

Supplementary information for

Elucidation of the noncovalent interactions driving enzyme activity guides branching enzyme engineering for α -glucan modification

Zhiyou Zong^{1,2,#,*}, Xuewen Zhang^{1,2,#}, Peng Chen^{1,2,#}, Zhuoyue Fu^{1,2}, Yan Zeng^{1,2}, Qian Wang^{1,2},
Christophe Chipot^{3,4,5}, Leila Lo Leggio⁶ & Yuanxia Sun^{1,2,*}

¹ National Engineering Research Center of Industrial Enzymes, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China.

² National Center of Technology Innovation for Synthetic Biology, Tianjin 300308, China.

³ Laboratoire International Associé CNRS and University of Illinois at Urbana–Champaign, LPCT, UMR 7019 Université de Lorraine CNRS, Vandœuvre-lès-Nancy F-54500, France

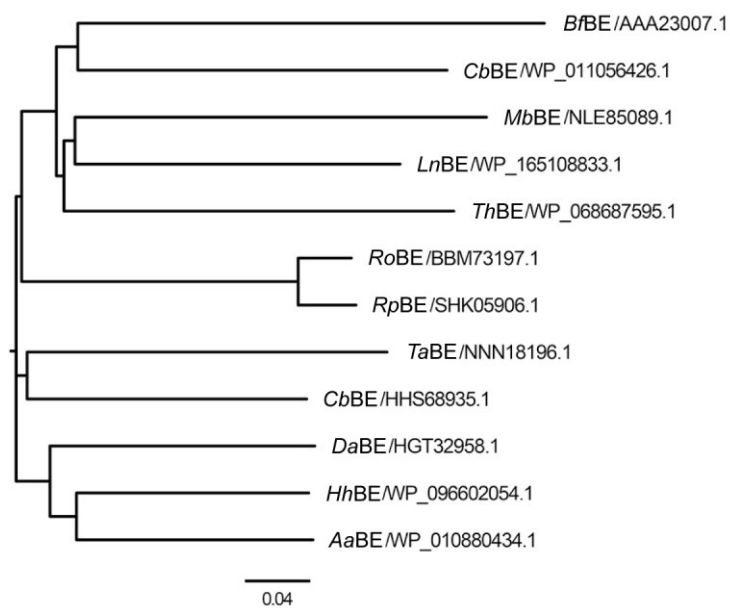
⁴ Department of Physics, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801, United States

⁵ Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, Illinois 60637, United States

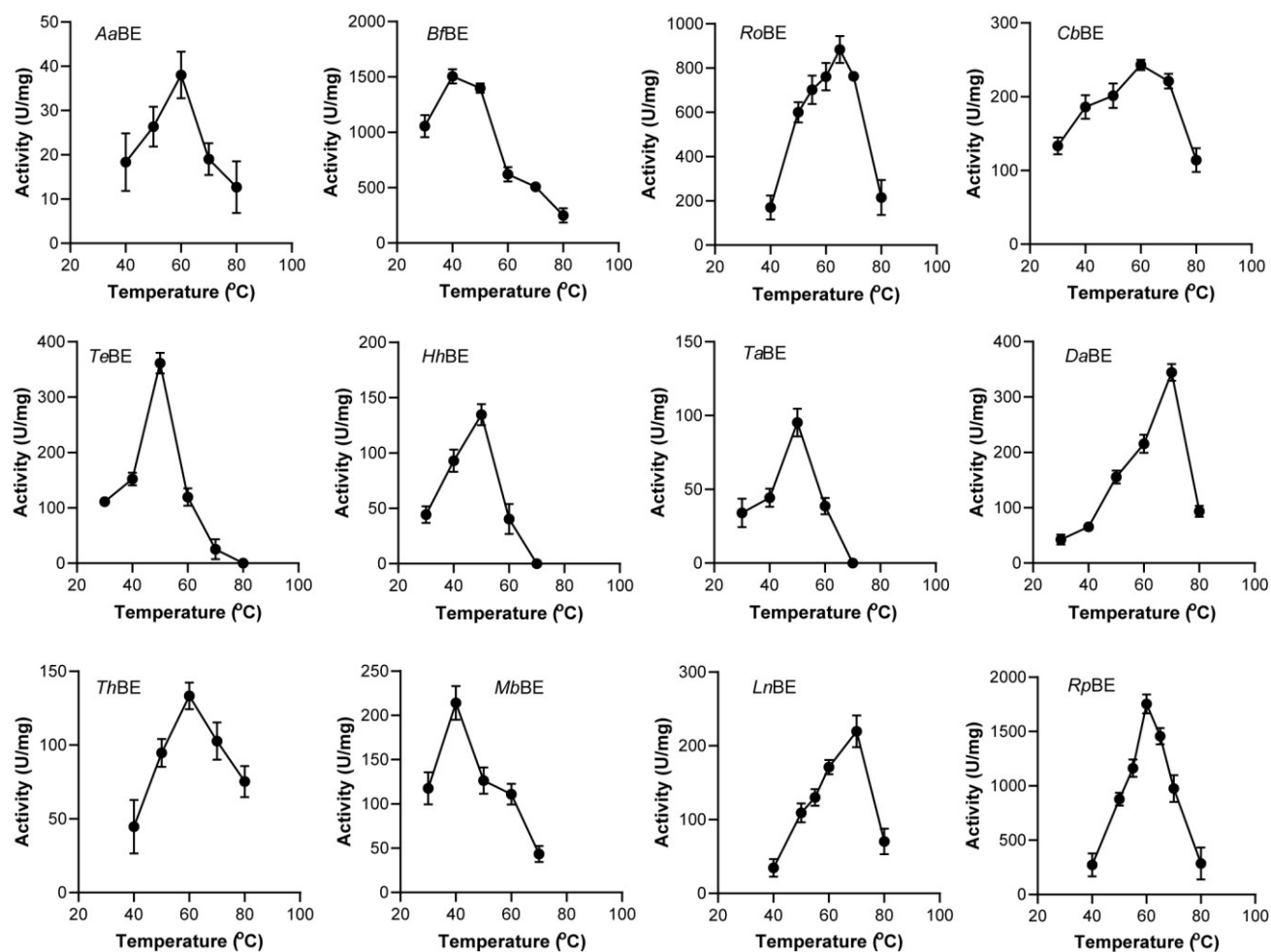
⁶ Department of Chemistry, University of Copenhagen, DK-2100 Copenhagen, Denmark.

