Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Sci.*, 2023, **14**, 6467

All publication charges for this article have been paid for by the Royal Society of Chemistry

Received 14th March 2023
Accepted 17th May 2023

DOI: 10.1039/d3sc01355c

rsc.li/chemical-science

Computer-assisted multistep chemoenzymatic retrosynthesis using a chemical synthesis planner†

Karthik Sankaranarayanan and Klavs F. Jensen *

Chemoenzymatic synthesis methods use organic and enzyme chemistry to synthesize a desired small molecule. Complementing organic synthesis with enzyme-catalyzed selective transformations under mild conditions enables more sustainable and synthetically efficient chemical manufacturing. Here, we present a multistep retrosynthesis search algorithm to facilitate chemoenzymatic synthesis of pharmaceutical compounds, specialty chemicals, commodity chemicals, and monomers. First, we employ the synthesis planner ASKCOS to plan multistep syntheses starting from commercially available materials. Then, we identify transformations that can be catalyzed by enzymes using a small database of biocatalytic reaction rules previously curated for RetroBioCat, a computer-aided synthesis planning tool for biocatalytic cascades. Enzymatic suggestions captured by the approach include ones capable of reducing the number of synthetic steps. We successfully plan chemoenzymatic routes for active pharmaceutical ingredients or their intermediates (*e.g.*, Sitagliptin, Rivastigmine, and Ephedrine), commodity chemicals (*e.g.*, acrylamide and glycolic acid), and specialty chemicals (*e.g.*, *S*-Metalochlor and Vanillin), in a retrospective fashion. In addition to recovering published routes, the algorithm proposes many sensible alternative pathways. Our approach provides a chemoenzymatic synthesis planning strategy by identifying synthetic transformations that could be candidates for enzyme catalysis.

1 Introduction

Enzymes are essential tools employed in synthetic and process chemistry. They catalyze stereo-, regio-, and enantio-specific reactions; as a result, biocatalysis enables more efficient synthetic routes with less need for protection and deprotection reactions.[1] Directed evolution has successfully enabled drastic changes to the substrate scope of enzymes to accept unnatural substrates to meet process needs.[2] Further, enzymatic reactions typically occur in water or benign organic solvents under mild conditions.[3–5] Moreover, multistep enzymatic reactions can be carried out in a single pot to avoid purifying intermediates and overcome equilibrium constraints.[1,3] Enzymes are also more amenable to economic modeling than precious metal catalysts, whose prices fluctuate.[6]

Organic synthesis is the workhorse of the modern chemical manufacturing industry. It has been refined over centuries to allow efficient routes to creating human-made compounds. A subset of the organic reactions used can also be catalyzed by enzymes, for example, to reduce cost or employ milder reaction conditions. However, organic chemists' educational experience has only a small overlap with that of enzymologists, making it challenging for some synthetic chemists to tap into benefits at the interface between organic- and enzyme chemistry. Chemoenzymatic synthesis planning tools can efficiently help identify biocatalysis opportunities in the manufacture of small molecules.

Retrosynthesis techniques in computer-aided synthesis planning (CASP) propose feasible multistep synthetic routes to a target from available starting materials by starting with the target and choosing appropriate disconnections recursively. From the early CASP tools in organic chemistry presented over 50 years ago,[7,8] the methods have improved to predict realistic organic synthesis routes to a desired target using rule-based methods and machine learning to generalize known reactions.[9–11] Recent developments in enzymatic retrosynthesis show a tremendous potential for developing similar CASP tools for enzymes.[12–15] Finnigan *et al.* recently curated a small set of expertly encoded reaction rules to describe the enzyme toolbox for biocatalysis.[13] These reaction rules implicitly reflect the established substrate promiscuity of the different enzyme classes. Enzymes that these rules represent have been shown to be amenable to enzyme engineering in many cases, for the acceptance of novel substrates.[13] Furthermore, they were successfully employed to plan biocatalytic cascades to target molecules. Although RetroBioCat successfully plans multistep enzymatic routes, it cannot propose chemoenzymatic routes toward a desired target that synergistically involves both organic and enzymatic approaches.

Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. E-mail: kfjensen@mit.edu

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3sc01355c

Recently, Levin *et al.* created a chemoenzymatic synthesis planning tool that identifies enzymatic opportunities while performing retrosynthesis.[14] This method employs the 'BKMS-react' database of enzymatic reactions mainly used for natural product biosynthesis to identify enzymatic opportunities. Limitations in this dataset result in two imperfections. First, it reduces the algorithm's capability to identify chemoenzymatic routes toward human-made molecules that are not similar to natural products. Specifically, the algorithm only identified a limited number of synthesis routes and did not capture the precedent chemoenzymatic routes for several model human-made compounds tested (Fig. S1†). However, this molecule class plays a significant role in pharmaceutical, commodity chemical, and agrochemical industries. Second, BKMS database is biased towards metabolic enzymes; this might lead to retrosynthetic suggestions with reduced chances of engineerability of the enzymes. Therefore, notwithstanding this impressive advance in chemoenzymatic synthesis planning, there remains a strong need for planning algorithms to identify enzymatic opportunities within synthetic routes toward human-made molecules.

Herein, we demonstrate a method facilitating chemoenzymatic synthesis by identifying enzymatic transformations within pathways proposed by the open-source CASP tool, ASKCOS.[4] The procedure is illustrated in Fig. 1 using Duloxetine[16] as an example target. First, a user specifies the target compound, duloxetine, by its SMILES string. An existing CASP algorithm proposes many possible synthetic routes from commercial starting materials to the target. We use ASKCOS for this step, but in principle, other CASP approaches would work. Second, a small database of biocatalytic reaction rules previously curated by Finnigan *et al.* enables the identification of synthetic reactions that are candidates for enzyme catalysis, including the possibility of shortened reaction pathways.

