

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



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This paper describes a combination of biochemical and computational studies of *Rhodothermus profundus* (RhBE) branching enzyme. The authors have chosen this enzyme due to its superior activity relative to the commercial enzyme (from *Rhodothermus Obamamensis* RoBE) which is over 93% identical to RhBE. They use the alphafold homology model as the structure for their docking studies, which shows high structural similarity with other BE's whose structures are known, though there are some curious differences, especially with regard to the structure of RoBE (PDBID 6JOY). There are in fact three loops, Loop 105-114, Loop 257-274 and Loop 314-328 that differ substantially between the two structures, even though the sequences are almost identical for each, which is curious. The RhBE alphafold structure closely follows the Cyanothecae structure in the later two loops, though it differs in loop 105-114. The reason these differences are important is that the first two loops are both involved in donor strand binding in the cyanothecae structure, and presumably in RhBE as well. In fact a residue affirmed to be critical in the present study, D269, is part of the second loop and is in a completely different position in RoBE. However, it is important to note that the B factors for all three of these loops is quite large in RoBE. The essential question here is, is there a large conformational change that occurs upon donor strand binding in RoBE, and could the same conformational change happen in RhBE? Or, alternatively, do the authors maintain that the RoBE structure is simply wrong relative to this loop? Given the incredibly high sequence identity of these regions, it is highly unlikely that they are just mysteriously different. The authors really need to address these issues.

The authors then conduct a massive docking effort using maltoheptaose (M7) for these studies. The goal is to identify glucan binding sites on the surface of the enzyme, and hopefully identify the binding that leads to productive reaction. The conclusion in these studies is that there is a "prebinding" step ("pose i" in Figure 4) with the glucan then moving into "pose iv" which is a conformation more or less competent for reaction. My issue with these studies is the use of a short oligosaccharide like M7. To my knowledge, no BE will react with a glucan of this length as either an acceptor or donor chain, indicating that cooperative interactions over a longer stretch of sugar chain is necessary for productive binding, perhaps involving binding both upstream and downstream of the glycosylation site for example. If M7 is too short to be a bonafide substrate, how can it accurately capture initial binding steps if interactions are required beyond its length, which seems likely? Have authors investigated the minimum length of glucan necessary for reaction for RhBE? Docking studies with a longer glucan seem essential. In figure 5, the authors suggest that there is a passing of the glucan from D269 to the catalytic residues in the active site using free energy calculations. The recipient of this passing is D305, which is one of the three absolutely conserved residues in GH13 enzymes involved in catalysis. It is the residue responsible for covalent attack of the glucan in all GH13 enzymes. Here, the authors are suggesting quite a different role from that proposed for this family of enzymes. Though the authors show mutation results in loss of activity, loss of activity upon mutation of this residue has been seen in many other GH13 enzymes and BE's as well. Thus mutation of this residue does not speak to the binding mechanism proposed here. Strangely, the authors present quite a different reaction mechanism for RhBE compared to all other GH13 enzymes (Figure 3c). Here, bond breakage results in oxocarbenium cation formation, and subsequently nucleophilic attack by D424, which is not the nucleophile in other GH13 enzymes and is considered to be involved in proton transfer events, not in covalent linkage to the glucan. The

authors will need to provide a lot of new information to back up this quite controversial theory of how this reaction goes. This mechanism was firmly established by the work of the Dijkstra lab in 1999 in a study of cyclodextrin transferase ((1999) Nat Struct Biol 6: 432-436).

The authors do not compare their glucan bound productively in the active site with the M7-bound cyanothese structure either. The structures, as described above, are quite similar, and many of the residues involved in binding are identical or strongly conserved in the two enzymes. It is essential to compare the docking result to the experimentally determined one. Which residues make similar contacts? Where are there differences etc.?

Having identified D269 as a key residue in the “prebinding” interaction, the authors mutate this residue to ala and indeed over 80% of the activity is lost, confirming its importance in the reaction. Though an interesting result, it does not prove that D269 is working the way the authors suggest. It could be affecting either donor or acceptor chain binding, serving as an important residue for conformational change of the loop etc.

I’m not sure what the mutagenesis of P268 proves in support of this mechanism. It is a proline, which is often important to maintain structure, the structures of the cyano and RhBE enzymes are quite similar here.

The authors then use their docking results to propose another mechanism involving prebinding for the acceptor chains. My concerns on this are similar to that above. Can this enzyme transfer a glucan to an acceptor chain as short as M7? If not, then there are probably essential interactions missing. The strongest part of the paper is depicted in Figure 8, where Rosetta was used to predict mutations that would increase binding and reactivity. However, Figure 8 is very difficult to understand. The authors show a glucan spanning A243 and K436, but in the alphafold model these residues are almost 30Å away from each other. How does an M7 span this distance? Now looking at Figure 4a and b versus c and d, I am having an impossible time understanding the relationship. In fact Figure 4a is essentially incomprehensible. In short I can’t tell where site A and site B are in the structure, nor do I understand the relationship between them with respect to the poses. This must be made clear. It is clear that improvement in activity has happened using Rosetta. However, where these positions are relative to the “poses” is not clear at all. For several of these positions, the cyano BE M7 is interacting. The authors should compare what they have with what the experiment shows.

Reviewer #2 (Remarks to the Author):

In the manuscript by Zhiyou Zong et al., the discovery of a new type of branching enzyme (BE) with superior properties for industrial production is presented, supported by modeling, docking, and simulation studies to explore the enzyme's mechanism. However, several areas require clarification and improvement to solidify the findings. Therefore, I recommend rejecting the manuscript in its current form but encourage resubmission following the solution of the outlined concerns:

1. How were the structures shown in Fig. 2A obtained? From homology modeling or AlphaFold2 database? It needs to be clarified in the main text.
2. Reference is necessary to prove the responsibility of using HADDOCK to dock oligosaccharides to RpBE since HADDOCK is always used in protein-protein docking rather than glycan docking. Maybe a specific docking software designed for glycans can be applied to compare with the current docking pose.

3. In the Methods part, MD simulations were only run for 1 ns. Why did the authors claim microsecond-level simulations in the main text?

4. In fact, the Methods part needs a lot of supplements. For MD simulations, how the temperature and pressure are controlled? Is there any constraint on hydrogens? How is the interaction energy calculated (that's a key point!)? Also, how Rosetta was used for generating mutation strategies needs to be explained. Suggest the authors take care of the reproducibility of their research.

5. To capture the catalysis process, metadynamics is not enough. I suggest the author provide QM/MM simulations to show the reaction process and prove the residue importance claimed by them. Cheap QM levels, especially semi-empirical ones, will not cost too much but provide a more direct view of such a process. But if permits, I suggest a standard DFT QM in QM/MM calculation.

Reviewer #3 (Remarks to the Author):

In this work the enzymatic activity of Branching Enzyme (BE) from *Rhodothermus profundus* is compared with other twelve BEs. The authors exploited the structural model obtained with AlphaFold program to run docking calculations and MD simulations in order to characterize the substrate binding and transfers steps (SBTs).

The molecular dynamic simulations are well performed and provide possible structural models of the SBT steps of the reaction, they can be used in combination with experimental methods. The work indeed misses experimental evidences as site directed mutagenesis experiments are not sufficient to explain the mechanism. Information about the mechanism of a reaction are generally provided by chemical kinetics studies, NMR spectroscopy of the substrate could be also used to observe the reaction intermediates. ITC measurements can provide valuable information about the thermodynamic parameters of binding.

Furthermore, the authors ought to demonstrate the proper folding of the mutated BEs. This entails confirming that any observed reduction in activity stems from the mutation of the key residue rather than a loss of the protein's three-dimensional structure.

Finally, It is not clear how the authors selected the 138 residues used for HADDOCK ? Which are the active residues used as input to run docking calculations? Which data are used to select them ?

Minor issues:

The authors describe that the 1D NMR experiment was recorded through a Varian Inova 600 spectrometer at 413 K. I think the authors misspelled the acquisition temperature.

Figure 4 is very difficult to interpret. The orientation of the 3D structure should be equal.

In conclusion I think that this is a preliminary work where the computational aspects need to be validated by experimental studies. The authors should investigate deeper the mechanism of the BE activity before considering the study acceptable for publication in Nature Communications

Response to the comments of the reviewers

We are grateful to the reviewers for their constructive criticisms of our manuscript. Below, we reply point by point to their comments, and have highlighted for editorial purposes all the changes in one of the copies of the revised manuscript.

Response to Reviewer #1:

1) This paper describes a combination of biochemical and computational studies of *Rhodothermus profundus* (RhBE) branching enzyme. The authors have chosen this enzyme due to its superior activity relative to the commercial enzyme (from *Rhodothermus Obamamensis* RoBE) which is over 93% identical to RhBE. They use the alphafold homology model as the structure for their docking studies, which shows high structural similarity with other BE's whose structures are known, though there are some curious differences, especially with regard to the structure of RoBE (PDBID 6JOY). There are in fact three loops, Loop 105-114, Loop 257-274 and Loop 314-328 that differ substantially between the two structures, even though the sequences are almost identical for each, which is curious. The RhBE alphafold structure closely follows the Cyanothecae structure in the later two loops, though it differs in loop 105-114. The reason these differences are important is that the first two loops are both involved in donor strand binding in the cyanothecae structure, and presumably in RhBE as well. In fact a residue affirmed to be critical in the present study, D269, is part of the second loop and is in a completely different position in RoBE. However, it is important to note that the B factors for all three of these loops is quite large in RoBE. The essential question here is, is there a large conformational change that occurs upon donor strand binding in RoBE, and could the same conformational change happen in RhBE? Or, alternatively, do the authors maintain that the RoBE structure is simply wrong relative to this loop? Given the incredibly high sequence identity of these regions, it is highly unlikely that they are just mysteriously different. The authors really need to address these issues.

To maintain consistency with the main text, we continue to abbreviate *Rhodothermus profundus* branching enzyme as *RpBE*.

To do everything we possibly could to address the concerns of the Reviewer, during these three months, we have done our utmost to conduct crystallographic experiments in the hope of providing experimental evidence for the conformation of loop 2. Finally, we obtained the *RpBE* crystal with a 2.9 Å resolution (PDB code 8ZQA). Our findings are in line with the structural prediction of AlphaFold v2.0 (see Fig. r1a); the conformation of loop 2 in *RpBE* crystal is similar to that in AlphaFold modeling *RpBE* and *CceBE1*, but quite different from that in *RoBE* (see Fig. r1b). Therefore, even though the sequence similarity between *RpBE* and *RoBE* is high, the conformation of loop 2 is different between the two structures.

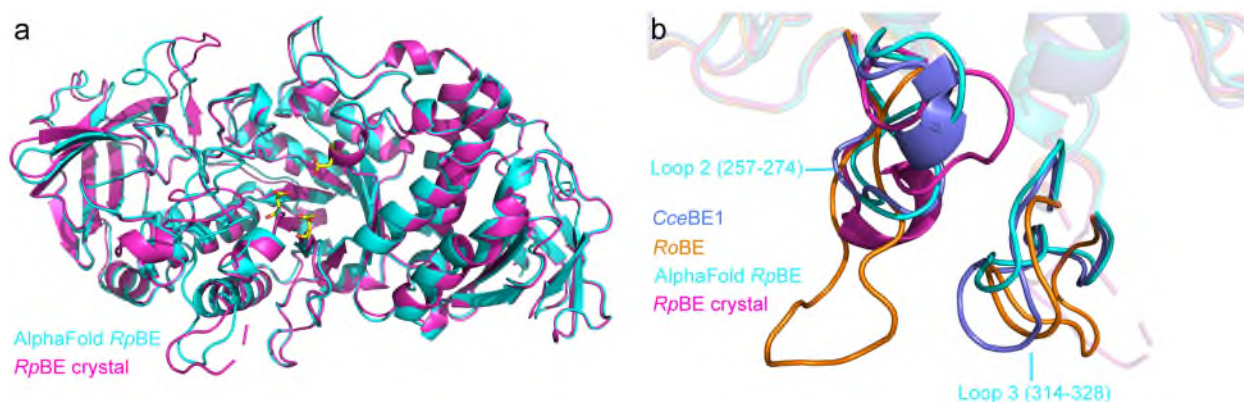
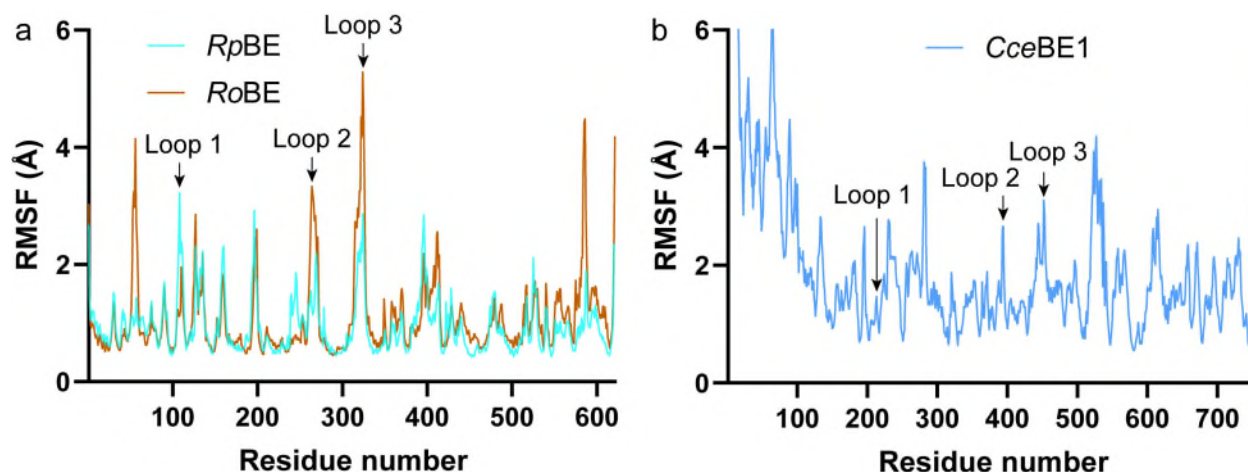


Fig. r1 Structure alignment. **a** Alignment between AlphaFold modeling *RpBE* and *RpBE* crystal (PDB code 8ZQA). **b** Alignment of the loop 2 between *CceBE1*, *RoBE*, AlphaFold modeling *RpBE*, and *RpBE* crystal. The catalytic residues are displayed as a licorice representation, and their carbon atoms are colored in yellow.

