



# Engineering and screening of novel $\beta$ -1,3-xylanases with desired hydrolysate type by optimized ancestor sequence reconstruction and data mining



Bo Zeng<sup>a</sup>, ShuYan Zhao<sup>a</sup>, Rui Zhou<sup>a</sup>, YanHong Zhou<sup>a</sup>, WenHui Jin<sup>b</sup>, ZhiWei Yi<sup>b</sup>, GuangYa Zhang<sup>a,\*</sup>

<sup>a</sup> Department of Bioengineering and Biotechnology, Huaqiao University, Xiamen, Fujian Province 361021, China

<sup>b</sup> Technology Innovation Center for Exploitation of Marine Biological Resources, Third Institute of Oceanography, Ministry of Natural Resources, Xiamen, Fujian Province 361005, China

## ARTICLE INFO

### Article history:

Received 21 April 2022

Received in revised form 22 June 2022

Accepted 22 June 2022

Available online 27 June 2022

### Keywords:

Hydrolase engineering and screening

Optimized ancestor protein reconstruction

Data mining

$\beta$ -1,3-xylo-oligosaccharides production

## ABSTRACT

Engineering of hydrolases to shift their hydrolysate types has not been attempted so far, though computer-assisted enzyme design has been successful. A novel integrative strategy for engineering and screening the  $\beta$ -1,3-xylanase with desired hydrolysate types was proposed, with the purpose to solve problems that the separation and preparation of  $\beta$ -1,3-xylo-oligosaccharides was in high cost yet in low yield as monosaccharides existed in the hydrolysates. By classifying the hydrolysate types and coding them into numerical values, two robust mathematical models with five selected attributes from molecular docking were established based on LogitBoost and partial least squares regression with overall accuracy of 83.3% and 100%, respectively. Then, they were adopted for efficient screening the potential mutagenesis library of  $\beta$ -1,3-xylanases that only product oligosaccharides. The virtually designed AncXyl10 was selected and experimentally verified to produce only  $\beta$ -1,3-xylobiose (60.38%) and  $\beta$ -1,3-xylotriose (39.62%), which facilitated the preparation of oligosaccharides with high purity. The underlying mechanism of AncXyl10 may associated with the gap processing and ancestral amino acid substitution in the process of ancestral sequence reconstruction. Since many carbohydrate-active enzymes have highly conserved active sites, the strategy and their biomolecular basis will shield a new light for engineering carbohydrates hydrolase to produce specific oligosaccharides.

© 2022 Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

As a new semi-rational design method, ancestral sequence reconstruction (ASR) has many successful examples of protein engineering. The resurrected ancestor proteins often exhibit “unusual” or “extreme” properties to some extent properties [1]. Recent experimental and computational works have specifically discussed the altered patterns of interaction with other subcellular components [2], enhanced stability [3], conformational flexibility/diversity [4] and catalytic promiscuity [5]. However, altered types of hydrolysates of hydrolases have not been reported among the remarkable properties of ancestral protein reconstructions up to now. The hydrolysates shifting were very different from the reported catalytic promiscuity and enantio-selectivity, which was rarely studied yet. In our previous work, six ancestral protein sequences of  $\beta$ -1,3-xylanases were reconstructed by optimizing

the ancestral sequence reconstruction strategy [6]. The ancestral protein AncXyl09 with unique properties was characterized and we keenly observed that there is no xylose in the hydrolysate of AncXyl09. So, we desire to find out whether there is no xylose in the hydrolysates of the remaining five ancestral  $\beta$ -1,3-xylanases. If this work, the ancestral protein resurrection may become an effective strategy to engineer of hydrolases to shift their hydrolysate types and provided an efficient mutation library. To the best of our knowledge, there is no related reports presently. Thus,  $\beta$ -1,3-xylanases, as a research model, were selected to propose a novel strategy for the engineering and screening the hydrolysate types of glycoside hydrolases.

As for computational aided enzyme engineering, there have been many successful examples based on the sequences or structures of the target enzymes. Computational techniques can be used to engineer enzymatic reactivity, substrate specificity and ligand binding, access pathways and ligand transport, and global properties like protein stability, solubility, and flexibility [7]. However, the engineering for shifting the hydrolysate types based on current

\* Corresponding author.

E-mail address: [zhgyghh@hqu.edu.cn](mailto:zhgyghh@hqu.edu.cn) (G. Zhang).

strategies is far from successful. Thus, other effective tools were urgently needed to deal with this problem. As we know, data mining is a powerful tool in dealing with many biological complex problems, such as discriminate thermophilic and mesophilic of proteins from their primary structure (sequence) information[8]; discriminating acidic and alkaline enzymes using a random forest model with secondary structure amino acid composition[9].

$\beta$ -1,3-xylan was found in the cell walls of some red and green algae[10].  $\beta$ -1,3-xylanase (EC3.2.1.32) can hydrolyze  $\beta$ -1,3-xylan to produce  $\beta$ -1,3-xylo-oligosaccharides with different xylose units. The main hydrolysates of  $\beta$ -1,3-xylanase were xylose,  $\beta$ -1,3-xylobiose and  $\beta$ -1,3-xylotriose after collecting and analyzing the experimental results[6,11–14]. Of particular interest was the observation that it can be classified into two types about the main hydrolysates: xylose and oligosaccharides and oligosaccharides (or almost without xylose). Investigating the underlying mechanism of such phenomenon is of great interests for  $\beta$ -1,3-xylanases designing or engineering. Besides, it is also meaningful in the application of the  $\beta$ -1,3-xylanase hydrolysates. As we knew, xylose has been reported to be a raw material for conversion of produce value-added products, such as xylitol, 2,3-butanediol[15]. While the  $\beta$ -1,3-xylo-oligosaccharides ( $\beta$ -1,3-xylobiose and  $\beta$ -1,3-xylotriose) have been reported variety of biological activities, such as anticoagulant activity, antioxidant activity[16], and antitumor activity [17]. However, there are a lot of monosaccharides in the hydrolysis products of existing enzyme. For example, the proportion of xylose in the hydrolyzates of the extremely thermophilic  $\beta$ -1,3-xylanase (ID: B9K760) and the most efficient  $\beta$ -1,3-xylanase (FlaGM003088, ID: MK253053.1) were 32.17% and 20.89%, respectively. The preparation of pure  $\beta$ -1,3-xylo-oligosaccharides is complex and costly, as the separation of  $\beta$ -1,3-xylo-oligosaccharides from xylose requires preparative-size size-exclusion chromatography. So, it is advantageous to the preparation of  $\beta$ -1,3-xylo-oligosaccharides if the hydrolysates of  $\beta$ -1,3-xylanase contains no xylose for application purposes. However, no relevant studies have been reported so far. Thus, we proposed a novel strategy for engineering and screening the hydrolase with desired hydrolysates (Scheme 1). The first challenge was classifying the hydrolysate types and coding them into numerical values to facilitate the establishment of mathemat-

