Electronic Supplementary Information

CONTENT

Experimental section

Mutagenesis 2
ProSAR modelling 2
HPAEC/HPLC 2
Screening 2
Detailed characterisation of enzymatic reaction 3
Docking 3
Production and downstream processing 4

Figures

Figure S1. Residues targeted for mutagenesis 5
Figure S2. HPAEC-PAD chromatograms of enzymatic reactions 6
Figure S3. Evaluation of the prediction model 7
Figure S4. Michaelis-Menten curves for the wild-type and mutant enzymes 8
Figure S5. Docking of D-glucose in the wild-type enzyme and double mutant 9
Figure S6. HPAEC-PAD and HPLC-RI profiles of the purified kojibiose 10
Figure S7. ^{1}H NMR and ^{13}C NMR spectrum of the purified kojibiose 11

Tables

Table S1. Predicted and measured selectivity and activity of the wild-type and mutant enzymes 12

References

EXPERIMENTAL DETAILS

Mutagenesis

All libraries were generated from the constitutive expression plasmid pCXP34h containing the sucrose phosphorylase gene from *Bifidobacterium adolescentis*[1], according to the method described by Coussement *et al*.[2] In brief, a backbone fragment is amplified in which a short fragment (50-60 bp) containing the degenerated codon (here NNK) is ligated by means of a Gibson assembly procedure.[3] Library DNA was transformed in *E. coli* CGSC 8974 and successful randomisation was checked by sequencing the pooled plasmids from the initial transformation.[4] Site-directed mutations were introduced with a modified two-stage megaprimer-based whole-plasmid PCR method[5] and individual clones were subjected to nucleotide sequencing to confirm that the correct mutations had been introduced, and to exclude the presence of undesirable mutations.

ProSAR modelling

Three recently developed statistical prediction models that are able to link protein sequence to activity/selectivity (AFSAR, SiPSAR and AFSiPSAR)[6, 7] were evaluated and prediction from the one that fitted the input data best (based on leave one out cross validation (LOOCV); Figure S3) were used to select the mutants for *in vitro* testing (here AFSAR). In brief, the AFSAR (Amino Acid Features in Sequence-Activity Relationships) model takes into account physico- and biochemical properties of amino acids for the encoding of the mutated positions, in contrast to the original ProSAR (protein sequence activity relationship) approach.[8, 9] Incorporation of (albeit different) amino acid properties is also included in the SipSAR (Signal-Processing in Sequence-Activity Relationships) method. Here, however, the focus is not only on the mutated positions, but the complete sequence is taken into account (by using signal processing techniques). Characteristics of both methods are finally combined in the AFSiPSAR model. This way, predictions for selectivity and activity can be made, based on both numerical and spectral information.

HPAEC/HPLC

All reactions were monitored by High Performance Anion Excange Chromatography (HPAEC) (Dionex ICS-3000, Thermo Scientific), using a CarboPac PA20 pH-stable column and Pulsed Amperometric Detection (PAD) and all samples (10 µl) were analysed at a constant flow rate of 0.5 ml/min at 30°C. The method used for screening consisted of an isocratic elution with 35 mM NaOH and 15 mM acetate for 12 min. That way trehalose, kojibiose, nigerose, maltose and isomaltose (products) could be separated from one another and from sucrose, glucose and fructose (which are not separated from each other). For the detailed characterisation, separation of all possible substrate and (potential) product peaks (sucrose, glucose, fructose, trehalose, kojibiose, nigerose, maltose, isomaltose) was achieved with a 30 min protocol. After 13 min of isocratic elution with 30 mM NaOH, the concentration was gradually increased to 100 mM in 5 min, kept constant for 3 min and decreased again to 30 mM within 1 min, followed by an equilibration period of 8 min. Purity of the crystallised kojibiose was also analysed by HPLC with an Aminex HPX-87H column (Bio-Rad), equilibrated at 30°C, and Refractive Index (RI) detection, under isocratic elution with 5 mM H_2SO_4 at a constant flow rate of 0.6 ml/min.

Screening

Individual colonies were picked, grown and lysed according to the procedure described by De Groeve et al.[10], except for the use of a different lysis buffer (1 mg/ml lysozyme, 0.1 mM PMSF, 50 mM Na_2SO_4, 4 mM $MgSO_4$ and 1 mM EDTA in 50 mM MOPS buffer pH 7.0). Reactions were initiated by adding 50 µl of crude cell extract to 50 µl substrate solution in low well microtiter plates (Nunc) (final concentration of 100 mM donor (sucrose) and 200 mM acceptor (D-glucose)). Reactions were incubated for 16 hours at 37°C, after which 200 µl of 0.01 N NaOH was added to stop the reaction. Subsequently, microtiter plates were spun down for 30 min at 4500 rpm. Supernatant (12 µl) was transferred to deep well plates and diluted with mQ to a final dilution of 500x (total volume of 2000 µl/well). Samples were analysed with the short HPAEC-PAD method described above. For each library, two hundred clones were screened. Since NNK codon degeneracy was applied for each individual library (number of theoretical clones = 32), this accounted for around a six times oversampling or a coverage of 99%.[11]

Detailed characterisation of enzymatic reaction

Enzyme production and His-tag purification was done as previously described,[12] and samples were taken and analysed as during the screening. All reactions were performed at 55°C in 50 mM MOPS buffer pH 7.0. Specific activity and selectivity were calculated from initial reaction rates obtained from reactions of 0.25-2 mg/ml Ni-NTA purified enzyme in the presence of 100 mM donor substrate (sucrose) and 200 mM acceptor (D-glucose). One unit was defined as the amount of enzyme that produced one µmol of kojibiose per minute, under the specified conditions, and the selectivity as the fraction of kojibiose in total product formation.

