White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/201825/

Version: Accepted Version

Article:
Grogan, Gideon James, Gilio, Amelia, Steflik, Jeremy et al. (2023) Engineering of a Reductive Aminase to Enable the Synthesis of a Key Intermediate to a CDK 2/4/6 Inhibitor. ACS Catalysis. ISSN: 2155-5435

https://doi.org/10.1021/acscatal.3c01534

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

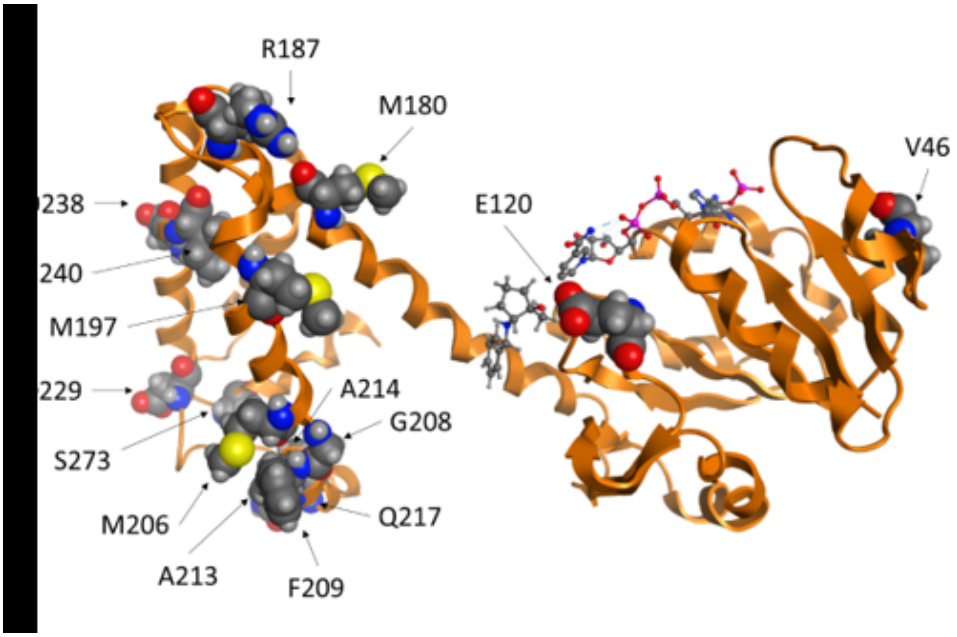
UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

This document is confidential and is proprietary to the American Chemical Society and its authors. Do not copy or disclose without written permission. If you have received this item in error, notify the sender and delete all copies.

Engineering of a reductive aminase to enable the synthesis of a key intermediate to a CDK 2/4/6 inhibitor

Journal:	*ACS Catalysis*
Manuscript ID	cs-2023-01534d.R2
Manuscript Type:	Article
Date Submitted by the Author:	05-Jul-2023
Complete List of Authors:	Steflik, Jeremy; Pfizer Global Research and Development, PSSM-CRD-Biocatalysis Gilio, Amelia; University of York, York Structural Biology Laboratories Burns, Michael; Pfizer, Inc., Worldwide Research and Developmet Grogan, Gideon; University of York, Department of Chemistry Kumar, Rajesh; Pfizer worldwide R & D, Chemical R & D Lewis, Russell; Pfizer, Chemical Research and Development Martinez, Carlos; Pfizer Global Research and Development, Chemical Development;

SCHOLARONE™
Manuscripts



60x38mm (220 x 220 DPI)

Engineering of a reductive aminase to enable the synthesis of a key intermediate to a CDK 2/4/6 inhibitor

Jeremy Steflik[1]*, Amelia Gilio[2], Michael Burns[1], Gideon Grogan[2], Rajesh Kumar[1], Russell Lewis[1], Carlos Martinez[1]

Affiliations:

[1]Chemical Research & Development, Pfizer Worldwide Research & Development, Groton, Connecticut 06340, United States.

[2] York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York, YO10 5DD (UK)

*Correspondence to: Jeremy.Steflik@pfizer.com

Abstract

Biocatalytic reductive amination reactions with reductive aminases (RedAms) are an emerging transformation with high potential value for pharmaceutical synthesis. Here, we report the identification and engineering of a RedAm to catalyze a reductive amination reaction making a key intermediate in the synthesis of an investigational cyclin-dependent kinase (CDK) inhibitor, using the relatively bulky benzylamine as a nucleophile. The engineered enzyme contains 6 mutations with respect to the wild-type and displays high productivity at high substrate concentrations (50-fold improved over wild-type). After identifying the optimized enzyme variant, crystal structures of both the wild-type and mutant enzymes were solved and used to rationalize how structural changes to the RedAm improved its performance under process conditions. Results suggest that mutations affecting both substrate binding and enzyme thermostability contribute to improved enzyme performance. By enabling multikilogram-

scale synthesis of the chiral intermediate, this work highlights the versatility and industrial utility of RedAm-catalyzed reductive amination.

Keywords

Biocatalysis

Enzyme Engineering

Imine Reductase

Reductive amination

Amines

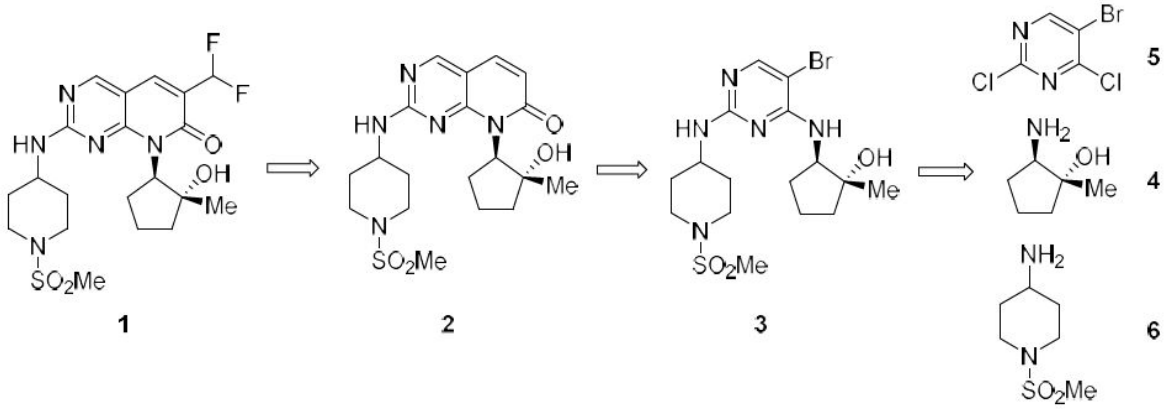
Introduction

The cyclin-dependent kinases (CDKs) are a 21-member family of serine–threonine kinases that are regulated by the formation of heterodimeric complexes with their cyclin-binding partners. CDK–cyclin complexes are involved in a wide array of cellular processes, including transcriptional regulation, splicing, DNA repair, and regulation of the cell cycle.[1] The ordered activation of the cell cycle CDKs (CDKs 1, 2, 4, and 6)[2] regulates progression through the four phases of the cell cycle (G1, S, G2, and M)[3] and their dysregulation can lead to sustained cell proliferation, a hallmark of cancer.[4]

Compound **1** (**Figure 1**) is a CDK2/4/6 inhibitor currently being developed by Pfizer as an oncology candidate.[5] The proposed manufacturing route of **1** involves consecutive S_NAr reactions of 5-bromo-2,4-dichloropyrimidine **5** with amine nucleophiles **4** and **6**, followed by a two-step Heck coupling/cyclization, and finally a difluoromethylation. The synthesis of the chiral amino alcohol **4** proved to be a challenge and several enabling synthetic routes were investigated, including an epoxide opening of **8** followed by a resolution and a transaminase reaction of **9** (**Figure 2**).[6] Ultimately, the cost

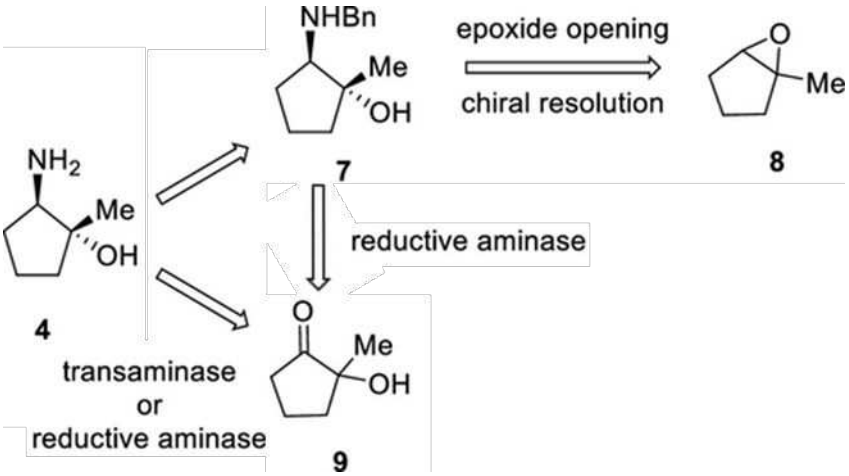
and yield limitations of the chemical and transaminase routes led to the investigation of a third route: a chiral resolution using an enzymatic reductive amination.

Figure 1. Retrosynthesis of CDK 2/4/6 Inhibitor 1



Reductive amination is frequently used to synthesize a variety of amines.[7] Chemical methods of reductive amination typically use stoichiometric amounts of reducing agents at low temperature or use transition metal catalysis, making them undesirable for use at commercial scale.[7, 8] More recently, a subset of imine reductase enzymes (IREDs), dubbed reductive aminases (RedAms), have been shown to catalyze intermolecular reductive amination directly[9], and numerous studies have been reported on the identification and characterization of enzymes with broad tolerance for carbonyl substrates and amine nucleophiles. [10-13]

Figure 2. Enabling Synthesis and Enzymatic Synthesis of 4



RedAms have the potential to be highly valuable catalysts for the synthesis of Active Pharmaceutical Ingredients (APIs), but their application to manufacturing processes[14, 15] is still maturing and many challenges remain. The reactions reported in academic labs predominantly use carbonyl compounds at relatively low concentrations paired with simple amines, often used in large excess to drive the reactions to completion.[10-13, 16, 17] Both of these limitations must be overcome for a viable commercial manufacturing process.

Herein we report the successful directed evolution of a RedAm to enable the synthesis of the key chiral intermediate **4**. Not only was this enzymatic reaction critical to delivering a manufacturing process, it was also the first demonstrated at kilogram scale using benzylamine as a co-substate.[6]

Results and Discussion

Route selection and identification of RedAm Enzyme.

The Pfizer in-house IRED screening panel consists of 88 wild type IRED enzymes. An initial screen of this panel was conducted to identify enzymes capable of the reductive amination of ketone **9** with benzylamine to give the chiral benzylamino-alcohol **7**. **Table 1** summarizes these screening results.

Table 1. Screen of RedAms for conversion of 9 to 7. Conditions: 100 mM Potassium phosphate buffer, 0.2 mL; RedAm, 1.0 mg; **9**, 20 mM; benzylamine HCl, 40 mM; glucose, 24 mM; glucose dehydrogenase, 0.5 mg/mL; NADP+, 0.5 mM; DMSO, 5%; 30 °C; pH8.0, 1000 rpm on a Thermomixer; 18 hours. Assayed by chiral SFC.

Enzyme Designation	Accession ID	% Conversion	%*ee*
IR006	YP_007358783	20	-27
IR007	WP_005159789	21	96
IR017	WP_010638508	11	-100
IR040	WP_018429927	18	1
IR044	WP_023809753	16	-1
IR048	WP_027512056	18	37
IR050	WP_029347354	20	27
IR056	WP_034788492	16	8
IR059	WP_037099331	18	0
IR061	WP_038576230	17	29
IR064	WP_046711811	20	-54
IR077	WP_053252429	18	10
IR083	Q1EQE0	22	-75
IR109	AUD39487	16	-71

Several enzymes were identified with the desired selectivity, but only one of these, IR007, a hypothetical protein originally identified in *Amycolatopsis azurea*, exhibited the desired selectivity to a high degree, showing 21% conversion with 96% enantiomeric excess (% *ee*) under the conditions tested. The reaction with IR007 was scaled to 0.5 g under pH control using 200 mM **9** (22.8 g/L) substrate loading and 4.6 g/L IR007 lyophilized cell-free lysate (20% by weight with respect to **9)**. This reaction gave 35% conversion to **7** with 97% *ee* after 50 hours. Further process improvements optimizing reaction temperature and DMSO concentration increased the conversion to 38% after 48 hours while maintaining the high % *ee*. During this reaction optimization it was observed that IR007 was sensitive to both reaction temperature and to benzylamine, both of which limit the possible process conditions. Though IR007 is a highly selective enzyme, its activity and stability were inadequate to deliver a scalable process.

Enzyme engineering.

After initial evaluation, it was estimated that over 50-aggregate fold improvement (**Table 2**) in enzyme performance over wild type (WT) IR007 was needed to enable a manufacturing process. We therefore embarked on a protein engineering effort to improve the performance of IR007, targeting increased tolerance to ketone **9** and benzylamine, and higher activity while retaining high selectivity.

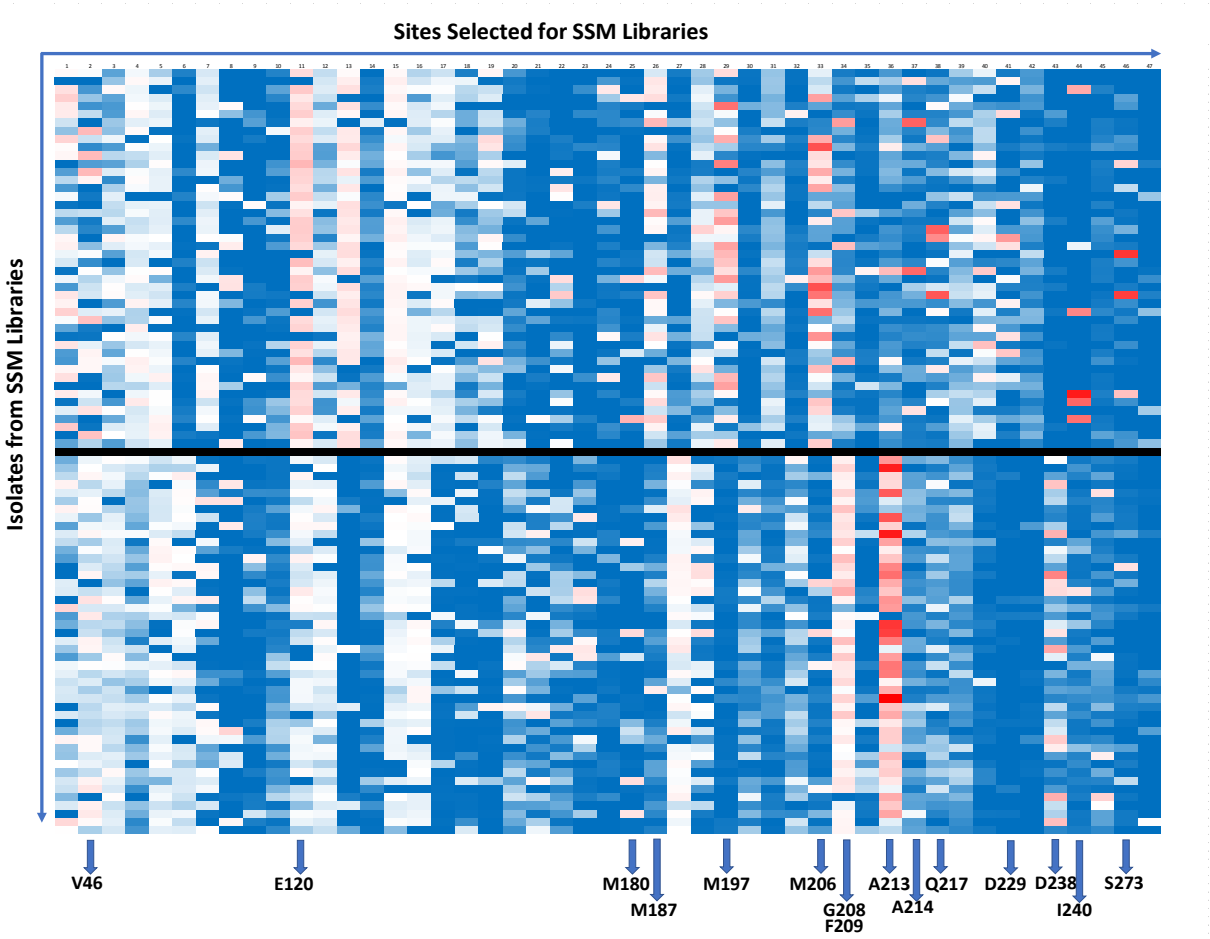
Table 2. Process targets for commercial scale manufacturing in comparison to initial performance of WT IR007

Parameter	Initial performance	Target	Fold Improvement
Substrate loading	200 mM	500 mM	2.5x
% Conversion	38%	45%	1.1x
Enzyme loading	20 wt%	2 wt%	10x
Time	48h	24h	2x

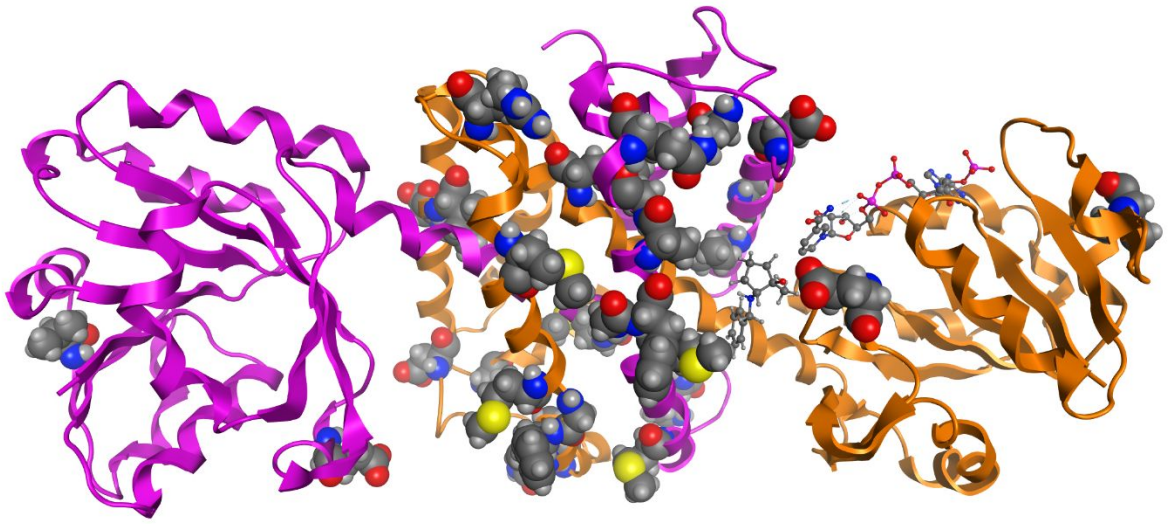
For Phase 1 of the engineering effort, initial site selection and library design was performed using a structural homology model as there was limited structural information available for this class of enzyme. A total of 94 out of 286 positions in IR007 were selected for single site saturation mutagenesis (SSM). These sites comprise both active site and secondary shell residues, as well as other residues identified by bioinformatic analysis. Phase 1 SSM libraries were initially screened at 35 g/L of ketone **9** with an enzyme load equivalent of 10 g/L of wet cell paste added as 40 µl of liquid expression broth (**Figure 3**) and assayed using achiral uPLC. This rapidly identified substitutions at 15 positions (**Table 3a**) that exhibited greater than 1.5 fold improvement over the parent (FIOP). The most active variants exhibited between 1.61 and 2.8 FIOP, and included the single substitutions M197W, M206H/R/K, A213N/R/F/H/K, A214H, I240L, Q217N, and S273G. Of note, upon rescreening using chiral SFC, the enantioselectivity of the variants remained unchanged.

Fig 3. Screening data from enzyme engineering phase 1 SSM libraries. a) Heat map of 94 sites tested, each column represents two sites (Supplementary Table S1 and Table S2 for additional details), sites with improved performance are shown in red color; b) Dimer view of protein homology model showing the location of sites exhibiting >1.5 FIOP; c) Protein monomer view showing the location of most active substitutions with product amine **7** and cofactor NADP+ shown.

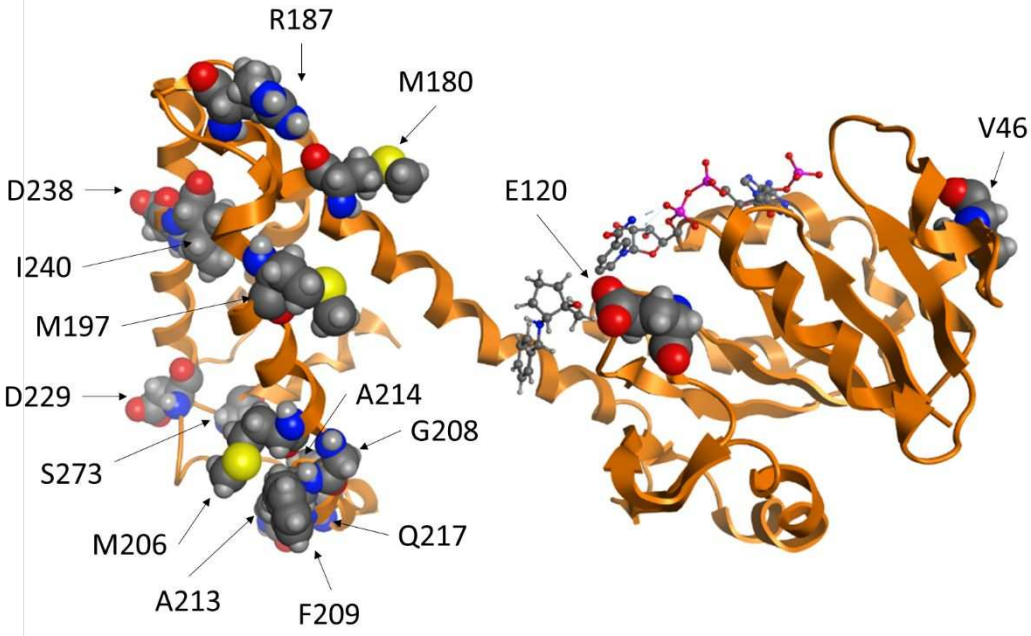
a)



b)



c)


R187
M180
V46
D238
E120
I240
M197
D229
A214
G208
S273
M206
Q217
A213
F209

Beneficial substitutions were identified at 24 positions (**Table 3b**) with greater than 1.3 FIOP. Interestingly, multiple substitutions were found at many of the positions where the characteristics of the substituted amino acids were vastly different. For instance, IR007 mutants exhibited significantly increased activity when residue A213 was substituted with negatively charged Glu, positively charged Lys, polar Asn, and other nonpolar amino acid of varying sizes (Val, Phe, Ile, Pro). The number and variety of beneficial substitutions at these 24 positions called for a more complex and multipronged approach for the recombination phases of protein engineering.

Table 3. A) Sites exhibiting greater than 1.5 FIOP in phase 1 screens. For each site, the degree of conservation based on an alignment of the closest 250 non-redundant sequences is given, as well as the average, median, and max FIOP values across the entire saturation library. The corresponding substitution with the max FIOP is shown on the right. B) Substitutions with greater than 1.3 FIOP in phase 1 screens.

