Figure S1: Trehalose detection assay spectrophotometrically quantifying the conversion of NADP+ to NADPH at 340 nm. The concentration of NADPH is stoichiometric to D-glucose, equating to twice the amount of trehalose.
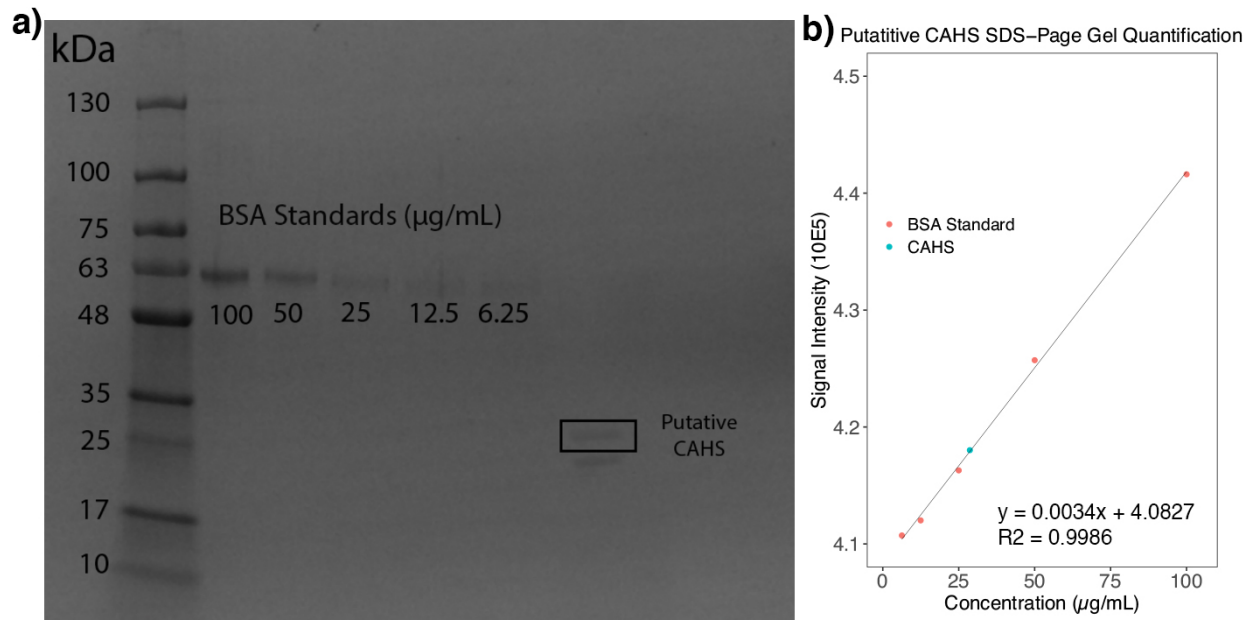
Figure S2: Concentration of putative CAHS protein was quantified by fitting band intensity of CAHS protein to a standard curve made from known BSA standards. **a)** SDS-PAGE gel with BSA standards and presumed CAHS protein band. **b)** Standard curve and CAHS protein concentration plot.
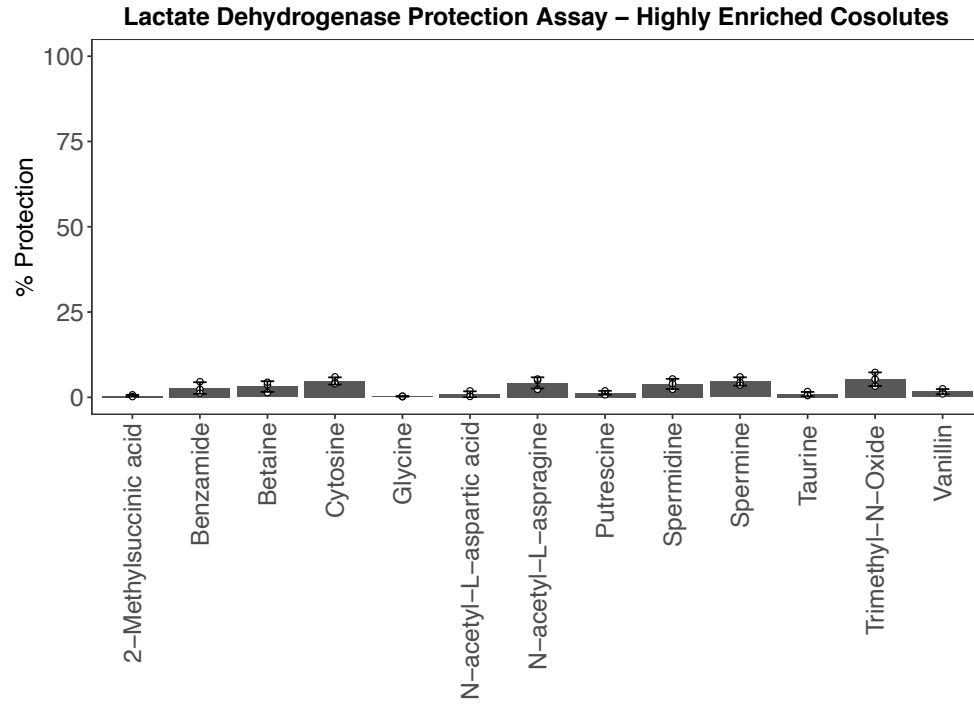
