

RESEARCH

Open Access



Codeine 3-*O*-demethylase catalyzed biotransformation of morphinan alkaloids in *Escherichia coli*: site directed mutagenesis of terminal residues improves enzyme expression, stability and biotransformation yield

Garrick W. K. Spencer^{1,2}, Xu Li¹, Kenny W. L. Lam¹, George Mutch², Fiona H. Fry² and Sally L. Gras^{1*}

Abstract

The cultivation of opium poppy is the only commercially viable source of most morphinan alkaloids. Bioproduction of morphinan alkaloids in recombinant whole-cell systems provides a promising alternate source of these valuable compounds. The enzyme codeine 3-*O*-demethylase can transform morphinan alkaloids by *O*-demethylation and has been applied in single step biotransformation reactions or as part of larger biosynthetic cascade, however, the productivity for these reactions remains low and suboptimal enzyme properties could be improved. This mutagenesis study targeted non-conserved N- and C-terminal residues, which were replaced with the equivalent residues from enzyme thebaine 6-*O*-demethylase. Whole cell biotransformation performance was significantly improved in *Escherichia coli* expressing codeine 3-*O*-demethylase mutants, with a ~2.8-fold increase in the production of oripavine from thebaine and ~1.3-fold increase in the production of morphine from codeine. Statistical analysis of biotransformation yield, enzyme expression and stability, predicted using changes in Gibbs free energy ($\Delta\Delta G$) with deep-learning-based model DDmut, suggested that altered enzyme stability and/or expression of soluble protein may contribute to the observed improvements in biotransformation. This approach could be beneficial for screening future codeine 3-*O*-demethylase mutations and for other enzymes.

Keywords Poppy, *E.coli*, Bioconversion, Oripavine, Morphine, DDmut., Overall stability

*Correspondence:

Sally L. Gras
sgras@unimelb.edu.au

¹The Department of Chemical Engineering and the Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, VIC 3010, Australia

²Sun Pharmaceutical Industries Australia Pty Ltd, Princes Highway, Port Fairy, VIC 3281, Australia



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

The morphinan alkaloids sourced from opium poppy (*Papaver somniferum* L.) and their semi-synthetic derivatives encompass some of the most effective pharmaceutical compounds available to treat pain, opioid poisoning and addiction. These include the agonists morphine, codeine and oxycodone, the mixed antagonist-agonist buprenorphine and antagonists naltrexone and naloxone [1]. Commercial production of these important alkaloids, or their precursors, relies upon poppy cultivation, as the presence of multiple chiral centers currently makes total chemical synthesis not feasible [2].

Whole cell based complete biosynthesis from simple sugars could be an alternative source of these valuable opioids, as has been demonstrated in *Saccharomyces cerevisiae* (*S. cerevisiae*) and *Escherichia coli* (*E. coli*) [3, 4]. Simpler whole cell systems that perform a single or small number of enzymatic steps to interconvert between

opioids are also being pursued [5, 6], such as an *E. coli* biotransformation system for the conversion of thebaine to oripavine and codeine to morphine by codeine 3-O-demethylase (CODM) [6]. These provide a means to better manage opioid supply [5, 6], as a further alternative to traditional agricultural production [7], whilst the yield of complete biosynthesis is being optimized.

CODM is a non-heme 2-oxoglutarate/Fe(II)-dependent dioxygenases, responsible for the 3-O-demethylation of thebaine to oripavine and codeine to morphine in plants [8] (Fig. 1A). This enzyme has been used in *E. coli* [6] and yeast based biotransformation [2, 9–11]. The performance of wild type (WT) CODM was less than optimal in yeast, with a productivity of $\sim 8.9 \times 10^{-6}$ g/(L·h) for codeine to morphine conversion, potentially due to low levels of enzyme expression or solubility or low enzyme activity [2, 9–11]. Productivity was improved with expression in *E. coli*, producing $\sim 2.2 \times 10^{-1}$ g/(L·h)

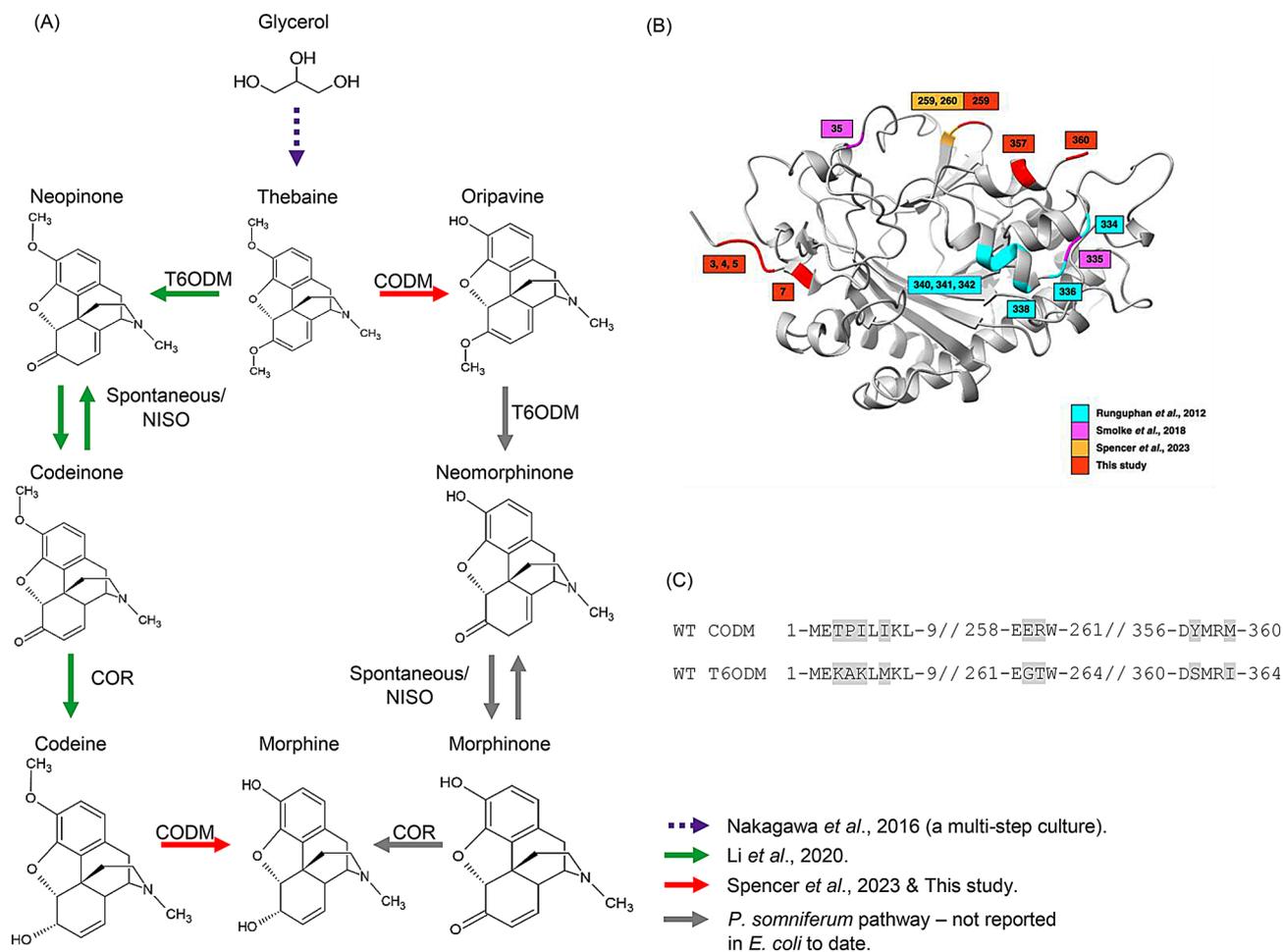


Fig. 1 Overview of biotransformation reaction steps and mutagenesis targets. (A) Overlay of the key enzymatic steps for the conversion of thebaine to morphine from within *Papaver somniferum* that have been expressed recombinantly in *E. coli* in the literature to date [4–6]. (B) 3D protein structure of WT CODM predicted by AlphaFold2 (assessed via EBML-EBI server), where residues were colored using ChimeraX. Residues with a notable impact on CODM performance in previous mutagenesis studies are highlighted (in cyan [13], magenta [14] and orange [6]) alongside the residues mutated in this study (in red). (C) Sequence alignment of the WT CODM (CODM.2) and WT T6ODM (T6ODM.4) variants found within *P. somniferum* HN1 [8, 15, 16] for the region of interest, where non-identical residues are highlighted in grey. Sequences were aligned using the MUSCLE methodology with SnapGene

morphine (at $t=1.25$ h when codeine achieved $>90\%$ conversion) [6] but was still at the lower end of the range reported for successful pharmaceutical biotransformations [12]).

