Protein consensus-based surface engineering (ProCoS): a computer-assisted method for directed protein evolution

Amol V. Shivange^{1,2,†}, Hans Wolfgang Hoeffken³, Stefan Haefner³, and Ulrich Schwaneberg^{1,2}

¹Lehrstuhl für Biotechnologie, RWTH Aachen University, Aachen, Germany, ²School of Engineering and Science, Jacobs University Bremen, Bremen, Germany, and ³BASF SE, Fine Chemicals and Biocatalysis Research, Ludwigshafen, Germany

[†]Present address: Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA

BioTechniques 61:305–314 (December 2016) doi 10.2144/000114483

Keywords: protein engineering method; protein surface engineering; directed evolution; ProCoS; phytase; pH stability; mutagenesis

Supplementary material for this article is available at www.BioTechniques.com/article/114483.

Protein consensus-based surface engineering (ProCoS) is a simple and efficient method for directed protein evolution combining computational analysis and molecular biology tools to engineer protein surfaces. ProCoS is based on the hypothesis that conserved residues originated from a common ancestor and that these residues are crucial for the function of a protein, whereas highly variable regions (situated on the surface of a protein) can be targeted for surface engineering to maximize performance. ProCoS comprises four main steps: (i) identification of conserved and highly variable regions; (ii) protein sequence design by substituting residues in the highly variable regions, and gene synthesis; (iii) in vitro DNA recombination of synthetic genes; and (iv) screening for active variants. ProCoS is a simple method for surface mutagenesis in which multiple sequence alignment is used for selection of surface residues based on a structural model. To demonstrate the technique's utility for directed evolution, the surface of a phytase enzyme from *Yersinia mollaretii* (Ymphytase) was subjected to ProCoS. Screening just 1050 clones from ProCoS engineering-guided mutant libraries yielded an enzyme with 34 amino acid substitutions. The surface-engineered Ymphytase exhibited 3.8-fold higher pH stability (at pH 2.8 for 3 h) and retained 40% of the enzyme's specific activity (400 U/mg) compared with the wild-type Ymphytase. The pH stability might be attributed to a significantly increased (20 percentage points; from 9% to 29%) number of negatively charged amino acids on the surface of the engineered phytase.

Directed protein evolution is a well-established, versatile, and successful algorithm for tailoring protein properties to industrial demands and advancing our understanding of structure–function relationships in biocatalysts (1,2). Directed evolution entails the accumulation of beneficial mutations in iterative cycles of mutagenesis and screening or selecting for improved enzyme variants. These accumulations of mutations mostly result in a downhill

path on the fitness landscape (fitness versus sequence) plot. A downhill mutational path eventually leads to an unfolded or inactive enzyme variant (3). It is becoming clear that there exist two pathways for directed evolution: (i) a widely known pathway of accumulating single amino acid changes at each cycle of mutagenesis and screening to select an improved enzyme (4) and (ii) a pathway in which cooperative effects between a combination of mutations

lead to synergistic or additive improvements (5–8).

In practice, substitutions resulting from cooperative effects are barely reported or studied. Recently, we showed that knowledge-based combinations of key residue substitutions identified in directed evolution yielded an improved enzyme (6). Traditional directed evolution approaches for improving enzyme properties are laborious and involve multiple steps of mutagenesis and screening. Screening

METHOD SUMMARY

Protein consensus-based surface engineering (ProCoS) is a simple method in which multiple sequence alignment is used to select surface residues based on a structural model. Synthetic gene variants are designed for surface mutagenesis and synthesized commercially. A mutant library generated by PCR-based in vitro recombination of synthetic genes is screened for active variants.

more than 9000 clones of random mutant libraries identified 5 key positions in a phytase (9), and the subsequent multi-site saturation mutagenesis of these sites and screening (1100 clones) yielded a pH-stable phytase variant (7). A hydrolase was evolved for higher pH stability and thermostability using error-prone PCR and DNA shuffling by screening >45,000 clones (10). Therefore, knowledge-based methods using sequence alignments and structural information are becoming an attractive alternative.

Here, we present a computer-assisted method for surface engineering of proteins employing sequence alignment and structural analysis to design and screen mutant libraries. Our goal was to determine if a large number of mutations could be incorporated into a protein to engineer its surface while also maintaining its functionality. This presumes that incorporation of a higher number of mutations in a protein will increase the probability of cooperative effects between the substitutions in the mutant library and could yield a functional protein due to increased functional diversity of the libraries.

Re-engineering the protein surface has become an interesting tool due to its wide applications in therapeutic protein delivery (11,12), immobilization (13–16), solubilization (17,18), stabilization of proteins in aqueous (19) or organic solvents (20), and preservation of enzyme activity in ionic liquids (21). Over the past few years, it has become apparent that modification of protein surface charge is a viable strategy for enhancing protein stability (22). Several computationally designed proteins with altered surface charge–charge interactions showed improvements in thermostability (23–25). A rational modification of 18 surface residues of carbonic anhydrase yielded a variant with extreme halotolerance that is active at >3 M NaCl (26). A replacement of charged residues on the surface with hydrophobic residues improved stability of a protease in organic solvents (20). Other approaches for protein surface engineering, including chemical modifications (amination or succinylation) (13), coupling of polymers (PEGylation) (11), and fusion of polypeptides (PASylation) (27), have been developed to alter the surface properties of proteins.

Here, we present protein consensus-based surface engineering (ProCoS), a

Figure 1. Schematic representation of protein consensus-based surface engineering (ProCoS). The workflow comprises four main steps: (i) identification of conserved and highly variable regions; (ii) protein sequence design by substituting residues in the highly variable regions, followed by gene synthesis; (iii) in vitro DNA recombination of synthetic genes; and (iv) mutant library screening for active variants.

method that can be used to incorporate >30 amino acid substitutions simultaneously in an enzyme to produce surface modifications. A computational analysis based on conservation of amino acids was used to identify functionally important regions and highly variable regions in a model protein, phytase from *Yersinia mollaretii* (Ymphytase). A combination of computational and molecular biology tools was used for surface engineering of Ymphytase. The Ymphytase variant retained ~40% of the wild-type's phytase activity after incorporation of 34 amino acid changes located on the surface of the protein. Interestingly, the pH stability of Ymphytase was improved 3.8-fold (pH 2.8) compared with the wild-type, which might be due to the significant increase (20 percentage points; from 9% to 29%) in negatively charged surface substitutions in the identified Ymphytase variant.

Materials and methods

Identification of conserved residues

As a first step toward identification of functionally important residues, available phytase amino acid sequences from the Enterobacteriaceae family of bacteria were retrieved from the ExPASy proteomics server (www.expasy.org). A total of 25 phytase enzyme sequences were obtained from different genera of the Enterobacteriaceae family of bacteria,

including *Escherichia* (6), *Yersinia* (7), *Klebsiella* (5), *Pectobacterium* (1), *Shigella* (1), *Obesumbacterium* (2), and *Citrobacter* (3). Sequence alignment of all 25 sequences was performed with VectorNTI suite 10 software (Invitrogen, Darmstadt, Germany) using the blosum62mt2 matrix with a gap-opening penalty of 10 and a gap-extension penalty of 0.05. A phylogenetic tree was built by applying the neighbor joining method implemented in VectorNTI to the sequence alignment of 25 Enterobacteriaceae phytases. The consensus sequence was generated by the AlignX module of VectorNTI (used for multiple sequence alignment). Amino acid residues that are conserved (blue), similar (green), or identical (yellow) in the multiple sequence alignment are shown in the consensus sequence (Supplementary Figures S1 and S2). The consensus sequence was used to identify conserved residues in the Enterobacteriaceae family. These sets of amino acids (blue, green, and yellow) were considered to be functionally important residues, whereas the residues colored in white (non-conserved) are considered variable regions.

Selection and substitution of sites

Amino acid sites in Ymphytase were selected based on the conservation of each residue within the Enterobacteriaceae family of bacteria. The protein

sequence was divided into functionally important regions (conserved residues) and variable regions (non-conserved). Each residue in the variable region of the sequence alignment was analyzed for the frequency of occurrence in all of the Enterobacteriaceae species included in this study. The positions of these residues were visualized in a homology model of Ymphytase (9) using VMD software (28). Residues belonging to loops and surface regions were selected. Selected residues were substituted with other amino acids based on three criteria: (i) frequently occurring residues in the sequence alignment were identified and selected for substitution; (ii) chemically similar amino acids were preferred, with swapping of the charged residues performed to avoid charge accumulation in a few areas; and (iii) sterically favorable residues were favored.

Gene synthesis and subcloning

Three synthetic genes were designed based on the above criteria, and codon optimization was performed using the GeneDesign server for *E. coli* expression (29). Ymphytase gene (*YmappA*) variants with restriction sites (*NdeI* at the 5' end and *NotI* at the 3' end) were synthesized commercially at GENEART (Regensburg, Germany). All synthetic constructs were digested with *NdeI* and *NotI* restriction enzymes (New England BioLabs). Digested synthetic genes were cloned into the pET-22b(+) vector (Merck Chemicals GmbH, Darmstadt, Germany) and transformed into the *E. coli* BL21-Gold(DE3) strain (Agilent Technologies Deutschland GmbH, Waldbronn, Germany) for protein expression.

In vitro DNA recombination

A PCR-based DNA recombination protocol was used to recombine the three synthetic genes of the *YmappA* variants. The template for recombination was generated from each pET-22b(+)-*YmappA* synthetic gene construct (1 ng/ μ L) by PCR (50 μ L reaction volume) (98°C for 3 min; 25 cycles of 98°C for 10 s, 58°C for 15 s, 72°C for 25 s; and then 72°C for 3 min) using the pET-22b(+) vector-specific primers F1 (5'-CGA CTC ACT ATA GGG GAA TGT TGA GCG GA-3') and R3 (5'-CGG GCT TTG TTA GCA GCC GGA TCT CAG-3') (0.4 μ M each), *Pfu* DNA polymerase (0.025 U/ μ L), and dNTP mix (0.2 mM each) in thin-walled PCR tubes. All PCR products were methylated using 8 U *dam* methyltransferase (New England BioLabs, Frankfurt, Germany) and column purified (NucleoSpin Extract II Kit; MACHEREY-NAGEL, Düren, Germany).