Figure S3: Highly enriched metabolites in desiccated tardigrades do not confer protection alone. Metabolites from the list of highly enriched metabolites were tested for LDH-protection under desiccation stress. Bar plots represent concentrations at 10 mg/ml. Error bars = standard deviation
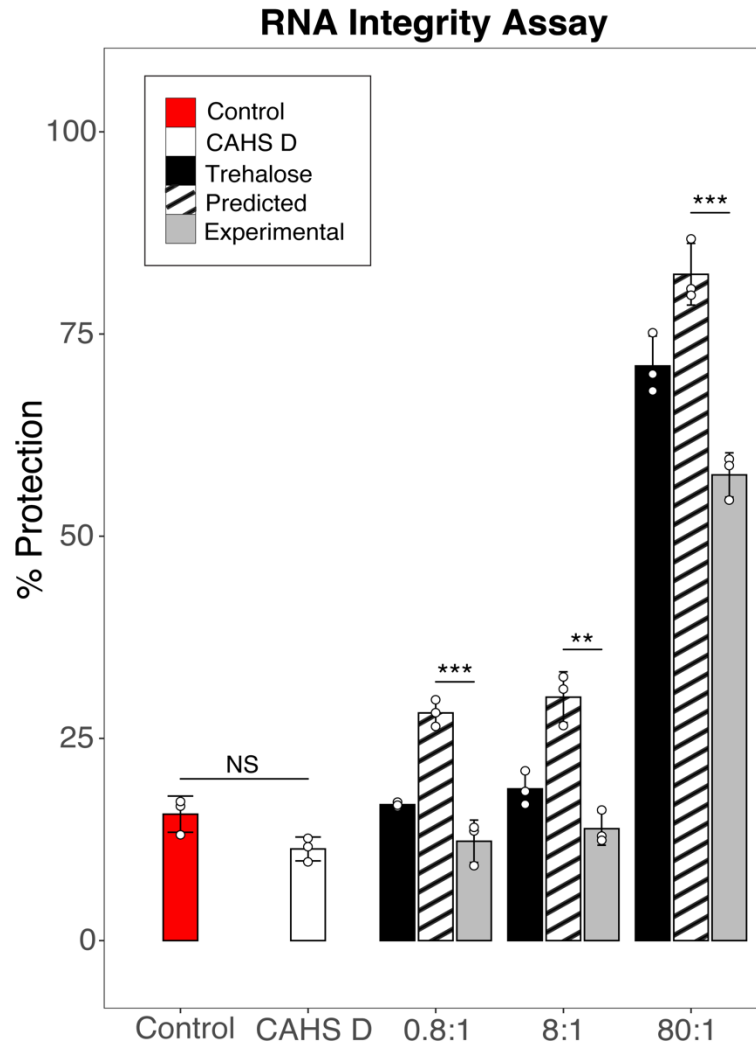
Figure S4. mRNA encoding eGFP was assayed with CAHS D and trehalose at experimental molar ratios to assess translation efficiency upon dedication stress. One way ANOVA was used for statistical comparisons where *** indicates p-value < 0.001. Error bars = standard deviation