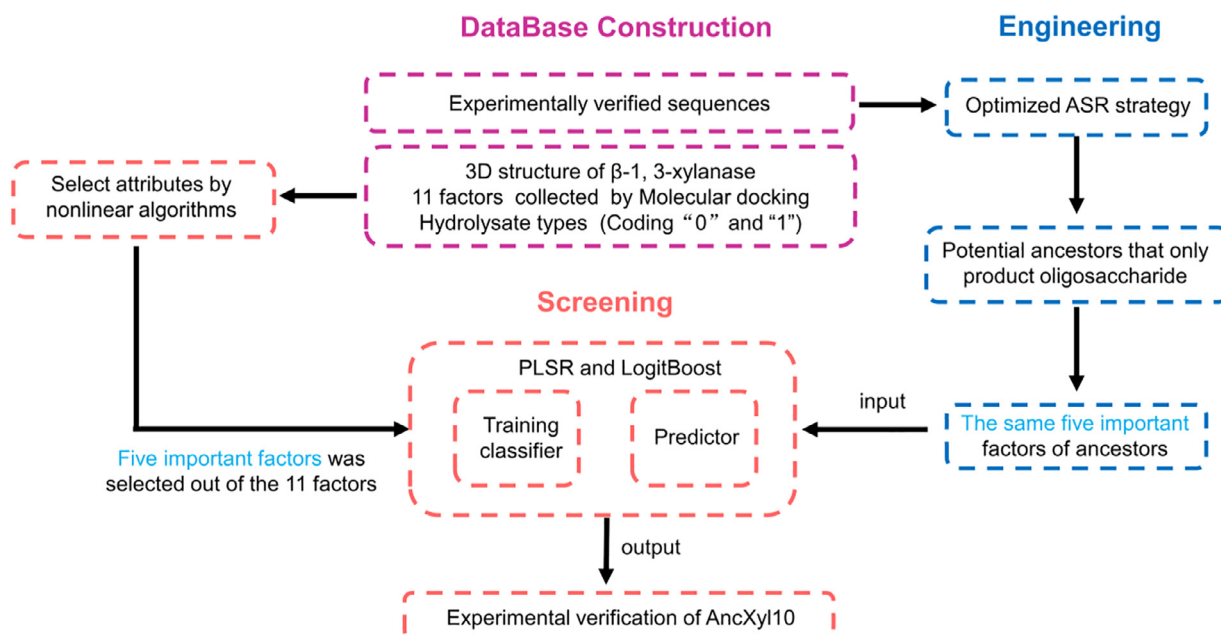
ical models for efficient screening the potential mutagenesis library of  $\beta$ -1,3-xylanases that only product oligosaccharides.

Herein, a novel  $\beta$ -1,3-xylanase (AncXyl10) that only produced oligosaccharides was designed by the strategy. The hydrolysates of AncXyl10 were  $\beta$ -1,3-xylobiose (60.38%) and  $\beta$ -1,3-xylotriose (39.62%), which facilitated the preparation of oligosaccharides with high purity. As the prediction accuracy of the LogitBoost for oligosaccharides was as high as 100%, and the Partial Least Squares Regression (PLSR) for the two types was 100%. The classifier with high prediction accuracy could avoid blindness and reduce experimental workload as much as possible. The obtained ancestral  $\beta$ -1,3-xylanase AncXyl10 is the first successful example of ASR to engineer the hydrolysate type, which broadens the application scope of ASR. The underlying mechanism for AncXyl10 to produce only  $\beta$ -1,3-xylobiose and  $\beta$ -1,3-xylotriose may shield lights for the development other tools for engineering the hydrolysate types of hydrolytic enzymes.

## 2. Materials and method

### 2.1. Dataset collection.

Databases UniProt, NCBI, CAZy and literature were searched using  $\beta$ -1,3-xylanase as keywords. A total of 6  $\beta$ -1,3-xylanases with experimental validated hydrolysate types were obtained. Their NCBI accession numbers were WP\_015919112.1 (TnB9K760), QDC28441.1 (FlaGM003088-T), AWH57212.1 (FlaGM004512), WP\_052432232.1 (FlaGM003092), MW915416 (AncXyl09) and D5MP61.1 (D5MP61). Among them,  $\beta$ -1,3-xylanase D5MP61 had crystal structures (PDB ID: 2ddx). Thin layer chromatography of  $\beta$ -1,3-xylanases was digitally analyzed through ImageJ software. Combining with the HPLC data of the hydrolysates, we draw the proportion of the hydrolysates as shown in [supplementary data](#). We found that the main hydrolysates were a combination of xylose,  $\beta$ -1,3-xylobiose and  $\beta$ -1,3-xylotriose (each contains >20%) and the major differences between them were the presence (defined as 0) or absence (or content <6%, defined as 1) of xylose in the hydrolysates. Thus, 0 and 1 were chosen as indicators for later classification with machine learning algorithms.



**Scheme 1.** Schematic flowchart indicating the engineering and screening the hydrolysate types of  $\beta$ -1,3-xylanase.

## 2.2. Molecular docking

To prepare 3D structure, the tertiary structures of  $\beta$ -1,3-xylanases (TnB9K760, FlaGM003088-T, FlaGM004512, FlaGM003092, AncXyl09) was predicted by the Robetta[18], Swiss-model[19] and I-TASSER servers[20] and then evaluated by PROCHECK, VERIFY 3D[21], ERRAT[22] and MolProbity servers [23]. The structures predicted by Robetta were chosen as they got the highest evaluation scores. To prepare ligands, we draw the 3D structure of  $\beta$ -1,3-xylobiose,  $\beta$ -1,3-xylotriose,  $\beta$ -1,3-xylotetraose and  $\beta$ -1,3-xylopentaose and performed CHARMM [24] to minimize ligands. The molecularly docking factors of  $\beta$ -1,3-xylotetraose and  $\beta$ -1,3-xylopentaose was neglected, because it cannot dock with most xylanases  $\beta$ -1,3-xylanase. Finally, we collected 11 factors generated in the molecular docking performed by CDOCKER implemented in Discovery Studio 2019. The first factor is the distance of the catalytic group (two glutamates) OE2 atom and marked as X1. The protein receptor radius and volume were derived binding sites from cavities in the structure of the receptor marked as X2 and X3. Others important parameters were collected in the process of molecular docking (X4-X5, binding energy[25] with  $\beta$ -1,3-xylobiose and  $\beta$ -1,3-xylotriose; X6-X7, -CDOCKER\_ENERGY with  $\beta$ -1,3-xylobiose and  $\beta$ -1,3-xylotriose; X8-X9, -CDOCKER\_INTERACTION\_ENERGY[26] with  $\beta$ -1,3-xylobiose and  $\beta$ -1,3-xylotriose; X10-X11, receptor surface area [27] of protein).

## 2.3. Determination of the factors influencing hydrolysate types of $\beta$ -1,3-xylanase

A total of 11 factors (attributes) were selected according to the molecular docking tools in DS 2019. Nonlinear algorithms were used for selecting the major factors (out of the 11 attributes) that affected the hydrolysate types of  $\beta$ -1,3-xylanase due to excellent performance. The five nonlinear algorithms (CfsSubsetEval(CSE), CorrelationAttributeEval(CAE), GainRatioAttributeEval (GRAE), ReliefAttributeEval (RAE), SymmetricalUncertAttributeEval (SUAE) were performed in WeKa software (3.9.0). Finally, we draw a Venn diagram to select major factors for the next prediction.

## 2.4. Prescreening samples for predicting the hydrolysate types of ancestral $\beta$ -1,3-xylanase

The optimized ancestral sequence reconstruction process has been explained in our previous work[6]. A total of six ancestral proteins have been reconstructed and the oldest ancestral protein (AncXyl09) with unique properties has been characterized. The hydrolysate of AncXyl09 was verified to be free of xylose by thin layer chromatography and liquid chromatography in our previous work. In order to prescreening the ancestral  $\beta$ -1,3-xylanases hydrolysate, The remaining five ancestral proteins (AncXyl10, AncXyl11, AncXyl12, AncXyl13, AncXyl14) were used as prescreening samples for predicting the hydrolysate types of  $\beta$ -1,3-xylanase. All ancestor  $\beta$ -1,3-xylanases performed domain prediction by Pfam [28] (<http://pfam.xfam.org/>). They belong to the GH26 family and have the conserved pattern ([RPVYLR]-xx-yE-x-[DE]-[nKP]-x-[fi]-x-E-xx-[Pqry], red E is an active site residue, x is any of the 20 natural residues). The GH26 catalytic domain exhibits the classical TIM ( $\beta/\alpha$ )<sub>8</sub>-barrel in a clan-GH-A member. The active cavity of  $\beta$ -1,3-xylanases is in the TIM barrel and the two Glu residues act as the catalytic acid/base and nucleophile in a double-displacement mechanism[12,13]. Based on these, we excluded ancestral  $\beta$ -1,3-xylanases (AncXyl12, AncXyl13, AncXyl14) whose receptor cavities were not within the range of TIM during molecular docking.

