Selin Kara
Florian Rudroff *Editors*

Enzyme Cascade Design and Modelling

Springer

Enzyme Cascade Design and Modelling

Selin Kara • Florian Rudroff
Editors

Enzyme Cascade Design and Modelling

Springer

Editors
Selin Kara
Department of Biological and Chemical Engineering
Aarhus University
Aarhus, Denmark

Florian Rudroff
Institute of Applied Synthetic Chemistry
TU Wien
Vienna, Austria

ISBN 978-3-030-65717-8 ISBN 978-3-030-65718-5 (eBook)
https://doi.org/10.1007/978-3-030-65718-5

Preface

In the last few decades, the use of natural catalysts—enzymes—in a cascading fashion has become of great interest to synthesize organic compounds. The mimic of nature, to prepare complex structures in a one-pot, is one of the main driving forces why the design of artificial enzyme cascade reactions is so popular. The recent developments in the field prompted us to highlight the advantages, disadvantages, and challenges of designed enzyme cascade reactions.

Although the proof-of-concept for multi-catalytic cascade reactions at laboratory scale looks promising, the number of examples that are applied at technical scales are still limited. From this perspective, we intend to join our forces in this textbook, which shall provide a fundamental background and comprehensive overview of recent developments achieved in the field of catalytic cascades (i.e., multi-enzymatic or chemo-enzymatic) from design (i.e., in vitro and in vivo applications) to kinetic and process modelling, reaction engineering as well as process control.

This textbook is written by experts from different backgrounds who develop and apply enzymatic cascade reactions in their research groups. They address various methodologies from design (i.e., retrosynthesis), kinetic modelling, and process modelling to analytical means for monitoring of enzymatic cascade reactions.

We hope that this book on Enzyme Cascade Design and Modelling will serve as a reference guide for academic and industrial researchers and provide a unique perspective on the design and modelling of enzymatic cascade reactions, both multi-enzymatic and chemo-enzymatic ones. With this, we anticipate to see more and more enzymatic cascades that arose from academic curiosity and were proven at small scales in the laboratories, to be implemented at industrial scales. This book shall be of assistance in the academic as well as in the industrial field for those who want to get an insight into the challenges of developing enzymatic cascade reactions. It is obvious that these challenges become more severe when more enzymes are involved in a synthetic route since more reaction parameters need to be optimized, more kinetic parameters need to be estimated, more degrees of freedom are available, and the complexity is increased. We hope to experience a trigger effect that makes it worthwhile for the readership, the authors, and the editors to have a second edition succeeding the first.

We sincerely hope that here presented knowledge and recent examples will inspire students and early-stage researchers to develop new enzymatic cascade applications providing our society with more sustainable solutions for the synthesis of compounds ranging from bulk chemicals to pharmaceutical intermediates.

We gratefully thank the authors who contributed to our joint "book project," who responded to our feedbacks promptly, who made this textbook ever possible. Finally, yet importantly, we thank our families for their support and tolerance during the time that we invested in the preparation of here presented textbook, which was partly completed during the COVID-19 pandemic that will certainly not be easily forgotten.

Aarhus, Denmark — Selin Kara
Vienna, Austria — Florian Rudroff
August 2020

Contents

Introduction

1

Selin Kara and Florian Rudroff

Abstract

Biocatalysis has become a key emerging field for the development of new synthesis routes for already existing or new-to-nature products to meet the dynamic needs and agendas of our globalizing society. Both fundamental and application-oriented insights obtained in different disciplines from the natural and engineering sciences have opened up new avenues for the use of enzymes in reaction cascades for organic synthesis. The great potential of enzymes becomes significant when they are coupled with other enzymes or with chemocatalysts to synthesize value-added products at industrially relevant product titers. Joining the forces from different disciplines for the use of enzymatic cascade reactions will pave the way for the use of nature's catalysts to produce products of our need.

The application of enzymes for the synthesis of chemicals is a key emerging field to meet the current and future needs of our society [1–3]. Nowadays, the challenges of using enzymes related to the establishment of single-step biotransformations can be overcome by means of molecular biotechnology techniques for designing biocatalysts with custom characteristics such as (1) high activity towards a non-natural substrate, (2) high stability at elevated substrate/product concentrations, and (3) high selectivity.

The design of multi-step cascade reactions has received great attention in the biocatalysis community. One of the main advantages of synthetic cascade reactions is the possibility to produce complex molecules that cannot be easily obtained from a single biotransformation. Additionally, cascade reactions offer solutions for a range of challenges faced in biocatalysis; the most important ones are:

1. reduced inhibition issues caused by intermediates (since they are formed in situ at low amounts and are immediately consumed),
2. no intermediate purification steps leading to low waste generation,
3. low consumption of resources (e.g., space, time, energy, and materials),
4. displacement of unfavorable thermodynamic equilibria, and
5. possibility to synthesize complex molecules from simple and readily available compounds.

Indeed, nature already uses an elegant and efficient synthetic strategy: Coupling enzymes in multi-step pathways without intermediate isolation and purification steps and with precise spatial

S. Kara (✉)
Department of Biological and Chemical Engineering, Aarhus University, Aarhus, Denmark
e-mail: selin.kara@eng.au.dk

F. Rudroff
Institute of Applied Synthetic Chemistry, TU Wien, Vienna, Austria

S. Kara, F. Rudroff (eds.), *Enzyme Cascade Design and Modelling*,
https://doi.org/10.1007/978-3-030-65718-5_1

control of catalysis. Hence, by mimicking nature the ultimate goal in multi-step biocatalysis is to design cascade reactions (multi-enzymatic and chemo-enzymatic) that run perfectly in balance with excellent yields and selectivities towards the final target product.

