

Evolutionary insights into the stereoselectivity of imine reductases based on ancestral sequence reconstruction

Corresponding Author: Professor Gao-Wei Zheng

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)

This m/s is an interesting exploration of how enantioselectivity can evolve in an enzyme. The authors study the evolution of R-stereoselectivity in a group of phylogenetically related imine reductases (IREDs), finding that while one mutation is paramount (D170Y), five others are required for complete inversion of enantioselectivity. They create all mutants between the wild-type ancestor and the mutant with all six residues changed and characterise the protein landscape.

While generally readable and well structured, this manuscript would benefit greatly from editing for English usage and grammar – the meaning of some wording is not completely clear and there are many grammatical and stylistic errors.

Overall the approach is sound and the data support the findings (with some issues raised below). However it is quite remarkable that none of the ee data on which the interpretations are based have been reported with any replicates or statistical analysis, i.e. there are no indications that repeat measurements were performed or estimates of experimental error in numerical values (SD, SEM etc.). Therefore the reader cannot determine which changes in ee are significant. While this reviewer accepts that mapping a landscape is a lot of work to do in duplicate, one would assume some of the initial characterisation of ee in site directed mutants could have been replicated. However, it appears that all significant conclusions are based on single measurements (e.g. see below) with no statistical analysis.

(The only data for which estimates of experimental error are cited are specific activity measurements in supp table 2, which are not referred to anywhere in the paper. Even here, however, there are no statistical analyses applied to determine, e.g., whether the results for the alternative ancestors differ from those obtained from the most probable ancestors.)

The other major point is that while the authors have quantified the specific activity of the ancestors and mutants derived from them, there is no comment on or discussion of the large variations in specific activity seen across ancestors, or between the most probable and second-most probable (alternative) ancestors.

Detailed comments:

Lines 56-58: This needs further qualification: an ancestor may not have as many epistatic complications as inserting a residue into an extant protein but it is not free of them. Please explain how an ancestor is 'to some extent capable of avoiding epistasis'.

Line 88-92: the Y-type sequences all form a relatively late branching clade in the tree shown in Supp Fig 1, not a 'deep node' as the authors report.

Line 100 and Fig 1B: Fig 1b appears to show that N1 exhibited selectivity towards the R enantiomer in contrast to what is written in the text and elsewhere.

Lines 551-552 in Fig 2 legend: the mutations listed do not match exactly: I120V vs V120A? It would help if these mutations were listed in the text for clarity at ~ line135.

Figure 3: Again one mutation shown in Fig 3B does not match the text: D170 A should presumably be D170Y?

Figure 4A: It is somewhat counter intuitive to show the orientation of the substrate in N559 in a structure of the mutant without any other information about the WT binding site. The reader is left to wonder how the shape of the WT active site differs. This comparison would be better done by showing side by side from the same perspective the way the substrate docks into the wt and M6 forms. The reader will be able to see the effect of the D170Y mutation in its proper context in this way.

Figure 4B: In a similar way it is confusing to show interactions from two different chains in the same figure. If the residues that interact differ between the two chains in the asymmetric unit of the crystal structure, please show them in separate images as the combined image is not necessarily reflective of either individual arrangement.

Fig 5 and lines 188-190: It is not clear why certain pathways are deemed 'inaccessible' simply because they lead to a decrease in enantioselectivity in the next node: unless selection pressure is particularly strong, i.e. so that a mutant with a decrease in enantioselectivity would be unlikely to survive to reproduce. There is no reason a modest decrease cannot then be counteracted by a more substantial increase in ee in a subsequent step.

Lines 185-188 and 194: The way these analyses were done is not mentioned in the methods. E.g., what substrate was used for the landscape analysis? Were the assay conditions the same as for the other assays? What inputs were provided for the MAGELLAN analysis

Supp Fig 8: What is the Y axis label?

Lines 198-201: The assumption made that the most likely fixation order is one that reflects a steady increase in ee assumes the only factor influencing evolution is the ee. However, this is an analysis of natural evolution and there is no information provided about the natural role of these enzymes and how essential they may or may not be to the fitness of the source organisms. Other factors such as protein stability, solubility etc. are also likely to change with such mutations and may influence the viability of the organisms and thereby the likelihood that mutations are fixed. Moreover, any amino acid in a given protein will be evolving under the influence of multiple selection pressures (see Dutton and Moser DOI: 10.1039/c005523a for a good review of this point). Some discussion of this issue is essential.

Lines 202-209: For the general audience of this paper, higher order epistasis should be defined. i.e. what is meant by e.g. sixth order epistasis.

Supp Table 2: The specific activity differed markedly between some ancestors and their alt versions as well as varying ~ an order of magnitude across the ancestors resurrected from the evolutionary tree. In particular the two newest ancestors were approximately an order of magnitude less active than the other ancestors. This point is completely ignored in the manuscript.

Supp Table 4: Please show the specific activity as well as the change in ee. It would be useful to quantify protein expression as well to determine whether there is a decrease in kcat.

Lines 243-244: the authors state "Therefore, high-order epistasis should be incorporated into protein design and protein engineering technologies in the near future". How would the authors propose to do this?

Lines 246- 248: The authors state "Furthermore, we demonstrate the D170Y is an evolvability-enhancing mutation—the ability to increase likelihood that subsequent substitutions are adaptive in evolving protein molecules—in this empirical evolutionary landscape." It is not clear what the authors mean here.

The rest of the paragraph seems to justify this statement by showing D170Y was essential to completely invert enantioselectivity, pointing out that the L19M/A119G double mutant (131% ee) is the only one that shows R selectivity but no mutation at residue 170. All other double mutants showing R-stereoselectivity contained D170Y. A reasonable question therefore is how significant is the inversion in this (L19M/A119G) mutant, i.e. what is the experimental error in these measurements. An alternative interpretation here is that the D170Y is simply essential to revert the stereoselectivity and the single measurement for L19M/A119G is an outlier!

Lines 274-276: The authors comment "We then deleted large gaps in the MSA results to decrease subsequent phylogeny reconstruction noise. Locations where >50% of sequences were gaps in MSA were considered large gaps."

This needs clarification and further justification. (How large? Some estimates of length and the threshold used are needed.) The authors appear to mean that where some sequences had large insertions but the majority of sequences had no sequence, the inserted sequences were edited out of the sequences in which they were found. Such insertions may arise from either real evolutionary events or as errors in virtual translation of sequences. If the former, this is removing a real factor in an enzyme's evolution. If the latter, it suggests the sequence may not be reliable (e.g. miscalled when the genomic DNA was 'translated'). Therefore care needs to be taken either way and a judgement made as to whether the sequence should be included in its entirety (if likely to be real, e.g. if found consistently in one lineage of a tree) or not (if likely to be erroneous). The authors should comment on the likely source of these insertions and justify why they edited the sequences – a subjective intervention - rather than simply removing or including them as is routine practice in the field. See ref 41 methods section for a comment on how these issues are usually handled.

Line 288: The authors use a marginal reconstruction but for comparisons across an evolutionary tree, joint reconstruction is more appropriate. GRASP can do both. The authors should justify why marginal was chosen instead of joint.

Lines 289-290: The authors say they constructed the 'second most likely ancestors for each ancestral sequence (alternative ancestors) to verify the robustness of ASR'. The only data shown for these alternative ancestors is in Supp Fig 4 (where ee values are compared) and Supp Table 2 where the specific activity (U/mg) is provided. (There is no citation of this latter table in the text however.) The specific activity obtained for the alt ancestors is markedly different from the values obtained for the most probable ancestors, suggesting that in this case the results obtained from the ASR are not robust as concerns specific activity, even if they were for ee (Supp Fig 4 vs Fig 1D)!

Lines 289-290: Can the authors comment on how different the second most probable ancestor sequences are from the respective most probable ones? What explains the large differences in specific activity (approximately an order of magnitude in many cases)?

Supp Info Pages 2-18: in the Supp info the sequences of ancestors characterised are presented as both nucleotide and amino acid sequences. However all appear to have been subcloned into a pET vector which added a hexa-His tag. Therefore it is slightly misleading to not include this tag in the OPRF of the resurrected enzymes.

P.10, bottom: The statement "For crystallization of N559 and N559-M6, truncated proteins were constructed by enzyme cut and link-up experiments" needs explanation. Do the authors just mean cut with restriction enzymes then ligated? Please provide details of where the proteins were truncated to.

p. 12: The assay for IRED activity relies on monitoring the loss of absorbance from NADPH at 340 nm. From this the number of μmol converted to product was extrapolated. However NADPH oxidation is not particularly specific to the reaction of interest. What quality controls were undertaken to ensure NADPH oxidation faithfully reflected product formation, i.e. imine reduction?

Minor points:

Line 299: The correct name of "LB medium" is "Lysogeny Broth".

Lines 306, 310, 313, 318, etc: x g values should be used rather than rpm values to describe centrifugation steps.

Lines 305-306: "Terrific Broth medium" is a tautology: the name of the medium (broth) is just "Terrific Broth".

Supp Fig 2: the top left SDS-PAGE gel is obscured by water droplets and bubbles. Please take another image of this gel for a publication quality figure. This figure is not referred to in the text.

Reviewer #2

(Remarks to the Author)

Zhu et al. report an ancestral sequence reconstruction study into the evolution of stereoselectivity in imine reductases. They claim to have identified a stereoselectivity inversion in the evolution of these enzymes, as well as the historical mutations responsible for the inversion.

One potential problem that I find with this work is that the authors appear to have relied heavily on computational tools and resources available from web servers. This is not necessarily wrong, of course. However, many of these tools involve assumptions that must be critically examined. For instance, the authors use the tool GRASP very recently developed by Gillam and coworkers (ref. 58 in the submitted manuscript). It is to be noted that GRASP is based on maximum likelihood, while most researchers in the ancestral reconstruction field consider Bayesian approaches to be more reliable. Certainly, maximum likelihood allows for the analysis of a larger number of sequences, which was one of the main reasons for developing GRASP. However, Gillam and coworkers have also emphasized the need to extensively curate the alignment of sequences used as starting point for ancestral reconstruction (Journal of Biological Chemistry 298(10) 102435, 2022), in order to eliminate incorrect sequences harboring different types of errors. Sequence curation is a process that is done manually to a substantial extent. The point I am trying to make is that the authors use an alignment of 1407 sequences as input for GRASP, but do not make it clear in the submitted manuscript how curation was carried out, except for a brief and obscure reference to the deletion of "large gaps in the MSA results to decrease subsequent phylogeny reconstruction noise" (top of page 10).

In addition, the authors do not clearly discuss the reliability of the phylogenetic tree used as basis for reconstruction. Plausible evolutionary interpretations of the biophysical properties of "resurrected" proteins require that the phylogenetic tree used resembles to some substantial extent the tree of life or, at least, the accepted phylogeny of the protein system under study (in case the evolutionary history of the protein family involves gene duplication events). I assume, although this does not appear to be clearly stated in the manuscript, that among modern imine reductases there are examples of enzymes with S- and R- stereoselectivity. Furthermore, I find it plausible that imine reductases with S- and R- stereoselectivity exist in the same organism, reflecting gene duplication events. Do the phylogenetic nodes targeted for reconstruction correspond to speciation events or to gene duplications events? Is there a correspondence between the phylogenetic tree used (Supplementary figure 1) and an accepted phylogeny for imine reductases?

The authors claim that a stereoselectivity inversion occurred during the evolution of imine reductases. As judged from the data in Figures 1 and Supplementary 4, it appears that the inversion was sharp instead of gradual. The manuscript does not seem to provide a convincing evolutionary narrative for the sudden (in evolutionary time scales) inversion. What kind of

change in intra-cellular or extra-cellular environment triggered the stereoselectivity inversion? Was there a change in metabolism that required imine reductases to change stereoselectivity? Do the stereoselectivities of modern imine reductases shed some light on this hypothetical ancestral change?

The authors use ibis2analyzer to perform the co-evolution analysis that leads to their identification of the historical mutations responsible for the stereoselectivity inversion. This also poses several issues. Co-evolution analysis is not straightforward and different approaches may lead to different results, in particular when using a limited number of sequences in the analysis (which seems to be the case here because, if I understand correctly, about 10 ancestral sequences were used). It is possible then, that other mutations, in addition to the 6 identified by the ibis2analyzer (page 5), also affect stereoselectivity. The authors do not describe which mutations are "rejected" by ibis2analyzer as potential causes of stereoselectivity inversion and how reliable such rejection is.

Based on the 6 mutations selected, the authors derive an adaptive landscape of stereoselective evolution. This obviously involves the determination of stereoselectivity for 64 combinations of mutations. Then the authors use the MAGELLAN to ascertain the impact of different types of epistasis on the determined landscape. MAGELLAN is, however, used essentially as a black box and no serious attempt to provide intuitive, convincing explanations for the results obtained is reported.

Overall, I find that many of the analyses, discussions and evolutionary interpretations (or lack of) provided in the manuscript are not convincing. Certainly, the identification of a small number of mutations that can be combined to switch stereoselectivity is interesting. However, it is not clear from the manuscript whether the same of equivalent information could have derived from the comparative analysis of the stereoselectivity of modern imine reductases.

Reviewer #3

(Remarks to the Author)

Zhu et al. present intriguing results aimed at providing insight into the evolution of stereoselectivity from ancestral imine reductases to modern proteins. They have resurrected ancient imine reductases using a single phylogenetic tree of extant descendants and determined the stereoselectivity of these ancient proteins through biochemical analyses. The results revealed six mutations responsible for changing stereoselectivity from S- to R-stereoselectivity. X-ray crystallography elucidated the structural mechanism for regulating stereoselectivity inversion. The authors then investigate the adaptive landscape of stereoselectivity evolution. From these findings, the authors assert that ancestral sequence reconstruction offers an efficient method for identifying mutations that alter protein functions.

The paper addresses an interesting question, and the approach used is appropriate. Overall, I believe the authors have done an excellent job of identifying the amino acids involved in the acquisition and transition of protein stereoselectivity. In particular, by expanding their research beyond ancestral sequence reconstruction and biochemical analysis of resurrected proteins to include crystal structure analysis and binding modeling, the authors' claims regarding the mechanism of stereoselectivity acquisition are strongly supported. However, I do have some concerns that should be addressed before my enthusiasm can be fully endorsed.

The results presented are exciting and provocative in the realm of protein chemistry and enzyme chemistry research. However, if discussing evolution, it is necessary to verify the reliability of the phylogenetic tree and the confidence in the estimated ancestral sequences. Especially, the branch connecting the target protein family and the outgroup is very long, and it cannot be said that the root position is sufficiently specified with confidence. Regarding the root position, careful interpretation is necessary.

The phylogenetic construction details are not sufficient. The tree robustness as well as the sequence curation needs more clarification. The manuscript does not sufficiently explain the branch lengths (unlabeled on the tree), root and out-group choice with satisfactory detail. In addition, it would be nice to see some measure of branch support, especially the nodes for resurrected proteins. Otherwise, no one can see if the tree topology is well or poorly supported. Because the algorithms for inferring ancestral sequences assume that the given phylogeny is true, this problem substantially undermines confidence in the inferred ancestral sequence. Providing information on the likelihood or posterior probability of each ancestral amino acid in the resurrected sequences would be also advisable.

The tree's architecture should also be described in more detail. Supplementary Fig. 1 depicts only the branches of the phylogenetic tree, making it difficult to determine the distribution of species. It should be labeled to at least distinguish between Archaea, Bacteria, and Eukarya, and to indicate the distribution of major phyla. Furthermore, it should be compared to known phylogenetic trees to assess its reliability. Without this comparison, the reliability of the phylogenetic tree cannot be evaluated.

It would be appreciated if a list of sequences (accession numbers) included in the phylogenetic tree could be provided as supplementary data.

Regarding the enzymatic characteristics, the relative stereochemical selectivity has been thoroughly investigated. However, while the catalytic activity of the ancestral proteins is listed in Supplementary Table 2, there is no comparison with wild-type (extant) enzymes. This is, in my opinion, very unfortunate as the properties of the ancestrally reconstructed variants cannot be directly compared with the starting sequences and that significantly impacts the relevance of the results. Additionally, it seems interesting to explore the K_m and k_{cat}/K_m of extant and reconstructed proteins for each substrate as well.

It is undeniable that the authors' epistasis analysis has provided crucial insights into the evolution of stereochemical selectivity. However, the impact of epistasis on protein evolution has been debated for the past 20 years. It would be appropriate to cite the following landmark papers and discuss them:

doi: 10.1038/nature05385

doi: 10.1038/nature08249

doi: 10.1038/nature23902

In the Discussion section, the authors describe, "This differs from previous studies in which inverting stereoselectivity through protein engineering involved mutations that typically interact directly with substrates." However, it is not fair. It is well known that amino acid substitutions remote from the active site often influence activity. ASR can be a very powerful tool for accessing these mutations - as outlined in the paper (which you should cite in this context):

doi.org/10.1042/BST20180614

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)

The authors have satisfactorily addressed my concerns.

Reviewer #3

(Remarks to the Author)

It is undeniable that the authors have made efforts to revise the manuscript in accordance with the reviewers' comments. However, it still seems that the revisions made do not satisfactorily address the initial review comments. Specifically, as the authors themselves state that "Inferring the last common ancestor (LCA) of all extant imine reductase (IREd) sequences is a significantly key part of this study," the placement of the root within the IREd phylogeny is critically important. Nonetheless, it appears that the issue of the root position's reliability has not been adequately addressed in the revised manuscript. Therefore, in my opinion, further revisions are necessary for publication.

Below, I provide additional comments on each response from the authors.

