ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/236652471

MAP(2.0)3D: A Sequence/Structure Based Server for Protein Engineering

ARTICLE *in* ACS SYNTHETIC BIOLOGY · APRIL 2012

Impact Factor: 4.98 · DOI: 10.1021/sb200019x · Source: PubMed

CITATIONS
9

READS
34

3 AUTHORS, INCLUDING:

Rajni Verma
Washington University in St. Louis
13 PUBLICATIONS **46** CITATIONS

SEE PROFILE

Danilo Roccatano
University of Lincoln
103 PUBLICATIONS **2,328** CITATIONS

SEE PROFILE

Available from: Rajni Verma
Retrieved on: 01 February 2016

ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

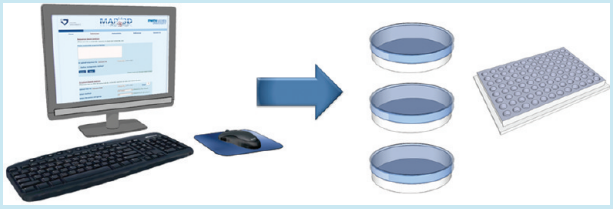
MAP$^{2.0}$3D: A Sequence/Structure Based Server for Protein Engineering

Rajni Verma,[†,‡] Ulrich Schwaneberg,[‡] and Danilo Roccatano*[,†]

[†]School of Engineering and Science, Jacobs University Bremen, Campus Ring 1, 28759 Bremen, Germany

[‡]Department of Biotechnology, RWTH Aachen University, Worringer Weg 1, 52074 Aachen, Germany

ABSTRACT: The Mutagenesis Assistant Program (MAP) is a web-based tool to provide statistical analyses of the mutational biases of directed evolution experiments on amino acid substitution patterns. MAP analysis assists protein engineers in the benchmarking of random mutagenesis methods that generate single nucleotide mutations in a codon. Herein, we describe a completely renewed and improved version of the MAP server, the MAP$^{2.0}$3D server, which correlates the generated amino acid substitution patterns to the structural information of the target protein. This correlation aids in the selection of a more suitable random mutagenesis method with specific biases on amino acid substitution patterns. In particular, the new server represents MAP indicators on secondary and tertiary structure and correlates them to specific structural components such as hydrogen bonds, hydrophobic contacts, salt bridges, solvent accessibility, and crystallographic B-factors. Three model proteins (D-amino oxidase, phytase, and *N*-acetylneuraminic acid aldolase) are used to illustrate the novel capability of the server. MAP$^{2.0}$3D server is available publicly at http://map.jacobs-university.de/map3d.html.



KEYWORDS: *directed evolution, random mutagenesis methods, mutational bias indicator, bioinformatics tools, mutagenesis assistant program, protein engineering*

Over the past two decades directed protein evolution has been proven to be a powerful algorithm to tailor protein properties through iterative rounds of random mutagenesis and screening for improved protein variants.[1,2] Directed evolution methods are especially useful for improving properties difficult to rationalize and, hence, to identify amino acids and protein regions that can lead to further enhancements using site directed and saturation mutagenesis methods.[3,4] The success of a directed evolution campaign depends highly on the quality of the mutant library and on the employed random mutagenesis method. Random mutagenesis methods are based on specific error prone polymerases (enzymatic methods), DNA modifying chemicals (*e.g.*, nitrous acid), or mutator strains (*e.g.*, *Escherichia coli* mutA).[5] The quality of a mutant library is determined by the generated genetic diversity and corresponding protein sequence space.[6] Since the number of muteins rises with the increasing number of amino acids exchanged in the protein, protein engineers are challenged with an astronomically vast sequence space.[7] Despite advances in high-throughput screening, it is very difficult to screen the theoretically generated diversity even in the case of a small protein.[8,9] Therefore, generating high quality mutant libraries enriched with functional traits is of high importance. To deal with the challenge to access and screen such a large sequence space, protein engineers usually adopt two strategic approaches.[10−12] The first approach consists in the random mutagenesis of the target protein and the subsequent identification of "mutagenic hot spots". Random mutagenesis can be followed by recombination of the best variants by site directed mutagenesis or saturation mutagenesis.[13] The second approach involves the identification of a subset of specific residues using rational or semirational design with the help of computational tools.[14] Up to five amino acid positions can be efficiently targeted with focused mutagenesis methods allowing the generation of focused mutant libraries of a number of variants that can be screened with the state of the art in flow cytometry methods.[13] Focused mutagenesis is normally employed to improve the properties of a target protein such as activity or selectivity, by mutating residues in close proximity to a specific protein region such as the active site. In this case, random mutagenesis methods are complementary to the rational design since they can identify important amino acid positions, especially in the second and third coordination sphere, which would have been overlooked rationally. Nevertheless, random mutagenesis methods are biased toward certain nucleotide exchanges (*e.g.*, many epPCR methods prefer transition mutations). The mutagenic preferences produced by biased random mutagenesis methods affect the generated diversity. The analysis of the effects of mutational bias on the amino acid diversity provides a useful indicator in the selection of the mutagenesis method with diverse and complementary amino acid substitution patterns. The generated complementary mutant libraries extend the sampling of the vast protein space and enhance the chance to obtain improved variants.[15,16]

