**Supplementary Information**

Predicting chronic copper and nickel reproductive toxicity to *Daphnia pulex-pulicaria* from whole-animal, metaboliic profiles

Nadine S. Taylora,d, Jennifer A. Kirwanb, Craig Johnsona, Norman D. Yan c, Mark R. Viantb, John M. Gunnd, James C. McGeera

a Laurier Institute for Water Science, Wilfrid Laurier University, 75 University Avenue West, Waterloo, ON, N2L 3C5, Canada

b School of Biosciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, U.K.

c Department of Biology, York University, 4700 Keele Street, Toronto, ON, M3J 1P3

d Vale Living with Lakes Centre, Laurentian University, 935 Ramsey Lake Road, Sudbury, ON, P3E 2C6, Canada

**Corresponding Author**:

Nadine Taylor

School of Biosciences, University of Birmingham,

Edgbaston, Birmingham, 2TT, U.K.

Phone: +44 121 4145409

FAX: +44 121 4145925

Email: n.s.taylor@bham.ac.uk

**Water chemistry**

*Methods*

Water chemistry analyses were performed throughout the exposure studies on samples collected at test initiation and at each media renewal (taken from the media both entering and exiting the system). At each stage, pH and dissolved organic carbon levels (DOC; using total organic carbon analyser; TOC-LCPH, Shimadzu) were also recorded. Water samples (10 mL) were acidified to 1 % with trace metal grade HNO3 (Fisher Scientific, Nepean ON) and analysed for metal concentrations using an atomic absorption spectrometer (AAS; Varian Spectra AA.880; 0.45 µm filtered (Pall Acrodisc, VWR, Mississauga ON). The availabilities of Cu and Ni in free ion forms were calculated using the Windermere Humic Aqueous Model (WHAM ver 7.02; Tipping et al. 2011) with the parameters set from the mean measured pH, DOC and dissolved Cu and Ni and the known concentrations of components of the FLAMES media. For the purposes of the model the DOC was assumed to be 100% active with a ratio of 90:10 fulvic:humic acids (Al Reasi et al. 2012).

Measured water chemistry was a satisfactory match for nominal levels. The pH was relatively constant, averaging 6.35 in the acute study and 6.52 in the chronic exposure. Entering the exposure systems, there was a consistent background level of DOC in the FLAMES media (acute: 1.79 ± 0.07 mg/L, chronic: 1.04 ± 0.06 mg/L), Cu (acute: 0.00 ± 0.00 µg/L, chronic: 0.28 ± 0.08 µg/L), and Ni (acute: 0.04 ± 0.03, chronic: 0.07 ± 0.03 µg/L) with up to 60% of these measured metals included as micronutrients in the FLAMES medium. Full water chemistry details are provided in Tables SI1 and SI2.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **pH** | **DOC in (mg/L)a** | **DOC out (mg/L)b** | **Measured dissolved Cu concentration in (µg/L)a** | **Measured dissolved Cu concentration out (µg/L)b** | **Free ion Cu (µg/L)c** | **Measured dissolved Ni concentration in (µg/L)a** | **Measured dissolved Ni concentration out (µg/L)b** | **Free ion Ni (µg/L)c** | **Treatment****IDd** |
| 6.35 | 2.11 | 2.28 | 0.00 | 0.17 | 0.00 | 0.11 | 0.02 | 0.10 | Control |
| 6.43 | 1.67 | 1.76 | 0.89 | 0.92 | 0.01 | 0.00 | 0.01 | 0.00 | Low Cu |
| 6.35 | 1.95 | 2.76 | 1.79 | 1.77 | 0.13 | 0.02 | 0.00 | 0.02 | High Cu |
| 6.36 | 1.65 | 2.89 | 0.00 | 0.22 | 0.00 | 2.60 | 1.84 | 2.18 | Low Ni |
| 6.37 | 1.61 | 2.40 | 0.00 | 0.23 | 0.00 | 18.82 | 17.76 | 12.97 | High Ni |
| 6.39 | 1.55 | 1.93 | 0.86 | 0.99 | 0.61 | 2.98 | 1.81 | 2.48 | Low Cu + Low Ni |
| 6.35 | 1.57 | 2.64 | 0.87 | 1.07 | 0.58 | 18.64 | 17.25 | 12.86 | Low Cu + High Ni |
| 6.26 | 1.97 | 2.80 | 1.87 | 1.85 | 1.25 | 2.70 | 1.77 | 2.26 | High Cu + Low Ni |
| 6.29 | 2.02 | 2.21 | 1.71 | 1.90 | 1.13 | 18.56 | 17.41 | 12.82 | High Cu + High Ni |

Table SI1. Water chemistry measurements from samples obtained at the start and end of an acute (48 h) exposure of *D. pulex-pulicaria*to varying single and binary mixtures of Cu and Ni.