Supplemental Materials and Methods

Metabolomics sample preparation: Hydrated and dried tardigrades were sent to Metabolon, Inc (Morrisville, NC) for sample preparation. Samples were prepared using the automated MicroLab STAR® system from Hamilton Company. Several recovery standards were added prior to the first step in the extraction process for QC purposes. To remove protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was divided into five fractions: two for analysis by two separate reverse phase (RP)/UPLC-MS/MS methods with positive ion mode electrospray ionization (ESI), one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, one for analysis by HILIC/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. Samples were placed briefly on a TurboVap® (Zymark) to remove the organic solvent. The sample extracts were stored overnight under nitrogen before preparation for analysis.

QA/QC: Several types of controls were analyzed in concert with the experimental samples: a pooled matrix sample generated by taking a small volume of each experimental sample (or alternatively, use of a pool of well-characterized human plasma) served as a technical replicate throughout the data set; extracted water samples served as process blanks; and a cocktail of QC standards that were carefully chosen not to interfere with the measurement of endogenous compounds were spiked into every analyzed sample, allowed instrument performance monitoring and aided chromatographic alignment. Tables S1 and S2 describe these QC samples and standards. Instrument variability was determined by calculating the median relative standard deviation (RSD) for the standards that were added to each sample prior to injection into the mass spectrometers. Overall process variability was determined by calculating the median RSD for all endogenous metabolites (i.e., non-instrument standards) present in 100% of the pooled matrix samples. Experimental samples were randomized across the platform run with QC samples spaced evenly among the injections, as outlined in Figure S5.

Table S1: Description of Metabolon QC Samples

Type	Description	Purpose
MTRX	Large pool of human plasma maintained by Metabolon that has been characterized extensively.	Assure that all aspects of the Metabolon process are operating within specifications.

CMTRX	Pool created by taking a small aliquot from every customer sample.	Assess the effect of a non-plasma matrix on the Metabolon process and distinguish biological variability from process variability.
PRCS	Aliquot of ultra-pure water	Process Blank used to assess the contribution to compound signals from the process.
SOLV	Aliquot of solvents used in extraction.	Solvent Blank used to segregate contamination sources in the extraction.

Table S2: Metabolon QC Standards

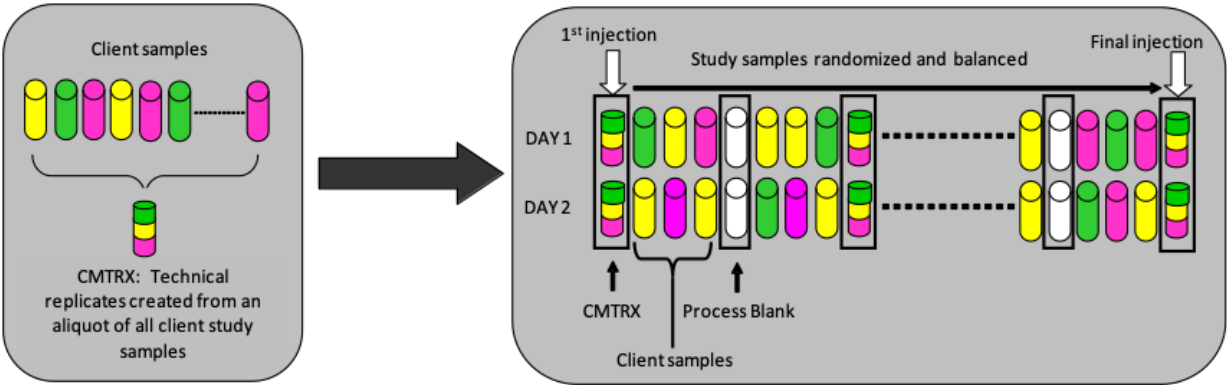| Type	Description	Purpose
RS	Recovery Standard	Assess variability and verify performance of extraction and instrumentation.
IS	Internal Standard	Assess variability and performance of instrument.

Figure S5: Preparation of client-specific technical replicates. A small aliquot of each client sample (colored cylinders) is pooled to create a CMTRX technical replicate sample (multi-colored cylinder), which is then injected periodically throughout the platform run.