Received: November 9, 2011
Published: February 13, 2012

ACS Publications © 2012 American Chemical Society

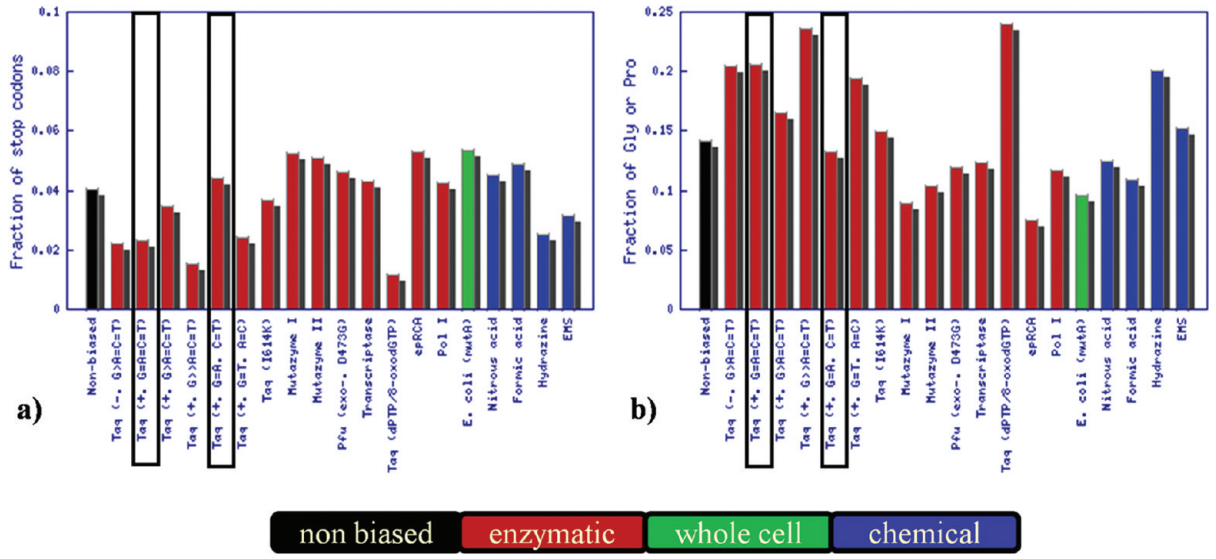


Figure 1. Statistical analysis of stop codon frequencies (a) and Gly/Pro substitutions (b) for D-amino acid oxidase (RgDAAO). The random mutagenesis methods enclosed in the black rectangles (epPCR (Taq ($MnCl_2$, G=A=C=T)) and epPCR (Taq ($MnCl_2$, G=A, C=T))) are used for the MAP$^{2.0}$3D analysis.

Recently, we introduced a freely available web-based statistical analysis tool (MAP[17]). The server statistically analyzes the effects of mutational bias of 19 different random mutagenesis methods on the level of amino acid substitutions for a given nucleotide sequence of the target protein. The analysis is returned in terms of MAP indicators that allow a rapid comparison of different random mutagenesis methods on the protein level. It has been shown that this approach can be used to predict the type, extent, and chemical nature of the genetic diversity generated by different mutagenesis methods.[17,18] Recently, Rasila and co-workers[19] reported a comparative evolution of commonly used random mutagenesis methods. They found the experimentally induced substitution patterns to be very similar to those obtained by MAP server and suggested the use of a combination of mutagenesis methods to generate high diversity.[19]

One of the limitations of the original MAP server is the absence of the analysis tools relating the MAP indicators to the structural properties of the target protein. The nature of the amino acid change in different regions of the protein can affect its global and local structural and thermodynamic properties.[14,20,21] