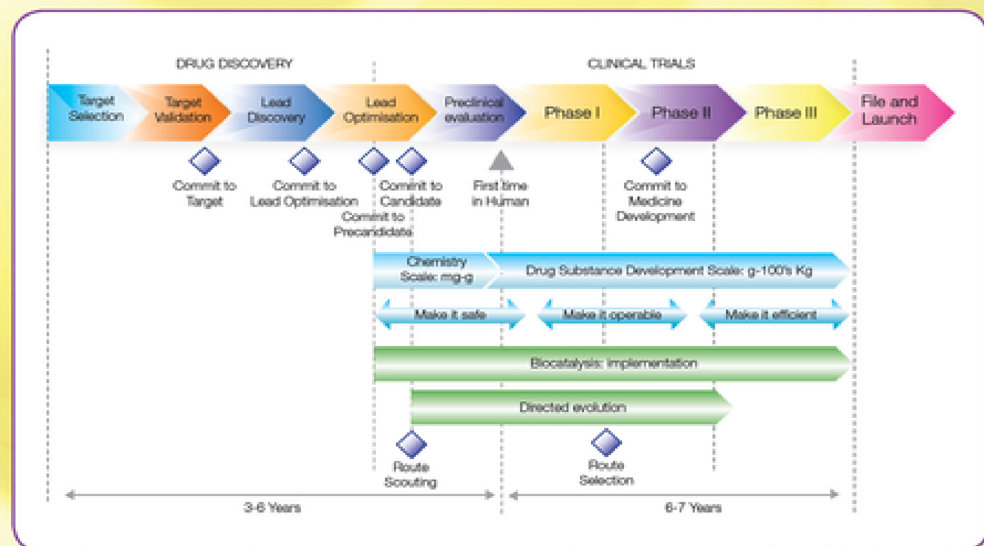
Edited by
John Whittall | Peter W. Sutton

Applied Biocatalysis

The Chemist's Enzyme Toolbox



WILEY

Applied Biocatalysis

Applied Biocatalysis

The Chemist's Enzyme Toolbox

Edited by

JOHN WHITTALL

University of Manchester
Manchester
UK

PETER W. SUTTON

Glycoscience S.L.
Barcelona
ES

WILEY

This edition first published 2021
© 2021 John Wiley & Sons Ltd

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at http://www.wiley.com/go/permissions.

The right of John Whittall and Peter W. Sutton to be identified as the authors of the editorial material in this work has been asserted in accordance with law.

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Limit of Liability/Disclaimer of Warranty
In view of ongoing research, equipment modifications, changes in governmental regulations, and the constant flow of information relating to the use of experimental reagents, equipment, and devices, the reader is urged to review and evaluate the information provided in the package insert or instructions for each chemical, piece of equipment, reagent, or device for, among other things, any changes in the instructions or indication of usage and for added warnings and precautions. While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organisation, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organisation, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication Data

Names: Whittall, John, editor. | Sutton, Peter (Peter W.), editor.
Title: Applied biocatalysis : the chemist's enzyme toolbox / edited by Dr
John Whittall, Dr Peter W. Sutton.
Description: First edition. | Hoboken, NJ : Wiley, 2021. | Includes
bibliographical references and index.
Identifiers: LCCN 2020015374 (print) | LCCN 2020015375 (ebook) | ISBN
9781119487012 (cloth) | ISBN 9781119487029 (adobe pdf) | ISBN
9781119487036 (epub)
Subjects: LCSH: Biocatalysis.
Classification: LCC TP248.65.E59 A677 2021 (print) | LCC TP248.65.E59
(ebook) | DDC 660.6/34–dc23
LC record available at https://lccn.loc.gov/2020015374
LC ebook record available at https://lccn.loc.gov/2020015375

Cover Design: Wiley
Cover Images: (Background) © Angelatriks/Shutterstock,
(Inset diagram) Courtesy of Gheorghe-Doru Roiban

Set in 10/12pt TimesLTStd by SPi Global, Chennai, India

Printed and bound by CPI Group (UK) Ltd, Croydon, CR0 4YY

10 9 8 7 6 5 4 3 2 1

Contents

Abbreviations *xi*

1 Directed Evolution of Enzymes Driving Innovation in API Manufacturing at GSK 1
1.1 Introduction 1
1.2 Drug Development Stages 3
1.3 Enzyme Panels 6
1.4 Enzyme Engineering 10
1.5 Case Studies 18
1.6 Outlook 22

2 Survey of Current Commercial Enzyme and Bioprocess Service Providers 27
2.1 Commercial Enzyme Suppliers/Distributors 28
2.2 Bioprocess Service Providers 92
2.3 Chemical Transformations of Selected Commercially Available Enzymes 103

3 Imine Reductases 135
3.1 Imine Reductase-Catalysed Enantioselective Reductive Amination for the Preparation of a Key Intermediate to Lysine-Specific Histone Demethylase 1 (LSD1) Inhibitor GSK2879552 135
3.2 Expanding the Collection of Imine Reductases Towards a Stereoselective Reductive Amination 138
3.3 Asymmetric Synthesis of the Key Intermediate of Dextromethorphan Catalysed by an Imine Reductase 143
3.4 Identification of Imine Reductases for Asymmetric Synthesis of 1-Aryl-tetrahydroisoquinolines 148
3.5 Preparation of Imine Reductases at 15 L Scale and Their Application in Asymmetric Piperazine Synthesis 156
3.6 Screening of Imine Reductases and Scale-Up of an Oxidative Deamination of an Amine for Ketone Synthesis 162

4 Transaminases 165
4.1 A Practical Dynamic Kinetic Transamination for the Asymmetric Synthesis of the CGRP Receptor Antagonist Ubrogepant 165
4.2 Asymmetric Biosynthesis of L-Phosphinothricin by Transaminase 168

4.3 Application of *In Situ* Product Crystallisation in the Amine Transaminase from *Silicibacter pomeroyi*-Catalysed Synthesis of (*S*)-1-(3-Methoxyphenyl)ethylamine 173
4.4 Enantioselective Synthesis of Industrially Relevant Amines Using an Immobilised ω-Transaminase 178
4.5 Amination of Sugars Using Transaminases 182
4.6 Converting Aldoses into Valuable ω-Amino Alcohols Using Amine Transaminases 187

5 Other Carbon–Nitrogen Bond-Forming Biotransformations 193
5.1 Biocatalytic *N*-Acylation of Anilines in Aqueous Media 193
5.2 Enantioselective Enzymatic Hydroaminations for the Production of Functionalised Aspartic Acids 196
5.3 Biocatalytic Asymmetric Aza-Michael Addition Reactions and Synthesis of L-Argininosuccinate by Argininosuccinate Lyase ARG4-Catalysed Aza-Michael Addition of L-Arginine to Fumarate 204
5.4 Convenient Approach to the Biosynthesis of C2,C6-Disubstituted Purine Nucleosides Using *E. coli* Purine Nucleoside Phosphorylase and Arsenolysis 211
5.5 Production of L- and D-Phenylalanine Analogues Using Tailored Phenylalanine Ammonia-Lyases 215
5.6 Asymmetric Reductive Amination of Ketones Catalysed by Amine Dehydrogenases 221
5.7 Utilisation of Adenylating Enzymes for the Formation of *N*-Acyl Amides 231