***Corresponding E-mail:** zongzhy@tib.cas.cn; sun_yx@tib.cas.cn

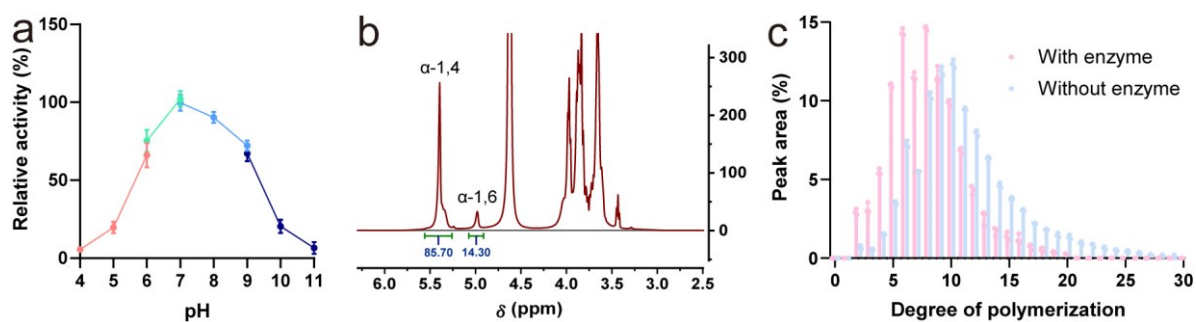
Supplementary Figures 1-17



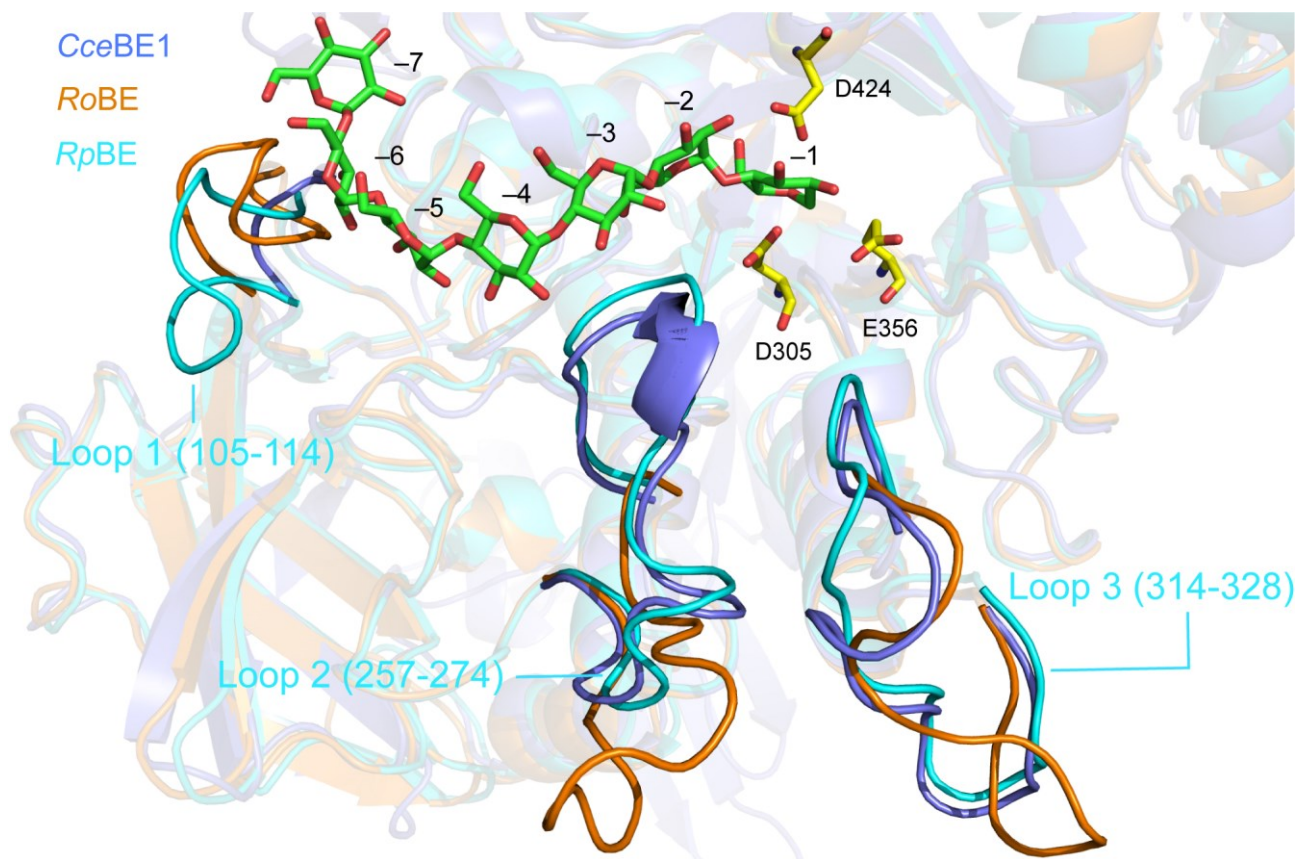
Supplementary Fig. 1 Phylogenetic tree of twelve BEs by Clustal Omega software. *Rp*BE has the highest homology with *Ro*BE.



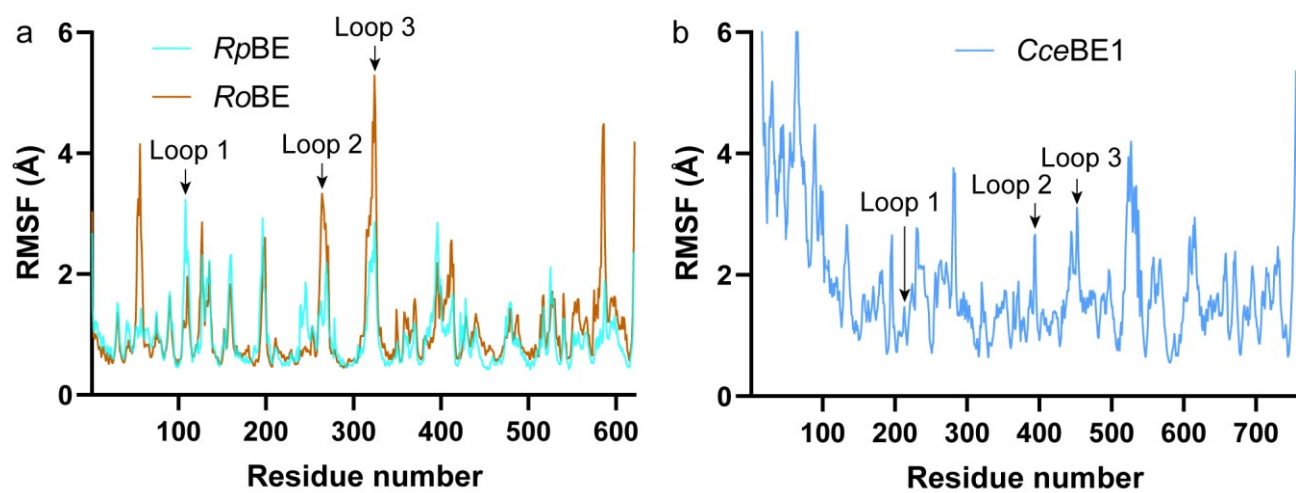
Supplementary Fig. 2 Thermal dependence of activity of the twelve BEs. The activity was measured using amylose for 20 minutes. The bars and the error bars are the average and standard deviation of triplicates ($n = 3$ independent experiments), respectively. Source data are provided as a Source Data file.



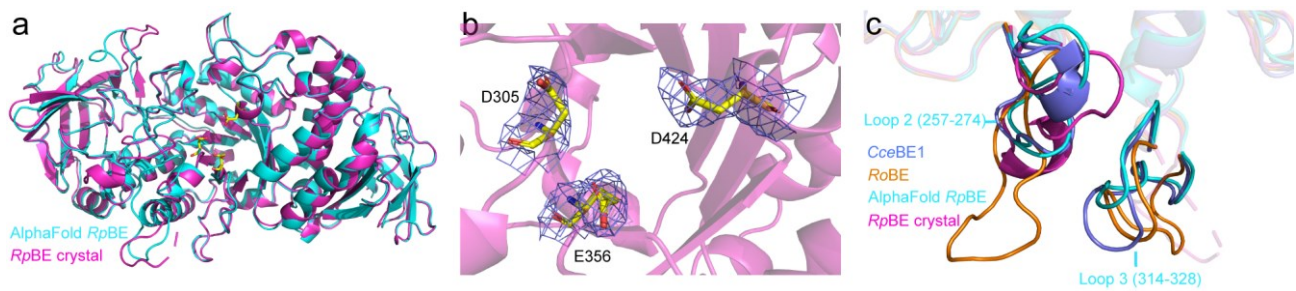
Supplementary Fig. 3 Determination of catalytic properties for *RpBE*. **a** pH-dependence of activity. **b** Nuclear magnetic resonance spectroscopy analysis of branched product from amylose. **c** Chain length distribution of the substrate before and after catalysis by *RpBE*. In **(a)**, 50 mM citric acid-sodium citrate buffer (pH 4.0-6.0), phosphate buffer (pH 6.0-7.0), Tris-HCl buffer (pH 7.0-9.0), and glycine-NaOH buffer (pH 9.0-11.0) were employed to measure the enzyme activity in different pH conditions. In **(b)**, α -1,4 and α -1,6 denote corresponding glucosidic linkages. The product of *RpBE* action on amylose possesses 14.3% branching degree. In **(c)**, the substrate is debranched amylopectin (i.e., linear dextrans or short-chain amylose). After enzyme catalysis, the degree of polymerization (DP) ≥ 10 glycan chains were consumed. G10 is, therefore, regarded as the minimum donor substrate required for *RpBE* catalysis. The bars and the error bars are the average and standard deviation of triplicates ($n = 3$ independent experiments), respectively. Source data for a, c are provided as a Source Data file.

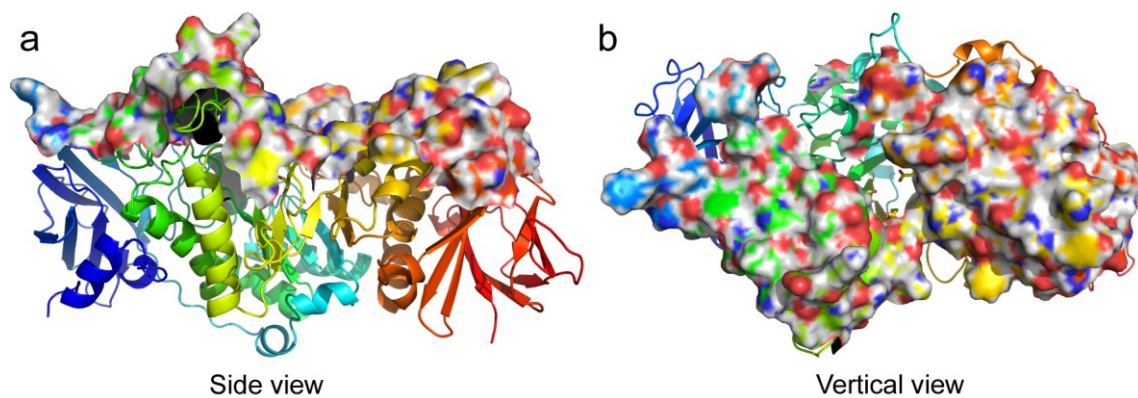


Supplementary Fig. 4 Alignment of the three loops (loop 1: 105-114, loop 2: 257-274, and loop 3: 314-328; *RpBE* numbering) between *CceBE1*, *RoBE*, and *RpBE*. The catalytic triad of *RpBE* is shown in a licorice representation and their carbon atoms are colored in yellow. *RpBE* loops 2 and 3 align well with the corresponding ones in *CceBE1*, whereas the conformation of loop 1 is distinct from that of the two others.

