

Data-Driven Construction of an Imine Reductase Library Capable of Broad-Scope Reductive Amination at Equimolar Substrate Concentrations

Corresponding Author: Dr Joerg Schrittwieser

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Schrittwieser et al are reporting on constructing an imine reductase panel which has broad substrate scope and very importantly catalyses reductive amination at equimolar ratio of carbonyl and amine, something which was originally thought is not feasible so broadly within the enzyme class. The advantage of equimolar ratios is particularly important if one of the the substrate is expensive or hard to source (e.g. in pharmaceutical intermediates).

The quality of the experimental and computational work is without a doubt very high.

I have few comments which should be considered:

1. p1, l6- could there be updated estimation from e.g. 2025, than from 2023? or at least commented that the trend remained the same.
2. p3 Screening Strategy - it took me a while to understand the paragraph and the process of the setting up different stages of screening. I suggest to make the high level description clearer and use e.g. table with steps - 1. overview screening (currently middle of the scheme, could that be on top of the scheme? logical to be on top if that is first step
3. p4 l211-225, nice approach to validate false positive and negative data. But would it not be faster to directly used LC/MS of all samples, than this 2-step validation with regrowing plates in between which takes lot of time too.
4. p4 - I am bit disappointed by the choice of substrates as they are very simple compounds previously published by others (e.g. Turner's group). only in the 2nd and 3rd round were used bit more complex amines and ketones. it could have been more time efficient and less experimentally demanding to have less of simple substrates and compensate with some of more difficult substrates in 1 total screen. could you comment if this would still allow you to collect the right size and quality of dataset which was required for AI activity prediction and deep-learning model?
- 5.p5 l278- can you comment in the manuscript and postulate hypothesis why most general IREDs were found in the starting panel while the enzymes recruited based on ML predictions tended towards high maximum activity. this finding deserves explanation.
6. p7 Fig 4 - see comment 4 above- very simple ketones chosen for testing
7. Fig 4 -can you comment on ketones c27-c33 as the carbonyl is in alpha position to aromatic ring- how easy is to convert such ketones in general. also any trends to learn from other enzymatic classes like transaminases?
8. p9 up to l418- can you comment in text on any 3D enzyme structure (catalophore) which explains the obtained data?
9. p9. l460- were there detected any enzymes with high conversion but low e.e.?
10. p10 Fig 5b - those products were obtained generally with low conversion, why is that?
- 11 p10, l468 - how about prediction of negative results - could the model predict which substrate-enzyme pair is bad and won't give good conversion with high confidence?
12. p11, l499-501 - thanks so much for providing this data set to the community! an exemplary gesture of sharing data and not hiding them for future use of others

overall, this is really great study, with high quality dataset. However, this study is maybe 2-3 years coming too late. other similar studies were published meanwhile. I am not questioning the content of the manuscript itself, but if the novelty is high enough for this journal. Maybe other journal would be more appropriate.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

In this study, Schrittwieser and collaborators presented a data-driven strategy to construct a broad-scope IRED library by combining large-scale experimental screening with machine-learning analysis. By generating a dataset of more than 20,000 enzyme–substrate reactions under standardized conditions, the authors establish correlations between IREDs and their catalytic performance toward different amine and carbonyl substrates and demonstrate that predictive models can help identify promising candidate enzymes.

The most noteworthy result of this work is the creation of a large, internally consistent activity dataset under unified assay conditions, together with a practical predictive framework for enzyme prioritization. This is important because the broader use of IREDs has long been constrained by two major bottlenecks: enzyme discovery/screening and enzyme engineering. The present study directly addresses the first of these challenges. In previous studies, progress in IRED research has often depended heavily on the scale of enzyme libraries available within individual laboratories, which makes the field relatively resource-intensive. In contrast, the framework proposed here has the potential to reduce the experimental burden substantially by allowing researchers to identify likely hits computationally and then test only a relatively small number of enzymes. From this perspective, the work is significant not only to the IRED field, but also to related areas of biocatalyst discovery and data-driven enzyme screening. Although the machine-learning models and feature representations used here are not conceptually new, their integration with a large and systematically generated experimental dataset constitutes a meaningful advance. Overall, the study is well executed, methodologically sound, and of clear practical value. The data generally support the authors' main conclusions, and the methods appear sufficiently detailed to enable reproduction. I therefore support publication in Nature Communications after the following points are addressed.

Major comments

1. While the predictive framework is convincing, the manuscript provides limited mechanistic insights into how enzyme structure influences catalytic activity and substrate scope. The authors are encouraged to include structural analyses of representative enzymes (e.g., comparison of active sites, molecular docking, or MD simulations) to identify key residues that may contribute to activity or promiscuity.
2. Because the dataset was generated under standardized conditions, it may serve as a valuable reference dataset for the field. However, it is unclear whether previously reported literature data—especially for substrate combinations not studied here—could be incorporated into the database. The authors may consider discussing whether literature data could be integrated through calibration or benchmarking strategies to expand the dataset in the future.
3. Enzyme discovery and enzyme engineering are the two main themes in IRED research. It would be valuable if the authors could briefly discuss whether the current framework could be extended to include engineered variants. In particular, more detailed substrate descriptors or interaction-based features might improve predictive performance for engineered enzymes.

Minor comments

1. The gradient-boosted random forest model is described as a “deep learning (DL) model”, which is not strictly accurate. The terminology should be corrected.
2. The authors may consider citing recent reviews and articles on Machine Learning-assisted protein engineering for improving stereoselectivity and/or activity.

