"The conversion of UDP-Glc to UDP-Man:
***In Silico* and Biochemical Exploration to Improve the Catalytic Efficiency of CDP-Tyvelose C2-Epimerases"**

Authors: Ulrike Vogel[1], Matthieu Da Costa[1], Carlos Alvarez Quispe[1], Robin Stragier[1], Henk-Jan Joosten[2], Koen Beerens[1], Tom Desmet[1§]

[1] Centre for Synthetic Biology (CSB) – Unit for Biocatalysis and Enzyme Engineering, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, 9000 Gent, Belgium

[2] Bio-Prodict BV, Nieuwe Marktstraat 54E, 6511, AA, Nijmegen, the Netherlands

[§] e-mail: tom.desmet@ugent.be (corresponding author)

Keywords (in alphabetical order including at least two from the core keyword list (see Section 3.6).): Nucleotide sugar, nucleotide sugar-active short-chain dehydrogenase/reductase (NS-SDR), C2-epimerase, Biocatalysis, Carbohydrates, …

Declaration of interests: none

Abstract

A promiscuous CDP-tyvelose 2-epimerase (TyvE) from *Thermodesulfatator atlanticus* (*Ta*TyvE) belonging to the Nucleotide Sugar active Short-Chain Dehydrogenase/Reductase superfamily (NS-SDRs) was recently discovered. *Ta*TyvE performs the slow conversion of NDP-glucose (NDP-Glc) to NDP-mannose (NDP-Man). Here, we present the sequence fingerprints that are indicative of the conversion of UDP-Glc to UDP-Man in TyvE-like enzymes based on the heptagonal box motifs. Our data-mining approach led to the identification of 11 additional TyvE-like enzymes for the conversion of UDP-Glc to UDP-Man. We characterized the top two wild-type candidates that show a 15- and 20-fold improved catalytic efficiency on UDP-Glc compared to *Ta*TyvE. Additionally, we present a quadruple variant of one of the identified enzymes with a 70-fold improved catalytic efficiency on UDP-Glc compared to *Ta*TyvE. These findings may help the design of new nucleotide production pathways starting from a cheap sugar substrate like glucose or sucrose.

Introduction

Nucleotide sugars are the substrates for glycosyltransferases, which are used to glycosylate a wide array of molecules that find different applications in the pharma, food and cosmetics industry. The glycosylation of small molecules can influence their biochemical and pharmacokinetic properties depending on the attached monosaccharide.[1] This concept has been described as glycodiversification of natural small molecules and has been employed in small molecule drug development.[1–3] Promiscuous

glycosyltransferases, their (rare) nucleotide sugar (NS) donor substrates and the enzymes involved in the NS biosynthesis pathways play essential roles for the glycodiversification.

Despite the good understanding of glycosyltransferases and the natural NS biosynthesis pathways, the biggest hurdle for biocatalytic *in vitro* glycodiversification remains the lack of (cheaply) available NSs.[4,5] Different multi-enzymatic cascades for different common nucleotide sugars, such as UDP-Glc, dTDP-Glc, GDP-Man and UDP-GlcNAc, as well as the rare GDP-L-fucose have been described.[4,5] However, these biocatalytic pathways usually pose many challenges for large scale NS production due to many unwanted by-products, inhibitory effects of the by-products on the pathway enzymes, expensive downstream processing, and difficult automation lead to high process costs and overall to extremely high prices per gram of NS.[4,5]

UDP-Glc is the NS with the best established commercial production pipeline starting from the cheap substrate sucrose, which is converted in presence of UDP by a sucrose synthase to UDP-Glc and fructose.[6,7] This short one-step reaction and the application of sucrose as cheap substrate significantly lowered the production costs and made it commercially available for a few hundred dollars per gram.[7] In this context, the conversion of NDP-Glc to NDP-Man is very attractive since the latter is the natural starting point for the biosynthesis of many rare NSs, such as GDP-L-fucose, GDP-perosamine and GDP-L-collitose.[4,5,8] A promiscuous CDP-tyvelose 2-epimerase (TyvE) from *Thermodesulfatator atlanticus* (*Ta*TyvE) was recently found to catalyze this reaction as side activity.[9]

TyvEs belong to the Nucleotide Sugar active Short-chain Dehydrogenase/Reductase superfamily (NS-SDRs), which contains different NS-active epimerases (C4-, C2-, and C3/C5-epimerases) that can invert the stereochemistry of the hydroxyl groups on the sugar ring at different positions.[10,11] The different members of the NS-SDR superfamily can be identified based on their amino acid sequence fingerprints in the heptagonal box surrounding the NS in the active site.[11] The heptagonal box model describes seven amino acid motifs surrounding the sugar substrate in the active site, which determines the substrate and product specificity of the NS-SDRs and allows the differentiation of the NS-SDRs based on their amino acid sequence.[11] The TyvEs from the NS-SDR superfamily are typically involved in the biosynthesis of 3,6-deoxy sugars and catalyze the C2-epimerization of CDP-paratose to CDP-tyvelose in *e.g.,Yersinia pseudotuberculosis* IVA and *Salmonella typhi*.[12,13] However, *Ta*TyvE was also found to perform the conversion of NDP-Glc to NDP-Man.[9]

In contrast to most other NS-SDR enzymes, the catalytic mechanism of the TyvEs does not start with the introduction of a keto-group at C4 (figure 1).[12] Instead, a proton is abstracted from the C2-OH by the conserved Tyr (Yx_3K motif, catalytic dyad, cyan wall of the heptagonal box) while a hydride is abstracted from the glycosyl C2 by the NAD^+-cofactor leading to the formation of a C2-keto-intermediate (figure 1).[12,14] The abstraction of the hydride at C2 has been identified as the rate limiting step during catalysis.[14] Subsequently, the sugar ring flips in the active site and the hydrogen is transferred back by NADH to the opposite site of the sugar ring, thereby allowing the epimerization of the C2-position.[12–14]

