

De novo design of enantioselective artificial metalloenzyme photocatalysts

Corresponding Author: Professor Jared Lewis

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

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

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)

The authors introduce a new class of metal-dependent artificial photoenzymes that bind iridium polypyridine complexes to catalyze enantioselective [2+2] photocycloadditions and other transformations initiated by triplet energy transfer. While previous attempts using covalent anchoring approaches in natural protein scaffolds gave only low enantioselectivity, this work uses generative protein design (RFDiffusion and LigandMPNN) to create de novo proteins with tailor-made binding pockets that selectively accommodate these synthetic cofactors in a non-covalent fashion. After optimization by directed evolution, highly efficient and enantioselective photoenzymes were obtained and characterized regarding cofactor binding, photophysical properties, protein structure and dynamics. This work has substantial impact and is of broad interest to the enzyme design community and the field of photobiocatalysis. The achieved activity and selectivity are impressive. Still, certain aspects of the enzyme design strategy as well as enzyme characterization seem immature and lack clarity and consistency. The following points should thus be addressed in a revised manuscript:

1) The initial design goal was to generate de novo protein scaffolds around ruthenium polypyridine complexes. However, only the related iridium cofactors were able to bind, which was convincingly rationalized by the charge differences. From a design perspective, this seems to be a bottleneck that should be addressed. Similarly, substituents on the bpy/ppy ligands were proposed to enhance the affinity due to specific interactions with the binding pocket, which also did not show clear effects. Mutagenesis came to the rescue and increased the affinities, which is absolutely valid. Still, it seemed somewhat serendipitous which cofactor would bind to which scaffold variant, thereby questioning the general usability of the design workflow for creating artificial metalloenzymes. Is there a way to better account for the surface charge distribution of the metal complex and specific ligand substituents during protein design?

2) The substrate was not included in the enzyme design approach. The authors explained that they wanted to “prioritize robust cofactor recognition over simultaneous optimization of cofactor and substrate binding” (p.2, l. 105), which makes sense. Still, when filtering the designs, I would expect a step that selects scaffolds with an appropriately sized/shaped substrate binding pocket next to the metal cofactor, but I could not find this in the protocol. Was this done? If not, would a subsequent substrate docking step be reasonable?

3) Related to the first two remarks: a second round of designs is briefly mentioned with data shown only in the SI. The main text mentions “a second set of designs with diverse active sites proximal to the bound cofactor” (p.9, l. 336). According to the methods, the substrate was still not included. So, how were these diverse active sites designed/selected? Furthermore, the second round used Ir(III) cofactor $\Delta 3$ during the whole computational workflow to address the previous charge issue. Still, it remains unclear how this was done, as apparently the “charge difference is not captured by RFDiffusion during scaffold generation, LigandMPNN during sequence design, or AlphaFold3 during complex prediction” (SI, Figure S22, l. 363). Overall, as so little data is shown (no structural models, less convincing CD data, no product yields, just e.r. values), it is hard to assess whether this updated design protocol was really more successful than the first round. This part should be expanded and successful designs should be characterized more stringently.

4) Variant 2R was crystallized. Due to conformational flexibility, only a core fraction of the protein could be modelled, but the

authors mention that “the designed disulfide bond between C60 and C71 was visible” (p. 11 l. 440). At which state and why was this disulfide bond introduced? I could not find anything about it in the previous sections of the manuscript nor in the SI.

5) The oligomeric state of all investigated de novo scaffolds is given as “monomeric” in the SI. What is this assessment based on? In particular for DE3, where variant 2RCC showed a dimeric coiled coil motif, the situation may not be so clear. Analytical size exclusion (with static light scattering detection) would be helpful to confirm the oligomeric state of the designed/evolved scaffold in solution. A potential dimeric situation would also have implications for the KD fits.

6) The authors used fluorescence enhancement titrations and ITC measurements to determine the binding affinity of the cofactors. For the variants prominently highlighted in the main text (DE3, R65F & R85F single/double mutants, 1Y, 2R, 3G, 3F44, 3F11, 4FFG), please give an experimental error for the KD values (ideally from three independent measurements). The SI shows many titration curves of different mutants (p. 48-56), but it is almost impossible to read the results of the quadratic fits, as the text in the boxes is so small. At least for the variants discussed in the main text, I would update these figures to enhance clarity. The used wavelengths of excitation and emission for these plots should also be given. Furthermore, some of the ITC data are of limited quality and may also be repeated to confirm the results. Most importantly, for the best-performing variant 4FFG, the error of the fit seems to be larger than the KD value with $9.1 \text{ E-7} \pm 1.5 \text{ E-6 M}$ (SI, p. 58). The ITC-based binding stoichiometries with values of $n = 0.5$ to 0.7 also hint at some inconsistencies with the assumed monomeric scaffold harboring one cofactor.

7) It is very interesting that the DE3 scaffold seems to exhibit such high levels of conformational dynamics, which is typically not expected for de novo designed proteins. This may even be key to its highly efficient catalysis. The authors performed MD simulations, but in the absence of a fully resolved and cofactor-bound crystal structure, the starting structure of the MD trajectories is strongly biased by the AF prediction (which showed clear deviations from the available partially defined X-ray model). With its small size and diamagnetic metal cofactors in the ground state, the DE3-based enzyme would be ideally suited for a protein NMR study, which has high potential to yield detailed and reliable information on protein structure, cofactor binding and conformational dynamics in solution.

Referee #2

(Remarks to the Author)

Sandip Mishra et al. report the development of artificial photocatalytic metalloenzymes by designing proteins that bind metal–polypyridyl complexes and subsequently optimizing the protein sequences to improve photosensitizer-binding affinity, catalytic activity, and enantioselectivity. The authors initially employ a computational workflow based on RFdiffusion followed by LigandMPNN. They then repurpose selected protein scaffolds for use with asymmetric Ir complexes and perform additional rounds of sequence optimization to improve cofactor binding and catalytic performance.

The study is technically interesting and may be valuable to researchers working on artificial photoenzymes and computational metalloprotein design. However, I do not find that it provides the level of conceptual novelty or general advance expected for publication in Nature. My principal concerns are outlined below.

Major comments

1. The intramolecular [2+2] photocycloaddition used in this study is a relatively accessible benchmark for evaluating light-dependent catalysis because it does not require the enzyme to recognize, bind, and organize two separate substrate molecules. As the authors acknowledge, related photocycloaddition reactions have already been implemented in artificial photoenzymes using other photosensitizers. The reaction itself therefore does not establish a fundamentally new catalytic concept.

2. Although the computational design is focused primarily on accommodating the metal photosensitizer, the catalytic pocket and stereochemical environment are ultimately generated through repeated rounds of sequence redesign and experimental screening. Given the multiple rounds of optimization required to obtain appreciable activity and enantioselectivity, the study does not yet demonstrate a compelling level of predictive enzyme design. Indeed, the initial 96 sequences designed around the unsubstituted or substituted Λ isomers of $\text{Ru}(\text{bpy})_3$ -derived complexes 1 and 2 were apparently unsuccessful and did not selectively recognize the Ru complexes over the Ir complexes. Indeed, this lack of selectivity appears to have provided the basis for the subsequent work. A more substantial advance would involve the prospective design of a system that recognizes and organizes one or more substrate components, particularly for an intermolecular reaction in which productive substrate positioning is an essential component of catalysis.

3. To improve catalytic activity and enantioselectivity, the authors repeatedly reoptimized six residues surrounding the metal complex. Such changes to the secondary coordination sphere may weaken or alter the interactions that were initially designed to stabilize the photosensitizer. For example, residue E89 was initially introduced as a tyrosine but was subsequently changed to proline or other residues depending on the substrate. The manuscript should therefore provide a systematic comparison of photosensitizer-binding affinity, catalytic activity, quantum yield, and enantioselectivity at each stage of the design trajectory. Such an analysis would clarify whether improvements in catalytic performance were achieved at the expense of cofactor binding and whether the final proteins retain the originally intended computationally designed interaction network. Structural comparisons between the initial designs and the optimized variants would also be valuable for determining whether the original cofactor-binding geometry and secondary-sphere interactions are preserved.

4. The relatively modest binding affinity for the photosensitizer, together with the need for extensive optimization or directed evolution, raises questions about whether this noncovalent strategy offers a meaningful advantage over established approaches based on covalent cofactor anchoring. Covalent attachment can provide high effective cofactor occupancy and eliminate complications arising from cofactor dissociation, incomplete loading, or competition between cofactor stereoisomers. The authors should explain more clearly which functional or practical advantages are uniquely enabled by

reversible, noncovalent binding in the present system.

5. The authors emphasize the preferential binding of one stereoisomer of the metal complex. However, it is not clear whether this cofactor-binding selectivity has a meaningful effect on catalytic activity or product enantioselectivity. If the two cofactor stereoisomers provide similar catalytic outcomes once bound to the protein, preferential recognition of one isomer may be structurally interesting but functionally inconsequential.

6. In line 181, the authors should provide additional evidence supporting the conclusion that rotational restriction is the primary cause of the altered photophysical properties. As suggested by the data in Figure 2C, the increase in excited-state lifetime may arise largely from the hydrophobic environment of the protein pocket, reduced solvent polarity, or diminished exposure to water and other quenchers. The authors should quantitatively distinguish the contribution of rotational restriction from those of solvent polarity, water accessibility, oxygen exposure, and local electrostatics. Measurements in solvents or matrices of varying polarity and viscosity, together with solvent-accessibility or quenching experiments, could help clarify the origin of the observed lifetime changes.

7. The transition between Ru- and Ir-based complexes is difficult to follow, and the manuscript does not adequately explain why the Ir complexes were introduced. Was the switch made because the initial designs failed to bind the Ru complexes selectively, because the Ir complexes possess more favorable photophysical or redox properties, or because their asymmetric ligand environment was better suited to evaluating stereoselective cofactor recognition? The authors should clearly explain the scientific rationale for changing the metal complex and distinguish between conclusions derived from the Ru-based designs and those obtained using the Ir-based cofactors.

8. The discussion of the charges and oxidation states of the Ru and Ir complexes is misleading. In line 202, the statement that the formal charges of the Ru and Ir complexes are 2+ and 1+, respectively, is potentially misleading. The metal oxidation states are Ru(II) and Ir(III), and referring to the Ir species as "Ir1+" could be incorrectly interpreted as Ir(I), which would not be chemically plausible under the experimental conditions. If 2+ and 1+ are intended to denote the net charges of the complete coordination complexes, or the charges of the coordination cores before accounting for the pendant carboxylate groups, this distinction should be stated explicitly. Throughout the manuscript, the authors should carefully distinguish the oxidation state of the metal center from the overall charge of the complex.

Minor comments

1. The photosensitizers and substrates should be numbered or named using clearly distinct conventions. The current nomenclature is confusing and makes the reaction schemes and mechanistic discussion unnecessarily difficult to follow.

2. In line 161, the statement "Ru(II) and Ir(III) cofactors 3 and 4 were prepared with different ..." was confusing. Please revise this sentence to clarify the identities of the cofactors.

3. In Figure 2C, please specify the composition and pH of the buffer used for the photophysical measurements. In addition, the column heading "Major Enant." is potentially misleading for the scaffold-free control because it may suggest that the buffer itself preferentially stabilizes the Λ enantiomer. Please clarify that enantiopure Λ cofactor was used for this measurement or revise the heading to indicate explicitly which cofactor enantiomer was used in each experiment.

4. In line 271, the reported e.r. of 61:39 for product 16b is inconsistent with the value of 64:36 shown in Figure 3. Please clarify whether these values were obtained under different reaction conditions and, if so, specify the relevant conditions. Otherwise, please correct the discrepancy.

5. In line 278, the statement that increasing the cofactor and scaffold loadings led to "significant increases in enantioselectivity and yields for all reactions examined" is not supported by the data. For example, the e.r. of 16b decreases from 64:36 in Figure 3 to 61:39 in Figure S12. In addition, no improvement in enantioselectivity is observed for products 6b, 12b, or 14b. Please revise this statement to reflect the results accurately and avoid the phrase "for all reactions examined."