Michaelis-Menten profiles were obtained from reactions with 200 mM sucrose and D-glucose concentrations ranging from 0-1500 mM. The apparent kinetic parameters for the glucose binding mode leading to kojibiose or maltose were derived from the respective quantification of kojibiose or maltose formation. For the determination of the (kinetic) stability of the wild-type enzyme, the single mutants L341I and Q345S as well as the double mutant L341I_Q345S, 0.5 mg/ml purified enzyme was incubated at 55°C and 60°C and samples were taken at regular intervals. The half-live values were calculated from first-order fits of the stability curves.

Docking

A D-glucose molecule was docked in the covalent glycosyl-enzyme intermediate of the wild-type enzyme, the single mutants L341I and Q345S, and the double mutant L341I_Q345S. Dockings were repeated with all side-chains fixed, side-chains of residues 341 and 345 flexible and all side-chains in the acceptor site flexible (except for the catalytic residues), for the four enzyme variants. All manipulations and the docking setup were performed with the molecular modeling program YASARA.[13] Docking was performed with the VINA[14] docking module implemented in YASARA, using default parameters, with the exception of the number of runs which was increased to 100. The D-glucose molecule was created with the built-in oligosaccharide builder and was treated fully flexible during docking (all internal degrees of freedom are taken into account and the molecule was allowed to freely rotate). The crystal structure of the covalent glycosyl-enzyme intermediate (pdb entry 2GDV chain A) served as the wild-type enzyme receptor molecule. A model of the L341I, Q345S and L341I_Q345S mutant was derived from this structure by *in silico* mutation of the required residues, followed by optimization of the side-chains of the respective and surrounding residues.[15, 16]

Production and downstream processing

Reaction at 1 liter scale was performed at 55°C in milliQ water. After 24 hours the reaction mixture was heated to 95°C for 10 min to inactivate the enzyme and contaminating carbohydrates were removed by yeast treatment (30g/l spray dried *Saccharomyces cerevisiae* (Algist Bruggeman), 30°C). After 8 hours the yeast was cleared out by centrifugation (5000 rpm, 4°C, 15 min), followed by vacuum microfiltration (0.22 micron). Finally, the solution was evaporated by means of a rotavapor (50°C, 50 mbar) to 1/3 of the initial volume, cooled to room temperature overnight and placed on ice to promote further crystal growth. In a last step, crystals were washed with ethanol and dried to the air.

FIGURES

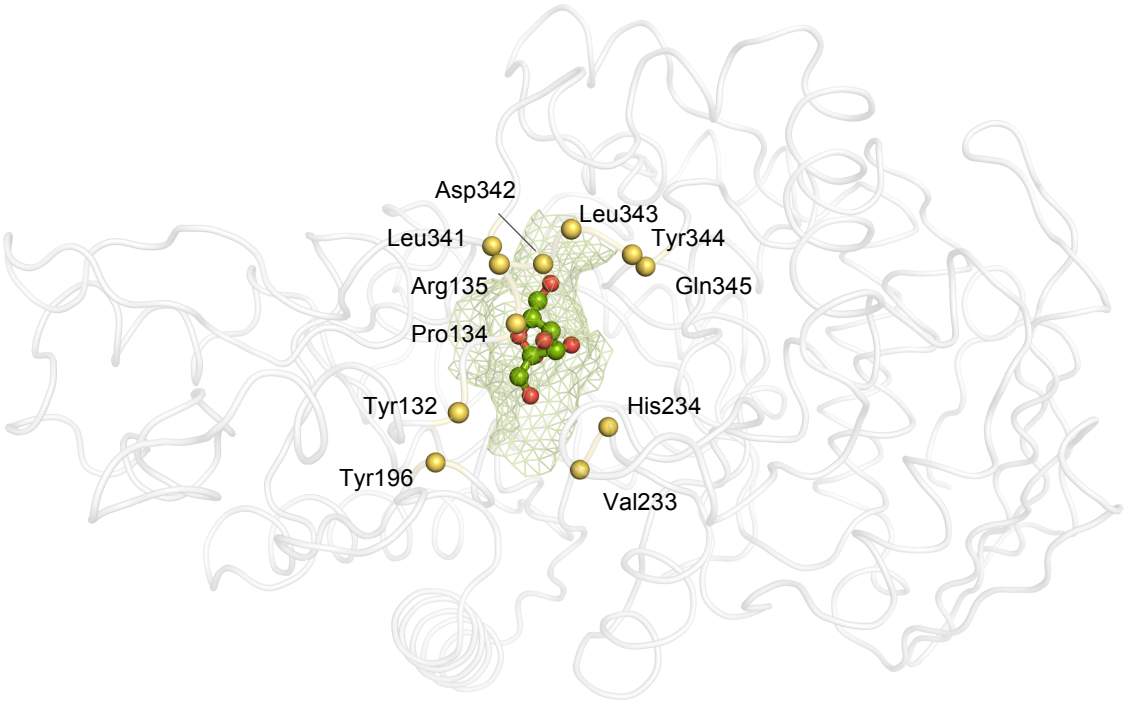

Figure S1. Residues targeted for mutagenesis, i.e. all residues within 5 Å from the fructose or phosphate moiety that are neither the nucleophile (Asp192), general acid/base catalyst (Glu232) or transition state stabiliser (Asp290), nor that are part of the donor site (Phe156) (eleven positions in total); see also references[17, 18] for additional information (yellow ball: Cα atoms of mutated residues; green/red ball and stick: fructose moiety; green mesh; active site pocket; pdb entry 2GDU).

