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

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

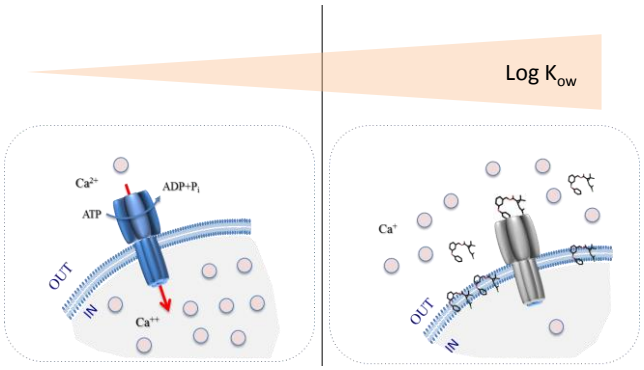
Phone: +1-510-642-1834; +44-151-795-4558. Fax: +1-510-643-3132; +44-151-795-4408

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

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



Introduction

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

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

a priori hypothesis.

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

Materials and Methods

Analysis strategy

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

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

Chemical Exposures and expression profiling dataset

Transcriptional data from exposure of DM to twenty six organic chemicals were initially selected from a previous study.[1] Briefly, this set represented gene expression profiles for 14 day old DM adults exposed to $\frac{1}{10}$ of the identified nLC_{50} (calculated using neonatal DM). At the validation stage of the project, an additional exposure to thapsigargin was performed. For more details on the compounds see Table S5. Exposures were performed with twenty two week old DM in 1L of modified COMBO media[8] containing $\frac{1}{10}$ LC_{50} concentration of each of the chemical in four replicates. Each beaker was carefully sealed with clingfilm to reduce volatility and to improve delivery. After a 24h exposure, total RNA was extracted and arrayed using a custom Agilent microarray (AMAID: 023710, GPL15139). (Further information on the selected chemicals can be found in Supplementary Table S5 and S8). To verify that the experimental precautions taken to reduce evaporation were effective chemical concentrations were measured during exposure to the 3 most volatile compounds. The analysis revealed that the concentration of 2 of the volatile compounds were reduced by 50% at the end of the exposure while the remaining chemical showed $< 1\%$ loss. This shows that compound loss caused by evaporation is not likely to affect LC_{50} determination (See Supplementary Table S9 for more details).

Basal toxicity model and excess toxicity

To establish whether the compounds in this dataset fit a baseline toxicity model, a comparison with an already developed model[9] was performed. First a model, based on this dataset,

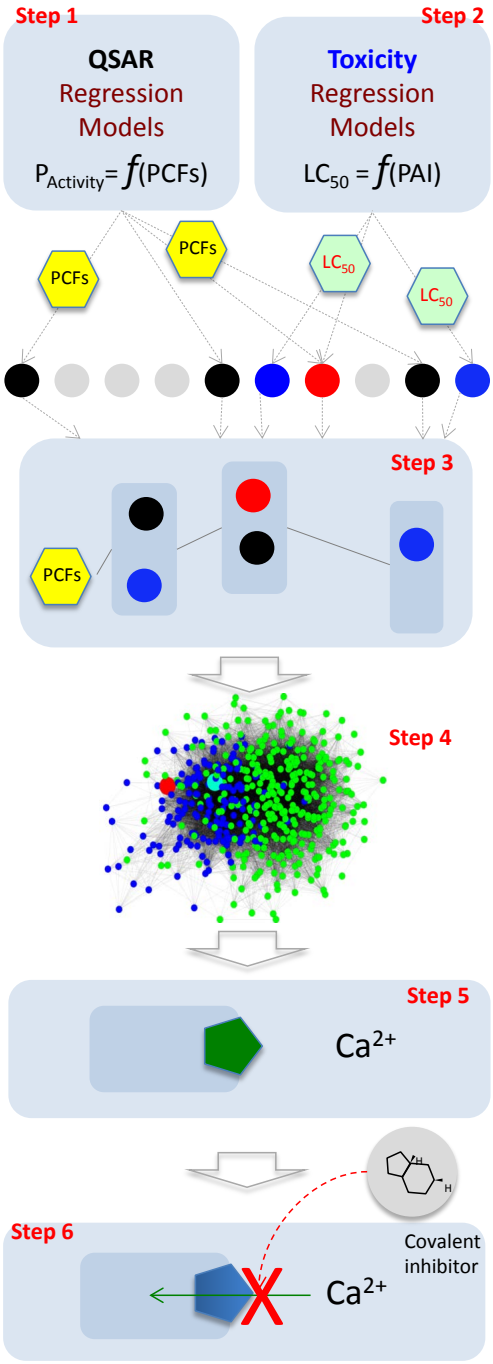


Figure 1: Step 1: Predicting pathway activity from PCFs. Step 2: Predicting toxicity from pathway activity. Step 3: Visualization of the results in a KEGG Pathway Interaction map. Step 4: Network reconstruction of identified genes and PCFs. Step 5: Hypothesis generation through integration of Step 3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

was derived:

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

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

$$Te = \frac{predictedLC_{50}}{experimentalLC_{50}} \quad (2)$$

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

Calculation of PCFs

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

Calculation of indexes of pathway activity (PAI)

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

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

Toxicity endpoint

Organismal toxicity was determined in the initial generated dataset.[1] Briefly, neonates ($< 24h$) were exposed to varying nominal concentrations of each chemical over 24h and nominal LC_{50}s generated. This neonate LC_{50} (nLC_{50}) was then used to link the transcriptional response of 14 day old adults at $\frac{1}{10}$ nLC_{50}.

Pathway activity as a function of chemical features

Using an advanced machine learning technique (GALGO[14]) optimized sub-sets of 3 PCFs were identified, which are able to predict each of the 285 PAIs (3 PCs x 95 pathways) using the following randomForest model:

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

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

Toxicity as a function of Pathway activity

For each pathway a random Forest[15] regression model was used to identify pathways predictive of toxicity (nLC_{50}). The regression model was defined as:

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

To identify statistically significant pathways, nLC $_{50}$ are randomized 1000 times and the model rerun. Twenty significant pathways linked to toxicity were identified with an $R^2 > 0.6$.

Analysis of Gene expression dataset

To develop a KEGG Pathway map links between KEGG pathways are represented as the Jaccard's Index of overlap, which is defined as the ratio between the size of the intersect over the size of the union of any 2 samples. To aid in interpretation, pathways were ordered into higher functional groups and coloured on the basis of their association (PCFs black, toxicity blue and both red). To further build a network representing the dependency between genes, PCFs and LC_{50} based on the KEGG map, ARACNE[16] was applied (p-value 10^{-8}). To identify highly interconnected sub-networks GLay (clusterMaker[17,18]) was applied. A force driven layout was used to represent the graph. Statistically significant correlation between genes and experimental log K_{ow} was identified using SAM (significance analysis for microarrays).[19] Significant over-representation was identified using a modified fisher test as described in.[20] To identify metabolites or genes highly represented by the log K_{ow} signature KEGG reference pathways with at least 5 represented members in this dataset were considered and the fisher test applied. To link exposure of thaspigargin and the remaining dataset genes are ranked by d-statistic from the above SAM analysis and used as ranked input to GSEA.[21] Genesets were defined as thapsigargin significantly differentially expressed genes (FDR $< 20\%$). To define contribution of specific vs. basal toxicity mechanisms a 2 factor ANOVA was used to

identify whether the majority of the variation observed in the dataset was associated with chemical class (as defined in[1]), log K_{ow} or their interaction.

Validation of observed Ca^{2+} effects

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

Development of a model predictive of toxicity and integrating calcium signalling expression signatures and lipophilicity

A machine learning approach (GALGO[14]) was used to assess whether inclusion of calcium associated genes to log K_{ow} could produce a better predictive model than log K_{ow} by itself.

Results

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

The first objective of this study was to identify putative AOPs representing a link between compound PCFs, pathway activity and toxicity outcome (Figure 1). Remarkably, 35 out of the 95 (36%) KEGG pathways could be identified whose activity can be predicted as a 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.

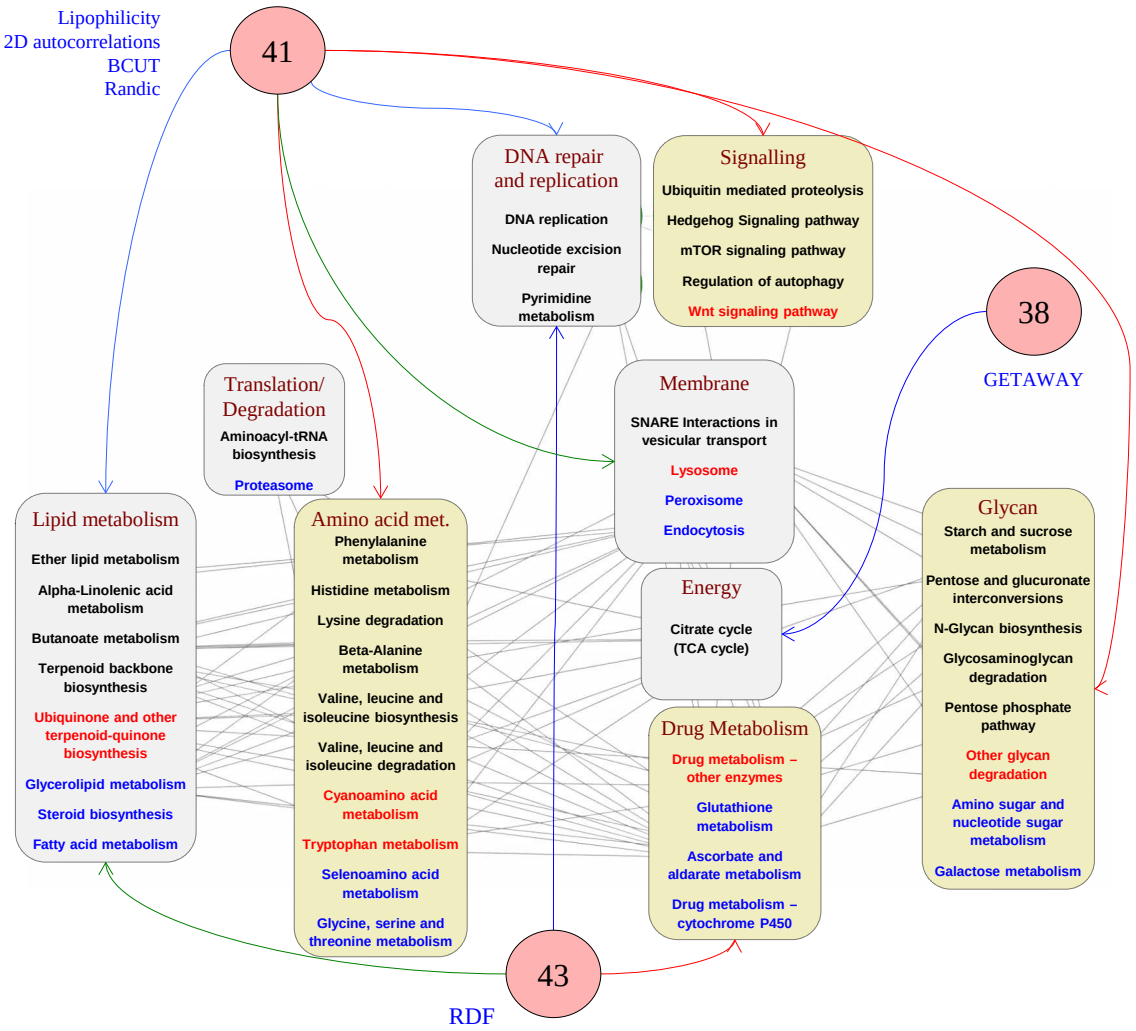


Figure 2: Representation of the interactions between functional clusters and groups of PCFs. Each functional cluster represents a number of related pathways associated to either PCFs (black), toxicity (blue) and both (red). Strength of interactions between PCFs and functional clusters is represented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