In *RpBE* crystal, however, the pronounced flexibility of loop 3 precluded access to its electron density. We, therefore, employed molecular dynamics simulations on AlphaFold modeling *RpBE* to decipher the conformational change of these three loops. As shown in Supplementary Fig. 5, in *RoBE* (PDB code 6JOY), loops 2 and 3, exhibit the highest flexibility in the three BEs, especially loop 3. We, therefore, reason that when the enzyme interacts with the substrate, *RpBE* and *CceBE1* loops 2 and 3 do not undergo significant conformational changes like the corresponding loops in *RoBE*. Furthermore, we found that the conformation and flexibility of loop 1 vary between the three enzymes, suggesting significant structural differences across species.



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

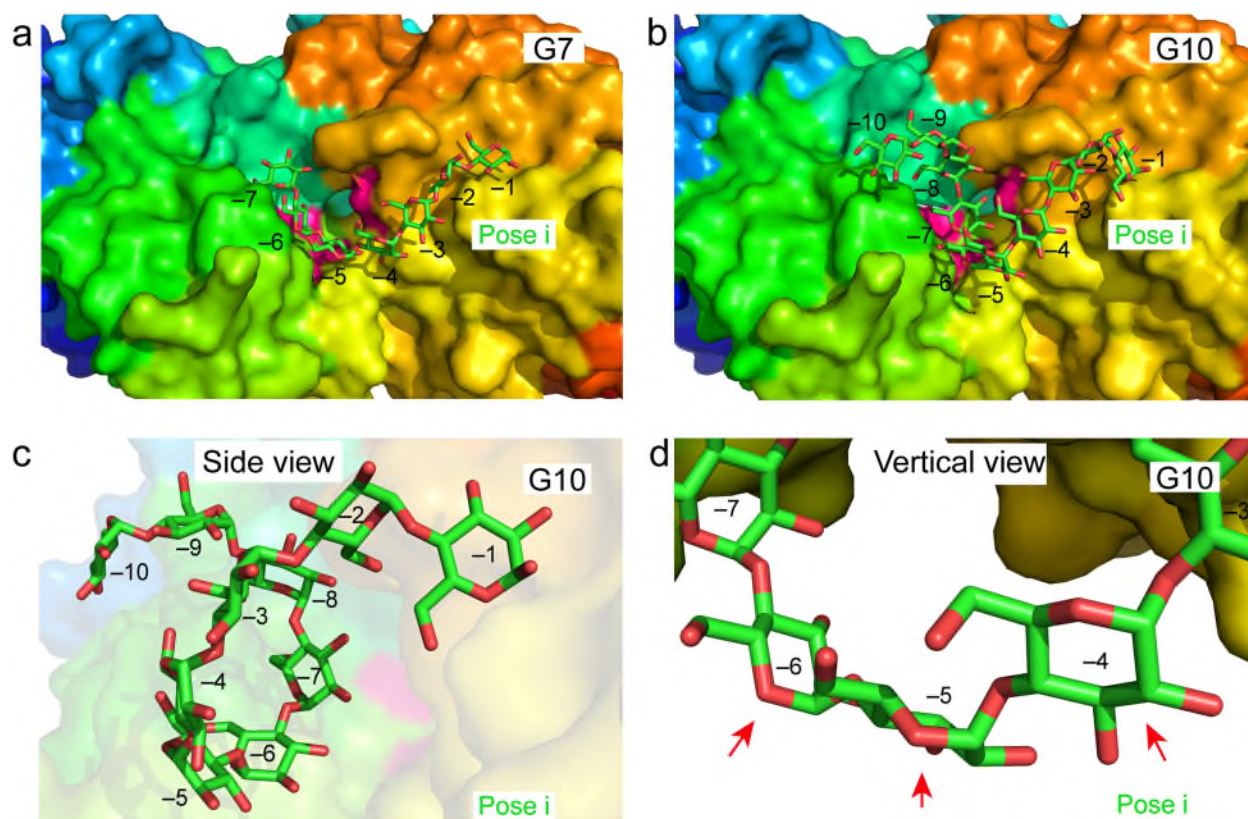
2) The authors then conduct a massive docking effort using maltohepatose (M7) for these studies. The goal is to identify glucan binding sites on the surface of the enzyme, and hopefully identify the binding that leads to productive reaction. The conclusion in these studies is that there is a “prebinding” step (“pose i in Figure 4) with the glucan then moving into “pose iv” which is a conformation more or less competent for reaction. My issue with these studies is the use of a short oligosaccharide like M7. To my knowledge, no BE will react with a glucan of this length as either an acceptor or donor chain, indicating that cooperative interactions over a longer stretch of sugar chain is necessary for productive binding, perhaps involving binding both upstream and downstream of the glycosylation site for example. If M7 is too short to be a bonafide substrate, how can it

accurately capture initial binding steps if interactions are required beyond its length, which seems likely? Have authors investigated the minimum length of glucan necessary for reaction for RhBE? Docking studies with a longer glucan seem essential.

To maintain consistency with the main text, we continue to abbreviate maltoheptaose as G7.

Regarding the substrate selection for docking, maltoheptaose (G7) is the most common linear substrate model obtained in BE-substrate crystal complexes, and the only polysaccharide that binds to the active-site cleft to date. Additionally, as shown in Supplementary Fig. 3c, maltodecaose (G10) is the minimum glucan length necessary for the *RpBE* reaction to be observed. We, therefore, employed both G7 and G10 as substrates to dock with *RpBE*, and compared their binding poses.

As illustrated in Supplementary Fig. 7a and 7b, the optimal conformation of G7 binding to *RpBE* is very similar to that of G10. Supplementary Fig. 7c and 7d, however, indicate that G10 adopts a twisted conformation, wherein three glucosyl rings do not interact with the enzyme. The electrostatic interaction energy in the optimal *RpBE*-G10 complex is much lower than that in the optimal *RpBE*-G7 complex (see Supplementary Table 4 and Table 5). These results suggest that compared to G10, using G7 as a substrate is more conducive to exploring enzyme-substrate association. Furthermore, in a previous work, Ban et al. employed G7 binding to *RoBE* in order to examine potential mutation sites that affect enzyme activity (Ban, X. et al., *Food Res. Int.*, 2022, 162, 112119). These theoretical investigations of G7 binding to BEs, together with the numerous BE-G7 crystallographic complexes (PDB codes 5GQX, 4LPC, 5CLW, to cite a few) obtained experimentally led us to focus on *RpBE*-G7 association in the present study.



Supplementary Fig. 7 Optimal poses of G7 and G10 binding to *RpBE*. **a** Optimal pose of G7 binding to *RpBE*. **b** Optimal pose of G10 binding to *RpBE*, and **c** its side and **d** vertical views. In the optimal *RpBE*-G10 complex, three glucosyl rings, which are denoted by red arrows, do not bind to the enzyme.

3) In figure 5, the authors suggest that there is a passing of the glucan from D269 to the catalytic residues in the active site using free energy calculations. The recipient of this passing is D305, which is one of the three absolutely conserved residues in GH13 enzymes involved in catalysis. It is the residue responsible for covalent attack of the glucan in all GH13 enzymes. Here, the authors are suggesting quite a different role from that proposed for this family of enzymes. Though the authors show mutation results in loss of activity, loss of activity upon mutation of this residue has been seen in many other GH13 enzymes and BE's as well. Thus mutation of this residue does not speak to the binding mechanism proposed here. Strangely, the authors present quite a different reaction mechanism for RhBE compared to all other GH13 enzymes (Figure 3c). Here, bond breakage results in oxocarbenium cation formation, and subsequently nucleophilic attack by D424, which is not the nucleophile in other GH13 enzymes and is considered to be involved in proton transfer events, not in covalent linkage to the glucan. The authors will need to provide a lot of new information to back up this quite controversial theory of how this reaction goes. This mechanism was firmly established by the work of the Dijkstra lab in 1999 in a study of cyclodextrin transferase ((1999) Nat Struct Biol 6: 432-436).

We acknowledge that the description of the covalent catalytic process was erroneous. We are, therefore, greatly appreciative of the Reviewer's comments. The description of the covalent catalytic process embodied in Fig. 3c and 3f, wherein residue D305 is the nucleophile, has been duly corrected.

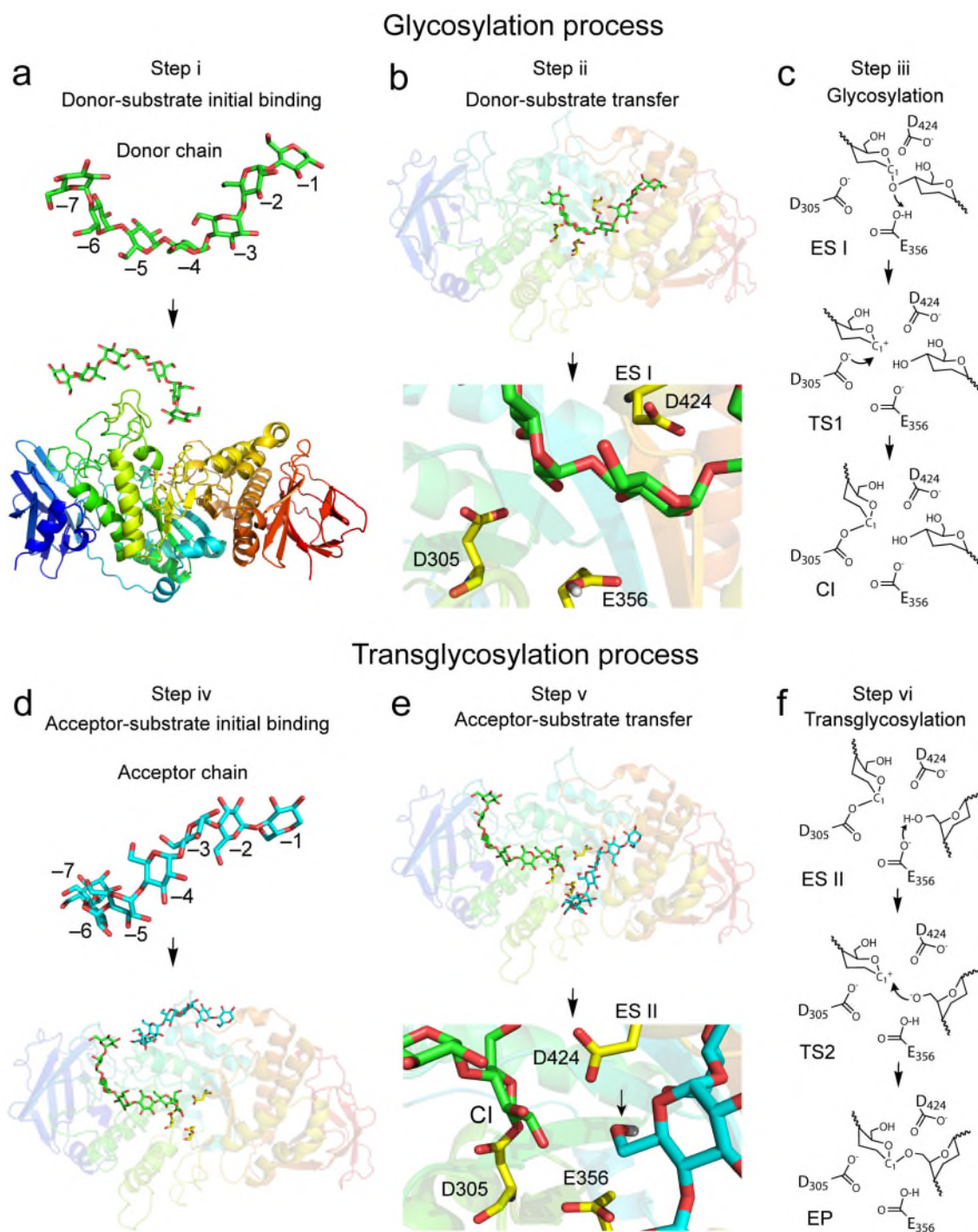


Fig. 3 Schematic diagram of the whole process for generating the branching product. In (a), step i, the substrate is G7. The constructed *RpBE* was employed as the structural model; For clarity, the enzyme is displayed with an opaque representation in a, but with a semi-transparent one in b, d, and e. In (b), step ii, D305, E356, and D424 form the catalytic triad in *RpBE*. ES I denotes the noncovalent enzyme-substrate complex before glycosylation. In (c), step iii, TS1 and CI denote the transition state in the course of glycosylation and the formed enzyme-substrate covalent intermediate (CI) after glycosylation, respectively. In (d), step iv, the acceptor chain (cyan) binds to the enzyme. In (e), step v, the acceptor chain moves to the catalytic position and