By identifying biocatalysis opportunities in planned synthesis routes, we demonstrate the capability of our approach to propose chemoenzymatic steps. Enzymatic suggestions captured by our approach include ones capable of reducing the number of synthetic steps. We successfully plan chemoenzymatic routes for active pharmaceutical ingredients or their intermediates (*e.g.*, Sitagliptin, Rivastigmine, and Ephedrine), commodity chemicals

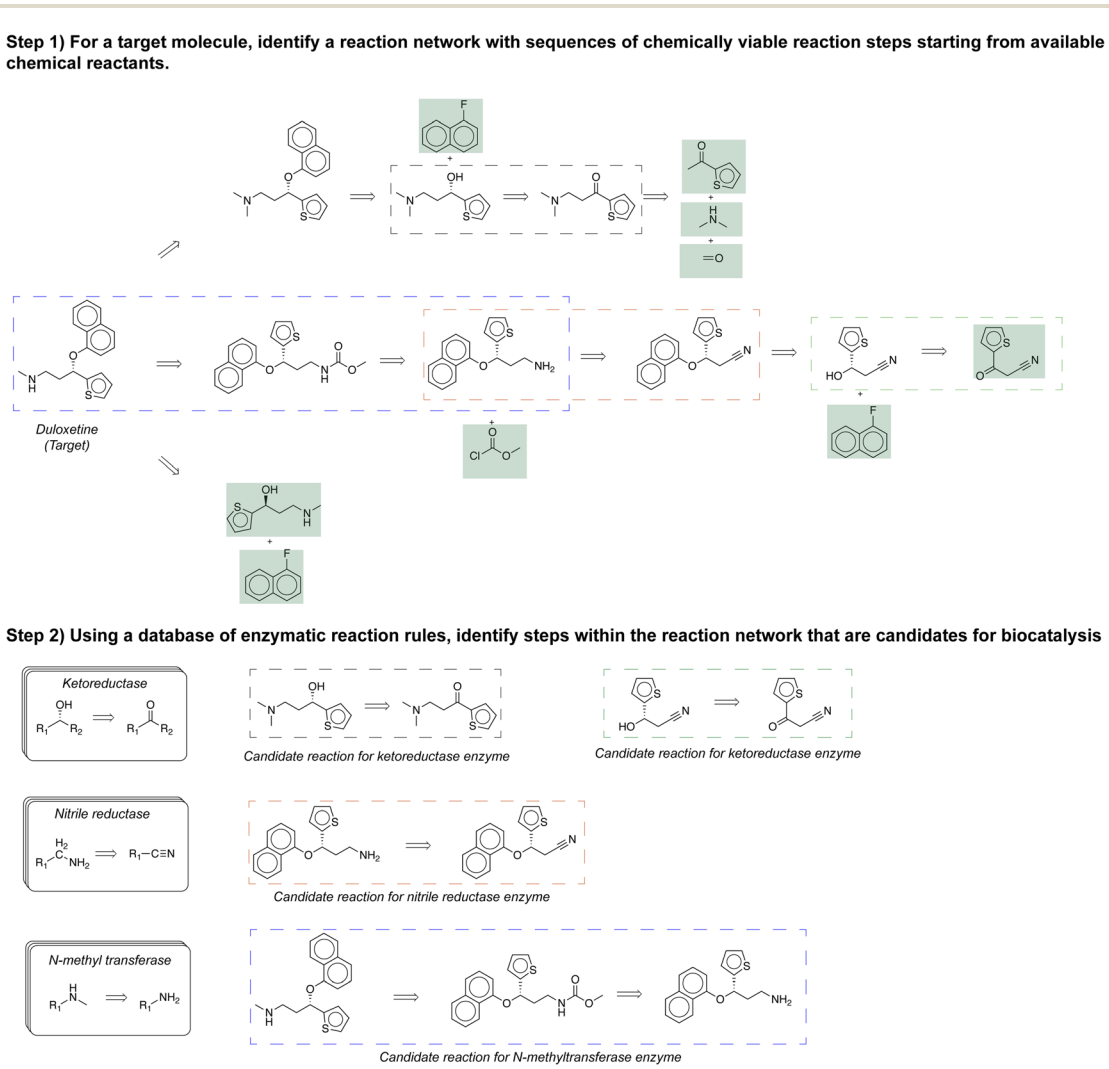

Fig. 1 Computer-assisted multistep chemoenzymatic retrosynthesis using a synthesis planner. Commercially available starting materials are highlighted in a green box.

(*e.g.*, acrylamide and glycolic acid), and specialty chemicals (*e.g.*, *S*-Metalochlor and Vanillin), in a retrospective fashion. In addition to recovering published routes, the algorithm proposes sensible alternatives. A primary disadvantage of employing enzymes is the long lead time necessary for directed evolution campaigns.[6] Our approach often proposes suitable chemocatalytic alternatives to the enzymes to serve as placeholders during evolution campaigns, alleviating a practical bottleneck for applying enzymes in process chemistry.

2 Methods

2.1. Single-step biocatalysis opportunity identifier

Our single-step biocatalysis opportunity identifier asks the question: can an enzyme be utilized to catalyze the single-step transformation proposed by the synthesis planner? First, for a specific target molecule, the algorithm generates a diverse set of chemically viable single-step synthetic transformations that could produce the target. Second, the algorithm identifies candidate reactions that can be catalyzed by enzymes. This analysis treats biocatalysis opportunity identification as a catalyst prediction problem after performing retrosynthesis. The following paragraphs explain the workflow in detail.

First, we employ the single-step retrosynthesis model in ASKCOS to generate a set of reactants that could produce the target.[9] Using information from 12.5 million single-step synthetic reactions tabulated in the Reaxys database, ASKCOS has approximately 160 000 generalized reaction rules summarizing retrosynthetic transformations commonly used in synthetic chemistry. A feedforward neural network predicts and ranks which of the 160 000 reaction rules most apply to a target based on its molecular structure. The top-1000 templates or top-N templates with a maximum cumulative predicted score of 0.999 (whichever set contains a lower number of templates) were considered for a typical analysis that could produce the target.

Second, for every retrosynthetic suggestion, we exhaustively apply the biocatalytic templates from RetroBioCat to determine whether enzymes could potentially catalyze the reaction. RetroBioCat's 135 biocatalytic templates summarize the transformations catalyzed by different classes of enzymes and implicitly describe their established substrate scopes.[13] Many templates are promiscuous with respect to stereochemical preferences of enzymes; this is intended to capture the potential to engineer enzymes for meeting process specific stereochemical needs.[17–20] First, the product SMILES string and a list of separated reactant SMILES strings are obtained for every ASKCOS proposed retrosynthetic suggestion. Then, biocatalytic templates with reaction rules that apply to the target product are applied to the product SMILES using RDChiral, an open-source tool for retrosynthetic template application.[21] For every applied template, the algorithm subsequently checks whether the resulting enzymatic reactants are present within the list of reactants associated with the original ASKCOS proposed retrosynthetic suggestion using an exact SMILES string match. If true, the enzyme class associated with the biocatalytic template is recorded as applicable to the ASKCOS proposed retrosynthetic suggestion. By ensuring the biocatalytic template applies to the proposed retrosynthetic suggestion, this approach intends to propose enzymes whose active site can catalyze the proposed synthetic transformation and whose binding pockets can accommodate the substrate through protein engineering.

