nature research

Corresponding author(s): Sarah E. O'Connor

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Experimental design

1. Sample size

   Describe how sample size was determined.

   Prior determination of sample size was not a consideration for our data. The majority of experiments we conducted did not involve statistical analyses: to check robustness we repeated experiments and ensured reported results were reproducible (see replication below). Sample size was a consideration for kinetic experiments: sufficient data points were collected to ensure fit parameters were estimated with $p<0.001$ (this was $n = 9$-11). For proteomics a single sample for each tissue type was sufficient: the fisher exact test is designed for $n=2$ comparisons. Furthermore the proteomics was primarily used to guide enzyme discovery rather than as an end in itself, so multiple samples were not required.

2. Data exclusions

   Describe any data exclusions.

   No data were excluded from the analyses.

3. Replication

   Describe the measures taken to verify the reproducibility of the experimental findings.

   The majority of our presented experimental findings are representative of two similar experiments performed independently on different days. We have attempted to verify the reproducibility of many of our key experimental findings by demonstrating similar effects in varying conditions. For example, findings in Figures 2B are shown to be reproduced with different enzymes and buffers in Supplementary Figures 3 and 4. Similarly the NEPS enzyme activities reported (Fig 3) are reproduced with different iridoid synthases (Fig S12) and with a different substrate (Fig 4). All attempts at replication were successful.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   For GC-MS analysis, samples were typically randomized prior to injection to minimize chances of artifacts. No other scenario was identified in which randomization would be required.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   Blinding investigators was not a requirement for our study. The data types collected (e.g. GC-MS chromatograms, X-ray crystallography, kinetics, proteomics etc) are not significantly impacted by investigator prior knowledge so blinding was not necessary.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☒	☐ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Proteomic analysis was performed using: MaxQuant (version 1.5.3.30), Mascot search engine (version 2.4.1, Matrixscience, London) and statistical analysis was performed in Scaffold4 software.
Kinetic analysis was performed using the R software environment. The LM (linear model) function was used to fit initial rates and the NLS (nonlinear least squares) function was used to fit the standard Michaelis-Menten parameters.
Protparam (ExPasy) was used to calculate MW and epsilon for determining protein concentrations.
Phylogenetic analysis was performed using ClustalW2, MUSCLE, IQTree v1.5.4 (with ModelFinder), FigTree v1.4.3 and Geneious v9.1.8.
Crystallography analysis was performed using: CCP4i2, XDS, XIA2, AIMLESS, SCULPTOR, PHASER, BUCCANEER, COOT, REFMAC5, MolProbity, USCF-Chimera v1.11.2, iTasser, Yasara and AutoDockVina v1.1.2.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

There are no restrictions on the availability of data. See the 'Data Availability' section in the main paper for details of deposited and available data.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used in this study.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used in this study.

b. Describe the method of cell line authentication used.

No eukaryotic cell lines were used in this study.

c. Report whether the cell lines were tested for mycoplasma contamination.

No eukaryotic cell lines were used in this study.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No eukaryotic cell lines were used in this study.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.	No animals were used in the study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	No human research participants were used in the study.