was performed using PLS-Toolbox version 7.8 (Eigenvector, USA).

3 | RESULTS AND DISCUSSION

3.1 | Chromatographic analysis of *D. magna* extracts

Chromatographic analysis of *D. magna* extracts from control and exposed salinity samples by GC-MS and GC \times GC-TOF-MS gave different information. Figure 2 shows two sample chromatograms of control and exposed *D. magna* extracts analyzed by GC-MS and GC \times GC-TOF-MS. Figures 2A and 2B depict the total ion chromatogram (TIC) for GC-MS data of one control sample and one salinity exposed *daphnia* sample, respectively. In addition, the TIC for GC \times GC-TOF-MS data of a control and a salinity exposed *daphnid* samples are shown in Figures 2C and 2D, respectively. As it is observed in Figure 2, TIC plot corresponding to the GC-MS data has a significantly lower number of chromatographic peaks compared with the TIC plot of GC \times GC-TOF-MS data that has a higher number of chromatographic peaks uniformly distributed along the whole 2D separation space (50 peaks for GC-MS and 100 peaks for GC \times GC-TOF-MS except for bleeding and derivatization agent's parts). This means that although *D. magna* extracts had a broad range of polar metabolites, they were not well separated using the 1D separation mechanism of GC-MS. Therefore, the analysis of the same samples by GC \times GC-TOF-MS provided higher peak capacities (9300 in this study) and better separations of the sample metabolites with diverse polarities. In this regard, whereas nonpolar compounds were separated in the first column according to their volatility, semipolar and polar compounds were better separated in the second nonpolar column according to their polarity. Another important aspect of GC-MS and GC \times GC-TOF-MS chromatograms was the presence of different chromatographic challenges, such as baseline/background contribution, column bleeding, S/N signal, and peak overlap. These problems were more severe in GC-MS compared with GC \times GC-TOF-MS. The possible reason for this phenomenon was the preconcentration of the analytes from the first chromatographic column in the modulator in GC \times GC-TOF-MS. All of these chromatographic challenges could be overcome favorably with the aid of chemometric techniques, increasing the number of metabolites detected by GC-MS and by GC \times GC-TOF-MS in the analysis of the 12 *D. magna* samples.

3.2 | Chemometric analysis of GC-MS data

GC-MS analysis of every sample extracts gave a data matrix with different elution times in the rows and m/z values in the columns. Chemometrics strategy based on wavelet compression was used to reduce the size of these individual data matrixes. For GC-MS data, a level-1 wavelet compression was enough to reduce the data volume and calculation time.

When a higher compression level was used, negative values appeared and some peaks with unreasonable chromatographic shapes were observed. Wavelet compressed individual data matrixes from the 12 different samples were then organized obtaining a new column-wise augmented data matrix, with the elution times for the different samples in the rows and m/z values in the columns of this data matrix. This new column-wise augmented wavelet compressed data matrix was then resolved using MCR-ALS method under application of non-negativity constraints of the resolved elution, spectra profiles, and spectra profiles normalization. Initial singular value decomposition (SVD) analysis [56] suggested the presence of at least 40 different components. Initial estimations of the spectra profiles of these 40 components were estimated from the purest retention times (like in the simple-to-use interactive self-modeling mixture analysis (SIMPLISMA) procedure [57]).

A summary of the MCR-ALS results obtained in the analysis of the GC-MS data from the 12 different control and salinity exposed *D. magna* extracts are shown in Figure 3. In Figure 3A, the column-wise augmented GC-MS data for the 12 *D. magna* samples are shown, the boundary between control and salt exposed samples are given. MCR-ALS resolved elution profiles and mass spectra profiles for the 40 components in the 12 sample extracts are shown in Figure 3B and 3C, respectively. To better follow the concentration changes of the metabolites in different control and exposed samples, resolved elution profiles were examined for more details. Figure 3D shows an example of the concentration changes (in terms of peak areas) of one of the metabolites found in the 12 simultaneously analyzed *D. magna* samples. As it is shown, peak area (concentration) of this metabolite was significantly reduced when *D. magna* samples were exposed to salinity conditions. To identify this metabolite, the corresponding MCR-ALS resolved mass spectrum (shown in Figure 3E) was compared with the template mass spectrum of the NIST Library and of Golm database (Figure 3F). Mass spectral matching confirmed that this resolved component was most probably glycerol (three trimethylsilane (3TMS) derivatized), considering its very good reverse match factor (RMF) of 904 obtained.

Identification of all resolved components by MCR-ALS method was attempted by comparison of their mass spectral profiles with the templates obtained from NIST and Golm databases. Among these 40 components, 22 of them had resolved profiles that could be clearly assigned to *D. magna* metabolites. The other components had resolved profiles that could be related to the derivatizing agents used, the solvent and baseline changes. Finally resolved components in GC-MS data using MCR-ALS and their corresponding information including RI, chemical name, derivatization order, KEGG identification, and RMF values are summarized in Supporting Information Table S1.

3.3 | Chemometric analysis of comprehensive GC \times GC-TOF-MS data

As it was mentioned before, due to the huge size of the data sets collected in the GC \times GC-TOF-MS analysis of the 12 *D. magna* samples, a data segmentation procedure and a wavelet compression strategy in time direction were additionally required. Before data segmentation, chromatographic sections related to column bleeding and derivatization agents were removed from the data. Supporting Information Figure S1a shows one raw GC \times GC-TOF-MS chromatogram for one of the *D. magna* control samples and Supporting Information Figure S1b shows the GC \times GC-TOF-MS chromatogram for the same sample after excluding chromatographic regions related to column bleeding and the derivatizing agents. Also, in Supporting Information Figure S1b, the corresponding GC \times GC-TOF-MS data segmentation into four parts (A–D) is given. Each chromatographic segment was compressed using level-3 wavelet compression to reduce its size before chemometric analysis and to reduce the computer storage requirements. For instance, using this compression strategy, the data size of one single sample obtained with the GC \times GC-TOF-MS analysis was reduced from 1.2 GB to 150 MB, from the raw data to the segmented and compressed data, respectively.

MCR-ALS investigation of full scan segmented/compressed data matrixes of GC \times GC-TOF-MS simultaneous analysis of 12 control and exposed *D. magna* samples allowed the simultaneous resolution of the elution and mass spectral profiles of a larger number of metabolites than in the case of the MCR-ALS analysis of the similar GC-MS simultaneous analysis of the same 12 control and exposed *D. magna* samples. As it was mentioned in Section 2, GC \times GC-TOF-MS data were segmented into four parts (A–D) and then were column-wise augmented and analyzed by MCR-ALS with the application of the constraints of nonnegativity of elution and spectra profiles and spectral normalization.

In Figure 4, an example of the MCR-ALS results for chromatographic segment B of the 12 samples (control and exposed samples) is shown. The number of resolved components in this segmented data matrix was 20. Figures 4A and 4B show their MCR-ALS resolved elution profiles in first- and second-chromatographic columns for control and salinity exposed samples, respectively. The boundary between the six control and six exposed samples is shown in these figures. Also, Figure 4C depicts the MCR-ALS resolved mass spectra of these 20 components in this data segment. In Figure 4D, first-column elution profiles for one of the resolved metabolite in 12 *D. magna* samples are plotted, and the second-column elution profiles are shown in Figure 4E. As observed

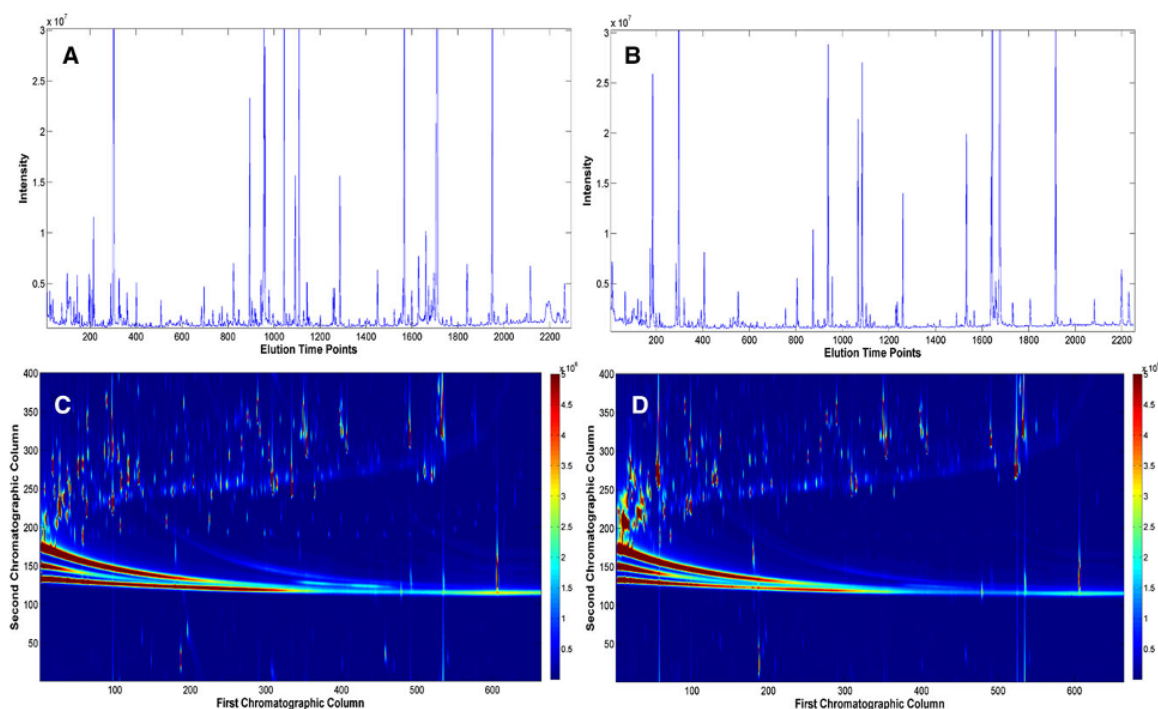


FIGURE 2 GC-MS and GC \times GC-TOF-MS of metabolite extracts from control and salinity exposed *D. magna* samples. (A) Plot of GC-MS TIC of one control sample and (B) of one salinity exposed sample. (C) Contour plots of GC \times GC-TOF-MS TIC of one control sample and (D) of one salinity exposed sample

in Figure 4E, whereas the concentration of this metabolite was very low in control samples (below LOD), this concentration increases considerably when samples were exposed to salinity. In Figure 4F, the MCR-ALS resolved mass spectrum of this metabolite is shown and compared with those in NIST library and Golm databases. This metabolite was identified as L-alanine (3TMS) with a RMF = 908. The lack of fit and R^2 of the MCR-ALS model were 5.59 and 99.69%, respectively, which were acceptable according to the noise level of the GC \times GC-TOF-MS raw data.

The same strategy was used for the other three chromatographic segments and additional metabolites were identified by comparing their MCR-ALS resolved mass spectra with NIST and Golm databases. In total, 90 components were resolved using MCR-ALS method. The values of R^2 for developed MCR-ALS models were higher than 99.60% for all segments. Also, lack of fit values were below 5.59% for all models. Among these components, 42 of them had resolved profiles that could be clearly assigned to *D. magna* metabolites. The other components had resolved profiles that could be related to the derivatizing agents used and with the solvent

and baseline changes. Components related to derivatized sub-products had resolved profiles with irregular chromatographic peak shapes and unreasonable mass spectra. All these additional components were finally discarded and not identified. Only those MCR-ALS resolved components showing reasonable chromatographic peak shapes and reliable mass spectra were finally considered for their identification and linked to *D. magna* metabolome.

Finally resolved peaks using MCR-ALS model of GC \times GC-TOF-MS data of *D. magna* samples and their corresponding information including RI, metabolite name, derivatization order, KEGG identification, and RMF are summarized in Supporting Information Table S2.

3.4 | Comparison of GC and GC \times GC

According to Figures 2 and 3, and 4, and Supporting Information Tables S1 and S2, several advantages of using GC \times GC-TOF-MS over GC-MS are observed. Firstly, the higher resolutions and increase of sensitivities of GC \times GC-TOF-MS compared to GC-MS are reasonable. In the GC-MS

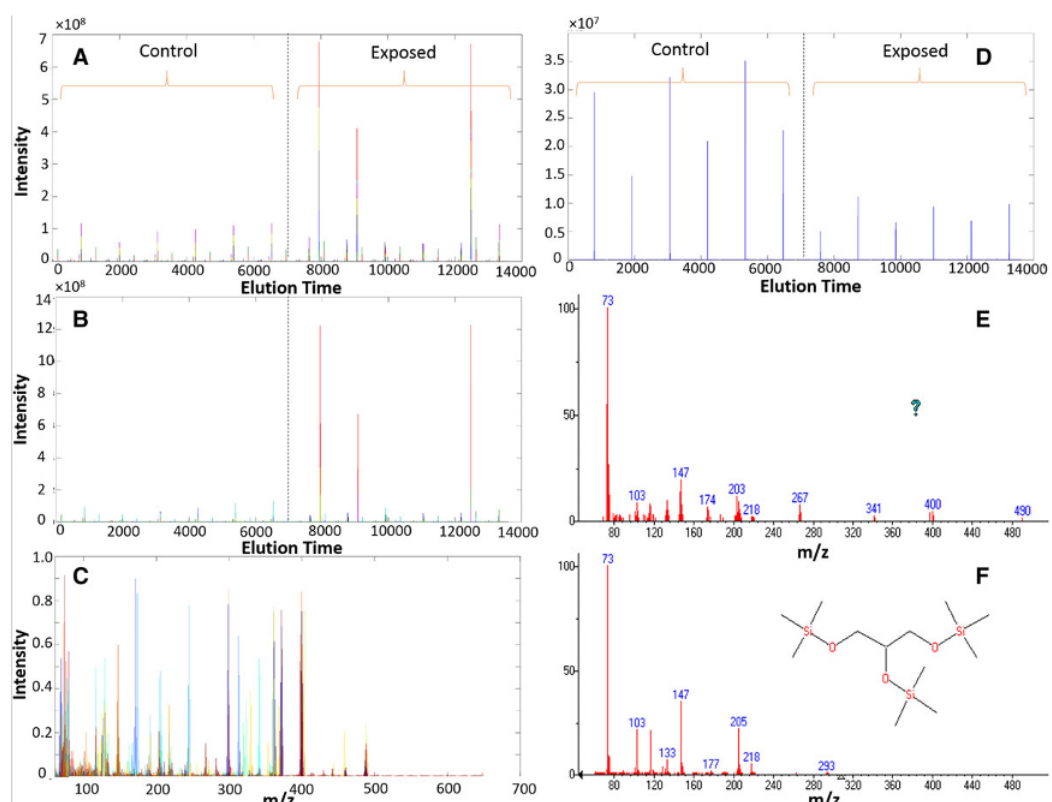


FIGURE 3 MCR-ALS analysis of GC-MS data of 12 *D. magna* samples (control and salinity exposed samples). The number of components was selected to be 40 in this case. (A) Plot of GC-MS raw data; (B) Plot of MCR-ALS resolved elution profiles; (C) Plot of the MCR-ALS resolved mass spectral profiles; (D) Plot of the resolved elution profiles for one metabolites in the 12 *D. magna* samples; (E) Plot of the resolved mass spectrum for the component shown in (D); and (F) Plot of the NIST library mass spectrum hint for glycerol (3TMS) with RMF = 904

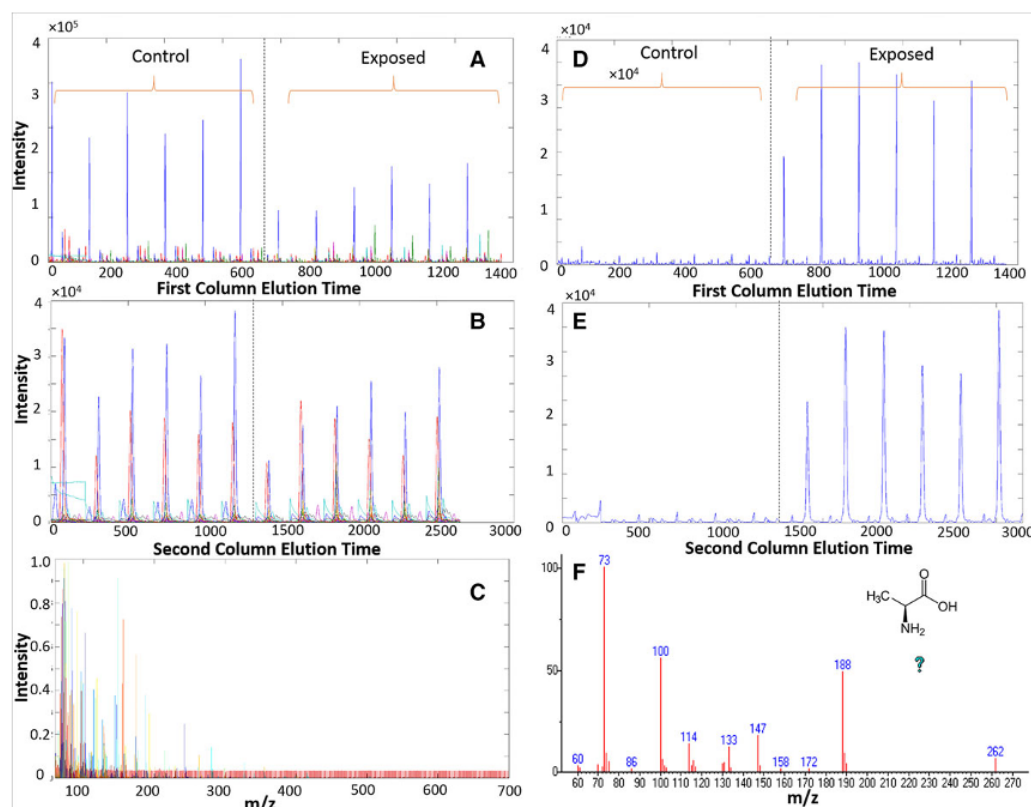


FIGURE 4 MCR-ALS analysis of segment B in GC \times GC-TOF-MS data of 12 *D. magna* samples (control and salinity exposed samples). The number of components was 20 in this data segment. (A) Plot of MCR-ALS resolved elution profiles in first chromatographic column, (B) Plot of MCR-ALS resolved elution profiles in second chromatographic column, (C) Plot of MCR-ALS resolved mass spectra profiles of the 20 components, (D) Plot of the MCR-ALS resolved first column elution profiles for one of the metabolites in the 12 samples, (E) Plot of the MCR-ALS resolved second column elution profiles for one of the metabolites in 12 samples and (F) Plot of the MCR-ALS resolved mass spectrum for the metabolite in subplots (D) and (F) which was identified as L-alanine (3TMS) with a RMF = 908 using NIST and Golm databases

analysis, many of the components were overlapped seriously when using a single column due to the diversity of the polarities of *D. magna* metabolites. In contrast, in the GC \times GC-TOF-MS analysis, orthogonal configuration of two chromatographic columns (nonpolar and polar) cause better resolution of *D. magna* metabolites and their MS spectra were cleaner than those obtained by GC-MS analysis. Secondly, as a result of this superior resolution and increased sensitivity, more peaks were identified in the GC \times GC-TOF-MS analysis (50 peaks in GC-MS and 100 peaks in GC \times GC-TOF-MS). For this particular *D. magna* extracted samples, the GC-MS analysis could only clearly identify 14 metabolites with RMF values above 800. Whereas with GC \times GC-TOF-MS analysis of the same samples, 30 metabolites were well resolved and identified with good RMF values above 800. Moreover, a well “structured” TIC chromatogram of GC \times GC-TOF-MS was achieved and further used to manually identify the metabolites of the samples. A significant advantage of GC \times GC-TOF-MS over GC-MS is the availability of a better baseline estimation

which can be then subtracted from peak integrals and compute more reliable peak areas associated to every metabolite. In GC \times GC-TOF-MS chromatograms, the background level is often easily estimated from a region close to the peak of interest. This leads to a more accurate and precise quantification of compounds in complex mixtures. On the contrary, baseline is often interfered by coeluting compounds in GC-MS chromatograms and more difficult to estimate and use for its subtraction.

3.5 | Classification of *D. magna* samples

After confirmation of the benefits of using GC \times GC-TOF-MS over GC-MS and the resolution and identification of 42 metabolites in control and exposed *D. magna* samples, the peak areas of the same MCR-ALS component profiles in every sample for all components were arranged in a new data table. This data table, with 42 metabolites peak areas in the 12 samples (**X** data block) was then correlated to the **y**-block

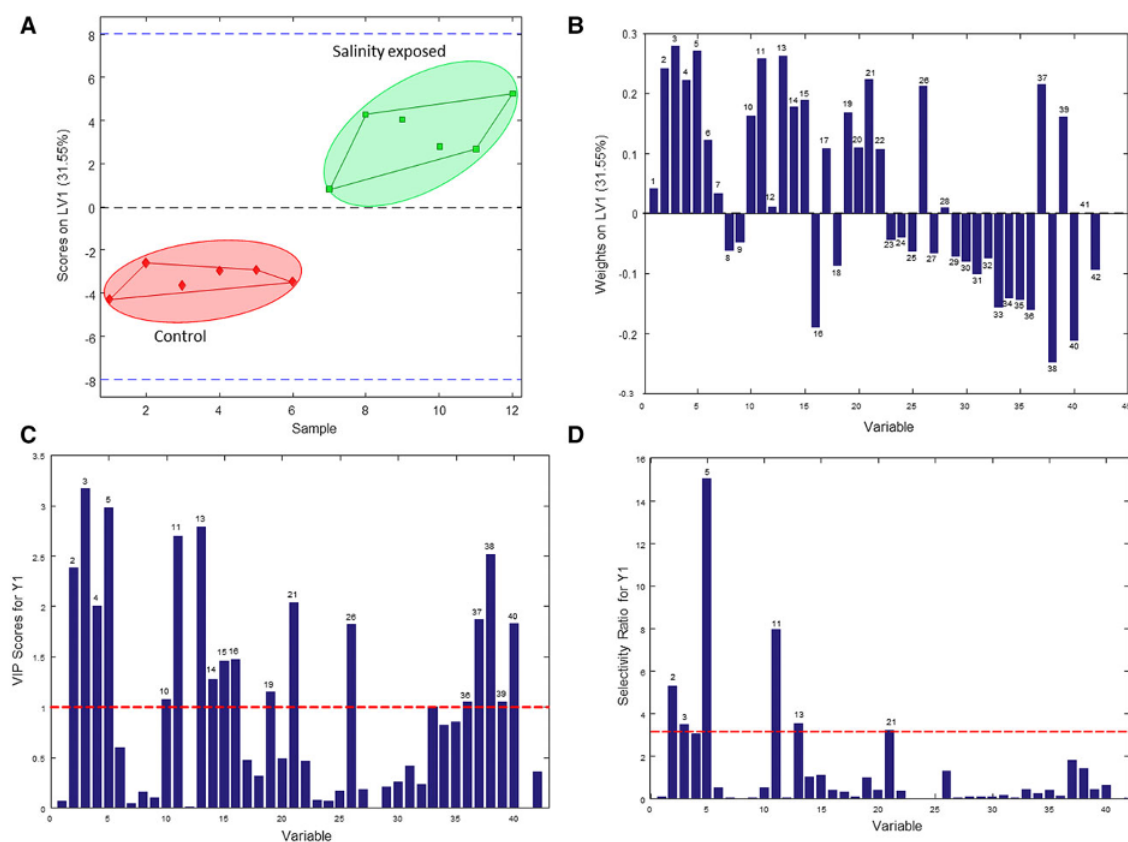


FIGURE 5 PLS-DA results in the analysis of the peak areas of the 42 MCR-ALS resolved metabolites in the 12 *D. magna* samples. (A) LV1 scores plot, (B) PLS weights vector on LV1, (C) PLS VIP scores plot and (D) PLS SR values plot

containing the sample codes for 12 control (0) and exposed (1) samples using PLS-DA. **X**-block data were autoscaled before PLS-DA analysis, and leave-one-out cross-validation was applied to determine the number of significant latent variables (LVs) needed in the PLS-DA model. Finally, one LV was finally considered necessary to capture 90.32% of the **y**-block variance and 31.55% of the **X**-block variance. MCC (see Section 2) for the PLS-DA model was equal to 1 indicating a very good classification between the two groups of samples (control and exposed *daphnids* samples). Correct positive (sensitivity = 1) and negative (specificity = 1) assignments were also achieved in this model.

Figure 5A shows the PLS-DA scores plot for the 12 analyzed samples. As it can be seen, all control and salinity exposed *D. magna* samples were well discriminated in this LV1 scores plot. PLS-DA weights, VIP scores, and SR showing the influence of each variable on the PLS-DA model are shown in Figures 5B–5D, respectively. As it is observed in Figure 5B, whereas PLS weights show the direction of the effect of each variable on PLS-DA model, VIP scores reflect what variables (metabolites) had the greatest influence on the discrimination among samples (Figure 5C). In this study, the

“greater than one rule” was used as a criterion for variable selection [46]. **X**-variables that had a VIPs above one had a higher than average influence in the PLS-DA model and therefore they were considered to be important in the discrimination between control and salinity exposed samples. In the case of SR values, the statistically selected SR threshold value (*F*-test, at $P = 0.05$) [58] was 3.14. A total of 18 metabolites were selected according to VIPs and, in a more restrictive way according to SR values, only six of these metabolites were selected. The number of selected variables by the SR method was lower than the selected by VIP scores method (Figure 5D). As described in [59], variables selected by VIPs method may be in some cases false positive candidates, while those selected by SR method may be false negative candidates. In this PLS-DA model, according to MCC, correct positive and negative assignments were achieved. The best approach should be performed according to the aim of the variable selection. Hence, in this case, for interpretative purposes, 18 metabolites (of the 42 included) of the *D. magna* metabolome were finally selected as the ones more significantly affected by the salinity exposure.

TABLE 1 Most disrupted KEGG metabolic pathways by salinity treatment

Metabolic pathways	^a N	^b Metabolites
ko02010 ABC transporters	6	PHO, ALA, SER, GLY, LEU, MAL
ko01110 Biosynthesis of secondary metabolites	6	SER, LEU, GLYC, BEN, LAC, AGLU
ko04978 Mineral absorption	4	PHO, ALA, SER, LEU
ko05230 Central carbon metabolism in cancer	4	ALA, SER, LEU, LAC
ko01200 Carbon metabolism	4	ALA, SER, GLYC, AGLU
ko04974 Protein digestion and absorption	3	ALA, SER, LEU
ko00970 Aminoacyl-tRNA biosynthesis	3	ALA, SER, LEU
ko01230 Biosynthesis of amino acids	3	ALA, SER, LEU

^aN: Number of altered metabolites.

^bMetabolites abbreviations: AGLU, alpha-D-glucose; ALA, Alanine; BEN, Benzoic acid; GLY, Glycerol; GLYC, Glycolic acid; LAC, Lactate; LEU, Leucine; MAL, Maltose; PHO, Phosphoric acid; SER, Serine.

3.6 | Metabolic pathways involved

Those metabolites whose concentrations changed significantly due to the salinity treatment and selected according to their VIP scores above 1 are shown in Figure 5C, with their fold changes relative to control treatments depicted in Supporting Information Table S3. Significant deregulated metabolites were then used in the KEGG metabolic pathways database to obtain enriched pathways.

Salinity upregulated 14 metabolites and downregulated four metabolites (Supporting Information Table S3). Eight of the 42 metabolites identified had a percentage of change higher than a 100% and showed significant differences respect to control samples. 4-Methylvaleric acid, 2-oxopentanoic acid, lactic acid, and glycolic acid were the metabolites with the highest percentage of change with a 369.1, 526.4, 536.6, and 298.7%, respectively (Supporting Information Table S3). Eight metabolic pathways were enriched with three or more metabolites (Table 1). Pathways with high impact (i.e., changes occurring in more important nodes in the network will have a higher impact on the pathway) included ATP-binding cassette (ABC) transporters, aminoacyl-tRNA biosynthesis (a key component in translation), carbon metabolism, and biosynthesis of secondary metabolites. Central carbon metabolism in cancer, biosynthesis of amino acids, protein digestion and absorption, and mineral absorption were also enriched, but had low impact, since changes occurred in marginal or relatively isolated positions in the pathway. Interestingly, salinity altered two metabolites of glycolysis, increasing fivefold lactic acid and decreased significantly glucose levels. The previous metabolic changes indicate that salinity probably impaired oxygen consumption

rates affecting *Daphnia* respiratory apparatus (i.e., pleopoda), and hence increase anaerobic metabolism that consumes glucose and generates lactic acid. There is reported evidence that salinity decreases respiration rates in *Daphnia* [60]. Specific metabolomic effects of salinity included increased levels of glycerol, which is related to cell osmoregulation [61].

4 | CONCLUDING REMARKS

As it was anticipated, GC × GC-TOF-MS outperformed GC-MS in the analysis of the complex metabolite extracts from *D. magna* samples analyzed in this work, with a significant increase in the number of detected and identified metabolites when the same samples were analyzed by both methods. This improvement can be related to the superior resolution and increased sensitivity of GC × GC-TOF-MS over GC-MS. Additionally, in the GC × GC-TOF-MS analysis, since the modulator focused and reinjected the extracts as a sharp chemical pulse into the second dimension GC column for further separation, *D. magna* metabolites were better resolved and their MS spectra were cleaner than those obtained by GC-MS. Combination of previously developed chemometric strategy (wavelet transform and MCR-ALS) [27] to GC-MS and to GC × GC-TOF-MS for metabolic profiling of *D. magna* samples allowed improved resolution of chromatographic and mass spectral profiles and therefore, allowed chromatographic challenges, such as baseline drift, elution time shift, and peak overlap to be better overcome. Thus, a clearer picture of all involved metabolites in control and salt exposed samples were obtained. Additionally, PLS-DA analysis of the peak areas of the MCR-ALS resolved GC × GC-TOF-MS elution profiles allowed the investigation of the most relevant metabolites changing their concentrations in control and exposed samples. These results also allowed the investigation of the most disrupted KEGG metabolic pathways following the exposure of the *D. magna* samples to salinity. In conclusion, the strategy proposed in this work confirmed the usefulness of the combination of GC-MS and GC × GC-TOF-MS with chemometric multivariate data analysis methods for untargeted metabolomics studies.

ACKNOWLEDGMENTS

Hadi Parastar would like to thank Sharif University of Technology (SUT) of Iran for financial support for a short sabbatical leave at IDAEA-CSIC. The research leading to these results received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement no. 320737.

ORCID

Hadi Parastar  <http://orcid.org/0000-0002-2552-6549>

REFERENCES

- Almstetter, M. F., Oefner, P. J., Dettmer, K., Comprehensive two-dimensional gas chromatography in metabolomics. *Anal. Bioanal. Chem.* 2012, *402*, 1993–2013.
- Liu, X., Locasale, J. W., Metabolomics: A primer. *Trends Biochem. Sci.* 2017, *42*, 274–284.
- Viant, M. R., Recent developments in environmental metabolomics. *Mol. Biosyst.* 2008, *4*, 980–986.
- Bundy, J. G., Davey, M. P., Viant, M. R., Environmental metabolomics: A critical review and future perspectives. *Metabolomics* 2009, *5*, 3–21.
- Ryan, D., Robards, K., Metabolomics: The greatest omics of them all? *Anal. Chem.* 2006, *78*, 7954–7958.
- Houten, S. M., Metabolomics: Unraveling the chemical individuality of common human diseases. *Ann. Med.* 2009, *41*, 402–407.
- Putri, S. P., Yamamoto, S., Tsugawa, H., Fukusaki, E., Current metabolomics: Technological advances. *J. Biosci. Bioengin.* 2013, *116*, 9–16.
- Garreta-Lara, E., Campos, B., Barata, C., Lacorte, S., Tauler, R., Metabolic profiling of *Daphnia magna* exposed to environmental stressors by GC-MS and chemometric tools. *Metabolomics* 2016, *12*, 86–95.
- Zhao, L., Huang, Y., Hu, J., Zhou, H., Adeleye, A. S., Keller, A. A., 1H NMR and GC-MS based metabolomics reveal defense and detoxification mechanism of cucumber plant under nano-cu stress. *Environ. Sci. Technol.* 2016, *50*, 2000–2010.
- Kruk, J., Doskocz, M., Jodłowska, E., Zacharzewska, A., Łakomic, J., Czaja, K., Kujawski, J., NMR techniques in metabolomic studies: A quick overview on examples of utilization. *Appl. Mag. Res.* 2017, *48*, 1–21.
- Naz, S., Moreira Dos Santos, D. C., García, A., Barbas, C., Analytical protocols based on LC-MS, GC-MS and CE-MS for non-targeted metabolomics of biological tissues. *Bioanalysis* 2014, *6*, 1657–1677.
- Wang, Y., Liu, S., Hu, Y., Li, P., Wan, J. B., Current state of the art of mass spectrometry-based metabolomics studies - A review focusing on wide coverage, high throughput and easy identification. *RSC Adv.* 2015, *5*, 78728–78737.
- Haggarty, J., Burgess, K. E., Recent advances in liquid and gas chromatography methodology for extending coverage of the metabolome. *Curr. Opin. Biotechnol.* 2017, *43*, 77–85.
- Han, P., Huang, Y., Hylands, P. J., Legido-Quigley, C., Assessment of *polygnum capitatum* Buch.-Ham. ex D. Don by metabolomics based on gas chromatography with mass spectrometry. *J. Sep. Sci.* 2016, *39*, 1979–1986.
- Špánik, I., Machyňáková, A., Recent applications of gas chromatography with high-resolution mass spectrometry. *J. Sep. Sci.* 2018, *41*, 163–179.
- Allwood, J. W., Goodacre, R., An introduction to liquid chromatography-mass spectrometry instrumentation applied in plant metabolomic analyses. *Phytochem. Anal.* 2010, *21*, 33–47.
- Muth, D., Kachlicki, P., Stobiecki, M., Mass spectrometry-Applications in the studies on plant metabolome. *Biotechnol.* 2007, *76*, 156–175.
- Wolfender, J. L., Glauser, G., Boccard, J., Rudaz, S., MS-based plant metabolomic approaches for biomarker discovery. *Nat. Prod. Commun.* 2009, *4*, 1417–1430.
- Yuan, W., Edwards, J. L., Capillary separations in metabolomics. *Bioanalysis* 2010, *2*, 953–963.
- Ibáñez, C., García-Cañas, V., Valdés, A., Simó, C., Novel MS-based approaches and applications in food metabolomics. *Trends Anal. Chem.* 2013, *52*, 100–111.
- Hirayama, A., Wakayama, M., Soga, T., Metabolome analysis based on capillary electrophoresis-mass spectrometry. *Trends Anal. Chem.* 2014, *61*, 215–222.
- Pocsfalvi, G., Stanly, C., Fiume, I., Vékey, K., Chromatography and its hyphenation to mass spectrometry for extracellular vesicle analysis. *J. Chromatogr. A* 2016, *1439*, 26–41.
- Kuehnbaum, N. L., Britz-Mckibbin, P., New advances in separation science for metabolomics: Resolving chemical diversity in a post-genomic era. *Chem. Rev.* 2013, *113*, 2437–2468.
- García, A., Barbas, C., Gas chromatography-mass spectrometry (GC-MS)-based metabolomics. *Methods Mol. Biol.* 2011, *708*, 191–204.
- de Lima, P. F., Furlan, M. F., de Lima Ribeiro, F. A., Pascholati, S. F., Augusto, F., In vivo determination of the volatile metabolites of saprotroph fungi by comprehensive two-dimensional gas chromatography. *J. Sep. Sci.* 2015, *38*, 1924–1932.
- Martins, C., Brandão, T., Almeida, A., Rocha, S. M., Metabolomics strategy for the mapping of volatile exometabolome from *Saccharomyces* spp. widely used in the food industry based on comprehensive two-dimensional gas chromatography. *J. Sep. Sci.* 2017, *40*, 2228–2237.
- Górecki, T., Harynuk, J., Panić, O., The evolution of comprehensive two-dimensional gas chromatography (GC×GC). *J. Sep. Sci.* 2004, *27*, 359–379.
- Mondello, L., Tranchida, P. Q., Dugo, P., Dugo, G., Comprehensive two-dimensional gas chromatography-mass spectrometry: A review. *Mass Spectrom. Rev.* 2008, *27*, 101–124.
- Seeley, J. V., Seeley, S. K., Multidimensional gas chromatography: Fundamental advances and new applications. *Anal. Chem.* 2012, *85*, 557–578.
- Pierce, K. M., Kehimkar, B., Marney, L. C., Hoggard, J. C., Synovec, R. E., Review of chemometric analysis techniques for comprehensive two dimensional separations data. *J. Chromatogr. A* 2012, *1255*, 3–11.
- Bro, R., PARAFAC. Tutorial and applications. *Chemom. Intell. Lab. Syst.* 1997, *38*, 149–171.
- Geladi, P., Kowalski, B. R., Partial least-squares regression: A tutorial. *Anal. Chim. Acta.* 1986, *185*, 1–17.
- Tauler, R., Multivariate curve resolution applied to second order data. *Chemom. Intell. Lab. Syst.* 1995, *30*, 133–146.
- Tauler, R., Smilde, A., Kowalski, B., Selectivity, local rank, 3-way data analysis and ambiguity in multivariate curve resolution. *J. Chemom.* 1995, *9*, 31–58.
- Zeng, Z. D., Hugel, H. M., Marriott, P. J., Chemometrics in comprehensive multidimensional separations. *Anal. Bioanal. Chem.* 2011, *401*, 2373–2386.

36. Amigo, J. M., Skov, T., Bro, R., Chromatography: Solving chromatographic issues with mathematical models and intuitive graphics. *Chem. Rev.* 2010, *110*, 4582–4605.
37. Parastar, H., Radović, J. R., Jalali-Heravi, M., Diez, S., Bayona, J. M., Tauler, R., Resolution and quantification of complex mixtures of polycyclic aromatic hydrocarbons in heavy fuel oil sample by means of GC×GC-TOFMS combined to multivariate curve resolution. *Anal. Chem.* 2011, *83*, 9289–9297.
38. Parastar, H., Jalali-Heravi, M., Tauler, R., Comprehensive two-dimensional gas chromatography (GC×GC) retention time shift correction and modeling using bilinear peak alignment, correlation optimized shifting and multivariate curve resolution. *Chemom. Intell. Lab. Syst.* 2012, *117*, 80–91.
39. Parastar, H., Radović, J., Bayona, J., Tauler, R., Solving chromatographic challenges in comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry using multivariate curve resolution–alternating least squares. *Anal. Bioanal. Chem.* 2013, *405*, 6235–6249.
40. Parastar, H., Tauler, R., Multivariate curve resolution of hyphenated and multidimensional chromatographic measurements: A new insight to address current chromatographic challenges. *Anal. Chem.* 2014, *86*, 286–297.
41. de Godoy, L. A. F., Hantao, L. W., Pedrosa, M. P., Poppi, R. J., Augusto, F., Quantitative analysis of essential oils in perfume using multivariate curve resolution combined with comprehensive two-dimensional gas chromatography. *Anal. Chim. Acta.* 2011, *699*, 120–125.
42. Bailey, H. P., Rutan, S. C., Chemometric resolution and quantification of four-way data arising from comprehensive 2D-LC-DAD analysis of human urine. *Chemom. Intell. Lab. Syst.* 2011, *106*, 131–141.
43. Winnike, J. H., Wei, X., Knagge, K. J., Colman, S. D., Gregory, S. G., Zhang, X., Comparison of GC-MS and GC×GC-MS in the analysis of human serum samples for biomarker discovery. *J. Prot. Res.* 2015, *14*, 1810–1817.
44. Wold, S., Sjöström, M., Eriksson, L., PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* 2001, *58*, 109–130.
45. Wold, S., Trygg, J., Berglund, A., Antti, H., Some recent developments in PLS modeling. *Chemom. Intell. Lab. Syst.* 2001, *58*, 131–150.
46. Chong, I. G., Jun, C. H., Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* 2005, *78*, 103–112.
47. Barata, C., Baird, D. J., Determining the ecotoxicological mode of action of chemicals from measurements made on individuals: Results from instar-based tests with *Daphnia magna* Straus. *Aqua. Toxicol.* 2000, *48*, 195–209.
48. Izadmanesh, Y., Garreta-Lara, E., Ghasemi, J. B., Lacorte, S., Matamoros, V., Tauler, R., Chemometric analysis of comprehensive two dimensional gas chromatography–mass spectrometry metabolomics data. *J. Chromatogr. A* 2017, *1488*, 113–125.
49. Shao, X., Cai, W., Pan, Z., Wavelet transform and its applications in high performance liquid chromatography (HPLC) analysis. *Chemom. Intell. Lab. Syst.* 1999, *45*, 249–256.
50. Walczak, B., Massart, D. L., Wavelets - something for analytical chemistry? *Trends Anal. Chem.* 1997, *16*, 451–462.
51. Matthews, B. W., Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *BBA Prot. Struct.* 1975, *405*, 442–451.
52. Wold, S., Sjöström, M., Eriksson, L., Partial least squares projections to latent structures (PLS) in chemistry. In: Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R. (Eds.), *Encyclopedia of Computational Chemistry*, John Wiley & Sons, Chichester, UK 1998, pp. 2006–2021.
53. Kanehisa, M., Goto, S. G. G., Kyoto encyclopedia of genes and genomes. *Nucl. Acids Res.* 2000, *28*, 27–30.
54. Jaumot, J., Gargallo, R., De Juan, A., Tauler, R., A graphical user-friendly interface for MCR-ALS: A new tool for multivariate curve resolution in MATLAB. *Chemom. Intell. Lab. Syst.* 2005, *76*, 101–110.
55. Jalali-Heravi, M., Parastar, H., Kamalzadeh, M., Tauler, R., Jaumot, J., MCRC software: A tool for chemometric analysis of two-way chromatographic data. *Chemom. Intell. Lab. Syst.* 2010, *104*, 155–171.
56. Wold, S., Esbensen, K., Geladi, P., Principal component analysis. *Chemom. Intell. Lab. Syst.* 1987, *2*, 37–52.
57. Windig, W., Lippert, J. L., Robbins, M. J., Kresinske, K. R., Twist, J. P., Snyder A. P., Interactive self-modeling multivariate analysis. *Chemom. Intell. Lab. Syst.* 1990, *9*, 7–30.
58. Rajalahti, T., Arneberg, R., Kroksveen, A. C., Berle, M., Myhr, K. M., Kvalheim, O. M., Discriminating variable test and selectivity ratio plot: Quantitative tools for interpretation and variable (biomarker) selection in complex spectral or chromatographic profiles. *Anal. Chem.* 2009, *81*, 2581–2590.
59. Farrés, M., Platikanov, S., Tsakovski, S., Tauler, R., Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *J. Chemom.* 2015, *29*, 528–536.
60. Arnér, M., Koivisto, S., Effects of salinity on metabolism and life history characteristics of *Daphnia magna*. *Hydrobiologia* 1993, *259*, 69–77.
61. Diamant, S., Eliahu, N., Rosenthal, D., Goloubinoff, P., Chemical chaperones regulate molecular chaperones in vitro and in cells under combined salt and heat stresses. *J. Biol. Chem.* 2001, *276*, 39586–39591.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Parastar H, Garreta-Lara E, Campos B, Barata C, Lacorte S, Tauler R. Chemometrics comparison of gas chromatography with mass spectrometry and comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry *Daphnia magna* metabolic profiles exposed to salinity. *J Sep Sci* 2018;41:2368–2379. <https://doi.org/10.1002/jssc.201701336>

Supporting Information

Chemometrics comparison of GC-MS and GC×GC-TOFMS *Daphnia magna* metabolic profiles exposed to salinity

Hadi Parastar^{a*}, Elba Garreta-Lara^b, Bruno Campos^b, Carlos Barata^b, Silvia Lacorte^b, Roma Tauler^b

^aDepartment of Chemistry, Sharif University of Technology, P.O. Box 11155-3516, Tehran, Iran

^bDepartment of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, 08034, Spain

Chromatographic conditions

GC-MS: The oven temperature program was set at 60°C and held for 2 minutes, increased at 7 °C min⁻¹ to 130 °C, and then to 310 °C at 5 °C min⁻¹ and held for 5 minutes, the delay time was 7.2 minutes and the time of analysis was 53 minutes. Other operating conditions were as follows: carrier gas was He, with a flow rate of 0.8 mL min⁻¹; source temperature, 250 °C; interface temperature, 280 °C; split ratio of 1:20; detector voltage, 440 V; 2 µL were injected in the split mode injection. The *m/z* values were recorded in full scan mode in the range of *m/z* 60-650 amu with a scan rate of 500 amu s⁻¹. The mass spectrometer was interfaced to a computer workstation running Xcalibur software (version 2.2, Thermo Scientific) for data acquisition and processing. Samples were run in a randomized order. For each detected peak, a linear retention index (RI) was calculated using GC retention index standards (hydrocarbons from C₇ to C₃₀ were used as internal standards). Therefore, an *n*-alkane standard mixture was prepared dissolved in hexane at a concentration of 10 µg µL⁻¹ and was injected at the same programmed conditions as the samples in the GC-MS instrument.

GC×GC-TOFMS: The first dimension column (¹D) consisted of ZB-5MS from Phenomenex (30 m × 0.25 mm Internal Diameter (ID) × 0.25 µm film thickness) coated with 5 % diphenyl, 95 %

dimethylpolysiloxane. The second dimension column (²D) was TRB-50 (2.0 m × 0.25 mm Internal Diameter (ID) × 0.25 μm film thickness) from Teknokroma coated with 50 % diphenyl, 50 % dimethylpolysiloxane.

Briefly, the primary oven temperature program was set at 60 °C for 2 minutes, ramped at 7 °C min⁻¹ to 130 °C, increased at 5 °C min⁻¹ to 260 °C and then to 295 °C at 5 °C min⁻¹ and held for 5 min. The secondary oven temperature was 5 °C higher than the preliminary oven. The helium carrier gas flow rate was 1 mL min⁻¹. The modulation period was 4 s, with 0.5 s hot pulse duration and a 15 °C modulator temperature offset versus the primary oven temperature. The MS transfer line was held at 250 °C, and TOFMS instrument, equipped with an electron ionization ion source, was set to scan mass range of 60-700 *m/z*. The ion source temperature was 200 °C. Detector operation was at 1800 V and applied electron energy was 70 eV. Acquisition rate was 100 spectra per second with unit mass resolution and 2.0 μL of each sample was injected via an auto-sampler (Agilent 7890) in splitless injection mode.

Chemometric strategies

GC-MS and GC×GC-TOFMS data of different *D. magna* samples (control and exposed) were transferred to the MATLAB numerical computing environment. The GC×GC-TOFMS data sets were in '.cdf' format and they were imported into MATLAB using Bioinformatics toolbox of MATLAB. Original full scan GC-MS data sets ".raw" from Xcalibur software (version 2.2, Thermo Scientific) were converted to ".cdf" data files by File Converter tool of Xcalibur Software to be further processed with MATLAB (The Mathworks, Inc., Natick, MA, USA). GC-MS data can be arranged in a two-way data matrix with elution time as rows and *m/z* values as columns. For multiple samples, GC-MS data will be arranged in a column-wise augmented data matrix with elution time for different samples as rows and *m/z* values as columns. Data from GC×GC-TOFMS

can be also column-wise super-augmented with elution times in first- and second-chromatographic columns as rows and m/z values as columns. For multiple samples, this column-wise super-augmented data matrix was extended for different samples but with the same structure [37,39,40,48]. It should be pointed out that the m/z values over 500 amu were removed from the data matrix for MCR-ALS analysis as they did not contain any information.

Due to the huge size of the data sets collected in GC-MS and GC×GC-TOFMS data sets of 12 *D. magna* samples, a data compression strategy based on the use of wavelets in time direction was used to make more feasible their chemometric analysis and reduce computer storage requirements [27]. Wavelet decomposition and compression were independently applied to every column (m/z) of original data matrix [49,50]. For the datasets under study in this paper, level-1 and level-3 wavelet compression were sufficient for GC-MS and GC×GC-TOFMS data, respectively without loss of relevant information in the elution time direction.

In addition, GC×GC-TOFMS data for 12 samples were segmented into four chromatographic sections (A-D) manually deciding the places to cut by visual inspection of the chromatograms. Furthermore, the bleeding parts of the chromatogram were excluded from the data before their analysis. GC-MS data were not segmented before MCR-ALS analysis as they can be analyzed simultaneously at a reasonable calculation time.

Since for all the chromatographic runs the same m/z range was selected (from 60 to 700 amu, a total of 640 m/z points), GC-MS and GC×GC-TOFMS data (segments A-D) for 12 *daphnids* samples were column-wise augmented with their m/z values in the common column mode. The column-wise augmented data matrix for GC-MS and the column-wise super-augmented data matrices for four segments (A-D) for GC×GC-TOFMS in 12 samples were then analyzed using Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS).

The MCR-ALS bilinear model for a data matrix is based on the description of the variation of the measurements as a linear mixture of the contributions of their pure components [25, 26]. This method can be straightforwardly extended to higher order data, i.e., to GC-MS and GC×GC-TOFMS data sets obtained in the analysis of different samples and arranged in an augmented data matrix. In linear algebra notation, the general MCR-ALS model applied to the augmented GC-MS data and to the super-augmented GC×GC-TOFMS data sets obtained in the simultaneous analysis of multiple samples is written as follows:

$$\mathbf{X}_{\text{aug}} = \mathbf{C}_{\text{aug}}\mathbf{S}^T + \mathbf{E}_{\text{aug}} \quad (\text{Eq. 1})$$

$$\mathbf{X}_{\text{supaug}} = \mathbf{C}_{\text{supaug}}\mathbf{S}^T + \mathbf{E}_{\text{supaug}} \quad (\text{Eq. 2})$$

where \mathbf{X}_{aug} is the column-wise augmented data matrix with dimensions $LK \times J$ (L samples, K elution times and J mass values) and $\mathbf{X}_{\text{supaug}}$ is the column-wise super-augmented data matrix with dimensions $LKI \times J$ (L samples, K data modulations taken from the first column, I second column elution time points and J m/z values) for GC-MS and GC×GC-TOFMS data, respectively. \mathbf{C}_{aug} ($KL \times N$) is the augmented matrix of resolved elution profiles for GC-MS data in different L samples. Also, $\mathbf{C}_{\text{supaug}}$ ($KLI \times N$) is the super-augmented matrix containing resolved second-dimension elution profiles for the different modulations in GC×GC-TOFMS data in different L samples. Note that the profiles of the same component in the different modulations may be different, both in shape and in peak position. The \mathbf{S}^T ($N \times J$) is the matrix of pure mass spectra profiles, and \mathbf{E}_{aug} and $\mathbf{E}_{\text{supaug}}$ are the residual matrices with the data variance unexplained by the bilinear model $\mathbf{C}_{\text{aug}}\mathbf{S}^T$ or by the model $\mathbf{C}_{\text{supaug}}\mathbf{S}^T$, respectively. In addition, N is the number of

chemical components considered in the factor matrices. MCR-ALS solves Eqs. 1 and 2 for \mathbf{C} and \mathbf{S}^T , using an iterative algorithm based on two constrained linear least-squares steps [25, 26].

Before starting MCR-ALS analysis, some prior knowledge was required. An initial estimation of the number of components is usually obtained from the size of the singular values of the analyzed data matrix and from the changes in explained variances (R^2) and of lack of fit (LOF) of MCR-ALS results (see Eqs. 3 and 4 below). When the size of singular values does not show any additional significant change (apart from experimental noise) and R^2 and LOF values already reached a good data fitting, no further components are added to the model. Another possible test is from the shapes of the profiles of the resolved components. Only those components having elution profiles with reasonable peak shapes and mass spectra are finally considered. In some cases, it is also convenient to consider some extra components to model solvent and background contributions to the measured signals. The final selection of the number of components is therefore a problem dependent choice and should be judged considering different aspects and also what is the final goal of the study. In this particular case, the goal is to resolve the elution and spectra profiles of those metabolites whose concentrations show major changes due to the stressing factor, in this case the salinity.

The explained variance (R^2) and the lack of fit (LOF) of the considered model are calculated as follows:

$$R^2 (\%) = \sqrt{\frac{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (\hat{x}_{aug,i,j,k})^2}{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K x_{aug,i,j,k}^2}} \times 100 \quad (\text{Eq. 3})$$

$$LOF(\%) = \sqrt{\frac{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x_{aug,i,j,k} - \hat{x}_{aug,i,j,k})^2}{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K x_{aug,i,j,k}^2}} \times 100 \quad (\text{Eq. 4})$$

where $x_{aug,i,j,k}$ is the element of the original data matrix and $\hat{x}_{aug,i,j,k}$ is the recovered value using MCR-ALS method.

Apart from the number of components, to start ALS optimization, an initial rough estimation of the profiles of the component in one of the two modes is required. In the case of GC-MS (or GC×GC-MS) data, the most obvious choice is to select the purest MS spectra measured experimentally which can be easily find using a procedure similar to the simple-to-use interactive self-modeling mixture analysis (SIMPLISMA) [39] method. In addition, some natural constraints fulfilled by the investigated system like non-negativity (concentration and spectral profiles), and other like pure spectra normalization (to unit length) to avoid scale ambiguities during ALS optimization were applied to get reliable solutions and minimum rotational ambiguity [40].

\mathbf{C}_{supaug} concentration matrix in Eq. 2 for GC×GC-TOFMS data has the second dimension elution profiles for all N components in K modulations in L samples. To get first dimension elution profiles of every component in each analyzed sample, every column of \mathbf{C}_{supaug} should be appropriately refolded to give a matrix for each analyzed sample. The column sum of this refolded data matrix gives an estimation of the corresponding first dimension elution profile. Therefore, for every sample, a matrix of first dimension elution profiles was additionally obtained. On the other side, MCR-ALS resolves the second dimension elution profiles (pure modulated peaks) of the metabolites in the presence of baseline/background contributions and of other interference signals. Thus, relative quantitative information for one target metabolite in different samples can be then directly derived from the comparison of the area of its MCR-ALS resolved second dimension

elution profile obtained in these different samples. After resolving the GC×GC-TOFMS data for 12 *D. magna* samples (control and exposed), the resolved elution profiles in two chromatographic columns along with their mass spectra were obtained. The pure mass spectra were used for the identification of the resolved components (i.e., metabolites) by comparing them with those of standard compounds in NIST *Mass Spectral Search 2.2* program distributed with the NIST 2014 library (<http://chemdata.nist.gov/mass-spc/ms-search>) and the Golm metabolome database (GMD) (<http://gmd.mpimp-golm.mpg.de/download>).

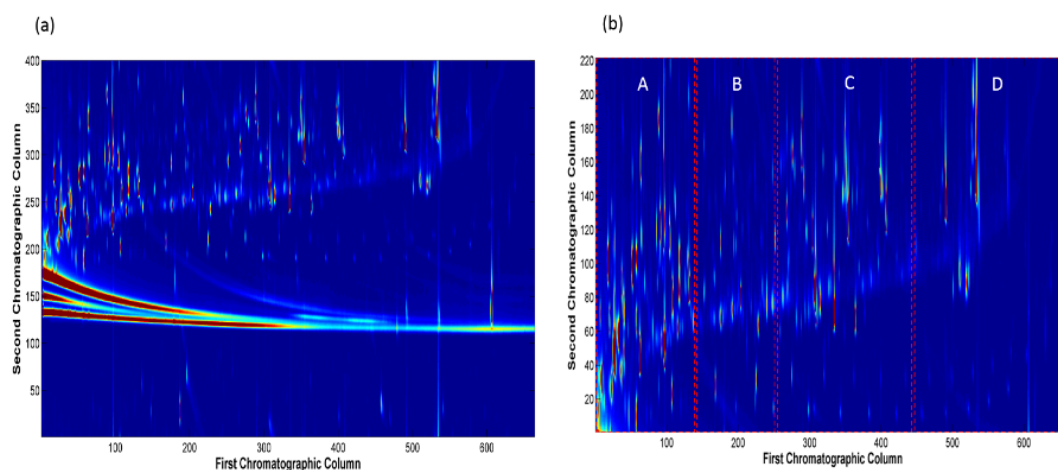


Figure S1. GC×GC-TOFMS total ion chromatogram (TIC) of one of control samples. Contour plots of (a) a raw GC×GC-TOFMS chromatogram and (b) of a GC×GC-TOFMS chromatogram after excluding the bleeding and derivatization parts. Four different chromatographic segments are marked in this figure.

The relative concentrations of the identified metabolites in different samples from GC×GC-TOFMS analysis were obtained using their peak areas by simple peak integration of resolved elution profiles in this second chromatographic dimension, for each component. In other words, after resolving the second-order data matrix, a matrix containing pure elution profiles for each analyte will be obtained (i.e., a vector containing the elution profile for each metabolite). To obtain

the peak area for each metabolite, the elements of this vector were summed and this procedure is equivalent to the peak integration.

These peak areas were used to build a new data matrix having the changes in the relative concentrations of all resolved metabolites for 12 *D. magna* samples (control and exposed). Then, this new data matrix was used to find those metabolites that better discriminate between the two groups of samples (control and exposed), by means of Partial Least Squares-Discriminant Analysis (PLS-DA). Matthew's Correlation Coefficient (MCC) [51] was calculated to validate the discrimination model. MCC is a correlation coefficient between the observed and predicted binary classifications; returned values are between +1 and -1. A coefficient of +1 indicates a perfect prediction, a value of 0 represents no better than a random prediction and a value of -1 indicates a total disagreement between prediction and observation.

Finally, for the interpretation of the PLS-DA model and of the more important metabolic changes produced by salinity changes, Variable Importance in Projection (VIP) scores [52] and Selectivity Ratio (SR) were used [46]. VIP scores give information about how the variables combine to form the quantitative relation between \mathbf{X} (the matrix with the peak areas of the resolved metabolites) and \mathbf{y} (a column vector designing the control and salinity stressed samples, labeled 0 and 1, respectively), thus providing a better evaluation of their relative importance in the model. Hence, these VIP scores are useful for the understanding of which metabolites are important (numerically large VIPs) in the discrimination between control and salinity exposed samples.

The SR index of each variable is defined as the ratio between the explained variance and its residual variance on target projected (TP) component. The variables with high SR index are expected to have good discrimination power. In order to find a reasonable threshold in SR for variable selection, F-test method is proposed [35].

Identification of metabolites

For each mass spectrum, 100 hits were retrieved by the NIST Mass Spectral Search 2.2 software. A reverse match factor (RMF) based on the correlation coefficient between the MCR-AS resolved and experimental mass spectra reported by NIST software was used for selection of the best-identified compounds for MCR-ALS resolved mass spectra. This match factor is reported between 0 (no match) and 1000 (perfect match). As a general guide, a value of 900 or greater was considered to be a very good matching; between 800 and 900, a good match; between 700 and 800, a fair match; and less than 600 a poor or very poor match.

Table S1. Identification of the 22 *D. magna* metabolites resolved by MCR-ALS including their GC-MS retention index (RI), metabolite name, derivatization order, KEGG identification and RMF values.

No.	RI _{Exp.} *	RI _{Theo.} **	Δ RI***	Metabolite	Derivatization	KEGG-ID	RMF
1	940	910	+30	Glycine	1TMS	C00037	895
2	1069	1081	-12	Glycolic acid	2TMS	C00160	915
3	1120	1094	+26	Valine	1TMS	C00183	766
4	1180	1136	+44	Oxalic acid	2TMS	C00209	814
5	1258	1266	-8	Ethanolamine	3TMS	C00189	805
6	1340	1286	+54	Phosphoric acid	3TMS	C00009	830
7	1347	1289	+58	Glycerol	3TMS	C00116	904
8	1350	1363	-13	Alanine	3TMS	C00041	808
9	1360	1390	-30	Malic acid	3TMS	C00149	876
10	1480	1558	-78	1-Dodecanol	1TMS	C02277	910
11	1591	1651	-60	Lyxose	4TMS	C00476	806
12	1690	1753	-63	Galactopyranose	4TMS	C00124	715
13	1880	1883	-3	Glucose	5TMS	C00031	868
14	2000	1997	+3	Galacturonic acid	6TMS	C00333	798
15	2070	2050	+20	Palmitic acid	1TMS	C00249	772
16	2130	2113	+17	Myo-Inositol	6TMS	C00137	712
17	2200	2218	-18	Oleic acid	1TMS	C00712	809
18	2272	2246	+26	Stearic acid	1TMS	C01530	920
19	2700	2715	-15	Trehalose	8TMS	C01083	780
20	2759	2749	+10	Alpha-D-Glucose	8TMS	C00267	763
21	2850	2806	+44	Glycerol monostearate	2TMS	D01947	793
22	3140	3150	-10	Cholestrol	1TMS	C00187	862

* RI based on NIST library

** RI based on the retention times of normal hydrocarbons (C₇-C₃₀)

*** Δ RI = RI_{Exp.} - RI_{Theor.}

Table S2. Identification of the 42 *D. magna* metabolites resolved by MCR-ALS including their GC×GC-TOFMS retention index (RI), metabolite name, derivatization order, KEGG identification and RMF values.

No.	RI _{Exp.} *	RI _{Theo.} **	ΔRI***	Metabolites	Derivatization	KEGG-ID	RMF
1	932	919	+13	Glycine	1TMS	C00037	730
2	1060	1040	+20	4-methylvaleric acid	1TMS	C00123	778
3	1062	1048	+14	2-Oxopentanoic acid	1TMS	C06255	831
4	1065	1066	-1	Lactic acid	2TMS	C00186	939
5	1075	1081	-6	Glycolic acid	2TMS	C00160	915
6	1090	<i>none</i>	<i>none</i>	Buthylamine	1TMS	C18706	794
7	1110	1094	+16	Valine	1TMS	C00183	766
8	1117	1125	-8	Hydroxylamine	3TMS	C00192	827
9	1150	1136	+14	Oxalic acid	2TMS	C00209	814
10	1192	1179	+13	Isoleucine	1TMS	C00407	713
11	1210	1249	-39	Benzoic acid	1TMS	C00180	953
12	1230	1252	-22	Serine	2TMS	C00065	783
13	1260	1266	-6	Ethanolamine	3TMS	C00189	853
14	1270	1282	-12	1-Toluic acid	1TMS	C01454	763
15	1280	1286	-6	Phosphoric acid	3TMS	C00009	874
16	1300	1289	+11	Glycerol	3TMS	C00116	897
17	1340	1322	+18	Thymol	1TMS	C09908	719
18	1345	1350	-5	Putrescine	4TMS	C00134	838
19	1350	<i>none</i>	<i>none</i>	1,3-diaminopropane	4TMS	C00986	808
20	1360	1355	+5	Nonanoic acid	1TMS	C01601	877
21	1380	1363	+17	Alanine	3TMS	C00041	908
22	1400	1390	+10	Malic acid	3TMS	C00149	876
23	1405	1397	+8	Threose	3TMS	C06463	806
24	1440	1407	+33	Decanoic acid	1TMS	C01571	935
25	1503	1538	-35	N-Acetyl-L-glutamic acid	2TMS	C00624	882
26	1540	1558	-18	1-Dodecanol	1TMS	C02277	910
27	1680	1651	+29	Lyxose	4TMS	C00476	806
28	1760	1756	+4	1-Tetradecanol	1TMS	D05097	787
29	1780	1775	+5	L-Glycerol-3-phosphate	4TMS	C03189	910
30	1843	<i>none</i>	<i>none</i>	Octadecadienoic acid	1TMS	C01595	828
31	1900	1928	-28	Methyl galactoside	4TMS	C03619	844
32	1990	1985	+5	Glucose	5TMS	C00031	828
33	2020	2037	-17	Glucopyranose	5TMS	C00031	914

34	2070	2050	+20	Palmitic acid	1TMS	C00249	926
35	2090	<i>none</i>	<i>none</i>	Inositol	6TMS	C00137	706
36	2108	2091	+17	N-Acetylgalactosamine	4TMS	C01074	708
37	2110	2113	-3	Myo-Inositol	6TMS	C00137	712
38	2240	2218	+22	Oleic acid	1TMS	C00712	809
39	2258	2246	+12	Stearic acid	1TMS	C01530	920
40	2700	2749	-49	Alpha-D-Glucose	8TMS	C00267	907
41	2753	2769	-16	Maltose methoxyamine	8TMS	C00208	820
42	2820	2806	+14	Glycerol monostearate	2TMS	D01947	866

* RI based on NIST library

** RI based on the retention times of normal hydrocarbons (C₇-C₃₀)

*** $\Delta RI = RI_{Exp.} - RI_{Theor.}$

Table S3. Fold changes, percentages of concentration changes and trends of the 42 *D. magna* metabolites selected in **Figure 5**.

No.	Metabolites	KEGG-ID	Fold changes	% changes	Trends
1	Glycine	C00037	1.0	4.7	-
2	4-Methylvaleric acid	C00123	4.7*	369.1*	UP
3	2-Oxopentanoic acid	C06255	6.3**	526.4**	UP
4	Lactic acid	C00186	6.4**	536.6**	UP
5	Glycolic acid	C00160	4.0**	298.7**	UP
6	Valine	C00183	1.2	18.8	-
7	Hydroxylamine	C00192	1.0	5.0	-
8	Oxalic acid	C00209	0.9	-15.0	-
9	Isoleucine	C00407	0.9	-13.5	-
10	Benzoic acid	C00180	2.4*	136.3*	UP
11	Serine	C00065	2.2**	117.4**	UP
12	Ethanolamine	C00189	1.0	3.2	-
13	1-Toluic acid	C01454	2.3**	131.0**	UP
14	Phosphoric acid	C00009	1.8*	83.9*	UP
15	Glycerol	C00116	1.7**	70.5**	UP
16	Thymol	C09908	0.7**	-30.3**	DOWN
17	Putrescine	C00134	1.4	35.1	-
18	Nonanoic acid	C01601	0.8	-20.1	-
19	Alanine	C00041	1.2*	21.8*	UP
20	Malic acid	C00149	1.2	21.5	-
21	Threose	C06463	2.7**	171.6**	UP
22	Decanoic acid	C01571	1.1	12.6	-
23	N-Acetyl-L-glutamic acid	C00624	0.9	-10.9	-
24	1-Dodecanol	C02277	0.9	-7.2	-
25	Lyxose	C00476	0.9	-10.8	-
26	1-Tetradecanol	D05097	1.9**	88.0**	UP
27	L-Glycerol-3-phosphate	C03189	0.9	-11.4	-
28	Methyl galactoside	C03619	1.0	0.8	-
29	Glucose	C00031	0.8	-24.6	-
30	Glucopyranose	C00031	0.7	-31.1	-
31	Palmitic acid	C00249	0.5	-47.6	-
32	N-Acetylgalactosamine	C01074	0.8	-17.5	-
33	Myo-Inositol	C00137	0.5	-52.3	-
34	Oleic acid	C00712	0.7	-30.3	-
35	Stearic acid	C01530	0.3	-71.3	-
36	Alpha-D-Glucose	C00267	0.8*	-15.2*	DOWN
37	Maltose methoxyamine	C00208	1.6**	55.3**	UP
38	Glycerol monostearate	D01947	0.5**	-49.4**	DOWN
39	Buthylamine	C18706	1.7*	65.9*	UP
40	1,3-diaminopropane	C00986	0.6**	-44.0**	DOWN
41	Octadecadienoic acid	C01595	1.0	0.1	-
42	Inositol	C00137	0.8	-19.3	-

2.6 DISCUSSIÓ DELS RESULTATS

En aquesta secció es discuteixen els resultats obtinguts en les publicacions incloses en aquest capítol. A més de les diferents plataformes analítiques emprades per a l'estudi del metaboloma de la *D. magna* s'ha posat de manifest l'estudi dels factors abiòtics relacionats amb el canvi climàtic. En la **Figura 2.2** es mostra un esquema de les diferents fases de treball que s'han seguit en els treballs presentats en aquest capítol, des de l'optimització de les condicions analítiques per a cada una de les plataformes analítiques emprades fins a la interpretació de la rellevància biològica dels resultats obtinguts, passant per l'anàlisi de les mostres i el pretractament i posterior tractament de les dades.

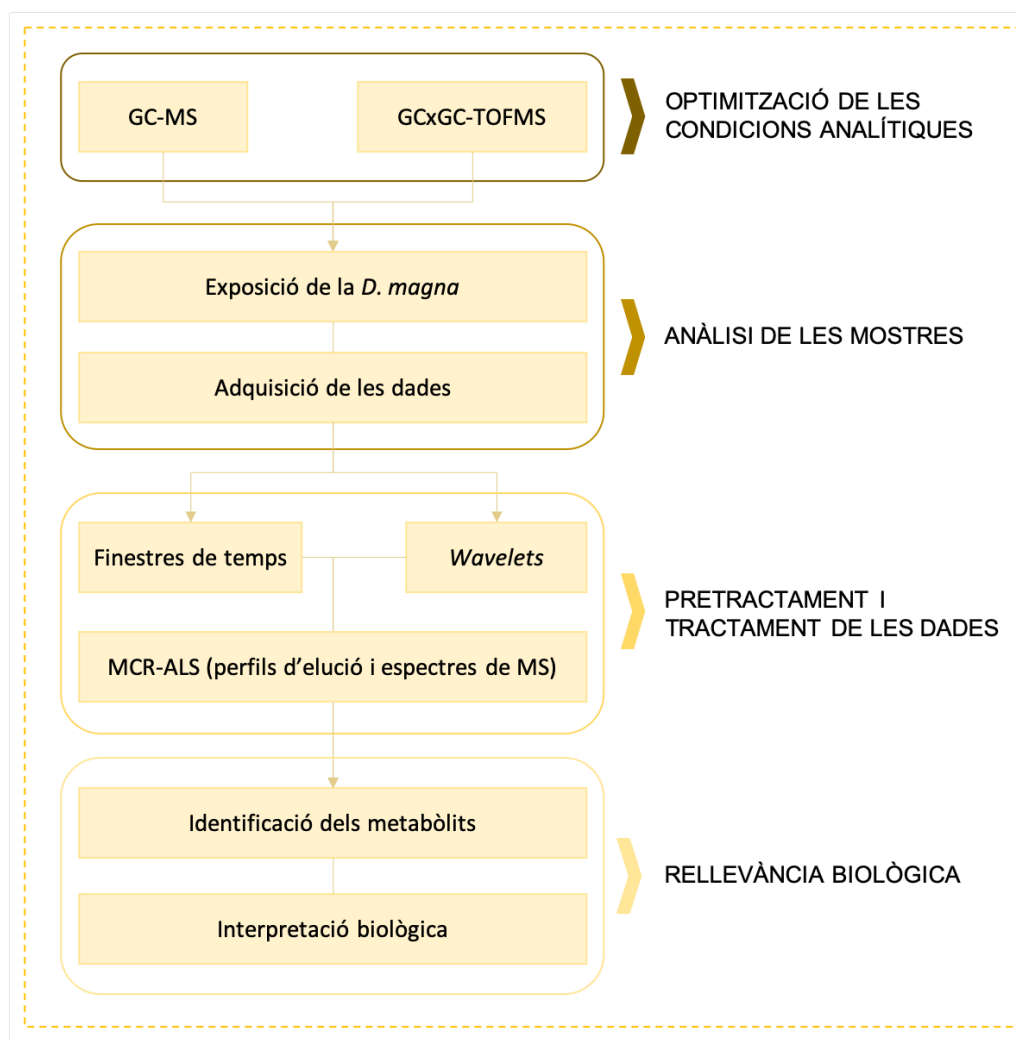


Figura 2.2 Flux de treball proposat pels estudis de metabolòmica no dirigida.

2.6.1 OPTIMITZACIÓ DE LES CONDICIONS ANALÍTIQUES

En la primera etapa del flux de treball es va portar a terme el desenvolupament dels mètodes analítics, GC-MS i GCxGC-TOFMS, per tal d'obtenir els perfils metabòlics de la *D. magna*. Les quatre publicacions presentades en aquest capítol es poden dividir en dos blocs segons la plataforma analítica utilitzada. Així, és en l'**article científic I** on s'optimitzen les diferents condicions experimentals, concretament el procediment d'extracció i de derivatització per a l'anàlisi dels metabòlits mitjançant la GC-MS, i per la GCxGC-TOFMS és en l'**article científic III** on s'avaluen diferents configuracions de columnes cromatogràfiques per a la determinació del perfil metabòlic de la *D. magna*.

Procediment d'extracció dels metabòlits

L'extracció dels metabòlits és un pas clau per obtenir dades útils i fiables per als estudis de metabolòmica. En la literatura s'han avaluat diferents protocols d'extracció per tal d'optimitzar aquest pas tant per a metabòlits polars com no polars, com per a diverses plataformes i sistemes biològics (Carneiro et al. 2014; Elie et al. 2015). La gran majoria de procediments es basen en un pas d'homogeneïtzació de la mostra seguit d'unes particions de cloroform, metanol i aigua. Les tècniques més habituals d'extracció són mitjançant l'ús de boles d'acer inoxidable (Liebeke and Bundy 2012), de microones (Khoomrung et al. 2015) o d'una sonda ultrasònica o d'un bany d'ultrasons (Vandenbrouck, et al. 2010).

Per tal de determinar les millors condicions per analitzar els metabòlits de la *D. magna* es van avaluar tres procediments d'extracció diferents: les boles d'acer inoxidable, la sonda ultrasònica i el bany d'ultrasons. En aquestes proves es van analitzar 3 replicats de mostres de *D. magna* per a cada una de les condicions amb les mateixes proporcions de cloroform, metanol i aigua (2:1:2). Els resultats no van determinar canvis significatius entre els tres procediments d'extracció avaluats. Tot i això, encara que només va consistir en una prova, l'avaluació de les condicions del procés d'extracció va permetre determinar unes condicions concretes i viables per a la determinació dels metabòlits presents en les mostres de *D. magna*. Així, en els **articles científics I, II i III** les condicions d'extracció dels metabòlits corresponen a unes proporcions de

cloroform, metanol i aigua (2:1:2), on s'utilitza un bany d'ultrasons per la homogeneïtzació de les mostres de *D. magna*. En canvi, en l'**article científic IV** es van poder emprar les boles d'acer inoxidable enlloc del bany d'ultrasons, ja que es tracta d'un procediment eficaç i ràpid.

El procés de derivatització dels metabòlits

Per als estudis de metabolòmica basats en GC el primer pas és ser capaços de detectar els metabòlits presents en les mostres biològiques. Per això, cal escollir l'agent derivatitzant adequat per a la formació dels derivats tèrmicament estables dels metabòlits. Entre els agents derivatitzant comunament utilitzats en la metabolòmica per GC destaca la sililació. Els agents sililants modifiquen una gran diversitat de grups funcionals, com hidroxils, àcids carboxílics, amines, amides, imines, tiols i fosfats. Els derivats sililats esdevenen menys polars, més volàtils i més estables tèrmicament.

Tal com es mostra en la **Figura 2.3**, la derivatització consta de dos passos, primer es duu a terme una MeOX en piridina on es formen derivats estables dels carbonils, molt útil pels metabòlits amb una cetona en la seva estructura, seguit de la introducció de l'agent sililant amb l'*N*-metil-*N*-(trimetilsilil) trifluoroacetamida (MSTFA) amb un 1% de trimetilclorosilà (TMCS) com a catalitzador, per tal de convertir els metabòlits en els seus derivats sililats ((metoxima)-trimetilsilil (-Si(CH₃)₃), (MeOX)-TMS) (**Figura 2.3**). Aquest procés de derivatització genera molècules tèrmicament estables adequades per la detecció mitjançant l'anàlisi per GC-MS. Cal tenir en compte que després de la derivatització, un únic metabòlit pot estar format per més d'un pic cromatogràfic degut a la generació d'artefactes (Xu et al. 2010).

Tot i que la GC-MS ha estat àmpliament utilitzada en metabolòmica de plantes (Hanifah et al. 2018; Schueuermann et al. 2019), el nombre de treballs on s'ha utilitzat per analitzar metabòlits d'organismes com la *D. magna* és escàs. És per això que es van agafar de referència diferents condicions descrites en la bibliografia, tenint en compte els factors més limitats del procés de derivatització (Gullberg et al. 2004; Kanani, et al. 2008; Kanani and Klapa 2007). Així, en la primera publicació d'aquesta Tesi es va avaluar aquest procés tenint en compte set factors experimentals per a l'anàlisi de metabòlits mitjançant GC-MS, els quals es poden dividir en dos grups segons si es troben en el

primer pas del procés de la derivatització o en el segon (**Figura 2.3**). Amb l'objectiu d'avaluar les diferents condicions es van analitzar una barreja de 13 metabòlits de diverses famílies (aminoàcids, sucres, àcids orgànics entre altres, veure secció 2.1 de l'**article científic I**) en 30 condicions experimentals on es van combinar set factors diferents en tres nivells (veure **Taula 2.1**): la concentració de MeOX (20, 30 i 40 mg mL⁻¹), el volum afegit de MeOX (30, 65 i 100 µL), la temperatura del procés de MeOX (20, 30 i 40 °C) i el temps de duració (90, 525 i 960 min), el volum afegit de MSTFA (30, 65 i 100 µL), la temperatura (20, 30 i 40 °C) i el temps de duració de la trimetilsililació amb MSTFA (TMS) (90, 525 i 960 min). Aquest estudi va permetre avaluar les millors condicions del procés de derivatització dels metabòlits presents en la *D. magna*, tal com es descriu en l'**article científic I** d'aquesta Tesi.

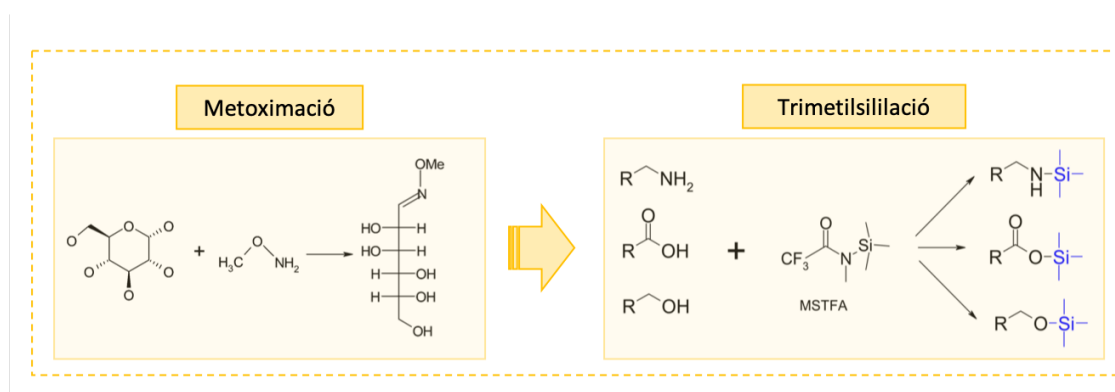


Figura 2.3| El procés de derivatització es porta a terme mitjançant dos agents derivatitzants. El primer pas és la MeOX en piridina, on es formen derivats estables dels carbonils. Llavors, es porta a terme una trimetilsililació amb MSTFA per tal de substituir els hidrògens actius dels compostos polars per un grup trimetilsilil [(MeOX)-TMS].

En general, els efectes dels diferents nivells i condicions aplicades en les mostres de metabòlits no van determinar un factor com el més rellevant. Tot i això, els diferents experiments van posar de manifest la importància de la realització d'una completa metoximació, sobretot per a la bona detecció dels sucres. Tanmateix, l'excés de l'agent derivatitzant permet que tots els metabòlits presents en la mostra es derivatitzin de manera completa. D'aquesta manera les condicions de derivatització escollides per a la determinació dels metabòlits de la *D. magna* van ser un volum de 65 µL de MeOX en piridina a una concentració de 20 mg mL⁻¹, seguit d'una incubació de 90 min a una temperatura de 30 °C. Després es procedeix a l'addició de 30 µL de MSTFA,

a l'1% de TMCS, amb una incubació de 30 min a 20 °C. Encara que no es va portar a terme un disseny experimental (DoE) factorial complet, aquesta avaluació de les condicions del procés de derivatització va permetre determinar les condicions més òptimes i viables per portar a terme una derivatització completa dels metabòlits presents en les mostres de *D. magna* per analitzar-les mitjançant la GC.

Taula 2.1 | Els diferents factors i nivells estudiats per avaluar el procés de derivatització dels metabòlits presents en la *D. magna*.

FACTORS	NIVELLS		
	Baix	Mig	Alt
Concentració MeOX (mg mL ⁻¹)	20	30	40
Volum MeOX (µL)	30	65	100
Temperatura MeOX (°C)	20	30	40
Temps MeOX (min)	90	525	960
Volum MSTFA (µL)	30	65	100
Temperatura TMS (°C)	20	30	40
Temps TMS (min)	30	60	90

Elecció de les columnes en GCxGC

Després de desenvolupar el mètode de detecció dels metabòlits presents en les mostres de *D. magna* mitjançant GC-MS, es va procedir a desenvolupar el mètode per GCxGC-TOFMS. Les condicions de treball instrumentals, com la rampa de temperatura i la velocitat del modulador, el diàmetre intern i el gruix de la pel·lícula de les columnes analítiques influeixen en els perfil i la capacitat dels pics dels metabòlits, així com en la correcta separació dels compostos i seva distribució al llarg del cromatograma (Koek, et al. 2008; Koek, et al. 2011b). És per això, que a l'hora de desenvolupar el mètode analític a part de tenir en compte la selecció de les columnes cromatogràfiques, el procés de modulació en la 2D del GCxGC-TOFMS també hi té un paper important. Així,

en l'**article científic III** d'aquesta Tesi es van avaluar la influència d'aquests paràmetres en la determinació dels perfils metabòlics de la *D. magna*.

En primer lloc, es va avaluar el comportament de tres configuracions de columnes diferents per l'anàlisi dels perfils metabòlics de la *D. magna* (**Taula 2.2**). Per a la 1D es van utilitzar dues columnes no polars diferents: TRB-5MS (*Teknokroma*) i ZB-5MS (*Phenomenex*), ambdues amb un 5 % difenil i 95 % dimetilpolisiloxà però de diferents llargades (m), diàmetres interns (mm) i gruix de pel·lícula (μm) (veure **Taula 2.2**). En canvi, en la 2D les columnes utilitzades eren de polaritat intermèdia amb un 50% difenil i 50% dimetilpolisiloxà, en concret es tractaven de les columnes TRB-50HT (*Teknokroma*) i TRB-50 (*Teknokroma*), també de característiques diferents, tal com es descriu en la **Taula 2.2**. A més, la columna TRB-50HT està indicada per arribar a temperatures elevades de fins a 370 °C, una especificació molt útil en la 2D. Amb la selecció d'aquestes columnes, en cada una de les configuracions estudiades es va poder avaluar la influència del diàmetre, ja que en cada set les columnes utilitzades tenen combinacions de diàmetres diferents. Així, en el set A la columna de la 1D té un diàmetre intern de 0.18 mm i la de la 2D un diàmetre intern de 0.10 mm. En canvi, en el set B tenen un diàmetre intern de 0.25 mm i de 0.10 mm en la 1D i en la 2D, respectivament, i un diàmetre intern de 0.25 mm per ambdues columnes en el set C (veure **Taula 2.2**).

Taula 2.2 | Columnes utilitzades en la primera (1D) i la segona dimensió (2D) per cada una de les configuracions o "sets" diferents.

	1D (5 % difenil, 95 % dimetilpolisiloxà)	2D (50 % difenil, 50 % dimetilpolisiloxà)
SET A	TRB-5MS (20 m x 0.18 mm x 0.18 μm)	TRB-50HT (2 m x 0.10 mm x 0.10 μm)
SET B	ZB-5MS (30 m x 0.25 mm x 0.25 μm)	TRB-50HT (2 m x 0.10 mm x 0.10 μm)
SET C	ZB-5MS (30 m x 0.25 mm x 0.25 μm)	TRB-50 (2 m x 0.25 mm x 0.25 μm)

Per altra banda, el procés de modulació també és un paràmetre que influeix en la separació dels metabòlits al llarg del cromatograma. El temps o període de modulació descriu el temps d'anàlisi utilitzat per a la completa separació i detecció de cada una de les fraccions que són enviades a la 2D i, al mateix temps, en la freqüència amb la que el modulador mostreja el que s'elueix en la 1D. És a dir, en augmentar la freqüència de mostreig, el temps d'anàlisi en la 2D disminueix. Per tant, ja que la resolució en la 1D està relacionada en funció del nombre de modulacions per ample del pic cromatogràfic, el temps de modulació ha de ser prou curt per tal de mantenir la resolució obtinguda, sense que al mateix temps produeixi una excessiva disminució en el temps d'anàlisi en la 2D (Dorman et al. 2008). La temperatura de modulació o temperatura *offset* correspon a la diferència de temperatura entre el modulador i cada un dels forns cromatogràfics, la seva funció és facilitar el procés de desorció dels compostos quan són enviats per al seu anàlisi.

Tenint en compte aquests paràmetres amb les tres configuracions de columnes, es van realitzar sis configuracions instrumentals diferents, és a dir, en cada un dels sets es van provar rampes de temperatura diferents, tant en la 1D com en la 2D, i diferents paràmetres del modulador per a la 2D, com el període de modulació i la temperatura del modulador, tal com es mostra en la Taula 1 de l'**article científic III**, on es descriuen els diferents paràmetres del mètode GCxGC-MS i les diferents configuracions de columnes utilitzades (veure també **secció 2.3** de l'**article científic III**). D'aquesta manera, es van analitzar diferents extractes de mostres de *D. magna* mitjançant GCxGC-TOFMS, on es van utilitzar les diferents configuracions de les columnes (set A, B i C) descrites anteriorment. En la Figura 4 de l'**article científic III** es mostren els cromatogrames contorn TIC (*Total Ion Chromatogram*). Les configuracions instrumentals realitzades amb el set A (S1, S2 i S3) i el set B (S4) mostren una pitjor resolució dels pics, així com unes formes de pic dels metabòlits poc definides en comparació a la mostra realitzada amb la configuració de columnes C (S5 i S6). Aquesta configuració on s'utilitza dues columnes amb el mateix diàmetre intern (0.25 mm) proporciona els millors resultats, ja que millora la distribució dels metabòlits al llarg del cromatograma, i també s'observa una millor resolució dels pics.

En la **Figura 2.4** es mostren els efectes de les diferents configuracions (set A, B i C) en la distribució de dos metabòlits: la treonina ($m/z = 219$) i l'àcid màlic ($m/z = 233$). En les seccions del cromatograma 2D, s'observa que en els sets A i B (**Figura 2.4A** i **2.4B**) el fet de tenir un diàmetre més petit en la columna de la 2D genera un coll d'ampolla entre les dues columnes. Això, produeix un endarreriment dels pics, amb un temps de retenció superior tant per la treonina, com per l'àcid màlic en aquests dos sets respecte del que s'obté en el set C. Aquest efecte és més evident en el set B on la diferència entre els diàmetres és major. Per altra banda, això provoca que els metabòlits que elueixen al principi del cromatograma es perdin i no sigui possible la seva detecció. També cal remarcar, que en la utilització d'una pel·lícula de fase estacionària més gruixuda s'observa que els pics dels metabòlits són més simètrics, amb pics estrets i menys distorsió en la forma dels perfils, fet que proporciona una millor resolució dels pics en aquesta configuració (**Figura 2.4C**).

En aquest context, la selecció de la millor configuració també es va realitzar a través de la comparació de la capacitat de pic (veure **secció 2.6** de l'**article científic III**). Les capacitats de pic calculades per a les configuracions S1, S2, S3 i S4 van ser inferiors a les calculades en les configuracions S5 i S6. En els sets A i B la separació dels metabòlits que s'aconsegueix en la 1D es perd en la 2D a causa de la excessiva durada del temps de modulació. En la columna de la 2D, la seva longitud és curta i amb un temps de modulació massa llarg s'augmenta la probabilitat del solapament dels pics cromatogràfics. Per contra, en el set C, la separació que s'aconsegueix en la 1D es manté en la 2D a causa del temps de modulació curt (2.5 s). La configuració S4 mostra una capacitat de pic més elevada que la obtinguda en el set A, aquest fet es pot relacionar amb la espessor de la fase de les columnes utilitzades en la 1D i en la 2D en la configuració S4 que permet una separació dels metabòlits més efectiva. Si es comparen la configuració S5 i S6, la temperatura final de la primera configuració és de 310 °C, la qual és superior a la temperatura tolerada per la columna en la 2D, per aquest motiu, la configuració S6 és la seleccionada com la millor configuració. Amb aquesta configuració s'obté una elevada capacitat de pic i això permet una bona distribució dels metabòlits en l'espai cromatogràfic.

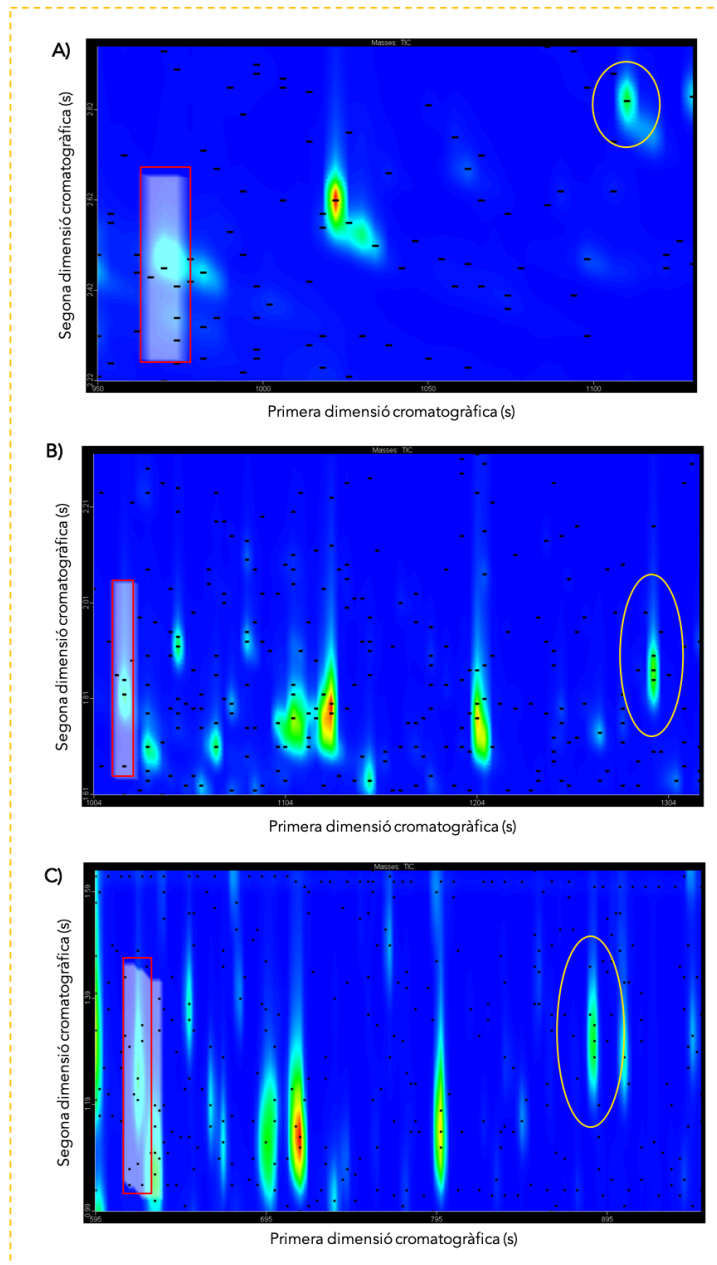


Figura 2.4| Seccions corresponents al cromatograma 2D del **A)** set A, **B)** set B i **C)** set C, on es mostra la distribució dels metabòlits: la treonina ($m/z = 219$) i l'àcid màlic ($m/z = 233$), els quals es troben assenyalats amb un rectangle vermell i amb un cercle groc, respectivament.

2.6.2 ANÀLISI DE LES MOSTRES

Una vegada s'han optimitzat les condicions del desenvolupament analític, el següent pas és realitzar l'exposició de les mostres. En els **articles científics I i II** els organismes són exposats a tres factors abiòtics: un increment de la salinitat, un augment de la temperatura i una manca d'oxigen o hipòxia. En el primer d'aquests treballs, els

efectes en el metaboloma de la *D. magna* s'estudien d'una manera independent per tal d'avaluar la viabilitat de l'anàlisi de les mostres exposades mitjançant la GC-MS i així observar els canvis en el metaboloma. En canvi, en el segon treball es determinen els efectes de les combinacions dels factors amb l'aplicació d'un DoE, a través d'un disseny *full factorial*, amb L^k combinacions de k factors a L nivells (Hibbert 2012). Tal com es pot veure en la **Taula 2.3**, les exposicions dels tres factors es van realitzar a dos nivells: nivell baix (0) i nivell alt (1) (per més detall veure **secció 2.2.2** de l'**article científic II**). D'aquesta manera, es van poder confirmar els efectes de manera individual, ja determinats en el primer treball publicat, i després avaluar els canvis produïts en el metaboloma de la *D. magna* a través de les interaccions entre els factors estudiats.

Taula 2.3 | Disseny *full factorial* de 3 factors a dos nivells 2^3 on es representen els 8 experiments realitzats a nivell baix (0) i nivell alt (1) (C: control, S: factor salinitat, T: factor temperatura i O: factor oxigen).

	Salinitat	Temperatura	Hipòxia
C	0	0	0
S	1	0	0
T	0	1	0
O	0	0	1
ST	1	1	0
OS	1	0	1
OT	0	1	1
OTS	1	1	1

Les condicions de l'anàlisi instrumental a través de la GC-MS per analitzar les mostres exposades estan descrites en la **secció 2.3** dels **articles científics I i II**. L'objectiu era tenir una bona resolució cromatogràfica i identificar tots els metabòlits amb fidelitat. En l'**article científic III** únicament es van utilitzar mostres control de *D. magna*, ja que l'objectiu de l'article era una optimització de les condicions analítiques de la tècnica GCxGC-TOFMS per a la determinació del metaboloma de la *D. magna*.

Les condicions de l'anàlisi mitjançant GCxGC-TOFMS estan descrites en la **secció 2.3** de l'**article científic III**. Per altra banda, en l'**article científic IV**, amb l'objectiu de comparar les tècniques GC-MS i GCxGC-TOFMS, els organismes van ser exposats únicament a un increment de la salinitat, ja que és el factor que produïa els efectes més significatius en el metaboloma. Tal com s'observa en els cromatogrames TIC que es mostren en la Figura 2 en l'**article científic IV**, els perfils dels extractes de les mostres de *D. magna* són diferents en cada una de les tècniques utilitzades.

La informació que proporciona la GCxGC-TOFMS és més àmplia que la que s'obté mitjançant la GC-MS. En l'anàlisi per GC-MS molts dels metabòlits coelueixen per falta de resolució en la tècnica analítica (Figura 2A i 2B). La utilització d'una columna cromatogràfica no permet separar correctament el gran nombre de metabòlits presents en les mostres de *D. magna*. Per contra, en analitzar els extractes mitjançant la GCxGC-TOFMS la configuració ortogonal de les dues columnes cromatogràfiques genera una molt bona resolució dels pics cromatogràfics (Figura 2C i 2D). D'aquesta manera, s'obté una millor separació amb una gran capacitat de pic que permet un increment en la detecció i la identificació dels metabòlits presents en les mostres (Koek et al. 2011a). Aproximadament es detecten un total de 50 i 100 pics mitjançant la GC-MS i la GCxGC-TOFMS, respectivament. Això significa que l'anàlisi GC-MS és molt adequat per determinar una gran varietat de metabòlits amb el pas previ de la derivatització. Tot i això, la tècnica GCxGC-TOFMS té un rendiment superior, ja que la separació en la 2D permet la determinació de compostos d'un ventall més ampli de polaritats, els quals a causa de la seva volatilitat no és possible detectar-los únicament en 1D. En conseqüència, el nombre de metabòlits detectats i identificats a través de l'anàlisi en 2D és significativament superior en comparar ambdues tècniques analítiques.

2.6.3 PRETRACTAMENT I TRACTAMENT DE LES DADES

El gran volum de dades generades mitjançant les tècniques *full scan* de metabolòmica no dirigida provoca la necessitat de reduir la mida d'aquesta informació per tal de facilitar el seu processament. D'aquesta manera, l'aplicació de les estratègies per a la compressió de les dades és de gran ajuda abans de l'anàlisi mitjançant MCR-

ALS. En els articles presentats en aquest capítol s'ha demostrat la utilitat tant de l'ús d'aquesta compressió de les dades unidimensionals i bidimensionals de GC en la direcció dels temps de retenció. El mètode MCR-ALS també s'aplica en les dades de GC, el qual permet l'obtenció dels perfils de concentració i dels espectres de masses dels metabòlits, i permet llur identificació i la caracterització dels canvis produïts a causa de les exposicions als factors ambientals.

Tal com s'ha explicat en la **secció 1.5.1** del **capítol 1** d'aquesta Tesi, existeixen varies estratègies per comprimir la mida de les dades obtingudes. En els **articles científics I i II** la compressió de les dades GC-MS es realitza mitjançant les finestres de temps (*time windowing*) (veure Figura 1.9 del **capítol 1** d'aquesta Tesi). Aquesta segmentació de les dades obtingudes va permetre analitzar els cromatogrames GC-MS de totes les mostres simultàniament mitjançant MCR-ALS.

