ChemCatChem

The European Society Journal for Catalysis

Chemistry Europe

European Chemical Societies Publishing

Supported by GECATS GERMAN CATALYSIS SOCIETY

# Accepted Article

**Title:** Multi-faceted set-up of a diverse ketoreductase library enables the synthesis of pharmaceutically-relevant secondary alcohols

**Authors:** Moritz Voss, Robin Küng, Takahiro Hayashi, Magdalena Jonczyk, Michael Niklaus, Hans Iding, Dennis Wetzl, and Rebecca Maria Ursula Buller



**To be cited as:** *ChemCatChem* 10.1002/cctc.202001871

**Link to VoR:** https://doi.org/10.1002/cctc.202001871

WILEY-VCH

10.1002/cctc.202001871

WILEY-VCH

FULL PAPER

# Multi-faceted set-up of a diverse ketoreductase library enables the synthesis of pharmaceutically-relevant secondary alcohols

Moritz Voss,[a] + Robin Küng,[a], [§] + Takahiro Hayashi,[a], [‡] Magdalena Jonczyk,[a] Michael Niklaus,[a] Hans Iding,[b] Dennis Wetzl,[b] and Rebecca Buller*[a]

[a] Dr. M. Voss, R. Küng, Dr. T. Hayashi, M. Jonczyk, Dr. M. Niklaus, Prof. Dr. R. Buller
Competence Center for Biocatalysis
Institute for Chemistry and Biotechnology
Zurich University of Applied Sciences
Einsiedlerstrasse 31, 8820 Wädenswil, Switzerland
E-mail: rebecca.buller@zhaw.ch

[b] Dr. Hans Iding, Dr. Dennis Wetzl
Process Chemistry & Catalysis
F. Hoffmann-La Roche Ltd.
CH-4070 Basel, Switzerland

[§] Present address:
Fisher Clinical Services
Thermo Fisher Scientific
Steinbühlweg 69, 4123 Allschwil, Switzerland

[‡] Present address:
Science & Innovation Center
Mitsubishi Chemical Corporation
1000 Kamoshidacho, Aoba ward, Yokohama,
Kanagawa 227-8502, Japan

+ These authors contributed equally to the work

Supporting information for this article is given via a link at the end of the document.

**Abstract:** Enzymes are valuable tools to introduce chirality into small molecules. Especially, ketoreductase (KRED)-catalyzed transformations of ketones to yield chiral secondary alcohols have become an established biocatalytic process step in the pharmaceutical and fine chemical industry. Development time, however, remains a critical factor in chemical process development and thus, the competitiveness of a biocatalytic reaction step is often governed by the availability of off-the-shelf enzyme libraries. To expand the biocatalytic toolbox with additional ketoreductases, we established a multi-faceted screening procedure to capture KRED diversity from different sources, such as literature, available genome data, and uncharacterized microbial strains. Overall, we built a library consisting of 51 KRED enzymes, 29 of which have never been described in literature before. Notably, 18 of the newly described enzymes exhibited anti-Prelog preference complementing the majority of ketoreductases which generally follow Prelog's rule. Analysis of the library's catalytic activity toward a chemically diverse ketone substrate set of pharmaceutical interest further highlighted the broad substrate scope and the complementing enantio-preference of the individual KREDs. Using the generated sequence-function data of the included short chain dehydrogenases in a bioinformatic analysis led to the identification of possible sequence determinants of the stereospecificity exhibited by these enzymes.

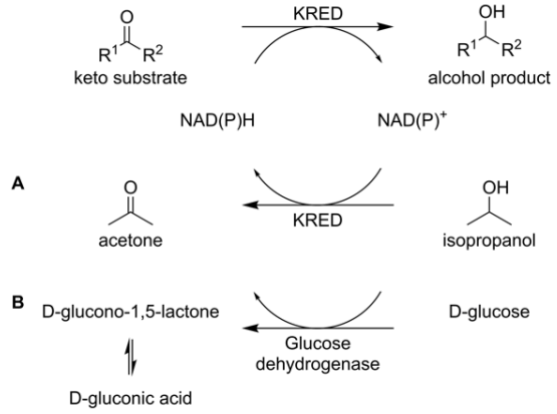
## Introduction

Optical pure alcohols are key chiral intermediates in the synthesis of fine chemicals, agrochemicals, and pharmaceuticals.[1] Unsurprisingly, many synthetic procedures have been developed to access these molecules, predominantly relying on asymmetric hydrogenation (AH) or transfer hydrogenation (ATH) of ketones, or the (dynamic) kinetic resolution of the alcohol racemates employing ruthenium-complexes as catalysts.[2] Especially, the asymmetric reduction of carbonyl compounds is of particular interest as this synthetic strategy allows to valorize all starting material (100 % theoretical yield) inherently allowing for a high atom efficiency. The biocatalytic synthesis of optically pure alcohols via ketoreductases (KREDs) is an appealing alternative to chemical synthesis, owing to the enzyme's intrinsically high regio- and stereoselectivity, the mild reaction conditions, the absence of possibly contaminating metal-based catalysts and the often reduced environmental burden of the overall reaction.[3] KREDs, also named alcohol dehydrogenases (ADH) or carbonyl reductases, are NAD(P)H-dependent enzymes that catalyze the reversible transfer of a hydride from the cofactor NAD(P)H onto the carbonyl function of the substrate (Scheme 1).

Today, KREDs commonly find industrial use in the asymmetric reduction of prochiral ketones (and aldehydes) or, less frequently, in the corresponding oxidation of the alcohols. KRED-catalyzed transformations are industrially used to access a range of (APIs) or chiral synthons thereof,[4] including ibrutinib,[4j] atorvastatin,[4h, 4i] and angiotensin-converting enzyme (ACE) inhibitors.[4k] The industrial importance of KREDs is further underscored by a number of innovative patents describing KRED-catalyzed procedures for preparing valuable compounds,[5] and the overall increasing number of KRED-related patents filed in recent years by academia and industry (Figure 1).

10.1002/cctc.202001871




**Scheme 1.** KRED catalyzed carbonyl reduction via the oxidation of the NAD(P)H cofactor. Regeneration of the NAD(P)H cofactor by (A) using isopropanol as cosubstrate for the KRED or (B) employing glucose dehydrogenase (GDH) in combination with D-glucose as cosubstrate in the reaction.