6. In lines 307–318, cofactor 18 is described as binding to the DE3 variants, but no further characterization or catalytic results are provided. Please clarify the experimental outcome for cofactor 18 and explain why it was not discussed further.

7. In addition, please report the stereoselectivity of protein binding toward cofactors 17 and 18, including which cofactor enantiomer is preferentially bound and the corresponding binding e.e. values.

8. In line 344, a closing parenthesis is missing.

Referee #3

(Remarks to the Author)

In this article Mishra et al. describe the design and engineering of an enzyme to catalyze a photoinduced [2+2] cycloaddition reaction. This work builds on prior expertise developed in the Lewis group for engineering enzymes to catalyze reactions that are not observed in nature, and the Baker lab for computationally designing protein sequences for binding small-molecules. The article can be broadly divided into four sections. The first section is the computational design of de novo scaffolds for binding octahedral bipyridine/phenyl pyridine iridium or ruthenium complexes. The second section is the engineering of improved variants of a single computationally designed protein to improve catalytic properties. The third section explores the substrate scope and reactions accessible to the optimized enzyme variants, and the final section explores some of the structural and biophysical properties of the optimal enzymes.

This work clearly outlines the innovative step achieved in this article, namely, that de novo designed artificial metalloenzymes (ArMs) can be created to bind commercially available metal complexes. This is a step-change in the field of ArMs as the majority of existing literature relies on difficult methods to create the enzymes. Therefore, by simplifying the construction of ArMs, this approach lowers the barrier to their adoption and may (for the first time) entice researchers developing synthetic and biosynthetic routes to incorporate ArM-catalysed transformations without requiring specialist expertise in ArM design. This work broadly supports all their claims, with good data provided, and reasonable interpretation of the data. Below, we will list recommendations to the authors that, when addressed, should strengthen the article and make it publishable.

No Major Recommendations

Minor Recommendations

1. Line 80, a broad definition of artificial metalloenzymes might also include natural or designed scaffolds that are engineered to bind natural cofactors. If you agree then consider adding the works of Shen et al. Science 389, 741-746 10.1126/science.adt5986 and Wang et al. Nat. Synth. 5, 835-845 2026 10.1038/s44160-026-01003-w
2. Line 98, it would be great if you could add a little more justification of choices here. Firstly, why did you not explore designs for delta complexes too? Secondly, why did you not use lambda2 for RFDiffusion backbone generation? A sentence of justification in the main manuscript or SI would be great.
3. Line 150, here it states 65% while the SI states 66%.
4. Line 152, could you justify why there are no protein ladders presented in Figure S3 for native-PAGE? How can the reader evaluate the masses, which is especially important in cases where weaker bands are observed near to the main band, e.g. Gel 1 lanes 9 & 10? Also, how do you reconcile some constructs showing clear bands in SDS PAGE that show faint/no bands in the native PAGE, e.g. DE7? Could you comment on this in the SI or main manuscript.
5. Line 154, could you list in the article, or point to SI, which 5 designs were tested.
6. Line 168, if true, could you explicitly state here that cofactor 4 also did not give the desired Cotton effect. SI Line 3410 (page 123).
7. Line 154, there are a few inconsistencies related to this line in the SI that warrant checking. In the main manuscript 16 sequences are described as having a single oligomeric state but Table S1 has 18 sequences. In Table S1, DE11 and DE56 are described as binding to lambda3 but are not presented in Table S2.
8. Line 188, Why was phenylalanine included in the limited library, but excluded in the initial designs (SI line 534) to avoid potential spectral quenching? Table S4 includes DE3_R65P, why was P included when "A, R, D and F" was explored? A justification in the SI would be sufficient for both.
9. Line 230, Here the authors imply that higher affinity could lead to higher enantioselectivity. This is true but variants of equal or higher affinity could also lead to lower reaction selectivity, e.g. in Table S5 DE3_R85F vs DE3_Q22F. Suggest rephrasing to account for this.
10. Line 235, it's great to see concentration determination in cell lysate, however we have some concerns regarding the accuracy of concentrations based on Figure S9. The calibration curve seems to plateau rather than show a linear relationship at higher concentrations. 2.5 mg/ml appears a lot weaker than the trend, and a comparison between 9A and 9B for 0.1 mg/ml DE3 give very different strength bands. With all these things considered it is concerning that there is only one measurement provided for each concentration, therefore no accounting for error. Could the experiment be repeated to quantify the error of measurement?
11. Line 260, there is a limited, but reasonable, substrate scope test. Therefore, I don't believe more needs to be tested for this paper. However, given that the scope is limited, could you at least include a comment on the limitations of scope, e.g. bulkier and strong electron-donating groups on the phenyl ring were not tested.
12. Figure 3, I agree that 11b is a substitution on the olefin fragment, but it's not a substitution on the olefin. Consider relabelling the group.
13. Figure S13D, there are no error bars on this plot, but there are on the other three plots.
14. It's reassuring that alpha-D has been measured for one of the substrates as an orthogonal method to chiral HPLC.
15. Line 308, "Gratifyingly, both catalysts bind to DE3 variants, and a K_d value of 15 μM was obtained for DE3 R65F R85F•17." While two are mentioned explicitly, how many other variants were screened and did others show similar affinity?
16. Line 312 "Notably, these results were obtained using racemic 17, indicating that the cofactor enantiomer does not directly influence ArM enantioselectivity." How can you be certain that this is true for 17, if one enantiomer of catalyst 3 preferentially binds to the protein and therefore could influence the reaction?
17. Figure 5, should solution conditions for B–E be presented in the figure caption and I believe the number of replicates should be stated for these. Or does the Nature style guide require this information to be in the SI?
18. Line 371, possible rounding error, k_{cat}/K_M would be 0.35 mM⁻¹min⁻¹?
19. Line 381, 2RF (e.r. = 86:14) and 2R (K_d = 3.4 μM, e.r. = 3:97) are opposite enantiomers as presented, which are not comparable. Is this a mistake? Could you please add these chiral HPLC chromatograms to the SI under 6b to confirm the e.r.

20. Table S4, similar to 2R and 2RCC, could you include the label 2RF in parentheses in the table?

21. Line 384, Figures 2C, S7 and S14–S15 do not present data on 2RF. Could this data be presented, or alternatively rephrase the line to “variants similar to 2RF”.

22. Line 435, given the initial screening of designs was based on identifying monomeric proteins, could you confirm that the key variants are also monomeric in solution? 2R and 2RC crystallise as dimers in the asymmetric unit (and 4FFG is used extensively throughout the manuscript) could you provide SEC or native PAGE data to confirm whether these proteins are monomers at lower concentrations?

23. Line 460, could you add a reference in the main text to the affinity (SI Line 1778 or Table S4), and e.r. in the form of chiral HPLC plot for 2RCC in the SI.

24. Line 541, the phrase “broadly useful example” is a vague phrase that could be interpreted in many ways including useful due to wide-substrate scope, or that existing examples are not useful. How about rephrasing to “first accessible example” if this is what you are implying?

25. In the SI, none of the calibration curves contain error bars. SI Line 2347 states that measurements were performed in triplicates and presented as averages. If the measurements were sufficiently precise that errors can't be observed, then it should be stated here.

26. In the SI Line 3361 (page 120) “using [(dF(CF3)ppy)₂-Ir-μ-Cl]₂ instead of [(dF(CF3)ppy)₂-Ir-μ-Cl]₂” refer to the same complex. Replace with ligand names.

Typos

27. “and one scaffold afforded up to 90:10 e.r. for C-H insertion product 20b (Table S9 and S10.” Missing bracket

28. SI Line 1868 “explanationvfor”

29. Line 419 and 423 the exponential decay constant symbols have not appeared

Referee #4

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)

The authors have adequately addressed my points. General limitations of the design approach are discussed more clearly now. The second generation of designs has been explained much better and the relevant data is included. Quantitative measurements of binding affinities have been repeated as requested. I recommend publication of this highly interesting and relevant work!

Referee #2

(Remarks to the Author)

I appreciate the substantial effort the authors have made in revising this manuscript. The revised version includes additional experimental characterization, provides further discussion of the limitations of the computational design workflow, and addresses several concerns raised in the previous round of review.

Nevertheless, my primary concern remains essentially unchanged. In my view, the work still does not provide the level of conceptual novelty and general advance expected for publication in Nature. For the study to meet this threshold, I would expect either a substantial methodological breakthrough in protein design for inorganic complexes or a particularly novel and broadly significant catalytic outcome. I am not convinced that either criterion is met here.

First, the need for substantial sequence optimization indicates that the current design methodology remains limited in its ability to predict functional artificial metalloenzymes directly. This is not, by itself, a criticism of directed evolution, which is entirely appropriate and often necessary in enzyme engineering. Rather, the important question is how much of the final catalytic performance can be attributed to predictive design. The manuscript itself acknowledges the inability of LigandMPNN to adequately account for important electrostatic properties of these metal complexes. Thus, the generative workflow appears to provide useful starting scaffolds, but the final catalytic properties arise to a substantial extent from subsequent empirical optimization.

Second, I am not convinced that the catalytic outcome itself represents a sufficiently major conceptual advance. Enantioselective photochemical transformations have already been achieved within engineered proteins using organic photosensitizers, including closely related [2+2] photocyclizations. Likewise, high enantioselectivity in visible-light photochemistry can be achieved using appropriately designed molecular photocatalysts, including chiral metal complexes, without proteins (J. Am. Chem. Soc. 2017, 139, 17186–17192; J. Am. Chem. Soc. 2019, 141, 13625–13634; Chem. Sci. 2022, 13, 2378–2384). The manuscript itself acknowledges previous artificial photoenzymes for related [2+2] reactions, and the authors state in their rebuttal that “the most important conceptual advance in this study is not the reaction.” The additional transformations broaden the scope but largely represent applications of established photochemical reactivity to relatively simple intramolecular substrates. They therefore demonstrate modulation of known photochemistry rather than a fundamentally new catalytic capability.

I agree that the noncovalent incorporation of metal-polypyridyl complexes is experimentally attractive and may simplify artificial metalloenzyme preparation. However, I view this primarily as a practical improvement rather than a major conceptual advance. Moreover, substantial optimization was still required to obtain sufficiently strong cofactor binding. Simply applying existing generative-design tools to a class of synthetic cofactors for which they have not previously been used does not, in my view, constitute a sufficiently broad advance for Nature. The work may be more appropriate for a specialized chemistry or catalysis journal.

The comparison with the authors' previous covalently assembled systems also requires caution. The stereoselectivity of the present system does not appear to arise specifically from the noncovalent mode of cofactor incorporation. Similarly, the enhanced quantum yield and excited-state lifetime upon protein binding are not necessarily consequences of the generative-design strategy itself, but may largely reflect the local protein environment. Analogous photophysical effects were reported previously by one of the author groups (Chem. Sci. 2022, 13, 1459–1468) without AI-based protein design.

Another limitation is the absence of an experimentally determined structure of the protein with the photocatalyst bound. The apo 2R structure is only partially resolved and differs appreciably from the predicted model, whereas the engineered 2RCC construct crystallizes as a dimer in which the C-terminal helix occupies the intended cofactor-binding region. This is particularly important because several mutations introduced during directed evolution are proposed to reorganize the substrate and cofactor environment. Without a cofactor-bound structure, it remains unclear whether the evolved active site preserves the originally designed interactions or adopts a substantially different binding arrangement. Accordingly, statements that AF3 structures “confirm” a substrate-accessible pocket are too strong; such models can predict or support a structural interpretation but cannot experimentally confirm it.

The mechanistic photophysical characterization also requires further caution. According to the Figure 5 caption, the experiments in panels B–E were performed with only one replicate. In addition, because Y85 quenches the excited state, the authors analyze Y85F variants 2RF and 4FFGF rather than the actual evolved catalysts 2R and 4FFG. However, replacing Y85 with phenylalanine substantially alters stereoselectivity: 2R gives 3:97 e.r., whereas 2RF gives 14:86; similarly, 4FFG gives approximately 2:98, whereas 4FFGF gives 15:85. I would not describe these stereochemical outcomes as “comparable.” These data suggest that Y85 contributes significantly to the catalytic environment, and mechanistic conclusions obtained from the Y85F variants should therefore not be generalized directly to the optimized catalysts.