En la **Figura 2.5** es representa l'organització de les dades GCxGC-MS, en primer lloc per cada modulació s'obté una separació cromatogràfica individual a la segona columna, i per tant, s'obté una matriu de dades per cada una de les modulacions ($X_{k,m}$) (**Figura 2.5A**). Les files d'aquesta matriu d'una única modulació contenen els espectres de masses a cada temps de retenció de la segona dimensió cromatogràfica, mentre que les columnes contenen els cromatogrames de la segona dimensió per cada valor de m/z . Quan aquestes matrius, $X_{k,m}$, es col·loquen una sota de l'altra, mantenint en comú els seus valors de m/z , les dades GCxGC-MS per una mostra s'organitzen en una estructura en forma de matriu augmentada en la direcció de les columnes (X_{aug}), tal com s'observa en la **Figura 2.5B**. Les files d'aquesta matriu augmentada contenen els espectres de MS per cada combinació de les dues dimensions cromatogràfiques i les seves columnes contenen els cromatogrames bidimensionals desplegats un darrera de l'altre, agrupats successivament segons les diferents modulacions a la segona dimensió. En els **articles científics III i IV** s'expliquen les estructures de les dades GCxGC-MS quan s'analitza simultàniament un conjunt de mostres. El conjunt de dades es pot representar per una matriu superaumentada en la direcció de les columnes, tal com es mostra en la **Figura 2.5C**. Aquesta matriu s'organitza col·locant cada una de les matrius augmentades de les diferents mostres una sota de l'altra, mantenint en comú els

espectres MS en les files d'aquesta matriu superaugmentada i amb els seus valor de m/z en comú a les columnes de la matriu.

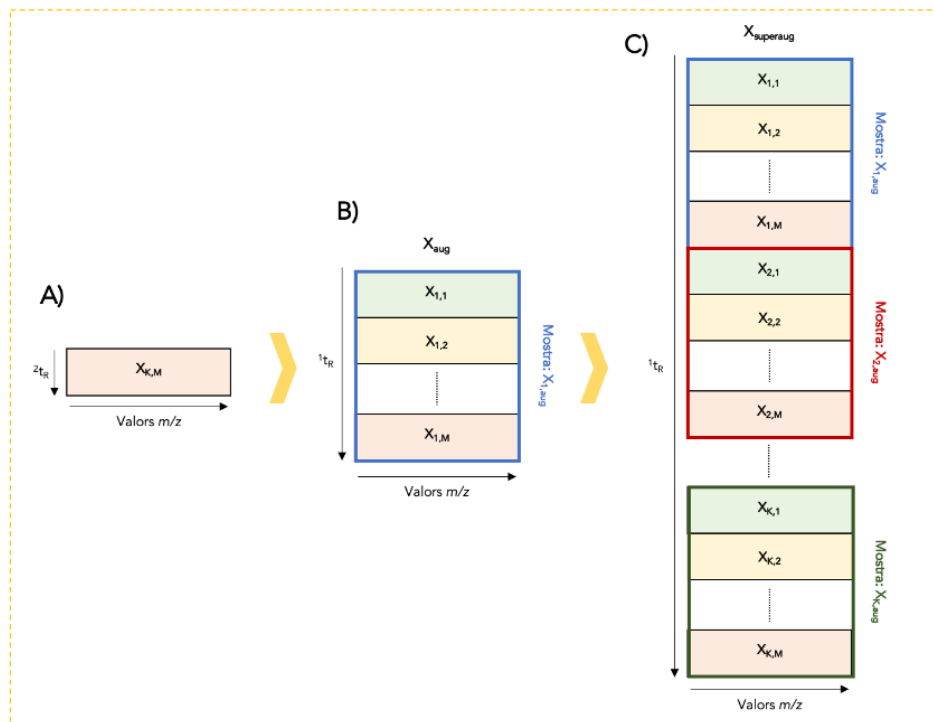


Figura 2.5] Descripció gràfica de l'organització de les dades GCxGC-MS. **A)** Les matrius $X_{K,M}$ són les matrius modulació que contenen la separació en la segona columna cromatogràfica per cada modulació provinent de la primera columna cromatogràfica, on K són el nombre de mostres en matrius augmentades de M modulacions. **B)** Una única mostra ordenada en forma de matriu augmentada en la direcció de les columnes (X_{aug}). **C)** Matriu superaugmentada de les dades GCxGC-MS per un conjunt de K mostres ($X_{superaug}$).

Un cop avaluada l'estructura de les dades GCxGC-MS, és necessària una estratègia de compressió de les dades que permeti analitzar la gran mida dels conjunts de dades GCxGC-TOFMS. És per això que es realitza la reducció de les dades a través de les transformades d'ones (*wavelet transform*), tal com es descriu en l'**article científic III**. En aquest article es va aplicar un nivell 4 de compressió per transformades d'ones (*level-4 wavelet compression*) (veure secció 3.2 de l'**article científic III**). També s'aplica aquest preprocessament en les dades GC-MS de l'**article científic IV**. En aquest cas, amb un nivell 1 de compressió és suficient per reduir la mida de les dades de les 12 mostres de *D. magna* analitzades (6 mostres control i 6 mostres exposades a salinitat). Per contra, en les dades obtingudes mitjançant la GC en 2D de l'**article científic IV**, a causa de les grans dimensions de les dades, és necessari aplicar les dues estratègies de

compressió en la direcció dels temps de retenció, una segmentació mitjançant finestres de temps i una posterior reducció de les dades crues. D'aquesta manera, és possible processar el conjunt de dades de les 12 mostres simultàniament. Així, es va aplicar una segmentació dels cromatogrames obtinguts a través de GCxGC-TOFMS en 4 finestres de temps diferents (*time windowing*). Després, es van comprimir les dades a través de *wavelets*, on es va aplicar un nivell 3 de compressió per tal de reduir-ne les dimensions i alhora reduir-ne els requisits d'emmagatzematge informàtic.

Una vegada s'ha realitzat la compressió de les dades, tant per les dades unidimensionals com bidimensionals, es porta a terme la resolució dels metabòlits presents en les mostres analitzades mitjançant el mètode MCR-ALS, tal com es detalla en la **secció 1.5.2** del **capítol 1**. En els articles publicats s'ha demostrat la gran eficàcia i la robustesa d'aquesta eina per la resolució dels perfils d'elució i dels espectres de masses dels metabòlits presents en les mostres de metabolòmica no dirigida de *D. magna* analitzades mitjançant GC-MS i GCxGC-TOFMS (**articles científics I i II, i III i IV**, respectivament). L'aplicació del mètode MCR-ALS permet determinar els perfils d'elució dels metabòlits presents en les mostres analitzades i associar-los a un espectre de masses concret, el qual s'utilitza per a la identificació del metabòlit en qüestió. A més, aquesta estratègia permet realitzar d'una manera ràpida i fàcil una discriminació entre els senyals cromatogràfics propis dels metabòlits i dels senyals dels agents derivatitzants i les contribucions del soroll instrumental i dels solvents que interfereixen en les mostres biològiques analitzades.

D'aquesta manera, tot i que el percentatge de la variància explicada (R^2) dels diferents models de MCR-ALS sigui superior al 90 %, el nombre de components estimats pel model és superior al nombre de pics resolts amb formes de pic cromatogràfic raonables. Aquest nombre addicional de components resolts per MCR-ALS descriuen altres tipus de variació en les dades experimentals, com per exemple el soroll instrumental de fons i els senyals dels agents derivatitzants. Així, només es tenen en compte aquells components resolts, els quals s'han obtingut mitjançant l'aplicació del mètode MCR-ALS, que mostren formes de pic cromatogràfic raonables (pics

unimodals estrets) per a realitzar la posterior identificació i relacionar-los amb el metaboloma de la *D. magna*.

2.6.4 RELLEVÀNCIA BIOLÒGICA

La identificació dels metabòlits a través dels perfils d'elució i dels espectres de masses obtinguts per l'aplicació del mètode MCR-ALS es realitza mitjançant els índex de retenció (RI) i un factor de coincidència dels espectres de masses teòrics i experimentals. Es poden obtenir dos valors diferents pes factors de coincidència dels espectres de masses teòrics i experimentals mitjançant el programa de la base de dades NIST 2014 (*NIST Mass Spectral Search 2.2 program*): un factor de coincidència per a l'espectre desconegut i el de la biblioteca anomenat factor de coincidència (*match factor*, MF), i un factor de coincidència inversa (*reverse match factor*, RMF) el qual correspon al factor de coincidència per a l'espectre de la biblioteca i l'espectre desconegut, sense tenir en compte els pics de masses de l'espectre desconegut que no coincideixen amb els pics de masses de l'espectre de la biblioteca. Aquests factors coincidents es troben entre 0 (sense coincidència) i 1000 (coincidència perfecta). Com a guia general, es va considerar un valor de 900 o superior com a molt bona coincidència; entre 800 i 900 una bona coincidència; entre 700 i 800 una coincidència justa; i menys de 600 una coincidència pobre o dolenta. En el càlcul del MF, l'espectre experimental s'usa com a plantilla, mentre que pel RMF la plantilla és l'espectre de la biblioteca.

En la base de dades NIST 2014 també és possible incloure-hi la base de dades GMD, amb els corresponents metabòlits derivatitzats. A més, per tal d'augmentar la fiabilitat de la identificació, s'inclouen els índex de retenció (*retention indices*, RI) en l'avaluació de la identificació dels metabòlits. Així, s'utilitza una barreja de patrons estàndards *n*-alcans (hidrocarburs des de C₇ fins a C₃₀) per calcular-los (Strehmel et al. 2008). El clar avantatge d'emprar aquesta estratègia de MCR-ALS és que permet una millor comparació dels espectres de MS dels pics resolts amb els espectres MS teòrics que proporciona la biblioteca NIST 2014 i la base de dades GMD a causa de l'ús simultani del temps de retenció dels pics d'elució i de la presència dels fragments GC-

MS en l'espectre de masses resolts assignats a aquests pics. Tal com es mostra en la **Figura 2.6**, aquest procediment permet la determinació dels metabòlits derivatitzats a partir de la comparació entre l'espectre MS experimental (en la part superior) i el teòric (en la part inferior). Per exemple, la identificació del metabòlit derivatitzat de l'àcid palmític (TMS) es realitza amb la determinació d'un valor de RMF d'aproximadament 800 entre els espectres MS i un RI experimental de 2070 respecte del valor RI teòric de 2050, els quals confirmen la identitat del metabòlit.

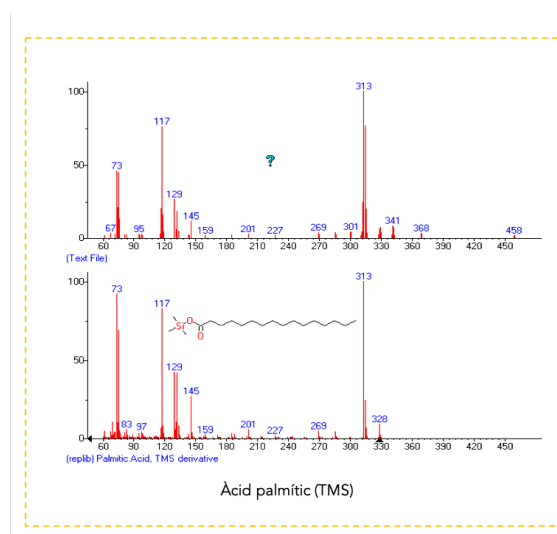


Figura 2.6| Identificació dels metabòlit derivatitzat de l'àcid palmític (TMS) mitjançant l'espectre MS obtingut mitjançant el procediment MCR-ALS i l'espectre teòric de les bases de dades NIST 2014 i GMD.

Per altra banda, l'ús de les eines quimiomètriques permet determinar els metabòlits rellevants per a la diferenciació entre les mostres control i les diferents mostres exposades. En l'**article científic I** l'aplicació del mètode PLS-DA als perfils d'elució autoescalats dels diferents components resolts mitjançant MCR-ALS permet la discriminació, a través del mètode de selecció de les variables importants en la projecció (VIPs) (veure selecció de variables en la **secció 1.5.2** del **capítol 1** d'aquesta Tesi), entre les mostres control i les mostres exposades de cada factor individual. D'aquesta manera, és possible determinar mitjançant l'estimació de les àrees/concentracions relatives dels metabòlits resolts, quins són els que canvien significativament entre les diferents exposicions.

En l'**article científic I** s'identifiquen com a potencials biomarcadors 40 metabòlits pel factor salinitat, 21 per la temperatura i 13 per la hipòxia. Una vegada s'ha portat a terme la identificació d'aquests metabòlits, és necessària l'exploració bioquímica corresponent per tal d'avaluar les rutes metabòliques afectades utilitzant la informació continguda en la base de dades KEGG. En aquest article es mostra que els tres tractaments comparteixen efectes en varis metabòlits relacionats amb el metabolisme del piruvat i del cicle de Krebs, els quals estan connectats amb les rutes metabòliques energètiques de la *D. magna*. El factor de la salinitat té un gran impacte en els metabòlits del glicerol i la trehalosa, els quals estan relacionats amb la regulació de la osmosi en les cèl·lules. Com que només es tracta l'exposició als factors d'una manera individual, únicament és possible avaluar si els metabòlits afectats coincideixen en les tres exposicions, tal com es mostra en el diagrama Venn de la Figura 4 (**article científic I**). D'aquesta manera, un total de 11 metabòlits es van veure afectats per la salinitat i la temperatura, i 2 metabòlits en l'exposició a la temperatura i la hipòxia i, a la salinitat i la hipòxia.

Seguint aquesta línia de recerca, en l'avaluació de la combinació dels efectes que es realitza en l'**article científic II** s'apliquen els mètodes PCA, ASCA i OPLS-DA. En l'aplicació del PCA als perfils d'elució cromatogràfica autoescalats resolts amb MCR-ALS es dedueix el gran impacte que té en el metaboloma de la *D. magna* l'increment de la salinitat en el medi. A mode exploratori, en la Figura 1 de l'**article científic II** s'observa la separació de les mostres exposades a salinitat (S, ST, OS i OTS) i de les mostres no exposades a aquest factor (C, O, T, OT). Es confirma aquesta tendència a través de l'aplicació del model ASCA sobre tot el conjunt de metabòlits analitzats (anàlisi multivariant), on els resultats mostren un efecte especialment significatiu de la salinitat en el metaboloma de la *D. magna*. En canvi, en les exposicions a la temperatures i la hipòxia no s'observa un efecte tant intens en els nivells del metabòlits respecte a les mostres control. No obstant això, cal considerar que els intervals de variació emprats en aquests paràmetres no s'aparten significativament dels que es poden trobar en condicions naturals.

Per tal de determinar d'una manera més exhaustiva els efectes dels factors, el mètode OPLS-DA permet determinar quins metabòlits tenen una major influència en la discriminació entre els tractaments, a través de la selecció dels paràmetres VIPs i SR (veure selecció de variables en la **secció 1.5.2** del **capítol 1** d'aquesta Tesi). A més, l'estratègia estadística univariada *one-way* ANOVA aplicada a cadascun dels metabòlits per separat és un mètode que permet avaluar i confirmar les diferències entre les àrees de pic dels metabòlits dels diferents tractaments per cada metabòlit. Les dues estratègies utilitzades per seleccionar els biomarcadors mostren diferències segons els nivells de significació o els valors llindar (*threshold values*) seleccionats (Taula 3, **article científic II**). Com que l'objectiu és investigar els possibles canvis dels perfils metabòlics de la *D. magna*, es decideix no seleccionar les condicions estadístiques més restrictives. Així, mitjançant el mètode *one-way* ANOVA, 24, 5 i 12 metabòlits resulten afectats d'una manera significativa per la salinitat, la temperatura i la hipòxia, respectivament. Aquests resultats corroboren els resultats obtinguts amb l'ASCA, on s'observa la poca influència dels tractaments de la temperatura i la hipòxia, en els nivells d'exposició estudiats. Per contra, mostren la salinitat com el factor dominant en les alteracions produïdes en el metaboloma de la *D. magna*. Tant en l'**article científic I** com en el **II**, un percentatge elevat dels metabòlits alterats a causa de la salinitat estan relacionats amb el metabolisme dels aminoàcids, de les proteïnes, dels sucres, dels àcids grassos i amb el metabolisme energètic, com el cicle de Krebs.

En l'**article científic IV** les mostres de *D. magna* es comparen mitjançant la GC-MS i la GCxGC-TOFMS, on es confirma els beneficis de la tècnica bidimensional respecte de la unidimensional. D'aquesta manera, en 2D es proporciona un augment de la resolució i de la sensibilitat en la detecció dels metabòlits amb un increment del nombre de metabòlits detectats que passa de 20 a 42 metabòlits, amb les tècniques unidimensional i bidimensional, respectivament. Després de la identificació, s'aplica el mètode PLS-DA als 42 metabòlits per tal de discriminar les mostres control de les mostres exposades a salinitat. La selecció de variables a través dels VIPs i del SR mostra un nombre diferent de biomarcadors segons el grau de restricció de cada una, on es determinen 18 i 6 metabòlits significatius, respectivament. Entre els metabòlits que canvien significativament la seva concentració a causa de la salinitat es troben l'àcid

làctic i la glucosa, els quals estan relacionats amb la glicòlisi. Això podria relacionar-se amb un consum d'oxigen que afectaria la respiració de la *D. magna*, on es produeix un increment del metabolisme anaeròbic, el qual consumeix glucosa i genera àcid làctic. Aquests resultats confirmen els obtinguts en els **articles científics I i II**, on s'identifica que l'increment de la salinitat afecta el metabolisme energètic en la *D. magna*, amb una influència directa en els nivells de glucosa.

Els articles científics d'aquest capítol han permès concloure la gran influència que tenen els increments de la salinitat en el metaboloma de la *D. magna*. L'aplicació del disseny experimental per a l'estudi dels factors d'exposició ha permès avaluar les possibles interaccions entre els diferents tractaments i llurs combinacions. No obstant, en aquest cas els factors de la temperatura i la hipòxia han resultat que afecten poc al metaboloma, en el rang de variació estudiat d'aquests dos paràmetres ja que els seus possibles efectes queden emmascarats per les variacions naturals de l'organisme i per l'efecte predominant que la salinitat produeix en el metabolisme de la *D. magna*.

CAPÍTOL 3



ESTUDIS DE METABOLÒMICA EN *D.*
MAGNA EXPOSADA A CONTAMINANTS
AMBIENTALS MITJANÇANT LC-MS

3.1 INTRODUCCIÓ

Durant el segle XX i principis del XXI, els ecosistemes han patit canvis apreciables com a conseqüència directa de l'activitat humana. A més, milers de nous productes químics són alliberats i acumulats en l'entorn els quals produeixen efectes desconeguts sobre els organismes vius i els sistemes ecològics. Aquestes alteracions químiques i físiques de l'entorn i dels ecosistemes aquàtics produeixen variacions en els sistemes metabòlics dels organismes vius. En aquest context, l'estudi i l'anàlisi d'aquests canvis proporcionen informació significativa sobre les respostes biològiques a aquestes alteracions ambientals.

Les tecnologies metabolòmiques i els enfocaments basats en MS proporcionen plataformes complementàries per l'avaluació dels efectes dels contaminants ambientals en els sistemes biològics. La utilització de la cromatografia de líquids d'alta eficàcia (*high performance liquid chromatography*, HPLC) per estudis de metabolòmica és relativament nova comparada amb altres tècniques com ara la GC o la RMN. En les últimes dècades, el gran poder de separació i l'elevada resolució cromatogràfica que permeten la determinació de metabòlits d'un rang ample de polaritats i de característiques estructurals diferents ha donat lloc a un creixement exponencial en l'ús d'aquesta tècnica en estudis de metabolòmica (Tournois et al. 1998; Zhou, et al. 2012). Entre els mètodes de LC més comuns es troba la cromatografia de fase invertida (*reversed-phase liquid chromatography*, RPLC) i la cromatografia d'interacció hidrofílica (*hydrophilic interaction liquid chromatography*, HILIC). En la RPLC generalment s'utilitzen columnes amb fases estacionàries C₁₈ o C₈ degut a la seva versatilitat a l'hora d'analitzar un ampli ventall de compostos, tant metabòlits semipolars com lípids (Zhou and Yin 2016). Tanmateix, en els estudis de metabolòmica, aquestes columnes presenten desavantatges a causa de la baixa retenció dels metabòlits més polars i hidròfils que no queden retinguts en fases de C₈ o C₁₈. És per això, que han sorgit altres mètodes de separació cromatogràfica que utilitzen tipus de columnes com les HILIC, les quals permeten millorar la determinació de compostos d'elevada polaritat (Navarro-Reig et al. 2017b; Ortiz-Villanueva et al. 2017c; Tang et al. 2016).

Les tècniques LC-MS han esdevingut molt exitoses quan s'utilitza l'estratègia no dirigida donada llur gran capacitat d'obtenir informació per caracteritzar adequadament els perfils metabòlics dels sistemes biològics. A més, les anàlisis no dirigides ofereixen la possibilitat de detectar nous biomarcadors que permeten elucidar els efectes dels contaminants ambientals. En aquest context, el creixement exponencial dels treballs de metabolòmica recau en els avenços en els mètodes cromatogràfics, tant en la instrumentació com en les fases estacionàries de les columnes analítiques emprades, i en l'augment de la sensibilitat i la precisió en l'espectrometria de masses (Gika et al. 2019). D'aquesta manera, s'han desenvolupat nombrosos mètodes i protocols que cobreixen l'anàlisi de diverses classes de metabòlits: compostos hidròfils com aminoàcids, àcids orgànics, sucres, etc., i compostos lipòfils com ara lípids, eicosanoids, oxilipines, àcids biliars, etc. Recentment, s'han publicat varies revisions on es descriuen aquests nous procediments que intenten determinar una àmplia gamma de metabòlits, a través del desenvolupament de noves fases estacionàries i l'aplicació de metodologies bidimensionals (Begou et al. 2017a; Zhou and Yin 2016).

Les tècniques basades en LC-MS permeten la caracterització extensa del metaboloma, a més han permès identificar metabòlits desconeguts. La informació que s'obté sobre els perfils metabòlics i les seves variacions han estat clau per avaluar els efectes dels contaminants ambientals sobre els organismes aquàtics. Per exemple, Ortiz-Villanueva et al. (2018) van utilitzar una estratègia metabolòmica no dirigida de LC-HRMS per tal d'avaluar els efectes de les exposicions de les dosis subletals de tres disruptors endocrins (TBT, BPA i PFOS) a diferents períodes de temps en els perfils metabòlics d'embrions de peixos zebra, on es van determinar canvis en el metabolisme dels glicerofosfolípids, aminoàcids, purines i àcids carboxílics. També, Elie, et al. (2015) van avaluar els efectes dels compostos PAHs i els seus derivats oxigenats (oxy-PAHs) en els perfils metabòlics de peixos zebra en creixement mitjançant la utilització de HPLC-Q-TOF-MS, on es van determinar un total de 63 metabòlits alterats després d'una exposició de 24 hores a una concentració ambiental de 4 μ M.

Per altra banda, entre una ampla gamma de contaminants ambientals la presència de productes farmacèutics en el medi ambient és un tema de gran rellevància social, és per això que cada vegada s'investiguen més els seus efectes en el metabolisme dels organismes aquàtics. Song et al. (2018) van determinar 45 potencials biomarcadors, incloent aminoàcids, nucleòtids i els seus derivats, amb una anàlisi no dirigida mitjançant UPLC-Q-TOF-MS, on es van avaluar els efectes estereoselectius de l'ibuprofè a concentracions ambientals en el metaboloma dels peixos zebra adults.

En aquesta Tesi es proposa l'avaluació dels efectes dels contaminants ambientals (fluoxetina, diazepam, carbamazepina i BPA) on es presenten les estratègies dirigida i no dirigida per a l'estudi del metaboloma de la *D. magna*.

En aquest capítol s'inclouen les següents publicacions:

- **Article científic V:** *Effects of psychiatric drugs on Daphnia magna oxylipin profiles*. E. Garreta-Lara, A. Checa, D. Fuchs, R. Tauler, S. Lacorte, C.E. Wheelock, C. Barata. *Science of the Total Environment* 644 (2018), 1101 – 1109.

En aquest article es va utilitzar una estratègia analítica dirigida per quantificar els mediadors lipídics dels àcids grassos poliinsaturats (PUFAs). Es va desenvolupar un procediment que permetia avaluar els canvis en el perfil de les oxilipines i dels eicosanoids derivats dels PUFAs com a conseqüència de l'exposició a tres fàrmacs psiquiàtrics, fluoxetina, diazepam i carbamazepina, a concentracions ambientals (100 ng L⁻¹). La identificació i la quantificació dels metabòlits es va realitzar amb els softwares comercials de Waters: *MassLynx* i *TargetLynx*.

- **Article científic VI:** *Novel combined approach for the target/untarget characterization of Daphnia magna metabolome*. E. Garreta-Lara, B. Campos, C. Barata, R. Tauler, S. Lacorte. En preparació.

En aquest article s'explora la combinació de dos tipus d'estratègies analítiques, la dirigida i la no dirigida, per determinar els canvis que es produeixen en el metaboloma de la *D. magna* quan és exposada a BPA. Ambdues estratègies són complementàries, ja que per una banda, l'aproximació dirigida permet una quantificació d'un grup de

metabòlits prèviament seleccionats a través del desenvolupament del mètode analític. Mentre que per altra banda, l'estratègia d'anàlisi no dirigida permet una quantificació relativa dels canvis globals en el perfil metabòlic de la *D. magna*. La combinació d'aquestes dues estratègies permet una caracterització del metaboloma de la *D. magna* més exhaustiva, tant qualitativa com quantitativa a partir de l'anàlisi mitjançant LC-HRMS. Aquest mètode d'anàlisi genera grans conjunt de dades, els quals requereixen la seva filtració i compressió, a partir de la utilització del concepte de les regions d'interès ROIs i del procediment ROIMCR (veure **secció 1.5.2** en la introducció d'aquesta Tesi, **capítol 1**).

3.2 ARTICLE CIENTÍFIC V

Effects of psychiatric drugs on Daphnia magna oxylipin profiles.

E. Garreta-Lara, A. Checa, D. Fuchs, R. Tauler, S. Lacorte, C.E. Wheelock, C. Barata.

Science of the Total Environment 644 (2018), 1101 – 1109.



Contents lists available at ScienceDirect

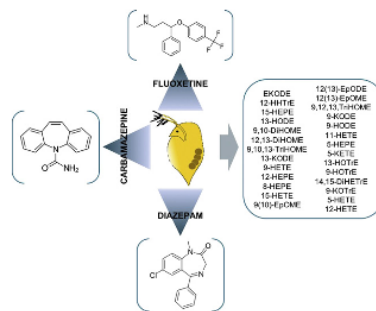
Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenvEffect of psychiatric drugs on *Daphnia magna* oxylipin profilesElba Garreta-Lara^a, Antonio Checa^b, David Fuchs^b, Romà Tauler^a, Silvia Lacorte^a, Craig E. Wheelock^b, Carlos Barata^{a,*}^a Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18–26, Barcelona 08034, Catalonia, Spain^b Division of Physiological Chemistry 2, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-17177 Stockholm, Sweden

HIGHLIGHTS

- Oxylipin profiles in *D. magna* adults upon exposure to low concentrations of psychiatric drugs were studied.
- Up to 28 different oxylipins belonging to arachidonic, linoleic and eicosapentaenoic acid metabolic pathways were altered.
- Fluoxetine increased the accumulation of the prostaglandin product 12-HHTrE.
- Diazepam increased the concentration of 12 eicosanoids and other oxylipins.
- Carbamazepine had little effects only affecting one oxylipin from the linoleic acid metabolism.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 9 May 2018
 Received in revised form 26 June 2018
 Accepted 27 June 2018
 Available online xxxx

Editor: D. Barcelo

Keywords:

Oxylipin profiles
 Eicosanoids
Daphnia magna
 Target analysis
 UPLC-MS/MS
 Psychiatric drugs

ABSTRACT

Neuro-active pharmaceuticals have been reported to act as endocrine disruptors enhancing reproduction in the model crustacean *Daphnia magna* at environmental concentrations of ng/L. Oxylipins and more specifically eicosanoids, which are lipid mediators formed from polyunsaturated fatty acids (PUFAs), are known to regulate reproduction together with other physiological processes in insects. In *D. magna*, the biosynthesis of eicosanoids and their putative role in the regulation of reproduction has been studied using transcriptomics, genomics and exposures to cyclooxygenase inhibitors. Quantification of eicosanoids and oxylipins derived from PUFAs upon exposure to pharmaceuticals is therefore crucial for a better understanding of the mode of action of neuro-active pharmaceuticals on aquatic invertebrates. The aim of this study was to investigate shifts in the oxylipin profile in *D. magna* adults upon exposure to environmental concentrations of the three psychiatric drugs, fluoxetine, diazepam and carbamazepine, with known effects of enhancing offspring production. Oxylipin profiles were determined in whole organism tissues using ultra performance liquid chromatography–tandem mass spectrometry (UPLC-MS/MS). Up to 28 different oxylipins belonging to arachidonic (AA), linoleic acid (LA), α -linoleic acid (α -LA) and eicosapentaenoic acid (EPA) pathways were detected and quantified in *D. magna* adults. Exposure to the selected psychiatric drugs showed that fluoxetine enhanced the accumulation of the cyclooxygenase (COX) product 12-

Abbreviations: AA, arachidonic acid; COX, cyclooxygenases; CYP, cytochrome 450; DiHETrE, dihydroxy-eicosatrienoic acids; DiHOMEs, dihydroxy-octadecenoic acids; EKODEs, epoxy-keto-octadecenoic acids; EPA, eicosapentaenoic acid; EpETrEs, epoxy-eicosatrienoic acids; EpODEs, epoxy-octadecadienoic acids; EpOMEs, epoxy-octadecenoic acids; ESI, electrospray ion source; HEPEs, hydroxyl-pentaenoic acids; HETEs, hydroxyl-eicosatetraenoic acids; HODEs, hydroxyl-octadecadienoic acids; HOTrEs, hydroxyl-octadecatrienoic acids; KODEs, oxo-octadecadienoic acids; KOTrEs, oxo-octadecatrienoic acids; LA, linoleic acid; LOX, lipoxygenases; LTB₄, leukotriene B₄; PGE₂, prostaglandin E₂; PGH₂, prostaglandin H₂; PGs, prostaglandins; PUFA, polyunsaturated fatty acid; QC, quality control; SPE, solid phase extraction; SSRI, selective serotonin re-uptake inhibitor; TriHOMEs, trihydroxy-octadecenoic acids; TXs, thromboxanes; α -LA, alpha-linoleic acid.

* Corresponding author at: Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18, 08034 Barcelona, Spain.

E-mail address: cbmqam@cid.csic.es (C. Barata).

<https://doi.org/10.1016/j.scitotenv.2018.06.333>
 0048-9697/© 2018 Elsevier B.V. All rights reserved.

hydroxyheptadecatrienoic acid (12-HHTrE), whereas diazepam increased the concentration of eicosanoids belonging to the lipoxygenase (LOX) and cytochrome P450 (CYP) pathways (HETEs, EpOMes, HODEs, HOTrEs and HEPes) from the AA, LA, α LA and EPA pathways. Carbamazepine had little effect and only one LA-derived compound from the LOX pathway (13-HODE) increased significantly. This means that despite having different modes of action in humans, fluoxetine and diazepam up-regulated eicosanoid pathways in *D. magna*, closely related to known biologically active products that regulate reproduction in insects.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

There is an urgent need for assessing the risks of long-term exposure to low doses of human prescribed pharmaceuticals (Fent et al., 2006). Waste water treated and untreated effluents are the main source of pharmaceuticals and/or their metabolites to water. Consequently, pharmaceuticals are continuously released into the environment and thus their negative effects are independent from their persistence in the environment (Fent et al., 2006; Petrović et al., 2003).

Antidepressants and anxiolytic drugs deserve special attention due to their usage and environmental persistence. Surveys in US have reported levels of 12–540 ng/L of fluoxetine, the active ingredient of Prozac, in surface waters and effluents (Kolpin et al., 2002) and total concentrations of serotonin reuptake inhibitors in aquatic systems were measured in the range of 840 ng/L to 3.2 μ g/L (Metcalf et al., 2010; Vasskog et al., 2008). Diazepam, is widely used to treat anxiety (Riss et al., 2008). Concentrations of diazepam ranging from 4 to 40 ng/L have been found in Spanish urban rivers (Valcárcel et al., 2012). Carbamazepine is a medication used primarily in the treatment of epilepsy and neuropathic pain (Ambrósio et al., 2002). Carbamazepine is fairly persistent in water and hence can be found at concentrations ranging from 1 to up to 3000 ng/L in rivers receiving waste water treatment effluents (Muñoz et al., 2009; Tixier et al., 2003).

Recent studies have reported that, at environmentally relevant concentrations, antidepressants and anxiolytic drugs alter the behavior of fish, mollusks and crustaceans (Ford and Fong, 2016). The targets of these compounds are highly conserved across vertebrates and they are also found in the invertebrate crustacean *D. magna* (Gunnarsson et al., 2008). Previous studies have reported that the selective serotonin re-uptake inhibitor fluoxetine increased brain serotonin activity in *D. magna*, and increased reproduction and positive phototactic behavior at as low concentrations as 100 ng/L (Garreta-Lara et al., 2016; Rivetti et al., 2016). There is a cross-talk between serotonin receptors and arachidonic/prostaglandin (AA/PGs) metabolism (Tournois et al., 1998), which regulate reproduction in insects (Spracklen and Tootle, 2015). Thus, it is possible that fluoxetine, by increasing serotonin activity up-regulates AA/PGs metabolism, and hence increases reproduction. In decapod crustaceans, exposure to serotonin or serotonergic drugs increased the levels of the crustacean hyperglycemic hormone, which regulates energy metabolism (Santos et al., 2001), anxiety-like and aggressive behavior (Fossat et al., 2014; Kravitz, 2000) and stimulated ovarian maturation (Sarojini et al., 1995). Furthermore, it has been recently reported that the psychiatric drugs carbamazepine and diazepam at low concentrations (100 ng/L) also increased reproduction and alter phototactic behavior in *D. magna* (Rivetti et al., 2016). The main target of carbamazepine is blocking voltage-dependent sodium channels (Ambrósio et al., 2002), but there is reported information indicating that carbamazepine also causes increases in extracellular serotonin levels (Dailey et al., 1997). Accordingly, carbamazepine may also act like fluoxetine increasing serotonin activity and hence PGs synthesis and reproduction. Diazepam acts at the GABA signaling pathway, which has been related to the expression of predatory induced life-history defenses in *D. magna* like growth, timing and reproductive

output (Weiss et al., 2012). In rats, diazepam alters dopamine (Biswas and Carlsson, 1978), which also modulates reproduction and behavior in crustacean (Tierney et al., 2003), and PGs (Labhsetwar and Zolovick, 1973).

Oxygenated lipids, collectively known as oxylipins, are endogenous signaling molecules that are highly diverse in terms of their structure, metabolism and biological function, including cell proliferation, inflammation, immune actions and blood pressure regulation (Buczynski et al., 2009). Oxylipins are derived from polyunsaturated fatty acids (PUFAs) such as arachidonic acid (AA), linoleic acid (LA), α -linoleic acid (α -LA) and eicosapentaenoic acid (EPA by COX (cyclooxygenases) and LOX enzymes (lipoxygenases) or by cytochrome P450 (CYP) forming a complex pool of bioactive components (Barquissau et al., 2017; Lundstrom et al., 2011; Zhao et al., 2014). Oxylipins derived from the different AA, LA, α -LA and EPA pathways are shown in Fig. 1. Prostanoids, including prostaglandins (PGs) and thromboxanes (TX), are generated from AA via the initial oxidation through COX pathways (Stanley, 2006). Leukotrienes are generated via the LOX pathways. LOX-dependent products include LOX hydroperoxidases families such as the AA-derived hydroxy-eicosatetraenoic acids (HETEs), as well as the LA-derived hydroxy-octadecadienoic acids (HODEs), α -LA-derived hydroxy-octadecatrienoic acids (HOTrEs) and EPA-derived hydroxy-pentaenoic acids (HEPEs) (Strassburg et al., 2012) (Fig. 1). CYP enzymes can convert AA and LA via epoxygenase activity into epoxy-eicosatrienoic acids (EpETrEs) and epoxy-octadecenoic acids (EpOMes), respectively. These epoxygenase products from AA and LA are rapidly metabolized via soluble epoxide hydrolase (sEH) to their corresponding diols, dihydroxy-eicosatrienoic acids (DiHETrEs) and dihydroxy-octadecenoic acids (DiHOMEs) (Schmelzer et al., 2005). In addition, CYP epoxygenase can also metabolize α -LA-derived resulting in epoxy-octadecadienoic acids (EpODEs) (Gabbs et al., 2015). In *D. magna*, eicosanoids act as important lipid mediators in reproduction, the immune system and ion transport (Heckmann et al., 2008). In addition to the well-known eicosanoids derived from AA, LA together with EPA are essential fatty acids for growth and reproduction in *D. magna* (Martin-Creuzburg et al., 2010).

To date, few studies have investigated PUFA-derived *D. magna* oxylipins levels. The quantification of these chemically and structurally related oxylipins derived from different PUFAs has become a challenge, but the development of LC-MS/MS-based methods is enabling this possibility. Analytical methods were developed in order to detect and quantify a large number of eicosanoids and other oxylipins in several biological matrices (Checa et al., 2015; González-Peña et al., 2017; Lundström et al., 2011; Strassburg et al., 2012). Additional biological information can be obtained by performing chiral-based analyses on the S and R isoforms of oxylipids (Balgoma et al., 2013). Equal amounts of isoforms S and R or an excess of the former means, respectively, that oxylipins are produced enzymatically or by autooxidation (Willenberg et al., 2015). Until now, no mass spectrometric verification has been performed for eicosanoids and other oxylipins in *daphnids*. These lipid mediators are present at very low levels in endogenous systems (Balgoma et al., 2013) and therefore require targeted methods for

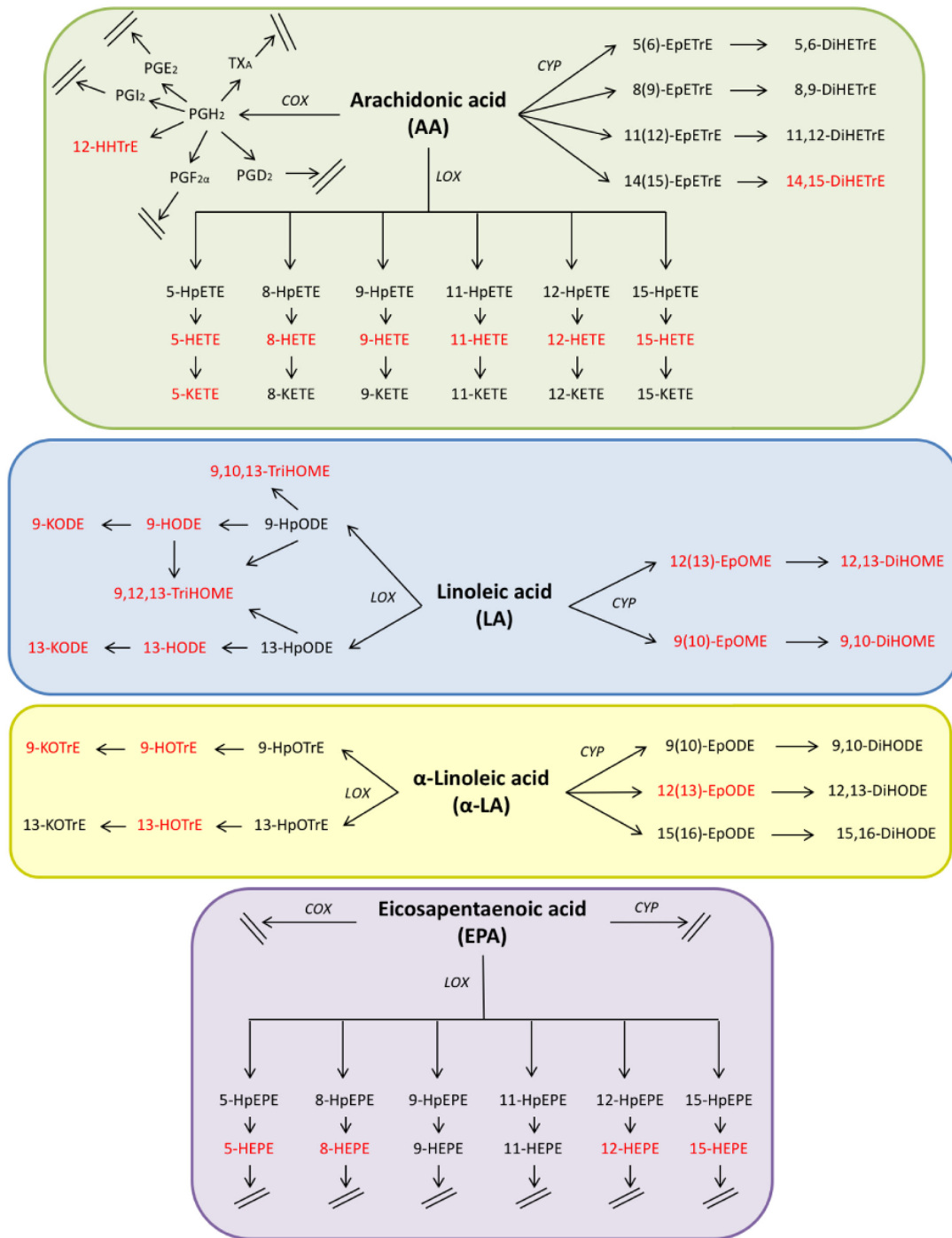


Fig. 1. Outline of the oxylipins biosynthesis pathways derived from (AA, green), linoleic acid (LA, blue), α-linoleic acid (α-LA, yellow) and eicosapentaenoic acid (EPA, purple) via cyclooxygenase (COX), lipoxygenase (LOX) and cytochrome 450 (CYP) pathways. Oxylipins marked in red were detected and quantified in *D. magna*. Diagram modified from Gabbs et al. (2015). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

quantification. Determining their biological activities, as well as their profiles in *D. magna*, will increase our understanding and could have very important implications for several areas of ecological risk assessment.

Using ultra performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS), the aim of the present study was to characterize the alterations induced by fluoxetine, carbamazepine and diazepam in the *D. magna* eicosanoid and other oxylipin profiles.

2. Materials and methods

2.1. *D. magna* growth conditions

All experiments were performed using the well-characterized clone F of *D. magna* maintained indefinitely as pure parthenogenetic cultures (Barata and Baird, 1998). Photoperiod was set to 16 h light: 8 h dark cycle and temperature at 20 ± 1 °C. The culture medium was changed every other day. Bulk cultures of 10 animals in 1 L were maintained in ASTM hard water at high food ration levels (5×10^5 cells mL⁻¹ of *Chlorella vulgaris*), as described in Barata and Baird (1998).

2.2. Exposure assays

Exposure assays were performed using fluoxetine hydrochloride (CAS-No. 56296-78-7; analytical standard, purity 100%, Sigma-Aldrich), carbamazepine (CAS-No. 298-46-4; analytical standard, purity 99%, Sigma-Aldrich) and diazepam (CAS-No. 439-14-5; analytical standard, purity 99%, Sigma-Aldrich). Due to the high number of animals needed to carry out the tests (60 adults per treatment) two independent assays were performed, with one control group for each assay: (i) fluoxetine and carbamazepine and (ii) diazepam. Experiments were initiated with neonates <14 h old collected from third brood females from the bulk cultures described above. Animals were exposed to the studied drugs in triplicate in groups of 20 individuals in 2 L glass vessels at a concentration of 100 ng/L in ASTM at high food. The studied drugs were added directly in water from stock solutions of 100 µg/L prepared in nanopure Milli Q water. An additional three 2 L glass vessels filled with 20 animals without drugs were used as control treatments. Media was changed every other day. Remaining culture conditions were similar to those depicted above. Animals were maintained until they released the first brood of eggs into their brood pouch (8–10 days). Then, animals were brooded by gently flushing water into their brood pouch, pooled into eppendorfs, snap-frozen in liquid N₂ and stored at –80 °C until extraction.

2.3. Extraction protocol

Oxylipins and their respective internal standards were purchased from Cayman (Ann Arbor, MI, USA). For the extraction of oxylipins, 10 µL of internal standards solution was spiked to the pools of 20 *daphnids* and dissolved in 400 µL methanol (MeOH). Homogenization was performed using stainless steel beads (0.9–2.0 mm of diameter) added to each sample, and shaken for 6 min with a Digital Cell Disruptor supplied by Scientific Industries. Samples were sonicated in an ultrasound bath for 20 min, centrifuged at 10,000g for 5 min at 4 °C and the supernatant was withdrawn. *Daphnids* were then re-extracted with 350 µL of a mixture of MeOH:CHCl₃:H₂O (4:2:1), following the above described procedure. Finally, the supernatants were combined and evaporated to dryness under nitrogen gas. The sample extract was then purified using solid phase extraction as previously described by Balgoma et al. (2016) with some modifications. Briefly, 1 mL of extraction buffer (Na₂HPO₄:citric acid pH = 5.6) was added to the dry extracts and automated solid phase extraction (SPE) on an Extrahera™ automated extraction system (Biotage, Uppsala, Sweden) using ABN Biotage cartridge columns (60 mg, 3 mL, Biotage). The resulting fraction was concentrated under nitrogen gas and reconstituted with 70 µL of MeOH and 10 µL of double-deionized water. Extracts were filtered through 0.1 µm membrane spin filters (Merck Millipore, Darmstadt, Germany) before UPLC-MS/MS analysis. In addition, two quality controls (QC) were pooled from *daphnids* samples, extracted as the samples, and then used to control for quantification reproducibility and batch extraction effects.

2.4. UPLC-MS/MS quantification of oxylipins

The chromatographic separation was performed on an Acquity UPLC BEH C18 column equipped with a pre-column (ACQUITY UPLC BEH C18 VanGuard Pre-column, 1.7 µm, 2.1 mm × 5 mm) (Mildford, US). Oxylipins were determined using a Xevo TQ-S mass spectrometer (Waters) equipped with an Electrospray Ion Source (ESI) operated in negative ionization mode. The details of the chromatographic and MS conditions have been previously published (Balgoma et al., 2016). For quantification purposes, a calibration curve was prepared by spiking 10 µL of the internal standard mixture to 10 calibration levels. A calibration curve was injected at the beginning and at the end of the sequence. Linear regression curves with 1/x weighting were built using the ratio between each compound and its respective internal standard. A list of detected oxylipins and their selected reaction monitoring (SRM) transitions is provided in Table S11.

A chiral LC-MS/MS method was used for the separation and identification of the 9-hydroxy-octadecadienoic acid (9-HODE) and 13-hydroxy-octadecadienoic acid (13-HODE) enantiomers to determine if the origin was enzymatic (S) or non-enzymatic (R) in the samples analyzed. For chirality determination, separation was carried out on an Acquity UPLC separation system using a Chiralpak® AD-RH column (150 × 2.1 mm, 5 µm) from Daicel Corporation (France). A linear gradient was set between 0 and 6.3 min from 40% of mobile phase A (0.1% of acetic acid in water) to 98% of mobile phase B (Acetonitrile:Isopropanol, 90/10 (v/v)). The flow rate and column oven were set to 0.6 mL min⁻¹ and 40 °C, respectively. The MS was operated in the negative ionization mode. The SRM transitions for the compounds measured are reported in Table S11.

2.5. Statistical methods

Effects of fluoxetine and carbamazepine treatments relative to unexposed controls on concentration levels of oxylipins were assessed by one-way ANOVA analyses followed by Dunnett's post-hoc tests. For diazepam Student's *t*-test were used to compare treatment effects. Hierarchical clustering of fold changes relative to controls of oxylipins across treatments was performed using the Euclidean distance. The software used in this work includes MassLynx, TargetLynx (Waters, Milford, MA) and Graph Pad Prism 5 for Windows (GraphPad Software, San Diego, CA, USA). ANOVA, Student's *t*-test and clustering analyses were performed using the Statistics toolbox under MATLAB 2015a computer and visualization environment (The Mathworks Inc. Natick, MA, USA).

3. Results

3.1. Detection of oxylipins in *D. magna*

A total of 28 oxylipins derived from different PUFAs were detected and quantified in *D. magna* controls as well as in samples exposed to fluoxetine, carbamazepine and diazepam (Table 1). Most oxylipins detected and quantified were biosynthesized via the COX, LOX and CYP enzymatic pathways (Fig. 1). These included seven oxylipins from the AA: 12-hydroxy-heptadecatrienoic acid (12-HHTrE) formed by COX enzymes (Gabbs et al., 2015) and HETEs derived from the LOX pathways (e.g., 5-HETE, 8-HETE, 9-HETE, 11-HETE, 12-HETE and 15-HETE). From the LA, α-LA and EPA pathways measured oxylipins included derivatives of the LOX pathways: HODEs, oxo-octadecadienoic acids (KODEs), HOTrEs -octadecatrienoic acids (KOTrEs) and HEPEs (e.g., 5-HEPE, 8-HEPE, 12-HEPE and 15-HEPE). In addition, epoxide products from the CYP epoxigenase-dependent metabolism of AA (e.g., 14,15-DiHETrE), LA (e.g., 12(13)-EpOME to 12,13-DiHOME, and 9(10)-EpOME to 9,10-DiHOME) and α-LA (e.g., 12(13)-EpODE) were detected (Table 1). A epoxy-keto-octadecenoic acid derivative (EKODE) belonging to the La pathway was also detected. The abundance for LA-derived oxylipins was found to be higher than AA-derived oxylipins.

