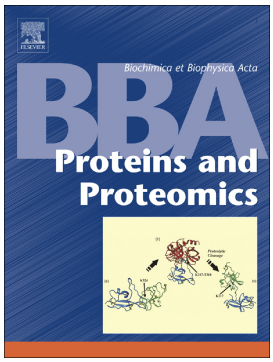


Identification and application of threonine aldolase for synthesis of valuable α-amino, β-hydroxy-building blocks

Mathieu Ligibel, Charles Moore, Robert Bruccoleri, Radka Snajdrova

PII: S1570-9639(19)30208-0

DOI: https://doi.org/10.1016/j.bbapap.2019.140323

Reference: BBAPAP 140323

To appear in: *BBA - Proteins and Proteomics*

Received date: 28 July 2019

Revised date: 31 October 2019

Accepted date: 1 November 2019

Please cite this article as: M. Ligibel, C. Moore, R. Bruccoleri, et al., Identification and application of threonine aldolase for synthesis of valuable α-amino, β-hydroxy-building blocks, *BBA - Proteins and Proteomics*(2019), https://doi.org/10.1016/j.bbapap.2019.140323

Biochimica et Biophysica Acta - Proteins and Proteomics (BBA-Proteins and Proteomics)

Identification and Application of Threonine Aldolase for Synthesis of Valuable α-amino, β-hydroxy-Building Blocks

Mathieu Ligibel,[1] Charles Moore,[1] Robert Bruccoleri,[2] Radka Snajdrova[1,*]

[1]Global Discovery Chemistry, Novartis Institute for BioMedical Research, Basel, Switzerland

[2]Congenomics, LLC, Glastonbury CT USA

*corresponding author: radka.snajdrova@novartis.com

Keywords: biocatalysis, threonine aldolase, enzymatic synthesis, amino acids, *in-silico* gene mining

Abstract

Chiral β-hydroxy α-amino acid structural motifs are interesting and common synthons present in multiple APIs and drug candidates. To access these chiral building blocks either multistep chemical syntheses are required or the application of threonine aldolases, which catalyze aldol reactions between an aldehyde and glycine. Bioinformatics tools have been utilized to identify the gene encoding threonine aldolase from *Vanrija humicola* and subsequent preparation of its recombinant version from *E. coli* fermentation. We planned to implement this enzyme as a key step to access the synthesis of our target API. Beyond this specific application, the aldolase was purified, characterized and the substrate scope of this enzyme further investigated. A number of enzymatic reactions were scaled-up and the products recovered to assess the diastereoselectivity and scalability of this asymmetric synthetic approach towards β-hydroxy α-amino acid chiral building blocks.

1. Introduction

Biocatalysis is gaining ever more attention in the pharmaceutical industry, in part due to recent break-through applications of diverse enzymes offering shorter, greener and more atom- and cost-efficient routes to drug candidates or APIs.[1] Industrial biocatalysis is largely recognized for its power to interconvert functional groups, however bond

formation, especially C-C and C-N, remain under-represented.[2] As mentioned in a recent publication[2], aldolases are discrete enzymes capable of catalyzing C-C bond formation, however, broader use of these biocatalysts has yet to become commonplace industrially. Despite a handful of successful industrial scale processes applying aldolases, the limited number of compatible reactants has prevented wide spread uptake of this technology. Accelerated progress in enzyme engineering has further advanced aldolase application by broadening the scope for acceptors and donors,[3] increasing their stereospecificity[4] and improving their thermostability.[5]

In our case, threonine aldolases (TAs) were of particular interest because of their potential to synthesize β-hydroxy α-amino acids in an enantio and diastereoselective manner. This class of aldolases has been studied extensively by multiple academic groups.[6] The advantage of TAs is their inherent potential to generate two chiral centers in one step starting from two prochiral substrates; inexpensive glycine and an aldehyde (Figure 1). The resulting β-hydroxy, α-amino acid structural motif is found in many natural products (vancomycin, cyclosporin, polyoxin D)[7] and represents an interesting synthon of many pharmaceutical intermediates (e.g. DOPS[8], a peptide mimetic of RNA[9], monocyclic β-lactams[10]). In our case, we were looking to find an asymmetric route towards such a synthon with trans configuration (target compound **1a**). Despite the risk that stereochemistry at the β-center might not be perfect, we were looking for initial starting activity to form our target compound **1a**. Such an enzyme could be further improved by enzyme evolution if required. High-throughput screening and even selection to assist enzyme engineering of PLP-dependent enzymes, such as TAs, has been described previously.[11]

Unfortunately, during the course of our early work the target compound was deselected from our internal program. However, we recognized the utility of the enzymatic work carried out for this project and sought to develop it further for future applications.

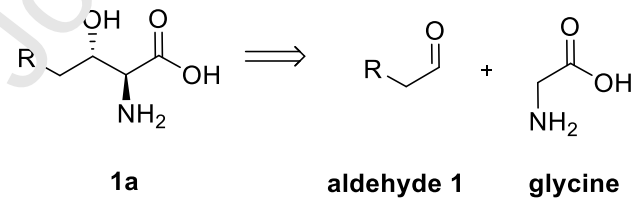


*Figure 1: retrosynthetic planning for synthesis of β-hydroxy α-amino acids using threonine aldolase towards target compound **1a**.*

A number or reports have been published on using glycine-dependent aldolases in the synthesis of such L-*anti* analogues*;* mainly by threonine aldolases (TAs) or serine hydroxymethyltransfereases (SHMT).[12] One such catalyst with good biochemical characterization, L-TA from *Vanrija humicola,* was reported, purified and crystallized almost 50 years ago,[13] however no gene accession number or DNA sequence is known, limiting its application.

Here we report on the identification of the gene encoding this aldolase from *Vanrija humicola* by *in-silico* gene mining. *Vanrija humicola* was originally taxonomically classified as *Candida humicola* and later changed to *Cryptococcus humicola* with its current classification residing in the Genus *Vanrija;* making these organism names homotypic. We describe the subsequent gene cloning, purification and application to the synthesis of β-hydroxy, α-amino acids which are valuable building blocks in the pharmaceutical industry. Publishing this gene gives other researchers the ability to tap into the biocatalytic potential of this aldolase or use it as starting point for subsequent enzyme evolution for a specific application.

2. Materials and Methods

2.1. General

All chemicals and solvents were purchased from commercial suppliers (Sigma-Aldrich, VWR and Brenntag) and were used without further purification.
All NMR spectra were recorded on a Bruker 400 MHz Avance III spectrometer in Methanol-d_4 or DMSO-d_6 and are referenced to the residual solvent peak.

High-resolution mass spectra were recorded on an electrospray ionization quadrupole time-of-flight (ESI-Q-TOF) MaXis system (Bruker Daltonics, Bremen, Germany) coupled to an Agilent 1200 HPLC system (Agilent Technology, California, USA). Analytical HPLC was performed using an Agilent 1260 HPLC system with a photo-diode array UV/Vis detector and a single quadrupole MSD system (Agilent technology, California, USA) or UPLC Acquity (Waters, Massachusetts, USA).

