

Article

A systems biology approach reveals a novel calcium-dependent mechanism for basal toxicity in *Daphnia magna*

Philipp Antczak, Thomas A White, Anirudha Giri, Francesco Michelangeli,
Mark R. Viant, Mark Cronin, Chris D Vulpe, and Francesco Falciani

Environ. Sci. Technol., **Just Accepted Manuscript** • DOI: 10.1021/acs.est.5b02707 • Publication Date (Web): 05 Aug 2015

Downloaded from <http://pubs.acs.org> on August 7, 2015

Just Accepted

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.

A systems biology approach reveals a novel calcium-dependent mechanism for basal toxicity in *Daphnia magna*

Philipp Antczak,[†] Thomas A. White,[‡] Anirudha Giri,[¶] Francesco Michelangeli,[‡]
Mark R. Viant,[‡] Mark T. D. Cronin,[§] Chris Vulpe,^{*,||,⊥} and Francesco
Falciani^{*,†,⊥}

*Centre for Computational Biology and Modelling, Institute for Integrative Biology,
University of Liverpool, L69 7ZB, Liverpool, UK, School of Biosciences, University of
Birmingham, B15 2TT, Birmingham, UK, Assam University, Silchar, 788011, India,
School of Pharmacy and Chemistry, Liverpool John Moores University, Liverpool L3 3AF,
UK, and Nutritional Sciences and Toxicology & Berkeley Institute of the Environment,
University of California, Berkeley, CA 94720, USA*

E-mail: vulpe@berkeley.edu; f.falciani@liv.ac.uk

Phone: +1-510-642-1834; +44-151-795-4558. Fax: +1-510-643-3132; +44-151-795-4408

Abstract

The expanding diversity and ever increasing amounts of man-made chemicals discharged to the environment pose largely unknown hazards to ecosystem and human

*To whom correspondence should be addressed

[†]University of Liverpool

[‡]University of Birmingham

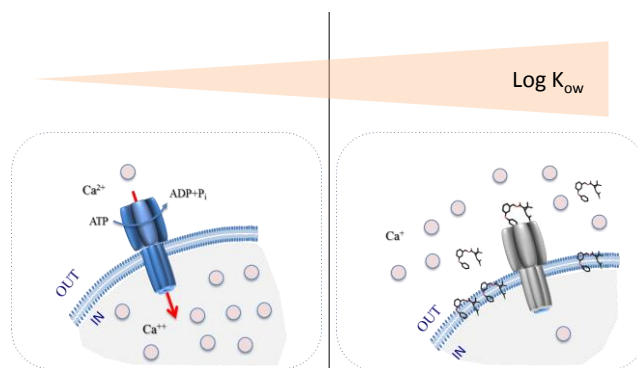
[¶]Assam University

[§]Liverpool John Moores University

^{||}University of California

[⊥]Authors wish to be considered joint senior and corresponding authors

4 health. The concept of adverse outcome pathways (AOPs) emerged as a comprehensive
5 framework for risk assessment. However, the limited mechanistic information available
6 for most chemicals and a lack of biological pathway annotation in many species rep-
7 resent significant challenges to effective implementation of this approach. Here, a
8 systems level, multi-step modeling strategy demonstrates how to integrate information
9 on chemical structure with mechanistic insight from genomic studies, and phenotypic
10 effects to define a putative adverse outcome pathway. Results indicated that tran-
11 scriptional changes indicative of intracellular calcium mobilization were significantly
12 overrepresented in *Daphnia magna* (DM) exposed to sub-lethal doses of presumed nar-
13 cotic chemicals with $\log K_{ow} \geq 1.8$. Treatment of DM with a calcium ATPase pump
14 inhibitor substantially recapitulated the common transcriptional changes. We hypoth-
15 esize that calcium mobilization is a potential key molecular initiating event in DM
16 basal (narcosis) toxicity. Heart beat rate analysis and metabolome analysis indicated
17 sub-lethal effects consistent with perturbations of calcium preceding overt acute toxic-
18 ity. Together, the results indicate that altered calcium homeostasis may be a key early
event in basal toxicity or narcosis induced by lipophilic compounds.



20 Introduction

21 The release of an increasingly large number of anthropogenic chemicals into the environ-
22 ment represents a formidable challenge in ecological risk assessment. The potential toxicity
23 of chemicals to ecologically relevant organisms must be considered in this process. However,

24 the historic lack of a regulatory imperative, the number of relevant ecosystems, the multiple
25 bio-indicator species for each ecosystem, and the cost of acute and chronic toxicity tests,
26 has resulted in ecotoxicity data being available for only a minority of chemicals in com-
27 merce. In addition, these traditional ecotoxicity measurements do not consider sub-lethal
28 effects nor provide insight into the mechanisms underlying any observed toxicity. Alternative
29 rapid, predictive, mechanism based and cost effective approaches for ecological risk assess-
30 ment of chemicals are urgently needed to preserve the integrity of the natural environment.
31 Quantitative structure activity relationships (QSARs) provide an established alternative to
32 traditional toxicity tests for the identification of toxic chemicals.^{1,2} However, they generally
33 do not provide a mechanistic link between physical chemical features and the observed tox-
34 icity. Recently, a conceptual framework for environmental risk assessment termed adverse
35 outcome pathways (AOPs) has been proposed and adopted by the OECD. AOPs repre-
36 sent the causal relationships between the molecular initiating event(s) of chemical(s) action
37 through biological processes to organism and population level adverse effects.³ Its application
38 in a regulatory context is likely to revolutionise the way we understand environmental tox-
39 icity. A significant barrier to implementation of this strategy is the dearth of well-annotated
40 biological pathways in many eco-relevant species. Genomic information can provide mech-
41 anistic insight into chemical action but methods to utilize this information in an adverse
42 outcome pathway framework remain limited. Here we propose a computational approach
43 that integrates traditional QSAR, expression profiling following exposure to sub-lethal (or
44 NOEC no observable effect concentrations) chemical concentrations, and toxicity endpoints
45 to generate a model of putative AOPs. This postulates that the physical-chemical features
46 (PCFs) of a chemical can explain an organisms' transcriptional response at a dose and at
47 a time of exposure where no overt toxicity is observed, and second, that such a response is
48 informative and predictive of the molecular mechanisms underlying toxicity at higher expo-
49 sure endpoints. **A key characteristic of this approach is that a hypothesis for the**
50 **AOP is generated from the computational analysis rather than testing a specific**

51 **a priori hypothesis.**

52 Application of this methodology to a dataset comprised of 24 environmentally relevant
53 chemicals revealed that structural features linked to compound lipophilicity are able to ex-
54 plain a considerable fraction of sub-lethal (NOEC) transcriptional response of DM to each
55 compound. Baseline toxicity, also termed narcosis, has been linked to the lipophilicity of
56 chemicals^{4,5} and has been subdivided into narcosis attributable to non-polar and polar com-
57 pounds which are proposed to have related if distinct mechanisms.^{6,7} However, despite in-
58 tense efforts, the precise mechanism of narcosis and the relative importance of basal versus
59 target-specific toxicity in the environment are still largely unknown.⁴ Analysis of the inferred
60 transcriptional network linked to both compound PCFs and toxicity outcomes support the
61 hypothesis that intracellular calcium release triggered by lipophilic chemicals may be one
62 of the initiating events that underlie basal toxicity of these compounds. More generally,
63 the approach shows for the first time how a computational approach integrating traditional
64 QSAR with advanced systems biology approaches can help define an AOP. The widespread
65 application of the approach developed here is therefore expected to have significant impact
66 on the development of AOPs in the field of chemical and environmental hazard assessment.