Table 1

Concentrations of quantified oxylipins in *D. magna*. Control (i) refers to the fluoxetine and carbamazepine control samples, and control (ii) refers to diazepam control samples. An oxylipin nomenclature list is provided in Table S11.

PUFA	Oxylipins	Control (i) (ng/daphnia ± SD)	Fluoxetine (ng/daphnia ± SD)	Carbamazepine (ng/daphnia ± SD)	Control (ii) (ng/daphnia ± SD)	Diazepam (ng/daphnia ± SD)
AA	5-KETE	2.94 ± 0.17	3.03 ± 0.18	3.88 ± 0.32	2.03 ± 0.05	3.89 ± 0.13
	12-HHTrE	3.24 ± 0.23	7.27 ± 0.15	4.08 ± 0.33	6.83 ± 0.35	5.59 ± 0.26
	14,15-DiHETrE	0.55 ± 0.03	0.74 ± 0.03	0.58 ± 0.04	0.35 ± 0.02	0.70 ± 0.04
	9-HETE	2.54 ± 0.25	2.80 ± 0.07	3.01 ± 0.13	2.52 ± 0.06	3.10 ± 0.26
	8-HETE	2.95 ± 0.05	3.27 ± 0.18	4.08 ± 0.15	1.82 ± 0.05	2.62 ± 0.12
	5-HETE	5.35 ± 0.11	5.91 ± 0.10	5.88 ± 0.33	3.21 ± 0.11	6.97 ± 0.15
	11-HETE	2.08 ± 0.04	2.45 ± 0.08	2.38 ± 0.18	1.42 ± 0.01	2.28 ± 0.09
	15-HETE	2.50 ± 0.02	4.06 ± 0.22	3.85 ± 0.25	1.99 ± 0.12	3.30 ± 0.05
	12-HETE	1.34 ± 0.05	1.57 ± 0.05	1.87 ± 0.10	0.54 ± 0.02	1.82 ± 0.07
	EKODE	10.83 ± 0.59	33.19 ± 2.85	18.75 ± 0.80	15.84 ± 0.59	33.93 ± 2.88
	9-KODE	83.32 ± 1.63	116.64 ± 5.39	93.69 ± 3.19	84.39 ± 1.82	139.36 ± 4.22
	LA	13-KODE	40.58 ± 1.87	51.23 ± 1.37	47.07 ± 2.04	37.52 ± 1.13
9(10)-EpOME		13.99 ± 0.50	25.69 ± 3.12	17.56 ± 0.83	20.84 ± 0.47	31.87 ± 0.42
12(13)-EpOME		11.60 ± 0.54	17.61 ± 2.15	13.27 ± 0.53	14.05 ± 0.15	19.73 ± 0.14
9,12,13-TriHOME		5.27 ± 0.33	7.39 ± 0.27	5.02 ± 0.30	4.54 ± 0.46	7.28 ± 0.73
9,10-DiHOME		15.27 ± 0.30	18.68 ± 0.64	16.34 ± 0.14	16.48 ± 0.57	17.14 ± 0.14
9,10,13-TriHOME		12.94 ± 1.25	16.42 ± 0.19	12.90 ± 0.57	12.87 ± 1.19	16.53 ± 0.56
12,13-DiHOME		13.19 ± 0.34	14.88 ± 0.77	13.78 ± 0.21	14.14 ± 0.37	13.91 ± 0.19
9-HODE		89.03 ± 4.85	121.73 ± 6.05	102.48 ± 2.49	84.67 ± 0.85	133.19 ± 2.54
13-HODE		72.02 ± 1.58	106.28 ± 1.87	92.11 ± 1.77	72.70 ± 0.94	78.79 ± 1.44
13-HOTrE		27.67 ± 1.76	39.14 ± 1.58	34.88 ± 0.91	24.62 ± 0.46	45.20 ± 1.03
9-KOTrE		14.25 ± 1.82	19.39 ± 1.80	14.08 ± 1.10	13.77 ± 0.40	28.61 ± 0.42
9-HOTrE		47.12 ± 6.21	62.72 ± 5.48	49.34 ± 3.13	42.44 ± 1.95	83.02 ± 0.73
α-LA	12(13)-EpODE	2.17 ± 0.08	3.98 ± 0.23	2.75 ± 0.12	3.11 ± 0.10	4.27 ± 0.10
	8-HEPE	1.37 ± 0.05	1.44 ± 0.08	2.04 ± 0.12	0.83 ± 0.05	0.96 ± 0.05
	15-HEPE	1.19 ± 0.01	1.97 ± 0.08	1.70 ± 0.10	1.31 ± 0.10	1.23 ± 0.09
EPA	12-HEPE	0.93 ± 0.05	1.09 ± 0.09	1.37 ± 0.16	0.88 ± 0.06	0.97 ± 0.03
	5-HEPE	2.55 ± 0.07	2.50 ± 0.11	2.94 ± 0.25	1.64 ± 0.03	2.90 ± 0.06

The quantified oxylipins are mapped in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Kanehisa et al., 2014) for AA, LA and α-LA in *Daphnia pulex* (see Supplementary Figures S11, S12 and S13, respectively), in order to contextualize them in the biology of *D. magna*. As it can be observed, most of the oxylipins are end products of the pathways. However, not all the oxylipins appear in the KEGG database and therefore could not be represented, such as 12-HHTrE for AA pathway, 9-KOTrE for the LA pathways and 12(13)-EpODE for the α-LA pathway.

The abundance of some *S* and *R* stereoisomers of monohydroxy lipid mediators was determined using chiral chromatography. Due to the low abundance of most of these compounds, chirality could only be determined for 9- and 13-HODE. For both compounds, the *S* was the more predominant form, with an enantiomeric excess between 44% and 66% for 9-HODE and between 14% and 30% for 13-HODE. Results for the specific treatments are detailed in Table 2.

3.2. Effects of anti-depressant pharmaceuticals

Fold changes of quantified oxylipins are given in Fig. 2, where more significant ones ($p < 0.05$) have been marked. Of the three tested drugs, diazepam produced the most pronounced changes in oxylipin levels (Fig. 2). Increased levels ($p < 0.05$, Student *t*-tests) were found for the AA-derivatives, 5-KETE, 5-HETE and 12-HETE, LA-derivatives, 9-KODE, 9(10)-EpOME, 12(13)-EpOME and 9-HODE, α-LA derivatives, 9-HOTrE, 13-HOTrE, 9-KOTrE and 12(13)-EpODE and the EPA-derivative

5-HEPE. Exposure of *D. magna* with fluoxetine led only to increases ($p < 0.05$, ANOVA) in 12-HHTrE. Only the LA-derivative 13-HODE increased after carbamazepine exposure ($p < 0.05$, ANOVA) (Fig. 2).

Heat map representation of fold changes of oxylipins across treatments is shown in Fig. 3. As illustrated in Fig. 3, hierarchical clustering separated exposure treatments fluoxetine and carbamazepine from diazepam. The oxylipins EKODE and 12-HETE were represented in red, with fold changes values higher than 3, for fluoxetine and diazepam treatments, respectively. Fluoxetine showed elevated fold changes for EKODE, 12-HHTrE, 9(10)-EpOME and 12(13)-EpODE, whereas carbamazepine did so for EKODE and 15-HETE. Diazepam treatment increased levels of some AA, LA, EPA and most of the detected α-LA-derivatives being the exception 12(13)-EpODE, whose fold change was around 1.

4. Discussion

In the present study, we used a comprehensive UPLC-MS/MS method to determine changes in eicosanoids and other oxylipins in *D. magna* exposed to three different psychiatric drugs. Free PUFAs are oxygenated by three families of enzymes (COX, LOX and epoxygenase CYP) into distinct classes of oxylipins (Massey and Nicolaou, 2013).

Enantiomeric excess of 9-(*S*) HODE and 13-(*S*) HODE over 9-(*R*) HODE and 13-(*R*) HODE, respectively (as determined by chiral chromatography), suggests that there is an enzymatic activity involved in the production of these compounds in *D. magna*. Other LOX-derived

Table 2

Contribution of the *R* or *S* enantiomers of 9-HODE and 13-HODE resolved by the chiral method in the *D. magna* samples. Also, the enantiomeric excess was calculated.

	% 9(<i>S</i>)-HODE	% 9(<i>R</i>)-HODE	Enantiomeric excess	% 13(<i>S</i>)-HODE	% 13(<i>R</i>)-HODE	Enantiomeric excess
Control (i)	75	25	50	60	40	20
Fluoxetine	83	17	66	65	35	30
Carbamazepine	74	26	48	62	38	24
Control (ii)	72	28	44	58	42	16
Diazepam	78	22	56	57	43	14

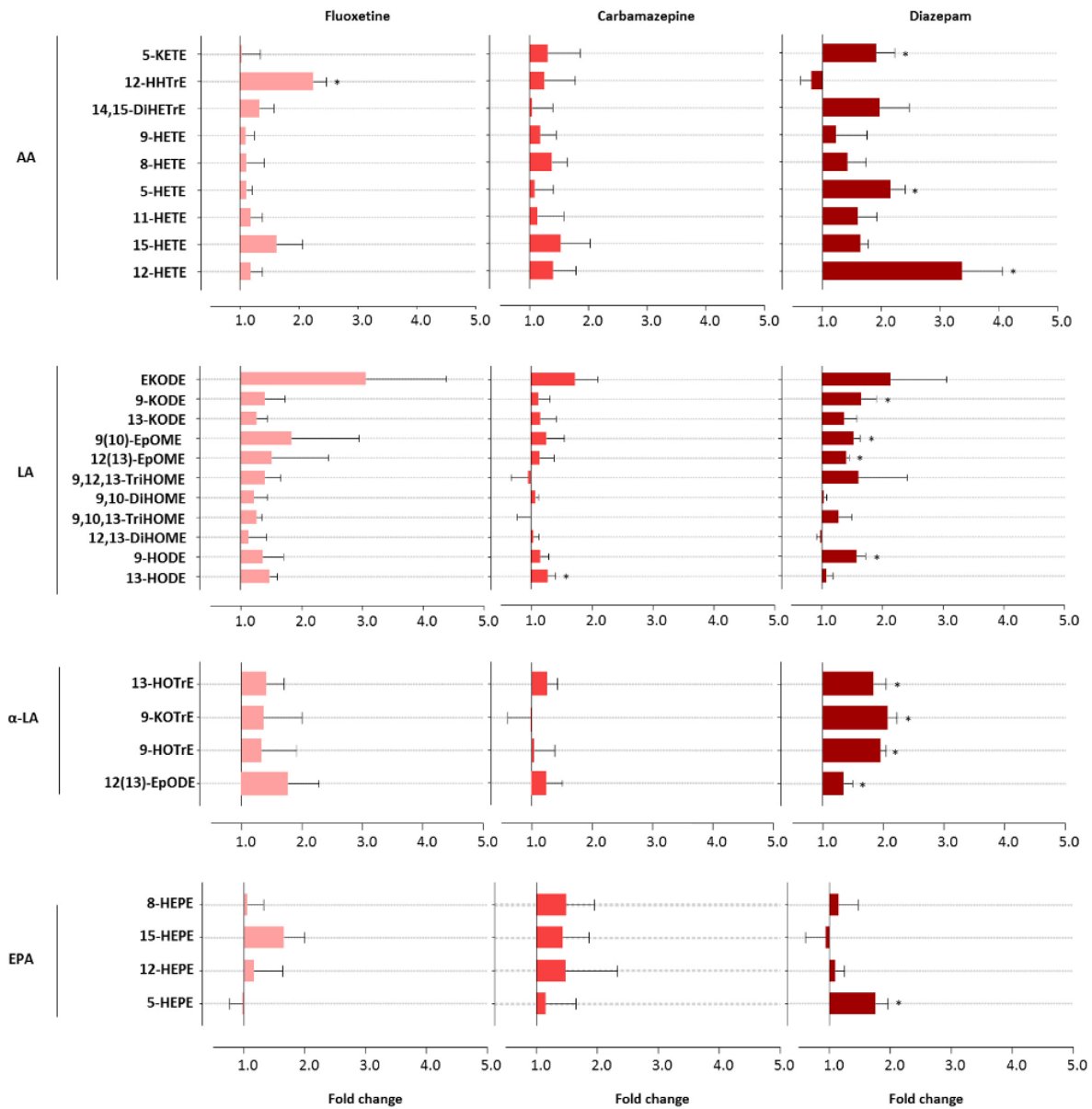


Fig. 2. Representation of the fold changes of each oxylipin in each exposure treatment respect to the control samples. Oxylipins with significant changes relative to controls ($p < 0.05$, ANOVA or Student's *t*-tests) are marked with an asterisk.

oxylipins quantified in *D. magna* (Table S12), such as 5-HETE and 15-HETE, could not be detected using the chiral method due to their lower abundance.

The most detected oxylipins were final products of the enzymatic pathways involved, which may indicate a high rate of turnover of intermediated products (Fig. 1). Within AA-pathways related oxylipins, HETEs (e.g., 8-HETE, 9-HETE, 11-HETE, 12-HETE and 15-HETE), as well as 5-KETE and 14,15-DiHETrE are end-products for LOX and CYP pathways, respectively. In addition, KODEs (9-KODE and 13-KODE) and TriHOMEs (9,12,13-TriHOME and 9,10,13-TriHOME) are LA-derived endproducts for the LOX pathway, and DiHOMEs (12,13-DiHOME and 9,10-DiHOME) are LA-derived end-products for the CYP pathway.

Besides, HOTrEs (9-HOTrE and 13-HOTrE) are also final products in the LOX pathway for α -LA.

Unfortunately, PGs could not be detected in *D. magna* samples. Previous studies indicated that prostanoids are mostly present in the ovary and are more abundant in early stages of oocyte maturation in decapods (Sumpownon et al., 2015). We sampled *D. magna* adults just after the release of eggs into the brood pouch when the ovaries have their minimum size and start a new egg provisioning cycle (Baird et al., 1991). This may have prevented our ability to detect prostanoids.

In *D. magna* the total amount of lipids is dependent upon dietary lipid content, feeding and ontogeny. Under high food ratios, within a single instar interval, the proportion of lipids can vary between 16%

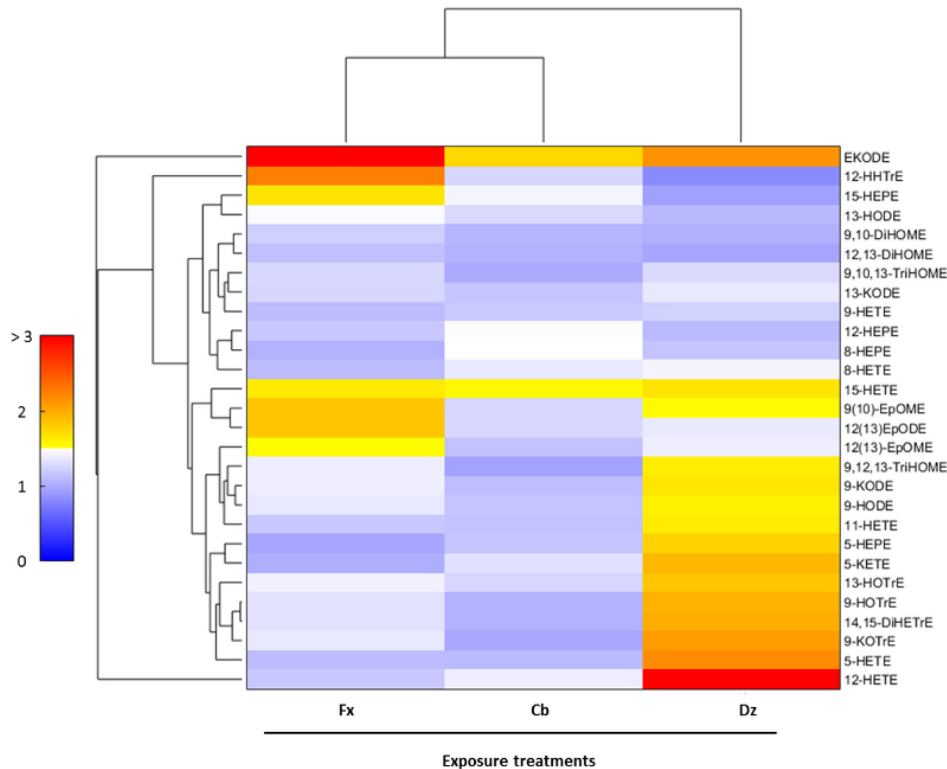


Fig. 3. Heat map representation with hierarchical clustering of fold changes of oxylipins across treatments. Data has been autoscaled relative to controls, so 0 value (blue) corresponds to no change relative to controls. Fx: fluoxetine exposure, Cb: carbamazepine exposure and Dz: diazepam exposure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and 22% of their dry mass (Barata et al., 2005; Goulden and Place, 1993). In the present study, the use of constant food conditions and a sampling scheme that consists in sampling adult females at the beginning of the intermolt interval, respectively, prevented large differences in total lipid to be found across animals. Reported fatty acid composition in *D. magna* individuals, cultured in similar conditions as those used here, indicated that among PUFAs, the C18 group is the most common including LA (18:2n-6) and α -LA (18:3n-3). PUFAs of 20 carbon chain are found only in small amounts, being AA (20:4n-6) followed by EPA (20:5n-3) the most abundant ones (Barata et al., 2005). Endogenous fatty acid synthesis contributes <5% of the accumulated total fatty acids in *D. magna* with most of them being incorporated from food (Cowgill et al., 1984; Sundbom and Vrede, 1997). The number of oxylipins quantified in whole *D. magna* tissues agrees with their relative abundance, being in decreasing order LA, AA, α -LA and EPA, respectively.

The biological role of eicosanoids in *D. magna* is still unclear. For AA, in silico and transcriptomic studies suggest that only the COX and the LOX pathways are present in *D. magna* (Heckmann et al., 2008). Other genomic studies (Baldwin et al., 2009; Kim et al., 2018), contradict the bioinformatic arguments of Heckmann et al. (2008), who reported that there is no current evidence of CYP pathway in *D. magna* and other arthropods. Detection of 14,15-DiHETrE, which is a downstream product of the CYP produced 14(15)-EpETrE, and of 12(13)-EpOME and 9(10)-EpOME provides an evidence for the existence of eicosanoid CYP epoxygenase pathway in the AA and LA pathways in *D. magna*. Our results, thus, are in line with the findings of Baldwin et al. (2009) and Kim et al. (2018). Recently, sEH activities and epoxy fatty acids have been found in the mosquito *Culex quinquefasciatus* (Xu et al., 2015). In

the present study, the detection and quantification of DiHOMEs in *D. magna*, related with the sEH activities, confirmed the presence of an eicosanoid CYP pathway. However, further research is needed to confirm that finding.

Fluoxetine enhances serotonin activity in *D. magna* (Campos et al., 2016). There is also a cross-talk between serotonin receptors and AA/PG metabolism (Tournois et al., 1998), which in insects regulates reproduction (Stanley, 2006). The increase of 12-HHTrE upon fluoxetine exposure, which was identified as a product of prostaglandin H₂ (PGH₂) (Matsunobu et al., 2013), suggests that this psychiatric drug may increase, and consequently, up-regulate PG synthesis and, hence, offspring production. Little is known about the physiological role of LOX and CYP derivate oxylipins in insects and crustaceans (Heckmann et al., 2008; Stanley, 2006). Medeiros et al. (2004) found that LOX inhibitors decreased yolk uptake in the insect *Rhodnius prolixus*, whereas a specific LOX product, leukotriene B₄ (LTB₄), stimulated its uptake. Conversely, in an early study, the same authors found that the prostaglandin E₂ (PGE₂) down-regulates the ovarian yolk uptake in the same species (Medeiros et al., 2002). Thus, the previous results indicate that LOX may be involved in reproduction. Accordingly, the observed increased levels of several LOX and CYP products by diazepam and, to a lesser extent by carbamazepine, also supports the hypothesis that the studied psychiatric drugs up-regulated oxylipins metabolism and as such reproduction in *D. magna*. In a previous work, it was reported that the three studied drugs enhanced reproduction in *D. magna* at similar low exposure levels (Rivetti et al., 2016). Our results may thus provide additional evidence for a link between observed up-regulation of COX and LOX related oxylipins products and enhanced offspring production in *D. magna*, upon exposure to psychiatric drugs.

5. Conclusions

In this study, we report for the first time, a total of 28 different oxylipins belonging to the COX, LOX and CYP pathways for the AA, LA, α -LA and EPA pathways in *D. magna* individuals. All detected oxylipins were final oxidized products of the enzymatic pathways involved. Some of the eicosanoids detected, such as 12-HHTrE, are involved in PG and leukotriene biosynthetic pathways, respectively (Hamberg and Samuelsson, 1974; Stanley, 2006), which regulate reproduction in insects (Medeiros et al., 2004; Medeiros et al., 2002; Stanley, 2006).

Exposure to environmental concentrations of psychiatric drugs enhanced the differential accumulation of specific oxylipins. Fluoxetine increased the accumulation of the PG product 12-HHTrE, whereas diazepam increased the concentration of 12 eicosanoids and other oxylipins belonging to the LOX and CYP pathways of AA, LA, α -LA and EPA. Some of them (5-HETE and 5-KEPE) are closely related to the leukotriene metabolic pathway. Carbamazepine only increased significantly one LOX metabolite from the LA pathway (13-HODE). Previous studies in insects indicated that both COX and LOX products, PGs and leukotrienes, respectively, regulate reproduction (Medeiros et al., 2004; Medeiros et al., 2002; Stanley, 2006). This means that despite having different modes of action, fluoxetine and diazepam up-regulated eicosanoid pathways, closely related to known biologically active products that regulate reproduction. Results of carbamazepine are difficult to interpret, as there is no evidence in invertebrates that oxylipins from the LA-pathway affected reproduction.

Acknowledgments

The research leading to these results received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. 320737 and the Spanish project (CTM2017-83242-R).

Compliance with ethical standards

The authors declare no financial conflict of interest. This article does not contain any studies with human participants or animals performed by any of the authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2018.06.333>.

References

Ambrósio, A.F., Soares-da-Silva, P., Carvalho, C.M., Carvalho, A.P., 2002. Mechanisms of action of carbamazepine and its derivatives, oxcarbazepine, BIA 2-093, and BIA 2-024. *Neurochem. Res.* 27, 121–130.

Baird, D.J., Barber, I., Soares, A.M.V.M., Calow, P., 1991. An early life-stage test with *Daphnia magna* straus: an alternative to the 21-day chronic test? *Ecotoxicol. Environ. Saf.* 22, 1–7.

Baldwin, W.S., Marko, P.B., Nelson, D.R., 2009. The cytochrome P450 (CYP) gene superfamily in *Daphnia pulex*. *BMC Genomics* 10, 169.

Balgoma, D., Checa, A., Sar, D.G., Snowden, S., Wheelock, C.E., 2013. Quantitative metabolic profiling of lipid mediators. *Mol. Nutr. Food Res.* 57, 1359–1377.

Balgoma, D., Yang, M., Sjödin, M., Snowden, S., Karimi, R., Levänen, B., et al., 2016. Linoleic acid-derived lipid mediators increase in a female-dominated subphenotype of COPD. *Eur. Respir. J.* 47, 1645–1656.

Barata, C., Baird, D.J., 1998. Phenotypic plasticity and constancy of life-history traits in laboratory clones of *Daphnia magna* straus: effects of neonatal length. *Funct. Ecol.* 12, 442–452.

Barata, C., Carlos Navarro, J., Varo, I., Carmen Riva, M., Arun, S., Porte, C., 2005. Changes in antioxidant enzyme activities, fatty acid composition and lipid peroxidation in *Daphnia magna* during the aging process. *Comp. Biochem. Physiol. B: Biochem. Mol. Biol.* 140, 81–90.

Barquissau, V., Ghandour, R.A., Ailhaud, G., Klingenspor, M., Langin, D.F., Amri, E.-Z., et al., 2017. Control of adipogenesis by oxylipins, GPCRs and PPARs. *Biochimie* 136, 3–11.

Biswas, B., Carlsson, A., 1978. On the mode of action of diazepam on brain catecholamine metabolism. *Naunyn Schmiedeberg's Arch. Pharmacol.* 303, 73–78.

Buczynski, M.W., Dumlao, D.S., Dennis, E.A., 2009. An integrated omics analysis of eicosanoid biology. *J. Lipid Res.* 50, 1015–1038.

Campos, B., Rivetti, C., Kress, T., Barata, C., Dirksen, H., 2016. Depressing antidepressant: fluoxetine affects serotonin neurons causing adverse reproductive responses in *Daphnia magna*. *Environ. Sci. Technol.* 50, 6000–6007.

Checa, A., Holm, T., Sjödin, M.O.D., Reinke, S.N., Alm, J., Scheynius, A., et al., 2015. Lipid mediator profile in vernix caseosa reflects skin barrier development. *Sci. Rep.* 5.

Cowgill, U.M., Williams, D.M., Esquivel, J.B., 1984. Effects of maternal nutrition on fat content and longevity of neonates of *Daphnia magna*. *J. Crustac. Biol.* 42, 173–190.

Dailey, J.W., Reith, M.E.A., Yan, Q.S., Li, M.Y., Jobe, P.C., 1997. Carbamazepine increases extracellular serotonin concentration: lack of antagonism by tetrodotoxin or zero Ca^{2+} . *Eur. J. Pharmacol.* 328, 153–162.

Fent, K., Weston, A.A., Caminada, D., 2006. Ecotoxicology of human pharmaceuticals. *Aquat. Toxicol.* 76, 122–159.

Ford, A.T., Fong, P.P., 2016. The effects of antidepressants appear to be rapid and at environmentally relevant concentrations. *Environ. Toxicol. Chem.* 35, 794–798.

Fossat, P., Bacqué-Cazenave, J., De Deurwaerdère, P., Delbecq, J.P., Cattaeert, D., 2014. Anxiety-like behavior in crayfish is controlled by serotonin. *Science* 344, 1293–1297.

Gabbs, M., Leng, S., Devassy, J.G., Monirujaman, M., Aukema, H.M., 2015. Advances in our understanding of oxylipins derived from dietary PUFAs. *Advances in nutrition: an international review. Journal* 6, 513–540.

Garreta-Lara, E., Campos, B., Barata, C., Lacorte, S., Tauler, R., 2016. Metabolic profiling of *Daphnia magna* exposed to environmental stressors by GC-MS and chemometric tools. *Metabolomics* 12.

González-Peña, D., Checa, A., de Ancos, B., Wheelock, C.E., Sánchez-Moreno, C., 2017. New insights into the effects of onion consumption on lipid mediators using a diet-induced model of hypercholesterolemia. *Redox Biol.* 11, 205–212.

Goulden, C.E., Place, A.R., 1993. Lipid accumulation and allocation in Daphniid-Cladocera. *Bull. Mar. Sci.* 53, 106–114.

Gunnarsson, L., Jauhainen, A., Kristiansson, E., Nerman, O., Larsson, D.G.J., 2008. Evolutionary conservation of human drug targets in organisms used for environmental risk assessments. *Environ. Sci. Technol.* 42, 5807–5813.

Hamberg, M., Samuelsson, B., 1974. Prostaglandin endoperoxides. Novel transformations of arachidonic acid in human platelets. *Proc. Natl. Acad. Sci. U. S. A.* 71, 3400–3404.

Heckmann, L.-H., Sibly, R.M., Timmermans, M.J., Callaghan, A., 2008. Outlining eicosanoid biosynthesis in the crustacean *Daphnia*. *Front. Zool.* 5, 11.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42 (D199–D205).

Kim, R.-O., Jo, M.-A., Song, J., Kim, I.-C., Yoon, S., Kim, W.-K., 2018. Novel approach for evaluating pharmaceuticals toxicity using *Daphnia* model: analysis of the mode of cytochrome P450-generated metabolite action after acetaminophen exposure. *Aquat. Toxicol.* 196, 35–42.

Kolpin, D.W., Furlong, E.T., Meyer, M.T., Thurman, E.M., Zaugg, S.D., Barber, L.B., et al., 2002. Pharmaceuticals, hormones, and other organic wastewater contaminants in U.S. streams, 1999–2000: a national reconnaissance. *Environ. Sci. Technol.* 36, 1202–1211.

Kravitz, E.A., 2000. Serotonin and aggression: insights gained from a lobster model system and speculations on the role of amine neurons in a complex behavior. *Journal of Comparative Physiology - A Sensory, Neural, and Behavioral Physiology* 186, 221–238.

Labhsetwar, A.P., Zolovick, A., 1973. Hypothalamic interaction between prostaglandins and catecholamines in promoting gonadotrophin secretion for ovulation. *Nat. New Biol.* 246, 55–56.

Lundstrom, S.L., Balgoma, D., Wheelock, A.M., Haeggstrom, J.Z., Dahlen, S.E., Wheelock, C.E., 2011. Lipid mediator profiling in pulmonary disease. *Curr. Pharm. Biotechnol.* 12, 1026–1052.

Lundström, S.L., Levänen, B., Nording, M., Klepczynska-Nyström, A., Sköld, M., Haeggström, J.Z., et al., 2011. Asthmatics exhibit altered oxylipin profiles compared to healthy individuals after subway air exposure. *PLoS One* 6, e23864.

Martin-Creuzburg, D., Wacker, A., Basena, T., 2010. Interactions between limiting nutrients: consequences for somatic and population growth of *Daphnia magna*. *Limnol. Oceanogr.* 55, 2597–2607.

Massey, K.A., Nicolaou, A., 2013. Lipidomics of oxidized polyunsaturated fatty acids. *Free Radic. Biol. Med.* 59, 45–55.

Matsunobu, T., Okuno, T., Yokoyama, C., Yokomizo, T., 2013. Thromboxane A synthase-independent production of 12-hydroxyheptadecatrienoic acid, a BLT2 ligand. *J. Lipid Res.* 54, 2979–2987.

Medeiros, M.N., Oliveira, D.M.P., Paiva-Silva, G.O., Silva-Neto, M.A.C., Romeiro, A., Bozza, M., et al., 2002. The role of eicosanoids on Rhodnius heme-binding protein (RHBP) endocytosis by *Rhodnius prolixus* ovaries. *Insect Biochem. Mol. Biol.* 32, 537–545.

Medeiros, M.N., Mendonça, L.H., Hunter, A.L., Paiva-Silva, G.O., Mello, F.G., Henze, I.P., et al., 2004. The role of lipoxygenase products on the endocytosis of yolk proteins in insects: participation of cAMP. *Arch. Insect Biochem. Physiol.* 55, 178–187.

Metcalfe, C.D., Chu, S.G., Judt, C., Li, H.X., Oakes, K.D., Servos, M.R., et al., 2010. Antidepressants and their metabolites in municipal wastewater, and downstream exposure in an urban watershed. *Environ. Toxicol. Chem.* 29, 79–89.

Muñoz, I., López-Doval, J.C., Ricart, M., Villagrasa, M., Brix, R., Geislinger, A., et al., 2009. Bridging levels of pharmaceuticals in river water with biological community structure in the Llobregat river basin (northeast Spain). *Environ. Toxicol. Chem.* 28, 2706–2714.

Petrović, M., Gonzalez, S., Barceló, D., 2003. Analysis and removal of emerging contaminants in wastewater and drinking water. *TrAC - Trends in Analytical Chemistry* 22, 685–696.

Riss, J., Cloyd, J., Gates, J., Collins, S., 2008. Benzodiazepines in epilepsy: pharmacology and pharmacokinetics. *Acta Neurol. Scand.* 118, 69–86.

- Rivetti, C., Campos, B., Barata, C., 2016. Low environmental levels of neuro-active pharmaceuticals alter phototactic behaviour and reproduction in *Daphnia magna*. *Aquat. Toxicol.* 170, 289–296.
- Santos, E.A., Keller, R., Rodriguez, E., Lopez, L., 2001. Effects of serotonin and fluoxetine on blood glucose regulation in two decapod species. *Braz. J. Med. Biol. Res.* 34, 75–80.
- Sarojini, R., Nagabhusanam, R., Fingerman, M., 1995. In vivo inhibition by dopamine of 5-hydroxytryptamine-stimulated ovarian maturation in the red swamp crayfish, *Procambarus clarkii*. *Experientia* 51, 156–158.
- Schmelzer, K.R., Kubala, L., Newman, J.W., Kim, I.-H., Eiserich, J.P., Hammock, B.D., 2005. Soluble epoxide hydrolase is a therapeutic target for acute inflammation. *Proc. Natl. Acad. Sci. U. S. A.* 102, 9772–9777.
- Spracklen, A.J., Tootle, T.L., 2015. *Drosophila*: a model for studying prostaglandin signaling. *Bioactive Lipid Mediators. Current Reviews and Protocols*, pp. 181–197.
- Stanley, D., 2006. Prostaglandins and other eicosanoids in insects: biological significance. *Annu. Rev. Entomol.* 51, 25–44.
- Strassburg, K., Huijbrechts, A.M.L., Kortekaas, K.A., Lindeman, J.H., Pedersen, T.L., Dane, A., et al., 2012. Quantitative profiling of oxylipins through comprehensive LC-MS/MS analysis: application in cardiac surgery. *Anal. Bioanal. Chem.* 404, 1413–1426.
- Sumpownon, C., Engsusophon, A., Siangcham, T., Sugiyama, E., Soonklang, N., Meeratana, P., et al., 2015. Variation of prostaglandin E₂-inf-2-/inf-concentrations in ovaries and its effects on ovarian maturation and oocyte proliferation in the giant fresh water prawn, *Macrobrachium rosenbergii*. *Gen. Comp. Endocrinol.* 223, 129–138.
- Sundbom, M., Vrede, T., 1997. Effects of fatty acid and phosphorous content of food on the growth, survival and reproduction of *Daphnia*. *Freshw. Biol.* 38, 665–674.
- Tierney, A.J., Kim, T., Abrams, R., 2003. Dopamine in crayfish and other crustaceans: distribution in the central nervous system and physiological functions. *Microsc. Res. Tech.* 60, 325–335.
- Tixier, C., Singer, H.P., Oellers, S., Müller, S.R., 2003. Occurrence and fate of carbamazepine, clofibrac acid, diclofenac, ibuprofen, ketoprofen, and naproxen in surface waters. *Environ. Sci. Technol.* 37, 1061–1068.
- Tournois, C., Mutel, V., Manivet, P., Launay, J.M., Kellermann, O., 1998. Cross-talk between 5-hydroxytryptamine receptors in a serotonergic cell line. Involvement of Arachidonic Acid Metabolism. *Journal of Biological Chemistry* 273, 17498–17503.
- Valcárcel, Y., Martínez, F., González-Alonso, S., Segura, Y., Catalá, M., Molina, R., et al., 2012. Drugs of abuse in surface and tap waters of the Tagus River basin: heterogeneous photo-Fenton process is effective in their degradation. *Environ. Int.* 41, 35–43.
- Vasskog, T., Anderssen, T., Pedersen-Bjergaard, S., Kallenborn, R., Jensen, E., 2008. Occurrence of selective serotonin reuptake inhibitors in sewage and receiving waters at Spitsbergen and in Norway. *J. Chromatogr. A* 1185, 194–205.
- Weiss, L.C., Kruppert, S., Laforch, C., Tollrian, R., 2012. Chaoborus and gasterosteus anti-predator responses in *Daphnia pulex* are mediated by independent cholinergic and gabaergic neuronal signals. *PLoS One* 7.
- Willenberg, I., Ostermann, A.I., Schebb, N.H., 2015. Targeted metabolomics of the arachidonic acid cascade: current state and challenges of LC-MS analysis of oxylipins. *Anal. Bioanal. Chem.* 407, 2675–2683.
- Xu, J., Morisseau, C., Yang, J., Mamatha, D.M., Hammock, B.D., 2015. Epoxide hydrolase activities and epoxy fatty acids in the mosquito *Culex quinquefasciatus*. *Insect Biochem. Mol. Biol.* 59, 41–49.
- Zhao, Y.Y., Cheng, X.L., Lin, R.C., 2014. Lipidomics applications for discovering biomarkers of diseases in clinical chemistry. *Int. Rev. Cell Mol. Biol.* 313, 1–26.

Supplementary Information

Effect of psychiatric drugs in *Daphnia magna* oxylipin profiles

Elba Garreta-Lara^a, Antonio Checa^b, David Fuchs^b, Silvia Lacorte^a, Romà Tauler^a, Craig E. Wheelock^b, Carlos Barata^{a*}

^aDepartment of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, 08034, Catalonia, Spain

^bDivision of Physiological Chemistry 2, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-17177 Stockholm, Sweden

Corresponding author:

Carlos Barata

E-mail address:

cbmqam@cid.csic.es

This file contents:

Table SI1	2
Figure SI1	4
Figure SI2	5
Figure SI3	6

TABLES

Table SII. List of the quantified oxylipins in *D. magna* samples, with their nomenclature and MS-specific parameter SRM transitions.

	Oxylipins	Nomenclature	SRM transition	
Standards	5-HETE	5-hydroxy-eicosatetraenoic acid	319.1 > 114.9	
	8-HETE	8-hydroxy-eicosatetraenoic acid	319.0 > 154.8	
	9-HETE	9-hydroxy-eicosatetraenoic acid	319.2 > 179.1	
	11-HETE	11-hydroxy-eicosatetraenoic acid	319.0 > 166.8	
	12-HETE	12-hydroxy-eicosatetraenoic acid	319.0 > 178.8	
	15-HETE	15-hydroxy-eicosatetraenoic acid	319.1 > 174.8	
	5-KETE	5-oxo-eicosatetraenoic acid	317.2 > 203.1	
	14,15-DiHETrE	14,15-dihydroxy-eicosatrienoic acid	337.3 > 206.9	
	12-HHTrE	12-hydroxy-heptadecatrienoic acid	279.2 > 179.0	
	9-HODE	9-hydroxy-octadecadienoic acid	295.3 > 171.0	
	13-HODE	13-hydroxy-octadecadienoic acid	295.2 > 195.0	
	9-KODE	9-oxo-octadecadienoic acid	293.2 > 185.0	
	13-KODE	13-oxo-octadecadienoic acid	293.2 > 112.9	
	9-HOTrE	9-hydroxy-octadecatrienoic acid	293.2 > 170.9	
	13-HOTrE	13-hydroxy-octadecatrienoic acid	293.2 > 195.2	
	9-KOTrE	9-oxo-octadecatrienoic acid	291.2 > 121.0	
	EKODE	epoxy-keto-octadecenoic acid	309.2 > 291.0	
	9(10)-EpOME	9(10)-epoxy-octadecenoic acid	295.2 > 170.9	
	12(13)-EpOME	12(13)-epoxy-octadecenoic acid	295.2 > 195.0	
	9,10-DiHOME	9,10-dihydroxy-octadecenoic acid	313.2 > 201.1	
	12,13-DiHOME	12,13-dihydroxy-octadecenoic acid	313.1 > 183.1	
	9,10,13-TriHOME	9,10,13-trihydroxy-octadecenoic acid	329.0 > 139.0	
	9,12,13-TriHOME	9,12,13-trihydroxy-octadecenoic acid	329.1 > 211.0	
	12(13)-EpODE	12(13)-epoxy-octadecadienoic acid	293.2 > 183.0	
	5-HEPE	5-hydroxy-eicosapentaenoic acid	317.3 > 114.8	
	8-HEPE	8-hydroxy-eicosapentaenoic acid	317.0 > 154.9	
	12-HEPE	12-hydroxy-eicosapentaenoic acid	317.0 > 178.8	
	15-HEPE	15-hydroxy-eicosapentaenoic acid	317.1 > 219.1	
	Internal Standards	d4-PGE2	d4-prostaglandin E2	355.3 > 275.1
		d6-5-KETE	d6-5-oxo-eicosatetraenoic acid	323.2 > 209.1
		d4-9-HODE	d4-9-hydroxy-octadecadienoic acid	299.3 > 172.0
		d8-15-HETE	d8-5-hydroxy-eicosatetraenoic acid	327.3 > 182.0
d4-9(10)-EpOME		d4-9(10)-epoxy-octadecenoic acid	299.2 > 172.1	
d4-9(10)-DiHOME		d4-9,10-dihydroxy-octadecenoic acid	317.2 > 202.9	
d11-14,15-DiHETrE		d11-14,15-dihydroxy-eicosatrienoic acid	348.3 > 207.1	

FIGURES

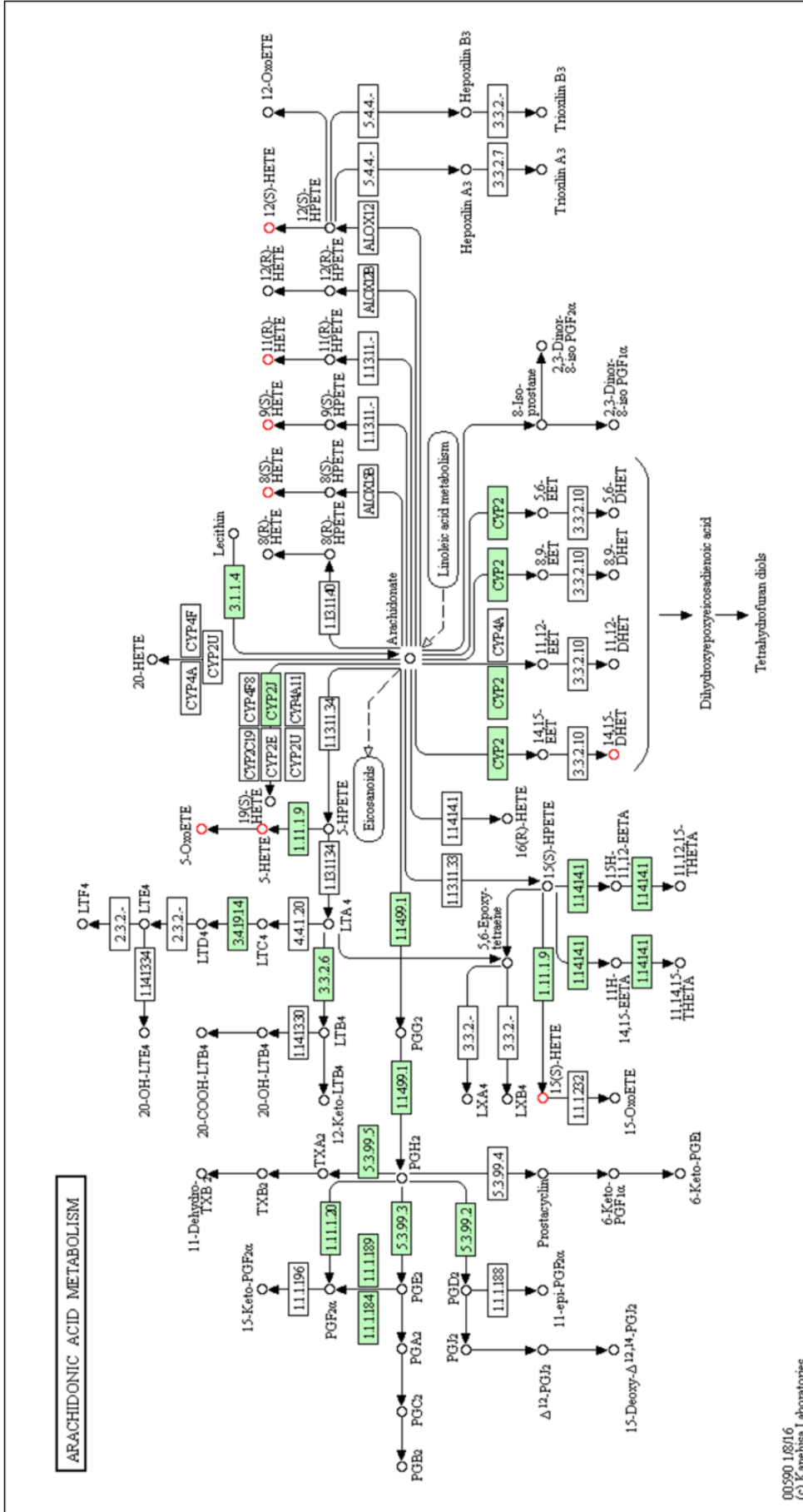


Figure S11. Oxylipins quantified in the *D. magna* samples (colored in red circles) mapped to the KEGG arachidonic acid pathways of *Daphnia pulex*.

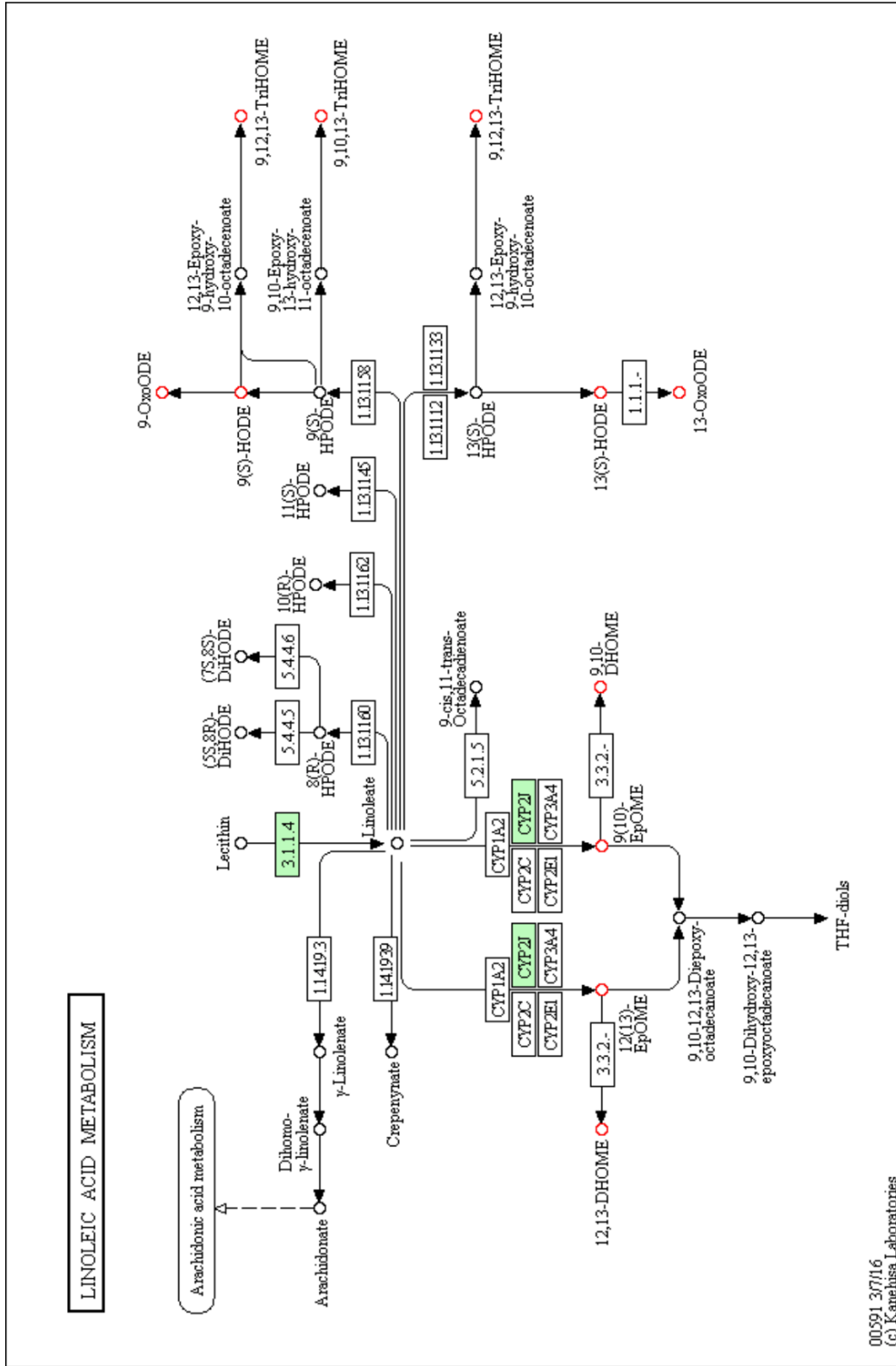


Figure S12. Oxylipins quantified in the *D. magna* samples (colored in red circles) mapped to the KEGG linoleic acid pathways of *Daphnia pulex*.

