A biocatalytic platform for the synthesis of 2′-functionalized nucleoside analogues.

Matthew Willmott[1], William Finnigan[1], William R. Birmingham[1], Rachel S. Heath[1], Sasha R. Derrington[1], Christian Schnepel[1], Martin A. Hayes[2], Peter D. Smith[3], Francesco Falcioni[4], Nicholas J. Turner[1*]

1 Department of Chemistry, University of Manchester, Manchester Institute of Biotechnology, 131 Princess Street Manchester, UK

2 Compound Synthesis and Management, Discovery Sciences, BioPharmaceuticals, R&D, AstraZeneca, Gothenburg, Sweden

3 Early Chemical Development, Pharmaceutical Sciences, R&D, AstraZeneca, Macclesfield, UK

4 Early Chemical Development, Pharmaceutical Sciences, R&D, AstraZeneca, Cambridge, UK

ABSTRACT: Nucleosides functionalized at the 2′-position play a crucial role in therapeutics, serving as both small molecule drugs and modifications in therapeutic oligonucleotides. However, the synthesis of these molecules often presents significant synthetic challenges. In this study, we present an approach to the synthesis of 2′-functionalized nucleosides based on enzymes from the purine nucleoside salvage pathway. Initially active-site variants of DERA aldolase were generated for the highly stereoselective synthesis of D-ribose-5-phosphate analogs with a broad range of functional groups at the 2-position. Thereafter these 2-modified pentose phosphates were converted into 2′-modified purine analogs by construction of one-pot multi-enzyme cascade reactions, leading to the synthesis of guanosine (2′-OH) and adenosine (2′-OH, 2′-Me, 2′-F) analogues. Our findings demonstrate the capability of these biocatalytic cascades to efficiently generate 2′ functionalized nucleosides, starting from simple starting materials.

Introduction

Nucleoside analogues are an important class of pharmaceutical agent showing activity towards a range of biological targets.[1] For therapeutic applications nucleosides are typically modified at a variety of different positions including the ribose sugar, the phosphate backbone or the nucleobase, depending on the desired pharmacological properties[2]. Particularly important and synthetically challenging is the 2′-modification of the ribose sugar, being utilized in both nucleoside analogue drugs (Figure 1A), and also nucleoside building blocks of therapeutic oligonucleotides (Figure 1B). Therapeutic oligonucleotides are an important new modality that facilitate a precision medicine approach for patients.[3,4] Modifications to the 2'-position of the nucleoside, e.g. 2′-methoxy, 2′-methoxyethoxy (MOE), and 2′-fluoro are all commonly deployed substitutions in therapeutic oligonucleotides (Figure 1B) which impart both resistance to nucleases and improvements in RNA binding affinity.

The preparation of 2′-modified nucleosides often requires multi-step synthesis, with extensive protection/deprotection sequences or generates complex mixtures of regioisomers, making the synthesis of these molecules challenging. We considered a biocatalytic route to access 2′-modified nucleosides since enzymes are often able to carry out complex transformations in an atom efficient manner, under relatively mild conditions, and without the use of protecting groups. Moreover, since different biocatalysts often function under similar temperature and pH in aqueous solution, they can be combined into multi-enzyme cascades allowing for two or more reactions to be carried out in one pot without the need to isolate intermediates. This approach leads to a reduction in waste from intermediate/product isolation and also allows reactions to be coupled together, thereby overcoming potential thermodynamic barriers of individual reactions.[5] These advantages have resulted in biocatalysis becoming increasingly deployed in the synthesis of active pharmaceutical ingredients (APIs).[6]

Recent reports have shown that enzymes from the nucleoside salvage pathway[7] (Figure 1C) provide a useful platform for the generation of functionalized nucleosides. In the retrosynthetic direction, this pathway is responsible for the degradation of 2′-deoxynucleosides. Starting with a nucleoside phosphorylase (NP) the nucleobase is removed resulting in the generation of a C-1 phosphate sugar (Figure 1C). In the following step phosphopentomutase (PPM) catalyses transfer of the phosphate group from the 1' position to the 5' position, and finally deoxyribose-5-phosphate aldolase (DERA) breaks down the sugar into acetaldehyde and D-glyceraldehyde-3-phosphate via a retroaldol reaction. As all three enzymes catalyze reversible reactions

they can be combined in the synthetic direction to generate nucleosides. An excellent example of this approach is the synthesis of the antiviral nucleoside analogue islatravir (Figure 1D)[8]. There are additional examples of enzymes on this pathway being used to generate nucleoside analogues such as dideoxyinosine[9] and molnupiravir[10]. Furthermore, there are also examples of other biocatalysts being employed in the synthesis of nucleosides, nucleotide analogues[11,12] and oligonucleotides[13].

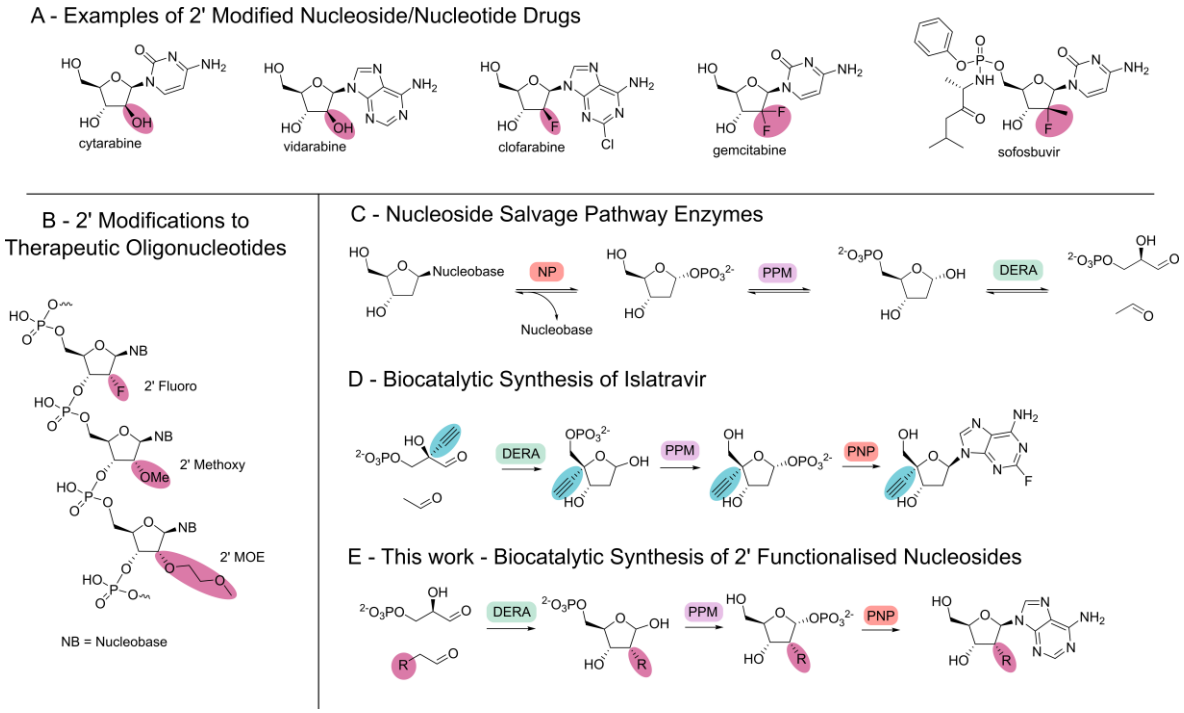


Figure 1: A – Examples of 2′-modified nucleoside and nucleotide analogues. B – Examples of common 2′-modifications made to therapeutic oligonucleotides. C- Nucleoside salvage pathway enzymes break down and recycle nucleosides. D – Biocatalytic synthesis of islatravir employing the salvage pathway enzymes. E – This work - proposed synthesis of 2′-functionalized nucleosides using non-natural aldol donors resulting in modifications (R group) being installed at the 2′-position.