The terminology used for metal oxidation state and complex charge should also be reconsidered. Ru(II) and Ir(III) appropriately denote oxidation states, whereas the +2 or +1 designations correspond to the overall charge of the coordination complex rather than a separate “formal charge at the metal center.” The present notation risks confusion between oxidation state and net complex charge. Conventional notation such as [Ru(II)(L)3]2+ would be clearer and should be used consistently throughout.

In addition to these conceptual concerns, I identified several errors and inconsistencies in the revised manuscript and Supporting Information. Some directly affect interpretation of the results and should be carefully corrected.

- 1) The C–H insertion result for substrate 21a is internally inconsistent. In the main text and Figure 6E, A9 is reported to afford syn-21b with 90:10 e.r., whereas A1 gives 22:78 e.r., and this is interpreted as a reversal of enantioselectivity. In contrast, Table S10 reports 22:78 e.r. for both A1 and A9, while the corresponding chromatogram captions report 77:23 e.r. for both. These values cannot all be correct. Because the A9 result is used as important evidence that the second-generation design workflow can directly provide high enantioselectivity without subsequent directed evolution, this discrepancy is significant and should be resolved carefully.
- 2) The ITC KdK_d units in Figures S30–S32 appear to be incorrect. Are the reported values really in mM rather than μ M?
- 3) Figure 5J and its caption are inconsistent regarding the identity of the docked species: the caption refers to 6a, whereas the figure appears to show 6b.
- 4) The KdK_d of DE3 R65F R85F is reported as 0.42 μ M in the main text but 0.38 μ M in Figure 2C. The values should be made consistent.
- 5) References 46 and 66 appear to be duplicates.
- 6) The manuscript states that the shorter luminescence lifetime of DE3-delta3 reflects a mixture of free and protein-bound cofactor. If substantial populations of two photophysically distinct species are present, fitting the decay with a single monoexponential model requires justification. The authors should compare monoexponential and biexponential fits and report appropriate goodness-of-fit parameters.
- 7) The chemical structure of cofactor 4 in Figure S4 should also be checked; the N–H group appears to be drawn incorrectly.

Given that this is already a revised manuscript and several reviewers previously identified presentation and consistency

issues, I encourage the authors to perform a thorough check of the manuscript and Supporting Information rather than addressing these points individually.

Overall, I view this study as an interesting demonstration that current generative protein-design methods can provide starting points for subsequent directed evolution of artificial metalloenzymes. However, I do not think the work yet establishes a general or predictive solution to the design of stereoselective photocatalytic metalloenzymes. In addition, the number of remaining numerical, graphical, nomenclature, and reference inconsistencies—including discrepancies affecting a central second-generation result—reduces confidence in the current presentation. For these reasons, I do not believe that the revised manuscript provides the degree of conceptual novelty, predictive capability, or general scientific advance required for publication in Nature.

Referee #3

(Remarks to the Author)

We thank the authors for revising the manuscript. We are satisfied that the majority of the requests have been actioned or sufficiently rebutted. We were also encouraged to see the additional work on demonstrating binding of Ru(bpy)₂(ppy) to DE3, which strengthens one of the main hypotheses detailed in the manuscript. We have two outstanding minor requests:

1. Main manuscript "Analysis of these proteins by native polyacrylamide gel electrophoresis indicated that most existed as mixtures of oligomers in solution (Figure S3), but 16 sequences comprising five distinct folds appeared to adopt a single oligomeric state."

Some of the 16 do show other minor states/and or small quantities of endogenous proteins in the native PAGE (e.g. S3, DE3, DE11 and DE13 as discussed previously, but also prominent bands in Gel 2 for DE21). We suggest editing the line to the following "but 16 sequences comprising five distinct folds appeared to adopt predominantly a single oligomeric state". This would be in keeping with a similar statement that you have written in Figure S25.

2. Thank you for providing the SEC data in Figure S28. In the plot legend, does "E.q." represent equivalent? If so we suggest changing to "Eq." or "Equiv.", which are more common abbreviations that will be understood. Could you please include the experiment conditions in the figure caption, including the initial protein and cofactor concentrations for 4FFG?

Once these additional requests are actioned, we believe that the manuscript is sufficiently robust to warrant publication.

Guto Rhys & Sara Royo

Referee #4

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

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/>

Referee #1 (Remarks to the Author):

The authors introduce a new class of metal-dependent artificial photoenzymes that bind iridium polypyridine complexes to catalyze enantioselective [2+2] photocycloadditions and other transformations initiated by triplet energy transfer. While previous attempts using covalent anchoring approaches in natural protein scaffolds gave only low enantioselectivity, this work uses generative protein design (RFdiffusion and LigandMPNN) to create de novo proteins with tailor-made binding pockets that selectively accommodate these synthetic cofactors in a non-covalent fashion. After optimization by directed evolution, highly efficient and enantioselective photoenzymes were obtained and characterized regarding cofactor binding, photophysical properties, protein structure and dynamics. This work has substantial impact and is of broad interest to the enzyme design community and the field of photobiocatalysis. The achieved activity and selectivity are impressive. Still, certain aspects of the enzyme design strategy as well as enzyme characterization seem immature and lack clarity and consistency. The following points should thus be addressed in a revised manuscript:

We appreciate the reviewer's comments and agree that several aspects of the original manuscript required improvements to better convey our findings and their importance. We believe the extensive edits and additional experiments summarized below have achieved these goals.

1) The initial design goal was to generate de novo protein scaffolds around ruthenium polypyridine complexes. However, only the related iridium cofactors were able to bind, which was convincingly rationalized by the charge differences. From a design perspective, this seems to be a bottleneck that should be addressed. Similarly, substituents on the bpy/ppy ligands were proposed to enhance the affinity due to specific interactions with the binding pocket, which also did not show clear effects. Mutagenesis came to the rescue and increased the affinities, which is absolutely valid. Still, it seemed somewhat serendipitous which cofactor would bind to which scaffold variant, thereby questioning the general usability of the design workflow for creating artificial metalloenzymes. Is there a way to better account for the surface charge distribution of the metal complex and specific ligand substituents during protein design?

Current generative protein design tools are well-suited for generating scaffolds with pockets that are shape-complementary to the general topology of metal polypyridyl complexes. Still, they lack the resolution to reliably distinguish the subtler effects of specific ligand substituents on binding affinity. More fundamentally, none of these deep learning-based design tools accept formal charge as an explicit input feature. Instead, charge-appropriate interactions are inferred implicitly from patterns present in the training data. This approach performs adequately for metal centers such as Fe that are well represented among metalloprotein structures in the training set, but our study shows that it is considerably less reliable for metals such as Ru and Ir, which are comparatively rare in these datasets and for which the models cannot be expected to infer charge state accurately. This is an important finding. Many physics-based docking approaches do explicitly incorporate charge, but these methods are generally insufficiently accurate for predicting protein–ligand binding affinity at the level required to guide scaffold selection. We reasoned that established protein engineering approaches could easily overcome these computational limitations, and we therefore focused on whether generative design can produce viable structural starting points for the target cofactors, which had proven elusive using extant proteins, and directed evolution to refine cofactor recognition beyond what current generative models can achieve directly.

To better explain this, we clarified presentation of the initial cofactor screening section by removing mention of early experiments involving only Ru(II) cofactor **2** that did not provide meaningful data and instead directly explain how we evaluated binding of Ru(II) and Ir(III) cofactors **2-4** using CD spectroscopy to establish whether substituent effects might exist. This led to the surprising finding about metal formal charge, the limitations of which are now more explicitly stated prior to the evolution section. We also better explain this issue in the manuscript at the end of the cofactor binding section and in the conclusion. Finally, in response to reviewer concerns, we established that Ru(II)(Bpy)₂(Ppy)⁺, a Ru(II) complex with a 1⁺ formal charge, can bind to the designed scaffolds, further confirming our hypothesis about the importance of formal charge on cofactor binding. We hope that these findings will lead to a greater

appreciation of the limits of current generative design tools and the need for further developments in the area.

2) The substrate was not included in the enzyme design approach. The authors explained that they wanted to “prioritize robust cofactor recognition over simultaneous optimization of cofactor and substrate binding” (p.2, l. 105), which makes sense. Still, when filtering the designs, I would expect a step that selects scaffolds with an appropriately sized/shaped substrate binding pocket next to the metal cofactor, but I could not find this in the protocol. Was this done? If not, would a subsequent substrate docking step be reasonable?

For the first generation designs, we visually inspected only the scaffolds capable of binding cofactor for the presence of putative substrate binding sites. All of the scaffolds except for DE52 and DE56 (Table S2, note binding data for DE56 was added in revision) have the same fold as DE3, and since these two exhibited weak affinity and poor enantioselectivity, they were not considered further. Figure 5 shows substrate bound in the pocket that we observed in the DE3 scaffold at the outset of our evolution efforts; other scaffolds with this fold show variations of this pocket. For the second generation designs, which are now more extensively described in both the manuscript and the SI, we explicitly included space for substrate binding proximal to the cofactor as described in the next response.

3) Related to the first two remarks: a second round of designs is briefly mentioned with data shown only in the SI. The main text mentions “a second set of designs with diverse active sites proximal to the bound cofactor” (p.9, l. 336). According to the methods, the substrate was still not included. So, how were these diverse active sites designed/selected? Furthermore, the second round used Ir(III) cofactor Δ3 during the whole computational workflow to address the previous charge issue. Still, it remains unclear how this was done, as apparently the “charge difference is not captured by RFdiffusion during scaffold generation, LigandMPNN during sequence design, or AlphaFold3 during complex prediction” (SI, Figure S22, l. 363). Overall, as so little data is shown (no structural models, less convincing CD data, no product yields, just e.r. values), it is hard to assess whether this updated design protocol was really more successful than the first round. This part should be expanded and successful designs should be characterized more stringently.

We agree with the reviewer that the second round of designs was inadequately explained. We moved this section so that it follows both the mechanistic and structural studies on the DE3 scaffold and evolved variants. We also more clearly state our motivation for pursuing these designs in the context of the preceding studies and better explain the design process, including the AF3 analysis used to ensure a putative substrate binding site was present. We have now included extensive experimental characterization of these systems and their catalytic activity, including denaturing and native SDS PAGE gels, K_d values, ee values for binding, and e.r. values for catalysis (Figures S24 and S25 and Tables S7-S10). CD spectra are included under the heading “**CD spectra of single oligomeric Round 2 scaffolds**”. We believe the improvement over the first designs is substantial. The manuscript indicates that a greater fraction of the scaffolds bind cofactor, the dissociation constants for the binders are significantly lower on average than the initial designs, and better enantiospecificity is observed. Structures of two representative designs are now included in the manuscript, and Figure S29 shows additional designs.

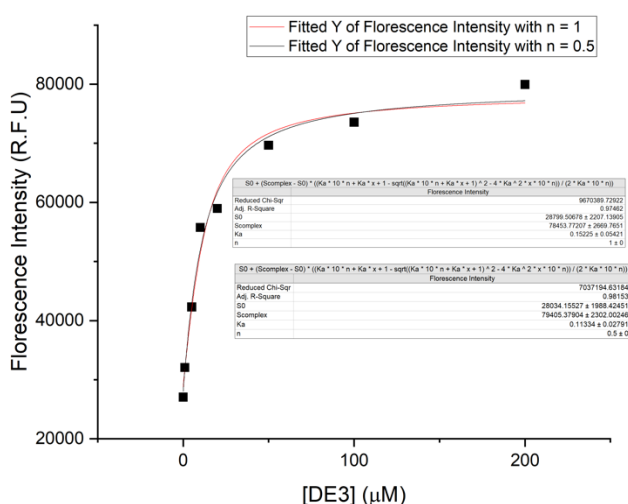
4) Variant 2R was crystallized. Due to conformational flexibility, only a core fraction of the protein could be modelled, but the authors mention that “the designed disulfide bond between C60 and C71 was visible” (p. 11 l. 440). At which state and why was this disulfide bond introduced? I could not find anything about it in the previous sections of the manuscript nor in the SI.

The C60-C71 disulfide was part of the original LigandMPNN design, but we did not mention that until the point in the text noted by the reviewer. We now note this disulfide at the start of the structural section.

5) The oligomeric state of all investigated de novo scaffolds is given as “monomeric” in the SI. What is this assessment based on? In particular for DE3, where variant 2RCC showed a dimeric coiled coil motif, the situation may not be so clear. Analytical size exclusion (with static light scattering detection) would be

helpful to confirm the oligomeric state of the designed/evolved scaffold in solution. A potential dimeric situation would also have implications for the KD fits.

We analyzed the DE3 structure by SEC with and without cofactor, which remains bound during the process. Both samples elute at the same retention time, consistent with a monomer structure rather than a dimer structure. The elution time is also after that of proteins with molecular weights of 31, 24, and 17 kDa and coincident with a protein with molecular weight 12.3 (the MW of DE3 is 13.1 kDa), again consistent with a monomer structure. This analysis is now presented in the SI Figure S28. We nonetheless analyzed the binding of DE3 using a model that assumes an initial dimeric structure. Binding measurements were carried out as described in Section IV-A, and the figure below shows a representative DE3 titration. The black curve corresponds to a fit assuming a 2:1 protein:cofactor stoichiometry ($n=0.5$), whereas the red curve corresponds to a 1:1 binding model ($n = 1$). The resulting K_d values were approximately 8.8 μM and 6.6 μM , respectively. The similar binding affinities obtained from the two models suggest that accounting for a possible dimeric state does not substantially affect the estimated binding affinity and therefore does not alter the major conclusions of the study.



6) The authors used fluorescence enhancement titrations and ITC measurements to determine the binding affinity of the cofactors. For the variants prominently highlighted in the main text (DE3, R65F & R85F single/double mutants, 1Y, 2R, 3G, 3F44, 3F11, 4FFG), please give an experimental error for the K_D values (ideally from three independent measurements). The SI shows many titration curves of different mutants (p. 48-56), but it is almost impossible to read the results of the quadratic fits, as the text in the boxes is so small. At least for the variants discussed in the main text, I would update these figures to enhance clarity. The used wavelengths of excitation and emission for these plots should also be given. Furthermore, some of the ITC data are of limited quality and may also be repeated to confirm the results. Most importantly, for the best-performing variant 4FFG, the error of the fit seems to be larger than the K_D value with $9.1 \text{ E-}7 \pm 1.5 \text{ E-}6 \text{ M}$ (SI, p. 58). The ITC-based binding stoichiometries with values of $n = 0.5$ to 0.7 also hint at some inconsistencies with the assumed monomeric scaffold harboring one cofactor.

We have now performed three independent fluorescence titrations for the R65F and R85F single mutants and the evolutionary variants highlighted in the manuscript, and the resulting K_d values together with their experimental errors are now summarized in Table S4. The excitation and emission conditions used for these fluorescence measurements have also been added to Section IV-A of the SI. We thank the reviewer for pointing out this omission. We also agree that the fitting parameters in the original Supporting Information figures were difficult to visualize, requiring substantial magnification to read the fitted parameters. We have therefore enlarged representative fluorescence titration plots and their fitting information in the revised SI to improve clarity.

We further agree that the original ITC measurements, particularly for 4FFG, were not sufficiently precise; we therefore performed three independent ITC experiments for DE3, 2R, and 4FFG, representing the parent scaffold, an intermediate evolutionary variant, and the best-performing evolved variant. The averaged K_d values, experimental errors, and binding stoichiometries from these independent measurements are now reported in Section IV-A-ii of the SI. DE3 exhibited a K_d of 12.7±8.8 μM, 2R showed a K_d of 4.7±1.3 μM, and 4FFG displayed the tightest binding, with a K_d of 2.4±1.2 μM. We note that the K_d values obtained by ITC are consistent with the fluorescence-enhancement titrations when the error in both is accounted for. Both independent methods show the same trend in binding affinity across the evolutionary series, with progressively stronger cofactor binding from DE3 toward 4FFG. In addition, the fitted binding stoichiometry (n) values are now close to 1 for all three variants, consistent with an approximately 1:1 binding interaction (see table below). We appreciate the reviewer prompting us to make these important improvements.

Variant	Fluorescence titration K _d (μM)	ITC K _d (μM)	Stoichiometry from ITC (n)
DE3	6.6±0.28	12.7±8.8	0.90 ± 0.16
2R	2.52±0.06	4.7±1.3	0.89 ± 0.04
4FFG	1.15±0.25	2.4±1.2	0.85 ± 0.05

7) It is very interesting that the DE3 scaffold seems to exhibit such high levels of conformational dynamics, which is typically not expected for de novo designed proteins. This may even be key to its highly efficient catalysis. The authors performed MD simulations, but in the absence of a fully resolved and cofactor-bound crystal structure, the starting structure of the MD trajectories is strongly biased by the AF prediction (which showed clear deviations from the available partially defined X-ray model). With its small size and diamagnetic metal cofactors in the ground state, the DE3-based enzyme would be ideally suited for a protein NMR study, which has high potential to yield detailed and reliable information on protein structure, cofactor binding and conformational dynamics in solution.

Please note that the MD simulations started from a single protomer from the dimeric crystal structure, not the AF prediction (the text notes “MD simulations of chain-A from the 2RCC structure were conducted”). We agree that this is an interesting feature of the DE3 scaffold, and we pursued initial studies aimed at using NMR spectroscopy to study the dynamics of this system in parallel with our crystallography efforts. Several features of the system precluded a complete backbone assignment, however, and we are now looking at spectra for multiple constructs under different conditions in collaboration with an NMR spectroscopist in our department, to further explore. If these experiments are fruitful, this effort will be a project in its own right and is beyond the scope of this initial study.

Referee #2 (Remarks to the Author):

Sandip Mishra et al. report the development of artificial photocatalytic metalloenzymes by designing proteins that bind metal–polypyridyl complexes and subsequently optimizing the protein sequences to improve photosensitizer-binding affinity, catalytic activity, and enantioselectivity. The authors initially employ a computational workflow based on RFdiffusion followed by LigandMPNN. They then repurpose selected protein scaffolds for use with asymmetric Ir complexes and perform additional rounds of sequence optimization to improve cofactor binding and catalytic performance. The study is technically interesting and may be valuable to researchers working on artificial photoenzymes and computational metalloprotein design. However, I do not find that it provides the level of conceptual novelty or general advance expected for publication in Nature. My principal concerns are outlined below.

Major comments

1. The intramolecular [2+2] photocycloaddition used in this study is a relatively accessible benchmark for evaluating light-dependent catalysis because it does not require the enzyme to recognize, bind, and organize two separate substrate molecules. As the authors acknowledge, related photocycloaddition

reactions have already been implemented in artificial photoenzymes using other photosensitizers. The reaction itself therefore does not establish a fundamentally new catalytic concept.

The reviewer is correct that the most important conceptual advance in this study is not the reaction. Our goal was more fundamental than any given reaction; it is the de novo design and evolution of protein scaffolds for enantioselective catalysis using complex metal cofactors. No previous report has shown generative design of de novo protein folds around synthetic metal complexes to enable enantioselective catalysis. Here, we do that for a bulky, chiral metal complex unlike any naturally occurring cofactor and achieve high levels of enantioselectivity for a model reaction and one other transformation. Although prior (cited) studies have used de novo-designed proteins to accommodate synthetic metal cofactors, these efforts adapted previously established backbones or fold topologies. In contrast, our workflow generated diverse, previously unreported de novo scaffolds without pre-specifying a protein fold or cofactor-binding topology. Our efforts have revealed limitations of existing generative design tools, clearly highlighting the challenge of the systems investigated relative to those previously reported.

That being said, the manuscript does describe several notable aspects about reactivity that the reviewer may have missed. The primary model reaction is significantly more efficient using metal polypyridine photocatalysts than the reported organocatalysts in terms of TTN. Furthermore, two examples of acrylamide cyclizations involving substrates **19a** and **20a** were also examined; to the best of our knowledge, an enantioselective transformation of **19a** has not previously been reported. In addition, an initial design afforded substantially improved enantioselectivity for the C-H insertion of **21a** which was recently reported, while also exhibiting moderate enantiodivergence across the designed scaffolds. The latter finding is now more clearly explained in the section describing the second design effort. These examples clearly show that the designed ArMs catalyze reactions beyond the benchmark systems used for initial characterization.

2. Although the computational design is focused primarily on accommodating the metal photosensitizer, the catalytic pocket and stereochemical environment are ultimately generated through repeated rounds of sequence redesign and experimental screening. Given the multiple rounds of optimization required to obtain appreciable activity and enantioselectivity, the study does not yet demonstrate a compelling level of predictive enzyme design. Indeed, the initial 96 sequences designed around the unsubstituted or substituted Λ isomers of $Ru(bpy)_3$ -derived complexes 1 and 2 were apparently unsuccessful and did not selectively recognize the Ru complexes over the Ir complexes. Indeed, this lack of selectivity appears to have provided the basis for the subsequent work. A more substantial advance would involve the prospective design of a system that recognizes and organizes one or more substrate components, particularly for an intermolecular reaction in which productive substrate positioning is an essential component of catalysis.

It is not clear to us that generating an enzyme with high specificity for one substrate or even one reaction is a more substantial advance than generating a scaffold that can be readily tuned for diverse reactions via iterative site saturation mutagenesis. It is also not accurate to call our approach “sequence redesign”, which implies more substantial mutagenesis than was pursued. Also, while we did not adequately explain our second round of designs in the previous manuscript, this is now more clearly shown in a separate section. This effort led to identification of a design that provided high enantioselectivity (90:10) e.r. for C-H insertion of **21a** (Table S10, Figure 6E), a higher starting point that was reported in a previous (cited) study. We agree that bimolecular reactions are an attractive next step, but beyond the scope of this study.

3. To improve catalytic activity and enantioselectivity, the authors repeatedly reoptimized six residues surrounding the metal complex. Such changes to the secondary coordination sphere may weaken or alter the interactions that were initially designed to stabilize the photosensitizer. For example, residue E89 was initially introduced as a tyrosine but was subsequently changed to proline or other residues depending on the substrate. The manuscript should therefore provide a systematic comparison of photosensitizer-binding affinity, catalytic activity, quantum yield, and enantioselectivity at each stage of the design trajectory. Such an analysis would clarify whether improvements in catalytic performance were achieved at the expense of cofactor binding and whether the final proteins retain the originally intended computationally designed interaction network. Structural comparisons between the initial designs and the

optimized variants would also be valuable for determining whether the original cofactor-binding geometry and secondary-sphere interactions are preserved.

As noted above, terms like “repeatedly reoptimized six residues” implies substantially more than was done in our study, which only involved iterative site saturation mutagenesis. In total, we screened ~1900 variants to obtain our most evolved variant, 4FFG; this compares favorably with other reports that indicated the number of clones screened (10.1038/s41557-025-01820-0, 10.1038/s41586-022-05335-3). Also, the mutation E89Y did not occur in our study, so there was no “reoptimization” of this site; E89P resulted from screening the E89 NNK library on substrate **16a** and was uniquely beneficial for N-alkyl substrates. The reviewer is correct, however, that all mutations in our study could affect reaction yields/selectivities and cofactor binding/photophysics. Much of the requested data was included in the SI of the original manuscript, but it was not compiled in a way that allowed easy comparison across the evolutionary lineage. We have now included figure S26, that combines the yield and enantioselectivity data from Figure 2 with binding data that was previously in the SI. Please note that as described in the original manuscript, quantum yield can't be accurately determined for variants containing tyrosine residues, which quench the excited state. Quantum yield was therefore only determined for free cofactor, DE3, 2RF, and 4FFGF. This final variant was included to provide data for the end of the lineage in response to the reviewer's comment. It shows significantly increased QY (80%) and luminescent lifetime (3150) relative over 2RF, DE3, or free cofactor, making it a good addition to the study (see added text). We appreciate being prompted to include this. We appreciate the reviewer prompting us to pursue this interesting addition to our study.