Respond #3-1:

Upon examining Supplementary Fig. 1, it is evident that the branch connecting the outgroup to the IREd clade is significantly longer compared to the branches within the IREd clade. This suggests that the reliability of the root position within the IREd clade is not beyond doubt. Additionally, the number of outgroup sequences is extremely limited compared to the IREd clade sequences, which raises further concerns about the accuracy of the root position. It seems entirely possible that another position within the IREd clade could be the LCA. Therefore, additional evidence supporting the current root position is necessary, such as verifying the root position using alternative methods.

Respond #3-2:

Please include the phylogenetic construction details in either the main text or the supplementary text. Additionally, while the average values for the entire sequence are shown in Table R5, it is recommended to provide a figure or table in the Supporting Online Materials that shows the posterior probability for each ancestral amino acid residue.

Respond #3-3:

In my first-round review comments, I requested that the authors show the distribution of major phyla on the phylogenetic tree. While I realized that almost all sequences are derived from Bacteria, there is no discussion on the distribution of bacterial phyla or how this compares to previously reported phylogenetic trees of all organisms or specifically bacterial phylogenies. Furthermore, if the authors aim to demonstrate the robustness of the tree topology, it would be more convincing to compare it with a phylogenetic tree constructed using Bayesian methods, rather than another tree based on a maximum likelihood method. Please note, I am not expressing concerns about the tree topology, but I would like to see a comparison with other well-known phylogenetic trees.

Respond #3-4:

What are the values in column B of the Excel file? For outgroup sequences, it seems to be the UniProt ID, but there is no explanation for the IREd sequences, making it unclear.

Respond #3-5:

It appears that the authors have not addressed the comment from the first-round review requesting a comparison of the catalytic activity of ancestral enzymes with that of wild-type enzymes (extant forms, especially ScIR and SIR). Values for the wild-type enzymes are not provided in Supplementary Table 2. According to Supplementary Table 7, the *k_{cat}* value for N560 is 1/10 compared to ScIR, but it would be interesting to know how this translates to the specific activity.

Additionally, no explanation for AltN1-N5 and Alt557-560 is found in the revised manuscript. Please add an explanation in both the main text and Supplementary Table 2.

How can the significant difference in specific activity between N560 and Alt560 be explained? Please provide a discussion

from the perspective of the reliability of the ancestral sequence inference.

Respond #3-6:

I believe that the reaction rate kinetics parameters added by the authors would be helpful for understanding this paper. However, it seems that the new Supplementary Table 7 is not cited in the main text, and there is no description of these results in the main text either. Additionally, there is no explanation of how the reaction rate kinetics parameters were determined, including the reaction temperature and composition of the reaction mixture.

Respond #3-7 :

I think it has been appropriately addressed.

Response #3-8:

I think it has been appropriately addressed.

I looked at the comments from Reviewer 1 and the authors' responses, as per [Editors] request. While the authors have adequately addressed many of Reviewer 1's comments, there are several instances where the responses appear to be insufficient, as I will point out below.

Response #1-3: Even if the paper's main focus is on the inversion of stereoselectivity, the significant differences in specific activity observed between the reconstructed ancestral proteins should be mentioned and discussed in the revised manuscript. The authors suggest that these differences might be due to variations in optimal pH; however, this explanation lacks evidence and remains speculative.

Response #1-4: The authors' discussion seems to remain speculative, and no evidence has been provided to support their claims. Specifically, which amino acids in the extant proteins reduce the effect?

Response #1-15: The explanation appears to be insufficient. Does "six-order epistasis" refer to the interaction among six mutations? Please clarify.

Response #1-16: While the low activity of some Alt forms is somewhat understandable based on the explanation provided in response #1-3, it is necessary to include a reasonable explanation in the manuscript for the significantly lower specific activities of N559 and N560, especially N559. Since the enantiomeric excess selectivity is inverted between N559 and N560, the low specific activities of these forms must be thoroughly examined. Additionally, the manuscript should discuss the potential relationship between the inversion of stereoselectivity and the reduction in specific activity. Failing to do so could raise concerns about the reliability of the results, particularly regarding the accuracy of the ancestral sequences.

Response #1-17: The results of the additional experiments should be explained and discussed in the main text of the revised manuscript.

Response #1-18: The authors' response remains abstract and lacks specificity. Additionally, it has not been reflected in the main text.

Response #1-20: The response merely states that the authors have mimicked previous studies and does not provide a satisfactory answer to the reviewer's suggestions or concerns.

Response #1-21: The authors should also state their response to the reviewer in the main text of the revised manuscript.

Response #1-22: I agree with the concerns raised by Reviewer 1. Even if similar results have been reported in other ASR studies, this does not support the robustness of the inferred ancestral sequences (and their catalytic properties). At the very least, the authors need to provide such an explanation in the main text.

Response #1-22: The authors should explain in the main text that the ambiguous ancestral amino acid sites in the ancestral sequences significantly influence the specific activity. In other words, while the evolution of stereoselectivity from the ancestral sequences can be reproduced, the magnitude of specific activity may not be reproducible. This might explain the significantly lower specific activity observed in N559.

Version 2:

Reviewer comments:

Reviewer #3

(Remarks to the Author)

I have carefully reviewed the revised manuscript and found that the authors have adequately addressed all of my previous comments. I believe the revisions enhance both the clarity and quality of the manuscript. I have no further concerns at this time.

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/>

Point-by-Point Responses

Contents

Response to Reviewer #1 Page 2-26

Response to Reviewer #2 Page 27-37

Response to Reviewer #3 Page 38-48

Response to Reviewer #1

Reviewer #1 (Remarks to the Author):

Comments: This m/s is an interesting exploration of how enantioselectivity can evolve in an enzyme. The authors study the evolution of R-stereoselectivity in a group of phylogenetically related imine reductases (IREDs), finding that while one mutation is paramount (D170Y), five others are required for complete inversion of enantioselectivity. They create all mutants between the wild-type ancestor and the mutant with all six residues changed and characterise the protein landscape.

We thank you for dedicating time to assessing our manuscript and providing invaluable feedback on our work. We really appreciate your invaluable suggestions for helping us improve this manuscript.

1. While generally readable and well structured, this manuscript would benefit greatly from editing for English usage and grammar – the meaning of some wording is not completely clear and there are many grammatical and stylistic errors.

Response #1-1:

Thanks for your advice. All authors have made careful modifications to the original manuscript. Additionally, English language of this manuscript has been polished by the professional team available at International Science Editing.

2. Overall the approach is sound and the data support the findings (with some issues raised below). However it is quite remarkable that none of the ee data on which the interpretations are based have been reported with any replicates or statistical analysis, i.e. there are no indications that repeat measurements were performed or estimates of experimental error in numerical values (SD, SEM etc.). Therefore the reader cannot determine which changes in ee are significant. While this reviewer accepts that mapping a landscape is a lot of work to do in duplicate, one would assume some of the initial

characterisation of ee in site directed mutants could have been replicated. However, it appears that all significant conclusions are based on single measurements (e.g. see below) with no statistical analysis. (The only data for which estimates of experimental error are cited are specific activity measurements in supp table 2, which are not referred to anywhere in the paper. Even here, however, there are no statistical analyses applied to determine, e.g., whether the results for the alternative ancestors differ from those obtained from the most probably ancestors.)

Response #1-2:

Thanks for your kind suggestions, which is valuable for improving the accuracy of the manuscript. We conducted the experiments for the stereoselectivity analysis of all enzymes three times (Figure R1–R5, Table R1). The error bars have also been added to the Fig. 1 and Fig. 2 (average value $\pm$ SEM) in the revised manuscript. In addition, the original spectra of gas chromatograms and liquid chromatograms of these repeated experiments have been present in revised supporting information as Supplementary Fig. 3-6 and Supplementary Fig.10

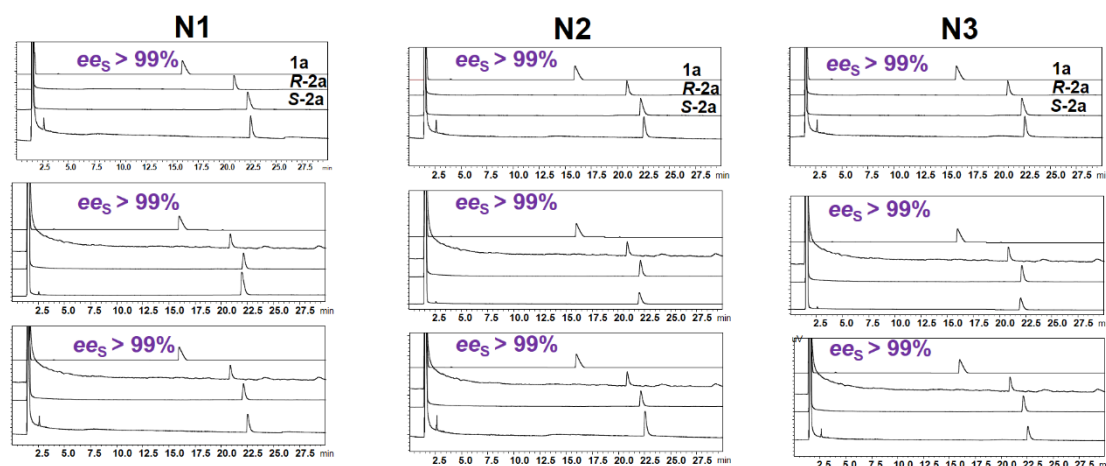


Figure R1. Original chromatograms for the stereoselectivity analysis of N1-N3.

The numbers on the graph denote the corresponding ee in each measurement. The experiments were conducted three times for each enzyme.

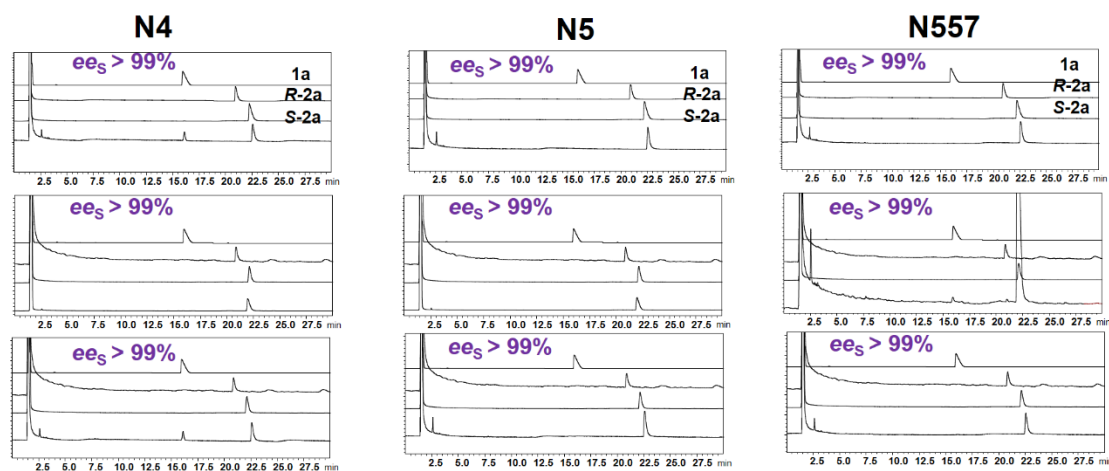


Figure R2. Original chromatograms for the stereoselectivity analysis of N4-N557.

The numbers on the graph denote the corresponding ee in each measurement. The experiments were conducted three times for each enzyme.

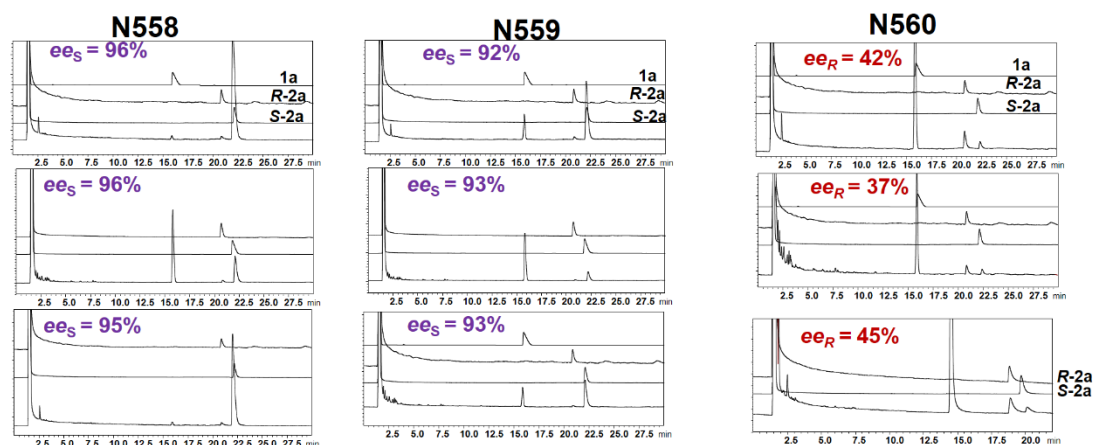


Figure R3. Original chromatograms for the stereoselectivity analysis of N558-

N560. The numbers on the graph denote the corresponding ee in each measurement. Some measurements showed different retention times due to the change in column efficiency during this study. The experiments were conducted three times for each enzyme.

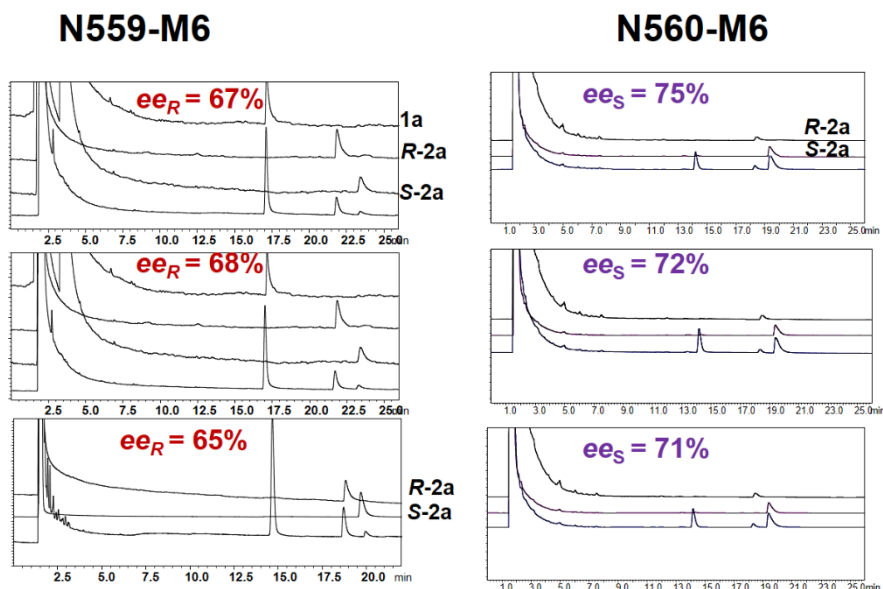


Figure R4. Original chromatograms for the stereoselectivity analysis of N559-M6 and N560-M6. The numbers on the graph denote the corresponding ee in each measurement. The experiments were conducted three times for each enzyme.

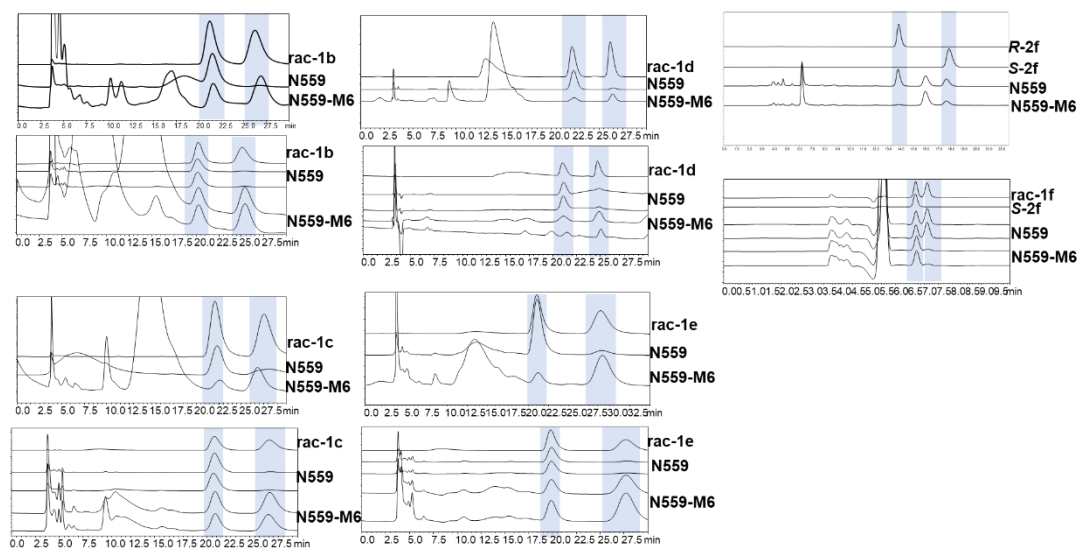


Figure R5. Original chromatograms for the stereoselectivity analysis of N559 and N559-M6. *R*- and *S*-2f are products of 1f. The product 2f was derivatized by acetic anhydride, however, we found later that the configuration of 2f could also be analyzed by liquid chromatography without derivatization. Therefore, 1f has two types of chromatogram.

Table R1 Stereoselectivities of N559 and N560 toward **1b-1f**.

Substrates	Enzymes	<i>ee</i> (%)		
1b	N559	89	90	89
	N559-M6	-40	-42	-41
1c	N559	86	87	88
	N559-M6	-19	-17	-18
1d	N559	84	89	86
	N559-M6	-29	-32	-24
1e	N559	84	90	82
	N559-M6	-41	-43	-48
1f	N559	-17	-18	-25
	N559-M6	88	89	80

Notes: The negative numbers denote *R*-stereoselectivity. Due to structural similarity, the products' stereoselectivity of **1b-1e** referred to **1a** and our previous publication. (Zhang et al. *Org. Lett.* **2020**, 22, 3367-3372).

3. The other major point is that while the authors have quantified the specific activity of the ancestors and mutants derived from them, there is no comment on or discussion of the large variations in specific activity seen across ancestors, or between the most probable and second-most probable (alternative) ancestors.

Response #1-3:

We thank your attention to this aspect very much. Given that all ancestral enzymes have a lower specific activity level, we decided to repeat the activity measurement using gas chromatography (Response Table R2). These data are updated in Supplementary Table 2. Our results show that the difference in a specific activity is not due to experimental errors. For the difference across ancestors (between the most probable and alternative ancestors), we inferred that the conformation and optimal pH difference are mainly responsible for it. Each ancestral enzyme will accumulate new mutations relative to the last common ancestor (N1) during evolution. In spite of the fact that these mutations do not result in essential function differences, they may affect the conformation of enzymes when they react with substrates. Due to the fact that conformation differences

play a crucial role in enzyme reactions, ancestors with several new mutations may exhibit differences in the values of specific activities. Additionally, the optimal pH may be different for each ancestral ancestor. Because the measurements of specific activities were conducted under the same conditions rather than under their respective optimal conditions, we would observe great differences across ancestors.

Additionally, the difference of activity between the most probable and alternative ancestors was also found in relevant studies. For example, in the first research investigating the evolution of stereoselectivity using ASR, we found that the conversion of maximum-likelihood and alternative enzymes presented in Fig. 3 shows a significant difference (Chiang, C-H. et al. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, 120(15), e2218248120). In a recent study investigating the evolution of phosphorylase activity in ancestral glycosyltransferase, the maximum-likelihood ancestral enzyme shows up to 10^2 time more K_{cat}/K_m than the alternative ancestral enzyme (Figure 1d, Franceus J. et al. *ACS catal.* **2024**, 14, 3103-3114). More importantly, although there is a significant difference in specific activity between maximum likelihood and alternative enzymes, stereoselectivity remains a consistent trend. Therefore, we argue that our reconstruction is robust. Considering stereoselectivity is the focused property in this manuscript and each ancestral enzyme can catalyze the reduction of **1a**, we don't perform extra discussion on specific activity. We hope these data and answers will help you better evaluate our work. Many thanks again for your comments.

Response Table R2 The specific activity of each intermediate and the corresponding alternative ancestral sequences from N1 to N560 toward **1a**.

Enzymes	Specific activity (mU/mg)
N1	56 ± 5
N2	46 ± 2
N3	43 ± 6
N4	46 ± 4
N5	60 ± 0.9
N557	62 ± 1
N558	30 ± 1
N559	1.4 ± 0.1
N560	14 ± 0.6
AltN1	38 ± 4
AltN2	20 ± 1
AltN3	53 ± 3
AltN4	39 ± 3
AltN5	80 ± 0.5
AltN557	18 ± 1
AltN558	60 ± 3
AltN559	4.2 ± 0.09
AltN560	0.34 ± 0.1

Specific activity was measured by gas chromatography.

Detailed comments:

4. Lines 56-58: This needs further qualification: an ancestor may not have as many epistatic complications as inserting a residue into an extant protein but it is not free of them. Please explain how an ancestor is ‘to some extent capable of avoiding epistasis’.

Response #1-4:

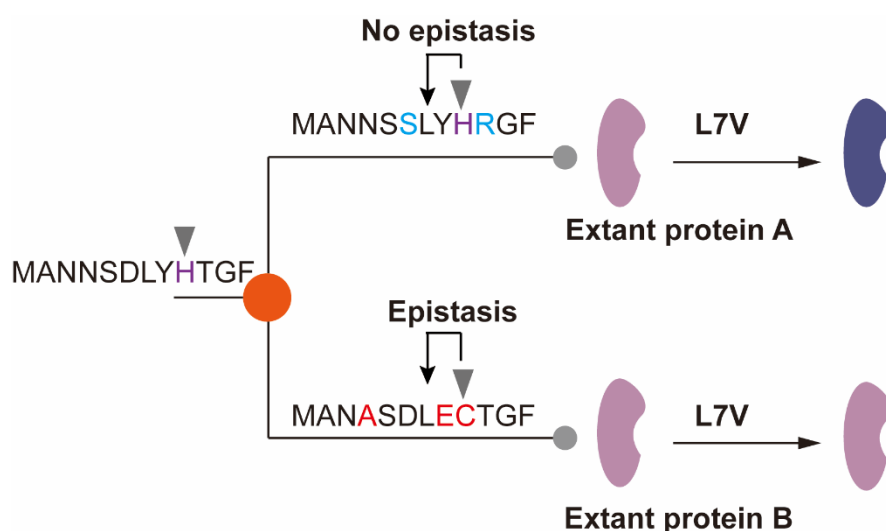
We greatly appreciate the insightful suggestions raised by you. As you suggested, we have herein provided an explanation. **In brief, a mutation that affects function fails**

to produce the desired effect in some homologous sequences. In this scenario, the mutation cannot produce the desired effect because other mutations have been accumulated in the extant sequence background. These mutations form epistatic interactions with this mutation. However, the ancestral sequence has not yet acquired those epistatic mutations; therefore, it can be used to investigate mutations that are functionally critical for the entire protein family.

In a review by Gumulya et al. (Gumulya Y. and Gillam EMJ, *Biochem. J.* **2017**, 474, 1-19), researchers suggested that ancestral enzymes have the ability to avoid epistasis to some extent. Their main viewpoints are as follows: In traditional biochemical research to investigate residues responsible for protein functions, the residue-swapping experiment usually fails to produce expected results by reason of epistasis. Researchers argue that the extant sequences evolved from the recent common ancestors; however, it is possible that epistasis does not appear in each evolution pathway. We further explained its implication with the help of an illustration (Figure R6). In Figure R6, the extant protein A does not accumulate historical substitutions that contribute to epistasis, while B does. In this case, the L7V mutation has different effects on proteins A and B. Despite the fact that the L7V mutation plays an important role in the evolution of this protein family, epistasis in protein B prevents a conclusion from being drawn. Therefore, when we set out to investigate the residues important for protein function from a whole protein family perspective, a horizontal biochemistry study is not appropriate. This is due to epistasis common in extant sequences.

In this review, the researchers wrote that “*ancestral proteins can be used as a suitable background for studying the effect of mutations on functional diversification in the context in which they originally arose, and prior to the introduction of any subsequence, confounding epistatic interactions.*” Moreover, Hochberg & Thornton (2017) also expressed the same viewpoint. Because historical mutations are reintroduced into ancestral sequence backgrounds identical to or very similar to those in which they actually occurred, they suggested that ASR would reduce the effect of epistasis. However, a horizontal swap often leads to a nonfunction protein if a mutation

from one homolog is incompatible with the sequence background of the homolog into which it is introduced (Hochberg G. K. A. et al. *Annu.Rev. Biophys.* **2017**, 46, 247-269).



Historical substitution H9C exert epistasis effect on the L7V

Figure R6. An illustration used to explain why ancestors can avoid epistasis to some extent. Extant protein A does not have an epistasis effect on the key mutation L7V, and extant protein B does. The epistasis stems from accumulated historical mutations, and A just does not obtain this type of mutation. The orange circle denotes the recent common ancestor of A and B. Purple and blue represent different functions of proteins.

5. Line 88-92: the Y-type sequences all form a relatively late branching clade in the tree shown in Supp Fig 1, not a 'deep node' as the authors report.

Response #1-5:

We sincerely thank you for providing valuable suggestions and conducting a meticulous review of the manuscript. We sincerely appreciate and acknowledge your excellent suggestions. We have revised this sentence in lines 87-90 in the revised manuscript.

6. Line 100 and Fig 1B: Fig 1b appears to show that N1 exhibited selectivity towards the R enantiomer in contrast to what is written in the text and elsewhere.

Response #1-6:

We sincerely thank you for conducting a meticulous review of the manuscript and pointing out this important mistake. The N1 indeed exhibited *S*-stereoselectivity; we have revised this mistake in Fig. 1b of revised manuscript.

7. Lines 551-552 in Fig 2 legend: the mutations listed do not match exactly: I120V vs V120A? It would help if these mutations were listed in the text for clarity at ~ line135.

Response #1-7:

We truly appreciate the time and effort you put into reviewing the manuscript. We sincerely appreciate your dedication to helping us improve our manuscript. We have changed V120A to V120I, and these mutations (L19M, C67S, T94S, A119G, I120V, and D170Y) have been listed in the text at line 136.

8. Figure 3: Again one mutation shown in Fig 3B does not match the text: D170 A should presumably be D170Y?

Response #1-8:

We gratefully thank you for your careful check. We have corrected D170A to D170Y in Fig. 3b of the revised manuscript.

9. Figure 4A: It is somewhat counter intuitive to show the orientation of the substrate in N559 in a structure of the mutant without any other information about the WT binding site. The reader is left to wonder how the shape of the WT active site differs. This comparison would be better done by showing side by side from the same perspective the way the substrate docks into the wt and M6 forms. The reader will be able to see the effect of the D170Y mutation in its proper context in this way.

Response #1-9:

We sincerely appreciate the excellent suggestion from you. We have revised Fig. 4 according your kind suggestion in the revised manuscript.

10. Figure 4B: In a similar way it is confusing to show interactions from two different chains in the same figure. If the residues that interact differ between the two chains in the asymmetric unit of the crystal structure, please show them in separate images as the combined image is not necessarily reflective of either individual arrangement.

Response #1-10:

We appreciate the suggestions provided by you. We have revised this problem according your kind suggestion in the revised manuscript.

11. Fig 5 and lines 188-190: It is not clear why certain pathways are deemed ‘inaccessible’ simply because they lead to a decrease in enantioselectivity in the next node: unless selection pressure is particularly strong, i.e. so that a mutant with a decrease in enantioselectivity would be unlikely to survive to reproduce. There is no reason a modest decrease cannot then be counteracted by a more substantial increase in ee in a subsequent step.

Response #1-11:

We greatly appreciate the insightful query raised by you. More importantly, we sincerely thank you for the crucial suggestion you provided. **We have revised the relevant expression on lines 255-258.** Given that stereoselectivity has not yet been used as a proxy for fitness in the construction of empirical fitness landscapes, we wondered whether analyzing a fitness landscape constructed from stereoselectivity could lead to similar conclusions as other related studies. We hope that such a study demonstrates that a variety of enzyme properties can be explored from an evolutionary biology perspective. At the same time, we hope to support some evolutionary hypotheses based on our experimental data.

Through statistical analysis, we found that sign epistasis is pervasive in our landscape. Additionally, sign epistasis often limits the number of evolutionarily accessible trajectories in many empirical fitness landscapes. Therefore, we wonder if sign epistasis plays the same role in our landscape constructed by using stereoselectivity as a proxy for fitness. Interestingly, although selection pressure was not considered in this fitness landscape, we would draw the same conclusion as in other empirical fitness landscapes. Additionally, as one review pointed out, although these fitness landscapes don't address actual natural evolution in the presence of selection pressure, these findings would help us obtain a glimpse of actual biological landscapes (Poelwijk F. J. et al. *Nature*, **445**, 25, 383-386). More importantly, our conclusions resulting from this research without selection pressure are in accordance with some viewpoints of Darwinian theory. For example, we found that the formation of *R*-stereoselectivity in N559-M6 is a gradual optimization process, which is consistent with Darwinian gradualism. From these perspectives, our study can provide experimental support for Darwinian theory at the molecular level. It would be greatly appreciated if you could take some time to reassess our discussions.

12. Lines 185-188 and 194: The way these analyses were done is not mentioned in the methods. E.g., what substrate was used for the landscape analysis? Were the assay conditions the same as for the other assays? What inputs were provided for the MAGELLAN analysis.

Response #1-12:

We greatly appreciate your valuable suggestions. We have added the method for constructing the fitness landscape to the method section in lines 367–376 in the revised manuscript.

In the revised manuscript, we used two statistical approaches to analyze epistasis: the MAGELLAN tool and the 'epistasis' Python package. The revised result section was listed on lines 187–219. We used the value at each circle in Fig. 5 as input to

perform statistical analysis. Before running statistic models, we converted integers from Fig. 5 to decimals.

13. Supp Fig 8: What is the Y axis label?

Response #1-13:

We thank you for this query. The Y-axis label is a decimal form of value in Fig. 5., We have provided the definition in Supplementary Fig. 15 in the revised Supplementary Information

14. Lines 198-201: The assumption made that the most likely fixation order is one that reflects a steady increase in ee assumes the only factor influencing evolution is the ee. However, this is an analysis of natural evolution and there is no information provided about the natural role of these enzymes and how essential they may or may not be to the fitness of the source organisms. Other factors such as protein stability, solubility etc. are also likely to change with such mutations and may influence the viability of the organisms and thereby the likelihood that mutations are fixed. Moreover, any amino acid in a given protein will be evolving under the influence of multiple selection pressures (see Dutton and Moser DOI: 10.1039/c005523a for a good review of this point). Some discussion of this issue is essential.

Response #1-14:

We sincerely appreciate the excellent suggestion from you for helping us consider how to felicitously discuss certain theories in evolutionary biology. We made this simple assumption about the order of mutation fixation without carefully considering the complexity of natural evolution. We would like to thank the reviewers again for providing us with such important feedback, and we would also like to thank the reviewers again for selflessly guiding us to discuss relevant evolutionary hypotheses in the context of rigorous evolutionary biology knowledge. **Based on an insightful**

suggestion from the reviewer, we have determined to delete this assumption from the manuscript to avoid misuse of evolutionary concepts. And we have revised relevant content in Result section of the revised manuscript (lines 188-220).

According to Yi and Dean's perspective (Yi X. et al. *Mol. Biol. Evol.* **2019**, 36(5), 890-907), this adaptive landscape we constructed here belongs to static landscapes. In a static landscape, these dynamic factors, such as selection, are not considered. These empirical landscapes, constructed with all possible combinations of small sets of mutations, play a crucial role in systemically investigating the relationship between the sequence and functions of proteins. The important effect epistasis exerts on a range of protein functions is easily uncovered via the landscape (Miton CM et al. *Curr. Opin. Struc. Biol.*, **2021**, 69, 160-168). Moreover, empirical landscapes are also used to discuss the predictability of evolution in a review (de Visser JAGM et al., *Nat. Rev. Genet.*, **2014**, 15(7), 480-490). Therefore, we have two main aims for constructing this adaptive landscape in this study. Numerous studies use enzyme activity as a proxy to investigate the role of epistasis in the evolution of function. In an attempt to demonstrate the equivalence of other enzymatic properties, such as stereoselectivity, we determined to construct the adaptive landscape and investigate the role of epistasis in the evolution of stereoselectivity. Meanwhile, the first author has always believed that the recognition and understanding of epistasis are significantly important for direct evolution. For traditional protein engineers, knowing the importance of epistasis in the direct evolution campaign is the first step, and then considering how epistasis helps to realize as rational protein engineering as possible is the next level. Another goal is that we hope our landscape can contribute to relevant theoretical research (Buda K. et al. *Nat. Commun.* **2023**, 14, 8508) and even provide meaningful data for deep learning, assisting in the engineering of stereoselectivity.

15. Lines 202-209: For the general audience of this paper, higher order epistasis should be defined. i.e. what is meant by e.g. sixth order epistasis.

Response #1-15:

Many thanks to you for highlighting this aspect for helping to improve the quality of our manuscript. We have added a definition of high-epistasis to line 217 of the revised manuscript.

The changed sentence:

Based on these results, 32% of variation can be explained without epistasis, 24% by second-order epistasis, and 44% by high-order epistasis (interactions between three or more mutations).

16. Supp Table 2: The specific activity differed markedly between some ancestors and their alt versions as well as varying ~ an order of magnitude across the ancestors resurrected from the evolutionary tree. In particular the two newest ancestors were approximately an order of magnitude less active than the other ancestors. This point is completely ignored in the manuscript.

Response #1-16:

We greatly appreciate you highlighting this aspect. We have provided response in Response #1-3. Given the difference in the specific activity does not affect the pattern of stereoselectivity evolution in ancestors and their alternative ancestors, we didn't highlight the difference in the manuscript. Moreover, several empirical studies have also found a difference in enzyme activity between ancestors and alternative ancestors. (Chiang, C-H. et al. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, 120(15), e2218248120, Franceus J. et al. *ACS catal.* 2024, 14, 3103-3114). We hope these answers are helpful for you. Many thanks for your contribution to our manuscript.

17. Supp Table 4: Please show the specific activity as well as the change in ee. It would be useful to quantify protein expression as well to determine whether there is a decrease in kcat.

Response #1-17:

We greatly appreciate the excellent suggestion from you. We have measured the specific activity of each mutant, and the corresponding results have been added to Supplementary Table 5 of the revised Supplementary Information.

18. Lines 243-244: the authors state “Therefore, high-order epistasis should be incorporated into protein design and protein engineering technologies in the near future”. How would the authors propose to do this?