A)

		FIOP			
Site	**Site Conservation**	**Average**	**Median**	**Max**	**Substitution**
V46	0.07	0.6	0.7	1.51	Ser
E120	0.38	1.15	1.29	1.52	Arg
M180	0.3	0.13	0	1.55	Leu
R187	0.28	0.89	1.02	1.51	Lys
M197	0.04	0.92	0.99	2.03	Trp
M206	0.32	1.02	0.99	2.26	Lys
G208	0.18	0.4	0.34	1.71	Pro
F209	0.2	1.14	1.13	1.6	Tyr
A213	0.37	1.54	1.47	2.84	Phe
A214	0.54	0.29	0.12	2.14	His
Q217	0.43	0.45	0.19	2.22	Asn
D229	0.05	0.35	0.1	1.73	Asn
D238	0.88	0.69	0.66	1.9	Gly
I240	0.85	0.24	0	2.65	Leu
S273	0.53	0.26	0.04	2.46	Gly

B)

E20A	L169T	G208P	E227Q
E20Q	N174A	F209G	D229N
E20T	M180L	F209K	D238G
N21A	V186F	F209N	D238N
V46A	R187H	F209Q	D238S
V46S	R187K	F209S	I240L
D47K	R187Q	F209Y	I240V
A112C	R187S	E212P	S273G
E120A	R187W	A213E	
E120F	M197A	A213F	
E120G	M197F	A213I	
E120I	M197L	A213K	
E120K	M197V	A213N	
E120L	M197W	A213P	
E120N	M197Y	A213R	
E120P	M206G	A213S	
E120Q	M206H	A213V	
E120R	M206K	A214H	
E120T	M206Q	Q217N	
E120V	M206R	Q217R	

In phase 2 of enzyme engineering, targeted recombinations were first made to determine interactions between positions in close proximity based on the structural homology model available at the time. It was predicted that substitutions at E120, A213, A214, and Q217 could all change the charge landscape of a predicted channel leading into the binding site (**Figure S2**), with the highest performing substitutions adding positively charged amino acids. Double substitution variants at these positions showed that many of the possible combinations were additive. Modest improvements in activity were seen when combining substitutions close in the primary sequence. The variant IR007-41 contained the substitutions A213R and A214H and exhibited 3.17 FIOP. Similarly, the A214H with Q217N substitutions in IR007-42 produced a variant with 3.57 FIOP. The larger gains in activity, however, were seen with combinations involving both sides of the proposed access channel with the incorporation of substitutions at E120. The highest performing variant from this series of substitutions was IR007-49 (E120R, A213F), which exhibited 5.53

FIOP. Unfortunately, triple substitution variants using templates from this series showed decreases in activity. Combinations of substitutions at positions M206, G208, and F209, all within the active site, showed only very small improvements in activity compared to the single substitution variants used as templates; the maximum FIOP observed was 2.38 FIOP for IR007-57 (M206H, F209S). Combinations at D238 and I240, both predicted to alter the position of the highly conserved L169 residue in the active site, were found to be additive, with the best combination being D238G and I240L in IR007-53, showing 4.63 FIOP.

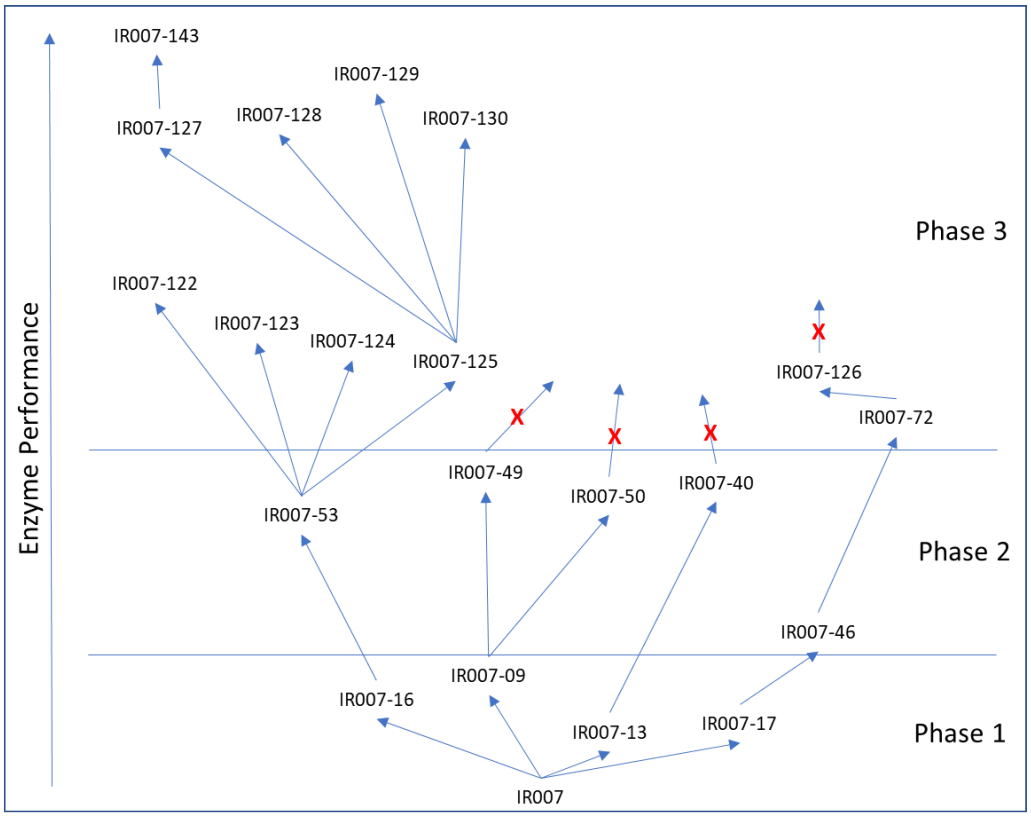
The initial phase 2 screening results revealed that double substitutions in close proximity to one another were largely additive when combined. However, combining these beneficial access channel and active site substitutions pairs resulted in decreased activity compared to the double substitution variants used as templates. In particular, neither substitutions at positions 206/209 nor 238/240 recombined favorably with substitutions at positions of 213/214/217.

Phase 3 of enzyme engineering initially utilized several single and double variants identified in phase 1 and early phase 2 as templates for targeted substitutions, iterative saturation mutagenesis (ISM) and combinatorial active site saturation testing (CASTing) libraries, as well as libraries randomly incorporating all the beneficial substitutions identified in phase 1 (**Table 3b**).

As we progressed and identified improved variants, these were in turn used as templates for additional mutagenesis (**Figure 4a**). As enzyme activity improved, the screening conditions used became more stringent, gradually decreasing the enzyme load and increasing the substrate concentration (**Figure 4b**).

Figure 4. Evolution of enzyme and screening conditions. A) Enzyme parent template, showing top hits from each phase of enzyme engineering. B) Screening conditions for different phases of enzyme engineering.

A)



B)

	Condition			
Screening Parameters	**1**	**2**	**3**	**4**
Substrate Loading (g/L)	35	35	50	50
Enzyme Loading (µL expression broth)	40	20	20	10
Enzyme Loading (wt%)	2.57	1.29	0.90	0.45
Engineering Phase	1	2	2/3	3

Due to the number of beneficial substitutions and templates being investigated, as well the large number of possible combinations with CASTing and random recombination libraries, an exhaustive screen of the constructed libraries was not feasible. Instead, a sample set of approximately 400 variants was screened from each library, with additional variants screened if any improvements were seen in the library. This approach undoubtedly missed beneficial combinations but was required based on assay and resource limitations. Table 4 summarizes variant performance, highlighting several of the most improved variants from across the three phases of engineering. The highest performing variant, IR007-143, contains six substitutions compared to the WT enzyme: E120A, M197W, D238G, I240L, M206S, and A213P. Of note, that although the M206 position was identified as having substitutions that improved activity in phase 1, the M206C/N substitutions seen in IR007-128 and IR007-129 had deleterious effects on activity (0.88 and 0.90 FIOP), highlighting the need to revisit substitutions previously observed not to be beneficial when working with advanced variant backbones.