Mutagenesis of CODM has helped to establish the importance of key CODM features. Non-conserved residues, including the N- or C-terminal regions, have typically been altered (Fig. 1B and Supplementary Figure S1) [6, 13, 14]. For example, mutations in the N- and C-terminus of CODM led to ~ 1.4 -times more morphine (no productivity was given) but also ~ 2.6 -times more side product neomorphine [14]. Two CODM C-terminal mutants, created by interchanging non-conserved residues with the paralogous enzyme thebaine 6-*O*-demethylase (T6ODM), had reduced activity relative to the WT CODM [13].

Regions outside the N- and C-terminus have also displayed improved productivity. Mutant E259G (Fig. 1B) gave $>22\%$ greater oripavine and morphine production than WT CODM in an *E. coli* biosynthetic system [6]. This mutation involves the only residue that differs between the three CODM open reading frames present within the *P. somniferum* L. HN1 genome [6]. It is also non-conserved between CODM and many homologues, including T6ODM [8]. While the E259G mutant showed higher productivity, this could still be enhanced to increase economic viability.

An opportunity exists to further increase CODM performance and improve our understanding of this enzyme by systematically replacing other non-conserved residues with those found in T6ODM, which showed a higher substrate affinity for thebaine in purified enzyme assays (K_m of $20 \pm 7 \mu\text{M}$ for T6ODM compared to $42 \pm 8 \mu\text{M}$ for CODM) [8, 17, 18]. Although no crystal structure has yet been reported for CODM, comparisons can be drawn to published structures for T6ODM (PDB entry of 5O9W and 5O7Y), which share $\sim 73\%$ identity to CODM [19]. The two enzymes differ by ~ 105 residues but to date only 23 residues in CODM have been replaced with the equivalent residues in T6ODM, either as single or multiple changes [6, 13, 14].

Enzyme stability and the level of soluble protein expression could impact whole cell biotransformation, together with enzyme selectivity and turnover rate. Suboptimal stability and expression are proposed to have played a role in early reports of poor CODM performance in yeast [2, 9–11]. A greater concentration of soluble enzyme was observed for CODM mutants E259G and E259D compared to WT CODM, indicating that enhanced solubility may contribute to improved biotransformation [6].

In recent years, significant progress has been made towards developing computational tools to aid our understanding of the effects of mutations on protein stability and soluble expression [20–22]. DDMut (a deep

learning model for predicting protein stability), for example, was used among other stability prediction tools to interrogate how mutations to the CSMD1 (Cub and Sushi Multiple Domains 1) protein, an important component of the innate immune response, may cause pathogenicity through changes in protein stability [23]. An assessment of these properties may therefore help to better understand the many factors contributing to CODM biotransformation performance.

This study aimed to improve the conversion of thebaine to oripavine and codeine to morphine in an *E. coli* biotransformation system through CODM mutagenesis. We hypothesized that replacing six non-conserved residues throughout the N- and C-terminal regions of CODM with the equivalent residue from T6ODM would increase biotransformation productivity, by increasing enzyme activity, soluble enzyme expression level, enzyme stability, or a combination of these factors. These residues have not been studied to date. Combinations of enhanced single-site mutations were also explored. Correlations between the normalized biotransformation yield, predicted enzyme stability and soluble enzyme expression level were determined for this study, providing insight into the underlying mechanisms that may lead to successful improvements in yield and a possible method for guiding the future rational selection of mutant enzymes.

Results and discussion

Non-conserved CODM residue changes impact biotransformation

This study sought to explore how 6 non-conserved residues contribute to CODM enzyme performance during the whole cell biotransformation of the opiates thebaine and codeine (Fig. 1A). Within the C-terminal extension region of CODM, 2 residues at position 357 and 360 were chosen (Fig. 1B), as this region is known to impact on CODM performance [13]. In addition, 4 residues were selected at the N-terminus at positions 3, 4, 5 and 7 (Fig. 1B), as only one limited previous study has explored the N-terminal region of CODM with nothing known about how mutagenesis directly at the beginning of N-terminus may impact CODM biotransformation [14]. The residues present at these sites in CODM were replaced with the equivalent residue present in T6ODM, as shown in Fig. 1C, creating: T3K, P4A, I5K, I7M, Y357S or M360I. Biotransformation was conducted following a previously described protocol, where strains were screened after 30 min for morphine yield and 4 h for oripavine yield [5, 6]. Both time points were demonstrated as satisfying points to screen CODM mutants with enhanced biotransformation performance [6].

Improved yield was observed for two of the six single mutation CODM strains investigated, I5K and M360I, with ~ 1.8 -fold and ~ 1.3 -fold more oripavine from

thebaine and ~1.6-fold and ~1.3-fold more morphine from codeine, respectively (Fig. 2A, Supplementary Table S4). An increase in soluble enzyme appears to contribute to the enhanced activity of the I5K strain but not that of M360I (Supplementary Fig. S2). Interestingly, the P4A strain improved oripavine yield (~1.4-fold) but not morphine yield, while T3K had no significant effect and yield was reduced for both the I7M and Y357S strains, most notably for oripavine production (I7M yield was 12% of WT while Y357S yield was 58% of WT) rather than morphine production (I7M yield was 72% of WT while Y357S

yield was 78% of WT) (Fig. 2A, Supplementary Table S4). Notably, there was no significant difference in soluble protein expression of I7M and Y357S (Supplementary Fig. S2), suggesting the observed effects may be due to a change in enzyme catalytic function.

It is noteworthy that both mutations in the C-terminal extension have an impact on CODM activity, one positive (M360I) and the other negative (Y357S), indicating the importance of this region, which has previously been shown to be associated with CODM substrate specificity [13, 19]. Mutations in the N-terminal domain of

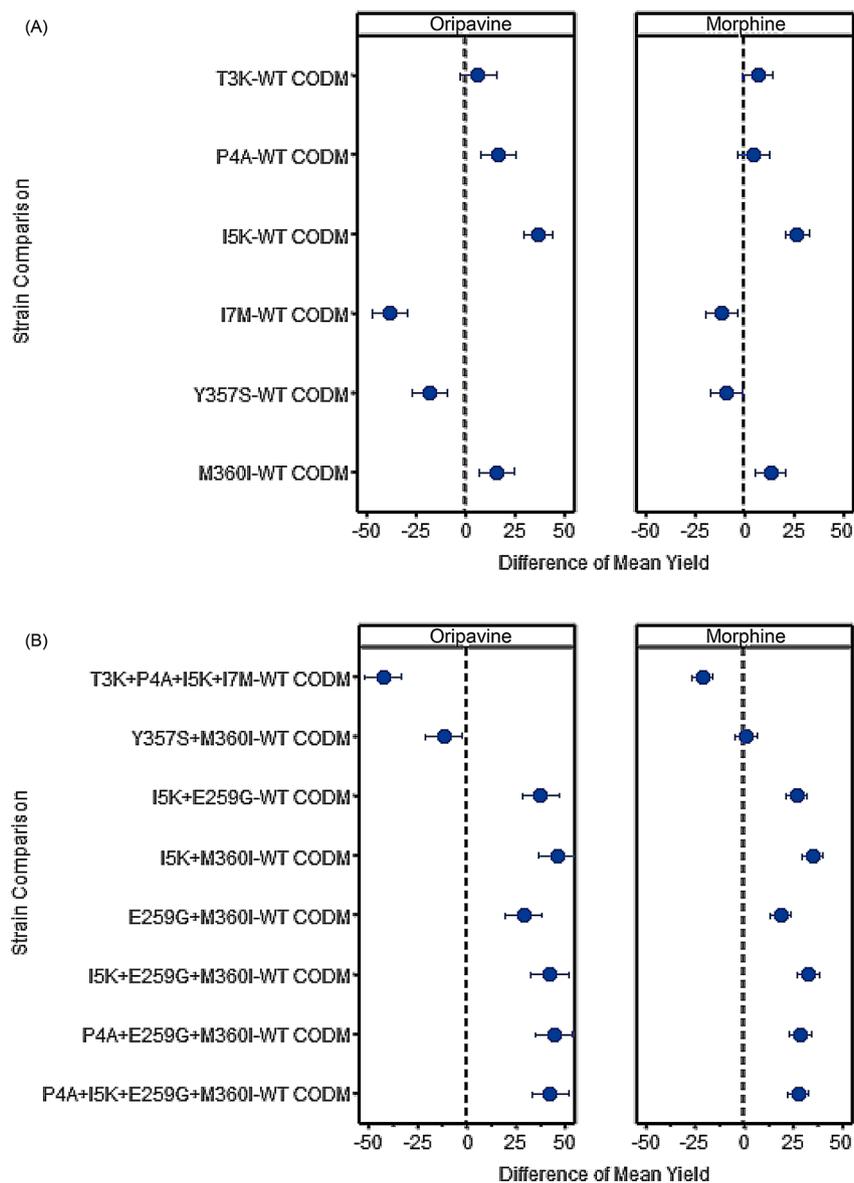


Fig. 2 The production of oripavine and morphine with mutant strains in comparison to WT CODM. **(A)** Mean difference in oripavine and morphine yield produced by the *E. coli* strains expressing CODM mutants with varying amino acids at positions 3, 4, 5, 7, 357 or 360. **(B)** Mean difference in oripavine and morphine yield produced by the *E. coli* strains expressing CODM mutants with combinations of amino acids varying at positions 3, 4, 5, 7, 259, 357 or 360. Biotransformation was assessed after four hours for oripavine or 30 min for morphine. WT CODM and I5K data are the mean difference of six independent replicates \pm 95% confidence interval, all other data are the mean difference of three independent replicates \pm 95% confidence interval (see Supplementary Tables S4 and S5 for oripavine and morphine percentage yield)

CODM, which has been less explored, also had a mixed effect, with two having a positive impact (I5K and P4A), one negative (I7M) and one no significant effect (T3K). This is consistent with this region also affecting substrate binding [24], with the location and nature of the mutation affecting the impact of substitution.

Improving CODM biotransformation by combining mutations

In a second strategy, eight additional CODM strains were generated based on three complementary approaches. The first approach combined both mutations in the C-terminus (Y357S + M360I) or combined all four mutations in the N-terminus (T3K + P4A + I5K + I7M) to investigate the effect of mixed mutations (with observed neutral, positive and negative effects) within a localized region. The second approach combined the most successful strains from the single mutation screen, I5K and M360I (Fig. 2B), with the mutant E259G, which was previously found to give good performance for biotransformation of both purified thebaine and codeine, as well as raw thebaine poppy extract, in a screen of mutants at sites 259 and 260 [6]. The additive effect of these mutants was assessed in all four combinations (I5K + E259G, I5K + M360I, E259G + M360I and I5K + E259G + M360I). In the third approach, two additional mutants were screened (P4A + E259G + M360I and P4A + I5K + E259G + M360I), to assess the impact of adding the mutant P4A to select combinations of mutants, as P4A was the only site in the single mutant screen that had a positive effect on one opiate and a neutral effect on the other: increasing oripavine yield and not altering morphine yield relative to WT CODM. E259G + M360I was selected as the base strain for this last strategy, as it was the best performing strain without changes to the N-terminus; P4A was also tested in combination with I5K, the other successful single mutation at the N-terminus to determine if there was any interaction between two N-terminal mutations.

The CODM strains containing all four mutations in the N-terminus (T3K + P4A + I5K + I7M) or both mutations in the C-terminus (Y357S + M360I) from the first approach did not improve performance relative to WT CODM. The T3K + P4A + I5K + I7M strain produced only 5% of the oripavine of WT and 51% the morphine of WT (Fig. 2B, Supplementary Table S5), while the Y357S + M360I strain produced similar morphine yield to WT CODM but only 70% of the oripavine yield of WT. These experiments illustrate how beneficial single mutations (e.g., I5K or M360I) can lose or have their activity masked when combined with other mutations from the same region of CODM. An increase in soluble enzyme concentration noted for T3K + P4A + I5K + I7M (Supplementary Fig. S3) did not increase oripavine or morphine yield, illustrating the importance of screening for both

activity, and expression level to understand the impact of mutations on whole cell biotransformation.

All four combined strains (I5K + E259G, I5K + M360I, E259G + M360I and I5K + E259G + M360I) from the second strategy improved the yield of morphine and oripavine relative to WT CODM (Fig. 2B). The combination of beneficial single mutations was not additive, with performance difficult to predict. In this instance, a higher concentration of soluble mutant enzyme was expressed compared to WT CODM for all mutant combinations examined (Supplementary Fig. S3), although this coincided with higher yield for only some of the mutant strains examined. There was also no positive impact on oripavine or morphine yield observed when the N-terminal mutation P4A was tested alone or together with the N-terminal mutation I5K in the strain containing the mutations E259G + M360I, highlighting how combining mutations to improve enzyme activity is not straightforward, as observed widely in the literature [25].

CODM stability and soluble protein expression

Whilst the mutagenesis approach employed generated improved mutants, it was not easy to predict mutant activity, including when mutants were combined. We hypothesized that greater insights might be obtained by a systematic consideration of protein stability and soluble enzyme expression across the datasets obtained from this and our prior study [6].

A measure of protein stability was considered first, as a loss of protein stability is often linked to loss of function [26], so we sought to investigate if the mutations tested altered protein stability, potentially impacting enzyme activity. We used DDmut [20], a deep-learning-based model that predicts changes in Gibbs free energy ($\Delta\Delta G$) upon single and multiple amino acid mutations, as a global measure of stability. This method has been trained and validated for up to 3 simultaneous mutations. Structure prediction and modelling tools were used to generate the input 3D structure, as homology modelling and AlphaFold2 prediction are considered comparable to X-ray crystallography structures for predicting $\Delta\Delta G$ [27] and the CODM crystal structure has not yet been reported.

The predicted $\Delta\Delta G$ was then paired with the measured soluble enzyme expression level and compared to the experimental biotransformation yield to assess whether structural stability and/or the soluble enzyme expression level may influence biotransformation yield. The first analysis considered mutants described in this study which ranged in mutation site, as well as complementary mutants E259G, E260T, and E259G + E260T from our previous study [6], where residues were replaced with the equivalent residue in T6ODM. The second analysis included mutants only occurring at site 259, where the

amino acids were systematically varied (also from our previous study [6]).

Both soluble enzyme expression level and enzyme stability appear to influence CODM mutant whole cell biotransformation performance for the enzymes where mutations were made using the amino acids present in T6ODM (Fig. 3A-B). A moderate but statistically significant positive correlation was observed between the normalized soluble expression level and normalized

experimental yield for both oripavine and morphine yield for this set of mutants (Fig. 3A; oripavine: $r=0.63$, one-sided 95% CI= (0.45, 1.00), $p=1.00 \times 10^{-5}$; morphine: $r=0.69$, one-sided 95% CI= (0.52, 1.00), $p=1.00 \times 10^{-5}$). A weaker but still statistically significant positive correlation was also observed between the Gibbs free energy ($\Delta\Delta G$) predicted by DDMut and the normalized experimental yield (Fig. 3B; oripavine: $r=0.48$, one-sided 95% CI= (0.24, 1.00), $p=1.00 \times 10^{-4}$; morphine: $r=0.50$,

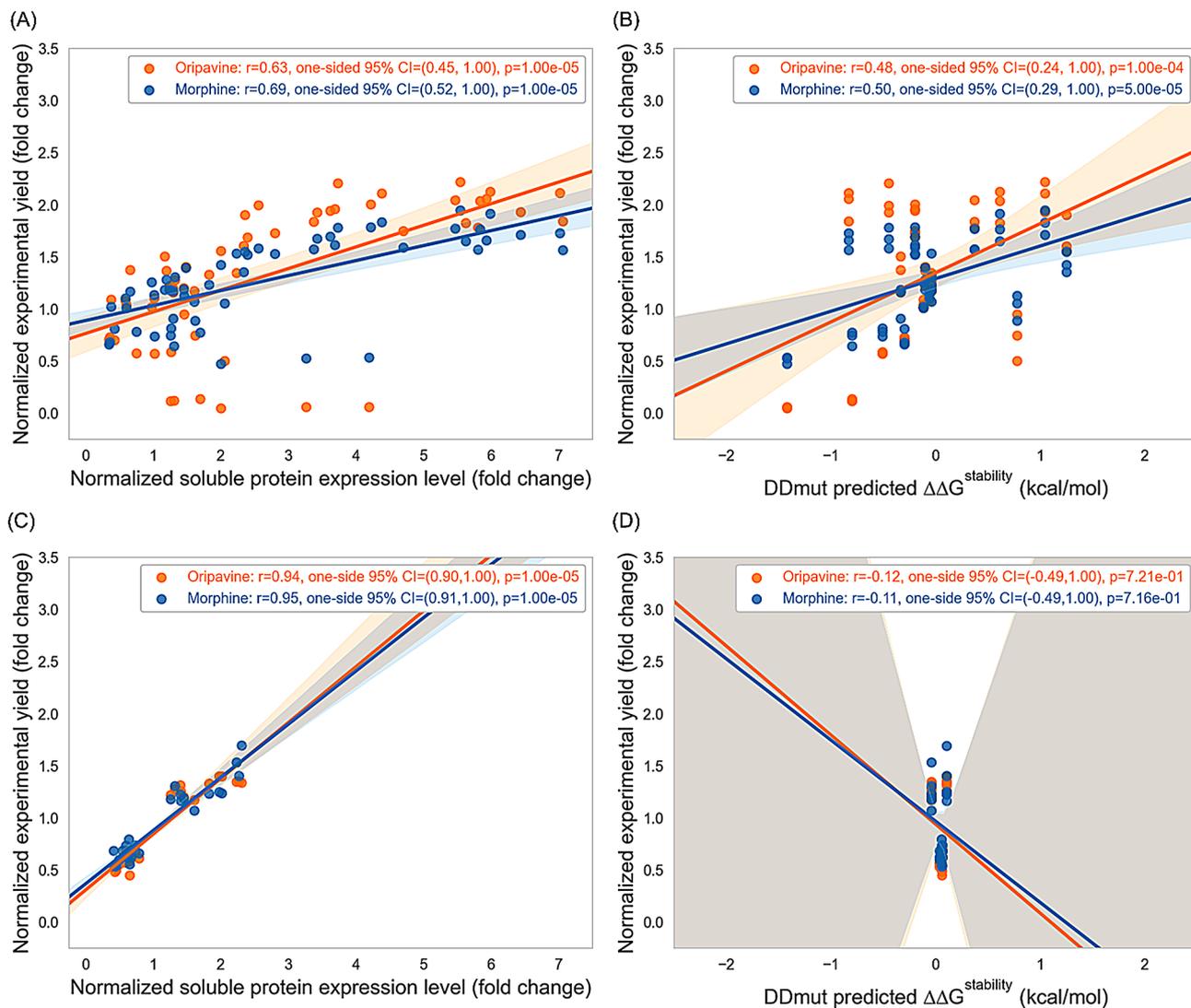


Fig. 3 Correlations between soluble protein expression and biotransformation yield and between predicted enzyme stability and biotransformation yield. Normalized experimental yield for *E. coli* strains expressing CODM mutants using amino acids from T6ODM (including mutant data from our previous study) at positions 3, 4, 5, 7, 259, 260, 357, 360 as a function of **(A)** normalized soluble enzyme expression level or **(B)** predicted enzyme stability. Normalized experimental yield for *E. coli* strains expressing CODM mutants with amino acid substitution at position 259 from our previous study, as a function of **(C)** normalized soluble enzyme expression level or **(D)** predicted enzyme stability. DDMut stability predictions are the mean for predictions made using the input of four different 3D modelling methods (MODELLER, SWISS-MODEL, AlphaFold2, ColabFold-AlphaFold2 (see Supplementary Table S8 and Fig. S4). Positive $\Delta\Delta G$ values denote stabilization and negative values denote destabilization. All experimental replicates are represented as separate data points, with three replicates for all mutants, except for E259G and I5K, for which there were six replicates. SciPy was used for statistical analysis: Pearson's correlations (r) and permutation test for calculating the p -values, where the null hypothesis indicates less than or zero correlation between variables, and the bootstrap method for calculating the one-sided 95% confidence intervals (CI). Seaborn was used to plot the regression line and the two-sided 95% CI for the regression line as the shaded area in the graphs, also based on bootstrap estimation

one-sided 95% CI= (0.29, 1.00), $p=5.00 \times 10^{-5}$), indicating that mutations giving increased enzyme stability may contribute to better performance. No correlation was observed between the predicted $\Delta\Delta G$ and normalized soluble expression level (Supplementary Figure S4C), indicating this did not contribute to the observed correlations. The moderate and fair correlations observed here for soluble enzyme expression and enzyme stability with experimental yield suggest that enzyme selectivity, which was not considered in this analysis, also potentially differs between these mutants and contributes to the differences in biotransformation performance observed (Fig. 2).

In contrast, the mutations previously described at residue 259 in CODM, where the amino acid was systematically varied at the same location [6], appear to differ in soluble enzyme expression but not enzyme stability (Fig. 3C-D). A very strong correlation was observed between the normalized soluble expression level and normalized experimental yield when all amino acids in the previous study were considered (Fig. 3C; oripavine: $r=0.94$, one-sided 95% CI= (0.90, 1.00), $p=1.00 \times 10^{-5}$); morphine: $r=0.95$, one-sided 95% CI= (0.91, 1.00), $p=1.00 \times 10^{-5}$), with a correlation stronger than that observed here in the current study where mutations were examined across multiple site locations. There was no correlation, however, between the $\Delta\Delta G$ predicted by DDMut and the normalized experimental yield for mutants considered just at the 259 site (oripavine: $r=-0.12$, one-sided 95% CI= (-0.49, 1.00), $p=0.72$; morphine: $r=-0.11$, one-sided 95% CI= (-0.49, 1.00), $p=0.72$). This suggests that the mutants at residue 259 have retained the same stability as the WT CODM. Again, no correlation was also observed between the predicted $\Delta\Delta G$ and normalized soluble expression level (Supplementary Fig. S4B).

The difference in enzyme properties between the two groups of mutants analyzed is significant, as it can potentially explain the mechanisms behind improved CODM performance for the two studies. In the prior study, altering the amino acid at site 259 increased soluble enzyme expression and did not alter mutant stability relative to the WT enzyme [6]. In contrast introducing different mutants using residues based on the T6ODM sequence across multiple locations in this study led to a greater performance improvement, likely through several mechanisms; this could include an increase in both soluble enzyme expression and enzyme stability but it seems that other mechanisms may also be involved, such as changes in enzyme selectivity.

This study illustrates the potential benefit of using a $\Delta\Delta G$ prediction tool to explore and understand the impact of mutagenesis. Computational tools that predict mutant impact on enzyme expression could also be used to simplify the design space for enzyme engineering. For

example, when the tool MPEPE (a Mutation Predictor for Enhanced Protein Expression) [21], was applied to the sequences examined here, the soluble enzyme expression level could be effectively predicted, with a moderate correlation between predicted and measured expression levels (Supplementary Fig. S5) suggesting both tools may be combined to optimize selection *in silico* prior to bench experimentation. Future work that could consider additional mechanisms, such as enzyme selectivity, could also provide more comprehensive predictions.

Strain performance during a biotransformation reaction

Having identified CODM mutants with better capacity for biotransformation, we next sought to investigate their performance over the course of an entire biotransformation reaction. Single mutation I5K from the N-terminus, single mutant M360I from the C-terminus or the double mutant I5K + M360I, which had a high yield but was the combination with the fewest mutations from the second strategy, were examined and compared to WT CODM. Oripavine was assayed every hour for the first nine hours, then every three hours for the following fifteen hours. While the quicker codeine *O*-demethylation reaction was assessed every fifteen minutes over two hours.

The I5K + M360I strain performed the fastest for thebaine, achieving near complete conversion ($\geq 95\%$) after ~ 5 h, with strains I5K, M360I or WT CODM achieving similar conversions after ~ 7 h, ~ 12 h or ~ 18 h, respectively (Fig. 4A, Supplementary Table S6). For codeine, the I5K and I5K + M360I strains displayed similar kinetic behavior, achieving near complete conversion ($\geq 94\%$) after ~ 1 h, with strains M360I or WT CODM achieving similar conversions after ~ 1.75 h and ~ 2 h, respectively (Fig. 4B, Supplementary Table S7). These data suggest that the I5K + M360I strain would be most desirable for biosynthetic applications.

The productivity, or time space yield (TSY), is important for the economic viability of opiate biotransformation and the I5K + M360I strain reported here is one of the most productive systems reported to date. This strain displayed an average TSY of $5.1 \times 10^{-2} \pm 6.5 \times 10^{-3}$ g/(L·h) for oripavine measured at $t=5$ h when thebaine achieved near complete conversion ($\geq 95\%$). This is a ~ 3.4 -fold improvement compared to WT CODM (oripavine TSY of $1.5 \times 10^{-2} \pm 6.2 \times 10^{-4}$ g/(L·h) at $t=18$ h when thebaine achieved $\geq 95\%$ conversion, see Supplementary Table S6). This is also a ~ 2.8 -fold improvement compared to the CODM mutant expressing strain E259D reported recently (oripavine TSY of $1.8 \times 10^{-2} \pm 4.4 \times 10^{-4}$ g/(L·h) at $t=15$ h when thebaine achieved $\geq 95\%$ conversion) [6]. The I5K + M360I strain also displayed a high TSY of $2.9 \times 10^{-1} \pm 8.2 \times 10^{-3}$ g/(L·h) for morphine measured at $t=1$ h when codeine achieved near complete conversion ($\geq 94\%$). This is a ~ 1.9 -fold improvement compared to

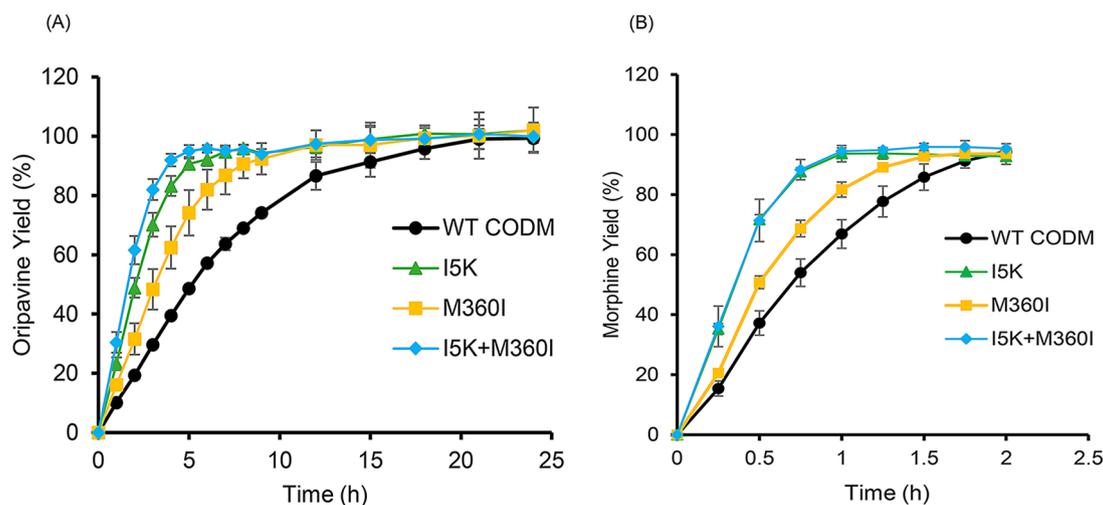


Fig. 4 Production of oripavine and morphine as a function of biotransformation time. Yield of oripavine (A) or morphine (B) as a function of time produced using either thebaine or codeine as substrate respectively and *E. coli* strains expressing I5K, M360I, I5K+M360I, or the WT CODM. Data are the mean \pm standard deviation of three independent replicates (see Supplementary Table S6 and S7)

WT CODM (morphine TSY of $1.5 \times 10^{-1} \pm 6.3 \times 10^{-3}$ g/(L·h) at $t=2$ h when codeine achieved $\geq 94\%$ conversion, see Supplementary Table S7). This is also a ~ 1.3 -fold improvement compared to the E259D strain reported recently (morphine TSY of $2.3 \times 10^{-1} \pm 4.4 \times 10^{-3}$ g/(L·h) at $t=1.25$ h when codeine achieved $\geq 94\%$ conversion) [6].

Future opportunities

The productivity of this CODM I5K+M360I variant complements existing biotransformation strategies. For example, incorporation of this mutant into the *E. coli* total biosynthesis system developed by Nakagawa and colleagues (Fig. 1) would expand the capabilities of this platform to produce the commercially valuable compound oripavine from glycerol [4]. Moreover, adoption of this strain within the stepwise process demonstrated in our earlier study (Fig. 1A), would allow for the complete biosynthetic production of morphine from thebaine [5]; this could reduce the timescales for morphine production using alternative feedstocks and potentially provide greater flexibility for opioid production on demand. The mutagenesis performed also identified I7M and Y357S as reducing oripavine yield relative to morphine yield, making these enzymes valuable tools within the synthetic biologist's toolbox that could be used to direct flux through desired metabolic pathways (i.e. to direct flux towards morphine production in Fig. 1), as has been suggested for other CODM mutants with altered substrate specificity [13]. Further insight into CODM mutants could also be obtained by quantifying and optimizing the relative proportions of soluble and insoluble protein expression for the current and any future mutants.

Several other mutation tools may be complementary in future enzyme studies, to the $\Delta\Delta G$ prediction

tool used here for stability and the MPEPE tool applied to confirm soluble enzyme expression levels. Potential enzyme mutations can be designed using web servers such as HotSpot Wizard v3.1 [28, 29], FireProt v2.0 [30] and PROSS (the Protein Repair One-Stop Shop) [31, 32]. These tools make protein design more accessible and can direct and simplify protein design activities [33–38]. While the underlying methods differ between algorithms, they commonly make use of evolutionary conserved analysis and/or stability calculations to compute promising mutation sites [38]. PROSS for example, optimizes the energy of the native protein to identify stabilizing mutations, subject to constraints inferred from multiple sequence alignments, to improve protein expression and heterologous expression [31, 32]. Interestingly, several of the mutations identified and tested in this and the previous study were also identified by retrospective analysis using these automated design tools (Supplementary Table S10), illustrating their broad potential.

Apart from improving thermostability and heterologous expression computational enzyme design can also be used to fast-track efforts to re-engineer enzyme specificity [39]. For example, the cofactor specificity of the *E. coli* K12 malic enzyme (MaeB), which has an unsolved three-dimensional structure, was changed from NADP⁺-dependent to NAD⁺-dependent form by combining a logistic regression model with an amino acid database of structurally homologous enzymes possessing different substrate/cofactor specificity, allowing residues to be ranked for their contribution to cofactor specificity [39]. A total of 10 mutant enzymes were then created, each progressively containing 10 further mutations with 20–30 mutations found sufficient to switch cofactor specificity without greatly affecting expression levels [39]. While

this strategy requires a sufficient database of homologues, it illustrates the potential for machine learning to accelerate optimization of enzyme performance.

Conclusions

Microbial biosynthetic systems are increasing our capacity to sustainably produce essential medications. A common bottleneck that has been encountered in opioid biosynthesis systems is the suboptimal performance of enzyme CODM when expressed in heterologous systems. In this work, we demonstrate how residues found in the paralogue enzyme T6ODM can be used to generate mutant CODM enzymes with a variety of characteristics including increased yield. The *E. coli* strain expressing the CODM I5K+M360I mutant is one of the most productive strains reported to date, with a productivity of $5.1 \times 10^{-2} \pm 6.5 \times 10^{-3}$ g/(L·h) oripavine and productivity of $2.9 \times 10^{-1} \pm 8.2 \times 10^{-3}$ g/(L·h) morphine. The mutant CODM strain could be applied to expand the diversity of opioids that can be produced, advancing on the low yields previously reported for WT CODM in yeast. This study also illustrated how the computational tool DDMut can potentially be used to better understand enzyme engineering strategies and could be used in the future to identify promising residues for mutagenesis.

Methods

Chemicals and reagents

Cell culture reagents were purchased from Merck, USA (antibiotics, analytical grade glycerol, isopropyl β -D-1-thiogalactopyranoside (IPTG), sodium chloride) or Oxoid, Thermo Scientific UK (yeast extract and tryptone). Biotransformation reagents (sodium ascorbate, glucose, and iron sulfate heptahydrate) and high-performance liquid chromatography (HPLC) reagents (HPLC grade acetonitrile, trifluoroacetic acid (TFA) and acetic acid) were purchased from Merck, USA. Cloning reagents (restriction enzymes, Phusion High Fidelity DNA Polymerase, NEBuilder HiFi DNA Assembly, *E. coli* 5-alpha competent cells and *E. coli* BL21(DE3) competent cells) were purchased from New England BioLabs Inc., USA. Sun Pharmaceutical Industries Australia Pty Ltd. provided the opioids ($\geq 98\%$ w/w purity on a wet basis); thebaine, oripavine, codeine and morphine. Water used was purified to a resistivity of ≥ 18.2 M Ω cm.

Plasmids and strains

The plasmids, primers and synthetic sequences used are listed in Supplementary Tables S1–S3. Synthetic gBlock sequences (codon optimized for expression in *E. coli* B strains with sufficient N- and C-terminal homology for assembly by NEBuilder HiFi DNA Assembly) were purchased from Integrated DNA Technologies (IDT), USA. Open reading frames (ORF) were inserted into the

pET24b-6H-MBP vector (Merck, USA), N-terminally fused to a six-histidine tag (6H) and a maltose binding protein (MBP). Constructs were verified by Sanger Sequencing (Australian Genome Research Facility, Australia) (see Supplementary Table S2). *E. coli* 5-alpha competent cells were used for plasmid cloning and maintenance, while *E. coli* BL21(DE3) competent cells were used for biotransformation and protein expression.

The CODM.2 expression plasmid pGWKS134 was constructed using the CODM.2 sequence UniProtKB: D4N502, referred to here as WT CODM for ease of comparison [5, 8, 15]. The plasmid was created by assembling the 1.1 kb CODM.2 gBlock (UMg14), employing *E. coli* codon usage, into the BamHI- and XhoI-linearized pET24b-6H-MBP backbone using NEBuilder.

Other expression vectors (pGWSK127-128, 135–146) (see Supplementary Table S1) were created using the same procedure; by assembling the respective 1.1 kb ORF gBlocks (UMg8-26) (see Supplementary Table S3) into the linearized pET24b-6H-MBP backbone using NEBuilder.

Cell culture and protein expression

Seed cultures were grown overnight at 37 °C in Lysogeny Broth medium (0.5% yeast extract, 1% tryptone and 1% NaCl) containing kanamycin (50 μ g/mL). Protein expression cultures contained 1 mL of the seed cell culture, 50 mL of 2-YT medium (1% yeast extract, 1.6% tryptone and 0.5% NaCl), kanamycin (50 μ g/mL) and 1 mL of glycerol. When the optical density measured at 600 nm (OD_{600}) reached 0.4–0.8 after culturing at 37 °C, protein expression was induced by IPTG addition (0.1 mM) and growth continued for an additional 22 h at 18 °C. Following protein expression, cells were pelleted by centrifugation (3,000 g for 15 min), then resuspended in 15% glycerol to an OD_{600} of 100 in preparation for biotransformation.

Whole cell biotransformation

Biotransformation (20 mL total volume) consisted of the indicated alkaloid (1 mM thebaine or 1 mM codeine), 100 mM phosphate buffer (pH 6.0), 0.5% w/w glucose, 10 μ M iron (II) sulfate ($FeSO_4$) and 10 mM sodium ascorbate and CODM expressing *E. coli* cells of OD_{600} of 10. Reactions were incubated at 24 °C with shaking (220 rpm). Samples were centrifuged (16,000 g for 7 min) with the alkaloid content of the supernatant determined by HPLC. At least three independent replicates were conducted for biotransformation.

Soluble protein expression analysis

SDS-PAGE analysis was used to investigate the soluble protein expressed following cell protein expression. Frozen cell pellets (OD_{600} of 1) were resuspended in Bug-Buster™ protein extraction reagent containing benzonase

(25 units/mL of BugBuster™ reagent) (Merck, USA). Soluble protein containing supernatant was separated from the insoluble components by centrifugation (16,000 g for 20 min at 4 °C). The protein fractions were run on a pre-cast Bolt 8% Bis-Tris Plus Gels (Thermo Fisher Scientific, USA) in MOPS running buffer at 120 V for 60 min. The Precision Plus protein™ Kaleidoscope™ Prestained Protein Standard (Bio-Rad Laboratories, USA) was used to estimate protein sizing. Protein band intensity was calculated by subtracting the background intensity and normalizing the data to the 75 kDa molecular weight band using the Image Lab software (Bio-Rad Laboratories, USA). Soluble protein expression was determined for each independent replicate.