(Remarks on code availability)

I did not independently install or execute the code, so I cannot fully verify computational reproducibility through direct re-running. However, I carefully reviewed the codebase and, based on its organization and implementation quality, the code appears to be generally usable and appropriate as a community resource. The structure is coherent, the workflow is understandable, and the implementation seems consistent with the computational analyses described in the manuscript. From my inspection, the code appears likely to support reproduction of the main modeling procedures, provided that the required software environment and input data are made available.

Reviewer #3

(Remarks to the Author)

Schrittwieser and coworkers present a comprehensive and systematic substrate-profiling study of 175 IREDs/RedAms across multiple reductive amination substrate combinations. This represents the most extensive structure–activity relationship investigation of imine reductases to date. The work is thorough, well executed, and clearly written.

In this reviewer's opinion, many of the general activity trends observed are consistent with what has been previously reported in the IRED/RedAm literature. Despite the impressive scale of the screening, no enzymes in the panel demonstrate high activity toward the most challenging substrates (e.g., aromatic ketones such as 1-tetralone, 4-chromanone), which aligns with earlier findings.

Nevertheless, the study provides several valuable contributions to the biocatalysis community:

Identification of sequence features associated with top-performing enzymes, offering guidance for future IRED/RedAm discovery and mining efforts.

A unified dataset combining extensive substrate and sequence information, enabling rapid decision-making when selecting IREDs/RedAms for synthetic applications.

Screening at high ketone concentrations (50 mM), which is highly relevant for assessing industrial applicability.

A framework for iterative machine-learning-guided enzyme discovery, which will be useful for future computational–experimental workflows for other enzyme classes.

These contributions strengthen the impact of the work, and I would recommend the manuscript for publication after the authors address the following points:

1. Lysate-based NADPH depletion assay and false positives

The use of a lysate-based NADPH depletion assay is known to introduce background activity and potential false positives. For highly reactive substrate combinations, the IRED-dependent activity is often sufficiently strong to distinguish from background. However, for weakly reactive substrates, the difference between background and true enzymatic activity can be minimal.

The authors should clarify what measures were taken to ensure that background activity does not obscure low-level IRED activity. Additionally, a control containing amine + IRED + NADPH but no ketone should be included as one of the controls.

2. Expression level variability

As a limitation of the lysate-based approach, the authors should acknowledge that differences in protein expression levels can influence apparent activity. Without normalization of enzyme expression, activity comparisons across the panel may be affected by variation in expression levels.

3. Non-canonical IREDs

The authors should reference recent work on non-canonical imine reductases, and provide rational for excluding these enzymes in their screening.

(Remarks on code availability)

Not reviewed

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

thank you for responding and addressing all comments and shortcomings of this manuscripts.
I have no further points and I look forward to seeing this manuscript published.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

The author has provided a thorough explanation and made revisions in response to my comments, and these new evidence answers my questions, so I recommends its publication in Nature Communications.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

The revised manuscript by Schrittwieser and colleagues addresses all of my previously raised concerns and represents a significant improvement. An impressive amount of data has also been included in the Supporting Information. The analysis of stereoselectivity is particularly interesting (Table S11).

Minor suggestions:

The authors should consider including reaction schemes for Table 1 and Figure 5.

They should consider using a colour scheme to distinguish fragments of the product derived from the carbonyl and amine coupling partners.

(Remarks on code availability)

Not reviewed

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Response to the Reviewers' Comments

Reviewer 1

Schrittwieser et al are reporting on constructing an imine reductase panel which has broad substrate scope and very importantly catalyses reductive amination at equimolar ratio of carbonyl and amine, something which was originally thought is not feasible so broadly within the enzyme class. The advantage of equimolar ratios is particularly important if one of the the substrate is expensive or hard to source (e.g. in pharmaceutical intermediates).

The quality of the experimental and computational work is without a doubt very high. I have few comments which should be considered:

1 // p1, l6- could there be updated estimation from e.g. 2025, than from 2023? or at least commented that the trend remained the same.

— *At the time of submitting this revision, the data from 2025 are not yet available on the Njardarson lab's web page. However, we expect them to be posted some time in June and will update the reported percentages when submitting the final version of our manuscript.*

2 // p3 Screening Strategy - it took me a while to understand the paragraph and the process of the setting up different stages of screening. I suggest to make the high level description clearer and use e.g. table with steps - 1. overview screening (currently middle of the scheme, could that be on top of the scheme? logical to be on top if that is first step

— *Thank you for this suggestion. Figure 1 is designed to summarise both the computational and the experimental contributions to the overall study. The first step in our overall workflow was the parametrisation of substrate and enzyme properties (box labelled "In-Silico Analysis of Substrates and Enzymes" in Figure 1), so we prefer to keep the current layout of the figure. However, we have added a table to the SI (Table S27) that provides further details on the screening phases in a step-wise format, as suggested by the reviewer. This table also contains information on time lines, coverage and hit rates, and is accompanied by text that explains the computational and experimental iterations more clearly. A cross-reference to this new SI section has been added to the main text.*

3 // p4 l211-225, nice approach to validate false positive and negative data. But would it not be faster to directly used LC/MS of all samples, than this 2-step validation with regrowing plates in between which takes lot of time too.