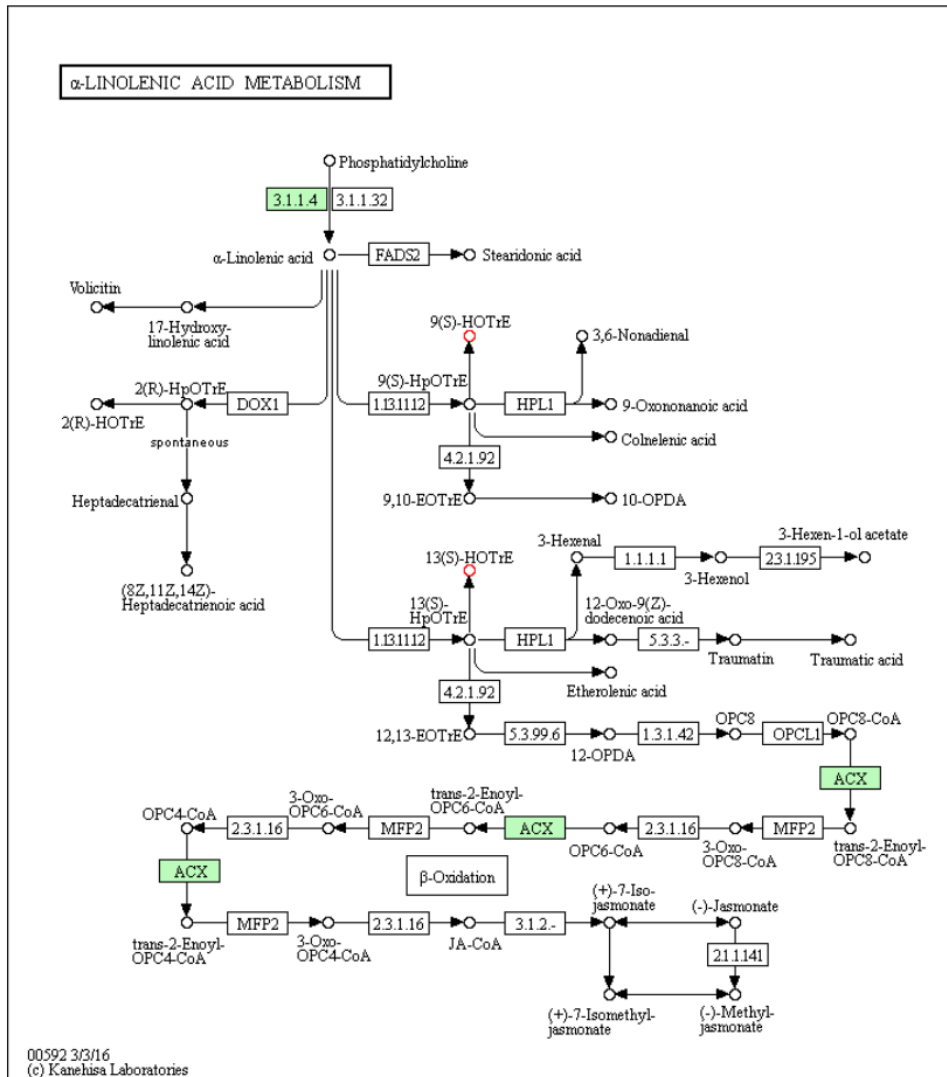


Figure SI3. Oxylipins quantified in the *D. magna* samples (colored in red circles) mapped to the KEGG α -linoleic acid pathways of *Daphnia pulex*.

3.3 ARTICLE CIENTÍFIC VI

Novel combined approach for the target/untarget characterization of Daphnia magna metabolome.

E. Garreta-Lara, B. Campos, C. Barata, R. Tauler, S. Lacorte.

(En preparació)

Novel combined approach for the targeted/untargeted characterization of *Daphnia magna* metabolome

Elba Garreta-Lara, Bruno Campos, Carlos Barata, Romà Tauler, Sílvia Lacorte

Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona,
08034, Catalonia

Corresponding author:

Sílvia Lacorte

E-mail address:

slbqam@cid.csic.es

Telephone:

+34-934006133

ABSTRACT

Targeted and untargeted data analysis strategies provide different alternatives for metabolomic studies. In this study, the freshwater flea *Daphnia magna* was exposed to bisphenol, a well-known endocrine disrupter, to elucidate the major changes of the metabolomic profile and to quantify these changes. In a first step, the extraction conditions were optimized using different number of *daphnids* and different extractions procedures. Second, liquid chromatography-high resolution mass spectrometry (LC-HRMS) data was exploited using (i) targeted analysis for analytical protocol validation and for the quantitative determination of a preselected set of 23 metabolites and (ii) untargeted analysis to fully characterize the metabolites that were de-regulated in exposed *Daphnia* which were then used for the identification of the biomarkers of exposure. This study proposes the combination of these two approaches to enhance the complete quantitative and qualitative metabolomic characterization of *D. magna* by LC-HRMS.

Keywords:

Targeted analysis, Untargeted analysis, Metabolomics, *Daphnia magna*, Liquid chromatography-high resolution mass spectrometry, Bisphenol A.

Abbreviations:

ASTM: American society for testing and materials; BPA: Bisphenol A; CE: Capillary electrophoresis; ESI: Electrospray ionization; F: Slope; FDR: False recovery rate; FT-ICR/MS: Fourier transform ion cyclotron resonance mass spectrometry; FWHM: Full widths at half maximum; GC-MS: Gas chromatography-mass spectrometry; HCD: High energy collision dissociation; HMDB: Human metabolome database; IDL: Instrumental detection limit; IS: Internal standard; KEGG: Kyoto encyclopedia of genes and genomes; LC-HRMS: Liquid chromatography-mass spectrometry; LC-MS: Liquid chromatography-mass spectrometry MCR-ALS: Multivariate curve resolution-alternating least squares; MDL: Method detection limits; METLIN: Metabolite and tandem MS Database; MS: Mass spectrometry; NMR: Nuclear magnetic resonance; PC: Principal component; PCA: Principal component analysis; PIPES: Piperazine-N,N'-bis (2-ethanesulfonic acid); QTRAP: Quadrupole ion trap; R: Recovery; R²: Regression coefficient; ROI: Regions of interest; RSD: Relative standard deviation; SD: Standard deviation; SI: Supplementary information; TQMS: Tandem quadrupole mass spectrometry.

1. Introduction

Metabolomics has emerged as an efficient method for assessing and understanding sub-lethal toxic responses on an organism [1] and it is used as a sensitive indicator of stress. Under stress conditions, either physical or chemical, the organisms physiologically react to overpass or accommodate to new situations, and several metabolites or lipids are de-regulated either increasing or decreasing their concentrations within organism. The ability to identify the biomolecules responsible for these changes represents a new challenge in analytical chemistry and in the field of metabolomics. Current techniques used to characterize the metabolome in several organisms span from RAMAN spectroscopy capable to reveal biochemical alterations in different biological tissues [2], nuclear magnetic resonance (NMR) which can successfully analyze the metabolomic profiles affected by exposure to various toxicants [3] and liquid chromatography coupled with High Resolution Mass Spectrometry (LC-HRMS), which has revolutionized the field of metabolomics as it extends from quantification of a predefined number of metabolites (target analysis) to untargeted analysis of hundreds to thousands of metabolites [4-6].

Targeted approaches select a predefined group of already characterized metabolites and require the optimization of sample preparation and LC-MS or HRMS acquisition method to provide a sensitive and accurate detection for large-scale metabolomics quantification [7]. This strategy has been used for the analysis of endogenous and exogenous metabolites in biofluids, cells and tissues [8], metabolites in plant samples [9], biomolecules in the freshwater crustacean *Gammarus pulex* [10], neurotransmitters and related metabolites in the water flea *Daphnia magna* [11,12].

On the other hand, untargeted metabolomic strategies focus on the simultaneous measurement of as many previously unknown metabolites as possible in a biological system, enabling a more exhaustive picture on the changes of the complete metabolic profile based on the identification of compounds affected by an external stressor. This strategy has been used to identify biomolecules that respond to a stress condition in *Daphnia magna* [13], in rice [14] and in zebrafish embryos [15] or muscle [16].

The main analytical challenges to perform large, reliable and meaningful metabolite profiling: (i) the quantification of the changes in the expression and absolute concentrations of metabolites induced in stressed organisms; (ii) the characterization of the whole composition of the metabolome; (iii) the discovery of the chemical structure of the potential unknown biomarkers and (iv) metabolite unequivocal identification [17].

Most of these gaps can be resolved using efficient extraction methods combined with advanced analytical and data treatment processes.

Fast and accurate statistical and bioinformatics tools have emerged to exploit the large data sets generated using LC-HRMS, containing gigabits of information. New software and chemometric tools have expedited data processing [18-21] and allowed their deeper interpretation [6] together with raw data pre-processing [22]. Finally, in untargeted analysis, metabolite identification is based on reliable mass spectral and metabolite libraries and databases such as NIST Research Library (<https://www.nist.gov/nist-research-library>), METLIN (<http://metlin.scripps.edu>), MassBank (<http://www.massbank.jp>), HMDB (<http://www.hmdb.ca>) and KEGG databases (<http://www.genome.jp/kegg>). In addition, MetaboAnalyst (<http://www.metaboanalyst.ca>) is a comprehensive web application for metabolomics data analysis and interpretation and permits to discriminate metabolites by applying the false discovery rate (FDR) tool.

The freshwater flea *Daphnia magna* is widely used as aquatic toxicity model [23-25], for its relevance in aquatic food webs [26] and that its genome is highly responsive to environmental stresses [27]. This organism has been shown to be affected by exposures of copper [28], pyrene and fluoranthene [29], cadmium [30] and abiotic factors, such as salinity, temperature and hypoxia [13, 31]. Previous studies highlighted that bisphenol A (BPA), a monomer used in the production of polycarbonate and epoxy resins, has endocrine disrupting properties [32], alters the storage of lipids in *D. magna* and increases the concentrations of amino acids with a concomitant decrease of glucose and lactate at sub-lethal concentrations [33,34].

Within this context, the aim of this study was to combine a target and untargeted approach to assess the global changes of *D. magna* metabolome produced by the exposure of the endocrine disruptor BPA. Extraction conditions, LC-HRMS analysis together with advanced chemometric tools were optimized and used for: (i) targeted analysis and quantitative characterization of a set of representative metabolites (sugars, amino acids, nucleotides, nucleosides, neurotransmitters and organic acids) in the *D. magna* metabolome and (ii) untargeted analysis using several chemometric tools to reveal the metabolic profile of *D. magna* in control and BPA exposed conditions. The combined approach of targeted and untargeted HRMS analysis and chemometric tools is intended to provide an accurate, quantitative and selective integrated strategy to be used in metabolomic studies.

2. Materials and methods

2.1 Chemicals and standard solutions

Twenty-three metabolites (including sugars, amino acids, nucleotides, nucleosides, neurotransmitters and organic acids) were purchased from Sigma-Aldrich (Steinheim, Germany) and Fluka (Buchs, Switzerland). These target compounds represent main biomolecules detected in metabolomics studies [35]. **Table S11** displays the main physicochemical properties of the target compounds. Ammonium acetate ($\geq 98.0\%$) was obtained from Sigma-Aldrich (Steinheim, Germany). Ultra-pure water was supplied from Millipore Milli-Q water purification system with a specific resistance of $18.2 \text{ M}\Omega \text{ cm}$ or higher (Millipore, Bedford, MA, USA). Acetonitrile (ACN), methanol (MeOH) and chloroform were supplied by Merck (Darmstadt, Germany). For accurate quantitative analysis, L-methionine sulfone was used as surrogate standard while piperazine-N,N'-bis (2-ethanesulfonic acid) (PIPES) ($\geq 99.0\%$) as Internal Standard (IS) (Sigma-Aldrich, Steinheim, Germany). Stock solutions ($1000 \mu\text{g mL}^{-1}$) and serially diluted solutions of metabolite standards were prepared in MeOH. BPA (CAS 80-05-7) used in the *D. magna* experiments was purchased from Sigma Aldrich (Steinheim, Germany). A standard stock of BPA was prepared at a concentration of 200 mg L^{-1} in acetone.

2.2 *D. magna* growth and BPA exposure conditions

All experiments were performed using the well-characterized clone F of *D. magna* maintained indefinitely as pure parthenogenetic cultures as described in a previous study by Barata and Baird [36]. Photoperiod was set to 16 h light: 8 h dark cycle and temperature at $20 \pm 1 \text{ }^\circ\text{C}$. Bulk cultures of $10 \text{ animals L}^{-1}$ were maintained in American Society for Testing and Materials (ASTM) International hard water at high food conditions ($5 \times 10^5 \text{ cells mL}^{-1}$ of *Chlorella vulgaris*). The culture medium was changed every other day, and neonates were removed within 24 h. In this study, four days old *D. magna* juveniles having a body length of $2.19 \pm 0.29 \text{ mm}$ (mean \pm SD) were used.

BPA is a particular environmental stressor affecting the metabolome of *D. magna* [37]. Eight *daphnids* were exposed to sublethal concentrations of BPA at $4.5 \mu\text{M}$ (low-dose) and at $45 \mu\text{M}$ (high-dose), following a previous study on the effects of BPA in the lipids storage after 24 hours exposition where the EC_{50} was of $99.4 \mu\text{M}$ [33]. *D. magna* juveniles were placed in 300 mL glass vessels filled with 200 mL of test medium (ASTM). Four replicate samples were used per each treatment (control, low- and high-dose). The same concentration of the carrier solvent (acetone, $20 \mu\text{L L}^{-1}$) was added to all

treatments. After 24 hours of exposure, *daphnids* were collected from each jar and transferred into eppendorfs. Finally, water was removed, and samples were snap-frozen in liquid N₂ and stored at – 80 °C.

Two different solid-liquid extraction protocols (A and B) using MeOH, Milli-Q water and chloroform were investigated. Before the extraction and on a basis, L-methionine sulfone was added at a concentration of 100 ng as surrogate standard. In the first extraction protocol (A), metabolites of *D. magna* were extracted with 900 µL of MeOH. After mixing for 15 s in vortex, samples were sonicated for 15 min and vortexed again 15 s. Then, 500 µL of Milli-Q water and 300 µL of chloroform (used to remove lipids) were added. After this, samples were vortexed for 15 s, placed on ice for 10 min and centrifuged at 23500 g for 10 min at 4 °C. Finally, the aqueous fraction was transferred to a new eppendorf, evaporated to dryness under nitrogen gas and reconstituted with 100 µL of 1:1 ACN:water (v/v) containing PIPES (IS) at a concentration of 5 µg mL⁻¹. Extracts were filtered through 0.22 µm filters (Ultrafree[®]-MC, Millipore) and stored at -80 °C until LC-Orbitrap analysis. In the second extraction protocol (B), *D. magna* metabolites were extracted with 900 µL of 90:10 MeOH:water mixture following the same protocol as indicated above excluding the addition of chloroform (thus, without lipid removal).

2.3 LC-Orbitrap MS

An LC-Orbitrap MS (Exactive, Thermo Scientific, San Jose, CA) equipped with heated electrospray ionization (H-ESI) source was used. The system was equipped with an HTC PAL autosampler and a Surveyor MS Plus pump. First, an optimization of the ionization parameters was performed using the direct infusion to set up the most intense ions of the studied metabolites (see the section *Optimization of the ionization parameters* in the **Supplementary Information (SI)**). Chromatographic separation was done on a TSK Gel-Amide 80 HPLC column (250 x 2 mm, 5 µm) from Tosoh Bioscience GmbH (Stuttgart, Germany) following the conditions of Gómez-Canela et al. [10]. Mobile phases were binary mixtures of (A) ACN and (B) 5 mM ammonium acetate (pH = 5) in HPLC-grade water. Gradient elution started at 75% A and 25% B, which was decreased to 70% A in 8 min and to 40% A in 12 min, with a holding for 5 min. Initial conditions were reached in 3 min and the system was stabilized for 10 min. The total run time was 30 min. Flow rate was set at 150 µL min⁻¹ and the injection volume was 5 µL. Metabolites were analyzed under negative electrospray ionization (ESI) mode because better sensitivity was obtained (see **Table SI2**) for the set of preselected metabolites. Spray voltage was set at 3.00 kV, the capillary voltage at 30 V, skimmer voltage at 25 V and tube lens voltage at 75 V. Sheath gas flow rate was set at 45 arbitrary units (au), the auxiliary gas

flow rate at 10 $\mu\text{L}/\text{min}$ and the capillary temperature at 300 $^{\circ}\text{C}$. Data acquisition conditions were optimized to enhance the compound identification and acquisition over a mass range of 100-700 m/z was performed at 50,000 full widths at half maximum (FWHM) using an accurate mass error below 5 ppm. High energy collision dissociation (HCD) voltage was set at 25 eV to maximize structural information for the selected metabolites and enhance the identification potential. Identification criteria followed the European Directive 2002/657/EC [38] where: (i) precursor ions were acquired at a resolving power of 50,000 FWHM with four decimals, restricting the number of candidates; (ii) mass accuracy error < 5 ppm of both precursor and product ions; (iii) retention time shift < 2% between the standards and the samples and (iv) fragmentation patterns recorded at a HCD voltage of 25 eV to confirm the mass spectral signature.

3. Data analysis strategies and identification criteria

LC-HRMS datasets of Daphnids exposed to BPA were processed using a target approach with the Thermo Xcalibur 2.1 (Thermo Scientific, San Jose, CA) software to quantify the 23 target metabolites. Raw LC-HRMS datasets were also processed using advanced chemometric tools with the ROIMCR procedure [6, 19, 39] which combines the regions of Interest (ROI) methodology [19] and the Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) method [40] to identify metabolites affected by BPA exposure. Finally, PLS Toolbox 8.2 (Eigenvector Research Inc., Wenatchee, WA, USA) [41] under MATLAB 2015 (The Mathworks Inc. Natick, MA, USA) computer environment was used for Principal Component Analysis (PCA) in target and untargeted analysis. **Figure 1** shows the different steps required for target and untargeted strategies.

3.1. Targeted analysis

Target analysis is aimed to set up the analytical conditions to monitor the levels of a preselected number of biomolecules with certainty at very low concentration. In this study, 23 metabolites were preselected and represent characteristic metabolites of biological systems. Quality control analysis were undertaken to obtain a good precision of the method, sensitivity and extraction efficiency. Internal standard calibration using L-methionine sulfone was performed over the concentration range from 0.01 to 7 $\text{ng } \mu\text{L}^{-1}$ using eight calibration points. Intra-day and inter-day precision were assessed at 5 $\text{ng } \mu\text{L}^{-1}$, using a standard solution injected on the same day and on 3 different days, respectively. Solvent blanks did not contain any of the investigated metabolites, indicating no carryover effect in any of the LC-HRMS runs. Extraction efficiency was

performed with sets of 2, 3, 5, 8 and 10 pools of *daphnids* spiked with 100 ng of the metabolite standards and the method was optimized to obtain high extraction yields using the optimum number of *daphnids*. For each pool, five replicates of spiked *D. magna* samples and two replicates of non-spiked *D. magna* samples were extracted to subtract the contribution of their own metabolites. Finally, instrumental detection limit (IDL) was calculated as the concentration that gives a signal intensity of 10^3 and method detection limits (MDL) were calculated considering the concentration that gives a signal intensity of 10^3 using the spiked *D. magna* samples, with pools of 8 individuals, spiked at 100 ng. Quantitative analysis permitted to determine the extent of metabolite recovery considering the matrix effect and provided an initial step to be used in the untargeted analysis.

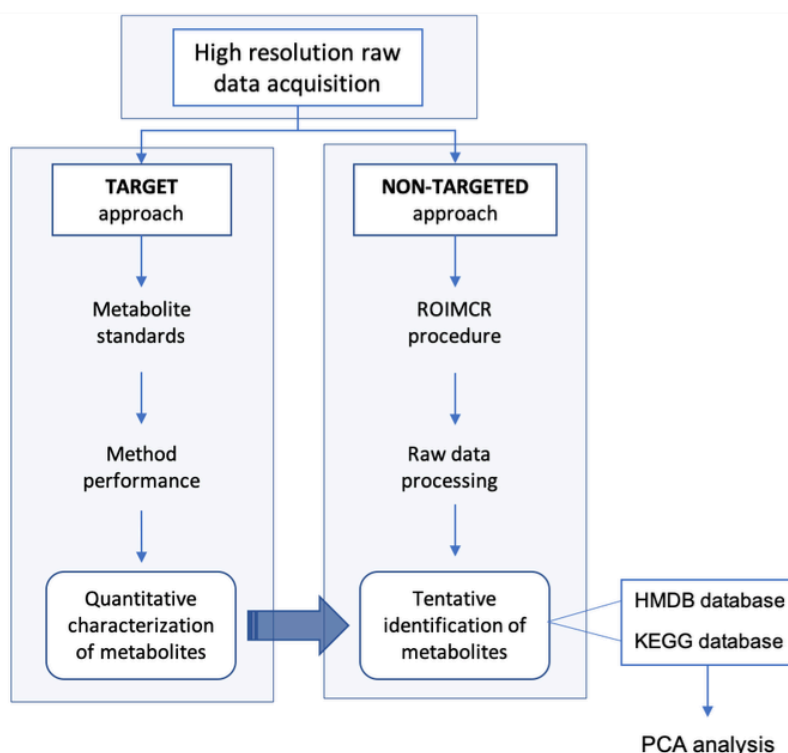


Figure 1. Flowchart listing the steps involved in the target and non-target approaches.

3.2 Untargeted analysis

As shown in **Figure 1**, untargeted analysis requires different data processing steps of the complex experimental data sets acquired by LC-HRMS. Each LC-HRMS chromatographic run recorded for every sample resulted in a raw data file, which was converted to NetCDF format through Thermo Xcalibur 2.1 (Thermo Scientific, San Jose, CA) software and further imported into MATLAB (The Mathworks, Inc, US) computational environment using its Bioinformatics toolbox. A reduced size data matrix containing only

relevant LC-HRMS signals was constructed using the ROIMCR procedure recently developed which on its first step searches the regions of interest (ROI) [19]. Briefly, ROIs are intended to compress the MS data filtering the noise from the mass traces with significant MS intensity values and are set according to an optimized signal-to-noise ratio threshold. References [6, 14, 19, 39, 42] describe a more detailed explanation of the ROIMCR method. Different ROI parameters need to be selected in order to apply correctly the ROIMCR methodology [19]. The first one is the ROI threshold value which was first optimized between 0.1 and 1% of the maximum measured MS peak intensity to filter only significant MS intensities. The second parameter is the m/z error or acceptable mass accuracy, which according to the known instrument spectral resolution was varied from 0.005, 0.010, 0.025 and 0.05 Da/e. Finally, the third parameter to optimize is the number of consecutive high-density data values to be present in the same ROI defining a chromatographic elution peak, which was tested at 10, 20 and 25. From the simultaneous optimization of these three parameters, different data matrices were built. Once these parameters were optimized, a ROI column-wise augmented data matrix was built to compare control and exposed samples.

In the second step of the ROIMCR procedure is the analysis ROI column-wise augmented data matrix was analyzed by the MCR-ALS method [43]. The quality of MCR-ALS models was evaluated according to the lack of fit, which is the difference among the input data and the data reproduced from the product obtained by MCR-ALS and the percent of explained variance (R^2). More details on how this method is implemented can be found elsewhere [40]. Due to the high resolution of MS data, a large number of components can be proposed to describe the variance observed in the experimental data. Each MCR-ALS component is defined by its chromatographic elution profile and by its HRMS spectrum. From these spectra, MCR-ALS resolved components are assigned to different metabolites (see below). Resolution of the elution profiles of the different components provided also the peak areas of the metabolites for each of the different analyzed samples, as well as their associated mass spectra. From the changes in these peak areas, relative concentrations of the metabolites in the different samples (control and treated) can be retrieved.

Metabolites identification represents an analytical challenge in untargeted analysis due to the lack of previous knowledge about many of the metabolites present in the samples. Hence, their identification was based on their accurate mass analysis and fragmentation patterns. Identification of the relevant MCR-ALS resolved components was performed online from the mass spectra profiles using METLIN [44] and Human Metabolome Databases (HMDB) [45] as described in the *Mass Spectral Characterization*

section in the SI of this work. Kyoto Encyclopedia of Genes and Genomes (KEGG) [46] has been used to achieve a reliable identification of the metabolites present in the investigated biological system, *D. magna* (**Figure 1**).

4. Results and discussion

In this study, the analytical conditions were optimized and HRMS datasets were processed using both target and untargeted strategies to identify a large number of compounds and achieve the quantitative characterization of the changes in the *D. magna* metabolome exposed to BPA. Target analysis ensured optimal analytical performance with regards to extraction conditions and method efficiency enabling low level determination of biomolecules. The untargeted approach provided a wide visualization of the whole metabolic profiles and changes observed in exposed *D. magna*.

4.1 Method performance

To obtain reliable efficiency of the analytical protocol and to guarantee the optimal feasible culture conditions, extraction conditions and HRMS acquisition were optimized in a quantitative way. This permitted visualization of the changes in the metabolome of exposed and non-exposed organisms could be visualized. In a first step, the identification criterion was done using the conventional method of mass spectral characterization (see **Table 1**) where the precursor ion $[M-H]^-$ for each target metabolite and one fragment ion were used as quantification ions and confirmation, respectively, at the optimized ionization conditions. **Figure 2** shows the extracted ion chromatograms of the selected metabolites at a concentration of $5 \text{ ng } \mu\text{L}^{-1}$. Twenty-three metabolites including sugars, amino acids, nucleotides, nucleosides, neurotransmitters and organic acids were resolved using a TSK Gel-Amide 80 HILIC column. Maltose and trehalose coeluted at a m/z 341.108 $[M-H]^-$ (**Figure 2**) but were completely resolved at specific fragment ions (**Table 1**). On the other hand, galactose, glucose and myo-inositol are structural isomers with the molecular ion at $m/z = 179.0561 [M-H]^-$ but different fragmentation pattern and were chromatographically resolved (**Figure 2**) (**Table 1**). The LC-Orbitrap MS method proved to be highly sensitive and linear over a mass range of 0.01 to $7 \text{ ng } \mu\text{L}^{-1}$ ($R^2 \geq 0.99$), except for sugar compounds, L-citrulline, ADP, L-aspartic acid, L-methionine and pimelic acid (**Table 2**). Intra- and inter-day precision ranges were in the range of 0.4 to 16% and between 1.4 and 20%, respectively, at a concentration level of $5 \text{ ng } \mu\text{L}^{-1}$.

In a second step, extraction conditions were optimized to ensure quantitative trace analysis of metabolites. Table SI3 shows the recoveries of the target compounds

in spiked *D. magna* samples using the two extraction protocols. The use of chloroform in method A permitted the separation of the organic phase from the aqueous phase, obtaining cleaner extracts with less interferences. With method A between 20 and 23 metabolites were recovered being some metabolites lost using 2, 3 and 5 *daphnids* probably because of the small mass of sample and difficulties in the extraction process. The recoveries with 8 and 10 individuals was similar but we chose pools of 8 as the optimum conditions as it requires less cultivation effort. Recoveries were acceptable except for glucose, L-aspartic acid, L-methionine, L-serine, L-valine, and fumaric acid which were <50% (**Table 2**). Despite these low recoveries, these compounds were quantified in the BPA exposed experiments since recovery losses were corrected using the surrogate standard. On the other hand, with method B without the use of chloroform not all metabolites were recovered with the sets of 2, 3, 5, 8 and 10 *daphnids* which was attributed to the presence of interferences as lipids were not removed. Finally, method A with a pool of 8 *daphnids* was chosen for BPA exposure studies for its better efficiency and capability to applied for the target and untargeted approach.

Method detection limits ranged from 3.24 to 109 ng g⁻¹ which ensures the detection of these compounds according to the levels detected in biological systems [10]. In addition, the low RSD among replicates warrants that the extraction and analytical procedure is reproducible and robust.

4.2 Metabolic response of *D. magna* exposed to BPA

4.2.1 Target analysis

HRMS database were first processed using a target approach with the Xcalibur software. Metabolite concentrations in *D. magna* were evaluated for control and 24 hours low- and high-dose BPA exposed samples. All target metabolites were detected using 8 *daphnids* and this is in line with the presence of these main metabolites in biological systems. The concentration in whole *Daphnia* pools ranged from 0.20 ± 0.01 (hypoxanthine) to 105 ± 2.77 ng daphnia⁻¹ (L-Tyrosine) (**Table S15**). The levels detected are in line to those reported in [10] and [12]. The method proved to be highly sensitive and robust considering that the concentration of the different metabolites can be low. The general low SD proves the precision of the method. The target analysis permitted also to distinguish differences in metabolite concentrations in control, low and high does exposure experiment, which are further explored in next sections.

Table 1. Mass spectral characterization of the target compounds. *n.d.*: not detected.

Target compounds	Molecular formula	Exact mass	Precursor ion			Fragmentation ion			
			[M-H] ⁻	[M-H] ⁻ experimental	Error <i>m/z</i> (ppm)	Molecular formula	Theoretical	Experimental	Error <i>m/z</i> (ppm)
Maltose	C ₁₂ H ₂₂ O ₁₁	342.1162	341.1089	341.1087	0.55	[C ₈ H ₁₃ O ₇] ⁻	221.0666	221.0664*	1.03
Trehalose	C ₁₂ H ₂₂ O ₁₁	342.1162	341.1089	341.1088	0.25	[C ₅ H ₅ O ₃] ⁻	113.0244	113.0241*	2.38
Myo-Inositol	C ₆ H ₁₂ O ₆	180.0634	179.0561	179.0560*	0.35	[C ₆ H ₅ O ₄] ⁻	141.0193	141.0170	16.19
Galactose	C ₆ H ₁₂ O ₆	180.0634	179.0561	179.0559	0.91	[C ₅ H ₅ O ₃] ⁻	113.0244	113.0242*	1.49
Glucose	C ₆ H ₁₂ O ₆	180.0634	179.0561	179.0558	1.47	[C ₄ H ₅ O ₃] ⁻	101.0244	101.0243*	0.68
L-Aspartic acid	C ₄ H ₇ NO ₄	133.0375	132.0302	132.0298*	2.90	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>
L-Citrulline	C ₆ H ₁₃ N ₃ O ₃	175.0957	174.0884	174.0883*	0.38	[C ₅ H ₁₁ O ₂ N ₂] ⁻	131.0826	131.0824	1.16
L-Phenylalanine	C ₉ H ₁₁ NO ₂	165.0790	164.0717	164.0717*	-0.28	[C ₉ H ₇ O ₂] ⁻	147.0451	147.0449	1.39
L-Methionine	C ₅ H ₁₁ NO ₂ S	149.0510	148.0437	148.0437*	0.16	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>
L-Tryptophan	C ₁₁ H ₁₂ N ₂ O ₂	204.0899	203.0826	203.0824*	0.75	[C ₈ H ₆ N] ⁻	116.0505	116.0505	0.21
L-Tyrosine	C ₉ H ₁₁ NO ₃	181.0739	180.0666	180.0664*	0.93	[C ₈ H ₇ O] ⁻	119.0502	119.0500	1.59
L-Serine	C ₃ H ₇ NO ₃	105.0426	104.0353	104.0351*	1.61	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>
L-Valine	C ₅ H ₁₁ NO ₂	117.0790	116.0717	116.0716*	0.46	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>
ADP	C ₁₀ H ₁₅ N ₅ O ₁₀ P ₂	427.0294	426.0221	426.0215*	1.38	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>
NADH	C ₂₁ H ₂₉ N ₇ O ₁₄ P ₂	665.1248	662.1018	662.1021*	-0.45	[C ₁₅ H ₂₀ O ₁₃ N ₅ P ₂] ⁻	540.0538	540.0536	0.34
Cytidine	C ₉ H ₁₃ N ₃ O ₅	243.0855	242.0782	242.0780*	0.81	[C ₇ H ₆ O ₂ N ₂] ⁻	150.0434	150.0421	8.85
Inosine	C ₁₀ H ₁₂ N ₄ O ₅	268.0808	267.0734	267.0732*	0.91	[C ₅ H ₃ ON ₄] ⁻	135.0312	135.0312	-0.11
Uridine	C ₉ H ₁₂ N ₂ O ₆	244.0695	243.0622	243.0623*	-0.37	[C ₅ H ₄ O ₂ N] ⁻	110.0247	110.0247	0.03
Adenosine	C ₁₀ H ₁₃ N ₅ O ₄	267.0968	266.0894	266.0895*	-0.27	[C ₅ H ₄ N ₅] ⁻	134.0472	134.0471	0.52
Guanosine	C ₁₀ H ₁₃ N ₅ O ₅	283.0917	282.0843	282.0846*	-0.91	[C ₅ H ₄ ON ₅] ⁻	150.0421	150.0420	0.56
Hypoxanthine	C ₅ H ₄ N ₄ O	136.0385	135.0312	135.0312*	-0.11	[C ₄ H ₂ N ₃] ⁻	92.0254	92.0253	0.78
Fumaric acid	C ₄ H ₄ O ₄	116.0110	115.0036	115.0037*	-0.58	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>
Pimelic acid	C ₇ H ₁₂ O ₄	160.0736	159.0662	159.0660*	1.47	[C ₆ H ₁₁ O ₂] ⁻	115.0764	115.0761	2.65

*Quantification ions.

Table 2. Instrumental method performance of the selected target compounds (ordered by metabolic group) by LC-HRMS. F: slope; R²: regression coefficient; IDL: instrumental detection limit; MDL: method detection limits in fresh weight *daphnids*; %R ± RSD: % recovery ± relative standard deviation.

Chemical families	Target compounds	Retention time (min)	Linearity (ng µL ⁻¹)	F	R ²	IDL (ng)	MDL (ng g ⁻¹)	Intra-day precision (5 ng µL ⁻¹)	Inter-day precision (5 ng µL ⁻¹)	% R ± RSD
Sugars	Maltose	14.95	0.05-7	0.046	0.9942	0.050	105.05	3.6	9.4	89 ± 15
	Trehalose	14.95	0.05-7	0.342	0.9989	0.018	80.34	1.5	4.1	94 ± 6
	Myo-Inositol	15.27	0.05-7	0.244	0.9983	0.018	73.92	0.9	1.4	79 ± 4
	Galactose	12.77	0.05-7	0.008	0.9932	0.785	66.10	3.0	3.9	57 ± 8
	Glucose	11.89	0.05-7	0.082	0.9911	0.166	69.95	2.9	6.4	33 ± 9
Amino acids	L-Aspartic acid	15.29	0.05-7	1.187	0.9950	0.087	109.92	7.8	14.6	43 ± 9
	L-Citrulline	15.58	0.01-7	0.472	0.9897	0.014	65.63	1.7	4.9	51 ± 5
	L-Phenylalanine	9.61	0.01-7	2.21	0.9980	0.009	8.00	2.3	10.7	62 ± 15
	L-Methionine	11.89	0.05-7	0.559	0.9940	0.048	44.73	6.3	15	25 ± 13
	L-Tryptophan	8.32	0.01-7	2.368	0.9985	0.005	13.50	4.1	7.7	55 ± 14
	L-Tyrosine	11.06	0.01-7	1.354	0.9981	0.009	11.43	2.1	9.0	75 ± 16
	L-Serine	15.72	0.01-7	0.140	0.9939	0.106	42.10	1.0	5.2	24 ± 9
	L-Valine	12.79	0.01-7	0.901	0.9945	0.022	17.80	5.6	7.9	41 ± 11
Nucleotides	ADP	15.20	1-7	0.026	0.9882	1.027	43.06	16	20	111 ± 2
	NADH	13.46	0.01-7	0.152	0.9949	0.074	26.58	1.9	4.8	94 ± 4
Nucleosides	Cytidine	8.69	0.01-7	0.832	0.9993	0.008	28.07	1.8	3.6	82 ± 7
	Inosine	7.67	0.01-7	1.682	0.9990	0.004	3.24	1.9	10	50 ± 12
	Uridine	6.76	0.01-7	1.257	0.9969	0.005	6.44	4.3	9.4	79 ± 12
	Adenosine	6.66	0.01-7	0.229	0.9977	0.032	12.91	1.9	2.1	89 ± 11
	Guanosine	8.83	0.01-7	1.092	0.9988	0.005	18.22	2.3	4.5	68 ± 11
Neurotransmitters	Hypoxanthine	6.61	0.01-7	1.445	0.9951	0.003	16.29	1.5	6.8	71 ± 5
Organic acids	Fumaric acid	5.72	0.01-5	0.218	0.9901	0.059	37.50	3.3	5.8	25 ± 2
	Pimelic acid	4.85	0.05-7	1.278	0.9985	0.006	10.66	0.4	3.9	67 ± 3

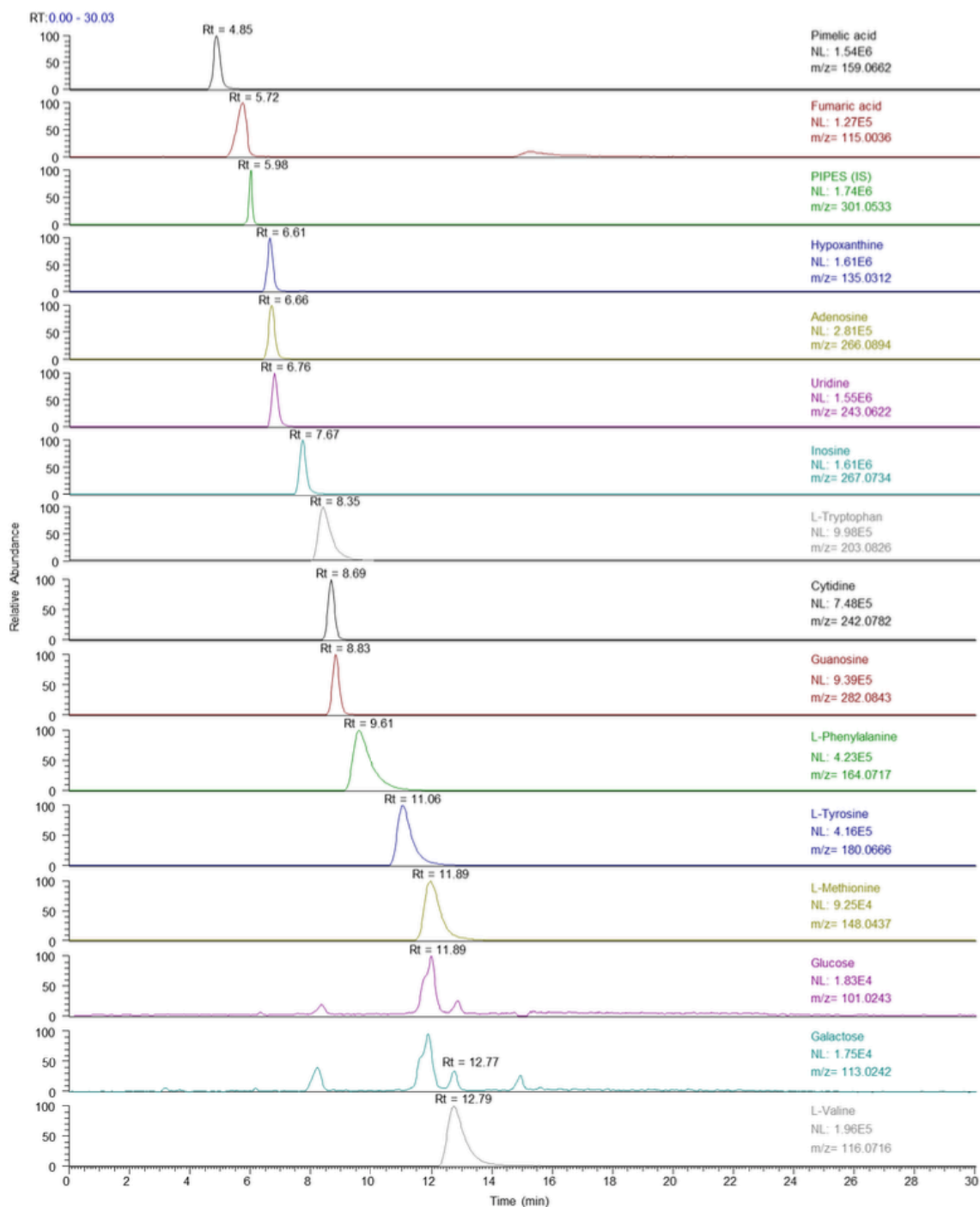


Figure 2. LC-HRMS extracted ion chromatogram of the 23 target compounds, the surrogate and the internal standard (L-methionine sulfone and PIPES) from a full scan LC-HRMS spectrum.

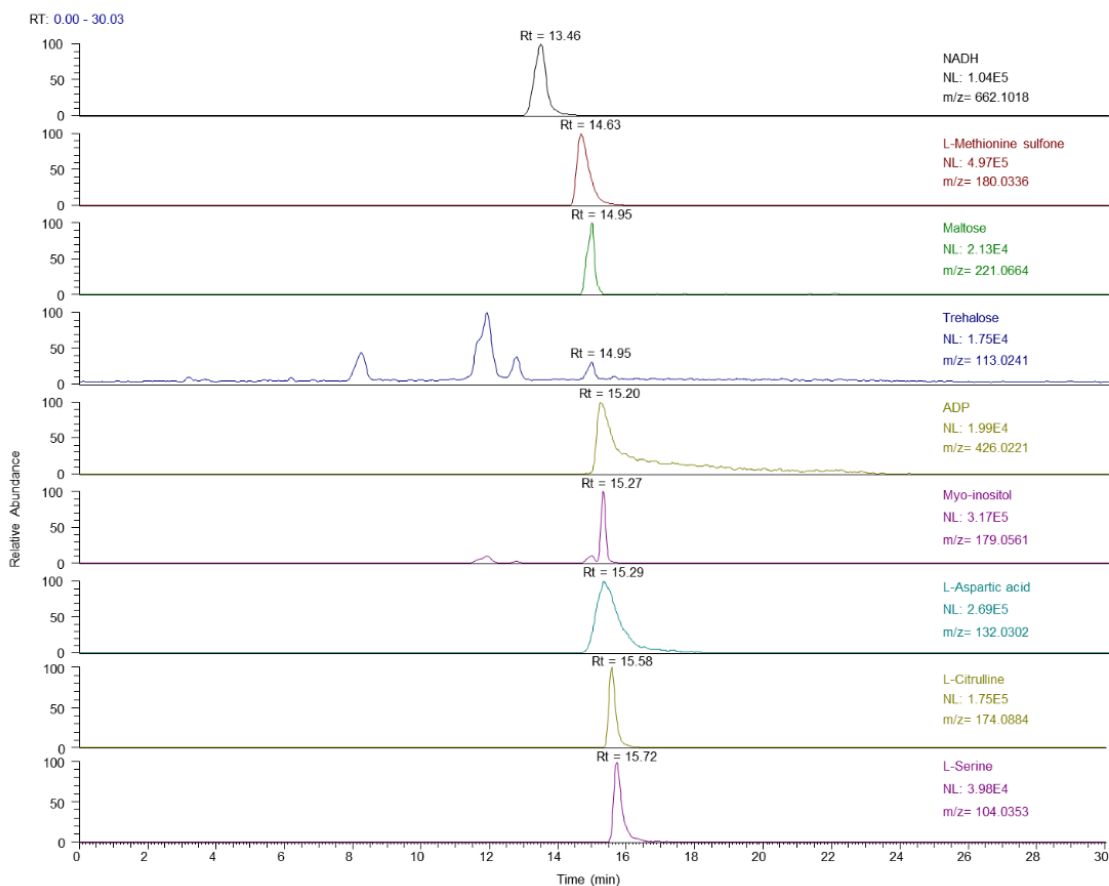


Figure 2. (continued)

4.2.2 Untargeted analysis

As described in the method section, untargeted HRMS data were analyzed using the ROIMCR procedure [42] to fully characterize *D. magna* metabolome. Among the different ROI conditions tested, to filter significant MS intensities the threshold value of < 0.1% of the maximum MS intensity was selected. In the definition of every ROI the mass accuracy selected was 0.025 Da/e and the minimum number of 20 consecutive high-density data points were considered. Only those ROIs with a reasonable chromatographic peak shape and having multiple individual measurement along the same peak were finally considered for the subsequent MCR-ALS analysis [19]. MCR-ALS was then applied to the ROIs augmented data matrix obtained from control and exposed samples at a low- and a high-level dose of BPA (see Section 3.2). **Table S14** of the SI provides the information indicating the number of MCR-ALS components tested and the results obtained according to R^2 and lack of fit of the different models. A total number of 97 components were finally resolved by MCR-ALS, explaining 99.6% of the total data variance. As mentioned in section 3.3, the identification of the MCR-ALS

components was performed using HMDB and KEGG online databases allowing a putative identification of most of the mass signals obtained. Thus, of these 97 resolved components 42 were correctly identified as *D. magna* metabolites, which are reported in **Table 3**, including sugars, amino acids, nucleotides, nucleosides, nucleobases, organic acids, oligosaccharides, fatty acids, flavonoids, dipeptides and several organic compounds.

Because ROI values have all the measured MS data with an intensity higher than a fixed signal-to-noise ratio threshold [6], untargeted strategy resolved a large number of metabolites in comparison to target analysis. Those metabolites with a relevant signal in target analysis were also detected in the untargeted approach, such as L-citrulline, L-phenylalanine, L-tryptophan, L-tyrosine, L-valine, uridine, inosine and ADP (**Table 3**). However, some compounds with intensity signals lower than the ROI signal threshold value were not detected by the ROIMCR procedure and these are maltose, trehalose, myo-inositol, galactose, glucose, L-aspartic acid, L-methionine, L-serine, NADH, cytidine, adenosine, guanosine, hypoxanthine, fumaric and pimelic acids. In this case, the ROIMCR analysis filtered the MS signal at a specific threshold value which did not allow the detection of some of the compounds for which the signal or sensitivity was too low. The signal threshold values can be lowered using a more accurate ROI selection to distinguish metabolites from the experimental noise of the LC-HRMS system and from the solvent and sample matrix contribution. Therefore, the parameters selected in the untargeted approach is a compromise between sensitivity and number of detected metabolites.

4.3 PCA analysis to elucidate BPA effects on *D. magna*

The exploration of the information provided by PCA of metabolites is shown in **Figure 3**. First, PCA results obtained using the autoscaled metabolite concentrations obtained by the targeted approach for control, low-dose and high-dose BPA exposed samples are shown in **Figure 3a**. PC1 and PC2 explained 26.7 % and 20.0 % of the total data variance, respectively. PCA was also applied to the autoscaled peak areas obtained by the ROIMCR untargeted approach (see *Section 3.2* and **Table S16**) where 42 metabolites identified (6 in common to the target analysis) (**Figure 3b**). PC1 and PC2 explained 26.9 % and the 23.0 % of the total data variance, similar to the target approach. Likewise, two sample clusters along PC1 were distinguished in both PCA score plots for targeted and untargeted analysis. Control and low-dose exposed samples appeared overlapped in the left side of the plot, whereas, samples related to the high-dose exposure of BPA were observed separately in the right side of the plot. Sample behavior

was similar in both cases, even though the metabolites studied in both PCA analysis were not all coincident.