2.2. Multistep chemoenzymatic retrosynthesis

Complex target molecules often require a series of organic and enzymatic transformations, which sequentially build molecular complexity starting from simpler building blocks. Herein, we developed two complementary methods for performing multistep chemoenzymatic retrosynthesis. First, we recursively apply the combined single-step retrosynthesis model and biocatalysis opportunity identifier to plan multistep chemoenzymatic synthesis routes. This first workflow simply extends the single-step biocatalysis opportunity identifier described in Section 2.1. Second, we developed an automated computer-aided synthesis planning tool that plans multistep chemoenzymatic reaction pathways in two sequential steps. Given a target compound as input, ASKCOS proposes multistep synthetic pathways starting from commercial starting materials. Then, we identify candidate reactions within ASKCOS proposed routes well suited for biocatalysis. The following paragraphs detail the second workflow.

First, we employ the MCTS-based multistep chemical synthesis planning tool in ASKCOS to generate a reaction network comprising different pathways to synthesize the target starting from commercially available materials.[9] Retrosynthetic expansion occurs recursively for a predetermined amount of time (typically about 2 minutes) or up to a specified depth (typically about 5 steps) before all resulting pathways are returned. ASKCOS uses the Monte Carlo tree search algorithm to balance exploitation of branches thought to be promising and exploration of less frequently visited branches. Our database of buyables contains commercially available compounds from eMolecules, LabNetwork, or Sigma Aldrich. This buyables database can be customized for individual user needs.[22] In this study, all compounds with an average price per gram listed at $100 or lower were included in the database. This analysis yielded a reaction network with different routes to produce the target staring from commercially available materials.

Second, we developed a method to process the resulting reaction network for identifying opportunities to employ biocatalysts. Because of the chemo-, regio-, and enantio-selectivity of enzyme-catalyzed reactions, enzymes offer opportunities to reduce the overall number of steps in the synthesis. This reduction was encouraged by allowing the algorithm to identify multistep synthetic transformations that a single enzymatic reaction can replace. Since a typical reaction network comprises O(10)–O(100) possible pathways to produce a target from commercial materials, the algorithm first parses the reaction network to identify the individual routes before each one is subsequently further analyzed. For every molecule in the pathway, the algorithm identifies single-step and multistep synthetic transformations to produce the molecule using

precursors in the route. Then, the algorithm uses the single-step biocatalysis opportunity identifier described previously to determine whether the molecule can be created using the initial reactants of the identified single-step or multistep synthetic transformations. Importantly, this technique is a 'post-processing method' that uses pathways proposed by a synthesis planning tool; therefore, it can be integrated into other open-source or proprietary retrosynthetic workflows efficiently to identify biocatalysis opportunities.

3 Results

3.1. Single-step biocatalysis opportunity identifier

We tested our algorithm on diverse model compounds or intermediates produced using enzymes in different chemical manufacturing industries, and it successfully recovered their recorded precursors. *Tert*-butyl (*R*)-3-hydroxyl-5-hexenoate,[23] Esomeprazole,[24] (*S*)-salsolidine,[25] L-(*S*)-*tert*-leucine,[26] and Clopidogrel intermediate[27] are pharmaceutical intermediates. Vanillin,[28] *trans*-2-hexen-1-al,[29] and decanal[30] are flavor/food compounds. Acrylonitrile is a commodity chemical.[31] These compounds employ a range of biocatalysts for their production, including ketoreductase (Fig. 2A),[23] carboxylic acid reductase (Fig. 2B),[28] transaminase (Fig. 2C),[26] prazole sulfide monooxygenase (Fig. 2D),[24] alcohol oxidase (Fig. 2E),[29] ene reductase (Fig. 2F),[30] imine reductase (Fig. 2G),[25] hydroxynitrile lyase (Fig. 2H),[27] and nitrile hydratase (Fig. 2I).[31] The approach correctly proposes the recorded reactants and identifies suitable enzymes in all cases. The diversity of these transformations highlights the power of using the collective knowledge contained in a synthetic retrosynthesis prediction tool (*i.e.*, ASKCOS single-step predictor) and a biocatalysis template database to identify strategic retrosynthetic steps and enzyme catalysts that might otherwise be overlooked, particularly by someone less familiar with enzymatic synthesis.

3.2. Interactive multistep chemoenzymatic pathway planning

We then sought to recursively apply the combined single-step retrosynthesis model and biocatalysis opportunity identifier to plan multistep chemoenzymatic synthesis routes for industrially relevant small molecule targets starting from commercial materials. The complex pharmaceutical target Montelukast (Fig. 3A),[32] chiral pharmacophore of Atorvastatin (Fig. 3B),[33] dipeptide Ala–Gln (Fig. 3C),[34] and commodity chemical glycolic acid (Fig. 3D)[35] served as model targets. This interactive pathway planning approach first employs the newly developed single-step chemoenzymatic retrosynthesis approach for brainstorming synthesis strategies for every step and subsequently relies on human expertise to guide the overall multistep chemoenzymatic synthesis plan towards commercial starting materials. The suggested disconnections are consistent with published paths, except for one slight difference. The published biocatalytic route to the side chain of Atorvastatin by Codexis starts with ethyl 4-chloroacetoacetate rather than methyl 4-chloroacetoacetate;[33] however, the proposed route (Fig. 3B) is sensible and captures the overall synthesis strategy.

3.3. Computer-aided chemoenzymatic synthesis planning using MCTS

Chemically diverse compounds produced in multiple steps using chemocatalysts and enzymes served as model compounds to test the algorithm. Sitagliptin (Fig. 4A),[2,36]

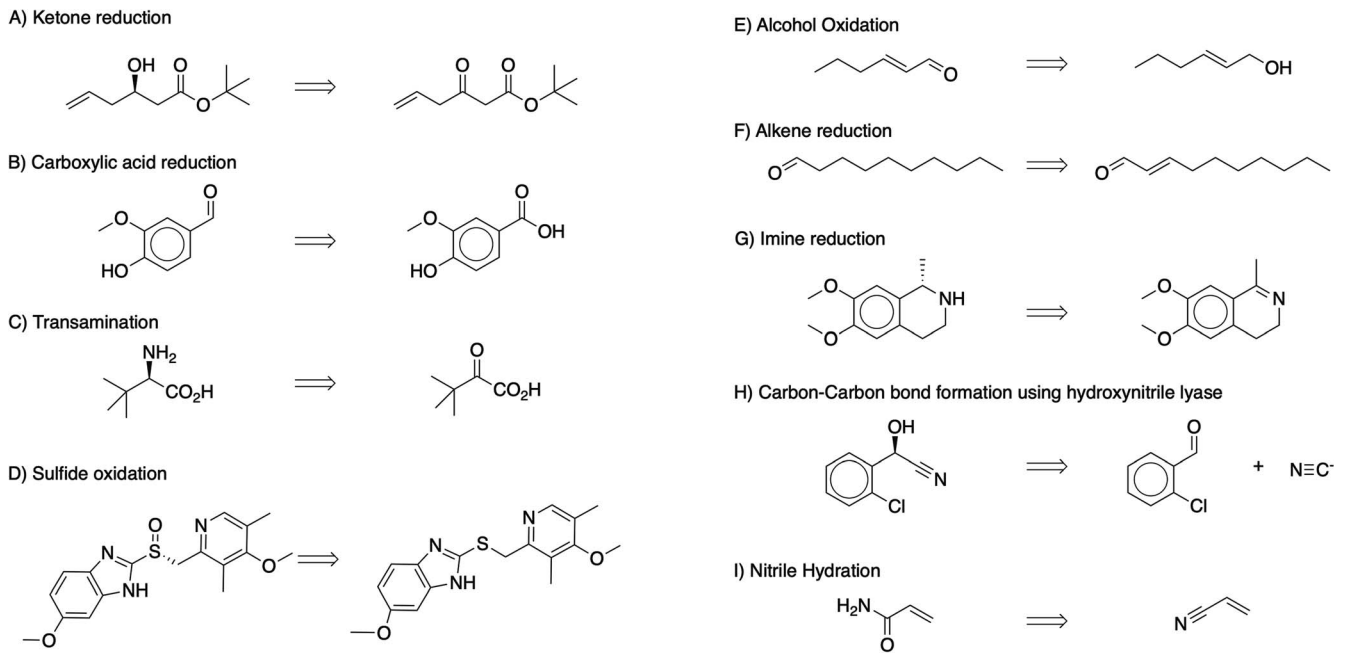

Fig. 2 Example biocatalysis opportunities identified by combining ASCKOS single-step retrosynthesis model and the newly developed biocatalysis opportunity identifier.

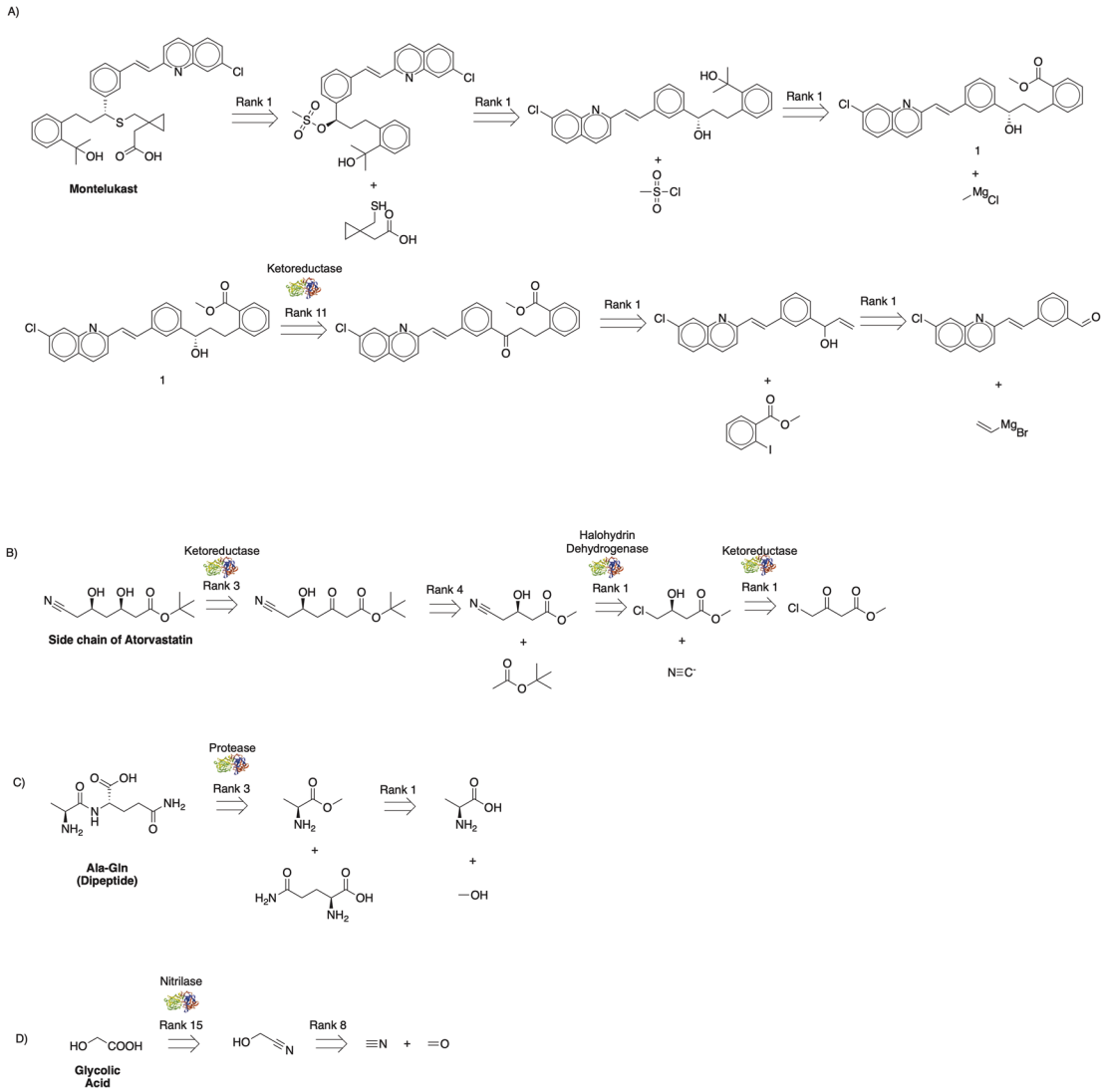

Fig. 3 Multistep chemoenzymatic synthesis planning with human intervention. Routes are constructed by recursively applying the single-step chemoenzymatic synthesis methodology to (a) Montelukast, (b) the key side chain of Atorvastatin, (c) dipeptide Ala–Gln, and (d) commodity chemical glycolic acid. The suggested disconnections are consistent with published pathways, with slight differences discussed in the text.

Rivastigmine (Fig. 4B),[37] a chiral intermediate for angiotensin 1 converting enzyme inhibitor (Fig. 4C),[31] Ephedrine (Fig. 4D),[38] (*R*)-*o*-chloromandelic acid (Fig. 4E),[31] Rasagiline (Fig. 4F),[39] (*S*)-Metalochlor (Fig. 4G)[40] served as model compounds in the pharmaceutical and agrochemical industry. The algorithm capitalized on the ability of enzymes to catalyze stereoselective transformations to introduce chirality into the molecules. In all cases, we recovered the literature pathways successfully. We removed all intermediates in the literature pathways from our buyable database to prevent premature termination at expensive starting materials. In practice, users can recursively run the algorithm if proposed routes start from costly but commercially available starting materials.

In addition to recovering the known literature precedent, the algorithm suggests other sensible chemoenzymatic routes to produce the target. (*S*)-duloxetine is a versatile pharmaceutical in treating major depressive disorders, stress urinary incontinence, and obsessive-compulsive disorder.[16] In a few minutes, the algorithm identified 260 different pathways starting from commercially available materials that could produce this compound. 195 out of 260 pathways were identified as candidates for using enzymes as catalysts (Fig. 5A). To begin with, the known literature precedent to produce this compound was successfully identified by the algorithm (Fig. 5B).[16] Second, the algorithm identifies many alternative routes for producing this molecule, and one such pathway is presented (Fig. 5C). This pathway starts with commercially available starting material 2-theonylacetonitrile (**3**). In the first synthetic step, a ketoreductase converts the ketone (**3**) to the chiral alcohol (**4**). Then, a nucleophilic aromatic substitution reaction transforms 1-fluoronaphthalene (**4**) and (**5**) into (**6**). Third, a nitrile reductase converts the nitrile (**6**) into the amine (**7**). Finally, an *N*-methyltransferase methylates the amine (**7**) to produce the target (**8**). A commercial process could potentially

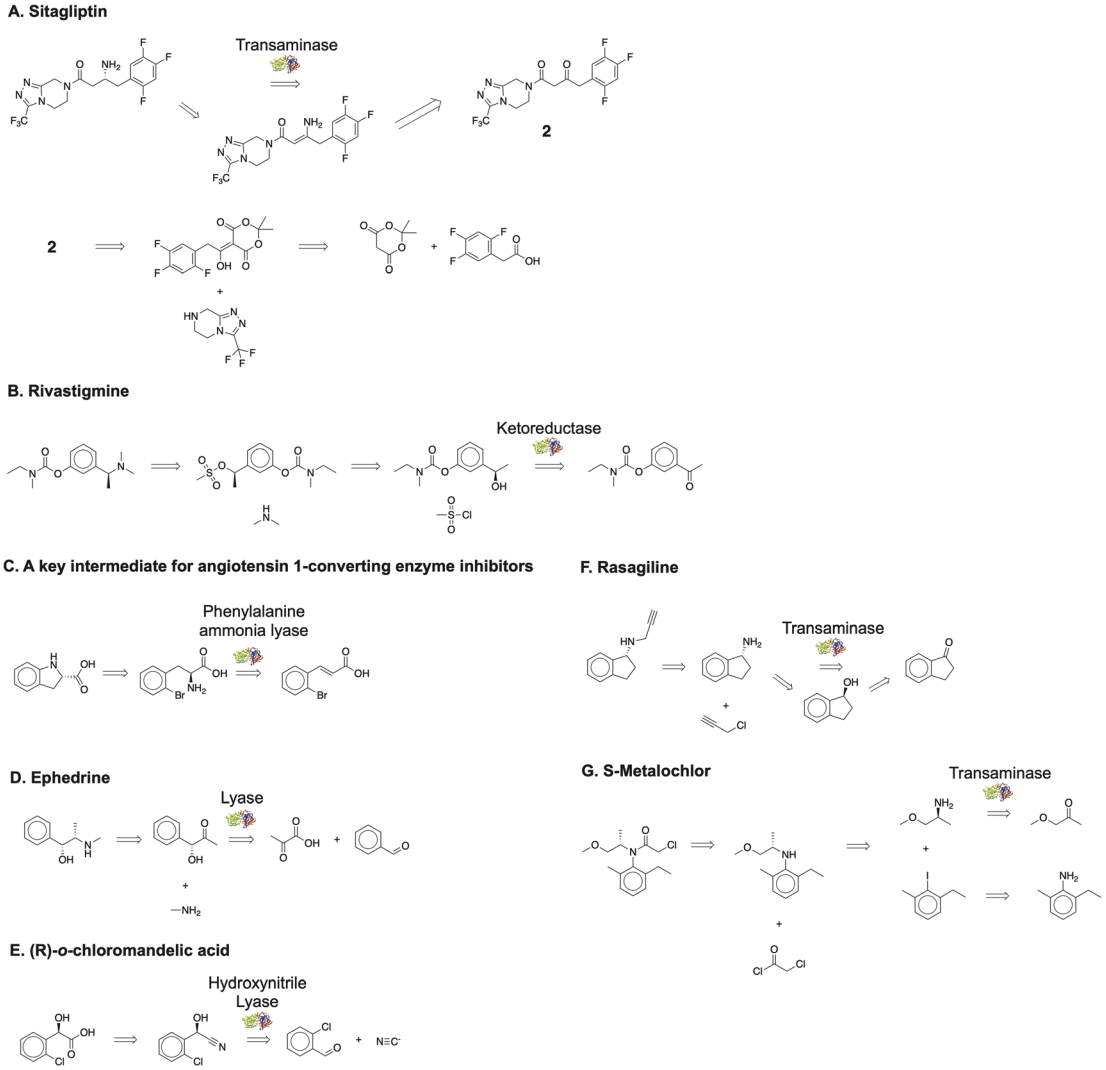

Fig. 4 Multistep chemoenzymatic synthesis planning with minimal human intervention. Using our algorithm, we plan chemoenzymatic synthetic routes for (A) Sitagliptin (B) Rivastigmine (C) a key intermediate for angiotensin 1-converting enzyme inhibitors (D) Ephedrine (E) (*R*)-*o*-chloromandelic acid, (F) Rasagiline, and (G) *S*-Metalochlor. The suggested disconnections are consistent with published pathways.

develop a biocatalytic cascade around the last two transformations; these reactions could be run in a single pot to avoid purifying the intermediate and reduce the overall number of steps in the process. Compared to the published route (Fig. 5B), this alternative pathway (Fig. 5C) is predominantly catalyzed by enzymes and reduces the overall number of steps. Similarly, the algorithm proposes additional chemoenzymatic alternatives for producing duloxetine. Ultimately, the final choice of production process comes down to many factors, including costs, impurity profiles, kinetics, and available equipment resources.

4 Discussion

In this study, we have developed a tool for identifying opportunities to employ biocatalysis in organic syntheses toward synthetic human-made molecules. Chemists tasked with producing a new molecule can utilize this platform as a brainstorming tool. In a few minutes, this planning tool identifies O(10)–O(100) chemoenzymatic synthesis pathways to the target starting from commercial materials. Enzymes are successfully proposed to build the necessary stereochemical and molecular complexity associated with the target. Further, this synthesis planning tool identifies multistep synthetic reactions that could potentially be replaced by a single enzymatic step because of the selectivity of enzyme-catalyzed transformations.

Two complementary multistep chemoenzymatic retrosynthesis tools were developed in this study. The interactive multistep chemoenzymatic pathway planning tool leverages human expertise to guide the synthesis planner. It allows human experts to control the biocatalytic transformations employed in the synthesis. As a result, these human experts can choose to prioritize biocatalytic transformations that their organizations have significantly invested in previously. On the other hand, computer-aided chemoenzymatic synthesis

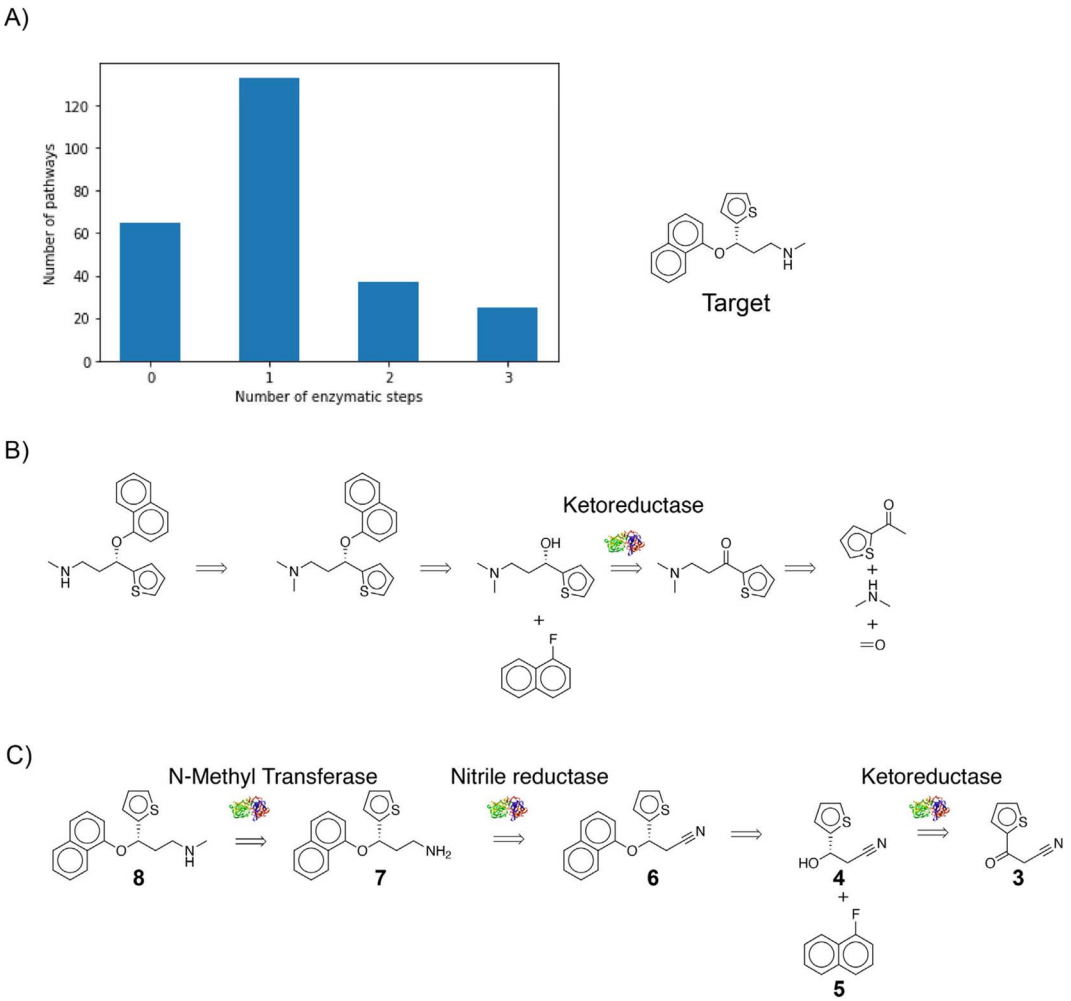

Fig. 5 (A) Chemoenzymatic synthesis planning for the exemplary pharmaceutical compound duloxetine yields 260 total routes that start from commercial materials, including 195 routes that employ at least one enzyme. (B) The published pathway to duloxetine is recovered by the approach. (C) A chemically sensible alternative route to duloxetine proposed by the approach. In this alternative route, three out of four proposed reactions are enzyme-catalyzed; therefore, enzymes play a dominant role in producing the target.

planning using MCTS automatically considers the use of 135 commonly employed biocatalytic transformations. This second tool can algorithmically generate many different chemoenzymatic routes to the target with minimal human intervention. Further, this task can easily be parallelized to compute chemoenzymatic synthesis routes to O(100)–O(1000) different targets. Therefore, this CASP platform can be employed for long-term investment planning to identify critical biocatalytic transformations needed to produce a portfolio of small molecules. As a result, strategic biocatalysis investments can be made to promote the use of enzymes in the manufacture of human-made small molecules.

The presented chemoenzymatic algorithms are open-source and widely available to practitioners. Identifying biocatalysis opportunities can promote collaborations among organic chemists, chemical and protein engineers, and enzymologists to facilitate the sustainable manufacture of small molecules. Furthermore, it could complement many textbooks and review articles on this topic,[31,39,41–43] with an emphasis on solving real-world synthesis challenges important to individual practitioners.

To our knowledge, there is currently only one other approach in the literature that synergistically employs both organic and enzymatic transformations for planning multistep synthesis of complex target compounds. Levin *et al.* has introduced enzymatic transformations during the retrosynthetic analysis using reaction templates associated with natural product biosynthesis.[14] In our method, we focus on identifying synthetic transformations that can potentially be catalyzed by enzymes, which complements the accomplishment of Levin *et al.* by capturing suggestions towards synthetic, human-made chemicals. We performed a comparison between the two methods for the three model compounds Sitagliptin, (*S*)-Duloxetine, and (*S*)-Metalochlor (Fig. S2†). In this comparison, the primary difference is the technique employed to identify biocatalytic transformations in organic syntheses. In all three case studies, our method exceeds that of Levin *et al.* by successfully recovering the known literature precedents. Further, our approach also proposes a greater number of synthetic routes to the target. We

hypothesize that by limiting the biocatalysis search to a small expert-curated biocatalytic template set after completion of the synthetic tree search, we avoid exploring a large number of algorithmically extracted enzymatic reaction templates commonly used in the biosynthesis of natural products during the tree search. Such templates are less likely to be productive when the target is dissimilar to natural products. In the future, computer-aided synthesis planners integrating both approaches could exploit the advantages of the individual methods.

This algorithm can identify candidate transformations for enzyme catalysis but does not propose specific enzyme sequences for directed evolution efforts. Large biocatalysis reaction databases are not widely available like metabolic reaction datasets[44] or organic reaction datasets.[45] Therefore, chemical similarity-based methods for selecting enzyme sequences for directed evolution[12] are less applicable to biocatalysis opportunities associated with human-made compounds. The expert curated reaction rules from RetroBioCat are associated with enzymes that generally have a track record for showing promiscuous substrate specificity.[13] Therefore, identifying these biocatalysis opportunities can be valuable on their own even if specific enzyme sequences are not proposed by the algorithm; enzyme screening panels can be employed to find the right enzyme for a specific reaction.

This tool relies on the synthesis planner ASKCOS to propose retrosynthetic suggestions.[9] ASKCOS predictions are driven by an underlying database of synthetic transformations from Reaxys, and most of these transformations are organic. At every step in a retrosynthetic analysis, beneficial enzymatic reaction templates that do not overlap with ASKCOS predicted templates are not considered. Biocatalytic and chemoenzymatic routes that do not align with the same order of chemical transformations as classical synthetic approaches might not be favored by our tool. On the other hand, suggestions resulting from our tool might have alternative classical synthetic approaches for the enzyme catalyzed steps. Because of the possibility of accomplishing some of the enzyme catalyzed steps through alternative approaches, the inherent risk of pursuing risk-prone and time-intensive biocatalysis opportunities is greatly mitigated. This could encourage risk-averse practitioners in highly regulated industries (*e.g.* pharmaceutical manufacturing) to explore biocatalysis opportunities.

5 Conclusion

We have developed a computer-assisted chemoenzymatic synthesis planner for synthetic human-made molecules. Our approach can capture selective enzymatic transformations that reduce the number of synthetic steps. Our route modification strategy identifies synthetic transformations that could be candidates for enzyme catalysis to plan chemoenzymatic syntheses. This post-processing step could, in principle, be integrated with other open source or proprietary organic synthesis planners to identify opportunities for biocatalysis. Efforts to incorporate beneficial enzymatic reaction templates into the retrosynthetic analysis that don't overlap with synthetic templates already in Reaxys will further improve the synthetic efficiency of the proposed routes. Notwithstanding this limitation, the stage is set for computational chemoenzymatic synthesis planning for human-made compounds.

Data availability

ASKCOS is publicly available at **https://askcos.mit.edu/**. The tool's algorithms are available at **https://github.com/karthiksankar93/ChemoEnzymaticSynthesis**. Additional information on the supporting tables and figures are provided in the ESI.†

Author contributions

KS: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing – original draft, review, and editing. KFJ: conceptualization, writing – review and editing. supervision and funding acquisition.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work is supported by the DARPA PANACEA program grant HR0011-19-2-0022. This work is also supported by the consortium for Machine Learning in Pharmaceutical Discovery and Synthesis (MLPDS). We thank Deeptak Verma, Colin Lam, Willow Ross Carretero Chavez, Andrew Zahrt, Jason Mustakis, and Ania Fryszkowska for helpful discussions.

References

1 M. A. Huffman, A. Fryszkowska, O. Alvizo, M. Borra- Garske, K. R. Campos, K. A. Canada, P. N. Devine, D. Duan, J. H. Forstater, S. T. Grosser, H. M. Halsey, G. J. Hughes, J. Jo, L. A. Joyce, J. N. Kolev, J. Liang, K. M. Maloney, B. F. Mann, N. M. Marshall, M. McLaughlin, J. C. Moore, G. S. Murphy, C. C. Nawrat, J. Nazor, S. Novick, N. R. Patel, A. Rodriguez-Granillo, S. A. Robaire, E. C. Sherer, M. D. Truppo, A. M. Whittaker, D. Verma, L. Xiao, Y. Xu and H. Yang, *Science*, 2019, **366**, 1255–1259.

2 C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, P. N. Devine, G. W. Huisman and G. J. Hughes, *Science*, 2010, **329**, 305–309.

3 K. Sankaranarayanan, X. X. Antaris, B. A. Palanski, A. El Gamal, C. M. Kao, W. L. Fitch, C. R. Fischer and C. Khosla, *J. Am. Chem. Soc.*, 2019, **141**, 9474–9478.

4 B. Lowry, T. Robbins, C.-H. Weng, R. V. O'Brien, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 2013, **135**, 16809–16812.