a Water samples taken from media entering the exposure system at test initiation

b Water samples taken from media exiting the exposure system at test end

c Calculated using WHAM VII, based on measured pH, measured DOC (in), measured dissolved Cu and Ni (in) and the components in the FLAMES media

d Treatment identifiers for metabolomics analyses

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **pH** | **DOC in (mg/L)a** | **DOC out (mg/L)b** | **Measured dissolved Cu concentration in (µg/L)a** | **Measured dissolved Cu concentration out (µg/L)b** | **Free ion Cu (µg/L)c** | **Measured dissolved Ni concentration in (µg/L)a** | **Measured dissolved Ni concentration out (µg/L)b** | **Free ion Ni (µg/L)c** | **Treatment****IDd** |
| 6.59 | 1.23 | 1.81 | 0.32 | 0.32 | 0.008 | 0.00 | 0.10 | 0.000 | Control |
| 6.53 | 1.21 | 1.63 | 0.97 | 0.97 | 0.315 | 0.11 | 0.17 | 0.101 | Low Cu |
| 6.46 | 0.92 | 1.66 | 1.82 | 1.89 | 0.788 | 0.09 | 0.09 | 0.083 | High Cu |
| 6.58 | 0.73 | 1.76 | 0.27 | 0.32 | 0.215 | 1.58 | 1.17 | 1.370 | Low Ni |
| 6.58 | 0.99 | 1.35 | 0.24 | 0.34 | 0.164 | 17.96 | 17.50 | 12.460 | High Ni |
| 6.54 | 1.23 | 1.63 | 0.99 | 1.00 | 0.701 | 1.50 | 1.27 | 1.302 | Low Cu + Low Ni |
| 6.49 | 0.94 | 1.42 | 1.03 | 1.02 | 0.701 | 17.71 | 17.42 | 12.310 | Low Cu + High Ni |
| 6.53 | 0.99 | 1.72 | 1.85 | 1.84 | 1.390 | 1.46 | 1.21 | 1.270 | High Cu + Low Ni |
| 6.56 | 1.12 | 1.63 | 1.87 | 1.92 | 1.271 | 17.90 | 17.68 | 12.420 | High Cu + High Ni |

Table SI2. Water chemistry measurements from samples obtained throughout a chronic (21 d) exposure of *D. pulex-pulicaria*to varying single and binary mixtures of Cu and Ni.

a Water samples taken from media entering the exposure system at test initiation and media changes (mean values)

b Water samples taken from media exiting the exposure system at media changes and test end (mean values)

c Calculated using WHAM VII, based on measured pH, measured DOC (in), measured dissolved Cu and Ni (in) and the components in the FLAMES media

d Treatment identifiers for metabolomics analyses

**Metabolomics**

*Methods*

The metabolomics data (acute and chronic) were collected and filtered as detailed in the manuscript. Following the data filtering steps, further data processing included normalising each sample to total ion count, imputing missing values with half of the lowest detectable value for that peak, and autoscaling to reduce leverage on the model due to peaks with high variance. Metabolomics data from the acute and chronic exposure studies were filtered and analysed separately, and data were processed and analysed in MATLAB (version R2013a, The Mathworks, Natick, MA) and PLS\_Toolbox (version 7.8.2, Eigenvector Research Inc., Manson, WA).

An optimal PLS-R model was built and the root mean square error (RMSE) of the model was calculated. This was cross validated using a venetian blinds approach with 9 data splits and an average of 11 % of samples removed from the model to give a RMSE cross validated (RMSECV) value. Both the RMSE and RMSECV values are quoted in the units of the y axis, with an average error rate (equivalent to a standard deviation) equivalent to the number of neonates per sample. Three outlying samples (defined as sitting outside the Hotelling’s t2 statistic and with excessive leverage on the model) were removed for the final optimised model. The statistical significance of the predictive power of this optimal model was calculated by comparing the cross-validated RMSECV value against RMSECV values derived from 1000 random models, i.e. each calculated using randomly permuted class labels. The number of times the RMSECV value of the randomly permuted model was greater than the RMSECV of the optimal model was determined, and divided by the number of permuted models (1000) to generate a *p*-value which we used for inference.

A custom written script (MATLAB) was utilised for improved feature selection, and employed a forward selection strategy to generate a model that included only the mass spectral peaks that were deemed the most important predictors of reproductive output. Briefly, the variable importance of the projection (VIP) values for all of the peaks in the PLS-R model were ranked in order of absolute value, and then peaks were incorporated into the model building process in rank order (i.e. first the most important peak, then the first two important peaks, then three, etc.) and the predictive capability of the model was calculated based on the r2 value generated each time a new peak was added. The optimal model consisted of only those peaks deemed necessary to create a regression model with the highest r2 value, and hence the greatest capability in predicting reproductive output.

*References*

Tipping E, Lofts S, Sonke JE. 2011. Humic ion binding model VII: a revised parameterisation of cation-binding by humic substances. *Environ Chem* 8:225-235

Al-Reasi HA, Smith DS, Wood CM. 2012. Evaluating the ameliorative effect of natural dissolved organic matter (DOM) quality on copper toxicity to *Daphnia magna*: improving the BLM. *Ecotoxicology* 21:524-537

Dieterle F, Ross A, Schlotterbeck G, Senn H. 2006. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal Chem* 78: 4281–4290.