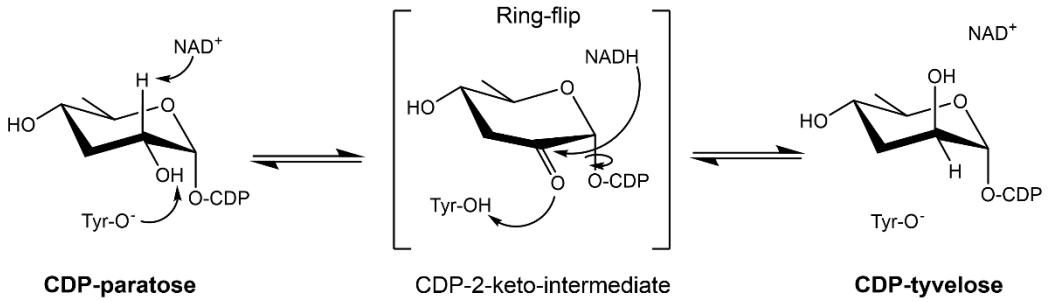


Figure 1 Catalytic mechanism of the CDP-tyvelose 2-epimerases. *TyvEs convert CDP-paratose to CDP-tyvelose. A hydride from glycosyl C2 and a proton from the C2-OH are abstracted by the NAD⁺ cofactor and by the catalytic Tyr (cyan wall, Yx3K motif), respectively, forming a 2-keto intermediate. The 2-keto-intermediate flips in the active site and the hydride is transferred back to the opposite side of the ring yielding a 2-epimer.*

Figure based on Rapp and Nidetzky, 2022.[14]

In this article, we present the sequence fingerprints that are indicative of the conversion of UDP-Glc to UDP-Man in TyvE-like enzymes based on the heptagonal box motifs.[11] Our in-depth data-mining approach led to the identification of 11 TyvE-like enzymes for the conversion of UDP-Glc to UDP-Man. We characterized the top two wild-type candidates, which showed a 15- and 20-fold improved catalytic efficiency on UDP-Glc compared to the previously described *Ta*TyvE. Additionally, we present a quadruple variant of one of the identified hits with a 70-fold improved catalytic efficiency on UDP-Glc compared to *Ta*TyvE.

Results and Discussion

The CDP-tyvelose C2-epimerase subfamily comprises different heptagonal-box motifs

Members of the TyvE subfamily were collected from the publicly available NS-SDR alignment from the 3DM database [11,15]. This resulted in 2395 putative CDP-tyvelose/paratose 2-epimerases that were subsequently subjected to phylogenetic analysis (Supplementary Figure 1). We combined our heptagonal box model with sequence similarity networks (SSN) to identify and cluster similar tree sections. At stringent SSN thresholds (high pairwise identity, >50% identity), the sequences in the TyvE subfamily disperse into disconnected groups generating clusters of very similar proteins. At a threshold of 100 (55% pairwise sequence identity, Figure 2) and of 120 (60% pairwise sequence identity), we observed the appearance of distinct islets. Next, all sequence IDs of each cluster were manually scrutinized for their heptagonal box motifs via the 3DM database. This process allowed the mapping of the TyvE subfamily into 8 different clades (i.e., Clade I-VIII) based on their different heptagonal box motifs (Table 1). Notably, some of the clusters from the SSN show identical heptagonal box motifs and were therefore clustered in the same clades (i.e., Ia-Ib and IIa-IIb).

Three TyvE representatives from *Salmonella typhi* (*St*TyvE), *Yersinia pseudotuberculosis* (*Yp*TyvE) and *Thermodesulfatator atlanticus* (*Ta*TyvE) have been characterized to date [9,12–14]. All three enzymes cluster

in clade III, despite that only *Ta*TyvE has been shown to convert NDP-Glc to NDP-Man[9]. This leaves most of the phylogenetic subfamily clusters unexplored.

The three characterized TyvEs show high motif similarity to Clades I-II but with a distinct orange wall motif (alignment pos. 191-194, HSSM instead of MSCI/QSCI, Table 1). The yellow wall (alignment pos. 124-126) is an important structural element in the active site of the NS-SDRs that often dictates an enzyme's specificity. The clades I-III contain the TNK yellow wall motifs and additionally display the classical Yx_3K dyad (cyan wall, alignment pos. 164-167, Table 1), which in this combination could be an important indicator for the C2-epimerization activity. In addition, novel yellow wall motifs (TIH, TSR and TRQ, Table 1) have been identified in the clades IV, V and VI, respectively. To date, they are not associated with any of the known enzymatic activities in the NS-SDR superfamily. [11] Furthermore, clade VII shows the TCE yellow wall motif, which has been identified in TunA from *Streptomyces chartreusis*, a 5,6-dehydratase. [8,16] However, the other heptagonal box motifs of clade VII do not correspond with those identified in 5,6-dehydratases. [8,11]

Clade VIII is the second largest cluster of the TyvE subfamily (528 sequences) and also contains the TNK motif but contains a Phe instead of the catalytic Tyr in the catalytic dyad (Fx_3K instead of Yx_3K, pos. 164-167, cyan wall, Table 1). Similarly, clades IV and VI bear a non-standard catalytic dyad motif with Lx_3K and Nx_3K (Table 1), respectively. The presence of an alternative dyad (Mx_3K) has been observed in other members of the NS-SDR superfamily, more specifically in UDP-GlcNAc 4,6-dehydratases (WbpM from *Pseudomonas aeruginosa* and PglF from *Campylobacter jejuni*). [17,18] Both enzymes still showed the typical 4,6-dehydratase activity on UDP-GlcNAc but likely employ a different mechanism. [17,18] However, in our analysis, none of the clades (IV, VI or VIII) display the TDK/E motif, nor the necessary acid/base residues (H, R or K) that are required for the 4,6-dehydratase activity. Unfortunately, assigning a specific function to these divergent catalytic dyad residues lays outside of the scope of this study.

Also worth mentioning is the presence of a structurally conserved beta-hairpin in the positions 133 - 151 in *St*TyvE. This super-secondary structure emerged as common feature of TyvEs from the structure-based multiple sequence alignment and is absent in other NS-SDR reactivities and specificities[11]. However, its function is currently not elucidated. Only the clades VI and VII do not contain the beta-hairpin (Figure 2).