Variability among consistently detected biochemicals can be used to calculate an estimate of overall process and platform variability.

Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy (UPLC-MS/MS): All methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. The sample extract was dried then reconstituted in solvents compatible to each of the four methods. Each reconstitution solvent contained a series of standards at fixed concentrations to ensure injection and chromatographic consistency. One aliquot was analyzed using acidic positive ion conditions, chromatographically optimized for more hydrophilic compounds. In this method, the extract was gradient eluted from a C18 column (Waters UPLC BEH C18-2.1x100 mm, 1.7 μm) using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). Another aliquot was also analyzed using acidic positive ion conditions, however it was chromatographically optimized for more hydrophobic compounds. In this method, the extract was gradient eluted from the same aforementioned C18 column using methanol, acetonitrile, water, 0.05% PFPA and 0.01% FA and was operated at an overall higher organic content. Another aliquot was analyzed using basic negative ion optimized conditions using a separate dedicated C18 column. The basic extracts were gradient eluted from the column using methanol and water, however with 6.5mM Ammonium Bicarbonate at pH 8. The fourth aliquot was analyzed via negative ionization following elution from a HILIC column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 μm) using a gradient consisting of water and acetonitrile with 10mM Ammonium Formate, pH 10.8. The MS analysis alternated between MS and data-dependent MS^n scans using dynamic exclusion. The scan range varied slightly between methods but covered 70-1000 m/z. Raw data files are archived and extracted as described below.

Bioinformatics: The informatics system consisted of four major components, the Laboratory Information Management System (LIMS), the data extraction and peak-identification software, data processing tools for QC and compound identification, and a collection of information interpretation and visualization tools for use by data analysts. The hardware and software foundations for these informatics components were the LAN backbone, and a database server running Oracle 10.2.0.1 Enterprise Edition.

LIMS: The purpose of the Metabolon LIMS system was to enable fully auditable laboratory automation through a secure, easy to use, and highly specialized system. The scope of the Metabolon LIMS system encompasses sample accessioning, sample preparation and instrumental analysis and reporting and advanced data analysis. All of the subsequent software systems are grounded in the LIMS data structures. It has been modified to leverage and interface with the in-house information extraction and data visualization systems, as well as third party instrumentation and data analysis software.

Data Extraction and Compound Identification: Raw data was extracted, peak-identified and QC processed using Metabolon's hardware and software. These systems are built on a web-service platform utilizing Microsoft's .NET technologies, which run on high-

performance application servers and fiber-channel storage arrays in clusters to provide active failover and load-balancing. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. Metabolon maintains a library based on authenticated standards that contains the retention time/index (RI), mass to charge ratio (*m/z*), and chromatographic data (including MS/MS spectral data) on all molecules present in the library. Furthermore, biochemical identifications are based on three criteria: retention index within a narrow RI window of the proposed identification, accurate mass match to the library +/- 10 ppm, and the MS/MS forward and reverse scores between the experimental data and authentic standards. The MS/MS scores are based on a comparison of the ions present in the experimental spectrum to the ions present in the library spectrum. While there may be similarities between these molecules based on one of these factors, the use of all three data points can be utilized to distinguish and differentiate biochemicals. More than 3300 commercially available purified standard compounds have been acquired and registered into LIMS for analysis on all platforms for determination of their analytical characteristics. Additional mass spectral entries have been created for structurally unnamed biochemicals, which have been identified by virtue of their recurrent nature (both chromatographic and mass spectral). These compounds have the potential to be identified by future acquisition of a matching purified standard or by classical structural analysis.

Curation: A variety of curation procedures were carried out to ensure that a high quality data set was made available for statistical analysis and data interpretation. The QC and curation processes were designed to ensure accurate and consistent identification of true chemical entities, and to remove those representing system artifacts, mis-assignments, and background noise. Metabolon data analysts use proprietary visualization and interpretation software to confirm the consistency of peak identification among the various samples. Library matches for each compound were checked for each sample and corrected if necessary.

Metabolite Quantification and Data Normalization: Peaks were quantified using area-under-the-curve. Total protein concentration (determined by Bradford assay) was used for normalization to account for differences in metabolite levels due to differences in the amount of material present in each sample.