Table 4

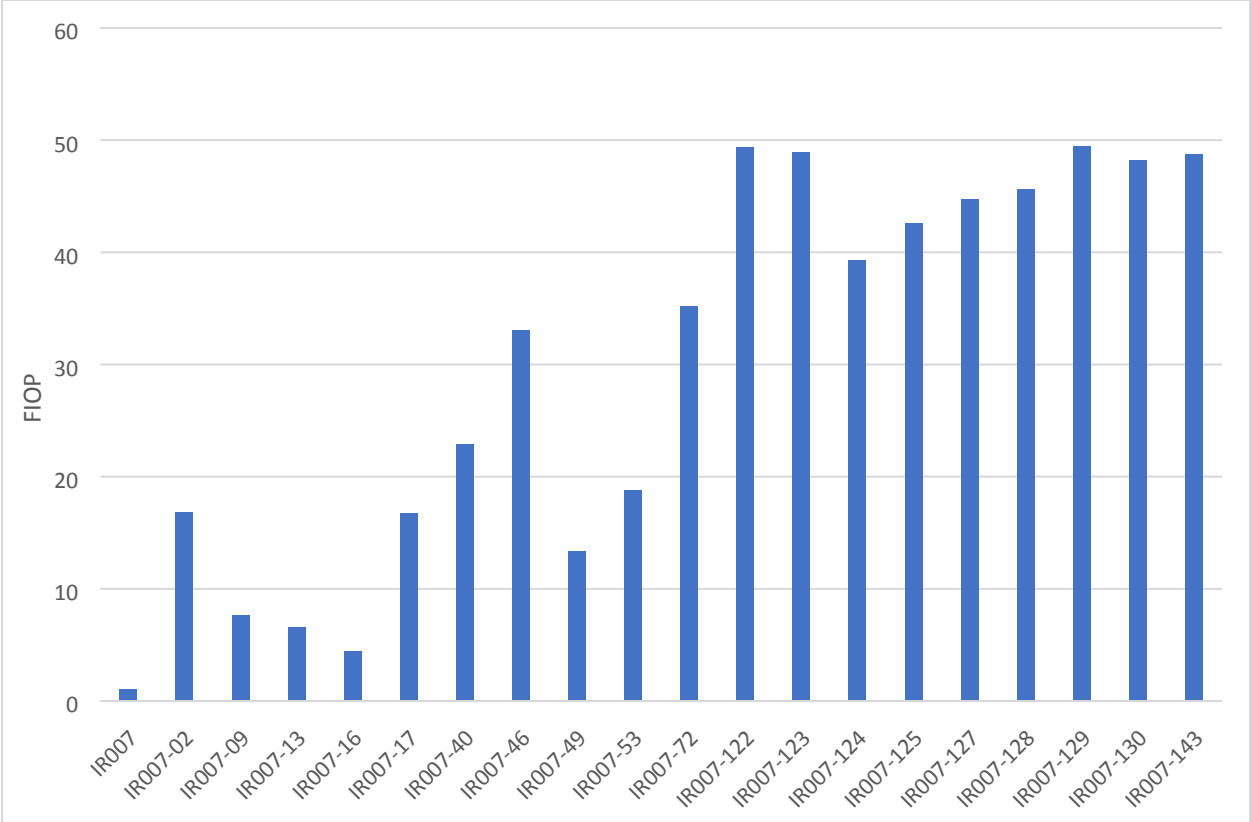
Variant	Substitutions	Screening Condition	FIOP
IR007-E120A	E120A	1	1.35
IR007-E120V	E120V	1	1.31
IR007-01	E120K	1	1.43
IR007-02	E120R	1	1.52
IR007-09	A213F	1	2.84
IR007-13	A214H	1	2.14
IR007-16	I240L	1	2.65
IR007-17	Q217N	1	2.22
IR007-26	M206H	1	2.26
IR007-37	E120R, A213R	1	4.35
IR007-40	E120R, A214H	1	4.89
IR007-46	E120R, Q217N	1	2.92
IR007-49	E120R, A213F	1	5.53
IR007-50	E120Q, A213F	1	4.81
IR007-53	D238G, I240L	1	4.63
IR007-57	M206H, F209S	1	2.38
IR007-72	V46A, E120R, Q217N	2	5.13

IR007-72	V46A, E120R, Q217N	3	24.6*
IR007-122	E120K, M197W, D238G, I240L	3	29.7*
IR007-123	V46A, E120R, R187W, D238G, I240L	3	28.9*
IR007-124	E120V, M197K, M206S, D238G, I240L	3	27.2*
IR007-125	E120A, M197W, D238G, I240L	3	26.4*
IR007-126	V46A, E120R, Q217N, D238G	3	23.0*
IR007-127	E120A, M197W, D238G, I240L, M206S	4	32.7*
IR007-128	E120A, M197W, D238G, I240L, M206C	4	33.7*
IR007-129	E120A, M197W, D238G, I240L, M206N	4	36.1*
IR007-130	E120A, M197W, D238G, I240L, M206R	4	33.3*
IR007-143	E120A, M197W, D238G, I240L, M206S, A213P	4	39.1*

Data represents n≥2.
*FIOP calculated using variants from previous phases of engineering as the comparator.

As the screening conditions changed over the course of the program, we moved from using the WT to using advanced variants as the comparator, as the WT has very little activity under the final screening conditions. FIOP was calculated using the comparator's performance versus the WT in milder conditions as a multiplier to obtain the overall FIOP. After the final variant was identified, a selection of variants from across the program were prepared to compare directly under the final screening conditions (**Figure 5**).

Figure 5. Selected Variant Performance under the final screening conditions: 100 mM potassium phosphate buffer, 0.2 mL; RedAm culture, 10 µl; **9**, 50 g/L; benzylamine HCl, 0.6 equivalents; DMSO, 7.5%; NADP+, 0.25 mg/mL; GDH, 0.25 mg/mL; 30 °C; pH 8.0, 1000 rpm on a Thermomixer; 18 h. Assayed using GC.



Though overall trends in FIOP largely agreed with those calculated using the comparator, IR007-02 (E120R) is an obvious exception. Under the final lower enzyme and higher substrate loading conditions, this variant was dramatically improved when compared to other phase 1 variants, showing 16.79 FIOP. Similarly, IR007-17 (Q217N), stands out among the other phase 1 variants. Neither E120R nor Q217N are found in the final variant, leading one to question if these substitutions were dropped from the engineering program prematurely or if the strategies employed would have been different had this information been available earlier in the program. A closer examination of the variants screened revealed this is not the case. Due to the large number of beneficial substitutions identified at E120 in phase 1 screening, this site was routinely saturated in subsequent phases of engineering when the templates used did not already contain a substitution at E120. Q217N was similarly investigated in all templates used in phases 2 and 3. Additionally, IR007-46, which contains both E120R and Q217N, was investigated with CASTing and random recombination libraries, from which IR007-72 was identified. This variant was

extensively mutated in phase 3, eventually leading to an evolutionary dead end at IR007-126, with no improvements identified upon further mutagenesis. As noted previously, substitutions at E120 and Q217 are among those predicted to change the charge landscape of a proposed channel leading to the binding site, with the two sites on opposite sides of the channel.

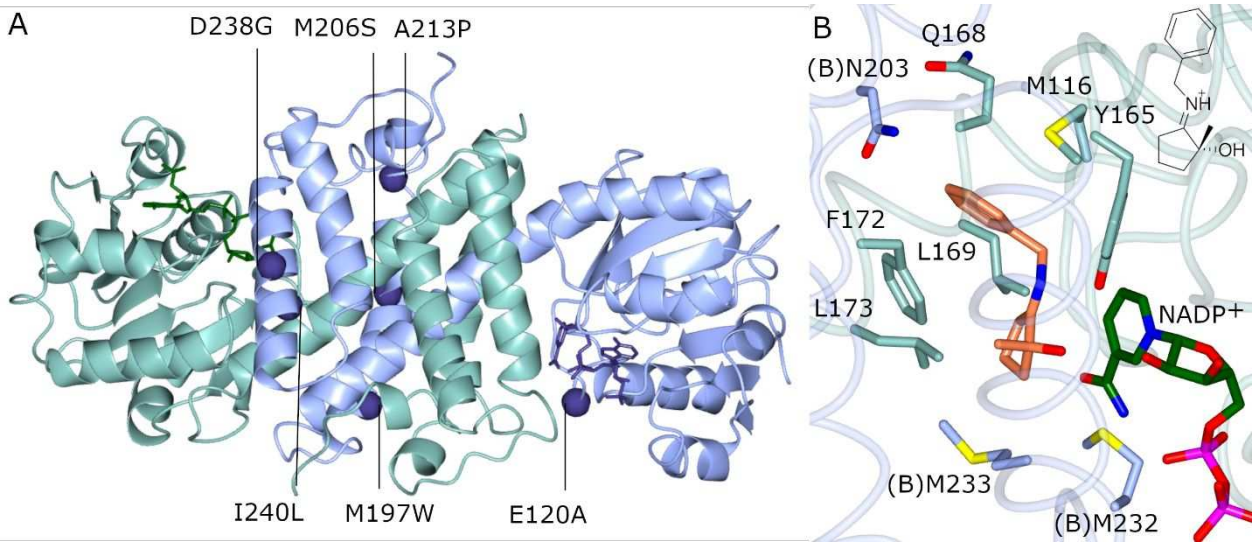
Final Variant Analysis

To shed light on the molecular context of the beneficial mutations, the structures of IR007 and the mutant IR007-143 were solved using X-ray crystallography (PDB IDs 8BJ5 and 8BK1, respectively). The wild-type IR007 displays the canonical IRED-fold;[18] the monomer comprises an N-terminal Rossmann cofactor-binding domain linked through a long helix to a C-terminal helical bundle. Two monomers associate with reciprocal domain sharing to form two active sites per dimer at the interface of the N-terminal domain of one monomer and the C-terminal domain of its partner **Figure 6a**. Each active site contained density for the cofactor $NADP^+$. Of IRED structures in the PDB, the IR007 monomer is most similar to the (*S*)-selective imine reductase enzymes from *Stackebrandtia nassauensis* (6JIT; 56% id; 1.0 Å over 292 C-alphas) and *Nocardiopsis halophila* (4D3S;[19] 45% id; 1.8 Å over 285 C-alphas), as determined by the DALI server.[20] Compared to the homology model used in the initial site selection, the RSMD of the IR007 monomer is 1.9 Å over 292 C-alphas.

The IR007 active site is characterized by residue Y165 which projects down from the interdomain helix into the active site in close proximity to the nicotinamide ring of the cofactor. Equivalent tyrosine residues have been implicated in catalysis in IREDs, especially those for which the selectivity for the model substrate 2-methylpyrroline is (*S*) [19,21,22]. Mutation of these tyrosine residues to F or A can render the enzymes inactive.[19, 21, 22] The helix containing Y165 also contributes L169, F172 and L176 to the left-hand wall of the active site as seen in **Figure 6b**, together providing a largely hydrophobic environment for substrate binding. The front of the active site, which is formed in the closed conformation of the

enzyme, is dominated by two methionine residues from the 'B' chain, M232 and M233, the side chains of which project towards the nicotinamide ring.

Figure 6: A) IR007 dimer showing location of mutation sites in variant IR007-143. B) IR007-143 active site with iminium ion intermediate docked with Autodock Vina.[23]



While we were not able to obtain a ternary complex with both cofactor and substrate **9** or product **7**, the intermediate **(R,Z)-N-(2-hydroxy-2-methylcyclopentylidene)-1-phenylmethanaminium** was modelled into the active site using Autodock Vina (**Figure 6b**).[23] The docking results suggest an aromatic binding pocket for the benzyl group is formed by the side chains of M116 and F172 at the rear of the active site as shown. It also places the methyl group of the favored (*R*)-enantiomer of the chiral alcohol in a pocket formed by F172, L173 and M233. The imine is bound in the correct orientation to receive hydride from the C4 atom of the cofactor nicotinamide ring at the *si*-face of the imine, which would give the experimentally observed *(R*)-enantiomer at the amine stereocenter. The recently deposited structure of 6JIT in complex with **1-(2-phenylethyl)-3,4-dihydroisoquinoline**[24] affords an informative comparison with IR007 as much of the hydrophobic character of the substrate binding sites within the two enzymes is retained. Residues L169 (IR007) (L175 in 6JIT), L176 (L182), M232 (M245) M233 (M246) are conserved,

but F172 of IR007 is replaced by smaller L178 in 6JIT (**Figure S1**). The substitution of Y210 in IR007 for M216 in 6JIT at the rear of the active site appears to be crucial for substrate specificity determination, as superimposition of the two structures shows Y210 in IR007 clashing with the aromatic ring of the phenylethyl side chain in the 6JIT complex, reflective of a smaller active site overall in IR007.

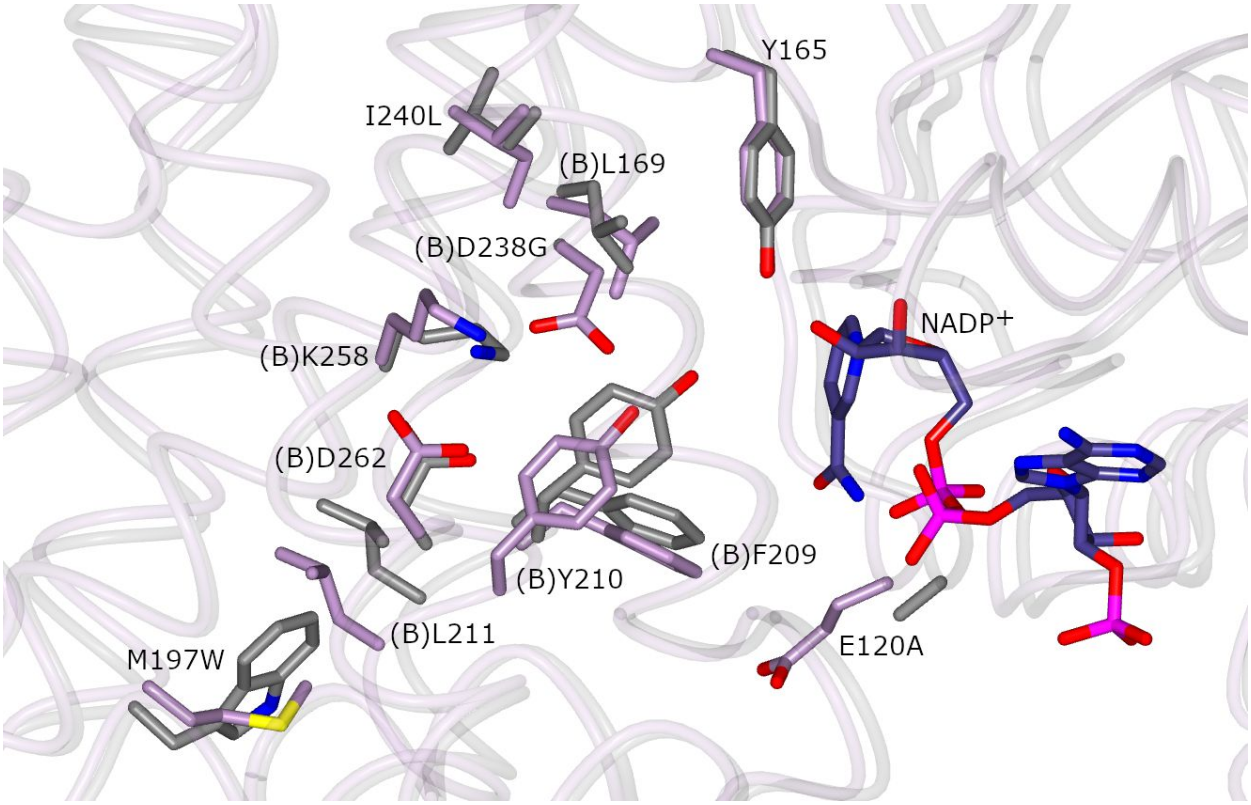
Structural Consequences of Mutation in the IR007-143 variant

The structures of both IR007 and IR007-143 variant permit analysis of the molecular context of these improving mutations: E120A is on the periphery of the enzyme at the entrance to the active site access channel; although E120 appears to make no close interactions with other residues, its mutation to hydrophobic alanine contributes to a change in the charge landscape of the access channel (**Figure S2**). A213 is also part of the active site access channel and while again making no close interactions with other amino acids, its mutation to proline indirectly changes the charge landscape by moving adjacent E212 away from the channel opening, further reducing the negative charge at this site.