## 2.5. Establishment of a non-linear and linear model for predicting the hydrolysate types of ancestral $\beta$ -1,3-xylanases

Five important factors were selected as the training samples through the attribute selection and an assembled classifier named LogitBoost[29] was adopted. The performance and robustness of the model was evaluated by three different validation check approaches. Firstly, Back-check prediction (or self-consistency test) method was performed to train the model. We have used the five factors of six  $\beta$ -1,3-xylanases to predict these same proteins whether each protein hydrolysate was xylose and oligosaccharides or oligosaccharides. Secondly, leave-one-out cross-validation was carried out. Finally, the reliability of the method was evaluated with an independent testing datasets where no information was used in training data.

As the non-linear algorithms could not tell us the exact relationship between the five important factors and the hydrolysate types, a linear predictor named partial least squares regression (PLSR)[30] was established with the threshold of 0.5 to discriminate the two hydrolysate types (0 and 1). Then, the ancestor sequence samples were screened and predicted by LogitBoost and PLSR. The type of hydrolysate of AncXyl10 was predicted to be only oligosaccharide by two algorithms, while the type of hydrolysate of AncXyl11 was predicted to be xylose and oligosaccharide. As the hydrolysates only contained oligosaccharide were more valuable, we decided to experimentally verify the AncXyl10.

## 2.6. Cloning and expression of the AncXyl10

The coding gene of the AncXyl10 was optimized for *Escherichia coli* (Accession: BankIt2538175 beta-1\_3-xylanase OM287162), and cloned into the pET-22b(+) vector via the *Nde*I and *Hind*III sites by GeneScript (Nanjing, China) with a 6xHis-tag at the N-terminus. The plasmid was co-transformed into *E. coli* BL21 (DE3) which cultured in 200 mL TB broth at 37 °C for 4 h. Expression was carried out in TB supplemented with 0.1 mmol/L IPTG at 20 °C overnight.

## 2.7. Chromatographic identification of the hydrolyzed products

The hydrolysates of AncXyl10 were detected by High performance liquid chromatography (HPLC) as described previously [6]. Using  $\beta$ -1,4-xylobiose as the standard, we detected the hydrolysates of AncXyl10 after incubated at 0 h, 8 h, and 24 h, respectively. In addition, the hydrolysates of AncXyl09 were also detected by HPLC under the same conditions after incubated at 24 h, which serves as reference as it has the results of HPLC and thin layer chromatography in our previous work[6]. The results were analyzed using the Empower chromatographic workstation.

## 2.8. Enzyme activity determination

The  $\beta$ -1,3-xylanase (AncXyl10) activity was measured by the Somogyi-Nelson method[31], which determining the amount of reduced sugar released from  $\beta$ -1,3-xylan. One unit of enzyme activity was defined as the amount of enzyme that liberated 1  $\mu$ mol of D-xylose per min under the below conditions. The reaction system (400  $\mu$ L) was incubated at 55 °C for 5 mins, containing 1%  $\beta$ -1,3-xylan (300  $\mu$ L), an appropriate amount of  $\beta$ -1,3-xylanases (100  $\mu$ L).

## 2.9. Optimization of the preparation conditions for $\beta$ -1,3-xylo-oligosaccharides

To investigate the preparation conditions of  $\beta$ -1,3-xylo-oligosaccharides, we explored the optimal temperature and optimal pH of AncXyl10 according to the enzyme activity determina-

tion method. The optimal temperature of AncXyl0 was measured in the range 40–70 °C at intervals of 10 °C and the optimal pH was measured in the pH range 4.0–8.0. We also determined the effects of metal ions on the AncXyl10 enzymatic production of  $\beta$ -1,3-xylo-oligosaccharides at a final concentration of 10 mmol/L (Na<sup>+</sup>, K<sup>+</sup>, Ca<sup>2+</sup>, Mg<sup>2+</sup>, Cu<sup>2+</sup>, Zn<sup>2+</sup> and Ba<sup>2+</sup>).

### 2.10. Bioinformatical analysis of the underlying mechanism on the shift of $\beta$ -1,3-Xylanases hydrolysate types

There is no ready-made reference method to study the potential mechanism on the shift of the hydrolysate types. However, we still found the idea from the algorithm of ancestor sequence reconstruction, which inferred ancestors was often quite different from existing sequences (<30%) due to sequence gap handling and ancient amino-acid replacements [32–34]. Firstly, AncXyl10 and TnB9K760 were selected as the pairwise  $\beta$ -1,3-xylanases as they have the highest sequence identity (79.5%) while their hydrolysate types were different. Secondly, their sequence gap handling and ancient amino-acid replacements were analyzed based on the sequence and structure through multiple sequence alignment and protein superimposed performed by DS2019. Finally, combining the relationship between the five important factors and ASR processes (sequence gap handling and ancient amino-acid replacements) as the main starting point to comprehensively explore the molecular mechanism that altering the hydrolysate type of AncXyl0.

## 3. Results and discussion

### 3.1. Selection of the major factors influencing the hydrolysate types of $\beta$ -1,3-xylanase.

After characterizing four  $\beta$ -1,3-xylanases by our laboratory, we found that some  $\beta$ -1,3-xylanases hydrolysates contained more xylose (content >20%) while others with few xylose or almost no xylose [6,11–14] (Supplementary Fig. S1). Fishbone diagram method can help us analyze the factors of the interesting scientific question (Supplementary Fig. S2A). There are six possible causes: the distance of active site, receptor cavity, receptor surface area, free energy of binding with substrate, -CDOCKER\_ENERGY with substrate, and -CDOCKER\_INTERACTION\_ENERGY with substrate. In the above parameters, the factor of -CDOCKER\_ENERGY was used like a score, where a higher value indicates more favorable binding. It includes the internal ligand strain energy and the receptor-ligand interaction energy. At the same time, the receptor-ligand interaction energy was assigned as -CDOCKER\_INTERACTION\_ENERGY. Therefore, we collected the data based on molecular docking as shown in the material method (Table S1).

Weka has been used for automated protein annotation [35,36], probe selection for gene-expression arrays [37] and automatic cancer diagnosis [38], which can assist users in extracting useful information from data and enable them to easily identify a suitable algorithm for generating an accurate predictive model [39]. Thus, after collecting all the potential factors, we adopted five nonlinear data mining algorithms to select the major factors influencing the hydrolysate types of  $\beta$ -1,3-xylanase by WeKa (3.9.0). The top six factors selected by the five algorithms were listed: CSE (X1, X2, X3, X8, X9, X10); CAE (X1, X2, X8, X7, X6, X10); GRAE (X1, X2, X3, X5, X4, X11); RFAE (X1, X2, X8, X10, X6, X3); SUAE (X1, X3, X2, X5, X4, X11). The Venn diagram for the factors selected by the five algorithms was generated using an online tool (<https://hiplot.com.cn/basic/venn>). Venn diagram indicated there were two factors (X1, X2) selected by all of the five algorithms, one factor

(X3) selected by four of the algorithms, two factors (X8, X10) selected by three of the algorithms, four factors (X4, X5, X6, X11) selected by two of the algorithms, and two factor (X7, X9) selected by one of the algorithm (Supplementary Fig. S2B). Therefore, these top five major factors (X1, X2, X3, X8, X10) were selected for the later classification or regression analysis.

### 3.2. The hydrolysate type predictor based on LogitBoost and PLSR

We selected the dataset containing of the five factors (X1, X2, X3, X8, X10) and adopted an assemble classifier named LogitBoost to predict the hydrolysate types of ancestral  $\beta$ -1,3-xylanases. To evaluate the method, three validation check methods were used. Firstly, we conducted a consistency check using the training set as a test set for testing, it achieved the 100% overall accuracy. The overall correct rate shows that the LogitBoost algorithm has mastered the complicated relationship between the hydrolysate types and the various factors. LogitBoost algorithm could be considered a useful predictor while poor self-consistency certainly cannot useful predictor. Secondly, to truly reflect the power of the predictor, leave-one-out cross-validation (also named 6-fold cross-validation here) was carried out. The accuracy rates for “xylose and oligosaccharides” is 66.66% while for “oligosaccharides” is 100%, and the overall accuracy rate is 83.33%. This means that the algorithm has a higher correct rate (100%) of recognition of the type “oligosaccharides” after learning the complex relationships between the factor and the hydrolysate types which were consistent with our purpose. Finally, the reliability of the method was evaluated with an independent testing datasets where no information was used in training data. AncXyl10 and AncXyl11 were collected the same five factors (X1, X2, X3, X8, X10) as screening samples. The predictor predicted that the hydrolysate types of AncXyl10 was the type “oligosaccharides” and the AncXyl11 was the type “xylose and oligosaccharides”.