To transfer nature's synthetic strategy into the laboratory and industry has become a major focus of the biocatalysis community in recent years [4–12]. In fact, the potential application fields of multi-enzymatic reaction cascades are extremely diverse; they range from the synthesis of fine chemicals and active pharmaceutical intermediates (APIs) [13]—which is an established field of biocatalysis—to the conversion of renewable raw materials into platform chemicals.

Nature's wealth in biologically active and complex molecules has inspired synthetic chemists for centuries. They strived to imitate natural chemical processes and synthesize biomaterials through biomimetic synthesis, asymmetric catalysis, and natural product synthesis. The way in which chemists have designed chemical routes towards complex target molecules has been changed in the mid of the last century by the introduction of the concept of retrosynthetic analysis [14–16]. This theoretical approach is based on the systematic disconnection of strategic chemical bonds that link major components of a specific target molecule until simple or readily available starting materials are obtained. Crucial for each retrosynthetic step is the existence of a feasible chemical transformation in the synthetic forward direction. Consequently, this approach heavily depends on the knowledge of all organic reactions. Synthetic chemists are well grounded in retrosynthesis, as it is part of their education to design novel synthetic strategies for the preparation of complex molecules. This process has been aided by the formalization of how to generate "synthons" by homolytic or heterolytic cleavage of C–C or C–X bonds in the reverse direction. In addition, functional group interconversions (FGIs) have been introduced to prepare the molecule for the best retrosynthetic disconnection. This concept served as the basis for the development of a vast number of novel transformations in organic chemistry and constantly increases the toolbox of synthetic methodologies (e.g., metal- and organo-catalysis). On top of that, nature can serve as an additional pool for novel chemical transformations, especially with respect to exceptional chemo-, regio-, diastereo-, and enantioselectivity. The catalytic ability, natural diversity, and evolvability of enzymes extend the chemical space and complement the synthetic chemists' transformation portfolio [16–18]. Therefore, the continuously expanding biocatalytic toolbox has become an indispensable part of the chemical wealth.

Thanks to the pioneering work in molecular biology, microbiology, genetics, and biotechnology, tremendous progress in DNA technologies (e.g., directed evolution), high-throughput screening methods, and bioinformatics has been achieved, which led to the development of novel and tailor-made biocatalysts and enabled their production on industrial scale. During the past 15 years, more than 20 different types of enzymes (e.g., hydrolytic-, reverse hydrolytic-, oxidative-, reductive-, C–C and C–X bond-forming enzymes) have become commercially available and have been applied for the synthesis of chiral building blocks, pharmaceuticals, agrochemicals, polymers, and biofuels. These biocatalysts have been exploited in the targeted synthesis of complex molecules since they are exceptionally effective at catalyzing FGIs and often offer an appealing alternative to chemocatalysts. With this catalytic portfolio in hands, guidelines for the so-called biocatalytic retrosynthesis have been established (Chap. 2). Thereby biocatalysts as well as chemocatalysts and reagents are considered for key bond-forming steps in the targeted synthesis of complex molecules. Applying the principles of biocatalytic retrosynthesis, novel disconnections can be proposed that are not accessible by classical chemical catalysis, which facilitates the design of alternative synthetic strategies.

Nature's synthetic strategy is applied in a (bio)chemical laboratory by using either microorganisms (i.e., whole-cells) (in vivo) or isolated enzymes (in vitro). The main advantages of the in vivo approach are: (1) it is inexpensive as no enzyme isolation or purification is required,

(2) it offers high enzyme stability as enzymes are in their natural environment, and (3) there is no need for the addition of external cofactors as they are directly provided from the cell metabolism. However, there may be some major disadvantages, such as: (1) the molecular design of microorganisms (i.e., "designer bugs") [19–25] can be material and time-intensive, (2) the competitive reactions catalyzed by other intracellular enzymes in the host can lead to low selectivity and productivity, (3) high substrate and product concentrations—which are prerequisites at industrial scales—can be toxic to the cells, (4) adjusting the enzymes' expression levels is not straightforward, and (5) controlling and scaling up the bioprocess can be difficult.

Unlike the in vivo approach, in vitro reaction systems can be easily controlled and optimized, and high product yields and purities can be achieved due to the absence of aforementioned competing side-reactions. In principle, the advantages of the in vivo approach can be regarded as the disadvantages of the in vitro strategy. However, with the current technology and knowledge, it is possible to optimize in vitro systems by means of methods such as (1) automated protein purification, (2) reuse of expensive cofactors with a broad spectrum of regeneration techniques, (3) alteration of enzyme pH and temperature profiles, (4) enzyme stabilization (*via* immobilization, protein engineering, or use of additives or cosolvents), and (5) compartmentalization of incompatible enzymes and/or their reaction conditions, as nature does in cellular organelles or compartments.

Biocatalytic cascades are defined as two- or multi-step transformations carried out in the same reaction vessel (the so-called one-pot) using at least one enzyme; hence, they cover multi-enzymatic [5, 10, 26, 27], chemo-enzymatic [28–30], photo-enzymatic [31, 32], electro-enzymatic [33, 34], and enzyme-initiated spontaneous [35, 36] reactions. Among those, multi-enzymatic and chemo-enzymatic cascade reactions have been leading to higher productivities and hence will be of high interest for organic synthesis. However, it does not necessarily mean that the cascade reactions run concurrently; they can run in a sequential fashion as well. As long as a series of reactions takes place in a one-pot system, the term "cascade" can be used as a definition in the field of biocatalysis [37]. It is important to note that the term "reactor cascades" is commonly used in the field of engineering, whereby a reaction occurs in a series of linked reactors with each processing the output of the previous one. The biocatalytic cascade reactions reported so far have been categorized by their different designs, for which we hereby refer to excellent reviews summarizing them [4–12].