Response #1-18:

Many thanks for pointing out this aspect. First, the crucial role of epistasis in protein engineering campaigns has received extensive attention in recent years (Marcel W. et al. *ACS Catal.* **2022**, 12, 14243-14263; Miton C.M. *Protein Sci.* **2016**, 25(7), 1260-1272; Cadet X.F. et al. *Method in Molecular Biology*, **2022**, 2461, pp: 225-275). Moreover, a plethora of predictive models incorporating epistasis have been successfully applied in a range of protein engineering campaigns. The EVmutation would predict the effect of a mutation with specific respect to two-body epistatic interactions. Hence, it would obtain better results compared to other strictly additive tools (Hopf T. A. et al. *Nat. Biotechnol.* **2017**, 35(2), 128-135). Zaugg et al. developed computer models that could capture low-order epistatic interactions and successfully realized the improvement of the enantioselectivity of an epoxide hydrolase using this model (Zaugg J. et al. *J.comput. Aided Mol. Des.* **2017**, 31(12), 1085-1096). In addition, these computer models provide better prediction results in protein engineering campaigns when they take into account high-order epistasis. The Innov'SAR approach, which is able to capture high-order epistatic interaction, would identify better-performing variants than another simple

model without considering epistasis (Cadet F. et al. *Sci. Rep.* **2018**, 8(1), 16757; Li G. et al. *Chembiochem*, **2021**, 22(5), 904-914; Ostafe R. et al. *Biotechnol. Bioeng.* **2020**, 117(1), 17-29). Recently, more computer approaches incorporating high-order epistasis have been developed, such as ECNet, which has the ability to model the high-order epistasis of multiple residues (Luo Y. et al. *Nat. Commun.* **2021**, 12(1), 5743). With more and more research on the effect of epistasis on protein function and the development of machine learning, we believe that more computational tools will be available to help protein engineering in the near future. Herein, we are supposed to provide more experimental data investigating the relationship between epistasis and protein functions to assist in the development of computer models. The primary purpose of establishing the empirical fitness landscape in this publication is to reveal the significant impact of epistasis on stereoselectivity.

19. Lines 246- 248: The authors state “Furthermore, we demonstrate the D170Y is an evolvability-enhancing mutation—the ability to increase likelihood that subsequent substitutions are adaptive in evolving protein molecules—in this empirical evolutionary landscape.” It is not clear what the authors mean here. The rest of the paragraph seems to justify this statement by showing D170Y was essential to completely invert enantioselectivity, pointing out that the L19M/A119G double mutant (131% ee) is the only one that shows R selectivity but no mutation at residue 170. All other double mutants showing R-stereoselectivity contained D170Y. A reasonable question therefore is how significant is the inversion in this (L19M/A119G) mutant, i.e. what is the experimental error in these measurements. An alternative interpretation here is that the D170Y is simply essential to revert the stereoselectivity and the single measurement for L19M/A119G is an outlier!

Response #1-19:

We really appreciate your efforts in reviewing our manuscript and your precious comments. For the L19M/A119G mutant, we re-measured its stereoselectivity in three

parallel experiments, which reproduced the previous results. In the revised manuscript, we have removed the discussion on the significance of the D170Y mutation. After we thought it over, we realized that the supposed importance of the D170Y mutation depended on other mutations. The fact is that stereoselectivity inversion is not a major role of a single site, but a result of multiple sites working together. In the present study, we cannot emphasize the importance of individual mutations alone. By reanalyzing the results, we suggest that stereoselectivity inversion mainly results from epistatic interactions between multiple mutations. The results also imply that a single site mutation may not be successful for inverting stereoselectivity of IREDs in protein engineering campaign. Once again, we thank you for dedicating time to assess our manuscript and providing valuable feedback on our work.

20. Lines 274-276: The authors comment “We then deleted large gaps in the MSA results to decrease subsequent phylogeny reconstruction noise. Locations where >50% of sequences were gaps in MSA were considered large gaps.” This needs clarification and further justification. (How large? Some estimates of length and the threshold used are needed.) The authors appear to mean that where some sequences had large insertions but the majority of sequences had no sequence, the inserted sequences were edited out of the sequences in which they were found. Such insertions may arise from either real evolutionary events or as errors in virtual translation of sequences. If the former, this is removing a real factor in an enzyme’s evolution. If the latter, it suggests the sequence may not be reliable (e.g. miscalled when the genomic DNA was ‘translated’). Therefore care needs to be taken either way and a judgement made as to whether the sequence should be included in its entirety (if likely to be real, e.g. if found consistently in one lineage of a tree) or not (if likely to be erroneous). The authors should comment on the likely source of these insertions and justify why they edited the sequences – a subjective intervention - rather than simply removing or including them as is routine practice in the field. See ref 41 methods section for a comment on how these issues are usually handled.

Response #1-20:

We sincerely thank your valuable comments. Several previous studies used the Gblock tool to delete gaps in multiple sequence alignment. In these studies, if the columns with more than 50% gap, researchers will delete the gap (Hager M. et al. *Mol. Biol. Evol.* **2022**, 39: msac250, Ding Q. et al. *Nat. Commun.* **2024**, 15:2492). We used the same protocol to perform gap delete, if a column possesses >704 gaps (more than >50%). Thank you again for your comments, and we hope the answers are helpful for your review.

21. Line 288: The authors use a marginal reconstruction but for comparisons across an evolutionary tree, joint reconstruction is more appropriate. GRASP can do both. The authors should justify why marginal was chosen instead of joint.

Response #1-21:

We greatly appreciate you highlighting this aspect and totally agree with your suggestions. According to the definition of marginal and joint reconstruction, joint reconstruction seems more appropriate. However, these GRASP developers pointed out that the ancestral sequence resulting from marginal reconstruction is almost identical to that from joint reconstruction on the Guide and Tutorial page. To verify the identity of reconstructed sequences resulting from marginal and joint reconstruction, we performed sequence alignment using ancestral sequences from the two reconstruction strategies (Figure R7). We found that the identity of the global sequence has no significant difference; importantly, the six key historical substitutions are totally the same. Additionally, a range of ASR research would use the second probable ancestor (alternative ancestor) to support the robustness of reconstruction (Dishman A. F. et al. *Science*, **2021**, 374(6524), 86-90; Natarajan C. et al. *Curr. Biol.*, **2023**, 33, 98–108; Nicoll C. R. et al. *Nat. Struct. Mol. Biol.* **2020**, 27(1), 14-24). But the joint reconstruction doesn't generate the alternative ancestors. Recently, the marginal reconstruction was also applied in several studies investigating the evolution of

different enzyme functions in an evolution trajectory (Franceus J. et al. *ACS catal.* **2024**, 14, 3103-3114; Kostelac. Et al. *Appl. Environ. Microbiol.* **2024**, 90(1), e01676-23).

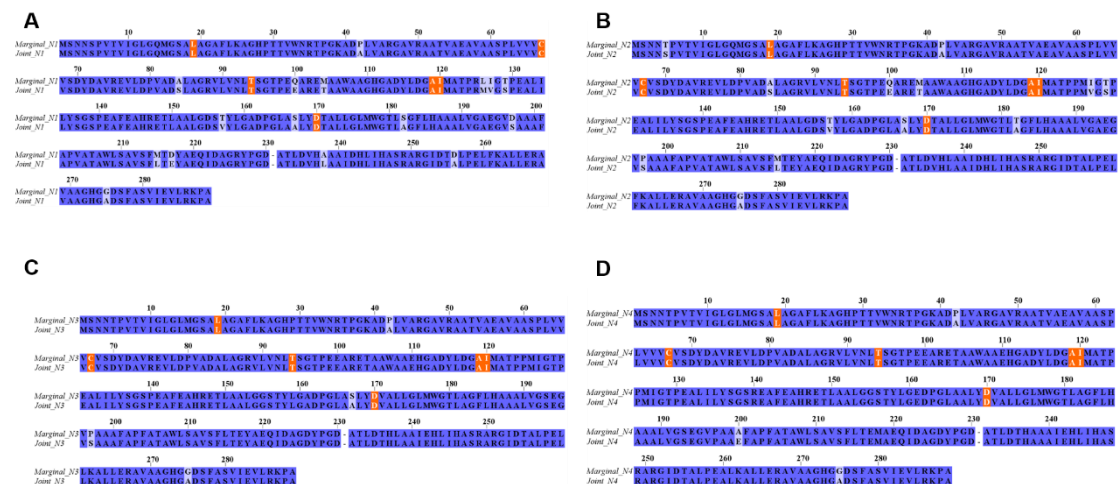


Figure R7. Sequence comparison between marginal and joint reconstruction. A N1 and AltN1. **B** N2 and AltN2. **C** N3 and AltN3. **D** N4 and AltN4. Dark blue shade denotes identical sites, light blue shade represents the different sites between marginal and joint reconstruction, and orange shade denotes the six crucial historical substitutions in our manuscript.

22. Lines 289-290: The authors say they constructed the ‘second most likely ancestors for each ancestral sequence (alternative ancestors) to verify the robustness of ASR’. The only data shown for these alternative ancestors is in Supp Fig 4 (where ee values are compared) and Supp Table 2 where the specific activity (U/mg) is provided. (There is no citation of this latter table in the text however.) The specific activity obtained for the alt ancestors is markedly different from the values obtained for the most probable ancestors, suggesting that in this case the results obtained from the ASR are not robust as concerns specific activity, even if they were for ee (Supp Fig 4 vs Fig 1D)!

Response #1-22:

Many thanks to you for highlighting this aspect. We have cited Supplementary Table 2 at line 119 in the revised manuscript. We have also listed relevant responses regarding the differences between specific activities in Response #1-3. Additionally, we inferred

that these mutations between ancestor and alternative ancestor may influence enzyme conformations, so that a significant difference in activity was observed. Moreover, the difference of activity is also observed in several ASR studies (Franceus J. et al. *ACS catal.* **2024**, 14, 3103-3114; Chiang, C-H. et al. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, 120(15), e2218248120).

23. Lines 289-290: Can the authors comment on how different the second most probable ancestor sequences are from the respective most probable ones? What explains the large differences in specific activity (approximately an order of magnitude in many cases)?

Response #1-23:

We thank you for raising these pertinent concerns. We have listed the sequence difference between the most probable and second most probable ancestor (Table R3) in the Supplementary Information. As for the second question, we would like to use the responses in Response #1-3 and #1-22.

Table R3. Ancestral sequence reconstruction statistics.

Ancestral enzyme	Mean posterior probability ^a	Uncertain sites ^b	Sequence differences between alternative ^c
N1	0.89	70 (24%)	45 (15%)
N2	0.93	40 (13%)	31 (10%)
N3	0.95	30 (10%)	23 (8%)
N4	0.95	29 (10%)	19 (6%)
N5	0.95	26 (9%)	15 (5%)
N557	0.94	35 (12%)	23 (8%)
N558	0.94	39 (13%)	26 (9%)
N559	0.92	46 (16%)	34 (11%)
N560	0.94	36 (12%)	30 (10%)

^aMean posterior probability of the maximum-likelihood version of the ancestral sequence using the WAG mode.

^bNumber (percentage) of sites at which the maximum-likelihood reconstruction sequence has posterior probability < 80%.

^cNumber (percentage) of sites at which the alternative sequences of the ancestral enzyme differ in residues.

24. Supp Info Pages 2-18: in the Supp info the sequences of ancestors characterised are presented as both nucleotide and amino acid sequences. However all appear to have been subcloned into a pET vector which added a hexa-His tag. Therefore it is slightly misleading to not include this tag in the OPRF of the resurrected enzymes.

Response #1-24:

We greatly appreciate your valuable suggestions. We have added hexa-His tags to all sequences in the revised Supplementary Information.

25. P.10, bottom: The statement “For crystallization of N559 and N559-M6, truncated proteins were constructed by enzyme cut and link-up experiments” needs explanation. Do the authors just mean cut with restriction enzymes then ligated? Please provide details of where the proteins were truncated to.

Response #1-25:

We greatly appreciate the insightful query raised by you. There are several amino acids located between the hexa-His tag and the target enzyme. These amino acids often form an unordered loop at the N-terminal of the target protein. Therefore, we used restriction enzymes (*Xho* I and *Nco* I) to remove the amino acids and improve crystallization success rates (Figure R8). The detailed method has been added to lines 316 to 320 in the revised manuscript. Moreover, we found that the truncated enzyme possesses the same stereoselectivity as a normal enzyme.



Figure R8. Plasmid details. **a** N559 sequence with several extra amino acids of vector. **b** N559-cut sequence without the amino acids of vector. Red box denotes the truncated amino acids. Brown band denotes the residues of vector. Yellow band denotes the amino acid sequence of N559.

26. p. 12: The assay for IRED activity relies on monitoring the loss of absorbance from NADPH at 340 nm. From this the number of μmol converted to product was extrapolated. However NADPH oxidation is not particularly specific to the reaction of

interest. What quality controls were undertaken to ensure NADPH oxidation faithfully reflected product formation, i.e. imine reduction?

Response #1-26:

Thank you for your thorough review of our manuscript and for providing valuable feedback. Theoretically, consuming one molecule of NADPH produces one molecule of product. Indeed, NADPH is partially oxidized by oxygen in the air, but this is usually slower and in smaller quantities. The change in absorption is approximately 0.002 in a reaction system without substrates, and background values were subtracted to calculate specific activity. In the revised supplementary information, we also used the gas chromatograph to detect products for measuring specific activity and replaced the previous data in Supplementary Table 2.

Minor points:

27. Line 299: The correct name of “LB medium” is “Lysogeny Broth”.

Response #1-27:

We sincerely appreciate your valuable suggestion. We have revised it at line 313 of the revised manuscript.

28. Lines 306, 310, 313, 318, etc: x g values should be used rather than rpm values to describe centrifugation steps.

Response #1-28:

We sincerely appreciate your valuable suggestion. We have replaced rpm with x g on lines 327, 330, 335 of the revised manuscript.

29. Lines 305-306: “Terrific Broth medium” is a tautology: the name of the medium (broth) is just “Terrific Broth.

Response #1-29:

We sincerely appreciate your valuable suggestion. We have revised it on lines 322-323 of the revised manuscript.

30. Supp Fig 2: the top left SDS-PAGE gel is obscured by water droplets and bubbles. Please take another image of this gel for a publication quality figure. This figure is not referred to in the text.

Response #1-30:

We sincerely appreciate your valuable feedback. We have carried out the new sodium dodecyl sulfate-polyacrylamide gel electrophoresis and replaced the obscured SDS-PAGE gel in the revised Supplementary Information (Figure R9 and Supplementary Fig. 8). The corresponding citation has added on lines 119–120 of the revised manuscript.

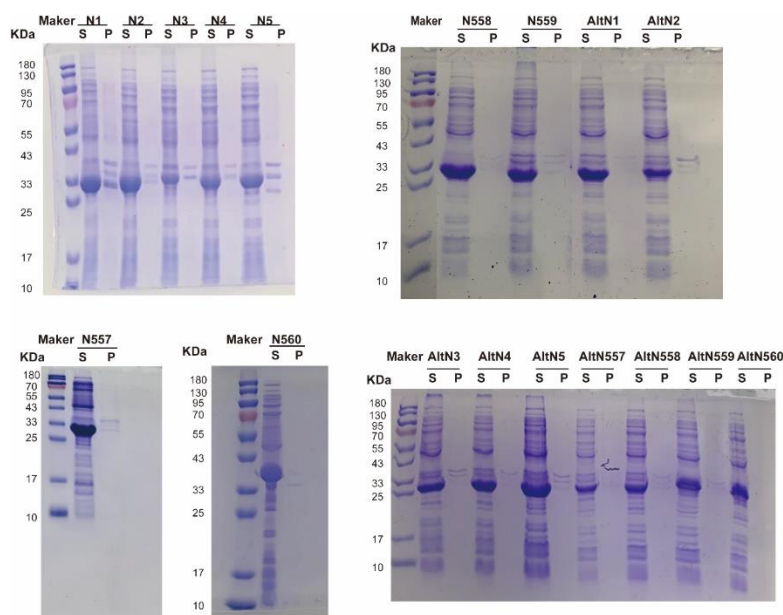


Figure R9

SDS-PAGE analyses of ancestral IREDs (N1-N560), and their alternative protein (AltN1-AltN560). The molecular weight of the proteins is between 29 to 35 kDa.

Response to Reviewer #2

Reviewer #2 (Remarks to the Author):

Zhu et al. report an ancestral sequence reconstruction study into the evolution of stereoselectivity in imine reductases. They claim to have identified a stereoselectivity inversion in the evolution of these enzymes, as well as the historical mutations responsible for the inversion.

We thank you for dedicating time to assess our manuscript and providing valuable comments. We hope these answers will help you further assess the credibility of our study.

Comments:

1. One potential problem that I find with this work is that the authors appear to have relied heavily on computational tools and resources available from web servers. This is not necessarily wrong, of course. However, many of these tools involve assumptions that must be critically examined. For instance, the authors use the tool GRASP very recently developed by Gillam and coworkers (ref. 58 in the submitted manuscript). It is to be noted that GRASP is based on maximum likelihood, while most researchers in the ancestral reconstruction field consider Bayesian approaches to be more reliable. Certainly, maximum likelihood allows for the analysis of a larger number of sequences, which was one of the main reasons for developing GRASP. However, Gillam and coworkers have also emphasized the need to extensively curate the alignment of sequences used as starting point for ancestral reconstruction (Journal of Biological Chemistry 298(10) 102435, 2022), in order to eliminate incorrect sequences harboring different types of errors. Sequence curation is a process that is done manually to a substantial extent. The point I am trying to make is that the authors use an alignment of 1407 sequences as input for GRASP, but do not make it clear in the submitted manuscript how curation was carried out, except for a brief and obscure reference to the deletion of “large gaps in the MSA results to decrease subsequent phylogeny reconstruction noise” (top of page 10).