Clearly to exploit this pathway for the synthesis of 2′ modified nucleosides, all enzymes in the pathway would need to possess sufficient promiscuity to tolerate a wide range of 2′-modifications. The initial challenge posed by this synthetic cascade is the limited substrate scope of DERA. While PPM and PNP enzymes have been shown to tolerate a small number of 2′-modifications[14,15], thus far the approach has been limited to 2′-deoxy nucleosides when starting from DERA, with acetaldehyde as the donor. Engineering DERA would allow the synthesis of 2-modified D-ribose-5-phosphate analogues starting from D-glyceraldehyde-3-phosphate and an aldehyde donor. The products could then be further modified with the remaining enzymes in the pathway to generate 2′-functionalized nucleosides (Figure 1E).

In general aldolases tend to have a more limited substrate scope for the donor rather than the acceptor, consequently they are often classified based on their donor selectivity[16] with the acetaldehyde-dependent enzyme DERA being a prime example of this [17]. DERA catalyses the reversible aldol condensation of acetaldehyde and D-glyceraldehyde-3-phosphate and is typically only able to accept a small range of donor analogues, such as propanal or glycolaldehyde, often coming at the cost of significantly reduced activity[18]. A recent report demonstrated that screening a diverse panel of DERAs resulted in activity toward six different aldehydes[19]. Recent work has also involved engineering DERA to expand the activity of the wild-type enzyme, resulting in variants able to catalyze Michael addition of nitromethane to α,β-unsaturated ketones[20].

In this work we demonstrate that mutations in the active site of *E. coli* DERA result in significantly broadened donor substrate scope allowing for the synthesis of a wide range of 2-modified-5-phosphate sugars. We have used these DERA variants to develop a one-pot synthesis of 2-functionalized D-ribose and L-lyxose-5-phosphate analogues with a range of functional groups including fluoro, MOE and OBn. Finally addition of PPM and PNP results in a one

pot synthesis of adenosine analogues functionalized at the 2′-position with OH, araOH, Me and F.

Broadening the Donor Substrate Scope of DERA

Previous work suggested that the restriction in donor substrate scope of DERA was possibly due to steric constraints in the active site[17]. Moreover, the substrate scope of fructose-6-phosphate aldolase (FSA) was greatly expanded by a "minimalist protein engineering" approach increasing space in the active site[21]. We initially decided to focus on a similar strategy, generating mutations in the active site of DERA with the intention of enlarging the binding pocket for the donor substrate. The previously reported crystal structure of *E. coli* DERA (PDB =1JCL)[18], which contains the linear aldol product bound to the catalytic lysine residue, was used to identify six active site residues in close proximity to the 2-position of the final product. These residues were mutated to alanine and the resulting variants screened for activity against a broad panel of aldehyde donor substrates (Figure 2).

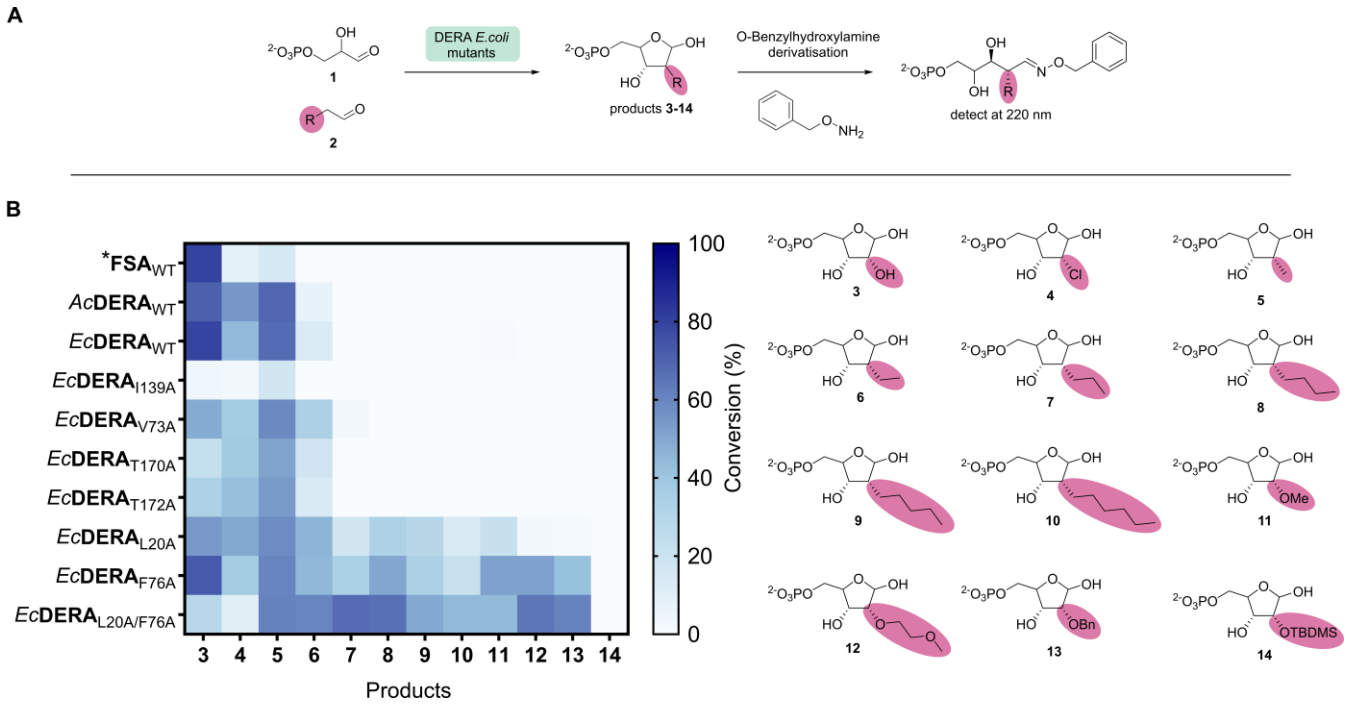


Figure 2: A - Assay of DERA variants towards a range of aldehyde donors using DL-G3P as acceptor. Conversion was monitored by HPLC using UV at 220nm after derivatisation with *O*-benzylhydroxylamine. B - Heatmap showing the average conversions of triplicate repeats for each of the aldolase variants with a range of aldehyde donors generating products (3-14).

*For FSA the opposite stereochemistry at C2 is generated to the stereochemistry shown.

Alongside the *Ec*DERA WT and variants we also screened the DERA from *A. chlorophenolicus* (*Ac*DERA$_{WT}$), which was recently shown to have the broadest donor substrate scope of a panel of DERAs[19], and fructose-6-phosphate aldolase (EcFSA$_{WT}$). The panel of aldehyde donors included halogen (Cl)-substituted, two functional groups found in therapeutic oligonucleotides (OMe and MOE) and two typical protecting groups used in nucleoside chemistry (OBn and OTBDMS). These aldehydes were all screened with DL-glyceraldehyde-3-phosphate (DL-G3P, **1**) as the acceptor.

The *Ec*DERA L20A and F76A variants showed significant improvement over both *Ec*DERA$_{WT}$ and FSA$_{WT}$. For both variants, conversions ranging from 34-91% were obtained towards all substrates screened with the exception of *O*-TBDMS which showed no activity towards any of the enzyme variants. For the majority of substrates *Ec*DERA$_{F76A}$ gave higher conversion than *Ec*DERA$_{L20A}$. The double alanine mutation at these two positions *Ec*DERA$_{L20A/F76A}$ showed increased conversion with some of the larger aldehyde substrates (MOE, OBn, heptanal) at the cost of decreased conversion with a selection of other functional groups (OH, Cl, OMe).