Notably, when surveying recent patent activity on KRED-facilitated transformations, we found that more than half of KRED-related patents filed in the last ten years had been registered at the Chinese patent register, either by Chinese companies or universities. Possible reasons include China's "National Patent Development Strategy (2011-2020)", which was implemented in 2011 to enhance China's core competitiveness and which led to a massive increase of patents filed.[6]

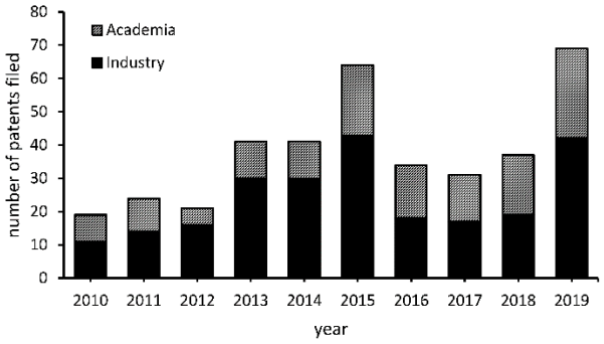



**Figure 1.** Patents filed per year in the period from 2010 – 2019 describing novel KREDs or KRED-catalyzed reactions.

The increasing academic and industrial interest in KREDs engenders the need to discover novel wildtype enzymes in order to build versatile enzyme libraries capable of accepting a wide range of substrates.[4d] Thanks to the growing numbers of genome sequences becoming available in the frame of whole genome shotgun sequencing studies,[7] an increasing amount of protein sequences are identified and annotated,[8] including putative KRED sequences. The experimental verification of these potential KREDs and the exploration of their synthetic potential as biocatalysts enable the implementation of enzymes in routine synthesis route scoutings and expand the available intellectual property space.

The screening of a rich KRED library can serve as the entry point for applying these biocatalysts industrially in the synthesis of valuable chiral entities. Subsequent protein-engineering campaigns to improve the identified starting scaffolds allow to develop new process routes in a relatively short time frame.[9] Notably, an imbalance in enantioselectivity of the applied whole-cell biocatalysts was found in the carbonyl reduction, where most described biocatalysts follow Prelog' rule, typically obtaining the (*S*)-alcohol when the sterically smaller substituent is consistent with lower priority according to CIP nomenclature.[10] Therefore, the availability of diverse and stereo-complementary enzyme libraries is vital for the fast implementation of biocatalytic processes in the production of fine chemicals, agrochemicals or pharmaceutical ingredients and to strengthen biocatalysis as a viable complement to classical chemical synthesis.[11]

With this goal in mind, we embarked on the establishment of a comprehensive KRED toolbox. Here, we describe the set-up and profiling of a diverse KRED library composed of 51 enzymes, including 29 novel KREDs sourced via protein data mining combined with strain screening. Within this project, we focused our screening effort on strains stemming solely from Swiss sources to facilitate industrialization within the framework of the Nagoya protocol. Our KRED library showed promising synthetic potential as highlighted by our capability to selectively produce a diverse set of secondary alcohols, including valuable active pharmaceutical intermediates, with high selectivity. Notably, we newly identified 18 KREDs with anti-Prelog preference and added them to the biocatalytic reduction toolbox. A close bioinformatic inspection of the sequence and activity data of KRED enzymes belonging to the group of short chain dehydrogenases indicated determinants for the observed enantiopreference.

## Results and Discussion

Set-up of the KRED library

At the outset of our project, we targeted to build a KRED library capable of acting on different substrates in order to facilitate early-stage process scouting projects or screenings in the frame of 2nd generation processes. With this in mind, three approaches were followed for the acquisition of appropriate KRED protein sequences. Firstly, we screened available literature and patent data to identify KREDs with a diverse substrate scope. We chose enzymes known to accept substrates ranging, for example, from ethyl-4-chloroacetate[4i] to cyanocyclopentenone[12] and anthrahydroquinone.[13] Secondly, protein data mining was performed to explore the continually growing protein sequence space and identify novel KREDs. Here, structure-sequence alignments in the commercial 3DM database based on the KRED from *L. kefir* (KRED 5)[14] was performed and promising protein sequences were selected. Additionally, protein sequences from the NCBI database fitting the adapted motif KxAxxxGxxxGxG of the common Rossmann-fold motif for short-chain dehydrogenases (SDR) were retrieved,[15] and the sequences with high similarity towards literature described KREDs were chosen. Lastly, a strain screening using the model substrate acetophenone as well as an undisclosed ketone of pharmaceutical interest was performed to identify novel KREDs from bacteria, fungi, and yeasts (33 organisms in total; Table S1). All selected strains, sampled from the *Culture Collection of Switzerland* (CCOS), were Swiss in origin and chosen such as to

10.1002/cctc.202001871




increase the likelihood to complement our library with completely novel KRED sequences. To do so, the following criteria were applied in the selection of the strains: 1) microorganism not yet sequenced; 2) sequence of microorganism available but not annotated; 3) sequence of organism available and annotated but included KREDs not published or patented; 4) sequence of microorganism available, annotated, and included KREDs commercialized. In the screening, the two bacteria *Weissella hellenica* and *Pseudomonas tolaasii*, and the fungus *Aspergillus oryzae* showed interesting KRED-activity. Thus, the deposited genomes (NCBI database) of these three organisms were analyzed to identify putative KRED sequences and ranked according to similarity with literature described KREDs. Overall, we incorporated 13 soluble and active KRED enzymes from these three strains in our KRED library.

All genes encoding the KRED protein sequences acquired through our above-delineated sourcing strategy were ordered codon-optimized for the expression in *E. coli* and cloned in the pET22b+ vector for experimental characterization. Overall, our final KRED library consisted of 51 active and soluble expressed enzymes, 22 of which had been described in literature and patents, while 29 KREDs represent to our knowledge novel enzymes without prior mentioning in literature (Table S2).

Activity screening of the KRED library

For the characterization of the KRED substrate spectrum and as an assessment of the enzyme's synthetic potential, a wide set of ketone substrates were selected, including aromatic and cyclic ketones, and α- and β-ketoesters. All members of the enzyme library were recombinantly expressed in *E. coli* BL21(DE3) in 96-deep well plate format, and the crude-cell lysates were applied in biocatalysis reactions along with a NAD(P)H recycling-system, consisting of D-glucose and glucose dehydrogenase (GDH). The enzymes were screened towards the reduction of acetophenone (**1a**), tetralone (**2a**), benzylacetone (**3a**), ethyl-2-oxo-4-phenylbutyrate (**4a**), 1-boc-3-piperidone (**5a**), and ethyl 4-chloro-acetoacetate (**6a**).