Response #2-1:

Many thanks to you for highlighting this method used to reconstruct ancestral enzymes. We totally agree with your comments. We want to present two main explanations for the ASR approach used in our manuscript. Maximum likelihood and Bayesian reconstruction are the most reliable methods for reconstructing ancestral sequences. Bayesian analysis is typically limited to a small number of sequences and requires more computational resources. We chose the maximum likelihood approach to reconstruct ancestral sequences for more than 1400 sequences. Furthermore, researchers demonstrated that Bayesian has not substantially improved accuracy compared to maximum likelihood in a study using a simulated data set to reconstruct ancestral sequences (Hanson-Smith V. et al. *Mol. Biol. Evol.* **2010**, 27, 1988-1999). In addition, Randall et al. randomly mutated a red fluorescent protein sequence to generate 19 descendant sequences. They then used these ‘modern’ sequences to infer ancestral sequences via several software based on the maximum likelihood approach. The results showed that these inferred ancestral proteins possess the same functions as the actual ancestral protein (Randall R. N. et al. *Nat. Commun.* **2016**, 7: 12847). Additionally, the maximum-likelihood approach is widely used in a range of research on ancestral enzymes (Kaltenbach M. et al. *Nat. Chem. Biol.* **2018**, 14(6): 548-555, Chiang C-H. et al. *Proc. Natl. Acad. Sci. U.S.A.*, **2023**, 120(15): e2218248120, Jones B. S. et al. *Angewandte Chemie*, **2024**, e202314869).

We greatly appreciate your insightful comments that contribute to assessing the reliability of our study. For these modern sequences, we gathered them from a public IREDs database and performed a sequence curation operation involving excluding a sequence with an X character in it and deleting these gaps within multiple sequence alignment. In view of the criteria used to build the database, we didn't conduct extra sequence curation. In this database, the sequences with a global sequence identity of >98% are assigned to a single protein entry; in this case, the longest sequence was taken as the reference sequence for representing the sequences with >98% identity. Furthermore, all sequences in this database are obtained from NCBI based on two seed

sequences, which have been biochemically characterized in previous studies (Fademrecht S. et al. *Protein*. **2016**, 84:600-610). These sequences have been curated for setting up this database. According to their standard for building the database, we argued that these sequences are suitable for performing ancestral sequence reconstruction.

2. In addition, the authors do not clearly discuss the reliability of the phylogenetic tree used as basis for reconstruction. Plausible evolutionary interpretations of the biophysical properties of “resurrected” proteins require that the phylogenetic tree used resembles to some substantial extent the tree of life or, at least, the accepted phylogeny of the protein system under study (in case the evolutionary history of the protein family involves gene duplication events). I assume, although this does not appear to be clearly stated in the manuscript, that among modern imine reductases there are examples of enzymes with S- and R- stereoselectivity. Furthermore, I find it plausible that imine reductases with S- and R- stereoselectivity exist in the same organism, reflecting gene duplication events. Do the phylogenetic nodes targeted for reconstruction correspond to speciation events or to gene duplications events? Is there a correspondence between the phylogenetic tree used (Supplementary figure 1) and an accepted phylogeny for imine reductases?

Response #2-2:

Many thanks for the excellent suggestions from you. As Harms and Thornton pointed out in a Review, the aim of evolutionary biochemistry is to solve these questions that cannot be addressed by biochemists and evolution biologists alone (Harms M. J. & Thornton J. W. *Nat. Rev. Genet.* **2013**, 14(8), 559-571). Furthermore, ASR has proven to be a powerful means for identifying important mutations that change protein functions. ASR has illuminated a range of mechanisms behind protein functions that were not addressed in traditional biochemistry studies (Hochberg G. K. A. et al. *Annu.Rev. Biophys.* **2017**, 46, 247-269; Dishman A. F. et al. *Science*, **2021**, 371, 86-90).

Our main aim is to investigate how stereocomplementarity is regulated with the help of evolutionary biochemistry. Hence, we hope to provide some biochemical experimental data to support the reliability of our ancestral sequence reconstruction. There is a universally accepted hypothesis that ancestral proteins possess higher thermostability than extant proteins. Moreover, this hypothesis has been verified in a range of ancestral proteins resurrected by ASR. Therefore, we would like to share some of our unpublished data in another study to support the robustness of inferred ancestral enzymes (Figure R10A). We further established an ancestral imine reductase panel based on this work for performing relevant biocatalysis studies. Our results showed that the T_m (measuring the thermostability) of reconstructed ancestral imine reductases was significantly higher than that of extant enzymes (Figure R10B, $p < 0.05$). Based on this evidence, we argue that our reconstruction is reliable.

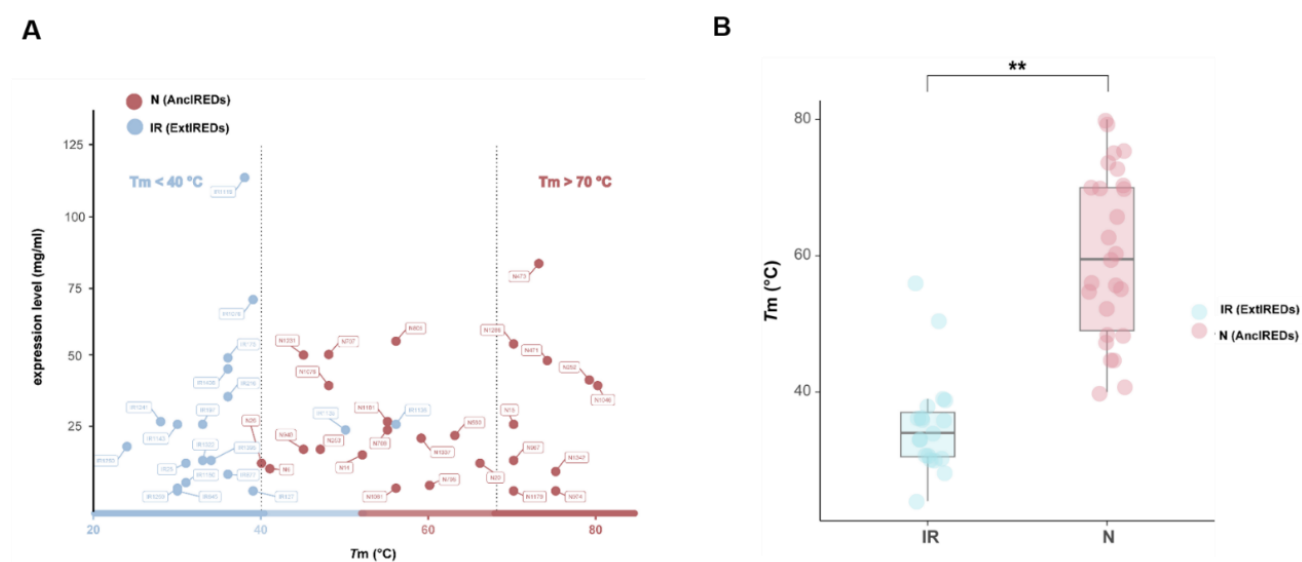


Figure R10. Comparison of thermostability between ancestral and extant imine reductases. **A** The distribution of specific T_m value of extant and ancestral imine reductases. Blue and pink dots denote extant and ancestral imine reductases, respectively. **B** The statistical analysis of T_m with t-test. ** denotes $p < 0.05$.

To support the reliability of this study, we have also used IQtree to reconstruct a phylogenetic tree. We found that the topological structure of the phylogenetic tree resulting from PhyML 3.0 and IQtree is almost identical (Figure R11). Therefore, we argued that our reconstruction is reliable. In addition, the main aim of this evolutionary

biochemistry research is to investigate the regulation mechanism of stereocomplementarity by virtue of a vertical research strategy (with the help of appropriate evolutionary knowledge), and the native physiological functions of IREDs are unknown. Thus, we are unable to discuss more information on the effect of the evolution of stereoselectivity in imine reductases on organisms' evolution. We would like to thank you again for taking the time to review our manuscript. Meanwhile, we also hope these answers will help you understand our work.

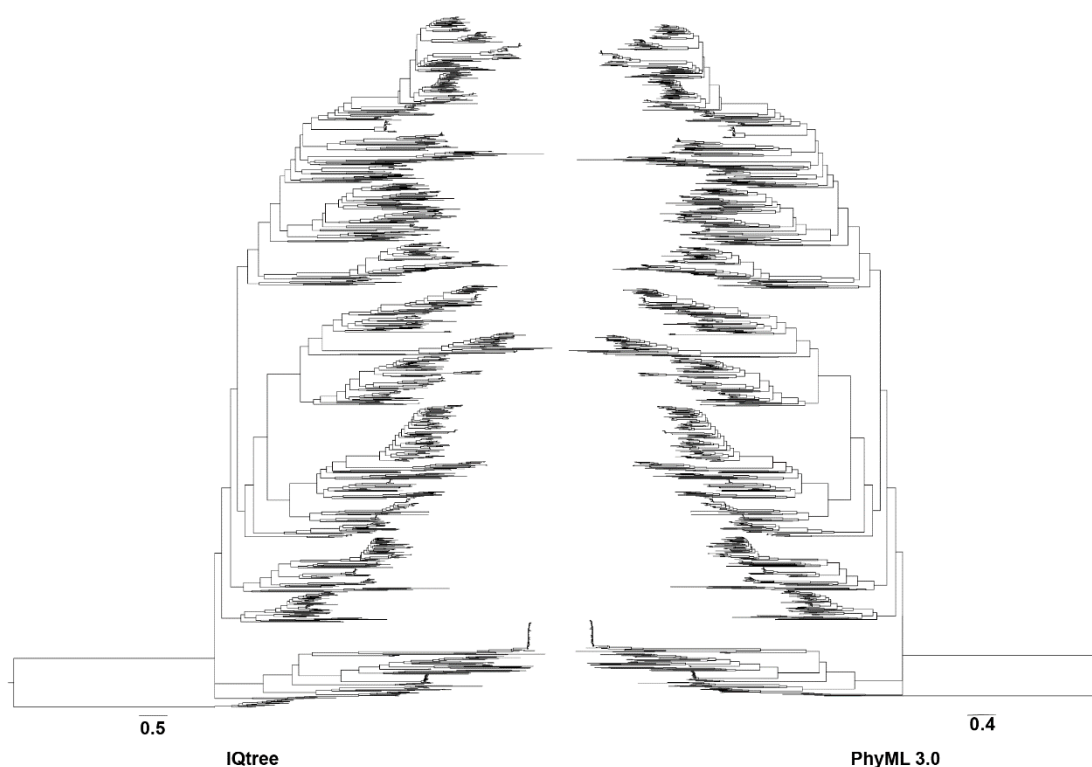


Figure R11. Topology comparison of phylogenetic tree resulting from IQtree and PhyML 3.0.

3. The authors claim that a stereoselectivity inversion occurred during the evolution of imine reductases. As judged from the data in Figures 1 and Supplementary 4, it appears that the inversion was sharp instead of gradual. The manuscript does not seem to provide a convincing evolutionary narrative for the sudden (in evolutionary time scales) inversion. What kind of change in intra-cellular or extra-cellular environment triggered the stereoselectivity inversion? Was there a change in metabolism that required imine reductases to change stereoselectivity? Do the stereoselectivities of modern imine

reductases shed some light on this hypothetical ancestral change?

Response #2-3:

We would appreciate your important inquiry. It is seemingly true that the formation of *R*-stereoselectivity is a sudden event, according to our description of the experiment data. In fact, *R*-stereoselectivity is a gradual event. As shown in Fig. 1d, the *R*-enantiomer has already appeared in N558 and N559. But according to the definition of stereoselectivity, the enzyme is called *R*-/*S*- stereoselectivity only when the proportion of *R*-/*S*- enantiomers in products is greater than the proportion of another enantiomer. In other words, the N560 shows *R*-stereoselectivity because the proportion of *R*-enantiomers first exceeds that of *S*-enantiomers, which does not represent that the *R*-enantiomers first appear in the N560. Therefore, the evolution of stereoselectivity is gradual. More importantly, the phenomenon of the evolution of stereoselectivity in our study is in accordance with the first study investigating the evolution of stereoselectivity with ASR (Chiang C-H. et al. *Proc. Natl. Acad. Sci. U.S.A.*, **2023**, 120(15): e2218248120).

While we hope our work will make a contribution to our understanding of how stereoselectivity is related to species evolution, the effect of stereoselectivity on organism fitness is not yet clear. This gap limits further discussions. The main purpose of this study is to understand the biophysical mechanism of stereoselectivity by borrowing the idea of evolutionary biochemistry. This is a very rare research area in evolutionary biochemistry, and only one related study has been reported so far (Chiang C-H. et al. *Proc. Natl. Acad. Sci. U.S.A.*, **2023**, 120(15): e2218248120). Moreover, the investigation of the mechanism underlying stereoselectivity is not easy in traditional biochemical research. Therefore, we hope to apply multiple methods to study stereoselectivity. Importantly, we also argue that the environment changes may play an important role in stereoselectivity inversion. To answer this question, a suitable research system will be needed in the future. Unfortunately, these interesting discussions are beyond the capabilities of our research system. We hope our answers are clear. Thank

you for your important suggestions again.

4. The authors use ibis2analyzer to perform the co-evolution analysis that leads to their identification of the historical mutations responsible for the stereoselectivity inversion. This also poses several issues. Co-evolution analysis is not straightforward and different approaches may lead to different results, in particular when using a limited number of sequences in the analysis (which seems to be the case here because, if I understand correctly, about 10 ancestral sequences were used). It is possible then, that other mutations, in addition to the 6 identified by the ibis2analyzer (page 5), also affect stereoselectivity. The authors do not describe which mutations are “rejected” by ibis2analyzer as potential causes of stereoselectivity inversion and how reliable such rejection is.

Response #2-4:

We greatly appreciate your invaluable comments. In our manuscript, we efficiently identified these historical substitutions contributing to stereoselectivity inversion with the help of the co-evolution analysis. We totally agree with your comments. However, validating all 45 mutations in N560 compared to N559, including single mutations, two-site mutations, and multiple-site random combination mutations, is experimentally impossible (2^{45} mutants). Therefore, we need to seek some strategies for assisting in the determination of the key mutations. A biochemical experiment has already revealed the significance of the D170Y mutation. But the single mutation D170Y doesn't invert stereoselectivity, so we speculate that extra mutations are necessary for stereoselectivity inversion. Furthermore, the stereoselectivity of all sequences in our study has been experimentally determined. On the basis of these prior knowledge bases, we surmised that the sites coevolving with D170Y might invert stereoselectivity. Hence, we performed the co-evolution analysis with the ibis2analyzer tool. It is true that ibis2analyzer gives multiple coevolution groups, but our goal is to identify the coevolution group containing D170Y. That is, finding a coevolution group containing

the D170Y mutation is our null hypothesis. If the co-evolution analysis gives such a result, we will verify it through a biochemical experiment. If it fails to give such a result, we will use a new approach to identify the key historical substitutions. We hope these responses will help you evaluate the reliability of our co-evolution analysis. Furthermore, many thanks for your insightful comments.

5. Based on the 6 mutations selected, the authors derive an adaptive landscape of stereoselective evolution. This obviously involves the determination of stereoselectivity for 64 combinations of mutations. Then the authors use the MAGELLAN to ascertain the impact of different types of epistasis on the determined landscape. MAGELLAN is, however, used essentially as a black box and no serious attempt to provide intuitive, convincing explanations for the results obtained is reported.

Response #2-5

We greatly appreciate your invaluable suggestions. We have performed new statistical analyses with two methods, involving a Python package and the MAGELLAN tool. The corresponding results have been added to lines 209-214 in the revised manuscript. Additionally, the statistical models behind MAGELLAN have relevant publication (Ferretti L. et al. *J. Theor. Biol.* **2016**, 396, 132-143). Moreover, the MAGELLAN tool was cited in a paper (Bank C. et al. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, 113, 14085-14090). We hope you find these answers helpful. Thank you for your comments.

6. Overall, I find that many of the analyses, discussions and evolutionary interpretations (or lack of) provided in the manuscript are not convincing.

Response #2-6:

We feel great thanks for your professional review work on our manuscript. We have revised the manuscript according to each reviewer's valuable comments and properly modified the discussion. We hope these modifications are useful for providing

convincing conclusions to this study.

Furthermore, we would like to provide some reasons for supporting this manuscript. First, the regulation mechanism behind stereoselectivity of enzymes has rarely been investigated from an evolutionary biochemistry perspective, besides the only study (Chiang C-H. et al. *Proc. Natl. Acad. Sci. U.S.A.*, **2023**, 120(15): e2218248120). Our and Chiang's study shows that enzyme stereoselectivity, an enzymatic property often understood from a chemical perspective, can also be studied from an evolutionary perspective. On this basis, we hope to establish a research system in the future in which enzyme stereoselectivity correlates with organisms' fitness. In this research system, the importance of enzyme stereoselectivity for the origin of life is explored by means of ancestral sequence reconstruction. Then, we argue that the stereoselectivity of enzymes is also a suitable agent for investigating the effect of epistasis on the evolution of protein functions. This relevant study is important. Because basic research often would derive many important and general knowledges, which has huge potential for helping to understand the regulation mechanism behind stereoselectivity and instruct protein engineering of stereoselectivity.

Moreover, our and Chiang's study demonstrates that there are two advantages. 1) It is relatively easy to find key mutations that change stereoselectivity by looking at the evolution path from *R*-/*S*- stereoselectivity to the opposite one. 2) These identified historical substitutions often exert the expected effect on stereoselectivity inversion in ancestral sequence backgrounds. However, this same experiment strategy usually is useless in extant sequences background. For example, researchers performed site-saturation mutagenesis at these positions, which were reported to be essential for *S*-stereoselectivity, but failed to reverse the *R*-stereoselectivity of an extant IRED (Stockinger P. et al. *ChemCatChem*, **2021**, 13, 1-7). That is, the assumed key sites obtained from horizontal multiple sequence alignment have no expected effect on stereoselectivity inversion. Additionally, labor-intensive saturation mutagenesis is necessary in traditional biochemical research. Interestingly, our and Chiang's work showed that these shortcomings could be avoided to a certain extent. Therefore, we

argue that our results are valuable in the specific areas of research.