As the LogitBoost algorithms could not tell us the exact relationship between the five important factors and the hydrolysates types, PLSR, as a linear regression method, has the function of predicting and finding the basic relationship between the five important factors and the type of hydrolysate. For example, Burnett et al. adopted PLSR models to predict the leaf traits from spectral data [40]. It can be clearly seen from the Ti and Ui diagrams that the model can separate the types of hydrolysates “0” and “1” (Fig. 1A). This implies that the model is effective. Model effects and dependent variable weights of the five factors were calculated by PLSR model, where X1 is the most important factor and followed by X2, X8, X10, X3. X1 is the most important factor among the five factors, which was consistent with the results obtained by nonlinear models, as X1 was the only factor that was selected by all the five nonlinear methods (Fig. 1B).

We also conducted a consistency check using the training set as a test set for testing, it achieved the 100% overall accuracy. The high accuracy shows that the PLSR predictor has mastered the complicated relationship between the hydrolysate types and the various factors. After calculating the regression equation of the linear regression model, we defined the threshold as 0.5 to discriminate the hydrolysate types. This means the hydrolysate type was oligosaccharides if the calculated value >0.5, while the hydrolysate types was xylose and oligosaccharides if the value <0.5. Thus, the accuracy of the training set was 100% with two hydrolysates types (Table 1). We tested the ancestral protein with the regression equation of the PLSR. The result predicted that the hydrolysate types of AncXyl10 was oligosaccharides, and the hydrolysate types of AncXyl11 was xylose and oligosaccharides, which was consistent with the results of the LogitBoost classifier. The performances of the predictor based on LogitBoost and PLSR were both excellent

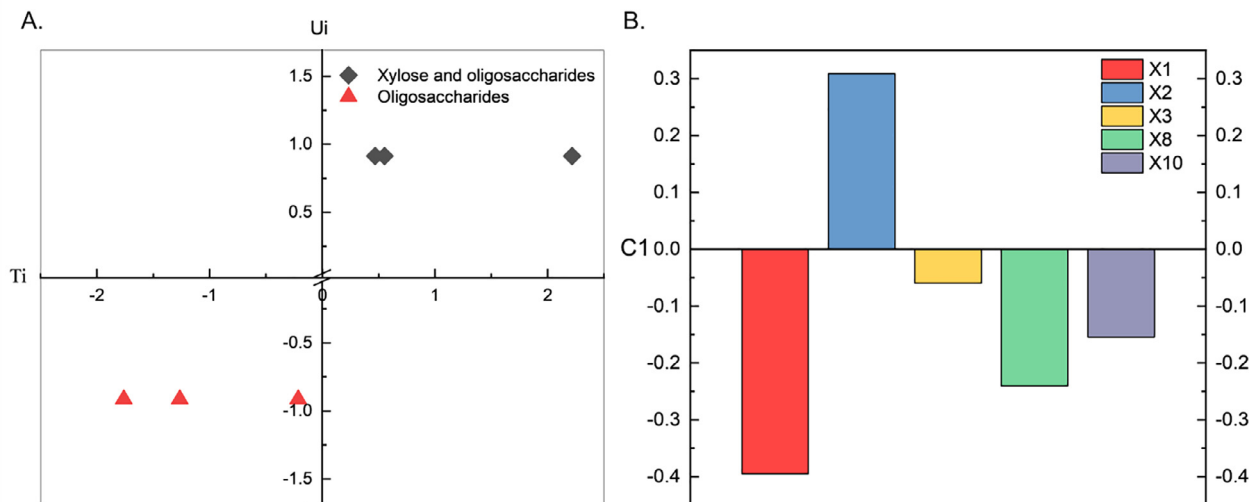


Fig. 1. The PLSR model based on the 5 important factors. A. The two dimensional map of Ti and Ui of PLSR. B. The dependent variable weights of PLSR.

Table 1  
The calculated result of PLSR.

| Sample  | Function           | Y   | Calculated value |
|---|--------------------|-----|------------------|
| TnB9K760  | Training set       | 0   | 0.37             |
| FlaGM003088-T   | Training set       | 0   | 0.40             |
| FlaGM000512   | Training set       | 0   | -0.16            |
| FlaGM003092   | Training set       | 1.0 | 0.94             |
| AncXyl09  | Training set       | 1.0 | 0.61             |
| VsD5MP61  | Training set       | 1.0 | 1.10             |
| <b>AncXyl10</b>   | <b>Testing set</b> |     | <b>0.97</b>      |
| <b>AncXyl11</b>   | <b>Testing set</b> |     | <b>0.10</b>      |
| Regression equation $Y = 4.16 - 0.44x_1 + 0.108x_2 - 0.000012x_3 - 0.035x_4 - 0.0035x_5$ ( $R^2 = 0.69$ ) |                    |     |                  |

with an accuracy of 100% in predicting “oligosaccharides” type (Fig. 2).

As we all know, functional oligosaccharides have a variety of physiological activities and broad application prospects in the field of medical and food [41,42]. However, the complicated and costly separation and preparation process limit their development. Our original intention of establishing the predictor was to mine the potential  $\beta$ -1,3-xylanases that only produces functional oligosaccharides and explore the potential mechanism that ASR shift the hydrolysate types of  $\beta$ -1,3-xylanases. Therefore, we experimentally characterized the hydrolysate type of AncXyl10 to further verify the reliability of the predictor based on LogitBoost and PLSR.

### 3.3. HPLC identification of the hydrolysate types of AncXyl10

AncXyl10 was successfully constructed and expressed. It had a molecular mass of approximately 33.1 kDa as estimated by SDS-PAGE (Supplementary Fig. S3A and S3B). The hydrolysate of AncXyl10 was detected by high performance liquid chromatography. We detected the hydrolysis products at 0 h, 8 h, and 24 h (Fig. 3A). May be due to product inhibition, the hydrolysis products did not increase significantly over time. The baseline is smooth and has no peaks. Taking the hydrolysates at 24 h as an example, two peaks were detected by HPLC and the peak times were 4.597 min and 5.165 min, respectively. We detected the hydrolysates of AncXyl09 and AncXyl10 by HPLC and the peak times of  $\beta$ -1,3-xylobiose and  $\beta$ -1,3-xylotriose were the same. Meanwhile,  $\beta$ -1,4-xylobiose was used as a standard, the concentration was 2 mg/mL and the peak time was 4.901 min, further confirming the hydrolysates were  $\beta$ -1,3-xylobiose and  $\beta$ -1,3-xylotriose (Fig. 3B). Therefore, the pattern of hydrolysates of AncXyl10 was  $\beta$ -1,3-xylobiose and  $\beta$ -1,3-xylotriose and proportions of them were 60.38% and 39.62% after 24 h of hydrolysis, respectively. Separation peaks with larger spacing are beneficial to the preparation of  $\beta$ -1,3-xylo-oligosaccharides. The hydrolysate type of AncXyl10 belongs to the type “oligosaccharides” defined by us. This is consistent with the predicted results, which further illustrates the reliability of the classifier based on LogitBoost and PLSR. AncXyl10 is an effective  $\beta$ -1,3-xylanase for enzymatic preparation of  $\beta$ -1,3-xylo-oligosaccharide to obtain pure oligosaccharides without further

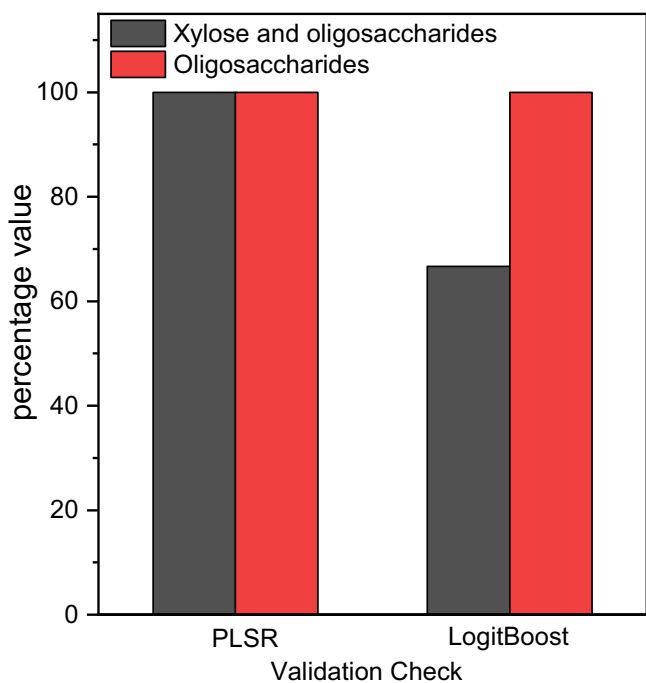
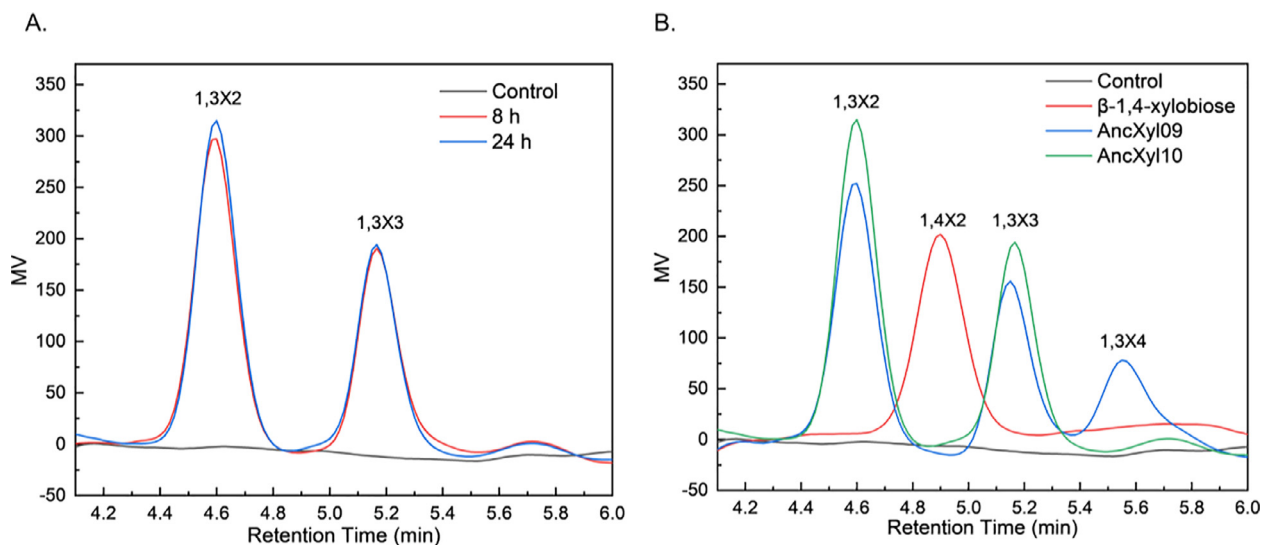


Fig. 2. Performances of PLSR and LogitBoost in predict the hydrolysate types.



**Fig. 3.** Hydrolysates identification of AncXyl10 by HPLC. A. The hydrolysates of AncXyl10 incubated for 0 h, 8 h, 24 h, respectively. B. HPLC of  $\beta$ -1,4-xylobiose as a standard and the hydrolysates of AncXyl09 and AncXyl10 incubated for 24 h.

separation or a one-step separation to obtain  $\beta$ -1,3-xylobiose and  $\beta$ -1,3-xylotriose with high purity.

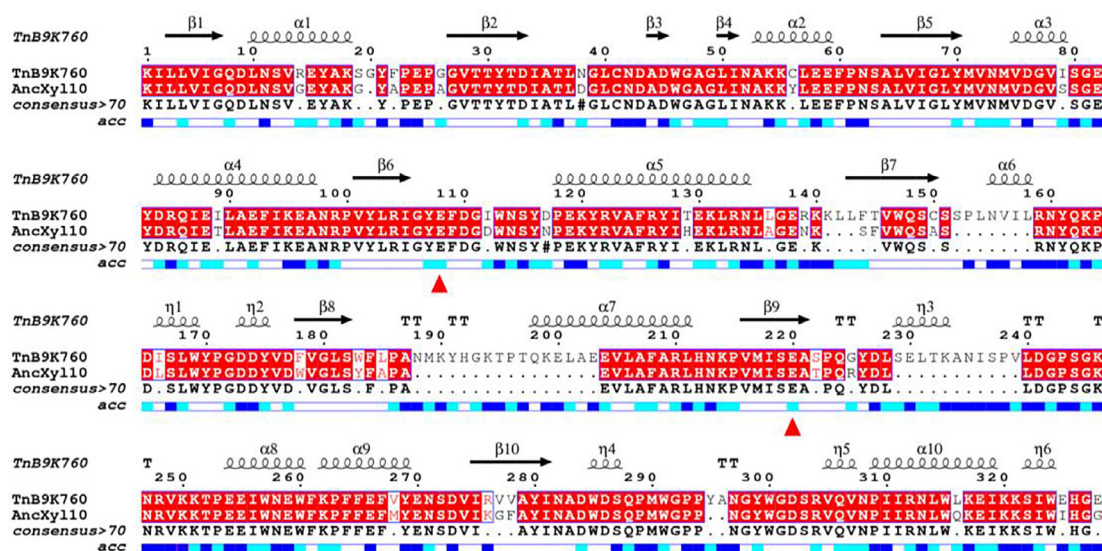
Furthermore, we also optimized the conditions for preparing  $\beta$ -1,3-xylo-oligosaccharides by AncXyl10. The optimal pH and temperature of AncXyl10 was 6.0 and 60 °C, respectively (Supplementary Fig. S4A and B). The activity was almost not affected by  $K^+$ , and 50–90% inhibition was observed with  $Na^+$ ,  $Zn^{2+}$ ,  $Mg^{2+}$ ,  $Ca^{2+}$ ,  $Ba^{2+}$  at 10 mmol/L. On the other hand,  $Cu^{2+}$  improved the activity ranging from 130% to 150% at the same concentration (Supplementary Fig. S4C).

### 3.4. The molecular mechanism of ASR altered the hydrolysate types of AncXyl10

There are many types of oligosaccharides currently studied, such as pectin oligosaccharides[43], Some typical feruloylated oligosaccharides[44], Chitosan oligosaccharides[45],  $\beta$ -1,3-xylo-oligosaccharides production et al. Current methods for obtaining synthetic oligosaccharides were chemical synthesis or chemo-enzymatic synthesis[42]. The enzymes for enzymatic preparation of oligosaccharides were screening from natural microorganisms and no report on the engineering of enzymes for preparing oligosaccharides.

To explore the molecular mechanism of the shift of the hydrolysate types of the ancestor AncXyl10 (without xylose), we selected TnBK760 (with xylose, 20.89%), the original existing enzyme with the closest evolutionary distance with AncXyl10 as the reference object after pair sequence alignment with other  $\beta$ -1,3-xylanases performed by online server. The pair sequence alignment showed that identity, similarity and gap of AncXyl10 and TnB9K760 were 79.5%, 82.0%, 12%, respectively. There are more gaps and amino acid substitutions between AncXyl10 and TnB9K760 (Fig. 4). Ancestral sequence reconstruction (ASR) has emerged as the leading technique to determine sequences of ancient proteins and identify ancient amino-acid replacements that led to functional changes across evolutionary lineages[32]. The inferred ancestors are often quite different from existing sequences (<30%) due to sequence gap handling and ancient amino-acid replacements. Therefore, we will combine the five important factors, the experimental results and ASR processes (sequence gap handling and ancient amino-acid replacements) to explore the molecular mechanism that altering the hydrolysate type of AncXyl10.