PCFs clusters associate with functional groups

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

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

The previously developed high level map already provided indications on the possible points of interactions between compounds (based on PCFs) and pathway activity. However, it does not provide a detailed representation of the interaction between PCFs, gene expression and the toxicity endpoint. Features and genes identified in the models summarized in Figure 2 were therefore extracted and used as an input to the well-validated network inference algorithm ARACNE[16] to reconstruct the underlying structure of a biological network. This results in a higher resolution map of the interaction between PCFs, individual gene expression and toxicity outcome. To be able to interpret this network, highly interconnected regions were identified which yielded two larger modules (503 and 469 nodes) and 3 smaller modules (less than 10 nodes). One of the larger modules (Figure S2, module 1) contained

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

Genes correlated to log K_{ow} define a calcium response signature

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

Ca^{2+} mobilization recapitulates the log K_{ow} transcriptional signature and explains a significant proportion of the response to single chemical exposure

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

nLC_{50} did not show an accompanied statistical significant change in metabolites (Figure S5) suggesting that calcium mobilization may precede toxicity rather than being a consequence. In fact, a metabolic response characterised by an increase in formate, alanine, lactic acid and glycerophosphocholine and a decrease of glucose, tyrosine and trimethyl-N-oxide was only detected at the much higher dose (nLC_{50}).

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

	Significant Genes	$logK_{ow}$ (%)	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%
Pyripoxyfen	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%

Integration of Ca^{2+} dependent transcriptional signatures and compound lipophilicity is required for optimal prediction of chemical toxicity

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

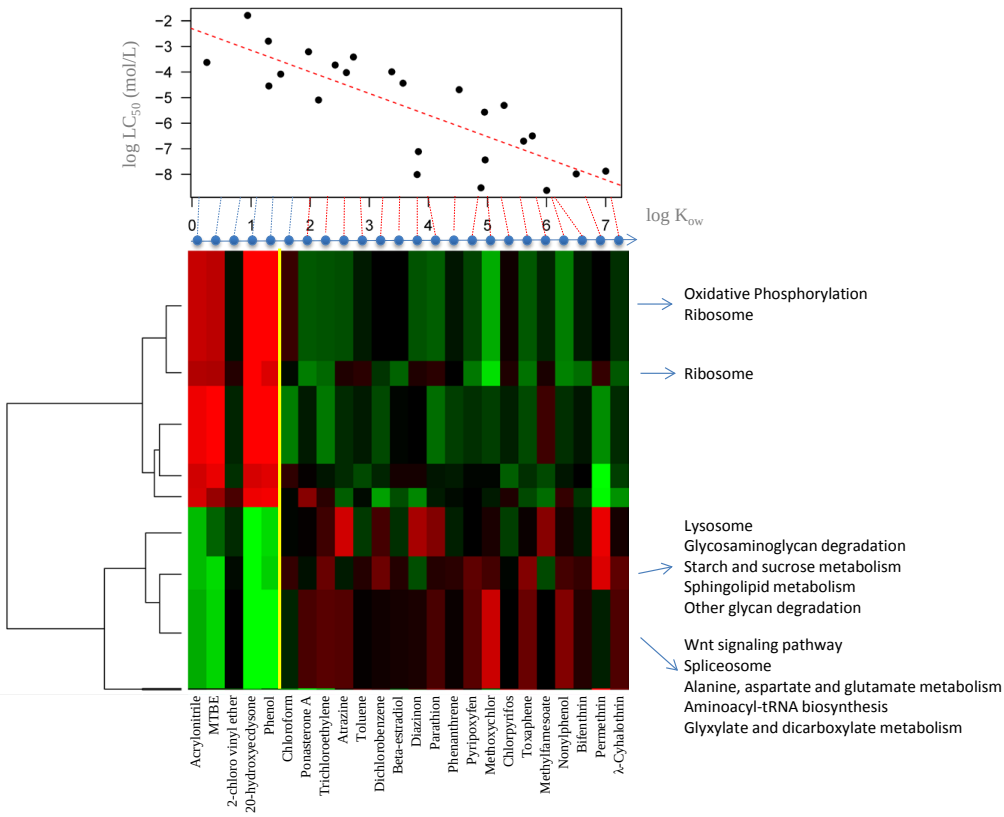


Figure 3: Heatmap of clusters of genes associated to ALOGPS_logP ordered by increasing lipophilicity. The graph above the heatmap shows the relationship between LC_{50} and experimental log K_{ow} with lines indicating the position of the compound in both plots. A transcriptional inversion at an ALOGPS_logP 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