Statistical Methods and Terminology

Statistical Calculations: Two types of statistical analysis are usually performed: (1) significance tests and (2) classification analysis. Standard statistical analyses are performed in ArrayStudio on log transformed data. For those analyses not standard in ArrayStudio, the programs R (http://cran.r-project.org/) or JMP were used.

Welch's two-sample *t*-test

Welch's two-sample *t*-test is used to test whether two unknown means are different from two independent populations.

This version of the two-sample *t*-test allows for unequal variances (variance is the square of the standard deviation) and has an *approximate t*-distribution with degrees of freedom

estimated using Satterthwaite's approximation. The test statistic is given by
$t = (\bar{x}_1 - \bar{x}_2)/\sqrt{s_1^2/n_1 + s_2^2/n_2}$,and the degrees of freedom is given by

$$\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2 \bigg/ \left(\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}\right)$$

, where $\bar{x}_1, \bar{x}_2$ are the sample means, s_1, s_2, are the sample standard deviations, and n_1, n_2 are the samples sizes from groups 1 and 2. We typically use a two-sided test (tests whether the means are different) as opposed to a one-sided test (tests whether one mean is greater than the other).

Principal Components Analysis (PCA)

Principal components analysis is an unsupervised analysis that reduces the dimension of the data. Each principal component is a linear combination of every metabolite and the principal components are uncorrelated. The number of principal components is equal to the number of observations.

The first principal component is computed by determining the coefficients of the metabolites that maximizes the variance of the linear combination. The second component finds the coefficients that maximize the variance with the condition that the second component is orthogonal to the first. The third component is orthogonal to the first two components and so on. The total variance is defined as the sum of the variances of the predicted values of each component (the variance is the square of the standard deviation), and for each component, the proportion of the total variance is computed. For example, if the standard deviation of the predicted values of the first principal component is 0.4 and the total variance = 1, then 100*0.4*0.4/1 = 16% of the total variance is explained by the first component. Since this is an unsupervised method, the main components may be unrelated to the treatment groups, and the "separation" does not give an estimate of the true predictive ability.

Random Forest Analysis (RFA)

Random forest is a supervised classification technique based on an ensemble of decision trees (Breiman 2001). For a given decision tree, a random subset of the data with identifying true class information is selected to build the tree ("bootstrap sample" or "training set"), and then the remaining data, the "out-of-bag" (OOB) variables, are passed down the tree to obtain a class prediction for each sample. This process is repeated thousands of times to produce the forest. The final classification of each sample is determined by computing the class prediction frequency ("votes") for the OOB variables over the whole forest. For example, suppose the random forest consists of 50,000 trees and that 25,000 trees had a prediction for sample 1. Of these 25,000, suppose 15,000 trees classified the sample as belonging to Group A and the remaining 10,000 classified it as belonging to Group B. Then the votes are 0.6 for Group A and 0.4 for Group B, and hence the final classification is Group A. This method is unbiased since the prediction for each sample is based on trees built from a subset of samples that do not include that

sample. When the full forest is grown, the class predictions are compared to the true classes, generating the "OOB error rate" as a measure of prediction accuracy. Thus, the prediction accuracy is an unbiased estimate of how well one can predict sample class in a new data set. Random forest has several advantages – it makes no parametric assumptions, variable selection is not needed, it does not overfit, it is invariant to transformation, and it is easy to implement with R.

To determine which variables (biochemicals) make the largest contribution to the classification, a "variable importance" measure is computed. We use the "Mean Decrease Accuracy" (MDA) as this metric. The MDA is determined by randomly permuting a variable, running the observed values through the trees, and then reassessing the prediction accuracy. If a variable is not important, then this procedure will have little change in the accuracy of the class prediction (permuting random noise will give random noise). By contrast, if a variable is important to the classification, the prediction accuracy will drop after such a permutation, which we record as the MDA. We will note that in our top 30 metabolites reported by RFA were three xenobiotics: PFOA, tartarate, and benzoate. We have removed these compounds from our presentation of this data, as likely these are man-made chemical contaminates introduced inadvertently somewhere in the metabolomics pipeline.