Computational Rationalization of Substrate Scope

To rationalize the increase in activity of the three active variants, models were generated *in silico* and compared to

that of the *E. coli* wild type DERA. The binding pockets volumes for all four enzymes were analyzed using fpocket[22] and compared. As predicted, mutation of the selected residues to alanine resulted in an increase in the volume of the active site for all variants. The WT enzyme has a binding pocket volume of 317 Å^3, this increases to 416 Å^3 for L20A, 447 Å^3 for F76A and 529 Å^3 for L20A/F76A. These increases in binding pocket size seem to correlate with the trend of activity observed, with only smaller donor molecules like glycolaldehyde and propanal being active with the WT, whereas the variants with significantly increased binding pocket volumes are able to accommodate larger substrates. Particularly bulky substrates such as benzoxyacetaldehyde or methoxy(ethoxy)acetaldehyde also show better activity with the larger binding pockets of F76A and L20A/F76A.

To determine whether the expected products are able to fit inside the active site and occupy a reasonable binding mode an example product functionalized with 2-MOE was representatively covalently docked into the three variants using autodock[23] and compared to WT ligand in the crystal structure (Figure 3). For all three variants the 2-MOE functionalized products were able to occupy a binding mode similar to that of the wild type product (2-deoxy) and in all cases the larger functional group added at the 2-position was able to occupy the newly generated binding pockets. This further supports the hypothesis that an increase in the volume of the active site is responsible for the increase in substrate scope.

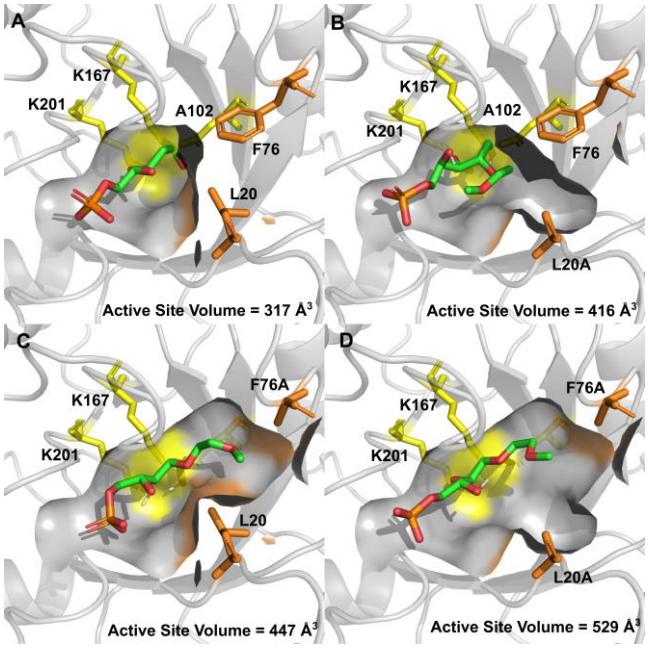
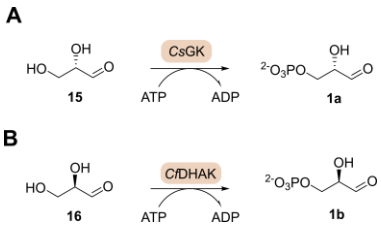


Figure 3: The binding pockets for DERA and the three variants. Each of the variants are shown with the 2-MOE substrate covalently docked into the active site. Catalytic residues are shown in yellow, residues targeted for mutation are shown in orange, binding pocket sizes calculated by fpocket are shown for each enzyme. A – Wild Type *E.coli* crystal structure from 1JCL. B – L20A variant. C – F76A variant. D- L20A/F76A variant.

Synthesis of 2-Modified D-Ribose-5-Phosphates and L-Lyxose-5-Phosphates

Glyceraldehyde-3-phosphate is both expensive and relatively unstable, degrading to methyl glyoxal and free phosphate at neutral pH[24]. One way to offset these problems is to generate D- or L-glyceraldehyde-3-phosphate from D- or L- glyceraldehyde respectively *in situ* using a kinase, an approach that has been employed previously when using FSA aldolase [25,26]. L-G3P (**1a**) can be generated from L-glyeraldehyde (**15**) using glycerokinase from *Cellumonas. sp*[27] (scheme 1A) and D-G3P (**1b**) can be generated from D-glyceraldehyde (**16**) using dihydroxyacetone kinase from *Citrobacter freundii*[28] (Scheme 1B). These kinase enzymes were therefore combined into two step cascades with the active *Ec*DERA variants.

Scheme 1



Scheme 1: A – Generation of L G3P (**1a**) from L-glyceraldehyde (**15**) via glycerokinase from *Cellumonas sp* B- Generation of D-G3P (**16**) from D-glyceraldehyde (**1b**) via dihydroxyacetonekinase from *Citrobacter freundii*

For both cascades, enzyme loadings were optimized (supplementary 8.2.1) and the cascades were then screened towards a selection of the substrates from the initial panel (Figure 4). In general a similar pattern of activity for the WT and the two variants was seen compared to the initial screen. However, for both cascades, when starting from enantiomerically pure glyceraldehyde only a single product peak was observed in the HPLC traces (compared to two peaks for DL-G3P) which suggest the formation of a single product diastereomer at the 2-position as desired (supplementary 11.15.1).

Interestingly when comparing the two reactions, the cascade starting from L-glyceraldehyde performed better than with D-glyceraldehyde. One possible explanation for this difference is due to the presence of native *E. coli* enzymes co-expressed with the *Ec*DERA variants. For example, D-G3P is a substrate for triose phosphate isomerase (TIM), a highly active enzyme that is rate-limited only by diffusion

of G3P into the active site[29]. Even small quantities of this enzyme present from *E. coli* remaining after purification are likely to be sufficent to isomerize a significant proportion of D-G3P into dihydroxyacetone phosphate (DHAP, **20**) thus lowering conversions (Scheme 3). Despite this side reaction, the D-ribose-5-phosphate analogues were all generated with conversions ranging from 22-66% for all the substrates screened.

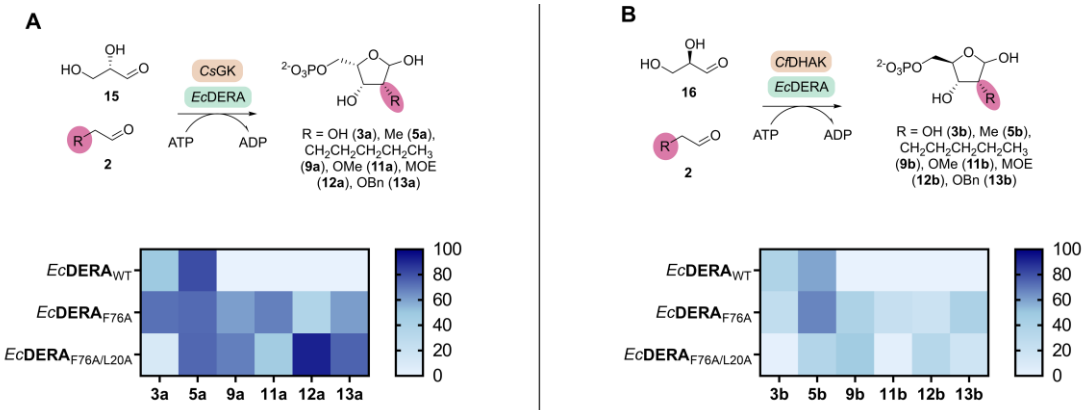
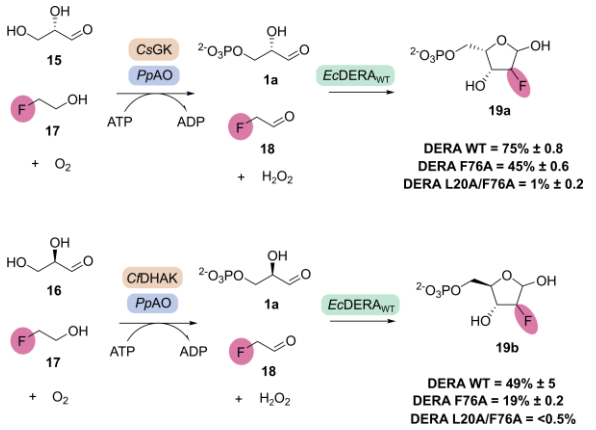


Figure 4: A Synthesis of L-lyxose-5-phosphate analogues via two step biocatalytic cascades using *Cs*GK and *Ec*DERA variants. B Synthesis of D-ribose-5-phosphate analogues via two step biocatalytic cascades using *Cf*DHAK and *Ec*DERA variants. Heatmaps show average conversions for triplicate repeats.

Synthesis of 2-Fluoro Pentose-5-Phosphates

One modification of particular interest for therapeutic oligonucleotides is 2-fluoro, which is used in several currently approved therapeutic oligonucleotides (lumasiran, inclisiran, etc.)[30]. Synthesis of the 2-F modified R5P requires fluoroacetaldehyde **18** as the donor substrate. In view of the difficulty of handling aldehyde **18** we decided instead to generate it *in situ* using an oxidase.

Initial work showed that fluoroacetaldehyde **18** could be generated from fluoroethanol **17** using methanol oxidase from *Pichia pastoris* - *Pp*AO$_{WT}$ (supplementary 9.1). This system was then combined with both kinase enzymes, *Cf*DHAK or *Cs*GK, and screened with the *Ec*DERA$_{WT}$ alongside the two best performing variants (Scheme 2).



Scheme 2: A – Synthesis of 2-F-L-lyxose-5-phosphate via one pot biocatalytic cascade using *Pp*AO, *Cs*GK and *Ec*DERA variants. B – Synthesis of 2-F D-ribose-5-phosphate via one pot biocatalytic cacade using *Pp*AO, *Cf*DHAK and *Ec*DERA variants.

Scheme 2

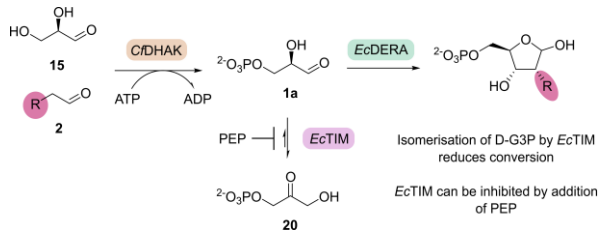
Combining the oxidase, kinase and aldolase enabled the synthesis of fluorinated products **19a** and **19b** (Scheme 2). In contrast to the previous aldehyde substrates, with fluoroacetaldehyde **17** the WT enzyme gave higher conversion than both the F76A and L20A/F76A variants presumably due to the small size of the fluorine substituent. The HPLC traces for this cascade (supplementary 9.2.2, 9.3.2) contained multiple peaks, suggesting the presence of a

mixture of diastereomers at the 2-position. For **19a** ratio = 15:85 for **19b** ratio = 50:50).

Inhibition of TIM Activity by Addition of phosphoenolpyruvate

As highlighted above, the isomerization of D-G3P **1a** to DHAP **20** catalyzed by TIM presented a problem for the the cascade to generate D-ribose-5-phosphate analogues. (Scheme 3). Phosphoenolpyruvate (PEP) has been previously shown to inhibit TIM[31] and can also be used as a phosphate donor substrate for pyruvate kinase in order to implement an ATP recycling system. We reasoned that addition of phosphoenolpyruvate to the cascade could serve to both recycle ATP for the kinase step as well as increase overall conversion to product by inhibiting the formation of DHAP via TIM.

Scheme 3



Scheme 3: Isomerization of D-G3P (**1a**) into DHAP (**19**) via TIM reduces conversion.

Reactions were screened using the F76A variant, both with and without the PEP/PK recycling system present (Figure 5). A 2-fold excess of PEP was added to ensure efficient inhibition of TIM throughout the reaction while still enabling recycling of ATP. Addition of the PEP/PK recycling system resulted in an increase in conversion for all substrates screened, more closely matching conversions seen for the L cascade. In addition to the increase in conversion to product, the amount of DHAP side product also decreased (supplementary 10.4).

Addition of the PEP/PK recycling system enables not only the use of catalytic amounts of ATP but also significantly reduces the effects of TIM on the cascade, enabling the generation of D-R5P analogues in higher conversions.

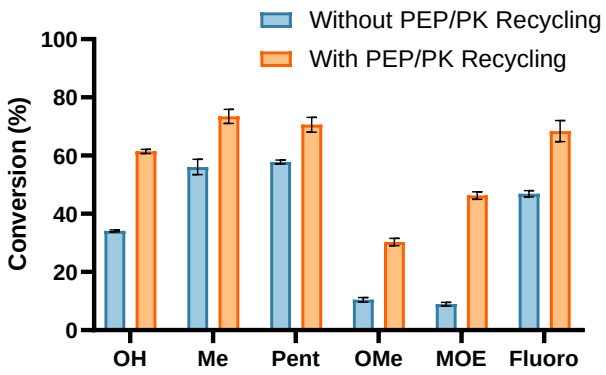


Figure 5: Conversions for a range of D-ribose-5-phosphate analogues with and without the addition of PEP/PK recycling system. Error bars represent the standard deviation of triplicate repeats.

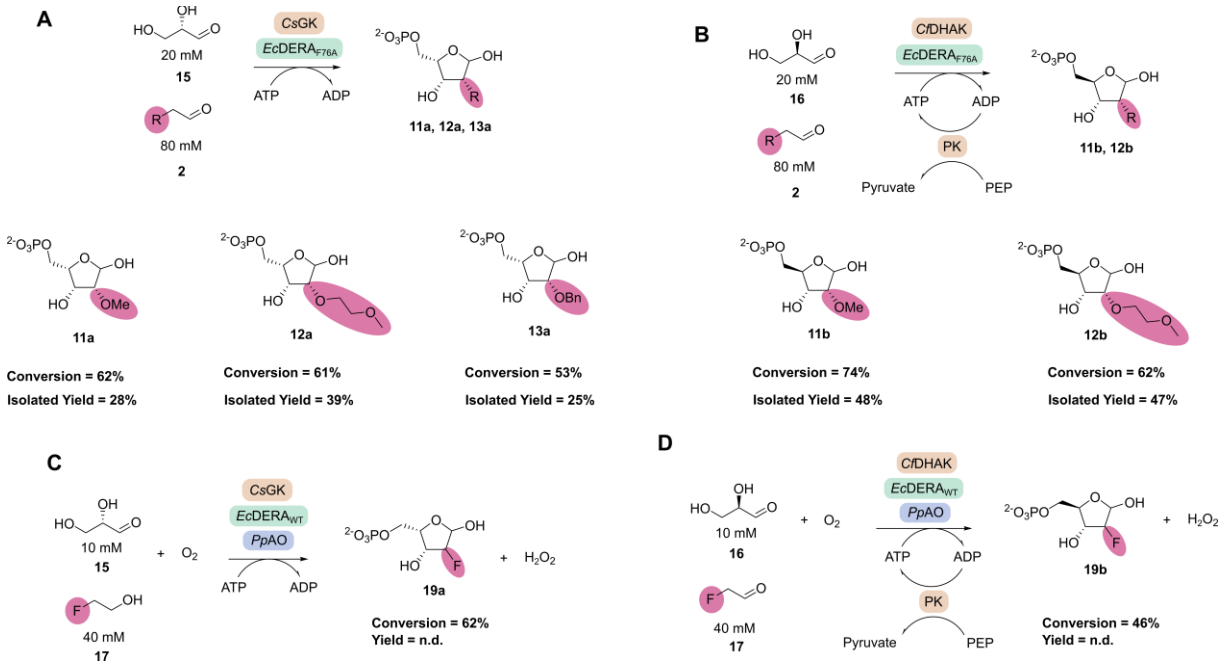
Semi-Preparative Scale Synthesis of D-Ribose-5-Phosphate and L-Lyxose-5-Phosphate Analogues

Following the development of successful analytical scale reactions a number of key substrates were then carried through on a semi-preparative scale using *Ec*DERA$_{F76A}$ with a substrate loading of 20 mM glyceraldehyde.

For the synthesis of L-lyxose-5-phosphates **11a-13a** (Scheme 4A) the ATP recycling system was omitted since the PEP based inhibition of TIM was not needed for L-G3P. Reactions were instead carried out at 20 mM substrate loading with stoichiometric amounts of ATP. The initial substrate chosen to analyse was the 2-OMe, due to its diagnostic singlet peak in ^{1}H NMR. Conditions for the scale-up reaction of L-lyxose-5-phosphates were optimized via design of experiment (DOE) (supplementary 11.2.1). The most significant factors were found to be both aldolase and donor substrate concentrations.

For the synthesis of D-ribose-5-phosphates **11b,12b** (Scheme 4B) reactions were again carried out at 20 mM substrate loading with the PEP/PK recycling system. The reaction conditions were again optimized by DOE, (supplementary 11.2.2) For all reactions of both cascades products were purified by anion exchange chromatography.

Scheme 4



Scheme 4: Semi-preparative scale synthesis of L-lyxose-5-phosphate analogues via a two step cascade with *Cs*GK and *Ec*DERA$_{F76A}$. B – Semi-preparative scale of D-ribose-5-phosphate analogues via two step cascade with *Cf*DHAK and *Ec*-DERA$_{F76A}$. C – Semi-preparative scale synthesis of 2-F-L-lyxose-5-phosphate D – Semi-preparative scale synthesis of 2-F-D-ribose-5-phosphate. Products purified by anion exchange chromatography and isolated as ammonium salts.

For the synthesis of L-lyxose-5-phosphate analogues, 2-OMe and MOE products **11a** and **12a** were generated in good conversions (62%, 61% respectively) and isolated in reasonable yields (28%, 39% respectively). For the benzylated product **13a**, conversion of 53% was obtained alongside an isolated yield of 25%.

For the D-cascade, 2-OMe and 2-MOE products **11b** and **12b** were obtained in high analytical yields (74%, 62% respectively) and good isolated yields (46%, 47%). Interestingly for **11b**, the scale up analytical yield was higher than any of the previous optimization reactions carried out in 200 µL volumes. When compared to the L-lyxose-5-phosphate products, D-ribose-5-phosphate analogues contained a small amount (<10%) of pyruvate present in the purified sample, which is a by-product of the pyruvate kinase recycling system.

Preparative scale syntheses of the 2-F products were also carried out (Scheme 4C/D). For these reactions several alterations had to be made to the reaction conditions. To overcome issues with oxygen limitation posed by the oxidase, the concentration of substrate was decreased to 10 mM and the reaction was carried out as a series of smaller volume biotransformations (10 x 500 µL). For 2-F-L-lyxose-5-phosphate **19a** a conversion of 62% was obtained whereas the 2-F-D-ribose-5-phosphate **19b** was generated with a conversion of 46%.

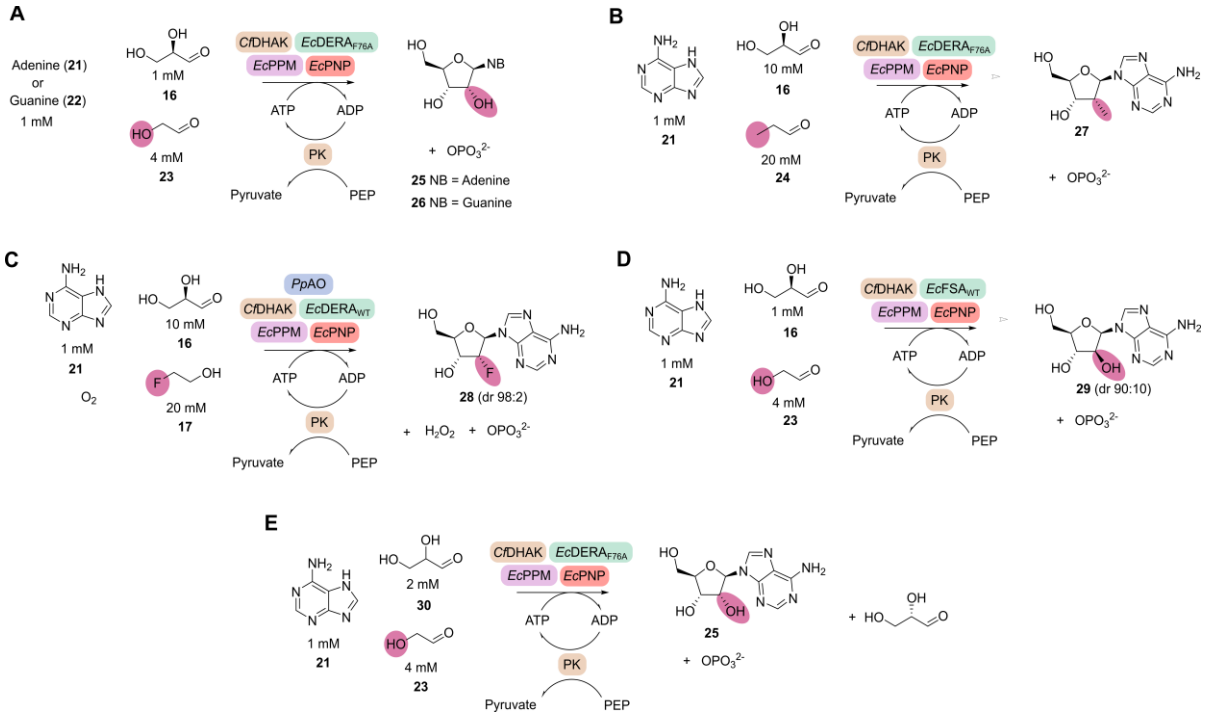
Mass spectrometry (ESI-) analysis confirmed the presence of desired $[M-H]^-$ ions in both products. While HPLC analysis seemed to show a mixture of diastereomers in a ratio of 80:20 for 2-F-L-lyxose-5-phosphate and 50:50 for 2-F-D-ribose-5-phosphate, the 1H NMR spectra showed a more complex mixture of products. The fluorine NMR analysis showed the expected four diastereomers (supplementary 11.3.4, 11.4.4), alongside multiple unknown fluorine peaks. These extra peaks may correspond to diastereomers at the 3' position or small amounts of fluorinated side products that were not eliminated during purification.

The stereochemistry of the isolated products was determined by comparison to previous work and by the use of ^{1}H-^{1}H NOESY NMR (supplementary 11.5.2). All products were assigned the 2*R*, 3*R* stereochemistry, in agreement with previous work.[19]

Biocatalytic Synthesis of 2′-Functionalised Adenosine and Guanosine Analogues

Finally, to demonstrate the application of this cascade to generate nucleosides functionalized at the 2′-position the aldolase step was combined with phosphopentomutase (*Ec*PPM$_{WT}$) and purine nucleoside phosphorylase (*Ec*PNP$_{WT}$) to synthesize guanosine and adenosine analogues in a one pot reaction (Scheme 5).

Scheme 5



Scheme 5: Synthesis of 2′-purine analogues via a one pot multi-enzyme cascade. A - synthesis of 2′-OH adenosine/guanosine B - Synthesis of 2′-Me adenosine/guanosine C – Synthesis of 2′-fluoro adenosine/guanosine. D – Replacement of DERA aldolase with FSA allows synthesis of 2′-ara-OH adenosine (vidarabine). E – Synthesis of adenosine starting from DL-glyceraldehyde.

