

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| ☐ | ☒ | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☒ | ☐ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Qualitative and quantitative analysis data (NMR, GC, HPLC, photometric absorbance) were collected using commercial software provided by the manufacturers of the instruments, mainly Agilent ChemStation B.04.03, Agilent MassHunter B.07.03, Agilent OpenLab 2.8, Shimadzu LabSolutions 5.97, Bruker TopSpin (unknown version), Molecular Devices Softmax Pro 6.4. Custom Python code (available on GitHub: https://github.com/saberger/Screening_for_Generality) was used to extract data from individual result files generated by these software packages and combine the data into CSV files for further processing.
Data analysis	Processing of GC and HPLC chromatograms was carried out using commercial software provided by the manufacturers of the instruments, mainly Agilent ChemStation B.04.03, Agilent MassHunter B.07.03, Agilent OpenLab 2.8, Shimadzu LabSolutions 5.97. Processing of NMR spectra was carried out using MestReNova 15.0.1. Conversion of time-series absorbance readings into slopes was carried out using commercial software provided by the manufacturer of the instrument (Molecular Devices Softmax Pro 6.4) or using custom Python code (available on GitHub: https://github.com/saberger/Screening_for_Generality). Simple arithmetic transformations were carried using Microsoft Excel (various versions). Data visualisation was performed using Microsoft Excel (various versions), R 4.40 (using the ggplot library), and custom Python code available on GitHub: https://github.com/saberger/Screening_for_Generality .

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The data generated in this study (consolidated screening data, NMR spectra, protein and DNA sequences, chemoinformatic descriptors of the substrates) are openly available in the Mendeley Data repository (<https://data.mendeley.com/>) at <https://doi.org/10.17632/jn4sfsrz8x.2>. Source data for Figures 2, 3, S7, S9, S14, S15, S16, S17, S18, S22, S24 and S25 are provided with the paper. All other figures can be directly reproduced from the deposited data set.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.

Reporting on race, ethnicity, or other socially relevant groupings

Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status). Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.) Please provide details about how you controlled for confounding variables in your analyses.

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Sample size for the screenings and for biotransformations were chosen as follows: Overview Screening (photometric): triplicates for all reactions including negative controls, allowing identification of hits based on mean and standard deviation. Diversification Screening, Amine Scope (photometric): single determinations for IRED reactions, n=9 for negative controls, making experimental replicates an additive rather than multiplicative effort, while still allowing identification of hits based on mean and standard deviation of negatives. Diversification Screening, Carbonyl Scope (FIA-MS): single determinations for IRED reactions, n=9 for negative controls, making experimental replicates an additive rather than multiplicative effort, while still allowing identification of hits based on mean and standard deviation of negatives. Analytical-scale biotransformations: single determinations, as we prioritised collecting data on a larger set of enzyme-substrate combinations over determining random variation for a smaller number of combinations. Preparative-scale biotransformations: single determinations, as material input is considerable and the outcome of a single experiment already serves to demonstrate practical applicability.

Data exclusions

Photometric reads that were not interpretable (e.g., due to formation of air bubbles) were excluded from analysis. If replicate reads for the same experimental condition were available, these were used for further analysis. If no replicates were available, no data are reported for the

respective experimental condition. The same approach was followed for non-interpretable FIA–MS runs (e.g., insufficient signal in all recorded mass channels).

Replication

Although many of the reported data points arise from single determinations, most of these data have been qualitatively replicated (even by different operators), as the original biotransformation data set was obtained using non-calibrated GC–MS analysis (data still available at: <https://chemrxiv.org/doi/full/10.26434/chemrxiv-2024-0sq49?redirectToLatest=false>, <https://doi.org/10.17632/jn4sfsrz8x.1>) and was later "upgraded" to calibrated GC–FID. All GC–MS positives could be replicated by re-running the biotransformations and analysing them by GC–FID. High-throughput screening data have been consolidated by replicating selected samples (see Supporting Information, Tables S14 and S16–S19) with GC–MS or HPLC–MS analysis. Where high-throughput data could not be replicated by chromatographic re-analysis, they were manually flagged as negative and are reported as negatives in the final data set.

Randomization

Randomisation is only relevant in the current study in the context of selecting substrate–enzyme combinations for chromatographic retesting to estimate false positive and negative rates (Supporting Information, Tables S14, S15). Combinations were picked using a random-number generator in Microsoft Excel.

Blinding

Blinding is not relevant in the chemical sciences, as operator expectations are unlikely to influence reaction outcome.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.