Firstly, we will discuss the sequence gap. AncXyl10 and TnB9K760 was superimposed by DS2019 and their RMSD was 3.198. Structural change may be caused by Sequence gap and amino acids replacements. Therefore, we display all the gaps on the superimposed structure and outline the radius of receptors sites sphere (X2, Receptors radius) at the same time. There are a total of 6 gap fragments distributed in the sequence. The radius of receptors site sphere of TnB9K760 and AncXyl10 was 12.2 and 10.7 Å, respectively (Supplementary Fig. S5). The reason for the decrease in the radius of receptors site sphere was due to the negative affected of gapping the fragment YA (295–296) and fragment NMKYHGKTPTQKELAE (188–203). At the same time, this may also have negative impacts on the factors X3 and X10. From the data of the receptor surface volume (X3) and surface area (X10) obtained by docking TnB9K760 and AncXyl10 with  $\beta$ -1,3-xylobiose, the receptor volume and the receptor surface area have been reduced from 42259.1 Å<sup>3</sup> to 38552.1 Å<sup>3</sup>, 297.09 Å<sup>2</sup> to 270.24 Å<sup>2</sup>, respectively. So far, the  $\beta$ -1,3-xylanase were reported to be *endo*-type  $\beta$ -1,3-xylanase. The production of xylose is likely to be the secondary hydrolysis of the already produced  $\beta$ -1,3-xylo-oligosaccharides. Thus, smaller receptors radius (X2), the receptor surface volume (X3) and surface area (X10) may blocked  $\beta$ -1,3-xylo-oligosaccharides (multiple xylose units) entering the active cavity. For example,  $\beta$ -1,3-xylo-tetraose can be successfully docked with TnB9K760 but cannot be successfully docked with AncXyl10. And from the docking data of TnB9K760 and AncXyl09 with  $\beta$ -1,3-xylo-triose, its -CDOCKER\_ENERGY changed from 6.01 to -3.27. The docking score reported as the positive value (-CDOCKER\_ENERGY), where a higher value indicates a more favorable binding. This means that  $\beta$ -1,3-xylo-triose are not conducive to binding in the active cavity of AncXyl10, while TnB9K760 is favorable binding with  $\beta$ -1,3-xylo-triose. So,  $\beta$ -1,3-xylo-triose continue to be hydrolyzed to produce xylose and  $\beta$ -1,3-xylobiose by TnB9K760, while xylose was not detected in the hydrolysates of AncXyl10 by HPLC after a long incubation (24 h). Therefore, this may be one of the reasons that xylose does not exist in the hydrolysates of AncXyl10. In the process of ancestral sequence reconstruction, the sequence gap negatively affected the radius of receptors site sphere (X2, Receptors radius), the receptor surface volume (X3) and surface area (X10), resulting in blocking the entry of  $\beta$ -1,3-oligosaccharides (multiple xylose units). This may explain the absence of  $\beta$ -1,3-xylo-tetraose and  $\beta$ -1,3-xylopentaose in the



**Fig. 4.** Multiple sequence alignment of TnB9K760 and AncXyl10. The red triangle indicates the active site, the top indicates the secondary structure of the corresponding sequence, and the bottom indicates the solvent accessibility of the corresponding sequence. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

hydrolysates of AncXyl10. In terms of the calculation results of CDOCKER\_ENERGY and the experimental phenomenon, the generated  $\beta$ -1,3-xylotriase was not conducive to binding in the active cavity of AncXyl10, resulting in no further hydrolysis of  $\beta$ -1,3-xylotriase.

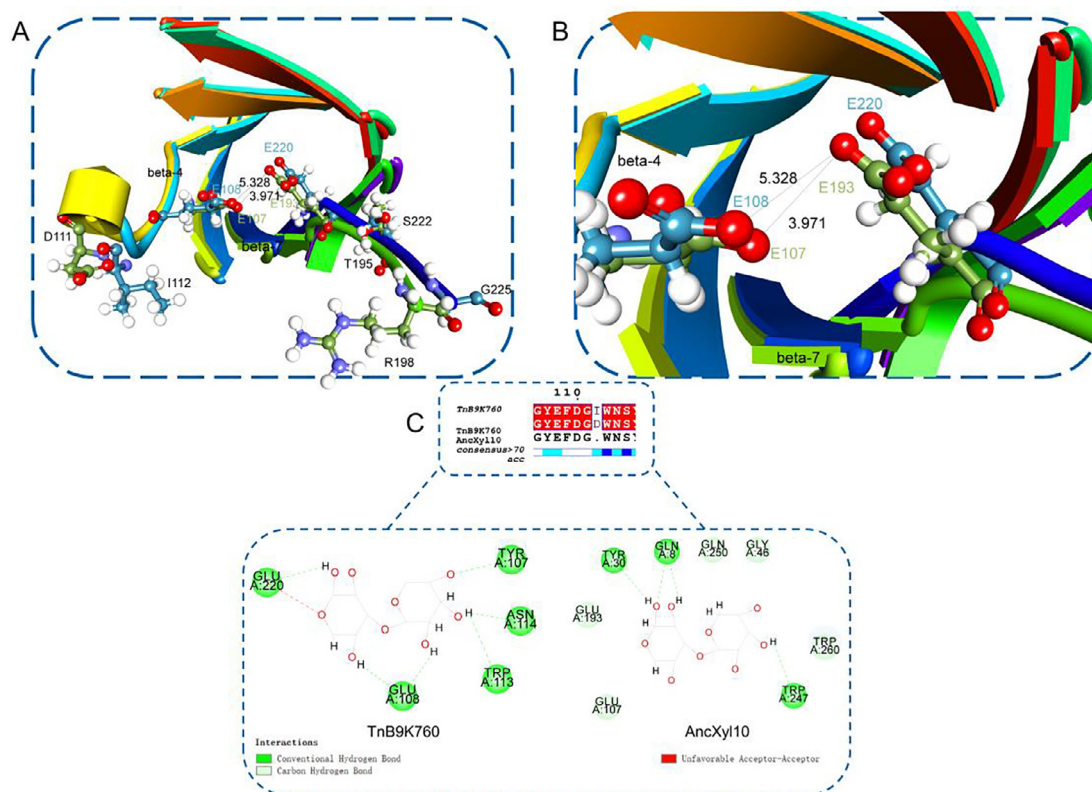
Secondly, ancient amino acid replacements were another important one. ASR uses alignment of extant sequences, phylogenetic tree and evolutionary models to calculate marginal posterior probability for each sequence position and each ancestral node. Based on the definite evolutionary model, the ancestor protein and the existing protein have some ancestral amino acid substitutions in the sequence of some position. After careful checking, we found that the ancestor amino acid substitution greatly influenced the factor X1 and X8. In terms of factor X8 (-CDOCKER\_INTERACTION\_ENERGY), the interaction energies of TnB9K760 and AncXyl0 with  $\beta$ -1,3-xylobiose are 34.15 and 33.91 (Kcal/mol), respectively. The number of conventional hydrogen bond (H-H) of TnB9K760 and AncXyl0 with  $\beta$ -1,3-xylobiose were six and five, respectively. The weakening of the interaction energy is due to the amino acid substitution (I112D) in the ancestor of TnB9K760, resulting in the original three aromatic amino acids (W107, W113, 114 N) being unable to form an interaction bone (H-H) with the substrate  $\beta$ -1,3-xylobiose in space. As aromatic residues have strong hydrophobic interactions with oligosaccharides, the number of aromatic residues and their spatial arrangement were both critical to the binding capacity of oligosaccharides [46,47]. More aromatic residues interact with  $\beta$ -1,3-xylobiose by aromatic face in TnB9K760, while more aromatic residues interact with  $\beta$ -1,3-xylobiose by aromatic edge in AncXyl10 (Supplementary Fig. S6). Compared with TnB9K760, the interaction between AncXyl10 and  $\beta$ -1,3-xylobiose was weakened, resulting in unstable binding of  $\beta$ -1,3-xylobiose to the active cavity of AncXyl10, whereas the stable binding is very important for the enzymatic hydrolysis of substrates.