6 Carbon–Carbon Bond Formation or Cleavage 237
6.1 An Improved Enzymatic Method for the Synthesis of (*R*)-Phenylacetyl Carbinol 237
6.2 Tertiary Alcohol Formation Catalysed by a Rhamnulose-1-Phosphate Aldolase : Dendroketose-1-Phosphate Synthesis 241
6.3 Easy and Robust Synthesis of Substituted L-Tryptophans with Tryptophan Synthase from *Salmonella enterica* 247
6.4 Biocatalytic Friedel–Crafts-Type *C*-Acylation 250
6.5 MenD-Catalysed Synthesis of 6-Cyano-4-Oxohexanoic Acid 256
6.6 Production of (*R*)-2-(3,5-Dimethoxyphenyl)propanoic Acid Using an Aryl Malonate Decarboxylase from *Bordetella bronchiseptica* 259

7 Reductive Methods 263
7.1 Synthesis of Vibegron Enabled by a Ketoreductase Rationally Designed for High-pH Dynamic Kinetic Reduction 263
7.2 Synthesis of a GPR40 Partial Agonist Through a Kinetically Controlled Dynamic Enzymatic Ketone Reduction 265
7.3 Lab-Scale Synthesis of Eslicarbazepine 267
7.4 Direct Access to Aldehydes Using Commercially Available Carboxylic Acid Reductases 270

7.5 Preparation of Methyl (*S*)-3-Oxocyclohexanecarboxylate Using an Enoate Reductase 277

8 Oxidative Methods 281
8.1 Macrocyclic Baeyer–Villiger Monooxygenase Oxidation of Cyclopentadecanone on 1 L Scale 281
8.2 Regioselective Lactol Oxidation with O_2 as Oxidant on 1 L Scale Using Alcohol Dehydrogenase and NAD(P)H Oxidase 286
8.3 Synthesis of (3*R*)-4-[2-Chloro-6-[[(*R*)-Methylsulfinyl]methyl]-Pyrimidin-4-yl]-3-Methyl-Morpholine Using BVMO-P1-D08 291
8.4 Oxidation of Vanillyl Alcohol to Vanillin with Molecular Oxygen Catalysed by Eugenol Oxidase on 1 L Scale 295
8.5 Synthesis of Syringaresinol from 2,6-Dimethoxy-4-Allylphenol Using an Oxidase/Peroxidase Enzyme System 301
8.6 Biocatalytic Preparation of Vanillin Catalysed by Eugenol Oxidase 308
8.7 Vanillyl Alcohol Oxidase-Catalysed Production of (*R*)-1-(4′-Hydroxyphenyl)ethanol 312
8.8 Enzymatic Synthesis of Pinene-Derived Lactones 319
8.9 Enzymatic Preparation of Halogenated Hydroxyquinolines 326

9 Hydrolytic and Dehydratase Enzymes 333
9.1 Synthesis of (*S*)-3-(4-Chlorophenyl)-4-Cyanobutanoic Acid by a Mutant Nitrilase 333
9.2 Nitrilase-Mediated Synthesis of a Hydroxyphenylacetic Acid Substrate via a Cyanohydrin Intermediate 337
9.3 Production of (*R*)-2-Butyl-2-Ethyloxirane Using an Epoxide Hydrolase from *Agromyces mediolanus* 339
9.4 Preparation of (*S*)-1,2-Dodecanediol by Lipase-Catalysed Methanolysis of Racemic Bisbutyrate Followed by Selective Crystallisation 344
9.5 Biocatalytic Synthesis of *n*-Octanenitrile Using an Aldoxime Dehydratase from *Bacillus* sp. OxB-1 349
9.6 Access to (*S*)-4-Bromobutan-2-ol through Selective Dehalogenation of *rac*-1,3-Dibromobutane by Haloalkane Dehalogenase 354

10 Glycosylation, Sulfation and Phosphorylation 363
10.1 Rutinosidase Synthesis of Glycosyl Esters of Aromatic Acids 363
10.2 Biocatalytic Synthesis of Kojibiose Using a Mutant Transglycosylase 369
10.3 Biocatalytic Synthesis of Nigerose Using a Mutant Transglycosylase 377
10.4 Easy Sulfation of Phenols by a Bacterial Arylsulfotransferase 381
10.5 Shikimate Kinase-Catalysed Phosphorylations and Synthesis of Shikimic Acid 3-Phosphate by AroL-Catalysed Phosphorylation of Shikimic Acid 386
10.6 Kinase-Catalysed Phosphorylations of Ketohexose Phosphates and LacC-Catalysed Synthesis of D-Tagatose-1,6-Diphosphate Lithium Salt 393

10.7 Kinase-Catalysed Phosphorylations of Xylulose Substrates and Synthesis of Xylulose-5-Phosphate Enantiomers 397
10.8 Phosphoramidates by Kinase-Catalysed Phosphorylation and Arginine Kinase-Catalysed Synthesis of N_{ω}-Phospho-L-Arginine 401

11 Enzymatic Cascades 409
11.1 Redox-Neutral Ketoreductase and Imine Reductase Enzymatic Cascade for the Preparation of a Key Intermediate of the Lysine-Specific Histone Demethylase 1 (LSD1) Inhibitor GSK2879552 409
11.2 Asymmetric Synthesis of α-Amino Acids through Formal Enantioselective Biocatalytic Amination of Carboxylic Acids 413
11.3 Enantioselective, Catalytic One-Pot Synthesis of γ-Butyrolactone-Based Fragrances 420
11.4 Synthesis of Six out of Eight Carvo-Lactone Stereoisomers via a Novel Concurrent Redox Cascade Starting from (*R*)-and (*S*)-Carvones 426
11.5 One-Pot Biocatalytic Synthesis of D-Tryptophan Derivatives from Substituted Indoles and L-Serine 435
11.6 *Escherichia coli* Lysate Multienzyme Biocatalyst for the Synthesis of Uridine-5′-Triphosphate from Orotic Acid and Ribose 441
11.7 Aerobic Synthesis of Aromatic Nitriles from Alcohols and Ammonia Using Galactose Oxidase 449
11.8 Hydrogen-Borrowing Conversion of Alcohols into Optically Active Primary Amines by Combination of Alcohol Dehydrogenases and Amine Dehydrogenases 455
11.9 Ene-Reductase-Mediated Reduction of C=C Double Bonds in the Presence of Conjugated C≡C Triple Bonds: Synthesis of (*S*)-2-Methyl-5-Phenylpent-4-yn-1-ol 468

12 Chemo-Enzymatic Cascades 475
12.1 Synergistic Nitroreductase/Vanadium Catalysis for Chemoselective Nitroreductions 475
12.2 Chemo-Enzymatic Synthesis of (*S*)-1,2,3,4-Tetrahydroisoquinoline Carboxylic Acids Using D-Amino Acid Oxidase 482
12.3 Amine Oxidase-Catalysed Deracemisation of (*R,S*)-4-Cl-Benzhydrylamine into the (*R*)-Enantiomer in the Presence of a Chemical Reductant 488
12.4 Asymmetric Synthesis of 1-Phenylpropan-2-amine from Allylbenzene through a Sequential Strategy Involving a Wacker–Tsuji Oxidation and a Stereoselective Biotransamination 497
12.5 Chemoenzymatic Synthesis of (2*S*,3*S*)-2-Methylpyrrolidin-3-ol 504

13 Whole-Cell Procedures 509
13.1 Semipreparative Biocatalytic Synthesis of (*S*)-1-Amino-1-(3′-Pyridyl)methylphosphonic Acid 509

13.2 Practical and User-Friendly Procedure for the Regio- and Stereoselective Hydration of Oleic, Linoleic and Linolenic Acids, Using Probiotic *Lactobacillus* Strains as Whole-Cell Biocatalysts 515
13.3 Clean Enzymatic Oxidation of 12α-Hydroxysteroids to 12-Oxo-Derivatives Catalysed by Hydroxysteroid Dehydrogenase 521
13.4 Whole-Cell Biocatalysis Using PmlABCDEF Monooxygenase and Its Mutants: A Versatile Toolkit for Selective Synthesis of Aromatic *N*-Oxides 528

Index ***535***

Abbreviations

a/a	Area/area
AA	Amino acid
AADH	Amino acid dehydrogenase
ABTS	2,2′-Azino-bis(3-ethylbenzthiazoline-6-sulfonic acid
ABS	Amide bond synthetase
ACN	Acetonitrile (Also MECN)
AcOEt	Ethyl acetate (Also EtOAc)
ADH	Alcohol dehydrogenase (Also KRED)
ADP	Adenosine diphosphate
AEAA	(*S*)-*N*-(2-Aminoethyl)aspartic acid
AHAS	Acetohydroxyacid synthase
AHTC	Anhydrotetracycline
AMDase	Aryl malonate decarboxylase
AmDH	Amine dehydrogenase
AP	Area percent
API	Active pharmaceutical ingredient
ArgK	Arginine kinase
ArgK-LP	ArgK from *Limulus Polyphemus*
AroL	*E. coli* K12 shikimate kinase
AST	Arylsulfotransferase
ATA	Amine transaminase (Also TA)
ATase	Acetyl transferase
atm	Atmosphere (pressure)
ATP	Adenosine triphosphate
BCA	Bicinchoninic acid assay
BCL	Agarose beads crosslinked
BA	Benzaldehyde
BaSP	Sucrose phosphorylase from *Bifidobacterium adolescentis*
Bis-tris	2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol
BL21	Competent *Escherichia coli* BL21
BME	β-Mercaptoethanol
BMGH	Buffered minimal glycerol medium
BMGY	Buffered glycerol-complex medium
BMMH	Buffered minimal methanol medium
BMMY	Buffered methanol-complex medium
$BnONH_2$	*O*-Benzylhydroxylamine

Boc	t-Butoxycarbonyl
BSA	Bovine serum albumin
BSM	Basal salt medium
BVMO	Baeyer–Villiger monooxygenase
BVO	Baeyer–Villiger oxidation
c	Conversion
CA	Cholic acid
CAD	Charged aerosol detection
CALB	*Candida antarctica* lipase B
CAR	Carboxylic acid reductase
Cbz-OSu	N (Benzyloxycarbonyloxy)succinimide
CDCA	Chenodeoxycholic acid
CDW	Cell dry weight
CDMO	Cyclododecanone monooxygenase
CFE	Cell free extract
c.f.u	Colony-forming unit
CHMO	Baeyer–Villiger monooxygenase (cyclohexanone monoxygenase)
CLR	Controlled laboratory reactor
CoG	Cost of goods
CP	Citrate phosphate (buffer)
CPD	Cyclopentadecanone
CPME	Cyclopentyl methyl ether
CSU	Catalytic subunit
CSU-GST	Catalytic subunit – glutathione S-transferase tag
CV	Column volume
DAAO	D-Amino acid oxidase
DAAT	D-Amino acid transferase
DAD	Diode-array detection
DAPG	2,4-Diacetylphloroglucinol
DCC	Dicyclohexylcarbodiimide
DCM	Dichloromethane
ddH_2O	Double distilled water
DEA	Diethyl amine
DEAE	Diethylaminoethanol (group in ion-exchange resin)
DERA	2-Deoxyribose-5-phosphate aldolase
dH_2O	Distilled water
DHA	Dihydroxyacetone
DHAK	Dihydroxyacetone kinase
DHAP	Dihydroxyacetone phosphate
DHIQ	Dihydroisoquinoline
dI	Deoxyinosine
DI	Deionised
DIBAL-H	Diisobutyl aluminium hydride
DK(R)	Dynamic kinetic (resolution)
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide

DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DoE	Design of experiments
DSP	Downstream processing
DPPA	3,3-Diphenylpropionic acid
DTT	Dithiothreitol
E	Enantioselectivity
EDDS	Ethylenediamine-N,N′-disuccinic acid
EDTA	Ethylenediaminetetraacetic acid
ee	Enantiomeric excess
EH	Epoxide hydrolase
ELSD	Evaporative light scattering detector
ER	Ene-reductase (often same as ERED)
ERED	Enoate reductase (often same as ER)
Et_2O	Diethyl ether
EtOAc	Ethyl acetate (Also AcOEt)
EUGO	Eugenol oxidase
EV	Expansion vessel
EWG	Electron-withdrawing group
FA	Formic acid
FAD	Flavin adenine dinucleotide
$FADH_2$	Flavin adenine dinucleotide reduced form
FDH	Formate dehydrogenase
F&F	Flavour and Fragrance
FID	Flame ionisation detector
FMN	Flavin mononucleotide
FMO	Flavin monooxygenase enzyme
Fre	Flavin reductase
FSA	D-Fructose-6-phosphate aldolase
*Fs*DAAO	D-Amino acid oxidase from *Fusarium solani*
FTIR	Fourier-transform infrared spectroscopy
g	Gram (× g is centrifuge unit)
G6P	Glucose-6-phosphate
G6PDH	Glucose-6-phosphate dehydrogenase
GC	Gas chromatography
GC-FID	Gas chromatography/flame ionisation detection
GC-MS	Gas chromatography/mass spectrometry
GDH	Glucose dehydrogenase
GHMP	Galactokinase, homoserine kinase, mevalonate kinase
GMP	Good manufacturing practice
Gox	Galactose oxidase
GPDH	α-Glycerophosphate dehydrogenase
GPR40	G-protein-coupled receptor 40
GRAS	Generally recognised as safe
GST	Glutathione S-transferase
GTP	Guanosine-5′-triphosphate