67 **Materials and Methods**

68 **Analysis strategy**

69 The overall computational strategy to identify AOPs is to link the transcriptional state of
70 an organism following sub-lethal chemicals exposures to both PCFs and organismal toxicity
71 and can be conceptualised as six interconnected steps (Figure 1). First, PCFs are identified
72 that predict the transcriptional activity of KEGG pathways (Figure 1, Step 1). In parallel,
73 each pathway is tested for its ability to predict neonate LC_{50} (nLC_{50}) (Figure 1, Step 2). The
74 relationship between these objects can be represented in a multi-level map (Figure 1, step
75 3) defining the interaction of PCFs with pathways, and pathways with organismal toxicity.

76 Together these linkages form a network between chemicals (PCFs), biological pathways and
77 toxicity, which allows generation of AOP hypotheses (Figure 1, Step 4-5). Targeted studies
78 can be used to test these hypotheses (Figure 1, Step 6). Details of the individual analysis
79 for each step can be found in the sections below.

80 **Chemical Exposures and expression profiling dataset**

81 Transcriptional data from exposure of DM to twenty six organic chemicals were initially
82 selected from a previous study.¹ Briefly, this set represented gene expression profiles for 14
83 day old DM adults exposed to $\frac{1}{10}$ of the identified nLC₅₀ (calculated using neonatal DM). At
84 the validation stage of the project, an additional exposure to thapsigargin was performed.
85 For more details on the compounds see Table S5. Exposures were performed with twenty
86 two week old DM in 1L of modified COMBO media⁸ containing $\frac{1}{10}$ LC₅₀ concentration of
87 each of the chemical in four replicates. Each beaker was carefully sealed with clingfilm to
88 reduce volatility and to improve delivery. After a 24h exposure, total RNA was extracted
89 and arrayed using a custom Agilent microarray (AMAID: 023710, GPL15139). (Further
90 information on the selected chemicals can be found in Supplementary Table S5 and S8). To
91 verify that the experimental precautions taken to reduce evaporation were effective chemical
92 concentrations were measured during exposure to the 3 most volatile compounds. The
93 analysis revealed that the concentration of 2 of the volatile compounds were reduced by 50%
94 at the end of the exposure while the remaining chemical showed < 1% loss. This shows
95 that compound loss caused by evaporation is not likely to affect LC₅₀ determination (See
96 Supplementary Table S9 for more details).

97 **Basal toxicity model and excess toxicity**

98 To establish whether the compounds in this dataset fit a baseline toxicity model, a compari-
99 son with an already developed model⁹ was performed. First a model, based on this dataset,

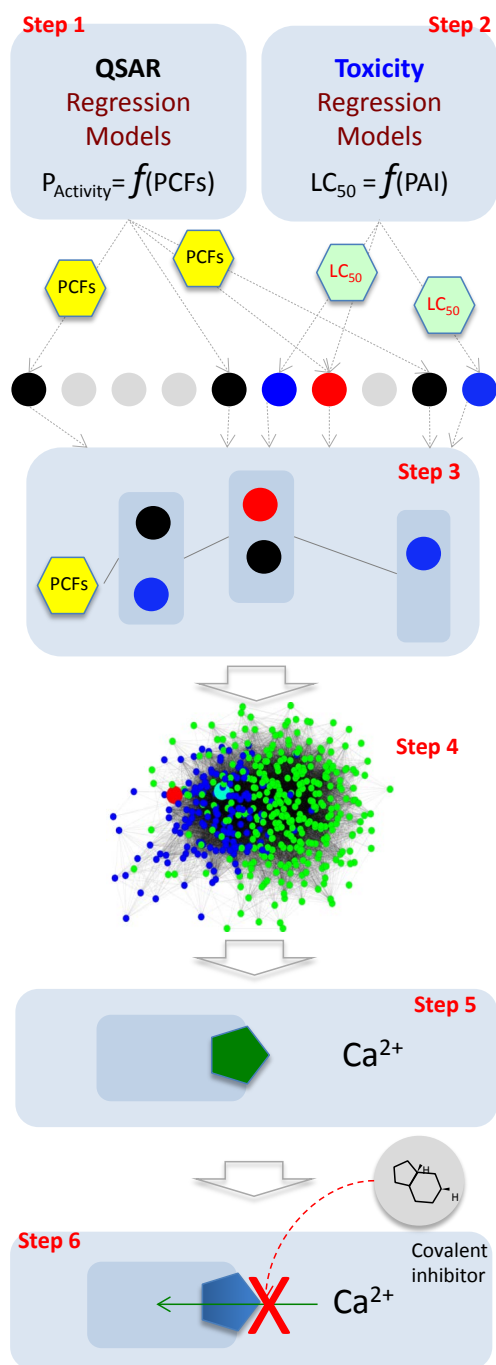


Figure 1: Step 1: Predicting pathway activity from PCFs. Step 2: Predicting toxicity from pathway activity. Step 3: Visualization of the results in a KEGG Pathway Interaction map. Step 4: Network reconstruction of identified genes and PCFs. Step 5: Hypothesis generation through integration of Step 3 and Step 2 results. Step 6: Validation of hypotheses generated in Step 5.

100 was derived:

$$\log(LC_{50}) = -0.8438 * \log(K_{ow}) - 2.3078. \quad (1)$$

101 Comparing this with the von der Ohe et al.⁹ baseline toxicity model verified that the es-
102 timated parameters were within the confidence intervals and hence the two models were
103 considered to be indistinguishable. To further assess this datasets' compounds, an excess
104 toxicity index (Equ. 2) was calculated based on equations developed by von der Ohe et al.⁹.

$$Te = \frac{\textit{predicted}LC_{50}}{\textit{experimental}LC_{50}} \quad (2)$$

105 Plotting this index as a function of LC_{50} resulted in no compounds with excess toxicity (> 2)
106 and 4 compounds with a slightly lower Te value than expected. These 4 compounds include
107 Phenol, Acrylonitrile, Ponasterone A and 20-hydroxyecdysone. The analysis shows that, in
108 the conditions of this experimental system, most of lethal toxicity in these chemicals can
109 be expected to follow a narcosis based mechanism (see Supplementary Figure S12). These
110 results are consistent with the initial assessment which shows minimal loss of highly volatile
111 chemical during exposure.