As mentioned above, the active site distance (X1) was the most important factor, to investigate the factor, we displayed the TIM barrel core and active site regions of TnB9K760 and AncXyl10 in DS2019 (Fig. 5A). The distances between the catalytic group (two glutamates) OE2 atom of TnB9K760 and AncXyl0 were 5.238 and 3.971, respectively (Fig. 5B). Ancient amino-acid had replacements

(I112D, S222T and G225R) on the TnB9K760 protein after ancestral sequence reconstruction. The replacement of longer branched chain residues with shorter branched ones shortened the distance between the active groups (X1). The shortening of the active site distances may change the spatial distribution of the active site in the  $\beta$ -1,3-xylobiose. The two active sites were distributed on two xylose units of  $\beta$ -1,3-xylobiose in the  $\beta$ -1,3-xylanase TnB9K760, while the two active sites were distributed on one xylose units of  $\beta$ -1,3-xylobiose in AncXyl10 (Fig. 5C). This can also be explained with subsites proposed by Nakamichi et al. study [48], the cleavage site of TnB9K760 for  $\beta$ -1,3-xylan was located between -1 and +1 subsites ( $\beta$ -1,3-xylobiose), while the active sites (Glu) of AncXyl10 cannot contact two xylose units at the same time due to the shorter distance in space. Therefore, the weakening of the  $\beta$ -1,3-xylobiose interaction (X8) and the shorter active site distance (X1) may lead AncXyl10 unable to hydrolyze  $\beta$ -1,3-xylobiose, which is consistent with the experimental phenomenon that the content of  $\beta$ -1,3-xylobiose does not decrease during the 24 h long-term incubation.

### 3.5. A novel integrative tool for engineering, screening the hydrolysates types of hydrolase

According to our recent studies, the hydrolysates of AncXyl09 were 75.2%  $\beta$ -1,3-xylobiose, 14.5%  $\beta$ -1,3-xylotriase, 4.4%  $\beta$ -1,3-xylotetraose and 5.9%  $\beta$ -1,3-xylopentose. Herein, the hydrolysates of AncXyl10 were 60.38%  $\beta$ -1,3-xylobiose and 39.62%  $\beta$ -1,3-xylotriase. As two ancestral mutants (AncXyl09 and AncXyl10) from the six ancestral  $\beta$ -1,3-xylanases mutagenesis library were proved to product only oligosaccharide, indicating the optimized ASR strategy was effective in engineer the hydrolysate types of  $\beta$ -1,3-xylanases. After rigorous cross-validation with experimentally verified training sample, the prediction accuracy of LogitBoost and PLSR model for oligosaccharides was 100%. Robust mathematical models ensured that the target enzymes can be screened from the ancestral  $\beta$ -1,3-xylanases library. Thus, we proposed a novel strategy for engineering and screening the hydrolase with desired hydrolysate types which will provide an effective tool for mining enzymes for preparing functional  $\beta$ -1,3-xylo-oligosaccharides (Scheme 1).



**Fig. 5.** Effects of amino acid substitution in AncXyl09 and AncXyl10. Blue represents the residues in TnB9K760. Green represents the residues in AncXyl10. A. Five pairs of amino acid substitutions existed in the active cavity. B. Ancient amino-acid replacements shorten the distances between active sites of AncXyl10. C. Ancient amino-acid replacements changed the interaction modes of TnB9K760 and AncXyl10 with  $\beta$ -1,3-xylobiose. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We have a concept in engineering  $\beta$ -1,3-xylanase to avoid the blindness of experiments and reduce the workload of experiments. To achieve this goal, we firstly constructed potential mutagenesis library of  $\beta$ -1,3-xylanases by collected all the reported dataset of  $\beta$ -1,3-xylanase and define the two hydrolysate types as 0 and 1. Secondly, 11 potential factors were generated by digital acquisition of  $\beta$ -1,3-xylanases characterization through fishbone and the molecular docking performed by CDOCKER[26] implemented in Discovery Studio 2019, which might affect the hydrolysate types. Thirdly, 5 important factors were screened out from the 11 potential factors through five non-linear attribute selection methods. Then, a LogitBoost classifier (non-linear algorithm) and a partial least squares regression model (PLSR, linear algorithm) were established to predict the hydrolysate types of  $\beta$ -1,3-xylanase base on the selected five factors. As the prediction accuracy of the LogitBoost for oligosaccharides was as high as 100%, and the PLSR for the two types was 100%, the two models were adopted to predict the remaining five ancestral proteins and the AncXyl10 was predicted as the type oligosaccharides. Finally, The experimental validation proved the hydrolyzed products of AncXyl10 were only  $\beta$ -1,3-xylobiose and  $\beta$ -1,3-xylotriose, which verified the reliability of the classifiers. This novel combinatorial strategy was successfully applied to  $\beta$ -1,3-xylanases. As a research model,  $\beta$ -1,3-xylanase has a highly conserved  $(\beta/\alpha)_8$  barrel structure and the active sites are experimentally verified to be glutamate. By searching through the CAZy database, we found many glycoside hydrolases with such similar characteristics. For example, the xylanases in GH10 family and the  $\beta$ -amylases in GH14 family. This indicated that they have similar catalytic mechanisms. So, the same factors can be collected to construct robust model for screening mutant libraries. On the other hand, the ancestral sequence

reconstruction has many successful examples, and two of the six ancestral proteins constructed by our optimized ancestral sequence reconstruction strategy were experimentally verified to have altered hydrolysis product types. This implied that the ancestral sequence reconstruction strategy allows the efficient construction of mutation libraries. Perhaps further research will be conducted for more carbohydrate-active enzymes as well as other related enzymes to prove the versatility of this strategy.

#### 4. Conclusions

By integrating the optimized ASR and data mining tools, we proposed a new strategy for engineering and screening novel  $\beta$ -1,3-xylanase for  $\beta$ -1,3-xylo-oligosaccharides production, which could significantly avoided the blindness and experimental workload comparing with the natural enzyme screening or traditional engineering methods. The hydrolytic products of the obtained AncXyl10 were only  $\beta$ -1,3-xylobiose (60.38%) and  $\beta$ -1,3-xylotriose (39.62%), which could facilitate the preparation of oligosaccharides with high purity. Since many carbohydrate-active enzymes have highly conserved active sites, the strategy may be an effective tool for mining or designing other carbohydrates hydrolase that produce desired functional oligosaccharides.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.



## Funding

This work was funded by the Natural Science Foundation of Fujian Province of China (2020J01079).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2022.06.050>.