prepares for transglycosylation, wherein the donor chain (green) will be linked as a branch to the hydroxyl on C6 of the cyan acceptor chain. ES II denotes the enzyme-substrate complex before transglycosylation. In (f), step vi, TS2 and EP denote the transition state in transglycosylation and the noncovalent enzyme-product complex after catalysis, respectively; the product is the α -1,6-glucosidic branching glucan.

4) The authors do not compare their glucan bound productively in the active site with the M7-bound cyanothece structure either. The structures, as described above, are quite similar, and many of the residues involved in binding are identical or strongly conserved in the two enzymes. It is essential to compare the docking result to the experimentally determined one. Which residues make similar contacts? Where are there differences etc.?

As depicted in Fig. 4a, we align the *RpBE*-G7 complex (pose i) with the *CceBE1*-G7 complex. Additionally, Fig. 4b indicates that in the *RpBE*-G7 complex, seven residues interact with the polysaccharide chain, wherein residue W270 (*RpBE* numbering) is the conserved residue, corresponding to residue W399 in the *CceBE1*-G7 complex. Ban et al. also aligned their docking *RoBE*-G7 complex with the *CceBE1*-G7 complex, and deduced a possible conformation of a long-chain substrate binding to *RoBE* through the poses of the two G7 substrates (Ban, X. et al., *Food Res. Int.*, 2022, 162, 112119). In light of the result of our alignment, the overall conformation of the two G7 substrates (named *CceBE1* G7-pose i here) suggests an ideal initial binding pose of the donor substrate on *RpBE*.

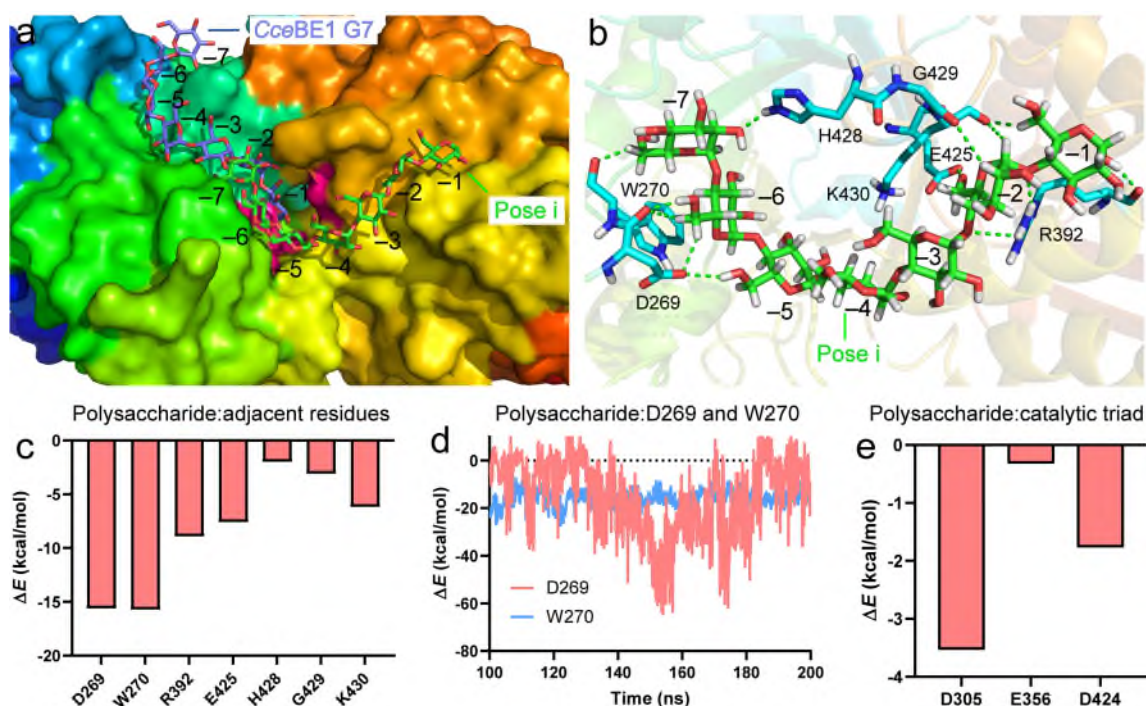


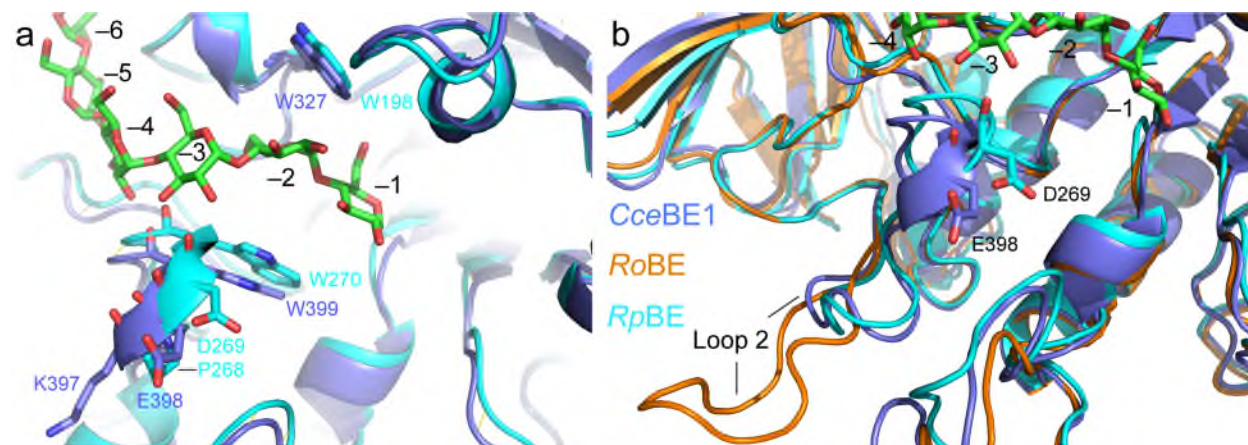
Fig. 4 SBTs for glycosylation. **a** Alignment of pose i with the substrate in the *CceBE1*-G7 complex. The carbon atoms of the substrates in the *CceBE1*-G7 complex (PDB code 5GQX) and the *RpBE*-G7 complex (pose i) are colored in blue and green, respectively. In the center of *RpBE*, the catalytic triad is highlighted in pink. **b** Enzyme-substrate interaction in pose i. The polysaccharide chain and its adjacent residues are displayed as a licorice representation, the carbon atoms of which are colored in green and cyan, respectively. **c** Average interaction energy of the polysaccharide chain with the adjacent residues in pose i. **d** Time series of the interaction energies of residues D269 and W270 with the polysaccharide chain in pose i. **e** Interaction energies of the catalytic triad with the polysaccharide chain in pose i.

5) Having identified D269 as a key residue in the “prebinding” interaction, the authors mutate this residue to ala and indeed over 80% of the activity is lost, confirming its importance in the reaction. Though an interesting result, it does not prove that D269 is working the way the authors suggest. It could be affecting either donor or acceptor chain binding, serving as an important residue for conformational change of the loop etc.

Our theoretical results supported by experiment demonstrate that D269 is the key residue that affects the interaction of the donor chain with *RpBE*. We agree with the Reviewer’s viewpoint that D269 could also affect the acceptor chain binding with *RpBE*. However, in this work, we aim at identifying the most crucial residue responsible for substrate binding and transfer. Given the results of docking, simulations, as well as free-energy calculations, we found that E425 is the most crucial residue that affects the binding of the acceptor chain binding with *RpBE*.

6) I’m not sure what the mutagenesis of P268 proves in support of this mechanism. It is a proline, which is often important to maintain structure, the structures of the cyano and RhBE enzymes are quite similar here.

As shown in Supplementary Fig. 11, in the aligned structures, P268 in *RpBE* corresponds to K397 in *CceBE1*. This difference led us to explore whether the mutagenesis of P268 could affect our proposed mechanism and the activity of *RpBE*.



Supplementary Fig. 11 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.

7) The authors then use their docking results to propose another mechanism involving prebinding for the acceptor chains. My concerns on this are similar to that above. Can this enzyme transfer a glucan to an acceptor chain as short as M7? If not, then there are probably essential interactions missing.

Please see the response to point 2). The theoretical investigations of G7 binding to BEs, together with the numerous BE-G7 crystallographic complexes obtained experimentally led us to focus in the present study on *RpBE*-G7 association.

8) The strongest part of the paper is depicted in Figure 8, where Rosetta was used to predict mutations that would increase binding and reactivity. However, Figure 8 is very difficult to understand. The authors show a glucan spanning A243 and K436, but in the alphafold model these residues are almost 30Å away from each other. How does an M7 span this distance?

To make the results of enzyme engineering easier to understand, we have modified Fig. 8 and the corresponding descriptions. In our response to point 4), we state that the overall conformation of the two G7 substrates, i.e., the CceBE1 G7-pose i, suggests an ideal initial binding pose of the donor substrate on R_pBE. Additionally, as depicted in Fig. 8b, seven mutation sites, i.e., A243, K436, G431, G429, P420, L421, and H385, lie around the CceBE1 G7-pose i, which includes four of the five top variants, i.e., A243T (1.7-fold increase), K436R (1.5-fold increase), G431N (1.5-fold increase), and G429M (1.4-fold increase). Our enzymatic kinetic experiments indicate that in the top two variants (A243T and K436R), the *K_m* values decrease by 44.2% and 46.5%, respectively, suggesting enhanced enzyme-substrate affinity (see Supplementary Table 10). The overall conformation of the two G7 spans well the distance between A243 and K436.