112 Calculation of PCFs

113 PCFs describing each chemical were identified using the e-Dragon web service available at
114 www.vcclab.org.¹⁰ Only features that were available across all chemicals (1260/2352) were
115 retained. Chemicals which showed outlier PCFs profiles, or for which calculations of PCFs
116 lead to errors due to structural peculiarities were removed (2 compounds). The final dataset
117 therefore contained 24 chemicals.

118 Calculation of indexes of pathway activity (PAI)

119 To reduce the complexity of the expression profiling dataset, the individual gene expression
120 profiles were grouped according to biological pathways defined in the KEGG database and

121 then indexes of pathway activity (PAI) were computed using principal component analysis
122 (PCA). The first three principal components (representing at least 70% of the variance),
123 were retained for further analysis. This procedure reduces the initial set of 1425 genes to
124 285 pathway components (95 KEGG pathways and 3 principal components (PCs) each).
125 Although this approach eliminates the non-annotated genes, biological interpretability and
126 statistical power are greatly enhanced.¹¹⁻¹³

127 **Toxicity endpoint**

128 Organismal toxicity was determined in the initial generated dataset.¹ Briefly, neonates (<
129 24h) were exposed to varying nominal concentrations of each chemical over 24h and nominal
130 LC₅₀s generated. This neonate LC₅₀ (nLC₅₀) was then used to link the transcriptional
131 response of 14 day old adults at $\frac{1}{10}$ nLC₅₀.

132 **Pathway activity as a function of chemical features**

133 Using an advanced machine learning technique (GALGO¹⁴) optimized sub-sets of 3 PCFs
134 were identified, which are able to predict each of the 285 PAIs (3 PCs x 95 pathways) using
135 the following randomForest model:

$$PAI_{j,k} = a\theta_1 + b\theta_2 + c\theta_3 + d + \epsilon \quad (3)$$

136 Here $PAI_{j,k}$ represents the pathway activity index for pathway j and component k, θ_{1-3}
137 represent 3 PCFs and d and ϵ the intercept and error of the model. For 35 pathways (at
138 least one of its component) a highly significant association to PCFs with an $R^2 > 0.75$ could
139 be identified.

140 Toxicity as a function of Pathway activity

141 For each pathway a random Forest¹⁵ regression model was used to identify pathways predic-
142 tive of toxicity (nLC₅₀). The regression model was defined as:

$$\log(LC_{50}) = a\theta_1 + b\theta_2 + c\theta_3 + d + \epsilon \quad (4)$$

143 To identify statistically significant pathways, nLC₅₀ are randomized 1000 times and the
144 model rerun. Twenty significant pathways linked to toxicity were identified with an R² >
145 0.6.

146 Analysis of Gene expression dataset

147 To develop a KEGG Pathway map links between KEGG pathways are represented as the
148 Jaccard's Index of overlap, which is defined as the ratio between the size of the intersect over
149 the size of the union of any 2 samples. To aid in interpretation, pathways were ordered into
150 higher functional groups and coloured on the basis of their association (PCFs black, toxicity
151 blue and both red). To further build a network representing the dependency between genes,
152 PCFs and LC₅₀ based on the KEGG map, ARACNE¹⁶ was applied (p-value 10⁻⁸). To identify
153 highly interconnected sub-networks G_{Lay} (clusterMaker^{17,18}) was applied. A force driven
154 layout was used to represent the graph. Statistically significant correlation between genes
155 and experimental log K_{ow} was identified using SAM (significance analysis for microarrays).¹⁹
156 Significant over-representation was identified using a modified fisher test as described in.²⁰
157 To identify metabolites or genes highly represented by the log K_{ow} signature KEGG reference
158 pathways with at least 5 represented members in this dataset were considered and the fisher
159 test applied. To link exposure of thaspigargin and the remaining dataset genes are ranked
160 by d-statistic from the above SAM analysis and used as ranked input to GSEA.²¹ Genesets
161 were defined as thaspigargin significantly differentially expressed genes (FDR < 20%). To
162 define contribution of specific vs. basal toxicity mechanisms a 2 factor ANOVA was used to

163 identify whether the majority of the variation observed in the dataset was associated with
164 chemical class (as defined in¹), log K_{ow} or their interaction.

165 **Validation of observed Ca^{2+} effects**

166 To establish whether compounds are likely to inhibit SERCA ATPase, IC_{50} data was sourced
167 from the public domain and complemented with new measurements as described in.²² Fur-
168 thermore, semi-targeted and relative quantitative measurements using 1H nuclear magnetic
169 resonance (NMR) spectroscopy was performed. Hydrophilic metabolites were extracted from
170 DM as described in.²³ Data was normalized, g-logged (generalized logarithm) and analysed
171 by PCA. Metabolites were identified using an online database.²⁴ Lastly, heart rate was mea-
172 sured in two week old DM following 1h and 24h exposure to 8 compounds of varying log K_{ow}
173 through video monitoring. Data was collected using 15 individuals as described in.²⁵

174 **Development of a model predictive of toxicity and integrating cal- 175 cium signalling expression signatures and lipophilicity**

176 A machine learning approach (GALGO¹⁴) was used to assess whether inclusion of calcium
177 associated genes to log K_{ow} could produce a better predictive model than log K_{ow} by itself.

178 **Results**

179 **Statistical modeling reveals an interaction between compound PCFs, 180 whole organism transcriptional response and toxicity outcome**

181 The first objective of this study was to identify putative AOPs representing a link between
182 compound PCFs, pathway activity and toxicity outcome (Figure 1). Remarkably, 35 out
183 of the 95 (36%) KEGG pathways could be identified whose activity can be predicted as a
184 function of a subset of PCFs (Table S1) thus linking PCFs to pathway activity. In addition,

transcriptional activity of 20 out of 95 (21%) KEGG pathways were found to be predictive of toxicity (Table S1) and thereby completing a link between PCFs, biological pathways, and toxicity. Grouping of all of these pathways resulted in eight functional groups: 1) amino acid metabolism, 2) glycan metabolism, 3) lipid metabolism, 4) signaling, 5) DNA repair and replication, 6) membrane, 7) protein translation/degradation and 8) energy metabolism. A functional network representing the inferred complex relationship between PCFs, pathways and toxicity is shown in Figure 2.

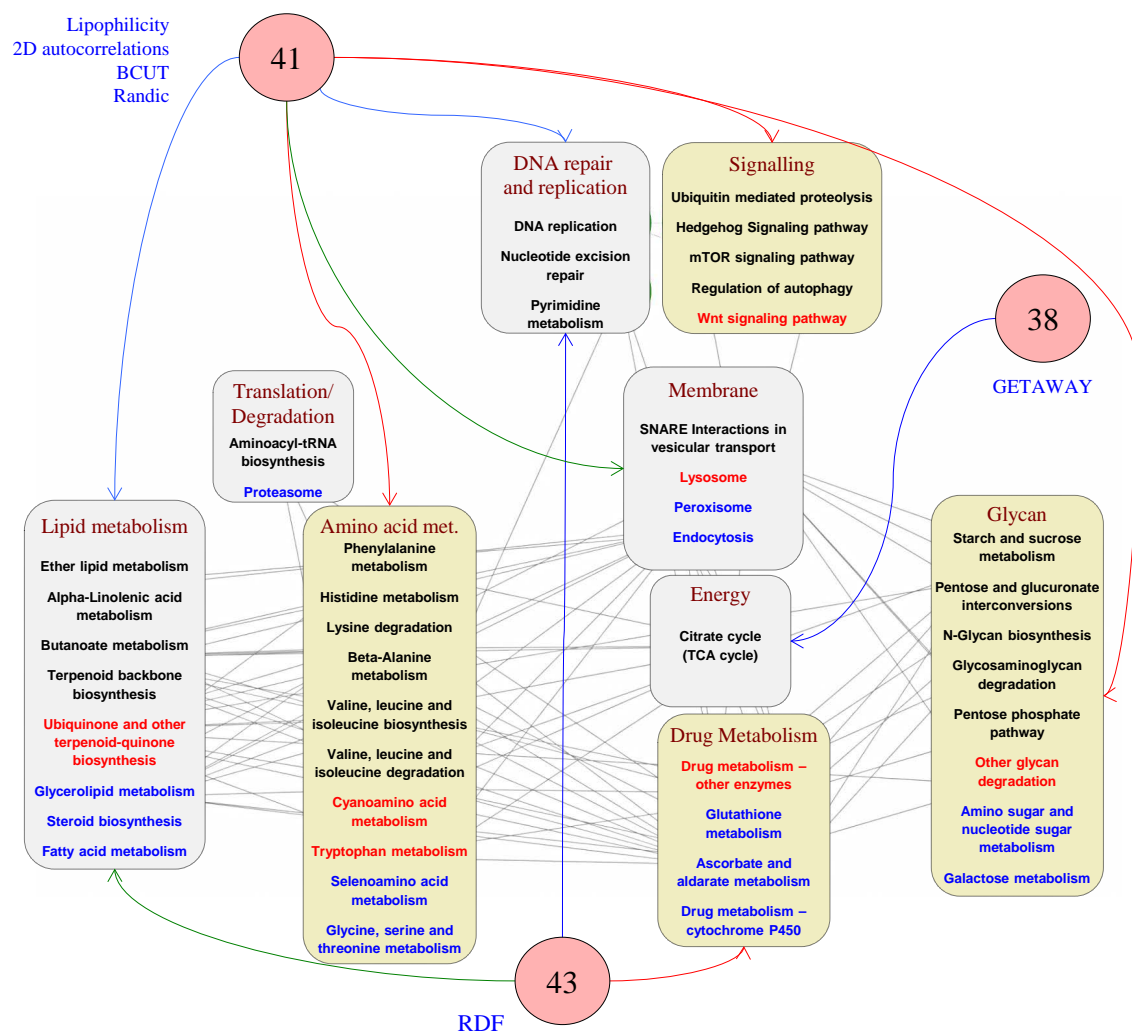


Figure 2: Representation of the interactions between functional clusters and groups of PCFs. Each functional cluster represents a number of related pathways associated to either PCFs (black), toxicity (blue) and both (red). Strength of interactions between PCFs and functional clusters is represented by colored arrows (red > 0.15, blue > 0.1, and green < 0.1).

192 **PCFs clusters associate with functional groups**

193 Visual inspection of the network shows that a cluster of PCFs, representing mainly com-
194 pound lipophilicity, 2D autocorrelations (representing the shape of a molecule by topological
195 distance weighted by molecular properties) and BCUT descriptors (representing atomic prop-
196 erties relevant to intermolecular interactions) were connected with six out of the 8 defined
197 functional pathway groups (for additional descriptor groups see Figure S1). A second PCFs
198 cluster, which mainly represented features from the RDF descriptor group (Radial distribu-
199 tion function: which represent the probability of an atom to be present at a given radius from
200 the center of the molecule), and a third cluster mainly comprised of GETAWAY descriptors
201 (geometry, topology, and atoms-weighted assembly; descriptors representing the molecular
202 structure based on atomic coordinates calculated with respect to the geometrical center of
203 the molecule) were connected to three and one functional group, respectively (Figure 2, for
204 more detail see Figure S1).

205 **Inference of a biological network integrating PCFs, gene expression** 206 **and toxicity**

207 The previously developed high level map already provided indications on the possible points
208 of interactions between compounds (based on PCFs) and pathway activity. However, it
209 does not provide a detailed representation of the interaction between PCFs, gene expres-
210 sion and the toxicity endpoint. Features and genes identified in the models summarized in
211 Figure 2 were therefore extracted and used as an input to the well-validated network infer-
212 ence algorithm ARACNE¹⁶ to reconstruct the underlying structure of a biological network.
213 This results in a higher resolution map of the interaction between PCFs, individual gene
214 expression and toxicity outcome. To be able to interpret this network, highly interconnected
215 regions were identified which yielded two larger modules (503 and 469 nodes) and 3 smaller
216 modules (less than 10 nodes). One of the larger modules (Figure S2, module 1) contained

217 all but 1 PCFs (DISPp: Displacement value weighted by polarizability) and a subset of
218 341 genes (41% of the total number of genes). Within this sub-network ALOGPS_logP , a
219 representative feature of $\log K_{ow}$, was the node with the highest correlation to nLC_{50} (see
220 Figure S3) and lay at the interface between PCFs and gene sub-clusters. Extensive previous
221 work has demonstrated a link between $\log K_{ow}$ which provides a measure of lipophilicity of a
222 compound and toxicity in a variety of organisms including DM.^{5,9,26} Toxicity attributable to
223 this relationship has been designated baseline toxicity or narcosis although the underlying
224 mechanism(s) remain unclear.^{4,5}

225 **Genes correlated to $\log K_{ow}$ define a calcium response signature**

226 The mechanism(s) underlying narcosis which link lipophilicity and toxicity remain contro-
227 versial. We reasoned that the expression and the functional profile of genes correlated with
228 $\log K_{ow}$ might be informative of such mechanisms. A subsequent analysis of the dataset
229 revealed that 1846 and 2438 genes respectively were positively and negatively correlated
230 with $\log K_{ow}$ ($< 1\%$ FDR). These were grouped into 10 clusters ($r > 0.75$, Figure 3 and
231 Supplementary File 2 for more details). Surprisingly, the analysis of their expression profile
232 as a function of $\log K_{ow}$ reveals an inversion in the transcriptional response to chemical
233 exposure at approximately $\log K_{ow} \geq 1.8$ (Figure 3 and S58-S249). Functional enrichment
234 analysis of this gene expression signature (Figure 3) was consistent with the high level model
235 described above (Figure 2). Identification of the most represented molecular component in
236 these KEGG Pathways revealed that calcium pathways and related kinases such as ERK,
237 PKA and MAP2K1 were most represented (25, 21, 20 and 17 respectively out of 97 pathways,
238 Table S6).