Using gas- or liquid-chromatography coupled to mass spectrometry analysis (GC-MS or LC-MS, respectively), we analyzed the biotransformation with respect to activity, enantiopreference and enantioselectivity (Figure 2 and Table S4). The absolute configuration of the product alcohols was assigned through a comparison of their retention times with those of commercial product standards in the chiral LC- or GC- methods. With the single exception of product (*R*)-tetralol (**2b**), the KRED library was shown to harbor active and selective enzymes for the enantiopure synthesis of all investigated alcohols. Remarkably, the synthesis of all desired chiral alcohols was also feasible solely by the action of the 29 novel KREDs, highlighting the broad synthetic potential of these previously undescribed biocatalysts. Notably, the alcohol products of the three ketones **4a**-**6a** represent pharmaceutical relevant synthons.[4h-k] The KRED-catalyzed asymmetric reduction of the α-ketoester **4a** to (*R*)-ethyl 2-hydroxy-4-phenylbutanoate (**4b**), for example, is literature-known,[16] since (*R*)-**4b** is pharmaceutically relevant for synthesizing several ACE-inhibitors.[4k] KREDs exhibiting this industrially desirable enantiospecificity were also identified in our KRED library, with KRED 23,[17] 30,[18] 31,[19] 47,[20] as well as the novel KREDs from *P. tolaasii* 40 and 50 being able to catalyze the selective reduction of **4a**. Remarkably, the novel KRED 50

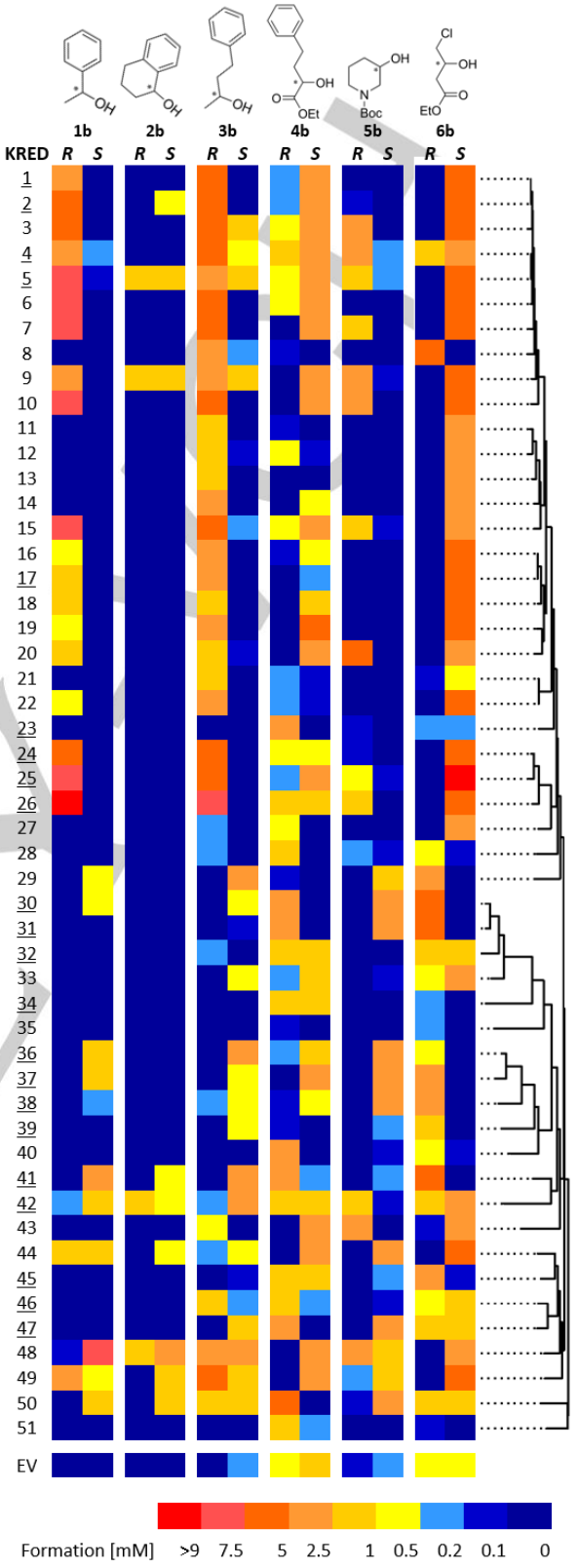



**Figure 2.** Characterization of the KRED library regarding the activity towards substrates **1a**-**6a** in the crude-cell lysate biocatalysis reaction with 10 mM ketone substrate, 1 mM NADH, 1 mM NADPH, 20 mM D-glucose, 0.02 mg $mL^{-1}$ glucose dehydrogenase (GDH-105), and 2 mM $MgCl_2$ in 30 mM MES buffer pH 6.6. A colour gradient shows the concentration of the formed (*S*)- and (*R*)-alcohol products. The background of the *E. coli* cell lysate has been subtracted in all cases and is additionally presented for clarity (EV). KREDs are ordered based on their phylogenetic relationship and literature described KREDs are underlined.

showed the highest activity for the selective reduction of **4a** to (*R*)-**4b** within the library, expanding the biocatalytic toolbox of known KREDs for the synthesis of this chiral reaction intermediate.

The structural motif of piperidine is present in various natural products and biologically active molecules, including APIs.[21] (*S*)-*N*-Boc-3-hydroxylpiperidine (**5b**), for example, serves as the chiral synthon for the lymphoma treating drug ibrutinib,[4j] and biocatalytic processes for the KRED-catalyzed asymmetric reduction of this difficult-to-reduce ketone[22] **5a** to (*S*)-**5b** were reported.[4j, 22-23] In our KRED toolbox, we identified the novel KREDs 44 and 50 from *P. tolaasii* to be selective and moderately active towards the synthesis of (*S*)-**5b.** In addition, we found that the literature-known KREDs 30,[18] 31,[19] 36,[24] 37,[25] 38,[26] and 47[20]are capable to reduce **5a**. To our knowledge, however, none of these KREDs have been reported for the selective synthesis of (*S*)-**5b** before, with exception of KRED 38 [23e] underlining the value of enzyme mining approaches when searching to convert selected substrates with specific enantiopreference.

The chiral synthon ethyl (*S*)-4-chloro-3-hydroxybutyrate (**6b**) is a key intermediate in the synthesis of atorvastatin, a cholesterol-lowering drug.[4i, 27] As an intermediate of high importance, several KREDs have been reported to facilitate the enantiopure and preparative synthesis of (*S*)-**6b** with high substrate loadings.[4i, 28] To complement the existing KRED selection with enzymes present in our library, we screened for the stereo-selective conversion of **6a**. In our study, several KREDs showed the desired activity and produced (*S*)-**6b**, with KRED 25 being most active under the screening conditions. The high activity of KRED 25 towards **6a** is in-line with the initial enzyme characterization by Nie *et al*.[29] However, even though KRED 25 has been described as an anti-Prelog KRED, the (*S*)-selectivity for the synthesis of **6b** has not been reported before, only for homolog enzymes.[30] In addition, 12 novel KREDs were found to exhibit high activity (>50% product formation) for the conversion of **6a** while simultaneously exhibiting high enantiopreference toward the desired S-stereoisomer forming a large pool of alternative KREDs for this interesting biocatalytic reaction. Besides the application of (*S*)-**6b** as chiral synthon, the (*R*)-enantiomer was described as a chiral precursor for the synthesis of L-carnitine.[23a, 31] When targeting this product, the novel KRED 8 could for example serve as an active and selective biocatalyst.