## References

- Risso VA, Sanchez-Ruiz JM, Ozkan SB. Biotechnological and protein-engineering implications of ancestral protein resurrection[J]. *Curr Opin Biotechnol* 2018;51:106–15.
- Candel AM, Luisa Romero-Romero M, Gamiz-Arco G, et al. Fast folding and slow unfolding of a resurrected Precambrian protein[J]. *P Natl Acad Sci USA* 2017;114(21):4122–3.
- Nguyen V, Wilson C, Oemberger HM, et al. Evolutionary drivers of thermoadaptation in enzyme catalysis[J]. *Science* 2017;355(6322):289–94.
- Clifton BE, Jackson CJ. Ancestral protein reconstruction yields insights into adaptive evolution of binding specificity in solute-binding proteins[J]. *Cell Chem Biol* 2016;23(2):236–45.
- Nakano S, Niwa M, Asano Y, et al. Following the evolutionary track of a highly specific L-arginine oxidase by reconstruction and biochemical analysis of ancestral and native enzymes[J]. *Appl Environ Microbiol* 2019;85(12):1–12.
- Zeng B, Zhou Y, Yi Z, et al. Highly thermostable and promiscuous beta-1,3-xylanases designed by optimized ancestral sequence reconstruction[J]. *Bioresour Technol* 2021;340:125732–41.
- Planas-Iglesias J, Marques SM, Pinto GP, et al. Computational design of enzymes for biotechnological applications[J]. *Biotechnol Adv* 2021;47:107696–718.
- Zhang C, Fang B. LogitBoost classifier for discriminating thermophilic and mesophilic proteins[J]. *J Biotechnol* 2007;127(3):417–24.
- Zhang G, Li H, Fang B. Discriminating acidic and alkaline enzymes using a random forest model with secondary structure amino acid composition[J]. *Process Biochem* 2009;44(6):654–60.
- Iriki Y, Suzuki T, Nisizawa K, et al. Xylan of siphonaceous green algae[J]. *Nature* 1960;187:82–3.
- Cai Z, Ge H, Yi Z, et al. Characterization of a novel psychrophilic and halophilic beta-1,3-xylanase from deep-sea bacterium, *Flammeovirga pacifica* strain WPAGA1[J]. *Int J Biol Macromol* 2018;118(B):2176–84.
- Kiyohara M, Sakaguchi K, Yamaguchi K, et al. Molecular cloning and characterization of a novel beta-1,3-xylanase possessing two putative carbohydrate-binding modules from a marine bacterium *Vibrio sp* strain AX-4[J]. *Biochem J* 2005;388(3):949–57.
- Okazaki F, Nakashima N, Ogino C, et al. Biochemical characterization of a thermostable beta-1,3-xylanase from the hyperthermophilic eubacterium, *Thermotoga neapolitana* strain DSM 4359[J]. *Appl Microbiol Biot* 2013;97(15):6749–57.
- Yi Z, Cai Z, Zeng B, et al. Identification and characterization of a novel thermostable and salt-tolerant beta-1,3-xylanase from *Flammeovirga pacifica* Strain WPAGA1[J]. *Biomolecules* 2020;10(9):1287–304.
- Kim S, Sim H, Kim J, et al. Enhanced production of 2,3-butanediol from xylose by combinatorial engineering of xylose metabolic pathway and cofactor regeneration in pyruvate decarboxylase-deficient *Saccharomyces cerevisiae*[J]. *Bioresour Technol* 2017;245:1551–7.
- Liang WS, Liu TC, Chang CJ, et al. Bioactivity of  $\beta$ -1,3-xylooligosaccharide Extracted from *Caulerpa lentillifera* by Using *Escherichia coli* Clear Coli BL21 (DE3)- $\beta$ -1,3-xylanase XYLIII[J]. *J Food Nutr Res* 2015;7(3):437–44.
- Maeda R, Ida T, Ihara H, et al. Induction of apoptosis in MCF-7 Cells by beta-1,3-Xylooligosaccharides prepared from *Caulerpa lentillifera*[J]. *Biosci Biotech Bioch* 2012;76(5):1032–4.
- Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server[J]. *Nucleic Acids Res* 2004;32(2):526–31.
- Arnold K, Bordoli L, Kopp J, et al. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling[J]. *Bioinformatics* 2006;22(2):195–201.
- Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction[J]. *Nat Protoc* 2010;5(4):725–38.
- Bowie JU, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure.[J]. *Science* 1991;253(5016):164–70.
- Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions[J]. *Protein sci* 1993;2(9):1511–9.
- Davis IW, Leaver-Fay A, Chen VB, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids[J]. *Nucleic Acids Res* 2007;35(2):375–83.
- Miller BT, Singh RP, Klauda JB, et al. CHARMMing: A new, flexible web portal for CHARMM[J]. *J Chem Inf Model* 2008;48(9):1920–9.
- Tirado-Rives J, Jorgensen WL. Contribution of conformer focusing to the uncertainty in predicting free energies for protein-ligand binding[J]. *J Med Chem* 2006;49(20):5880–4.
- Wu GS, Robertson DH, Brooks CL, et al. Detailed analysis of grid-based molecular docking: A case study of CDOCKER – A CHARMM-based MD docking algorithm[J]. *J Comput Chem* 2003;24(13):1549–62.
- Meslamani J, Li J, Sutter J, et al. Protein-ligand-based pharmacophores: Generation and utility assessment in computational ligand profiling[J]. *J Chem Inf Model* 2012;52(4):943–55.
- Finn RD, Mistry J, Schuster-Bockler B, et al. Pfam: clans, web tools and services [J]. *Nucleic Acids Res* 2006;34(1):247–51.
- Friedman J, Hastie T, Tibshirani R. Additive logistic regression: A statistical view of boosting[J]. *Ann Stat* 2000;28(2):337–74.
- Wold S, Sjostrom M, Eriksson L. PLS-regression: a basic tool of chemometrics [J]. *Chemometr Intell Lab* 2001;58(2):109–30.
- Smogyi M. Notes on sugar determination[J]. *J Biol Chem* 1952;195(1):19–23.
- Bershtein S, Kleiner D, Mishmar D. Predicting 3D protein structures in light of evolution[J]. *Nat Ecol Evol* 2021;5(9):1195–8.
- Gumulya Y, Gillam EMJ. Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the 'retro' approach to protein engineering[J]. *Biochem J* 2017;474(1):1–19.
- Thomas A, Cutlan R, Finnigan W, et al. Highly thermostable carboxylic acid reductases generated by ancestral sequence reconstruction[J]. *Commun Biol* 2019;2(429):1–12.
- Bazzan A, Engel PM, Schroeder LF, et al. Automated annotation of keywords for proteins related to mycoplasmatocaeae using machine learning techniques[J]. *Bioinformatics* 2002;18(2):35–43.
- Kretschmann E, Fleischmann W, Apweiler R. Automatic rule generation for protein annotation with the C4.5 data mining algorithm applied on SWISS-PROT[J]. *Bioinformatics* 2001;17(10):920–6.
- Tobler JB, Molla MN, Nuwaysir EF, et al. Evaluating machine learning approaches for aiding probe selection for gene-expression arrays[J]. *Bioinformatics* 2002;18(1):164–71.
- Li J, Liu H, Ng S, et al. Discovery of significant rules for classifying cancer diagnosis data[J]. *Bioinformatics* 2003;19(2):93–102.
- Frank E, Hall M, Trigg L, et al. Data mining in bioinformatics using Weka[J]. *Bioinformatics* 2004;20(15):2479–81.
- Burnett AC, Anderson J, Davidson KJ, et al. A best-practice guide to predicting plant traits from leaf-level hyperspectral data using partial least squares regression[J]. *J Exp Bot* 2021;72(18):6175–89.
- Xu Q, Chao Y, Wan Q. Health benefit application of functional oligosaccharides [J]. *Carbohydr Polym* 2009;77(3):435–41.
- Zhao C, Wu Y, Liu X, et al. Functional properties, structural studies and chemo-enzymatic synthesis of oligosaccharides[J]. *Trends Food Sci Tech* 2017;66:135–45.
- Gullon B, Gomez B, Martinez-Sabajanes M, et al. Pectic oligosaccharides: Manufacture and functional properties[J]. *Trends Food Sci Tech* 2013;30(2):153–61.
- Ou J, Sun Z. Feruloylated oligosaccharides: Structure, metabolism and function [J]. *J Funct Foods* 2014;7:90–100.
- Zou P, Yang X, Wang J, et al. Advances in characterisation and biological activities of chitosan and chitosan oligosaccharides[J]. *Food Chem* 2016;190:1174–81.
- Eide KB, Stockinger LW, Lewin AS, et al. The role of active site aromatic residues in substrate degradation by the human chitotriosidase[J]. *BBA-Proteins Proteom* 2016;1864(2):242–7.
- Mattinen ML, Kontteli M, Kerovuo J, et al. Three-dimensional structures of three engineered cellulose-binding domains of cellobiohydrolase I from *Trichoderma reesei*[J]. *Protein Sci* 1997;6(2):294–303.
- Nakamichi Y, Fouquet T, Ito S, et al. Mode of action of GH30-7 reducing-end xylose-releasing exoxylanase A (Xyn30A) from the filamentous fungus *Talaromyces cellulolyticus*[J]. *Appl Environ Microb* 2019;85(13):1–17.