Table 3. Tentative identification of *D. magna* metabolites resolved by ROIMCR procedure [6, 18] using measured m/z values, database (KEGG or HMDB) identification, exact mas and mass error in ppm

	Measured m/z	Metabolite	Database ID	Ion assignment	Exact mass	Mass error (ppm)
1	111.0193	Uracil	C00106	[M-H] ⁻	111.0200	0.0126
2	114.0561	L-Proline	C00148	[M-H] ⁻	114.0561	0.4165
3	115.0764	Caproic acid	C01585	[M-H] ⁻	115.0765	0.4658
4	116.0718	L-Valine ξ	C00183	[M-H] ⁻	116.0717	0.9391
5	118.0510	L-Threonine	C00188	[M-H] ⁻	118.0510	0.2779
6	128.0353	Glutamic acid	C00025	[M-H ₂ O-H] ⁻	128.0348	4.1567
7	130.0873	Leucine/Isoleucine	C00407	[M-H] ⁻	130.0874	0.4051
8	134.0472	Adenine	C00147	[M-H] ⁻	134.0472	0.1432
9	143.1077	Octanoic acid	C06423	[M-H] ⁻	143.1078	0.3759
10	145.0618	Glutamine	C00064	[M-H] ⁻	145.0619	0.4564
11	157.1233	Nonanoic acid	C01601	[M-H] ⁻	157.1234	0.6606
12	164.0387	DL-Methionine S-oxide	C02989	[M-H] ⁻	164.0387	0.0738
13	164.0717	L-Phenylalanine ξ	C00079	[M-H] ⁻	164.0717	0.0158
14	171.1390	Capric acid	C01571	[M-H] ⁻	171.1391	0.3150
15	180.0667	L-Tyrosine ξ	C00082	[M-H] ⁻	180.0666	0.4598
16	201.0275	Nicotinuric acid	C05380	[M+Na-H] ⁻	201.0282	3.2737
17	203.0810	L-Tryptophan ξ	C00078	[M-H] ⁻	203.0826	4.8865
18	218.1036	2-(2-methylbutanamido) acetic acid	HMDB00339	[M+HAc-H] ⁻	218.1034	0.9431
19	224.1044	Prolyl-Glutamine	HMDB29015	[M-H ₂ O-H] ⁻	224.1035	3.9446
20	227.2017	Myristic acid	C06424	[M-H] ⁻	227.2017	0.2020
21	234.1078	Citrulline ξ	C00327	[M+HAc-H] ⁻	234.1095	4.4423
22	243.0624	Uridine ξ	C00299	[M-H] ⁻	243.0623	0.5756
23	255.2332	Palmitic acid	C00249	[M-H] ⁻	255.2330	0.9627
24	267.0739	Inosine ξ	C00294	[M-H] ⁻	267.0735	1.5221
25	273.0380	Deoxyribose 5-phosphate	C00673	[M+HAc-H] ⁻	273.0381	0.3307
26	274.1047	L-gamma-Glutamylcysteine	C00669	[M-H] ⁻	274.1045	0.8781
27	281.2486	Oleic acid	C00712	[M-H] ⁻	281.2486	0.0153
28	284.0683	L-Asparaginyl-L-methionine	HMDB28737	[M+Na-2H] ⁻	284.0686	1.2064
29	302.5336	UDP-N-acetylglucosamine	C00043	[M-2H] ⁻	302.5335	0.3016
30	346.0561	Adenosine monophosphate	C00020	[M-H] ⁻	346.0558	0.8429
31	362.0510	Guanosine monophosphate	C00144	[M-H] ⁻	362.0507	0.7651
32	425.0808	Cysteine-glutathione disulfide	HMDB00656	[M-H] ⁻	425.0806	0.4020
33	426.0227	ADP ξ	C00008	[M-H] ⁻	426.0221	1.3175
34	426.9683	Phosphohydroxypyruvic acid	C03232	[2M+HAc-H] ⁻	426.9684	0.3021
35	467.1287	L-Hydroxypropyl-L-Cysteine	HMDB28860	[2M-H] ⁻	467.1276	2.3993
36	503.1623	Maltotriose	C01835	[M-H] ⁻	503.1618	1.0752
37	539.1395	3'-O-Methyl(-)-epicatechin 7-O-glucuronide	HMDB41659	[M+HAc-H] ⁻	539.1406	2.0907
38	558.0652	Adenosine diphosphate ribose	C00301	[M-H] ⁻	558.0644	1.4380
39	606.0748	UDP-N-acetylgalactosamine	C00203	[M-H] ⁻	606.0743	0.8357
40	611.1458	Oxidized glutathione	C00127	[M-H] ⁻	611.1447	1.8230
41	628.0570	UDP-N-acetyl-D-mannosamine	C01170	[M+Na-2H] ⁻	628.0562	1.2172
42	665.2159	Glycogen	C00182	[M-H] ⁻	665.2146	1.9807

Metabolites also present in the target approach.

PC1 loading plots showed that metabolites with high positive loadings (represented in dark orange color) have larger concentrations in high-dose exposed

samples, whereas, metabolites depicted in light orange (negative loadings in PC1) have higher concentrations in control and low-dose exposed samples. Moreover, metabolites identified using both targeted and untargeted approaches (marked in **Figure 3**), showed the same tendency in PC1 loading plots for both strategies. Hence, L-citrulline, L-phenylalanine, L-tryptophan, L-tyrosine, L-valine, uridine, inosine and ADP will have higher concentrations in high-dose exposed samples. Thus, metabolomic information provided by both strategies was concordant.

Student t-test compared the metabolite changes for target (23 compounds) and untargeted analyses (42 compounds) using concentration and peak areas, respectively, between control and BPA exposed samples. In **Figure 4**, fold changes and Standard Deviation (SD) of all detected metabolites in the target/untargeted approach were represented. A total number of 28 metabolites changed significantly ($*p<0.05$, $**p<0.1$ and $***p<0.005$) across BPA treatments (see underlined metabolites in **Figure 4**). Results show that 7 amino acids were up-regulated and 3 were down-regulated; 3 fatty acids were down-regulated, 8 sugars and 4 nucleosides were up-regulated and NADH was down-regulated.

4.4 Biological interpretation of BPA exposure

Enriched analysis was performed with the 28 statistically significant de-regulated metabolites using MetaboAnalyst platform analysis [47]. It identified up to 20 significant ($p<0.05$) enriched metabolic pathways containing at least 3 metabolites as Hits (**Figure 4**), from which only 7 were still significant after the application of FDR. These most enriched metabolic pathways were related with carbon metabolism (amino acid, urea, ammonia) and with glutamate playing a central role in amino acid metabolism. Basically, it acts as a collector of the amino group of the amino acids [48]. On the other hand, glutamate is involved on the formation of gamma-aminobutyric acid (GABA) [48]. Whereas glutamate concentrations decreased upon exposure to BPA, concentrations of most of the amino acids increased.

According to Nagato *et al.* [37], increased levels of amino acids in *D. magna* individuals exposed to BPA can be interpreted as a reduced protein synthesis under stressing conditions to sustain cellular energy repositories. These arguments are also consistent with the observed increased concentrations of sugars and of other key nucleosides/nucleotides (AMP, NADH) involved in many cellular metabolic processes including energy metabolism. Thus, animals exposed to BPA would accumulate amino acids, sugars and other fuel cell components such as lipids, and would have reduced growth and reproduction.

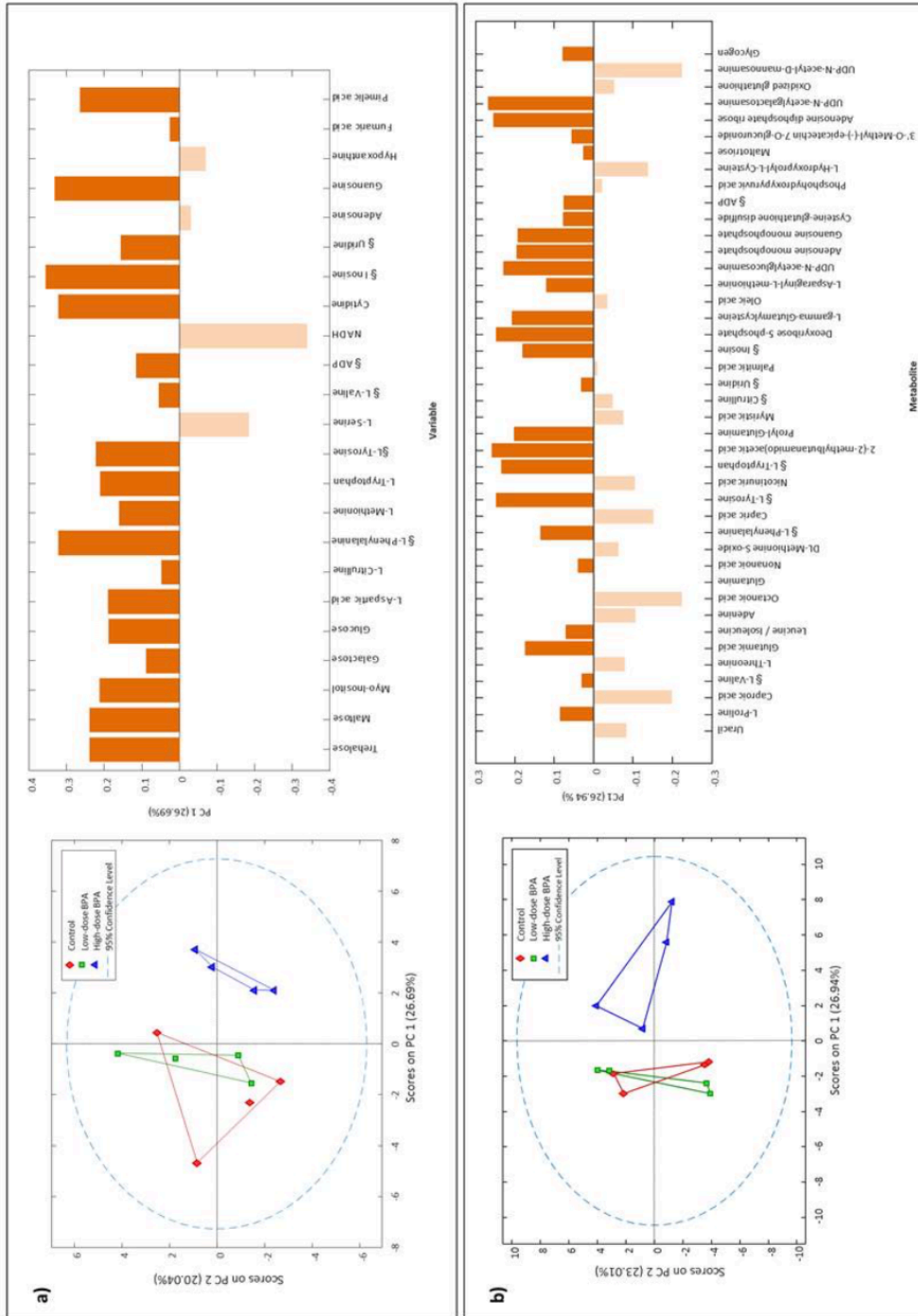


Figure 3. PCA analysis of *D. magna* samples exposed to BPA. **a)** PCA model of the metabolite concentration obtained with the target approach. **b)** PCA model of the relative peak areas provided by the untargeted approach. Control, low- and high-dose exposed samples are represented in red, green and blue colors, respectively. Positive loadings are displayed in dark orange and negative loadings in light orange. *§Metabolites detected in the target and non-target analysis.*

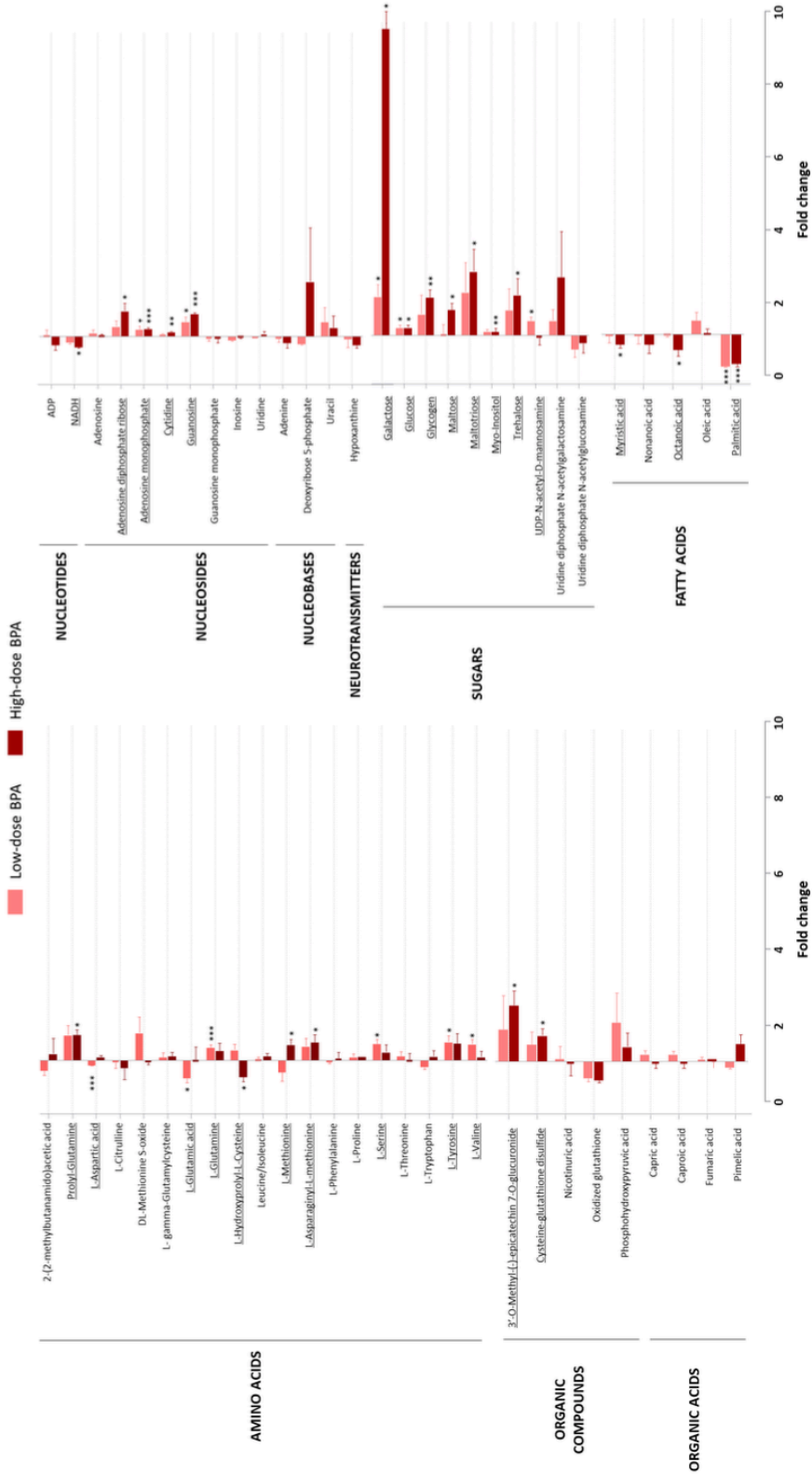


Figure 4. Fold changes and the relative standard deviation of the BPA exposition respect to controls. All the metabolites detected in the target/untargeted approach were listed and grouped by families. Identified significant changes relative to controls ($p < 0.05$, $**p < 0.01$, $***p < 0.005$) following Student's *t*-test. De-regulated metabolites appear underlined.

Table 4. Enriched metabolic pathways and associated metabolites following MetaboAnalyst platform analysis. Red and green metabolites mean up and down regulated ones. *Abbreviations:* ADPR: adenosine diphosphate ribose; AMP: adenosine monophosphate; ASP: aspartic acid; GAL: galactose; GLUC: glucose, GLU: glutamic acid; GLUM: glutamine; GLY: glycogen; GUA: guanosine; MAL: maltotriose; MET: methionine; INO: myo-inositol; MYR: myristic acid; NADH: nicotinamide adenine dinucleotide reduced; OCT: octanoic acid; PAL: palmitic acid; SER: serine; TYR: tyrosine; VAL: valine.

	Total metabolites	Hits	p-value	FDR	Associated Metabolites						
Ammonia Recycling	32	6	<0.001	0.003	AMP	GLUC	SER	ASP	GLUM	NADH	
Urea Cycle	29	5	<0.001	0.014	AMP	NADH	ASP	GLUM	GLU		
Glutamate Metabolism	49	6	<0.001	0.015	AMP	GLU	ASP	GLUM	ADPR	NADH	
Purine Metabolism	74	7	<0.001	0.017	AMP	GUA	GLU	ASP	GLUM	NADH	ADPR
Nicotinate and Nicotinamide Metabolism	37	5	<0.001	0.017	NADH	GLU	GLUM	ADPR	AMP		
Malate-Aspartate Shuttle	10	3	0.001	0.017	NADH	ASP	GLU				
Glucose-Alanine Cycle	13	3	0.002	0.034	GLUC	GLU	NADH				
Aspartate Metabolism	35	4	0.006	0.078	GLUM	GLU	ASP	AMP			
Glycine and Serine Metabolism	59	5	0.008	0.084	AMP	GLU	SER	MET	NADH		
Galactose Metabolism	38	5	0.008	0.084	NADH	INO	MAL	GLUM	GAL		
Propanoate Metabolism	42	4	0.012	0.109	NADH	GLU	VAL	AMP			
Methionine Metabolism	43	4	0.013	0.109	AMP	SER	MET	NADH			
Phenylacetate Metabolism	9	2	0.015	0.119	AMP	GLUM					
Cysteine Metabolism	26	3	0.018	0.129	NADH	GLU	AMP				
Mitochondrial Beta-Oxidation of Short Chain Saturated Fatty Acids	27	3	0.020	0.133	AMP	OCT	NADH				
Phenylalanine and Tyrosine Metabolism	28	3	0.022	0.138	AMP	GLU	TYR				
Arginine and Proline Metabolism	53	4	0.027	0.157	AMP	GLU	ASP	NADH			
Glycogen and Sucrose Metabolism	31	3	0.029	0.161	MAL	GLY	NADH				
Beta-Alanine Metabolism	34	3	0.037	0.184	GLU	ASP	NADH				
Fatty Acid Biosynthesis	35	3	0.040	0.189	PAL	OCT	MYR				

In our previous studies, BPA was shown to act as a metabolic disruptor enhancing the accumulation of storage lipids into lipid droplets without affecting growth, neither reproduction [33]. The production of lipid droplets requires the mobilization of large amounts of fatty acids into triacylglycerols [49]. Interestingly, BPA down-regulated the biosynthesis of fatty acids probably because the fatty acids involved in this pathway were incorporated into triacylglycerols (**Table 4**). This means that, observed metabolic changes upon BPA exposure are likely to be related to major disruptive metabolic changes in sugar and lipid metabolism rather than to a general response to toxic stress. In relation to this, it is also possible that protein constituents (amino acids) were used as additional carbon sources for sugar synthesis [50]. Conversion of amino acids into sugars involves that most of their carbon skeleton is converted to glucose. The nitrogen is converted to urea and, to a lesser extent, to glutamine. The integration of the urea cycle with gluconeogenesis ensures that the bulk of the reducing power (NADH) required in the cytosol for gluconeogenesis can be provided by reactions of the urea cycle [50], and glutamate is at the middle of these metabolic events [32]. Interestingly is that 19 out of the 20 enriched pathways (**Table 4**) were closely related with central carbon metabolism and gluconeogenesis. Furthermore, three of the de-regulated metabolites that could not be annotated using the MetaboAnalyst platform (L-Hydroxypropyl-L-Cysteine, L-Asparaginyll-L-methionine and Prolyl-Glutamine) were dipeptides coming from the incomplete degradation of proteins (according to HMDB database), which would further support our previous arguments.

5. Conclusions

This study proposes the combination of target/untarget analysis for the discriminant quantitative profiling of the *D. magna* metabolites under BPA exposure stressing conditions. The optimum number of *daphnids* was investigated and two distinct extraction protocols were tested to achieve optimal extraction efficiency. The extraction protocol of 8 *daphnids* with MeOH and chloroform allowed lipid removal and cleaner extracts with higher recoveries and with a larger number of identified metabolites.

We show that the combination of the targeted and untargeted approaches to ensure optimum analytical performance and complete characterization of the *D. magna* metabolome under BPA exposure condition. While the targeted analysis is limited to a reduced number of preselected compounds, compounds at low concentrations can be more safely determined. In the untargeted approach, a considerably larger number of compounds can be properly identified and allow expanding the toxicological knowledge and biomarker identification, at expenses that minor compounds can be lost during the

data processing steps. In these cases, to minimize the loss of metabolites requires ROI optimization. Based in the combined target-untargeted approach, this study demonstrates that *Daphnia* exposed to BPA produced the de-regulation of 28 metabolites. Most enriched metabolic pathways of *D. magna* by BPA exposure were related with amino acid, urea, ammonia and glutamate metabolism.

The potential of target and untargeted approach can expand the potential of metabolic analysis in terms of both enhanced identification and quantification of biomolecules affected by environmental stressors.

Acknowledgments

The research leading to these results received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement no. 320737. Romà Tauler acknowledges the Ministerio de Economía y Competividad, Spain for the grant CTQ2015-66254-C2-1-P.

Compliance with Ethical Standards

All authors have accepted the content of this paper and the principles of ethical and professional conduct. This research has no conflict of interest and has not involved humans or animals.

References

1. Viant, M.R., Metabolomics of aquatic organisms: The new 'omics' on the block. *Marine Ecology Progress Series*, 2007. 332: p. 301-306.
2. Butler, H.J., et al., Using Raman spectroscopy to characterize biological materials. *Nature Protocols*, 2016. 11(4): p. 664-687.
3. Reo, N.V., NMR-based metabolomics. *Drug and Chemical Toxicology*, 2002. 25(4): p. 375-382.
4. Dunn, W.B., et al., Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature protocols*, 2011. 6(7): p. 1060-1083.
5. Rochat, B., From targeted quantification to untargeted metabolomics: Why LC-high-resolution-MS will become a key instrument in clinical labs. *TrAC Trends in Analytical Chemistry*, 2016. 84(Part B): p. 151-164.
6. Gorrochategui, E., et al., Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow. *TrAC Trends in Analytical Chemistry*, 2016. 82: p. 425-442.

7. Zhou, J. and Y. Yin, Strategies for large-scale targeted metabolomics quantification by liquid chromatography-mass spectrometry. *Analyst*, 2016. 141(23): p. 6362-6373.
8. Yuan, M., et al., A positive/negative ion-switching, targeted mass spectrometry-based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue. *Nature protocols*, 2012. 7(5): p. 872-881.
9. Sawada, Y., et al., Widely Targeted Metabolomics Based on Large-Scale MS/MS Data for Elucidating Metabolite Accumulation Patterns in Plants. *Plant and Cell Physiology*, 2009. 50(1): p. 37-47.
10. Gómez-Canela, C., et al., Targeted metabolomics of *Gammarus pulex* following controlled exposures to selected pharmaceuticals in water. *Science of The Total Environment*, 2016. 562: p. 777-788.
11. Rivetti, C., et al., Characterization of neurotransmitter profiles in *Daphnia magna* juveniles exposed to environmental concentrations of antidepressants and anxiolytic and antihypertensive drugs using liquid chromatography–tandem mass spectrometry. *Analytical and bioanalytical chemistry*, 2019. 411(22): p. 5867-5876.
12. Gómez-Canela, C., et al., Analysis of neurotransmitters in *Daphnia magna* affected by neuroactive pharmaceuticals using liquid chromatography-high resolution mass spectrometry. *Environmental Pollution*, 2019. 254: p. 113029.
13. Garreta-Lara, E., et al., Metabolic profiling of *Daphnia magna* exposed to environmental stressors by GC–MS and chemometric tools. *Metabolomics*, 2016. 12(5): p. 86.
14. Navarro-Reig, M., et al., Evaluation of changes induced in rice metabolome by Cd and Cu exposure using LC-MS with XCMS and MCR-ALS data analysis strategies. *Analytical and Bioanalytical Chemistry*, 2015. 407(29): p. 8835-8847.
15. Ortiz-Villanueva, E., et al., Metabolic disruption of zebrafish (*Danio rerio*) embryos by bisphenol A. An integrated metabolomic and transcriptomic approach. *Environmental Pollution*, 2017. 231(Part 1): p. 22-36.
16. Gómez-Canela, C., et al., Assessment of chlorpyrifos toxic effects in zebrafish (*Danio rerio*) metabolism. *Environmental Pollution*, 2017. 220: p. 1231-1243.
17. Dettmer, K., P.A. Aronov, and B.D. Hammock, Mass spectrometry-based metabolomics. *Mass spectrometry reviews*, 2007. 26(1): p. 51-78.
18. Ortmayr, K., et al., Increasing selectivity and coverage in LC-MS based metabolome analysis. *TrAC Trends in Analytical Chemistry*, 2016. 82 (Supplement C): p. 358-366.
19. Gorrochategui, E., J. Jaumot, and R. Tauler, A protocol for LC-MS metabolomic data processing using chemometric tools. *Protocol Exchange*, 2015.
20. Uppal, K., et al., xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. *BMC Bioinformatics*, 2013. 14(1): p. 15.
21. Takahashi, H., et al., AMDORAP: Non-targeted metabolic profiling based on high-resolution LC-MS. *BMC Bioinformatics*, 2011. 12.

22. Yi, L., et al., Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. *Analytica Chimica Acta*, 2016. 914: p. 17-34.
23. Garreta-Lara, E., et al., Metabolic profiling of *Daphnia magna* exposed to environmental stressors by GC-MS and chemometric tools. *Metabolomics*, 2016. 12(5).
24. Jordão, R., et al., Obesogens beyond vertebrates: Lipid perturbation by tributyltin in the crustacean *Daphnia magna*. *Environmental Health Perspectives*, 2015. 123(8): p. 813-819.
25. Rivetti, C., B. Campos, and C. Barata, Low environmental levels of neuro-active pharmaceuticals alter phototactic behaviour and reproduction in *Daphnia magna*. *Aquatic Toxicology*, 2016. 170: p. 289-296.
26. Shaw, J.R., et al., *Daphnia* as an emerging model for toxicological genomics, in *Advances in Experimental Biology*, C. Hogstrand and P. Kille, Editors. 2008, Elsevier. p. 165-328.
27. Colbourne, J.K., et al., The Ecoresponsive Genome of *Daphnia pulex*. *Science*, 2011. 331(6017): p. 555.
28. Taylor, N.S., et al., A new approach to toxicity testing in *Daphnia magna*: Application of high throughput FT-ICR mass spectrometry metabolomics. *Metabolomics*, 2009. 5(1): p. 44-58.
29. Vandenbrouck, T., et al., Mixtures of similarly acting compounds in *Daphnia magna*: From gene to metabolite and beyond. *Environment International*, 2010. 36(3): p. 254-268.
30. Poynton, H.C., et al., Metabolomics of Microliter Hemolymph Samples Enables an Improved Understanding of the Combined Metabolic and Transcriptional Responses of *Daphnia magna* to Cadmium. *Environmental Science & Technology*, 2011. 45(8): p. 3710-3717.
31. Garreta-Lara, E., et al., Combined effects of salinity, temperature and hypoxia on *Daphnia magna* metabolism. *Science of the Total Environment*, 2018. 610-611: p. 602-612.
32. Mu, X., et al., Covert signal disruption: Anti-ecdysteroidal activity of bisphenol A involves cross talk between signaling pathways. *Environmental Toxicology and Chemistry*, 2005. 24(1): p. 146-152.
33. Jordão, R., et al., Compounds altering fat storage in *Daphnia magna*. *Science of The Total Environment*, 2016. 545-546: p. 127-136.
34. Fuertes, I., et al., Allocation of glycerolipids and glycerophospholipids from adults to eggs in *Daphnia magna*: Perturbations by compounds that enhance lipid droplet accumulation. *Environmental Pollution*, 2018. 242: p. 1702-1710.
35. Wishart, D.S., et al., HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Research*, 2009. 37(suppl_1): p. D603-D610.
36. Barata, C. and D.J. Baird, Phenotypic plasticity and constancy of life-history traits in laboratory clones of *Daphnia magna* straus: Effects of neonatal length. *Functional Ecology*, 1998. 12(3): p. 442-452.

37. Nagato, E.G., A.J. Simpson, and M.J. Simpson, Metabolomics reveals energetic impairments in *Daphnia magna* exposed to diazinon, malathion and bisphenol-A. *Aquatic Toxicology*, 2016. 170: p. 175-186.
38. Official Journal of the European Communities, O., Implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results. 2002. p. 29.
39. Dalmau, N., C. Bedia, and R. Tauler, Validation of the Regions of Interest Multivariate Curve Resolution (ROIMCR) procedure for untargeted LC-MS lipidomic analysis. *Analytica Chimica Acta*, 2018. 1025: p. 80-91.
40. Jaumot, J., A. de Juan, and R. Tauler, MCR-ALS GUI 2.0: New features and applications. *Chemometrics and Intelligent Laboratory Systems*, 2015. 140: p. 1-12.
41. Wold, S., K. Esbensen, and P. Geladi, Principal component analysis. *Chemometrics and intelligent laboratory systems*, 1987. 2(1-3): p. 37-52.
42. Gorrochategui, E., J. Jaumot, and R. Tauler, ROIMCR: a powerful analysis strategy for LC-MS metabolomic datasets. *BMC Bioinformatics*, 2019. 20(1): p. 256.
43. Farrés, M., B. Piña, and R. Tauler, LC-MS based metabolomics and chemometrics study of the toxic effects of copper on *Saccharomyces cerevisiae*. *Metallomics*, 2016. 8(8): p. 790-798.
44. Smith, C.A., et al., METLIN: A metabolite mass spectral database. *Therapeutic Drug Monitoring*, 2005. 27(6): p. 747-751.
45. Wishart, D.S., et al., HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Research*, 2013. 41(D1): p. D801-D807.
46. Kanehisa, M., et al., KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 2017. 45(D1): p. D353-D361.
47. Xia, J. and D.S. Wishart, Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis, in *Current Protocols in Bioinformatics*. 2016, John Wiley & Sons, Inc.
48. Newsholme, P., et al., Glutamine and glutamate - Their central role in cell metabolism and function. *Cell Biochemistry and Function*, 2003. 21(1): p. 1-9.
49. Olofsson, S.O., et al., Lipid droplets as dynamic organelles connecting storage and efflux of lipids. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, 2009. 1791(6): p. 448-458.
50. Brosnan, J.T., Glutamate, at the interface between amino acid and carbohydrate metabolism. *Journal of Nutrition*, 2000. 130 (4 SUPPL.): p. 988S-990S.

Supplementary Information

Novel combined approach for the target/untarget characterization of *Daphnia magna* metabolome

Elba Garreta-Lara, Bruno Campos, Carlos Barata, Romà Tauler, Sílvia Lacorte

Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, Barcelona, 08034, Catalonia

Corresponding author: Sílvia Lacorte

E-mail address: slbqam@cid.csic.es

This file contents:

Optimization of the ionization parameters	2
Mass spectral characterization	2 - 5
Table SI1	6
Table SI2	7
Table SI3	8
Table SI4	9
Table SI5	10
Table SI6.....	11
References.....	12

Optimization of the ionization parameters

These ionization conditions were performed using direct infusion to configure the most intense ions of the analyzed metabolites. These experiments were carried out with individual standard solutions of the metabolites at a capillary voltage from 25 to 70 V, skimmer voltage from 15 to 25 V, and tube lens voltage from 30 to 100 V under positive and negative ESI mode. Negative ionization mode was chosen because its better signal sensitivity of all the metabolites studied (**Table S12**). Sheath gas flow rate was optimized from 20 to 45 au and auxiliary gas flow rate from 2 to 45 au. Different mass spectral conditions were obtained depending on the metabolite analyzed, so the optimum parameters were described to obtain the best sensitivity and selectivity for most of compounds. The best resolution was obtained in negative ionization mode with a spray voltage of 3.00 kV, capillary voltage of 30 V, skimmer voltage of 25 V and tube lens voltage of 75 V. Also, sheath gas flow rate at 45 au, auxiliary gas flow rate at 10 au and a capillary temperature at 300 °C were selected.

Mass spectral characterization

The mass spectral characterization of target compounds was achieved at the optimum working conditions specified in the present work, as shown in **Table 1** in the manuscript. This characterization was performed according to the fragmentation patterns of each compound checked in METLIN (<http://metlin.scripps.edu>) [1] and HMDB [2] online metabolomics databases. In a previous study, Gómez-Canela et al. [3] studied the changes of several metabolites (including amino acids, nucleotides, nucleosides and sugars) in *Gammarus pulex* after the exposure of selected pharmaceuticals, where some precursor ions were concurrent in the present work.

Sugars

Belonging to this group, maltose, trehalose, myo-inositol, galactose and glucose were studied. Maltose and trehalose are isobaric compounds. These metabolites have the same exact mass. However, distinct fragment ions were obtained by Orbitrap MS for each metabolite for a better identification. Maltose showed the deprotonated molecule at m/z 341.1087 $[M-H]^-$ as precursor ion and one fragment was observed at m/z 221.0664 $[C_8H_{13}O_7]^-$. Otherwise, trehalose gave also the deprotonated molecule at m/z 341.1088 as base peak. However, using HCD cell at 25 eV, one fragment ion at m/z 113.0241

$[\text{C}_5\text{H}_5\text{O}_3]^-$ was observed. No studies about the fragmentation by LC-MS based techniques of maltose and trehalose were found in the literature.

Glucose, galactose and myo-inositol are structural isomers. These molecules gave the same deprotonated molecule at m/z 179.0559 $[\text{M}-\text{H}]^-$. Glucose showed one fragment ion at m/z 101.0243 $[\text{C}_4\text{H}_5\text{O}_3]^-$ and galactose at m/z 113.0242 $[\text{C}_5\text{H}_5\text{O}_3]^-$. No studies about the fragmentation by LC-MS based techniques of glucose and galactose were found in the literature. Myo-inositol formed the same deprotonated molecule at m/z 179.0560 $[\text{M}-\text{H}]^-$ as precursor ion. Using the HCD cell at 25 eV one fragment ions was formed at m/z 141.0193 $[\text{C}_6\text{H}_5\text{O}_4]^-$ that corresponds to the loss of two hydroxyl groups. In a previous study, an analytical method based on LC-MS/MS to differentiate the myo-inositol and chiro-inositol isomers was optimized [4]. Authors reported that the m/z 179 $[\text{M}-\text{H}]^-$ to m/z 87 transition was selected to determine the concentration of myo-inositol, but no information of the fragment ion formed at m/z 141.0193 was showed. In this work, the mass range of the studied metabolites was from 100 to 700 Da.

Amino acids

L-aspartic acid, L-citrulline, L-phenylalanine, L-methionine, L-tryptophan, L-tyrosine, L-serine and L-valine were the amino acids studied in the present work. L-aspartic acid, L-methionine, L-serine and L-valine gave the deprotonated molecules $[\text{M}-\text{H}]^-$ at m/z 132.0298, at m/z 148.0437, at m/z 104.0351 and at m/z 116.0716, respectively. Due to the low percentage of abundance ($< 10\%$) in the fragmentation pattern for these molecules, no fragment ions could be identified. In these cases, retention time and exact mass with four decimals of the precursor ion and with an error < 5 ppm were used as identification points, following the Directive 96/23/EC [5]. On the other hand, L-citrulline showed the deprotonated molecule at m/z 174.0883 as base peak. Using the HCD cell at 25 eV, one fragment ion was formed at m/z 131.0824 $[\text{C}_5\text{H}_{11}\text{O}_2\text{N}_2]^-$. In the case of L-phenylalanine, the deprotonated molecule was formed at m/z 164.0717 $[\text{M}-\text{H}]^-$ and one fragment ion was observed at m/z 147.0449 $[\text{C}_9\text{H}_7\text{O}_2]^-$ corresponding to the loss of the amine group. The deprotonated molecule of L-tryptophan was shown at m/z 203.0824 $[\text{M}-\text{H}]^-$ as base peak. One fragment ion was formed at m/z 116.0505 $[\text{C}_8\text{H}_6\text{N}]^-$, corresponding to the indole of the molecule. Finally, L-tyrosine showed the deprotonated molecule at m/z 180.0664 $[\text{M}-\text{H}]^-$ as precursor ion and using HCD cell at 25 eV, one fragment ion was observed at m/z 119.0500 $[\text{C}_8\text{H}_7\text{O}]^-$, that corresponds to the loss of the amino and the carboxylic groups. In a previous study of Gómez-Canela et al., authors reported the same precursor ion for the coincident amino acids studied in the present work [3].

Nucleotides

Belonging to this group, adenosine diphosphate (ADP) and nicotinamide adenine dinucleotide reduced (NADH) were described. ADP showed the deprotonated molecule at m/z 426.0215 [M-H]⁻ as precursor ion. However, no fragment ions could be identified due to the low percentage of abundance (< 10%) in the fragmentation pattern for this molecule. NADH consists of two nucleotides joined through their phosphate groups. NADH formed the deprotonated molecule at m/z 662.1021 [M-H]⁻ as precursor ion. Using HCD cell at 25 eV, one fragment ion was obtained at m/z 540.0536 [C₁₅H₂₀O₁₃N₅P₂]⁻ that corresponds to the loss of the nicotinamide group.

Nucleosides

Cytidine, inosine, uridine, adenosine and guanosine were the nucleosides described in this work. Cytidine showed the deprotonated molecule at m/z 242.0780 [M-H]⁻ as the precursor ion and using HCD cell at 25 eV, one fragment ion was obtained at m/z 150.0421 [C₇H₆O₂N₂]⁻. Otherwise, inosine formed the deprotonated molecule at m/z 267.0732 [M-H]⁻ as precursor ion. The most fragment ion was exhibited at m/z 135.0312 [C₅H₃ON₄]⁻ corresponding to the loss of the sugar group of the molecule. The deprotonated molecule of uridine was shown at m/z 243.0620 [M-H]⁻ as base peak. Using HCD cell at 25 eV, one fragment ion was obtained at m/z 110.0247 [C₅H₄O₂N]⁻. The deprotonated molecule of adenosine was formed at m/z 266.0895 [M-H]⁻. One fragment ion was produced at m/z 134.0471 [C₅H₄N₅]⁻ that corresponds to the loss of the sugar group. Finally, guanosine formed the deprotonated molecule at m/z 282.0846 [M-H]⁻ as precursor ion and using the HCD cell at 25 eV, one fragment ion was shown at m/z 150.0420 [C₅H₄ON₅]⁻ corresponding to the nitrogenous base of the molecule.

Neurotransmitters

Neurotransmitters are also known as chemical messengers and are endogenous chemicals that enable neurotransmission. They transmit signals across a chemical synapse, such as a neuromuscular junction, from one neuron (nerve cell) to another "target" neuron, muscle cell, or gland cell. Belonging to this group, hypoxanthine has been described. Hypoxanthine formed the deprotonated molecule at m/z 135.0312 [M-H]⁻ as precursor ion and using HCD cell at 25 eV, one fragment ion was obtained at m/z 92.0253 [C₄H₂N₃]⁻.

Organic acids

Fumaric acid showed the deprotonated molecule at m/z 115.0037 $[M-H]^-$ as base peak with no fragment ions due to the low percentage of abundance (< 10%) in the fragmentation pattern. Pimelic acid gave the deprotonated molecule at m/z 159.0660 as precursor ion. Using the HCD cell at 25 eV, one fragment ion was formed at m/z 115.0761 $[C_6H_{11}O_2]^-$ corresponding to the loss of a carboxylic group.

Table S11. Physicochemical properties of the target compounds studied.

Chemical families	Target compounds	KEGG-ID	CAS no.	Molecular formula	Molecular weight	pKa
Sugars	Maltose	C00208	69-79-4	C ₁₂ H ₂₂ O ₁₁	342.30	11.25
	Trehalose	C01083	99-20-7	C ₁₂ H ₂₂ O ₁₁	342.30	11.91
	Myo-Inositol	C00137	87-89-8	C ₆ H ₁₂ O ₆	180.16	12.29
	Galactose	C01825	59-23-4	C ₆ H ₁₂ O ₆	180.16	11.30
	Glucose	C00031	50-99-7	C ₆ H ₁₂ O ₆	180.16	11.30
Amino acids	L-Aspartic acid	C00049	56-84-8	C ₄ H ₇ NO ₄	133.10	1.70
	L-Citrulline	C00327	372-75-8	C ₆ H ₁₃ N ₃ O ₃	175.19	2.27
	L-Phenylalanine	C00079	63-91-2	C ₉ H ₉ NO ₂	165.19	2.47
	L-Methionine	C00073	63-68-3	C ₅ H ₁₁ NO ₂ S	149.21	2.53
	L-Tryptophan	C00078	73-22-3	C ₁₁ H ₁₂ N ₂ O ₂	204.23	2.54
	L-Tyrosine	C00082	60-18-4	C ₉ H ₉ NO ₃	181.19	2.00
	L-Serine	C00065	56-45-1	C ₃ H ₇ NO ₃	105.09	2.03
	L-Valine	C00183	72-18-4	C ₅ H ₁₁ NO ₂	117.15	2.72
Nucleotides	ADP	C00008	58-64-0	C ₁₀ H ₁₅ N ₅ O ₁₀ P ₂	427.21	1.77
	NADH	C00004	53-84-9	C ₂₁ H ₂₉ N ₇ O ₁₄ P ₂	664.44	-7.00
Nucleosides	Cytidine	C00475	65-46-3	C ₉ H ₁₃ N ₃ O ₅	243.22	12.55
	Inosine	C00294	58-63-9	C ₁₀ H ₁₂ N ₄ O ₅	268.23	6.94
	Uridine	C00299	58-96-8	C ₉ H ₁₂ N ₂ O ₆	244.21	9.70
	Adenosine	C00212	58-61-7	C ₁₀ H ₁₃ N ₅ O ₄	267.25	12.45
	Guanosine	C00387	118-00-3	C ₁₀ H ₁₃ N ₅ O ₅	283.25	10.16
Neurotransmitters	Hypoxanthine	C00262	68-94-0	C ₅ H ₄ N ₄ O	136.11	8.72
Organic acids	Fumaric acid	C00122	110-17-8	C ₄ H ₄ O ₄	116.07	3.55
	Pimelic acid	C02656	111-16-0	C ₇ H ₁₂ O ₄	160.17	4.05

Table S12. Intensities of the target compounds in the positive and negative mode of LC-Orbitrap MS. *n.d.* not detected.

Target compounds	Molecular formula	Exact mass	Precursor ion			
			[M-H] ⁻	Intensity (u.a.)	[M+H] ⁺	Intensity (u.a.)
Maltose	C ₁₂ H ₂₂ O ₁₁	342.1162	341.1089	6.56E+05	343.1235	<i>n.d.</i>
Trehalose	C ₁₂ H ₂₂ O ₁₁	342.1162	341.1089	6.56E+05	343.1235	<i>n.d.</i>
Myo-Inositol	C ₆ H ₁₂ O ₆	180.0634	179.0561	1.42E+06	181.0707	<i>n.d.</i>
Galactose	C ₆ H ₁₂ O ₆	180.0634	179.0561	1.42E+06	181.0707	<i>n.d.</i>
Glucose	C ₆ H ₁₂ O ₆	180.0634	179.0561	1.42E+06	181.0707	<i>n.d.</i>
L-Aspartic acid	C ₄ H ₇ NO ₄	133.0375	132.0302	2.52E+04	134.0448	4.72E+05
L-Citrulline	C ₆ H ₁₃ N ₃ O ₃	175.0957	174.0884	2.06E+06	176.1030	2.60E+06
L-Phenylalanine	C ₉ H ₁₁ NO ₂	165.079	164.0717	7.37E+06	166.0863	5.77E+06
L-Methionine	C ₅ H ₁₁ NO ₂ S	149.051	148.0437	1.15E+06	150.0584	3.03E+06
L-Tryptophan	C ₁₁ H ₁₂ N ₂ O ₂	204.0899	203.0826	1.16E+07	205.0972	8.47E+06
L-Tyrosine	C ₉ H ₁₁ NO ₃	181.0739	180.0666	5.80E+06	182.0812	3.44E+06
L-Serine	C ₃ H ₇ NO ₃	105.0426	104.0353	1.58E+06	106.0499	1.61E+06
L-Valine	C ₅ H ₁₁ NO ₂	117.079	116.0717	4.09E+06	118.0863	5.75E+06
ADP	C ₁₀ H ₁₅ N ₅ O ₁₀ P ₂	427.0294	426.0221	1.16E+05	428.0367	6.25E+04
NADH	C ₂₁ H ₂₉ N ₇ O ₁₄ P ₂	665.1248	662.1018	8.84E+04	666.1321	<i>n.d.</i>
Cytidine	C ₉ H ₁₃ N ₃ O ₅	243.0855	242.0782	5.31E+06	244.0928	4.95E+06
Inosine	C ₁₀ H ₁₂ N ₄ O ₅	268.0808	267.0734	1.10E+07	269.0881	3.84E+06
Uridine	C ₉ H ₁₂ N ₂ O ₆	244.0695	243.0622	1.19E+07	245.0769	2.80E+04
Adenosine	C ₁₀ H ₁₃ N ₅ O ₄	267.0968	266.0894	9.58E+05	268.1041	1.51E+07
Guanosine	C ₁₀ H ₁₃ N ₅ O ₅	283.0917	282.0843	5.91E+06	284.0990	2.89E+06
Hyoxanthine	C ₅ H ₄ N ₄ O	136.0385	135.0312	1.30E+07	137.0458	8.69E+06
Fumaric acid	C ₄ H ₄ O ₄	116.011	115.0036	2.80E+06	117.0183	3.17E+02
Pimelic acid	C ₇ H ₁₂ O ₄	160.0736	159.0662	1.15E+07	161.0809	1.13E+03

Table S13. Recoveries of the target compounds obtained with extraction protocols A and B using spiked *D. magna* samples. *n.r.*: not recovered.

Target compounds	Molecular formula	Quantification ion	% Recoveries extraction \pm RSD A					% Recoveries extraction \pm RSD B				
			2	3	5	8	10	2	3	5	8	10
Maltose	C ₁₂ H ₂₂ O ₁₁	221.0664	71 \pm 9	46 \pm 6	89 \pm 21	89 \pm 15	109 \pm 18	n.r.	31 \pm 13	n.r.	60 \pm 20	156 \pm 17
Trehalose	C ₁₂ H ₂₂ O ₁₁	113.0241	100 \pm 4	88 \pm 7	98 \pm 12	94 \pm 6	102 \pm 11	73 \pm 15	87 \pm 14	66 \pm 14	102 \pm 8	166 \pm 13
Myo-Inositol	C ₆ H ₁₂ O ₆	179.0561	87 \pm 2	82 \pm 1	80 \pm 7	79 \pm 4	74 \pm 6	87 \pm 9	86 \pm 1	81 \pm 10	77 \pm 5	82 \pm 4
Galactose	C ₆ H ₁₂ O ₆	113.0242	28 \pm 3	n.r.	n.r.	57 \pm 8	70 \pm 25	n.r.	n.r.	50 \pm 7	55 \pm 14	n.r.
Glucose	C ₆ H ₁₂ O ₆	101.0243	n.r.	n.r.	n.r.	33 \pm 9	27 \pm 38	n.r.	n.r.	n.r.	n.r.	n.r.
L-Aspartic acid	C ₄ H ₇ NO ₄	132.0302	34 \pm 9	34 \pm 4	33 \pm 0	43 \pm 9	44 \pm 4	28 \pm 14	26 \pm 18	29 \pm 10	27 \pm 6	n.r.
L-Citrulline	C ₆ H ₁₃ N ₃ O ₃	174.0884	41 \pm 5	46 \pm 3	47 \pm 2	51 \pm 5	47 \pm 8	44 \pm 11	44 \pm 15	51 \pm 5	55 \pm 6	n.r.
L-Phenylalanine	C ₉ H ₁₁ NO ₂	164.0717	63 \pm 8	57 \pm 5	48 \pm 6	62 \pm 15	56 \pm 8	39 \pm 8	32 \pm 13	44 \pm 5	38 \pm 6	55 \pm 7
L-Methionine	C ₅ H ₁₁ NO ₂ S	148.0437	n.r.	n.r.	17 \pm 6	25 \pm 13	40 \pm 10	n.r.	n.r.	n.r.	n.r.	n.r.
L-Tryptophan	C ₁₁ H ₁₂ N ₂ O ₂	203.0826	62 \pm 4	53 \pm 6	53 \pm 4	55 \pm 14	50 \pm 3	51 \pm 6	42 \pm 10	42 \pm 4	36 \pm 8	31 \pm 18
L-Tyrosine	C ₉ H ₁₁ NO ₃	180.0666	65 \pm 8	56 \pm 4	49 \pm 2	75 \pm 16	74 \pm 15	n.r.	n.r.	62 \pm 8	33 \pm 10	64 \pm 7
L-Serine	C ₃ H ₇ NO ₃	104.0353	n.r.	16 \pm 6	13 \pm 2	24 \pm 9	20 \pm 5	14 \pm 10	n.r.	19 \pm 8	17 \pm 6	17 \pm 6
L-Valine	C ₅ H ₁₁ NO ₂	116.0717	30 \pm 6	28 \pm 2	30 \pm 0	41 \pm 11	40 \pm 6	18 \pm 8	16 \pm 11	20 \pm 3	18 \pm 4	26 \pm 4
ADP	C ₁₀ H ₁₅ N ₅ O ₁₀ P ₂	426.0221	133 \pm 7	156 \pm 1	102 \pm 12	111 \pm 2	103 \pm 1	134 \pm 23	105 \pm 19	68 \pm 17	43 \pm 19	25 \pm 18
NADH	C ₂₁ H ₂₈ N ₇ O ₁₄ P ₂	662.1018	110 \pm 1	115 \pm 4	103 \pm 10	94 \pm 4	91 \pm 13	121 \pm 9	116 \pm 5	94 \pm 6	87 \pm 6	63 \pm 14
Cytidine	C ₉ H ₁₃ N ₃ O ₅	242.0782	101 \pm 3	96 \pm 4	91 \pm 7	82 \pm 7	83 \pm 5	101 \pm 8	92 \pm 1	84 \pm 5	72 \pm 14	91 \pm 7
Inosine	C ₁₀ H ₁₂ N ₄ O ₅	267.0734	97 \pm 7	62 \pm 4	57 \pm 6	50 \pm 12	59 \pm 7	66 \pm 7	49 \pm 4	26 \pm 5	55 \pm 6	110 \pm 4
Uridine	C ₉ H ₁₂ N ₂ O ₆	243.0622	104 \pm 5	102 \pm 4	91 \pm 11	79 \pm 12	93 \pm 5	89 \pm 4	81 \pm 5	46 \pm 5	54 \pm 7	115 \pm 4
Adenosine	C ₁₀ H ₁₃ N ₅ O ₄	266.0894	90 \pm 24	110 \pm 2	95 \pm 9	89 \pm 11	89 \pm 14	108 \pm 10	99 \pm 20	92 \pm 6	78 \pm 14	64 \pm 18
Guanosine	C ₁₀ H ₁₃ N ₅ O ₅	282.0843	90 \pm 3	92 \pm 2	83 \pm 10	68 \pm 11	81 \pm 7	75 \pm 8	79 \pm 5	51 \pm 8	65 \pm 10	88 \pm 16
Hypoxanthine	C ₅ H ₄ N ₄ O	135.0312	77 \pm 4	76 \pm 3	74 \pm 2	71 \pm 5	68 \pm 8	82 \pm 14	n.r.	76 \pm 7	70 \pm 4	70 \pm 9
Fumaric acid	C ₄ H ₄ O ₄	115.0036	29 \pm 1	27 \pm 3	25 \pm 6	25 \pm 2	27 \pm 2	29 \pm 13	29 \pm 2	28 \pm 3	26 \pm 7	26 \pm 13
Pimelic acid	C ₇ H ₁₂ O ₄	159.0662	74 \pm 6	65 \pm 19	70 \pm 3	67 \pm 3	67 \pm 4	76 \pm 12	79 \pm 2	76 \pm 1	69 \pm 8	65 \pm 15
Total number of recovered metabolites			20	20	21	23	23	18	17	20	21	18

Table SI4. MCR-ALS fitting results are shown for the analysis of ROIs augmented data matrix.