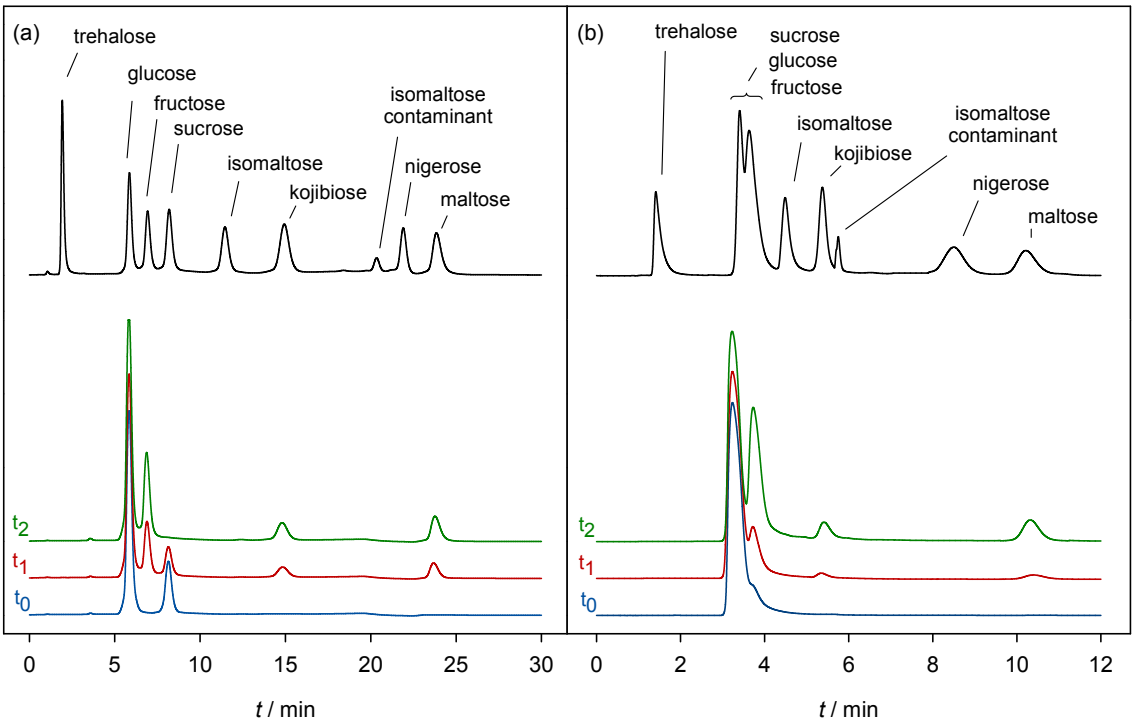

Figure S2. High performance anion exchange chromatography (HPAEC) profiles for (a) detailed characterisation and (b) initial screening with sucrose as donor substrate (top: standard series; below: reaction at different time points for wild-type enzyme incubated with 100 mM sucrose and 200 mM D-glucose).