239 **Ca²⁺ mobilization recapitulates the log K_{ow} transcriptional signa-**
240 **ture and explains a significant proportion of the response to single**
241 **chemical exposure**

242 Since functional analysis of the genes correlated with log K_{ow} suggested a link with Ca²⁺,
243 chemicals with high log K_{ow} might change membrane permeability causing an imbalance of
244 Ca²⁺ exchange within the endoplasmic reticulum and mitochondria. This perturbation may
245 ultimately lead to a change in cytoplasmic Ca²⁺ concentration and ultimately the observed
246 transcriptional response (Figure S10A-B). Indeed chemicals with log K_{ow} greater than 1.8
247 were more effective inhibitors of the Sarcoplasmic reticulum Ca²⁺ ATPase (SERCA) than
248 less lipophilic chemicals (p-value < 0.03, Figure S10C). In order to further evaluate the rela-
249 tionship between intracellular calcium levels, transcription and toxicity, DM were exposed to
250 a SERCA non-competitive inhibitor (thapsigargin) at a concentration (100nM) that has been
251 shown to be highly specific.^{27,28} This chemical induces a large increase in intracellular calcium
252 concentration by blocking the active transport of calcium in the endoplasmic reticulum. The
253 results show that thapsigargin is an effective inducer of transcription (Figure S6) and that
254 it is able to recapitulate 43% of the transcriptional signature linked to log K_{ow} (Figure 4).
255 Further analysis showed that on average 45% of the transcriptional response following single
256 chemical exposure can be explained by a log K_{ow} signature and that on average 35% can be
257 directly linked to the Ca²⁺ mobilization signature defined by thapsigargin (Table 1 , S2 and
258 Figure S7). To further validate these observations we reasoned that calcium release, induced
259 by highly lipophilic compounds might affect specific functions, which are highly dependent
260 on a tightly controlled Ca²⁺ concentration. This was explored through two separate ap-
261 proaches, 1) evaluation of calcium dependent myocardial contraction and found that highly
262 lipophilic compounds indeed change heart rates at concentrations below that which cause
263 any overt toxicity (Figure S8) and 2) through a semi-targeted metabolomics analysis of thap-
264 sigargin exposure. The metabolomics analysis revealed that exposure to thapsigargin at $\frac{1}{10}$

265 nLC₅₀ did not show an accompanied statistical significant change in metabolites (Figure S5)
 266 suggesting that calcium mobilization may precede toxicity rather than being a consequence.
 267 In fact, a metabolic response characterised by an increase in formate, alanine, lactic acid and
 268 glycerophosphocholine and a decrease of glucose, tyrosine and trimethyl-N-oxide was only
 269 detected at the much higher dose (nLC₅₀).

Table 1: Table showing the percentage of associated genes to $\log K_{ow}$ and Thapsigargin.

	Significant Genes	$\log K_{ow}$ (%)	Thapsigargin (%)
PonasteroneA	4822	43.05%	31.48%
Trichloroethylene	5540	43.03%	28.23%
Toluene	1868	39.56%	40.10%
Atrazine	3099	38.50%	40.88%
Dichlorobenzene	4269	40.55%	34.15%
Beta-estradiol	4270	40.40%	31.69%
Parathion	5330	43.28%	35.55%
Diazinon	5208	38.50%	32.05%
Phenanthrene	3395	37.47%	33.46%
Pyriproxyfen	4793	45.52%	28.90%
Methoxychlor	6533	43.52%	30.90%
Chlorpyrifos	2787	38.72%	31.40%
Toxaphene	5127	46.93%	31.79%
Methylfarnesoate	4810	38.77%	30.91%
Bifenthrin	4889	40.52%	30.37%
Lamda-Cyhalothrin	4891	43.14%	31.75%
Nonylphenol	3011	46.43%	28.30%
Permethrin	2504	41.13%	41.81%

270 **Integration of Ca²⁺ dependent transcriptional signatures and com-**
 271 **pound lipophilicity is required for optimal prediction of chemical**
 272 **toxicity**

273 The observed link between lipophilicity, calcium signalling and toxicity suggested that a pre-
 274 dictive model including calcium signalling transcriptional response may be a better predictor
 275 of toxicity than a model based on $\log K_{ow}$ on its own. First, a QSAR linking toxicity and

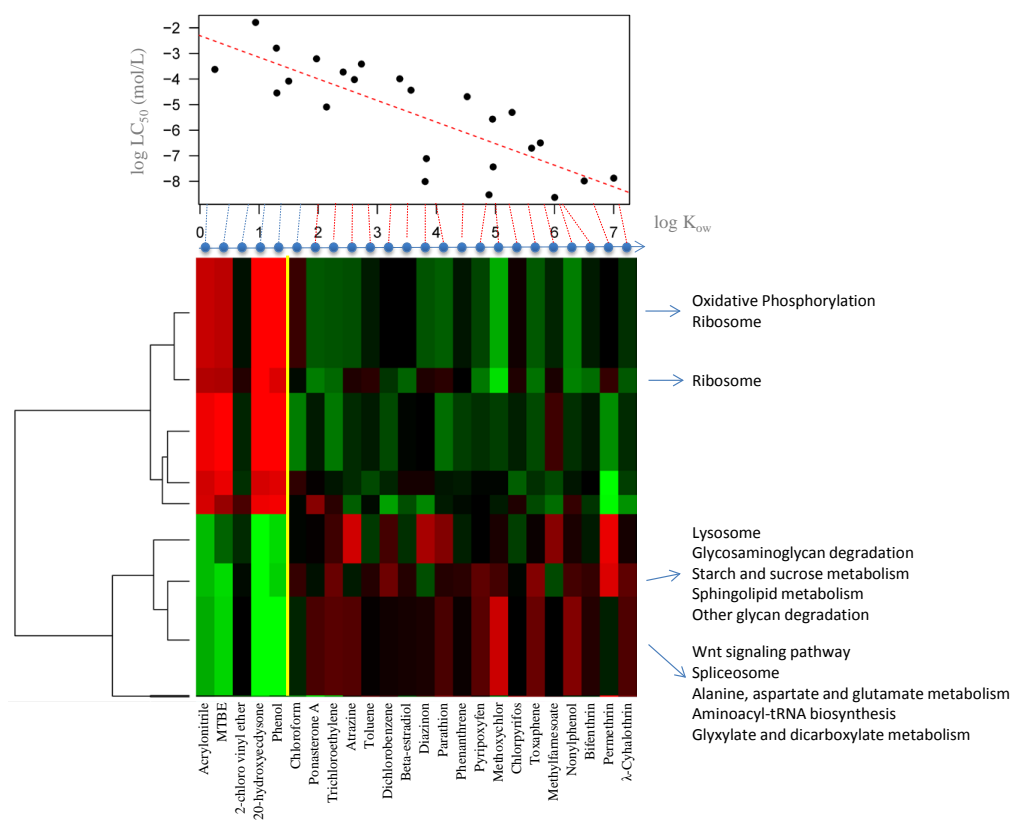


Figure 3: Heatmap of clusters of genes associated to ALOGPS_logP ordered by increasing lipophilicity. The graph above the heatmap shows the relationship between LC_{50} and experimental $\log K_{ow}$ with lines indicating the position of the compound in both plots. A transcriptional inversion at an ALOGPS_logP value of 1.8 is visible. The height of the heatmap block is representative to the number of genes within that cluster. Significantly enriched functional groups are represented for every given cluster.

lipophilicity only in this dataset was developed, which resulted in an intercept of -2.3078 and a $\log K_{ow}$ coefficient of -0.8438 ($R^2 = 0.65$), which is within the boundaries of QSAR developed by von der Ohe et al.⁹. To explore whether addition of Ca^{2+} associated genes will improve the predictive model a genetic algorithm, to select optimal sets of Ca^{2+} associated genes, was applied. The resulting model included 7 genes and their interaction with $\log K_{ow}$ ($R^2 = 0.966$, Table S5), which is a significant improvement in the prediction of toxicity as compared to lipophilicity on its own ($R^2 = 0.65$). The seven genes identified represented 3 functional groups, energy: phosphoglycerate kinase (PGK), methylenetetrahydrofolate reductase (metF), signalling: F-box and WD-40 domain protein (FBXW1_11), collagen type IV alpha (COL4A), and metabolism: large subunit ribosomal protein L44e (RP-L44e), serine palmitoyltransferase (E2.3.1.50) and sodium-dependent inorganic phosphate cotransporter (SLC17A5) (Table S5). A more detailed look at the coefficients revealed that metF, SLC17A5 and RP-L44e all contribute greater than $\log K_{ow}$ alone. Interestingly, interactions between $\log K_{ow}$ and gene expression only added little information towards the final model.

Discussion

This manuscript describes the first example of an experimentally validated integration of traditional QSAR analysis, functional genomics and ecotoxicology in a quantitative and predictive computational framework. This approach led to formulating a working model explaining the molecular basis of the basal toxicity of lipophilic chemicals in DM which could have broad implications in toxicology. Basal toxicity, also commonly termed narcosis, has classically been attributed to two related lipophilic compounds, type 1 (non-polar lipophilic compounds) and type 2 (polar lipophilic compounds), with similar but distinct toxicities. On the mechanistic level this difference has been hypothesized to be due to the physical characteristics of polar and non-polar compounds and their interaction with cellular membranes. Polar compounds (type 2) are hypothesized to disrupt membranes by binding between the

Thapsigargin

	up	down
Correlated with log K_{ow} +	757/1846 130 (7%)	3045/1846 189 (10%)
Correlated with log K_{ow} -	757/2438 69 (3%)	3045/2438 879(36%)

Figure 4: Gene-level comparison of the gene lists between thapsigargin and log K_{ow} FL showed a 43% overlap.

301 polar functional groups on the lipophilic compound and the polar head groups of membrane
302 lipids (e.g phosphatidyl choline head group). In contrast, non-polar compounds (type 1)
303 disrupt membrane integrity through direct interactions with the hydrophobic membrane in-
304 terior.⁶ Compounds with lower log K_{ow} do not generally partition into the lipid phase and
305 therefore may not disrupt membranes directly.

306 **Intracellular calcium mobilization a mechanism for basal toxicity**

307 Basal toxicity or narcosis is believed to result from alterations in membrane integrity due to
308 the partition of toxic chemicals into biological membranes. The findings presented in this
309 manuscript support the hypothesis that an early event in basal toxicity is disrupted calcium
310 homeostasis perhaps secondary to disrupted membrane integrity or direct inhibition of cal-
311 cium transport. In summary, the key findings supporting the calcium hypothesis are; 1) a
312 considerable fraction (an average of 50%) of the transcriptional response across all chemicals
313 correlates with log K_{ow} , 2) increase in intracellular calcium reproduces this transcriptional
314 response, 3) lipophilic chemicals (log $K_{ow} \geq 1.8$) are better SERCA inhibitors, and 4) ex-
315 pression signatures linked to calcium release are predictive of toxicity. Due to the ambiguous
316 nature of baseline toxicity however it is unclear as to how many mechanisms may be repre-
317 sented in the explored chemical space. Another limiting factor, which mainly apply to points
318 1 and 2 of these key findings are the timing and dose that were used in the thapsigargin
319 exposure. It is conceivable that the observed transcriptional response is likely to follow dif-
320 ferent dynamics depending on the compound and concentration used in the exposure. The
321 results here however provide additional strong evidence that supports the hypothesis that
322 calcium release is a molecular initiating event in the identified narcosis mechanism.

323 Ca^{2+} movement from the ER to the cytoplasm and mitochondria can trigger biological
324 processes leading to cell death. For example, increased mitochondria calcium levels can ac-
325 tivate the mitochondrial permeability transition pore (MPTP)²⁹ leading to mitochondrial
326 swelling and cell death through apoptosis or necrosis.³⁰ Moreover, depletion of Ca^{2+} in the

ER can result in inhibition of the entire protein translation machinery via a wide range of mechanisms.^{31–33} Several findings support the hypothesis that the Ca^{2+} dependent transcriptional signature is a "molecular initiating event" rather than a consequence of an already on-going tissue degeneration process. First, the calcium transcriptional response signature appear at a much lower dose and at an earlier time than any effect can be detected in the DM immobilization/toxicity assay. Moreover, NMR metabolomics analyses confirms that despite the transcriptional response observed after exposure to thapsigargin at $\frac{1}{10}$ LC_{50} up to 24 hours, there is no observable difference in metabolite concentrations (Figure S5A and C). It is only at much higher dose (LC_{50}) that metabolic responses consistent with transcriptional effect are measured (Figure S5). This finding suggests that the transcriptional signature observed at 24h is a molecular event likely to precede toxicity manifestation since metabolism can be considered highly sensitive. Heart rate analysis revealed greater perturbations in response to highly lipophilic chemicals (Figure S8) which supports this hypothesis of early effects on calcium prior to significant mortality.