R² explained data variance of the model.

Number estimated components	R ² (%)	Lack of fit (%)
200	97.8	14.62
190	98.5	11.8
180	98.9	7.9
175	99.2	9.2
170	99.4	7.9
160	99.5	8.5
165	99.5	6.9
160	99.5	7.3
150	99.5	6.7
125	99.3	7.9
110	99.4	6.8
105	99.3	7.9
104	99.5	6.8
103	99.6	6.5
102	98.8	10.9
101	99.6	9.3
100	99.5	6.9
99	99.6	6.8
98	99.6	6.7
97	99.6	6.4
96	99.6	6.7
95	99.5	6.8
94	99.5	6.7
93	99.5	7.1
92	99.2	8.8
91	98.8	10.9
90	99.4	7.5

Table S15. Metabolite concentration (ng daphnia⁻¹) in *D. magna* pools of 8 individuals and the relative standard deviation (RSD) in the distinct studied samples: control, low- and high-dose of BPA exposed samples.

Target compounds	Control	Low-dose BPA	High-dose BPA
	ng daphnia ⁻¹ ± RSD	ng daphnia ⁻¹ ± RSD	ng daphnia ⁻¹ ± RSD
Maltose	22.67 ± 1.20	23.40 ± 1.25	39.94 ± 0.88
Trehalose	7.50 ± 0.35	13.07 ± 0.94	16.42 ± 0.71
Myo-Inositol	4.19 ± 0.03	4.65 ± 0.05	4.70 ± 0.01
Galactose	8.27 ± 0.42	17.76 ± 4.76	31.16 ± 0.02
Glucose	37.69 ± 0.55	46.21 ± 0.46	46.25 ± 0.57
L-Aspartic acid	30.78 ± 0.22	26.31 ± 0.06	33.31 ± 0.24
L-Citrulline	2.79 ± 0.15	2.70 ± 0.10	2.18 ± 0.17
L-Phenylalanine	32.52 ± 0.65	30.80 ± 0.27	34.19 ± 1.01
L-Methionine	12.44 ± 0.38	8.44 ± 0.63	16.91 ± 0.41
L-Tryptophan	5.22 ± 0.12	4.23 ± 0.06	5.77 ± 0.16
L-Tyrosine	74.47 ± 2.64	105.04 ± 2.77	97.39 ± 4.96
L-Serine	35.22 ± 2.11	50.71 ± 0.83	42.89 ± 1.41
L-Valine	63.09 ± 1.99	90.47 ± 1.40	69.02 ± 1.76
ADP	27.62 ± 1.15	28.52 ± 0.85	20.34 ± 0.83
NADH	38.11 ± 0.83	31.41 ± 0.42	28.48 ± 0.54
Cytidine	8.71 ± 0.07	9.23 ± 0.04	9.91 ± 0.01
Inosine	20.06 ± 0.31	17.75 ± 0.24	19.79 ± 0.18
Uridine	62.37 ± 1.84	60.59 ± 0.38	66.65 ± 0.92
Adenosine	5.11 ± 0.15	5.64 ± 0.09	5.22 ± 0.04
Guanosine	1.58 ± 0.04	2.26 ± 0.04	2.64 ± 0.01
Hypoxanthine	0.27 ± 0.01	0.25 ± 0.01	0.20 ± 0.01
Fumaric acid	4.75 ± 0.05	4.95 ± 0.07	5.13 ± 0.08
Pimelic acid	0.36 ± 0.02	0.29 ± 0.01	0.53 ± 0.02

Table S16. Metabolites mean areas of untarget approach in the studied samples: control, low- and high-dose BPA exposed samples.

	Database ID	Metabolites	Mean area control	Mean area low-dose	mean area high-dose
1	C00106	Uracil	26348366	33481732	29379772
2	C00148	L-Proline	73754232	62214819	67443594
3	C01585	Caproic acid	35869324	38566451	32970393
4	C00183	L-Valine §	204065662	195167532	214621833
5	C00188	L-Threonine	54987782	50778865	49884861
6	C00025	Glutamic acid	103202891	45286597	91481097
7	C00407	Leucine/Isoleucine	760197939	674660074	779926723
8	C00147	Adenine	16443366	14411115	14222662
9	C06423	Octanoic acid	88542093	57979353	44251773
10	C00064	Glutamine	147348414	143500777	150884422
11	C01601	Nonanoic acid	66107277	49549133	51002620
12	C02989	DL-Methionine S-oxide	71240457	83231114	55892894
13	C00079	L-Phenylalanine §	195351614	170118225	201517950
14	C01571	Capric acid	19774087	17976177	14905333
15	C00082	L-Tyrosine §	42156506	31591820	58254065
16	C05380	Nicotinuric acid	57977307	66280339	49541666
17	C00078	L-Tryptophan §	37530046	28194901	42524375
18	HMDB00339	2-(2-methylbutanamido)acetic acid	83001973	32679792	84646575
19	HMDB29015	Prolyl-Glutamine	75517599	69091981	98476990
20	C06424	Myristic acid	252813338	198567861	189819672
21	C00327	Citrulline §	24083640	13973315	16015481
22	C00299	Uridine §	146138089	139621359	141607635
23	C00249	Palmitic acid	415118586	36654252	92447354
24	C00294	Inosine §	137437753	119178430	151437654
25	C00673	Deoxyribose 5-phosphate	31496862	5020154	22971860
26	C00669	L-gamma-Glutamylcysteine	20485398	14644587	18537154
27	C00712	Oleic acid	47576768	55530991	52588271
28	HMDB28737	L-Asparaginy-L-methionine	17966170	15491779	19453272
29	C00043	UDP-N-acetylglucosamine	212200046	80877221	167977080
30	C00020	Adenosine monophosphate	441947030	342494437	415373117
31	C00144	Guanosine monophosphate	32519925	21839193	25830284
32	HMDB00656	Cysteine-glutathione disulfide	16420333	14973076	21123127
33	C00008	ADP §	120114057	94546118	94297354
34	C03232	Phosphohydroxypyruvic acid	19262553	25337241	24762322
35	HMDB28860	Hydroxypropyl-L-Cysteine	47471468	56150606	34190334
36	C01835	Maltotriose	30893544	45705143	56927334
37	HMDB41659	3'-O-Methyl(-)-epicatechin 7-O-glucuronide	38431133	39273842	46842386
38	C00301	Adenosine diphosphate ribose	21928864	14446841	26226843
39	C00203	UDP-N-acetylgalactosamine	66789846	23668484	82928020
40	C00127	Oxidized glutathione	18733400	10023768	10141687
41	C01170	UDP-N-acetyl-D-mannosamine	114546064	159534418	99015468
42	C00182	Glycogen	27220904	33717151	41490993

§metabolites also present in the target approach.

References

- [1] C.A. Smith, G. O'Maille, E.J. Want, C. Qin, S.A. Trauger, T.R. Brandon, D.E. Custodio, R. Abagyan, G. Siuzdak, METLIN: A metabolite mass spectral database, *Therapeutic Drug Monitoring*, 27 (2005) 747-751.
- [2] D.S. Wishart, T. Jewison, A.C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorn Dahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner, A. Scalbert, HMDB 3.0—The Human Metabolome Database in 2013, *Nucleic Acids Research*, 41 (2013) D801-D807.
- [3] C. Gómez-Canela, T.H. Miller, N.R. Bury, R. Tauler, L.P. Barron, Targeted metabolomics of *Gammarus pulex* following controlled exposures to selected pharmaceuticals in water, *Science of The Total Environment*, 562 (2016) 777-788.
- [4] B.H. Kim, J.Y. Park, J.B. Jang, D.C. Moon, LC-MS/MS method for the quantification of myo- and chiro-inositol as the urinary biomarkers of insulin resistance in human urine, *Biomedical Chromatography*, 26 (2012) 429-433.
- [5] Directive, Implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results., Directive 2002/657/CE, (2002).

3.4 DISCUSSIÓ DELS RESULTATS

A continuació es descriuen els resultats obtinguts en aquest capítol on es presenten els efectes de la fluoxetina, el diazepam i la carbamazepina en els nivells de les oxilipines de la *D. magna* mitjançant una anàlisi dirigida. També es mostren les pertorbacions a nivell metabòlic a causa de l'exposició a BPA mitjançant les estratègies dirigida i no dirigida, la combinació de les quals permet obtenir una interpretació més global dels efectes d'aquest contaminant ambiental en el metaboloma de la *D. magna*.

3.4.1 ESTRATÈGIA DIRIGIDA

En els estudis de metabolòmica, generalment s'utilitzen dues estratègies analítiques diferents per la determinació dels metabòlits en les mostres biològiques. En la metabolòmica dirigida la utilització dels mètodes basats en LC-MS/MS permeten obtenir una gran sensibilitat per la detecció i la quantificació dels metabòlits prèviament seleccionats, els quals es troben a concentracions baixes.

Per a la determinació dels PUFAs i els seus derivats, aquests estan presents a concentracions molt baixes en els sistemes biològics. En aquest cas, l'adquisició de les dades a través dels procediments de MRM permet la selecció específica d'aquells metabòlits que es volen determinar (Sawada et al. 2008). En aquest context, Balgoma et al. (2013) van desenvolupar un mètode LC-MS/MS quantitatiu, robust i sensible per determinar i quantificar el perfil de les oxilipines, com els eicosanoids i els derivats dels mediadors lipídics, relacionades amb les rutes dels PUFAs. El desenvolupament d'aquesta anàlisi dirigida ha permès l'avaluació de més de 100 oxilipines en diverses matrius biològiques (Checa et al. 2015; Strassburg et al. 2012). Tot i així, el nombre d'estudis que investiguen les alteracions dels PUFAs en la *D. magna* encara són escassos. És per això, que en l'**article científic V** es van optimitzar les condicions analítiques per realitzar una detecció acurada dels mediadors lipídics predeterminats presents en la *D. magna* a uns nivells baixos, de l'ordre de nM i pM.

Les dades que s'obtenen dels estudis dirigits proporcionen informació d'un subgrup de metabòlits coneguts del metaboloma. L'avantatge d'aquesta estratègia recau en primer lloc, en què els possibles artefactes analítics (interferències) seran

eliminats en la determinació dels metabòlits i, en segon lloc, en què els metabòlits seleccionats poden determinar-se de forma quantitativa absoluta mitjançant una corba de calibratge amb patrons de concentració coneguda i l'ús de patrons interns. L'anàlisi de les dades es realitza mitjançant els softwares comercials, els quals faciliten aquesta determinació i quantificació dels metabòlits presents en les mostres. En aquest treball, l'anàlisi de les dades es va portar a terme mitjançant els softwares *MassLynx* i *TargetLynx*, els quals van permetre la determinació i la posterior quantificació de 28 oxilipines derivades de diferents PUFAs en les mostres de *D. magna* (veure **secció 3.1** de l'**article científic V**).

La quantificació dels nivells d'oxilipines permet identificar les seves alteracions relatives en la *D. magna* a causa de l'exposició als tres fàrmacs, fluoxetina, diazepam i carbamazepina. A més, això permet avaluar els riscos ambientals associats a aquests fàrmacs. La majoria d'oxilipines detectades estan relacionades amb les rutes enzimàtiques COX (*cyclooxygenase*), LOX (*lipooxygenase*) i CYP (*cytochrome 450*), les quals regulen la biosíntesis d'oxilipines derivades de l'àcid araquidònic (AA), l'àcid linoleic (LA), l'àcid α -linoleic (α -LA) i l'àcid eicosapentaenoic (EPA), tal com es descriu en la Figura 1 de l'**article científic V**.

Entre els PUFAs, el grup que té una cadena de carbonis C18 és el més freqüent, en particular el LA (18:2n-6) i l' α -LA (18:3n-3). Els PUFAs de cadenes de 20 carbonis només es troben en petites concentracions, essent els relacionats amb l'AA (20:4n-6) seguit dels EPA (20:5n-3) els més abundants (Barata et al. 2005). Així, el nombre d'oxilipines quantificades en aquest treball coincideix amb la seva abundància relativa, que són en ordre creixent: LA, AA, α -LA i EPA. En total, 28 metabòlits derivats dels PUFAs van ésser quantificats en la *D. magna*, tots ells productes finals oxidats de les rutes enzimàtiques involucrades. La gran majoria de les oxilipines estudiades mostren un augment de la concentració d'aquests metabòlits a causa de les exposicions als compostos farmacèutics, fluoxetina, carbamazepina i diazepam. Per altra banda, es van determinar 14 oxilipines que mostraven un canvi significatiu en les seves concentracions (amb un nivell de significació p inferior a 0.05) a causa de les tres exposicions estudiades. En concret, es va identificar una oxilipina per l'exposició a la fluoxetina

corresponent a la ruta de l'AA, una altra que es va veure alterada per l'exposició a la carbamazepina que forma part de la ruta del LA i 12 oxilipines que es trobaven afectades per l'exposició al diazepam les quals estan relacionades amb les rutes de l'AA, LA, α -LA i EPA (veure **Taula 3.1**). Aquestes alteracions en els nivells de les oxilipines posen de manifest l'efecte considerable que produeix l'exposició del diazepam en el metaboloma de la *D. magna*. Les oxilipines que mostren aquest canvi significatiu en les concentracions estan resumides en la **Taula 3.1**.

Taula 3.1 | Oxilipines amb canvis significatius en les concentracions a causa de l'exposició als productes farmacèutics estudiats.

EXPOSICIONS			
	FLUOXETINA	DIAZEPAM	CARBAMAZEPINA
AA	12-HHTrE	5-KETE, 5-HETE, 12-HETE	-
LA	-	9-KODE, 9(10)-EpOME, 12(13)-EpOME, 9-HODE	13-HODE
α -LA	-	13-HOTrE, 9-KOTrE, 9-HOTrE, 12(13)-EpODE	-
EPA	-	5-HEPE	-

Fins ara, el paper biològic dels eicosanoids en la *D. magna* únicament s'ha estudiat mitjançant estudis de transcriptòmica (Heckmann et al. 2008b) i de genòmica (Kim et al. 2018). Tot i això, es coneix l'existència d'una possible relació entre els receptors de la serotonina i el metabolisme de l'AA i de les prostaglandines (Tournois, et al. 1998), les quals es coneix que regulen la reproducció en els insectes (Stanley 2006). En aquest context, Jordão, et al. (2016) van estudiar l'efecte de l'exposició de la *D. magna* a la fluoxetina, on es va determinar que es produïa un augment de l'activitat de la serotonina en aquest organisme. En l'**article científic V**, a causa de l'exposició a la fluoxetina, es va observar un increment dels nivells de l'oxilipina 12-HHTrE, la qual s'ha identificat com a producte de degradació de la prostaglandina H₂ (PGH₂) en la *D. magna*. Com a conseqüència dels resultats d'aquest treball, es suggereix que aquest fàrmac psiquiàtric pot incrementar, i per tant, regular la síntesi de les prostaglandines.

Per altra banda, tot i que el coneixement que es té sobre les rutes LOX i CYP és menor, aquestes rutes estan relacionades també amb el metabolisme de les oxilipines i amb la reproducció (Medeiros et al. 2002). A més, en un estudi anterior, Rivetti, et al. (2016) van observar una millora en la reproducció de la *D. magna* a causa de l'exposició als compostos farmacèutics com la fluoxetina, el diazepam i la carbamazepina. Per això, l'augment de la concentració en diversos productes de la ruta LOX com a resultat de la seva exposició al diazepam (veure **Taula 3.2**), i en menys mesura de l'exposició a la carbamazepina (13-HODE), pot explicar la influència d'aquests fàrmacs en les rutes de les oxilipines i, en conseqüència, en una possible regulació dels mecanismes de reproducció de la *D. magna*.

Taula 3.2 Oxilipines de la ruta enzimàtica LOX que presenten canvis significatius en les mostres exposades al diazepam.

AA	LA	α -LA	EPA
5-KETE	9-KODE	13-HOTrE	5-HEPE
5-HETE	9-HODE	9-KOTrE	
12-HETE		9-HOTrE	

3.4.2 ESTRATÈGIA DIRIGIDA I NO DIRIGIDA

L'aplicació de l'estratègia no dirigida permet la possibilitat de considerar tot el conjunt de metabòlits presents en el sistema biològic. En aquest cas, és necessària la utilització d'una instrumentació més selectiva i amb una gran capacitat de resolució la qual permeti la determinació d'un ventall molt ampli de metabòlits mitjançant un tractament de les dades exhaustiu. Així, tal com es mostra en l'**article científic VI**, l'aplicació de l'anàlisi no dirigida va permetre caracteritzar de manera extensa el perfil metabòlic de la *D. magna*. Una vegada s'han resolt i identificat els metabòlits rellevants a les exposicions als contaminants ambientals, la determinació de les rutes metabòliques afectades permet descriure millor els canvis que presenta el metaboloma de la *D. magna*. Concretament, mitjançant l'estratègia no dirigida es van poder determinar un nombre més extens de metabòlits, però no se'n va poder realitzar la seva

quantificació en termes absoluts ja que no es disposava de patrons estàndards de tots els metabòlits detectats. Tot i això, el tractament de les dades mitjançant el procediment ROIMCR va permetre realitzar una quantificació relativa en les diferents mostres analitzades (mostres control i mostres exposades), i sobretot avaluar els seus canvis de concentració en aquests dos tipus de mostres estudiades. D'aquesta manera, l'anàlisi conjunta de les dades corresponents a les dues concentracions de BPA estudiades va permetre determinar els canvis que aquest contaminant produeix en el metaboloma de la *D. magna*.

Per altra banda, l'anàlisi dirigida permet obtenir una quantificació de les concentracions dels metabòlits seleccionats, encara que això suposa deixar de tenir una perspectiva global del conjunt del metaboloma. Així, amb l'anàlisi dirigida es van determinar les concentracions absolutes (mitjançant un calibratge analític) dels metabòlits a causa de l'exposició a BPA a dues concentracions, amb l'inconvenient que únicament es van poder quantificar aquells metabòlits que s'han seleccionat prèviament i dels quals es disposen patrons estàndards. L'avaluació de les exposicions a BPA mitjançant les dues estratègies permet obtenir resultats complementaris, ja que d'aquesta manera s'aconsegueix una visió global del conjunt del metaboloma i es poden descobrir els potencials biomarcadors dels efectes a partir d'un disseny analític, el qual integra la identificació i la quantificació dels metabòlits alterats.

Els espectres d'alta resolució obtinguts en metabolòmica no dirigida mitjançant LC-HRMS generen una gran quantitat de dades i d'informació. Per això, és necessària una filtració i una reducció prèvia que permeti gestionar-les, i així poder-les analitzar i treure'n la màxima informació. D'aquesta manera, en l'**article científic VI** es realitza l'anàlisi de les dades mitjançant el mètode ROIMCR (Gorrochategui et al. 2016; Gorrochategui et al. 2019), el qual utilitza el procediment de selecció de les regions d'interès, ROI, juntament amb l'aplicació del mètode MCR-ALS (veure **secció 1.5.2** de la introducció de la Tesi, **capítol 1**). Aquest procediment s'ha aplicat en altres treballs per LC-HRMS (Gómez-Canela, et al. 2017), CE-HRMS (electroforesis capil·lar acoblada a l'espectrometria de masses d'alta resolució) (Ortiz-Villanueva et al. 2017a) o MSI (espectroscòpia d'imatges de MS) (Bedia, et al. 2017). L'aplicació de les ROIs va

comportar una gran reducció de la dimensió espectral sense cap pèrdua de resolució de m/z , ja que conserva la resolució espectral original de les dades, i d'aquesta manera elimina aquella informació no rellevant (veure secció 3.2 de l'**article científic VI**).

La utilització del procediment de les ROIs requereix la determinació de les condicions òptimes que permeten la identificació dels metabòlits presents en les mostres. Per això, és necessària l'optimització dels paràmetres de selecció de les ROIs. Per obtenir els senyals MS rellevants, es va optimitzar el valor llindar de la relació senyal/soroll fixat a uns valors de 0.1, 0.5 i 1 % del valor màxim d'intensitat de pics MS mesurats. També es va determinar l'error m/z de l'instrument on es van aplicar diferents valors des de 0.005, 0.010, 0.025 i 0.05 Da/e. A més, es va optimitzar el paràmetre que determina el nombre mínim de punts que defineixen un pic d'elució cromatogràfica en el procediment emprat, on es va provar amb 10, 20 i 25 punts per definir un pic. Finalment, es va seleccionar un valor llindar de les ROIs superior al 0.1% del valor màxim d'intensitat de pic per tal de filtrar entre les intensitats MS significatives i aconseguir una bona resolució dels pics i una bona precisió de la massa. A més, atès que cada una de les ROIs està relacionada a un possible perfil d'elució del metabòlit, es va considerar un nombre mínim de 20 punts consecutius de dades d'alta intensitat. Així, únicament es van considerar les ROIs amb una forma de pic cromatogràfica raonable. Una vegada realitzada la selecció de les ROIs es va aplicar el mètode MCR-ALS a les dades obtingudes, el qual va permetre resoldre de manera satisfactòria els perfils d'elució i els espectres de masses purs d'un gran nombre de metabòlits presents en la *D. magna*. El model de MCR-ALS obtingut va mostrar un percentatge de variància explicada (R^2) superior al 99 % i un percentatge de falta d'ajust (*lof*) inferior al 6.4 %. En concret, es van resoldre 97 components MCR quan s'analitzen conjuntament les mostres control i les exposades a BPA. No obstant això, cal comentar que no tots aquests components de MCR-ALS resolts poden ser assignats a metabòlits. Dels 97 components determinats, 42 van ésser identificats com a metabòlits i la resta representaven altres senyals de soroll que provenien de les contribucions desconegudes, del solvent o del soroll de fons. Aquest procediment ROIMCR és una eina eficaç per la resolució dels perfils d'elució i dels espectres purs dels metabòlits, tot i la complexitat inicial de les dades obtingudes per HRMS (múltiples coelucions i solapament de pics cromatogràfics). A més, permet

realitzar les estimacions dels canvis en les concentracions relatives dels metabòlits estudiats i llur identificació. Per tant, esdevé una eina molt útil per determinar els biomarcadors alterats per l'exposició a contaminants ambientals.

En l'**article científic VI** es van avaluar les concentracions dels metabòlits en mostres control i mostres exposades a una dosi baixa i alta de BPA. L'anàlisi dirigida de les mostres va permetre la quantificació d'aquestes variacions en els nivells de 23 metabòlits prèviament seleccionats. D'altra banda, a través de l'estratègia no dirigida es van poder determinar i identificar un total de 42 metabòlits. En aquest article, es va aplicar el mètode PCA a les concentracions autoescalades dels 23 metabòlits determinats en l'aproximació dirigida i les àrees de pic autoescalades dels 42 metabòlits identificats prèviament resolts pel mètode ROIMCR, amb l'aproximació no dirigida. En tots dos casos les mostres control i les mostres exposades a una dosi baixa de BPA mostren un comportament bastant similar, mentre que les mostres sotmeses a una dosi alta de BPA es distingeixen perfectament dels altres dos grups de mostres (control i de baixa concentració de BPA). Aquesta diferenciació de les mostres es va observar en el primer PC, el qual explicava un 26.69 % i un 26.94 % de la variància explicada de les dades en cada model, respectivament (veure Figura 3 de l'**article científic VI**). Tal com es mostra en la Figura 3 de l'**article científic VI**, els metabòlits que presenten *loadings* positius tenen unes concentracions més elevades en les mostres exposades a una dosi alta de BPA. Els metabòlits coincidents amb les dues estratègies són: la citrul·lina, la fenilalanina, el triptòfan, la tirosina, la valina, la uridina, la inosina i l'ADP. Cal remarcar que alguns dels metabòlits quantificats en l'estratègia dirigida: maltosa, trehalosa, mioinositol, galactosa, glucosa, àcid aspàrtic, metionina, serina, NADH, citidina, adenosina, guanosina, hipoxantina, àcid fumàric i àcid pimèlic, no es van poder determinar mitjançant el procediment ROIMCR a causa de la seva molt baixa concentració i del valor determinat com a llindar de la relació senyal/soroll en el procediment de selecció de les ROIs. Cal tenir en compte que l'optimització dels paràmetres per la selecció de les ROIs han de ser un compromís entre la sensibilitat del procediment i el nombre de metabòlits determinats.

Finalment, es van seleccionar els metabòlits determinats en les dues estratègies que presentaven un canvi (*fold change*) significatiu amb una exposició de dosi baixa i alta a BPA. En total, es van detectar 28 metabòlits que mostraven diferències significatives en les seves concentracions a través de les exposicions ($p < 0.05$, $p < 0.01$, $p < 0.005$). Entre les famílies dels metabòlits que van patir canvis es troben aminoàcids, àcids grassos, sucres, nucleòsids i el NADH (veure Figura 4 de l'**article científic VI**), els quals van permetre realitzar una interpretació biològica dels possibles efectes produïts pel BPA en el metaboloma de la *D. magna*. Per esbrinar les possibles rutes metabòliques afectades es va utilitzar la plataforma *online* MetaboAnalyst (veure **secció 1.6** de la introducció de la Tesi, **capítol 1**). Aquesta plataforma permet analitzar els metabòlits alterats en una anàlisi de l'enriquiment, la qual proporciona una interpretació exhaustiva de les rutes metabòliques involucrades. En concret, aquesta anàlisi indica que les alteracions produïdes pel BPA en el metaboloma de la *D. magna* estan relacionades amb el metabolisme del carboni (aminoàcids, urea i amoníac) i del glutamat, el qual juga un paper central en el metabolisme dels aminoàcids. En la **Taula 3.3** es mostra un resum dels metabòlits afectats segons si es tracta d'un increment o d'una reducció dels seus nivells en el metaboloma.

Taula 3.3| Metabòlits rellevants els nivells dels quals es veuen modificats a causa de l'exposició a BPA.

METABÒLITS	
↑ NIVELLS	Àcid aspàrtic, àcid glutàmic, glucosa.
↓ NIVELLS	AMP, NADH, serina, glutamat, ADP-ribosa.

Entre els efectes més destacats, s'observa que mentre les concentracions de glutamat disminueixen a causa de l'exposició, les concentracions dels aminoàcids augmenten. Això, concorda amb els resultats obtinguts en estudis previs, on es relacionen els increments dels nivells dels aminoàcids a causa de l'exposició a BPA amb la reducció de la síntesis de proteïnes, quan es sotmet la *D. magna* a condicions d'estrès (Nagato, et al. 2016). D'aquesta manera, els organismes exposats a BPA, generarien

una acumulació d'aminoàcids, sucres i lípids per tal d'accelerar els processos cel·lulars del metabolisme energètic, en detriment del creixement i la reproducció.

Els resultats obtinguts en els dos articles presentats en aquest capítol mostren els efectes que provoquen diversos contaminants ambientals en el metaboloma de la *D. magna*. Per una banda, l'anàlisi dirigida ha permès una quantificació absoluta de les alteracions produïdes en les oxilipines a causa de l'exposició a tres compostos farmacèutics, la fluoxetina, el diazepam i la carbamazepina, i elucidar com aquests canvis afecten a les prostaglandines. Per altra banda, les estratègies dirigida i no dirigida han caracteritzat de manera quantitativa i qualitativa el metaboloma de la *D. magna*, on han generat informació complementaria i més exhaustiva dels efectes del BPA. Així, s'ha realitzat una interpretació més detallada i quantitativa dels nivells dels metabòlits afectats per l'exposició a BPA a través de l'estratègia dirigida. I es confirma que l'anàlisi no dirigida, mitjançant l'aplicació del procediment ROIMCR, permet obtenir una informació global del perfil metabòlic i caracteritzar els canvis produïts en el metaboloma de la *D. magna* a causa de l'exposició al BPA.

Tal com s'ha mostrat en aquest capítol, cada una d'aquestes dues estratègies presenta una finalitat diferent que permet obtenir informació complementaria per la determinació dels efectes dels contaminants ambientals en el metaboloma. En la **Taula 3.4** es resumeixen les característiques específiques de les dues estratègies, les quals s'han anat descrivint al llarg d'aquest capítol.

Taula 3.4 | Característiques específiques de les dues estratègies: anàlisi dirigida i anàlisi no dirigida.

ANÀLISI DIRIGIDA	ANÀLISI NO DIRIGIDA
Quantificació absoluta	No quantificació absoluta
Patrons de referència	Identificació de metabòlits desconeguts
Metabòlits preseleccionats	Perfil global dels metabòlits
Software cases comercials	Tractament de les dades
Gran sensibilitat	Gran capacitat de resolució

CONCLUSIONS



En aquesta Tesi s'han avaluat les capacitats de diverses plataformes analítiques basades en la cromatografia i l'espectrometria de masses per realitzar una anàlisi completa del metaboloma de la *D. magna*. Tant la utilització de les estratègies analítiques dirigides com les no dirigides han permès explorar i avaluar els canvis produïts en els perfils metabòlics de la *D. magna* a causa de l'exposició a diferents estressants ambientals, factors abiòtics i contaminants ambientals.

A continuació, es detallen les conclusions específiques en referència a les estratègies analítiques i quimiomètriques emprades, i els efectes toxicològics dels estressants ambientals avaluats en aquesta Tesi:

ESTRATÈGIES ANALÍTQUES

- S'ha demostrat el gran potencial de les tècniques de cromatografia de gasos unidimensional i bidimensional acoblades a espectrometria de masses per realitzar estudis de metabolòmica no dirigida. L'optimització de les condicions analítiques ha permès la determinació i la identificació dels metabòlits polars presents en les mostres de *D. magna*. En concret, s'ha determinat un procediment d'extracció amb unes proporcions de cloroform, metanol i aigua (2:1:2) adequades per a l'anàlisi dels metabòlits presents en les mostres biològiques. A més, la utilització d'un procés de derivatització amb unes condicions de MeOX seguides de l'addició de MSTFA en TMCS ha permès obtenir molècules tèrmicament estables per a la seva detecció mitjançant GC.
- La metodologia bidimensional de GCxGC-MS proposada en aquesta Tesi ha estat capaç de resoldre mostres de metabolòmica complexes d'una forma més completa que la metodologia unidimensional GC-MS. L'avaluació de diferents configuracions de columnes mitjançant GCxGC-MS ha permès determinar unes condicions cromatogràfiques idònies per obtenir un perfil metabòlic més detallat de la *D. magna*.
- L'aplicació de les estratègies dirigides i no dirigides han permès un estudi exhaustiu del metaboloma de la *D. magna* a través de dues perspectives analítiques complementàries. La combinació de les dues estratègies, dirigides i no dirigides,

ha permès una caracterització més completa del conjunt dels metabòlits de la *D. magna*. Per una banda, l'ús de l'estratègia dirigida on s'han optimitzat les condicions analítiques LC-MS ha permès l'anàlisi quantitativa dels metabòlits prèviament escollits. En canvi, en els estudis de metabolòmica no dirigida s'obté un perfil global qualitatiu del metaboloma de la *D. magna*, a la vegada que permet avaluar els efectes dels contaminants ambientals sobre aquest metaboloma a partir dels canvis de concentració relativa dels metabòlits observats.

- L'aplicació de l'aproximació dirigida mitjançant LC-MS ha permès una interpretació quantitativa acurada dels baixos nivells de concentracions de les oxilipines en les mostres de *D. magna*, on es van determinar i quantificar 28 metabòlits derivats dels PUFAs.

ESTRATÈGIES QUIMIOMÈTRIQUES

- La utilització del mètode MCR-ALS ha permès resoldre un gran nombre de metabòlits en l'anàlisi de mostres complexes a partir d'estratègies no dirigides. S'ha confirmat que és un mètode robust pel processament de les dades generades en estudis d'òmica ambiental mitjançant GC-MS i GCxGC-MS.
- Es proposa la utilització dels mètodes de pretractament per la compressió de les dades a partir de procediments basats en les finestres de temps i en les transformades d'ones pels grans volums de dades generades en les anàlisis per GC-MS i GCxGC-MS, respectivament. En concret, aquests procediments de pretractament han permès filtrar, comprimir i reduir la complexitat de les dades en la direcció del temps cromatogràfic. Així, s'ha facilitat l'aplicació del procediment MCR-ALS per resoldre els metabòlits presents en la *D. magna*.
- En el cas d'emprar dades de gran resolució espectral com ara en les anàlisis no dirigides LC-HRMS, la combinació de les metodologies de selecció de les regions d'interès, ROIs, i de la resolució multivariant mitjançant el mètode MCR-ALS, ha facilitat la detecció, resolució i identificació d'un elevat nombre de metabòlits del perfil metabòlic de la *D. magna*. L'aplicació de la metodologia combinada ROIMCR ha permès el filtratge, la compressió i la resolució de les dades de metabolòmica

no dirigida LC-HRMS mantenint l'exactitud en la resolució espectral de les dades originals d'alta resolució (HRMS).

EFFECTES TOXICOLÒGICS DELS ESTRESSANTS AMBIENTALS

- Les exposicions individuals a la temperatura i la hipòxia no han resultat produir efectes amb tanta intensitat en el metaboloma de la *D. magna* com l'augment de la salinitat en el medi. A més, a partir d'un disseny experimental adequat ha estat possible l'avaluació dels efectes de diferents estressants abiòtics i les seves possibles interaccions. Es conclou que els efectes de la temperatura i la hipòxia queden emmascarats per les variacions naturals de l'organisme i per l'efecte predominant que la salinitat produeix en el metabolisme de la *D. magna*. Aquest factor abiòtic té un gran impacte en un elevat nombre de metabòlits relacionats amb el metabolisme dels aminoàcids, de les proteïnes, dels sucres, dels àcids grassos i amb el metabolisme energètic, com el cicle de Krebs. Entre els metabòlits més afectats es troben l'àcid làctic i la glucosa, que estan relacionats amb la glicòlisi.
- A causa de l'exposició a fluoxetina els nivells de les oxilipines EKODE, 12-HETE, 9(10)-EpOME i 12(13)-EpODE es van veure alterats, mentre que la carbamazepina va afectar l'EKODE i la 15-HETE. En canvi, el diazepam va afectar els nivells de diversos derivats de l'àcid araquidònic, l'àcid linoleic, l'àcid α -linoleic i l'àcid eicosapentaenoic. S'ha observat la influència de les exposicions a la fluoxetina, el diazepam i la carbamazepina en les rutes de les oxilipines i, en conseqüència, la possible regulació dels mecanismes de reproducció de la *D. magna*.
- Les alteracions produïdes a causa de l'exposició a bisfenol A en el metaboloma de la *D. magna* estan relacionades amb el metabolisme del carboni, en concret els aminoàcids, la urea i l'amoníac, i amb una disminució dels nivells de glutamat el qual juga un paper central en el metabolisme dels aminoàcids. D'aquesta manera, s'ha demostrat que els organismes exposats a bisfenol A generen una acumulació d'aminoàcids, producte de la reducció dels nivells de glutamat, i un increment dels nivells de sucres i lípids que acceleren les funcions cel·lulars en detriment dels processos fisiològics com el creixement i la reproducció.

REFERÈNCIES

ADEMA, D. M. M. *Daphnia magna* as a test animal in acute and chronic toxicity tests. *Hydrobiologia*, 1978, 59(2), 125-134.

ALMSTETTER, M. F., OEFNER, P. J. AND DETTMER, K. Comprehensive two-dimensional gas chromatography in metabolomics. *Analytical and Bioanalytical Chemistry*, 2012, 402(6), 1993-2013.

ALTMAIER, E., RAMSAY, S. L., GRABER, A., MEWES, H. W., WEINBERGER, K. M. AND SUHRE, K. Bioinformatics analysis of targeted metabolomics - uncovering old and new tales of diabetic mice under medication. *Endocrinology*, 2008, 149(7), 3478-3489.

ALTSHULER, I., DEMIRI, B., XU, S., CONSTANTIN, A., YAN, N. D. AND CRISTESCU, M. E. An integrated multi-disciplinary approach for studying multiple stressors in freshwater ecosystems: *Daphnia* as a model organism. *Integrative and Comparative Biology*, 2011, 51(4), 623-633.

BALGOMA, D., CHECA, A., SAR, D. G., SNOWDEN, S. AND WHEELOCK, C. E. Quantitative metabolic profiling of lipid mediators. *Molecular Nutrition & Food Research*, 2013, 57(8), 1359-1377.

BARATA, C., NAVARRO, J. C., VARO, I., RIVA, M. C., ARUN, S. AND PORTE, C. Changes in antioxidant enzyme activities, fatty acid composition and lipid peroxidation in *Daphnia magna* during the aging process. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 2005, 140(1), 81-90.

BARKER, M. AND RAYENS, W. Partial least squares for discrimination. *Journal of Chemometrics*, 2003, 17(3), 166-173.

BARTEL, J., KRUMSIEK, J. AND THEIS, F. J. Statistical methods for the analysis of high-throughput metabolomics data. *Computational and Structural Biotechnology Journal*, 2013, 4(5).

BEDIA, C., TAULER, R. AND JAUMOT, J. Analysis of multiple mass spectrometry images from different *phaseolus vulgaris* samples by multivariate curve resolution. *Talanta*, 2017, 175, 557-565.

BEGOU, O., GIKA, H., WILSON, I. AND THEODORIDIS, G. Hyphenated MS-based targeted approaches in metabolomics. *Analyst*, 2017a, 142(17), 3079-3100.

BEGOU, O., GIKA, H. G., WILSON, I. D. AND THEODORIDIS, G. Hyphenated MS-based targeted approaches in metabolomics. *Analyst*, 2017b, 142(17), 3079-3100.

BENNETT, D. A. AND WATERS, M. D. Applying biomarker research. *Environmental Health Perspectives*, 2000, 108(9), 907-910.

BENSKIN, J. P., IKONOMOU, M. G., LIU, J., VELDHOEN, N., DUBETZ, C., HELBING, C. C. AND COSGROVE, J. R. Distinctive metabolite profiles in in-migrating sockeye salmon suggest sex-linked endocrine perturbation. *Environmental Science and Technology*, 2014, 48(19), 11670-11678.

BILES, J. E., MCNEAL, T. P., BEGLEY, T. H. AND HOLLIFIELD, H. C. Determination of bisphenol-a in reusable polycarbonate food-contact plastics and migration to food-simulating liquids. *Journal of Agricultural and Food Chemistry*, 1997, 45(9), 3541-3544.

BONNEFILLE, B., GOMEZ, E., ALALI, M., ROSAIN, D., FENET, H. AND COURANT, F. Metabolomics assessment of the effects of diclofenac exposure on *mytilus galloprovincialis*: Potential effects on osmoregulation and reproduction. *Science of the Total Environment*, 2018, 613-614, 611-618.

BOUDAH, S., PARIS, A. AND JUNOT, C. Liquid chromatography coupled to mass spectrometry-based metabolomics and the concept of biomarker. 2013. *Advances in Botanical Research*.

BOXSHALL, G. *Daphnia magna* straus, 1820. Accessed through: World Register of Marine Species at <http://www.marinespecies.org/aphia.php?p=taxdetails&id=148372>, 2013.

BRENNAN, S. J., BROUGHAM, C. A., ROCHE, J. J. AND FOGARTY, A. M. Multi-generational effects of four selected environmental oestrogens on *Daphnia magna*. *Chemosphere*, 2006, 64(1), 49-55.

BRO, R. PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, 1997, 38(2), 149-171.

BRO, R. AND DE JONG, S. A fast non-negativity-constrained least squares algorithm. *Journal of Chemometrics*, 1997, 11(5), 393-401.

BROWN, S. C., KRUPPA, G. AND DASSEUX, J. L. Metabolomics applications of FT-ICR mass spectrometry. *Mass Spectrometry Reviews*, 2005, 24(2), 223-231.

BUNDY, J. G., DAVEY, M. P. AND VIANT, M. R. Environmental metabolomics: A critical review and future perspectives. *Metabolomics*, 2009, 5(1), 3-21.

BYLESJÖ, M., RANTALAINEN, M., CLOAREC, O., NICHOLSON, J. K., HOLMES, E. AND TRYGG, J. OPLS discriminant analysis: Combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics*, 2006, 20(8-10), 341-351.

CALISTO, V. AND ESTEVES, V. I. Psychiatric pharmaceuticals in the environment. *Chemosphere*, 2009, 77(10), 1257-1274.

CAMPOS, B., GARCIA-REYERO, N., RIVETTI, C., ESCALON, L., HABIB, T., TAULER, R., TSAKOVSKI, S., PIÑA, B. AND BARATA, C. Identification of metabolic pathways in *Daphnia magna* explaining hormetic effects of selective serotonin reuptake inhibitors and 4-nonylphenol using transcriptomic and phenotypic responses. *Environmental Science & Technology*, 2013, 47(16), 9434-9443.

CAMPOS, B., PIÑA, B. AND BARATA, C. Mechanisms of action of selective serotonin reuptake inhibitors in *Daphnia magna*. *Environmental Science & Technology*, 2012, 46(5), 2943-2950.

CARNEIRO, S., PEREIRA, R. AND ROCHA, I. Yeast metabolomics: Sample preparation for a GC/MS-based analysis. 2014. *Methods in Molecular Biology*.

CHECA, A., HOLM, T., SJÖDIN, M. O., REINKE, S. N., ALM, J., SCHEYNIUS, A. AND WHEELLOCK, C. E. Lipid mediator profile in *vernix caseosa* reflects skin barrier development. *Scientific reports*, 2015, 5, 15740.

CHEN, S., KONG, H., LU, X., LI, Y., YIN, P., ZENG, Z. AND XU, G. Pseudotargeted metabolomics method and its application in serum biomarker discovery for hepatocellular carcinoma based on ultra high-performance liquid chromatography/triple quadrupole mass spectrometry. *Analytical Chemistry*, 2013, 85(17), 8326-8333.

CHONG, I. G. AND JUN, C. H. Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, 2005, 78(1), 103-112.

CHRISTIN, C., SMILDE, A. K., HOEFSLOOT, H. C. J., SUITS, F., BISCHOFF, R. AND HORVATOVICH, P. L. Optimized time alignment algorithm for LC-MS data: Correlation optimized warping using component detection algorithm-selected mass chromatograms. *Analytical Chemistry*, 2008, 80(18), 7012-7021.

CLARKE, C. J. AND HASELDEN, J. N. Metabolic profiling as a tool for understanding mechanisms of toxicity. *Toxicologic Pathology*, 2008, 36(1), 140-147.

CLENDINEN, C. S., MONGE, M. E. AND FERNÁNDEZ, F. M. Ambient mass spectrometry in metabolomics. *Analyst*, 2017, 142(17), 3101-3117.

COLBOURNE, J. K., HEBERT, P. D. AND TAYLOR, D. J. Evolutionary origins of phenotypic diversity in *daphnia*. *Molecular evolution and adaptive radiation*, 1997, 163-188.

COMBER, M. H. I., WILLIAMS, T. D. AND STEWART, K. M. The effects of nonylphenol on *Daphnia magna*. *Water Research*, 1993, 27(2), 273-276.

COURANT, F., ARPIN-PONT, L., BONNEFILLE, B., VACHER, S., PICOT-GROZ, M., GOMEZ, E. AND FENET, H. Exposure of marine mussels to diclofenac: Modulation of prostaglandin biosynthesis. *Environmental Science and Pollution Research*, 2018, 25(7), 6087-6094.

DALMAU, N., BEDIA, C. AND TAULER, R. Validation of the regions of interest multivariate curve resolution (ROIMCR) procedure for untargeted LC-MS lipidomic analysis. *Analytica Chimica Acta*, 2018, 1025, 80-91.

DE GELAS, K. AND DE MEESTER, L. Phylogeography of *Daphnia magna* in europe. *Molecular Ecology*, 2005, 14(3), 753-764.

DE JUAN, A., JAUMOT, J. AND TAULER, R. Multivariate curve resolution (MCR). Solving the mixture analysis problem. *Analytical Methods*, 2014, 6(14), 4964-4976.

DE JUAN, A. AND TAULER, R. Comparison of three-way resolution methods for non-trilinear chemical data sets. *Journal of Chemometrics*, 2001, 15(10), 749-771.

DE JUAN, A. AND TAULER, R. Multivariate curve resolution (MCR) from 2000: Progress in concepts and applications. *Critical Reviews in Analytical Chemistry*, 2006, 36(3-4), 163-176.

DORMAN, F. L., OVERTON, E. B., WHITING, J. J., COCHRAN, J. W. AND GARDEA-TORRESDEY, J. Gas chromatography. *Analytical Chemistry*, 2008, 80(12), 4487-4497.

DUNN, W. B. AND ELLIS, D. I. Metabolomics: Current analytical platforms and methodologies. *TrAC - Trends in Analytical Chemistry*, 2005, 24(4), 285-294.

EBERT, D. The effect of size at birth, maturation threshold and genetic differences on the life-history of *Daphnia magna*. *Oecologia*, 1991, 86(2), 243-250.

EBERT, D. Ecology, epidemiology, and evolution of parasitism in *daphnia*. Bethesda (MD): National Center for Biotechnology Information (US). 2005, Chapter 2, Introduction to *Daphnia* Biology.

ELIE, M. R., CHOI, J., NKRUMAH-ELIE, Y. M., GONNERMAN, G. D., STEVENS, J. F. AND TANGUAY, R. L. Metabolomic analysis to define and compare the effects of pahs and oxygenated pahs in developing zebrafish. *Environmental Research*, 2015, 140, 502-510.

ESBENSEN, K. H. AND GELADI, P. Principal component analysis: Concept, geometrical interpretation, mathematical background, algorithms, history, practice. In *Comprehensive chemometrics*. 2010, vol. 2, p. 211-226.

ESCOBAR-ARNANZ, J., MEKNI, S., BLANCO, G., ELJARRAT, E., BARCELÓ, D. AND RAMOS, L. Characterization of organic aromatic compounds in soils affected by an uncontrolled tire landfill fire through the use of comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry. *Journal of Chromatography A*, 2018, 1536, 163-175.

FALANGA, A., MERCURIO, F. A., SICILIANO, A., LOMBARDI, L., GALDIERO, S., GUIDA, M., LIBRALATO, G., LEONE, M. AND GALDIERO, E. Metabolomic and oxidative effects of quantum dots-indolicidin on three generations of *Daphnia magna*. *Aquatic Toxicology*, 2018, 198, 158-164.

FALCIANI, F., DIAB, A. M., SABINE, V., WILLIAMS, T. D., ORTEGA, F., GEORGE, S. G. AND CHIPMAN, J. K. Hepatic transcriptomic profiles of european flounder (*platichthys flesus*) from field sites and computational approaches to predict site from stress gene responses following exposure to model toxicants. *Aquatic Toxicology*, 2008, 90(2), 92-101.

FENT, K., WESTON, A. A. AND CAMINADA, D. Ecotoxicology of human pharmaceuticals. *Aquatic Toxicology*, 2006, 76(2), 122-159.

FICKE, A. D., MYRICK, C. A. AND HANSEN, L. J. Potential impacts of global climate change on freshwater fisheries. *Reviews in Fish Biology and Fisheries*, 2007, 17(4), 581-613.

FIEHN, O. Metabolomics – the link between genotypes and phenotypes. *Plant Molecular Biology*, 2002, 48(1), 155-171.

FIEHN, O., KOPKA, J., DÖRMANN, P., ALTMANN, T., TRETHERWEY, R. N. AND WILLMITZER, L. Metabolite profiling for plant functional genomics. *Nature Biotechnology*, 2000, 18(11), 1157-1161.