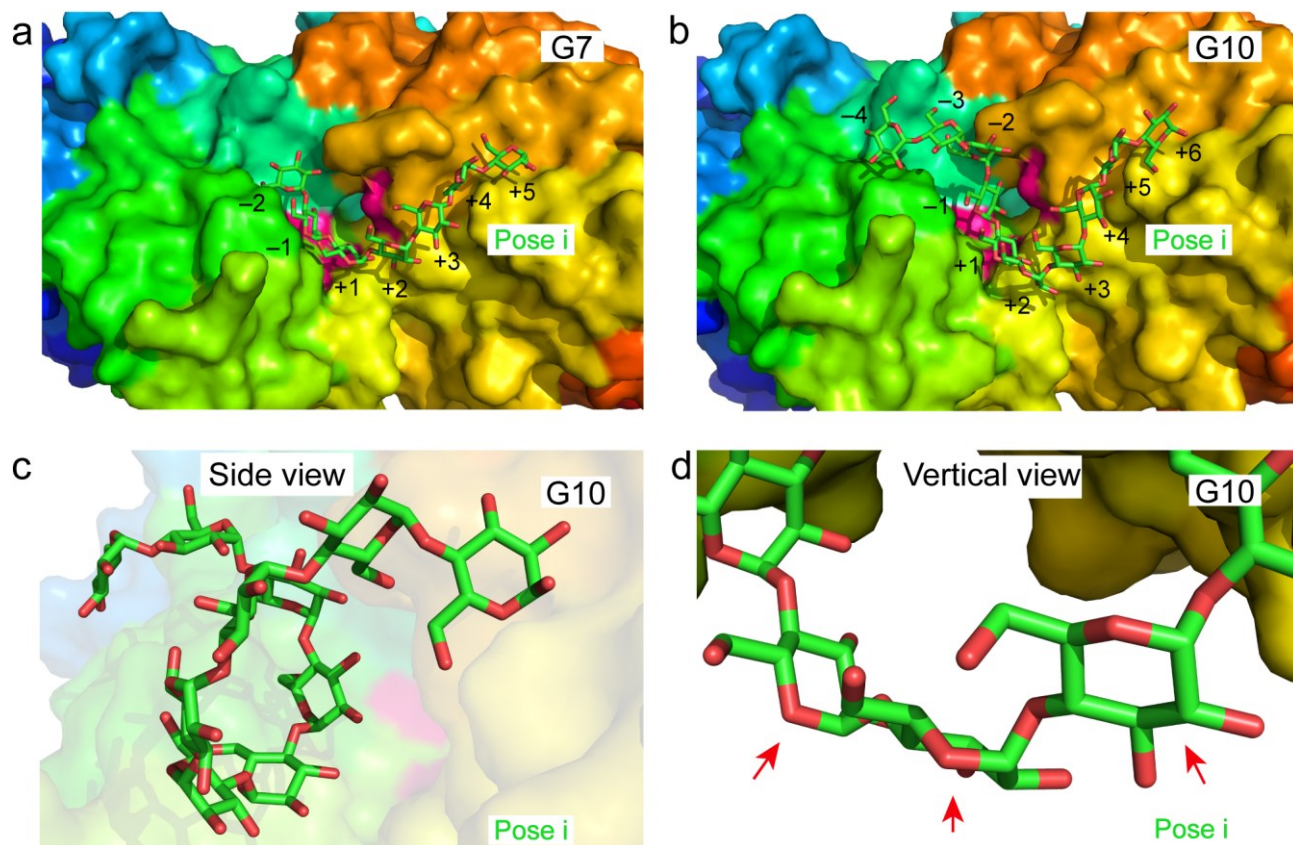


Supplementary Fig. 5 RMSF values. **a** RMSF of *RpBE* with *RoBE*. **b** RMSF of *CceBE1*. RMSF were obtained from each 500-ns MD trajectory. The flexibility of *RoBE* loops 2 and 3 is the highest in the three BEs.

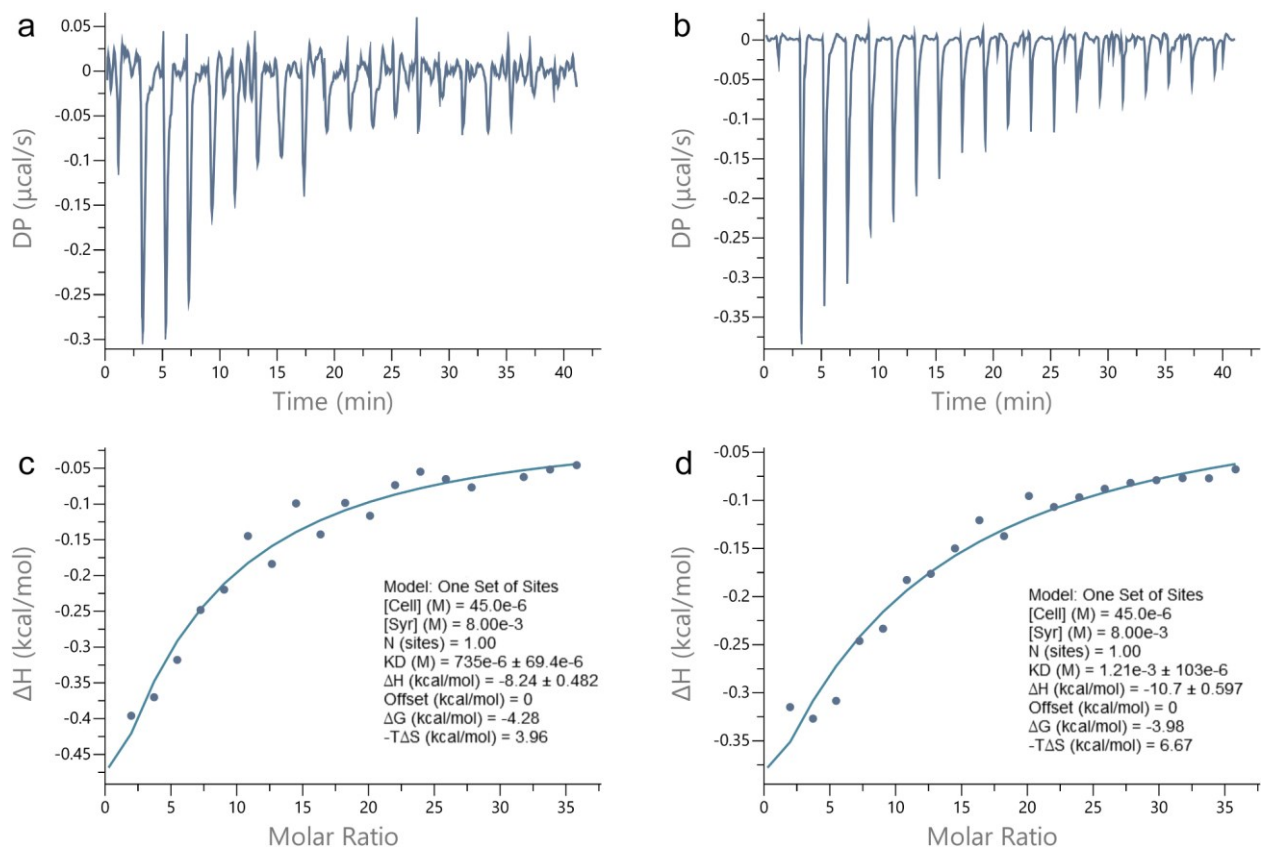




Supplementary Fig. 7 Residues used for docking before glycosylation. 138 residues, which lie on the side of the active-site cleft, were selected as the docking surface. The residues are presented as the electrostatic potential surface using PyMOL software, and shown in (a) side and (b) vertical views, respectively. Catalytic residues are displayed in a licorice representation and colored in yellow (carbon atoms).

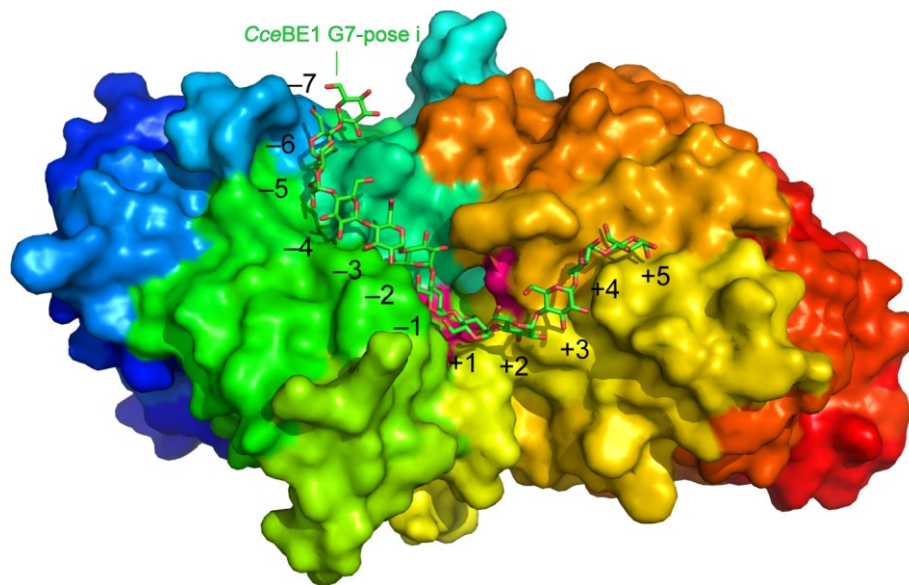


Supplementary Fig. 8 G7 and G10 binding to *RpBE*. **a** G7 binding to *RpBE*. **b** G10 binding to *RpBE*, and **c** its side and **d** vertical views. In the *RpBE*-G10 complex, three glucosyl rings, which are denoted by red arrows, do not bind to the enzyme.