In addition, several studies used a range of published empirical fitness landscapes to derive significant evolutionary conclusions (Wagner A. *Nat. Commun.* **2023**, 14, 3624; Buda K. et al. *Nat. Commun.* **2023**, 14, 8508). For that reason, we suggest that our empirical fitness landscape could also provide important experimental data for that research as well. Once again, we thank you for your careful review of this study and hope that our responses will help you assess the credibility of our research.

7. Certainly, the identification of a small number of mutations that can be combined to switch stereoselectivity is interesting. However, it is not clear from the manuscript whether the same of equivalent information could have derived from the comparative analysis of the stereoselectivity of modern imine reductases.

Response #2-7:

We greatly appreciate your valuable feedback. We have conducted relevant experiments to answer this comment. The results showed that the six coevolution substitutions can't invert stereoselectivity on extant sequences but have the ability to significantly decrease the *ee* of wild types (Table R4). Moreover, 4 out of 7 extant enzymes will be inactive. These results indicate that the six historical substitutions can affect IREDs' stereoselectivity regulation. Meanwhile, it suggests that these extant sequences acquire other mutations that produce epistatic interactions with the six mutations. Thus, the six mutations alone can't invert stereoselectivity.

Table R4 The effect of six historical substitutions on extant IREDs.

Enzymes	<i>ee</i> (%)
<i>SIR</i> ^a	-99
<i>SIR</i> -M6	-45
<i>SvIR</i> ^a	99
<i>SvIR</i> -M6	n.a.
<i>StIR</i> ^a	99
<i>StIR</i> -M6	n.a.
<i>RIR</i> ^a	99
<i>RIR</i> -M6	n.a.
<i>AaIR</i> ^a	99
<i>AaIR</i> -M6	48
<i>AdIR</i> ^a	99
<i>AdIR</i> -M6	n.a.
<i>SeIR</i> ^a	99
<i>SeIR</i> -M6	26

Note: ^aThe *ee* from our previous study (Zhang Y. H. *Org. Lett.* **2020**, 22, 3367-3372). Positive and negative number denote *S*- and *R*- stereoselectivity, respectively. The n.a. denotes that the enzymes are inactive.

Response to Reviewer #3

Reviewer #3 (Remarks to the Author):

Zhu et al. present intriguing results aimed at providing insight into the evolution of stereoselectivity from ancestral imine reductases to modern proteins. They have resurrected ancient imine reductases using a single phylogenetic tree of extant descendants and determined the stereoselectivity of these ancient proteins through biochemical analyses. The results revealed six mutations responsible for changing stereoselectivity from S- to R-stereoselectivity. X-ray crystallography elucidated the structural mechanism for regulating stereoselectivity inversion. The authors then investigate the adaptive landscape of stereoselectivity evolution. From these findings, the authors assert that ancestral sequence reconstruction offers an efficient method for identifying mutations that alter protein functions.

The paper addresses an interesting question, and the approach used is appropriate. Overall, I believe the authors have done an excellent job of identifying the amino acids involved in the acquisition and transition of protein stereoselectivity. In particular, by expanding their research beyond ancestral sequence reconstruction and biochemical analysis of resurrected proteins to include crystal structure analysis and binding modeling, the authors' claims regarding the mechanism of stereoselectivity acquisition are strongly supported. However, I do have some concerns that should be addressed before my enthusiasm can be fully endorsed.

The results presented are exciting and provocative in the realm of protein chemistry and enzyme chemistry research.

We thank you for dedicating time to assessing our manuscript and providing positive feedback on our work. Thank you very much again.

Comments:

1. However, if discussing evolution, it is necessary to verify the reliability of the phylogenetic tree and the confidence in the estimated ancestral sequences. Especially, the branch connecting the target protein family and the outgroup is very long, and it

cannot be said that the root position is sufficiently specified with confidence. Regarding the root position, careful interpretation is necessary.

Response #3-1:

We greatly appreciate your excellent suggestion on providing confidence in the estimated ancestral sequences. To support the reliability of the ancestral sequences we reconstructed, we listed these relevant statistical parameters in Table R5. Overall, our reconstructed sequences have the lowest posterior probability > 89%. Therefore, we argued that the reconstructed ancestral sequences are capable of supporting the experiment results and conclusions in the present manuscript.

Inferring the last common ancestor (LCA) of all extant imine reductase (IRED) sequences is a significantly key part of this study, which acts as a starting point in this study. We first need to investigate whether the LCA possesses stereoselectivity toward the target reaction. If a positive result was obtained, we then determined what the specific stereoselectivity was. Then, a suitable evolution trajectory needs to be selected for investigating the divergence of stereoselectivity. Once this evolution pathway is identified, we just need to investigate at which evolutionary node stereoselectivity inversion occurs. We then combine biochemistry, bioinformatics, and X-rays to interrogate the mechanism of stereoselectivity inversion. Rooting a phylogenetic tree is the first step in inferring LCA. Recently, there have been four rooting methods: molecular clock, Bayesian molecular clock, outgroup rooting, and midpoint rooting (Kinene T et al. *Encyclopedia of Evolutionary Biology*, pp 489-493). Herein, we used the outgroup rooting method to determine the LCA of IREDs. According to the position of the outgroup, we chose a position that divides the ingroup and outgroups into two distinct clusters as the root of the tree in Figtree software. We selected a sequence (Uniprot ID: A1B8Z0) as a seed to conduct a blast, and 18 sequences were downloaded for use as outgroup sequences after the blast run.

There are two reasons for the selection of outgroup sequences. (1) The enzymes with Rossman-fold originate from a common ancestor (Laurino P. et al. *PLoS Biol.* **2016**,

14, e1002396). (2) The seed sequence of the outgroup has been reported to possess the ability to reduce imine and belongs to NADH-dependent oxidoreductase (von Borzyskowski L. S. et al. *Nature*, **2019**, 575, 500-504). The above-mentioned evidence shows that these sequences are distantly related to IRED sequences, making them suitable for use as outgroup sequences. We hope these answers are useful to support the reliability of ancestral sequence reconstruction.

Table R5. Ancestral sequence reconstruction statistics.

Ancestral enzyme	Mean posterior probability ^a	Uncertain sites ^b	Sequence differences between alternative ^c
N1	0.89	70 (24%)	45 (15%)
N2	0.93	40 (13%)	31 (10%)
N3	0.95	30 (10%)	23 (8%)
N4	0.95	29 (10%)	19 (6%)
N5	0.95	26 (9%)	15 (5%)
N557	0.94	35 (12%)	23 (8%)
N558	0.94	39 (13%)	26 (9%)
N559	0.92	46 (16%)	34 (11%)
N560	0.94	36 (12%)	30 (10%)

^aMean posterior probability of the maximum-likelihood version of the ancestral sequence using the WAG model.

^bNumber (percentage) of sites at which the maximum-likelihood reconstruction sequence has posterior probability < 80%.

^cNumber (percentage) of sites at which the alternative sequences of the ancestral enzyme differ in residues.

2. The phylogenetic construction details are not sufficient. The tree robustness as well as the sequence curation needs more clarification. The manuscript does not sufficiently explain the branch lengths (unlabeled on the tree), root and out-group choice with satisfactory detail. In addition, it would be nice to see some measure of branch support, especially the nodes for resurrected proteins. Otherwise, no one can see if the tree topology is well or poorly supported. Because the algorithms for inferring ancestral sequences assume that the given phylogeny is true, this problem substantially undermines confidence in the inferred ancestral sequence. Providing information on the likelihood or posterior probability of each ancestral amino acid in the resurrected sequences would be also advisable.

Response #3-2:

We used the PhyML 3.0 online server (<http://www.atgc-montpellier.fr/phyml/>) to reconstruct the phylogenetic tree. Firstly, a multiple sequence alignment (MSA) file containing outgroup and ingroup (phylp format) was input data. Then, these default parameters in the substitution model option, optimization options, and branch support were selected. The matrix of amino acid substitution WAG with parameters + G + I + F was selected by Smart Model Selection (SMS) according to the Akaike Information Criterion (AIC), which was the best-fitting evolutionary model. Given more than 1400 sequences, the phylogenetic tree reconstructed by PhyML 3.0 was used to further calculate the branch support with BOOSTER (Lemoine F. et al., *Nature*, **2018**, 556: 452-456).

These sequences used to reconstruct a phylogenetic tree were derived from an IREDs database (Fademrecht S. et al., *Proteins*, **2016**; 84: 600–610). The researchers established the data source based on reliable bioinformatic analysis, and the *S*-IRED-*Ss* (GI: 460838082) and *R*-IRED-*Sk* (GI: 524933133) were used as seed sequences to set up this database. In this database, these sequences with a global sequence identity of >98% were assigned to a single protein entry. Therefore, the researchers have already performed a redundancy analysis. Importantly, this database contains plenty of IREDs

that have been characterized by biochemistry experiments. Therefore, we chose all sequences in this database to reconstruct a phylogenetic tree, except for one sequence starting with the X character.

We initially thought that if more than 1000 values were displayed, it would make the picture look cluttered. We really appreciate the suggestion, and these bootstrap values at the nodes for resurrected proteins have been displayed (Figure R11). These posterior probabilities for each ancestral amino acid in the resurrected sequences are listed in Table R5.

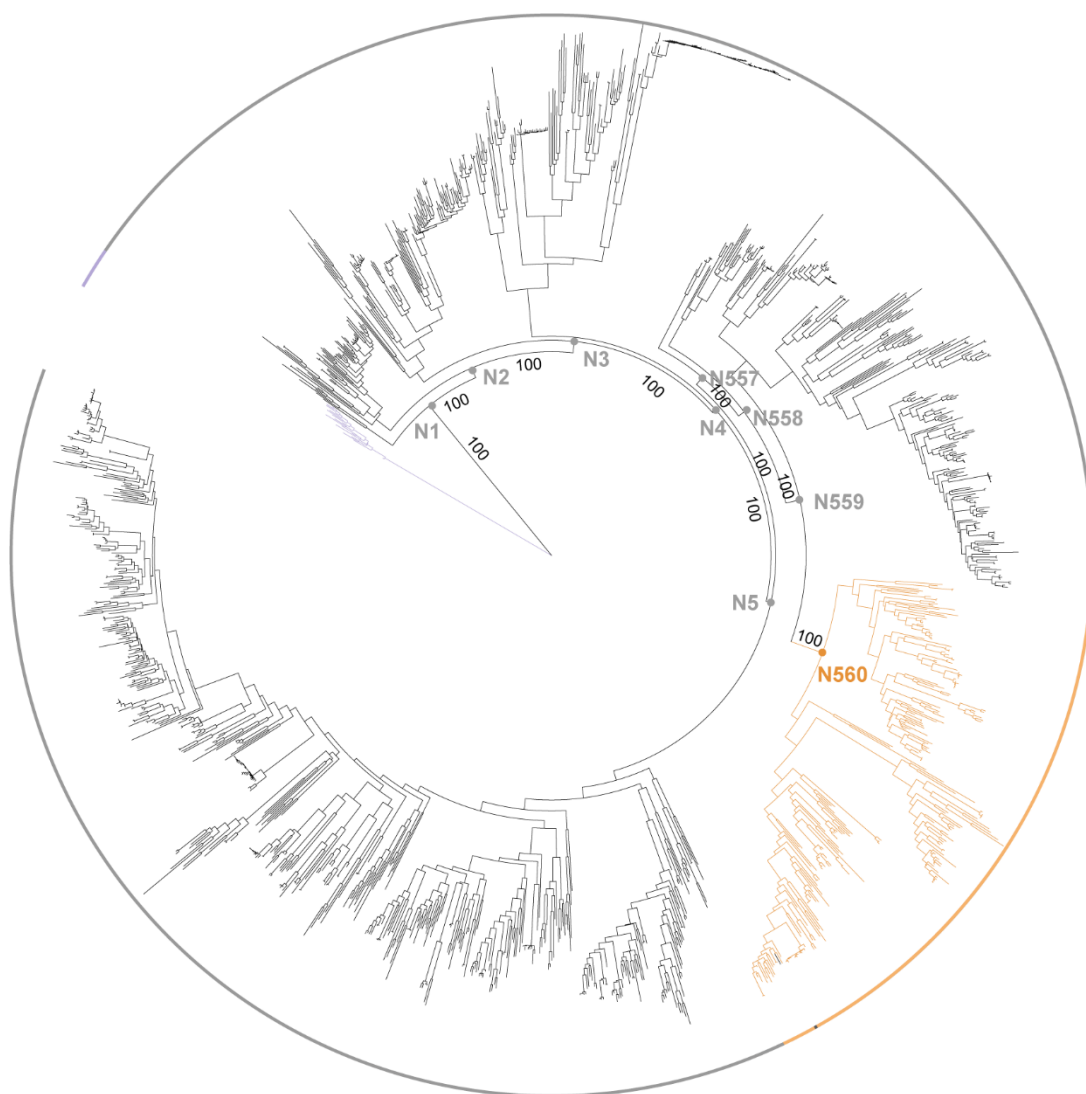


Figure R12 The phylogenetic tree of imine reductases in this study. The yellow clade comprises all Y-type sequences, except for one sequence that is not Tyr at standard position 187, which is also included in this clade. The purple clade is the outgroup we

select for determining the root of the ingroup. Moreover, the labels N1, N2, N3, N4, N5, N557, N558, N559, and N560 are reconstructed ancestral sequences in this research. These numbers in the branch are the transfer bootstrap expectation values calculated with BOOSTER (<https://booster.pasteur.fr/>).

3. The tree's architecture should also be described in more detail. Supplementary Fig. 1 depicts only the branches of the phylogenetic tree, making it difficult to determine the distribution of species. It should be labeled to at least distinguish between Archaea, Bacteria, and Eukarya, and to indicate the distribution of major phyla. Furthermore, it should be compared to known phylogenetic trees to assess its reliability. Without this comparison, the reliability of the phylogenetic tree cannot be evaluated.

Response #3-3:

We gratefully thank you for the precious time the reviewer spent making constructive remarks. According to the description in the paper (Fademrecht S. et al., *Proteins*, **2016**; 84: 600–610), more than 99% of the IREDs sequences in this phylogenetic tree are of bacterial origin, and only ten are of fungal origin. Thus, this study mainly focuses on the IREDs originating from prokaryotes.

To support reliability of this study, we have also used IQtree to reconstruct a phylogenetic tree. We found that the topological structure of the phylogenetic tree resulting from PhyML 3.0 and IQtree is almost identical (Figure R13). Therefore, we argued that our reconstruction is reliable.

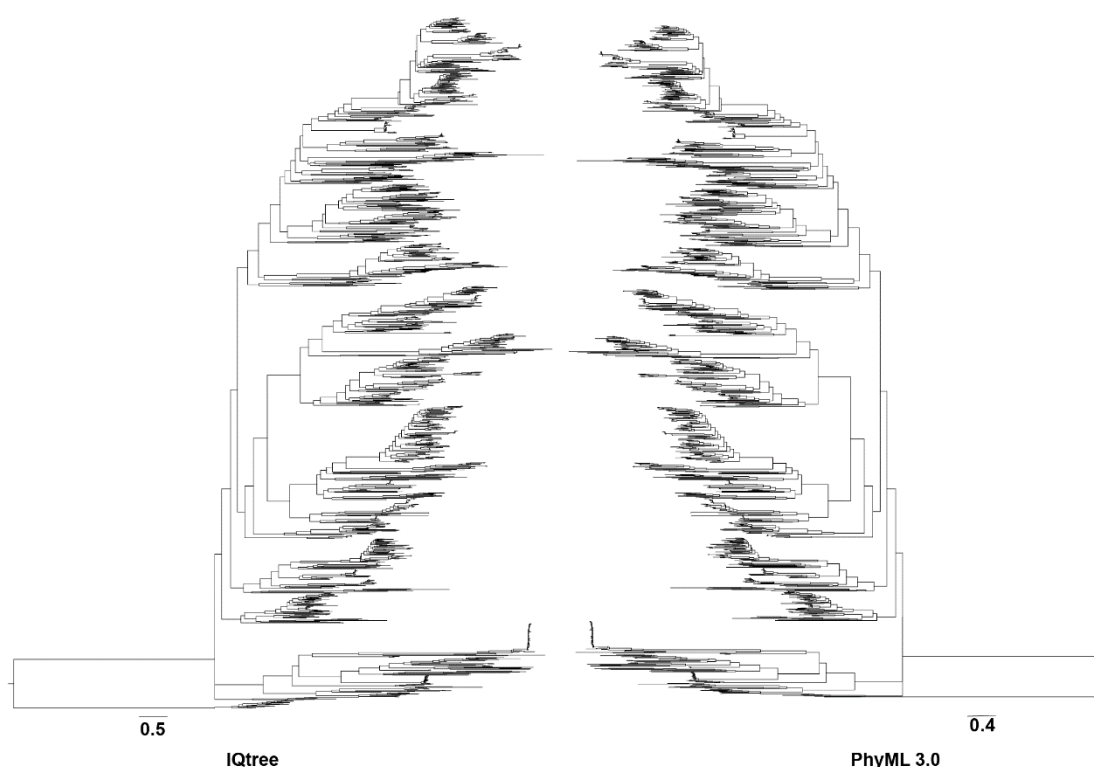


Figure R13. Topology comparison of phylogenetic tree resulting from IQtree and PhyML 3.0.

4. It would be appreciated if a list of sequences (accession numbers) included in the phylogenetic tree could be provided as supplementary data.

Response #3-4:

We are very pleased to provide this raw data. We have provided the data in Excel format because there are more than 1400 sequences. The file was named 'IRED sequence GI number'.