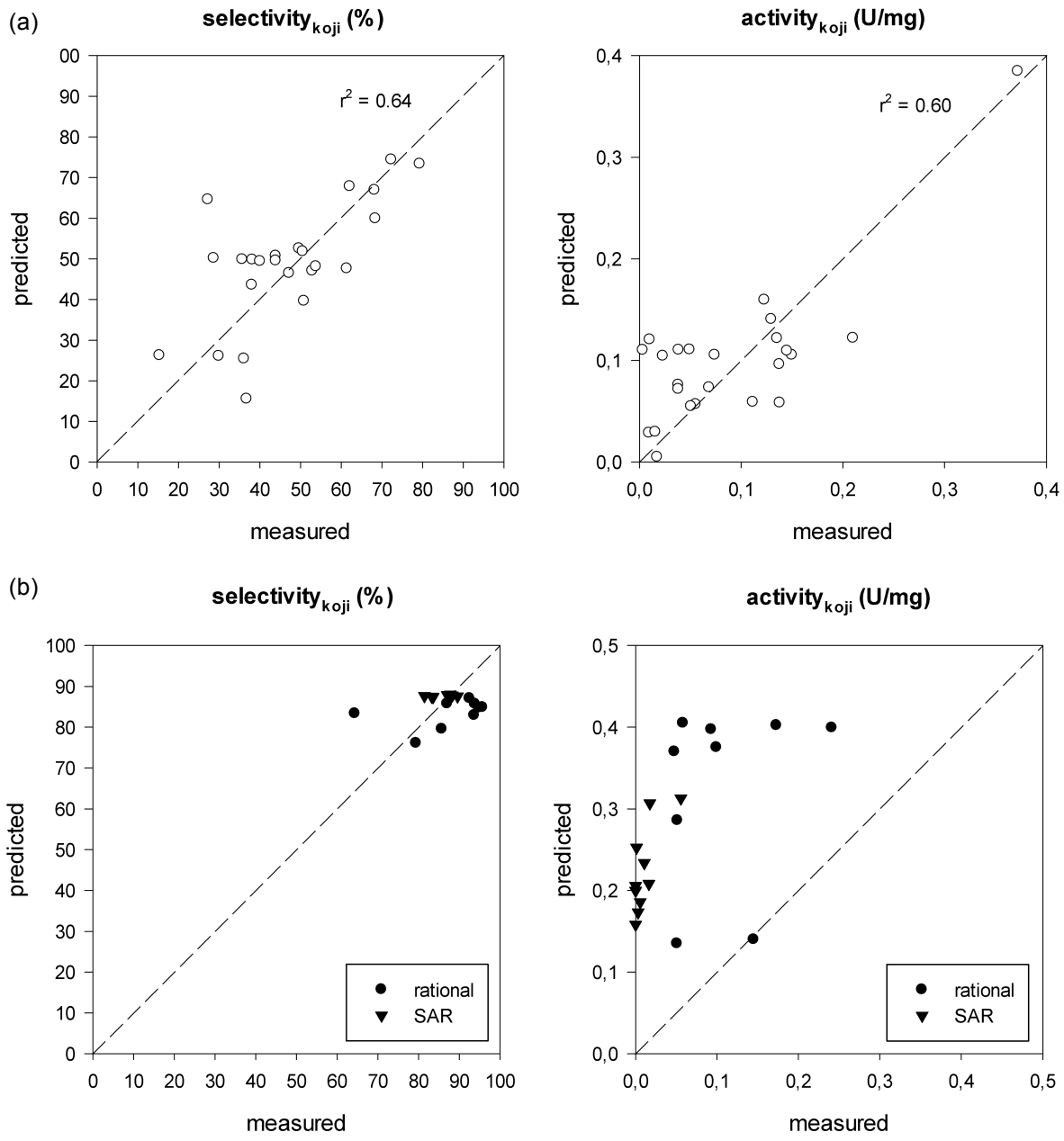

Figure S3. Evaluation of the prediction model that fitted the input data best (AFSAR). Predicted versus *in vitro* measured values for (a) the training set of single mutants (leave one out cross validation (LOOCV)) and (b) predicted best combinations and rational combinations. Despite the high r-squared value in the LOOCV (meaning a good fit of the model to the input data) the activity was often largely overestimated for new combinations. Note that in contrast to classical sequence-activity relationship (SAR) prediction, the algorithm applied here does not necessarily need large training sets (extra parameters like e.g. amino acid properties are taken into account). For selectivity for instance, predictions were fairly accurate. For activity on the contrary this was not the case, but this is most likely due to the complexity of the variable (see also docking results), rather than the size of the training set (although it is not excluded that increasing the size could help to improve the prediction).