Vector-specific primers F1 and R3 were used for in vitro recombination. Each generated template was mixed together in equimolar amounts. Three PCRs were performed using different annealing/extension times at 55°C. PCR was performed using the amplified template (20 ng), 0.15 μ M of each primer (F1 and R3), 1 $\times$ *Taq* buffer, dNTP (0.2 mM each), and 2.5 U *Taq* polymerase. The PCR program consisted of 94°C for 30 s (denaturation), and 55°C for 1 s, 5 s or 10 s (annealing/extension), performed on a Mastercycler gradient (Eppendorf AG, Hamburg, Germany). PCR products (~1.4 kb) were gel-extracted and purified with the NucleoSpin Extract II Kit. PCR products obtained with 5 s and 10 s annealing/extension times were mixed together (Supplementary Figure S3). Following the PCR, 20 U *DpnI* (New

England BioLabs) was added, and the mixture was incubated overnight at 37°C and then column purified. Both purified PCR products (1 s annealing/extension time or combined 5/10 s annealing/extension times) were digested separately with the *NdeI* and *NotI* restriction enzymes. Digested PCR products were cloned into the *E. coli* expression vector pET-22b(+) and transformed into *E. coli* BL21-Gold(DE3) for expression. We will refer to the mutant library obtained by 1 s annealing/extension as the ProCoS mutant library-A and the library obtained by 5 s and 10 s annealing/extension as the ProCoS mutant library-B.

Screening of ProCoS variants

ProCoS mutant libraries were expressed in 96-well microtiter plates (37°C, 900 rpm, 70% relative humidity), and a 96-well microtiter plate-based AMol (Ammonium Molybdate) screening system was used as reported previously (9). Briefly, 10 μ L cell lysate was incubated with 140 μ L substrate solution (0.6% phytic acid in 250 mM acetate buffer, 0.01% Tween-20, pH 5.5) for 1 h at 37°C. The reaction was stopped by the addition of 150 μ L 15% trichloroacetic acid. Inorganic phosphate release was quantified by addition of 20 μ L stopped reaction mixture to 280 μ L color mix solution (0.27% w/v ammonium molybdate and 1.08% w/v ascorbic acid in 0.32 M H_2SO_4), and the absorption was measured at 820 nm using a Tecan Infinite M1000 microtiter plate reader (Tecan Group AG, Männedorf, Switzerland).

Purification and characterization of wild-type and mutant Ymphytase

Purification and kinetic parameter determination for Ymphytase wild-type and the ProCoS-2 variant were performed as described previously (9).

pH stability

The pH stability of the wild-type and mutant phytases was determined at 37°C using 4 different buffers: 0.25 M glycine-HCl buffer for pH 2.0–3.2; 0.25 M sodium acetate buffer for pH 3.6–5.6; 0.25 M imidazole-HCl buffer for pH 6.0–7.0; and 0.25 M Tris-HCl for pH 7.4–9.0. Purified enzymes were diluted in the specified buffers to 40 ng/mL and incubated at 37°C for 3 h. The activity assay for Ymphytase was carried out with 1 mM phytate at pH 4.5 and at 37°C using the 96-well-plate format colorimetric AMol screening assay as reported

Figure 2. pH activity and stability profiles of Ymphytases (A) pH-activity profile of the ProCoS-2 variant and wild-type Ymphytase. (B) Residual enzyme activity of Ymphytase wild-type (Wt) and the ProCoS-2 variant after incubation at different pH values, ranging from pH 2.8 to pH 9.1, at 37°C for 3 h. The activity was measured after 3 h and compared with non-pH treated Ymphytase (the initial activity measured at pH 4.5 before the incubation).

previously (9). Non pH-treated Ymphytase activity (the initial activity measured at pH 4.5 before the incubation) was considered to be 100% activity, and residual (relative) activity was calculated.

Results and discussion

ProCoS method concept

Homologous enzymes harbor a common feature for catalysis in their primary sequences, the sequence motifs, which reflect their functionality. These functional sequence motifs are conserved throughout different species. Amino acid residues located in the interior (especially the hydrophobic residues), buried inside the protein, are important for correct folding. The underlying concept for ProCoS engineering is the hypothesis that the highly conserved residues in a protein sequence belong to the ancestors of a family of species and, therefore, these residues (non-variable) are important for protein function. On the other hand, the highly variable regions that are not conserved during protein evolution can be targeted for mutagenesis to alter the surface properties of enzymes.

In our ProCoS method (Figure 1), a target protein sequence is first divided into functional and variable (sequence variable or non-conserved) regions based on a multiple sequence alignment of the homologous proteins from a species family. The sequences of these homologous proteins are retrieved from a protein database, and a phylogenetic tree is then constructed based on the multiple sequence alignment to study the relatedness between the homologous sequences. The sequence alignment is used to identify the most variable residue positions. The locations of these residues are then visualized using a protein structure or homology model. Residues situated on the surface of the protein and in the loop regions are targeted for substitution with another residue based on three criteria: (i) frequently occurring residues in the sequence alignment; (ii) chemical similarity, where charged amino acids are introduced either by swapping two charged amino acids that might be involved in salt bridge interactions or by analyzing the surrounding area to avoid the accumulation of positive or negative charge; and (iii) sterically favorable, where the size of the amino acid is considered for substitution to avoid steric clashes after substitution. Synthetic genes were

Table 1. Kinetic parameters of wild-type Ymphytase and the ProCoS-2 variant.

Enzyme	Specific activity (U mg ⁻¹)	K _{half} (μM)	*k _{cat} (S ⁻¹)	Hills coefficient (h)
Wild-type	1043.5 (±26.5)	268.1 (±18.2)	823.47 (±20.91)	2.24 (±0.12)
ProCoS-2	401.08 (±10.92)	260.55 (±9.55)	316.51 (±8.62)	2.93 (±0.2155)

Phytase kinetics were determined in 0.25 M sodium acetate containing different concentrations of sodium phytate (7.8–750 μM) at 37°C, pH 4.5. Values in parentheses represent standard error (3 N).
*Catalytic center activity: enzyme concentration (moles) multiplied by the number of active centers (four).

designed based on the above criteria and synthesized commercially. All of the synthetic gene constructs are recombined using in vitro PCR recombination, and the resulting mutant library is screened for active variants.

Ymphytase surface engineering by ProCoS

Validation of the ProCoS method was done by engineering the surface of the phytase from *Y. mollaretii*. Functionally important conserved residues most likely belonging to common ancestors of the Enterobacteriaceae family of bacteria were identified by a multiple sequence alignment (Supplementary Figure S1). The relatedness of the seven groups of Enterobacteriaceae species was evaluated using a phylogenetic tree (Supplementary Figure S2). Phytase sequences from these species were 25%–85% identical with Ymphytase. On average, the selected sequences were ~50% identical to Ymphytase, except for *Yersinia* (80%–86%) and *Klebsiella* (25%–30%) species. The phylogenetic tree revealed phytases to be grouped according to their activity, with highly active phytases from *Yersinia* species (specific activities ranging 1000–3900 U/mg) (30) grouped together, whereas

Klebsiella phytases, which are known to possess a lower specific activity (99 U/mg) (31), were grouped together. Despite large differences in the activities of the phytases, all 25 sequences, including 7 sequences from *Yersinia* and 5 sequences from *Klebsiella* species, were used in multiple sequence alignment and consensus sequence calculation. Although the sequence identity (25%–85%) between the compared sequences was very low, significant regions in the phytase sequence were highly conserved in Enterobacteriaceae species. Based on the consensus sequence, highly variable regions in the Ymphytase sequence were identified (white-colored residues in Supplementary Figure S1). Compared with a previously published consensus protein design method (32), where the consensus sequence was synthesized as a novel phytase, the ProCoS method uses multiple sequence alignment to identify highly variable regions that can be targeted for mutagenesis. In the present study, highly variable regions situated on the surface and in the loop regions of Ymphytase were targeted for amino acid substitution.

Three synthetic genes with 31 (SyntheticGene-1), 34 (SyntheticGene-2), and 40 (SyntheticGene-3) amino acid substitu-

Table 2. Analysis of the amino acid composition of the 34 substitutions obtained for the ProCoS-2 variant.

Amino acid	Wild-type		ProCoS-2		Chemical Property
	Total number	Substitutions (%)	Total number	Substitutions (%)	
Glu	3	9	10	29	Negative, polar
Lys	6	18	8	24	Positive, polar
Ala	2	6	2	6	Neutral, non-polar
Arg	1	3	2	6	Positive, polar
Asp	2	6	2	6	Negative, polar
Leu	2	6	2	6	Neutral, non-polar
Pro	1	3	2	6	Neutral, non-polar
Ser	2	6	2	6	Neutral, polar
Gln	7	21	1	3	Neutral, polar
His	1	3	1	3	Positive, polar
Thr	3	9	1	3	Neutral, polar
Tyr	0	0	1	3	Neutral, polar
Gly	2	6	0	0	Neutral, non-polar
Ile	2	6	0	0	Neutral Non-polar

tions were synthesized commercially, and mutations introducing the most favorable substitutions were randomly assigned to all 3 genes. As expected, due to a high mutational load on the protein, all 3 synthetic Ymphytase variants were inactive after protein expression at 37°C. Barely detectable phytase activity was observed in SyntheticGene-2 when it was expressed at a lower temperature (25–30°C). Recombination of all three synthetic genes, omitting wild-type Ymphytase, was then performed. PCR-based recombination methods have successfully been used to recombine point mutations from multiple templates to improve the thermostability of a subtilisin E (33) and the enantioselectivity of a lipase (34). Sequence information within the parental (template) sequences is exchanged due to template switching events in the PCR recombination. The unique sequence information (mutations) in a parental sequence can be passed on to other parental sequences, and the mutations are distributed in the chimeras (ProCoS mutant library) obtained from the experiment. DNA templates were amplified using vector-specific primers to incorporate two universal primer binding sites (F1 and R3) in all three synthetic genes to be recombined. Three different annealing/extension times were used for PCR, resulting in a distinct electrophoresis band of the correct size (~1.4 kb) after 20 and 40 cycles in all 3 cases. Higher numbers of cycles resulted in a smear that increased with the number of PCR cycles (Supplementary Figure S3).

The cloned ProCoS mutant libraries of Ymphytase variants were screened for enzyme activity. Almost 1050 clones were screened using a 96-well microtiter plate-based AMol assay. A total of 552 and 460 mutant clones were screened from ProCoS mutant library-A and ProCoS mutant library-B, respectively. Upon initial screening, mutants showing at least 20% relative activity to wild-type Ymphytase were selected. ProCoS mutant library-A was found to have more active clones (26%) than ProCoS mutant library-B (20%). The high ratio of active clones in ProCoS mutant library-A might be due to the high recombination efficiency of a 1 s extension/annealing time compared with 5s or 10 s extension/annealing. Sampling of only 1050 clones produced 6 variants with phytase activity, suggesting that these were highly enriched libraries with active

Table 3. Amino acid substitutions in the ProCoS-2 variant grouped according to their chemical property and shown in comparison to wild-type Ymphytase.

Position	Wild-type	ProCoS-2
166	Ala	Ser
136	Ala	Thr
127	Gly	Asp
81	Gly	Lys
401	Ile	Leu
403	Ile	Leu
115	Leu	Lys
78	Lue	Tyr
85	Asp	Glu
122	Asp	Lys
188	Glu	Arg
174	Glu	Gln
48	Glu	Lys
60	Gln	Glu
118	Gln	Glu
177	Gln	Glu
185	Gln	Glu
202	Gln	Glu
74	Gln	Lys
203	Gln	Lys
386	Ser	Glu
200	Ser	Lys
77	Thr	Ala
158	Thr	Lys
193	Thr	Pro
64	Pro	His
162	Arg	Glu
433	His	Ser
199	Lys	Ala
57	Lys	Arg
181	Lys	Asp
131	Lys	Glu
206	Lys	Glu
388	Lys	Pro

Amino acid color code: aliphatic (G, A, V, L, I; white), aromatic (F, Y, W; magenta), neutral (C, M, P, S, T, N, Q; green), charged positive (H, K, R; blue), charged negative (D, E; pink), polar (T, S, Y, H, N, Q, D, E, K, R; cyan), and non-polar (I, V, L, F, C, M, A, G, W, P; orange).

enzyme populations. Phylogenetic trees using multiple sequence alignments of the six ProCoS variants, the synthetic genes, and wild-type Ymphytase (Supplementary Figure S5) revealed the relatedness of the ProCoS variants. These sequences were found to be grouped according to activity, and the ProCoS-2 variant was grouped close to SyntheticGene-2, which is the only active synthetic gene among all of the synthetic genes (when expressed at a lower temperature). The ProCoS-2 variant was recombined to incorporate six amino acid substitutions at the C-terminus compared with SyntheticGene-2 Ymphytase (I401L, I403L, K407E, K411T,

E426Q, and H433S). The active clones obtained from the ProCoS mutant library screening were scattered in the phylogenetic tree, showing that the designed sequences (synthetic genes) are located quite close to the active clones (ProCoS variants). The six best ProCoS variants were retransformed, and their activity was assayed using test tube expression with cell lysates (Supplementary Figure S4). The sequences of Ymphytase variants ProCoS-3 and ProCoS-5 were identical. A ProCoS-2 variant with the highest activity among the six clones was selected for further purification and characterization.

Characterization of the Ymphytase ProCoS-2 variant

The pH activity profile of the ProCoS-2 variant showed a broad pH optimum, from pH 3.0 to pH 4.5, compared with the wild-type pH optimum of pH 4.5 (Figure 2A). The phytase kinetics parameters were determined at pH 4.5 using sodium phytate as a substrate, and the data generated were analyzed by the general allosteric model using the Hill equation. The Hill coefficient (*h*) value for the wild-type and ProCoS-2 enzymes were 2.2 and 2.9, respectively, indicating positive cooperativity (Table 1). The K_{half} for the wild-type and ProCoS-2 enzymes were 268 μM and 261 μM , respectively. The catalytic activity (k_{cat}) was calculated by assuming four active site centers per molecule of Ymphytase. The ProCoS-2 variant was found to possess ~40% of the wild-type enzyme's specific activity with a total of 34 amino acid substitutions. One of the main criteria for the use of phytases in feed is their stability at acidic pH as they pass through the stomach. Therefore, we analyzed the pH-activity profiles and pH stability of the Ymphytase wild-type and ProCoS-2 variants. Surprisingly, the ProCoS-2 variant showed an overall improved pH stability from pH 2.8 to pH 7.5 when compared with the wild-type and a 3.8-fold and 3.0-fold increased pH stability at pH 2.8 and pH 3.3, respectively (37°C/3 h; Figure 2B).

Sequence analysis of ProCoS variants

In the surface engineering of Ymphytase using ProCoS, most of the substitutions that are chemically non-similar were at the N terminus (Supplementary Figure S5; highlighted in magenta). These varia-

Figure 3. Structural model of the surface-engineered Ymphytase. (A) Amino acid substitutions (total of 34) in the Ymphytase ProCoS-2 variant show on the homology model of wild-type Ymphytase. The electrostatic potential surfaces of the wild-type (B,C) and the ProCoS-2 variant (D,E) showed a significantly increased negative surface for ProCoS-2 compared with the wild-type. The colors are mapped from red (for the most negatively charged surfaces), to white (uncharged), to blue (for the most positively charged surfaces).

tions were the result of the criterion used for designing synthetic genes, namely selecting the most frequently occurring residues from the alignment, for example P64H, which was a result of a recombination event (codons for the most frequently occurring residues, H or K, were present in the synthetic genes), and T77A, which was also the result of a recombination event (codons for the most frequently occurring residues, A or V, were present in the synthetic genes). Likewise, chemically non-similar substitutions were the result of incorporating the most frequently occurring residues (mentioned in the parentheses): L78Y (Y/N/E), G81K (R/H), L115K (K/A), G127D (D). D122K was introduced by swapping the charge of the residue in a synthetic gene.

Only the Ymphytase variable regions identified using sequence alignment were targeted to design the synthetic genes. The conserved regions in the Entero-

bacteriaceae family were scattered throughout the sequence (Supplementary Figure S1, regions highlighted in blue, green, and yellow). Some of the conserved regions might be highly important for the function of a phytase, and others may have been phylogenetic artifacts, due the enzymes being in the same family. To rule out this possibility, we used a species-specific UniProt blast of Ymphytase to retrieve acid phosphatase sequences from several other organisms (Archaea, eukaryotes, arthropods, fungi, nematodes, and mammals). Interestingly, sequence alignment of phosphatases from these distantly related species revealed several areas of Ymphytase to be highly conserved (Supplementary Figures S1 and S7). The active site sequence motifs RHGXRRP (residues 37–43) and HD (residues 326–327) were conserved in all of the organisms. Additionally, the motifs GXLT (residues 66–69), RTXXS/T (residues

112–116), and FXP (residues 145–147) in Ymphytase were highly conserved in all of the organisms. The four residues G123, P252, C406, and C415 of Ymphytase were also highly conserved. A disulfide bond between C406 and C415 might be essential for phytase function and is conserved throughout the organisms.

A total of 34 amino acid positions were substituted in the ProCoS-2 variant compared with the wild-type. In the case of the wild-type, 7 of these 34 amino acids were glutamine (a neutral and polar amino acid), while only 1 glutamine was preserved in the ProCoS-2 variant (Tables 2 and 3). Lysine, a positively charged amino acid (the blue-colored residues in Table 3), was increased from 6 residues in the wild-type to 8 in ProCoS-2, whereas the number of negatively charged amino acids (pink-colored residues in Table 3) was increased from 3 (9%) in the wild-type to 10 (29%) in the ProCoS-2 variant.

Table 4. Fraction of surface residues at the substituted positions in wild-type Ymphytase and the ProCoS-2 variant grouped by the chemical properties of the amino acids.

Chemical properties of the amino acids	Residues in wild-type (%)	Substitutions in ProCoS-2 (%)
Negative, Polar	15	35
Positive, Polar	24	32
Neutral, Polar	35	15
Neutral, Non-polar	26	18
Polar	74	82
Non-polar	26	18

These findings suggest that the increase in charged amino acids (Lys, Glu, Asp) occurred by decreasing the neutral surface represented by glutamine. This increased charged polar surface on the ProCoS-2 variant might contribute to the improved pH stability.

In order to obtain a molecular understanding of the type of substitutions on the Ymphytase surface, a thorough analysis was performed (Tables 3 and 4). In the wild-type enzyme, out of 34 substituted residues, a negative polar surface was represented by 5 residues (2 Asp and 3 Glu), which was significantly increased to 12 residues in the ProCoS-2 variant (2 Asp and 10 Glu). The residues substituted in the ProCoS-2 variant are shown in Figure 3A, which illustrates the location of the substitutions. The electrostatic potential calculated using the ABPS (Adaptive Poisson-Boltzmann Solver) module implemented in UCSF chimera software (35) showed a significant increase in the negative potential of the ProCoS-2 variant (Figure 3D and 3E) when compared with the wild-type (Figure 3B and 3C). As shown in Table 4, the amount of surface contributed by either positive polar or polar residues was increased by 8 percentage points (positive polar: from 24% to 32%; polar: from 74% to 82%). Both the neutral non-polar or non-polar residues were decreased by 8 percentage points (from 26% to 18%), and the neutral polar surface was decreased by 20 percentage points (from 35% to 15%) of substituted residues. Overall, the polar surface of the ProCoS-2 variant was increased by 8 percentage points, and the non-polar surface of wild-type Ymphytase was decreased by 8 percentage points (cyan and orange colored residues in Table 3). However, no significant change was observed between the hydrophobic surfaces of ProCoS-2 and the wild-type enzyme (Supplementary

Figure S6). The increase in pH stability of the ProCoS-2 variant at acidic pH might be due to the increased fraction of the polar surface and a significant increase in the charged surface (28 percentage points, negative and positive charge), especially an increase in the negative surface (20 percentage points; from 9% to 29%), that was achieved by decreasing overall neutral polar surface (20 percentage points) of the Ymphytase.

Here, we have described ProCoS, a simple computer-assisted method for protein surface engineering. The surface engineering of phytase shows that the highly variable regions (non-conserved) located on the protein surface can easily be targeted for protein surface modification, and the highly conserved residues belonging to an ancestor of a family of species are required for the function of a protein. Interestingly, a total of 34 sites in Ymphytase were successfully substituted with other amino acids, yielding an active variant with 3.8-fold improved pH stability (at pH 2.8 for 3 h). The improved pH stability might be due to increased negatively charged amino acid substitutions on the Ymphytase surface. The ProCoS method can be generally applied to other biocatalysts to improve pH stability and could potentially be applied for improving properties that are influenced by the amino acid residues on the surface of a protein, such as solubility, thermostability, and non-aqueous solvent stability.

Author contributions

A.V.S., H.W.H., S.H., and U.S. conceived and designed the experiments. A.V.S. performed the experiments. A.V.S. and H.W.H. contributed reagents, materials, and analysis tools. A.V.S. analyzed the data. A.V.S. and U.S. wrote the paper.

Acknowledgments

We thank BASF SE for financial support.

Competing interests

The authors declare no competing interests.

References

- Shivange, A.V., J. Marienhagen, H. Mundhada, A. Schenk, and U. Schwaneberg. 2009. Advances in generating functional

- diversity for directed protein evolution. *Curr. Opin. Chem. Biol.* 13:19-25.
- Goldsmith, M. and D.S. Tawfik. 2012. Directed enzyme evolution: beyond the low-hanging fruit. *Curr. Opin. Struct. Biol.* 22:406-412.
- Romero, P.A. and F.H. Arnold. 2009. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* 10:866-876.
- Tracewell, C.A. and F.H. Arnold. 2009. Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr. Opin. Chem. Biol.* 13:3-9.
- Salverda, M.L., E. Dellus, F.A. Gorter, A.J. Debets, J. van der Oost, R.F. Hoekstra, D.S. Tawfik, and J.A. de Visser. 2011. Initial mutations direct alternative pathways of protein evolution. *PLoS Genet.* 7:e1001321.
- Shivange, A.V., D. Roccatano, and U. Schwaneberg. 2016. Iterative key-residues interrogation of a phytase with thermostability increasing substitutions identified in directed evolution. *Appl. Microbiol. Biotechnol.* 100:227-242.
- Shivange, A.V., A. Dennig, and U. Schwaneberg. 2014. Multi-site saturation by OmniChange yields a pH- and thermally improved phytase. *J. Biotechnol.* 170:68-72.
- Shivange, A.V. and U. Schwaneberg. 2016. Recent advances in directed phytase evolution and rational phytase engineering. In M. Alcalde (Ed.), *Directed Enzyme Evolution: Advances and Applications*. Springer, New York, NY.
- Shivange, A.V., A. Serwe, A. Dennig, D. Roccatano, S. Haefner, and U. Schwaneberg. 2012. Directed evolution of a highly active *Yersinia mollaretii* phytase. *Appl. Microbiol. Biotechnol.* 95:405-418.
- Liu, Z., Z. Sun, and Y. Leng. 2006. Directed evolution and characterization of a novel D-pantothanolase from *Fusarium moniliforme*. *J. Agric. Food Chem.* 54:5823-5830.
- Veronese, F.M. 1994. Enzyme surface modification by polymers for improved delivery. *J. Control. Release* 29:171-176.
- Qi, Y. and A. Chilkoti. 2015. Protein-polymer conjugation-moving beyond PEGylation. *Curr. Opin. Chem. Biol.* 28:181-193.
- Montes, T., V. Grazu, F. Lopez-Gallego, J.A. Hermoso, J.M. Guisan, and R. Fernandez-Lafuente. 2006. Chemical modification of protein surfaces to improve their reversible enzyme immobilization on ionic exchangers. *Biomacromolecules* 7:3052-3058.
- Montes, T., V. Grazu, F. Lopez-Gallego, J.A. Hermoso, J.L. Garcia, I. Manso, B. Galan, R. Gonzalez, et al. 2007. Genetic modification of the penicillin G acylase surface to improve its reversible immobilization on ionic exchangers. *Appl. Environ. Microbiol.* 73:312-319.
- Mateo, C., V. Grazu, B.C. Pessela, T. Montes, J.M. Palomo, R. Torres, F. Lopez-Gallego, R. Fernandez-Lafuente, and J.M. Guisan. 2007. Advances in the design of new epoxy supports for enzyme immobilization-stabilization. *Biochem. Soc. Trans.* 35:1593-1601.
- Novak, M.J., A. Pattammattel, B. Koshmerl, M. Puglia, C. Williams, and C.V. Kumar. 2016. "Stable-on-the-Table" Enzymes: Engineering the Enzyme-Graphene Oxide Interface for Unprecedented Kinetic Stability of the Biocatalyst. *ACS Catal.* 6:339-347.
- Simeonov, P., R. Berger-Hoffmann, R. Hoffmann, N. Strater, and T. Zuchner. 2011.

Surface supercharged human enteropeptidase light chain shows improved solubility and refolding yield. *Protein Eng. Des. Sel.* 24:261-268.

18. Mosavi, L.K. and Z.Y. Peng. 2003. Structure-based substitutions for increased solubility of a designed protein. *Protein Eng.* 16:739-745.
19. Turunen, O., M. Vuorio, F. Fenel, and M. Leisola. 2002. Engineering of multiple arginines into the Ser/Thr surface of *Trichoderma reesei* endo-1,4-beta-xylanase II increases the thermotolerance and shifts the pH optimum towards alkaline pH. *Protein Eng.* 15:141-145.
20. Martinez, P. and F.H. Arnold. 1991. Surface Charge Substitutions Increase the Stability of a-Lytic Protease in Organic Solvents. *J. Am. Chem. Soc.* 113:6336-6337.
21. Zhao, J., N. Jia, K.E. Jaeger, M. Bocola, and U. Schwaneberg. 2015. Ionic liquid activated *Bacillus subtilis* lipase A variants through cooperative surface substitutions. *Biotechnol. Bioeng.* 112:1997-2004.
22. Strickler, S.S., A.V. Gribenko, A.V. Gribenko, T.R. Keiffer, J. Tomlinson, T. Reihle, V.V. Loladze, and G.I. Makhatadze. 2006. Protein stability and surface electrostatics: a charged relationship. *Biochemistry* 45:2761-2766.
23. Gribenko, A.V., M.M. Patel, J. Liu, S.A. McCallum, C. Wang, and G.I. Makhatadze. 2009. Rational stabilization of enzymes by computational redesign of surface charge-charge interactions. *Proc. Natl. Acad. Sci. USA* 106:2601-2606.
24. Chan, C.H., C.C. Wilbanks, G.I. Makhatadze, and K.B. Wong. 2012. Electrostatic contribution

of surface charge residues to the stability of a thermophilic protein: benchmarking experimental and predicted pKa values. *PLoS One* 7:e30296.

25. Schweiker, K.L., A. Zarrine-Afsar, A.R. Davidson, and G.I. Makhatadze. 2007. Computational design of the Fyn SH3 domain with increased stability through optimization of surface charge charge interactions. *Protein Sci.* 16:2694-2702.
26. Warden, A.C., M. Williams, T.S. Peat, S.A. Seabrook, J. Newman, G. Dojchinov, and V.S. Haritos. 2015. Rational engineering of a mesohalophilic carbonic anhydrase to an extreme halotolerant biocatalyst. *Nat. Commun.* 6:10278.
27. Schlapschky, M., U. Binder, C. Borger, I. Theobald, K. Wachinger, S. Kisling, D. Haller, and A. Skerra. 2013. PASylation: a biological alternative to PEGylation for extending the plasma half-life of pharmaceutically active proteins. *Protein Eng. Des. Sel.* 26:489-501.
28. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: Visual molecular dynamics. *J Mol Graph* 14:33-38.
29. Richardson, S.M., S.J. Wheelan, R.M. Yarrington, and J.D. Boeke. 2006. GeneDesign: rapid, automated design of multi-kilobase synthetic genes. *Genome Res.* 16:550-556.
30. Huang, H., H. Luo, Y. Wang, D. Fu, N. Shao, G. Wang, P. Yang, and B. Yao. 2008. A novel phytase from *Yersinia rohdei* with high phytate hydrolysis activity under low pH and strong

pepsin conditions. *Appl. Microbiol. Biotechnol.* 80:417-426.

31. Sajidan, A., A. Farouk, R. Greiner, P. Jungblut, E.C. Muller, and R. Borris. 2004. Molecular and physiological characterisation of a 3-phytase from soil bacterium *Klebsiella* sp. ASR1. *Appl. Microbiol. Biotechnol.* 65:110-118.
32. Lehmann, M., L. Pasamontes, S.F. Lassen, and M. Wyss. 2000. The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta* 1543:408-415.
33. Zhao, H., L. Giver, Z. Shao, J.A. Affholter, and F.H. Arnold. 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.* 16:258-261.
34. Eggert, T., S.A. Funke, N.M. Rao, P. Acharya, H. Krumm, M.T. Reetz, and K.E. Jaeger. 2005. Multiplex-PCR-based recombination as a novel high-fidelity method for directed evolution. *ChemBioChem* 6:1062-1067.
35. Baker, N.A., D. Sept, S. Joseph, M.J. Holst, and J.A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* 98:10037-10041.

Received 26 July 2016; accepted 23 September 2016.

Address correspondence to Amol V. Shivange, RWTH Aachen University, Lehrstuhl für Biotechnologie, Worringer Weg 3, 52074 Aachen, Germany. E-mail: shivange@caltech.edu

To purchase reprints of this article, contact: biotechniques@fosterprinting.com

RAININ
Pipetting 360°

Introducing BenchSmart 96!

A smarter approach to automated 96/384-well pipetting.

- Three interchangeable heads – pipette from 0.5 µL to 1000 µL
- Programmable features, such as mixing and multi-dispense
- Automated tip loading for perfect seals on all 96 channels
- Manual x-y-z movement of the pipetting head
- Large touchscreen – so easy and intuitive, everyone in the lab will want to use it

See it for yourself!

Let an application specialist demonstrate how BenchSmart can speed your workflow.

► Set up a live online virtual demo at www.mt.com/BenchSmartDemo

METTLER TOLEDO