LC-MS analysis of alkaloids

A Shimadzu LCMS-2020 liquid chromatograph mass spectrometer and Onyx Monolithic C18 column (100×4.6 mm, Phenomenex Australia Pty Ltd) was used for opioid quantification. A linear gradient of 0–20% buffer B and at a flow rate increased from 1 mL/min to 2.5 mL/min over 10 min at 28 °C was employed [buffer A: 0.1% TFA in water; buffer B: 0.1% TFA in acetonitrile]. A wavelength of 285 nm was used to detect compounds, with the reference wavelength set at 360 nm. Shimadzu LabSolutions software was used for opioid identification where the opioid retention time was compared to a standard (morphine at 4.0 min, codeine at 6.8 min, oripavine at 7.5 min and thebaine at 10.1 min), peak area was then compared to the standard with a series of concentrations ranging from 25 µg/mL to 500 µg/mL. A 5 µL injection volume was used for sample analysis. The product percentage yield was calculated by dividing the product molar concentration (in mol/L) obtained at a particular experimental time point by the starting substrate molar concentration (in mol/L).

Statistical analysis for mutant CODM comparison

Minitab® was employed for statistical analysis of whole cell biotransformation. Statistical significance with a *p*-value of 0.05 was assessed using a general linear model ANOVA (analysis of variance). A confidence interval of 95% was applied for comparison with the control strain using Dunnett's method.

Enzyme stability predictions

DDmut [20], a deep-learning-based model, was used to predict changes in Gibbs free energy ($\Delta\Delta G$) following introduction of single or multiple amino acid mutations in CODM. As no crystal structure has yet been reported for CODM, structure models were generated using four different tools; MODELLER (using the T6ODM crystal structure [PDB entry 5O7Y] as a template) [19, 40], and using align2d alignment method (best mode was the

one with the least DOPEvalue) [41], SWISS-MODEL (using the T6ODM crystal structure [PDB entry 5O7Y] as a template) [42], AlphaFold2 (accessed via EMBL-EBI's server) [43, 44] and ColabFold-AlphaFold2 (template mode using pb100, using AMBER force fields [45] for protein structure relaxation. The best model was chosen according to the rankings) [46]. The DDmut output from each model was then averaged and employed in correlation analysis.

Statistical analysis for correlation between yield, expression and predicted stability

The relationship between oripavine or morphine yield, DDmut predicted stability, and enzyme expression levels were investigated using Pearson correlation coefficients (*r*). Statistical significance for the correlation coefficient was then assessed using permutation test (with 100,000 resamples per test) with the significance threshold set at *p*<0.05 using SciPy v1.11.4 [47]. For the correlation coefficient between predicted stability versus yield and soluble protein expression level versus yield, one-sided test was used, and the null hypothesis indicates less than or zero correlation between variables. For the correlation coefficient between predicted stability versus soluble protein expression level, two-sided test was used, and the null hypothesis indicates zero correlation between variables. In addition, a 95% confidence interval for the correlation coefficient were calculated using bootstrapping method with 100,000 resamples using SciPy.

For Fig. 3, Supplementary Figs. S4 and S5, Seaborn v0.13.2 [48] was used to generate the regression line and the bootstrap two-sided 95% CI for the regression line as the shaded area in the graphs, based on 100,000 resamples per plot.

Two different analyses were performed, the first included the CODM mutants where select amino acids were replaced with the respective residues from T6ODM, this included mutants from this study and select mutants from our previous study (E259G, E260T, and E259G + E260T) where residues were replaced with the equivalent residue in T6ODM [6]. The second analysis included mutants where the amino acid was systematically varied at only residue 259 from our previous study [6].

Descriptive language used for describing correlation strength is based on [49] and [50].

Abbreviations

WT	Wild type
IPTG	Isopropyl β-D-1-thiogalactopyranoside
PCR	Polymerase chain reaction
HPLC	High-performance liquid chromatography
TFA	Trifluoroacetic acid
6H	Six-histidine tag
MBP	Maltose binding protein
2-YT	2× yeast extract tryptone medium

CODM	Codeine 3-O-demethylase
T6ODM	Thebaine 6-O-demethylase
h	Hour
d	Day
CI	Confidence interval

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13036-025-00477-0>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements

HPLC technical support was provided by the Bio21 Melbourne Mass Spectrometry and Proteomics Facility. We thank Dr. Jeremy Silver of the Statistical Consulting Centre at the University of Melbourne for statistical guidance.

Author contributions

GWKS conceptualized the study, designed and performed the experiments and drafted the manuscript. XL conceptualized the study, assisted with the experiment design and execution and data analysis, revised manuscript and co-supervised GWKS. KWLL performed the computational analysis and drafted the manuscript. GM assisted with conceptualization, data analysis and co-supervision of GWKS. FHF assisted with conceptualization and data analysis, revised manuscript and co-supervised GWKS. SLG acquired funding, conceptualized the study, assisted with the experiment design and data analysis, drafted/revised the manuscript and supervised GWKS.

Funding

Sun Pharmaceutical Industries Australia Pty Ltd and The University of Melbourne supported this research. G.W.K. Spencer is an Ingenium PhD Scholar in the Faculty of Engineering and Information Technology.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 6 September 2024 / Accepted: 6 January 2025