Relationship between basal and target-specific toxicity

Minimizing adverse impact on biodiversity and human health is the focus of chemical risk assessment. To aid in understanding such effects work by us and others has focused on identifying chemical specific mechanisms of toxicity.^{1,2} Basal toxicity may represent the primary mechanism of some chemicals and at the least can contribute significantly to the toxicity of a chemical. This work indicates that a common calcium-dependent mechanism may underlie the basal toxicity of lipophilic chemicals. In contrast, specific toxicity mechanism (endocrine disruption) shows little overlap with basal toxicity and involves a relatively small number of genes when compared to a basal toxicity response, even at sub-lethal doses (Figure S9). This suggests that the toxicity of some chemicals will result from both the calcium dependent basal toxicity mechanism and additional specific toxicity mechanisms, i.e. receptor-mediated.

352 **Predictive ecotoxicology as a means towards understanding chem-** 353 **ical toxicity**

354 Omics technologies have dramatically increased the ability to characterize the molecular
355 responses of virtually any species to chemical exposure. While there have been concerns
356 about the utility of this approach for discovering informative biomarkers,^{34,35} a number of
357 groups have shown that the use of sophisticated computational approaches to link molecu-
358 lar response to toxicity endpoints can be a very effective tool to discover mechanism based
359 biomarkers.^{2,3,36,37} Biales et al, for example, developed predictive models to inform classical
360 toxicity identification evaluation and have shown that gene expression profiles are highly sen-
361 sitive even at sub-lethal concentrations. This work provides further evidence of the potential
362 of predictive toxicology in the ecotoxicology arena and suggests that molecular responses
363 linked to disrupted calcium homeostasis secondary to membrane perturbation may be com-
364 mon in molecular ecotoxicology studies of diverse species. Moreover, it is possible that such
365 mechanisms are similarly relevant in human drug toxicity. Inter-species conservation of basal
366 toxicity could provide an improved AOP framework for inter species extrapolation of toxicity.

367 **Funding Sources**

368 This work was supported by a Natural Environment Research Council (NERC) Grant
369 [NE/1028246/2] to FF, a NSF (CBET-1066358) grant to CV, by the Department of Biotech-
370 nology, Govt. of India under CRESTprogramme to AG, and support to TW by the NERC
371 and Cefas seedcorn funding through project DP247.

372 Supporting Information Available

373 This information is available free of charge via the Internet at <http://pubs.acs.org>.

374 **References**

- 375 (1) Antczak, P.; Jo, H. J.; Woo, S.; Scanlan, L.; Poynton, H.; Loguinov, A.; Chan, S.;
376 Falciani, F.; Vulpe, C. Molecular Toxicity Identification Evaluation (mTIE) Approach
377 Predicts Chemical Exposure in *Daphnia magna*. *Environmental Science & Technology*
378 **2013**, *47*, 11747–11756.
- 379 (2) Biales, A. D.; Kostich, M.; Burgess, R. M.; Ho, K. T.; Bencic, D. C.; Flick, R. L.; Por-
380 tis, L. M.; Pelletier, M. C.; Perron, M. M.; Reiss, M. Linkage of Genomic Biomarkers
381 to Whole Organism End Points in a Toxicity Identification Evaluation (TIE). *Environ-*
382 *mental Science & Technology* **2013**, *47*, 1306–1312.
- 383 (3) Ankley, G. T.; Bennett, R. S.; Erickson, R. J.; Hoff, D. J.; Hornung, M. W.; John-
384 son, R. D.; Mount, D. R.; Nichols, J. W.; Russom, C. L.; Schmieder, P. K.; Ser-
385 rrano, J. A.; Tietge, J. E.; Villeneuve, D. L. Adverse outcome pathways: a conceptual
386 framework to support ecotoxicology research and risk assessment. *Environmental Tox-*
387 *icology and Chemistry* **2010**, *29*, 730–741.
- 388 (4) Escher, B. I.; Hermens, J. L. Modes of action in ecotoxicology: their role in body
389 burdens, species sensitivity, QSARs, and mixture effects. *Environmental Science &*
390 *Technology* **2002**, *36*, 4201–4217.
- 391 (5) Verhaar, H. J.; Urrestarazu Ramos, E.; Hermens, J. L. Classifying environmental pol-
392 lutants. 2: Separation of class 1 (baseline toxicity) and class 2 ('polar narcosis') type
393 compounds based on chemical descriptors. *Journal of Chemometrics* **1996**, *10*, 149–162.
- 394 (6) Roberts, D. W.; Costello, J. F. Mechanisms of action for general and polar narcosis: A
395 difference in dimension. *QSAR & Combinatorial Science* **2003**, *22*, 226–233.
- 396 (7) Bradbury, S. P.; Carlson, R. W.; Henry, T. R. Polar narcosis in aquatic organisms.
397 *Aquatic Toxicology and Hazard Assessment* **1989**, *12*, 59–73.

- 398 (8) Kilham, S. S.; Kreeger, D. A.; Lynn, S. G.; Goulden, C. E.; Herrera, L. COMBO: a
399 defined freshwater culture medium for algae and zooplankton. *Hydrobiologia* **1998**, *377*,
400 147–159.
- 401 (9) von der Ohe, P. C.; Kühne, R.; Ebert, R.-U.; Altenburger, R.; Liess, M.;
402 Schüürmann, G. Structural Alerts A New Classification Model to Discriminate Ex-
403 cess Toxicity from Narcotic Effect Levels of Organic Compounds in the Acute Daphnid
404 Assay. *Chemical research in toxicology* **2005**, *18*, 536–555.
- 405 (10) Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Pa-
406 lyulin, V. A.; Radchenko, E. V.; Zefirov, N. S.; Makarenko, A. S.; Vsevolod, Y. T.;
407 Prokopenko, V. V. Virtual computational chemistry laboratory—design and description.
408 *Journal of Computer-Aided Molecular Design* **2005**, *19*, 453–463.
- 409 (11) Lee, E.; Chuang, H.-Y.; Kim, J.-W.; Ideker, T.; Lee, D. Inferring pathway activity
410 toward precise disease classification. *PLoS Computational Biology* **2008**, *4*, e1000217.
- 411 (12) Ekins, S.; Andreyev, S.; Ryabov, A.; Kirillov, E.; Rakhmatulin, E. A.; Sorokina, S.;
412 Bugrim, A.; Nikolskaya, T. A combined approach to drug metabolism and toxicity
413 assessment. *Drug Metabolism and Disposition* **2006**, *34*, 495–503.
- 414 (13) Antczak, P.; Ortega, F.; Chipman, J. K.; Falciani, F. Mapping Drug Physico-Chemical
415 Features to Pathway Activity Reveals Molecular Networks Linked to Toxicity Outcome.
416 *PloS ONE* **2010**, *5*, e12385.
- 417 (14) Trevino, V.; Falciani, F. GALGO: an R package for multivariate variable selection using
418 genetic algorithms. *Bioinformatics* **2006**, *22*, 1154–1156.
- 419 (15) Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32.
- 420 (16) Margolin, A. A.; Nemenman, I.; Basso, K.; Wiggins, C.; Stolovitzky, G.; Favera, R. D.;