hr	hour
HA	Hydroxyacid
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
HEWT	Amine transaminase from *Halomonas elongata* DSM 2581
HFA	Hydroxy fatty acid
HIC-DH	Hydroxyisocaproate dehydrogenase
HMM	Hidden Markov model
HMU	Worldwide PDB entry aminotransferase from *Silicibacter pomeroyi*
HPAEC	High-performance anion exchange chromatography
HPLC	High-performance liquid chromatography
HPLC-DAD	High-performance liquid chromatography with diode-array detection
HPLC-PAD	High-performance liquid chromatography with photodiode-array detection
HPLC-RI	High performance liquid chromatography with refractive index
HRMS	High-resolution mass spectrometry
HRP	Horse radish peroxidase
HSDH	Hydroxysteroid dehydrogenases
HTP	High throughput
HWE	Horner–Wadsworth–Emmons reaction
iBAT	Ileal bile acid transport
IBX	o-Iodoxybenzoic acid
ID	Internal diameter
IMAC	Immobilised metal affinity chromatography
IMB	Immobilised
IP	Intellectual property
IPA	Isopropyl alcohol
IPA	Isopropyl amine
IPAc	Isopropyl acetate
IPEA	Isopropenyl acetate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Imine reductase
IRED	Imine reductase
IS	Internal standard
ISM	Iterative saturation mutagenesis
ISPC	*In situ* product crystallisation
ISPR	*In situ* product removal
kDA	Kilodalton
KHK	Ketohexokinase
KIRED	Ketimine reductase
KPB	Potassium phosphate buffer
KPi	Potassium phosphate buffer
KR	Kinetic resolution (Also ADH)
KRED	Ketoreductase
LAAD	L-Amino acid deaminase
LacC	D-Tagatose 6-phosphate kinase
LB	Lysogenic broth (also known as Luria–Bertani medium)
LC-MS	Liquid chromatography/mass spectrometry

LDH	Lactate dehydrogenase
LE-AmDH	Lysine amine dehydrogenase
LSD	Lysine specific histone demethylase
MALDI	Matrix-assisted laser desorption/ionisation
MAO-N	Monoamine oxidase from *Aspergillus niger*
MAP	Methoxyacetophenone
MeCN	Acetonitrile (Also ACN)
MenD	2-Succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase
MeOH	Methanol
MES	2-(*N*-Morpholino)ethanesulfonic acid
MeTHF	2-Methyl tetrahydrofuran
MDH	Malate dehydrogenase
MFC	Mass flow controller
min	Minute
MOPS	4-Morpholinepropanesulfonic acid
MPLC	Medium-pressure liquid chromatography
m.pt	Melting point
MRS	de Man, Rogosa and Sharpe broth
MS	Mass spectroscopy
$mS.cm^{-1}$	Millisiemens per centimeter
MTBE	Methyl *tert*-butyl ether
MWCO	Molecular weight cut-off
m/z	Mass-to-charge ratio
NA	Nutrient agar
NAC	*N*-Acetyl-L-cysteine
NAD^+	β-Nicotinamide adenine dinucleotide
NADH	β-Nicotinamide adenine dinucleotide, reduced form
NADPH	β-Nicotinamide adenine dinucleotide 2′-phosphate, reduced form
$NADP^+$	β-Nicotinamide adenine dinucleotide 2′-phosphate
$NH_3{\cdot}BH_3$	Ammonia-borane complex
Ni-NTA	Nickel-nitrilotriacetic acid
nm	Nanometre
NMAADH	*N*-Methylamino acid dehydrogenase
NMP	Nucleoside monophosphate
NMR	Nuclear magnetic resonance spectroscopy
NOX	NAD(P)H oxidase
NP	Nitrophenol
NP	Normal phase (chromatography)
NPS	Nitrophenyl sulfate
NR	Nitroreductase
nr	Nonredundant
NTP	Nucleoside triphosphate
OD	Optical density
Omd	Orotidine-5′-monophosphate decarboxylase
OMP	Orotidine-5′-monophosphate

OPA	o-Phthalaldehyde
Opt	Orotate phosphoribosyl transferase
OYE	Old yellow enzyme
OUR	Oxygen uptake rate
P	*ortho*-Phosphate
P450	Cytochrome P450
P5CR	Δ^1-Pyrroline-5-carboxylate reductase
PA	Phenyl acetate
PAD	Pulsed amperometric detection
PAGE	Polyacrylamide gel electrophoresis
PAL	Phenylalanine ammonia lyase
PAPS	Adenosine-3′-phospho-5′-phosphosulfate
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PD	Potato-dextrose
PDa	Potato dextrose agar
PDA	Photodiode array
PDb	Potato dextrose broth
PDB	Protein data bank
PDC	Pyruvate decarboxylase
PDL	Pentadecanolide
PE	Petroleum ether
PEP	Phosphoenolpyruvic acid
PFAM	Protein family
Pf-TA	Transaminase from *Pseudomonas fluorescens*
Phe-DH	Phenylalanine dehydrogenase
PK	Pyruvate kinase
Pk	Porcine kidney
PLIF	Protein–ligand interaction fingerprint
PLP	Pyridoxal 5′-phosphate
PMP	Pyridoxamine 5-phosphate
PMSF	Phenylmethylsulfonyl fluoride
PNPase	Purine nucleoside phosphorylase
polyP	Polyphosphate
PP	Pyrophosphate
PpATaseCH	Acylase from *Pseudomonas protegens*
PPB	Potassium phosphate buffer
PPK	Polyphosphate kinase
PPO	4-(Hydroxy(methyl)phosphoryl)-2-oxobutanoic acid
Pps	PRPP synthetase
PPT	Phosphinothricin
PRPP	Phosphoribosyl pyrophosphate
PTM	*Pichia* trace metal salt supplement
PTV	Programmed temperature vaporisation
PWM	Position weight matrix
PYR	Pyruvate

QbD	Quality by design
rac	Racemic
R&D	Research and Development
RCF	Relative centrifugal force
Rdc2	Halogenase from *Pochonia chlamydosporia*
Rf	Retention factor
RhuA	Rhamnulose-1-phosphate aldolase
ROH	Generic alcohol
ROP	Ring-opening polymerisation
RP	Reverse phase
R-PAC	(*R*)-Phenylacetyl carbinol
rpm	Revolutions per minute
RSU	Regulatory subunit
rt	Room temperature
Rt	Retention time (also t_R)
SASA	Solvent accessible surface area
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS–polyacrylamide gels
sec	Seconds
SeH	Soluble epoxide hydrolase
SFC	Supercritical fluid chromatography
SLM	Supported liquid membrane
sLpm	Standard litre per minute
SOC	Super optimal broth with catabolite repression
STRAP	Structure based sequences alignment program
TA	Transaminase
TA-CV	ξ-Transaminase from *Chromobacterium violaceum*
TATP	Tri-acetone-triperoxide
TB	Terrific broth
TBAP	Tetrabutylammonium phosphate
TBDAc	2,3-Di-*O*-acetyl-6-*O*-t-butyldimethylsilyl)-β-cyclodextrin
TEA	Triethylamine
TEoA	Triethanolamine
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
ThDP	Thiamine diphosphate
THIQ	Tetrahydroisoquinoline
TLC	Thin layer chromatography
TMS	Tetramethyl silane
$TMSCHN_2$	Trimethylsilyl diazomethane
TOF	Time-of-flight
TPI	Triosephosphate Isomerase
t_R	Retention time (also Rt)
Tris	Tris(hydroxymethyl)aminomethane
TrpS	Tryptophan synthase
TY	Tryptone yeast extract broth

U	Units
UAB	Universitat Autònoma de Barcelona
UDCA	Ursodeoxycholic acid
UFA	Unsaturated fatty acid
UHPLC	Ultra-high-performance liquid chromatography
UPLC	Ultra-performance liquid chromatography
UTP	Uridine-5′-triphosphate
UV	Ultraviolet
UV-Vis	Ultraviolet/visible
VAO	Vanillyl alcohol oxidase
v/v	Volume/volume
vvm	Gas volume flow per unit of liquid volume per minute (vessel volume per minute)
wcp	Wet cell pellet
wcw	Wet cell weight
wrt	With respect to
wt	Wild type
w/v	Weight/volume
w/w	Weight/weight
x g	Centrifugal force (relative centrifugal force more precise than rpm)
YNB	Yeast nitrogen base
YPD	Yeast extract peptone dextrose medium
YT	Yeast extract tryptone
ZmPDC	*Zymomonas mobilis* pyruvate decarboxylase

1

Directed Evolution of Enzymes Driving Innovation in API Manufacturing at GSK

*Jonathan Latham,[1] Anne A. Ollis,[2] Chris MacDermaid,[3] Katherine Honicker,[2] Douglas Fuerst[2] and Gheorghe-Doru Roiban,[*1]*

[1]Synthetic Biochemistry, Medicinal Science & Technology, GlaxoSmithKline Medicines Research Centre, Stevenage, Hertfordshire, UK

[2]Synthetic Biochemistry, Medicinal Science & Technology, GlaxoSmithKline, Collegeville, PA, USA

[3]Molecular Design, Computational and Modelling Sciences, GlaxoSmithKline, Collegeville, PA, USA

1.1 Introduction

Biocatalysis has had a significant impact on the synthesis of active pharmaceutical ingredients (APIs) in recent years. The main driver for this is the ability to harness the regio- and stereoselectivity of enzymes to improve the efficiency of synthetic routes. For example, enzymes can offer direct access to enantiopure products, where traditional organic synthesis would require either resolution or the use of auxiliary groups [1], whilst enzymes applied in manufacture have improved syntheses or generated molecules that would otherwise be either impossible or impractical to synthesise. Other factors supporting the adoption of enzymes in the synthesis of APIs include:

- **Reduced manufacturing costs:** The selectivity of biocatalytic processes often results in fewer overall processing and purification steps, reducing labour and material requirements.

Applied Biocatalysis: The Chemist's Enzyme Toolbox, First Edition. Edited by John Whittall and Peter W. Sutton.
© 2021 John Wiley & Sons Ltd. Published 2021 by John Wiley & Sons Ltd.

- **Environmental sustainability:** Traditional organic solvents and reagents raise environmental concerns. Enzymes often function in aqueous solutions using natural biochemical co-factors as reagents, providing environmentally friendly and operationally safe alternatives to traditional chemical transformations.
- **Sustainable supply:** Enzymes fulfil sustainability principals [2]. They are produced from microbial fermentation – a renewable feedstock requiring simple sugars to grow – and are themselves biodegradable.
- **Simplified drug manufacturing:** The increasing complexity of APIs and the need to implement new products has ramped up the pressure to find new synthetic strategies to simplify the way drugs are made. Additionally, GlaxoSmithKline's (GSK) focus on making medicines broadly accessible across the globe requires economical manufacture of APIs and contributes to the drive for simplification of manufacturing processes.
- **Quality:** Regulatory pressures from government agencies to maintain high API quality standards whilst reducing carbon emissions drive the use of biocatalysis to harness the high regio-, chemo- and stereoselectivity offered.
- **Accessibility:** Technology advances such as next-generation sequencing and directed evolution have simplified biocatalysis adoption.

Early attempts to embrace biocatalysis relied on identification of native microorganisms and enzymes capable of catalysing the desired transformation with exquisite selectivity under the required process conditions. This represented a significant barrier, as an enzyme's natural sensitivity and substrate specificity often were not compatible with manufacturing process conditions, where solvent concentration, temperature and pH, for example, are often out of a typical physiological range. Early protein engineering attempts to overcome these challenges largely involved generating structure-guided, rational mutations to an enzyme's primary sequence through site-directed mutagenesis.

Advances in the directed evolution of proteins, however, have facilitated, greatly accelerated and enabled the wider-scale implementation of biocatalysis by providing an accessible means of producing fit-for-purpose enzymes and increasing the overall speed of enzyme engineering [3]. Directed evolution can be used to tailor multiple enzyme properties that historically challenged the uptake of biocatalysis, rapidly alleviating problems with properties such as activity, specificity, expression, thermostability and tolerance to process conditions. This ability to engineer biocatalytic enzymes has been a boon to all fields of chemical manufacture, including pharmaceuticals.

Despite all the technological advances and positive trends, adoption of industrial-scale biocatalytic processes has been generally slower than expected. This is mainly due to the fact that small-molecule drug manufacturing processes are far from simple and efficient; development of a small-molecule screening hit to a commercial product takes around 10–15 years [4]. Recently, in GSK, there has been a focus on adopting biocatalysis by first intent, rather than as a second- or third-generation process. This aids in the reduction of costs associated with filing a new process post-approval, as well as reducing the resource requirement over the lifecycle of the product [5]. As such, within GSK, we employ a unified technology platform to deliver our portfolio [6].

In addition to flow chemistry (continuous primary) and chemical catalysis, biocatalysis was identified as step-changer technology that would boost GSK's ability to manufacture APIs. To accelerate internal biocatalysis and enzyme engineering capability, in 2014,

GSK in-licensed the CodeEvolver® platform from Codexis, a California, USA-based biotechnology company. Internal investment in a team of scientists with expertise in biocatalysis, directed evolution, molecular biology, sequencing and computational chemistry provided dedicated support and operation of this platform. In addition, specialised theoretical and practical biocatalysis courses and workshops were organised for medicinal, organic and process chemists, to increase awareness of biocatalysis and directed evolution for synthetic chemistry applications. This training enabled scientists to identify potential enzymatic opportunities in chemistry routes and creen available biotransformations.

This chapter highlights how the directed evolution of biocatalysts has delivered impact for GSK, and contextualises this in the API manufacturing and drug development environment, enabling understanding of the timelines required to deliver a robust and manufacturable biocatalytic process. The chapter also provides examples of success stories in implementing directed evolution at GSK, and discusses the hurdles currently associated with embedding biocatalysis and how the process may be further accelerated.

1.2 Drug Development Stages

To understand how biocatalysis fits into the pipeline, knowledge of the drug development process is required. The journey from small-molecule screening to commercialisation of a medicine starts with *target selection* and *validation* (Figure 1.1). During this process, scientists gather evidence to support the role of a target (e.g. an enzyme in a biochemical pathway or a receptor) in a given disease, and the potential therapeutic benefit of modulating its function. The second stage is *lead discovery*, where the scientists seek molecules capable of interacting with the target (hits), which can then be used as a starting template for further optimisation. These hits are derivatised into synthetic small substrates called lead molecules, which interact with the target and have additional qualities that give the team confidence that they can be optimised to deliver a medicine (e.g. favourable target binding strength and selectivity). This process, from target selection to identification of lead molecules, can take between 4 and 24 months.

During the lead discovery and optimisation stages, the attrition rate is very high, with most compounds being discarded because of poor biochemical or biophysical properties. The focus, therefore, is on the quick delivery of a large number of diverse compounds using whichever chemistry works – making the adoption of biocatalysis more difficult at this stage. As only small quantities of compound are required (mgs), techniques like chromatography and chiral resolution are considered acceptable, meaning that the selectivity and process advantages offered by biocatalysis are less likely to be harnessed. In the identification of compounds with the desired pharmacological properties, however, biocatalysis can still be a powerful tool at this stage, as it enables the synthesis of drug entities inaccessible by other chemistries.

Once several pre-candidate molecules have been selected, medicinal chemists can begin to focus their effort on identifying biocatalysts that may provide a more efficient synthetic step. Quick read-outs and fast delivery are required, meaning that enzyme hits (those identified as capable of performing the desired chemistry) need to be easily scaled to deliver grams of material and allow delivery of product for further toxicological studies. Reactions utilising non-process-ready biocatalysts (having low activity, poor stability or suboptimal

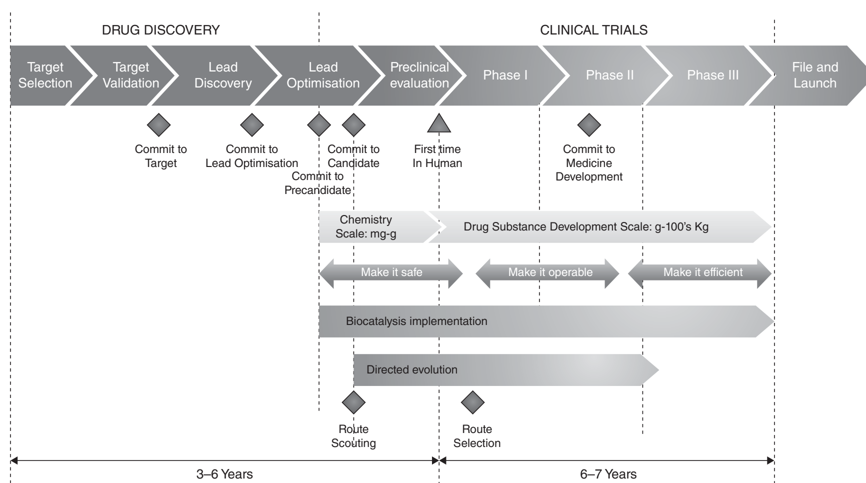


Figure 1.1 *A general drug discovery path overlapped with biocatalytic opportunities.*

selectivity) do not pose a significant problem at this stage, provided that the biotransformation affords product that meets the minimal quality criteria (e.g. chiral purity), since the quantity of product that must be prepared is small. Due to the relatively small scale, several of the downsides of non-process-ready enzymes can be easily mitigated (e.g. purification to meet quality requirements or centrifugation to deal with high enzyme loadings), affording some flexibility around the biocatalyst properties. However, if the catalyst presents challenges at this small scale (e.g. very low activity that cannot be reproduced on gram scale, or poor selectivity) that would require engineering to overcome, it is unlikely that the enzymatic step will be pursued as the high attrition of compounds means the resource commitment to enzyme evolution is difficult to justify. One solution to this problem is to increase the quality of the enzyme panels (collections of enzymes from various transformation classes that are initially screened for desired biotransformation), and therefore the likelihood of success on scale-up, by expanding the number and diversity of enzymes within the panels – either by acquisition of new enzymes or through panel expansion with engineered enzyme variants.

Once the few promising candidates have been selected, the drug journey continues with preclinical evaluation, where compounds are assessed for toxicity and efficacy using a combination of *in vitro* and *in vivo* animal models. With this data in hand, a decision is made as to whether or not a compound will progress to phase 1 clinical trials – also called 'commit to first time in human'. During phase 1, which typically takes between 12 and 18 months, batches of API are prepared for later dosing. The challenges that occur during API preparation can have a knock-on effect on clinical trials, causing them to slow or halt if drug supply is inadequate. Keeping these trials on time is key to timely assessment of drug candidates and ultimately to the delivery of approved medicines to patients. At this point, an ideal process for delivery of API is not required, provided that product quality is maintained. For assets in phase 1, the route employed by the medicinal chemistry team is usually scaled up to provide API for toxicology studies. Although these processes are usually not suitable or scalable for commercial-scale manufacturing, the time-critical nature of compound delivery and the small quantities required mean significant time isn't usually invested in process development at this stage.

Whilst API is being supplied using a non-ideal manufacturing process, route scouting activities are undertaken concurrently to identify more appropriate long-term syntheses. This is often where biocatalysis opportunities are identified and screened. At this point, the hits from enzyme panel screening must provide a significant advantage over the previous chemistry and provide API compound on a reasonable scale. Enzyme engineering becomes feasible at this stage and can play an important part in assessing a route's feasibility, impacting route scouting and selection.

As development progresses and the asset heads towards phase 2, delivery of API for clinical studies must occur concurrent with other process development activities. These include:

- Discovery of new routes of synthesis and chemistries to facilitate this (route scouting)
- Optimisation and understanding of the chemical process
- Thorough understanding of parameters impacting drug substance and intermediates quality

- Consideration of supply chain security for the process – that is, availability of starting materials and reagents on the required scale
- Transfer of the process to a full-scale manufacturing plant
- Preparation of regulatory documentation to support file and launch

If found to be effective during phase 1, an asset reaches phase 2 trials – a process that can take 2–3 years before commitment to medicine development. In addition to drug production for future trials, it is also important that scale-up process strategy is considered during this time. By now, the number of candidates has been reduced and route improvements driven by introducing biocatalysis have been linked directly with API production. The biocatalyst has been optimised to deliver a manufacturable process, but there is still room for further improvement and fine tuning if certain criteria (such as cost of goods) have not yet been met.

In phase 3, extensive work is undertaken to identify the logistics surrounding the distribution of clinical supply to the investigator sites and to develop a robust commercial, end-to-end supply chain to ensure continuity of both launch and long-term drug supply to patients. By this time, in addition to optimisation of enzymes through engineering and biocatalytic process development, a fermentation process for enzyme supply and the manufacturing chain must be established.

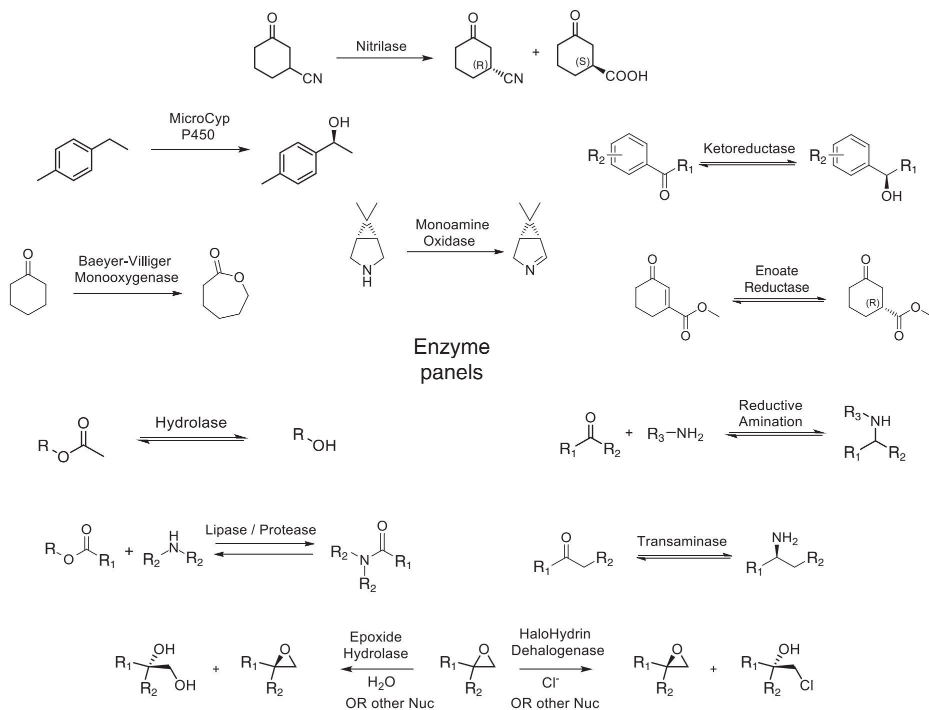
Throughout the drug development phases, the aim is to generate a process that is safe, operable and ultimately efficient. In this context, the decision to engineer an enzyme for use in an API manufacturing route is complex, due to the time and resource commitment required for protein engineering. It is often the case that either wild-type or panel enzymes will satisfy the requirements of early, small-scale manufacturing campaigns of nascent assets. However, once an asset is significantly advanced along the drug development pipeline, it becomes increasingly difficult for enzymes to provide a manufacturable process for larger-scale clinical supply campaigns. This is where directed evolution has a large impact. As strategies and technologies develop to expedite the evolution process, it becomes increasingly possible to adopt directed evolution earlier in the drug development cycle. In the next section, we discuss how recent developments have helped with this aim, and what future work is required to fully realise this vision.

1.3 Enzyme Panels

GSK has a significant number of panels, produced both internally and acquired from external sources, which are continuously enhanced through evolution and addition of new enzyme classes (Scheme 1.1).

A portfolio analysis reveals which transformations account for the most frequently used chemistries within GSK [6]. Heteroatom alkylation and arylations, together with aromatic heterocycle formations, make up approximately 40% of the portfolio. Functional group interconversion, C-C bond formation and reductions, oxidations and protections are other types of transformations frequently encountered. Focusing on enzymatic alternatives for these transformations would have the most impact.

New enzyme panels are assembled using a diverse set of enzymes which have been identified through previously demonstrated activities or using predictive tools (Figure 1.2). In some instances (e.g. lipases), a significant number of enzymes are already commercially available, and therefore these panels comprise mostly enzymes from a commercial source.



Scheme 1.1 *Selected enzyme classes currently in GSK's collection.*

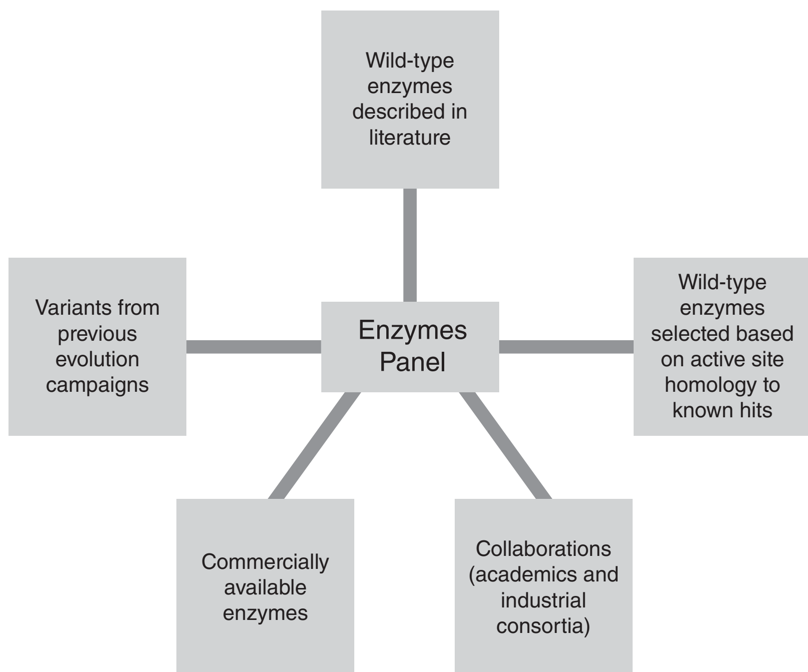


Figure 1.2 *Strategies for generating an enzyme panel.*

Although it is usually trivial to acquire larger quantities of commercial enzymes to facilitate scale-up of any hits for a particular transformation, commercial enzymes can carry intellectual property (IP) restrictions around their use for commercial-scale manufacture and can cause supply-chain concerns as most are single-source. Additionally, it is not usually possible to evolve a commercially-available enzyme without licensing agreements with the source company.

For most other enzyme classes, there are very few, if any, commercially-available enzymes available to screen. Although significant efforts are underway to change the situation, this often means that enzymes of the desired class must be identified by other means. Panels are often assembled by acquisition of the genes encoding wild-type enzymes reported to catalyse the reaction class of interest in the literature. In some cases, mutational studies have also been conducted, and variants of these enzymes may also be acquired.

Bioinformatic tools are the key to the assembly of an enzyme panel, allowing identification of additional putative enzymes from sequence databases based on similarity to known enzymes of the desired class. The goal of a search is straightforward: to identify naturally occurring enzymes that perform the same transformations but maximise their sequence and structural diversity in hopes of maximising the substrate scope, as well as the pliability of the enzymes to be evolved.

As with hit expansion for small molecules, enzymes identified in the literature are often used as seeds for subsequent similarity searching against large annotated sequence datasets, including those from Interpro [7], Uniprot [8], NCBI, PDB [9], CATH [10] and some metagenomics collections [11]. In a standard approach, homologous sequences are identified and clustered, and each cluster's functional annotations are examined. Homologous protein structures are often included in the clustering step as they help in the identification of the relevant clusters from which to sample. Selecting exemplars from the various clusters is often the most challenging part of the process. In a typical scenario, candidates are prioritised based on the availability of experimental annotations, starting with those with experimental data or structures or coming from extremophile organisms. In the absence of additional data, the remaining exemplars are chosen from a diverse set of clades that maximise coverage of each cluster's diversity. The HH-Suite toolset [12], maintained by the Soeding group, provides great tools for identifying homologues, as well as Mmseqs2 [11] and CLANS [13] for clustering. Phylogenetic reconstruction can be performed using MEGA [14] or ETE3 [15] and their corresponding tree-reconstruction and evolutionary-analysis workflows.