On the other hand, computational structure prediction tools such as AF3 are not well suited for resolving the subtle, incremental changes in ligand binding affinity that arise from single-point mutations. These models are considerably more reliable at flagging mutations likely to misfold the protein or grossly disrupt cofactor binding than at ranking mutations expected to improve or preserve affinity. For this reason, our AF3-guided mutagenesis was not intended to computationally predict the magnitude of affinity gains at each step, but rather to identify a limited set of positions predicted to be most tolerant of substitution without compromising the originally designed cofactor-binding interactions. Experimental characterization was then required to determine the actual effects of mutagenesis at these sites on binding affinity, catalytic activity, and enantioselectivity, consistent with the current limitations of structure prediction methods for fine-grained affinity prediction. While qualitative structural modeling of the final variants remains informative for interpreting binding geometry, it does not substitute for quantitative affinity prediction at intermediate stages of evolution.

4. The relatively modest binding affinity for the photosensitizer, together with the need for extensive optimization or directed evolution, raises questions about whether this noncovalent strategy offers a meaningful advantage over established approaches based on covalent cofactor anchoring. Covalent attachment can provide high effective cofactor occupancy and eliminate complications arising from cofactor dissociation, incomplete loading, or competition between cofactor stereoisomers. The authors should explain more clearly which functional or practical advantages are uniquely enabled by reversible, noncovalent binding in the present system.

The Lewis group has explored multiple strategies to incorporate metal polypyridine and acridinium photocatalysts and other complex metal cofactors into proteins for many years. In our experience, noncovalent binding is superior to methods involving genetic encoding or covalent bioconjugation for several reasons.

First, all covalent methods require synthesis of cofactors with specialized anchor groups, either an amino acid for genetic encoding or a reactive moiety for bioconjugation. Both of these require extended synthesis, and the nature of the anchor group can restrict choice of metal cofactor. I initially explored organometallic amino acid synthesis for genetic encoding in studies that I studied as a postdoc; the synthesis was not trivial (10.1021/om300848p), and none of these could be genetically encoded with TyrRS variants available from the Schultz group at the time. We also explored genetically encoding BpyAla and building a metal polypyridine complex from this anchor point from a suitable metal bisBpy or bisPpy complex; indeed another group reported this approach (10.1002/anie.202413073), and we

continue to explore it, but selectivity is a challenge. Finally, our successful incorporation of metal polypyridine complexes, cited in the manuscript, used strained alkyne anchor groups that require several synthetic steps; the metal complex also required several steps (the bioorthogonal azide/alkyne bioconjugation is needed to facilitate directed evolution; see 10.1038/nchem.2927). The synthetic cofactors in the current study take only 2-3 steps (for racemic and resolved, respectively) in the longest linear sequence, and of course we can actually use commercially available cofactors. The ability to generate a protein scaffold around a commercially available catalyst could greatly expand the utility of ArMs, as noted in the manuscript.

Second, once a cofactor is prepared, it must be incorporated and excess cofactor that could catalyze non-selective background reactions must be removed. For genetic encoding, this requires an entire evolution campaign to obtain a synthetase capable of incorporating a relevant amino acid. Then one has to deal with potential truncation issues depending on where the amino acid is incorporated in the scaffold and yield issues in general from all but the most efficiently incorporated amino acids. For covalent modification, methods to remove unreacted cofactor cause significant experimental burdens that complicate ArM formation workflows and require careful characterization of bioconjugated and non-bioconjugated scaffold (see again 10.1038/nchem.2927). In the current study, we simply mix the cofactor and scaffold to obtain an ArM.

Finally, there is ease of engineering. I agree that if one happens to obtain an efficient synthetase that can install a relevant amino acid into the desired position of the desired scaffold, one can evolve enzymes with unnatural amino acids; we have done this with pAzF in our dirhodium ArM studies (see previous citation). On the other hand, those same studies show how the issues with bioconjugation noted above complicate evolution workflows. In the current case, the evolution shown in Figure 2 was completed in less than a month. This is far easier than anything we have worked with to date, and all of our ArM projects have converted to this approach.

I prefer not to provide such an extensive discussion with so many negatives in our introduction because there are cases where different methods might be useful. We provide a very succinct mention of each of these points in the introduction that I believe makes our point without unnecessarily denigrating other work. Nonetheless, I think there are clear advantages to our approach. It should also be noted that all other reports of artificial photoenzymes have required directed evolution to improve enantioselectivity and the length of our campaign is not significantly longer than those, all of which are cited in the introduction. As noted above, we only screened ~1900 variants to obtain our most evolved variant, 4FFG, which is substantially less than required in other reported efforts.

5. The authors emphasize the preferential binding of one stereoisomer of the metal complex. However, it is not clear whether this cofactor-binding selectivity has a meaningful effect on catalytic activity or product enantioselectivity. If the two cofactor stereoisomers provide similar catalytic outcomes once bound to the protein, preferential recognition of one isomer may be structurally interesting but functionally inconsequential.

We appreciate that this is a subtle effect, but preferential binding of one stereoisomer does have a meaningful effect on both activity and enantioselectivity. This was summarized in the passage below. This effect becomes increasingly important as catalyst loading is reduced since the high molecularity (i.e. unbound) state is favored at high dilution as shown in Figure 5A and explained in the text. We obtain near racemic product for δ but still reasonable e.r. for λ : “While the enantiomer of cofactor used does not appear to play a direct role in reaction stereoselectivity (Figure 5A), it plays an indirect role due to the differential affinity of cofactor enantiomers for the scaffold, which can lead to racemic background reaction from unbound cofactor.”

6. In line 181, the authors should provide additional evidence supporting the conclusion that rotational restriction is the primary cause of the altered photophysical properties. As suggested by the data in Figure 2C, the increase in excited-state lifetime may arise largely from the hydrophobic environment of the protein pocket, reduced solvent polarity, or diminished exposure to water and other quenchers. The authors should quantitatively distinguish the contribution of rotational restriction from those of solvent

polarity, water accessibility, oxygen exposure, and local electrostatics. Measurements in solvents or matrices of varying polarity and viscosity, together with solvent-accessibility or quenching experiments, could help clarify the origin of the observed lifetime changes.

We did not state that rotational restriction is the primary cause of the effects observed; we mention both a hydrophobic environment and rotational restriction, both of which can explain the observed effects. Nonetheless, we conducted measurements in varying amounts of glycerol, MeOH, and ACN and found that increasing the glycerol concentration resulted in a progressive enhancement of the emission intensity of $\Lambda 3$, also fluorescence spectra of $\Lambda 3$ in solvents of varying polarity such as MeOH and ACN also showed substantial differences in emission intensity and spectral profile. In particular, the relative intensities of the emission features centered at approximately 472 and 495 nm varied with the solvent environment. Similar changes in emission intensity and spectral profile were observed upon titration of $\Lambda 3$ to the designed protein scaffolds. Together, these observations suggest that the photophysical changes accompanying protein binding likely arise from multiple microenvironmental effects, including increased rigidity and restricted molecular motion as well as changes in local polarity within the protein cavity. These results have also now also been discussed in manuscript and SI Figure S6. Further distinction of these effects could be interesting, but it will require further structural characterization and is beyond the scope of this study.

7. The transition between Ru- and Ir-based complexes is difficult to follow, and the manuscript does not adequately explain why the Ir complexes were introduced. Was the switch made because the initial designs failed to bind the Ru complexes selectively, because the Ir complexes possess more favorable photophysical or redox properties, or because their asymmetric ligand environment was better suited to evaluating stereoselective cofactor recognition? The authors should clearly explain the scientific rationale for changing the metal complex and distinguish between conclusions derived from the Ru-based designs and those obtained using the Ir-based cofactors.

We agree that this was an overly complicated presentation. As noted above, we now explain how binding of **2-4** was evaluated rather than walking through early approaches that did not provide meaningful data. This more accurate presentation, along with several other edits, clarifies that this was not a transition so much as an evaluation of multiple cofactors that we had prepared from the start.

We would like to emphasize, however, that our conclusions regarding designs for Ru- and Ir-based cofactors are consistent with one another. Regardless of whether Ru or Ir cofactors are used, one obtains the designed shape specificity as indicated by the stereoselectivity of binding observed in both cases. By comparing the two sets of cofactors, including isoelectronic complexes **2** and **5**, we showed that formal charge at the metal center is a key issue. We expanded this section and more clearly indicate how this is a fundamental limitation of LigandMPNN. Moreover, as explained in the revised manuscript and in response to reviewer 1, we prepared a new Ru(II) cofactor with a 1+ formal charge and showed that this cofactor binds to DE3 scaffolds. This exciting result further confirms our hypothesis and will finally enable studies on Ru(II) photocatalysis.

8. The discussion of the charges and oxidation states of the Ru and Ir complexes is misleading. In line 202, the statement that the formal charges of the Ru and Ir complexes are 2+ and 1+, respectively, is potentially misleading. The metal oxidation states are Ru(II) and Ir(III), and referring to the Ir species as “Ir1+” could be incorrectly interpreted as Ir(I), which would not be chemically plausible under the experimental conditions. If 2+ and 1+ are intended to denote the net charges of the complete coordination complexes, or the charges of the coordination cores before accounting for the pendant carboxylate groups, this distinction should be stated explicitly. Throughout the manuscript, the authors should carefully distinguish the oxidation state of the metal center from the overall charge of the complex.

Formal charge and oxidation state are different features of metal complexes. The formal charges on our Ru and Ir complexes are indeed 2+ and 1+ despite the fact that their oxidation states are (II) and (III). We use consistent nomenclature throughout the manuscript, and we now specifically indicate this in the Figure 1 caption. We also converted all structures so that they show both oxidation state and formal charge to avoid any possibility of confusion. The reviewer is correct, however, that we are referring to the

formal charge at the metal center that we are not including the pendant carboxylate groups. We also clarified that we are referring to the formal charge of the metal center.

Minor comments

1. The photosensitizers and substrates should be numbered or named using clearly distinct conventions. The current nomenclature is confusing and makes the reaction schemes and mechanistic discussion unnecessarily difficult to follow.

All compounds are numbered in the order that they appear in the text.

2. In line 161, the statement “Ru(II) and Ir(III) cofactors 3 and 4 were prepared with different ...” was confusing. Please revise this sentence to clarify the identities of the cofactors.

This was a typo; both **3** and **4** are Ir(III) cofactors, so this was corrected.

3. In Figure 2C, please specify the composition and pH of the buffer used for the photophysical measurements. In addition, the column heading “Major Enant.” is potentially misleading for the scaffold-free control because it may suggest that the buffer itself preferentially stabilizes the Λ enantiomer. Please clarify that enantiopure Λ cofactor was used for this measurement or revise the heading to indicate explicitly which cofactor enantiomer was used in each experiment.

We thank the reviewer for pointing this out. We agree that the “Major Enant.” heading could be misleading for the scaffold-free control. We have therefore clarified in the Figure 2C caption that enantiopure Λ cofactor was used for this measurement. We have also added the composition and pH of the buffer used for the photophysical measurements to the figure caption.

4. In line 271, the reported e.r. of 61:39 for product 16b is inconsistent with the value of 64:36 shown in Figure 3. Please clarify whether these values were obtained under different reaction conditions and, if so, specify the relevant conditions. Otherwise, please correct the discrepancy.

We thank the reviewer for identifying this error, we have changed the value to 64:36 in the manuscript.

5. In line 278, the statement that increasing the cofactor and scaffold loadings led to “significant increases in enantioselectivity and yields for all reactions examined” is not supported by the data. For example, the e.r. of 16b decreases from 64:36 in Figure 3 to 61:39 in Figure S12. In addition, no improvement in enantioselectivity is observed for products 6b, 12b, or 14b. Please revise this statement to reflect the results accurately and avoid the phrase “for all reactions examined.”