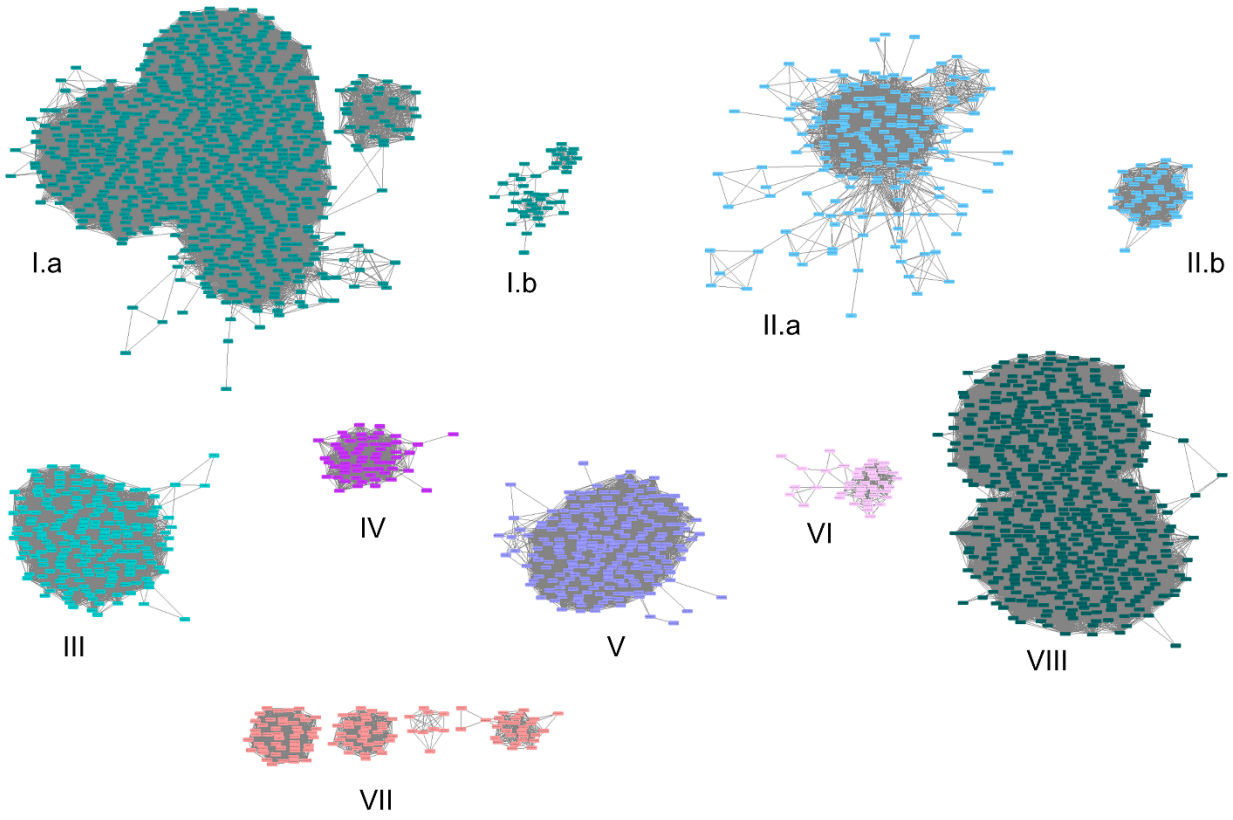


Figure 2 The sequence similarity network (threshold= 100) shows the fragmentation and clustering of the TyvE's subfamily. *The clades I-VIII were annotated according to the heptagonal-box motifs.*

Table 1 The consensus sequences of heptagonal-box motifs were identified for each of the TyvE's subfamily clades. *The biochemically characterized StTyvE, YpTyvE and TaTyvE are contained in clade III. The TNK motif of clade I, II, and III in combination with the NS-SDR-typical catalytic dyad (cyan wall) may therefore be an important indicator for the C2-epimerase activity. The divergent catalytic dyad motifs of clade IV, VII and VIII have been marked with bold letters. (All included sequences were sorted in a structural alignment targeted to StTyvE (PDB: 1ORR, UniProt: P14169))*

	Heptagonal Box							
	red	yellow	cyan	orange	purple	grey	green	
Position	82-86	124-126	164-167	191-194	199-207	234-238	301-305	No. of seq.
Family consensus	QVAVT	TNK	YGCSK	MSCI	QFGTEDQSGW	QVRDI	DQRYY	
Clade I	QVAVT	TNK	YGCSK	MSCI	QFGTEDQGW	QVRDI	DQRYF	887
Clade II	QVAVT	TNK	YGCSK	QSCI	QFGVEDQGW	QVRDL	DQKVY	230
Clade III	QVAMT	TNK	YGCSK	HSSM	QFATYDQGW	QVRDV	DQRVF	175
Clade IV	QPAMT	TIH	**L**HASK	LTGI	QFGGEDHGW	QVRDI	DLRYF	55
Clade V	EPSVL	TSR	YGTTK	CGVL	QMGKVDQGV	QVRDF	DIPIY	208
Clade VI	QVSHI	TRQ	**N**GINK	LTNV	QLIKHNRGF	QLRDF	DIGDF	43
Clade VII	NINVD	TCE	YGASK	PFNI	QKTGRFGA	QTRDY	EVARF	107
Clade VIII	QPSHD	TNK	**F**GASK	GGCL	QHSAVQLG	QVRDN	DHICY	528

Clades I to III exhibit C2-epimerase activity towards UDP-glucose

We hypothesized that TyvEs containing the TNK motif (alignment pos. 124-126, yellow wall, typical for TyvE, Table 1) and/or the traditional catalytic dyad (Yx_3K, alignment pos. 164-167, cyan wall, Table 1) are the most likely to perform the C2-epimerization of UDP-Glc (i.e., Clades I-III, V, and VIII, Table 1). Clades IV, VI and VII were excluded due to their divergent yellow motifs and catalytic dyads. In order to evaluate the activities of the different clades, representatives of each clade with expected C2-epimerase activity were expressed in *E. coli* BL21(DE3) and tested for activity on UDP-Glc (Table 2).

The clades I and II mainly comprise enzymes belonging to mesophilic organisms, while clade III contains mostly enzymes of thermophilic organisms. All tested representatives from clade I, II and III showed C2-epimerase activity on UDP-Glc, with clade III comprising the enzymes with the highest activity under the tested conditions (Table 2). *Mv*TyvE from *Methanocaldococcus villosus* (*Mv*TyvE) showed the highest specific activity on UDP-Glc with 27 $mU \cdot mg^{-1}$, followed by *Pm*TyvE from *Persephonella marina* with a specific activity of 4 $mU \cdot mg^{-1}$, which is comparable to that of *Ta*TyvE (Table 2).

Additional representatives from clades V and VIII were tested with NDP-Glc (GDP, CDP, ADP, dTDP and UDP), GDP-L-fucose, dTDP-L-rhamnose and UDP-glucuronic acid, but no activity was observed. Representatives from clade V carry a TSR motif in the yellow wall (alignment position 124-126, Table 1) in combination with the typical Yx_3K catalytic dyad. The representatives from clade VIII carry the TNK motif in combination with a divergent catalytic dyad (Px_3K instead of the typical Yx_3K, cyan wall, Table 1). Neither combination of motifs leads to C2-epimerization activity on UDP-Glc.

Based on our findings, it can be concluded that the specific combination of the heptagonal box motifs observed in clade I-III in combination with the typical NS-SDR catalytic dyad (Yx_3K, cyan wall) can be used as indicators for an inherent promiscuous C2-epimerization activity towards UDP-Glc of putative CDP-tyvelose 2-epimerases.

Table 2 The different representatives of the clades I-III, V and VIII were tested with UDP-Glc. The enzymes from clade I-III showed an inherent C2-epimerase activity towards UDP-Glc*. The representative from* Persephonella marina *and* Methanocaldococcus villosus *were shown to have equal and 7-fold improved specific activities on UDP-Glc compared to the previously characterized TyvE from* Thermodesulfatator atlanticus*. (N.D. = activity not detected)*

Clade	Species	UniProt Accession	Specific activity (mU/mg)	Temperature
I.a	*Sphingobium ummariense*	T0IXA2	0.09	30 °C
II.a	*Oscillochloris trichoides*	E1IC11	1.6	30 °C
	Isosphaera pallida	E8R6X6	1.71	30 °C
III	*Persephonella marina*	C0QSV4	4	60 °C
	Geobacter daltonii	B9M8Y6	0.8	30 °C
	Methanocaldococcus villosus	N6VPH4	27	60 °C
	Thermodesulfatator atlanticus	WP_022854415 (NCBI)	3.8 [9] 5.1 [This study]	60 °C
V	*Dyadobacter fermentans*	C6W226	N.D.	30 °C
	Cyanosarcina cf. burmensis	A0A2T1FLB3 (3DM)	N.D.	30 °C
	Microcystis aeruginosa	I4I2L3	N.D.	30 °C
VIII	*Rhodopirellula sallentina*	M5U8X5	N.D.	30 °C
	Fibrisoma limi	I2GR76	N.D.	30 °C
	Oscillatoria nigro-viridis	K9VM34	N.D.	30 °C

Further exploration of clades I-III for an inherent C2-epimerase activity on UDP-Glc

To identify substitutions that positively impact activity on UDP-Glc, several rational engineering strategies were followed. First, conserved positions were targeted through alanine scanning (N125A, K126A, S167A, H191A, S192A, S193A, M194A, R142A, Y143A, S272A, S272A and W207A). Only *Mv*TyvE S274A showed an increased relative activity of 156 ± 17 % compared to the wild-type (WT) *Mv*TyvE (Figure 3, Supplementary Figure 2A). Subsequently, mutations mimicking the active site of the UDP-Glc 4-epimerase (GalE) were introduced, leading to a drastic decrease in activity in all variants (Supplementary Figure 2B). In the next step, homologous TyvEs from clade III were analyzed and substitutions within the heptagonal box motifs were introduced into *Mv*TyvE (V83T, M85V, N125C, S193T, H191L, M194I, V83T/S193, Supplementary Figure 2C). Only, *Mv*TyvE S193T and *Mv*TyvE S193/V83T showed a relatively increased activity of 150 ± 17 and 219 ± 21 % compared to WT *Mv*TyvE (Figure 3, Supplementary Figure 2C). Additional 16 mutations were introduced based on the interpretation of *Mv*TyvE's three-dimensional model docked with CDP-Glc (Supplementary 2D), none of which led to an increased activity on UDP-Glc. Lastly, combinations of mutations in position 83, 124, 193 and 203 were tested. This led to a quadruple mutant (*Mv*TyvE T124S/V83T/S193T/Y203V) with a 3-fold increased relative activity compared to WT *Mv*TyvE (Figure 3).

The natural substrate of *Mv*TyvE is CDP-paratose, a 3,6-deoxy hexose. In comparison to paratose, glucose is bulkier due to the presence of the C3-OH and C6-OH groups. The engineering challenge for the conversion of UDP-Glc is to allow the bulkier sugar to fit into the active site without restraining the rotation of the sugar ring during catalysis. The smaller and slightly more flexible Ser in position 124 may allow an easier passage of the C6-OH during the flipping motion, similarly to the UDP-Glc 4-epimerases (GalEs). Mutations V83T and S193T may introduce additional interactions with the sugar ring. Y203 likely interacts with the nucleotide anker of the substrate. The introduction of a Val in this position may abolish this interaction and subsequently allow more mobility during catalysis. To further elucidate how these substitutions improved activity of *Mv*TyvE a structure-function analysis based on crystal structures with bound UDP-Glc and molecular dynamics simulations of the quadruple mutant will be necessary. Unfortunately, these analyses lay outside of the scope of this article.

Subsequently, we searched for TyvE-like enzymes (containing the 'TNK'-motif and the Yx_3K catalytic dyad in the cyan wall) that naturally contain these substitutions with the EnzymeMiner tool (Figure 3). This search yielded 303 sequences (Supplementary Figure 3), which are present in branches I-IV of the constructed phylogenetic tree. Branch II shows the greatest variety of substitutions in the positions that led to an improved activity on UDP-Glc in *Mv*TyvE (Supplementary Table 1). Therefore, only branch II was further manually screened for TyvE homologues. Six sequences from thermophilic organisms were identified that contained combinations of the orange heptagonal box motifs from clade II, the TNK motif, the Yx_3K catalytic dyad and at least one substitution in either position 83, 124, 142, 193, 203 or 274.

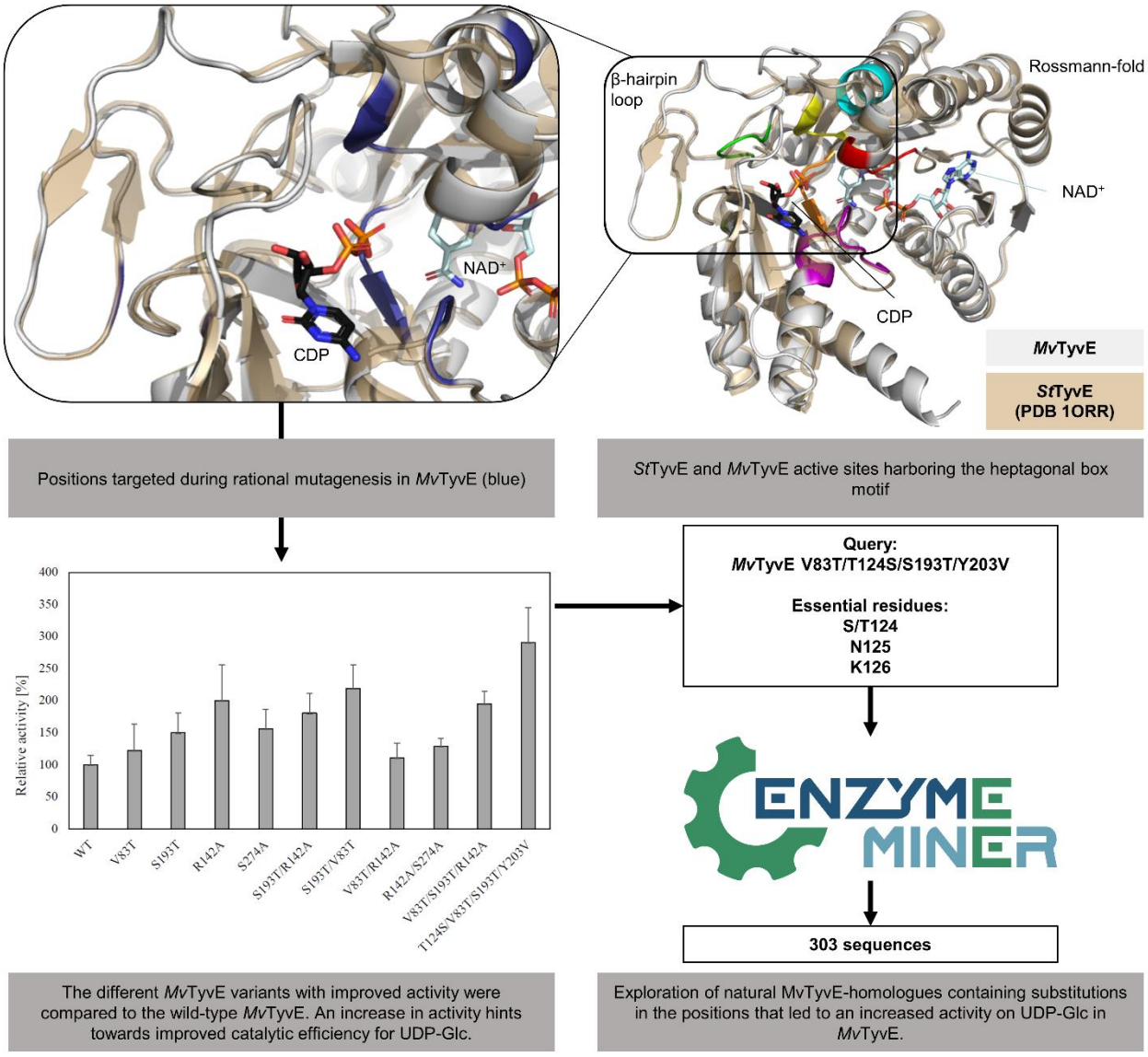


Figure 3 Data mining strategy to further explore the natural diversity of TyvEs: *The structural analysis and the creations of multiple MvTyvE variants led to the identification of a number of variants with slightly increased activity on UDP-Glc compared to the wild-type. The best variant (MvTyvE T124S/V83T/S193T/Y203V) was used as query for the EnzymeMiner 1.0 tool and the 'SNK/TNK' motif were chosen as essential residues, resulting in 303 sequences that were subjected to further manual analysis.*

These six enzymes were expressed in *E. coli* BL21(DE3) and tested with UDP-Glc (Table 3). All six enzymes showed activity on UDP-Glc under the tested conditions (Table 3). *Ab*TyvE from *Ardenticatenia bacterium* showed the highest specific activity of 16.3 mU/mg, which is a 4-fold increase compared to *Ta*TyvE but lower than the 27 mU/mg of *Mv*TyvE, the best wildtype found so far.

Based on these results, it can be concluded that the combination of the TNK motif with a traditional Yx_3K catalytic dyad and the orange wall motif of heptagonal box from clade II and III can be used to identify TyvE-like enzymes with a promiscuous activity towards UDP-Glc (Supplementary Table 2).

Table 3 All six identified TyvE-like sequences showed activity on UDP-Glc under the tested conditions.

Clade	Species	UniProt	Specific activity (mU/mg)	Temperature
II	*Desulfobacteraceae bacterium* 4572_187	A0A1W9U078	0.07	30 °C
II	*Candidatus Diapherotrites archaeon*	A0A497JHV9	0.59	60 °C
II.a	*Anaerolinea thermophila*	A0A117LGQ0	0.71	60 °C
II.a	*Chloroflexi bacterium* (sequence 1)	A0A497C4L0	0.25	60 °C
II	*Chloroflexi bacterium* (sequence 2)	A0A661VU89	0.17	60 °C
II.a	*Ardenticatenia bacterium*	A0A3M1BUZ3	16.3	50 °C

Kinetic characterization of *Mv*TyvE, *Mv*TyvE T124S/V83T/S193T/Y203V and *Ab*TyvE

To evaluate how much the enzymatic activity of *Mv*TyvE, the best *Mv*TyvE mutant (*Mv*TyvE T124S/V83T/S193T/Y203V), and *Ab*TyvE improved compared to the previously described *Ta*TyvE the kinetic parameters were determined for all three enzymes. Subsequently, the catalytic efficiencies on UDP-Glc were compared to *Ta*TyvE (Table 4).