Whereas major attention has been put on the development of enzymatic cascade reactions in vitro as well as in vivo, the combination of two disciplines, chemocatalysis and biocatalysis, is surprisingly underrepresented in the literature, although both research fields cover a significantly different chemical space in terms of reactivity, selectivity, and productivity. In the last decade, intensive investigations for the introduction of bioorthogonal functionalities were made to gain deeper insights into cellular mechanisms. Different types of reactions, so far unknown in nature, such as metal assisted C–C couplings (e.g., Suzuki, Negishi, Sonogashira), cross-metathesis or copper-catalyzed [2+3] dipolar cycloaddition (Huisgen reaction), were explored. Conversely, biocatalysis offers the possibility of chemical transformations either unknown to or poorly understood by chemists—like the C–H activation of unactivated C–H bonds—or enables increased yields by improved regio-, stereo-, and chemoselectivity in already known reactions.

The combination of the two worlds of bio- and chemocatalysis would open a completely new way to synthesize complex molecules by taking advantage of their individual assets. Nevertheless, cascade type reactions involving both disciplines suffer from incompatibilities of their totally different windows of operation. Whereas many metal- or organocatalytic systems demand water-free conditions, the absence of oxygen or high substrate concentrations, enzymes mainly work in aqueous conditions at ambient temperature with a significant lower substrate load. Most critical is the inactivation of the catalytic system

by either the chemocatalyst or the enzyme, which has a severe effect on the overall reaction performance. Recent advances in protein engineering provided the community with tailor-made enzymes with improved stability, substrate scope, and selectivity. Also chemists improved the stability of chemocatalysts towards water and oxygen significantly. However, combination of both catalytic systems in a cascade fashion is still a challenging task (Chap. 5) [28–30, 37].

Spatial control plays a particularly important role for chemo-enzymatic cascades since biocatalytic and chemocatalytic reactions often run under significantly different reaction conditions, e.g., temperature, pH, and medium. It is also worth mentioning that spatial arrangements may also be required in enzymatic cascades when different enzymes need different optimal reaction conditions to work. In this context, the cell is perhaps the most prominent spatially separated organization, but a myriad of other systems like organelles or micro-compartments, anchoring or assembling mega-enzyme complexes, facilitate the control of complex multi-step reaction cascades. Spatial separation (the so-called modularization or compartmentalization) [38–42] of enzyme pathways allows suppressing side-reactions, alleviating the effect of reactive species and avoiding inhibition. Conversely, if the biocatalysts' optimal conditions are not significantly divergent, bringing the (bio)catalysts spatially together (the so-called co-localization) [9, 43] would provide efficient channeling of substrates between the catalytic steps.

Having highlighted the attractiveness of designing multi-step cascade reactions for organic synthesis, it has to be pointed out that complexity increases with the number of enzymes, chemocatalysts, and compounds in a reacting system since the number of dependencies between different variables multiplies. For such complex and interconnected multi-catalytic systems to work and to become fully applicable on a technical scale, it is important to view cascade reactions from a reaction engineering perspective. To understand and optimize them, kinetic modelling of all the numerous interdependencies will guide us in facing the challenges towards implementation of cascade reactions on a larger scale [44, 45]. By elucidating the kinetics bottlenecks and developing models to describe cascade reactions, it is possible to implement multi-step cascade reactions in their best suitable reactor and operation mode. For this purpose, enzyme kinetics modelling based proper bioreactor selection is necessary to operate a cascade reaction with high productivity and selectivity towards the target product. Challenges of cascade reactions related to cross-reactivity, cross inhibition, optimization as well as reaction control highly suggest that compartmentalization (or modularization), rather than one-pot synthesis, is a particularly attractive route for technical implementation [46, 47]. After the enzyme kinetics modelling guided reactor engineering, the next step is the cascade process design for the optimization of individual cascade modules and the complete multi-step multi-catalytic process.

For monitoring and controlling of multi-step enzymatic cascade reactions, process analytical technologies (PAT) are a powerful tool allowing immediate *online* or *inline* data collection, data processing, and fast response to (re)optimize the process. During the last decades, the developments in analytical techniques and (portable) devices has led to new levels of precision in detection of analytes, even at high dilutions. Multi-variate *inline* or *online* data are processed *via* chemometric models for the calculation of concentration data, which is sometimes provided with the analytical technique as an imbedded software tool. In order to get industrially relevant product titers in multi-step reaction cascades (multi-enzymatic or chemo-enzymatic) the use of non-aqueous media has attracted a great deal of interest in the biocatalysis community. Organic media and neoteric solvents are explored to enhance the productivity as well as sustainability of biocatalytic cascades as the perception of water as a green- and mild reaction medium has changed. This is simply because wastewater generated in enzymatic reactions is/should be considered as a bottleneck for the E-factor (=mass of waste generated per mass of product synthesized) [48, 49].