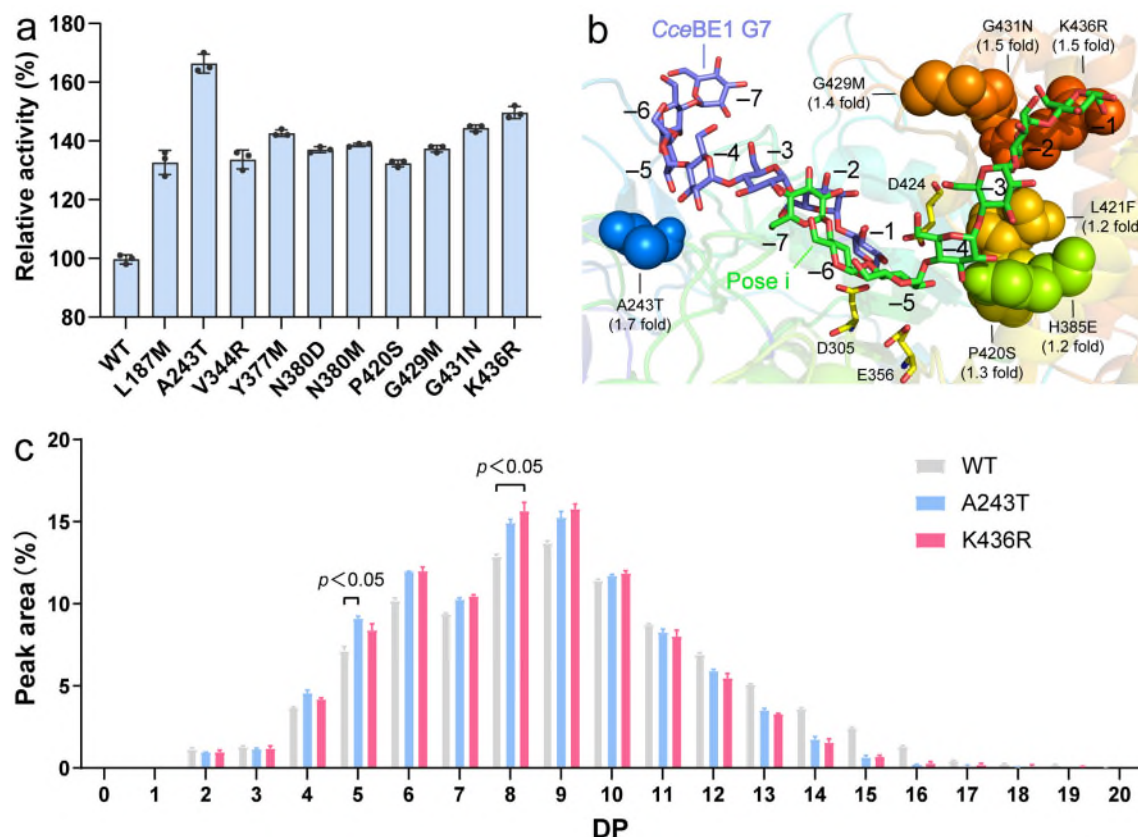


Fig. 8 Enzyme engineering. **a** Top ten variants in the forty-seven ones. The WT R_pBE is regarded to possess 100% enzyme activity. **b** Mutation sites (shown in sphere) around the CceBE1 G7-pose i. The catalytic triad is shown in a licorice representation, and colored in yellow (carbon atoms). **c** Distribution of the glucan chain length. The data without enzyme catalysis are not shown for clarity. The bars and the error bars are the average and standard deviation of triplicates (*n* = 3 independent experiments), respectively.

9) Now looking at Figure 4a and b versus c and d, I am having an impossible time understanding the relationship. In fact Figure 4a is essentially incomprehensible. In short I can't tell where site A and site B are in the structure, nor do I understand the relationship between them with respect to the poses. This must be made clear.

Please, see our response to point 4). Considering the concerns expressed by the Reviewers, we now only discuss the optimal pose i in the present manuscript.

10) It is clear that improvement in activity has happened using Rosetta. However, where these positions are relative to the “poses” is not clear at all. For several of these positions, the cyano BE M7 is interacting. The authors should compare what they have with what the experiment shows.

Please, see our response to point 8)

Response to Reviewer #2:

In the manuscript by Zhiyou Zong et al., the discovery of a new type of branching enzyme (BE) with superior properties for industrial production is presented, supported by modeling, docking, and simulation studies to explore the enzyme's mechanism. However, several areas require clarification and improvement to solidify the findings. Therefore, I recommend rejecting the manuscript in its current form but encourage resubmission following the solution of the outlined concerns:

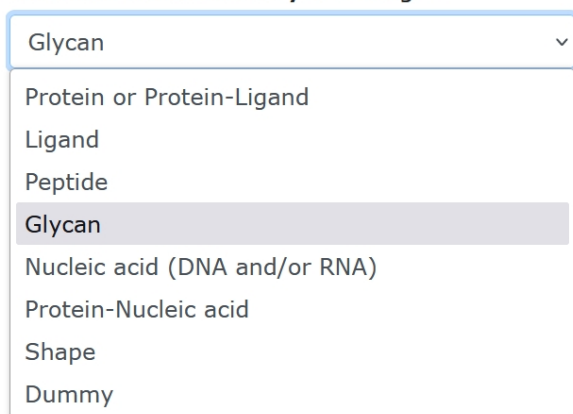
1) How were the structures shown in Fig. 2A obtained? From homology modeling or AlphaFold2 database? It needs to be clarified in the main text.

In the ‘**Structural feature**’ section of the ‘**Results**’, we state that ‘...the three-dimensional structure of *RpBE* constructed with AlphaFold v2.0 possesses a remarkable average model confidence of 96.4.’

2) Reference is necessary to prove the responsibility of using HADDOCK to dock oligosaccharides to *RpBE* since HADDOCK is always used in protein-protein docking rather than glycan docking. Maybe a specific docking software designed for glycans can be applied to compare with the current docking pose.

HADDOCK version 2.4 is not only used for protein-protein docking, but also applies, among others, to peptides, glycans, and DNA/RNA (see Fig. r2). HADDOCK version 2.4 features specific docking algorithms for the above ligands. In our docking procedure, after 50,000 docking trials, the top 1,000 enzyme-substrate complexes were selected and further refined by short molecular dynamics (MD) simulations in explicit solvent. The refinement step using MD simulations greatly contributes to optimizing the binding pose of the substrate on the enzyme; however, this function is not available in other docking software. Additionally, the clustering information provided by HADDOCK version 2.4 reflects the number of polysaccharides lying at a similar location, which is an important indicator to evaluate different poses. In this spirit, and from our perspective, HADDOCK version 2.4 is an ideal software for studying the *RpBE*-G7 binding.

What kind of molecule are you docking? *



A screenshot of a web-based docking interface. At the top, there is a label "What kind of molecule are you docking? *". Below it is a dropdown menu. The menu is currently open, showing a list of options: "Glycan", "Protein or Protein-Ligand", "Ligand", "Peptide", "Glycan" (highlighted with a grey background), "Nucleic acid (DNA and/or RNA)", "Protein-Nucleic acid", "Shape", and "Dummy". A small downward arrow is visible in the top right corner of the dropdown box.

Fig. r2 Screenshot of the docking process using HADDOCK version 2.4.

3) In the Methods part, MD simulations were only run for 1 ns. Why did the authors claim microsecond-level simulations in the main text? In fact, the Methods part needs a lot of supplements. For MD simulations, how the temperature and pressure are controlled? Is there any constraint on hydrogens? How is the interaction energy calculated (that's a key point!)? Also, how Rosetta was used for generating mutation strategies needs to be explained. Suggest the authors take care of the reproducibility of their research.

1 ns was used to pretreat the protein-substrate assemblies. A 2.2- μ s MD simulation was performed, devoid of geometric restraints, for the five assemblies. Please see Supplementary Table 12, as well as the details of the simulations. All the simulations were carried out at 300 K. The temperature and the pressure were controlled by Langevin dynamics and the Langevin piston method, respectively. The Shake/Rattle and Settle algorithms were utilized to constrain the length of the covalent bonds involving hydrogen atoms to their equilibrium value. The interaction energy is defined as the sum of the electrostatic interaction energy and the van der Waals interaction energy. Additionally, to enhance the enzyme-substrate binding affinity, we employed the Rosetta Cartersian_ddg module to screen the mutations following the software protocols. As shown in Supplementary Table 9, the experiments which were repeated three times independently, are completely reproducible.

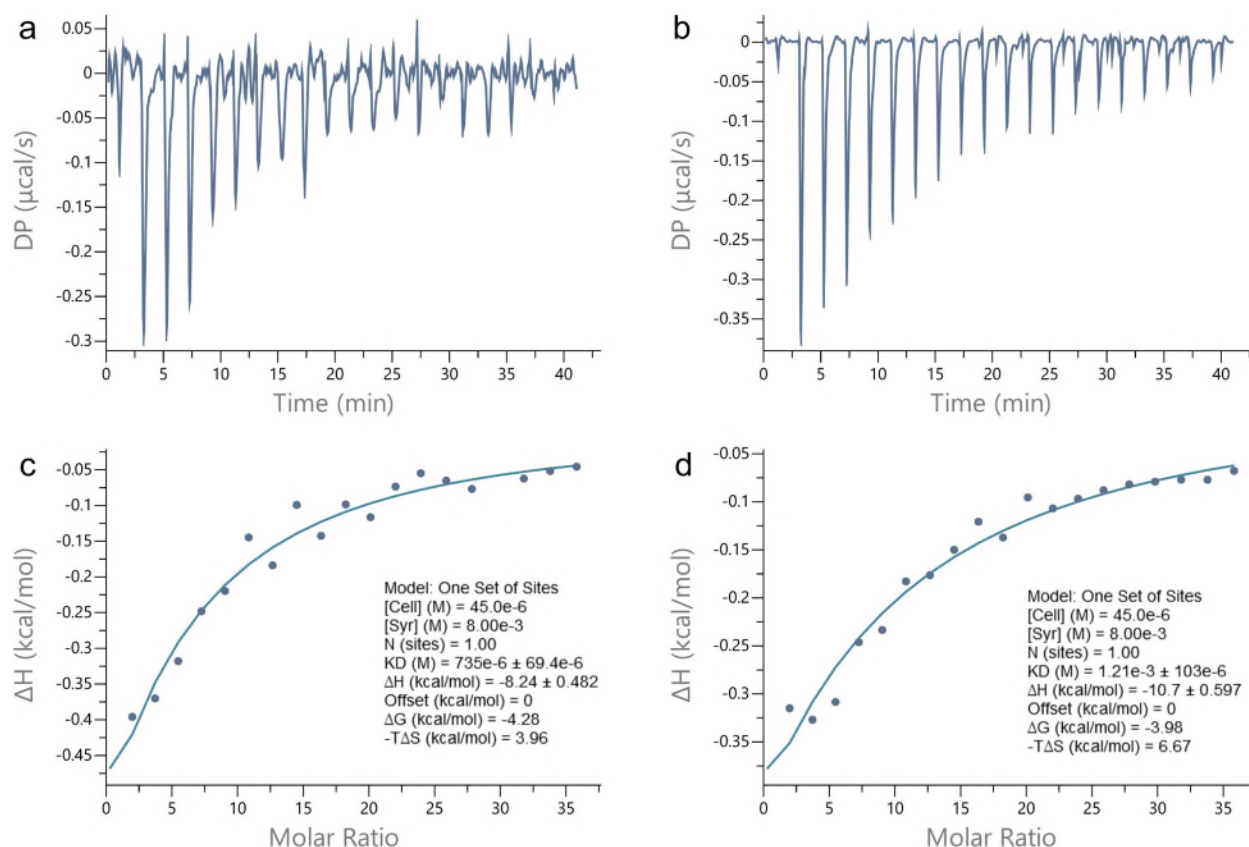
4) To capture the catalysis process, metadynamics is not enough. I suggest the author provide QM/MM simulations to show the reaction process and prove the residue importance claimed by them. Cheap QM levels, especially semi-empirical ones, will not cost too much but provide a more direct view of such a process. But if permits, I suggest a standard DFT QM in QM/MM calculation.

We are grateful to the Reviewer for their suggestion. Our work, however, focuses on a noncovalent enzyme-substrate interaction, in sharp contrast with covalent catalysis. Therefore, using QM/MM simulations to investigate the catalytic process goes beyond the scope of this article. Additionally, D269, D305, and E425 are considered to modulate the noncatalytic process, and their roles, therefore, cannot be investigated by QM/MM simulations.

Response to Reviewer #3:

1) In this work the enzymatic activity of Branching Enzyme (BE) from *Rhodothermus profundus* is compared with other twelve BEs. The authors exploited the structural model obtained with AlphaFold program to run docking calculations and MD simulations in order to characterize the substrate binding and transfers steps (SBTs). The molecular dynamic simulations are well performed and provide possible structural models of the SBT steps of the reaction, they can be used in combination with experimental methods. The work indeed misses experimental evidences as site directed mutagenesis experiments are not sufficient to explain the mechanism. Information about the mechanism of a reaction are generally provided by chemical kinetics studies, NMR spectroscopy of the substrate could be also used to observe the reaction intermediates. ITC measurements can provide valuable information about the thermodynamic parameters of binding.