- 421 Califano, A. ARACNE: an algorithm for the reconstruction of gene regulatory networks
422 in a mammalian cellular context. *BMC Bioinformatics* **2006**, *7*, S7.
- 423 (17) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.;
424 Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models
425 of biomolecular interaction networks. *Genome Research* **2003**, *13*, 2498–2504.
- 426 (18) Morris, J. H.; Apeltsin, L.; Newman, A. M.; Baumbach, J.; Wittkop, T.; Su, G.;
427 Bader, G. D.; Ferrin, T. E. clusterMaker: a multi-algorithm clustering plugin for Cy-
428 toscape. *BMC Bioinformatics* **2011**, *12*, 436.
- 429 (19) Tusher, V. G.; Tibshirani, R.; Chu, G. Significance analysis of microarrays applied to
430 the ionizing radiation response. *Proceedings of the National Academy of Sciences* **2001**,
431 *98*, 5116–5121.
- 432 (20) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. Systematic and integrative analysis
433 of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **2008**, *4*,
434 44–57.
- 435 (21) Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.;
436 Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.; Mesirov, J. P.
437 Gene set enrichment analysis: a knowledge-based approach for interpreting genome-
438 wide expression profiles. *Proceedings of the National Academy of Sciences of the United*
439 *States of America* **2005**, *102*, 15545–15550.
- 440 (22) Michelangeli, F.; Munkonge, F. M. Methods of reconstitution of the purified sarcoplas-
441 mic reticulum (Ca²⁺- Mg²⁺)-ATPase using bile salt detergents to form membranes of
442 defined lipid to protein ratios or sealed vesicles. *Analytical Biochemistry* **1991**, *194*,
443 231–236.
- 444 (23) Southam, A. D.; Lange, A.; Hines, A.; Hill, E. M.; Katsu, Y.; Iguchi, T.; Tyler, C. R.;
445 Viant, M. R. Metabolomics reveals target and off-target toxicities of a model

- 446 organophosphate pesticide to roach (*Rutilus rutilus*): Implications for biomonitoring.
447 *Environmental Science & Technology* **2011**, *45*, 3759–3767.
- 448 (24) Cui, Q.; Lewis, I. A.; Hegeman, A. D.; Anderson, M. E.; Li, J.; Schulte, C. F.; West-
449 ler, W. M.; Eghbalnia, H. R.; Sussman, M. R.; Markley, J. L. Metabolite identification
450 via the Madison metabolomics consortium database. *Nature Biotechnology* **2008**, *26*,
451 162–164.
- 452 (25) Dzialowski, E. M.; Turner, P. K.; Brooks, B. W. Physiological and reproductive effects
453 of beta adrenergic receptor antagonists in *Daphnia magna*. *Archives of Environmental*
454 *Contamination And Toxicology* **2006**, *50*, 503–510.
- 455 (26) Claeys, L.; Iaccino, F.; Janssen, C. R.; Van Sprang, P.; Verdonck, F. Development and
456 validation of a quantitative structure–activity relationship for chronic narcosis to fish.
457 *Environmental Toxicology and Chemistry* **2013**, *32*, 2217–2225.
- 458 (27) Thastrup, O.; Cullen, P. J.; Drobak, B.; Hanley, M. R.; Dawson, A. P. Thapsigargin, a
459 tumor promoter, discharges intracellular Ca²⁺ stores by specific inhibition of the endo-
460 plasmic reticulum Ca²⁺ (+)-ATPase. *Proceedings of the National Academy of Sciences*
461 **1990**, *87*, 2466–2470.
- 462 (28) Jackson, T. R.; Patterson, S.; Thastrup, O.; Hanley, M. A novel tumour promoter,
463 thapsigargin, transiently increases cytoplasmic free Ca²⁺ without generation of inositol
464 phosphates in NG115-401L neuronal cells. *Biochem. J* **1988**, *253*, 81–86.
- 465 (29) Bernardi, P.; Scorrano, L.; Colonna, R.; Petronilli, V.; Di Lisa, F. Mitochondria and
466 cell death. *European Journal of Biochemistry* **1999**, *264*, 687–701.
- 467 (30) Qian, T.; Herman, B.; Lemasters, J. J. The mitochondrial permeability transition medi-
468 ates both necrotic and apoptotic death of hepatocytes exposed to Br-A23187. *Toxicology*
469 *and Applied Pharmacology* **1999**, *154*, 117–125.

- 470 (31) Brostrom, M. A.; Brostrom, C. O. Calcium dynamics and endoplasmic reticular function
471 in the regulation of protein synthesis: implications for cell growth and adaptability. *Cell*
472 *Calcium* **2003**, *34*, 345–363.
- 473 (32) Berridge, M. J. The endoplasmic reticulum: a multifunctional signaling organelle. *Cell*
474 *Calcium* **2002**, *32*, 235–249.
- 475 (33) Jacobson, J.; Duchen, M. R. Mitochondrial oxidative stress and cell death in astro-
476 cytesrequirement for stored Ca²⁺ and sustained opening of the permeability transition
477 pore. *Journal of Cell Science* **2002**, *115*, 1175–1188.
- 478 (34) Fent, K.; Sumpter, J. P. Progress and promises in toxicogenomics in aquatic toxicology:
479 Is technical innovation driving scientific innovation? *Aquatic Toxicology* **2011**, *105*, 25–
480 39.
- 481 (35) Sumpter, J. P. Endocrine disruption in wildlife: the future? *Pure and applied chemistry*
482 **2003**, *75*, 2355–2360.
- 483 (36) Poynton, H. C.; Varshavsky, J. R.; Chang, B.; Cavigliolo, G.; Chan, S.; Holman, P. S.;
484 Loguinov, A. V.; Bauer, D. J.; Komachi, K.; Theil, E. C.; Perkins, E. J.; Hughes, O.;
485 Vulpe, C. D. *Daphnia magna* ecotoxicogenomics provides mechanistic insights into
486 metal toxicity. *Environmental Science & Technology* **2007**, *41*, 1044–1050.
- 487 (37) Williams, T. D. et al. Towards a system level understanding of non-model organisms
488 sampled from the environment: a network biology approach. *PLoS Computational Bi-*
489 *ology* **2011**, *7*, e1002126.