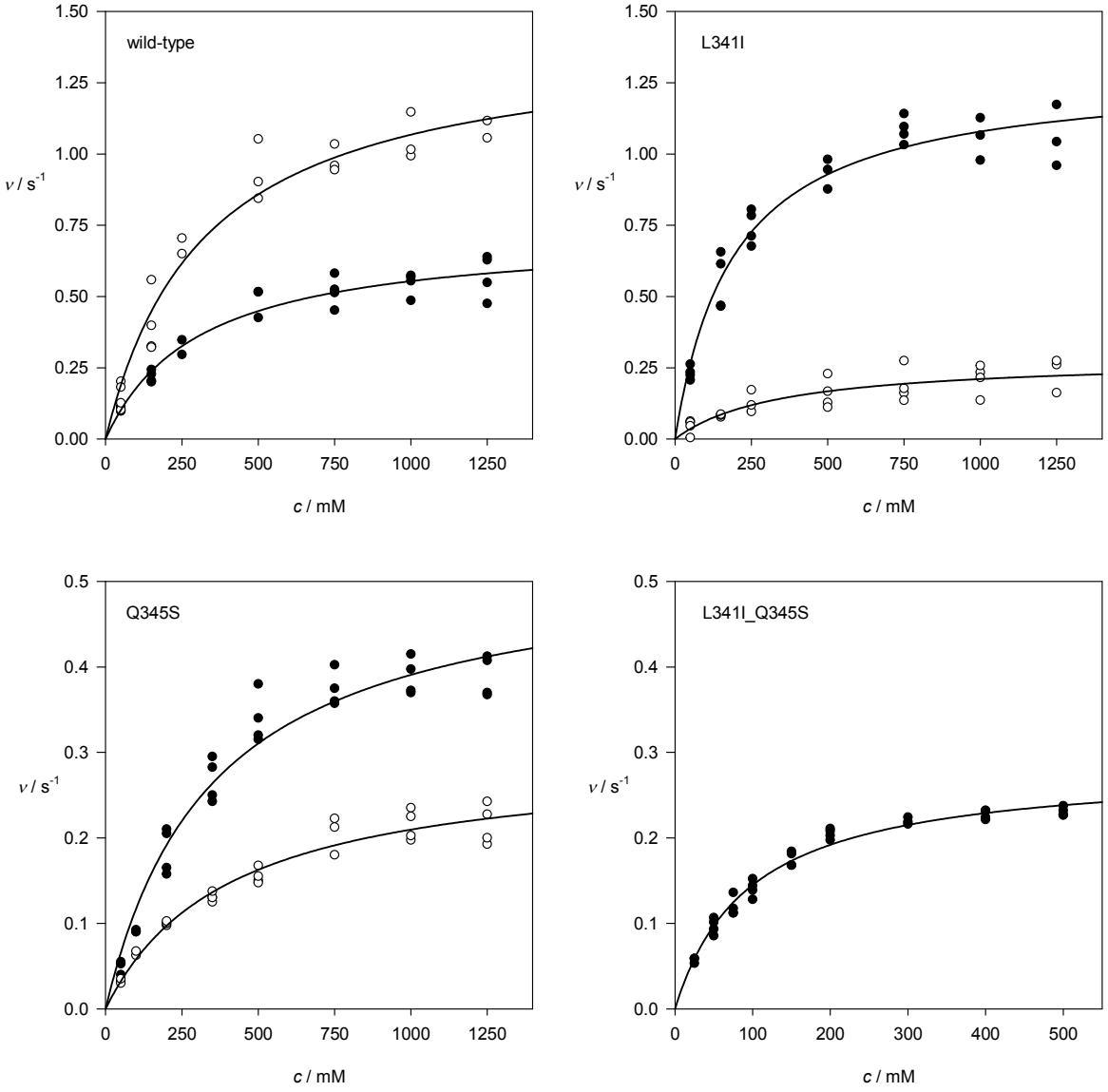

Figure S4. Michaelis-Menten curves for the wild-type enzyme, single mutants L341I and Q345S and double mutant L341I_Q345S. (formation of ● kojibiose and ○ maltose).