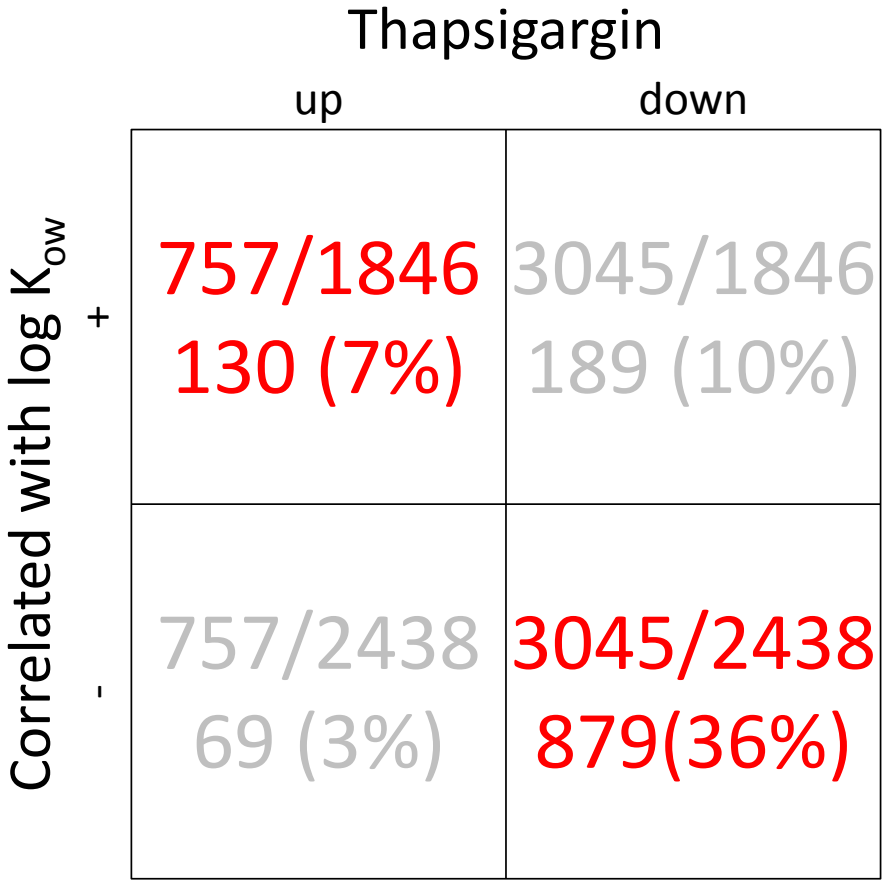


Figure 4: Gene-level comparison of the gene lists between thapsigargin and log K_{ow} FL showed a 43% overlap.

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

Intracellular calcium mobilization a mechanism for basal toxicity

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

Ca^{2+} movement from the ER to the cytoplasm and mitochondria can trigger biological processes leading to cell death. For example, increased mitochondria calcium levels can activate the mitochondrial permeability transition pore (MPTP) [29] leading to mitochondrial 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.

Predictive ecotoxicology as a means towards understanding chemical toxicity

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

Funding Sources

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

Supporting Information Available

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

References

(1) Antczak, P.; Jo, H. J.; Woo, S.; Scanlan, L.; Poynton, H.; Loguinov, A.; Chan, S.; Falciani, F.; Vulpe, C. Molecular Toxicity Identification Evaluation (mTIE) Approach Predicts Chemical Exposure in Daphnia magna. *Environmental Science & Technology* **2013**, *47*, 11747–11756.

(2) Biales, A. D.; Kostich, M.; Burgess, R. M.; Ho, K. T.; Bencic, D. C.; Flick, R. L.; Portis, L. M.; Pelletier, M. C.; Perron, M. M.; Reiss, M. Linkage of Genomic Biomarkers to Whole Organism End Points in a Toxicity Identification Evaluation (TIE). *Environmental Science & Technology* **2013**, *47*, 1306–1312.

(3) Ankley, G. T.; Bennett, R. S.; Erickson, R. J.; Hoff, D. J.; Hornung, M. W.; Johnson, R. D.; Mount, D. R.; Nichols, J. W.; Russom, C. L.; Schmieder, P. K.; Serrrano, J. A.; Tietge, J. E.; Villeneuve, D. L. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environmental Toxicology and Chemistry* **2010**, *29*, 730–741.

(4) Escher, B. I.; Hermens, J. L. Modes of action in ecotoxicology: their role in body burdens, species sensitivity, QSARs, and mixture effects. *Environmental Science & Technology* **2002**, *36*, 4201–4217.

(5) Verhaar, H. J.; Urrestarazu Ramos, E.; Hermens, J. L. Classifying environmental pollutants. 2: Separation of class 1 (baseline toxicity) and class 2 ('polar narcosis') type compounds based on chemical descriptors. *Journal of Chemometrics* **1996**, *10*, 149–162.

(6) Roberts, D. W.; Costello, J. F. Mechanisms of action for general and polar narcosis: A difference in dimension. *QSAR & Combinatorial Science* **2003**, *22*, 226–233.

(7) Bradbury, S. P.; Carlson, R. W.; Henry, T. R. Polar narcosis in aquatic organisms. *Aquatic Toxicology and Hazard Assessment* **1989**, *12*, 59–73.

(8) Kilham, S. S.; Kreeger, D. A.; Lynn, S. G.; Goulden, C. E.; Herrera, L. COMBO: a defined freshwater culture medium for algae and zooplankton. *Hydrobiologia* **1998**, *377*, 147–159.

(9) von der Ohe, P. C.; Kühne, R.; Ebert, R.-U.; Altenburger, R.; Liess, M.; Schüürmann, G. Structural Alerts A New Classification Model to Discriminate Excess Toxicity from Narcotic Effect Levels of Organic Compounds in the Acute Daphnid Assay. *Chemical research in toxicology* **2005**, *18*, 536–555.

(10) 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.; Vsevolod, Y. T.; Prokopenko, V. V. Virtual computational chemistry laboratory–design and description. *Journal of Computer-Aided Molecular Design* **2005**, *19*, 453–463.

(11) Lee, E.; Chuang, H.-Y.; Kim, J.-W.; Ideker, T.; Lee, D. Inferring pathway activity toward precise disease classification. *PLoS Computational Biology* **2008**, *4*, e1000217.

(12) Ekins, S.; Andreyev, S.; Ryabov, A.; Kirillov, E.; Rakhmatulin, E. A.; Sorokina, S.; Bugrim, A.; Nikolskaya, T. A combined approach to drug metabolism and toxicity assessment. *Drug Metabolism and Disposition* **2006**, *34*, 495–503.

(13) Antczak, P.; Ortega, F.; Chipman, J. K.; Falciani, F. Mapping Drug Physico-Chemical Features to Pathway Activity Reveals Molecular Networks Linked to Toxicity Outcome. *PloS ONE* **2010**, *5*, e12385.

(14) Trevino, V.; Falciani, F. GALGO: an R package for multivariate variable selection using genetic algorithms. *Bioinformatics* **2006**, *22*, 1154–1156.

(15) Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32.

(16) Margolin, A. A.; Nemenman, I.; Basso, K.; Wiggins, C.; Stolovitzky, G.; Favera, R. D.;

Califano, A. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **2006**, *7*, S7.

(17) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **2003**, *13*, 2498–2504.

(18) Morris, J. H.; Apeltsin, L.; Newman, A. M.; Baumbach, J.; Wittkop, T.; Su, G.; Bader, G. D.; Ferrin, T. E. clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics* **2011**, *12*, 436.

(19) Tusher, V. G.; Tibshirani, R.; Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences* **2001**, *98*, 5116–5121.

(20) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **2008**, *4*, 44–57.

(21) Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.; Mesirov, J. P. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 15545–15550.

(22) Michelangeli, F.; Munkonge, F. M. Methods of reconstitution of the purified sarcoplasmic reticulum (Ca^{2+}- Mg^{2+})-ATPase using bile salt detergents to form membranes of defined lipid to protein ratios or sealed vesicles. *Analytical Biochemistry* **1991**, *194*, 231–236.