— *The direct use of "regular" LC-MS (with chromatographic separation) for screening the same number of substrate-enzyme combinations would take more time. Please note that the photometric screening was exhaustive and performed in triplicates, while the chromatographic re-analysis using GC-MS and HPLC-MS was not, and that all assays (photometric and chromatographic) were set up using the same lyophilised stocks of cell-free extract, with no re-growing of cells required. Using this approach, assay plates could be easily set up from "master plates" containing the required substrate stock solutions and freshly prepared IRED stocks. Two plates, translating to 2 h of instrument time and approx. 4 h of*

hands-on time, were required for collecting triplicate photometric data for one IRED on all 56 substrate combinations (192 assay reactions, including 24 neg. controls). Analysing the same number of samples by HPLC-MS would require a similar hands-on time and approx. 28 h of instrument time, assuming 10-min runs, triplicate determinations, and no negative controls. Moreover, we considered the kinetic data obtained by the photometric assay a more informative basis for structure-activity correlations than the endpoint data obtained by chromatographic analysis.

All that said, we found that FIA-MS (ESI-MS without chromatographic separation) is a screening method that can compete with a photometric assay in terms of generality and throughput. However, we established this method in response to the turbidity problems observed in the Diversification Screening (carbonyl scope plate), so it was not yet available at the time of the Overview Screening, and this method, too, provides endpoint data.

4 // p4 - I am bit disappointed by the choice of substrates as they are very simple compounds previously published by others (e.g. Turner's group). only in the 2nd and 3rd round were used bit more complex amines and ketones. it could have been more time efficient and less experimentally demanding to have less of simple substrates and compensate with some of more difficult substrates in 1 total screen. could you comment if this would still allow you to collect the right size and quality of dataset which was required for AI activity prediction and deep-learning model?

— Our main goals for the Overview Screening were to (1) quickly identify poorly performing enzymes that should be replaced with new candidates, and (2) evaluate the reductive amination performance of the IRED panel against a set of amines and carbonyls that represent all major substrate classes well: simple primary amines (**a1-a3**), activated primary amines (**a4**), secondary amines (**a5**), aryl-aliphatic amines (**a6**), aromatic amines (**a7**); aldehydes (**c1**), simple alicyclic and open-chain aliphatic ketones (**c2, c6**), aliphatic ketones with additional functionality (**c4, c7**), aromatic ketones (**c5, c8**). Relatively simple representatives with limited additional functionality were selected to avoid compromising the screening results by undesired cross-reactivity. This risk is illustrated by ketone **c5**, whose fast hydrolysis allowed reductive amination only with the most favourable amine substrate, **a4**, and even that only in a few cases. Had such a problem occurred with a larger number of substrates, the screening would have provided much less information on the general reductive amination scope of the enzyme panel, making modelling and activity prediction difficult. Moreover, we deemed it important to include at least a small number of “positive controls” (very favourable substrate combinations, such as **c2-a2** and **c2-a4**) to distinguish between IREDs with limited scope and those that are entirely inactive. Practical considerations, such as solubility and commercial availability, also played a role in substrate selection.

5 // p5 l278- can you comment in the manuscript and postulate hypothesis why most general IREDs were found in the starting panel while the enzymes recruited based on ML predictions tended towards high maximum activity. this finding deserves explanation.

— We believe that two main factors contribute to this trend: (1) The inherent characteristics of the *k*-nearest neighbour model, and (2) the process of clustering and selecting sequences for experimental testing based on their predicted activities.

As discussed in the SI (section “Activity Prediction Using Machine-Learning Models”, subsection “Prioritisation of IRED candidates for the Overview Screening”), the *k*-nearest

neighbour model produces a complete set of predicted activity values for all 54 interpretable substrate combinations of the Overview Screening (see Figure S12 for an illustrative example). This model, being based on weighted averages, does not predict activity values outside the experimentally observed maximum and minimum. Consequently, it tends towards conservative predictions for high values and optimistic predictions for low values. In other words, it is biased to underestimate maximum activity and overestimate substrate scope. We have rephrased the last paragraph of the SI section “Prioritisation of IRED candidates for the Overview Screening” to discuss this characteristic more directly.

The generated predictions were used to rank candidate enzymes by expected substrate scope (predicted active combinations) and maximum predicted activity, respectively, in two separate lists. From these two lists, top-ranked candidates were selected for experimental testing in such a way as to avoid overrepresentation of highly homologous clusters in the sequence space. Both the substrate scope rank and the maximum activity rank were taken into account in this selection, but since our Overview Screening data indicate an inherent trade-off between scope and maximum activity, any attempt to balance these two parameters will necessarily introduce a bias for one or the other unless the chosen weighting factor happens to be just right. Apparently, in our workflow, which weighted both parameters equally, the bias is towards maximum activity.

We have added the statement “We attribute the observed trend towards high maximum activity and moderate generality to characteristics of the k-nearest-neighbour model and the clustering and selection process (for detailed discussion, see Activity Prediction Using Machine-Learning Models, Supplementary Information)” to the paragraph of the Results section where the recruiting of additional enzymes is discussed.

6 // p7 Fig 4 - see comment 4 above- very simple ketones chosen for testing

— The substrate panels for the Diversification Screening were designed to cover all major substrate classes of the “secondary” substrate with a minimum number of examples. Thus, for the amine scope part of the screening (Figure 4a), we chose a simple alicyclic and open-chain aliphatic ketone (**c2** and **c9**), an aliphatic ketone with additional functionality (**c10**), an aryl-aliphatic ketone (**c8**), a simple aldehyde (**c11**), and an aromatic ketone (**c12**) as reaction partners for the broad range of amines tested. More complex examples of the same carbonyl substrate classes were covered in the carbonyl scope part (Figure 4b), using simple representative amines as coupling partners. The choice of relatively simple compounds as “secondary” substrates, including known beneficial cases (**c2** and **c11** for the amine scope), was made to avoid the loss of information on “primary” substrate acceptance due to exceedingly challenging reaction partners. Practical considerations also played a role in substrate selection: Keto-ester **c10** was expected to give a mixture of open-chain (amino-ester) and cyclic (lactam) products, which would have complicated analysis by FIA–MS, and hence it was included as carbonyl substrate in the amine scope, analysed photometrically. Amphetamine derivatives are important motifs in medicinal chemistry, and hence phenylacetone (**c8**) was included in the amine scope part of the screening to cover a broad range of its possible reductive amination products.

7 // Fig 4 -can you comment on ketones c27-c33 as the carbonyl is in alpha position to aromatic ring- how easy is to convert such ketones in general. also any trends to learn from other enzymatic classes like transaminases?

— Ketones in which the carbonyl group is conjugated to an aromatic ring are known to be difficult substrates for IREDs in general (for an overview, see: G. A. Aleku, ACS Catal. 2024, 14, 19, 14308–14329; <https://doi.org/10.1021/acscatal.4c04756>). This is in contrast to enzyme classes such as transaminases, alcohol dehydrogenases, and amine dehydrogenases, which often show high activities towards these compounds. Since steric demand alone cannot explain this phenomenon (substrates **c16**, **c18**, and **c19**, for instance, are accepted well), we believe that conjugation of the carbonyl group to the aromatic ring lowers the electrophilicity of the ketone sufficiently to prevent efficient imine formation, irrespective of whether this occurs spontaneously in solution or inside the IRED's active site. Indeed, pre-forming imines derived from such ketones chemically and supplying them to the IRED for reduction reveals them to be viable substrates. We have filed a patent on this imine pre-formation approach (WO2025017089A1, <https://register.epo.org/application?number=EP24740944&tab=main>) and are currently preparing a journal publication in which we discuss the strategy more broadly.

We have added a brief discussion of this aspect of IRED substrate scope to the SI in a new section titled “Protein Structure–Activity and Structure–Selectivity Analysis”, which we also reference in the main article text.

8 // p9 up to l418- can you comment in text on any 3D enzyme structure (catalophore) which explains the obtained data?

— Thank you for this suggestion. While their multi-dimensional nature (3D coordinates enriched with several physicochemical descriptors) makes Catalophore point clouds an excellent, data-rich basis for machine learning, that same characteristic makes them too complex to be readily human-interpretable. However, to better rationalise the obtained screening data, we have carried out other sequence- and structure-based analyses, which are discussed in a new section of the Supplementary Information (Protein Structure–Activity and Structure–Selectivity Analysis). At the indicated position in the main text, we now reference this SI section to explain the poor acceptance of aromatic ketones and amines (see response to previous comment). In the following paragraph, we reference analyses that pinpoint sequence features linked to broad substrate acceptance and to (R)- or (S)-enantiopreference.

9 // p9. l460- were there detected any enzymes with high conversion but low e.e.?

— Yes, such cases were also observed, as shown in Tables S23–S26. For instance, IR2128 gave 22% conversion but only 40% ee for **c16-a6**, IR0063 gave 30% conversion but only 8% ee for **c23-a10**, and IR0202 gave 20% conversion but only 14% ee for **c24-a6**. Hence the wording “comprised positives with substantial conversion and ee in all cases” to indicate that such useful hits were found in all four examples, but not all positives combined substantial conversion and high selectivity.

10 // p10 Fig 5b - those products were obtained generally with low conversion, why is that?

— We attribute the moderate to low conversions achieved for the substrate combinations shown in Figure 5b to a combination of two factors: (1) To create valid test cases for the DL model, we selected comparably challenging substrates, as indicated by their predominantly negative median z-scores (**c16**: –0.567, **c23**: –0.356, **c24**: 0.205, **a6**: –0.130, **a10**: 0.675,

a15: -0.179). Achieving (near-)quantitative conversions with any pairs of these substrates is inherently difficult, even if the entire IRED panel had been screened in each case. (2) The DL model achieves practically useful levels of predictivity (enrichment of positives), but like any model, it is necessarily imperfect. Some of the top-15 ranked IREDs showed no or only trace conversion, while some enzymes with high performance might have been ranked outside the top 15 predictions.

Both factors taken together, it would be unrealistic to expect high levels of conversion from experimental testing of only 15 top-predicted IREDs for the chosen substrate combinations. Indeed, we consider the identification of enzymes that achieve substantial conversion and high selectivity in three of the four test cases, along with an overall hit rate of 70% across all four examples (now reported explicitly in the new Table S27), a positive outcome.

11 // p10, l468 - how about prediction of negative results - could the model predict which substrate-enzyme pair is bad and won't give good conversion with high confidence?

— Yes, both the leave-one-out cross-validation data (SI, p. S18) and the experimental data (Tables S23–S26) indicate that low confidence scores correlate with low or non-detectable activity. Activities <10 mU/mg are reduced 10-fold and activities <50 U/mg 6-fold compared to random selection.

12 // p11, l499-501 - thanks so much for providing this data set to the community! an exemplary gesture of sharing data and not hiding them for future use of others

— Thank you for your appreciative comment. Data sets of this size are hard to obtain and can be analysed in many different ways to extract actionable insights. By sharing the complete data set for re-use, we hope to maximise its value for the biocatalysis community.

overall, this is really great study, with high quality dataset. However, this study is maybe 2-3 years coming too late. other similar studies were published meanwhile. I am not questioning the content of the manuscript itself, but if the novelty is high enough for this journal. Maybe other journal would be more appropriate.

— We are aware that our study is not the first large-scale IRED/RedAm screening campaign reported in the literature. However, we believe that our study is the most systematic IRED screening performed to date, covering significant ground in all three dimensions of reaction space (amine substrate, carbonyl substrate, and enzyme). This makes it possible for the first time to answer general questions concerning the underlying chemistry, such as: Can the substrate reactivity trends reported in the previous literature be generalised for the entire IRED/RedAm sequence space? (In large parts, yes. Aromatic ketones are generally poor substrates, as are ammonia, secondary amines, and aniline derivatives.) Are RedAms consistently better at reductive amination than “regular” IREDs? (Yes, but the difference is gradual rather than categorical.) Is a high excess of amine required for efficient reductive amination? (No, in many cases it is not. Supplying both reaction partners in equimolar, substantial concentrations works well for many substrate–enzyme combinations.) Besides, as pointed out by the other reviewers, the large data set we provide represents a practically useful resource for enzyme selection in synthetic campaigns and for training models that predict reductive amination performance. The new analyses we have added during revision identify sequence features linked to high activity and broad scope beyond the three-residue

RedAm motif (Asn93, Asp169, Tyr177), and the stereoselectivity key residues deduced from docking and conservation analysis make it possible for the first time to predict the enantioselectivity of IREDs in reductive amination with high confidence. We are convinced that these aspects make our work relevant and interesting for the broad readership of Nature Communications.

Reviewer 2

In this study, Schrittwieser and collaborators presented a data-driven strategy to construct a broad-scope IRED library by combining large-scale experimental screening with machine-learning analysis. By generating a dataset of more than 20,000 enzyme–substrate reactions under standardized conditions, the authors establish correlations between IREDs and their catalytic performance toward different amine and carbonyl substrates and demonstrate that predictive models can help identify promising candidate enzymes.

The most noteworthy result of this work is the creation of a large, internally consistent activity dataset under unified assay conditions, together with a practical predictive framework for enzyme prioritization. This is important because the broader use of IREDs has long been constrained by two major bottlenecks: enzyme discovery/screening and enzyme engineering. The present study directly addresses the first of these challenges. In previous studies, progress in IRED research has often depended heavily on the scale of enzyme libraries available within individual laboratories, which makes the field relatively resource-intensive. In contrast, the framework proposed here has the potential to reduce the experimental burden substantially by allowing researchers to identify likely hits computationally and then test only a relatively small number of enzymes. From this perspective, the work is significant not only to the IRED field, but also to related areas of biocatalyst discovery and data-driven enzyme screening. Although the machine-learning models and feature representations used here are not conceptually new, their integration with a large and systematically generated experimental dataset constitutes a meaningful advance. Overall, the study is well executed, methodologically sound, and of clear practical value. The data generally support the authors' main conclusions, and the methods appear sufficiently detailed to enable reproduction. I therefore support publication in Nature Communications after the following points are addressed.

Major Comments

1 // While the predictive framework is convincing, the manuscript provides limited mechanistic insights into how enzyme structure influences catalytic activity and substrate scope. The authors are encouraged to include structural analyses of representative enzymes (e.g., comparison of active sites, molecular docking, or MD simulations) to identify key residues that may contribute to activity or promiscuity.

— *Thank you for this suggestion. We have carried out several new analyses of previously collected data and one additional experiment, with results that we find genuinely exciting. Sequence conservation analysis has identified twelve positions that are conservatively enriched in the top-20 performing IREDs relative to the entire enzyme family. With one exception, these residues are located outside the active site and thus represent sequence*

features linked to high reductive amination performance that go beyond the three-residue RedAm motif (Asn93, Asp169, Tyr177) already discussed in the original version of our manuscript. Protein–ligand docking was used to investigate the origins of the poor acceptance of aromatic substrates (mainly in response to a comment by Reviewer 1) and as a first step in identifying active-site residues that influence stereoselectivity. For this latter aspect, the docking was followed up by sequence analysis, and the two methods combined with literature data pinpoint three active-site positions that seem to govern (R)- vs. (S)-enantiopreference. Based on these three residues, we could rationalise our previous data and also make robust predictions about IREDs whose stereoselectivity we had not previously characterised.

These analyses are presented and discussed in a substantial new subsection of the Supplementary Information titled “Protein Structure–Activity and Structure–Selectivity Analysis”, referenced at several points in the Results section of the main manuscript, and briefly summarised in the Discussion section.

2 // Because the dataset was generated under standardized conditions, it may serve as a valuable reference dataset for the field. However, it is unclear whether previously reported literature data—especially for substrate combinations not studied here—could be incorporated into the database. The authors may consider discussing whether literature data could be integrated through calibration or benchmarking strategies to expand the dataset in the future.

— Yes, we are convinced that literature data can be integrated into the data set in ways that improve predictive performance. Literature-reported specific activity values, for instance, can be integrated directly with our activity data. Conversion values from literature, on the other hand, can be translated into z-scores based on the conv./z-score relationship we provide, while z-scores calculated independently from other systematic screening efforts will also represent a valuable addition to the current data set.

The main limitation in all these efforts, as hinted at by the reviewer, will be differences in the chosen reaction conditions, which translate into differences in the observed activities or conversions. However, we believe that this issue can be addressed by introducing correction factors that could be determined using a small, representative sub-set of reactions performed under the two conditions that need to be compared.

We have added the statement “We also note that the z-score transformation introduced in our study provides a framework for integrating performance data from other sources (e.g., previous literature or future screening campaigns) into the present data set.” to the second paragraph of the Discussion section to emphasise this aspect.

3 // Enzyme discovery and enzyme engineering are the two main themes in IRED research. It would be valuable if the authors could briefly discuss whether the current framework could be extended to include engineered variants. In particular, more detailed substrate descriptors or interaction-based features might improve predictive performance for engineered enzymes.

— The IRED features in the model are based on Catalophore point clouds, that is, on 3-dimensional, physicochemical descriptors of the space in an enzyme’s active-site pocket. Hence, they are agnostic to whether an enzyme is a wild type or an engineered variant (indeed, they are agnostic to protein sequence as a whole), and directly reflect any changes to the amino acid composition of the active site. The model, therefore, can easily deal with

active-site variants. Distant mutations, on the other hand, will only be reflected in the point cloud features if they lead to structural changes in the active site that can be accurately predicted by homology modelling. Overall, this means that our model will be useful for some, but not all, engineered IRED variants.

Minor Comments

1 // The gradient-boosted random forest model is described as a “deep learning (DL) model”, which is not strictly accurate. The terminology should be corrected.

— *Indeed, the gradient-boosted random forest (XGBoost) does not qualify as a DL model, but it represents only the last step in the activity prediction workflow. Deep learning was used for feature extraction from the IRED point clouds to generate the embeddings that the prediction is based on. The whole process can hence be described as a deep-learning pipeline.*

We have rephrased the sentence of the Results section that describes the workflow to make this aspect clearer: “We chose a gradient-boosted random forest for this task, using physico-chemical descriptors of the substrates, deep learning-derived representations of the IREDs’ active-site point clouds, and z-transformed screening data as inputs (for details, see Activity Prediction Using Machine-Learning Models, Supplementary Information).” In Figure 1 and the caption of Figure 5, we have replaced the wording “Deep-Learning Model” with “Deep-Learning Pipeline”. In the abstract, which was shortened to meet the journal’s formatting guidelines, we now use the more general term “machine learning”.

2 // The authors may consider citing recent reviews and articles on Machine Learning-assisted protein engineering for improving stereoselectivity and/or activity.

— *Thank you for this suggestion. We have cited two recent review articles on ML-assisted protein engineering (Kouba et al., ACS Catal. 2023, 13, 13863–13895; Yang et al., ACS Cent. Sci. 2024, 10, 226–241) in the Discussion section where the aspect of engineering for specific target reactions is mentioned.*

Reviewer 3

Schrittwieser and coworkers present a comprehensive and systematic substrate-profiling study of 175 IREDs/RedAms across multiple reductive amination substrate combinations. This represents the most extensive structure–activity relationship investigation of imine reductases to date. The work is thorough, well executed, and clearly written.

In this reviewer’s opinion, many of the general activity trends observed are consistent with what has been previously reported in the IRED/RedAm literature. Despite the impressive scale of the screening, no enzymes in the panel demonstrate high activity toward the most challenging substrates (e.g., aromatic ketones such as 1-tetralone, 4-chromanone), which aligns with earlier findings.

Nevertheless, the study provides several valuable contributions to the biocatalysis community:

Identification of sequence features associated with top-performing enzymes, offering guidance for future IRED/RedAm discovery and mining efforts.

A unified dataset combining extensive substrate and sequence information, enabling rapid decision-making when selecting IREDs/RedAms for synthetic applications.

Screening at high ketone concentrations (50 mM), which is highly relevant for assessing industrial applicability.

A framework for iterative machine-learning-guided enzyme discovery, which will be useful for future computational–experimental workflows for other enzyme classes.

These contributions strengthen the impact of the work, and I would recommend the manuscript for publication after the authors address the following points:

1 // Lysate-based NADPH depletion assay and false positives

The use of a lysate-based NADPH depletion assay is known to introduce background activity and potential false positives. For highly reactive substrate combinations, the IRED-dependent activity is often sufficiently strong to distinguish from background. However, for weakly reactive substrates, the difference between background and true enzymatic activity can be minimal.

— *Assay validity was our main concern during early development of the photometric assay. In particular, we wanted to ensure a minimal risk for false positives, so as not to report any IRED activities that do not really exist. Therefore, we decided to triplicate all assay reactions (including negative controls) and use a relatively conservative statistical criterion (mean > 3 standard deviations combined with a cut-off for minimum activity; SI, equation 3, p. S71) for hit identification. As reported in Table S15, this resulted in an estimated false-positive rate of 8% (affecting only low activity levels), which we reduced further by chromatographic re-analysis of selected photometric positives (Tables S16 and S17).*

To better illustrate how low-level activities can be distinguished from background absorbance decrease based on experimental replicates, we have added exemplary absorbance curves to the section “Photometric IRED Activity Assay: Overview Screening” of the SI.

The authors should clarify what measures were taken to ensure that background activity does not obscure low-level IRED activity. Additionally, a control containing amine + IRED + NADPH but no ketone should be included as one of the controls.

— *Compared to false positives (see above), false negatives were of minor concern to us, as they would generally affect activities that are likely too low to be synthetically useful. Nevertheless, we tried to minimise background (and thus maximise sensitivity) by choosing assay conditions that suppress undesired NADPH depletion. For instance, the slightly alkaline pH makes carbonyl reduction by endogenous ADHs less favourable, and bicine was chosen as buffer compound because it leads to lower background NADPH depletion compared to phosphate, for example. Under the chosen conditions, and based on the above-mentioned hit criterion, we estimate a false-negative rate of 4% (Table S15), and we were able to reliably detect IRED activities as low as 5–10 mU/mg for low-background carbonyl substrates.*

‘No carbonyl’ controls were included during early method development (data not shown), but were indistinguishable in all tested cases from controls containing both substrates but no IRED. Of all three types of negative controls (no amine, no carbonyl, no IRED), the ‘no amine’ condition gave the highest levels of background (with marked differences between carbonyls, according to how easily they are reduced by endogenous ADHs), which is why we decided to use this condition for background correction.