FLAHERTY, C. M. AND DODSON, S. I. Effects of pharmaceuticals on *Daphnia* survival, growth, and reproduction. *Chemosphere*, 2005, 61(2), 200-207.

FOCANT, J.-F., SJÖDIN, A. AND PATTERSON JR, D. G. Improved separation of the 209 polychlorinated biphenyl congeners using comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry. *Journal of Chromatography A*, 2004, 1040(2), 227-238.

FONG, P. P. AND FORD, A. T. The biological effects of antidepressants on the molluscs and crustaceans: A review. *Aquatic Toxicology*, 2014, 151, 4-13.

GHASTE, M., MISTRİK, R. AND SHULAEV, V. Applications of fourier transform ion cyclotron resonance (FT-ICR) and orbitrap based high resolution mass spectrometry in metabolomics and lipidomics. *International Journal of Molecular Sciences*, 2016, 17(6).

GIKA, H., VIRGILIOU, C., THEODORIDIS, G., PLUMB, R. S. AND WILSON, I. D. Untargeted LC/MS-based metabolic phenotyping (metabonomics/metabolomics): The state of the art. *Journal of Chromatography B*, 2019, 1117, 136-147.

GIL-SOLSONA, R., NÁCHER-MESTRE, J., LACALLE-BERGERON, L., SANCHO, J. V., CALDUCH-GINER, J. A., HERNÁNDEZ, F. AND PÉREZ-SÁNCHEZ, J. Untargeted metabolomics approach for unraveling robust biomarkers of nutritional status in fasted gilthead sea bream (*sparus aurata*). *PeerJ*, 2017, 2017(2).

GOLUB, G. H. AND REINSCH, C. Singular value decomposition and least squares solutions. *Numerische Mathematik*, 1970, 14(5), 403-420.

GÓMEZ-CANELA, C., MILLER, T. H., BURY, N. R., TAULER, R. AND BARRON, L. P. Targeted metabolomics of *gammarus pulex* following controlled exposures to selected pharmaceuticals in water. *Science of the Total Environment*, 2016, 562, 777-788.

GÓMEZ-CANELA, C., PRATS, E., PIÑA, B. AND TAULER, R. Assessment of chlorpyrifos toxic effects in zebrafish (*danio rerio*) metabolism. *Environmental Pollution*, 2017, 220, 1231-1243.

GÓMEZ-CANELA, C., TORNERO-CAÑADAS, D., PRATS, E., PIÑA, B., TAULER, R. AND RALDÚA, D. Comprehensive characterization of neurochemicals in three zebrafish chemical models of human acute organophosphorus poisoning using liquid chromatography-tandem mass spectrometry. *Analytical and Bioanalytical Chemistry*, 2018, 1-14.

GORROCHATEGUI, E., CASAS, J., PORTE, C., LACORTE, S. AND TAULER, R. Chemometric strategy for untargeted lipidomics: Biomarker detection and identification in stressed human placental cells. *Analytica Chimica Acta*, 2015a, 854, 20-33.

GORROCHATEGUI, E., JAUMOT, J., LACORTE, S. AND TAULER, R. Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow. *TrAC Trends in Analytical Chemistry*, 2016, 82, 425-442.

GORROCHATEGUI, E., JAUMOT, J. AND TAULER, R. A protocol for LC-MS metabolomic data processing using chemometric tools 2015b.

GORROCHATEGUI, E., JAUMOT, J. AND TAULER, R. ROIMCR: A powerful analysis strategy for LC-MS metabolomic datasets. *BMC Bioinformatics*, 2019, 20(1), 256.

GRIFFITHS, W. J., KOAL, T., WANG, Y., KOHL, M., ENOT, D. P. AND DEIGNER, H. P. Targeted metabolomics for biomarker discovery. *Angewandte Chemie - International Edition*, 2010, 49(32), 5426-5445.

GUIJAS, C., MONTENEGRO-BURKE, J. R., DOMINGO-ALMENARA, X., PALERMO, A., WARTH, B., HERMANN, G., KOELLENSPERGER, G., HUAN, T., URITBOONTHAI, W., AISPORN, A. E., WOLAN, D. W., SPILKER, M. E., BENTON, H. P. AND SIUZDAK, G. Metlin: A technology platform for identifying knowns and unknowns. *Analytical Chemistry*, 2018, 90(5), 3156-3164.

GULLBERG, J., JONSSON, P., NORDSTRÖM, A., SJÖSTRÖM, M. AND MORITZ, T. Design of experiments: An efficient strategy to identify factors influencing extraction and derivatization of *arabidopsis thaliana* samples in metabolomic studies with gas chromatography/mass spectrometry. *Analytical Biochemistry*, 2004, 331(2), 283-295.

GÜNDEL, U., KALKHOF, S., ZITZKAT, D., VON BERGEN, M., ALTENBURGER, R. AND KÜSTER, E. Concentration-response concept in ecotoxicoproteomics: Effects of different phenanthrene concentrations to the zebrafish (*danio rerio*) embryo proteome. *Ecotoxicology and Environmental Safety*, 2012, 76(1), 11-22.

GUNNARSSON, L., JAUHAINEN, A., KRISTIANSSON, E., NERMAN, O. AND LARSSON, D. G. J. Evolutionary conservation of human drug targets in organisms used for environmental risk assessments. *Environmental Science & Technology*, 2008, 42(15), 5807-5813.

HALKET, J. M., WATERMAN, D., PRZYBOROWSKA, A. M., PATEL, R. K. P., FRASER, P. D. AND BRAMLEY, P. M. Chemical derivatization and mass spectral libraries in metabolic profiling by GC-MS and LC-MS/MS. *Journal of Experimental Botany*, 2005, 56(410), 219-243.

HANIFAH, A., MAHARIJAYA, A., PUTRI, P. S., LAVIÑA, A. W. AND SOBIR Untargeted metabolomics analysis of eggplant (*solanum melongena*) fruit and its correlation to fruit morphologies. *Metabolites*, 2018, 8(3).

HEATH, A. G. *Water pollution and fish physiology*. Edition ed.: CRC press, 1995. ISBN 0873716329.

HEBERT, P. D. N. The population biology of *Daphnia* (crustacea, daphnidae). *Biological Reviews*, 1978, 53(3), 387-426.

HECKMANN, L. H., SIBLY, R. M., CONNON, R., HOOPER, H. L., HUTCHINSON, T. H., MAUND, S. J., HILL, C. J., BOUETARD, A. AND CALLAGHAN, A. Systems biology meets stress ecology: Linking molecular and organismal stress responses in *Daphnia magna*. *Genome Biology*, 2008a, 9(2).

HECKMANN, L. H., SIBLY, R. M., TIMMERMANS, M. J. AND CALLAGHAN, A. Outlining eicosanoid biosynthesis in the crustacean *daphnia*. *Frontiers in Zoology*, 2008b, 5.

HENDRIKS, M. M. W. B., EEUWIJK, F. A., JELLEMA, R. H., WESTERHUIS, J. A., REIJMERS, T. H., HOEFSLOOT, H. C. J. AND SMILDE, A. K. Data-processing strategies for metabolomics studies. *TrAC - Trends in Analytical Chemistry*, 2011, 30(10), 1685-1698.

HIBBERT, D. B. Experimental design in chromatography: A tutorial review. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 2012, 910, 2-13.

HOEFSLOOT, H. C. J., VIS, D. J., WESTERHUIS, J. A., SMILDE, A. K. AND JANSEN, J. J. Multiset data analysis: ANOVA simultaneous component analysis and related methods. In *Comprehensive chemometrics*. 2010, vol. 2, p. 453-472.

HOPE, J. L., PRAZEN, B. J., NILSSON, E. J., LIDSTROM, M. E. AND SYNOVEC, R. E. Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection: Analysis of amino acid and organic acid trimethylsilyl derivatives, with application to the analysis of metabolites in *rye grass* samples. *Talanta*, 2005, 65(2), 380-388.

HOWLETT, R. M., DAVEY, M. P. AND KELLY, D. J. Metabolomic analysis of *campylobacter jejuni* by direct-injection electrospray ionization mass spectrometry. 2017. *Methods in Molecular Biology*.

HUMMEL, J., STREHMEL, N., BÖLLING, C., SCHMIDT, S., WALTHER, D. AND KOPKA, J. Mass spectral search and analysis using the Golm metabolome database. In *The handbook of plant metabolomics*. Wiley-VCH Verlag GmbH & Co. KGaA, 2013, p. 321-343.

HURTADO, C., PARASTAR, H., MATAMOROS, V., PIÑA, B., TAULER, R. AND BAYONA, J. M. Linking the morphological and metabolomic response of *lactuca sativa* exposed to emerging contaminants using GC × GC-MS and chemometric tools. *Scientific Reports*, 2017, 7(1), 6546.

JANSEN, J. J., HOEFSLOOT, H. C. J., GREEF, J. V. D., TIMMERMAN, M. E., WESTERHUIS, J. A. AND SMILDE, A. K. ASCA: Analysis of multivariate data obtained from an experimental design. *Journal of Chemometrics*, 2005, 19(9), 469-481.

JAUMOT, J., GARGALLO, R., DE JUAN, A. AND TAULER, R. A graphical user-friendly interface for MCR-ALS: A new tool for multivariate curve resolution in MATLAB. *Chemometrics and Intelligent Laboratory Systems*, 2005, 76(1), 101-110.

JENKINS, S., SWENSON, T. L., LAU, R., ROCHA, A. M., AARING, A., HAZEN, T. C., CHAKRABORTY, R. AND NORTHEN, T. R. Construction of viable soil defined media using quantitative metabolomics analysis of soil metabolites. *Frontiers in Microbiology*, 2017, 8.

JEONG, T. Y., YOON, D., KIM, S., KIM, H. Y. AND KIM, S. D. Mode of action characterization for adverse effect of propranolol in *Daphnia magna* based on behavior and physiology monitoring and metabolite profiling. *Environmental Pollution*, 2018, 233, 99-108.

JI, C., WEI, L., ZHAO, J. AND WU, H. Metabolomic analysis revealed that female mussel *mytilus galloprovincialis* was sensitive to bisphenol a exposures. *Environmental Toxicology and Pharmacology*, 2014, 37(2), 844-849.

JONES, O. A. H., SWAIN, S. C., SVENDSEN, C., GRIFFIN, J. L., STURZENBAUM, S. R. AND SPURGEON, D. J. Potential new method of mixture effects testing using metabolomics and *caenorhabditis elegans*. *Journal of Proteome Research*, 2012, 11(2), 1446-1453.

JORDÃO, R., GARRETA, E., CAMPOS, B., LEMOS, M. F. L., SOARES, A. M. V. M., TAULER, R. AND BARATA, C. Compounds altering fat storage in *Daphnia magna*. *Science of The Total Environment*, 2016, 545–546, 127-136.

KALIVODOVÁ, A., HRON, K., FILZMOSE, P., NAJDEKR, L., JANEČKOVÁ, H. AND ADAM, T. PLC-DA for compositional data with application to metabolomics. *Journal of Chemometrics*, 2015, 29(1), 21-28.

KANANI, H., CHRYSANTHOPOULOS, P. K. AND KLAPA, M. I. Standardizing GC-MS metabolomics. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 2008, 871(2), 191-201.

KANANI, H. H. AND KLAPA, M. I. Data correction strategy for metabolomics analysis using gas chromatography-mass spectrometry. *Metabolic Engineering*, 2007, 9(1), 39-51.

KANEHISA, M., SATO, Y., KAWASHIMA, M., FURUMICHI, M. AND TANABE, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 2016, 44(D1), D457-D462.

KANERVA, M., VEHMAS, A., NIKINMAA, M. AND VUORI, K. A. Spatial variation in transcript and protein abundance of atlantic salmon during feeding migration in the baltic sea. *Environmental Science and Technology*, 2014, 48(23), 13969-13977.

KARAMAN, I. Preprocessing and pretreatment of metabolomics data for statistical analysis. 2017. *Advances in Experimental Medicine and Biology*.

KARIUKI, M. N., NAGATO, E. G., LANKADURAI, B. P., SIMPSON, A. J. AND SIMPSON, M. J. Analysis of sub-lethal toxicity of perfluorooctane sulfonate (PFOS) to *Daphnia magna* using ¹H nuclear magnetic resonance-based metabolomics. *Metabolites*, 2017, 7(2).

KATO, H., IZUMI, Y., HASUNUMA, T., MATSUDA, F. AND KONDO, A. Widely targeted metabolic profiling analysis of yeast central metabolites. *Journal of Bioscience and Bioengineering*, 2012, 113(5), 665-673.

KHOOMRUNG, S., MARTINEZ, J. L., TIPPMANN, S., JANS-ARD, S., BUFFING, M. F., NICASTRO, R. AND NIELSEN, J. Expanded metabolite coverage of *saccharomyces cerevisiae* extract through improved chloroform/methanol extraction and tert-butyltrimethylsilyl derivatization. *Analytical Chemistry Research*, 2015, 6, 9-16.

KIM, H.-Y. Analysis of variance (ANOVA) comparing means of more than two groups. *Restorative dentistry & endodontics*, 2014, 39(1), 74-77.

KIM, H. J., KOEDRITH, P. AND SEO, Y. R. Ecotoxicogenomic approaches for understanding molecular mechanisms of environmental chemical toxicity using aquatic invertebrate, *Daphnia* model organism. *International Journal of Molecular Sciences*, 2005, 16(6):12261-87.

KIM, R.-O., JO, M.-A., SONG, J., KIM, I.-C., YOON, S. AND KIM, W.-K. Novel approach for evaluating pharmaceuticals toxicity using *Daphnia* model: Analysis of the mode of *cytochrome p450*-generated metabolite action after acetaminophen exposure. *Aquatic Toxicology*, 2018, 196, 35-42.

KIM, T. K. T test as a parametric statistic. *Korean Journal of Anesthesiology*, 2015, 68(6), 540-546.

KIND, T., WOHLGEMUTH, G., LEE, D. Y., LU, Y., PALAZOGLU, M., SHAHBAZ, S. AND FIEHN, O. Fiehnlib: Mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Analytical Chemistry*, 2009, 81(24), 10038-10048.

KOEK, M. M., JELLEMA, R. H., VAN DER GREEF, J., TAS, A. C. AND HANKEMEIER, T. Quantitative metabolomics based on gas chromatography mass spectrometry: Status and perspectives. *Metabolomics*, 2011a, 7(3), 307-328.

KOEK, M. M., MUILWIJK, B., VAN STEE, L. L. P. AND HANKEMEIER, T. Higher mass loadability in comprehensive two-dimensional gas chromatography–mass spectrometry for improved analytical performance in metabolomics analysis. *Journal of Chromatography A*, 2008, 1186(1), 420-429.

KOEK, M. M., VAN DER KLOET, F. M., KLEEMANN, R., KOOISTRA, T., VERHEIJ, E. R. AND HANKEMEIER, T. Semi-automated non-target processing in GC × GC-MS metabolomics analysis: Applicability for biomedical studies. *Metabolomics : Official journal of the Metabolomic Society*, 2011b, 7(1), 1-14.

KOPKA, J., SCHAUER, N., KRUEGER, S., BIRKEMEYER, C., USADEL, B., BERGMÜLLER, E., DÖRMANN, P., WECKWERTH, W., GIBON, Y., STITT, M., WILLMITZER, L., FERNIE, A. R. AND STEINHAUSER, D. GMD@CSB.DB: The Golm metabolome database. *Bioinformatics*, 2005, 21(8), 1635-1638.

KOVACEVIC, V., SIMPSON, A. AND SIMPSON, M. Evaluation of *Daphnia magna* metabolic responses to organic contaminant exposure with and without dissolved organic matter using ^1H nuclear magnetic resonance (NMR)-based metabolomics. *Ecotoxicology and Environmental Safety*, 2018a, 164, 189-200.

KOVACEVIC, V., SIMPSON, A. AND SIMPSON, M. Investigation of *Daphnia magna* sub-lethal exposure to organophosphate esters in the presence of dissolved organic matter using ^1H NMR-based metabolomics. *Metabolites*, 2018b, 8(2), 34.

KOVACEVIC, V., SIMPSON, A. J. AND SIMPSON, M. J. ^1H NMR-based metabolomics of *Daphnia magna* responses after sub-lethal exposure to triclosan, carbamazepine and ibuprofen. *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics*, 2016, 19, 199-210.

KVALHEIM, O. M. AND KARSTANG, T. V. Interpretation of latent-variable regression models. *Chemometrics and Intelligent Laboratory Systems*, 1989, 7(1-2), 39-51.

LAWSON, C. L. AND HANSON, R. J. *Solving least squares problems*. Edition ed.: Siam, 1995. ISBN 0898713560.

LI, L., WU, H., JI, C., VAN GESTEL, C. A. M., ALLEN, H. E. AND PEIJNENBURG, W. J. G. M. A metabolomic study on the responses of *Daphnia magna* exposed to silver nitrate and coated silver nanoparticles. *Ecotoxicology and Environmental Safety*, 2015, 119, 66-73.

LIEBEKE, M. AND BUNDY, J. G. Tissue disruption and extraction methods for metabolic profiling of an invertebrate sentinel species. *Metabolomics*, 2012, 8(5), 819-830.

LINDON, J. C., HOLMES, E. AND NICHOLSON, J. K. Chapter 19 - global systems biology through integration of "omics" results. In *The handbook of metabonomics and metabolomics*. Amsterdam: Elsevier Science B.V., 2007, p. 533-555.

LIU, Y., WANG, X., LI, Y. AND CHEN, X. Metabolomic analysis of short-term sulfamethazine exposure on marine medaka (*oryzias melastigma*) by comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry. *Aquatic Toxicology*, 2018, 198, 269-275.

LU, W., BENNETT, B. D. AND RABINOWITZ, J. D. Analytical strategies for LC-MS-based targeted metabolomics. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 2008, 871(2), 236-242.

MALIK, A., JORDAO, R., CAMPOS, B., CASAS, J., BARATA, C. AND TAULER, R. Exploring the disruptive effects of tbt on lipid homeostasis of *Daphnia magna* using chemometric methods. *Chemometrics and Intelligent Laboratory Systems*, 2016, 159, 58-68.

MALLAT, S. G. A theory for multiresolution signal decomposition: The wavelet representation. In *Fundamental papers in wavelet theory*. 2009, p. 494-513.

MANSILHA, C., SILVA, P., ROCHA, S., GAMEIRO, P., DOMINGUES, V., PINHO, C. AND FERREIRA, I. M. Bisphenol A migration from plastic materials: Direct insight of ecotoxicity in *Daphnia magna*. *Environmental Science and Pollution Research*, 2013, 20(9), 6007-6018.

MARÍN-GARCÍA, M., IOELE, G., FRANQUET-GRIELL, H., LACORTE, S., RAGNO, G. AND TAULER, R. Investigation of the photodegradation profile of tamoxifen using spectroscopic and chromatographic analysis and multivariate curve resolution. *Chemometrics and Intelligent Laboratory Systems*, 2018, 174, 128-141.

MATTHEWS, B. W. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *BBA - Protein Structure*, 1975, 405(2), 442-451.

MCLACHLAN, J. A. Environmental signaling: What embryos and evolution teach us about endocrine disrupting chemicals. *Endocrine Reviews*, 2001, 22(3), 319-341.

MEDEIROS, M. N., OLIVEIRA, D. M., PAIVA-SILVA, G. O., SILVA-NETO, M. A., ROMEIRO, A., BOZZA, M., MASUDA, H. AND MACHADO, E. A. The role of eicosanoids on *rhodnius hemebinding protein* (rhbp) endocytosis by *rhodnius prolixus* ovaries. *Insect biochemistry and molecular biology*, 2002, 32(5), 537-545.

MEYER, Y. *Wavelets and operators*. Edtion ed.: Cambridge university press, 1992. ISBN 0521458692.

MISHRA, P., GONG, Z. AND KELLY, B. C. Assessing biological effects of fluoxetine in developing zebrafish embryos using gas chromatography-mass spectrometry based metabolomics. *Chemosphere*, 2017, 188, 157-167.

MOYLE, P. B. AND CECH, J. J. *Fishes: An introduction to ichthyology*. Edtion ed., 2004. ISBN 0131008471.

NAGATO, E. G., D'EON, J. C., LANKADURAI, B. P., POIRIER, D. G., REINER, E. J., SIMPSON, A. J. AND SIMPSON, M. J. ¹H NMR-based metabolomics investigation of *Daphnia magna*

responses to sub-lethal exposure to arsenic, copper and lithium. *Chemosphere*, 2013, 93(2), 331-337.

NAGATO, E. G., SIMPSON, A. J. AND SIMPSON, M. J. Metabolomics reveals energetic impairments in *Daphnia magna* exposed to diazinon, malathion and bisphenol-A. *Aquatic Toxicology*, 2016, 170, 175-186.

NAVARRO-REIG, M., JAUMOT, J., GARCÍA-REIRIZ, A. AND TAULER, R. Evaluation of changes induced in rice metabolome by cd and cu exposure using LC-MS with XCMS and MCR-ALS data analysis strategies. *Analytical and Bioanalytical Chemistry*, 2015, 407(29), 8835-8847.

NAVARRO-REIG, M., JAUMOT, J., PIÑA, B., MOYANO, E., GALCERAN, M. T. AND TAULER, R. Metabolomic analysis of the effects of cadmium and copper treatment in: *Oryza sativa* L. Using untargeted liquid chromatography coupled to high resolution mass spectrometry and all-ion fragmentation. *Metallomics*, 2017a, 9(6), 660-675.

NAVARRO-REIG, M., ORTIZ-VILLANUEVA, E., TAULER, R. AND JAUMOT, J. Modelling of hydrophilic interaction liquid chromatography stationary phases using chemometric approaches. *Metabolites*, 2017b, 7(4).

NICHOLSON, J. K., LINDON, J. C. AND HOLMES, E. 'Metabonomics': Understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, 1999, 29(11), 1181-1189.

NIELSEN, N. P. V., CARSTENSEN, J. M. AND SMEDSGAARD, J. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *Journal of Chromatography A*, 1998, 805(1-2), 17-35.

OECD Test no. 211: *Daphnia magna* reproduction test. Edition ed.: OECD Publishing, 2012.

ORATA, F. Derivatization reactions and reagents for gas chromatography analysis. In *Advanced gas chromatography-progress in agricultural, biomedical and industrial applications*. InTech, 2012.

ORTIZ-VILLANUEVA, E., BENAVENTE, F., PIÑA, B., SANZ-NEBOT, V., TAULER, R. AND JAUMOT, J. Knowledge integration strategies for untargeted metabolomics based on MCR-ALS analysis of CE-MS and LC-MS data. *Analytica Chimica Acta*, 2017a, 978, 10-23.

ORTIZ-VILLANUEVA, E., JAUMOT, J., BENAVENTE, F., PIÑA, B., SANZ-NEBOT, V. AND TAULER, R. Combination of CE-MS and advanced chemometric methods for high-throughput metabolic profiling. *Electrophoresis*, 2015, 36(18), 2324-2335.

ORTIZ-VILLANUEVA, E., JAUMOT, J., MARTÍNEZ, R., NAVARRO-MARTÍN, L., PIÑA, B. AND TAULER, R. Assessment of endocrine disruptors effects on zebrafish (*danio rerio*) embryos by untargeted LC-HRMS metabolomic analysis. *Science of the Total Environment*, 2018, 635, 156-166.

ORTIZ-VILLANUEVA, E., NAVARRO-MARTÍN, L., JAUMOT, J., BENAVENTE, F., SANZ-NEBOT, V., PIÑA, B. AND TAULER, R. Metabolic disruption of zebrafish (*danio rerio*) embryos by bisphenol A. An integrated metabolomic and transcriptomic approach. *Environmental Pollution*, 2017b, 231, 22-36.

ORTIZ-VILLANUEVA, E., NAVARRO-REIG, M., JAUMOT, J. AND TAULER, R. Chemometric evaluation of hydrophilic interaction liquid chromatography stationary phases: Resolving complex mixtures of metabolites. *Analytical Methods*, 2017c, 9(5), 774-785.

PAAIJMANS, K. P., HEINIG, R. L., SELIGA, R. A., BLANFORD, J. I., BLANFORD, S., MURDOCK, C. C. AND THOMAS, M. B. Temperature variation makes ectotherms more sensitive to climate change. *Global change biology*, 2013, 19(8), 2373-2380.

PANG, L., LIU, J., YIN, Y. AND SHEN, M. Evaluating the sorption of organophosphate esters to different sourced humic acids and its effects on the toxicity to *Daphnia magna*. *Environmental Toxicology and Chemistry*, 2013, 32(12), 2755-2761.

PATTI, G. J., YANES, O. AND SIUZDAK, G. Metabolomics: The apogee of the omic trilogy. *Nature reviews. Molecular cell biology*, 2012, 13(4), 263-269.

PENG, J., PENG, S., JIANG, A., WEI, J., LI, C. AND TAN, J. Asymmetric Least Squares for multiple spectra baseline correction. *Analytica Chimica Acta*, 2010, 683(1), 63-68.

PIERCE, K. M., HOPE, J. L., HOGGARD, J. C. AND SYNOVEC, R. E. A principal component analysis based method to discover chemical differences in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC×GC-TOFMS) separations of metabolites in plant samples. *Talanta*, 2006, 70(4), 797-804.

PIQUERAS, S., BEDIA, C., BELEITES, C., KRAFFT, C., POPP, J., MAEDER, M., TAULER, R. AND DE JUAN, A. Handling different spatial resolutions in image fusion by multivariate curve resolution-alternating least squares for incomplete image multisets. *Analytical Chemistry*, 2018, 90(11), 6757-6765.

PORTE, C., JANER, G., LORUSSO, L. C., ORTIZ-ZARRAGOITIA, M., CAJARAVILLE, M. P., FOSSI, M. C. AND CANESI, L. Endocrine disruptors in marine organisms: Approaches and perspectives. *Comparative Biochemistry and Physiology - C Toxicology and Pharmacology*, 2006, 143(3), 303-315.

POYNTON, H. C., TAYLOR, N. S., HICKS, J., COLSON, K., CHAN, S., CLARK, C., SCANLAN, L., LOGUINOV, A. V., VULPE, C. AND VIANT, M. R. Metabolomics of microliter hemolymph samples enables an improved understanding of the combined metabolic and transcriptional responses of *Daphnia magna* to cadmium. *Environmental Science and Technology*, 2011, 45(8), 3710-3717.

PUIG-CASTELLVÍ, F., ALFONSO, I., PIÑA, B. AND TAULER, R. A quantitative ^1H NMR approach for evaluating the metabolic response of *saccharomyces cerevisiae* to mild heat stress. *Metabolomics*, 2015, 11(6), 1612-1625.

RAJALAHTI, T., ARNEBERG, R., BERVEN, F. S., MYHR, K. M., ULVIK, R. J. AND KVALHEIM, O. M. Biomarker discovery in mass spectral profiles by means of selectivity ratio plot. *Chemometrics and Intelligent Laboratory Systems*, 2009a, 95(1), 35-48.

RAJALAHTI, T., ARNEBERG, R., KROKSVEEN, A. C., BERLE, M., MYHR, K. M. AND KVALHEIM, O. M. Discriminating variable test and selectivity ratio plot: Quantitative tools for interpretation and variable (biomarker) selection in complex spectral or chromatographic profiles. *Analytical Chemistry*, 2009b, 81(7), 2581-2590.

RIVETTI, C., CAMPOS, B. AND BARATA, C. Low environmental levels of neuro-active pharmaceuticals alter phototactic behaviour and reproduction in *Daphnia magna*. *Aquatic Toxicology*, 2016, 170, 289-296.

ROCHAT, B. From targeted quantification to untargeted metabolomics: Why LC-high-resolution-MS will become a key instrument in clinical labs. *TrAC Trends in Analytical Chemistry*, 2016, 84(Part B), 151-164.

ROLING, J. A., BAIN, L. J., GARDEA-TORRESDEY, J., KEY, P. B. AND BALDWIN, W. S. Using mummichog (*fundulus heteroclitus*) arrays to monitor the effectiveness of remediation at a superfund site in charleston, south carolina, USA. *Environmental Toxicology and Chemistry*, 2007, 26(6), 1205-1213.

RYAN, D. AND ROBARDS, K. Metabolomics: The greatest omics of them all? *Analytical Chemistry*, 2006, 78(23), 7954-7958.

SALGUEIRO-GONZÁLEZ, N., TURNES-CAROU, I., VIÑAS, L., BESADA, V., MUNIATEGUI-LORENZO, S., LÓPEZ-MAHÍA, P. AND PRADA-RODRÍGUEZ, D. Occurrence of alkylphenols and bisphenol A in wild mussel samples from the spanish atlantic coast and bay of biscay. *Marine Pollution Bulletin*, 2016, 106(1-2), 360-365.

SAWADA, Y., AKIYAMA, K., SAKATA, A., KUWAHARA, A., OTSUKI, H., SAKURAI, T., SAITO, K. AND HIRAI, M. Y. Widely targeted metabolomics based on large-scale MS/MS data for

elucidating metabolite accumulation patterns in plants. *Plant and Cell Physiology*, 2008, 50(1), 37-47.

SAYOOD, K. *Introduction to data compression*. Edtion ed.: Elsevier, 2005. ISBN 0080509258.

SCANLAN, L. D., LOGUINOV, A. V., TENG, Q., ANTCZAK, P., DAILEY, K. P., NOWINSKI, D. T., KORNBLUH, J., LIN, X. X., LACHENAUER, E., ARAI, A., DOUGLAS, N. K., FALCIANI, F., STAPLETON, H. M. AND VULPE, C. D. Gene transcription, metabolite and lipid profiling in eco-indicator *Daphnia magna* indicate diverse mechanisms of toxicity by legacy and emerging flame-retardants. *Environmental Science and Technology*, 2015, 49(12), 7400-7410.

SCHALLENBERG, M., HALL CATHERINE, J. AND BURNS CAROLYN, W. Consequences of climate-induced salinity increases on zooplankton abundance and diversity in coastal lakes. *Marine Ecology Progress Series*, 2003, 251, 181-189.

SCHINDLER, D. W. The cumulative effects of climate warming and other human stresses on canadian freshwaters in the new millennium. *Canadian Journal of Fisheries and Aquatic Sciences*, 2001, 58(1), 18-29.

SCHUEUERMANN, C., STEEL, C. C., BLACKMAN, J. W., CLARK, A. C., SCHWARZ, L. J., MORAGA, J., COLLADO, I. G. AND SCHMIDTKE, L. M. A GC-MS untargeted metabolomics approach for the classification of chemical differences in grape juices based on fungal pathogen. *Food Chemistry*, 2019, 270, 375-384.

SHEPARD, J. L., OLSSON, B., TEDENGREN, M. AND BRADLEY, B. P. Protein expression signatures identified in *mytilus edulis* exposed to PCBs, copper and salinity stress. *Marine Environmental Research*, 2000, 50(1-5), 337-340.

SIMMONS, D. B. D., BENSKIN, J. P., COSGROVE, J. R., DUNCKER, B. P., EKMAN, D. R., MARTYNIUK, C. J. AND SHERRY, J. P. Omics for aquatic ecotoxicology: Control of extraneous variability to enhance the analysis of environmental effects. *Environmental Toxicology and Chemistry*, 2015, 34(8), 1693-1704.

SIMÓN-MANSO, Y., LOWENTHAL, M. S., KILPATRICK, L. E., SAMPSON, M. L., TELU, K. H., RUDNICK, P. A., MALLARD, W. G., BEARDEN, D. W., SCHOCK, T. B., TCHEKHOVSKOI, D. V., BLONDER, N., YAN, X., LIANG, Y., ZHENG, Y., WALLACE, W. E., NETA, P., PHINNEY, K. W., REMALEY, A. T. AND STEIN, S. E. Metabolite profiling of a NIST standard reference material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources. *Analytical Chemistry*, 2013, 85(24), 11725-11731.

SMILDE, A. K., JANSEN, J. J., HOEFSLOOT, H. C. J., LAMERS, R. J. A. N., VAN DER GREEF, J. AND TIMMERMAN, M. E. ANOVA-simultaneous component analysis (ASCA): A new tool for analyzing designed metabolomics data. *Bioinformatics*, 2005, 21(13), 3043-3048.

SMOLDERS, R., BAILLIEUL, M. AND BLUST, R. Relationship between the energy status of *Daphnia magna* and its sensitivity to environmental stress. *Aquatic Toxicology*, 2005, 73(2), 155-170.

SNAPE, J. R., MAUND, S. J., PICKFORD, D. B. AND HUTCHINSON, T. H. Ecotoxicogenomics: The challenge of integrating genomics into aquatic and terrestrial ecotoxicology. *Aquatic Toxicology*, 2004, 67(2), 143-154.

SONG, Y., CHAI, T., YIN, Z., ZHANG, X., ZHANG, W., QIAN, Y. AND QIU, J. Stereoselective effects of ibuprofen in adult zebrafish (*danio rerio*) using UPLC-TOF/MS-based metabolomics. *Environmental Pollution*, 2018, 241, 730-739.

SOTTO, R. B. D., MEDRIANO, C. D., CHO, Y., KIM, H., CHUNG, I.-Y., SEOK, K.-S., SONG, K. G., HONG, S. W., PARK, Y. AND KIM, S. Sub-lethal pharmaceutical hazard tracking in adult zebrafish using untargeted LC-MS environmental metabolomics. *Journal of Hazardous Materials*, 2017, 339, 63-72.

STANLEY, D. Prostaglandins and other eicosanoids in insects: Biological significance. *Annu. Rev. Entomol.*, 2006, 51, 25-44.

STICKNEY, R. Encyclopedia of aquaculture. *Encyclopedia of Aquaculture*, 2000.

STOLT, R., TORGRIP, R. J. O., LINDBERG, J., CSENKI, L., KOLMERT, J., SCHUPPE-KOISTINEN, I. AND JACOBSSON, S. P. Second-order peak detection for multicomponent high-resolution LC-MS data. *Analytical Chemistry*, 2006, 78(4), 975-983.

STØRSETH, T. R. AND HAMMER, K. M. Environmental metabolomics of aquatic organisms. 2013. eMagRes.

STRASSBURG, K., HUIJBRECHTS, A. M., KORTEKAAS, K. A., LINDEMAN, J. H., PEDERSEN, T. L., DANE, A., BERGER, R., BRENKMAN, A., HANKEMEIER, T. AND VAN DUYNHOVEN, J. Quantitative profiling of oxylipins through comprehensive LC-MS/MS analysis: Application in cardiac surgery. *Analytical and Bioanalytical Chemistry*, 2012, 404(5), 1413-1426.

STREHMEL, N., HUMMEL, J., ERBAN, A., STRASSBURG, K. AND KOPKA, J. Retention index thresholds for compound matching in GC-MS metabolite profiling. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 2008, 871(2), 182-190.

SUMNER, L. W., AMBERG, A., BARRETT, D., BEALE, M. H., BEGER, R., DAYKIN, C. A., FAN, T. W. M., FIEHN, O., GOODACRE, R., GRIFFIN, J. L., HANKEMEIER, T., HARDY, N., HARNLY, J., HIGASHI, R., KOPKA, J., LANE, A. N., LINDON, J. C., MARRIOTT, P., NICHOLLS, A. W., REILY, M. D., THADEN, J. J. AND VIANT, M. R. Proposed minimum reporting standards for chemical analysis: Chemical analysis working group (CAWG) metabolomics standards initiative (MSI). *Metabolomics*, 2007, 3(3), 211-221.

SUZUKI, T., NAKAGAWA, Y., TAKANO, I., YAGUCHI, K. AND YASUDA, K. Environmental fate of bisphenol A and its biological metabolites in river water and their xeno-estrogenic activity. *Environmental Science & Technology*, 2004, 38(8), 2389-2396.

TANG, C., XIAO, H., ZHANG, L., CAO, D., FAN, W., LIANG, Y., ZENG, Y., TAN, B. AND ZENG, M. Strategies for structure elucidation of small molecules using gas chromatography-mass spectrometric data. *TrAC - Trends in Analytical Chemistry*, 2013, 47, 37-46.

TANG, D.-Q., ZOU, L., YIN, X.-X. AND ONG, C. N. HILIC-MS for metabolomics: An attractive and complementary approach to RPLC-MS. *Mass Spectrometry Reviews*, 2016, 35(5), 574-600.

TATARAZAKO, N., ODA, S., WATANABE, H., MORITA, M. AND IGUCHI, T. Juvenile hormone agonists affect the occurrence of male *daphnia*. *Chemosphere*, 2003, 53(8), 827-833.

TAULER, R. Multivariate curve resolution applied to second order data. *Chemometrics and Intelligent Laboratory Systems*, 1995, 30(1), 133-146.

TAULER, R. AND BARCELÓ, D. Multivariate curve resolution applied to liquid chromatography-diode array detection. *Trends in Analytical Chemistry*, 1993, 12(8), 319-327.

TAULER, R., KOWALSKI, B. AND FLEMING, S. Multivariate curve resolution applied to spectral data from multiple runs of an industrial process. *Analytical Chemistry*, 1993, 65(15), 2040-2047.

TAULER, R., SMILDE, A. AND KOWALSKI, B. Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *Journal of Chemometrics*, 1995, 9(1), 31-58.

TAYLOR, N., GAVIN, A. AND VIANT, M. Metabolomics discovers early-response metabolic biomarkers that can predict chronic reproductive fitness in individual *Daphnia magna*. *Metabolites*, 2018, 8(3), 42.

TAYLOR, N. S., WEBER, R. J. M., SOUTHAM, A. D., PAYNE, T. G., HRYDZIUSZKO, O., ARVANITIS, T. N. AND VIANT, M. R. A new approach to toxicity testing in *Daphnia magna*: Application of high throughput FT-ICR mass spectrometry metabolomics. *Metabolomics*, 2009, 5(1), 44-58.

TAYLOR, N. S., WEBER, R. J. M., WHITE, T. A. AND VIANT, M. R. Discriminating between different acute chemical toxicities via changes in the *daphnid* metabolome. *Toxicological Sciences*, 2010, 118(1), 307-317.

TETKO, I. V., GASTEIGER, J., TODESCHINI, R., MAURI, A., LIVINGSTONE, D., ERTL, P., PALYULIN, V. A., RADCHENKO, E. V., ZEFIROV, N. S., MAKARENKO, A. S., TANCHUK, V. Y. AND PROKOPENKO, V. V. Virtual computational chemistry laboratory - design and description. *Journal of Computer-Aided Molecular Design*, 2005, 19(6), 453-463.

TOURNOIS, C., MUTEL, V., MANIVET, P., LAUNAY, J.-M. AND KELLERMANN, O. Cross-talk between 5-hydroxytryptamine receptors in a serotonergic cell line involvement of arachidonic acid metabolism. *Journal of Biological Chemistry*, 1998, 273(28), 17498-17503.

TRYGG, J., GULLBERG, J., JOHANSSON, A. I., JONSSON, P. AND MORITZ, T. Chemometrics in metabolomics — an introduction. 2006. *Biotechnology in Agriculture and Forestry*.

TRYGG, J. AND WOLD, S. Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics*, 2002, 16(3), 119-128.

TSUGAWA, H., TSUJIMOTO, Y., SUGITATE, K., SAKUI, N., NISHIUMI, S., BAMBA, T. AND FUKUSAKI, E. Highly sensitive and selective analysis of widely targeted metabolomics using gas chromatography/triple-quadrupole mass spectrometry. *Journal of Bioscience and Bioengineering*, 2014, 117(1), 122-128.

VAN AGGELEN, G., ANKLEY, G. T., BALDWIN, W. S., BEARDEN, D. W., BENSON, W. H., CHIPMAN, J. K., COLLETTE, T. W., CRAFT, J. A., DENSLOW, N. D., EMBRY, M. R., FALCIANI, F., GEORGE, S. G., HELBING, C. C., HOEKSTRA, P. F., IGUCHI, T., KAGAMI, Y., KATSIADAKI, I., KILLE, P., LIU, L., LORD, P. G., MCINTYRE, T., O'NEILL, A., OSACHOFF, H., PERKINS, E. J., SANTOS, E. M., SKIRROW, R. C., SNAPE, J. R., TYLER, C. R., VERSTEEG, D., VIANT, M. R., VOLZ, D. C., WILLIAMS, T. D. AND YU, L. Integrating omic technologies into aquatic ecological risk assessment and environmental monitoring: Hurdles, achievements, and future outlook. *Environmental Health Perspectives*, 2010, 118(1), 1-5.

VANDENBERG, L. N., MAFFINI, M. V., SONNENSCHNEIN, C., RUBIN, B. S. AND SOTO, A. M. Bisphenol A and the great divide: A review of controversies in the field of endocrine disruption. *Endocrine Reviews*, 2009, 30(1), 75-95.

VANDENBROUCK, T., JONES, O. A. H., DOM, N., GRIFFIN, J. L. AND DE COEN, W. Mixtures of similarly acting compounds in *Daphnia magna*: From gene to metabolite and beyond. *Environment International*, 2010, 36(3), 254-268.

VEN, K. V. D., KEIL, D., MOENS, L. N., LEEMPUT, K. V., REMORTEL, P. V. AND COEN, W. M. D. Neuropharmaceuticals in the environment: Mianserin-induced neuroendocrine disruption

in zebrafish (*danio rerio*) using cDNA microarrays. *Environmental Toxicology and Chemistry*, 2006, 25(10), 2645-2652.

VENDEUVRE, C., RUIZ-GUERRERO, R., BERTONCINI, F., DUVAL, L. AND THIÉBAUT, D. Comprehensive two-dimensional gas chromatography for detailed characterisation of petroleum products. *Oil & Gas Science and Technology - Rev. IFP*, 2007, 62(1), 43-55.

VETTUKATTIL, R. Preprocessing of raw metabolomic data. 2015. *Methods in Molecular Biology*.

VIANT, M. R. Recent developments in environmental metabolomics. *Molecular BioSystems*, 2008, 4(10), 980-986.

VIANT, M. R., PINCETICH, C. A., HINTON, D. E. AND TJEERDEMA, R. S. Toxic actions of dinoseb in medaka (*oryzias latipes*) embryos as determined by in vivo ³¹P NMR, HPLC-UV and ¹H NMR metabolomics. *Aquatic Toxicology*, 2006, 76(3-4), 329-342.

VILLENEUVE, D. L., ANKLEY, G. T. AND MARTINOVIĆ, D. Toxicogenomics applied to ecological risk assessment. In *Applications of toxicogenomics in safety evaluation and risk assessment*. John Wiley & Sons, Inc., 2011, p. 207-235.

WACHSMUTH, C. J., VOGL, F. C., OEFNER, P. J. AND DETTMER, K. Gas chromatographic techniques in metabolomics. *Chromatographic Methods in Metabolomics*; Hyotylainen, T., Wiedmer, S., Eds, 2013, 87-105.

WAGNER, N. D., LANKADURAI, B. P., SIMPSON, M. J., SIMPSON, A. J. AND FROST, P. C. Metabolomic differentiation of nutritional stress in an aquatic invertebrate. *Physiological and Biochemical Zoology*, 2015, 88(1), 43-52.

WAGNER, N. D., SIMPSON, A. J. AND SIMPSON, M. J. Metabolomic responses to sublethal contaminant exposure in neonate and adult *Daphnia magna*. *Environmental Toxicology and Chemistry*, 2017, 36(4), 938-946

WAGNER, N. D., SIMPSON, A. J. AND SIMPSON, M. J. Sublethal metabolic responses to contaminant mixture toxicity in *Daphnia magna*. *Environmental Toxicology and Chemistry*, 2018, 37(9), 2448-2457.

WATANABE, H. AND IGUCHI, T. Using ecotoxicogenomics to evaluate the impact of chemicals on aquatic organisms. *Marine Biology*, 2006, 149(1), 107-115.

WATANABE, H., TAKAHASHI, E., NAKAMURA, Y., ODA, S., TATARAZAKO, N. AND IGUCHI, T. Development of a *Daphnia magna* DNA microarray for evaluating the toxicity of environmental chemicals. *Environmental Toxicology and Chemistry*, 2007, 26(4), 669-676.

WESTERHUIS, J. A., VAN VELZEN, E. J. J., HOEFSLOOT, H. C. J. AND SMILDE, A. K. Multivariate paired data analysis: Multilevel PLSDA versus OPLSDA. *Metabolomics*, 2010, 6(1), 119-128.

WINDIG, W. The use of the durbin-watson criterion for noise and background reduction of complex liquid chromatography/mass spectrometry data and a new algorithm to determine sample differences. *Chemometrics and Intelligent Laboratory Systems*, 2005, 77(1-2 SPEC. ISS.), 206-214.

WINDIG, W. AND GUILMENT, J. Interactive self-modeling mixture analysis. *Analytical Chemistry*, 1991, 63(14), 1425-1432.

WISHART, D. S. Computational strategies for metabolite identification in metabolomics. *Bioanalysis*, 2009, 1(9), 1579-1596.

WISHART, D. S., JEWISON, T., GUO, A. C., WILSON, M., KNOX, C., LIU, Y., DJOUMBOU, Y., MANDAL, R., AZIAT, F., DONG, E., BOUATRA, S., SINELNIKOV, I., ARNDT, D., XIA, J., LIU, P., YALLOU, F., BJORND AHL, T., PEREZ-PINEIRO, R., EISNER, R., ALLEN, F., NEVEU, V., GREINER, R. AND SCALBERT, A. HMDB 3.0-the human metabolome database in 2013. *Nucleic Acids Research*, 2013, 41(D1), D801-D807.

WOLD, S., ANTTI, H., LINDGREN, F. AND ÖHMAN, J. Orthogonal signal correction of near-infrared spectra. *Chemometrics and Intelligent Laboratory Systems*, 1998, 44(1-2), 175-185.

WOLD, S., ESBENSEN, K. AND GELADI, P. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 1987, 2(1-3), 37-52.

WOLD, S., JOHANSSON, E. AND COCCHI, M. PLS - Partial Least-Squares projections to latent structures. *3D QSAR in Drug Design*, 1993, 523-550.

XIA, J., SINELNIKOV, I. V., HAN, B. AND WISHART, D. S. Metaboanalyst 3.0-making metabolomics more meaningful. *Nucleic Acids Research*, 2015, 43(W1), W251-W257.

XU, F., ZOU, L. AND ONG, C. N. Experiment-originated variations, and multi-peak and multi-origination phenomena in derivatization-based GC-MS metabolomics. *TrAC Trends in Analytical Chemistry*, 2010, 29(3), 269-280.

YANG, C., HAO, R., DU, X., WANG, Q., DENG, Y. AND SUN, R. Response to different dietary carbohydrate and protein levels of pearl oysters (*pinctada fucata martensii*) as revealed by GC-TOF/MS-based metabolomics. *Science of the Total Environment*, 2019, 650, 2614-2623.

YI, L., DONG, N., YUN, Y., DENG, B., REN, D., LIU, S. AND LIANG, Y. Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. *Analytica Chimica Acta*, 2016, 914, 17-34.

ZAFFAGNINI, F. Reproduction in *daphnia*. *Mem. Ist. Ital. Idrobiol*, 1987, 45, 245-284.

ZEAITER, M. AND RUTLEDGE, D. 3.04 - Preprocessing Methods A2 - Brown, Steven D. *Comprehensive chemometrics*. R. TAULER AND B. WALCZAK Eds. Oxford: Elsevier, 2009, p. 121-231.

ZENDONG, Z., MCCARRON, P., HERRENKNECHT, C., SIBAT, M., AMZIL, Z., COLE, R. B. AND HESS, P. High resolution mass spectrometry for quantitative analysis and untargeted screening of algal toxins in mussels and passive samplers. *Journal of Chromatography A*, 2015, 1416, 10-21.

ZHANG, H. AND ZHAO, L. Influence of sublethal doses of acetamiprid and halosulfuron-methyl on metabolites of zebra fish (*brachydanio rerio*). *Aquatic Toxicology*, 2017, 191, 85-94.

ZHANG, W., TAN, N. G. J., FU, B. AND LI, S. F. Y. Metallomics and NMR-based metabolomics of *chlorella sp.* Reveal the synergistic role of copper and cadmium in multi-metal toxicity and oxidative stress. *Metallomics*, 2015, 7(3), 426-438.

ZHOU, B., XIAO, J. F., TULI, L. AND RESSOM, H. W. LC-MS-based metabolomics. *Molecular BioSystems*, 2012, 8(2), 470-481.

ZHOU, J. AND YIN, Y. Strategies for large-scale targeted metabolomics quantification by liquid chromatography-mass spectrometry. *Analyst*, 2016, 141(23), 6362-6373.

ZROSTLÍKOVÁ, J., HAJŠLOVÁ, J. AND ČAJKA, T. Evaluation of two-dimensional gas chromatography-time-of-flight mass spectrometry for the determination of multiple pesticide residues in fruit. *Journal of Chromatography A*, 2003, 1019(1), 173-186.