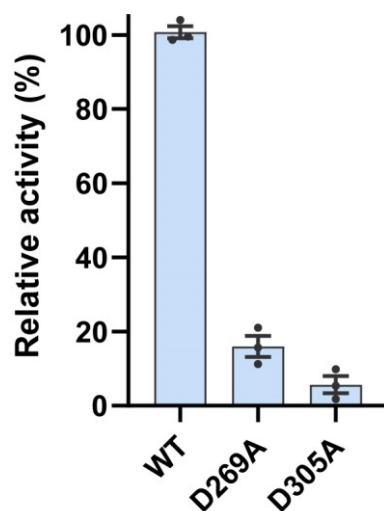


Supplementary Fig. 9 ITC test. a b Titration and **c d** fitting curves. The *RpBE-G7* affinity is approximately -4.1 ± 0.2

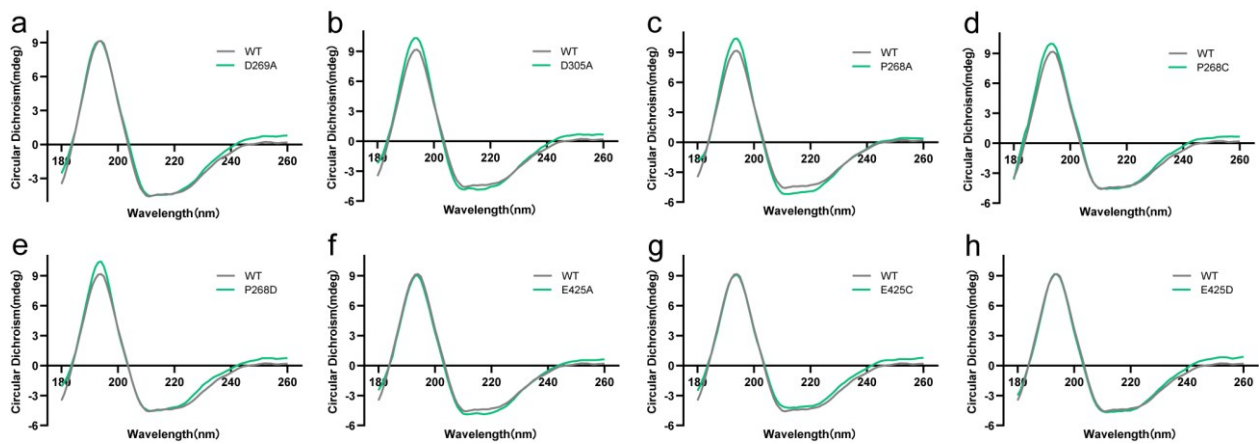
kcal/mol. Data are presented as mean value and standard error (\pm) inferred from two independent runs.



Supplementary Fig. 10 An assumed conformation of a long-chain donor substrate binding to *RpbE*. The -2 to +5 glucosyl rings and -7 to -3 glucosyl rings of *CceBE1* G7-pose i are modeled by referring to the substrate conformations in pose i (-2 to +5 glucosyl rings) and *CceBE1*-G7 (-7 to -3 glucosyl rings), respectively.



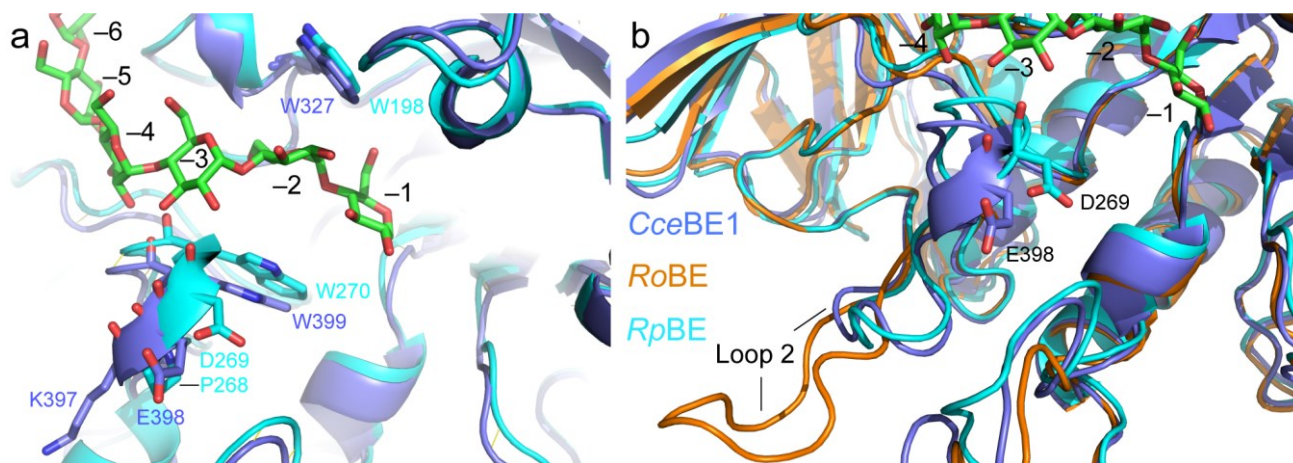
Supplementary Fig. 11 Effect of replacement of D269 and D305 of *RpBE* by alanine. The WT *RpBE* is regarded to possess 100% enzyme activity. The results suggest that residues D269 and D305 are critical to enzyme activity. The bars and the error bars are the average and standard deviation of triplicates ($n = 3$ independent experiments), respectively. Source data are provided as a Source Data file.



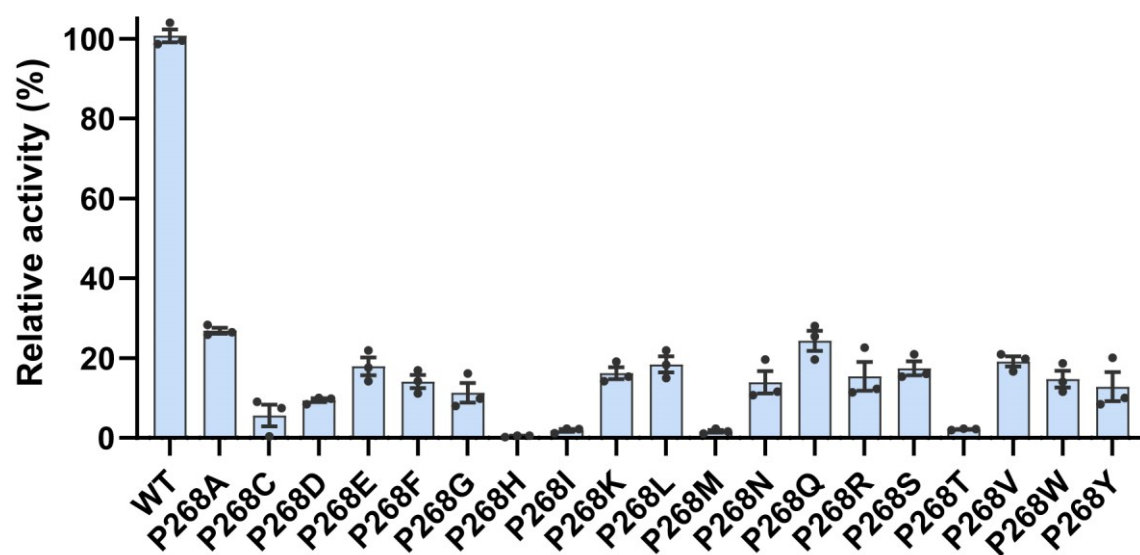
Supplementary Fig. 12 CD spectroscopy experiments for the WT *RpBE* and the variants. a D269A. b D305A.

c P268A. d P268C. e P268D. f E425A. g E425C. h E425D. The results indicate that the variants are properly folded.