One way to improve this approach is to select for enzymes that exhibit similar active sites, and thereby maintain activity and selectivity, by developing 'fingerprints' or protein-based pharmacophores, sometimes referred to as PLIFs, which describe the residue composition and positionality within the active site and the interactions with bound ligands [16]. Once a fingerprint is created, it can be used to identify clusters of sequences that exhibit similar active site makeups, which can then be prioritised for acquisition. The goal of this approach is often to acquire enzymes that minimally perturb the active site but which sample diversity throughout the remainder of the enzyme. The method can also be extended to sample diversity at specific positions within the active site. The success of this approach hinges on the availability of experimentally determined structures from the structural families of interest. The approach also assumes that the active site can be unambiguously identified.

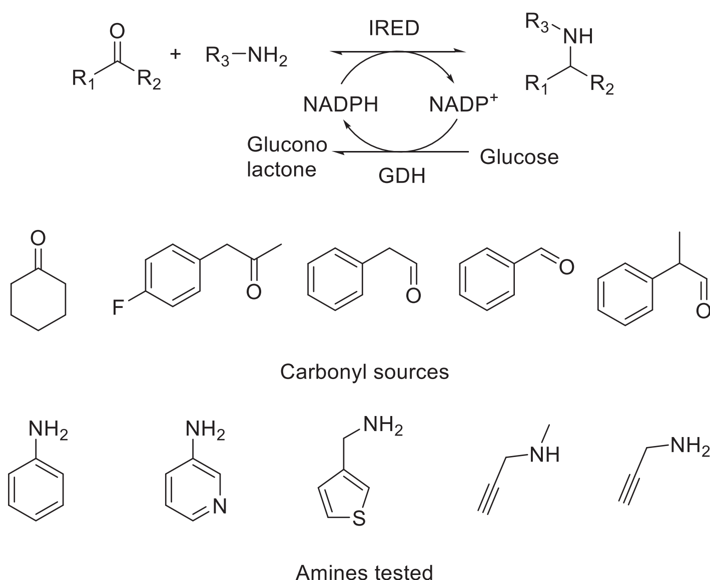
Beyond the acquisition of wild-type enzymes, it is also possible to leverage variants produced in previous enzyme evolution campaigns to enrich a panel. As these evolved enzymes are likely to have generally improved properties (e.g. activity, stability and expression) compared to wild-type enzymes, they can often provide hits more amenable to immediate scale-up. These enzymes can be the result of in-house evolution, in-licensing of other companies' enzyme panels or from collaborations.

As these panels are used for the initial assessment of a biocatalytic transformation, it is important to maintain a significant amount of sequence diversity within them, in order to increase the likelihood that enzymes with the desired activity and selectivity can be identified for a wide range of substrates.

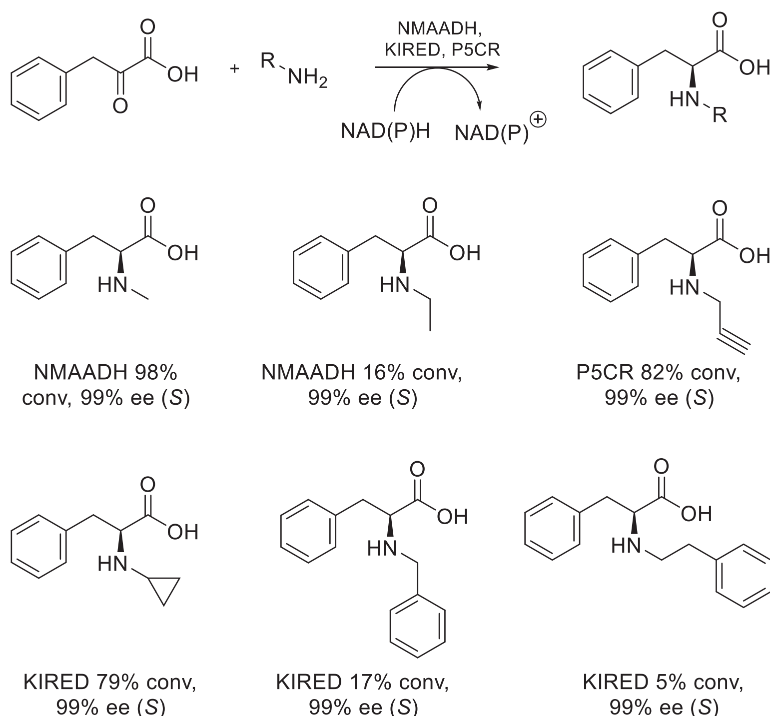
One example is the GSK IRED panel, where 85 IREDs were assembled from different sources [17]. These wild-type variants were tested under industrially relevant conditions using equimolar loadings of amine. Screening a diverse set of substrates, selected to cover a broad chemical space, showed most enzymes were capable of driving reactions to completion (Scheme 1.2). Enzyme-dependent stereoselectivity was observed for many products and successful scale-up was performed for several enzymes and substrates.

Reductive amination of keto acids represents another area where we focused our efforts and successfully generated a small collection of enzymes that are very promising for the synthesis of *N*-alkylated amino acids [18, 19]. Enzyme such as *N*-methylamino acid dehydrogenases (NMAADHs, EC 1.5.1.1 and EC 1.5.1.21), ketimine reductases (KIREDs, EC 1.5.1.25) and Δ^1-pyrroline-5-carboxylate reductases (P5CR, EC 1.5.1.2) have been shown to deliver highly enantioselective alkylated products, further extending the scope of these substrates (Scheme 1.3).

Following several successful in-house enzyme evolution campaigns, we have also constructed enzyme panels incorporating the best variants produced therein. One of the drivers



Scheme 1.2 *Reductive amination of ketones performed by imine reductases.*



Scheme 1.3 *Successful synthesis of alkylation products via enzymatic reductive amination of keto acids.*

for this practice is that, by creating panels of robust enzymes with diverse activities, it becomes more likely that a panel can be used directly without the need for further enzyme engineering – greatly increasing the speed at which the biocatalytic process can be implemented, and therefore increasing uptake across the portfolio.

GSK is also collaborating with various academics in the field with the purpose of continuously expanding its portfolio of enzymes (e.g. halogenases, nitration enzymes, unspecific peroxygenases, etc).

1.4 Enzyme Engineering

With high-quality enzyme panels in hand, it becomes possible to develop a manufacturable process for an API using the hit from panel screening without the need for further enzyme optimisation by either engineering or evolution. The definition of a manufacturable process changes throughout the drug development cycle and becomes more stringent as the asset matures and the quantity of API required increases – making it more likely that engineering of a panel enzyme will be necessary to meet these needs. The specific process parameters required for assets at each stage of development are complex, but the following list summarises some key aspects that must be considered before committing to the evolution of an enzyme: