Sequence homolog-based molecular engineering for shifting the enzymatic pH optimum

Fuqiang Ma a,1, Yuan Xie a,1, Manjie Luo a, Shuhao Wang a, You Hu b, Yukun Liu b, Yan Feng a, Guang-Yu Yang a, c, *

a State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, 200240, China
b School of Statistics, East China Normal University, Shanghai 200241, China
c Shanghai Collaborative Innovation Center for Biomanufacturing (SCICB), East China University of Science and Technology, 130 Meilong Road, Shanghai 200237, China

ABSTRACT

Cell-free synthetic biology system organizes multiple enzymes (parts) from different sources to implement unnatural catalytic functions. Highly adaption between the catalytic parts is crucial for building up efficient artificial biosynthetic systems. Protein engineering is a powerful technology to tailor various enzymatic properties including catalytic efficiency, substrate specificity, temperature adaptation and even achieve new catalytic functions. However, altering enzymatic pH optimum still remains a challenging task. In this study, we proposed a novel sequence homolog-based protein engineering strategy for shifting the enzymatic pH optimum based on statistical analyses of sequence-function relationship data of enzyme family. By two statistical procedures, artificial neural networks (ANNs) and least absolute shrinkage and selection operator (Lasso), five amino acids in GH11 xylanase family were identified to be related to the evolution of enzymatic pH optimum. Site-directed mutagenesis of a thermophilic xylanase from Caldicellulosiruptor bescii revealed that four out of five mutations could alter the enzymatic pH optima toward acidic condition without compromising the catalytic activity and thermostability. Combination of the positive mutants resulted in the best mutant M31 that decreased its pH optimum for 1.5 units and showed increased catalytic activity at pH < 5.0 compared to the wild-type enzyme. Structure analysis revealed that all the mutations are distant from the active center, which may be difficult to be identified by conventional rational design strategy. Interestingly, the four mutation sites are clustered at a certain region of the enzyme, suggesting a potential "hot zone" for regulating the pH optima of xylanases. This study provides an efficient method of modulating enzymatic pH optima based on statistical sequence analyses, which can facilitate the design and optimization of suitable catalytic parts for the construction of complicated cell-free synthetic biology systems.

© 2016 The Authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

In recent years, considerable progress has been made in the field of cell-free synthetic biology in both conception and practice [1–3]. Building up multi-enzyme catalytic cascades in vitro is beginning to provide an improved toolbox and more efficient processes in medical diagnostics, synthesis of fuel compounds, drug molecules and proteins [4–6]. With the excellence of easy to be constructed, tested and optimized, cell-free synthetic biology system can facilitate enzymatic catalysis under non-physiological condition and is compatible with cytotoxic components. Thus, cell-free synthetic biology system possesses unique advantages over conventional in vivo system. However, the construction of cell-free synthetic biology system relies on recruiting and organizing multiple enzymes from diverse organisms to work synergistically, in which case the overall efficiency is depended on the compatibility of each
enzyme to the reaction condition. Especially, employing enzymes with different pH optima requires a compromise in choosing pH value of the reaction buffering system, which would severely sacrifrice the biosynthetic efficiency [78]. Therefore, shifting enzymatic pH optima towards desired pH value is crucial for the high performance of a cell-free synthetic biology system.

Protein engineering is powerful in optimizing various enzymatic catalytic properties, such as catalytic activity, stability, reaction temperature, substrate specificity, etc. [9—12]. Scientists also try to use this technique to change the enzymatic pH optima to fulfill the requirement of harsh industrial conditions [13—17]. Theoretically, the enzymatic pH optimum is governed by the pKa value of key catalytic residues, which can be tuned by the mutations adjacent to the active center. Nielsen and co-workers developed pKD webserver to predict the pKa changes caused by amino acid substitutions [18,19]. Ohara et al. described that unique extended hydrogen bond network in the active site was important for the physical chemistry requirement for shifting the pH value, bringing a risk of losing catalytic activity. In addition, since pH value is expected interference on substrate binding or catalysis process, which ever, due to the complicated amino acid interaction networks in the active center, Nielsen and co-workers [16,20—22], sequence alignment analysis [17,23—26] and charged residues implantation [13,27]. For instance, Sugino and co-workers shifted the pH optimum of Acremonium ascorbate oxidase upward 0.5—1.0 unit by randomly mutagenesis [20]. Li et al. improved the pH optimum of Aspergillus niger xylanase by site-directed mutagenesis to charged residues [17]. Qiu et al. tailored the pH dependency of human non-pancreatic secretory phospholipase A2 by surface charge replacements [13]. Although some successes have been reported, most of them only achieved little change in pH optima, probably due to the effect of mutations was neutralized by the complicated interactions within the enzyme. Therefore, developing novel efficient protein engineering strategies for shifting enzymatic pH optima is an extremely challenging task.

To overcome these issues, strategies require less knowledge about the structure-function relationship have been employed to engineer enzymatic pH optima, such as directed evolution [16,20—22], sequence alignment analysis [17,23—26] and charged residues implantation [13,27]. For instance, Sugino and co-workers shifted the pH optimum of Acremonium ascorbate oxidase upward 0.5—1.0 unit by randomly mutagenesis [20]. Li et al. improved the pH optimum of Aspergillus niger xylanase by site-directed mutagenesis to charged residues [17]. Qiu et al. tailored the pH dependence of human non-pancreatic secretory phospholipase A2 by surface charge replacements [13]. Although some successes have been reported, most of them only achieved little change in pH optima, probably due to the effect of mutations was neutralized by the complicated interactions within the enzyme. Therefore, developing novel efficient protein engineering strategies for shifting enzymatic pH optima is an extremely challenging task.

The long history of evolution demonstrates the relationship between amino acid sequences of enzymes and their optimal catalytic conditions. In this study, we developed a novel strategy for shifting the pH adaptation of enzymes by means of biomathematics and biostatistics (Scheme 1). From more than one thousand GH11 xylanase sequences, 113 non-redundant, well-characterized enzymes were collected to construct a database with annotated pH optima. After digitizing the amino acid sequences according to isoelectric point (pI value) and hydrophathy index (Hy value), the relationship between the sequences and their pH optima was analyzed by artificial neural networks (ANNs) and least absolute shrinkage and selection operator (Lasso) algorithms. Five potential residues related to pH adaptation were identified. By introducing these single-site mutations to the catalytic domain of a neutral xylanase from thermophilic bacterium Caldicellulosiruptor bescii DSM 6725 (C�X-CD), four out of five mutants showed significant shift in pH optima without disrupting the catalytic activity. Combination of the positive sites further increased the shift of pH optimum by 1.5 units. The mechanism on how the mutations affect enzymatic pH optimum was also discussed.

2. Material and methods

2.1. Material

Beech wood xylan was purchased from Sigma-Aldrich (St. Louis, USA). Restriction enzyme and T4 ligase were purchased from New England Biolabs (Ipswich, MA). PrimeSTAR polymerase was purchased from Takara (Dalian, China). The QIAquick PCR purification kit was purchased from Qiagen (Hilden, Germany). The pET-28a vector was purchased from Novagen (Darmstadt, Germany). E. coli BL21-CodonPlus (DE3)-RII strains were purchased from Invitrogen (Carlsbad, CA, USA) and was used for DNA manipulation and recombinant protein production. Caldicellulosiruptor bescii DSM 6725 was obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ).

2.2. Sequences collection, alignment and construction of phylogenetic tree

Based on the classification of a preliminary release of the CAZy database (http://www.cazy.org/GH11.html) [39], GH11 xylanase with characterized data were collected and their sequences were extracted from NCBI database. According to the analysis of BlastP and signalP 3.0, carbohydrate binding module (CBM), carbohydrate binding domain (CBD) and signal peptides were deleted from xylanase sequences, thus only the catalytic domain (CD) were reserved. Then, the samples which showed more than 95% sequence identity were further removed using Blastclust (http://toolkit.tuebingen.mpg.de/blastclust). By investigating literature that reported the enzymatic pH-activity profiles, the pH optima of enzymes were determined as the pH values where the enzymes showed their highest activities. After filtrating the literature information, all the collected xylanases have been annotated a pH optimum, and database containing the data was constructed (Appendix B). Sequences were assembled manually, aligned by Clustal X 1.83, [40] and the result was shown by ESPript3.0 [41]. Mega 5.0 was employed to construct the Neighbor-Joining phylogenetic tree of 113 GH11 xylanases.

2.3. Data analysis by ANNs and Lasso

Each amino acid in the sequence was digitized by their isoelectric point (pI value) and hydrophathy index (Hy value), respectively (Table 2). Both pI value and Hy value are experimental
The isoelectric point (pI value) is the pH value at which the zwitterion predominates, but coexists in dynamic equilibrium with small amounts of net negative and net positive ions (https://en.wikipedia.org/wiki/Amino_acid). The pI value can be determined by the pK_a values of \( \alpha\)-COOH, \( \alpha\)-NH_2 and side-chains, which are measured by titration method, thus the pl value is an experimental parameter. On the other hand, the hydropathy index of an amino acid is a number representing the hydrophobic or hydrophilic properties of its side-chain. It was proposed in 1982 by Jack Kyte and Russell F. Doolittle [42]. The Hy value can be determined by measuring the water-vapor transfer free energies of side-chains (\( \Delta r Gm \)) and the burying fractions of side-chains when forming into proteins. Therefore, the Hy value is also an experimental parameter.

The pl value represents the dissociation ability of amino acids and the Hy value represents the polarity degree of amino acid side-chains. Both of them reflect the influence of a particular residue on the pK_a values of adjacent amino acid through molecular interactions, which in turn determine the enzymatic pH optima [43]. Therefore, the effects of amino acids on enzymatic pH optimum can be evaluate in a more reliable manner, which both parameters were considered simultaneously.

By digitalization employing pl and Hy values, the amino acid sequences were converted into sequences of digits. The gaps in the multi-sequence alignments were filled by figure zero. For example, the amino acid sequence AGD-HNEKEAA in the alignment can be converted into pl values as 6.11, 6.06, 2.85, 0, 7.6, 5.41, 3.15, 6.11 and 6.11. The conversion of hydropathy index is similar to pl values, so the sequence can be converted into hydropathy index 1.8, −0.4, −3.5, 0, −3.2, −3.5, −3.5, −3.9, −3.5, 1.8, and 1.8.

### 2.3.1. ANNs analysis

According to the pH optima, numerical sequences of alkaline, neutral and acidic xylanases were assigned as 2, 1 and 0,
respectively. Using the Analysis of Variance test (ANOVA) in the statistics software R to carry out the analysis towards every site in all GH11 xylanases in the library, the sites with p-value<0.01 were selected. The selected sites were further analyzed by NNET program in the statistics software R. The preliminary analysis result was imported into the input layer of NNET, and the parameters were selected to carry out the training of the classifier. After training, the weight of each layer was checked. The weight of each edge was calculated using a simple multiplication method, the impact of each site on the final classification was calculated, and the important sites for enzymatic pH adaptation were selected following these results.

2.3.2. Lasso analysis

(1) Data preprocessing: we eliminated nearly gap sites at which most sequences have figure zero or variances across all sequences are < 0.01. (2) Correlation analysis: correlation coefficient was used to quantify the correlation between pH optimum and pi value or hydropathy index at each site by statistical software R. The sites with absolute correlation coefficients less than 0.01 were removed, and the remaining sites were kept for subsequent feature screening under a linear regression model. (3) Feature screening through Lasso: based on a linear regression model between pH optimum and pi value or between pH optimum and a hydropathy index, we performed feature screening for key sites through the Lasso by the LARS software package in statistics software R.

The source codes used in ANNs and Lasso analysis were summarized in Appendix A.

2.4. Recombinant protein expression and purification

Wild type xylanase CbX-CD gene was amplified from genomic DNA of C. bescii DSM 6725 using primers containing the restriction sites of Ncol and Xho I (Appendix A, Table A3). Polymerase chain reaction (PCR) amplification was carried out with PrimeSTAR polymerase and a temperature program consisting of 98 °C for 2 min; 30 cycles of 10 s at 98 °C, 15 s at 55 °C, and 1 min at 72 °C; and a final 10-min extension at 72 °C. The PCR product was digested with Ncol and XhoI and subsequently cloned into pET-28a vector (pET28a-WT), which was then mutagenized into the BL21-CodonPlus (DE3)-RIL cells by electroporation. The mutants were prepared by whole-plasmid PCR using the primers containing mutagenesis at those target sites (Appendix A, Table A3). PCR was performed with PrimeSTAR polymerase and a temperature program consisting of 98 °C for 2 min; 30 cycles of 10 s at 98 °C, 15 s at 55 °C, and 7 min at 72 °C; and a final 10-min extension at 72 °C. The PCR products were digested with DpnI to remove the parent plasmid and purified with a PCR purification kit. The PCR products were electroporated into BL21-CodonPlus (DE3)-RIL cells. The expression and purification of recombinant protein followed the described method [44]. In brief, cells were grown at 37 °C in 2YT medium supplemented with 50 μg/ml kanamycin until the optical density at 600 nm (OD600) reached 0.6 to 0.8. Gene expression was induced for 16 h at 26 °C by the addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The cells were harvested and suspended in 30 mM Tris-HCl buffer (pH 8.0) containing 150 mM NaCl and 30 mMimidazole, and then disrupted by sonication. The recombinant protein was purified by Ni-NTA affinity chromatography (Qiagen, Hilden, Germany). Protein concentration was measured by Bradford method using bovine serum albumin as standard (Thermo Scientific, Waltham, USA).

2.5. Determination of enzymatic activities and properties

The standard assay for xylanase activity was performed at 65 °C in 40 mM pH 6.8 sodium phosphate buffer in the presence of 1.0% (w/v) beech wood xylan for 5 min. The amount of reducing sugars released was determined with the 3,5-dinitrosalicylic acid (DNS) reagent, using xylose as standard. After incubation, DNS reagent was added and the samples were heated in a boiling water bath for 5 min followed by cooling on ice. The absorbance was then measured at 540 nm. Each assay was performed in triplicate. One unit of xylanase activity is defined as the amount of enzyme required to release one μmol of reducing-sugar equivalents per minute at 65 °C, pH 6.8.

The effects of pH on enzyme activity were determined at 65 °C under pH ranging from 4 to 8 using 1% (w/v) beech wood xylan as substrate. The reaction buffer contained 30 mM each of 4-[2-hydroxymethyl]–1-piperazinethanesulfonic acid (HEPES), 3-[[1,3-dihydroxy-2-[(hydroxymethyl)propan-2-yl]amino]propane-1-sulfonic acid (TAPS), 3-[(Cyclohexylamino)-1-propanesulfonic acid (CAPS), 2-(N-morpholino)ethanesulfonic acid (MES) and acetic acid, and was adjusted to the appropriate pH at 65 °C with 1M NaOH.

2.6. Calculation of the apparent pKa1 and pKa2 values

The apparent pKa1 and pKa2 values were calculated by fitting the pH-activity profiles of wild-type CbX-CD and its mutants using non-linear fitting software 1stOpt. The fitting equation is shown in Eq. (1).

\[
V = \frac{V_{\text{max}}}{1 + 10^{(pK_{a1} - pH)}} + 10^{(pH - pK_{a2})}
\]

(1)

Where \(V_{\text{max}}\) is the pH-independent maximum reaction rate, and apparent pKa1 and pKa2 are the dissociation constants of the key catalytic residues, respectively.

2.7. Homology modeling

The 3D structure model of CbX-CD and its mutants were generated using SWISS-MODEL (http://swissmodel.expasy.org/). A xylanase from Dictyoglomus thermophilum Rt468.1 (PDB ID 1FS) was selected as the template with identity of approximately 88%. The geometry of the loop regions was corrected using Refine Loop/ MODELER. The quality of the model was evaluated by PROCHECK [45] and Profile-3D of DS 3.0. Structural figures were generated using PyMOL (http://www.pymol.org/).

3. Results

3.1. Sequence analyses of GH11 xylanase family

More than one thousand GH11 family xylanase sequences were reported on CAZy database, and more than 260 of them had basic function annotations. After literature investigation, 113 xylanase sequences (<95% sequence identity) from eukaryotic or bacterial species with experimentally measured pH optima were collected. The pH optima of the enzymes vary widely from pH 2.0 to 9.0 [46]. Among those sequences, 67 were annotated as acidic xylanases (pH optimum < 6.5), 36 were annotated as neutral xylanases (7.5 ≥ pH optimum ≥ 6.5) and 10 were annotated as alkaline xylanases (pH optimum > 7.5).

To study the evolutionary relationship within GH11 xylanase family, signal peptides, carbohydrate binding modules (CBM) and carbohydrate binding domains (CBD) were predicted using online software Signal 3.0 and SMART, respectively. These sequences were removed and only the catalytic domain (CD) of xylanases was remained for sequence alignment. The multi-sequence alignment
was performed by using Clustal X1.83 and the alignment result was shown in Appendix C. Based on the alignment, Neighbor-Joining phylogenetic tree was constructed using Mega 5.0. As shown in Fig. 1, the enzymes were clustered in branches with different pH optima. The majority of zone A was neutral xylanases, with a small portion of acidic and alkaline enzymes. Acidic and neutral xylanases were the major part in zone C, with several alkaline enzymes sporadically distributed. While in zone B and D, all the members were acidic xylanases. These results indicated that the pH optima of GH11 xylanase were relevant to divergent evolution, but it was hard to get a clear evolutionary relationship from the phylogenetic tree. Instead, the fact that alkaline xylanases are distributed among acidic and neutral enzymes suggested that some site mutations occurred during evolution might have significant impact on enzymatic pH adaptation.

3.2. Statistical sequence analysis by ANNs and Lasso

Statistical procedures are playing an increasing important role in extracting evolutionary rules from the vast ocean of protein sequences. Before employing statistical procedures, it is necessary to digitize amino acid sequences. We used two amino acids physical chemistry characteristics, isoelectric point (pI value) and hydropathy index (Hy value), for the digitalization, because they are important for determining the protonation state of an amino acid. After this conversion, the effect of each residue in protein sequence can be quantitatively evaluated by statistics algorithms.

Artificial neural networks (ANNs) are statistical machine learning models. By emulating the processing estimate or approximate functions that depend on a large number of inputs and are generally unknown, ANNs can analyze data in a way which generalizes its mapping to new data [47,48]. Therefore, ANNs are suitable for searching the possible sequence-function divergence ‘hot spots’, especially for navigating the possible mutagenesis for enzymatic pH adaptation among the numerous evolution data.

After analysis by ANNs towards pH-pl values and pH-Hy index correlation, 51 sites from pl library and 57 sites from Hy library were considered to be important for pH adaptation, and those sites were ranked by their weights (Appendix A, Table A1).

Least absolute shrinkage and selection operator (Lasso) can provide sparsity inducing estimation of regression coefficients by adding penalty functions to the trait-associated subset of markers into the model when the number of markers is larger than the number of individuals in the sample [49,50]. By feature screening through Lasso under linear regression models, it was noticed that pl or Hy values of some amino acid sites differed obviously at xylanases with different pH optima. In this case, these sites (independent variable) could better explain the model and their coefficients estimated by Lasso turned out to be nonzero even at higher Lasso constraints. On the contrary, sites whose pl or Hy values showed little difference at different pH optima were considered to have weak explanation to the model. And the coefficients of this kind of sites were prone to be zero at higher Lasso constraints. When loosening the Lasso restriction conditions gradually, the number of nonzero independent variables (sites) increased continually. To find key sites related to pH optima, we chose restriction conditions \( K = 40 \) and \( K = 26 \) to perform Lasso analysis on pl and Hy database, respectively. The suitable restriction conditions ensured that about 20 target sites could be selected on ChX-CD. As shown in Appendix A, Table A1, 23 sites for pl value and 21 sites for Hy value were identified and were ranked by their weights.

Then, the prediction results from ANNs and Lasso were comprehensively analyzed. The amino acid sites identified simultaneously by both algorithms were considered as important sites that may contribute to the pH adaptation of GH11 xylanase family. As shown in Appendix A, Table A1, some amino acid sites located at N terminal and C terminal (marked by asterisk) were also identified. These terminal sites were considered as false positive and were removed for further consideration. Eventually, six sites (Trp21, Thr102, Thr120, Thr137, Ala166 and Asp176) were identified by Hy

![Fig. 1. Neighbor-Joining phylogenetic tree of GH11 xylanases (red, xylanases with acidic pH optima; grey, xylanases with neutral pH optima; blue, xylanases with alkaline pH optima). For each enzyme, both the numbering and the pH optimum were provided. The detailed information for each enzyme is listed in Appendix B.](image-url)
index and eight sites (Asp23, Leu48, Ser56, Trp100, Thr102, Ala108, Ser153 and Gln177) were identified by pI values (Appendix A, Table A1).

3.3. Design of the mutations

After identification of key sites in GH11 family, the next step is to investigate the mutation mode of each site. To figure out amino acid species with potential effects on pH optimum, we first summarized the amino acid abundance on each site in the family (Appendix A, Table A2). The amino acid abundance was further analyzed by online software Weblogo 3 to make the data visible. As shown in Fig. 2, enzymes with different pH optima exhibited different amino acid preference on some sites. For instance, on site 166, the alkaline enzymes preferred Glu and Arg while the acidic and neutral enzymes preferred Ala. However, on some sites, the amino acid species was consensus. For example, on site 21, Trp was the consensus residue for most of the enzymes.

By analysis of the amino acid abundance in GH11 family, we designed the mutations of CbX-CD with the expectation of shifting pH optimum. Firstly, the consensus residue at a particular site was regarded as optimized in GH11 family, so it was chosen for mutation if CbX-CD had a different residue other than the consensus one. Secondly, ionizable amino acids might have greater effect on pH optima than aliphatic ones, so the mutations to ionizable amino acids were preferable. Based on these two criteria, the design of each mutation site was listed as following (also summarized in Table 3):

For site 21, 23, 102, 120, 137, they showed strong consensus in GH11 family, and the consensus were the same as their corresponding residues in CbX-CD. So these sites were not considered for further mutagenesis. Site 100 had consensus (Trp) in neutral and alkaline enzymes, but was equally preferred as Tyr and Trp in acidic enzymes. Since site 100 in CbX-CD was also Trp and an aliphatic mutation was not likely change the pKa value too much, this site was not considered for mutagenesis. Site 108 and 153 had no obvious consensus and their top three amino acid species were all aliphatic or non-ionizable amino acids, so they were not considered for further mutagenesis.

Site 48 had obvious consensus as Asn in neutral and acidic enzymes, and was equally preferred as Lys and Asn in alkaline enzymes. In CbX-CD, site 48 was Leu, which was different from the consensus Asn. So we designed a mutation at this site, i.e. L48N.

Fig. 2. Amino acid distribution of 21/23/48/56/100/102/120/137/153/166/176/177 sites in all, alkaline, neutral, and acidic groups GH11 family xylanases. In the weblog chart, the word size of the amino acid was proportional to the abundance of this amino acid species in the database. If one site was dominated by one or two amino acids, this site was considered to be conserved (e.g. site 21 and 23). For comparison, amino acids at corresponding sites of CbX-CD were also listed on top of the chart.
Similar case were also found at site 166 and 177, and their mutations were designed as A166E and Q177E, respectively. Amino acids at site 56 had no consensus in the family. The top three amino acid preference at this site were Asn, Thr and Asp, respectively. According to the criteria, the Ser56 at CbX-CD was mutated to be an ionizable Asp (S56D). Similar considerations were posted on site 176, which was designed as D176Y. In order to check the accuracy of sequence-based algorithm, S137I and S153T consistent with consensus amino acids were also selected as negative controls. Therefore, five mutants with the potential change in pH adaptation were designed as L48N, S56D, A166E, D176Y and Q177E.

### 3.4. Biochemical characterization of CbX-CD xylanase mutants

The designed mutants of CbX-CD were generated and constructed into pET28a plasmid. All the recombinant proteins were abundantly expressed in E. coli with similar expression level with the wild type enzyme, suggesting that the mutations didn’t affect protein folding. After purification by Ni-NTA affinity chromatography, the molecular weight and the purity (95%) of recombinant proteins was evaluated by SDS-PAGE (Appendix A, Fig. A5).

The pH-activity profile of each enzyme was measured using beech wood xylan as substrate. Compared with the pH optimum of wild type CbX-CD ($pH_{opt} = 6.5$), S56D, A166E, D176Y and Q177E exhibited obvious acidic shift in pH-activity curves (Fig. 3), especially D176Y and Q177E had the shift up to 0.75 unit. Moreover, the overall pH-activity curve of each mutant shifted to the acidic side, indicating that the mutations do not just simply improve the enzymatic tolerance against acidic pH, but authentically altered the $pK_a$ values of the catalytic residue. On the other hand, L48N didn’t cause much change in the pH-activity profile (Appendix A, Fig. A.6). As expected, the control mutants S137I and S153T also didn’t change the pH optimum (Appendix A, Fig. A.6). Remarkably, all the positive mutants retained a considerable level of catalytic activity.

| AA$^a$ site   | Description of AA$^a$ distribution | AA$^a$ on CbX-CD | Mutation on CbX-CD | Effect on pH optimum | Possible mechanism on pH optimum shifting                     |
|---------------|-----------------------------------|-------------------|--------------------|----------------------|-------------------------------------------------------------|
| 21 Conserved  | residue W                        | None              | L                  | No shift             | -                                                           |
| 23 Conserved  | residue D                        | None              | D                  | No shift             | -                                                           |
| 48 Consensus  | as N or K                        | L                 | L48N               | -                    | -                                                           |
| 56 Prefer N, | T and D                          | S                 | S56D               | 0.5 unit acidic      | Deprotonating residues surrounding catalytic residues      |
| 100 Conserved | residue W                        | None              | L                  | -                    | -                                                           |
| 102 Conserved | residue T                        | None              | T                  | -                    | -                                                           |
| 108 No obvious | No obvious consensus              | A                 | A166E              | 0.5 unit acidic      | Same as S56D                                                |
| 120 Conserved | residue T                        | None              | T                  | -                    | -                                                           |
| 137 Conserved | residue T                        | None              | T                  | -                    | -                                                           |
| 153 No obvious | No obvious consensus              | S                 | None               | -                    | -                                                           |
| 166 Consensus | as E or A                        | A                 | A166E              | 0.5 unit acidic      | Same as S56D                                                |
| 176 Prefer N, | Y and D                          | D                 | D176Y              | 0.75 unit acidic     | Undefined indirect long-distance effects                    |
| 177 Consensus | as E or Y                        | Q                 | Q177E              | 0.75 unit acidic     | Deprotonating a residue adjacent to catalytic residues by a salt bridge |

$^a$ AA, amino acid.

Fig. 3. The pH-activity profile of the wide-type CbX-CD and its mutants. The activity data were obtained from triplicate (at least) assays using 1% (w/v) beech wood xylan as substrate at 70 °C and defined pH ranging from 4.0 to 8.0. The activity data were obtained from triplicate (at least) assays using 1% (w/v) beech wood xylan as substrate at 70 °C.
Next, the positive mutants were combined to further explore their functions on pH adaptation. Because mutants S56D, A166E and Q177E introduced charged amino acid substitutions, they were firstly taken as a group for the combination and resulted in four mutants: M21 (A166E/Q177E), M22 (S56D/Q177E), M23 (S56D/A166E/Q177E) and M24 (S56D/A166E). Enzymatic characterization showed that their pH-activity profiles further shifted to the acidic side (Fig. 3). Later, D176Y was introduced and resulted in a mutant M31 (S56D/A166E/D176Y/Q177E) with pH optimum of 5.0, which was 1.5 units shift towards acidic side compared to the wild type (Fig. 3).

The enzymatic property changes along with molecular design process was shown in Fig. 4. The best mutant M31 maintained approximately 90% maximal activity of the wild type. While at pH 5.0, it exhibited even higher specific activity (3500 U/mg) than the wild type enzyme (3100 U/mg). In fact, several of the mutants showed a higher catalytic activity at pH 4.0 than the wild type enzyme (Appendix A, Fig. A.7), which clearly showed the efficiency of our design. In addition, same as the wild type enzyme, all the mutants showed high thermostability that with half-lives more than 24 h at 70 °C (data not shown). Therefore, we obtained the mutants possessing altered pH optima without sacrificing catalytic activity and thermostability.

3.5. Calculations of the pKa changes

The enzymatic pH optimum is determined by the pKs values of the key catalytic residues. To accurately measure the pKs changes of the catalytic residues (Glu93 and Glu183), the ΔpKs was validated by comparing the experimental ΔpKa values of the wild-type CbX-CD and its mutants. According to the pH-activity profile of wild type and mutants, the apparent pKs1 and pKs2 were calculated by non-linear fitting using Eq. (1). As shown in Table 4, the pKs1 of four single-site mutants S56D, A166E, D176Y and Q177E decreased 0.2, 0.2, 0.28, 0.15, 0.35 pH unit, respectively. Whereas only D176Y and Q177E had obvious decline in pKs2, with 0.49 and 0.27 unit shift, respectively. Combinational mutants exhibited similar shift in pKs1. However, except M24 (S56D/A166E), other four mutants showed obvious shift in pKs2 for 0.66, 0.66, 0.67 and 0.49 pH units, respectively. This was consistent with the observation that combinatorial mutagenesis further changed the enzymatic pH optima.

3.6. Structural modeling and mutational analysis

The modeling structure of CbX-CD was constructed by SWISS-MODEL using Rt46B.1 from Dictyoglomus thermophilum as a template (PDB ID IF5J, 88% sequence identity). The quality of the modelled structure was analyzed by PROCHECK and Profile-3D (data not shown). Calculated Ramachandran plot suggested 98% and 2% residues in the derived model are in favored and allowed regions. To further validate the reliability of this modelled structure, we performed a parallel modeling using another xylanase from Bacillus sp. 41M – 1 (PDB ID 2DCJ) with lower identity of ~64%. The comparison of two structures was illustrated in Appendix A, Fig. A.8. The two structures exhibited remarkable consistency with each other (with a root mean square deviations (RMSD) of 0.63 Å), confirming the reliability of the modelled structure of CbX-CD. The model of CbX-CD showed a canonical β-jelly roll structure, the acid/base catalyst is Glu183 (pKs 7.20) and the nucleophile is Glu93 (pKs 5.36), the pKs value were predicted using Propka software. Nucleophile Glu93 was surrounded by a series of protonated residues, including Arg50 (pKa 14.6, distance between Arg50NH1 and Glu93OE2 is 6.9 Å), Tyr84 (pKa 16.52, distance between Tyr84OH and Glu93OE2 is 2.6 Å), Tyr95 (pKa 21.30, distance between Tyr95OH and Glu93OE2 is 4.0 Å) and Arg128 (pKa 13.15, distance between Arg128NH1 and Glu93OE1 is 3.0 Å). Those protonated residues form hydrogen bonds and salt bridges with nucleophile Glu93, thus could stabilize the deprotonated state of Glu93. This interaction might be in favor of nucleophile attack of Glu93 during the catalysis. Comparing to nucleophile Glu93, only 2 hydrogen bonds were formed between Glu183-Ans46 and Glu83-Tyr95, which endowed Glu183 suitable pKa as acid/base catalyst (Fig. 5).

The structures of the mutants were constructed by similar method. Superimposing the mutant structure onto the wild type structure revealed that the root mean square deviations (RMSD) were less than 0.01 Å (data not shown). In spite of L48N, all the other mutations were close to each other, located on the protein surface, and far away from catalytic residues. The L48N mutation
located near the acid/base catalyst 183, with distance between Asn48ND2 to Glu183OE1 ~ 4.4 Å. The substitution of Ser56 by Asp forms a new hydrogen bond between Asp56OD1 and Arg88NH1 (3.1 Å). And A166E mutation introduced a charged side chain, which resulted in the formation of a new hydrogen bond Glu166OE2 - Leu172NH (2.6 Å). The original hydrogen bond Asp176OD2 - Gln177NE2 was replaced by Glu177OD2 - Lys54NZ (3.6 Å) due to the Q177E mutation (Fig. 5).

4. Discussion

Employing protein engineering to shift enzymatic pH optima would facilitate the highly adaption between catalytic parts and further improve the operating efficiency of cell-free synthetic
Cell-free synthetic biology systems often involve non-physiological pH conditions, therefore, reprogramming pH adaption of enzymes has always been an important goal. Since pH optimum is mostly governed by the ionization states of the side chains of the catalytic residues, common strategies for changing pH-activity profiles was done by introducing mutations around the active site based on structural analysis [15,35,51,52] and computational predictions [33,53]. Although these studies have successfully identified some mutations that shifting enzymatic pH optimum, there is risk of losing catalytic activity to introduce mutations near the catalytic residues. Besides, low accuracy is usually happened during the design of pH optimum. Qiu et al. tailored the pH-active profile through replacing its surface charged residues, only three out of nine candidates showed shift pH optimum [54]; Yang et al. changed pH optimum of Bacillus circulans xylanase based on molecular modeling and sequences alignment, three out of six mutants showed changes in pH optimum, but each mutant losing almost 70% of the wild-type activity [55]. Impeded by the difficulty of rational design, most successful cases on shifting enzymatic pH optima were based on random mutagenesis and screening [56–60], which is also quite limited because it could not provide a general guideline for pH engineering.

Recently, with the dramatic accumulation of genomic sequencing data, successful molecular engineering cases guided by natural evolutionary information are booming [61–64]. In this study, we used abundant sequence information of GH11 xylanase family for the design of pH adaptation for the first time. By means of ANNs and Lasso linear regression analyses of a well-characterized, digitalized GH11 database, we identified several key residues with significant effects on the pH optimum of a thermophilic xylanase CbX-CD. The biostatistics method developed in this study possesses high success rate, in which four positive sites out of five significantly shift the enzymatic pH optimum. Moreover, since our method is based on natural sequence analyses, the mutation sites designed for each site are all naturally occurred in GH11 family and thus they are relatively unlikely compromising the enzymatic activity. Most importantly, prediction based on this method doesn’t rely on the structural information, making the extracted information is applicable to many other xylanase family members.

Xylanases are always important models for investigating pH adaption due to their industrial valuable applications. In previously reported works, sites with effects on enzymatic pH optimus are usually identified close to the active center. While in our work, by employing biostatistics method, we have identified four amino acid sites which are distant from active center, but have significant effects on enzymatic pH optimum. Indeed, data from this study showed that CbX-CD mutants with obvious shift to acidic limb could be generated by mutations far from the active site. All four mutation sites (S56D, A166E, D176Y and Q177E) are dispersed on the protein surface and influenced enzymatic pH optimum by different mechanism, which cannot be easily identified by structural based rational design approaches.

Replacement of Ser56 by Asp exhibits only modest influence on the enzyme backbone structure. The mutation introduces an extra hydrogen bond between Asp56 and Arg88, which is far away from catalytic Glu93 and Glu183. Therefore, the change of pH optimum is not directly caused by this hydrogen bond. Because electrostatic effects decrease proportional to the reciprocal of the radius, change of pH optimum may still affect the pK_a value of key residues. Moreover, the distance between Asp56OD1 to Glu93OE2 and Glu183OE1 is 13.2 Å and 17.3 Å, respectively. The elimination of charged side chain abolished the electrostatic repulsion between catalytic residues (Glu93 and Glu183) might cause pK_a raise of catalytic residues. In addition, a series of protonated residues around catalytic residues Glu93, including Arg50, Tyr84, Tyr95 and Arg128 (the distance between electrostatic interactions is 14.6 Å, 16.6 Å, 21 Å and 20.9 Å, respectively). The electrostatic attraction between Asp56 and these residues would endow them a more stable deprotonation state, which might decrease the pK_a value of Glu93 and Glu183. We reasoned that the latter effect is more remarkable than the former one because of its short distance and more interactions. Similar to the mutation Ser56Asp, Ala166Glu mutation introduced charged amino acid far away from active site, which might also provide more stable protonation state of protonated residues around catalytic amino acid and showed similar change in pK_a (0.28 unit decline) and pK_a values (0.03 unit decline).

To introduce Tyr at position 176 substitute the charged Asp side chain to Tyr. The distance from Tyr176OH to Glu9302E and Glu18301E is 13.2 Å and 17.3 Å, respectively. The elimination of charged side chain abolished the electrostatic repulsion between Asp176 and catalytic residues (Glu93 and Glu183), while increase the pK_a values of catalytic amino acids. Interestingly, distance between Tyr1760H and Glu18301E is longer than the distance between Tyr1760H and Glu9302E, which means that the influence of mutation toward Glu93 should be more remarkable. However, according to the results of ΔpK_a calculations, small extent change in pK_a (0.15 unit decline) but big extent change in pK_a values (0.49 unit decline) is evident, which demonstrated that the mutation may change the pH adaption by some other indirect long-distance effects or by some unexpected interactions that we did not consider.
not find.

The Q177E mutation shift enzymatic pH optimum downward significantly. The distance from Glu177OE2 to Glu930E2 and Glu183OE1 is 8.1 Å and 10.7 Å, respectively. The weak electrostatic repulsion between Glu177 and catalytic residues (Glu93 and Glu183) may cause the pKa upward shift. However, the mutation introduced a strong salt bridge between Glu177 and Arg50 (the distance between Glu177OE2 to Arg50NH2 is 2.8 Å), which might stabilize the protonated state of Arg50. Furthermore, Arg50 is the nearest positive charge around catalytic residues, the distance between Arg50NH2 and Glu930E2 and Glu1930E1 is 7.2 Å and 8.1 Å, respectively. The stronger protonated state of Arg50 further stabilizes the deprotonated state of catalytic residues, and results in modest change in pK_{a1} (0.35 unit decline) and pK_{a2} values (0.27 unit decline).

Above mentioned single point mutations altered enzyme pH optimum in different ways (also see Table 3 for summarization). F55W, S56D and A166E major downward shift toward pK_{a1}, D176Y and Q177E downward change both in pK_{a1} and pK_{a2}. Further investigation revealed multiple mutations influenced both pK_{a1} and pK_{a2} simultaneously, which means pK_{a} shift guided by different mechanism could work collaboratively.

In this study, we constructed an elaborate GH11 xylanase database with pH annotation, and developed a data driven protein engineering strategy to redesign the pH adaptation of xylanase. Based on the analysis of ANNs and Lasso, four out of five mutation sites located far away from catalytic residues showed modest capability in pH optimum shift, highlighting the robustness of this data driven protein engineering strategy. This would also be helpful in further understanding the pH regulation mechanism of this important enzyme family. However, it should be noticed that this biostatistics method rely on well-characterized data for a large number of enzymes, which is still lacking for most of the enzyme families. Therefore, besides further enriching sequence information of enzyme families, developing high throughput techniques for expression and characterization of new enzymes would also make this method feasible and consummate.

Acknowledgments

This work is supported by National High Technology Research and Development Program of China (863 Program, 2013AA102801), the National Basic Research Program of China (973 Program, and Development Program of China (863 Program, 2013AA102801).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.synbio.2016.09.001.

References

[1] Hodgman CE, Jewett MC. Cell-free synthetic biology: thinking outside the cell. Metab Eng 2012:14:261–9.
[2] Harris DC, Jewett MC. Cell-free biology: exploiting the interface between synthetic biology and synthetic chemistry. Curr Opin Biotechnol 2012:23: 67–8.
[3] Smith MT, Wilding KM, Hunt JM, Bennett AM, Bundy BC. The emerging age of cell-free synthetic biology. FEMS Lett 2014:588:2755–61.
[4] Zawada JF, Yin G, Steiner AR, Yang J, Naresh A, Roy SM, et al. Microscale to manufacturing scale-up of cell-free cytokine production—a new approach for shortening protein production development timelines. Biotechnol Bioeng 2011:108:1570–8.
[5] Wang HH, Huang P-Y, Xu G, Haas W, Marblestone A, Li J, et al. Multiplexed in vivo his-tagging of enzyme pathways for in vitro single-pot multienzyme catalysis. ACS Synth Biol 2012:1:43–52.
[6] Murray CJ, Baliga R. Cell-free translation of peptides and proteins: from high throughput screening to clinical production. Curr Opin Chem Biol 2013:17: 420–6.
[7] Furuya T, Muria M, Kino K. A coenzyme-independent decarboxylase/oxygenase cascade for the efficient synthesis of vanillin. ChemBioChem 2014:15: 2248–54.
[8] Muschol J, Peters C, Oberleitner N, Mihovilovic MD, Bornscheuer UT, Rudloff F. Cascade catalysis—strategies and challenges en route to preparative synthetic biology. Chem Commun 2015:51:3798–811.
[9] Kuchner O, Arnold FH. Directed evolution of enzyme catalysts. Trends Biochem Technol 1997:15:523–30.
[10] Hult K, Berglund P. Engineered enzymes for improved organic synthesis. Curr Opin Biotechnol 2003:14:395–802.
[11] Jaeger K-E, Eggert T. Enantioselective biocatalysis optimized by directed evolution. Curr Opin Biotechnol 2004:15:305–13.
[12] Bloom JD, Meyer MM, Meinhold P, Otey CR, MacMillan D, Arnold FH. Evolving strategies for enzyme engineering. Curr Opin Biotechnol 2005:15:447–52.
[13] Qu S, Lai L. Tailoring the pH dependence of human non-pancreatic secretory phospholipase a2 by engineering surface charges. Appl Biochem Biotechnol 2013:171:1454–64.
[14] Tishkov VI, Guskov AV, Cherkashina AS, Sinitsyn AP. Engineering the pH optimum of activity of the gh12 family endoglucanase by site-directed mutagenesis. Biochimie 2013:95:1704–10.
[15] Ohara K, Unno H, Oshima Y, Hosoya M, Fujino N, Hoiruka K, et al. Structural insights into the low pH adaptation of a unique carboxylesterase from fer- roplasma altering the pH optima of two carboxylesterases. J Biol Chem 2014:289:24499–510.
[16] Shi F, Xie Y, Jiang J, Wang N, Li Y, Wang X. Directed evolution and mutagenesis of glutamate decarboxylase from Lactobacillus brevis (LbB) to broaden the range of its activity toward a near-neutral pH. Enzyme Microb Technol 2014:61:35–43.
[17] Li F, Xie J, Zhang X, Zhao L. Improvement of the optimum pH of aspergillus Niger xylanase towards an alkaline pH by site-directed mutagenesis. J Microbiol Biotechnol 2015:25:11–7.
[18] Tynan-Conolly BM, Nielsen JE. Pkd: Re-designing protein pka values. Nucleic acids res 2006:34:W48–51.
[19] Tynan-Conolly BM, Nielsen JE. Redesigning protein pK_{a} values. Protein Sci 2007:16:239–49.
[20] Sugino M, Kajita S, Banno K, Shirai T, Yamane T, Kato M, et al. Upward shift of the pH optimum of acermonium arscobate oxidase. Biochimica Biophysica Acta (BBA)-General Subj 2004:1672:174–83.
[21] Zhou Y, Liu L, Liu W, Yang Q, Yang S, Jiang Z. Engineering a thermostable 1, 3, 1, 4-glucanase from Paeucylomyces thermophilia to improve catalytic efficiency at acidic pH. J Biol Chem 2012:159:50–5.
[22] Fan C, Xu W, Zhang T, Zhou L, Jiang B, Wu M. Engineering of alicyclobacillus hesperidium l-arabinose isomerase for improved catalytic activity and reduced pH optimum using random and site-directed mutagenesis. Appl Biochem Biotechnol 2015:177:1480–92.
[23] Shirai T, Ishida H, J-M Noda, Yamane T, Otsuki K, Hakamada Y, et al. Crystal structure of alkaline cellulase 9: insight into the alkaline adaptation of an industrial enzyme. J Mol Biol 2001:310:1079–87.
[24] Yasuda T, Takehisa H, Iida R, Ueki M, Nakajima T, Kaneko Y, et al. A single aspartic acid substitution can shift the optimum pH of dnanase for enzyme ac- tivity: biochemical and molecular analysis of the piscine dnanase i family. Biochimica Biophysica Acta (BBA)-Protein Struct Mol Enzym 2002:1596:36–46.
[25] Jia H, Li Y, Liu Y, Yan Q, Yang S, Jiang Z. Engineering a thermostable 1, 3, 1, 4-glucanase from Paeucylomyces thermophilia to improve catalytic efficiency at acidic pH. J Biol Chem 2012:159:50–5.
[26] Fan C, Xu W, Zhang T, Zhou L, Jiang B, Wu M. Engineering of alicyclobacillus hesperidium l-arabinose isomerase for improved catalytic activity and reduced pH optimum using random and site-directed mutagenesis. Enzyme Microb Technol 2015:61:82–92.
[27] Shirai T, Ishida H, J-M Noda, Yamane T, Otsuki K, Hakamada Y, et al. Crystal structure of alkaline cellulase 9: insight into the alkaline adaptation of an industrial enzyme. J Mol Biol 2001:310:1079–87.
[28] Yasuda T, Takehisa H, Iida R, Ueki M, Nakajima T, Kaneko Y, et al. A single aspartic acid substitution can shift the optimum pH of dnanase for enzyme ac- tivity: biochemical and molecular analysis of the piscine dnanase i family. Biochimica Biophysica Acta (BBA)-Protein Struct Mol Enzym 2002:1596:36–46.
[29] Jia H, Li Y, Liu Y, Yan Q, Yang S, Jiang Z. Engineering a thermostable 1, 3, 1, 4-glucanase from Paeucylomyces thermophilia to improve catalytic efficiency at acidic pH. J Biol Chem 2012:159:50–5.
[30] Fan C, Xu W, Zhang T, Zhou L, Jiang B, Wu M. Engineering of alicyclobacillus hesperidium l-arabinose isomerase for improved catalytic activity and reduced pH optimum using random and site-directed mutagenesis. Appl Biochem Biotechnol 2015:177:1480–92.
bonding and catalysis: a novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase. J Mol Biol 2000;299:255–79.

[33] Xu H, Zhang F, Shang H, Li X, Wang J, Qiao D, et al. Alkalophilic adaptation of xyln endoxylanase from Aspergillus Niger via rational design of pk_s of catalytic residues. J Biosci Bioeng 2013;115:618–22.

[34] de Lemos Esteves F, Ruelle V, Lamotte-Brasseur J, Quinting B, Frere JM. Acidophilic adaptation of family 11 endo-beta-1,4-xylanases: modeling and mutational analysis. Protein Sci Publ Protein Soc 2004;13:1209–18.

[35] De Lemos Esteves F, Gouders T, Lamotte-Brasseur J, Rigali S, Frere JM. Improving the alkalophilic performances of the xy1 xylanase from Streptomyces sp. S38: structural comparison and mutational analysis. Protein Sci Publ Protein Soc 2005;14:292–302.

[36] Qiu J, Han H, Sun B, Chen L, Yu C, Peng R, et al. Residue mutations of xylanase from Bacillus circulans endoxylanase improves its acid stability. Protein Eng Des Sel 2009:gzp024.

[37] Yang JH, Park JY, Kim SH, Yoo YJ. Shifting pH optimum of bacillus circulans xylanase based on molecular modeling. J Biotechnol 2008;135:190–5.

[38] Joshi MD, Sidhu G, Pot I, Brayer GD, Withers SG, McIntosh LP. Hydrogen bonding and catalysis: a novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase. J Mol Biol 2000;299:255–79.

[39] Lombard V, Ramulu HG, Drula E, Coutinho PM, Henriques B. The carbohydrate-active enzymes database (cazy) in 2013. Nucleic Acids Res 2014;42:D490–5.

[40] Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, et al. Multiple sequence alignment with the clustal series of programs. Nucleic Acids Res 2003;31:3497–500.

[41] Robert X, Gouet P. Deciphering key features in protein structures with the new endscript server. Nucleic Acids Res 2014;42:W320–4.

[42] Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157:105–32.

[43] Belien T, Joyce IJ, Delcour JA, Courtin CM. Computational design-based molecular engineering of the glycosyl hydrolase family 11 b. Subtilis xyna endoxylanase improves its acid stability. Protein Eng Des Sel 2009:gpz024.

[44] Xie Y, An J, Yang CT, Wu G, Zhang Y, Cui L, et al. Enhanced enzyme kinetic stability by increasing rigidity within the active site. J Biol Chem 2014;289:7994–8006.

[45] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. Procheck: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–91.

[46] Paes G, Berrin JG, Beaurgard J. Ghull xylanases: structure/function/properties relationships and applications. Biotechnol Adv 2012;30:564–92.

[47] Szaleniec M. Prediction of enzyme activity with neural network models based on electronic and geometrical features of substrates. Pharmaco Rep 2012;64:761–81.

[48] Manning T, Sleator RD, Walsh P. Biologically inspired intelligent decision making: a commentary on the use of artificial neural networks in bioinformatics. Bioengineered 2014;5:80–95.

[49] Li Z, Sillanpaa MJ. Overview of lasso-related penalized regression methods for quantitative trait mapping and genomic selection. Theor Appl Genet 2012;125:419–35.

[50] Pasanen L, Holstrom L, Sillanpaa MJ. Bayesian lasso, scale space and decision making in association genetics. PLoS One 2015;10:e0120017.

[51] Kim T, Mullaney EJ, Pores JM, Roneker KR, Crowe S, Rice S, et al. Shifting the pH profile of aspergillus Niger phyo phytase to match the stomach pH enhances its effectiveness as an animal feed additive. Appl Environ Microbiol 2006;72:4397–403.

[52] Shiraki K, Sakiyama F. Histidine 210 mutant of a tryps in-type achromobacter protease i shows broad optimum pH range. J Biosci Bioeng 2002;93:331–3.

[53] Bai WQ, Zhou C, Zhao YJ, Wang QH, Ma YH. Structural insight into and mutational analysis of family 11 xylanases: implications for mechanisms of higher pH catalytic adaptation. PLoS One 2015;10.

[54] Qiu S, Lai T. Tailoring the pH dependence of human non-pancreatic secretory phospholipase a2 by engineering surface charges. Appl Biochem Biotechnol 2013;171:1454–64.

[55] Yang JH, Park JY, Kim SH, Yoo YJ. Shifting pH optimum of bacillus circulans xylanase based on molecular modeling. J Biotechnol 2008;133:294–300.

[56] Qin YQ, Wei XM, Song X, Qu YB. Engineering endoglucanase ii from trichoderma reesei to improve the catalytic efficiency at a higher pH optimum. J Biotechnol 2008;135:190–5.

[57] Fan C, Xu W, Zhang T, Zhou L, Jiang B, Mu W. Engineering of alcyocobacillus hesperidum l-arabinose isomerase for improved catalytic activity and reduced pH optimum using random and site-directed mutagenesis. Appl Biochem Biotechnol 2015;177:1480–92.

[58] Iba H, Li Y, Liu Y, Yan Q, Yang S, Jiang Z. Engineering a thermostable beta-1,3-1,4-glucanase from paecilomyces thermomorphia to improve catalytic efficiency at acidic pH. J Biotechnol 2012;159:50–5.

[59] Sugino M, Kajita S, Banno K, Shira T, Yamane T, Kato M, et al. Upward shift of the ph optimum of acronemum ascorbate oxidase. Biochimica Biophysica Acta-PROtein Struct Mol Enzym 2002;1596:36–46.

[60] Balci H, Ozturk MT, Pijning T, Ozturk SI, Gumusel F. Improved activity and pH stability of E. coli atcc 11105 penicillin acylase by error-prone PCR. Appl Microbiol Biotechnol 2014;98:4467–77.

[61] Kourist R, Jochens H, Bartsch S, Kupfers R, Padhi SK, Gall M, et al. The xj-hydrolase fold 3dm database (abhdb) as a tool for protein engineering. ChemBioChem 2010;11:1635–43.

[62] DiTursi MK, Kwon S-J, Reeder PJ, Dordick JS. Bioinformatics-driven, rational engineering of protein thermostability. Protein Eng Des Sel 2006;19:517–24.

[63] Suplatov D, Voevodin V, Zerbs S, Giuliani S, Collart F. Small-scale expression of proteins in E. coli with high-throughput. Methods Mol Biol 2014;1091:33.

[64] Lenfant N, Hotelier T, Velluet E, Bourne Y, Marchot P, Chatonnet A. Esther, the database of the zj-hydrolase fold superfamilly of proteins: tools to explore diversity of functions. Nucleic Acids Res 2013;41:D423–5.

[65] Camilo CM, Polikarpov I. High-throughput cloning, expression and purification of glycosid hydrolases using ligation-independent cloning (lic). Protein Expr Purif 2014;99:35–42.

[66] Lenfant N, Hotelier T, Velluet E, Bourne Y, Marchot P, Chatonnet A. Esther, the database of the zj-hydrolase fold superfamilly of proteins: tools to explore diversity of functions. Nucleic Acids Res 2013;41:D423–5.

[67] Camilo CM, Polikarpov I. High-throughput cloning, expression and purification of recombinant proteins in E. coli. Methods Mol Biol 2014;1091:33–53.

[68] Vincentelli R, Romier C. Complex reconstitution and characterization by combining co-expression techniques in Escherichia coli with high-throughput. Adv Exp Med Biol 2016;896:43–58.

[69] Zerbs S, Giuliani S, Collart F. Small-scale expression of proteins in E. coli. Methods Enzym 2014;536:117–31.

[70] Saez NJ, Nozach B, Blemont M, Vincentelli R. High throughput quantitative expression screening and purification applied to recombinant disulfide-rich venom proteins produced in E. coli. J Vis Exp JoVE 2014:e51464.

[71] Festa F, Steel J, Bian X, Labaer J. High-throughput cloning and expression library creation for functional proteomics. Proteomics 2013;13:1381–99.