5 X. Yu, T. Liu, F. Zhu and C. Khosla, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 18643–18648.

6 M. D. Truppo, *ACS Med. Chem. Lett.*, 2017, **8**, 476–480.

7 E. J. Corey and W. T. Wipke, *Science*, 1969, **166**, 178–192.

8 G. É. Vléduts and V. K. Finn, *Inf. Storage Retr.*, 1963, **1**, 101–116.
9 C. W. Coley, D. A. Thomas, J. A. M. Lummiss, J. N. Jaworski, C. P. Breen, V. Schultz, T. Hart, J. S. Fishman, L. Rogers, H. Gao, R. W. Hicklin, P. P. Plehiers, J. Byington, J. S. Piotti, W. H. Green, A. J. Hart, T. F. Jamison and K. F. Jensen, *Science*, 2019, **365**, eaax1566.
10 M. H. S. Segler, M. Preuss and M. P. Waller, *Nature*, 2018, **555**, 604–610.
11 S. Genheden, A. Thakkar, V. Chadimová, J.-L. Reymond, O. Engkvist and E. Bjerrum, *J. Cheminf.*, 2020, **12**, 70.
12 K. Sankaranarayanan, E. Heid, C. W. Coley, D. Verma, W. H. Green and K. F. Jensen, *Chem. Sci.*, 2022, **13**, 6039–6053.
13 W. Finnigan, L. J. Hepworth, S. L. Flitsch and N. J. Turner, *Nat. Catal.*, 2021, **4**, 98–104.
14 I. Levin, M. Liu, C. A. Voigt and C. W. Coley, *Nat. Commun.*, 2022, **13**, 7747.
15 D. Probst, M. Manica, Y. G. Nana Teukam, A. Castrogiovanni, F. Paratore and T. Laino, *Nat. Commun.*, 2022, **13**, 964.
16 X. Chen, Z.-Q. Liu, C.-P. Lin and Y.-G. Zheng, *Bioorg. Chem.*, 2016, **65**, 82–89.
17 O. May, P. T. Nguyen and F. H. Arnold, *Nat. Biotechnol.*, 2000, **18**, 317–320.
18 K.-E. Jaeger and T. Eggert, *Curr. Opin. Biotechnol.*, 2004, **15**, 305–313.
19 M. T. Reetz, *J. Org. Chem.*, 2009, **74**, 5767–5778.
20 M. T. Reetz, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 5716–5722.
21 C. W. Coley, W. H. Green and K. F. Jensen, *J. Chem. Inf. Model.*, 2019, **59**, 2529–2537.
22 *Buyable Compounds*, **https://askcos.mit.edu/buyables/**, accessed 29 April 2023.
23 C. Hu, M. Liu, X. Yue, Z. Huang and F. Chen, *Org. Process Res. Dev.*, 2020, **24**, 1700–1706.
24 N. Xu, J. Zhu, Y.-Q. Wu, Y. Zhang, J.-Y. Xia, Q. Zhao, G.-Q. Lin, H.-L. Yu and J.-H. Xu, *Org. Process Res. Dev.*, 2020, **24**, 1124–1130.
25 F. Leipold, S. Hussain, D. Ghislieri and N. J. Turner, *ChemCatChem*, 2013, **5**, 3505–3508.
26 Y.-M. Seo and H. Yun, *J. Microbiol. Biotechnol.*, 2011, **21**, 1049–1052.
27 A. Glieder, R. Weis, W. Skranc, P. Poechlauer, I. Dreveny, S. Majer, M. Wubbolts, H. Schwab and K. Gruber, *Angew. Chem., Int. Ed.*, 2003, **42**, 4815–4818.
28 J. Park, H.-S. Lee, J. Oh, J. C. Joo and Y. J. Yeon, *Biochem. Eng. J.*, 2020, **161**, 107683.
29 T. P. de Almeida, M. M. C. H. van Schie, A. Ma, F. Tieves, S. H. H. Younes, E. Fernández-Fueyo, I. W. C. E. Arends, A. Riul Jr and F. Hollmann, *Adv. Synth. Catal.*, 2019, **361**, 2668–2672.
30 A. Papadopoulou, C. Peters, S. Borchert, K. Steiner and R. Buller, *Org. Process Res. Dev.*, 2022, **26**, 2102–2110.
31 S. Wu, R. Snajdrova, J. C. Moore, K. Baldenius and U. T. Bornscheuer, *Angew. Chem., Int. Ed.*, 2021, **60**, 88–119.
32 S. Bollikonda, S. Mohanarangam, R. R. Jinna, V. K. K. Kandirelli, L. Makthala, S. Sen, D. A. Chaplin, R. C. Lloyd, T. Mahoney, V. H. Dahanukar, S. Oruganti and M. E. Fox, *J. Org. Chem.*, 2015, **80**, 3891–3901.
33 N. J. Turner and E. O'Reilly, *Nat. Chem. Biol.*, 2013, **9**, 285–288.
34 M. Yagasaki and S. Hashimoto, *Appl. Microbiol. Biotechnol.*, 2008, **81**, 13–22.
35 A. Panova, L. J. Mersinger, Q. Liu, T. Foo, D. C. Roe, W. L. Spillan, A. E. Sigmund, A. Ben-Bassat, L. W. Wagner, D. P. O'Keefe, S. Wu, K. L. Petrillo, M. S. Payne, S. T. Breske, F. G. Gallagher and R. DiCosimo, *Adv. Synth. Catal.*, 2007, **349**, 1462–1474.
36 F. Ye, Z. Zhang, W. Zhao, J. Ding, Y. Wang and X. Dang, *RSC Adv.*, 2021, **11**, 4805–4809.
37 P.-C. Yan, G.-L. Zhu, J.-H. Xie, X.-D. Zhang, Q.-L. Zhou, Y.-Q. Li, W.-H. Shen and D.-Q. Che, *Org. Process Res. Dev.*, 2013, **17**, 307–312.
38 C. K. Winkler, J. H. Schrittwieser and W. Kroutil, *ACS Cent. Sci.*, 2021, **7**, 55–71.
39 N. J. Turner and L. Humphreys, *Biocatalysis in Organic Synthesis: The Retrosynthesis Approach*, Royal Society of Chemistry, 2018.
40 D. A. Jackson, in *Biocatalysis for Green Chemistry and Chemical Process Development*, John Wiley & Sons, Ltd, 2011, pp. 255–276.
41 N. J. Turner and E. O'Reilly, *Nat. Chem. Biol.*, 2013, **9**, 285–288.
42 R. N. Patel, *Bioorg. Med. Chem.*, 2018, **26**, 1252–1274.
43 M. T. Reetz, *J. Am. Chem. Soc.*, 2013, **135**, 12480–12496.
44 P. Bansal, A. Morgat, K. B. Axelsen, V. Muthukrishnan, E. Coudert, L. Aimo, N. Hyka-Nouspikel, E. Gasteiger, A. Kerhornou, T. B. Neto, M. Pozzato, M.-C. Blatter, A. Ignatchenko, N. Redaschi and A. Bridge, *Nucleic Acids Res.*, 2022, **50**, D693–D700.
45 D. M. Lowe, Doctor of Philosophy (PhD thesis), University of Cambridge, 2012.