M197W intercalates with L211 and L215 of the partner monomer at the interface of C-terminal domains at the base of the enzyme. Due to the increased size of the side chain, the M197W mutation pushes L211 1.8 Å from its position in the wild-type enzyme and also causes the position of F209 and Y210 to shift, changing the topology of the active site. M206S is at the interface of the C-terminal domain and the N-terminal domain of its neighbor, again, not forming any close interactions with other residues. However, the M206S mutation introduces an H-bond to the backbone carbonyl of N203 on the same helix through the hydroxymethyl side chain. D238 forms a salt bridge with K258 on the neighboring helix. This salt bridge interaction is disrupted by the D238G mutation, and as a result, the K258 side chain changes orientation and makes an alternative salt bridge with D262. I240 interacts with L169 and L173 on the interdomain helix of the neighboring monomer, and the I240L mutation permits the relocation of L169

into space made available by the smaller amino acid, thus making more space available within the active site. Changes within the active site are detailed in **Figure 7**.

Figure 7: Superimposition of active sites of WT IR007 (carbon atoms in purple) and IR007-143 (grey) illustrating the location and effects of mutations E120A, M197W, D238G and I240L



The mutations to IR007 reveal a mixture of substitutions both within and distant from the active site, with the implication that multiple effects may contribute to improved performance by the IR007-143 variant. An attempt was made to determine kinetic constants for both WT and variant (**Figure S3**). In each case, while kinetic plots reached saturation, pronounced substrate inhibition was observed, although data at substrate concentrations higher than 350 mM could not be obtained, so data were not sufficiently reliable to calculate values for *K*i. However, the plots strongly suggest that $V\text{max}_{app}$ was increased for the variant as much as 2-fold, and that the K_i would be considerably increased in the variant, perhaps as a result of

active site modifications, and presumably leading to improved performance at the higher concentration under process conditions. The melting temperature (T_m) of WT and variant were also measured in the presence and absence of added cofactor using nano-DSF (differential scanning fluorimetry, **Table S2**). The T_m of the variant was revealed to be increased by approximately 5°C in each case, indicative of greater process stability, and perhaps attributable to the modifications made to interdomain interactions in the variant. Increases in melting temperature, and also half-life, were cited as major reasons for improved enzyme performance in the creation of the 'M3' variant of 'IR46' from *Saccarothrix espanaensis*[14] and 'IRED88' from *Streptomyces spectabilis*[17] from researchers at GSK and Novartis respectively.

Conclusions

Herein, we described how the engineering of a RedAm from *Amycolatopsis azurea* enabled the development of a substantially improved enzyme catalyst. The final engineered enzyme, IR007-143, supported manufacturing efforts at kilogram scale with a substrate load of 50 g/L, 43% conversion with an isolated 98.4% *ee* [6], and reached process targets described in **Table 2**. The x-ray crystallography, molecular modelling, and structural analysis presented suggest the modulated distribution of charges leading into the active site (**Figure S2**), as well as the change in available space illustrated in **Figure 7**, led to improvements in the kinetic properties of the enzyme and enabled higher substrate concentrations, increasing Vmax twofold. The T_m of the final variant was increased by approximately 5°C and was evidenced by increased operational stability under process conditions. The enabled process is more sustainable and efficient compared chemical routes used previously.[6] This study represents the first enzyme engineered to support a reductive amination using benzylamine as amine partner, which is very relevant in the synthesis of amine derivatives that are highly water soluble.

Experimental

IRED Panel Screening

The Pfizer IRED/RedAm screening panel consists of 88 wild type enzymes described in multiple sources[11, 12, 25, 26] and prepared as lyophilized cell free extract (Prozomix, Haltwhistle, UK). 1.0 mg of each enzyme was arrayed in a 96-well deep well plate prior to conducting the screen. Final screening conditions: 100 mM Potassium phosphate buffer, 0.2 mL; RedAm, 1.0 mg; ketone **9**, 20 mM; benzylamine HCl, 40 mM; glucose, 24 mM; glucose dehydrogenase, 0.5 mg/mL; NADP+, 0.5 mM; DMSO, 5%; 30 °C; pH 8.0, 1000 rpm on a Thermomixer; 18 hours. After incubation, the reactions were quenched with 0.60 mL acetonitrile, centrifuged to remove solids, and analyzed by chiral SFC.

Template identification and homology model construction

The Chemical Computing Group's Molecular Operating Environment (MOE) suite of tools was used to calculate amino acid conservation rates at each position along the protein backbone using an alignment file containing the closest 250 non-redundant sequences identified through a BLAST search (E-value ≤4.6e-99). A template search for homology modeling was performed in MOE and a BLAST search of Protein Data Bank proteins. The selected template, 4D3D from *Bacillus cereus* BAG3X2, resulted in an alignment with 50% identity to IR007 and a Pairwise % Positive (BLSM62) of 68%. A homology model was constructed using MOE. For this model, the conserved cofactor binding motif GxGxxG was constrained between the template and IR007 sequences, as were the highly conserved active site residues corresponding to Tyr165 and Phe172.

Identification of positions for site saturation mutagenesis

Binding site residues in the homology model were identified using the Site Finder function in MOE. Secondary shell residues were identified by selecting residues within 4.5 Å of binding site residues. Additional tools in MOE were used to predict rigidifying substitutions as well as identify hydrophobic

patches on the protein surface. All residues were sorted and scored by calculated conservation rates, variability of amino acids seen in the alignment mentioned previously (number of amino acids seen and chemical characteristics), and distance from the binding site. 94 positions were selected for site saturation mutagenesis (SSM) library creation.

Data collection, structure solution and refinement

The datasets described in this report were collected at the Diamond Light Source, Didcot, Oxfordshire, U.K. on beamlines I03 and I04. Data were processed and integrated using XDS[27] and scaled using SCALA[28] included in the Xia2 processing system.[29] Data collection statistics are provided in **Table S3**. The crystals of WT IR007 and IR007-143 were obtained in space groups $H3_2$ and $P2_12_12_1$ respectively, with four molecules in the asymmetric unit in each case; the structures were solved by molecular replacement using MOLREP[30] with the monomer of the IRED from *Stackebrandtia nassauensis* (PDB ID 6JIZ) as the initial model, and were built and refined using iterative cycles in Coot[31] and REFMAC,[32] employing local NCS restraints in the refinement cycles. Following building and refinement of the protein and water molecules in this complex, residual density was observed in the omit maps at the dimer interfaces for each structure, which could be clearly modelled as $NADP^+$. The final structures of IR007 and IR007-143 exhibited % R_{cryst}/R_{free} values of 24.7/27.3 and 18.8/21.1 respectively Refinement statistics for the structures are presented in **Table S3.** The structures of IR007 and IR007-143 have been deposited in the Protein Databank (PDB) with accession codes 8BJ5 and 8BK1 respectively.

Supporting Information

Additional experimental details, materials, and methods, as well as additional data supporting the findings reported are available as supporting information. All other data is available from authors upon reasonable request.

References

[1] Lim S., Kaldis P. (2013), Cdks, cyclins and CKIs: roles beyond cell cycle regulation. *Development*, 3079-3093.

[2] Malumbres M. (2014), Cyclin-dependent kinases. *Genome Biol.* , 122.

[3] Schwartz G. K., Shah, M. A. (2005), Targeting the cell cycle: a new approach to cancer therapy. *J.Clin. Oncol.*, **23**, 9408−9421.

[4] Malumbres M., Barbacid M. (2009), Cell cycle, CDKs and cancer: a changing paradigm. *Nat. Rev. Cancer*, **9**, 153−166.