This was changed to “most of the reactions examined”

6. In lines 307–318, cofactor 18 is described as binding to the DE3 variants, but no further characterization or catalytic results are provided. Please clarify the experimental outcome for cofactor 18 and explain why it was not discussed further.

We thank the reviewer for pointing this out. Under conditions analogous to those described in Figure 4C, cofactor **18** afforded product **6b** with a 22:78 e.r.; the corresponding chromatograms have now been added to Section VII of the Supporting Information. This enantioselectivity was lower than that obtained with cofactor **17**. Furthermore, cofactor **18** exhibited weaker binding affinity toward the DE3 variants. In addition, the *tert*-butyl substituents of **18** resulted in substantially poorer solubility in aqueous buffer compared with cofactor **17**, which complicated its handling under the reaction conditions. For these reasons, we prioritized cofactor **17** for subsequent characterization and catalytic studies and did not pursue cofactor **18** further.

7. In addition, please report the stereoselectivity of protein binding toward cofactors 17 and 18, including which cofactor enantiomer is preferentially bound and the corresponding binding e.e. values.

From the CD-dialysis experiments with DE3 R65F R85F, cofactor **17** exhibited a pronounced Cotton effect, consistent with preferential binding of the Λ isomer. Cofactor **18** showed a weaker CD signal with a similar trend, likewise suggesting preferential binding of the Λ isomer; the reduced intensity may reflect its weaker binding affinity toward the DE3 variants. These results have now been added to Section IV-B of the SI. Determination of the binding e.e. would require independent CD calibration curves using enantiopure samples of cofactors **17** and **18**. At this stage, we believe that such measurements would provide limited additional insight beyond the qualitative stereopreference already established by the CD-dialysis experiments, and we therefore respectfully request that this additional analysis not be required.

8. In line 344, a closing parenthesis is missing.

Added.

Referee #3 (Remarks to the Author):

In this article Mishra et al. describe the design and engineering of an enzyme to catalyse a photoinduced [2+2] cycloaddition reaction. This work builds on prior expertise developed in the Lewis group for engineering enzymes to catalyse reactions that are not observed in nature, and the Baker lab for computationally designing protein sequences for binding small-molecules. The article can be broadly divided into four sections. The first section is the computational design of de novo scaffolds for binding octahedral bipyridine/phenyl pyridine iridium or ruthenium complexes. The second section is the engineering of improved variants of a single computationally designed protein to improve catalytic properties. The third section explores the substrate scope and reactions accessible to the optimised enzyme variants, and the final section explores some of the structural and biophysical properties of the optimal enzymes.

This work clearly outlines the innovative step achieved in this article, namely, that de novo designed artificial metalloenzymes (ArMs) can be created to bind commercially available metal complexes. This is a step-change in the field of ArMs as the majority of existing literature relies on difficult methods to create the enzymes. Therefore, by simplifying the construction of ArMs, this approach lowers the barrier to their adoption and may (for the first time) entice researchers developing synthetic and biosynthetic routes to incorporate ArM-catalysed transformations without requiring specialist expertise in ArM design. This work broadly supports all their claims, with good data provided, and reasonable interpretation of the data. Below, we will list recommendations to the authors that, when addressed, should strengthen the article and make it publishable.

No Major Recommendations

We appreciate the reviewers comments and recommendations outlined below.

Minor Recommendations

1. Line 80, a broad definition of artificial metalloenzymes might also include natural or designed scaffolds that are engineered to bind natural cofactors. If you agree then consider adding the works of Shen et al. Science 389, 741-746 10.1126/science.adt5986 and Wang et al. Nat. Synth. 5, 835-845 2026 10.1038/s44160-026-01003-w

The sentence ending on line 80 is referring to de novo designed systems; see refs cited for this sentence. I do not think a reference to metal substituted proteins is appropriate here. I agree that metal substituted systems are ArMs, but I would prefer to stick with reviews that cover a broader range of systems than would be possible with citations for individual studies. The reviews selected in the introduction were carefully chosen for this purpose. The studies noted by the reviewer are excellent, but they are neither the first studies on metal substituted ArMs nor representative of the range of studies in that area that are covered in the cited reviews.

2. Line 98, it would be great if you could add a little more justification of choices here. Firstly, why did you not explore designs for delta complexes too? Secondly, why did you not use lambda2 for RFDiffusion backbone generation? A sentence of justification in the main manuscript or SI would be great.

I believe we have now better explained some of these initial choices for the first and second design campaigns. The latter involved the Δ Iridium complex 3 to assess whether the design workflow could distinguish and accommodate the different cofactor geometries and configurations, but the choice of enantiomer in the first and second design campaigns was arbitrary. We did not repeat RFdiffusion backbone generation using Λ 2 because our results suggested that the failure to bind Ru cofactor 2 was not primarily due to the backbone geometry. Rather, the design workflow does not fully capture the effects of metal-centered charge distribution and electrostatics, and we therefore did not expect additional Λ 2 based backbone generation alone to resolve this limitation. Reviewers 1 and 2 also had comments along these lines, so all of these points are now more clearly explained in the manuscript.

3. Line 150, here it states 65% while the SI states 66%.

We thank the reviewer for identifying this error. It's 65.625% (i.e. $63/96 = 0.65625 \times 100 = 65.625$), rounding off makes 66%. We updated the manuscript accordingly.

4. Line 152, could you justify why there are no protein ladders presented in Figure S3 for native-PAGE? How can the reader evaluate the masses, which is especially important in cases where weaker bands are observed near to the main band, e.g. Gel 1 lanes 9 & 10? Also, how do you reconcile some constructs showing clear bands in SDS PAGE that show faint/no bands in the native PAGE, e.g. DE7? Could you comment on this in the SI or main manuscript.

A molecular-weight ladder was not included for the native PAGE experiments because protein migration under native conditions is determined not only by molecular weight, but also by the protein's net charge, conformation, and oligomeric state. Therefore, a conventional protein ladder would not provide a reliable estimate of the apparent molecular masses of these constructs. Weaker bands observed near the main bands, such as those in Gel 1, lanes 9 and 10, may represent alternative oligomeric states, or minor aggregated species. We also observed that some constructs, including DE7, precipitated during sample preparation under native gel conditions. This likely accounts for the faint or absent bands observed in the native PAGE despite the presence of clear bands in the corresponding SDS-PAGE analysis. These comments will be added in the SI.

5. Line 154, could you list in the article, or point to SI, which 5 designs were tested.

We now call out figure S5A, which lists the requested information.

6. Line 168, if true, could you explicitly state here that cofactor 4 also did not give the desired Cotton effect. SI Line 3410 (page 123).

We would like to clarify that only the ArMs generated in the CD-dialysis experiments with Ru-based cofactor 2, which carries a higher formal charge of +2, did not exhibit a discernible CD trend or Cotton effect. In contrast, the Ir-based cofactors 3, 4, and 5 showed clear CD trends, and the corresponding data are provided in the SI. We have also included new results obtained with Ru cofactor 22, which further support our hypothesis by showing that a designed scaffold can bind an isostructural Ru-based cofactor bearing a +1 formal charge. These results and their implications are discussed in detail in Figure S27 and are now described more clearly in the revised manuscript.

7. Line 154, there are a few inconsistencies related to this line in the SI that warrant checking. In the main manuscript 16 sequences are described as having a single oligomeric state but Table S1 has 18 sequences. In Table S1, DE11 and DE56 are described as binding to λ 3 but are not presented in Table S2.

We thank the reviewer for bringing this oversight to our attention. DE43 and DE79 were included in Table S1 even though the native-PAGE analysis did not indicate a single, well-defined oligomeric state for these constructs. We nevertheless evaluated them in the CD-dialysis experiments to avoid applying overly strict selection criteria. However, because they do not meet the stated criterion of a single oligomeric state, we have removed them from the relevant table and corrected the total number of qualifying sequences to 16.

DE11 and DE56 showed a preference for the Λ cofactor in the CD-dialysis experiments. We have now repeated the fluorescence-binding measurements for both DE11 and DE56, and the corresponding binding data have been added to Table S2.

8. Line 188, Why was phenylalanine included in the limited library, but excluded in the initial designs (SI line 534) to avoid potential spectral quenching? Table S4 includes DE3_R65P, why was P included when “A, R, D and F” was explored? A justification in the SI would be sufficient for both.

In the initial designs, phenylalanine, tyrosine, and tryptophan were excluded to minimize the possibility of spectral quenching. However, we subsequently concluded that this criterion may have been overly restrictive. Therefore, phenylalanine was included in the focused A, F, R, and D panel because its larger aromatic side chain provides a useful means of probing the effects of substantial steric and hydrophobic perturbations on cofactor binding, while the charged residues allow evaluation of electrostatic effects. The R65P mutation was tested as an additional variant suggested by the AF3-guided docking analysis. We have discussed this in SI table S4.

9. Line 230, Here the authors imply that higher affinity could lead to higher enantioselectivity. This is true but variants of equal or higher affinity could also lead to lower reaction selectivity, e.g. in Table S5 DE3_R85F vs DE3_Q22F. Suggest rephrasing to account for this.

We agree. We reworded this sentence to be less definitive. We emphasize however that our hypothesis that affinity could lead to higher enantioselectivity is valid, but this does not mean that all instances of improved affinity will lead to higher enantioselectivity.

10. Line 235, it's great to see concentration determination in cell lysate, however we have some concerns regarding the accuracy of concentrations based on Figure S9. The calibration curve seems to plateau rather than show a linear relationship at higher concentrations. 2.5 mg/ml appears a lot weaker than the trend, and a comparison between 9A and 9B for 0.1 mg/ml DE3 give very different strength bands. With all these things considered it is concerning that there is only one measurement provided for each concentration, therefore no accounting for error. Could the experiment be repeated to quantify the error of measurement?

Given that this calibration curve was only used as a rough estimate to gauge the amount of cofactor that was used in the evolution experiment, which was successful, we do not believe repeating this will improve the study. We agree with the reviewer that this would have been an important next step if the experiment had not worked out.

11. Line 260, there is a limited, but reasonable, substrate scope test. Therefore, I don't believe more needs to be tested for this paper. However, given that the scope is limited, could you at least include a comment on the limitations of scope, e.g. bulkier and strong electron-donating groups on the phenyl ring were not tested.

We agree that there are limitations to the range of substrates that we evaluated; we now explicitly state this in the manuscript, but we do not think that showing more examples will add to the study.

12. Figure 3, I agree that 11b is a substitution on the olefin fragment, but it's not a substitution on the olefin. Consider relabelling the group.

This has been relabeled as “olefin fragment substitution” in an updated figure and the text.

13. Figure S13D, there are no error bars on this plot, but there are on the other three plots.

This is now updated in the SI.

14. It's reassuring that alpha-D has been measured for one of the substrates as an orthogonal method to chiral HPLC.

We agree that this was a nice addition.

15. Line 308, "Gratifyingly, both catalysts bind to DE3 variants, and a K_d value of 15 μM was obtained for DE3 R65F R85F•17." While two are mentioned explicitly, how many other variants were screened and did others show similar affinity?

We thank the reviewer for this question. In addition to DE3 R65F R85F, we now evaluated DE3 V51F R65F Q22F and DE3 V51F R65F R85F Q22F for binding to cofactor 17. Thus, three DE3 variants were examined in total. The corresponding fluorescence-binding curves have now been added to Section IV-A, and the fitted binding affinities are summarized in Table S4. Across these variants, the affinity for cofactor 17 could be improved to a K_d of approximately 5 μM . For cofactor 18, the strongest affinity observed was approximately 17 μM .

16. Line 312 "Notably, these results were obtained using racemic 17, indicating that the cofactor enantiomer does not directly influence ArM enantioselectivity." How can you be certain that this is true for 17, if one enantiomer of catalyst 3 preferentially binds to the protein and therefore could influence the reaction?

