APPLICATION OF COMPUTATIONAL TOOLS FOR THE DESIGN OF ENZYME CASCADES

SABINE L FLITSCH*

*School of Chemistry and MIB, The University of Manchester, Manchester, M1 7DN, United Kingdom

Our view on the present state of research on biocatalysis

The combination of sequential biocatalytic reactions in non-natural synthetic cascades is a rapidly developing field and has led to the generation of complex valuable chemicals from simple precursors [1-5]. These enzyme cascades can often be telescoped into one single reactor, either using cell free enzymes, whole cells or a mixture thereof, because many enzymatic reactions use similar reaction condition. With the toolbox of natural and engineered biocatalysts increasing dramatically through metagenomics data and protein engineering, so do the options for biocatalytic retrosynthesis of a target molecule, leading to new routes employing enzymatic transformations [6]. Until recently, the retrosynthetic analysis and the design of enzyme cascades was performed manually and limited highly skilled and trained specialists who have intricate knowledge of the field of biocatalysis. As the field expands dramatically in its coverage of chemical reactions and processes and becomes increasingly data rich, computational tools are emerging and helping to capture information from the literature and design enzyme cascades. These computational tools have become useful for the expert, but are also designed to provide for a wider chemical community to find biocatalytic synthetic strategies for developing more efficient and green synthetic processes.

Our recent research contributions to biocatalysis

The planning of a synthetic strategy starts with considering the broad types of reactions (reaction rules) that could be used in a stepwise fashion towards the target from accessible starting materials. In organic chemistry, the strategy of 'retrosynthesis', *i.e.* planning backwards from the target has proven very useful. Based on the chemical retrosynthesis concept, a collection of tools for automated biocatalytic cascade design ('RetroBioCat'; https://retrobiocat.com) was developed [7] (Figure 1). A database of reaction rules was established, compiled of chemical reactions that have been used for biocatalytic transformations. In the first instance, RetroBioCat applies these reaction rules iteratively for retrosynthesis towards a suitable starting material. For example, for the piperidine target in Figure 1, RetroBioCat would suggest three strategies involving either three steps (carboxylic acid reduction by CAR; transamination by a TA; imine reduction by IRED) or two steps (alcohol oxidation by AlOx; reductive amination by RedAM; or

amine oxidation AmOx followed by imine reduction by IRED) from the different respective starting materials.

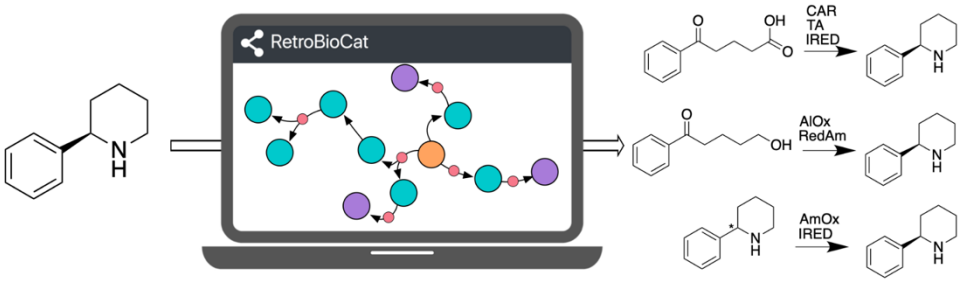

Fig. 1. RetroBioCat provides a collection of tools for automated biocatalytic cascade design. For example, three pathways towards the chiral piperidine target are suggested.

A number of pathways suggested by RetroBioCat have already been implemented successfully. Two recent examples from our laboratory are the chiral amino polyols **1** and **2** in Figure 2, both highly polar targets with dense functionality and stereochemistry that are challenging to access using synthetic chemistry. Target **1** and analogues were prepared from biorenewable and easily accessible amino polyol starting materials through an oxidation – cyclisation – reduction sequence that could all be performed in one pot [8]. Key to the success is finding suitable specific enzymes through protein engineering, directed evolution or metagenomics database analysis. Here, the first step was catalyzed by a mutant of galactose oxidase (F2) that had been engineered through directed evolution to accept a broad range of alcohol substrates [9, 10]. The second enzyme (pRed14) was identified from metagenomic database analysis. Interestingly, pRed14 had been annotated in protein databases as an alcohol dehydrogenase of the shikimic acid pathway, demonstrating how non-natural promiscuous activity can be used successfully in biocatalysis.

The sequence leading to target amino diol **2** (Figure 2) is an example of carbon-carbon bond forming reactions (such as aldol reactions) alongside functional group interconversions in biocatalytic cascades. Overall, the reaction represents a highly stereoselective three component synthesis from easily available prochiral starting materials in two steps [11] using biocatalysis mediated by aldolase FSA AS followed by imine reductase IR-259. The reversibility of the first aldol reaction and cross-reactivity of the carbonyl substrates prohibited one-pot reaction. Such issues of cross-reactivity can be overcome by using flow biocatalysis [12].

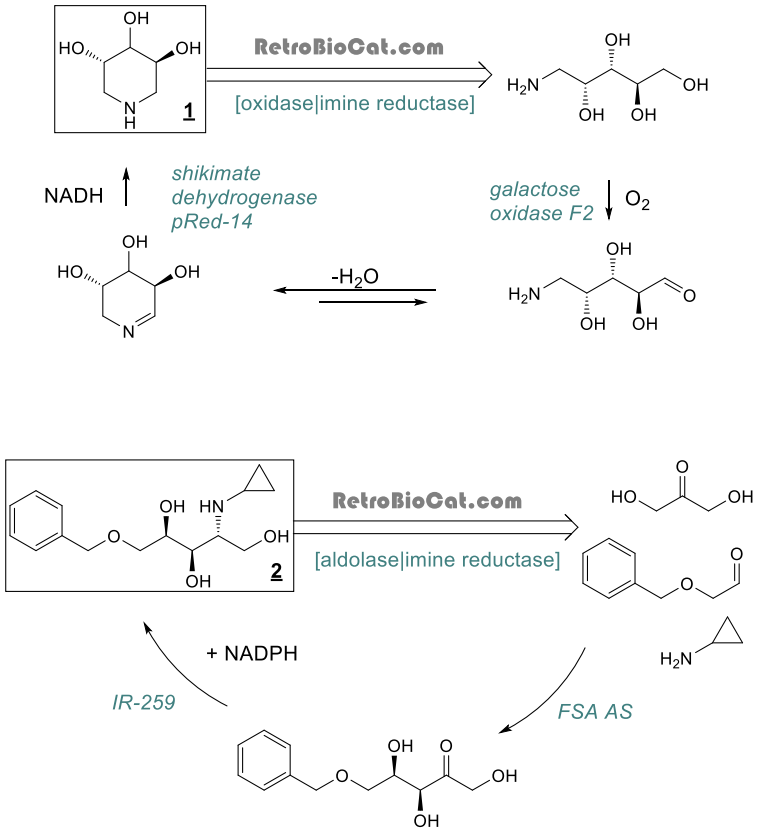

Fig. 2. Two recent examples of enzyme cascades designed using RetroBioCat

Outlook to future developments of research on biocatalysis

Many chemicals are synthesized through multistep processes using a diversity of reactions and biocatalysis will need to be equally diverse to make a broader impact as a green and sustainable alternative to chemical processes. There have been numerous examples of successful enzyme cascades, but the challenge remains to increase the 'reaction rules', the classes of transformations that enzymes can catalyse. For example, more biocatalysts are needed for C-C bond formations, C-H activations, halogenations and even catalytic amid bond formation, and finding these new activities and incorporating them into enzyme cascade are very active and exciting areas of research. There are several approaches that are being pursued at the moment – including studying biosynthesis of secondary metabolites to find new enzymatic reactions, looking for promiscuity to find non-natural enzymatic activity, protein engineering, directed evolution and *de novo* protein design. Given that most biocatalysts are proteins with common design, production methods and reaction conditions, biocatalysis lends itself to

the use of computational and automated tools that are increasingly able to deal with the numbers and complexity of protein sequences and structures. Protein structure prediction has been a bottleneck in biology for a long time and has recently made a large step forward through machine learning, although prediction of function at a precise molecular level remains challenging. The scientist interested in finding new enzyme activity has very impressive computational and experimental toolkits available that can generate and deal with very large datasets. However, what makes the subject particularly interesting intellectually, is the need for creativity in developing original mechanistic hypotheses that can then be explored through using these toolkits.

Acknowledgments

We kindly acknowledge financial support from the European Research Council 788231-ProgrES-ERC-2017-ADG.

References

1. E. L. Bell, W. Finnigan, S. P. France, A. P. Green, M. A. Hayes *et al., Nature Reviews Methods Primers* **1**, 46 (2021).
2. J. H. Schrittweiser, S. Velikogne, M. Hall, W. Kroutil, *Chem. Rev.* **118**, 270 (2018)
3. A. I. Benitez-Mateos, D. Roura Padrosa, F. Paradisi, *Nature Chemistry* **14**, 489 (2022).
4. S. P. France, L. J. Hepworth, N. J. Turner, S. L. Flitsch, *ACS Catalysis* **7***,* 710 (2017).
5. E. Romero, B. S. Jones, B. N. Hogg, A. R. Casamajo, M. A. Hayes *et al. Angew. Chem. Int. Ed.* **60**, 16824 (2021).
6. F. H. Arnold, *Angew. Chem. Int. Ed.* **58**, 14420 (2019).
7. W. Finnigan, L. P. Hepworth, S. L. Flitsch, N. J. Turner *Nature Catalysis* **4**, 98 (2021).
8. C. Swanson, G. J. Ford, A. P. Mattey, L. Gourbeyre *et al* ChemRxiv. (2022).
9. L. Sun, I. P. Petrounia, M. Yagasaki, G. Bandara, F. H. Arnold, *Protein Eng.* **14**, 699 (2001).
10. J. B. Rannes, A. Ioannou, S. C. Willies, G. Grogan, C. Behrens *et al.,J. Am. Chem. Soc.***133**, 8436 (2011)
11. G. J. Ford, C. R. Baldwin, R. T. Bradshaw Allen, J. R. Marshall, A. P. Mattey, *et al., J. Am. Chem. Soc. AU* (2022) .
12. A. P. Mattey G. J. Ford, J. Citoler, C. Baldwin, J. R. Marshall *et al. Angew. Chem. Int. Ed.* **60**, 18660 (2021).