2.2. Gene mining

DNA genomic assembly was performed using MIRA.[14] Gene calling was performed using EXONERATE.[15] Annotation of genes was performed using BLASTx taking the top hit to explain the likely function of each gene. Codon optimization was performed using JCat.[16] All sequence visualization was conducted in Geneious (Biomatters Ltd. Auckland, NZ). The genome assembly used for this work was: *Vanrija humicola* JCM_1457_assembly_v001 (NCBI). Multiple sequence alignments were performed to confirm BLAST results of identified genes and was performed with Clustal Omega.[17]

2.3. Gene cloning and expression

A gene encoding wild-type L-TA from *E. coli* was purchased from Genewiz (South Plainfield, NJ, USA) in pET28a (NdeI and HindIII restriction enzyme cloning). The plasmid was transformed into *E. coli* BL21(DE3) and the cells were cultivated on an LB agar plate containing 50 µg/mL kanamycin. A pre-culture from a single colony of freshly transformed cells was cultured in 20 mL of LB medium containing 50 µg/mL kanamycin at 37 °C. After the culture reached OD_{600}=0.5 it was induced by addition of 1 mM IPTG (final concentration), and further cultivated for 3 hours. The cells were pelleted by centrifugation

at 4000 x g, 4 °C, for 10 min, and the supernatant was discarded. Protein expression was confirmed by SDS-PAGE.

Five genes encoding for different variants of the L-TA from *Vanrija humicola* ATCC 20265 (reclassified into the Genus *Vanrija* at a later date) were purchased from Genscript in pET28a. NcoI and XhoI with a stop codon to the 3' side of the XhoI site for all five variants which are non-his tagged variants. Each plasmid was transformed into *E. coli* BL21(DE3) and the cells were plated on a LB agar plate containing 50 µg/mL kanamycin. A pre-culture from single colony of freshly transformed cells was cultured in 10mL of Terrific Broth medium containing 50 µg/mL kanamycin at 37 °C. After the culture reached OD_{600}=0.4 expression was induced by the addition of 1 mM IPTG (final concentration). The cultivation temperature was then reduced to 30 °C, and the culture was incubated overnight. The cells were pelleted by centrifugation at 7000 x g, at 4 °C, for 10 min, and the supernatant was discarded. Protein expression was determined by SDS-PAGE.

Proposed gene sequences of five potential aldolases in pET28a (NcoI and XhoI restriction enzyme cloning) were purchased to avoid N- and C-terminal HisTag for initial expression study). After the protein expression (as above) and activity study, the gene encoding L-TA_1 was recloned with either N- or C-terminal HisTag for protein purification -NdeI and XhoI restriction enzyme cloning was adopted for the N-his tagged Ald_1. NcoI and XhoI restriction enzyme cloning was adopted for the C-his tagged Ald-1 variant.

2.4. Protein purification

From 1 L culture grown in LB medium 10 g of cell pellet was obtained and suspended at 4 °C in 250 mL of the lysis buffer (100 mM NaH_2PO_4, 300 mM NaCl and 10 mM Imidazole, pH 8.0). The cell suspension was disrupted using a French press twice from 800 to 80 bar.

The suspension was then centrifuged at 10,000 x g for 45 min to obtain a clear supernatant which was subsequently filtered through a 0.22 µm filter. Clarified lysate was loaded onto a HisTrap excel column (GE) at 1 mL/min, previously equilibrated with buffer containing 250 mM imidazole. Protein was eluted with 350 mL of buffer containing 250 mM imidazole buffer and washed with 500 mM imidazole buffer. The fractions containing the aldolase were dialyzed at 4 °C against 2 L of PBS buffer (pH=7.3) with an YM-30 membrane. Protein concentration after dialysis was estimated as 2.9 mg/mL (at A280, f: 0.37). Aldolase was diluted to a final concentration of 0.60 mg/mL with buffer (100 mM phosphate buffer pH=7.0, 0.05 mM PLP and 25% glycerol), aliquoted into 0.5 mL portions, frozen on dry ice and stored at -80 ºC in a freezer until further use.

Large scale purification from 4 L of culture has been carried out by a third party (BLIRT) and purification method is described in supplementary information.

2.5. Enzymatic assays on analytical scale

Calculation of conversions using UPLC- or HPLC-UV analysis:
The conversion was calculated as follows:
Conversion = TAC peak area of the target compound / (TAC UV peak area of the target compound + TAC UV peak area of starting material)
TAC: Total Absorbance Chromatogram (210 – 450 nm)

The HPLC-MS estimated relative conversion was calculated as follows:
MS Estimated relative Conversion = MS peak area of the target compound with the chosen reaction conditions / (MS peak area of the target compound in the reference reaction conditions)

Pig liver serine hydroxymethyltransferase:

To a 2 mL Eppendorf tube was added 400 µL of phosphate buffer 100 mM (pH=7.0, 7.4 or 8.0), 5 µL of the aldehyde solution in a 0.1 M solution in DMSO (1 mM final concentration), 5 µL of glycine (1 M solution, 10 equiv.) and 100 µL of liver S9 homogenate (protein concentration 11.3 mg/mL estimated by Bradford method in triplicates).[18] The tube was incubated overnight at 30 °C and 1100 rpm in an Eppendorf Thermomixer. After 20 hours, the reaction was quenched with 5 mL ACN/MeOH (1:1), stirred 5 min at 1100 rpm and centrifuged at 21,130 x g for 3 min. The clear supernatant was analyzed using the UPLC monitoring method (see supplementary information).

L-TA from *E. coli*:
To a 2 mL Eppendorf tube was added 500 µL of the cell free extract (protocol described in Supplementary information) in 200 mM Tris buffer (pH=7.5), 5 µL of the aldehyde solution in a 0.1 M in DMSO stock (1 mM final concentration) and 5 µL of glycine 1 M (10 equiv.). The tube was incubated overnight at 37 °C and 1100 rpm on an Eppendorf Thermomixer. The reaction was quenched with 500 µL ACN/MeOH (1:1), stirred for 5 min at 1100 rpm and centrifuged at 21,130 x g for 3 min. The clear supernatant was analyzed using HPLC monitoring method (see supplementary information).

Crude lysate of *Vanrija humicola*:
To a 2 mL Eppendorf tube was added 500 µL of the cell free extract (protocol described in Supplementary information) in the reaction buffer (phosphate buffer 100 mM, pH=7.0, 0.05 mM PLP), 2.5 µL of the aldehyde solution from 0.1 M in DMSO stock (0.5 mM final concentration) and 12.5 µL of glycine 1 M (50 equiv.). The tube was incubated overnight at 30 °C and 1100 rpm on an Eppendorf Thermomixer. The reaction was quenched with 500 µL ACN/MeOH (1:1), stirred for 5 min at 1100 rpm and centrifuged at 21,130 x g for

3 min. The clear supernatant was analyzed using HPLC monitoring method (see supplementary information).

2.6. General procedure for screening aldehyde acceptors

To a 2 mL Eppendorf tube was added 10 µL of the aldehyde stock solution in DMSO (0.1 M , final concentration 10mM), 90 µL of buffer (100 mM phosphate buffer, 0.1 mM PLP, 2 M glycine, pH=7.0) and 2 µL of a 10.16 mg/mL enzyme solution in phosphate buffer (100 mM, 1 mM PLP, pH=7.0, 0.2 mg/mL final protein concentration). The mixture was incubated at 40 °C and 1100 rpm on an Eppendorf Thermomixer. After 1.5 hours, the reaction was quenched with 200 µL ACN/MeOH (1:1), stirred for 5 min at 1100 rpm and centrifuged at 21,130 x g for 3 min. The clear supernatant was analyzed using an UPLC monitoring method (see supplementary information).

The reaction with trifluoroacetone **13** was monitored using LC/CAD/MS (see supplementary information).

2.7. Preparative synthesis of the Cbz derivatized β-hydroxy α-amino acids

To a 15-50 mL conical centrifuge tube was added 1.5-15 mL of the reaction buffer (100 mM phosphate buffer, 0.1 mM PLP, 2 M glycine, pH=7.0) and 166-1570 µL of the aldehyde 1 M solution in DMSO (0.166-1.570 mmol, final concentration 100 mM). The solution was vortexed at RT. In the case that the aldehyde precipitated out of solution, another equivalent volume (1.5-15 mL) of the reaction buffer was added and the solution was once more vortexed at RT and/or dissolved with ultrasound (final aldehyde concentration 50 mM). L-TA was then added (16-320 µL of a 10.16 mg/mL stock solution in phosphate buffer (100 mM, 1 mM PLP, pH=7.0, to 0.2 mg/mL final concentration)). The reaction mixture was incubated at 40 °C, 750 rpm in a Eppendorf Thermomixer. After 0.5-17 hours, a conversion in range of 32 % to higher than 90 % was reached. The reaction mixture was quenched by the addition of one reaction volume of ethyl acetate under vigorous stirring over 2-5 minutes. The aqueous layer was concentrated on a rotary evaporator (40 °C, 10 mbar), filtered and directly injected on the reverse-phase column of preparative HPLC.

Each fraction was monitored using an UPLC monitoring method. Fractions containing pure product were collected and lyophilized.

To 1-3 mL of a saturated $NaHCO_3$ aqueous solution in a glass vial equipped with a magnetic stirrer was added 20-288 mg of the α-amino, β-hydroxy-building block and THF 1-3 mL. The reaction mixture was stirred at RT and 1 equiv. of Cbz-Cl was added dropwise and continued with stirring at room temperature during next 0.5-2 hours. After this period, a conversion higher than 90 % was observed.

The crude sample was filtered and purified with the corresponding HPLC preparative method.

For both steps, each fraction was monitored using an UPLC monitoring method.

The overall yield after two steps was 6-37 %

3. Results and Discussion

3.1. Biocatalyst identification

When looking for a potential biocatalyst to access our target compound, we considered, among other enzymatic approaches, an aldolase to perform the C-C bond formation. This strategy would greatly improve process efficiency by generating two chiral centers in one step. Furthermore, this was a particularly appealing approach since an aldolase would utilize two achiral precursors; the corresponding aldehyde and glycine. In addition, glycine is an inexpensive commodity chemical, which makes this approach cost-attractive.

Our search for an enzyme catalyzing the aldol reaction initiated with the screening of commercially available aldolase kit (threonine aldolase kit, Almac). However, none of the commercial enzymes showed activity towards our target compound (**1a**). Disappointed by this finding, we looked to the literature and searched for structurally similar compounds prepared by known aldolases, microbial strains or other biocatalysts. This revealed three aldolase substrate compounds below and their corresponding biocatalytic agents (Figure 2). Compound **2a and 3a** were prepared by *Vanrija humicola* AKU4586[19,20] and product **4a** by L-TA from *E. coli* and serine hydroxylmethyltransferase, respectively. [21,22]

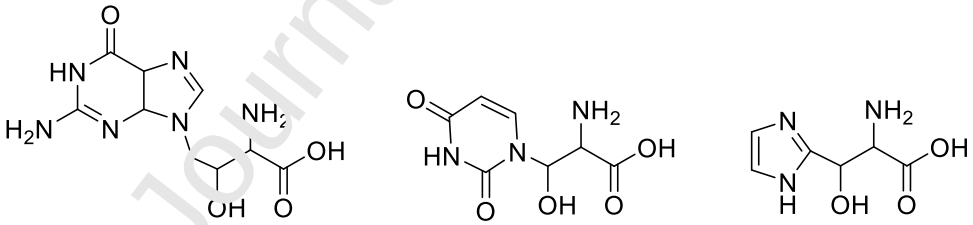


*Figure 2: Structurally similar compounds to target building block **1a** identified from literature search*

Encouraged by these findings, we proceeded to test the biocatalysts against target compound **1a** and the relevant literature substrate as positive control. Testing of serine hydroxymethyltransferase from pig liver was carried out in the form of a liver homogenate. In the case of a positive hit, a recombinant form would be prepared since three gene sequences were available. Unfortunately, no activity was observed towards formation of product **1a** and this approach was discontinued.

The gene encoding L-threonine aldolase (L-TA) from *E. coli* was DNA synthesized and cloned into an inducible expression vector for protein expression from *E. coli* BL21 (DE3). Clear bands with the expected size of L-TA (~45 kDa) were observed in the soluble protein fraction (see supplementary information). Unfortunately, we have not obtained any clear indication based on the activity data that this approach would be successful and so terminated any further effort at this point.

After we exhausted the two options immediately available to us we purchased the yeast strain of *Vanrija humicola* AKU4586 from ATCC (referenced as ATCC 20265,[23]). Upon receipt of the strain it was cultivated according to a patent publication wherein L-threonine was added to the minimal medium as the sole carbon source.[24] An activity assay with whole cells was performed, however, no formation of **1a** was detected, nor was activity against its natural substrate, **2a**, observed. In parallel, we prepared cell-free extract to test if the absence of activity could be due to poor transport of either the aldehyde or glycine into the yeast cells. This appeared to be the case since cell-free extract gave clear conversion to the target product as evidenced by LC/MS.

In order to achieve higher production of enzyme and subsequently increased yield to target product **1a**, and to make this approach scalable, it became clear that recombinantly expressed L-TA would be required. Since no DNA or protein sequence of the corresponding L-TA from *Vanrija humicola* has been published to date, we initiated identification of the the L-TA gene from its genome sequence.

3.2. *In-silico* mining for L-TA

Since we knew that the aldolase of interest was present in *Cryptococcus humicola* we searched in the genome of *Vanrija humicola* ATCC 20265; the final taxonomic name of this organism that had undergone two Genus level reclassifications. This genome entry was found as an unannotated draft quality genome. We were then challenged with identifying the coding exons in an unannotated genome with uncharacterized intron boundaries. By using protein sequences for aldolases which have a PLP binding domain, we used the Exonerate program to identify the genomic regions where the translated genomic regions match well and which have expected intron donor and acceptor sites.

The search of the *Vanrija humicola* genome using the 41 aldolase sequences only found two genomic regions with high scoring alignments of putative coding regions. We reassembled the genomic reads into a higher quality genome using MIRA[14] and annotated the ORFs with the gene caller Exonorate[15]. Next we applied BLASTx to each ORF using the threonine aldolase from *Cryptococcus neoformans* var. *neoformans* JEC21 as the query (see supplementary information). Two potential ORFs for the threonine aldolase were identified within these data with high sequence identity to threonine aldolases from phylogenetically related organisms (see supplementary information). Inspection of these ORFs revealed four possible start codons for one example (L-TA_2.1 through L-TA_2.4 are derived from the same gene with different possible start codons) and a singular likely start codon for the second example (L-TA_1).

Each of these potential full length ORFs was codon optimized for *E. coli* BL21 using JCat[16], which returned an optimized gene with a codon adaptability index of 1.0 in all cases, and ordered the genes as synthetic DNA cloned into pET28a using the NcoI and XhoI restriction sites. This approach provided LacI mediated induction of the cloned gene.

3.3. Expression and activity assays of identified enzyme variants

After receiving the 5 plasmids, the transformation with *E. coli* BL21 (DE3) strain was carried out and initial enzyme expression tested, initially in its native non-tagged version. Most of the constructs were expressed in a soluble form with the expected protein size based on SDS-PAGE (see supporting information). Regardless of expression quantity, all five enzyme variants were submitted to activity assays in the form of crude lysate. To our delight, all five of them showed activity at various levels. The most active variant was L-TA_1 whose DNA sequence was predicted with highest probability based on BLASTx results. The enzymes from expression of the second gene with various length at the N-terminus, were generally less active (approximately one third of activity, data not shown). No dramatic difference in activity was observed in relation to various length of N-terminal.

To characterize the L-TA_1 enzyme better and to probe its substrate scope beyond the target compound (product **1a**), protein purification was necessary to assist with accurate analytical detection and also to facilitate efficient recovery of the product.

3.4. Purification of L-TA_1

In order to simplify purification of the enzyme, the gene with His-tag on either its N- or C-terminus was recloned. Initial expression trials of both tagged enzyme variants indicated good expression of soluble protein. Interestingly, the C-terminal His-tagged variant (C-His L-TA_1) gave slightly higher conversion compared to its untagged version and also higher activity in comparison to its N-terminal His-tag counterpart (1). Both tagged variants were cultivated on larger scale (1 L culture) for protein purification purposes.

Since the initial enzyme expression trials showed a good level of soluble expression, no significant optimization was conducted prior to larger scale enzyme purification, with the exception of a slight modification of the cultivation medium and we directly proceeded using the initial parameters (see supplementary information). Protein purification on Ni-NTA and a subsequent dialysis afforded 72.8 mg of N-His L-TA_1 and 18.5 mg of C-His L-TA_1 as pure protein. The size of the protein was also confirmed by MS (see supplementary information). Based on this successful purification outcome, subsequent purification was outsourced to third party (BLIRT) using their slightly modified protocol (see supplementary information).

Biocatalyst	Relative conversion [a]

L-TA_1 CFE	1
C-His L-TA_1 CFE	1.10
N-His L-TA_1 CFE	0.83
[a] Conversion after 3 hours, based on the MS Peak intensity. CFE – cell free extract	

*Table 1: Conversion to **1a** with N- or C-terminal His Tag or no His tag version of CFE containing L-TA_1 after 3 hours.*

3.5. Substrate scope of L-TA_1

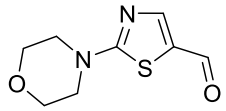
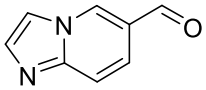
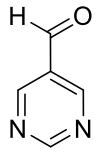
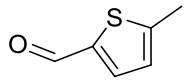
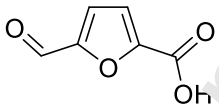
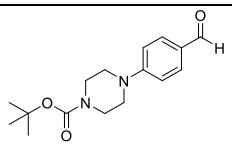
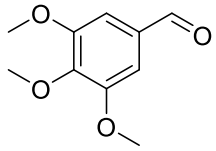
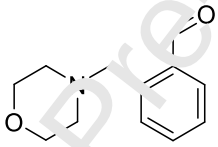
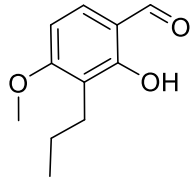
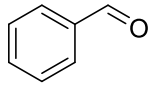
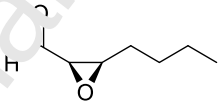
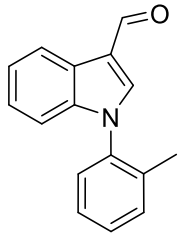
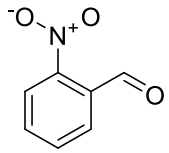
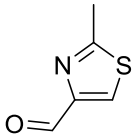
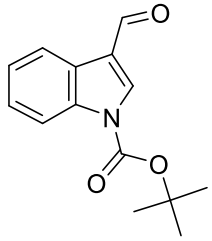
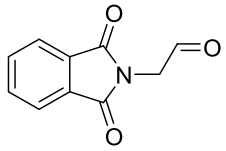
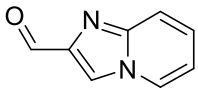
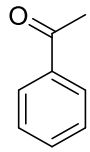
Having a pure enzyme in hand, we next examined the substrate scope of the enzyme. Specifically, substrates leading to desirable building blocks or reflecting, to some extent, intermediates of interest to a pharmaceutical portfolio. The search for suitable aldehydes was conducted based on β-hydroxy α-amino acids or acid derivative motifs. In this case, neither a comprehensive or systematic substrate finger printing was conducted, however, we examined opportunities where the aldolase enzymatic step fit in pharmaceutically relevant chemical space. The focus was on identifying situations that could impact the overall synthetic strategy and perhaps offer a shorter route to a drug or drug candidate. The second criteria was to have an aldehyde substrate with a chromophore for easy reaction monitoring, unlike in the case of target compound **1a**, where due to polarity and absence of a chromophore, analytical monitoring proved to be challenging or required derivatization of an amino group.

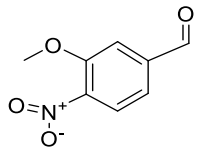
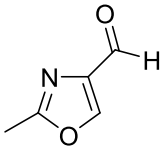
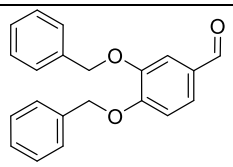
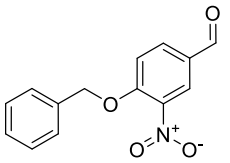
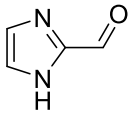
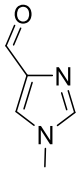
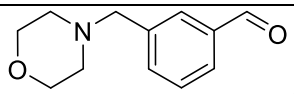
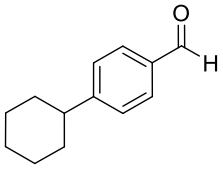
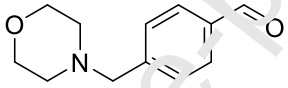
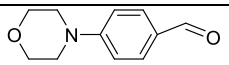
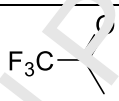
Based on these two criteria, approximately 30 aldehydes were selected for testing, including the literature known substrate **10** as a positive control. In addition to the set of aldehydes, two ketones were chosen to probe if the aldolase could give access to α-amino acids with tertiary alcohols in β-positions, which are otherwise difficult to obtain stereoselectively using classical chemical methods, but would open a door to many novel applications of aldolases.

Trifluroacetone **13** was selected as one of the ketones due to its high reactivity. Indeed, in this case, formation of a new product with the expected mass was observed with LC/CAD/MS as it was not possible to follow the reaction using our standard LC/MS method. To confirm that the new product formed was indeed the expected aldol product, an authentic sample of **13a** was spiked into the LC sample. This procedure revealed that the new product observed in LC/CAD/MS was not the anticipated product **13a** even though it had the same mass as the target compound. It would be difficult to prove the identity of the product from the enzymatic reaction since the conversion was very low and further investigation was deemed unwarranted.

In summary, this methodology has been shown to be beneficial towards accessing a range of molecules relevant to our discovery program. There is clearly a dependence on conversion to the steric and electronic environment of the aldehyde and it is our intention to further study this system to develop conditions that provide a more robust reaction with

broader substrate scope. For now however, the success generated has clearly demonstrated the potential that aldolases have for accessing complex, 'real world' molecules in a medicinal chemistry setting.

Accepted substrates				No conversion
Substrate	Conv. [a]	**Substrate**	Conv. [a]	**Substrate**
 **5**	1.4%	 **14**	~70%	 **24**
 **6**	30%	 **15**	~30%	 **25**
 **7**	43%	 **16**	~50%	 **26**
 **8**	25%	 **17**	N.M.	 **27**
 **9**	87%	 **18**	~90%	 **28**
 **10**	>95%	 **19**	~50%	 **29**

 **11**	91%	 **20**	N.M.	 **30**
 **12**	5%	 **21**	N.M.	 **31**
	0.7%*	 **22**	~10%	 **32**
		 **23**	~10%	 **33**
		 **13**	N.M.*	

[a]based on peak area in UPLC-UV. In some cases peaks were not fully resolved and in these cases the conversion is prefixed by approximation sign ~
* Reaction monitored by LC/CAD/MS, ketone could not be detected using this method, only formation of new product.
N.M. not measured because analyte peak co-elutes with other chromatographic artefacts

Table 2: Substrates tested with L-TA_1. Conversion estimated after 1.5 hours of reaction by UPLC-UV.

It was previously demonstrated by many groups that selectivity for the β-center is usually very low for threonine-dependent aldolase, however, this situation can be improved by enzyme engineering.[4] In our case, an initial hint of activity in the screening stage, would be considered as sufficient encouragement to progress, with activity and selectivity being improved later. For similar reasons, negative data (column III) are also reported to provide an overview and demonstrate where the scope of this aldolase ends with the wild type enzyme, but which could be perhaps expanded during further evolution programs.

3.6. Scale up and isolation of novel β-hydroxy α-amino acids

Based on the screening data, the reactions with the highest conversion (based on chromatography) were scaled up for isolation of novel products and determination of the diastereoselectivity of the enzyme. The well-known downside of this enzymatic reaction

is the necessity to add an excess of glycine to push the reaction towards completion over short reaction times. This equilibrium driving solution is useful on an analytical scale, but complicates the work up procedure on preparative scale. This is especially true if a crude enzyme preparation is used as the isolation of aldol product from reaction mixture is then even more complex due to the presence of additional components resulting from the microbial lysate. To make the preparative synthesis cost effective, a cell free extract was chosen as the biocatalyst formulation in the first instance. We observed a very quick reaction with phtalimidoacetaldehyde **10,** unfortunately the reverse reaction proceeded rapidly. For this reason, the reaction needed to be closely monitored by LC/MS, to determine the stop point of the reaction and hence minimizing the reverse reaction, when using crude lysate. Interestingly, when using purified protein, less of the reverse reaction was observed. Full conversion with purified protein has been reached in less than 20 minutes with concentration of substrate up to 0.95 g/l (5 mM) and no reverse reaction was observed following additional 40 minutes of incubation.

Time points	Conversion with CFE with L-TA_1 (%) [a]	Conversion with pure L-TA_1 at 0.2mg/mL (%) [b]
20min	88	100
40min	82	100
60min	69	100

[a] based on the peak area in UPLC-UV. Due to difficult chromatography monitoring, conversions are only estimates.
[a] based on peak area in UPLC-UV.

Table 3: Time course of the aldol reaction with pthalimidoacetaldehyde and glycine catalysed with CFE of L-TA_1 and purified L-TA_1

A second complication occurring in the reaction using cell-free extract was noticed during structure confirmation of the aldol product **10a.** Despite a high degree of purity according to chromatography under normal and reverse phase modes, a new product has been observed during NMR measurement. A mixture of byproduct (**10d**) and the expected aldol product (**10a**) was detected (byproduct is proposed to be diol base on ^{1}H-NMR and LC/MS analysis, Figure 3). Due to their very similar physico-chemical properties, the separation of both compounds proved elusive in our hands. The ratio between the target product and byproduct was determined as 38:62 (byproduct:amino alcohol) by NMR and 40:60 by chiral HPLC (byproduct:amino alcohol).

In the negative control, when using lysate prepared from the *E. coli* BL21 (DE3) strain which did not bear any plasmid for L-TA enzyme, no activity toward substrate **10** has been observed after even prolonged incubation time to 16.5 hours (Table 4). It is unclear how the diol was formed, since this unusual byproduct has not been observed

when using purified enzyme and the host background does not seem to contribute to this observed phenomenon

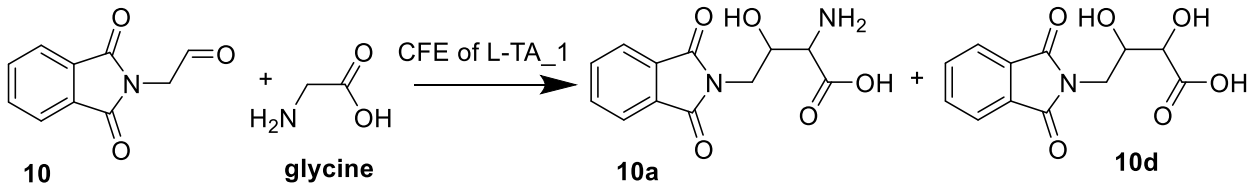


Figure 3: Aldol reaction between pthalimidoacetaldehyde and glycine catalysed by CFE of L-TA_1. Identity of by-product-10d in the reaction mixture is proposed to be diol.

Conditions	Conversion (%) [a]
0.2 mg/mL L-TA_1	>95
10 mg/mL of *E. coli* BL21 (DE3) CFE	0
Buffer only	0
[a] based on peak area in UPLC-UV	

*Table 4: Negative control experiment under screening conditions (10 mM aldehyde **10**, 2 M glycine, 40 °C)*

Because of these two reasons, reverse reaction and byproduct formation, preparative reactions were carried out using purified enzyme. Prior to the actual preparative synthesis starting from aldehyde **9**, **10**, **11**, **16**, **18** and **19**, reaction conditions (temperature, time, substrate and enzyme concentration) and the work up procedures were optimized.

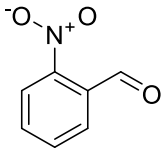
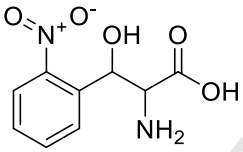
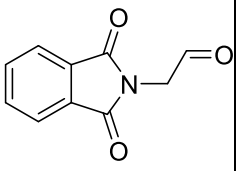
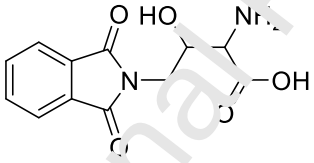
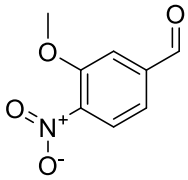
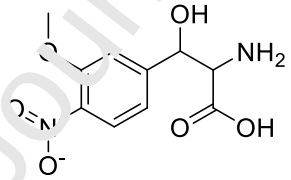
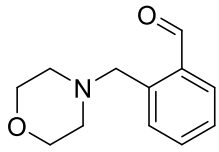
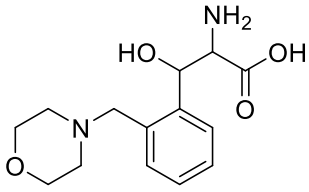
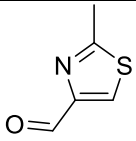
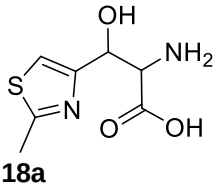
The final conditions applied were; 30 or 40 ºC, 50-100 mM substrate concentration (based on substrate solubility in reaction buffer), 0.14-0.27 mg/mL purified aldolase concentration in 100 mM phosphate buffer (pH=7.0) supplemented with 0.1 mM PLP and 2 M glycine. The highest conversion was observed in less than one hour in most cases. In the case, when conversion stalled, additional and an identical portion of enzyme has been added to boost product formation.

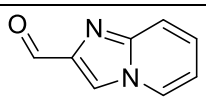
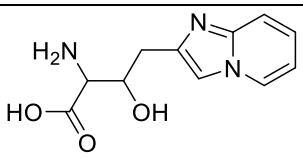
For isolation and purification of the β-hydroxy α-amino acids, initial extraction with ethyl acetate helped to remove unreacted aldehyde. The crude product, in the aqueous layer, was purified by preparative HPLC with reverse phase and fractions containing the product were combined and concentrated. The pure product was then derivatized to the CBz-protected aldol product for easier determination of diastereoisomeric excess values by either HPLC or NMR, except for **16a**, which decomposed during derivatization. The low isolated yields in some cases were attributed to losses in the work up procedure (6). Shortly after finalizing our scale up experiments, an elegant method for an easier work up

was published.[6b] Here a second enzyme, glycine oxidase from *Bacillius subtilis,* has been used to degrade the excess of glycine to glyoxylate after the reaction is completed and thus enabled higher isolated yields of aldol products. This approach will be very interesting to test in our future application.

In conclusion, threonine aldolases represent an appealing alternative to chemical synthesis of β-hydroxy α-amino acid chiral building blocks. Modern bioinformatic methods help to identify genes of novel or putative aldolases, which can be further improved by enzyme engineering. Thanks to their relatively broad substrate scope, these enzymes can serve as an ideal starting point for evolution in development of a particular chemical process and subsequent tailoring to a specific need (enhancement of activity, stability or enantioselectivity).

.

Entry	Aldehyde	Product	Conversion [a]	Reaction conditions	Isolated yield	d.r.[b]
1	 **9**	 **9a**	87%	Overnight, 40°C, 100mM substrate, 0.2mg/mL enzyme	72% (227 mg)	2.2[1] [2] [3] [4]
2	 **10**	 **10a**	>90 %	2.5 hours, 40°C, 50mM substrate, 0.14mg/mL enzyme	8 % (29.5 mg)	2.0 [2]
3	 **11**	 **11a**	84 %	3 hours, 30°C, 50mM substrate, 0.2mg/mL enzyme	58 % (97.4 mg)	1.5 [1] [2] [3] 1.6 [4]
4	 **16**	 **16a**	68 %	1.5 hours, 40°C, 27mM substrate, 0.27mg/mL enzyme	22 % (11.1 mg)	1.0 [1]
5	 **18**	 **18a**	80%	1 hour, 40°C, 100mM substrate, 0.2mg/mL enzyme	66% (288 mg)	1.9 [1]

6	 **19**	 **19a**	>70%	1.5 hours, 40°C, 100mM substrate, 0.2mg/mL enzyme	37% (159 mg)	2.1[1] [4]

[a] Conversion based on peak area in UPLC-UV.
[b] Diastereoisomeric ratio determined by H-NMR [1], by chiral-HPLC after Cbz derivatization [2], by H-NMR after Cbz derivatization [3] or by UPLC-UV after Cbz derivatization [4]

Table 5: Preparative scale aldol reaction catalyzed by purified L-TA_1 and 2 M glycine

Acknowledgment:

The authors would like to thank Liliane Porchet-Zemp for expression trials and SDS-PAGE analysis and David Plattner for developing method for purification of enzymes and supplying initial quantities of enzymes. BLIRT S.A. provided assistance on the purification of enzymes based on in-house protocol, which is gratefully acknowledged. Thanks to Finton Sirockin for his assistance on cheminformatics analysis. We are also very thankful to our colleagues supporting us with analytical (Chiral separation: Dan Huynh, Christophe Bury, Marie-Anne Lozach. Achiral separation: Céline Patissier) and spectroscopic methods (NMR: Thomas Lochmann, Emine Sager, Alexandre Luneau, Harald Schroeder. HRMS: Etienne Richard. LC/CAD/MS: Nicolas Lambert, Etienne Richard) and X-ray crystallography (Piechon Philippe)

Supporting information
Can be found as a separate word file

References:

Highlights for review

We would like to thank the reviewers for their thorough reading of our manuscript! We believe that thanks to their dedication, comments and suggestion, we could improve the manuscript's quality to the benefit or the readers.

Biochimica et Biophysica Acta - Proteins and Proteomics (BBA-Proteins and Proteomics)

Identification and Application of Threonine Aldolase for Synthesis of Valuable α-amino, β-hydroxy-Building Blocks

Mathieu Ligibel,[1] Charles Moore,[1] Robert Bruccoleri,[2] Radka Snajdrova[1,*]

[1]Global Discovery Chemistry, Novartis Institute for BioMedical Research, Basel, Switzerland

[2]Congenomics, LLC, Glastonbury CT USA

*corresponding author: radka.snajdrova@novartis.com

Conflict of interest

All authors declare no conflict of interest

[1] a. A. Crameri, M.L. Hill, S. L. Lovelock, M. Schober, D.G. Tew, P. J. Thomas, NOVEL PROCESSES FOR THE PRODUCTION OF OLIGONUCLEOTIDES, US 20180023122 A1. b. F.K. Kleinbeck-Riniker, B. Martin, G. Penn, F. Venturoni, T. Schlama, T. Ruch, T. Allmendinger, B. Wietfeld, P. Filipponi, NEW PROCESS AND INTERMEDIATES, WO2015098430. c. C. Plass, A. Hinzmann, M. Terhorst, W. Brauer, K. Oike, H. Yavuzer, Y. Asano, A. J. Vorholt, T. Betke, H. Groeger. Approaching Bulk Chemical Nitriles from Alkenes: A Hydrogen Cyanide-Free Approach through a Combination of Hydroformylation and Biocatalysis. *ACS Catal.* 96 (2019) 5198-5203.

[2] J.P. Adams, M.J.B Brown, A. Diaz-Rodriguez, R.C. Loyd, G.-D. Roiban, Biocatalysis: a pharma perspective. Adv. Synth. Catal. 361 (2019), 2421-2432.

[3] a. R. Roldán, I. Sanchez Moreno, T. Scheidt, V. Hélaine, M. Lemaire, T. Parella, P. Clapés, W. D. Fessner, C. Guérard-Hélaine, Breaking the Dogma of Aldolase Specificity: Simple Aliphatic Ketones and Aldehydes are Nucleophiles for Fructose-6-phosphate Aldolase, Chem. Eur. J. 2017, 23, 5005–5009; b. D. Güclü, A. Szekrenyi, X. Garrabou, M. Kickstein, S. Junker, P. Clapés, W.-D. Fessner, Minimalist Protein Engineering of an Aldolase Provokes Unprecedented Substrate Promiscuity ACS Catal. 6 (2016) 1848–1852;*c.* R. Roldán, K. Hernandez, J. Joglar, J. Bujons, T. Parella, I. Sánchez-Moreno, V. Hélaine, M. Lemaire, C. Guérard-Hélaine, W.-D. Fessner, Biocatalytic Aldol Addition of Simple Aliphatic Nucleophiles to Hydroxyaldehydes, *ACS Catal.* 8 (2018) 8804–8809. d. S. Junker, R. Roldan, H. J. Joosten, P. Clapés, W. D. Fessner, Complete Switch of Reaction Specificity of an Aldolase by Directed Evolution In Vitro: Synthesis of Generic Aliphatic Aldol Products. *Angew. Chem.* 130 *(*2018) 10310–10314.

[4] H. J. Gwon, S.H. Balik, Diastereoselective synthesis of L: -threo-3,4-dihydroxyphenylserine by low-specific L-threonine aldolase mutants, Biotechnol Lett 32 (2010) 143-149.

[5] L. Wieteska, M. Ionov, J. Szemraja, C. Fellerc, A. Kolinskid, D. Gront, Improving thermal stability of thermophilic l-threonine aldolase from Thermotoga maritima, J. Biotechnol 199 (2015) 69-76.

[6] a. S.L. Goldberg, A. Goswami, Z.Guo, Y. Chan, E.T. Lo, A. Lee, V. C. Truc, K. J. Natalie, C. Hang, L. T. Rossano, M. A. Schmidt. Preparation of β-hydroxy α-amino acids using recombinant D-Threonine aldolase. b. S. F. Beaudoin, M. P. Hanna, I. Ghiviriga, J.D. Stewart. Progress in using threonine aldolase for preparative synthesis. Enzyme and Microbial Technology 119 (2018) 1-9. c. K. Baer, N. Dueckers, T. Rosenbaum, C. Leggewie, S. Simon, M. Krausser, S. Osswald, W. Hummel, H. Groeger. A study towards efficient L-threonine aldolase-catalyzed enantio- and diastereoselective aldol reations of glycine with substituted benzaldehydes: biocatalyst production and process development. Tetrahedron Asym. 22 (2011) 925-928. d. N. Dueckers, K. Baer, S. Simon, H. Groeger, W. Hummel, Threoning aldolases-screening, properties and applications in the synthesis of non-proteinogenic β-hydroxy α-amino acids, Appl. Microbiol Biotechnol 88 (2010) 409-424.
[7] a. G. Yim, M.N. Thaker, K. Koteva, G. Wright, Glycopeptide antibiotic biosynthesis. J. Antibiot. 67 (2014) 31-41. b. S. Suzuki, K. isono, J. Nagatsu, T. Mizutami, Y. Kawashim, T. Mizuno, A new antibiotic pyloxin A. J. Antibiot. 18 (1965) 131. c. M.C. Pirrung, D.S. Nunn, A.T. McPhail, R.E. Mitchell, Synthesis and absolute configuration of hydroxythreonine, a biosynthetic precursor of rhizobitoxine in Pseudomonas, Bioorg. Med. Chem. Lett. 3 (1993) 2095-2098.
[8] S.H. Baik, H. Yoshioka, H. Yukawa, S. Harayama, Synthesis of L-threo-3,4-dihydroxyphenylserine (L-threo-dops with thermostabilised low-specific L-threonine aldolase from Streptomyces coelicolor A3(2), J. Microbiol. Biotechnol. 17 (2007) 721-727.
[9] T. Miura, M. Fujii, K. Shingu, I. Koshimizu, J. Naganoma, T. Kajimoto, Y. Ida, Application of L-Threonine Aldolase-Catalyzed Reaction for the Preparation of a Peptidic Mimetic of RNA: A Leading Compound of Vero-toxin Inhibitors, Tetrahedron Lett 39 (1998) 7313-7316.
[10] V. P. Vassilev, T. Uchiyama, T. Kajimoto, C.-H. Wong, An Efficient Chemo-Enzymatic Synthesis of α-Amino-β-Hydroxy- γ-Butyrolactone, Tetrahedron Lett 36 (1995) 5063-5064.
[11] L. Giger, M.D. Toscano, M. Bouzon, P. Marliere, D. Hilvert. A novel genetic selection system for PLP-dependent threonine aldolases. Tetrahedron 68 (2012) 7549-7557.
[12] Review and references therein: K. Fesko, M. Gruber-Khadjawi, Biocatalytic Methods for C-C Bond Formation, ChemCatChem 5 (2013) 1248-1272
[13] a. H. Yamada, H. Kumagai, T. Nagate, H. Yoshida, Formation of threonine aldolase by bacteria and yeasts, Agric. Biol. Chem. 35 (1971) 1340-1345. b. H. Kumagai, H. Yoshida, H. Yamada, T. Nagate, Threonine aldolase from Candida humicola 2. Purificaton, crystalization and properties, Biochim Biophys Acta 258 (1972) 779-790.
[14] B. Chevreux, T. Pfisterer, S. Suhai, Automatic assembly and editing of genomic sequences, in S. Suhai (Eds), Genomics and proteomics—functional and computational aspects, Kluwer Academic/Plenum Publishers, New York. 2000, pp. 51–65.
[15] G.S.C. Slater, E. Birney, Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6 (2005) 31.
[16] A. Grote, K. Hiller, M. Scheer, R. Munch, B. Nortemann, D.C. Hempel, D. Jahn, JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res. 33 (2005) 526-531.
[17] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7 (2011) 539.
[18] Preparation of liver homogenate: The liver S9 fractions were prepared as described in Kittelmann M. et al. (2003). M. Kittelmann, U. Reinegger, A. Espigat, L. Oberer, R. Aichholz, E. Francotte, O. Ghisalba, Preparative enzymatic synthesis of the acylglucuronide of mycophenolic acid., Adv. Synth. Catal., 345 (2003)s 825-829.
[19] T. Tanaka, C. Tsuda, T. Miura, T. Inazu, S. Tsuji, S. Nishihara, M. Hisamatsu, T. Kajimoto.Design and synthesis of peptide mimetics of GDP-fucose: targeting inhibitors of fucosyltransferases. Synlett 2 (2004) 243-246.
[20] T. Tanaka, M. Ozawa, T. Miura, T. Inazu, S. Tsuji, M. Hisamatsu, T. Kajimoto. Synthesis of novel mimetics of CMP-sialic acid as the Inhibitors of sialyltransferases. Synlett 9 (2002) 1487-1490.
[21] A. Saeed, D. W. Young. Synthesis of L- β-hydroxyaminoacids using serine hydroxymethytransferase. Tetrahedron 48 (1992) 2507-2514.
[22] T.Kimura, V. P. Vassilev, G.-J. Shenn, C.-H. Wong. Enzymatic synthesis of β-hydroxy-α-amino acids based on recombinant D- and L-Threonine Aldolases. J. Am. Chem. Soc. 119 (1997) 11734-11742.
[23] H. Yamade, Process for preparing L-serine. US 3755081 A patent, 1971.

[24] H. N. Kawasaki, H. E..; Zushi, K. K. Kawasaki, S. Okumura, Biological method of producing serine and serinol derivatives., S. O. US 3871958 patent, 1975.

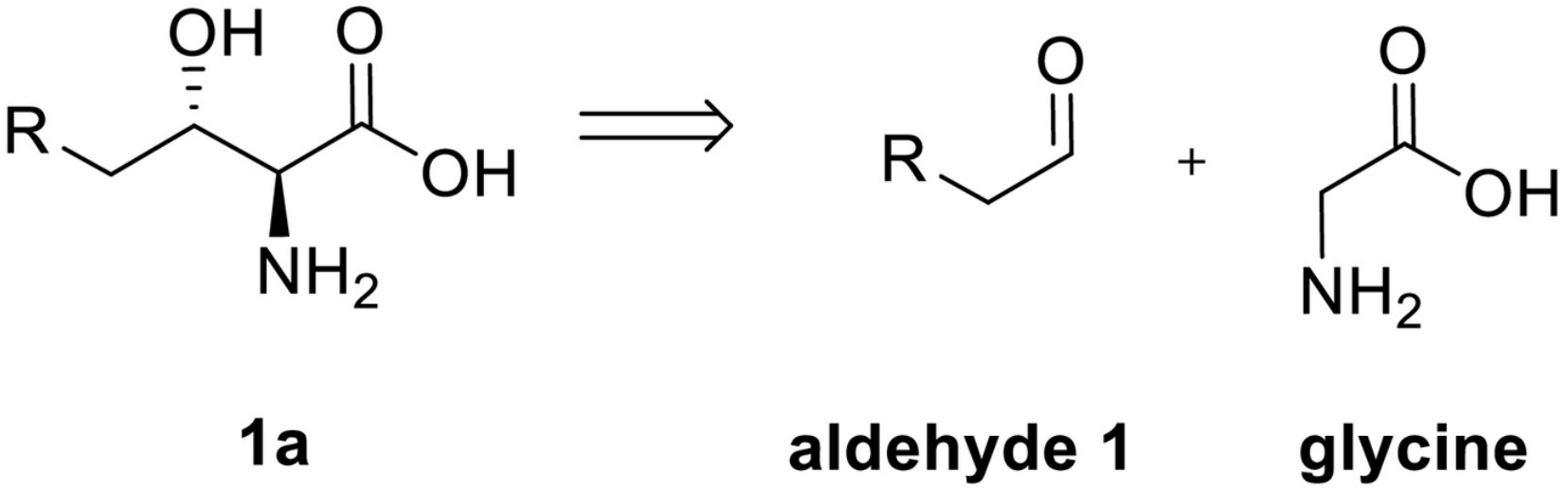
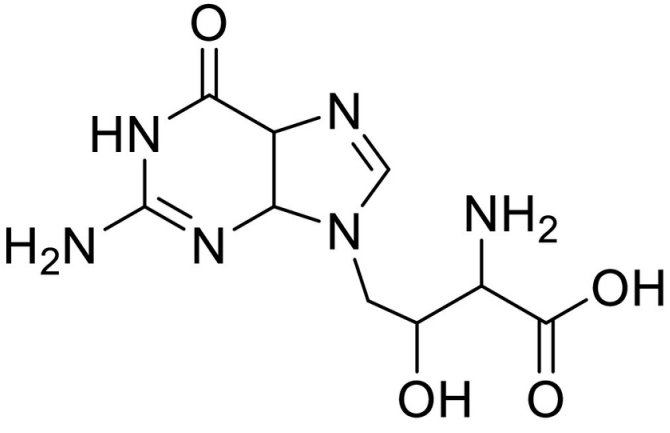
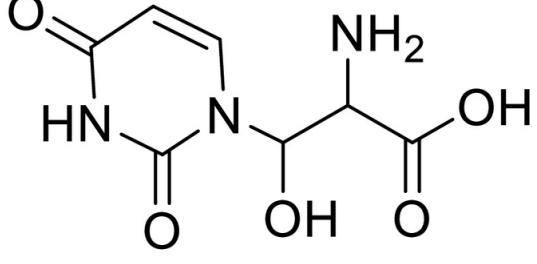


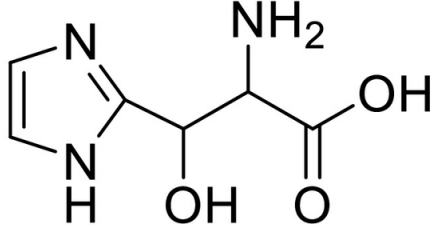
Figure 1



2a



3a



4a

Figure 2

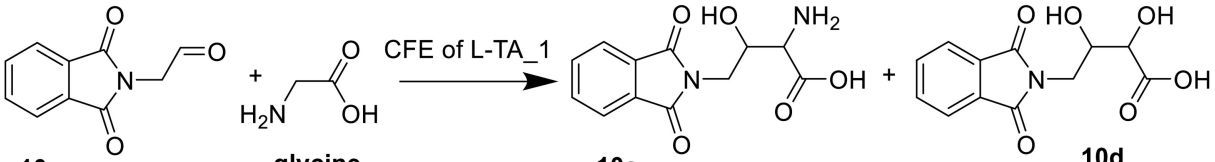


Figure 3