(23) Southam, A. D.; Lange, A.; Hines, A.; Hill, E. M.; Katsu, Y.; Iguchi, T.; Tyler, C. R.; Viant, M. R. Metabolomics reveals target and off-target toxicities of a model

organophosphate pesticide to roach (*Rutilus rutilus*): Implications for biomonitoring. *Environmental Science & Technology* **2011**, *45*, 3759–3767.

(24) Cui, Q.; Lewis, I. A.; Hegeman, A. D.; Anderson, M. E.; Li, J.; Schulte, C. F.; Westler, W. M.; Eghbalnia, H. R.; Sussman, M. R.; Markley, J. L. Metabolite identification via the Madison metabolomics consortium database. *Nature Biotechnology* **2008**, *26*, 162–164.

(25) Dzialowski, E. M.; Turner, P. K.; Brooks, B. W. Physiological and reproductive effects of beta adrenergic receptor antagonists in Daphnia magna. *Archives of Environmental Contamination And Toxicology* **2006**, *50*, 503–510.

(26) Claeys, L.; Iaccino, F.; Janssen, C. R.; Van Sprang, P.; Verdonck, F. Development and validation of a quantitative structure–activity relationship for chronic narcosis to fish. *Environmental Toxicology and Chemistry* **2013**, *32*, 2217–2225.

(27) Thastrup, O.; Cullen, P. J.; Drobak, B.; Hanley, M. R.; Dawson, A. P. Thapsigargin, a tumor promoter, discharges intracellular Ca2+ stores by specific inhibition of the endoplasmic reticulum Ca2 (+)-ATPase. *Proceedings of the National Academy of Sciences* **1990**, *87*, 2466–2470.

(28) Jackson, T. R.; Patterson, S.; Thastrup, O.; Hanley, M. A novel tumour promoter, thapsigargin, transiently increases cytoplasmic free Ca2+ without generation of inositol phosphates in NG115-401L neuronal cells. *Biochem. J* **1988**, *253*, 81–86.

(29) Bernardi, P.; Scorrano, L.; Colonna, R.; Petronilli, V.; Di Lisa, F. Mitochondria and cell death. *European Journal of Biochemistry* **1999**, *264*, 687–701.

(30) Qian, T.; Herman, B.; Lemasters, J. J. The mitochondrial permeability transition mediates both necrotic and apoptotic death of hepatocytes exposed to Br-A23187. *Toxicology and Applied Pharmacology* **1999**, *154*, 117–125.

(31) Brostrom, M. A.; Brostrom, C. O. Calcium dynamics and endoplasmic reticular function in the regulation of protein synthesis: implications for cell growth and adaptability. *Cell Calcium* **2003**, *34*, 345–363.

(32) Berridge, M. J. The endoplasmic reticulum: a multifunctional signaling organelle. *Cell Calcium* **2002**, *32*, 235–249.

(33) Jacobson, J.; Duchen, M. R. Mitochondrial oxidative stress and cell death in astrocytesrequirement for stored Ca2+ and sustained opening of the permeability transition pore. *Journal of Cell Science* **2002**, *115*, 1175–1188.

(34) Fent, K.; Sumpter, J. P. Progress and promises in toxicogenomics in aquatic toxicology: Is technical innovation driving scientific innovation? *Aquatic Toxicology* **2011**, *105*, 25–39.

(35) Sumpter, J. P. Endocrine disruption in wildlife: the future? *Pure and applied chemistry* **2003**, *75*, 2355–2360.

(36) Poynton, H. C.; Varshavsky, J. R.; Chang, B.; Cavigiolio, G.; Chan, S.; Holman, P. S.; Loguinov, A. V.; Bauer, D. J.; Komachi, K.; Theil, E. C.; Perkins, E. J.; Hughes, O.; Vulpe, C. D. Daphnia magna ecotoxicogenomics provides mechanistic insights into metal toxicity. *Environmental Science & Technology* **2007**, *41*, 1044–1050.

(37) Williams, T. D. et al. Towards a system level understanding of non-model organisms sampled from the environment: a network biology approach. *PLoS Computational Biology* **2011**, *7*, e1002126.