Initial studies showed that the five enzyme cascade (Scheme 5A) was able to generate the ribonucleoside adenosine **25** in good conversions (Table 1). Furthermore as *Cf*DHAK$_{WT}$ is enantiospecific, D-glyceraldehyde can also be substituted with 2 equivalents of DL-glyceraldehyde (Scheme 5E). While the 2′-substrate scope of wild-type PPM and PNP is likely to be restricted we found that in addition to 2′-OH both 2′-Me **27** and 2′-F **28** substitutions were accepted, by both enzymes, albeit with lower conversions (Table 1). Alongside this, a 10x excess of the D-glyceraldehyde starting material was required to give reasonable conversions. This is presumably due to the lower activity of the PPM and PNP steps, alongside the instability of the phosphorylated intermediates, particularly the corresponding 1' phosphate. The higher activity of the cascade towards the wild-type substrate 2′-OH suggests that, in principle, this excess would not be required for 2′-Me and 2′-F analogues if more active variants of PPM and PNP were obtained.

For the 2′-Me adenosine **27** changing the aldolase from *Ec*DERA$_{WT}$ to *Ec*DERA$_{F76A}$ gave an increase in conversion from 17% to 38%. DOE was used to determine the effects of enzyme loading on conversion for the three main enzymes (DERA, PPM, PNP) (supplementary 13). This revealed that the concentration of PNP was by far the most limiting of the three enzymes. By increasing the concentration of PNP in these reactions an increase in conversions for both 2′-Me (38% to 65%) and 2′-F (6% to 20%) was observed (Table 1). Addition of a recently published PNP variant which was engineered to give a small increase in activity to 2′-F adenosine[32] resulted in a further increase in conversion to 28% (Table 1). The lower conversions seen for the fluorinated analogues are in part due to a large proportion of this product being converted to 2′-F-inosine (presumably by contaminant adenosine deaminase) this issue is compounded by the relatively large amounts of protein required and highlights the need for further engineering, particularly of the PNP enzyme. In addition to 2′-OH adenosine 2′-OH guanine was synthesized using the optimal conditions from the adenosine reaction (Table 1). While some product was generated, the conversion for the synthesis of guanosine is notably lower than for adenosine, this is likely due in part to its poorer solubility under the reaction conditions.

As the stereochemistry of the 2′-position is set by the aldolase used, changing to a different aldolase allows for alteration of this stereochemistry. Thus replacement of DERA by *Ec*FSA$_{WT}$ [33] resulted in the synthesis of the nucleoside analogue vidarabine (2′-ara-OH adenosine, **29**) (Table 1). Vidarabine was generated in a diastereomeric ratio of 90:10, vidarabine:adenosine.

Table 1. Biocatalytic synthesis of 2′-functionalized nucleoside via one pot cascades

Entry	2′-Modification	Nucleobase	Sugar Equivalents[a]	Aldolase loading (variant)	PPM Loading (mg/ml)	PNP Loading (mg/ml)	Conversion (%)
1	OH	Adenine	1x	2 mg/ml (*Ec*-DERA$_{WT}$)	0.1	0.1	47 ± 1
2	OH	Adenine	1x	2 mg/ml (*Ec*-DERA$_{F76A}$)	0.1	0.1	61 ± 1
3[b]	OH	Adenine	2x	2 mg/ml (*Ec*-DERA$_{F76A}$	0.1	0.1	64 ± 1
4	Me	Adenine	1x	2 mg/ml (*Ec*-DERA$_{WT}$)	1	2	n.d.
5	Me	Adenine	10x	2 mg/ml (*Ec*-DERA$_{WT}$)	1	2	17 ± 1
6	Me	Adenine	10x	2 mg/ml (*Ec*-DERA$_{F76A}$)	1	2	38 ± 1
7	Me	Adenine	10x	2 mg/ml (*Ec*-DERA$_{F76A}$)	1	4	65 ± 8
8	F	Adenine	1x	2 mg/ml (*Ec*-DERA$_{WT}$)	1	2	n.d
9	F	Adenine	10x	2 mg/ml (*Ec*-DERA$_{WT}$)	1	2	6 ± 1
10	F	Adenine	10x	2 mg/ml (*Ec*-DERA$_{WT}$)	1	4	20 ± 1
11[c]	F	Adenine	10x	2 mg/ml (*Ec*-DERA$_{WT}$)	1	4	28 ± 1
12	OH	Guanine	1x	2 mg/ml (*Ec*-DERA$_{F76A}$)	0.1	0.1	24 ± 1
13	Ara-OH	Adenine	1x	1 mg/ml (*Ec*FSA$_{WT}$)	0.1	0.1	32 ± 1

a: sugar equivalents as determined by the equivalents of D-Glyceraldehyde relative to nucleobase

b: reaction carried out with 2 mM DL-Glyceraldehyde as opposed to 1 mM D-Glyceraldehyde, conversion reported with respect to limiting nucleobase.

c: reaction repeated with best performing PNP variant from recently published paper[32]

Finally, to characterize the 2′-OH, 2′-araOH, 2′-Me and 2′-F adenosine analogues by ^{1}H NMR 5 ml scale reactions were carried out and the products were purified by semi-preparative HPLC. For adenosine the ^{1}H NMR spectra obtained was identical to the commercial standard. For the 2′-OH and 2′-Me analogues, the 1H NMR spectra confirmed the presence of a single diastereomer at the 2' postion. For vidarabine, the 1H NMR spectrum was identical to the commercial standard. For 2′-F adenosine **28**, despite the aldolase step generating an epimeric mixture of 2′-F diastereomers, the 2′-F adenosine product was synthesized in a diasteromeric ratio of 98:2 favouring the desired "down" stereochemistry of the fluorine. These isomers were unable to be separated by semi-prep HPLC and were isolated together. This observation suggests that while the aldolase exhibits poor stereochemical control, one or both of the final two enzymes in the cascade are stereoselective for the desired fluorine diastereomer. Alongside the desired 2′-F adenosine we were also able to isolate the 2′-F inosine side product (supplementary 14.5), thereby demonstrating the potential to generate 2′ -functionalized inosine anlogues by addition of a deaminase enzyme to the cascade.

Conclusions

By targeting the active site of *Ec*DERA$_{WT}$, simple mutations (F76A, L20A) have been shown to significantly increase donor substrate scope and allow for the synthesis of a broad range of D-pentose-5-phosphate analogues functionalized at the 2-position. Addition of a kinase to generate glyceraldehyde-3-phosphate *in situ* resulted in the syn-

thesis of 2-functionalised D-ribose-5-phosphate and L-lyxose-5-phosphate analogues via a two-step cascade. Furthermore, addition of an alcohol oxidase led to the formation of fluoroacetaldehyde *in situ* and subsequent generation of 2-F analogues via a three step one pot reaction. To characterize the various pentose-5-phosphate products semi-preparative biotransformations were carried out. This resulted in the synthesis of 2-OMe, MOE, OBn and F analogues in good conversions and isolated yields. With the exception of fluorine, all products were generated as single diastereomers at 2- highlighting the excellent stereoselectivity of DERA.

To demonstrate the utility of the purine nucleoside salvage pathway enzymes for synthesizing nucleosides functionalized at the 2′-position the *Ec*DERA$_{F76A}$ variant was combined with *Ec*PPM$_{WT}$ and *Ec*PNP$_{WT}$. This allowed the synthesis of the ribonucleosides adenosine and guanine as well as 2′-Me and 2′-F adenosine analogues. These nucleoside analogues can be generated in a one pot reaction starting from relatively simple starting materials. By changing the aldolase used from *Ec*DERA$_{F76A}$ to *Ec*FSA$_{WT}$, stereochemical control can be imparted at the 2′-position allowing for the synthesis of the nucleoside analogue vidarabine.