Published online: 19 January 2025

References

- Hagel JM, Facchini PJ. Benzylisoquinoline Alkaloid metabolism: a Century of Discovery and a Brave New World. *Plant Cell Physiol*. 2013;54(5):647–72.
- Dastmalchi M, Chen X, Hagel JM, Chang L, Chen R, Ramasamy S, Yeaman S, Facchini PJ. Neopinone isomerase is involved in codeine and morphine biosynthesis in opium poppy. *Nat Chem Biol*. 2019;15(4):384–90.
- Galanie S, Thodey K, Trenchard JJ, Filsinger Interrante M, Smolke CD. Complete biosynthesis of opioids in yeast. *Science*. 2015;349(6252):1095–100.
- Nakagawa A, Matsumura E, Koyanagi T, Katayama T, Kawano N, Yoshimatsu K, Yamamoto K, Kumagai H, Sato F, Minami H. Total biosynthesis of opiates by stepwise fermentation using engineered *Escherichia coli*. *Nat Commun*. 2016;7:10390.
- Li X, Krysiak-Baltyn K, Richards L, Jarrold A, Stevens GW, Bowser T, Speight RE, Gras SL. High-efficiency biocatalytic conversion of thebaine to codeine. *ACS Omega*. 2020.
- Spencer GWK, Li X, Jarrold A, Gras SL. An industrially applicable *Escherichia coli* platform for bioconversion of thebaine to oripavine and codeine to morphine. *Chem Commun*. 2023;59(41):6251–4.
- Coop A, Lewis JW, Rice KC. Direct and simple O-Demethylation of Thebaine to Oripavine. *J Org Chem*. 1996;61(20):6774–6774.
- Hagel JM, Facchini PJ. Dioxygenases catalyze the O-demethylation steps of morphine biosynthesis in opium poppy. *Nat Chem Biol*. 2010;6(4):273–5.
- Dastmalchi M, Chang L, Torres MA, Ng KKS, Facchini PJ. Codeinone reductase isoforms with differential stability, efficiency and product selectivity in opium poppy. *Plant J*. 2018.
- Thodey K, Galanie S, Smolke CD. A microbial biomanufacturing platform for natural and semisynthetic opioids. *Nat Chem Biol*. 2014;10(10):837–44.
- Fossati E, Narcross L, Ekins A, Falgoutet J-P, Martin VJJ. Synthesis of Morphinan alkaloids in *Saccharomyces cerevisiae*. *PLoS ONE*. 2015;10(4):e0124459.
- Straathof AJ, Panke S, Schmid A. The production of fine chemicals by bio-transformations. *Curr Opin Biotechnol*. 2002;13(6):548–56.
- Runguphan W, Glenn WS, O'Connor SE. Redesign of a dioxygenase in morphine biosynthesis. *Chem Biol*. 2012;19(6):674–8.
- Smolke Christina D, Thodey C, Trenchard I. Methods of producing nor-opioid and nal-opioid benzylisoquinoline alkaloids, WO2018075670A1. In: US: ANTHEIA INC; 2018.
- Guo L, Winzer T, Yang X, Li Y, Ning Z, He Z, Teodor R, Lu Y, Bowser TA, Graham IA, et al. The opium poppy genome and morphinan production. *Science*. 2018;362(6412):343–7.
- Li Q, Ramasamy S, Singh P, Hagel JM, Dunemann SM, Chen X, Chen R, Yu L, Tucker JE, Facchini PJ, et al. Gene clustering and copy number variation in alkaloid metabolic pathways of opium poppy. *Nat Commun*. 2020;11(1):1190.
- Kudo H, Hayashi Y, Arai M. Identification of non-conserved residues essential for improving the hydrocarbon-producing activity of cyanobacterial aldehyde-deformylating oxygenase. *Biotechnol Biofuels*. 2019;12(1):89.
- Kudo H, Hayashi Y, Arai M. Improving hydrocarbon production by engineering cyanobacterial acyl-(acyl carrier protein) reductase. *Biotechnol Biofuels*. 2019;12(1):291.
- Kluza A, Niedzialkowska E, Kurpiewska K, Wojdyla Z, Quesne M, Kot E, Porebski PJ, Borowski T. Crystal structure of thebaine 6-O-demethylase from the morphine biosynthesis pathway. *J Struct Biol*. 2018;202(3):229–35.
- Zhou Y, Pan Q, Pires DEV, Rodrigues Carlos HM, Ascher David B. DDMut: predicting effects of mutations on protein stability using deep learning. *Nucleic Acids Res*. 2023;51(W1):W122–8.
- Ding Z, Guan F, Xu G, Wang Y, Yan Y, Zhang W, Wu N, Yao B, Huang H, Tuller T, et al. MPEPE, a predictive approach to improve protein expression in *E. coli* based on deep learning. *Comput Struct Biotechnol J*. 2022;20:1142–53.
- Jin J, Meng Q, Zeng M, Duan G, Wang E, Guo F. Rapid screening of multi-point mutations for enzyme thermostability modification by utilizing computational tools. *Future Generation Comput Syst*. 2024;160:724–38.
- Werren EA, Peirent ER, Jantti H, Guxholli A, Srivastava KR, Orenstein N, Narayanan V, Wiszniewski W, Dawidziuk M, Gawlinski P, et al. Biallelic variants in CSMD1 are implicated in a neurodevelopmental disorder with intellectual disability and variable cortical malformations. *Cell Death Dis*. 2024;15(5):379.
- Aik W, McDonough MA, Thalhammer A, Chowdhury R, Schofield CJ. Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr Opin Struct Biol*. 2012;22(6):691–700.
- Ostafe R, Fontaine N, Frank D, Ng Fuk Chong M, Prodanovic R, Pandjaitan R, Offmann B, Cadet F, Fischer R. One-shot optimization of multiple enzyme parameters: tailoring glucose oxidase for pH and electron mediators. *Biotechnol Bioeng*. 2020;117(1):17–29.
- Cagiada M, Johansson KE, Valanciute A, Nielsen SV, Hartmann-Petersen R, Yang JJ, Fowler DM, Stein A, Lindorff-Larsen K. Understanding the origins of loss of protein function by analyzing the effects of thousands of variants on activity and abundance. *Mol Biol Evol*. 2021;38(8):3235–46.
- Akdel M, Pires DEV, Pardo EP, Jänes J, Zalevsky AO, Mészáros B, Bryant P, Good LL, Laskowski RA, Pozzati G, et al. A structural biology community assessment of AlphaFold2 applications. *Nat Struct Mol Biol*. 2022;29(11):1056–67.
- Musil M, Stourac J, Bendl J, Brezovsky J, Prokop Z, Zendlulka J, Martinek T, Bednar D, Damborsky J. FireProt: web server for automated design of thermostable proteins. *Nucleic Acids Res*. 2017;45(W1):W393–9.
- Sumbalova L, Stourac J, Martinek T, Bednar D, Damborsky J. HotSpot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information. *Nucleic Acids Res*. 2018;46(W1):W356–62.

30. Musil M, Jezik A, Horackova J, Borko S, Kabourek P, Damborsky J, Bednar D. FireProt 2.0: web-based platform for the fully automated design of thermostable proteins. *Brief Bioinform* 2023, 25(1).
31. Goldenzweig A, Goldsmith M, Hill SE, Gertman O, Laurino P, Ashani Y, Dym O, Unger T, Albeck S, Prilusky J, et al. Automated structure- and sequence-based design of proteins for high bacterial expression and Stability. *Mol Cell* 2016;63(2):337–46.
32. Weinstein JJ, Goldenzweig A, Hoch S, Fleishman SJ. PROSS 2: a new server for the design of stable and highly expressed protein variants. *Bioinform* 2021;37(1):123–5.
33. Bandyopadhyay B, Goldenzweig A, Unger T, Adato O, Fleishman SJ, Unger R, Horovitz A. Local energetic frustration affects the dependence of green fluorescent protein folding on the chaperonin GroEL. *J Biol Chem* 2017;292(50):20583–91.
34. Wang X, Ma R, Xie X, Liu W, Tu T, Zheng F, You S, Ge J, Xie H, Yao B, et al. Thermostability improvement of a *Talaromyces leycettanus* xylanase by rational protein engineering. *Sci Rep* 2017;7(1):15287.
35. Vatansever R, Uras ME, Sen U, Ozyigit II, Filiz E. Isolation of a transcription factor DREB1A gene from *Phaseolus vulgaris* and computational insights into its characterization: protein modeling, docking and mutagenesis. *J Biomol Struct Dyn* 2017;35(14):3107–18.
36. Xia Y, Li X, Yang L, Luo X, Shen W, Cao Y, Peplowski L, Chen X. Development of thermostable sucrose phosphorylase by semi-rational design for efficient biosynthesis of alpha-D-glucosylglycerol. *Appl Microbiol Biotechnol* 2021;105(19):7309–19.
37. Kunka A, Marques SM, Havlasek M, Vasina M, Velatova N, Cengelova L, Kovar D, Damborsky J, Marek M, Bednar D, et al. Advancing enzyme's Stability and Catalytic Efficiency through Synergy of Force-Field calculations, evolutionary analysis, and machine learning. *ACS Catal* 2023;13(19):12506–18.
38. Kouba P, Kohout P, Haddadi F, Bushuiev A, Samusevich R, Sedlar J, Damborsky J, Pluskal T, Sivic J, Mazurenko S. Machine learning-guided protein Engineering. *ACS Catal* 2023;13(21):13863–95.
39. Sugiki S, Niide T, Toya Y, Shimizu H. Logistic regression-guided identification of cofactor specificity-contributing residues in enzyme with sequence datasets partitioned by Catalytic properties. *ACS Synth Biol* 2022;11(12):3973–85.
40. Webb B. A Sali 2016 Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinform* 54 1 561–5637.
41. Shen M-y, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci* 2006;15(11):2507–24.
42. Schwede T, Kopp N Jr, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 2003;31(13):3381–5.
43. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596(7873):583–9.
44. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 2021;50(D1):D439–44.
45. Ponder JW, Case DA. Force Fields for protein simulations. *Adv Protein Chem* Volume 66. Academic; 2003. pp. 27–85.
46. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods* 2022;19(6):679–82.
47. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, et al. Author correction: SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;17(3):352–352.
48. Waskom ML. Seaborn: statistical data visualization. *J Open Source Softw* 2021;6(60):3021–4.
49. Akoglu H. User's guide to correlation coefficients. *Turk J Emerg Med* 2018;18(3):91–3.
50. Chan YH. Biostatistics 104: correlational analysis. *Singap Med J* 2003;44(12):614–9.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.