Alongside insights into the scientific revolution in the field of enzymatic cascades, this book provides a fundamental background and

comprehensive overview of recent developments achieved in the field of enzymatic cascades (e.g., multi-enzymatic or chemo-enzymatic). This ranges from design (i.e., in vitro and in vivo applications) to kinetic and process modelling, reaction engineering as well as process control. In the first part of the book (Chaps. 2–5), our main goal is to discuss the opportunities and challenges of building multi-step cascade reactions, whereby the latter are critically assessed in each chapter. The second part of this book (Chaps. 6–9) provides the problem-solving methods to alleviate these challenges on the way to implementation. By doing so, the ultimate goal of this book is to deliver a road map from challenge identification to overcoming them. As aforementioned, not only multi-enzymatic cascades but also chemo-enzymatic cascades are presented with the motivation of combining the strengths of these two worlds that cover selectivity, activity, and robustness and to assess the associated challenges. Chapter 10 is dedicated to the application of enzymatic cascade reactions in non-conventional media from an industrial perspective focusing on industrially relevant product titers and recent achievements in a technical environment.

References

1. Vogel A, May O (2019) Industrial enzyme applications. Wiley, Weinheim. https://doi.org/10.1002/9783527813780
2. Sheldon RA, Woodley JM (2018) Chem Rev 118:801–838
3. Sheldon RA, Pereira PC (2017) Chem Soc Rev 46:2678–2691
4. Bruggink A, Schoevaart R, Kieboom T (2003) Org Process Res Dev 7:622–640
5. Ricca E, Brucher B, Schrittwieser JH (2011) Adv Synth Catal 353:2239–2262
6. García-Junceda E, Lavandera I, Rother D, Schrittwieser JH (2015) J Mol Catal B Enzym 114:1–6
7. Muschiol J, Peters C, Oberleitner N, Mihovilovic MD, Bornscheuer UT, Rudroff F (2015) Chem Commun 51:5798–5811
8. O'Reilly E, Turner NJ (2015) Perspect Sci 4:55–61
9. France SP, Hepworth LJ, Turner NJ, Flitsch SL (2017) ACS Catal 7:710–724
10. Schrittwieser JH, Velikogne S, Hall M, Kroutil W (2017) Chem Rev 118(1):270–348
11. Oroz-Guinea I, García-Junceda E (2013) Curr Opin Chem Biol 17:236–249
12. Santacoloma PA, Sin G, Gernaey KV, Woodley JM (2011) Org Process Res Dev 15:203–212
13. Woodley JM (2016) Green biocatalysis. Wiley, Hoboken, pp 503–518
14. Turner NJ, Humphreys L (2018) Biocatalysis in organic synthesis: the retrosynthesis approach. Royal Society of Chemistry, London
15. Green AP, Turner NJ (2016) Perspect Sci 9:42–48
16. Turner NJ, O'Reilly E (2013) Nat Chem Biol 9:285–288
17. de Souza ROMA, Miranda LSM, Bornscheuer UT (2017) A retrosynthesis approach for biocatalysis in organic synthesis. Chemistry 23:12040–12063
18. Honig M, Sondermann P, Turner NJ, Carreira EM (2017) Angew Chem Int Ed 56:8942–8973
19. Gröger H, Chamouleau F, Orologas N, Rollmann C, Drauz K, Hummel W, Weckbecker A, May O (2006) Angew Chem Int Ed 45:5677–5681
20. Berkessel A, Rollmann C, Chamouleau F, Labs S, May O, Gröger H (2007) Adv Synth Catal 349:2697–2704
21. Ema T, Ide S, Okita N, Sakai T (2008) Adv Synth Catal 350:2039–2044
22. Ni Y, Li C-X, Zhang J, Shen N-D, Bornscheuer UT, Xu J-H (2011) Adv Synth Catal 353:1213–1217
23. Ni Y, Pan J, Ma H-M, Li C-X, Zhang J, Zheng G-W, Xu J-H (2012) Tetrahedron Lett 53:4715–4717
24. Shen N-D, Ni Y, Ma H-M, Wang L-J, Li C-X, Zheng G-W, Zhang J, Xu J-H (2012) Org Lett 14:1982–1985
25. Eixelsberger T, Woodley JM, Nidetzky B, Kratzer R (2013) Biotechnol Bioeng 110:2311–2315
26. Lopez-Gallego F, Schmidt-Dannert C (2010) Curr Opin Chem Biol 14:174–183
27. Rudroff F (2019) Curr Opin Chem Biol 49:84–90
28. Gröger H, Hummel W (2014) Curr Opin Chem Biol 19:171–179
29. Verho O, Bäckvall J-E (2015) J Am Chem Soc 137:3996–4009
30. Wang Y, Zhao H (2016) Catalysts 6:194
31. Seel CJ, Gulder T (2019) Chembiochem 20:1871–1897
32. Macia-Agullo JA, Corma A, Garcia H (2015) Chemistry 21:10940–10959
33. Lütz S, Vuorilehto K, Liese A (2007) Biotechnol Bioeng 98:525–534
34. Kohlmann C, Märkle W, Lütz S (2008) J Mol Catal B Enzym 51:57–72
35. Kroutil W, Mayer SF, Faber K (2000) Proceedings of ECSOC-4. MDPI, Basel
36. Hollmann F, Arends IWCE (2012) Polymers 4:759–793
37. Rudroff F, Mihovilovic MD, Gröger H, Snajdrova R, Iding H, Bornscheuer UT (2018) Nat Catal 1:12–22
38. Wu C, Kraume M, Ansorge-Schumacher M (2011) ChemCatChem 3:1314–1319
39. von Langermann J, Wapenhensch S (2014) Adv Synth Catal 356:2989–2997
40. Sato H, Hummel W, Gröger H (2015) Angew Chem Int Ed 54:4488–4492
41. Schaaf P, Bayer T, Koley M, Schnürch M, Bornscheuer UT, Rudroff F, Mihovilovic MD (2018) Chem Commun 54:12978–12981