2 // Expression level variability

As a limitation of the lysate-based approach, the authors should acknowledge that differences in protein expression levels can influence apparent activity. Without normalization of enzyme expression, activity comparisons across the panel may be affected by variation in expression levels.

— *This is absolutely true, and for that reason we have normalised all enzyme activity data for IRED expression level, determined using a combination of the Bradford total protein assay and SDS-PAGE pixel densitometry. The workflow is described in the SI (section “Determination of Expression Level”, p. S66), but it is only thanks to the reviewer’s comment that we realise we haven’t mentioned this aspect anywhere in the main manuscript text. Therefore, we have now added the statement “To enable the reporting of specific activities, we normalised the apparent rates for IRED expression level, determined using the Bradford protein assay and SDS-PAGE pixel densitometry (see Determination of Expression Level, Supplementary Information).” to the paragraph of the Results section where the use of lysates is first mentioned. We have also added a table listing the expression levels and the parameters from which they were calculated to the SI (Table S32).*

3 // Non-canonical IREDs

The authors should reference recent work on non-canonical imine reductases, and provide rationale for excluding these enzymes in their screening.

— *With the exception of two pyrroline-5-carboxylate reductases (IR0096 and IR0097; UniProt identifiers: POA9L8 and G8BIF4) available within the Roche enzyme collection before the start of this study, we did not include any C=N-reducing enzymes from outside the canonical IRED/RedAm family in our screening, as a relatively homogeneous sequence basis was deemed important for extracting general insights from the collected data. Indeed, the RedAm key motif analysis we performed (SI, section “Protein Sequence–Activity Analysis”) and part of the new analyses added during revision (SI, section “Protein Structure–Activity and Structure–Selectivity Analysis”) rely on multiple-sequence alignment. Besides, the generation of Catalophore point-cloud descriptors of the enzymes’ active sites, which is central to the machine learning workflows used in our study, requires X-ray crystal structures or high-quality homology models of the proteins, which would not have been available for some families of non-canonical IREDs.*

We have added the statement “C=N-reducing enzymes outside the canonical IRED/RedAm family were not considered, as we deemed a relatively homogeneous sequence basis an important prerequisite for extracting generalisable insights from the screening data.”, along with a reference to a recent study focussing on non-canonical IREDs, to the paragraph of the Results section that discusses sequence selection and homology modelling.

Revision Overview

All changes are highlighted blue in the manuscript file. Changes made to the SI are listed here but not highlighted.

Changes requested by reviewers

Changes are grouped by location (sections of main text and SI).

Main article, Results

- *Changed the wording “Deep-Learning Model” in Figure 1 and in the caption of Figure 5 to “Deep-Learning Pipeline” (Reviewer 2).*
- *Added a cross-reference to the tabular overview of screening phases to the subsection “Screening Strategy” (Reviewer 1).*
- *Added a statement explaining the exclusion of non-canonical IREDs to the last paragraph of the subsection “In-Silico Analysis” (Reviewer 3).*
- *Added a statement explaining our approach for determining IRED expression levels to the first paragraph of the subsection “Overview Screening” (Reviewer 3).*
- *Added a statement regarding the origin of bias towards moderate substrate scope and high maximum activity in ML-recruited IREDs to the sixth paragraph of the subsection “Overview Screening” (Reviewer 1).*
- *Added a statement regarding the origin of low reactivity of aromatic substrates to the fifth paragraph of the subsection “Diversification Screening” (Reviewer 1).*
- *Added statements discussing correlations between enzyme structure/sequence and the observed activity and selectivity in reductive amination to the sixth paragraph of the subsection “Diversification Screening” (Reviewer 2).*
- *Rephrased the sentence that summarises our workflow for predicting IRED activity for new substrate combinations in the seventh paragraph of the subsection “Diversification Screening” to better explain the role of deep learning (DL) in this workflow (Reviewer 2).*

Main article, Discussion

- *Added a statement regarding the potential use of z-scores for incorporating additional screening data into the existing data set (Reviewer 2).*
- *Added a statement that summarises the findings of new sequence- and structure-based analyses (Reviewer 2).*

Main article, References

- *Added a reference to a recent study on non-canonical IREDs (ref. 52; Reviewer 3).*
- *Added two references to recent review articles on ML-assisted protein engineering (refs. 64 & 65; Reviewer 2).*

SI, Supplementary Data

- *Rephrased the last paragraph of the subsection “Prioritisation of IRED candidates for the Overview Screening” to provide a hypothesis regarding the origin of bias towards moderate substrate scope and high maximum activity in the recruited IREDs (Reviewer 1).*

— Added a subsection titled “Protein Structure–Activity and Structure–Selectivity Analysis”, where the identification of structural determinants for IRED activity and stereoselectivity through in-silico analysis is discussed (Reviewers 1 and 2).

SI, Supplementary Methods

— Added a subsection titled “Tabular Overview of Screening Phases”, containing Table S27 and accompanying text that summarise the computational and experimental contributions to the overall screening workflow in a step-wise format (Reviewer 1).

— Added Table S32, which lists the IRED expression levels (Reviewer 3).

— Added Figure S28 to the subsection “Photometric IRED Activity Assay: Overview Screening” to better illustrate how low-level activities can be distinguished from background absorbance decrease based on experimental replicates (Reviewer 3).

— Added a subsection titled “Protein Sequence/Structure–Activity and Structure–Selectivity Analysis”, which describes the methods used in the new sequence- and structure-based analyses added during revision (Reviewers 1 and 2).

Changes made to meet *Nat. Commun.* formatting instructions

Changes are grouped thematically.

Section names

— Changed “Supporting Information” to “Supplementary Information” throughout the main text and SI. Changed “Supporting Data” and “Supporting Methods” to “Supplementary Data” and “Supplementary Methods”, respectively, in the SI. Changed section headings in the main text from “Results and Discussion” to “Results” and from “Conclusions” to “Discussion”.

— Created a “Code Availability” section and moved the statement regarding availability of Python scripts there (from the “Data Availability” section).

— Created a “Competing Interests” section and declared no competing interests.

Length Restrictions

— Rephrased the abstract to meet the length limit of 150 words.

— Trimmed the list of references to 70 entries, mainly by removing older references from grouped citations.

Data Availability

— Created a Source Data file for all display items that cannot be directly reproduced from the deposited data set. Added a statement regarding this file to the “Data Availability” section.

Other changes

When we prepared the Source Data file, we discovered a minor data mismatch between the final, consolidated screening data and the median values and hit rates reported in Figure S14. Apparently, the figure had not been updated after the last data consolidation in autumn 2024.

— *Changed the median of non-zero activities for the Starting Panel, First Additions, and Second Additions from 0.029, -0.119, and -0.329 to 0.033, -0.132, and -0.330, respectively. Changed the number of positives for the Starting Panel, First Additions, and Second Additions from 3005, 1238, and 714 to 3003, 1237, and 711, respectively. The corresponding percentages were not affected by this change.*

When we prepared the Source Data file, we discovered that axis labels were missing from Figure S7. Moreover, we decided that data labels would be helpful, as the smallest bars of the chart are barely visible.

— *Prepared an updated version of Figure S7 with axis labels and data labels.*

Additional biotransformations were carried out to assess the predictive value of the stereoselectivity key residues discussed in the new section of the SI, raising the total number of reactions for which calibrated conversion and *ee* data are reported.

— *Changed number of calibrated conversion and *ee* values reported in Figure 1 (“Overall Outcome” box) from 299 to 328.*

The addition of new material in the Supplementary Information made the renumbering of Supplementary Figures and Tables necessary.

— *Updated the numbering of all affected Supplementary Figures and Tables and of cross-references to these elements in the main manuscript text and the SI.*

Response to the Reviewers' Comments

Reviewer 1

Thank you for responding and addressing all comments and shortcomings of this manuscript.

I have no further points and I look forward to seeing this manuscript published.

— *We thank the reviewer for the positive response.*

Reviewer 2

The author has provided a thorough explanation and made revisions in response to my comments, and these new evidence answers my questions, so I recommends its publication in Nature Communications.

— *We thank the reviewer for the positive response.*

Reviewer 3

The revised manuscript by Schrittwieser and colleagues addresses all of my previously raised concerns and represents a significant improvement. An impressive amount of data has also been included in the Supporting Information. The analysis of stereoselectivity is particularly interesting (Table S11).

— *We thank the reviewer for the positive response.*

Minor suggestions: The authors should consider including reaction schemes for Table 1 and Figure 5. They should consider using a colour scheme to distinguish fragments of the product derived from the carbonyl and amine coupling partners.

— *We have considered including reaction schemes in these display elements, but since the basic scheme of the reductive amination reaction remains the same for the entire article and is presented twice in the early sections (Figure 1a, Figure 2), we have opted against repeating it in later sections.*

While reformatting Table 1 and Figure 6 (previously numbered Figure 5) according to the Nature style, we have tried versions with black structures and versions with color-coded structures that indicate which parts of the molecule derive from which coupling partner. We have found the black versions more visually appealing, in particular against the ochre table background, and trust that the readers will be easily able to trace the products back to the starting materials from which they are derived based on the compound naming scheme.

Revision Overview

All changes are highlighted blue in the manuscript file. Changes made to the SI are listed here but not highlighted.

Changes made to meet *Nat. Commun.* formatting instructions

Changes are grouped thematically.

Text formatting

- *Removed underline format for weblinks.*

Mathematical terms

- *Revised formulae (2) and (3) in the SI (subscripts “NADPH”, “reaction” not italic).*

Figures and figure legends

- *Reformatted all figures in the main article according to the provided instructions (e.g., changed fonts to Arial, ensured proper dimensions, ensured proper formatting of chemical structures).*
- *Amended figure legends with statistical information and specifications of box plots. Added the statement “Source data are provided as a Source Data file.” to all relevant figure legends in the main article and SI.*
- *Split Figure 4 into two separate figures, duplicated and revised its figure legend, renumbered figures accordingly.*

Methods

- *Added information on relative centrifugal force ($\times g$) for all centrifugation steps.*

References

- *Added formal citations of the dataset deposited in Mendeley Data and the GitHub repository archived in Zenodo to the reference list (refs. 71 and 72).*

Other changes

The Njardarson group has recently published their latest *Top 200 Pharmaceuticals* posters with data from 2025, which has allowed us to update the estimates of nitrogen atom and chiral amine / amino acid prevalence in pharmaceutical drugs in the first paragraph of the Introduction.

- *Changed the second sentence of the Introduction to: “Over 90% of the 200 top-selling small-molecule drugs in 2025 contain at least one nitrogen atom, with 47% containing a non-racemic α -chiral amine or amino acid moiety.” Updated ref. 2 to cite the latest version of the Top 200 Small-Molecule Pharmaceuticals poster.*