ChemBioChem

Research Article
doi.org/10.1002/cbic.202400278

Chemistry Europe
European Chemical Societies Publishing

www.chembiochem.org

Broadening The Substrate Scope of Aldolases Through Metagenomic Enzyme Discovery

Andrea Rizzo,[a] Carmen Aranda,*[a] James Galman,[a] Annette Alcasabas,[a] Akash Pandya,[a] Amin Bornadel,[a] Bruna Costa,[a] Helen C. Hailes,[c] John M. Ward,[b] Jack W. E. Jeffries,[b] and Beatriz Dominguez[a]

Bio-processes based on enzymatic catalysis play a major role in the development of green, sustainable processes, and the discovery of new enzymes is key to this approach. In this work, we analysed ten metagenomes and retrieved 48 genes coding for deoxyribose-5-phosphate aldolases (DERAs, EC 4.1.2.4) using a sequence-based approach. These sequences were recombinantly expressed in *Escherichia coli* and screened for activity towards a range of aldol additions. Among these, one enzyme, DERA-61, proved to be particularly interesting and catalysed the aldol addition of furfural or benzaldehyde with acetone, butanone and cyclobutanone with unprecedented activity. The product of these reactions, aldols, can find applications as building blocks in the synthesis of biologically active compounds. Screening was carried out to identify optimized reaction conditions targeting temperature, pH, and salt concentrations. Lastly, the kinetics and the stereochemistry of the products were investigated, revealing that DERA-61 and other metagenomic DERAs have superior activity and stereoselectivity when they are provided with non-natural substrates, compared to well-known DERAs.

Introduction

The global chemical industry market is estimated to be approximately $5.4 trillion,[1] including materials derived from oil, molecules used in flavours and fragrances, active pharmaceutical ingredients, agrochemicals, and more. A portion of this market ($3.4 billion in 2016[1]) is represented by bio-based processes, which involve the use of enzymes or whole cells to produce chemicals and goods. Examples of this type of product are Sitagliptin,[2] Pregabalin,[3] and detergents for stain removal.[4] One of the considerable advantages of these processes is sustainability: they involve the use of low-cost, non-fossil, and renewable sources,[5] which can have a lower environmental impact when compared to traditional chemical processes. Moreover, enzymes are effective under mild conditions of pressure, pH and temperature, and have the great advantage of often being very specific in terms of stereochemistry of the reaction product.

In order to accelerate their implementation in industry, two enzyme-discovery techniques have become very popular in recent years: sequence and function-based metagenomics. Since metagenomic techniques have become available, they have changed the way enzyme discovery works, allowing scientists to discover more information about under-represented classes of microorganisms, and speeding up the process of enzyme discovery due to high-throughput screening approaches.[6] As a proof of concept, in 2016 Jeffries et al. extracted the DNA from an oral cavity metagenome and developed an *in silico* method for the analysis of the DNA sequence data to retrieve two fully functional lactate dehydrogenases, a malate dehydrogenase and five transketolases.[7]

In recent years, a sub-class of aldolases named deoxyribose-5-phosphate aldolases (DERAs, EC 4.1.2.4) has proved to be very interesting from an industrial perspective, due to its ability to catalyse the formation of carbon-carbon bonds through an aldol addition reaction between a range of substrates.[8] For example, DERAs were used in the production of Islatravir, an investigational HIV treatment drug.[9]

Aldol additions using traditional chemical processes present several disadvantages: they require expensive catalysts, extreme conditions, or the reactions proceed with poor chemo- and stereoselectivity.[10] Moreover, such processes often use transition-metal catalysts and chiral ligands which if not recycled and refined, can have a strong environmental impact associated with the materials' extraction.[11]

In this work, we explored ten metagenomes (five belonging to an NCBI database) and applied sequence-based techniques

[a] *A. Rizzo, Dr. C. Aranda, Dr. J. Galman, Dr. A. Alcasabas, Dr. A. Pandya, Dr. A. Bornadel, Dr. B. Costa, Dr. B. Dominguez*
Johnson Matthey
Unit 260, Cambridge Science Park
Cambridge, CB4 0PZ
E-mail: Carmen.Aranda@matthey.com

[b] *Prof. J. M. Ward, Dr. J. W. E. Jeffries*
Department of Biochemical Engineering
University College London
Gower Street, Bernard Katz Building
London, WC1E 6BT (UK)

[c] *Prof. H. C. Hailes*
Department of Chemistry
University College London
20 Gordon Street
London, WC1H 0AK (UK)

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cbic.202400278

to investigate the presence of DERAs. These genes were cloned, expressed, and characterised in terms of solubility and substrate scope. We decided to focus on the study of furfural (**1**), a green feedstock that can be obtained from lignocellulose and has attracted a lot of interest due to its potential as a building block in the production of valuable chemicals.[12] To the best of our knowledge, it has never been used as an electrophile for aldol addition reactions using DERAs as catalysts. The aldol addition product of **1** and the ketones selected in this study could be used as building blocks for the synthesis of molecules with commercial applications, such as chemicals with antifungal properties.[13] Moreover, this enzymatic activity could be potentially expanded to other furan derivates,[14] and the furan chain can be transformed through the Achmatowicz reaction into dihydropyrans,[15] which are valuable building blocks that found applications for the production of natural products.[16]

Results and Discussion

Metagenomic Library

Initially, ten metagenome libraries were explored, among which five belong to the NCBI "Metagenomic proteins (env nr)" database, while the remaining five are part of the UCL metagenomics and enzyme libraries (Jeffries, Ward & Hailes, Table 1). Among these, the "Drain" metagenome (entry 9) has been already investigated to retrieve transaminase sequences.[17] Some of these metagenomes belong to environments where the pH, temperature or salinity are so extreme that the enzymes develop unique adaptions to function in harsh conditions. These types of enzymes are interesting from an industrial point of view due to their improved stability at high temperatures or in the presence of high concentration of salts or solvents, making them a very good starting point for enzyme engineering.[18]

A total of 48 putative metagenomic DERA sequences were collected from the metagenomes, expressed in *E. coli*, and characterised using whole-cell assays (Figure 1) using the aldol addition reactions depicted in Scheme 1. Additional information on the selection method from the metagenomes is provided in the experimental section. Analysis revealed that 21 of them were not expressed, 15 were expressed but not present in the soluble fraction and 12 were expressed and soluble (25% of the total). Nevertheless, some activity could be detected with enzymes that presented low solubility (DERA-53 and DERA-65). It was also noted that all the enzymes between DERA-30 and DERA-56 are phylogenetically related and, in our hands, were difficult to express in soluble form.

Screening Of Novel Aldehydes As Electrophiles

The collection of metagenomic DERAs was initially tested as whole-cells in aldol addition assays with **1** or **2** (50 mM) as electrophiles, with **3**, **4** or **5** (100 mM) as nucleophiles, in TRIS buffer 100 mM pH 7.5 (Scheme 1). The reactions were incubated at 35 °C overnight and analysed by GC-FID against product standards (for **7** and **15**), or by characteristic fragmentography in GC-MS and/or NMR spectroscopy (Figures S9 to S22). When **4** was used as the nucleophile, products **8** and **10**, or **16** and **18** were formed. Notably, the main products were **8** (91% of total products) and **16** (81% of total products) (Figure 2), meaning that C-3 is preferred for the nucleophilic attack. This data highlights a difference between reported DERAs and another sub-class of well-known aldolases named D-Fructose-6-phosphate aldolase (FSA), which are able to accept **4** as a nucleophile, but activating preferentially C-1.[20]

Amongst all the enzymes, DERA-61 was the most active catalyst for the aldol reaction in all the conditions tested. Figure 2 shows a comparison between the yields obtained in the initial screens with DERA-61 and well-known DERAs from the literature (*Escherichia coli*, *Shewanella halifaxensis*,[21] *Arthrobacter chlorophenolicus*,[22] and the engineered *S. halifaxensis* DERA used in the Islatravir production process[9]). DERA-61 increased the formation of **6** to 59%, and of **7** to 5%. Moreover, it was able to accept **4** as a nucleophile reaching a yield of 24% towards **8** and 2% towards **10**. Lastly, the best results were

Table 1. The metagenomes explored in this work. Reference numbers are available for the NCBI database metagenomes, while the origin is indicated for the proprietary ones.

Entry	Metagenome name	Reference[a]	Number of DERAs
1	Compost metagenome	NCBI:txid702656	3
2	Hydrothermal vent metagenome	NCBI:txid652676	2
3	Bioreactor metagenome	NCBI:txid1076179	1
4	Invertebrate metagenome	NCBI:txid1711999	1
5	Marine sediment metagenome	NCBI:txid412755	1
6	Pilluana	Salt mine	16
7	MV16	Antarctic soil	10
8	MV17	Antarctic soil	4
9	Drain	Domestic drain	2
10	Maras6	Salt mine	8

[a] NCBI Taxonomy ID for NCBI metagenomic databases, and origin for proprietary metagenomic databases.

Research Article
doi.org/10.1002/cbic.202400278
Chemistry Europe
European Chemical Societies Publishing

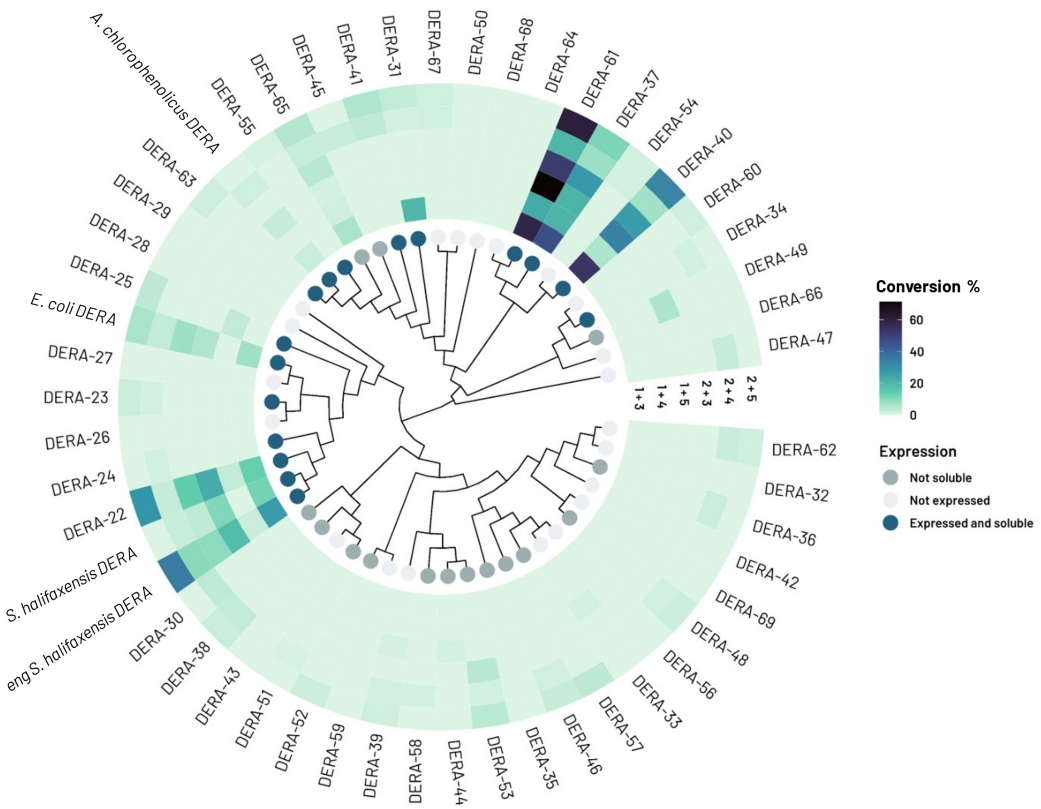
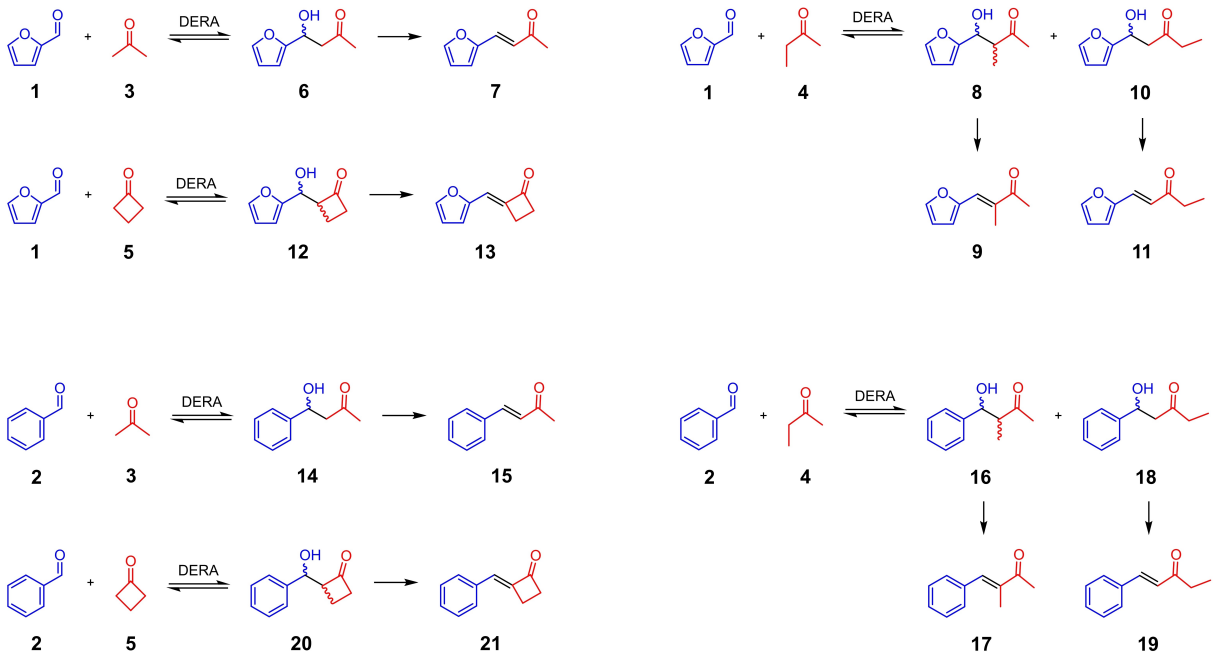