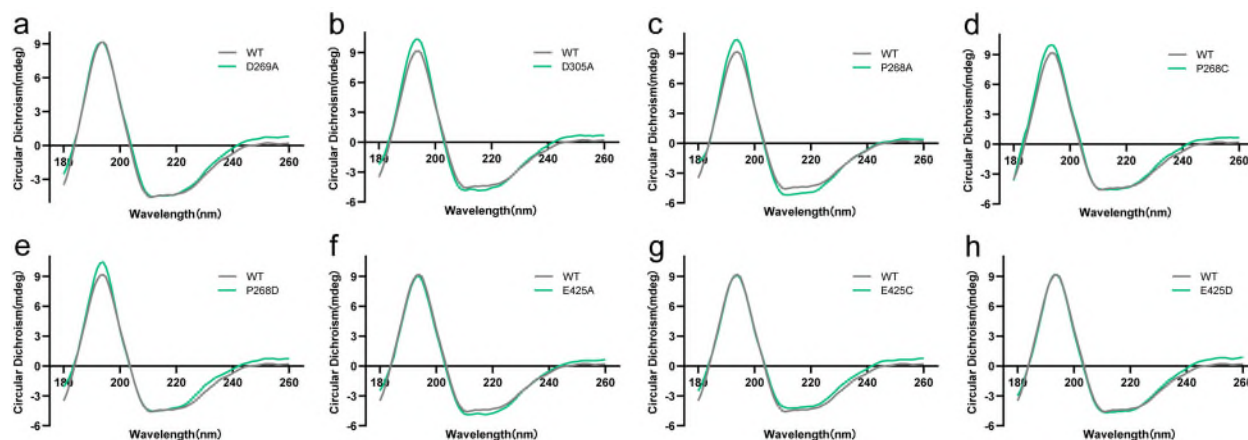
The authors appreciate the suggestion of methods by the Reviewer. However, the reaction intermediates of the substrates, which can be detected by NMR spectroscopy, are used to describe the catalytic process, which is unrelated to our research on the noncatalytic process. ITC is used to measure the binding affinity of an enzyme with another enzyme, or any substrate. Although we focus in this study on binding patterns, we measured the binding affinity of RpBE with G7 following the Reviewer's suggestion. Our results indicate that the RpBE-G7 binding affinity is -4.1 ± 0.2 kcal/mol (see Supplementary Fig. 8).



Supplementary Fig. 8 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.

2) Furthermore, the authors ought to demonstrate the proper folding of the mutated BEs. This entails confirming that any observed reduction in activity stems from the mutation of the key residue rather than a loss of the protein's three-dimensional structure.

As shown in Fig. 1, the wild-type (WT) *RpBE* has approximately twice the activity of the commercial *RoBE*. We, therefore, reason that the WT *RpBE* is correctly folded. Circular Dichroism (CD) spectropolarimetry, which can detect the secondary-structure elements of a protein, was employed to compare the difference between the WT *RpBE* and the variants. As shown in Supplementary Fig. 10, the CD result of the WT *RpBE* is similar to that of the D269A, D305A, P268A, P268C, P268D, E425A, E425C, and E425D. Additionally, in the field of enzyme engineering, we often evaluate whether proteins fold correctly by observing the level of soluble expression, which was the case during our experiment. Enzyme activity is another indicator that can be used to evaluate whether the proteins fold correctly. As shown in Supplementary Table 9, we obtained many variants, the activity of which is much higher than that of the WT *RpBE*. These evidences indicate that the proteins we produced through standardized processes are properly folded.



Supplementary Fig. 10 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.

3) Finally, It is not clear how the authors selected the 138 residues used for HADDOCK? Which are the active residues used as input to run docking calculations? Which data are used to select them?

The purpose of this work is to investigate the binding and transfer mechanism of a polysaccharide chain on the side of the active-site cleft. In order to not miss any docking site, as shown in Supplementary Fig. 6, we, therefore, selected all the residues lying on the side of the active-site cleft as the active residues to run docking calculations.

4) Minor issues:

The authors describe that the 1D NMR experiment was recorded through a Varian Inova 600 spectrometer at 413 K. I think the authors misspelled the acquisition temperature.

We have corrected 413 K to 313 K.

Figure 4 is very difficult to interpret. The orientation of the 3D structure should be equal.

Please see our response to point 4), Reviewer 1. Considering the concerns raised by the Reviewers, we now only discuss the optimal pose i in the present article.

5) In conclusion I think that this is a preliminary work where the computational aspects need to be validated by experimental studies. The authors should investigate deeper the mechanism of the BE activity before considering the study acceptable for publication in Nature Communications.

The authors would like to express a different opinion here. As mentioned in the main text, this work focuses on the enigmatic noncovalent enzyme-substrate binding and transfer mechanism before glycosylation and transglycosylation. Crystallographic data have heretofore only provided limited static images. For instance, the *CceBE1-G7* complex obtained by Hayashi et al. (*J. Biol. Chem.*, 2017, 5465) is the only experimental evidence of polysaccharide chain binding in the active-site cleft.

Furthermore, the process of substrate binding and transfer in the active-site cleft is dynamic, and reporting a faithful account of the dynamic process of enzyme-substrate interactions by crystallographic approach constitutes a daunting challenge. Providing a general means to decipher these enzyme-substrate interactions is, therefore, highly desirable.

In the present work, guided by large-scale state-of-the-art computer simulations, our enzyme kinetic experiments demonstrate that the enzyme-substrate affinity is enhanced by 46.5%. The mechanism-based enzyme engineering generated a host of variants with higher activity than the wild-type *RpBE* (see Supplementary Table 9; up to 70% increase). Additionally, we determined two crucial areas around residues of A243 and K436 (*RpBE* numbering); the engineering on these two residues increases the branching products with DP 5 and DP 8 by 27.3% ($p < 0.05$) and 21.4% ($p < 0.05$) respectively, which could help product customization with specific branching products.

By decrypting the mechanism at play, we further identified three key residues (i.e., P268, D269, and E425) responsible for the high activity of *RpBE*, which will contribute to guiding the family evolution of branching enzymes. The evidence supplied by these experiments fully supports mechanism that we inferred.

We are, therefore, convinced that the present work, which takes both the mechanism and the enzymatic industrial performance into account through the synergy of theory with experiment, will constitute a milestone for the future engineering of branching enzymes, and can be pivotal in future improvement of these enzymes for realization of more cost-effective α -glucan modifications.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Review: 482155_1.

Issue1: The authors have satisfactorily answered the issue regarding the flexible loops with the crystal structure of RpBE which shows the conformation of the second loop to be relatively similar to their alphafold prediction, though apparently not identical.

Issue 2: The major question here is the use of a G7 glucan for modeling. The authors justify this choice because most of the crystal structures of BE's and indeed other glucan-binding enzymes use G7 as the longest glucan, but of course the reason for this is availability. It has been very difficult to impossible to obtain homogeneous samples of glucans longer than G7 in amounts sufficient for structural work. Obviously a computational study suffers from no such limitation, I therefore don't find the rationale to be convincing. The authors then state that G10 is the shortest substrate that will react with RpBE, and refer to supplemental Fig 3c as evidence. However, Supplemental Fig 3c shows how the chain length distribution changes when RpBE reacts with amylopectin. I don't see how this figure speaks to the question of what the minimum chain length of a substrate is. The authors are going to have to clarify this point. More disturbing is the apparent complete lack of ability for their modeling software to properly dock G10 on the enzyme. Their docking shows G10 making a loop very near the active site, which is both completely inconsistent with any of the structures, and inconsistent with the docking of G7, which shows no such anomaly. This seems to indicate that the docking software is simply not up to the task of properly modeling and docking maltooligosaccharides onto the enzyme. Unfortunately I find the response to this issue begging more questions and issues than it answers. I understand that these docking experiments are difficult and time consuming, nonetheless they must be relevant to have meaning, the result with G10 compels me to question that.

Issue 3: The authors have properly responded to the incorrect identification of the catalytic aspartate. Fig 3 has been properly corrected.

Issue 4: The authors now compare their "pose i" with the donor chain G7-bound Cce structure as requested. This is where the paper is substantially altered relative to the original ms. In the previous version, they showed a number of docking poses, and hypothesized potential function for some of them. They now only show the pose i docking pose. The authors state:

"In light of the result of our alignment, the overall conformation of the two G7 substrates (named CceBE1 G7-pose i here) suggests an ideal initial binding pose of the donor substrate on RpBE."

From what I understand of this statement, the authors are suggesting that an "ideal initial binding pose" could be derived by making a continuous maltooligosaccharide using element of both pose i and the CceBE1 G7 structure. If the authors want to suggest such a structure, then they should model it, by modeling a continuous glucan that shows this. From what I can see in the figure, the closes residue to the active site in pose i is G6, which is nearest the G1 of Cce BE1, which is bound in the active site, oriented properly to react covalently with the aspartate that is equivalent to D305. This seems to suggest that the authors hypothesis is that the glucan in the "initial binding pose"

would reflect the binding of the Cce BE residues at the non-reducing end. If this is the case, the authors should model with a single glucan such an “initial binding pose.”

Issue 5: The issue of alternate functions for D269 is addressed as well as is possible here.

Issue 6: The issue regarding P268 is what, exactly do the authors believe this residue is doing. According to their calculations mutation of this residue would appear to increase the barrier between prebinding and reactive binding. The question is how? Does it have to do with required motions of the 269 loop? Does it alter the orientation of residue 269? The authors should clarify what the actual role of the residue is according to their mechanism.

Issue 7: Again, this reviewer is not convinced of the authors arguments to justify using shorter glucans in the modeling, this time with regard to the acceptor strand.

Issue 8: Using the Cce donor strand structure together with the proposed “pose I” now explains this issue. The figure is now more clear, though one continuous glucan modeled into the structure would again be germane.

Issue 9: this is satisfactorily answered by the new figure 4.

Issue 10: This is addressed by the Cce structure as discussed above.

Reviewer #2 (Remarks to the Author):

After such revision, the quality of this manuscript was improved but several key points in my comments were not answered.

1. The evidence supporting the claim that "HADDOCK can be applied to glycans" is insufficient. Providing a glycan docking option does not inherently validate the docking ability for glycans. I suggested that the authors compare HADDOCK with specialized glycan docking software (e.g., GlycanTorchVina or Vina-Carb), but this was not addressed. The efficacy of HADDOCK for glycan docking needs to be substantiated with comparative data rather than just saying 'It provides glycan mode'.

2. Reproducibility goes beyond stating that all experiments were repeated three times. It is crucial to provide detailed methodologies so others can replicate the experiments. For instance, the interaction energy calculation method appears unclear. Did the authors use MMPBSA.py, another tool, or did they manually extract residues to calculate the electrostatic and van der Waals energies? If the latter, the corresponding scripts should be provided. Additionally, the feasibility of this calculation method needs validation. Are there other publications employing this approach?

3. Regarding the catalytic process, I understand that the residue does not form a covalent interaction with the glycan. However, the glycan itself undergoes reactions, as depicted in the updated Figure 4 (right part). This mechanism is inferred rather than being directly modeled in silico.

4. The authors mention solving a crystal structure but did not include it in the revised manuscript. This crucial information should be incorporated into the manuscript rather than merely communicated to the reviewers.

5. In Figure 8, does G7 have two possible binding poses, one involving A243 and the other involving all other residues? If mutations result in less than a twofold change, this does not convincingly confirm the binding site. More substantial evidence is required to support this claim.

Reviewer #3 (Remarks to the Author):

As previously reported, this study investigates the enzymatic activity of the Branching Enzyme (BE) from *Rhodothermus profundus*, comparing it with twelve other BEs. The authors used a structural model obtained with the AlphaFold program to conduct docking calculations and molecular dynamics (MD) simulations, characterizing the substrate binding and transfer steps (SBTs). The well-executed MD simulations reinforce their conclusions.

In this revised version, the authors have enhanced the manuscript with new experimental data, as requested by the reviewer. They conducted circular dichroism (CD) experiments to provide evidence of the folding of the BE mutants and determined the X-ray structure of the RpBE protein. These additions significantly improve the quality of the manuscript, making it suitable for publication in *Nature Communications*.

The Response to the comments of the reviewers

We are grateful to the reviewers for their constructive criticisms of our manuscript. Below, we reply point by point to their comments, and, for editorial purposes, have highlighted in one of the copies of the revised manuscript all the changes brought to the text.

Response to Reviewer #1:

1) The authors have satisfactorily answered the issue regarding the flexible loops with the crystal structure of RpBE which shows the conformation of the second loop to be relatively similar to their alphafold prediction, though apparently not identical.

We are grateful to the Reviewer for accepting our revision regarding point 1).

2) The major question here is the use of a G7 glucan for modeling. The authors justify this choice because most of the crystal structures of BE's and indeed other glucan-binding enzymes use G7 as the longest glucan, but of course the reason for this is availability. It has been very difficult to impossible to obtain homogeneous samples of glucans longer than G7 in amounts sufficient for structural work. Obviously a computational study suffers from no such limitation, I therefore don't find the rationale to be convincing. The authors then state that G10 is the shortest substrate that will react with RpBE, and refer to supplemental Fig 3c as evidence. However, Supplemental Fig 3c shows how the chain length distribution changes when RpBE reacts with amylopectin. I don't see how this figure speaks to the question of what the minimum chain length of a substrate is. The authors are going to have to clarify this point. More disturbing is the apparent complete lack of ability for their modeling software to properly dock G10 on the enzyme. Their docking shows G10 making a loop very near the active site, which is both completely inconsistent with any of the structures, and inconsistent with the docking of G7, which shows no such anomaly. This seems to indicate that the docking software is simply not up to the task of properly modeling and docking maltooligosaccharides onto the enzyme. Unfortunately I find the response to this issue begging more questions and issues than it answers. I understand that these docking experiments are difficult and time consuming, nonetheless they must be relevant to have meaning, the result with G10 compels me to question that.

We are concerned that there might be some misunderstanding here. As mentioned by the Reviewer, the structural work of branching enzyme in complex with long glycans is limited by substrate availability. In the same way, the computational study also suffers from limitations associated to the substrate. There is a number of docking software capable of handling enzyme-glycan association, which includes HADDOCK version 2.4, GlycoTorch Vina, and Vina-Carb (unfortunately, the recent AlphaFold 3 does not support enzyme-glycan docking). HADDOCK version 2.4, which features a specific docking algorithm for glycans, effectively predicts the RpBE-G7 association. In the RpBE-G10 complex, the long glycan adopts a twisted conformation, wherein three glucosyl rings do not interact with the enzyme. We, therefore, reason that compared to G10, using G7 as a substrate is more appropriate to explore enzyme-substrate association.

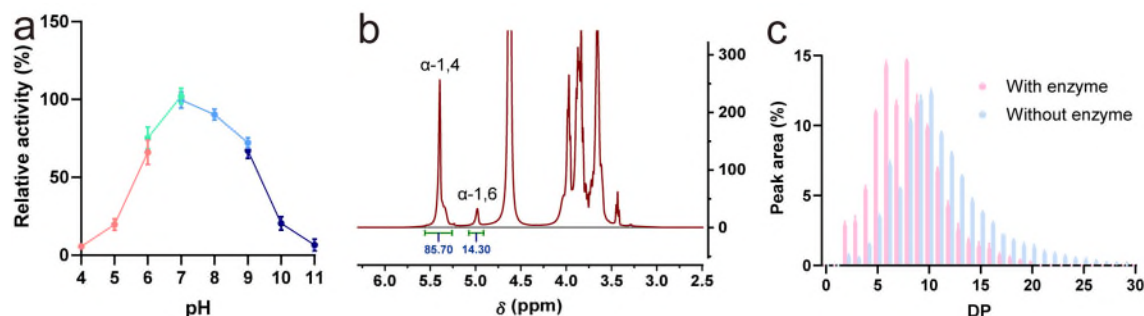
GlycoTorch Vina and Vina-Carb were developed to improve the glycosidic conformational equilibria, for instance, the glycosidic angle, during carbohydrate docking. For GlycoTorch Vina, Boittier et al. chose twelve enzyme-glycan crystals, ranging in length from tetra (G4)- to octasaccharides (G8), to optimize the docking algorithm (*J. Chem. Inf. Model.* 2020, 6328). The predictive accuracy of this software for longer polysaccharides (\geq G9) is, therefore, uncertain. We, therefore, have no other choice, but to compromise between docking accuracy and substrate length.

Ban et al. employed G7 binding to RoBE to examine potential mutation sites that affect enzyme activity (*Food Res. Int.*, 2022, 112119)—a docking study exploring branching enzyme-glycan interaction using G7, which the Reviewer probably knows. An additional value of docking is to provide working hypotheses that can be tested experimentally, as opposed to true representations. Computational results, therefore, still need proper experimental validation. In the present work, we performed extensive experimental verifications and validations in support of our calculations.

Our mechanism-guided mutation experiments result in a host of variants with higher activity than the wild-type RpBE; seven mutation sites lie around the CceBE1 G7-pose i, which includes four of the five top variants. The enzyme kinetic experiments for the optimal variants of A243T (1.7-fold increase) and K436R (1.5-fold increase) demonstrate that the residue replacements, as expected, enhance the enzyme-substrate affinity. Additionally, branching product determination experiments reveal that the products with the specific degree of polymerization (DP; i.e., DP 5 and DP 8) are increased by 27.3% (A243T, $p < 0.05$) and 21.4% (K436R, $p < 0.05$), respectively. We are earnestly convinced that our large-scale, state-of-the-art computational strategy supported by experiment is both robust and reliable for studying branching enzyme-glycan interaction.

Regarding the minimum substrate length for branching enzyme catalysis, in the work of Thiemann et al., upon incubation of the branching enzyme with a mixture of linear dextrans, the majority of larger oligosaccharides ($\geq \text{DP } 16$) were consumed, which indicates that DP 16 is the minimal chain length required to serve as donor for the branching enzyme of *A. gottschalkii* (*Appl. Microbiol. Biotechnol.*, 2006, 60). In our work, the amylopectin was debranched by 1 U/mL pullulanase and 2 U/mL isoamylase to obtain the debranched amylopectin, i.e., linear dextrin or short-chain amylose (please see the ‘DP analysis’ section of the ‘Methods’). Then, the debranched amylopectin was catalyzed by RpBE to add α -1,6-glucoside-linked branching chains. The branching products were further debranched. High-performance anion-exchange chromatography was employed to analyze which glycan chains were consumed and which ones were increased. In Supplementary Fig. 3c, after enzymatic catalysis, the $\text{DP} \geq 10$ glycan chains were consumed. G10 can, therefore, be regarded as the minimum donor substrate required for RpBE catalysis. This method of inferring the minimum substrate length for branching enzyme catalysis has been adopted by many scientists (*J. Mol. Microbiol. Biotechnol.*, 2016, 303; *Plant Cell Physiol.*, 2010, 776; *Int. J. Biol. Macromol.*, 2017, 156; *PLoS One*, 2019, e0219844, to cite a few).

We noticed as we were revising our manuscript that the legend for Supplementary Fig. 3c is insufficiently detailed. We are most grateful to the Reviewer for their concern regarding the minimum substrate length for branching enzyme catalysis. We have provided additional details in the legend of Supplementary Fig. 3c.



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

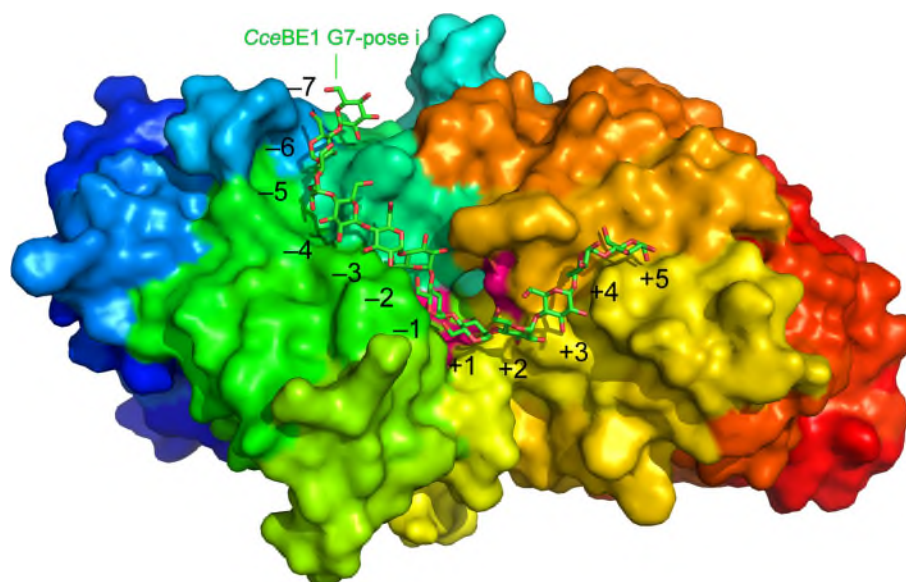
corresponding glucosidic linkages. The product of *RpBE* action on amylose possesses 14.3% branching degree. In (c), the substrate is debranched amylopectin (i.e., linear dextrin or short-chain amylose). After enzyme catalysis, the DP \geq 10 glycan chains were consumed. G10 is, therefore, regarded as 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.

3) The authors have properly responded to the incorrect identification of the catalytic aspartate. Fig 3 has been properly corrected.

We are most grateful to the Reviewer for accepting our revision regarding point 3).

4) The authors now compare their “pose i” with the donor chain G7-bound Cce structure as requested. This is where the paper is substantially altered relative to the original ms. In the previous version, they showed a number of docking poses, and hypothesized potential function for some of them. They now only show the pose i docking pose. The authors state: “In light of the result of our alignment, the overall conformation of the two G7 substrates (named CceBE1 G7-pose i here) suggests an ideal initial binding pose of the donor substrate on *RpBE*.” From what I understand of this statement, the authors are suggesting that an “ideal initial binding pose” could be derived by making a continuous maltooligosaccharide using element of both pose i and the CceBE1 G7 structure. If the authors want to suggest such a structure, then they should model it, by modeling a continuous glucan that shows this. From what I can see in the figure, the closest residue to the active site in pose i is G6, which is nearest the G1 of Cce BE1, which is bound in the active site, oriented properly to react covalently with the aspartate that is equivalent to D305. This seems to suggest that the authors hypothesis is that the glucan in the “initial binding pose” would reflect the binding of the Cce BE residues at the non-reducing end. If this is the case, the authors should model with a single glucan such an “initial binding pose.”

Ban et al. aligned their docking *RoBE*-G7 complex with the CceBE1-G7 complex, and deduced a possible conformation of a long-chain substrate binding to *RoBE* through the poses of the two G7 substrates (*Food Res. Int.*, 2022, 112119). To predict the possible conformation of a long-chain substrate binding to our *RpBE*, as shown in Supplementary Fig. 10, and following the Reviewer’s suggestion, we modeled a continuous glycan chain by referring to the substrate conformations in pose i (–2 to +5 glucosyl rings) and CceBE1-G7 (–3 to –7 glucosyl rings). Additionally, we amended the main text accordingly. In particular, the sentence “In light of the result of our alignment, the overall conformation of the two G7 substrates (named CceBE1 G7-pose i here) suggests an ideal initial binding pose of the donor substrate on *RpBE*” has been changed to “In light of the result of our alignment, we suggest the possible conformation of a long-chain donor substrate binding to *RpBE* (see Supplementary Fig. 10; named CceBE1 G7-pose i here)”.



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.

5) The issue of alternate functions for D269 is addressed as well as is possible here.

We are most grateful to the Reviewer for accepting our revision regarding point 5).

6) The issue regarding P268 is what, exactly do the authors believe this residue is doing. According to their calculations mutation of this residue would appear to increase the barrier between prebinding and reactive binding. The question is how? Does it have to do with required motions of the 269 loop? Does it alter the orientation of residue 269? The authors should clarify what the actual role of the residue is according to their mechanism.

We apologize to the Reviewer for not explaining in our previous revision the mechanism of P268 as clearly as we should have. Our motivation for studying P268 comes from the structural comparison between *RpBE* and CceBE1 (please see Supplementary Fig. 13a). In the aligned structures, P268 in *RpBE* corresponds to K397 in CceBE1. Furthermore, in the twelve branching enzymes investigated in this work, ten enzymes possess a proline at position 268 (*RpBE* numbering), while in the other two enzymes, the residues at that position are a lysine (same to CceBE1) and a glutamine, respectively (please see Supplementary Fig. 15). This difference led us to explore whether the mutagenesis of P268 could influence the 'G-R/C-S' actions, and further modulate enzyme activity.

Our experimental results suggest that P268 in *RpBE* is the optimal residue (please see Supplementary Fig. 14). The results of our simulation indicate that the 'G-R/C-S' actions mediated by residues D269 and D305 have been obliterated by mutagenesis (please see Fig. 5d). Fig. 5e implies that in the WT *RpBE*, the free energy is favorable for D269 approaching D305, which will help D269 deliver the polysaccharide to D305 to facilitate the subsequent releasing action of D269 and the competing action of D305. In contrast, in the P268A variant, the free energy is very unfavorable for D269 approaching D305. P268 is, therefore, considered to play a role in the process of D269 approaching D305.

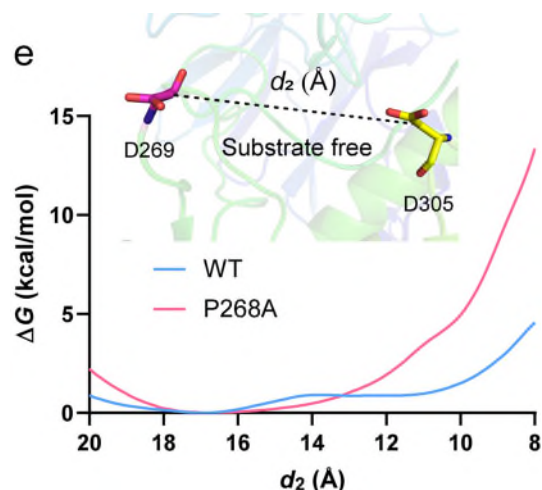


Fig. 5e One-dimensional free-energy profile, or potential of mean force, characterizing residue D269 approaching residue D305 in the polysaccharide-free model. d_2 denotes the distance between the centers of mass of residues D269 and D305.

We aligned the WT *RpBE* with the P268A variant, and found that the orientation of D269 is not altered by mutagenesis. Loop 2, including P268 and D269, plays an important role in enzyme-substrate association. Furthermore, as shown in Supplementary Fig. 5, loop 2 possesses a high flexibility. Proline, which has the feature of conferring structural stability, can stabilize loop 2. We, therefore, infer that P268 can modulate the motion of loop 2 in enzyme-substrate interaction to promote the approach of D269 towards D305. We have included in the revised manuscript our explanation on the role of P268.

7) Again, this reviewer is not convinced of the authors arguments to justify using shorter glucans in the modeling, this time with regard to the acceptor strand.

We refer the Reviewer to the response of point 2). Constrained by the current technological conditions, we had to compromise between the software's docking accuracy and the substrate length. In addition, there are many challenges in enzyme-glycan docking, whereby useful and testable experimental hypotheses generated by docking require experimental validation. In the present study, we show that our theoretical prediction agrees well with experiment.

8) Using the Cce donor strand structure together with the proposed “pose I” now explains this issue. The figure is now more clear, though one continuous glycan modeled into the structure would again be germane.

We are most grateful to the Reviewer for accepting our revision regarding point 8). Following the Reviewer’s suggestion, we have modeled a continuous glycan chain, i.e., the CceBE1 G7-pose i, and updated Fig. 8b.

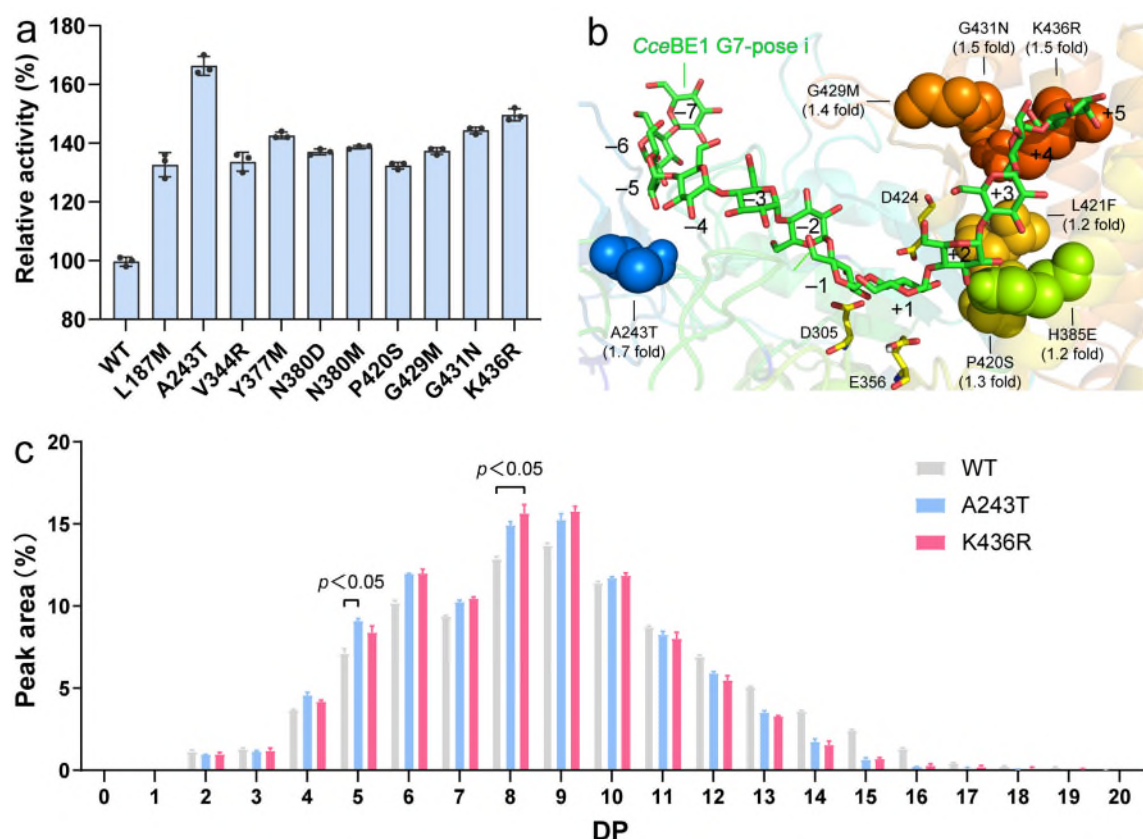


Fig. 8 Enzyme engineering. **a** Top ten variants amid the forty-seven ones. The WT *RpBE* is regarded as possessing 100% enzyme activity. **b** Mutation sites (shown in sphere) around the CceBE1 G7-pose i. The catalytic triad is shown in a licorice representation, and colored in yellow (carbon atoms). **c** Distribution of the glycan chain length. The data without enzyme catalysis are not shown for clarity. The bars and the error bars are the average and standard deviation of triplicates ($n = 3$ independent experiments), respectively.

9) this is satisfactorily answered by the new figure 4.

We are most grateful to the Reviewer for accepting our revision regarding point 9).

10) This is addressed by the Cce structure as discussed above.

We are most grateful to the Reviewer for accepting our revision regarding point 10).

Response to Reviewer #2:

After such revision, the quality of this manuscript was improved but several key points in my comments were not answered.

1) The evidence supporting the claim that "HADDOCK can be applied to glycans" is insufficient. Providing a glycan docking option does not inherently validate the docking ability for glycans. I suggested that the authors compare HADDOCK with specialized glycan docking software (e.g., GlycanTorchVina or Vina-Carb), but this was not addressed. The efficacy of HADDOCK for glycan docking needs to be substantiated with comparative data rather than just saying 'It provides glycan mode'.

Following the Reviewer's helpful suggestion, we read the papers on GlycoTorch Vina (*J. Chem. Inf. Model.*, 2020, 6328) and Vina-Carb (*J. Chem. Theory Comput.*, 2016, 892). The Python source code for GlycoTorch Vina is, however, not available at <https://github.com/EricBoittier/GlycoTorch-Online> (see Fig. r1); the website of www.glycotorch.com mentioned in the Supporting Information of the *J. Chem. Inf. Model.* article is not in service. Still, we found the source code and binaries for GlycoTorch Vina at <https://github.com/EricBoittier/GlycoTorch-Vina>. However, the uncompiled version provided on GitHub cannot be installed on Linux (see Fig. r2). We, therefore, used the Windows version for 1,000 docking trials.

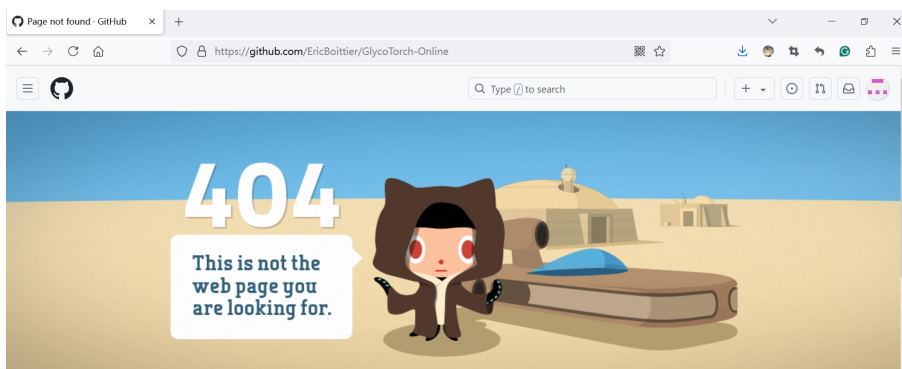


Fig. r1 Python source code for GlycoTorch is not available from the website.

```
wangq@Aaran4070Ti:~/tmp/GlycoTorch-Vina/autodock_vina_1_1_2/build/linux/release
(base)
# wangq @ Aaran4070Ti in ~/tmp/GlycoTorch-Vina/autodock_vina_1_1_2/build/linux/release on git:master
x [11:31:29]
$ make
nvcc --compiler-options="-g -fPIC -static -pthread -ansi -Wno-long-long -O3 -Ofast -DNDEBUG -std=c++11" -I /home/eric/Documents/boost_1_67_0 -I ../../src/lib/glylib/inc/ -I ../../src/lib -o main.o -c ../../src/main/main.cpp
../../src/main/main.cpp:36:10: fatal error: boost/program_options.hpp: No such file or directory
   36 | #include <boost/program_options.hpp>
      |          ^
compilation terminated.
make: *** [../../makefile_common:25: main.o] Error 1
(base)
# wangq @ Aaran4070Ti in ~/tmp/GlycoTorch-Vina/autodock_vina_1_1_2/build/linux/release on git:master
x [11:31:31] C:2
$
```

Fig. r2 GlycoTorch Vina error reported on Linux platform.

As shown in Fig. r3, the top two poses generated by GlycoTorch Vina are very similar to pose i in this study. These results indicate that both the HADDOCK web server and GlycoTorch Vina perform remarkably well in RpBE-G7 docking.

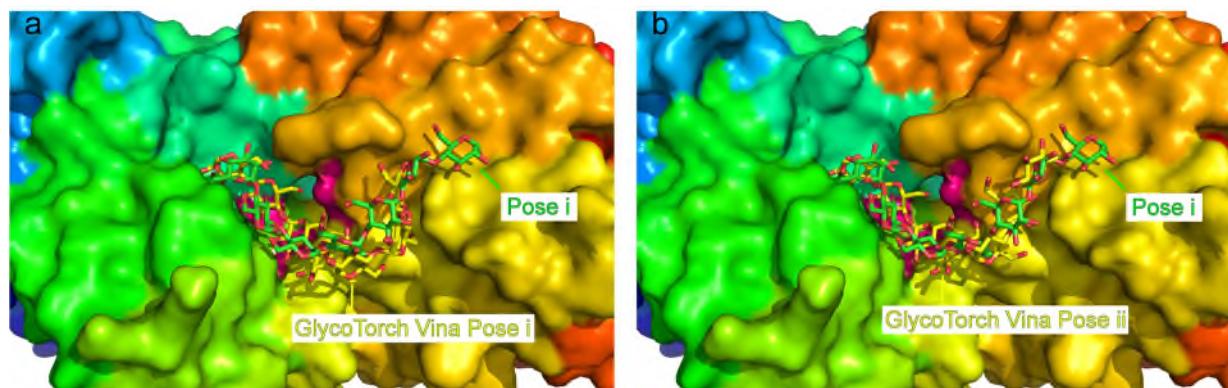


Fig. r3 Alignment between the pose i (green) and the top two poses (yellow) generated by GlycoTorch Vina.

As GlycoTorch Vina is an upgraded version of Vina-Carb, we, therefore, did not further try to use the latter for docking. In addition, please refer to our response to comment 2) of Reviewer 1, in which we reinforce the idea that computational results still need experimental validation. In the present work, our theoretical predictions agree well with the experimental results.

Apart from the important reasons spelled out in the previous revision, the HADDOCK web server has a user-friendly interface, and is constantly updated in real-time. Furthermore, protein engineers and biotechnologists represent, at least in part, the target audience of our work. Referring to the engineering strategy reported in this study, they can easily use the HADDOCK web server without specific programming skills, as well as without the need of local high-performance computers to search for potential mutation sites in the branching enzymes of interest. Ease of use is, therefore, one additional important reason for choosing the HADDOCK web server.

In the revised manuscript, prior to reporting the docking trials, we have included a paragraph on page 10 listing the current software able to achieve glycan docking, citing the relevant literature, and stating the reasons why we selected the HADDOCK web server to conduct our enzyme-substrate docking:

“Docking can generate useful and testable experimental hypotheses, albeit many challenges in enzyme-glycan docking remain to be addressed. The HADDOCK web server^{34,35}, GlycoTorch Vina³⁶, and Vina-Carb³⁷ are the main software currently available to perform glycan docking. We chose the HADDOCK web server based on the following reasons: i) the refinement step using MD simulations greatly contributes to optimizing the binding pose of the substrate on the enzyme, which is not available in other docking software; ii) the clustering information, which reflects the number of polysaccharides lying at a similar location, is an important indicator to evaluate different poses; and iii) the user-friendly interface enables protein engineers and biotechnologists to run docking procedure without the need of programming skills and local high-performance computers.”

2) Reproducibility goes beyond stating that all experiments were repeated three times. It is crucial to provide detailed methodologies so others can replicate the experiments. For instance, the interaction energy calculation method appears unclear. Did the authors use MMPBSA.py, another tool, or did they manually extract residues to calculate the electrostatic and van der Waals energies? If the latter, the corresponding scripts should be provided. Additionally, the feasibility of this calculation method needs validation. Are there other publications employing this approach?

In this work, we manually select target groups to calculate the electrostatic and van der Waals energies. Following the Reviewer’s suggestion, the corresponding scripts have been uploaded to Figshare.

The interaction energy, which is calculated as the sum of the van der Waals energy and the electrostatic energy (or Coulombic energy), has been widely used to analyze the strength of the interaction between two groups (*Nature* 2022, 743; *J. Am. Chem. Soc.*, 2019, 14451; *Nat. Commun.*, 2015, 6234; *Green Chem.*, 2021, 6020; *J. Chem. Inf. Model.*, 2023, 7837; *ACS Cent. Sci.*, 2022, 915, to cite a few). For instance, in the work of Velazhahan et al., the authors computed the sum of van der Waals and Coulombic interaction energies to evaluate the strength of the interaction of residues in a dimer (*Nature* 2022, 743).

3) Regarding the catalytic process, I understand that the residue does not form a covalent interaction with the glycan. However, the glycan itself undergoes reactions, as depicted in the updated Figure 4 (right part). This mechanism is inferred rather than being directly modeled in silico.

Fig. 4b highlights the noncovalent enzyme-substrate hydrogen-bonding interactions. The present work, however, does not involve a covalent catalytic process, which includes the formation and cleavage of covalent

intermediates under the catalysis of acid/base residue E356 and nucleophilic residue D305 (please see the diagrammatic sketch in Fig. 3c and 3f, i.e., the step iii and step vi). We state in the main text that steps (iii) and (vi) correspond to covalent catalysis obeying a well-defined mechanism, while the noncovalent enzyme-substrate interactions investigated in the present work involve the remaining four steps, i.e., (i), (ii), (iv), and (v). We need to clarify here that a discussion of the catalytic process is beyond the scope of this article.

4) The authors mention solving a crystal structure but did not include it in the revised manuscript. This crucial information should be incorporated into the manuscript rather than merely communicated to the reviewers.

Following the Reviewer's suggestion, we have included in the revised manuscript the information on the *RpBE* crystal (please see Supplementary Table 3), and incorporated a paragraph in the 'Structural feature' section of the 'Results'.

"We further conducted crystallographic experiments in the hope of providing experimental evidence for the conformation of the important loop 2. We have obtained the *RpBE* crystal with a 2.9 Å resolution (PDB code 8ZQA, see Supplementary Table 3). Our findings are in line with the structural prediction of AlphaFold v2.0 (see Supplementary Fig. 6a); the conformation of loop 2 in the *RpBE* crystal is similar to that from the AlphaFold modeling of *RpBE* and *CceBE1*, albeit quite different from that in *RoBE* (see Supplementary Fig. 6b). Therefore, despite the high sequence similarity between *RpBE* and *RoBE*, the conformation of loop 2 can prove to be different between the two structures. The pronounced flexibility of loop 3, however, precluded access to its electron density. Given the similarity between modeled and experimental structures, as well as the low resolution of the crystal structure, it is reasonable to perform computational work using the modeled structure."

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
α , β , γ (°)	73.59, 73.77, 65.25
Resolution (Å)	50 - 2.91 (2.94 - 2.91)
R_{merge}^1	0.26 (1.15)
Unique reflections	135389 (4521)
Redundancy	3.5 (3.5)
Completeness (%)	91.63 (89.54)
Mean <i>I</i> / σ <i>I</i>	3.6 (1.1)
CC _{1/2} (%)	93.3 (41.9)
Refinement	
R_{work} (95% data)	0.1971 (0.2818)
R_{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.

Moreover, we have added a paragraph in ‘Crystal structure determination’ section of the ‘Methods’:

“Crystal structure determination. Crystallization experiments were conducted at 25°C using the sitting-drop vapor-diffusion method. 10 mg/mL protein was mixed with reservoir solution in a 1:1 ratio. The crystals were obtained using the “1-28” condition of the Morpheus II screen kit (Molecular Dimensions; 0.1 M buffer System 4, 31% Precipitant Mix 8, and 4 mM Alkalis). Crystals were tested and X-ray diffraction data collected at beamlines BL10U2, BL17B, BL18U1, and BL19U1 of the National Facility for Protein Science in Shanghai (NFPS). The crystals were mounted in a cryoloop and soaked with 20% glycerol before data collection at 100 K. The diffraction images were processed using HKL2000⁴⁵. Crystals were solved by the molecular replacement (MR) method with the Phaser program⁴⁶ from the Phenix⁴⁷ suite using the AlphaFold modeling of *RpBE*. Further refinement was carried out using programs of Phenix Refine⁴⁸ and Coot⁴⁹. Before structural refinements, 5% of randomly selected reflections were set aside for calculating R_{free} ⁵⁰ as a monitor. The data collection, processing, and refinement statistics can be found in Supplementary Table 3.”

5) In Figure 8, does G7 have two possible binding poses, one involving A243 and the other involving all other residues? If mutations result in less than a twofold change, this does not convincingly confirm the binding site. More substantial evidence is required to support this claim.

In the previous version of Fig. 8b, the substrates colored in blue and green were from the *CceBE1*-G7 crystal complex and our docking pose i, respectively. Ban et al. aligned their docking *RoBE*-G7 complex with the *CceBE1*-G7 complex, and deduced a possible conformation of a long-chain substrate binding to *RoBE* through the poses of the two G7 substrates (*Food Res. Int.*, 2022, 112119). To avoid possible confusion caused by displaying the two glycan chains, as shown in Supplementary Fig. 10, we modeled a continuous glycan chain by referring to the substrate conformations in pose i (−2 to +5 glucosyl rings) and *CceBE1*-G7 (−7 to −3 glucosyl rings). As depicted in the revised version of Fig. 8b, seven mutation sites, i.e., A243, K436, G431, G429, P420, L421, and H385, lie around the *CceBE1* G7-pose i, which includes four of the five top variants, i.e., A243T (1.7-fold increase), K436R (1.5-fold increase), G431N (1.5-fold increase), and G429M (1.4-fold increase). We, therefore, conclude that residue replacement around the *CceBE1* G7-pose i is a potent strategy to improve enzyme activity.

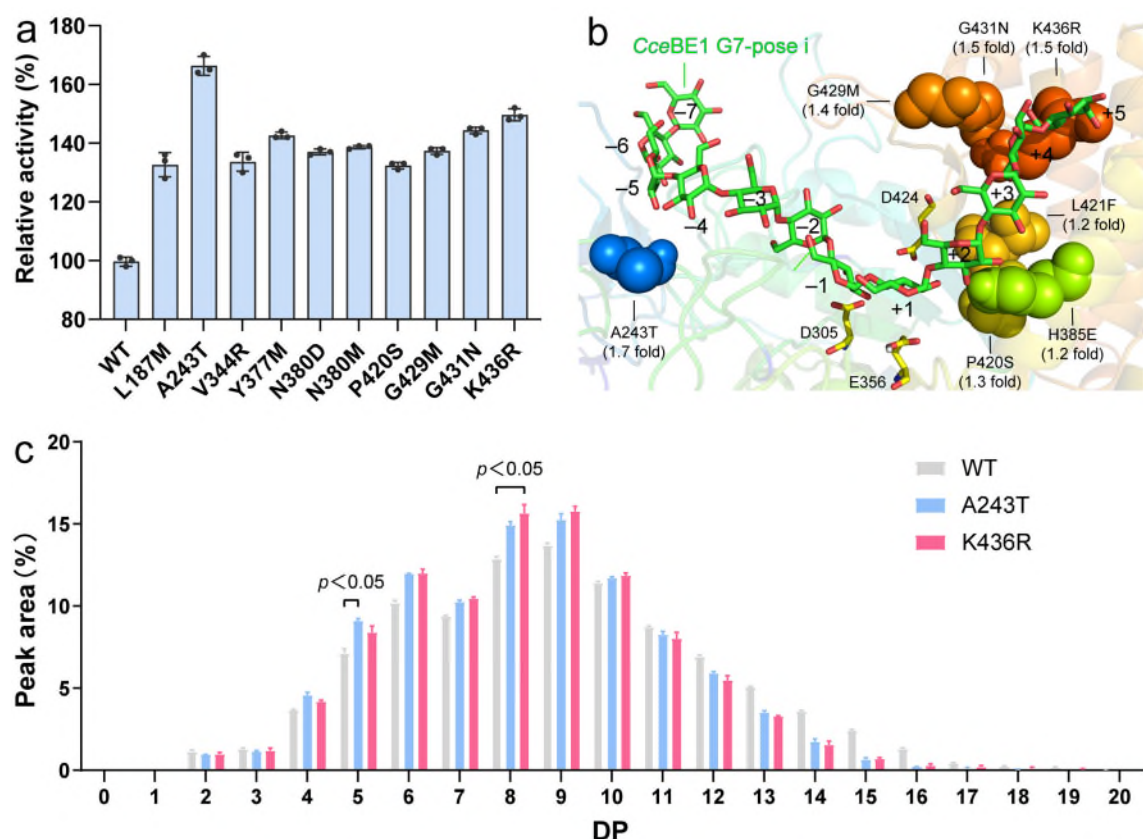


Fig. 8 Enzyme engineering. **a** Top ten variants amid the forty-seven ones. The WT *RpBE* is regarded as possessing 100% enzyme activity. **b** Mutation sites (shown in sphere) around the *CceBE1* G7-pose i. The catalytic triad is shown in a licorice representation, and colored in yellow (carbon atoms). **c** Distribution of the glycan chain length. The data without enzyme catalysis are not shown for clarity. The bars and the error bars are the average and standard deviation of triplicates ($n = 3$ independent experiments), respectively.

As regards the viewpoint of the Reviewer, i.e., “if mutations result in less than a twofold change, this does not convincingly confirm the binding site”, we humbly beg to differ, and would like to express a different opinion. In the field of enzyme engineering, it is difficult to achieve more than a twofold increase in activity through a single-point mutation. *RoBE* is a rather extensively studied branching enzyme. In the work of Ban et al., *RoBE* activity was increased by 5-10%, resulting from two optimal residue replacements at the Q489 site, i.e., Q489E and Q489R (*Food Res. Int.*, 2022, 112119). The G160F and G160R variants possess the highest reported increase in *RoBE* activity so far, with 1.6- and 1.8-fold improvements, respectively (*Food Chem.*, 2022, 132458).

In the present study, seven mutation sites lie around the *CceBE1* G7-pose i, for which enzymatic activity is increased 1.2- to 1.7-fold, resulting from residue replacements at these sites; this level of activity improvement is very reasonable and considered excellent in the field of enzyme engineering. The experimental data can, therefore, support our viewpoint, i.e., “residue replacement around the *CceBE1* G7-pose i is a potent strategy to improve enzyme activity.”

Response to Reviewer #3:

As previously reported, this study investigates the enzymatic activity of the Branching Enzyme (BE) from *Rhodothermus profundus*, comparing it with twelve other BEs. The authors used a structural model obtained with the AlphaFold program to conduct docking calculations and molecular dynamics (MD) simulations, characterizing the substrate binding and transfer steps (SBTs). The well-executed MD simulations reinforce their conclusions.

In this revised version, the authors have enhanced the manuscript with new experimental data, as requested by the reviewer. They conducted circular dichroism (CD) experiments to provide evidence of the folding of the BE mutants and determined the X-ray structure of the RpBE protein. These additions significantly improve the quality of the manuscript, making it suitable for publication in *Nature Communications*.

We are grateful to the Reviewer for their positive comments on our revision.

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have, to the best of their ability, and the ability of current modeling software, answered all of my questions. The effort to model the continuous substrate envisioned by the modeling is a definite plus. The addition of a new crystal structure is also quite enlightening and clarifies to best extent possible, the issue of the flexible loops. I have no other issues, the paper is recommended for publication.

Reviewer #2 (Remarks to the Author):

The authors have solved my issues and this manuscript has been suitable for publishing in Nature Communications. I'm grateful for their hard work in dealing with such comments.