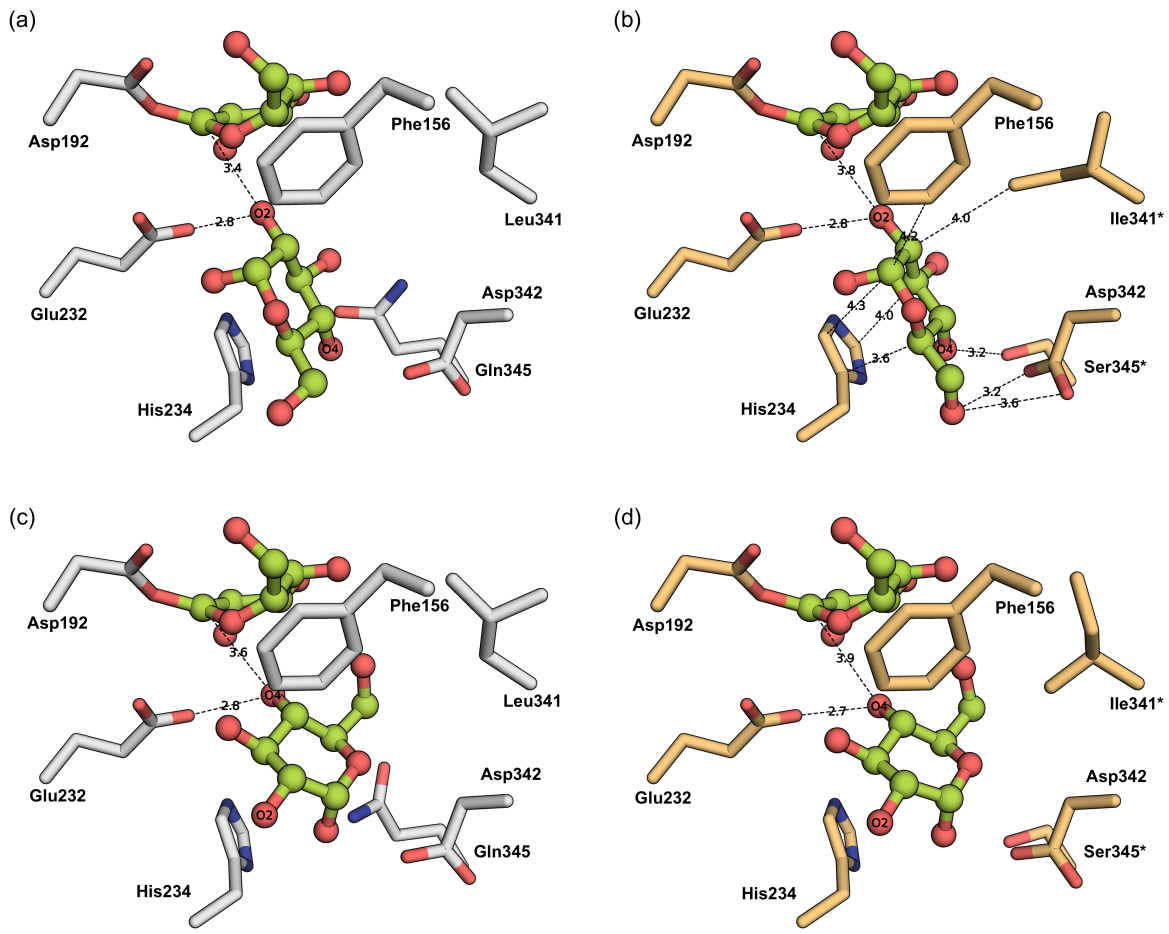

Figure S5. Docking of D-glucose into the covalent glycosyl-enzyme intermediate of the wild-type structure (a,c) and L341I_Q345S mutant model (b,d). The wild-type enzyme can accommodate glucose in an orientation suitable for the formation of either kojibiose (a) or maltose (c), with the C2-OH and C4-OH, respectively, within hydrogen bonding distance from the acid/base catalyst Glu232 and close to the anomeric carbon of the donor substrate. In the L341I_Q345S mutant, these two binding modes are essentially maintained. In the kojibiose-forming mode, the stacking and hydrophobic interaction with His234 and Phe156, respectively, are conserved and an additional hydrophobic interaction with the side chain of Ile341 is observed. The introduced Ser at position 345 moreover forms an extra hydrogen bond (with O4) and created space for the C6 hydroxyl group, which can now hydrogen bond with Asp342 (b). These extra interactions could be an explanation for the higher affinity of the L341I_Q345S mutant in the kojibiose-forming mode. In the maltose-forming mode, however, no changes in interaction are observed that could hamper binding and thus explain the lowered affinity compared to the wild-type enzyme (d) (green ball/stick: glucose; white sticks: wild-type enzyme; light orange sticks: L341I_Q345S mutant; catalytic residues: Asp192 (nucleophile) and Glu232 (general acid/base); asterisk: mutated residue).

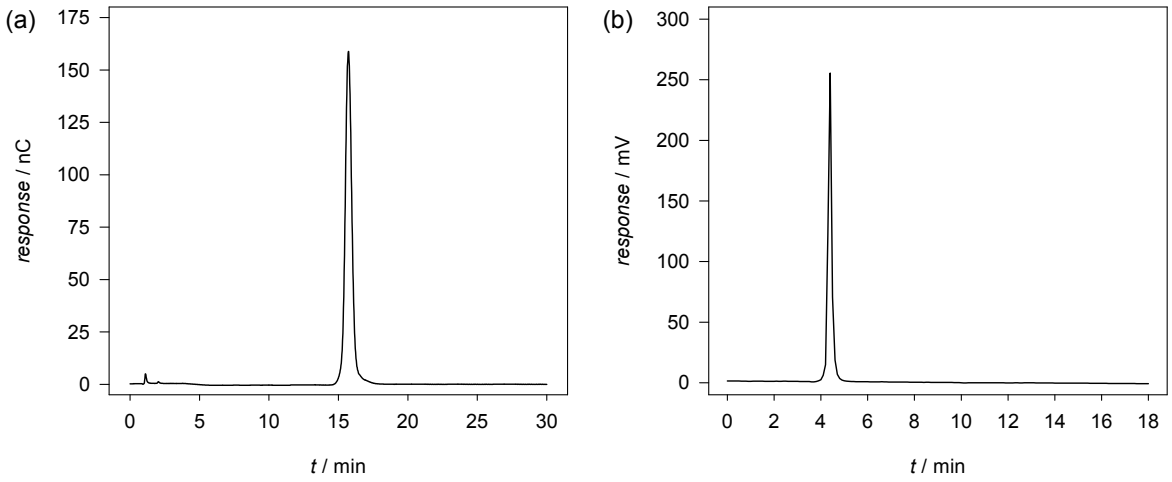

Figure S6. HPAEC-PAD (a) and HPLC-RI (b) profiles of the purified kojibiose (note that the small peak at 1 min in the HPAEC profile is the injection peak)

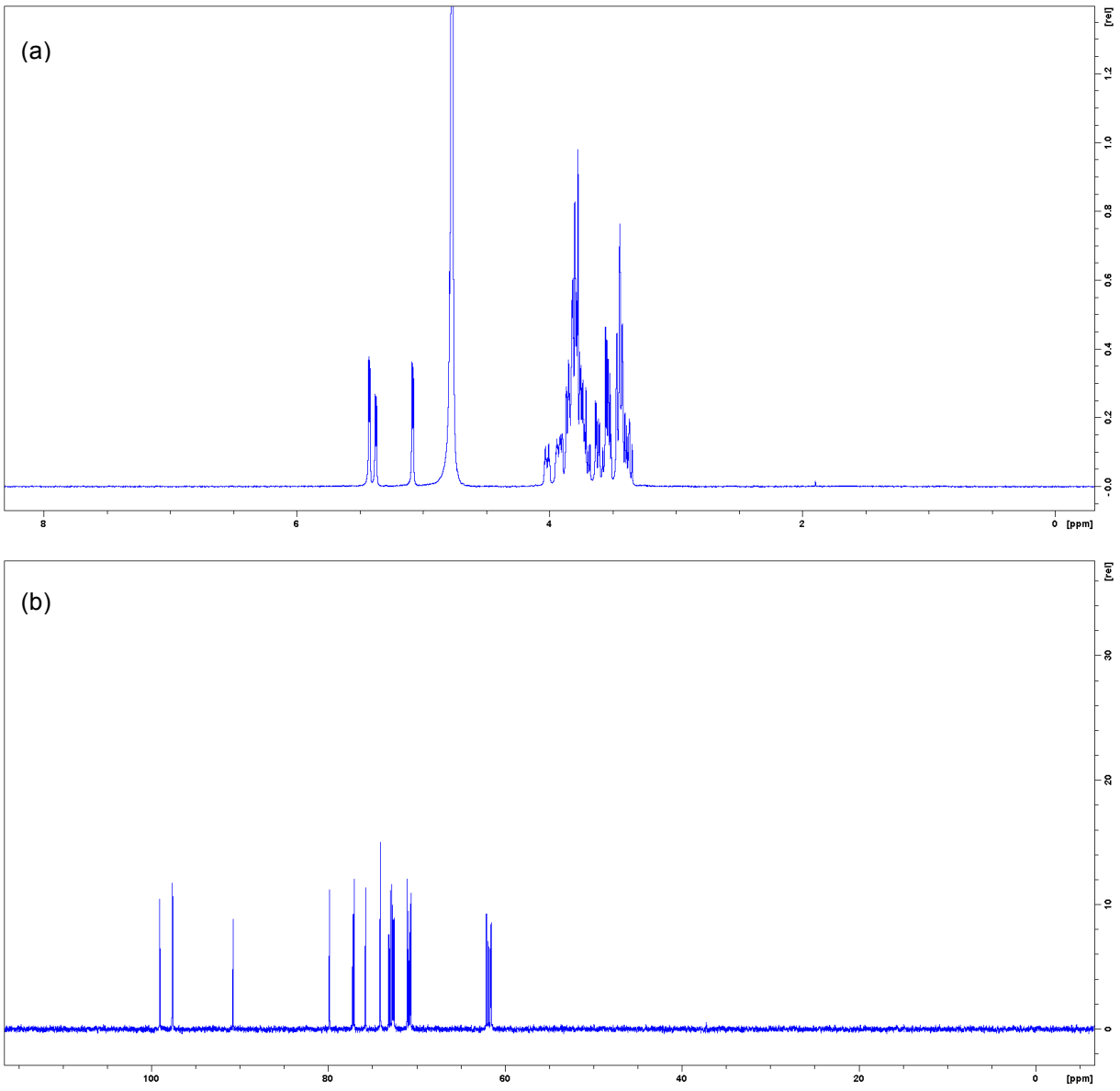

Figure S7. ^{1}H NMR (a) and ^{13}C NMR (b) spectrum of the purified kojibiose, consistent with reported literature[19] (^{1}H NMR spectra were recorded at 400 MHz and ^{13}C NMR spectra at 100.6 MHz (both in D_2O) using a BRUKER AVANCE III-400).

TABLES

Table S1. Predicted and measured selectivity and activity of the wild-type enzyme, single mutants obtained from screening, alanine mutants (previously created)[18], combinations predicted to be most promising (ProSAR_#) and rational recombinations (selectivity: fraction of kojibiose in total product formation; activity: formation of kojibiose; all reactions performed with His-tag purified enzyme, 100 mM sucrose, 200 mM glucose, 55°C, pH 7.0).

type	mutations										predicted		measured	
	132	134	135	233	234	341	342	343	344	345	selectivity (%)	activity ($U.mg^{-1}$)	selectivity (%)	activity ($U.mg^{-1}$)
wild-type[a]	Y	P	R	V	H	L	D	L	Y	Q	50%	0.11	36%	0.15
single[a]	.	.	.	.	.	I	.	.	.	.	79%	0.37	79%	0.37
single[a]	.	.	.	.	.	.	.	.	.	N	73%	0.11	72%	0.01
single[a]	.	V	.	.	.	.	.	.	.	.	68%	0.14	68%	0.14
single[a]	.	.	.	.	.	.	.	.	.	S	67%	0.11	68%	0.14
single[a]	.	R	.	.	.	.	.	.	.	.	62%	0.02	62%	0.02
single[a]	.	.	.	.	.	.	.	.	I	.	59%	0.03	61%	0.04
single[a]	.	.	.	.	.	.	.	.	V	.	54%	0.10	54%	0.07
single[a]	.	.	.	.	.	.	.	P	.	.	53%	0.03	53%	0.02
single[a]	.	.	.	.	T	.	.	.	.	.	51%	0.003	51%	0.00
single[a]	.	.	E	.	.	.	.	.	.	.	44%	0.01	44%	0.02
single[a]	.	.	.	.	.	.	.	.	F	.	42%	0.15	38%	0.12
single[a]	.	.	.	.	.	.	.	.	R	.	41%	0.07	44%	0.07
single[a]	.	.	V	.	.	.	.	.	.	.	38%	0.06	38%	0.04
single[a]	.	.	.	.	.	.	.	.	D	.	37%	0.04	40%	0.05
single[a]	.	.	P	.	.	.	.	.	.	.	15%	0.02	15%	0.01
alanine[a, b]	.	.	.	.	.	.	.	.	A	.	51%	0.14	50%	0.21
alanine[a, b]	.	.	.	.	.	.	.	.	.	A	48%	0.10	51%	0.11
alanine[a, b]	.	.	.	.	.	.	A	.	.	.	47%	0.05	47%	0.05
alanine[a, b]	.	.	A	.	.	.	.	.	.	.	37%	0.11	37%	0.14
alanine[a, b]	.	.	.	A	.	.	.	.	.	.	37%	0.14	36%	0.15
alanine[a, b]	A	.	.	.	.	.	.	.	.	.	27%	0.04	27%	0.04
alanine[a, b]	.	.	.	.	.	A	.	.	.	.	26%	0.06	30%	0.06
alanine[a, b]	.	.	.	.	.	.	.	A	.	.	26%	0.15	29%	0.13
ProSAR_1	.	V	.	.	.	I	.	P	I	N	88%	0.23	88%	0.01
ProSAR_2	.	V	.	.	.	I	A	P	I	N	88%	0.17	87%	0.003
ProSAR_3[c]	.	V	.	.	T	I	.	.	I	N	88%	0.21	-	-
ProSAR_4	.	V	.	.	.	I	.	.	I	N	88%	0.31	88%	0.06
ProSAR_5	.	V	.	.	.	I	A	.	I	N	88%	0.25	81%	0.001
ProSAR_6[c]	.	V	.	.	T	I	.	P	V	N	88%	0.20	-	-
ProSAR_7	.	V	.	.	.	I	.	P	V	N	88%	0.31	90%	0.02
ProSAR_8	.	V	E	.	.	I	.	.	I	N	87%	0.21	83%	0.02
ProSAR_9[c]	.	V	.	A	T	I	.	P	I	N	87%	0.16	-	-
ProSAR_10	.	V	V	.	.	I	.	P	I	N	87%	0.19	84%	0.01