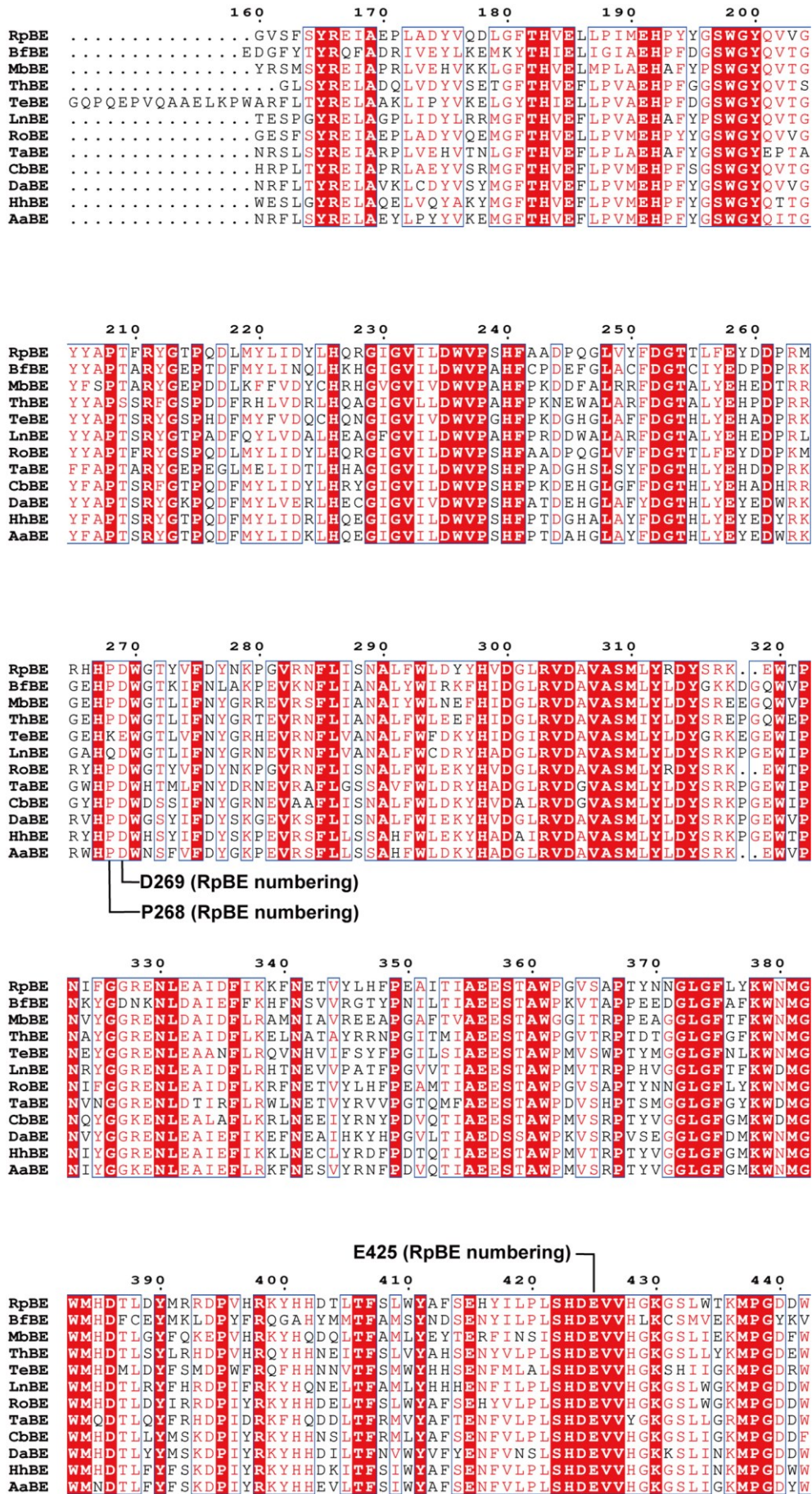
Source data are provided as a Source Data file.



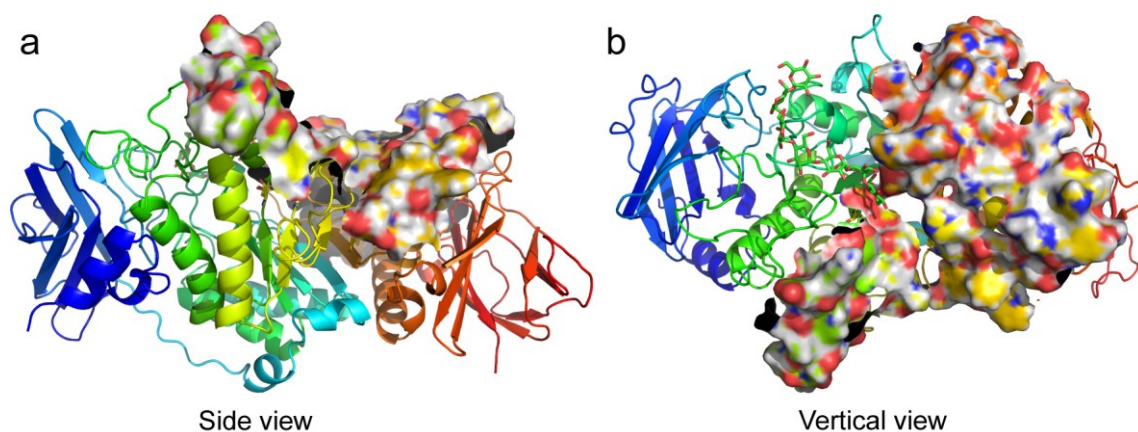
Supplementary Fig. 13 Alignment *RpBE* with *RoBE* and *CceBE1*. **a** Structure alignment of *RpBE* with *CceBE1* (PDB code 5GQU). **b** Alignment of *RpBE* (cyan) with *CceBE1* (blue) and *RoBE* (orange; PDB code 6JOY). The glucan chain, which is shown in a licorice representation and colored in green (carbon atoms), is from the enzyme-substrate complex of *CceBE1*, lying in the active-site cleft.



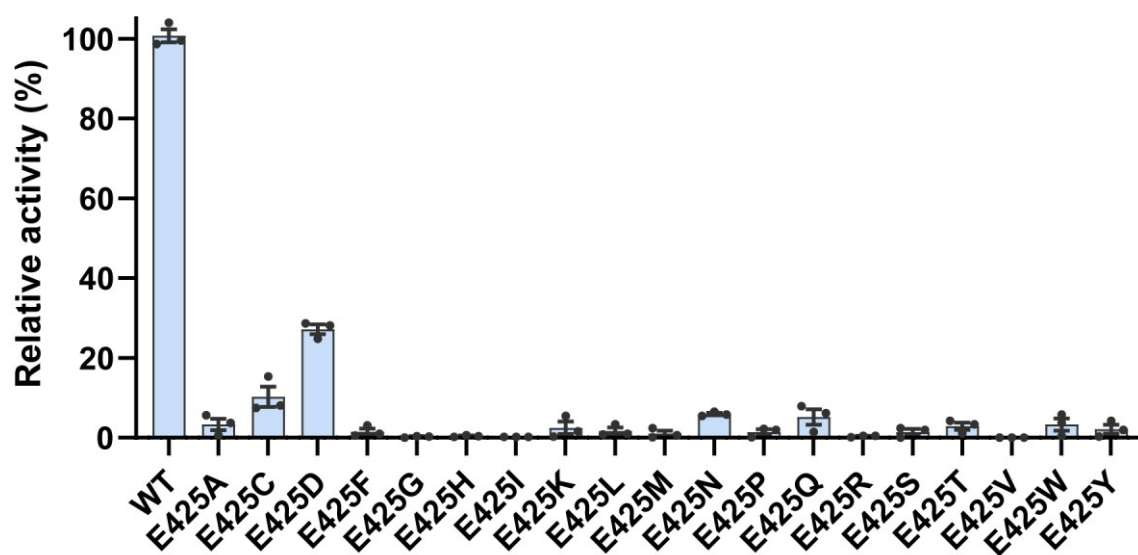
Supplementary Fig. 14 Saturation mutagenesis on the 268 site. The WT *RpbE* is regarded to possess 100% enzyme activity. The results suggest that the P268 in *RpbE* is the optimal residue. The bars and the error bars are the average and standard deviation of triplicates ($n = 3$ independent experiments), respectively. Source data are provided as a Source Data file.



Supplementary Fig. 15 Partial sequence alignments between the twelve BEs investigated. Residues P268, D269, and E425 (*RpBE* numbering) are labeled by black lines.



Supplementary Fig. 16 Residues used for docking before the transglycosylation step. The active-site cleft has been occupied by the donor polysaccharide chain, the carbon atoms of which are colored in green. The 85 residues located around the side of the active-site cleft were selected as the docking surface and shown in **(a)** side and **(b)** vertical views, respectively. The residues are presented as the electrostatic potential surface. Catalytic residues are displayed in a licorice representation and colored in yellow (carbon atoms).



Supplementary Fig. 17 Saturation mutagenesis on the 425 site. The WT *RpbE* is regarded to possess 100% enzyme activity. The results suggest that the E425 in *RpbE* is the optimal residue, which natural evolution has identified. The bars and the error bars are the average and standard deviation of triplicates ($n = 3$ independent experiments), respectively. Source data are provided as a Source Data file.

Supplementary Tables 1-14

Supplementary Table 1 Homology of *RpBE* with BE crystals^a.

No.	Coverage	Identity	Method	PDB ID	Origin
1	99%	93.4	X-ray, 2.4Å	6JOY	<i>Rhodothermus obamensis</i> STB05
2	85%	52.3	X-ray, 1.8Å	5GQU	<i>Cyanothece sp.</i> ATCC 51142

^a Top two BEs with the highest homology to *RpBE*.

Supplementary Table 2 Estimation of *B*-factor and residue-level quality^a.

Protein	r<i>B</i>-factor^b	n<i>B</i>-factor^c	TM-score^d	RMSD^e	Quality1^f	Quality2^g	Quality3^h
<i>RpBE</i>	25.89	0.03	0.91	3.23 Å	2.35 Å	2.03 Å	2.35 Å
<i>RoBE</i> ⁱ	25.85	0.03	0.91	3.36 Å	2.40 Å	2.10 Å	2.40 Å

^a The quality was defined as the distance deviation (Å) between residue positions in the model and the native structure. All data is displayed as an average.

^b Raw *B*-factor. Atoms with low *B*-factors are well ordered.

^c The value is defined as z-score-based normalization of the raw *B*-factor. Residues with n*B*-factor less than or equal to 0 are much stable.

^d TM-score is a metric for assessing the topological similarity of protein structures and has the value in 0 to 1, wherein the closer the value is to 1, the better.

^e Root mean square deviation.

^f Quality estimated by LOMETS threading templates. LOMETS is a local meta-threading-server for protein structure prediction.

^g Quality estimated by TM-align structure alignment templates.

^h Quality estimated by support vector regressions.

ⁱ Crystal structure of *RoBE* (PDB code 6JOY) as a control.

Supplementary Table 3 Crystallographic statistics.

Data Collection	
Wavelength (Å)	0.979
Space group	P1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	100.302, 112.266, 166.26
<i>α</i> , <i>β</i> , <i>γ</i> (°)	73.59, 73.77, 65.25
Resolution (Å)	50 - 2.91 (2.94 - 2.91)
<i>R</i> _{merge} ¹	0.26 (1.15)
Unique reflections	135389 (4521)
Redundancy	3.5 (3.5)
Completeness (%)	91.63 (89.54)
Mean I/σI	3.6 (1.1)
CC _{1/2} (%)	93.3 (41.9)
Refinement	
<i>R</i> _{work} (95% data)	0.1971 (0.2818)
<i>R</i> _{free} (5% data)	0.2560 (0.3527)
RMSD from ideal geometry ²	
Bond length (Å)	0.006
Bond angles (°)	1.12
Ramachandran statistics ³	
Favored (%)	96.59
Allowed (%)	3.41
Outliers (%)	0.00
No. atoms/B-factors	
Protein	39123/38.90
Water	272/35.10
PDB accession	8ZQA

¹ $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, wherein $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl*, and $\langle I(hkl) \rangle$ is the mean value of $I_i(hkl)$ for all the *i* measurements.

² Root mean square deviations from ideal geometry values¹.

³ Calculated by Phenix Refine².

Supplementary Table 4 Selected 138 residues in N, A, and C domains of *RpBE* for step i.

Domain	Residue number
N	88, 89, 90, 91, 92, 108, 109, 110, 111
A	152, 153, 196, 197, 211, 242, 243, 244, 245, 249, 250, 253, 254, 255, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 327, 328, 329, 330, 358, 359, 360, 361, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 447, 473, 474, 475, 476, 477, 478, 479
C	524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 549, 550, 551, 552, 553, 554, 555

Supplementary Table 5 Maltoheptaose binding by HADDOCK* for step i.

	H-Score^a	RMSD^b	VDW^c	Elec^d	Desol^e	Z-Score^f	Cluster^g
Pose i	16.7±13.3	0.7±0.4	-66.4±5.4	-127.3±29.9	1.0±1.1	-1.4	435
Pose ii	25.7±2.0	2.3±0.0	-54.9±4.3	-124.2±20.9	2.3±2.3	-0.8	268
Pose iii	30.1±4.0	1.9±0.1	-71.6±5.5	-85.5±26.2	-2.5±3.8	-0.5	117
Pose iv	36.5±4.5	2.2±0.1	-53.5±7.2	-144.5±38.8	6.8±3.4	-0.1	78

* The values are the optimal structure in top four clusters. Data are presented as mean value and standard error (±), inferred by HADDOCK software.

^a HADDOCK score.

^b RMSD (Å) from the overall lowest-energy structure.

^c Van der Waals energy (kcal/mol).

^d Electrostatic energy (kcal/mol).

^e Desolvation energy (kcal/mol).

^f Z-score represents how many standard deviations the HADDOCK score of a given cluster is separated from the mean of all clusters..

^g Number of poses in the Cluster.

Supplementary Table 6 Maltodecaose binding by HADDOCK* for step i.

	H-Score^a	RMSD^b	VDW^c	Elec^d	Desol^e	Z-Score^f	Cluster^g
Pose i	57.5±7.2	1.8±0.0	-66.1±4.7	-64.5±20.4	-2.2±2.7	-1.1	260
Pose ii	46.5±3.0	0.2±0.1	-61.1±2.6	-90.2±8.8	-3.5±1.2	-2.1	120
Pose iii	66.5±11.6	1.9±0.1	-61.5±4.3	-83.5±17.9	2.5±4.4	-0.3	85
Pose iv	61.3±3.6	1.7±0.0	-57.7±11.5	-126.9±6.3	3.0±1.9	-0.8	70

* The values are the optimal structure in top four clusters. Data are presented as mean value and standard error (\pm), inferred by HADDOCK software.

^a HADDOCK score.

^b RMSD (Å) from the overall lowest-energy structure.

^c Van der Waals energy (kcal/mol).

^d Electrostatic energy (kcal/mol).

^e Desolvation energy (kcal/mol).

^f Z-score represents how many standard deviations the HADDOCK score of a given cluster is separated from the mean of all clusters..

^g Number of poses in the Cluster.

Supplementary Table 7 Selected 85 residues in *RpBE* to probe the optimal substrate binding modes for setp iv.

Domain	Residue number
N	None
A	305,311,312,313,314,315,316,317,318,319,320,321,322,323,324,325,326,327,329,330,356,358,359,360,361,382,385,386,387,388,391,392,393,394,395,396,397,398,399,400,401,402,403,404,405,406,407,408,409,410,411,412,413,414,424,426,427,428,429,430,431,432,433,434,435,436,437,438,439,440,441,442,472,473,474,475,476,477,482,483,484,485,486,488,489
C	None

Supplementary Table 8 Polysaccharide binding by HADDOCK* for setp iv.

	H-Score^a	RMSD^b	VDW^c	Elec^d	Desol^e	Z-Score^f	Cluster^g
Pose I	-33.7±5.1	2.2±0.1	-58.6±2.0	-119.8±27.3	5.7±2.1	-1.0	479
Pose II	-37.3±5.6	0.7±0.5	-55.6±5.0	-128.9±21.9	1.6±1.3	-1.3	110
Pose III	-28.1±4.5	2.2±0.1	-61.6±5.3	-136.9±22.2	8.2±3.0	-0.6	74
Pose IV	-21.8±3.8	1.6±0.1	-66.4±2.0	-80.5±14.2	2.8±2.1	-0.1	44

* The values are the optimal structure in top four clusters. Data are presented as mean value and standard error (±), inferred by HADDOCK software.

^a HADDOCK score.

^b RMSD (Å) from the overall lowest-energy structure.

^c Van der Waals energy (kcal/mol).

^d Electrostatic energy (kcal/mol).

^e Desolvation energy (kcal/mol).

^f Z-score represents how many standard deviations the HADDOCK score of a given cluster is separated from the mean of all clusters.

^g Number of structures in the Cluster.

Supplementary Table 9 Mutation strategies provided by Rosetta software*.

AA ^a	Site	To AA ^b	AA	Site	To AA
L	187	M/C	L	421	F
Q	201	D/H	H	428	Y/E/D
L	234	M/I	G	429	M/T/N
V	237	I	G	431	N
S	239	G/M/I	K	436	R/Q/M
A	243	T/M/D/N	L	463	M/C
Y	377	M	I	334	K
N	380	D/W/M	D	122	K
M	381	L	A	115	R
H	385	N/E/C	V	203	R
L	388	M/W	V	344	R
P	420	S/V/L	M	190	H

* Rosetta Cartersian_ddg were employed to screen the mutations following the software protocols.

^a AA denotes the initial amino acid.

^b To AA denotes the mutation strategy. For instance, M/C denotes the L187M and L187C variants.

Supplementary Table 10 Experimental results for the variants designed by Rosetta.

Site	RA ^a	Site	RA ^a	Site	RA
A243T	1.66±0.03	H385E	1.16±0.01	K436M	0.46±0.06
K436R	1.50±0.02	P420V	1.11±0.01	H428D	0.43±0.02
G431N	1.45±0.01	L463C	1.10±0.01	K436Q	0.38±0.03
Y377M	1.43±0.01	L187C	1.02±0.03	H428E	0.37±0.02
N380M	1.39±0.00	H428Y	1.01±0.02	M381L	0.29±0.04
G429M	1.38±0.01	A243D	0.93±0.07	H385N	0.28±0.02
N380D	1.37±0.01	A243N	0.79±0.09	L463M	0.27±0.02
V344R	1.35±0.01	I354Y	0.78±0.11	Q201H	0.14±0.10
L187M	1.33±0.04	P420L	0.75±0.01	L234M	0.10±0.03
P420S	1.32±0.01	H385C	0.73±0.13	D122K	0.00±0.00
I334K	1.28±0.02	A115R	0.72±0.03	V203R	0.00±0.00
G429T	1.25±0.07	L234I	0.68±0.11	V237I	0.00±0.00
S239G	1.24±0.01	N380W	0.59±0.00	S239I	0.00±0.00
A243M	1.23±0.06	L187S	0.58±0.13	S239M	0.00±0.00
L421F	1.23±0.01	M190H	0.50±0.06	L388M	0.00±0.00
G429N	1.20±0.02	Q201D	0.49±0.02		

^a RA denotes the relative activity (fold) to that of the WT *RpBE*, which is regarded to possess 100% enzyme activity. Data are presented as mean value and standard error (±) inferred from three independent runs. Source data are provided as a Source Data file.

Supplementary Table 11 Enzyme kinetic experiments^a for the WT and variants of A243T and K436R.

	WT	A243T	K436R
V_{max} (min ⁻¹)	2199±72	2358±66	2547±69
K_m (mg/mL)	0.43±0.06	0.24±0.04	0.23±0.04

^a Data are presented as mean value and standard error (±) inferred from three independent runs. Source data are provided as a Source Data file.

Supplementary Table 12 Primers utilized for the creation of *RpBE* variants.

Gene	Primer 5'-3'sequence
A115R	F: TATTGAAGGCCTTCGTTCAATTATTACACGCCTTGATTA R: GTGTAATAATTGAACGAAGGCCTTCAATAGGTGAGCCTG
D122K	F: TATTACACGCCTTAAATATACATGGCATGATGATGCTTG R: CATGCCATGTATATTTAAGGCGTGTAAATAATTGAAGCAA
L187S	R: ACATGTTGAACTTAGCCCTATTATGGAACATCCTTATTA F: GTTCCATAATAGGGCTAAGTTCAACATGTGTAAAGCCAA
L187M	F: ACATGTTGAACTTATGCCTATTATGGAACATCCTTATTA R: GTTCCATAATAGGCATAAGTTCAACATGTGTAAAGCCAA
L187C	F: ACATGTTGAACTTTGCCCTATTATGGAACATCCTTATTA R: GTTCCATAATAGGGCAAAGTTCAACATGTGTAAAGCCAA
M190H	F: ACTTCTTCCTATTCATGAACATCCTTATTATGGCTCATG R: AATAAGGATGTTTCATGAATAGGAAGAAGTTCAACATGTG
Q201D	F: CTCATGGGGGTACGAAGTTGTTGGCTATTATGCTCCTAC R: AATAGCCAACAACCTTCGTACCCCCATGAGCCATAATAAG
Q201H	F: CTCATGGGGGTACCATGTTGTTGGCTATTATGCTCCTAC R: AATAGCCAACAACATGGTACCCCCATGAGCCATAATAAG
V203R	F: GGGGTACCAAGTTTCGTGGCTATTATGCTCCTACATTTTCG R: GAGCATAATAGCCACGAACCTTGGTACCCCCATGAGCCAT
L234M	F: CATTGGCGTTATTATGGATTGGGTTCCTTCACATTTTCGC R: AAGGAACCCAATCCATAATAACGCCAATGCCGCGCTGAT
L234I	F: CATTGGCGTTATTATTGATTGGGTTCCTTCACATTTTCGC R: AAGGAACCCAATCAATAATAACGCCAATGCCGCGCTGAT
V237I	F: TATTCTTGATTGGATTCCTTCACATTTCCGCCGCTGATCC R: CGAAATGTGAAGGAATCCAATCAAGAATAACGCCAATGC
S239G	F: TGATTGGGTTTCCTGGCCATTTCCGCCGCTGATCCTCAGGG R: CAGCGGCGAAATGGCCAGGAACCCAATCAAGAATAACGC
S239M	F: TGATTGGGTTTCCTATGCATTTCCGCCGCTGATCCTCAGGG R: CAGCGGCGAAATGCATAGGAACCCAATCAAGAATAACGC
S239I	F: TGATTGGGTTTCCTATTCATTTCCGCCGCTGATCCTCAGGG R: CAGCGGCGAAATGAATAGGAACCCAATCAAGAATAACGC
A243D	F: CTTACATTTCCGCCGATGATCCTCAGGGCCTTGTTTATT R: GGCCCTGAGGATCATCGGCGAAATGTGAAGGAACCCAAT
A243M	F: CTTACATTTCCGCCATGGATCCTCAGGGCCTTGTTTATT R: GGCCCTGAGGATCCATGGCGAAATGTGAAGGAACCCAAT
A243N	F: CTTACATTTCCGCCAATGATCCTCAGGGCCTTGTTTATT R: GGCCCTGAGGATCATTGGCGAAATGTGAAGGAACCCAAT
A243T	F: CTTACATTTCCGCCACCGATCCTCAGGGCCTTGTTTATT R: GGCCCTGAGGATCGGTGGCGAAATGTGAAGGAACCCAAT
I334K	F: GAATTTAGAGGCGAAAGACTTTATTAAGAAATTTAACGA R: GAATTTAGAGGCGAAAGACTTTATTAAGAAATTTAACGA
V344R	F: ATTTAACGAGACTCGTTATCTTCATTTCCCAGAGGCTAT R: GGAAATGAAGATAACGAGTCTCGTTAAATTTCTTAATAA
I354Y	F: AGAGGCTATTACATATGCTGAAGAATCAACAGCTTGGCC R: TTGATTCTTCAGCATATGTAATAGCCTCTGGGAAATGAA
Y377M	F: CCTTGGCTTTCTTATGAAATGGAACATGGGCTGGATGCA

	R: CCATGTTCCATTTTCATAAGAAAGCCAAGGCCGTTGTTGT
N380D	F: TCTTTATAAATGGGATATGGGCTGGATGCACGACACCTT
	R: GCATCCAGCCCATATCCCATTATATAAAGAAAGCCAAGGC
N380W	F: TCTTTATAAATGGTGGATGGGCTGGATGCACGACACCTT
	R: GCATCCAGCCCATCCACCATTATATAAAGAAAGCCAAGGC
N380M	F: TCTTTATAAATGGATGATGGGCTGGATGCACGACACCTT
	R: GCATCCAGCCCATCATCCATTATATAAAGAAAGCCAAGGC
M381L	F: TTATAAATGGAACCTGGGCTGGATGCACGACACCTTGGA
	R: CGTGCATCCAGCCCAGGTTCCATTATATAAAGAAAGCCAA
H385N	F: CATGGGCTGGATGAATGACACCTTGGACTACATGCGCCG
	R: AGTCCAAGGTGTCATTCATCCAGCCCATGTTCCATTAT
H385E	F: CATGGGCTGGATGGAAGACACCTTGGACTACATGCGCCG
	R: AGTCCAAGGTGTCCTTCATCCAGCCCATGTTCCATTAT
H385C	F: CATGGGCTGGATGTGCGACACCTTGGACTACATGCGCCG
	R: AGTCCAAGGTGTGCGACATCCAGCCCATGTTCCATTAT
L388M	F: GATGCACGACACCATGGACTACATGCGCCGCGATCCTGT
	R: GGCGCATGTAGTCCATGGTGTGTCGTCATCCAGCCCATGT
P420S	F: GCATTATATTCTTAGCCTTTCACATGATGAAGTTGTTCA
	R: CATCATGTGAAAGGCTAAGAATATAATGCTCACTGAAAG
P420V	F: GCATTATATTCTTGTGCTTTCACATGATGAAGTTGTTCA
	R: CATCATGTGAAAGCACAAGAATATAATGCTCACTGAAAG
P420L	F: GCATTATATTCTTCTGCTTTCACATGATGAAGTTGTTCA
	R: CATCATGTGAAAGCAGAAGAATATAATGCTCACTGAAAG
L421F	F: TTATATTCTTCCTTTTTTCACATGATGAAGTTGTTTCATGG
	R: CTTTCATCATGTGAAAAAGGAAGAATATAATGCTCACTGA
H428Y	F: TGATGAAGTTGTTTATGGCAAAGGCTCACTTTGGACAAA
	R: GTGAGCCTTTGCCATAAACAACCTTCATCATGTGAAAGAG
H428E	F: TGATGAAGTTGTTGAAGGCAAAGGCTCACTTTGGACAAA
	R: GTGAGCCTTTGCCTTCAACAACCTTCATCATGTGAAAGAG
H428D	F: TGATGAAGTTGTTGATGGCAAAGGCTCACTTTGGACAAA
	R: GTGAGCCTTTGCCATCAACAACCTTCATCATGTGAAAGAG
G429M	F: TGAAGTTGTTTCATATGAAAGGCTCACTTTGGACAAAGAT
	R: AAAGTGAGCCTTTCATATGAACAACCTTCATCATGTGAAA
G429T	F: TGAAGTTGTTTCATACCAAAGGCTCACTTTGGACAAAGAT
	R: AAAGTGAGCCTTTGGTATGAACAACCTTCATCATGTGAAA
G429N	F: TGAAGTTGTTTCATAATAAAGGCTCACTTTGGACAAAGAT
	R: AAAGTGAGCCTTTATTATGAACAACCTTCATCATGTGAAA
G431N	F: TGTTTCATGGCAAAAATTCACCTTTGGACAAAGATGCCAGG
	R: TTGTCCAAAGTGAATTTTTGCCATGAACAACCTTCATCAT
K436R	F: CTCACCTTTGGACACGATGCCAGGCGATGATTGGCAGAA
	R: CATCGCCTGGCATACTGTGCCAAAGTGAGCCTTTGCCAT
K436Q	F: CTCACCTTTGGACACAGATGCCAGGCGATGATTGGCAGAA
	R: CATCGCCTGGCATCTGTGTGCCAAAGTGAGCCTTTGCCAT
K436M	F: CTCACCTTTGGACAATGATGCCAGGCGATGATTGGCAGAA
	R: CATCGCCTGGCATCATTGTGCCAAAGTGAGCCTTTGCCAT
L463M	F: CCCTGGCAAGAAGATGTTATTCATGGGTGGCGAATTTGG
	R: CACCCATGAATAACATCTTCTTGCCAGGGTGCCCCCACA
L463C	F: CCCTGGCAAGAAGTGCTTATTCATGGGTGGCGAATTTGG
	R: CCCTGGCAAGAAGTGCTTATTCATGGGTGGCGAATTTGG

D269A	F: GCGCCATCATCCTGCATGGGGTACATACGTGTTTCGACTACAACAAACCT R: CGTATGTACCCCATGCAGGATGATGGCGCATGCGAGGATCATCGTACTC
D305A	F: TGGCCTTCGCGTTGCGGCAGTAGCATCTATGCTTTATCG R: TAGATGCTACTGCCGCAACGCGAAGGCCATCAACATGAT
P268A	F: CATGCGCCATCATGCGGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCCGCATGATGGCGCATGCGAGGATCAT
P268R	F: CATGCGCCATCATCGTGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCACGATGATGGCGCATGCGAGGATCAT
P268N	F: CATGCGCCATCATAATGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCATTATGATGGCGCATGCGAGGATCAT
P268D	F: CATGCGCCATCATGATGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCATCATGATGGCGCATGCGAGGATCAT
P268C	F: CATGCGCCATCAT TGTGATTGGGGTACATACGTGTTTCGACTACAACAAA R: ATGTACCCCAATC ACAATGATGGCGCATGCGAGGATCATCGTACTCGAA
P268E	F: CATGCGCCATCATGAAGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCTTCATGATGGCGCATGCGAGGATCAT
P268Q	F: CATGCGCCATCATCAGGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCCTGATGATGGCGCATGCGAGGATCAT
P268G	F: CATGCGCCATCATGGCGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCGCCATGATGGCGCATGCGAGGATCAT
P268H	F: CATGCGCCATCATCATGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCATGATGATGGCGCATGCGAGGATCAT
P268I	F: CATGCGCCATCATATTGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCAATATGATGGCGCATGCGAGGATCAT
P268L	F: CATGCGCCATCATCTGGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCCAGATGATGGCGCATGCGAGGATCAT
P268K	F: CATGCGCCATCATAAAGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCTTTATGATGGCGCATGCGAGGATCAT
P268M	F: CATGCGCCATCATATGGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCCATATGATGGCGCATGCGAGGATCAT
P268F	F: CATGCGCCATCATTTTGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCAAAATGATGGCGCATGCGAGGATCAT
P268S	F: CATGCGCCATCATAGCGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCGCTATGATGGCGCATGCGAGGATCAT
P268T	F: CATGCGCCATCATACCGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCGGTATGATGGCGCATGCGAGGATCAT
P268W	F: CATGCGCCATCATTGGGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCCCAATGATGGCGCATGCGAGGATCAT
P268Y	F: CATGCGCCATCATTATGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCATAATGATGGCGCATGCGAGGATCAT
P268V	F: CATGCGCCATCATGTGGATTGGGGTACATACGTGTTTCGA R: ATGTACCCCAATCCACATGATGGCGCATGCGAGGATCAT
E425A	F: TCTTTCACATGATGCAGTTGTTTCATGGCAAAGGCTCACT R: TGCCATGAACAACGCATCATGTGAAAGAGGAAGAATAT
E425C	F: TCTTTCACATGATTGTGTTGTTTCATGGCAAAGGCTCACT R: TGCCATGAACAACACAATCATGTGAAAGAGGAAGAATAT
E425D	F: TCTTTCACATGATGATGTTGTTTCATGGCAAAGGCTCACT R: TGCCATGAACAACATCATCATGTGAAAGAGGAAGAATAT
E425F	F: TCTTTCACATGATTTTGTTGTTTCATGGCAAAGGCTCACT

	R: TGCCATGAACAACAAAATCATGTGAAAGAGGAAGAATAT
E425G	F: TCTTTCACATGATGGTGTTGTTTCATGGCAAAGGCTCACT
	R: TGCCATGAACAACACCATCATGTGAAAGAGGAAGAATAT
E425H	F: TCTTTCACATGATCATGTTGTTTCATGGCAAAGGCTCACT
	R: TGCCATGAACAACATGATCATGTGAAAGAGGAAGAATAT
E425I	F: TCTTTCACATGATATTGTTGTTTCATGGCAAAGGCTCACT
	R: TGCCATGAACAACAATATCATGTGAAAGAGGAAGAATAT
E425K	F: TCTTTCACATGATAAAGTTGTTTCATGGCAAAGGCTCACT
	R: TGCCATGAACAACCTTATCATGTGAAAGAGGAAGAATAT
E425L	F: TCTTTCACATGATCTGGTTGTTTCATGGCAAAGGCTCACT
	R: TGCCATGAACAACCAGATCATGTGAAAGAGGAAGAATAT
E425M	F: TCTTTCACATGATATGGTTGTTTCATGGCAAAGGCTCACT
	R: TGCCATGAACAACCATATCATGTGAAAGAGGAAGAATAT
E425N	F: TCTTTCACATGATAATGTTGTTTCATGGCAAAGGCTCACT
	R: TGCCATGAACAACATTATCATGTGAAAGAGGAAGAATAT
E425P	F: TCTTTCACATGATCCGGTTGTTTCATGGCAAAGGCTCACT
	R: TGCCATGAACAACCGGATCATGTGAAAGAGGAAGAATAT
E425Q	F: TCTTTCACATGATCAGGTTGTTTCATGGCAAAGGCTCACT
	R: TGCCATGAACAACCTGATCATGTGAAAGAGGAAGAATAT
E425R	F: TCTTTCACATGATCGTGTTGTTTCATGGCAAAGGCTCACT
	R: TGCCATGAACAACACGATCATGTGAAAGAGGAAGAATAT
E425S	F: TCTTTCACATGATAGCGTTGTTTCATGGCAAAGGCTCACT
	R: TGCCATGAACAACGCTATCATGTGAAAGAGGAAGAATAT
E425T	F: TCTTTCACATGATACCGTTGTTTCATGGCAAAGGCTCACT
	R: TGCCATGAACAACGGTATCATGTGAAAGAGGAAGAATAT
E425V	F: TCTTTCACATGATGTTGTTGTTTCATGGCAAAGGCTCACT
	R: TGCCATGAACAACAACATCATGTGAAAGAGGAAGAATAT
E425W	F: TCTTTCACATGATTGGGTTGTTTCATGGCAAAGGCTCACT
	R: TGCCATGAACAACCAATCATGTGAAAGAGGAAGAATAT
E425Y	F: TCTTTCACATGATTATGTTGTTTCATGGCAAAGGCTCACT
	R: TGCCATGAACAACATAATCATGTGAAAGAGGAAGAATAT

Supplementary Table 13 Details of the molecular assemblies for MD simulation in this study.*

No.	Molecular assembly	Number of atom	Size of the simulation box (Å ³) ^a	Process	Simulation time (ns)
1	<i>Rp</i> BE pose i ^b	82,592	124×90×78	Glycosylation	1,000
2	<i>Rp</i> BE pose I ^c	80,948	123×89×78	Transglycosylation	100
3	<i>Rp</i> BE pose II ^d	84,116	124×89×81	Transglycosylation	100
4	<i>Cce</i> BE1 ^e	130,924	123×118×96	/	500
5	<i>Ro</i> BE ^f	114,143	125×98×98	/	500

* Assembly 1: *Rp*BE for glycosylation; assemblies 2-3: *Rp*BE for transglycosylation; assemblies 4-5: *Cce*BE1 and *Ro*BE for investigating the flexibility of the loops 1-3.

^a The size of the solvent boxes guarantees a minimum distance of 15 Å from any atom of the enzyme to any edge of the simulation box.

^b *Rp*BE-substrate complex is obtained by HADDOCK, which includes a polysaccharide chain (pose i), used to simulate the spontaneous process of the polysaccharide chain transfer.

^c *Rp*BE-substrate complex is obtained by HADDOCK, which includes a fixed polysaccharide chain lying in the active-site cleft and a new polysaccharide chain (pose I), used to simulate the spontaneous process of the new polysaccharide transfer.

^d *Rp*BE-substrate complex is obtained by HADDOCK, which includes a fixed polysaccharide chain lying in the active-site cleft and a new polysaccharide chain (pose II), used to simulate the spontaneous process of the new polysaccharide transfer.

^e *Cce*BE1 (PDB code 5GQU).

^f *Ro*BE (PDB code 6JOY).

Supplementary Table 14 Details of the molecular assemblies for free-energy calculations in this study.*

No.	Molecular assembly	Number of atom	Size of the simulation box (Å ³) ^a	Process	Simulation time (ns)
1	<i>Rp</i> BE pose i-2D ^b	78,188	124×88×77	Glycosylation	400
2	<i>Rp</i> BE pose i P268A-2D ^c	78,166	124×88×77	Glycosylation	400
3	<i>Rp</i> BE substrate free-1D ^d	77,916	123×86×77	Glycosylation	400
4	<i>Rp</i> BE P268A substrate free-1D ^c	77,875	123×86×77	Glycosylation	400
5	<i>Rp</i> BE pose I-1D ^f	81,994	125×89×79	Transglycosylation	400

* Assemblies 1, 2, and 5 for investigation of the polysaccharide approaching the catalytic position from different starting models; assemblies 3 and 4 for investigation of D269 approaching D305. Assemblies 1-4 for glycosylation; assembly 5 for transglycosylation. 1D and 2D denote one- and two-dimensional calculations.

^a The size of the solvent boxes guarantees a minimum distance of 15 Å from any atom of the enzyme to any edge of the simulation box.

^b The structure is generated from the trajectory of assembly 1 in Table 12, used to simulate the process of the polysaccharide chain transferring to the catalytic position.

^c The structure is constructed based on the assembly 1 in Table 13, used to simulate the process of the polysaccharide chain transferring to the catalytic position after alanine replacement.

^{d,e} Constructed *Rp*BE and P268A models with substrate free, used to simulate the process of D269 approaching D305.

^f The structure is generated from the trajectory of assembly 2 in Table 12, used to simulate the process of a new polysaccharide transferring to the catalytic position.

Supplementary References

1. Engh, R. A. & Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A* **47**, 392-400 (1991).
2. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861-877 (2019).