Figure 1. Phylogenetic tree of the 48 putative metagenomic DERA enzymes and four well-known wild-type DERAs. [19] Enzymes in light grey (44%) were not expressed, enzymes in grey (31%) were expressed but not soluble, while enzymes in blue (25%) were expressed and soluble. The heatmap shows the conversion %, calculated as the ratio between the concentration of all products and the initial concentration of aldehyde (**1** or **2**) (100 mM ketone and 50 mM aldehyde in 100 mM TRIS buffer pH 7.5, using whole-cells and incubating at 35 °C overnight). DERA-61 is the enzyme that provided the best conversions in all the tested reactions.



Scheme 1. DERA-catalysed reactions of furfural (**1**) or benzaldehyde (**2**) with acetone (**3**), butanone (**4**) or cyclobutanone (**5**). The aldol addition reaction products **6**, **8**, **10**, **12**, **14**, **16**, **18** and **20** can undergo dehydration that leads to **7**, **9**, **11**, **13**, **15**, **17**, **19** and **21**.

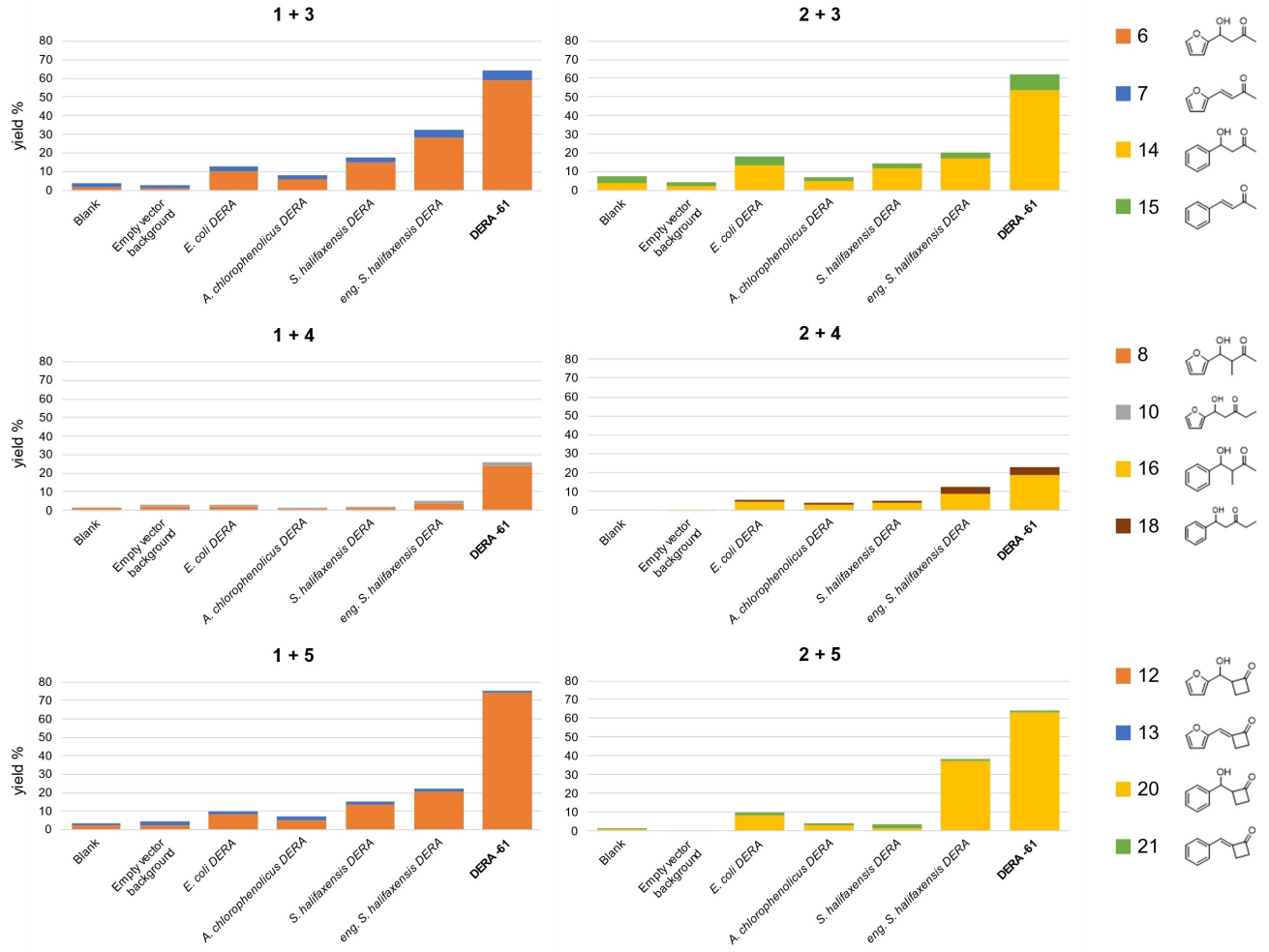


Figure 2. Yield (%) of products identified in the aldol addition of **1** or **2** (50 mM) in the presence of ketones **2**, **3**, or **4** (100 mM). The figures show reactions catalysed using four well-known DERA enzymes and DERA-61, in addition to two negative controls: one containing the empty vector and one containing only the buffer and the substrates (Blank). The data are extrapolated from the screening shown in Figure 1.

obtained with **5**, where the formation of **12** and **13** reached 74% and 1% respectively. Both the negative controls, a blank containing only the buffer and the substrates and a sample containing the empty vector, showed conversions far below these values.

The same analysis was performed for reactions that used **2** as an electrophile: the formation of **14**, **15**, **16**, **18**, **19** and **20** catalysed by DERA-61 were the highest when compared to the results obtained by all the other enzymes tested. Often, aldols can undergo dehydration leading to α,β-unsaturated ketones (enones). In this case, we identified enones in the reactions that used **3** and **5**, but not in the reaction with **4**, which conversion was probably too low to achieve a detectable amount of enones.

The promising data shown in Figure 2 prompted us to explore the reactions further and optimize the working conditions for this enzyme. We focused on the reaction between **1** and **3** investigating the effect of several parameters in the reaction: temperature (30-40 °C), buffer (TRIS, HEPES, GlyGly and phosphate, at a concentration of 100 mM), pH (7-9) and concentration of NaCl (0-0.5 M) used. This last variable was included as the sequence of DERA-61 comes from a metagenome extracted from a salt mine sample and enzymes from halophile microorganisms are adapted to work under concentrations of salt close to saturation, up to 4–5 M NaCl.[23] The experiment was designed as a full factorial, selecting three values for each variable and testing all the different combinations. Additional information on the conditions of each experiment is available in Tables S1 to S4. The results of the screening are shown in Figure 3.

In TRIS, HEPES and phosphate buffer the optimal conditions for the production of **6** were observed at pH 9 and 30 °C. The yield remained constant at approximately 40%. The production of **7**, on the other hand, increased with temperature as well as pH, ranging between 3% and 6%, which suggests that high temperature shifts the equilibrium towards the dehydrations of the aldol. When compared to the first screening data shown in Figure 1, the overall yield in this experiment was lower, decreasing from around 60% to 40%, due to the use of cell-free lysate instead of whole cells.

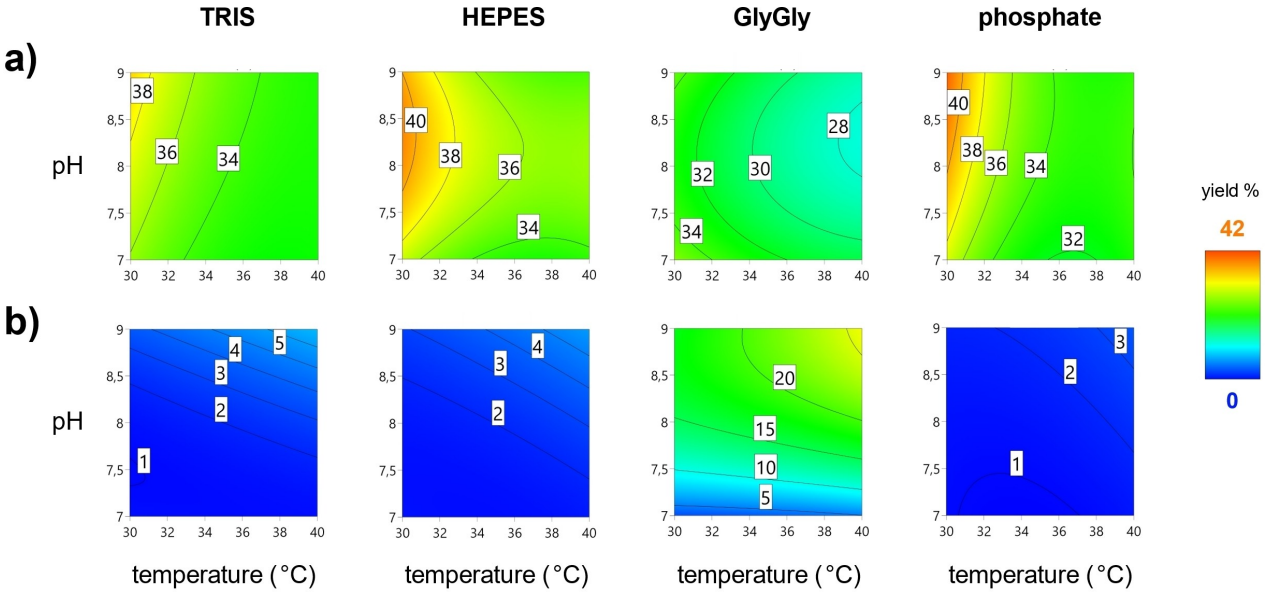


Figure 3. Variations of the activity of DERA-61 in the function of the pH, reaction temperature and buffer type towards the production of **6** (a) and **7** (b). The reactions were carried out using cell-free lysate (10 mg/mL), 50 mM of aldehydes, 100 mM of ketones and incubating for 18 hours. TRIS, HEPES and phosphate buffers had a similar effect on the course of the reaction, while GlyGly buffer shifted the reaction towards the dehydration of **6**.

A different behaviour could be seen in the GlyGly buffer, where the production of **6** did not exceed 34%, but the conversion towards **7** reached 25% at 40 °C and pH 9. This behaviour may be due to the facilitation of a base-catalysed reaction, as observed in the negative control in the absence of enzymes (Figure S4). This is probably due to the primary amine group in the GlyGly molecule forming a Schiff's base with **3** and subsequent reaction as has been demonstrated with other amino acids.[24] In all cases, the effect of the addition of up to 0.5 M of NaCl was shown to have no effect on the aldol addition reaction under any of the conditions used and a small negative effect on the dehydration towards **7** (data not shown).

After finding the improved conditions for the aldol addition using DERA-61, we studied the time course of the reaction (Figure 4). The reaction was carried out in phosphate buffer pH 8 at 30 °C. During the first two hours, five samples were collected at time 0 min, 15 min, 30 min, 60 min and 120 min. After that, a sample was collected every two hours up to 8 hours. Extra time points were collected at 26 and 32 hours. During the last sampling, additional reactions were also carried out in which, after 26 hours, 100 mM of **3** was added, 5 mg of fresh enzyme powder was added, or the sample was incubated at 100 °C.

The conversion of **1** into **6** reached 18% after 60 minutes, and slowed down after 8 h, plateauing at about 30%. The

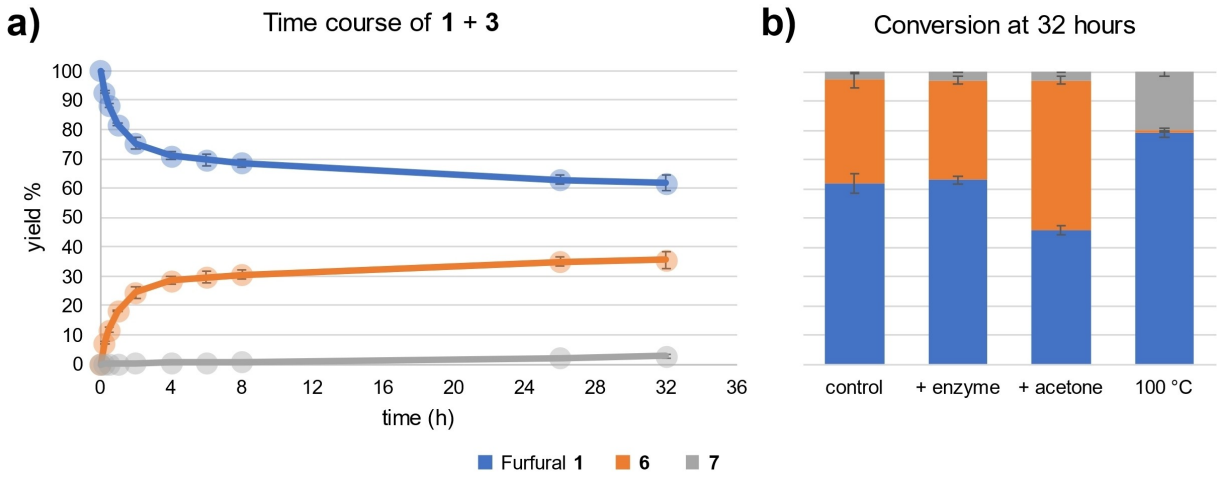


Figure 4. a) Time course of the conversion of **1** with DERA-61 to **6** and **7**, at the optimized reaction conditions (30 °C, phosphate buffer pH 8). b) Comparison between four samples at 32 h: a control (the reaction); a sample in which an extra 5 mg of enzyme was added at 26 h; a sample in which an extra 100 mM of **3** was added at 26 h; and a sample incubated at 100 °C at 26 h.

production of **7**, on the other hand, was much slower, never surpassing 3% of the total products (Fig. 4a). The negative controls (a blank containing only the buffer, **1** and **3**; and another containing the empty vector with **1** and **3**) showed a very small production of **6** that remained below 0.2% (data not shown). Figure 4b shows the effect on conversion of the addition of fresh enzyme, the addition of **3** and the incubation at 100 °C. The addition of fresh enzyme did not change the composition of the reaction, while the addition of **3** shifted the equilibrium towards the production of **6** to 51%. These two results indicate that the enzyme is still active after 26 hours of incubation, and evaporation of **3** could be the limiting factor in the progress of the reaction. This limitation was addressed by testing the reaction between **1** and **3** using concentrations of **3** between 0.1 M and 5 M, which increased the yield by seven-fold when using a concentration of **3** of 1.5 M (29% in 30 minutes, Figure S2). Lastly, the incubation at 100 °C shifted the equilibrium towards the dehydration of **6** to **7**. This approach could be exploited to easily obtain the corresponding enones, using for example a DERA with a high temperature optimum.

Further characterization of DERA-61 was performed by studying the kinetics of the reaction with **1**, using a constant concentration of 1.5 M of **3** (Figure S3). The results show that the K_m of DERA-61 towards **1** is 23.4 ± 1.7 mM, while its k_{cat} is 0.91 ± 0.02 s^{-1}. These values show improved catalytic properties than the corresponding previously reported data with DERA from *E. coli* with the non-natural substrate 2-deoxy-D-ribose (57 mM, 0.1 s^{-1}), but are worse that the kinetics for the reaction with the natural substrate (2-deoxy-D-ribose-5-phosphate (0.64 mM, 68 s^{-1}).[25]

Melting Temperature Analysis

For further characterization of DERA-61, we studied the effect of the type of buffer (TRIS, HEPES, GlyGly and phosphate, at a concentration of 100 mM), pH (between 6 and 8) and concentration of NaCl present (from 0 M to 1 M) on the melting temperature (Figure 5). As for the assays shown in Figure 3, this experiment was designed as a full factorial, and additional information is available in Tables S5 to S7.

The melting temperature of DERA-61 in all the buffers increased when the pH value increased. Moreover, in TRIS, HEPES and GlyGly it increased proportionally with the concentration of salt, up to values between 64 and 68 °C. In phosphate buffer, on the other hand, the T_m was less dependent on the concentration of salt. The maximum value for a T_m of 70 °C, was obtained in phosphate buffer at pH 8 with no addition of sodium chloride. These values are in line with other mesophilic DERAs, such as the DERA from *E. coli*, which in a previous study showed a T_m value close to 70 °C.[21] The same work revealed that DERAs from thermophile organisms have melting temperature values above 100 °C, but unfortunately no halophilic enzyme was involved in the study. High thermostability of DERAs from a halophilic organism (*Haloarcula japonica*) was disclosed in the past when researchers observed that 90% of its activity was retained after incubation at 70 °C for 10 minutes.[26]

Scale-Up And Stereochemistry Determination

All the aldol adducts described in this work possess at least one chiral centre (Scheme 1). In order to understand if the reactions catalysed by DERA-61 and the other metagenomic DERAs were stereoselective, a reaction was carried out between **1** (960 mg) and **3** (1.16 g) using DERA-61 as catalyst. The loading of enzyme was doubled (20 mg/mL), as well as the concentration of electrophile (100 mM), and nucleophile (200 mM). The product of the reaction was isolated to give a 31% isolated yield), and its identity was confirmed by NMR spectroscopy to be **6**. The resulting material was used to develop a chiral GC method and to determine the stereoselectivity of four DERAs from the literature and four DERAs from the metagenomes after 30 minutes of reaction, as shown in Table 2. Reactions were stopped after 30 minutes, in order to minimize the effects of the reversibility of the DERA reaction on the enantiomers produced.

The results revealed that DERA-37 and DERA-61 (entries 6 and 8) gave the highest yield and the highest ee among the enzymes tested, despite the concentration of DERA-61 in the reaction being lower (Table S8). Moreover, the stereoselectivity of DERA-37, DERA-40 and DERA-61 is the opposite of the stereoselectivity of DERA-22 and the DERAs from literature. This

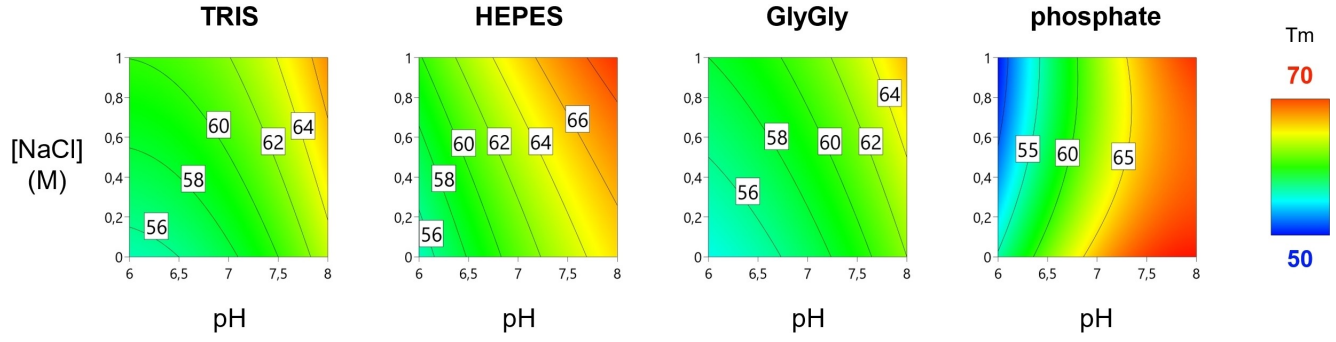


Figure 5. Variations in the melting temperature of DERA-61 with respect to pH, concentration of NaCl and buffer type. The enzyme (2 μM) seemed to be more stable the higher the pH, and the higher the concentration of NaCl, with a maximum value at 70 °C.

Table 2. List of four control DERAs and four novel metagenomic DERAs tested in the **1** + **3** reaction to investigate their enantioselectivity.

Entry	Enzyme name	Yield (%) at 30 min	**6** enantiomeric excess (ee) at 30 min[a]
1	*E. coli* DERA	0.2 ± 0.0	32%
2	*S. halifaxensis* DERA	0.2 ± 0.0	23%
3	eng *S. halifaxensis* DERA	3.3 ± 0.1	51%
4	*A. chlorophenolicus* DERA	0.2 ± 0.0	45%
5	DERA-22	1.4 ± 0.0	24%
6	DERA-37	7.8 ± 0.1	−83%
7	DERA-40	0.7 ± 0.0	−17%
8	DERA-61	6.7 ± 0.2	−72%

[a] Calculated as the ratio between the difference of the enantiomers' peak areas, and the sum of the enantiomers' peak areas.

data is particularly interesting when considering that inverting the stereoselectivity of the DERA from *E. coli* has proven to be a challenging process.[27] Considering the stereoselectivity of DERAs in other well-known reactions,[28] we hypothesize that the exceeding product in these reactions is (*S*)-**6** for entries 1 to 5, and (*R*)-**6** for entries 6 to 8 (Figure S6).

For the reactions between **1** + **5** and **2** + **5**, two chiral centres are present in **12** and **20**. GC-MS analysis of the reactions revealed two product peaks with identical *m/z* values from the mass spectra data, corresponding to two different sets of diastereomers (Figure S7). This suggests that DERA-61 and other metagenomic DERAs have a selectivity between the two diastereomeric forms generated, as shown in Table 3. Again, in order to minimize the effects of the reversibility of the DERA reaction, they were stopped after shorter reaction times (4 h). As in the previous case, DERA-37, and DERA-61 showed the highest diastereomeric excess, while DERA-22 demonstrated no diastereoselectivity. Further investigations are required to identify each stereoisomer produced in these experiments.

Conclusions

This work highlighted the potential of using metagenomic technology for enzyme discovery. By using sequence-based techniques, we have discovered several productive metagenomic DERAs with high activities compared to previously reported DERAs used for this study. DERA-61 showed the best activity in all the reactions tested, when using **1** and **2** as electrophiles for aldol additions. To the best of our knowledge, this is the first report of the acceptance of these substrates by DERAs, thus broadening the substrate scope of this sub-class of aldolases.

Due to their activity, DERA-61 and the other metagenomic enzymes have the potential to become suitable starting points for enzyme engineering, as they may require fewer rounds of mutagenesis to adapt to specific reaction conditions. Moreover, DERA-61's sequence, differs significantly from DERAs from *E. coli*, *S. halifaxensis* and *A. chlorophenolicus*, which share a percentage identity of only 31%, 32% and 42% respectively, suggesting that the diversity of DERAs is yet to be fully explored.

Further steps in this direction could be structural analyses of DERA-61 that take into consideration the differences between this enzyme and other well-known DERAs, to better understand

Table 3. List of four control DERAs and four novel metagenomic DERAs tested in the **1** + **5** and **2** + **5** reactions to investigate their diastereoselectivity. DERA-61 shows the highest de values for both reactions.

Entry	Enzyme name	Yield (%) at 4 h	**12** diastereomeric excess at 4 h[a]
1	*E. coli* DERA	0.8 ± 0.0	22%
2	*S. halifaxensis* DERA	1.4 ± 0.0	4%
3	eng *S. halifaxensis* DERA	2.9 ± 0.0	2%
4	*A. chlorophenolicus* DERA	0.2 ± 0.0	11%
5	DERA-22	3.6 ± 0.5	1%
6	DERA-37	2 ± 0.1	76%
7	DERA-40	0.9 ± 0.0	59%
8	DERA-61	3 ± 0.7	79%

[a] Calculated as the ratio between the difference of the diastereomers' peak areas, and the sum of the diastereomers' peak areas.

which residues are involved in enhancing its activity. Moreover, more extensive reaction optimization could enable a shift of the equilibrium towards the formation of the products, and help overcome the limitations of the aldol additions catalysed by the metagenomic DERAs discovered.

Experimental Section

Chemicals, Media And Reagents

Potassium phosphate monobasic, glycerol and cyclobutanone were purchased by Acros Organics, while HEPES, MTBE, benzaldehyde and butanone were purchased from Alfa Aesar. The SYBR-safe stain was obtained from Invitrogen, while the DNA ladders and analytical standards for **7** and **15** were purchased from Thermo Fisher. Potassium phosphate dibasic, magnesium sulphate, acetone, TRIS, GlyGly, silica gel 40–63, ethyl acetate, acetonitrile and heptane were purchased from Fisher. SOC media, LB, kanamycin sulphate, yeast extract, peptone and Isopropyl ß-D-1-thiogalactopyranoside (IPTG) were obtained from Formedium. Agarose, furfural, NaOH, HCl and deuterated chloroform for NMR spectroscopy were purchased from Sigma Aldrich.

Bioinformatic Analysis For Sequence Selection

The NCBI sequences were obtained through a BLASTp search against the metagenomic proteins database, using as a query the protein sequence of an engineered DERA from *Shewanella halifaxensis*, that has been used in the past for industrial applications.[9] The sequences with the highest degree of percentage identity, below 70%, were then selected.

The analysis of the proprietary metagenomes underwent a selection of all the sequences associated with the DeoC/LacD family aldolase PFAM sequence (PF01791), for a total of 25,934 predicted aldolases genes. After that, the resulting sequences were aligned using, as a query, the same DERA used for the NCBI database, and the protein sequence of a DERA from *Arthrobacter chlorophenolicus*. This DERA proved in the past to have accepted a broader range of substrates when compared to other common DERAs, thus it was a suitable choice to find enzymes with unique activities.[22] The resulting sequences with a global similarity above 40%, and the highest local similarity in the active site region, were selected to be part of the library. The value of 40% was selected as above this percentage of global similarity enzyme activity tends to be conserved.[29]

Cloning And Expression of Putative DERA Sequences

All the resulting sequences were codon-optimized for *E. coli* and purchased as DNA inserts from Twist Bioscience, then subcloned into a proprietary plasmid named pJlO by means of a Golden Gate method according to the manufacturer's protocol (NEB). To facilitate purification of the enzyme the DNA sequence of DERA-61, was subcloned into pJex-M, a proprietary vector that adds a His-tag to the N-terminus. Both vectors contain a kanamycin resistance and are lactose-induced. Correct insertion of genes was checked by sequencing method by Eurofins, using primers ordered from Thermo Fisher. The constructs were then transformed into chemically competent *E. coli* NEB 10β, which were then grown in LB supplemented with kanamycin (50 mg/L) and harvested to produce glycerol stocks (final glycerol concentration = 15% v/v) and extract the plasmid DNA using the Monarch® Plasmid Miniprep Kit by NEB according to the manufacturer's protocol. The 48 enzymes that were successfully cloned were assigned codes from DERA-22 to DERA-69.

The resulting plasmids were transformed into competent *E. coli* BL21 and inoculated into 2 mL of LB medium supplemented with kanamycin (50 mg/L), which were grown overnight at 37 °C and 250 rpm. 10 μL of the resulting medium were then inoculated into 500 μL of TB medium supplemented with kanamycin (50 mg/L). The cells were grown in a shaking incubator at 37 °C and 250 rpm to an optical density at 600 nm of 0.6 to 0.8, whereby protein expression was induced by the addition of IPTG (0.1 mM). Cell growth was prolonged at 20 °C and 250 rpm overnight. 100 μL of cells were harvested to analyse the expression. A "total protein" fraction was obtained by lysing the cells in a sonicator bath (2 mins total; 10 seconds on, 10 seconds off), and a "soluble protein" fraction was obtained by centrifuging the "total protein" fraction and collecting the supernatant. 3 μL of 6X SDS sample buffer (Thermo Fisher) was added to 15 μL of protein fraction, which was then boiled for 10 minutes and loaded on a precast polyacrylamide gel (Bio-Rad) together with a Precision Plus unstained protein ladder from Bio-Rad. Proteins that did not show any bands were not expressed, proteins that showed a band in the total fraction but not in the soluble fraction were expressed but not soluble, and lastly proteins that showed a band in both fractions were soluble. Cells were then harvested by centrifugation at 4,000 rpm for 20 min. The resulting whole cells were used for the initial library screening.

For larger-scale production of the enzymes, the expression procedure was accomplished in 200 mL of TB. Harvested cells were resuspended in 3 times their weight in phosphate buffer 100 mM pH 7.5 and lysis was performed by ultrasonication (2 mins total; 5 seconds on, 2 seconds off). The resulting suspension was centrifuged at 11,000 rpm for 30 min at 4 °C, and the supernatant was frozen at −80 °C until lyophilized and stored at −20 °C until use. Figure S1 shows an SDS-PAGE of the cell-free lysate of the 4 DERAs from literature and the 4 metagenomic DERAs used in this study, and in Table S8 data from a Bradford analysis are used to estimate DERA concentrations in each reaction mix.

For purification, up to 10 mL of the cell-free lysate was added into a column loaded with a nickel-charged affinity resin (Thermo Fisher). After a 1.5 hours incubation to allow the enzyme to bind, the column was washed to remove non-specific binding proteins and the enzyme with His-tag was eluted with an elution buffer containing 300 mM imidazole. Lastly, the imidazole was removed using a protein concentrator spin column (Sigma) for buffer exchange. The enzyme was stored at −20 °C until use.

Biocatalytic Aldol Addition Using Whole Cells Or Cell-Free Lysate

Reactions were always performed in 1.5 mL microcentrifuge tubes by adding 10 mg/mL of lyophilised cell-free lysate into the reaction mix, except in the initial screening in which whole-cells were harvested from a 500 μL TB culture and used for the reaction (Figure 1 and 2). The total volume of the reaction was 0.4 mL, unless otherwise stated. The reaction mixes contained 100 mM of buffer (TRIS, HEPES, GlyGly or phosphate) at a pH range between 7 and 9 from a 200 mM stock solution, to which NaCl could be added from a stock 5 M concentrated solution. Aldehyde **1** or **2** and one of the ketones **3**, **4** and **5** were then added at a concentration of 50 mM and 100 mM respectively, unless otherwise stated. Reactions were incubated at constant temperature (30-40 °C) and 900 rpm overnight, for approximately 18 hours, after which they were extracted using an epmotion liquid handling system (Eppendorf)

and ethyl acetate (2×750 μL) for GC-FID analysis or into D-chloroform (500 μL) for NMR analysis. Quantification was performed by relative percentage, using a response factor determined with standards.

Melting Temperature Analysis

The melting temperature was measured by means of a thermal shift assay using a Bio-Rad CFX Connect Real-Time PCR Detection System, loaded with 20 μL samples containing 2 μM of protein and a SYPRO® Orange Protein Gel Stain (Merck). The protocol started at 10 °C and increased up to 95 °C in increments of 0.5 °C for 10 seconds.

Scale-Up And Product Purification

The scaled-up reaction was performed in a 250 mL round-bottom flask, equipped with a magnetic stirrer, a temperature probe and a heater. To 100 mL of phosphate buffer (100 mM pH 8) was added 100 mM of **1**, 200 mM of **3** and 20 mg/mL of enzyme in the form of cell-free lysate. The reaction was carried out under mixing at 30 °C over the weekend for a total of 67 hours.

The resulting material was then quenched with three volumes of acetonitrile, mixed for 20 minutes, and decanted into a separating funnel. The bottom layer was discarded, while the top layer was extracted three times with one volume of MTBE. The obtained organic layer was washed twice with 0.5 volumes of NaCl brine, dried over magnesium sulphate and evaporated under reduced pressure. The product was then purified by silica chromatography (heptane and ethyl acetate, 7 : 3) to give **6** in 31 % yield.

Analytical Data

GC analysis was performed on an Agilent 7890B GC System equipped with an HP-5 ms Ultra Inert GC column, 15 m, 0.25 mm, 0.25 μm, 7-inch cage (19091S-431UI). Helium was used as a carrier gas, with a 6.5 mL/min flow rate. Gases were obtained from BOC Gas. The method used for the GC-FID analysis is available in Table S9.

Mass-spectra analysis was conducted on an Agilent 8890 GC System equipped with an Agilent 5977B GC/MSD and an HP-5 ms Ultra Inert GC column, 15 m, 0.25 mm, 0.25 μm, 7-inch cage (19091S-431UI). The method used for the GC-MS analysis is available in Table S10. The retention times of each compound analysed by means of GC are available in Table S11, and an example chromatogram is shown in Figure S5.

Supporting Information

Metagenomic DERA sequences, GC-MS chromatograms, ^{1}H NMR spectra and ^{13}C NMR spectra are available in the Supporting Information.

Acknowledgements

Authors want to thank Prof. Amparo I. Zavaleta (Facultad de Farmacia y Bioquímica, Laboratorio de Biología Molecular, Universidad Nacional Mayor de San Marcos, Lima, Peru), Dr. Carol N. Flores-Fernández, and Dr. Max Cardenas Fernandez for their help in sampling the Maras6 and Pilluana metagenomes. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956631.

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Keywords: DERA · Aldol reaction · biocatalysis · metagenome · furfural

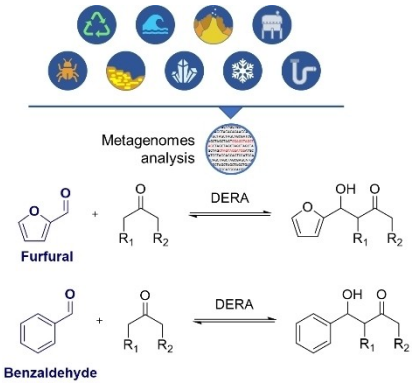
[1] S. O. Ramchuran, F. O'Brien, N. Dube, V. Ramdas, *Curr. Opin. Green Sustain. Chem.* **2023**, *41*, 100832.
[2] A. A. Desai, *Angew. Chem. Int. Ed.* **2011**, *50*, 1974–1976.
[3] C. A. Martinez, S. Hu, Y. Dumond, J. Tao, P. Kelleher, L. Tully, *Org. Process Res. Dev.* **2008**, *12*, 392–398.
[4] J. Chapman, A. E. Ismail, C. Z. Dinu, *Catalysts* **2018**, *8*, 238.
[5] K. Wowra, E. Hegel, A. Scharf, A. Grünberger, K. Rosenthal, *Trends Biotechnol.* **2023**.
[6] a) C. Simon, R. Daniel, *Appl. Microbiol. Biotechnol.* **2009**, *85*, 265–276; b) B. N. Hogg, C. Schnepel, J. D. Finnigan, S. J. Charnock, M. A. Hayes, N. J. Turner, *Angew. Chem. Int. Ed.* **2024**, *63*, 21.
[7] J. W. Jeffries, N. Dawson, C. Orengo, T. S. Moody, D. J. Quinn, H. C. Hailes, J. M. Ward, *ChemistrySelect* **2016**, *1*, 2217–2220.
[8] V. Hélaine, C. Gastaldi, M. Lemaire, P. Clapés, C. Guérard-Hélaine, *ACS Catal.* **2022**, *12*, 733–761.
[9] 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, *Science* **2019**, *366*, 1255–1259.
[10] a) H. N. Genc, U. Ozgun, A. Sirit, *Tetrahedron Lett.* **2019**, *60*, 1763–1768; b) V. Ashokkumar, C. Chithiraikumar, A. Siva, *Org. Biomol. Chem.* **2016**, *14*, 9021–9032; c) S. Shanmuganathan, L. Greiner, P. D. de María, *Tetrahedron Lett.* **2010**, *51*, 6670–6672; d) L. Feng, L. Xu, K. Lam, Z. Zhou, C. Yip, A. S. Chan, *Tetrahedron Lett.* **2005**, *46*, 8685–8689.
[11] S. McCarthy, D. C. Braddock, J. D. Wilton-Ely, *Coord. Chem. Rev.* **2021**, *442*, 213925.
[12] a) A. E. Eseyin, P. H. Steele, **2015**; b) G. Machado, S. Leon, F. Santos, R. Lourega, J. Dullius, M. E. Mollmann, P. Eichler, *Nat. Resour. J.* **2016**, *7*, 115–129.
[13] F. Zhao, H.-H. Dong, Y.-H. Wang, T.-Y. Wang, Z.-H. Yan, F. Yan, D.-Z. Zhang, Y.-Y. Cao, Y.-S. Jin, *Medchemcomm* **2017**, *8*, 1093–1102.
[14] A. R. Lingham, H. M. Hügel, T. J. Rook, *Aust. J. Chem.* **2006**, *59*, 340–348.
[15] C. Wei, R. Zhao, Z. Shen, D. Chang, L. Shi, *Org. Biomol. Chem.* **2018**, *16*, 5566–5569.
[16] K. Cheng, A. R. Kelly, R. A. Kohn, J. F. Dweck, P. J. Walsh, *Org. Lett.* **2009**, *11*, 2703–2706.
[17] L. Leipold, D. Dobrijevic, J. W. Jeffries, M. Bawn, T. S. Moody, J. M. Ward, H. C. Hailes, *Green Chem.* **2019**, *21*, 75–86.
[18] T. O. Akanbi, D. Ji, D. Agyei, *J. Food Biochem.* **2020**, 44, e13475.
[19] G. Yu, *Curr. Protoc. Bioinform.* **2020**, *69*, e96.
[20] R. Roldán, I. Sanchez-Moreno, T. Scheidt, V. Hélaine, M. Lemaire, T. Parella, P. Clapés, W. D. Fessner, C. Guérard-Hélaine, *Chem. Eur. J.* **2017**, *23*, 5005–5009.
[21] M. Dick, O. H. Weiergräber, T. Classen, C. Bisterfeld, J. Bramski, H. Gohlke, J. Pietruszka, *Sci. Rep.* **2016**, *6*, 17908.
[22] D. Chambre, C. Guérard-Hélaine, E. Darii, A. Mariage, J.-L. Petit, M. Salanoubat, V. de Berardinis, M. Lemaire, V. Hélaine, *Chem. Commun.* **2019**, *55*, 7498–7501.

[23] a) S. DasSarma, P. DasSarma, *Curr Opin Microbiol* **2015**, *25*, 120–126; b) S. Patel, M. Saraf, *Halophiles: biodiversity and sustainable exploitation* **2015**, 403–419.
[24] B. List, R. A. Lerner, C. F. Barbas, *J. Am. Chem. Soc.* **2000**, *122*, 2395–2396.
[25] G. DeSantis, J. Liu, D. P. Clark, A. Heine, I. A. Wilson, C. H. Wong, *Bioorg. Med. Chem.* **2003**, *11*, 43–52.
[26] T. Ohshida, J. Hayashi, T. Satomura, R. Kawakami, T. Ohshima, H. Sakuraba, *Protein Expr. Purif.* **2016**, *126*, 62–68.
[27] C. Bisterfeld, T. Classen, I. Küberl, B. Henßen, A. Metz, H. Gohlke, *PLOS ONE* **2016**, *11*, 6.
[28] J. Rouvinen, M. Andberg, J. Pääkkönen, N. Hakulinen, A. Koivula, *Applied Microbiology and Biotechnology* **2021**, *105*, 6215–6228.
[29] W. Tian, J. Skolnick, *J. Mol. Biol.* **2003**, *333*, 863–882.

Manuscript received: May 17, 2024
Revised manuscript received: June 28, 2024
Accepted manuscript online: July 2, 2024
Version of record online: ■■, ■■

RESEARCH ARTICLE

A sequence-based metagenomic approach was used to analyse ten metagenomes and extract putative DERA aldolases with unprecedented activity. We characterized the most promising enzyme by investigating its stability and activity under different conditions including temperature, pH and buffer type. Lastly, these enzymes were tested using furfural and benzaldehyde as acceptors, broadening the substrate scope of this class.



A. Rizzo, Dr. C. Aranda, Dr. J. Galman, Dr. A. Alcasabas, Dr. A. Pandya, Dr. A. Bornadel, Dr. B. Costa, Prof. H. C. Hailes, Prof. J. M. Ward, Dr. J. W. E. Jeffries, Dr. B. Dominguez*

1 – 11

Broadening The Substrate Scope of Aldolases Through Metagenomic Enzyme Discovery