5. Regarding the enzymatic characteristics, the relative stereochemical selectivity has been thoroughly investigated. However, while the catalytic activity of the ancestral proteins is listed in Supplementary Table 2, there is no comparison with wild-type (extant) enzymes. This is, in my opinion, very unfortunate as the properties of the ancestrally reconstructed variants cannot be directly compared with the starting sequences and that significantly impacts the relevance of the results.

Response #3-5:

We thank your attention to this aspect very much. Given that all ancestral enzymes have a lower specific activity level, we decided to repeat the activity measurement using gas chromatography (Table R6). These data are updated in Supplementary Table 2. Our results show that the difference in a specific activity is not due to experimental errors. For the difference across ancestors (between the most probable and alternative ancestors), we argued that the conformation and optimal pH difference are mainly responsible for it. Each ancestral enzyme will accumulate new mutations relative to the last common ancestor (N1) during evolution. In spite of the fact that these mutations do not result in essential function differences, they may affect the conformation of enzymes when they react with substrates. Due to the fact that conformation differences play a crucial role in enzyme reactions, ancestors with several new mutations may exhibit differences in the values of specific activities. Additionally, the optimal pH may be different for each ancestral ancestor. Because the measurements of specific activities were conducted under the same conditions rather than under their respective optimal conditions, we would observe great differences across ancestors.

Additionally, the difference of activity between the most probable and alternative ancestors was also found in relevant studies. For example, in the first research investigating the evolution of stereoselectivity using ASR, we found that the conversion of maximum-likelihood and alternative enzymes presented in Fig. 3 shows a significant difference (Chiang, C-H. et al. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, 120(15), e2218248120). In a recent study investigating the evolution of phosphorylase activity in ancestral glycosyltransferase, the maximum-likelihood ancestral enzyme shows up to 10^2 time more K_{cat}/K_m than the alternative ancestral enzyme (Figure 1d, Franceus J. et al. *ACS catal.* **2024**, 14, 3103-3114). More importantly, although there is a significant difference in specific activity between maximum likelihood and alternative ancestors, stereoselectivity remains a consistent trend. Therefore, we argue that our reconstruction is robust. Considering stereoselectivity is the focused property in this manuscript and each ancestral enzyme can catalyze the reduction of **1a**, we don't

perform extra discussion on specific activity. We hope these data and answers will help you better evaluate our work. Many thanks again for your comments.

Table R6 The specific activity of each intermediate and corresponding alternative ancestral sequences from N1 to N560 toward **1a**.

Enzymes	Specific activity (mU/mg)
N1	56 ± 5
N2	46 ± 2
N3	43 ± 6
N4	46 ± 4
N5	60 ± 0.9
N557	62 ± 1
N558	30 ± 1
N559	1.4 ± 0.1
N560	14 ± 0.6
AltN1	38 ± 4
AltN2	20 ± 1
AltN3	53 ± 3
AltN4	39 ± 3
AltN5	80 ± 0.5
AltN557	18 ± 1
AltN558	60 ± 3
AltN559	4.2 ± 0.09
AltN560	0.34 ± 0.1

Specific activity was measured by gas chromatography.

6. Additionally, it seems interesting to explore the K_m and k_{cat}/K_m of extant and reconstructed proteins for each substrate as well.

Respond #3-6:

We greatly appreciate your concerns. We herein present these kinetic parameters of maximum-likelihood ancestral enzymes (Table R7), which have been attached in Supplementary Information. We argued that the maximum-likelihood ancestors are the main object in this study, and the important conclusions come from them. Moreover, the other substrates were just used to verify the role of the six mutations in inverting the stereoselectivity of N559. Therefore, we believe that these kinetic parameters of maximum likelihood toward the model substrate are more important. We hope this data is informative for your review. Importantly, we sincerely appreciate your invaluable suggestions.

Table R7 Kinetic parameters of the ancestral enzymes toward model substrate **1a**.

Enzymes	K_m (mM)	K_{cat} (s^{-1})	K_{cat}/K_m ($s^{-1}mM^{-1}$)
N1	0.34	0.019	0.056
N2	0.030	0.025	0.83
N3	0.041	0.068	1.65
N4	0.013	0.012	0.92
N5	0.25	0.017	0.068
N557	2.70	0.027	0.010
N558	1.73	0.0054	0.003
N559	2.13	6.90×10^{-4}	3.20×10^{-4}
N560	0.11	0.014	0.13
<i>ScIR</i> ^a	0.173	0.16	0.90
<i>SvIR</i> ^a	0.064	8.40	131
<i>StIR</i> ^a	0.044	0.53	12.1
<i>RIR</i> ^a	0.086	0.77	8.90

^aKinetic parameters are cited our previous report (Zhang Y. H. *Org. Lett.* **2020**, *22*, 3367–3372).

7. It is undeniable that the authors' epistasis analysis has provided crucial insights into the evolution of stereochemical selectivity. However, the impact of epistasis on protein evolution has been debated for the past 20 years. It would be appropriate to cite the following landmark papers and discuss them:

doi: 10.1038/nature05385

doi: 10.1038/nature08249

doi: 10.1038/nature23902

Respond #3-7:

We sincerely thank you for your important recommendations. We have revised the Discussion section and cited these important papers in the revised manuscript.

8. In the Discussion section, the authors describe, "This differs from previous studies in which inverting stereoselectivity through protein engineering involved mutations that typically interact directly with substrates." However, it is not fair. It is well known that amino acid substitutions remote from the active site often influence activity. ASR can be a very powerful tool for accessing these mutations - as outlined in the paper (which you should cite in this context):

doi.org/10.1042/BST20180614

Response #3-8:

We sincerely appreciate the excellent suggestion from you and have revised it accordingly. We have cited this important review at line 240 of the revised manuscript.

The changed sentence:

Our structural analysis showed that five out of six substitutions did not directly interact with the substrate, indicating that ASR has significant potential to capture remote mutations critical for protein functions⁴³.

Point-by-point Responses

Contents

Response to Reviewer #3.....	page 2–18
------------------------------	-----------

Reviewer #3 (Remarks to the Author)

It is undeniable that the authors have made efforts to revise the manuscript in accordance with the reviewers' comments. However, it still seems that the revisions made do not satisfactorily address the initial review comments. Specifically, as the authors themselves state that "Inferring the last common ancestor (LCA) of all extant imine reductase (IREd) sequences is a significantly key part of this study," the placement of the root within the IREd phylogeny is critically important. Nonetheless, it appears that the issue of the root position's reliability has not been adequately addressed in the revised manuscript. Therefore, in my opinion, further revisions are necessary for publication.

Below, I provide additional comments on each response from the authors.

Respond #3-1:

Upon examining Supplementary Fig. 1, it is evident that the branch connecting the outgroup to the IREd clade is significantly longer compared to the branches within the IREd clade. This suggests that the reliability of the root position within the IREd clade is not beyond doubt. Additionally, the number of outgroup sequences is extremely limited compared to the IREd clade sequences, which raises further concerns about the accuracy of the root position. It seems entirely possible that another position within the IREd clade could be the LCA. Therefore, additional evidence supporting the current root position is necessary, such as verifying the root position using alternative methods.

Response #3-1:

We truly appreciate the time and effort you put into reviewing the manuscript. In order to demonstrate that the root location has no effect on our conclusions regarding the relationship between the six key historical substitutions and stereoselectivity evolution, we reconstructed a phylogenetic tree using new outgroup sequences. The root position of the updated phylogenetic tree is different from the previous tree; however, our bioinformatic analysis proved that the root position does not affect the main conclusion. In addition, to support the significance of the reconstruction, we also included experimental results from two other studies that used ancestral IREds reconstructed from the previous phylogenetic tree.

We selected two β -hydroxyacid dehydrogenase (β -HAD) sequences as an outgroup to rebuild the updated phylogenetic tree. The biochemically

characterized β -HADs and IREDs share considerable similarities in sequence and structure, indicating both enzyme families possess closest relative. We performed multiple sequence alignment (MSA) and gap deletion in accordance with previous methods in this manuscript. Then, we used IQtree (version 2.3.6 for Linux x86 64-bit) to the reconstruct phylogenetic tree with the following parameters: -m MFP -B 1000 -alrt 1000 -T AUTO. The best-fit model is chosen according to BIC: WAG+F+I+R10. The whole phylogenetic tree is present in Figure R1. According to the updated phylogenetic tree, we found that the root position is different from the tree in the manuscript. However, the sequences containing Tyr at 170 still clustered in a clade (Yellow part in Figure R1). We further reconstruct the ancestral sequences located in the focused evolution trajectory (from N1 to N548, corresponding to from N1 to N560 in the previous tree) with GRASP. Based on sequence analysis, the six historical mutations have not been affected by the root location (Figure R2 c and d). Meanwhile, the sequences at the crucial evolution interval (N559-N560 and N547-N548) are highly conserved (Figure R2 a, b, e, and f). Therefore, although the root location in the old tree may not be optimal, the new bioinformation analysis demonstrated that the findings in this manuscript are convincing. These results have been added to the Supplementary Discussion section in Supplementary Information.

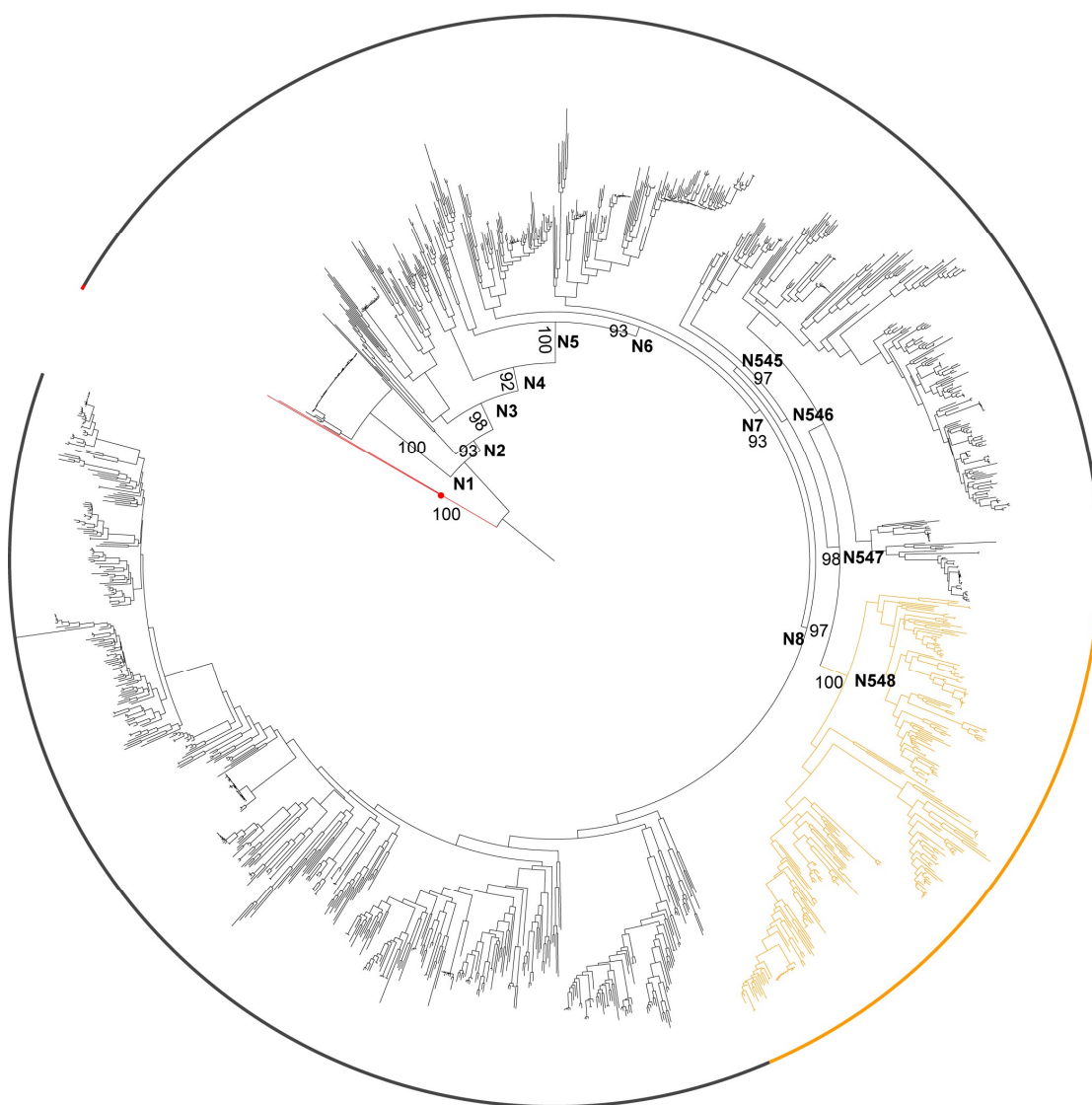


Figure R1. The updated phylogenetical tree. The red branches denote the two outgroup sequences (PDB ID: 4GBJ and 3PEF). The yellow clade corresponds to the N560 clade of the old phylogenetic tree. NXs represent the node name, and the numbers are support values calculated using the ultrafast bootstrap approximation method in IQtree. These support values suggest that the phylogenetic tree is reliable.

(1) We revealed a new reaction pattern catalyzed by N1 (the sequence located at the root of the previous phylogenetic tree), a trifunctional ancient enzyme that has never been reported (Figure R3). Our results demonstrated that ancestral enzymes possess high catalytic promiscuity and, importantly, enormous potential for developing biocatalysts performing new-to-nature reactions. Furthermore, a recent Review highlighted the importance of developing multifunctional biocatalysts (MFBs) for organic synthesis, and the research on MFBs is in its infancy (Thorpe et al. *J. Am. Chem. Soc.* **2024**, 146 (12), 7876-7884). This study is being prepared for submission.

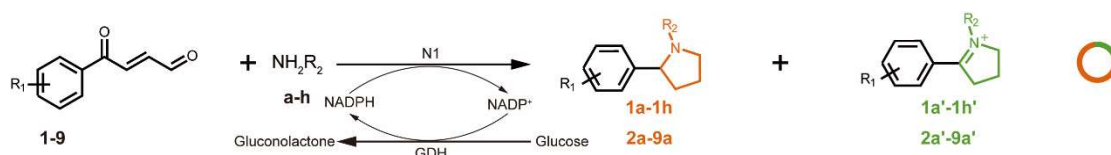


Figure R3. The new reaction pattern catalyzed by trifunctional ancestral IRED (N1). N1 catalyzes three types of reaction in sequence — reductive amination, the reduction of imine and alkene — in the reaction process.

(2) We recently developed a biocatalyst using an ancestral IRED to produce axial chiral benzylamine (Figure R4). This work has been submitted.

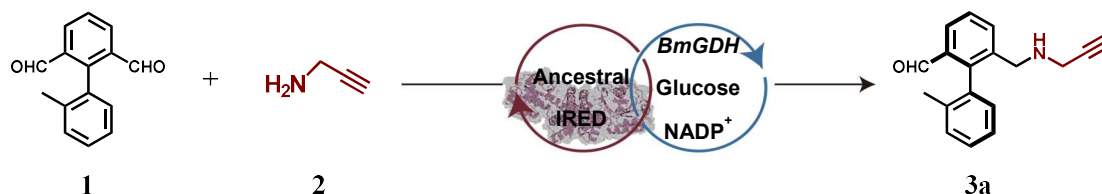


Figure R4. Ancestral IRED catalyze reductive amination of axisymmetric dialdehyde and amine donor.

Finally, several relevant studies demonstrated that the number of outgroups has not affected the root location. For example, although the ingroup contains hundreds of sequences, Ding et al and Chiang et al. used three sequences or one sequence as the outgroup to build phylogenetic tree for performing ancestral sequence reconstruction (Ding et al. *Nat. Commun.* **2024**, 15:2492, Chiang et al. *Proc. Natl. Acad. Sci.* **2023**, 120, e2218248120).

Respond #3-2:

Please include the phylogenetic construction details in either the main text or the supplementary text. Additionally, while the average values for the entire

sequence are shown in Table R5, it is recommended to provide a figure or table in the Supporting Online Materials that shows the posterior probability for each ancestral amino acid residue.

Response #3-2:

We greatly appreciate the insightful suggestions raised by you. We have described each step to construct a phylogenetic tree and the corresponding software in the method section (lines 328–360). Furthermore, we have provided a figure that shows the maximum posterior probability at each site of ancestral sequences in Supplementary Fig. 16 (Figure R5).

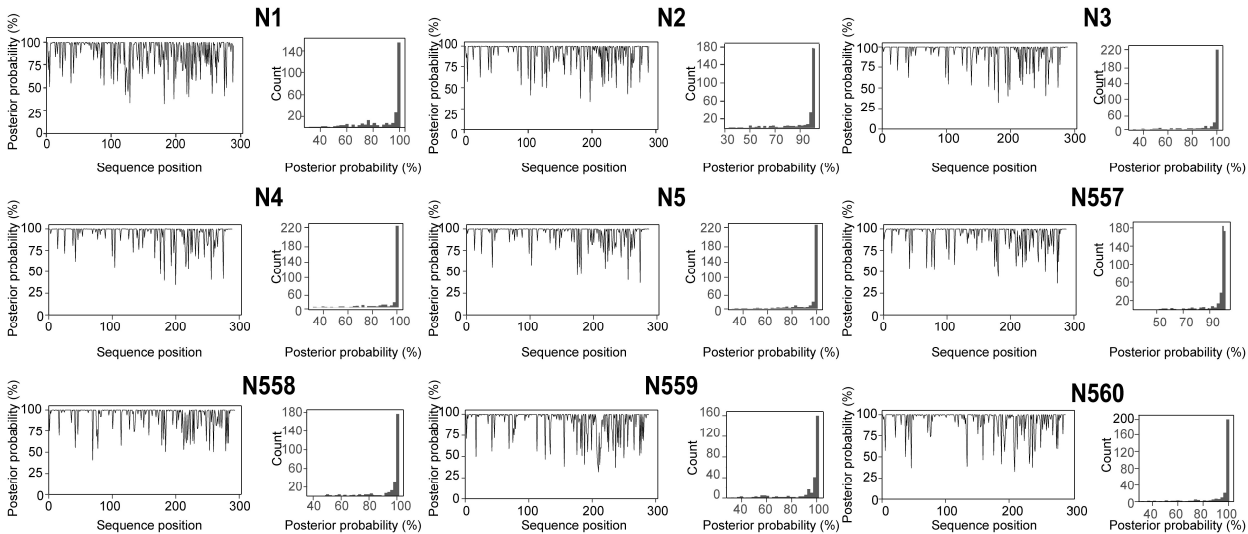


Figure R5. The distribution of maximum posterior probability of each site in all reconstructed ancestral sequences. Posterior probability respect to the position in the sequence is shown with line chart. Distribution of posterior probabilities is shown with histogram.