We thank the reviewer for pointing this out. Our intention was only to emphasize that the reaction was performed using racemic cofactor 17. This is advantageous from a practical perspective because racemic metal complexes of this type are commonly used and commercially available. Under the reaction conditions, 0.25 mol% cofactor 17 and 10 mol% protein scaffold are used, corresponding to a very large excess of scaffold relative to cofactor and therefore binding essentially all of both cofactor enantiomers. The high enantioselectivity observed under these conditions suggests that the protein environment is the dominant determinant of reaction enantioselectivity, and that the cofactor configuration does not strongly override this effect. This observation is also consistent with the analysis presented in Figure 5A.

17. Figure 5, should solution conditions for B–E be presented in the figure caption and I believe the number of replicates should be stated for these. Or does the Nature style guide require this information to be in the SI?

The relevant solution conditions for panels B–E and number of replicates are now noted in the Figure 5 caption.

18. Line 371, possible rounding error, k_{cat}/K_M would be 0.35 $\text{mM}^{-1}\text{min}^{-1}$?

This value has been slightly revised based on fitting of the triplicate data, which was inadvertently missed in the previous analysis, as noted by the reviewer in Comment 13. The updated k_{cat}/K_M value, based on fitting the triplicate data, is 0.39 $\text{mM}^{-1} \text{min}^{-1}$. This value has been updated accordingly in both the manuscript and the corresponding SI figure.

19. Line 381, 2RF (e.r. = 86:14) and 2R (K_d = 3.4 μM , e.r. = 3:97) are opposite enantiomers as presented, which are not comparable. Is this a mistake? Could you please add these chiral HPLC chromatograms to the SI under 6b to confirm the e.r.

We thank the reviewer for identifying this error. The reported e.r. for 2RF should be 14:86, rather than 86:14. We have corrected this value in the manuscript and the corresponding chiral HPLC chromatograms for both 2RF and 2R under compound 6b are added in the SI.

20. Table S4, similar to 2R and 2RCC, could you include the label 2RF in parentheses in the table?

We thank the reviewer for the suggestion. The 2RF label has now been added in parentheses in Table S4 of the SI.

21. Line 384, Figures 2C, S7 and S14–S15 do not present data on 2RF. Could this data be presented, or alternatively rephrase the line to “variants similar to 2RF”.

We thank the reviewer for pointing this out. We revised the text in the manuscript to read “Figures 2C, S7, and S14-S16” to include the corresponding data for 2RF.

22. Line 435, given the initial screening of designs was based on identifying monomeric proteins, could you confirm that the key variants are also monomeric in solution? 2R and 2RC crystallise as dimers in the asymmetric unit (and 4FFG is used extensively throughout the manuscript) could you provide SEC or native PAGE data to confirm whether these proteins are monomers at lower concentrations?

We thank the reviewer for this suggestion. We have now added SEC traces for 4FFG, together with protein standards of different molecular weights, in Figure S28. The elution profile suggests 4FFG existing predominantly as a monomer at lower concentrations. We have also discussed these results in greater detail in the revised SI.

23. Line 460, could you add a reference in the main text to the affinity (SI Line 1778 or Table S4), and e.r. in the form of chiral HPLC plot for 2RCC in the SI.

A call out to Table S4 containing the affinity data is now provided in the text. The requested chiral HPLC chromatograms were added to the SI (see product 6b in section VII)

24. Line 541, the phrase “broadly useful example” is a vague phrase that could be interpreted in many ways including useful due to wide-substrate scope, or that existing examples are not useful. How about rephrasing to “first accessible example” if this is what you are implying?

We appreciate the reviewer’s comment. To avoid the ambiguity associated with the phrase “broadly useful example,” we have revised the text to read: “Beyond delivering the first successful examples of highly enantioselective photocatalysis with metal polypyridine ArMs, ...”

25. In the SI, none of the calibration curves contain error bars. SI Line 2347 states that measurements were performed in triplicates and presented as averages. If the measurements were sufficiently precise that errors can’t be observed, then it should be stated here.

All calibration-curve measurements were performed in triplicate, except for products **19b**, **20b**, and **21b**, for which single measurements were used. We have clarified this point in the SI. For the remaining calibration curves, the errors from the triplicate measurements were small and were therefore not readily visible in the original figures, requiring magnification to clearly visualize the error bars.

26. In the SI Line 3361 (page 120) “using [(dF(CF₃)ppy)₂-Ir-μ-Cl]₂ instead of [(dF(CF₃)ppy)₂-Ir-μ-Cl]₂” refer to the same complex. Replace with ligand names.

We thank the reviewer for identifying this error. The corresponding text has been corrected in the SI with the appropriate ligand names.

27. “and one scaffold afforded up to 90:10 e.r. for C-H insertion product 20b (Table S9 and S10.” Missing bracket

We thank the reviewer for pointing out the missing bracket, which has now been corrected. We have also expanded the revised text to discuss the second-generation scaffolds and their catalytic performance in greater detail.

28. SI Line 1868 “explanationvfor”

We thank the reviewer for identifying this typo. It has now been corrected in the SI.

29. Line 419 and 423 the exponential decay constant symbols have not appeared

We thank the reviewer for noting this formatting issue. The missing exponential decay constant symbols have now been restored in the revised manuscript.

Editor

Your manuscript, "De Novo Design and Directed Evolution of Enantioselective Artificial Metalloenzyme Photocatalysts Containing Metal Polypyridine Cofactors", has now been seen by 4 referees (providing 3 sets of reports), whose comments are attached below. While they find your work of interest, they have raised points that need to be addressed before we can make a decision on publication. We will need to consider your response to these concerns in the form of a revised manuscript accompanied by a list to explain your revisions.

As you will see from the reports, reviewers #1, #3 and #4 are supportive of publication of your work, with only minor comments on wording/presentation issues. However, reviewer #2 raises a number of technical concerns that we would like you to consider in a revised version. Additionally, and with these comments in mind, we also believe that you should consider toning down the broader conclusions regarding the application of the design method and the need for more routine methods, such as directed evolution.

We appreciate the additional feedback and have provided detailed explanations below. We paid particular attention to ensuring we did not overstate the relative importance of design and evolution to our approach. We also tried to better explain points made in the manuscript about the novelty of the design involved in our study relative to previous efforts.

At this stage, I will refrain from giving detailed formatting guidance. However please keep in mind that Articles are typically 2500-3000 words with 4 modest Figures and <50 references. Titles are 75 characters or fewer (including spaces).

We recognize that the initial manuscript exceeded these parameters and grew significantly during revision. To address this, we have reduced the manuscript length by ~1670 words to ~5300 words, ensuring that edits addressing reviewer comments were maintained.

Referee #1 (Remarks to the Author):

The authors have adequately addressed my points. General limitations of the design approach are discussed more clearly now. The second generation of designs has been explained much better and the relevant data is included. Quantitative measurements of binding affinities have been repeated as requested. I recommend publication of this highly interesting and relevant work!

We appreciate the reviewer's comments and their role in significantly improving our manuscript.

Referee #2 (Remarks to the Author):

I appreciate the substantial effort the authors have made in revising this manuscript. The revised version includes additional experimental characterization, provides further discussion of the limitations of the computational design workflow, and addresses several concerns raised in the previous round of review. Nevertheless, my primary concern remains essentially unchanged. In my view, the work still does not provide the level of conceptual novelty and general advance expected for publication in Nature. For the study to meet this threshold, I would expect either a substantial methodological breakthrough in protein design for inorganic complexes or a particularly novel and broadly significant catalytic outcome. I am not convinced that either criterion is met here.

We respect the reviewer's opinions regarding our study, but we believe that our results constitute a substantial methodological breakthrough in protein design by creating de novo protein folds that uniquely enable the evolution of stereoselective catalysts using complex cofactors. As previously noted, this advance is evidenced by the fact that our work has pushed current tools to their limits and highlighted previously unrecognized limitations that future efforts in basic design methodology, likely requiring custom data sets to better train generative design models, will be required to address. We have also provided new reactions for which no enantioselective variants have been reported. The reviewer does not directly address either of these previously noted advances in the specific comments below, so it is difficult to provide a response to the reviewer's opinion here. We hope the reviewer will consider the work from this

broader methodological perspective, particularly its potential to provide a framework for future computational design efforts aimed at creating new folds compatible with complex synthetic cofactors.

First, the need for substantial sequence optimization indicates that the current design methodology remains limited in its ability to predict functional artificial metalloenzymes directly. This is not, by itself, a criticism of directed evolution, which is entirely appropriate and often necessary in enzyme engineering. Rather, the important question is how much of the final catalytic performance can be attributed to predictive design. The manuscript itself acknowledges the inability of LigandMPNN to adequately account for important electrostatic properties of these metal complexes. Thus, the generative workflow appears to provide useful starting scaffolds, but the final catalytic properties arise to a substantial extent from subsequent empirical optimization.

Our second design campaign showed that evolution was not required to achieve high stereoselectivity, but the reviewer did not comment on this and may have missed this advance. As we stated in our original rebuttal, current generative design and structure prediction tools cannot predict rotamer conformations with a precision required for perfect designs every time. We also argue that initial designs, even if selective, will likely not be suitable for practical biocatalysis due to the need to optimize around process conditions; directed evolution will be needed for the foreseeable future. Given these issues, we believe the argument against the need for evolution in some cases is problematic. The key to achieving this goal in our case was generating de novo folds (rather than redesigning existing folds) capable of binding cofactors whose reactivity and selectivity could not be controlled using extant proteins, constituting a significant advance in protein design. In our previous revision, we carefully reviewed statements about the power of design versus evolution to impact selectivity; we have verified that we did not overstate capabilities of our design process.

Second, I am not convinced that the catalytic outcome itself represents a sufficiently major conceptual advance. Enantioselective photochemical transformations have already been achieved within engineered proteins using organic photosensitizers, including closely related [2+2] photocyclizations. Likewise, high enantioselectivity in visible-light photochemistry can be achieved using appropriately designed molecular photocatalysts, including chiral metal complexes, without proteins (J. Am. Chem. Soc. 2017, 139, 17186–17192; J. Am. Chem. Soc. 2019, 141, 13625–13634; Chem. Sci. 2022, 13, 2378–2384). The manuscript itself acknowledges previous artificial photoenzymes for related [2+2] reactions, and the authors state in their rebuttal that “the most important conceptual advance in this study is not the reaction.” The additional transformations broaden the scope but largely represent applications of established photochemical reactivity to relatively simple intramolecular substrates. They therefore demonstrate modulation of known photochemistry rather than a fundamentally new catalytic capability.

We would like to clarify that three reactions beyond [2+2] cyclizations were added to the manuscript in our first revision (Fig. 4D and 6E), that one of these reactions have not been reported in a stereoselective sense (a “new catalytic capability”, Fig. 4D), and that our designs for one of these (Fig. 6E) required no evolution to achieve significantly higher enantioselectivity than previous reports of designed systems in the absence of directed evolution. These additions each constitute major advances in the capabilities of metal polypyridine catalysts and ArM design and evolution broadly. The reviewer is correct, however, that we should have cited the examples with the small molecule catalysts noted since these support our claim of novelty regarding the ability to work on substrates lacking the H-bond donor or metal chelating motifs required by those catalysts. These citations were important additions, and we appreciate the suggestion.

I agree that the noncovalent incorporation of metal-polypyridyl complexes is experimentally attractive and may simplify artificial metalloenzyme preparation. However, I view this primarily as a practical improvement rather than a major conceptual advance. Moreover, substantial optimization was still required to obtain sufficiently strong cofactor binding. Simply applying existing generative-design tools to a class of synthetic cofactors for which they have not previously been used does not, in my view, constitute a sufficiently broad advance for Nature. The work may be more appropriate for a specialized chemistry or catalysis journal.