[5] Freeman-Cook K.D., Hoffman R.L., Behenna D.C., Boras B., Carelli J., Diehl W., Ferre R.A., He Y., Hui A., Huang B., Husar N., Jones R., Kephart S.E., Lapek J., McTigue M., Miller N., Murray B.W., Nagata A., Nguyen L., Niessen S., Ninkovic S., O'Doherty I., Ornelas M.A., Solowiej J., Sutton S.C., Tran K., Tseng E., Visswanathan R., Xu M., Zehnder L., Zhang Q., Zhang C., Dann S. (2021), Discovery of PF-06873600, a CDK2/4/6 Inhibitor for the Treatment of Cancer. *Journal of Medicinal Chemistry*, **64** (13), 9056-9077.

[6] Duan S. Widlicka D.W., Burns M.P., Kumar R., Hotham I., Desrosiers J.N., Bowles P., Jones K.N., Nicholson L.D., Buetti-Weekly M.T., Han L., Steflik J., Hansen E., Hayward C.M., Strohmeyer H., Monfette S., Sutton, S.C., Morris C. (2022), Application of Biocatalytic Reductive Amination for the Synthesis of a Key Intermediate to a CDK 2/4/6 Inhibitor. *Org. Process Res. Dev.* **26** (3), 879–890.

[7] Afanasyev O.I., Kuchuk E., Usanov D. L., Chusov D. (2019), Reductive Amination in the Synthesis of Pharmaceuticals. *Chem. Rev*. **119**, 11857-11911.

[8] Nugent T.C., El-Shazly M. (2010), Chiral amine synthesis–recent developments and trends for enamide reduction, reductive amination, and imine reduction. *Adv. Synth. Catal*. **352**, 753-819.

[9] Aleku G., France S.P., Man H., Mangas-Sanchez J., Montgomery S.L., Sharma M., Leipold F., Hussain S., Grogan G., Turner N.J. (2017), A reductive aminase from Aspergillus oryzae. *Nature Chem* **9**, 961–969.

[10] France S.P., Aleku G., Sharma M., Mangas-Sanchez J., Howard R.M., Steflik J., Kumar R., Adams R.W., Slabu I., Crook R., Grogan G., Wallace T.W., Turner N.J. (2017), Biocatalytic Routes to Enantiomerically Enriched Dibenz[c,e]azepines. *Angew. Chem. Int. Ed.* **56**, 15589 –15593.

[11] Scheller P.N. Fademrecht S., Hofelzer S., Pleiss J., Leipold F., Turner N.J., Nestl B.M., Hauer B. (2014), Enzyme toolbox: novel enantiocomplementary imine reductases. *Chem. Bio. Chem*. **15**, 2201-2204.

[12] France S.P., Howard R.M., Steflik J., Weise N.W., Mangas-Sanchez J., Montgomery S.L., Crook R., Kumar R., Turner N.J. (2018), Identification of novel bacterial members of the imine reductase enzyme family that perform reductive amination. *Chem. Cat. Chem.* **10**, 510-514.

[13] Roiban G.D., Kern M., Liu Z., Hyslop J., Tey P.L., Levine M.S., Jordan L.S., Brown K.K., Hadi T., Ihnken L.A.F., Brown M.J.B. (2017), Efficient biocatalytic reductive aminations by extending the imine reductase toolbox. *Chem. Cat. Chem.* **9**, 4475-4479.

[14] Schober M. MacDermaid C., Ollis A., Chang S., Khan D., Hosford J., Latham J., Ihnken L., Brown M., Fuerst D., Sanganee M., Roiban D. (2019), Chiral synthesis of LSD1 inhibitor GSK2879552 enabled by directed evolution of an imine reductase. *Nat. Catal.* **2,** 909–915.

[15] Kumar R., Karmilowicz, M., Burke D., Burns M., Clark L., Connor C.,Cordi E., Do N., Doyle K., Hoagland S., Lewis C., Mangan D., Martinez C., McInturff E., Meldrum K., Pearson R., Steflik J., Rane A., Weaver J. (2021), Biocatalytic reductive amination from discovery to commercial manufacturing applied to abrocitinib JAK1 inhibitor. *Nat. Catal.* **4,** 775–782.

[16] Zhang K., He Y., Zhu J., Zhang Q., Tang L., Cui L., Feng Y. (2021) Engineering of Reductive Aminases for Asymmetric Synthesis of Enantiopure Rasagiline. *Front. Bioeng. Biotechnol.* **9**: 798147.

[17] Ma E.J., Siirola E., Moore C., Kummer A., Stoeckli M., Faller M., Bouquet C., Eggimann F., Ligibel M., Huynh D., Cutler G., Siegrist L., Lewis R.A., Acker A., Freund E., Koch E., Vogel M., Schlingensiepen H., Oakeley E.J., Snajdrova R.. (2021), Machine-Directed Evolution of an Imine Reductase for Activity and Stereoselectivity. *ACS Catalysis*. **11(20)**, 12433-12445.

[18] Rodriguez-Mata M., Frank A., Wells E., Leipold F., Turner N.J., Hart S., Turkenburg J.P., Grogan G. (2013) Structure and activity of NADPH-dependent reductase Q1EQE0 from Streptomyces kanamyceticus, which catalyses the R-selective reduction of an imine substrate. *Chembiochem.* **14**, 1372.

[19] Man H., Wells E., Hussain S., Leipold F., Hart S., Turkenburg J.P., Turner N.J., Grogan G. (2015) Structure, Activity and Stereoselectivity of NADPH-Dependent Oxidoreductases Catalysing the S-Selective Reduction of the Imine Substrate 2-Methylpyrrolidine. *Chembiochem.* **16**, 1052-1059.

[20] Holm L. (2020) DALI and the persistence of protein shape. *Protein Sci*. **29**, 128-140.

[21] Huber T. Schneider L., Präg A., Gerhardt S., Einsle O., Müller M. (2014) Direct Reductive Amination of Ketones: Structure and Activity of S-Selective Imine Reductases from Streptomyces. *Chemcatchem.* **6**, 2248-2252.

[22] Stockinger P., Schelle L., Schober B., Buchholz P.C.F., Pleiss J., Nestl B.M. (2014) Engineering of Thermostable β-Hydroxyacid Dehydrogenase for the Asymmetric Reduction of Imines. *Chembiochem.* **15**, 2201-2204.

[23] Trott O., Olson A. J. (2010) AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comp. Chem.* **31**, 455-461.

[24] Li H., Wu L., Zheng G.W., Zhou, J.H. (2020) Complex structure of an imine reductase at 2.05 Angstrom resolution. PDB DOI: 10.2210/pdb6JIT/pdb.

[25] Wetzl D., Berrera M., Sandon N., Fishlock D., Ebeling M., Müller M., Hanlon S., Wirz B., Iding H. (2015) Expanding the imine reductase toolbox by exploring the bacterial protein-sequence space. *Chembiochem*. **16**, 1749-1756.

[26] Li H., Luan Z., Zheng G., Xu J. (2015) Efficient synthesis of chiral indolines using an imine reductase from *Paenibacillus lactis*. *Adv. Synth. Catal.* **357**, 1692-1696.

[27] Kabsch, W. XDS. (2010) *Acta Crystallogr. Sect. D Biol. Crystallogr*. **66**, 125-132.

[28] Evans P.R. (2011) An Introduction to Data Reduction: Space-group Determination, Scaling and Density Statistics. *Acta Crystallogr. Sect. D Biol. Crystallogr*. **67**, 282-292.

[29] Winter G. (2010) An Expert System for Macromolecular Crystallography Data Reduction. *J. Appl. Crystallogr*. **43**, 186-190.

[30] Vagin A., Teplyakov, A. (1997) MOLREP: An Automated Program for Molecular Replacement. *J. Appl. Cryst*. **30**, 1022-1025.

[31] Emsley P., Cowtan K. (2004) Model-building Tools for Molecular Graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr*. **60**, 2126-2132.

[32] Murshudov G.N., Vagin, A.A., Dodson, E.J. (1997) Refinement of Macromolecular Structures by the Maximum Likelihood Method. *Acta Crystallogr. Sect. D Biol. Crystallogr*. **53**, 240-255.