Albeit not enantioselective in the synthesis of most investigated chiral synthons, KRED 48 from *P. tolaasii* showed interesting characteristics in the preparation of (*S*)-1-phenylethanol: within our library, it was the only enzyme candidate with high activity for the selective reduction of **1a** to (*S*)-1-phenylethanol (**1b**).

Bioinformatic analyses of the KRED library

With this wealth of sequence-activity information in hand, we performed bioinformatic analyses with the goal to identify global sequence determinants for the observed enzyme characteristics. Albeit the sole protein sequence is not sufficient to conclude the evolutionary relationship between the different enzymes included in the library (spanning from bacteria, yeasts, fungi, and plant origin), it enables the similarity clustering of the proteins. Using this strategy, we observed that the KRED sequences clustered in two groups (Figure 2), which roughly correlate with the recorded activity. Whereas group 1 (Figure 2, KRED 1-29) predominantly harbors KREDs with anti-Prelog specificity, group 2 (Figure 2, KRED 30-51) includes Prelog specific KREDs mainly (classification of Prelog and anti-Prelog configuration of the investigated alcohol products can be found in the SI).

As the sequence clustering (Figure 2) revealed a connection between sequence similarity and activity, we follow-up by comparing the amino acid distribution between Prelog and anti-Prelog specific KREDs in the hope to reveal the enantiopreference-defining sites. Due to the large sequence space reflected in the initial KRED library, we limited our analysis to enzymes belonging to the short-chain dehydrogenases (SDHs) by applying the Rossmann-fold sequence motif TGxxxGxG. In addition, we only considered enzymes showing a clear Prelog or anti-Prelog preference leading to a selection of 25 wild type enzymes from our library. The alignment of the selected SDR protein sequences showed, as expected, the conserved catalytic triad (S143, Y156, and K160, with the extended N114; Figure 3, box III, KRED 5 numbering) and the NAG and PG-motif (Figure 3, box I and IV, respectively).[15, 32] Subsequently, the SDR sequences were assigned to two subgroups of Prelog (4 sequences) and anti-Prelog KREDs (21 sequences) based on the specificity the enzymes exhibited for the substrates measured in this work. To increase the number of SDHs complying with Prelog's rule in the bioinformatic analysis, we complemented our database with literature known Prelog-specific enzymes (Table S3). By comparing the alignment of these two subgroups, three residues in the substrate-binding regions were identified as possible sequence determinants for the anti-Prelog selectivity of the KREDs, namely residues 94, 157, and 190 (Figure 3, KRED 5 numbering). These residues lie in the two substrate-binding pockets in the active site, which are essential to mediate the enzymes' stereoselectivity.

The alignment highlights the conservation of the large amino acid tyrosine in the anti-Prelog subgroup at position 190, where the Prelog subgroup predominantly harbors small amino acids (Figure 3, box V). The influence of position 190 regarding the stereoselectivity was discussed before: Niefind *et al*. rationalized the presence of the large tyrosine as an important structural feature to shape the corresponding binding pocket A in the KRED from *L. brevis* (identifier in this study: KRED 2). In this way, the binding pocket allows for accommodation of small moieties of the substrate only, accounting for the anti-Prelog specificity of the enzyme.[33] The same was found for the anti-Prelog KRED from *L. kefir* (identifier in this study: KRED 5),[14b] and a mutational study for the acceptance of halogenated acetophenones by *Lactobacillus fermentum* SDR (*Lf*SDR) variants.[32] The Prelog specific KRED 45, on the other hand, contains the smaller serine at the corresponding position (S192). This residue was previously described to be an important residue for controlling the stereoselectivity as mutation S192I affected the enantioselectivity of the investigated reaction.[34]

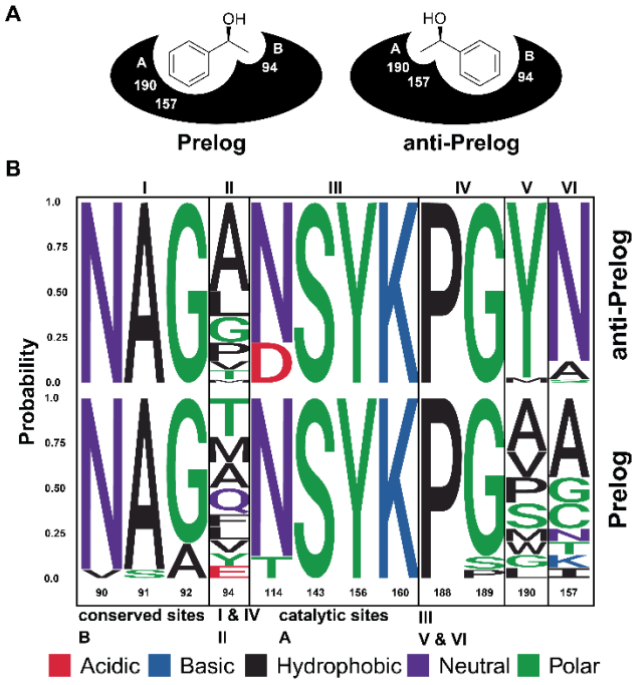
Complementary to position 190 in substrate-binding pocket A, Qin *et al*. identified position 94 as an important structural feature in the substrate-binding pocket B.[32] This finding is reflected in our study, as our sequence alignment shows the tendency of small amino acids for anti-Prelog and large residues for Prelog specific KREDs, shaping a large and small binding pocket, respectively. In addition, our study revealed for the first time a conserved asparagine at position 157, which is part of binding pocket A, and which seems to serve as an additional sequence determinant for anti-Prelog specificity (Figure 3, box VI).

It is of note, that the previous identification of site II and site V (residue 94 and 190 in KRED 5 numbering, respectively) as

10.1002/cctc.202001871


enantiopreference defining sequence motifs was carried out via the analysis of sequence-function data for single substrate types only. The structural study on *L. brevis* SDR, for example, was carried out with acetophenone,[33] whereas the biotransformation-based analyses were based on the conversion of substrates 3-thiacaclopentanone and 3-oxacyclopentanone by the *L. kefir* SDR[14b] or on the conversion of halogenated acetophenone derivatives as catalyzed by *Lf*SDR variants.[32] Conversely, when the sequence of the *Lf*SDR possessing anti-Prelog stereopreference was compared to 1000 further homologues, the residues at site II and V were shown to exhibit low conservation only. The observed diversity was attributed to the residues' possible involvement in substrate recognition[32] marking them as less indicative for enantiopreference prediction from sequence data.

Going one step beyond, our study included the activity data for six diverse substrates as transformed by 25 SDR enzymes (Figure 2) as the basis for bioinformatic analysis. Based on the differences in the amino acid distribution between Prelog and anti-Prelog specific enzymes, we conclude that the sites 94, 157 and 190 can be used as reliable indicators for the prediction of enantioselectivity. Anti-Prelog enzymes seem to prefer tyrosine and asparagine at position 190 and 157, respectively, as well as




a small residue at position 94 (Figure 3) independent of the nature of the transformed substrate.

**Figure 3.** A) Active site pocket model for the Prelog and anti-Prelog specific KREDs with discussed residues highlighted. B) Conservation patterns for Prelog and anti-Prelog KREDs belonging to the SDR subfamily. Residue numbering corresponds to KRED 5. The subgroups consist of 21 and 14 protein sequences for anti-Prelog and Prelog specific KREDs, respectively.

## Conclusion

We have successfully established a multi-factorial screening procedure to build a KRED library of broad synthetic application. By analyzing literature, available genome data in addition to screening microbial strains, we selected 51 KRED genes, including 29 genes encoding for novel enzymes, as the basis of our KRED library and expressed them recombinantly in *E. coli*. The comprehensive substrate scope and complementing enantiopreference of our KREDs in biotransformations of pharmaceutically relevant ketones bode well for our enzyme collection's utility in future academic or industrial route scoutings. Overall, the availability of diverse (wildtype) enzyme libraries is essential for implementing and establishing biocatalytic processes in the pharmaceutical industry. Through the facile and fast screening of readily accessible biocatalysts, promising enzyme candidates can be identified early in the process development, enabling the synthesis of initial amounts of the target compound.[9] In addition, the identified enzyme "hits" then may serve as starting scaffolds for protein engineering campaigns to further optimize the biocatalyst for the specific demands of the process.

## Experimental Section

Chemicals

All reagents were purchased from Sigma Aldrich (USA), VWR (USA) or Carl Roth (Germany) in analytical grade.

Generation and expression of the enzyme library

The codon-optimized genes were synthesized and cloned in the expression vector by Twist Bioscience (CA, USA). The genes including a C-terminal stop codon were cloned in the pET22b vector with the cloning sites NdeI/BamHI or NdeI/XhoI (nucleotide sequences in the SI). The plasmids were transformed in *E. coli* BL21(DE3) and cultured overnight in 1 mL well$^{-1}$. LB medium supplemented with 100 µg mL$^{-1}$ ampicillin in deep-well plates at 37 °C, 300 rpm (5 cm shaking diameter) in the Duetz-System (Kühner AG, Switzerland). For the expression, 0.9 mL well$^{-1}$ ZYM-5052 autoinduction medium[35] (without the addition of trace-metals) supplemented with 100 µg mL$^{-1}$ ampicillin was inoculated with 0.1 mL well$^{-1}$ preculture in a 96-deep well plate and cultured overnight at 20 °C, 300 rpm in the Duetz-System. The culture was harvested by centrifugation at 3,428 *g*, 4 °C, 20 min and the supernatant was discarded. The pellets were stored at -20 °C for the biocatalysis reaction.

Biocatalysis reactions

The cell pellets were chemically lysed for 20 min at 20 °C, 1000 rpm (ThermoMixer F0.5, Eppendorf, Germany) by resuspending the pellet with 100 µL well$^{-1}$ 30 mM MES buffer pH 6.6 containing 1 mg mL$^{-1}$ lysozyme (Sigma Aldrich), 0.5 mg mL$^{-1}$ polymyxin B-sulfate (Carl Roth, Germany), DNase (Sigma Aldrich, USA), and 2 mM $MgCl_2$. The biocatalysis reaction was initiated by the addition of 100 µL well$^{-1}$ 30 mM MES buffer pH 6.6 containing 20 mM substrate (resulting in 10 mM substrate in the biocatalysis reaction), 2 mM NADH, 2 mM NADPH, 40 mM D-glucose, 0.04 mg mL$^{-1}$ glucose dehydrogenase (GDH-105 with ~50 U mg$^{-1}$, Codexis, USA), and 2 mM $MgCl_2$. The plate was sealed, and the reaction incubated overnight at 20 °C, 800 rpm (ThermoMixer, Eppendorf).

Reaction work-up and HPLC analysis for biocatalysis with substrates **1a** and **2a**

The biocatalysis reaction was quenched by the addition of 800 µL well$^{-1}$ of 62.5% methanol in dd$H_2O$ and subsequently sealed, incubated for 10 min at 1300 rpm (ThermoMixer F0.5, Eppendorf), and centrifuged at 3,428 *g*, 4 °C, 20 min. The chiral HPLC analysis for the biocatalysis of substrate **1a**

was performed by injecting 4 uL of the supernatant in the Agilent 1260 HPLC system with a single quadrupole MSD (Agilent, USA) with the Chiralpak IE-3 (150 mm x 4.6 mm, 3.0 µm; Daicel Corporation, Japan) column at 30 °C. The mobile phases A and B were ddH2O + 5% acetonitrile + 0.2% formic acid and acetonitrile + 0.2% formic acid, respectively. The following method was used at a flow rate of 1.2 mL $min^{-1}$: 10% B for 6 min, gradient from 10 to 55% B in 1 min, and gradient from 55 to 85% B in 2.5 min. **1a** was detected in the DAD at 254 nm with at 8.94 min, (*S*)-**1b** was detected in the DAD at 200 nm with at 6.71 min and (*R*)-**1b** was detected in the DAD at 200 nm at 6.43 min. The chiral HPLC analysis for the biocatalysis of substrate **2a** was performed by injecting 4 µL of the supernatant in the Agilent 1290 InfinityLab LC/MSD HPLC system (Agilent, USA) with the Chiralpak IE-3 (150 mm x 4.6 mm, 3.0 µm; Daicel Corporation, Japan) column at 25 °C. The mobile phases A and B were ddH2O + 5% acetonitrile + 0.2% formic acid and acetonitrile + 0.2% formic acid, respectively. The following method was used at a flow rate of 0.4 mL $min^{-1}$: gradient from 10 to 40% B for 5 min, gradient from 40 to 60% B in 5 min, and gradient from 60 to 95% B in 7 min. **2a** was detected in the DAD at 254 nm with at 16.10 min, (*S*)-**2b** was detected in the DAD at 210 nm with at 11.28 min and (*R*)-**2b** was detected in the DAD at 210 nm at 11.52 min.

Reaction work-up and GC analysis for biocatalysis with substrates **3a**-**6a**

The biocatalysis reaction products were extracted by the addition of 1 mL ethyl acetate per well and shaking at 1300 rpm, 20 °C, 10 min (ThermoMixer, Eppendorf). For the subsequent GC-analysis, the plate was centrifuged at 3,428 *g*, 4 °C, 20 min and 350 uL $well^{-1}$ of the organic phase were transferred in a GC glass vial with inlet. The chiral GC analysis for the extracted biocatalysis reactions of substrates **3a**-**6a** were performed by injecting 1 µL in the Agilent GC-MS system equipped with the chiral MEGA-DEX ASX 2-92 (0.25 mm x 0.25 micron x 30 m) column (MEGA S.r.l., Italy). For all biocatalysis reactions, the flow-rate was set to 1.5 mL $min^{-1}$ (helium), the injector to 250 °C, the FID to 300 °C, and the MSD to 250 °C. For the analysis of the biocatalysis reactions with compounds **3a** and **6a**, the oven was set to 55 °C and a ramp with a rate of 1.5 °C $min^{-1}$ to 200 °C was applied, followed by a ramp with a rate of 40 °C $min^{-1}$ to 200 °C, which was on hold time for 3 min. The method was terminated 10 min after the retention of the last compound. The following retention times were detected: **3a** at 45.90 min, (*S*)-**3b** at 52.51 min, (*R*)-**3b** at 53.04 min, **6a** at 48.48 min, (*S*)-**6b** at 45.03 min, (*R*)-**6b** at 44.83 min. For the analysis of the biocatalysis reactions with compound **4a**, the oven was set to 120 °C for 1 min and a ramp with a rate of 2 °C $min^{-1}$ to 169 °C was applied and a subsequent ramp of 40 °C $min^{-1}$ to 200 °C which was on hold time for 3 min. The retention times were: **4a** at 22.37 min, (*S*)-**4b** at 24.56 min, and (*R*)-**4b** at 24.10 min. For the analysis of the biocatalysis reactions with compound **5a**, the oven was set to 120 °C for 1 min and a ramp with a rate of 2 °C $min^{-1}$ to 138 °C was applied, followed by a ramp of 0.5 °C $min^{-1}$ to 145 °C, and a subsequent ramp of 40 °C $min^{-1}$ to 200 °C which was on hold time for 3 min. The retention times were: **5a** at 18.48 min, (*S*)-**5b** at 23.29 min, and (*R*)-**5b** at 23.06 min.

Bioinformatics analysis of the KRED library

The selected subset of SDR sequences were aligned using MAFFT v7.450.[36] To achieve a high accuracy of the multiple sequence alignment the settings were set to *--maxiterate 1000 --localpair*. To create an approximate maximum-likelihood phylogenetic tree Fasttree 2.1.10 was used using the default options.[37] The generated phylogenetic tree was depicted with Dendroscope 3.[38]

## Acknowledgements

This study was sponsored by the Roche Basel PTDC Technology, Innovation & Science committee. In addition, we are very grateful to David Wunderlich and Mara Bönitz-Dulat (Roche Penzberg CPS DXRE) for performing the structure-sequence alignments in the commercial 3DM database.

**Keywords:** anti-Prelog stereoselectivity • biocatalysis • ketoreductase • multi-factorial screening • substrate scope

[1] R. N. Patel, *Biomolecules* **2013**, *3*, 741-777.

[2] a) S. Hashiguchi, A. Fujii, J. Takehara, T. Ikariya, R. Noyori, *J. Am. Chem. Soc.* **1995**, *117*, 7562-7563; b) T. Ohkuma, H. Ooka, S. Hashiguchi, T. Ikariya, R. Noyori, *J. Am. Chem. Soc.* **1995**, *117*, 2675-2676; c) R. Noyori, *Angew. Chem. Int. Ed.* **2002**, *41*, 2008-2022; d) S. Hashiguchi, A. Fujii, K. J. Haack, K. Matsumura, T. Ikariya, R. Noyori, *Angew. Chem. Int. Ed.* **1997**, *36*, 288-290; e) C. S. G. Seo, R. H. Morris, *Organometallics* **2019**, *38*, 47-65; f) K. B. Hansen, J. R. Chilenski, R. Desmond, P. N. Devine, E. J. J. Grabowski, R. Heid, M. Kubryk, D. J. Mathre, R. Varsolona, *Tetrahedron: Asymmetry* **2003**, *14*, 3581-3587; g) J. D. Hayler, D. K. Leahy, E. M. Simmons, *Organometallics* **2019**, *38*, 36-46; h) F. D. Klingler, *Acc. Chem. Res.* **2007**, *40*, 1367-1376; i) S. Rodriguez, B. Qu, K. R. Fandrick, F. Buono, N. Haddad, Y. B. Xu, M. A. Herbage, X. Z. Zeng, S. L. Ma, N. Grinberg, H. Lee, Z. X. S. Han, N. K. Yee, C. H. Senanayake, *Adv. Synth. Catal.* **2014**, *356*, 301-307.

[3] a) F. Hollmann, I. W. C. E. Arends, D. Holtmann, *Green Chem.* **2011**, *13*, 2285-2314; b) R. A. Sheldon, I. W. C. E. Arends, U. Hanefeld, in *Green Chemistry and Catalysis*, **2007**, pp. 91-131.

[4] a) J. Liang, J. Lalonde, B. Borup, V. Mitchell, E. Mundorff, N. Trinh, D. A. Kochrekar, R. N. Cherat, G. G. Pai, *Org. Process Res. Dev.* **2010**, *14*, 193-198; b) N. K. Modukuru, J. Sukumaran, S. J. Collier, A. S. Chan, A. Gohel, G. W. Huisman, R. Keledjian, K. Narayanaswamy, S. J. Novick, S. M. Palanivel, D. Smith, Z. Wei, B. Wong, W. L. Yeo, D. A. Entwistle, *Org. Process Res. Dev.* **2014**, *18*, 810-815; c) J. Liang, E. Mundorff, R. Voladri, S. Jenne, L. Gilson, A. Conway, A. Krebber, J. Wong, G. Huisman, S. Truesdell, J. Lalonde, *Org. Process Res. Dev.* **2010**, *14*, 188-192; d) S. Wu, R. Snajdrova, J. C. Moore, K. Baldenius, U. T. Bornscheuer, *Angew. Chem. Int. Ed.* **2020**, *59*, 2-34; e) X. Chen, Z. Q. Liu, C. P. Lin, Y. G. Zheng, *Bioorg. Chem.* **2016**, *65*, 82-89; f) C. Han, S. Savage, M. Al-Sayah, H. Yajima, T. Remarchuk, R. Reents, B. Wirz, H. Iding, S. Bachmann, S. M. Fantasia, M. Scalone, A. Hell, P. Hidber, F. Gosselin, *Org. Lett.* **2017**, *19*, 4806-4809; g) F. Xu, B. Kosjek, F. L. Cabirol, H. B. Chen, R. Desmond, J. Park, A. P. Gohel, S. J. Collier, D. J. Smith, Z. Q. Liu, J. M. Janey, J. Y. L. Chung, O. Alvizo, *Angew. Chem. Int. Ed.* **2018**, *57*, 6863-6867; h) X. M. Gong, Z. Qin, F. L. Li, B. B. Zeng, G. W. Zheng, J. H. Xu, *ACS Catal.* **2019**, *9*, 147-153; i) S. K. Ma, J. Gruber, C. Davis, L. Newman, D. Gray, A. Wang, J. Grate, G. W. Huisman, R. A. Sheldon, *Green Chem.* **2010**, *12*, 81-86; j) X. Ju, Y. Y. Tang, X. L. Liang, M. Q. Hou, Z. H. Wan, J. H. Tao, *Org. Process Res. Dev.* **2014**, *18*, 827-830; k) Y. Ni, Y. N. Su, H. D. Li, J. Y. Zhou, Z. H. Sun, *J. Biotechnol.* **2013**, *168*, 493-498.

[5] P. D. de Maria, G. de Gonzalo, A. R. Alcantara, *Catalysts* **2019**, *9*.

[6] a) D. Prud'homme, *J. Intellect. Prop. Law* **2015**, *10*, 619-625; b) Chinalawinfo Database, "Notice of the State Intellectual Property Office on Issuing the National Patent Development Strategy (2011-2020)", can be found under http://www.lawinfochina.com/Display.aspx?Lib=law&Id=8611, **2010**.

[7] A. Bateman, M. J. Martin, C. O'Donovan, M. Magrane, R. Apweiler, E. Alpi, R. Antunes, J. Ar-Ganiska, B. Bely, M. Bingley, C. Bonilla, R. Britto, B. Bursteinas, G. Chavali, E. Cibrian-Uhalte, A. Da Silva, M. De Giorgi, T. Dogan, F. Fazzini, P. Gane, L. G. Cas-Tro, P. Garmiri, E. Hatton-Ellis, R. Hieta, R. Huntley, D. Legge, W. D. Liu, J. Luo, A. MacDougall, P. Mutowo, A. Nightin-Gale, S. Orchard, K. Pichler, D. Poggioli, S. Pundir, L. Pureza, G. Y. Qi, S. Rosanoff, R. Saidi, T. Sawford, A. Shypitsyna, E. Turner, V. Volynkin, T. Wardell, X. Watkins, Watkins, A. Cowley, L. Figueira, W. Z. Li, H. McWilliam, R. Lopez, I. Xenarios, L. Bougueleret, A. Bridge, S. Poux, N. Redaschi, L. Aimo, G. Argoud-Puy, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, M. C. Blatter, B. Boeckmann, J. Bolleman, E. Boutet, L. Breuza, C. Casal-Casas, E. De Castro, E. Coudert, B. Cuche, M. Doche, D. Dornevil, S. Duvaud, A. Estreicher, L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, F. Jungo, G. Keller, V. Lara, P.

Lemercier, D. Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. Neto, N. Nouspikel, S. Paesano, I. Pedruzzi, S. Pilbout, M. Pozzato, M. Pruess, et al., *Nucleic Acids Res.* **2015**, *43*, D204-D212.

[8] A. Bateman, M. J. Martin, S. Orchard, M. Magrane, E. Alpi, B. Bely, M. Bingley, R. Britto, B. Bursteinas, G. Busiello, H. Bye-A-Jee, A. Da Silva, M. De Giorgi, T. Dogan, L. G. Castro, P. Garmiri, G. Georghiou, D. Gonzales, L. Gonzales, E. Hatton-Ellis, A. Ignatchenko, R. Ishtiaq, P. Jokinen, V. Joshi, D. Jyothi, R. Lopez, J. Luo, Y. Lussi, A. MacDougall, F. Madeira, M. Mahmoudy, M. Menchi, A. Nightingale, J. Onwubiko, B. Palka, K. Pichler, S. Pundir, G. Y. Qi, S. Raj, A. Renaux, M. R. Lopez, R. Saidi, T. Sawford, A. Shypitsyna, E. Speretta, E. Turner, N. Tyagi, P. Vasudev, V. Volynkin, T. Wardell, K. Warner, X. Watkins, R. Zaru, H. Zellner, A. Bridge, I. Xenarios, S. Poux, N. Redaschi, L. Aimo, G. Argoud-Puy, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, M. C. Blatter, J. Bolleman, E. Boutet, L. Breuza, C. Casals-Casas, E. de Castro, E. Coudert, B. Cuche, M. Doche, D. Dornevil, A. Estreicher, L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz, U. Hinz, C. Hulo, N. Hyka-Nouspikel, F. Jungo, G. Keller, A. Kerhornou, V. Lara, P. Lemercier, D. Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. B. Neto, S. Paesano, I. Pedruzzi, S. Pilbout, M. Pozzato, et al., *Nucleic Acids Res.* **2019**, *47*, D506-D515.

[9] G. W. Huisman, J. Liang, A. Krebber, *Curr. Opin. Chem. Biol.* **2010**, *14*, 122-129.

[10] P. X. Du, P. Wei, W. Y. Lou, M. H. Zong, *Microb. Cell Fact.* **2014**, *13*.

[11] M. D. Truppo, *ACS Med. Chem. Lett.* **2017**, *8*, 476-480.

[12] K. R. Campos, A. Klapars, Y. Kohmura, D. Pollard, H. Ishibashi, S. Kato, A. Takezawa, J. H. Waldman, D. J. Wallace, C. Y. Chen, N. Yasuda, *Org. Lett.* **2011**, *13*, 1004-1007.

[13] L. Fürtges, D. Conradt, M. A. Schätzle, S. K. Singh, N. Kraševec, T. L. Rižner, M. Müller, S. M. Husain, *Chembiochem* **2017**, *18*, 77-80.

[14] a) A. Weckbecker, W. Hummel, *Biocatal. Biotransform.* **2006**, *24*, 380-389; b) E. L. Noey, N. Tibrewal, G. Jimenez-Oses, S. Osuna, J. Park, C. M. Bond, D. Cascio, J. Liang, X. Y. Zhang, G. W. Huisman, Y. Tang, K. N. Houk, *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, E7065-E7072.

[15] Y. Kallberg, U. Oppermann, H. Jornvall, B. Persson, *Eur. J. Biochem.* **2002**, *269*, 4409-4417.

[16] a) N. D. Shen, Y. Ni, H. M. Ma, L. J. Wang, C. X. Li, G. W. Zheng, J. Zhang, J. H. Xu, *Org. Lett.* **2012**, *14*, 1982-1985; b) I. Lavandera, A. Kern, M. Schaffenberger, J. Gross, A. Glieder, S. de Wildeman, W. Kroutil, *Chemsuschem* **2008**, *1*, 431-436; c) H. M. Ma, L. L. Yang, Y. Ni, J. Zhang, C. X. Li, G. W. Zheng, H. Y. Yang, J. H. Xu, *Adv. Synth. Catal.* **2012**, *354*, 1765-1772; d) Y. N. Su, Y. Ni, J. C. Wang, Z. H. Xu, Z. H. Sun, *Chin. J. Catal.* **2012**, *33*, 1650-1660; e) J. Deng, Z. Q. Yao, K. L. Chen, Y. A. Yuan, J. P. Lin, D. Z. Wei, *J. Biotechnol.* **2016**, *217*, 31-40; f) R. Chen, J. Deng, J. P. Lin, X. P. Yin, T. Xie, S. L. Yang, D. Z. Wei, *Biotechnol. Appl. Biochem.* **2016**, *63*, 465-470; g) N. Li, Y. Ni, Z. H. Sun, *J. Mol. Catal. B: Enzym.* **2010**, *66*, 190-197.

[17] K. Perinbam, H. Balaram, T. N. G. Row, B. Gopal, *Protein Eng. Des. Sel.* **2017**, *30*, 265-272.

[18] G. C. Xu, Y. P. Zhang, Y. Wang, Y. Ni, *Bioresour. Technol.* **2018**, *247*, 553-560.

[19] S. Wang, Z. Cheng, L. Qin, Z. Tian, Y. Xu, Q. Liu, C. Zhang, CN111057725A ed., Abiochem Biotechnology Co., Ltd., Peop. Rep. China, **2020**.

[20] D. I. Liao, J. E. Thompson, S. Fahnestock, B. Valent, D. B. Jordan, *Biochemistry* **2001**, *40*, 8696-8704.

[21] E. Vitaku, D. T. Smith, J. T. Njardarson, *J. Med. Chem.* **2014**, *57*, 10257-10274.

[22] Z. T. Sun, R. Lonsdale, A. Ilie, G. Y. Li, J. H. Zhou, M. T. Reetz, *ACS Catal.* **2016**, *6*, 1598-1605.

[23] a) G. W. Zheng, Y. Y. Liu, Q. Chen, L. Huang, H. L. Yu, W. Y. Lou, C. X. Li, Y. P. Bai, A. T. Li, J. H. Xu, *ACS Catal.* **2017**, *7*, 7174-7181; b) N. Itoh, M. Matsuda, M. Mabuchi, T. Dairi, J. Wang, *Eur. J. Biochem.* **2002**, *269*, 2394-2402; c) M. Y. He, S. Zhou, M. L. Cui, X. L. Jin, D. Y. Lai, S. L. Zhang, Z. G. Wang, Z. M. Chen, *Appl. Biochem. Biotechnol.* **2017**, *181*, 1304-1313; d) L. F. Chen, H. Y. Fan, Y. P. Zhang, K. Wu, H. L. Wang, J. P. Lin, D. Z. Wei, *Tetrahedron Lett.* **2017**, *58*, 1644-1650; e) J. J. Chen, M. Yan, L. Xu, *World J. Microbiol. Biotechnol.S* **2017**, *33*; f) X. X. Ying, J. Zhang, C. Wang, M. J. Huang, Y. T. Ji, F. Cheng, M. L. Yu, Z. Wang, M. R. Ying, *Molecules* **2018**, *23*; g) G. P. Xu, H. B. Wang, Z. L. Wu, *Process Biochem.* **2016**, *51*, 881-885.

[24] B. Kosjek, W. Stampfer, M. Pogorevc, W. Goessler, K. Faber, W. Kroutil, *Biotechnol. Bioeng.* **2004**, *86*, 55-62.

[25] K. Abokitse, W. Hummel, *Appl. Microbiol. Biotechnol.* **2003**, *62*, 380-386.

[26] S. S. Wang, Y. Nie, Y. Xu, R. Z. Zhang, T. P. Ko, C. H. Huang, H. C. Chan, R. T. Guo, R. Xiao, *Chem. Commun.* **2014**, *50*, 7770-7772.

[27] R. O. M. A. de Souza, L. S. M. Miranda, U. T. Bornscheuer, *Chem. Eur. J.* **2017**, *23*, 12040-12063.

[28] a) S. Shah, R. Agera, P. Sharma, A. V. Sunder, H. Bajwa, H. M. James, R. P. Gaikaiwari, P. P. Wangikar, *Process Biochem.* **2018**, *70*, 71-78; b) F. J. Zhao, X. Q. Pei, Z. Q. Ren, Z. L. Wu, *Appl. Microbiol. Biotechnol.* **2016**, *100*, 3567-3575; c) G. C. Xu, M. H. Tang, Y. Ni, *J. Mol. Catal. B: Enzym.* **2016**, *123*, 67-72; d) M. Li, Z. J. Zhang, X. D. Kong, H. L. Yu, J. H. Zhou, J. H. Xu, *Appl. Environ. Microbiol.* **2017**, *83*.

[29] Y. Nie, Y. Xu, X. Q. Mu, H. Y. Wang, M. Yang, R. Xiao, *Appl. Environ. Microbiol.* **2007**, *73*, 3759-3764.

[30] Y. Nie, R. Xiao, Y. Xu, G. T. Montelione, *Org. Biomol. Chem.* **2011**, *9*, 4070-4078.

[31] B. N. Zhou, A. S. Gopalan, F. VanMiddlesworth, W. R. Shieh, C. J. Sih, *J. Am. Chem. Soc.* **1983**, *105*, 5925-5926.

[32] F. Y. Qin, B. Qin, W. H. Zhang, Y. L. Liu, X. Su, T. H. Zhu, J. P. Ouyang, J. Y. Guo, Y. X. Li, F. T. Zhang, J. Tang, X. Jia, S. You, *ACS Catal.* **2018**, *8*, 6012-6020.

[33] a) K. Niefind, J. Müller, B. Riebel, W. Hummel, D. Schomburg, *J. Mol. Biol.* **2003**, *327*, 317-328; b) N. H. Schlieben, K. Niefind, J. Müller, B. Riebel, W. Hummel, D. Schomburg, *J Mol. Biol.* **2005**, *349*, 801-813.

[34] P. Paananen, P. Patrikainen, P. Kallio, P. Mantsala, J. Niemi, L. Niiranen, M. Metsa-Ketela, *Biochemistry* **2013**, *52*, 5304-5314.

[35] F. W. Studier, *Protein Expression Purif.* **2005**, *41*, 207-234.

[36] K. Katoh, D. M. Standley, *Mol. Biol. Evol.* **2013**, *30*, 772-780.

[37] M. N. Price, P. S. Dehal, A. P. Arkin, *PLOS ONE* **2010**, *5*.

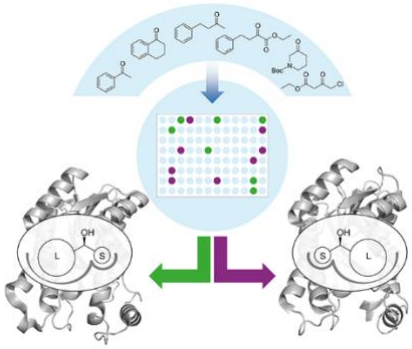
[38] D. H. Huson, C. Scornavacca, *Syst. Biol.* **2012**, *61*, 1061-1067.

10.1002/cctc.202001871





## Entry for the Table of Contents



The availability of ready-to-screen enzyme libraries enables the identification of suitable biocatalysts early in process design and accelerates the implementation of biocatalytic steps in industry. Here, we present a panel of 51 ketoreductases and showcase its ability to synthesize valuable building blocks. The identified protein sequence pattern linked to the enzymes' enantioselectivities is expected to guide future enzyme sourcing campaigns.