While PPM and PNP have been shown to tolerate a small range of 2′-functionalization such as 2′-NH_2, araOH and 2′-F, cascades utilizing DERA have thus far been limited to the synthesis of 2′-deoxy nucleosides. This work shows that a much broader range of 2′-functionalization is possible via engineering of the aldolase. We consider this to be the first step towards the repurposing of this pathway towards generating a much broader range of nucleoside analogues. Future work engineering PPM and PNP for increased substrate scope at the 2′-position will hopefully increase the range of nucleosides able to be generated via this approach.

ACKNOWLEDGMENT

This project has received funding from the EPSRC, BBSRC and AstraZeneca as part of the Prosperity Partnership grant (EP/S005226/1). The authors gratefully acknowledge the contributions of Rhys Barker, Sanaz Ahmadipour, Marielle Lemaire and collaborators, and Matt Cliff.

ABBREVIATIONS

ADP = Adenosine Diphosphate

AO = Alcohol Oxidase

ATP = Adenosine Triphosphate

DERA = 2-deoxy-d-ribose-5-phosphate aldolase

DHAK = Dihydroxyacetone kinase

DOE = Design of Experiment

ESI = Electrospray Ionisation

FSA = Fructose-6-Phosphate aldolase

G3P = Glyceraldehyde-3-phosphate

GK = Glycerokinase

HPLC = High Performance Liquid Chromatography

MOE = Methoxy(ethoxy)

NMR = Nuclear Magnetic Resonance

PPM = Phosphopentomutase

PNP = Purine Nucleoside Phosphorylase

R5P = Ribose-5-Phosphate

TIM = Triosephosphate Isomerase

WT = Wild Type

REFERENCES

(1) De Jonghe, S.; Herdewijn, P. An Overview of Marketed Nucleoside and Nucleotide Analogs. *Curr. Protoc.* **2022**, *2* (3). https://doi.org/10.1002/cpz1.376.

(2) *Advances in Nucleic Acid Therapeutics*; Agrawal, S., Gait, M. J., Eds.; The Royal Society of Chemistry, 2019. https://doi.org/10.1039/9781788015714.

(3) Blanco, M.-J.; Gardinier, K. M. New Chemical Modalities and Strategic Thinking in Early Drug Discovery. *ACS Med. Chem. Lett.* **2020**, *11* (3), 228–231. https://doi.org/10.1021/acsmedchemlett.9b00582.

(4) Smith, C. I. E.; Zain, R. Therapeutic Oligonucleotides: State of the Art. *Annu. Rev. Pharmacol. Toxicol.* **2019**, *59* (1), 605–630. https://doi.org/10.1146/annurev-pharmtox-010818-021050.

(5) Bell, E. L.; Finnigan, W.; France, S. P.; Green, A. P.; Hayes, M. A.; Hepworth, L. J.; Lovelock, S. L.; Niikura, H.; Osuna, S.; Romero, E.; Ryan, K. S.; Turner, N. J.; Flitsch, S. L. Biocatalysis. *Nat. Rev. Methods Prim.* **2021**, *1* (1), 46. https://doi.org/10.1038/s43586-021-00044-z.

(6) France, S. P.; Lewis, R. D.; Martinez, C. A. The Evolving Nature of Biocatalysis in Pharmaceutical Research and Development. *JACS Au* **2023**, *3* (3), 715–735. https://doi.org/10.1021/jacsau.2c00712.

(7) Tozzi, M. G.; Camici, M.; Mascia, L.; Sgarrella, F.; Ipata, P. L. Pentose Phosphates in Nucleoside Interconversion and Catabolism. *FEBS J.* **2006**, *273* (6), 1089–1101. https://doi.org/10.1111/j.1742-4658.2006.05155.x.

(8) Huffman, M. A.; Fryszkowska, A.; Alvizo, O.; Borra-Garske, M.; Campos, K. R.; Canada, K. A.; Devine, P. N.; Duan, D.; Forstater, J. H.; Grosser, S. T.; Halsey, H. M.; Hughes, G. J.; Jo, J.; Joyce, L. A.; Kolev, J. N.; Liang, J.; Maloney, K. M.; Mann, B. F.; Marshall, N. M.; McLaughlin, M.; Moore, J. C.; Murphy, G. S.; Nawrat, C. C.; Nazor, J.; Novick, S.; Patel, N. R.; Rodriguez-Granillo, A.; Robaire, S. A.; Sherer, E. C.; Truppo, M. D.; Whittaker, A. M.; Verma, D.; Xiao, L.; Xu, Y.; Yang, H. Design of an in Vitro Biocatalytic Cascade for the Manufacture of Islatravir. *Science (80-.).* **2019**, *366* (6470), 1255–1259. https://doi.org/10.1126/science.aay8484.

(9) Birmingham, W. R.; Starbird, C. A.; Panosian, T. D.; Nannemann, D. P.; Iverson, T. M.; Bachmann, B. O. Bioretrosynthetic Construction of a Didanosine Biosynthetic Pathway. *Nat. Chem. Biol.* **2014**, *10* (5), 392–399. https://doi.org/10.1038/nchembio.1494.

(10) McIntosh, J. A.; Benkovics, T.; Silverman, S. M.; Huffman, M. A.; Kong, J.; Maligres, P. E.; Itoh, T.; Yang, H.; Verma, D.; Pan, W.; Ho, H.-I.; Vroom, J.; Knight, A. M.; Hurtak, J. A.; Klapars, A.; Fryszkowska, A.; Morris, W. J.; Strotman, N. A.; Murphy, G. S.; Maloney, K. M.; Fier, P. S. Engineered Ribosyl-1-Kinase Enables Concise Synthesis of Molnupiravir, an Antiviral for COVID-19. *ACS Cent. Sci.* **2021**, *7* (12), 1980–1985. https://doi.org/10.1021/acscentsci.1c00608.

(11) McIntosh, J. A.; Liu, Z.; Andresen, B. M.; Marzijarani, N. S.; Moore, J. C.; Marshall, N. M.; Borra-Garske, M.; Obligacion, J. V.; Fier, P. S.; Peng, F.; Forstater, J. H.; Winston, M. S.; An, C.; Chang, W.; Lim, J.; Huffman, M. A.; Miller, S. P.; Tsay, F.-R.; Altman, M. D.; Lesburg, C. A.; Steinhuebel, D.; Trotter, B.

W.; Cumming, J. N.; Northrup, A.; Bu, X.; Mann, B. F.; Biba, M.; Hiraga, K.; Murphy, G. S.; Kolev, J. N.; Makarewicz, A.; Pan, W.; Farasat, I.; Bade, R. S.; Stone, K.; Duan, D.; Alvizo, O.; Adpressa, D.; Guetschow, E.; Hoyt, E.; Regalado, E. L.; Castro, S.; Rivera, N.; Smith, J. P.; Wang, F.; Crespo, A.; Verma, D.; Axnanda, S.; Dance, Z. E. X.; Devine, P. N.; Tschaen, D.; Canada, K. A.; Bulger, P. G.; Sherry, B. D.; Truppo, M. D.; Ruck, R. T.; Campeau, L.-C.; Bennett, D. J.; Humphrey, G. R.; Campos, K. R.; Maddess, M. L. A Kinase-CGAS Cascade to Synthesize a Therapeutic STING Activator. *Nature* **2022**, *603* (7901), 439–444. https://doi.org/10.1038/s41586-022-04422-9.

(12) Burke, A. J.; Birmingham, W. R.; Zhuo, Y.; Thorpe, T. W.; Zucoloto da Costa, B.; Crawshaw, R.; Rowles, I.; Finnigan, J. D.; Young, C.; Holgate, G. M.; Muldowney, M. P.; Charnock, S. J.; Lovelock, S. L.; Turner, N. J.; Green, A. P. An Engineered Cytidine Deaminase for Biocatalytic Production of a Key Intermediate of the Covid-19 Antiviral Molnupiravir. *J. Am. Chem. Soc.* **2022**, *144* (9), 3761–3765. https://doi.org/10.1021/jacs.1c11048.

(13) Moody, E. R.; Obexer, R.; Nickl, F.; Spiess, R.; Lovelock, S. L. An Enzyme Cascade Enables Production of Therapeutic Oligonucleotides in a Single Operation. *Science (80-.).* **2023**, *380* (6650), 1150–1154. https://doi.org/10.1126/science.add5892.

(14) Kamel, S.; Thiele, I.; Neubauer, P.; Wagner, A. Thermophilic Nucleoside Phosphorylases: Their Properties, Characteristics and Applications. *Biochimica et Biophysica Acta - Proteins and Proteomics*. Elsevier B.V. February 1, 2020, p 140304. https://doi.org/10.1016/j.bbapap.2019.140304.

(15) Fateev, I. V.; Kostromina, M. A.; Abramchik, Y. A.; Eletskaya, B. Z.; Mikheeva, O. O.; Lukoshin, D. D.; Zayats, E. A.; Berzina, M. Y.; Dorofeeva, E. V.; Paramonov, A. S.; Kayushin, A. L.; Konstantinova, I. D.; Esipov, R. S. Multi-Enzymatic Cascades in the Synthesis of Modified Nucleosides: Comparison of the Thermophilic and Mesophilic Pathways. *Biomolecules* **2021**, *11* (4), 586. https://doi.org/10.3390/biom11040586.

(16) Brovetto, M.; Gamenara, D.; Saenz Méndez, P.; Seoane, G. A. C-C Bond-Forming Lyases in Organic Synthesis. *Chem. Rev.* **2011**, *111* (7), 4346–4403. https://doi.org/10.1021/cr100299p.

(17) Haridas, M.; Abdelraheem, E. M. M.; Hanefeld, U. 2-Deoxy-d-Ribose-5-Phosphate Aldolase (DERA): Applications and Modifications. *Appl. Microbiol. Biotechnol.* **2018**, *102* (23), 9959–9971. https://doi.org/10.1007/s00253-018-9392-8.

(18) Heine, A.; DeSantis, G.; Luz, J. G.; Mitchell, M.; Wong, C. H.; Witson, I. A. Observation of Covalent Intermediates in an Enzyme Mechanism at Atomic Resolution. *Science (80-.).* **2001**, *294* (5541), 369–374. https://doi.org/10.1126/science.1063601.

(19) Chambre, D.; Guérard-Hélaine, C.; Darii, E.; Mariage, A.; Petit, J. L.; Salanoubat, M.; De Berardinis, V.; Lemaire, M.; Hélaine, V. 2-Deoxyribose-5-Phosphate Aldolase, a Remarkably Tolerant Aldolase towards Nucleophile Substrates. *Chem. Commun.* **2019**, *55* (52), 7498–7501. https://doi.org/10.1039/c9cc03361k.

(20) Kunzendorf, A.; Xu, G.; van der Velde, J. J. H.; Rozeboom, H. J.; Thunnissen, A.-M. W. H.; Poelarends, G. J. Unlocking Asymmetric Michael Additions in an Archetypical Class I Aldolase by Directed Evolution. *ACS Catal.* **2021**, *11* (21), 13236–13243. https://doi.org/10.1021/acscatal.1c03911.

(21) Güclü, D.; Szekrenyi, A.; Garrabou, X.; Kickstein, M.; Junker, S.; Clapés, P.; Fessner, W.-D. Minimalist Protein Engineering of an Aldolase Provokes Unprecedented Substrate Promiscuity. *ACS Catal.* **2016**, *6* (3), 1848–1852. https://doi.org/10.1021/acscatal.5b02805.

(22) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An Open Source Platform for Ligand Pocket Detection. *BMC Bioinformatics* **2009**, *10* (1), 168. https://doi.org/10.1186/1471-2105-10-168.

(23) Bianco, G.; Forli, S.; Goodsell, D. S.; Olson, A. J. Covalent Docking Using Autodock: Two-Point Attractor and Flexible Side Chain Methods. *Protein Sci.* **2016**, *25* (1), 295–301. https://doi.org/10.1002/pro.2733.

(24) Richard, J. P. Restoring a Metabolic Pathway. *ACS Chem. Biol.* **2008**, *3* (10), 605–607. https://doi.org/10.1021/cb800238s.

(25) Hélaine, V.; Mahdi, R.; Sudhir Babu, G. V.; de Berardinis, V.; Wohlgemuth, R.; Lemaire, M.; Guérard-Hélaine, C. Straightforward Synthesis of Terminally Phosphorylated L-Sugars via Multienzymatic Cascade Reactions. *Adv. Synth. Catal.* **2015**, 357 (8), 1703–1708. https://doi.org/10.1002/adsc.201500190.

(26) Sánchez-Moreno, I.; Hélaine, V.; Poupard, N.; Charmantray, F.; Légeret, B.; Hecquet, L.; García-Junceda, E.; Wohlgemuth, R.; Guérard-Hélaine, C.; Lemaire, M. One-Pot Cascade Reactions Using Fructose-6-Phosphate Aldolase: Efficient Synthesis of D-Arabinose 5-Phosphate, D-Fructose 6-Phosphate and Analogues. *Adv. Synth. Catal.* **2012**, *354* (9), 1725–1730. https://doi.org/10.1002/adsc.201200150.

(27) Gauss, D.; Schoenenberger, B.; Wohlgemuth, R. Chemical and Enzymatic Methodologies for the Synthesis of Enantiomerically Pure Glyceraldehyde 3-Phosphates. *Carbohydr. Res.* **2014**, *389* (1), 18–24. https://doi.org/10.1016/j.carres.2013.12.023.

(28) Gauss, D.; Sánchez-Moreno, I.; Oroz-Guinea, I.; García-Junceda, E.; Wohlgemuth, R. Phosphorylation Catalyzed by Dihydroxyacetone Kinase. *European J. Org. Chem.* **2018**, *2018* (23), 2892–2895. https://doi.org/10.1002/ejoc.201800350.

(29) Lim, W. A.; Raines, R. T.; Knowles, J. R. Triosephosphate Isomerase Catalysis Is Diffusion Controlled. *Biochemistry* **1988**, *27* (4), 1158–1165. https://doi.org/10.1021/bi00404a013.

(30) Metelev, V. G.; Oretskaya, T. S. Modified Oligonucleotides: New Structures, New Properties, and New Spheres of Application. *Russ. J. Bioorganic Chem.* **2021**, *47* (2), 339–343. https://doi.org/10.1134/S1068162021020175.

(31) Grüning, N. M.; Du, D.; Keller, M. A.; Luisi, B. F.; Ralser, M. Inhibition of Triosephosphate Isomerase by Phosphoenolpyruvate in the Feedback-Regulation of Glycolysis. *Open Biol.* **2014**, *4* (MARCH). https://doi.org/10.1098/rsob.130232.

(32) Teng, H.; Wu, Z.; Wang, Z.; Jin, Z.; Yang, Y.; Jin, Q. Site-Directed Mutation of Purine Nucleoside Phosphorylase for Synthesis of 2'-Deoxy-2'-Fluoroadenosine. *Process Biochem.* **2021**, *111*, 160–171. https://doi.org/10.1016/j.procbio.2021.10.028.

(33) Garrabou, X.; Castillo, J. A.; Guérard-Hélaine, C.; Parella, T.; Joglar, J.; Lemaire, M.; Clapés, P. Asymmetric Self- and Cross-Aldol Reactions of Glycolaldehyde Catalyzed by D-Fructose-6-Phosphate Aldolase. *Angew. Chemie Int. Ed.* **2009**, *48* (30), 5521–5525. https://doi.org/10.1002/anie.200902065.