Respond #3-3:

In my first-round review comments, I requested that the authors show the distribution of major phyla on the phylogenetic tree. While I realized that almost all sequences are derived from Bacteria, there is no discussion on the distribution of bacterial phyla or how this compares to previously reported phylogenetic trees of all organisms or specifically bacterial phylogenies. Furthermore, if the authors aim to demonstrate the robustness of the tree

topology, it would be more convincing to compare it with a phylogenetic tree constructed using Bayesian methods, rather than another tree based on a maximum likelihood method. Please note, I am not expressing concerns about the tree topology, but I would like to see a comparison with other well-known phylogenetic trees.

Response #3-3:

We greatly appreciate your suggestions. We have added descriptions on the distribution of bacterial phyla in lines 330–333 of the revised manuscript:

“These sequences mainly originate from three bacterial phyla, except ten sequences are of eukaryotic origin (Ascomycota)²⁷. More than 80% of sequences are from Actinobacteria, followed by Proteobacteria and Firmicutes. Interestingly, the first reported IRED sequences are also from Actinobacteria.”

We greatly appreciate your well-intentioned and professional suggestions. Until now, there has been no systematic phylogenetic research on the IRED family. Therefore, the other well-known phylogenetic trees are not available. Additionally, reconstructing phylogenetic trees using Bayesian methods is also a time-consuming process, particularly for large datasets. Meanwhile, we are also using the MrBayes software to construct the Bayesian tree. The program has run for 20 days, and the results indicated that more time is required for convergence of the Bayesian tree. When the software runs out, we'll add it to the Supplementary Information. Therefore, given the time-consuming process, we would like to send you the other responses first. Thank you again for your professional advice.

Respond #3-4:

What are the values in column B of the Excel file? For outgroup sequences, it seems to be the UniProt ID, but there is no explanation for the IRED sequences, making it unclear.

Response #3-4:

Many thanks for pointing out this aspect. We are so sorry for inconvenience caused by our mistake. We have added the annotation to the Excel file, and the numbers denote the GenBank Identifier (GI) of each IRED sequences.

Respond #3-5:

It appears that the authors have not addressed the comment from the first-round review requesting a comparison of the catalytic activity of ancestral enzymes with that of wild-type enzymes (extant forms, especially ScIR and SIR). Values for the wild-type enzymes are not provided in Supplementary Table 2.

Response # 3-5a:

We greatly appreciate your well-intentioned and invaluable suggestion. We have added the specific activity of ScIR and SIR to Supplementary Table 2. We're sorry for any inconvenience caused by our carelessness.

Respond #3-5b:

According to Supplementary Table 7, the k_{cat} value for N560 is 1/10 compared to ScIR, but it would be interesting to know how this translates to the specific activity.

Response # 3-5b:

We thank your attention to this aspect very much. As mentioned above, we have updated the specific activity of ScIR (180 mU/mg) in the revised Supplementary Table 2, which is 12.8 folds of the N560 specific activity (14 mU/mg). This is very close to what you estimated from their k_{cat} values.

Respond #3-5c:

Additionally, no explanation for AltN1-N5 and Alt557-560 is found in the revised manuscript. Please add an explanation in both the main text and Supplementary Table 2.

Response #3-5c:

We greatly appreciate your valuable suggestions. We have added the corresponding content in lines 120–121 and lines 123–125 of the revised manuscript:

“In addition, we also investigated the evolution pattern of stereoselectivity in alternative ancestral sequences (AltN1–AltN5 and AltN557–AltN560) to support the robustness of reconstruction (Supplementary Fig. 7).”

“We also measured the specific activity of ancestral enzymes and their corresponding alternative versions (Supplementary Table 2).”

The added explanation in the Supplementary Table2:

“AltN1–AltN5 and AltN557–AltN560 are the alternative ancestor versions of N1–N5 and N557–N560, respectively.”

Respond #3-5d:

How can the significant difference in specific activity between N560 and Alt560 be explained? Please provide a discussion from the perspective of the reliability of the ancestral sequence inference.

Response #3-5d

We sincerely appreciate the excellent suggestion from you. We have discussed possible reasons for the difference in specific activity between N560 and AltN560 in lines 278–293 of the revised manuscript:

“A plethora of ASR studies use maximum likelihood (ML) approach to infer ancestral sequences, and one benchmark research showed that the ML method is best estimate of the true ancestral sequences⁶³. It is possible, however, that the ML approach would incur statistical uncertainty at some sites, which means that the reconstructed residues at such sites would possess other

replaceable residues. Eick et al. conducted a systematic study to investigate the robustness of reconstructed ancestral protein functions to statistical uncertainty in the ML method⁶⁴. It is noteworthy that the alternative ancestors' protein functions are qualitatively unchanged from those of their maximum-likelihood ancestors, according to their results. However, they demonstrated that precise quantitative measurements of function, such as, enzyme kinetic parameters, vary among alternative ancestors. Although our reconstruction has a high average posterior probability for each ancestral sequence (Supplementary Fig 16), we also found the same experimental phenomenon in the present study. The specific activity of N560 is significantly higher than AltN560 (Supplementary Table 2), which is a consequence of reconstruction uncertainty. Additionally, we also found that the catalytic efficiency of ancestral enzymes is lower than extant enzymes (Supplementary Table 7), which is in accord with Jensen's hypothesis⁶⁵. Meanwhile, the results may indicate that the ancestral IREDs possess high catalytic promiscuity, and our study in progress supports this aspect.

Respond #3-6:

I believe that the reaction rate kinetics parameters added by the authors would be helpful for understanding this paper. However, it seems that the new Supplementary Table 7 is not cited in the main text, and there is no description of these results in the main text either. Additionally, there is no explanation of how the reaction rate kinetics parameters were determined, including the reaction temperature and composition of the reaction mixture.

Response #3-6:

Many thanks to you for highlighting this aspect to improve the quality of our manuscript. We have described and cited the results of Supplementary Table 7 in the main text in lines 290–291 of the revised manuscript: “Additionally, we also found that the catalytic efficiency of ancestral enzymes is lower than extant enzymes (Supplementary Table 7), which is in accord with Jensen's hypothesis.”

Moreover, we listed the detailed methods used to measure kinetics

parameters in the activity assay section of Supplementary Information.

Respond #3-7:

I think it has been appropriately addressed.

Response #3-7:

Thank for your positive feedback.

Response #3-8:

I think it has been appropriately addressed.

Response #3-8:

We greatly appreciate your positive feedback.

I looked at the comments from Reviewer 1 and the authors' responses, as per [Editors] request. While the authors have adequately addressed many of Reviewer 1's comments, there are several instances where the responses appear to be insufficient, as I will point out below.

We greatly appreciate your enormous contribution to our manuscript. Moreover, our sincere appreciation goes out to you for investing so much time and effort into improving this manuscript.

Response #1-3: Even if the paper's main focus is on the inversion of stereoselectivity, the significant differences in specific activity observed between the reconstructed ancestral proteins should be mentioned and discussed in the revised manuscript. The authors suggest that these differences might be due to variations in optimal pH; however, this explanation lacks evidence and remains speculative.

Response #1-3-new:

We greatly appreciate the insightful query raised by you. We've updated our responses to this comment and added the corresponding contents to the revised manuscript. Few mutations in the active pocket would cause a significant difference in specific activity in the directed evolution campaign. For example, introducing two-point mutations (V122C/F177W) in the active pocket would improve ScIRED's specific activity by 10-fold (Chen Q. et al., *ACS catal.* **2022**, 12, 14795-14803). The results indicate that positions 122 and 177 play an important role in regulating activity in IREDs. Based on the relevant research findings, we further investigated the residues located in the active pocket of

ancestral enzymes (Figure R6). Our findings showed that the active sites of N559 and N560 have suffered conspicuous mutations compared to those of other ancestral enzymes. Specifically, we found that residues at positions 122 and 177 have mutated among ancestral enzymes (Figure R6). The phenomenon bears similarities to the directed evolution study. Therefore, we speculated that differences in the active pocket residues caused a significant variation in activity among the reconstructed ancestral enzymes.

We have added relevant discussions in lines 294–303:

“Furthermore, we observed differences in specific activity between reconstructed ancestral enzymes (Supplementary Table 2). According to lessons learned from directed evolution campaigns, for example, two mutations (V122C and F177W) at the active pocket in SclR improve its specific activity 10-fold²¹. The results indicate that positions 122 and 177 play an important role in regulating activity in IREDs. Then, we looked at the residues in the active pocket of ancestral enzymes and discovered obvious mutations in the location, especially for N559 and N560 (Supplementary Fig. 17). Specifically, we observed that the two key positions (122 and 177) have mutated compared with other ancestral enzymes with different magnitudes of activity. Therefore, we speculated that differences in the active pocket residues caused a significant variation in activity among the reconstructed ancestral enzymes.”

Enzyme	95	120	121	122	123	124	170	173	174	177	178	208	231	232	236
N1	S	I	M	A	T	P	D	L	L	M	W	W	-	A	V
N2	S	I	M	A	T	P	D	L	L	M	W	W	-	A	V
N3	S	I	M	A	T	P	D	L	L	M	W	W	-	A	T
N4	S	I	M	A	T	P	D	L	L	M	W	W	-	A	T
N5	S	I	M	A	T	P	D	L	L	M	W	W	-	S	M
N557	S	I	M	A	T	P	D	L	L	M	W	W	-	S	M
N558	S	I	M	V	P	P	D	L	L	M	W	W	E	S	M
N559	S	I	M	V	P	P	D	L	L	F	W	I	E	S	M
N560	S	V	M	V	P	P	Y	Q	L	F	W	M	E	D	M

Figure R6. The residues around substrate 1a 6 Å in ancestral enzymes. Orange residues denote highly conservative sites. The ‘-’ presents gaps in multiple sequence alignment.

Response #1-4: The authors' discussion seems to remain speculative, and no evidence has been provided to support their claims. Specifically, which amino acids in the extant proteins reduce the effect?

Response #1-4-new:

Thank you so much for your comments. The opinion of 'the ancestral sequence is to some extent capable of avoiding epistasis' was cited from the review paper of Hochberg et al. (Hochberg G. K. A. & Thornton J. W. *Annu. Rev. Biophys.* **2017**, 22, 247-269). In this review, authors proposed an advantage of ancestral sequence reconstruction via the following statement: "*A second advantage of vertical evolutionary analysis is that it can overcome the obscuring effect that epistasis often has on sequence-structure-function studies that focus on present-day proteins.*" To make it clear, we modified our description in the revised manuscript in lines 55-59:

"Additionally, epistasis, the phenomenon in which the effect of mutations on functions depends on other mutations, often complicates the study of sequence-structure-function relationships. Empirical evidence has shown that ASR can help reduce the effect of epistasis by reintroducing historical substitutions into a genetic background that is identical or very similar to the one in which they originally occurred."

Response #1-15: The explanation appears to be insufficient. Does "six-order epistasis" refer to the interaction among six mutations? Please clarify.

Response #1-15-new:

We sincerely appreciate the excellent suggestion from you. Yes, the six-order epistasis denotes the interaction among six mutations. We have added a new explanation at line 221: "*Additionally, we used a Python package to detect epistasis and quantify the fraction of variation accounted for by first- to sixth-order (the interaction among six mutations) epistatic coefficients.*"

Response #1-16: While the low activity of some Alt forms is somewhat understandable based on the explanation provided in response #1-3, it is necessary to include a reasonable explanation in the manuscript for the significantly lower specific activities of N559 and N560, especially N559. Since the enantiomeric excess selectivity is inverted between N559 and N560, the

low specific activities of these forms must be thoroughly examined. Additionally, the manuscript should discuss the potential relationship between the inversion of stereoselectivity and the reduction in specific activity. Failing to do so could raise concerns about the reliability of the results, particularly regarding the accuracy of the ancestral sequences.

Response #1-16-new:

We greatly appreciate your valuable suggestions. We have made a discussion on the low specific activity of N559 and N560 in Response #1-3-new and added corresponding discussions in lines 293-302. Meanwhile, we have discussed the potential relationship between the inversion of stereoselectivity and enzyme activity in lines 303-313:

“An empirical study on the directed evolution of enantioselectivity revealed a trade-off between enantioselectivity and activity, resulting in improved enantioselectivity at the expense of activity⁶⁶. The other study on alcohol dehydrogenase showed that enantioselectivity inversion would decrease the k_{cat}/K_m ⁶⁷. The empirical evidence indicated that stereoselectivity and activity are in a trade-off relationship in protein engineering⁶⁶⁻⁶⁷. Our results on N559 and N559-M6 are consistent with this type of trade-off phenomenon (Supplementary Table 5). However, several studies also showed that further directed evolution would break the trade-off⁶. According to Yu et al., early stereoselective inversion reduces enzyme activity compared to the wild type, but further directed evolution simultaneously increases both stereoselectivity and enzyme activity⁶. Therefore, we speculated the stereoselectivity and activity of N559-M6 will be further improved with the help of more mutations.”

Response #1-17: The results of the additional experiments should be explained and discussed in the main text of the revised manuscript.

Response #1-17-new:

We greatly appreciate your invaluable suggestions. We have discussed the new results from Supplementary Table 5 in lines 313-316:

“Additionally, we examined the activities of the 64 mutants located in the empirical fitness landscape and observed 18 intermediates with improved activity than N559 (Supplementary Table 5). Moreover, the mutants with inverted stereoselectivity exhibit decreased activity compared to the start point (N559).”

Response #1-18: The authors' response remains abstract and lacks specificity. Additionally, it has not been reflected in the main text.

Response #1-18-new:

We really appreciate your efforts in reviewing our manuscript and your precious comments. We have described some prediction tools involved epistasis in lines 273-277. Due to our lack of experience developing statistical models, we are unable to provide a specific and detailed experimental plan. Consequently, we can only provide a summary of relevant research progress.

The added contents:

“Therefore, our results suggest that incorporating epistasis into predictive models contributes to improving the accuracy of predictions for the effect of mutations on protein functions. A plethora of epistasis-based predictive models have been developed in recent years, and these models outperform additive models without epistasis, such as EVmutation and Innov SAR^{61,62}. ”

Response #1-20: The response merely states that the authors have mimicked previous studies and does not provide a satisfactory answer to the reviewer's suggestions or concerns.

Response #1-20-new:

Many thanks for the excellent suggestions from you. We employed a structure-based method to perform multiple sequence alignments because structure is more evolutionary conservative. According to our analysis, two sources are responsible for the major indels. 1) These major indels that occurred in ingroup sequences were caused by the outgroup sequences. Because the structures of ingroup and outgroup sequences are not entirely identical, we manually removed these major indels. 2) The remaining major indels may indicate that these locations do not match the structures used to conduct MSA. Therefore,

we referred to Gblock tool parameter sets and empirical research as the primary standards to identify major indels. If more than 50% of ingroup sequences in a location show indels, we will remove these locations. Furthermore, our resolved structures demonstrate that these removals have no effect on the overall structure. The ancestral enzyme structures are consistent with previous reported structures.

We also added corresponding contents in lines 338-343 of the revised manuscript:

“Two principles are employed to modify major indels in the MSA: 1) We removed major indels that are caused by outgroup sequences. These indels result from differences in structure between ingroup and outgroup sequences. 2) We discard positions where more than 50% of sequences contain gaps in the MSA³⁹. An excess of indels at a location suggests that the location is not sufficiently conserved in the evolutionary process, since protein structure is more conservative in evolution.”

Response #1-21: The authors should also state their response to the reviewer in the main text of the revised manuscript.

Response #1-21-new:

We sincerely appreciate your valuable suggestion. We have added corresponding descriptions in lines 358-360:

“Additionally, we also used the joint reconstruction method to reconstruct the ancestral sequences, and the sequences from joint reconstruction are almost identical to those from marginal reconstruction (Supplementary Fig. 18–20).”

Response #1-22: I agree with the concerns raised by Reviewer 1. Even if similar results have been reported in other ASR studies, this does not support the robustness of the inferred ancestral sequences (and their catalytic properties). At the very least, the authors need to provide such an explanation in the main text.

Response #1-22: The authors should explain in the main text that the ambiguous ancestral amino acid sites in the ancestral sequences significantly

influence the specific activity. In other words, while the evolution of stereoselectivity from the ancestral sequences can be reproduced, the magnitude of specific activity may not be reproducible. This might explain the significantly lower specific activity observed in N559.

Response #1-22-new:

We greatly appreciate the excellent suggestion from you. We have discussed the relationship between the difference in activity and ambiguous reconstructed sites in lines 278-293. Meanwhile, we also added explanations for the low activity of N559 in lines 294-303. We hope these responses are helpful for your review. Thank you again.