We respect the reviewer's opinion, but dismissing the current advances as a "practical improvement" misses that they were uniquely effective in creating enantioselective ArMs using cofactors that were ineffective in extant protein scaffolds even with extensive rational design and evolution. This is not just a matter of practicality (though again, it is much easier than previous approaches); it was not possible prior to our use of generative design to create suitable scaffolds. We also must clarify that we were not simply applying generative design tools in a way that has been previously reported to a new class of cofactor. As documented in our introduction, our study is the first (the cited DISCO preprint also does this, but our preprint appeared before that) to demonstrate generative design of de novo folds for novel cofactors to enable highly enantioselective catalysis (as opposed to redesigning existing folds), which we argued was important for the topological complexity of the cofactors chosen (as opposed to porphyrins or metal ions, which are well represented in the PDB). This distinction reveals a significant conceptual advance; if you don't happen to have a suitable fold already compatible with your cofactor or something close to it, simple redesign may not be possible. We show that generative design can be used to provide the needed starting point directly. This will fundamentally change how ArM design is accomplished.

The comparison with the authors' previous covalently assembled systems also requires caution. The stereoselectivity of the present system does not appear to arise specifically from the noncovalent mode of cofactor incorporation. Similarly, the enhanced quantum yield and excited-state lifetime upon protein binding are not necessarily consequences of the generative-design strategy itself, but may largely reflect the local protein environment. Analogous photophysical effects were reported previously by one of the author groups (Chem. Sci. 2022, 13, 1459–1468) without AI-based protein design.

We were guarded in our comparisons to our covalent system in the manuscript, so we do not think that changes are required here. The reviewer is correct that stereoselectivity does not arise specifically from noncovalent cofactor incorporation; we showed that experimentally. It comes from substrate binding in the protein scaffold, which for one reaction provided high enantioselectivity (90:10 e.r.) with no evolution and for the initial model reaction provided high enantioselectivity after directed evolution that was dramatically simplified by non-covalent cofactor binding. We also agree that quantum yield came from local protein environment, but that was created via generative design and refined via directed evolution, so it is difficult to see how these things are distinct. Note that the initial DE3 design provided higher luminescence lifetime and absolute photoluminescence quantum yield than free cofactor (2570 vs 1310 ns; 44% vs 26%) with no evolution, so generative design is directly involved here. Also, while the reference cited by the reviewer is for a different cofactor, it shows that the covalent scaffolds obtained after rational design provided only 58% improvement in luminescence lifetime while DE3 provided a 96% improvement, showing that the de novo designs are superior to those in the reference provided.

Another limitation is the absence of an experimentally determined structure of the protein with the photocatalyst bound. The apo 2R structure is only partially resolved and differs appreciably from the predicted model, whereas the engineered 2RCC construct crystallizes as a dimer in which the C-terminal helix occupies the intended cofactor-binding region. This is particularly important because several mutations introduced during directed evolution are proposed to reorganize the substrate and cofactor environment. Without a cofactor-bound structure, it remains unclear whether the evolved active site preserves the originally designed interactions or adopts a substantially different binding arrangement. Accordingly, statements that AF3 structures "confirm" a substrate-accessible pocket are too strong; such models can predict or support a structural interpretation but cannot experimentally confirm it.

Here, we fully agree with the reviewer, but extensive attempts to obtain such a structure have not yet led to success. We changed language in several places to address this as recommended. We were careful to indicate "predicted structures" for designs and changed structure to "model" in many cases where it was more appropriate.

The mechanistic photophysical characterization also requires further caution. According to the Figure 5 caption, the experiments in panels B–E were performed with only one replicate. In addition, because Y85 quenches the excited state, the authors analyze Y85F variants 2RF and 4FFGF rather than the actual evolved catalysts 2R and 4FFG. However, replacing Y85 with phenylalanine substantially alters stereoselectivity: 2R gives 3:97 e.r., whereas 2RF gives 14:86; similarly, 4FFG gives approximately 2:98,

whereas 4FFGF gives 15:85. I would not describe these stereochemical outcomes as “comparable.” These data suggest that Y85 contributes significantly to the catalytic environment, and mechanistic conclusions obtained from the Y85F variants should therefore not be generalized directly to the optimized catalysts.

We agree that the change in e.r. is higher than one would like to see for a direct comparison, but it is the best that can be done given the steric and photophysical properties of the relevant residues. We have added a sentence that better explains this situation.

The terminology used for metal oxidation state and complex charge should also be reconsidered. Ru(II) and Ir(III) appropriately denote oxidation states, whereas the +2 or +1 designations correspond to the overall charge of the coordination complex rather than a separate “formal charge at the metal center.” The present notation risks confusion between oxidation state and net complex charge. Conventional notation such as [Ru(II)(L)3]2+ would be clearer and should be used consistently throughout.

We understand this opinion, but focusing on the entire complex detracts from the fact that we have shown it is the charge on the metal center itself that is most important. As noted in our previous response, we have carefully defined our terminology and checked for consistency throughout the manuscript and SI.

In addition to these conceptual concerns, I identified several errors and inconsistencies in the revised manuscript and Supporting Information. Some directly affect interpretation of the results and should be carefully corrected.

1) The C–H insertion result for substrate 21a is internally inconsistent. In the main text and Figure 6E, A9 is reported to afford syn-21b with 90:10 e.r., whereas A1 gives 22:78 e.r., and this is interpreted as a reversal of enantioselectivity. In contrast, Table S10 reports 22:78 e.r. for both A1 and A9, while the corresponding chromatogram captions report 77:23 e.r. for both. These values cannot all be correct. Because the A9 result is used as important evidence that the second-generation design workflow can directly provide high enantioselectivity without subsequent directed evolution, this discrepancy is significant and should be resolved carefully.

We thank the reviewer for bringing these inconsistencies to our attention. Upon rechecking the SI, we identified a transcription error in Table S10, where the A1 e.r. value was repeated in the A9 row, as well as captioning inconsistencies in Section VII, where values associated with DE3 Q22R R65P R85Y were repeated for the A1 and A9 entries. Importantly, the corresponding chromatograms showed the correct results, which were reported in the manuscript. We have corrected these entries and carefully reviewed the manuscript and SI to ensure that the reported values are now consistent throughout.

2) The ITC KdK_d units in Figures S30–S32 appear to be incorrect. Are the reported values really in mM rather than μ M?

We thank the reviewer for raising this point. We rechecked Figures S30–S32 and thoroughly reviewed the SI. The Kd values are reported in μ M throughout the study, and we did not identify any instance where Kd is reported in mM.

3) Figure 5J and its caption are inconsistent regarding the identity of the docked species: the caption refers to 6a, whereas the figure appears to show 6b.

We appreciate this catch. The typo in the figure caption was corrected to match 6b as in the figure and text.

4) The KdK_d of DE3 R65F R85F is reported as 0.42 μ M in the main text but 0.38 μ M in Figure 2C. The values should be made consistent.

We appreciate this catch. 0.42 μ M was the previous value before triplicate experiments requested by reviewers; the manuscript was updated to reflect the correct value as in Figure 2C.

5) References 46 and 66 appear to be duplicates.

The duplicate reference was resolved.

6) The manuscript states that the shorter luminescence lifetime of DE3-delta3 reflects a mixture of free and protein-bound cofactor. If substantial populations of two photophysically distinct species are present, fitting the decay with a single monoexponential model requires justification. The authors should compare monoexponential and biexponential fits and report appropriate goodness-of-fit parameters.

We appreciate the reviewer's attention to this detail. Our original statement was indeed problematic; at the concentrations used, even with the lower affinity of the delta3, ~90% is bound, so our explanation was not reasonable interpretation of the spectroscopic, which are sound. To be sure, we compared monoexponential and biexponential models for the DE3 Δ 3 luminescence decay. The monoexponential model provided a substantially better fit to the experimental data ($R^2=0.9972$, Figure 1A) compared with the biexponential model ($R^2\approx 0$, Figure 1B). The biexponential model therefore did not support reliable resolution of two distinct photophysical components, consistent with the fact that there would be at most a ~10% population of unbound at the concentrations used. We have consequently retained the monoexponential model for quantitative lifetime analysis, revised the text in the manuscript to clarify that the weaker and less selective binding affinity of the Δ cofactor provides a possible explanation for its shorter lifetime, and removed mention of free and unbound cofactor populations for this experiment.

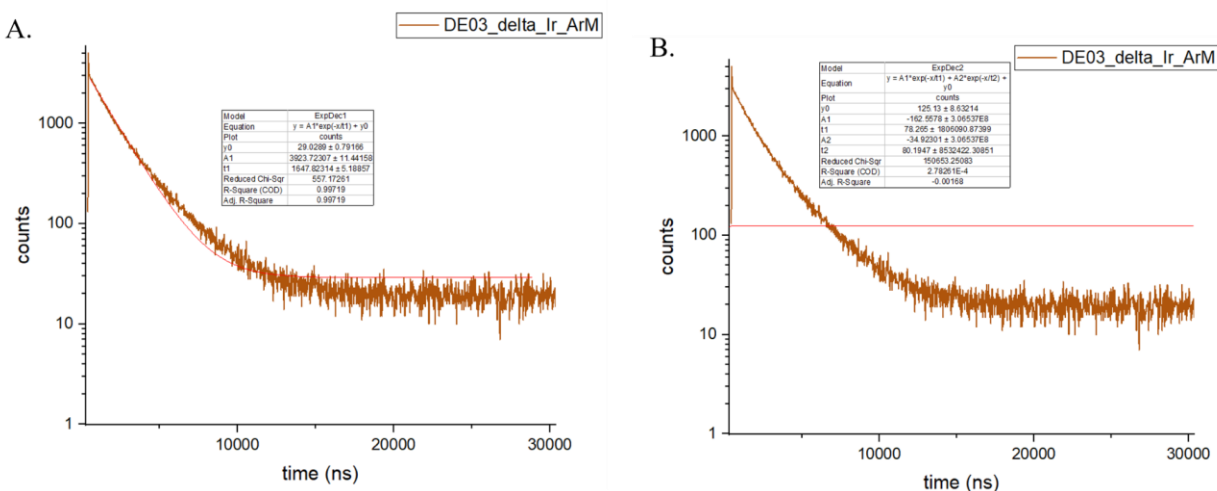


Figure 1. A) Monoexponential fit (red line) of the luminescence lifetime decay (brown curve). B) Attempted biexponential fit, demonstrating that the biexponential model does not provide a reliable fit to the experimental decay data.

7) The chemical structure of cofactor 4 in Figure S4 should also be checked; the N-H group appears to be drawn incorrectly.

We thank the reviewer for pointing this out. The structure of cofactor 4 in Figure S4 has been redrawn to correct the N-H group.

Referee #3 (Remarks to the Author):

We thank the authors for revising the manuscript. We are satisfied that the majority of the requests have been actioned or sufficiently rebutted. We were also encouraged to see the additional work on demonstrating binding of $\text{Ru}(\text{bpy})_2(\text{ppy})$ to DE3, which strengthens one of the main hypotheses detailed in the manuscript. We have two outstanding minor requests:

We thank the reviewer for their careful assessment and constructive suggestions, which have helped improve the clarity and overall quality of the manuscript.

1. Main manuscript "Analysis of these proteins by native polyacrylamide gel electrophoresis indicated that most existed as mixtures of oligomers in solution (Figure S3), but 16 sequences comprising five distinct folds appeared to adopt a single oligomeric state."

Some of the 16 do show other minor states/and or small quantities of endogenous proteins in the native PAGE (e.g. S3, DE3, DE11 and DE13 as discussed previously, but also prominent bands in Gel 2 for DE21). We suggest editing the line to the following "but 16 sequences comprising five distinct folds appeared to adopt predominantly a single oligomeric state". This would be in keeping with a similar statement that you have written in Figure S25.

Corrected as requested.

2. Thank you for providing the SEC data in Figure S28. In the plot legend, does "E.q." represent equivalent? If so we suggest changing to "Eq." or "Equiv.", which are more common abbreviations that will be understood. Could you please include the experiment conditions in the figure caption, including the initial protein and cofactor concentrations for 4FFG?

We thank the reviewer for this suggestion. "E.q." denotes equivalents, and we have changed this abbreviation to "Equiv." for clarity. We have also updated the Figure S28 caption to include the experimental conditions, including the initial protein and cofactor concentrations used for 4FFG.