rational	.	V	.	.	.	I	.	.	.	N	87%	0.40	93%	0.09
rational	.	V	.	.	.	I	.	.	.	S	86%	0.40	87%	0.17
rational	.	.	.	.	.	I	.	.	.	N	85%	0.37	95%	0.05
rational	.	V	.	.	.	I	.	.	.	.	83%	0.40	65%	0.24
rational	.	.	.	.	.	I	.	.	.	S	83%	0.38	94%	0.10
rational	.	V	.	.	.	.	.	.	.	N	80%	0.13	86%	0.05
rational	.	V	.	.	.	.	.	.	.	S	76%	0.14	78%	0.15

[a] training set for the statistical model, [b] created in previous research[18] and analysed here for activity on glucose, [c] no expression

REFERENCES

1. D. Aerts, T. Verhaeghe, M. De Mey, T. Desmet and W. Soetaert, *Eng. Life Sci.*, 2011, **11**, 10-19.
2. P. Coussement, J. Maertens, J. Beauprez, W. Van Bellegem and M. De Mey, *Metab. Eng.*, 2014, **23**, 70-77.
3. D. G. Gibson, G. A. Benders, K. C. Axelrod, J. Zaveri, M. A. Algire, M. Moodie, M. G. Montague, J. C. Venter, H. O. Smith and C. A. Hutchison, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 20404-20409.
4. B. Sullivan, A. Z. Walton and J. D. Stewart, *Enzyme Microb. Technol.*, 2013, **53**, 70-77.
5. J. Sanchis, L. Fernandez, J. Carballeira, J. Drone, Y. Gumulya, H. Hobenreich, D. Kahakeaw, S. Kille, R. Lohmer, J. Peyralans, J. Podtetenieff, S. Prasad, P. Soni, A. Taglieber, S. Wu, F. Zilly and M. Reetz, *Appl. Microbiol. Biotechnol.*, 2008, **81**, 387-397.
6. M. Berland, B. Offmann, I. André, M. Remaud-Siméon and P. Charton, *Protein Eng. Des. Sel.*, 2014, **27**, 375-381.
7. M. Berland, PhD PhD thesis, Université de la Réunion, 2013.
8. R. J. Fox, S. C. Davis, E. C. Mundorff, L. M. Newman, V. Gavrilovic, S. K. Ma, L. M. Chung, C. Ching, S. Tam, S. Muley, J. Grate, J. Gruber, J. C. Whitman, R. A. Sheldon and G. W. Huisman, *Nat. Biotechnol.*, 2007, **25**, 338-344.
9. R. J. Fox and G. W. Huisman, *Trends Biotechnol.*, 2008, **26**, 132-138.
10. M. R. M. De Groeve, M. De Baere, L. Hoflack, T. Desmet, E. J. Vandamme and W. Soetaert, *Protein Eng. Des. Sel.*, 2009, **22**, 393-399.
11. M. T. Reetz, D. Kahakeaw and R. Lohmer, *ChemBioChem*, 2008, **9**, 1797-1804.
12. D. Aerts, T. Verhaeghe, B. Roman, C. Stevens, T. Desmet and W. Soetaert, *Carbohydr. Res.*, 2011, **346**, 1860-1867.
13. E. Krieger and G. Vriend, *Bioinformatics*, 2014, **30**, 2981-2982.
14. O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455-461.
15. A. A. Canutescu, A. A. Shelenkov and R. L. Dunbrack, Jr., *Protein Sci.*, 2003, **12**, 2001-2014.
16. E. Krieger, K. Joo, J. Lee, J. Lee, S. Raman, J. Thompson, M. Tyka, D. Baker and K. Karplus, *Proteins: Struct., Funct., Bioinf.*, 2009, **77**, 114-122.
17. O. Mirza, L. K. Skov, D. Sprogoe, L. A. M. van den Broek, G. Beldman, J. S. Kastrup and M. Gajhede, *J. Biol. Chem.*, 2006, **281**, 35576-35584.
18. T. Verhaeghe, M. Diricks, D. Aerts, W. Soetaert and T. Desmet, *J. Mol. Catal. B: Enzym.*, 2013, **96**, 81-88.
19. D. Goffin, P. Bystricky, A. S. Shashkov, M. Lynch, E. Hanon, M. Paquot and A. V. Savage, *Bull. Korean Chem. Soc.*, 2009, **30**, 2535-2541.