42. Benítez-Mateos AI, Contente ML, Velasco-Lozano S, Paradisi F, López-Gallego F (2018) ACS Sustain Chem Eng 6(10):13151–13159
43. Kara S, Schrittwieser JH (2017) BIOspektrum 23:468–470
44. Pesci L, Kara S, Liese A (2018) Einführung in die enzymtechnologie. Springer, Berlin, pp 53–75
45. Scherkus C, Schmidt S, Bornscheuer UT, Gröger H, Kara S, Liese A (2017) Biotechnol Bioeng 114:1215–1221
46. Kara S, von Langermann J (2018) Einführung in die enzymtechnologie. Springer, Berlin, pp 225–242
47. Pesci L, Baydar M, Glueck S, Faber K, Liese A, Kara S (2016) Org Process Res Dev 21:85–93
48. Sheldon RA (2008) Chem Commun 2008:3352–3365
49. Tieves F, Tonin F, Fernández-Fueyo E, Robbins JM, Bommarius B, Bommarius AS, Alcalde M, Hollmann F (2019) Tetrahedron 75:1311–1314

Enzyme Cascade Design: Retrosynthesis Approach

2

William Finnigan, Sabine L. Flitsch, Lorna J. Hepworth, and Nicholas J. Turner

Abstract

Retrosynthetic analysis for the design of synthetic routes towards target molecules is well-established in organic chemistry, and has been extended to include biocatalysis in recent years. The increasing number of transformations known to be catalysed by enzymes, whilst ultimately rendering biocatalytic retrosynthesis more powerful, necessitates the use of computational tools if biocatalysis is to reach its full potential. In the following chapter, we outline the pipeline required to go from pathway generation towards a target molecule, to construction of selected optimal pathways in the laboratory and the techniques currently used to analyse them. We compare manual vs. computer-assisted approaches for each step of the workflow. Current computational tools used for automated identification of suitable enzymes, such as molecular fingerprinting and structure-based substrate docking, and the evaluation of metrics that can be used to rank order the generated pathways, will also be discussed. Finally, we discuss a number of recent high-throughput analytical techniques for the experimental validation of potential pathways, leveraging the design-build-test-analyse cycle for pathway improvement.

Keywords

Biocatalysis · Retrosynthesis · Enzyme cascade · CASP

2.1 Introduction

The application of retrosynthesis during the planning and execution of preparative routes to target molecules is now an established and indispensable tool in organic synthesis. The concept was first introduced in the 1960s by E.J. Corey and emerged from a recognition that the careful and logical analysis of possible modes of construction of bonds could reveal a path from the target molecule back to potential starting materials in a step-wise fashion [1–3]. Having identified potential 'synthons' via the process, the next step was to correlate these synthons with real chemical building blocks that could be employed in the synthetic direction. In some cases it was necessary to carry out Functional Group Interconversions (FGIs) in order to reveal potential disconnections that might otherwise not be apparent. Retrosynthesis is a powerful way of teaching organic chemistry to undergraduate students [4]. After the initial shock of being asked to think backwards from target molecule to starting material, rather than vice versa,

W. Finnigan · S. L. Flitsch · L. J. Hepworth · N. J. Turner (✉)
School of Chemistry, University of Manchester, Manchester Institute of Biotechnology, Manchester, UK
e-mail: nicholas.turner@manchester.ac.uk

S. Kara, F. Rudroff (eds.), *Enzyme Cascade Design and Modelling*,
https://doi.org/10.1007/978-3-030-65718-5_2

retrosynthetic analysis reinforces basic concepts such as mechanism and the nature of bond forming processes.

Implicit in the successful application of retrosynthetic analysis is a comprehensive understanding of exactly what reactions are available in the 'forward' direction once the required building blocks have been identified. In the early days of retrosynthesis, a highly skilled synthetic organic chemist may have been able to recall from memory, or from a quick read of the literature or a textbook, most of the important synthetic transformations that were required. However, the landscape of organic synthesis has changed dramatically during the past 50 years, with the invention of genuinely new reactions as well as reagents and, increasingly, catalysts for highly selective organic synthesis. Target molecules can also be structurally and stereochemically complex resulting in the generation of many possible routes for synthesis. As a consequence, in recent years there has been an attempt to apply computational methods to assist with both retrosynthetic analysis and also matching potential synthons with known building blocks and reactions/reagents/catalysts found in databases. Computational methods can of course handle very large data sets and can also incorporate machine learning and path finding algorithms in order to meet the challenge of 'computer-aided organic synthesis'.

2.1.1 Biocatalytic Retrosynthesis

In 2013 we proposed the concept of 'biocatalytic retrosynthesis' and suggested that the application of retrosynthetic analysis be extended to include biocatalysts [5]. By 2013 substantial developments had taken place globally in the discovery and development of new classes of biocatalysts for more general application in organic synthesis. In addition to the more established enzymes such as lipases, esterases, and dehydrogenases, many more (engineered) biocatalysts were becoming available including transaminases, oxidases, aldolases, oxygenases, P450 monooxygenases, and nitrilases [6]. This expansion of the biocatalytic toolbox has accelerated during the past few years with the arrival of new platforms such as imine reductases/reductive aminases, halogenases, P450 variants for C-H activation, peroxygenases, and carboxylic acid reductases amongst others [7]. Today there are *ca*. 100 different biocatalytic transformations that can be routinely carried out using commercially available biocatalysts [8] and hence the application of retrosynthetic tools in biocatalysis is very timely [9].

The power of biocatalytic retrosynthesis is that it can often result in the identification of completely new routes to target molecules when compared with the existing and more classical chemical routes. Some enzymes catalyse reactions that have no known (currently) non-enzymatic counterpart, e.g. ammonia lyases which catalyse the (enantioselective) addition of ammonia to the double bond of a cinnamic acid derivative to yield the corresponding L-aryl alanine product. P450 monooxygenases are able to insert oxygen into C–H bonds in ways which are unknown in organic synthesis. Recently these enzymes have been engineered to functionalise C–H bonds in a much broader sense enabling C–C and C–N bond forming processes, a development which will impact the way that biocatalysis can be employed in the future [10].

Interestingly, since biocatalysis is a relatively under-developed field, there are parallels with the early days of organic synthesis in the ways in which biocatalytic retrosynthesis is currently applied [11]. During the preparation of our recent book *Biocatalysis in Organic Synthesis—The Retrosynthesis Approach* we identified *ca.* 250 different 'retrosynthetic disconnections' which were possible using biocatalysis. This number is increasing and is probably closer to 300 at the time of writing (July 2020). With a good knowledge and understanding of organic chemistry and an awareness of what biocatalysts are available, it is currently possible to 'manually' apply biocatalytic retrosynthesis to analyse possible routes to target molecules. However, for reasons set out below, it is clear that biocatalytic retrosynthesis needs to embrace computational tools if it is to fully realise its potential and

widespread application. The application of biocatalysts in organic synthesis presents some special challenges where computational algorithms will become essential in the future. An obvious example is the increasing need to search large enzyme databases which are emerging from metagenomic and protein engineering programmes. Retrosynthetic analysis may identify the need for an imine reductase (IRED) in a key step. However, inspection of the database will reveal that there are >5000 different sequences known that potentially might catalyse the desired reaction! Another challenge, and indeed real opportunity, for biocatalytic synthesis/retrosynthesis is in the rapidly expanding area of enzymatic and chemo-enzymatic cascades where a multi-enzyme pathway needs to be designed, engineered, and optimised to convert the starting material to product.

2.1.2 Enzyme Cascade Design (Biosynthetic)

Nature exploits multi-enzyme cascade processes as a universal platform for creating the diversity of natural products that underpin life [12]. Compounds such as alkaloids, terpenes, and polyketides are all made from simple precursors by a sequence of enzyme-catalysed processes involving cofactors, coenzymes, and co-substrates. This process of natural product biosynthesis is remarkable from a synthetic chemistry perspective. Every step involves a (bio)-catalyst, no protecting groups are required, waste and by-products are minimised, and the final products often possess the molecular and stereochemical complexity necessary for biological activity (e.g. reserpine, penicillin, vitamin B12). The synthesis is also encoded (by genes) and can therefore be optimised through either engineering individual genes/biocatalysts or the whole pathway (or both). Reactions are carried out under mild conditions in aqueous environments, and are generally sustainable due to the fact that most reaction components are both biodegradable and renewable. Such an approach sounds like an excellent blueprint for organic synthesis!

The field of synthetic biology has sought to exploit this biosynthetic platform in order to improve the production levels of important and commercially valuable natural products (e.g. vanillin, resveratrol, D-methionine). Having established the pathway required for the biosynthesis of the target molecules in vivo, a range of different methods can then be applied to enhance production levels, e.g. gene duplication, gene deletion, altering circuitry, metabolic flux analysis, altering cofactor levels. Computational methods are widely applied in this field in order to search through the number of possible pathways and rank order the different options followed by experimental cycles of design-build-test-analyse (DBTA). Some spectacular successes have been achieved, notably the production of propane-1,3-diol (DuPont, United States).

2.1.3 Enzyme Cascade Design (Synthetic)

When moving to other sectors of the chemical industry, for example in the manufacture of intermediates for pharmaceuticals and agrochemicals, monomers for polymers, flavour components, fragrances, and fine chemicals, a major challenge emerges in that the target molecule is now ‘synthetic’, i.e. it is not ‘natural’, and hence no biosynthetic pathway will currently exist for its production. If a multi-enzyme cascade is to be used for its preparation, then a new pathway will have to be identified, constructed, and optimised. It may also be the case that the starting material required for its synthesis is ‘non-natural’ and hence the whole approach starts to resemble more closely that encountered in classical organic synthesis in which petrochemical-derived building blocks are converted via a series of reagents and catalysts to the target synthetic molecule.

Figure 2.1 sets out the workflow required to construct a multi-enzyme (chemo)biocatalytic cascade for the synthesis of any desired target

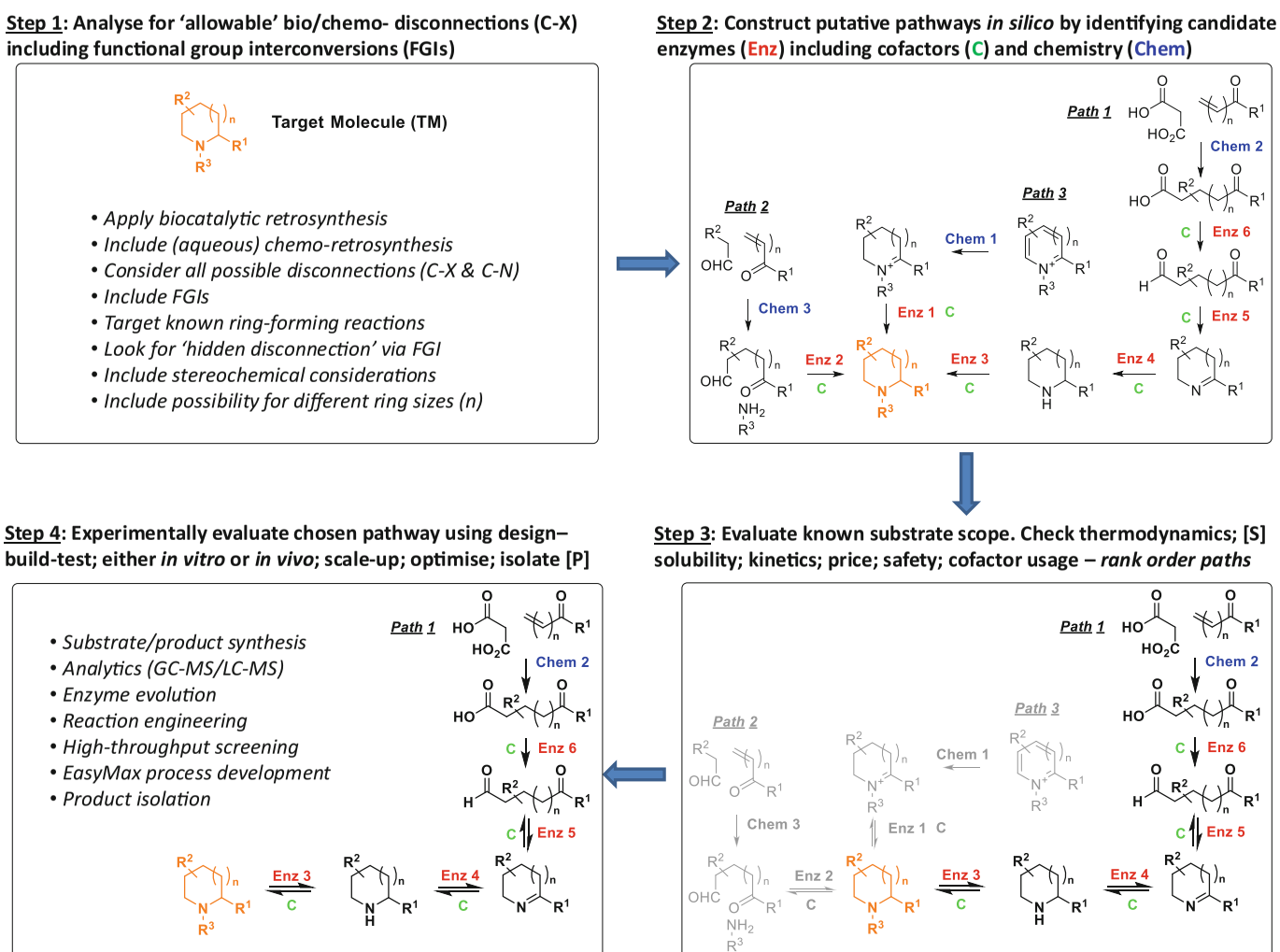

Fig. 2.1 Overview of workflow required for the design and validation of a biocatalytic pathway to a target molecule

molecule (TM). Step 1 involves the initial biocatalytic retrosynthesis in which the target molecule is subjected to a systematic analysis of how it might be disconnected into simpler precursors. At this stage both biocatalytic and aqueous compatible chemistries are included in order to potentially identify chemo-enzymatic as well as purely biocatalytic cascades. Note that it may also be necessary to introduce FGIs (e.g. alcohol to ketone, C–C to C=C) in order to generate additional possible retrosynthetic pathways by identifying 'hidden' disconnections. The data gathered from Step 1 is then fed into Step 2 where potential synthetic pathways are now constructed in silico for synthesis of the TM. For each step of the pathway either a biocatalyst (Enz) or chemical reagent (Chem) will need to be identified by interrogation of available databases. Cofactor requirements (C) can also be added at this stage together with any co-substrates required. Since several biocatalytic steps may be cofactor dependent it will be possible to generate cofactor networks to minimise overall net consumption of NADH, ATP, etc. The output from Step 2 is a complete set of putative pathways for further evaluation. In Step 3 additional parameters are added including, importantly, any known information concerning the substrate specificity and kinetics of the required enzymes. Other data such as substrate solubility, safety, commercial availability, and price can also be added. Each pathway will also be assessed for overall thermodynamics by adding known information concerning reversibility of each step together with the free energy of the reaction. At the end of Step 3 an overall assessment of each potential pathway will then be carried out enabling a rank order score to be assigned which will guide further work. In Step 4 the experimental begins in which pathways are now physically assembled in the laboratory, either in vitro or in vivo, to allow evaluation and screening for product and intermediate production. At this stage further enzyme evolution may be required in order to improve the kinetics of individual biocatalysts which are identified as rate limiting. Once a pathway has reached a proof-of-concept stage, in terms of meeting basic criteria for synthesis of the TM, it can be taken forward for further optimisation including reaction engineering and product isolation.

The purpose of this introductory chapter is (1) firstly, to map out the overall process and steps that need to be undertaken to design and construct a chemo-bio-catalytic cascade (Fig. 2.1) and (2) secondly, to highlight the various challenges and problems that need to be addressed in each of the individual Steps 1–4 in order to make the selection process as rapid and effective as possible. The subsequent sections in the book provide a comprehensive survey of the increasing number of chemo-bio-catalytic cascades (see Chaps. 3–5) that have been successfully designed and implemented and, in some cases, optimised for production on scale.

2.2 Retrosynthesis to Produce Pathways to the Target Molecule (Step 1)

2.2.1 Manual Retrosynthesis

We recently reported a comparison of six different biocatalytic routes for the preparation of the amino acid D-1 which is an intermediate in the synthesis of the antidiabetic drug sitagliptin [13]. In this example, 'manual retrosynthesis' was applied to generate the six pathways, all of which in this example used the same starting material, namely the aldehyde 2 (Fig. 2.2). Having conceptualised the different pathways, each was then constructed and executed experimentally in order to generate data to allow comparison between the routes in terms of overall yield and also green metrics. Interestingly, in this example, a pool of six different enzymes was used to create the six different pathways by appropriate choice of substrate and enzyme.

However, the example shown in Fig. 2.2 brings into focus the need to now move from 'manual retrosynthesis' to computer-assisted automated retrosynthesis for the following reasons: (1) the retrosynthetic analysis of D-1, and its subsequent synthesis from 2 using the six different routes shown, was greatly facilitated by

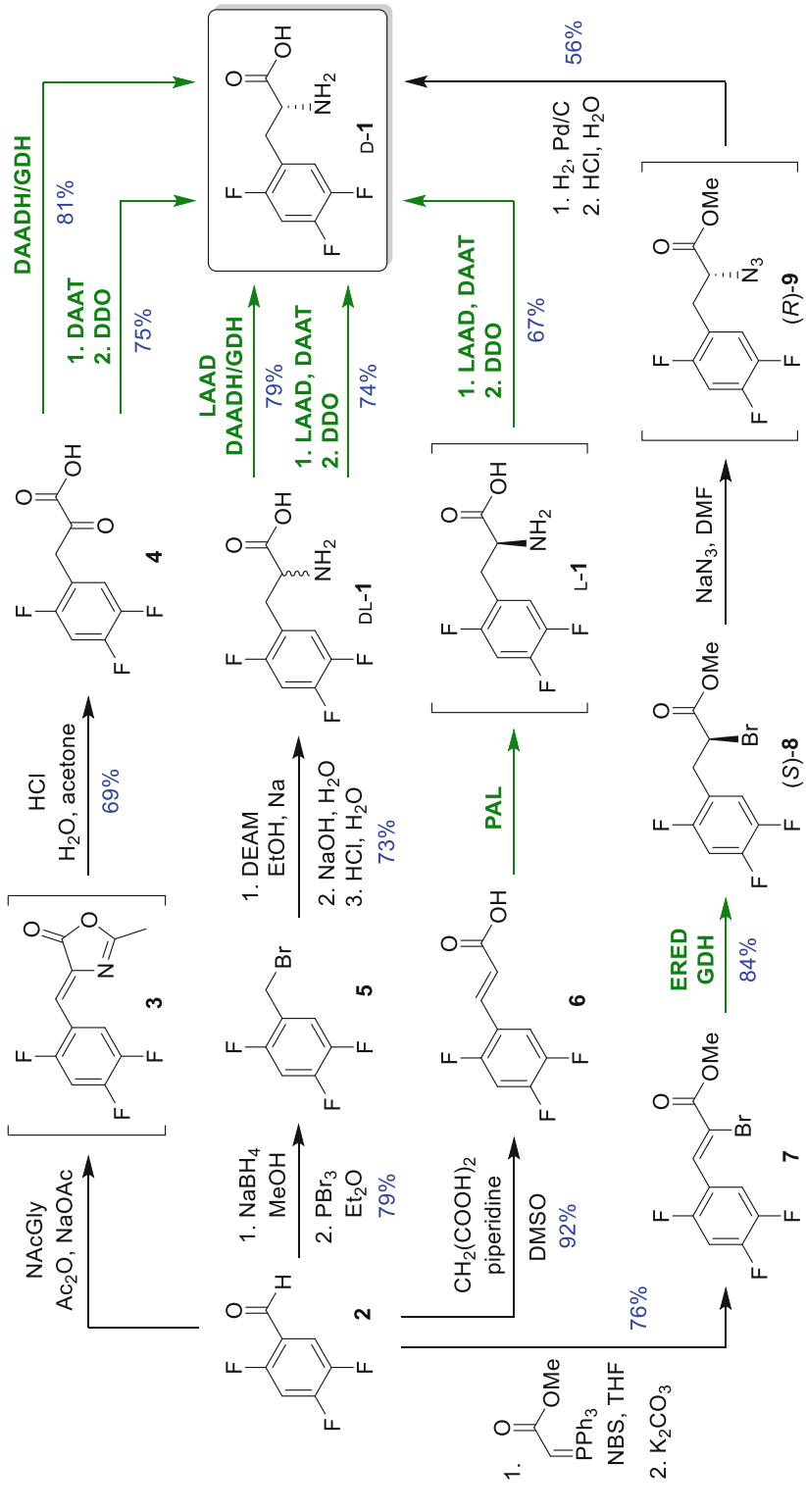

Fig. 2.2 Six complementary biocatalytic routes for the synthesis of α-amino acid D-1 from the aldehyde 2. (*DAADH* D-amino acid dehydrogenase, *GDH* glucose dehydrogenase, *DAAT* D-amino acid transaminase, *DDO* D-aspartate oxidase, *LAAD* L-amino acid deaminase, *ERED* ene reductase) [13]