Both *Mv*TyvE and *Ab*TyvE are thermophilic enzymes with a pH optimum of 8 and 7.5, and a temperature optimum of 60 and 45 – 55 °C, respectively (Table 4). *Mv*TyvE has a 7-fold improved V_{max}, 4-times lower K_M and 3-fold improved k_{cat} compared to *Ta*TyvE, which lead to a 15-fold improved catalytic efficiency (kcat/Km). Also *Ab*TyvE has an improved catalytic efficiency (20-fold), which results from both better affinity (10-fold lower K_M) and conversion (2-fold higher kcat) (Table 4). However, the best variant is by far the quadruple *Mv*TyvE variant (T124S/V83T/S193T/Y203V), which shows a 28-fold improved conversion (k_{cat}) on UDP-Glc, leading to a catalytic efficiency of 3.58 $min^{-1} \cdot mM^{-1}$, which is over 70-fold higher than that of *Ta*TyvE (Table 4). Compared to *Ta*TyvE the catalytic efficiencies were 15-fold, 70-fold and 20-fold improved for *Mv*TyvE, the *Mv*TyvE quadruple mutant (T124S/V83T/S193T/Y203V) and *Ab*TyvE, respectively. Both *Ta*TyvE and the *Mv*TyvE quadruple mutant (T124S/V83T/S193T/Y203V) show substrate inhibition with K_i-values of 45.7 mM[19] and 19 mM, respectively.

In conclusion, using our data-mining pipeline we were able to not only link the heptagonal box fingerprint motifs to the promiscuous conversion of UDP-Glc to UDP-Man, but were also able to identify and engineer enzymes with improved kinetic properties. However, it should critically be noted that the overall activity of these enzymes is still too low for efficient industrial applications. Nevertheless, the identification of these three enzymes is a significant step forward as we provide the means to identify more wild-type enzymes and screen for improved wild-type activities of the C2-epimerization on UDP-Glc. Furthermore, the engineering of newly identified representatives may hold the key to their applicability in biocatalytic processes. In-depth structural, mechanistic, and molecular dynamics studies of these new TyvE-like NS-SDRs will also be necessary to better understand which factors enable the conversion of NDP-Glc.

Table 4 The newly identified TyvE-like enzymes show a higher catalytic efficiency for the conversion of UDP-Glc to UDP-Man than the previously described TaTyvE. *TaTyvE and the MvTyvE quadruple mutant showed substrate inhibition at 45.7 and 18.6 mM of UDP-Glc, respectively.*

Enzyme	Accession N°	T_{opt} (°C)	pH_{opt}	V_{max} [mU/ mg]	K_M [mM]	k_{cat} [min^{-1}]	k_{cat}/K_M [$min^{-1} \cdot mM^{-1}$]	K_i [mM]	K_{eq} [UDP-Man/ UDP-Glc]	
*Ta*TyvE	WP_022854415 (NCBI)	70	7.5	3.8	5.6	0.33	0.05	45.7	-	[9,19]
*Mv*TyvE	N6VPH4 (UniProt)	60	8	27 ± 3.6	1.5 ± 0.6	1.1 ± 0.1	0.73	-	0.19	This study
*Mv*TyvE T124S/V83T/ S193T/Y203V	Quadruple mutant	60	8	243 ± 45	2.6 ± 0.8	9.3 ± 1.8	3.58	19 ± 9.0	0.36	
*Ab*TyvE	A0A3M1BUZ3 (UniProt)	45-55	7.5	16 ± 1.3	0.6 ± 0.2	0.6 ± 0.05	1.06	-	0.14	

Conclusion

In this study, we successfully linked heptagonal box sequence motifs to the conversion of UDP-Glc to UDP-Man. We thereby proved that this promiscuous activity is more common in nature than previously suspected and that it can be identified solely based on the amino acid sequence. The identified sequence fingerprints were used in a prove-of-concept to find more wild-type TyvE-like sequences with the ability to convert UDP-Glc to UDP-Man. Furthermore, using these fingerprint motifs we were able to characterize three enzymes with 15-, 20- and 70-fold increased catalytic efficiency for the conversion of UDP-Glc compared to the originally described *Ta*TyvE. Although the overall activity of these enzymes is still low, this study provides the means for the further screening of the sequence space and identification of more TyvE-like enzymes. Further screening of the sequence space in combination with in-depth structural, mechanistic, and molecular dynamics studies as well as enzyme engineering may eventually lead to TyvE-like enzymes that are interesting for industrial applications. Overall, we believe that this knowledge will help to pave the way to make rare NDP-sugar production more affordable by allowing the synthesis to start from cheap substrates like glucose or sucrose.

Computational Methods and Experimental Section

Materials

All chemicals were obtained from Sigma-Aldrich (Saint Lois, MO, USA) or Carbosynth (Compton, UK).

Transformation of the TyvE homologues into *E. coli* BL21 (DE3)

The codon-optimized genes of CDP-tyvelose 2-epimerase from the sequences mentioned in Tables 2 and 3 were synthesized by GeneArt Gene Synthesis (Thermo Fisher Scientific, Waltham, MA, USA) and subcloned into the pET21 vector at *Nde*I and *Xho*I restriction sites, providing a C-terminal His-tag. All ordered plasmids were transformed in *E. coli* BL21(DE3) electrocompetent cells for gene expression and protein production. Single point mutations were introduced in the plasmids of *Mv*TyvE following the PCR-

based protocol described by Sanchis et. al [20] and subsequently sequenced. The resulting vectors were subsequently transformed into *E. coli* BL21 (DE3).

Expression and purification of the TyvE homologues

The recombinant *E. coli* cells for the expression of TyvE homologues were cultivated in 250 mL of Lysogeny Broth (LB) medium containing 100 µg/mL of ampicillin (37 °C, 200 rpm). When the optical density of the cells reached 0.6, a final concentration of 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) was added and the culture was incubated for 20h (18 °C, 200 rpm). The produced biomass was harvested by centrifugation (4°C, 20 min at 9667 · g) and the resulting pellet was frozen at −20°C.

Then, cell pellets were thawed and dissolved in 8 mL lysis buffer (500 mM NaCl, 10 mM imidazole, 0.1 mM phenylmethane sulfonyl fluoride (PMSF) in 50 mM phosphate buffer; pH 7.5, 1 mg/mL lysozyme) and incubated on ice for 30 minutes. Subsequently, each sample was sonicated three times for 2 min while cooled on ice (Branson Sonifier 250, output control level 2.5, duty cycle 30%) and centrifuged (19 333 · g for 60 min, 4°C). The supernatant, comprising the soluble protein fraction, was subsequently filtered through a 0.2 µm polyethersulfone membrane filter and afterwards purified by nickel-nitrilotriacetic acid (Ni-NTA) chromatography as described by the supplier (HisPur™ Ni-NTA Resin, ThermoFisher Scientific, Waltham, United States). A buffer exchange was performed through ultracentrifugation with a 30 kDa cut-off (Amicon Ultra-15, Merck Millipore, Billerica, USA) using 3-(*N*-morpholino)propanesulfonic acid (MOPS, pH 7, 100 mM, Sigma-Aldrich, St. Louis, USA).

Lastly, the protein concentration was determined with the NanoDrop2000 Spectrophotometer (Thermo Scientific) by measuring the absorbance at 280 nm, considering the molecular weight and extinction coefficient of TyvE homologues (obtained from the Expasy ProtParam tool for each sequence). The molecular weight and purity of the proteins were verified by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE; 12% gels).

Activity tests

The activity of the different wild-type TyvEs and the *Mv*TyvE variants were assessed towards UDP-Glc in a reaction mixture containing 100 mM MOPS buffer (pH 7 or 7.5), 1 $mg \cdot mL^{-1}$ of enzyme and 5 or 10 mM UDP-Glc, respectively. The activity was determined at 60°C or 30°C, depending on the isolation source of the microorganism. All reactions were monitored by sampling every 10 minutes for 1 hour. For the determination of the optimal pH and temperature of *Mv*TyvE and *Ab*TyvE reaction mixtures with 0.75 and 0.5 $mg \cdot mL^{-1}$ enzyme, respectively, and 5 mM UDP-Glc were sampled over a time span of 40 min. To determine the kinetic parameters of *Mv*TyvE, *Mv*TyvE T124S/V83T/S193T/Y203V and *Ab*TyvE 0.75, 0.1 and 0.5 $mg \cdot mL^{-1}$ enzyme were used in a total volume of 150 µL with different concentrations of UDP-Glc (0-50 mM). The reactions were sampled over a time span of 9 min. Enzyme inactivation was achieved by transferring the sample to an equal amount of 50 % acetonitrile and subsequently diluting 10x in 100 mM acetic acid or by directly diluting 10x in 100 mM acetic acid. Acid hydrolysis was performed by incubating the samples for 1h at 95°C. Conversion of UDP-Glc to UDP-Man was evaluated by high-performance anion exchange chromatography-pulsed amperometric detection (HPAEC-PAD) using the Dionex ICS-3000 system (Thermo Fischer Scientific) (CarboPac PA20 column-3 × 150 mm) as described by Gevaert et al.[21] but with a 15-min isocratic method of 8 mM NaOH. The NDP-sugar conversion was quantified using a mannose standard curve.

For all activity tests, product concentration was plotted in function of time and a linear correlation was fitted representing the activity of the enzyme (in µmol/L/min) from which the specific activities (in mU/mg) were calculated taking into account the total amount of enzyme and reaction volumes. For the determination of the kinetic parameters, the specific activities at the different UDP-Glc concentrations were introduced in Sigma plot, which was then used to fit a ligand binding curve to the measured points to calculate V_{max}, K_M, k_{cat} and K_i. All experiments were performed in triplicates. The following equations were used in to fit the data in Sigma plot:

Michaelis-Menten kinetics

$$v = \frac{v_{max} \cdot [S]}{K_M + [S]}$$

and Michaelis-Menten kinetics for substrate inhibition

$$v = \frac{v_{max} \cdot [S]}{K_M + [S] \cdot \left(1 + \frac{[S]}{K_i}\right)}$$

In silico analysis

3DM

3DM is a software (https://3dm.bio-prodict.com)[15], for the construction of structure-based multiple sequence alignments with all sequences from a superfamily and known three-dimensional structures. Once created, the database provides several protein-related information such as the amino acid conservation or the correlated mutation data extracted from the alignments. A 3DM system was built based on the SDR family, designated NAD(P)-binding Rossmann-fold domains (2017), refined as "Targeted on 1ORRA" to use the TyvE structure from *Salmonella typhi* (*St*TyvE) as alignment template for all NS-SDR members, and made publicly available in a previous study. [11] Consequently, all sequence positions mentioned in this study are the positions of *St*TyvE in order to facilitate the comparison between structurally equivalent positions. Structural information was obtained by extracting family members from the 'Structural Classification of Proteins'-database (SCOP) and running BLAST against the PDB database using as queries representative members from each SCOP subfamily. Subsequently, a subset was constructed containing all 2395 putative TyvE sequences from the subfamily "1ORRA". This step was necessary to retrieve the alignment statistics of only TyvE close homologues.

Sequence similarity network

Generation of sequence similarity networks (SSN) was done using the EFI-EST web-tool [22,23]. Each network was visualized with Cytoscape[24]. First, the enzyme sequences were extracted from the 3DM database as a FASTA file and given as input to generate the initial dataset. Then, numerous sequence similarity thresholds (50 to 150), corresponding to sequence pairs connected by edges, were specified to generate and identify the final SSN. Given that thresholds determine the degree of segregation of proteins into clusters, a trial-and-error process was performed. Threshold values in terms of identity are visualised in Figure S1.

Enzyme miner for the identification of MvTyvE T124S/V83T/S193T/Y203V homologues

The EnzymeMiner 1.0 tool[25] was used to search for homologues of our best *Mv*TyvE variant (MvTyvE T124S/V83T/S193T/Y203V). The positions S/T124, N125 and K126 were defined as 'set residues'. The resulting 303 putative TyvE sequences were ordered in a phylogenetic tree using ClustalΩ. [26] Subsequently, all sequences from the branches I-IV were introduced into the 3DM database, which was used to assess the number of different substitutions in the positions 83, 124, 142, 193, 203 and 274. Then, only the sequences of branch II were manually scrutinized to identify sequences that originated from thermophilic organisms and contained at least one substitution in either position 83, 124, 142, 193, 203 or 274.

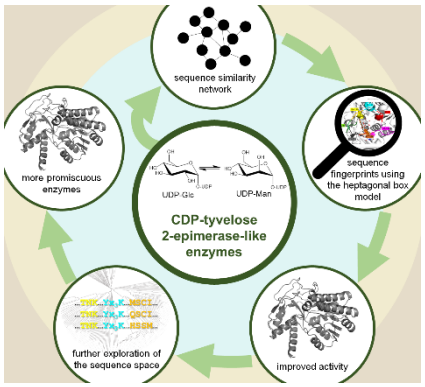
Acknowledgments

This research was funded by the Research Foundation – Flanders (FWO-Vlaanderen) via the EpiSwitch and DeoxyBioCat projects (grant n◦ G0F3417N and G0A7520N, respectively) as well as via a doctoral scholarship for M.D.C. (grant n◦ 1S03020N) and C.A.Q. (grant n◦ 1105922N). We would like to thank the UGent Core Facility 'HTS for SynBio' for training, support and access to the instrument park.

References

[1] C. J. Thibodeaux, C. E. Melançon, H. W. Liu, *Nature* **2007**, *446*, 1008–1016.
[2] C. J. Thibodeaux, C. E. Melançon, H. W. Liu, *Angewandte Chemie - International Edition* **2008**, *47*, 9814–9859.
[3] G. J. Williams, R. W. Gantt, J. S. Thorson, *Current Opinion in Chemical Biology* **2008**, *12*, 556–564.
[4] H. Frohnmeyer, L. Elling, *Carbohydrate Research* **2022**, 108727.
[5] J. P. Dolan, S. C. Cosgrove, G. J. Miller, *JACS Au* **2023**, *3*, 47–61.
[6] M. Diricks, F. D. Bruyn, P. V. Daele, M. Walmagh, T. Desmet, *Applied Microbiology and Biotechnology* **2015**, *99*, 8465–8474.
[7] K. Schmölzer, M. Lemmerer, A. Gutmann, B. Nidetzky, *Biotechnology and Bioengineering* **2017**, *114*, 924–928.
[8] U. Vogel, K. Beerens, T. Desmet, *Journal of Biological Chemistry* **2022**, *298*, 101809.
[9] C. Rapp, S. van Overtveldt, K. Beerens, H. Weber, T. Desmet, B. Nidetzky, *Appl Environ Microbiol* **2021**, *87*, e02131-20.
[10] K. L. Kavanagh, H. Jörnvall, B. Persson, U. Oppermann, *Cellular and Molecular Life Sciences* **2008**, *65*, 3895–3906.
[11] M. Da Costa, O. Gevaert, S. Van Overtveldt, J. Lange, H. J. Joosten, T. Desmet, K. Beerens, *Biotechnology Advances* **2021**, *48*, DOI 10.1016/j.biotechadv.2021.107705.
[12] T. M. Hallis, Z. Zhao, H. Liu, *J. Am. Chem. Soc.* **2000**, *122*, 10493–10503.
[13] N. M. Koropatkin, H. Liu, H. M. Holden, *Journal of Biological Chemistry* **2003**, *278*, 20874–20881.
[14] C. Rapp, B. Nidetzky, *ACS Catal.* **2022**, *12*, 6816–6830.
[15] R. K. Kuipers, H.-J. Joosten, W. J. H. van Berkel, N. G. H. Leferink, E. Rooijen, E. Ittmann, F. van Zimmeren, H. Jochens, U. Bornscheuer, G. Vriend, V. A. P. Martins dos Santos, P. J. Schaap, *Proteins* **2010**, 2101–2113.
[16] F. J. Wyszynski, S. S. Lee, T. Yabe, H. Wang, J. P. Gomez-Escribano, M. J. Bibb, S. J. Lee, G. J. Davies, B. G. Davis, *Nature Chemistry* **2012**, *4*, 539–546.
[17] C. Creuzenet, R. V. Urbanic, J. S. Lam, *Journal of Biological Chemistry* **2002**, *277*, 26769–26778.
[18] C. Creuzenet, J. S. Lam, *Molecular Microbiology* **2001**, *41*, 1295–1310.
[19] S. Van Overtveldt, Unlocking New Biocatalytic Pathways by Engineering of Carbohydrate Epimerases, Biology and Life Sciences, Technology and Engineering, Universiteit Gent. Faculteit Bio-ingenieurswetenschappen, **2020**.
[20] J. Sanchis, L. Fernández, J. D. Carballeira, J. Drone, Y. Gumulya, H. Höbenreich, D. Kahakeaw, S. Kille, R. Lohmer, J. J. P. Peyralans, J. Podtetenieff, S. Prasad, P. Soni, A. Taglieber, S. Wu, F. E. Zilly, M. T. Reetz, *Applied microbiology and biotechnology* **2008**, *81*, 387–397.
[21] O. Gevaert, S. Van Overtveldt, M. Da Costa, K. Beerens, T. Desmet, *International Journal of Biological Macromolecules* **2020**, *165*, 1862–1868.
[22] R. Zallot, N. Oberg, J. A. Gerlt, *Biochemistry* **2019**, *58*, 4169–4182.
[23] J. A. Gerlt, *Biochemistry* **2017**, *56*, 4293–4308.
[24] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, *Genome Res.* **2003**, *13*, 2498–2504.
[25] J. Hon, S. Borko, J. Stourac, Z. Prokop, J. Zendulka, D. Bednar, T. Martinek, J. Damborsky, *Nucleic Acids Research* **2020**, *48*, W104–W109.
[26] 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, *Mol Syst Biol* **2011**, *7*, 539.

Table of Contents entry



Promiscuous CDP-tyvelose 2-epimerase (TyvE) convert NDP-glucose to NDP-mannose. We present the sequence fingerprints that are indicative of this conversion in TyvE-like enzymes. 11 TyvE-like enzymes were identified and the top two wild-type candidates and a quadruple mutant were characterized. The improved catalytic efficiency of these enzymes may help the design of new nucleotide production pathways starting from a cheap sugar substrate like sucrose.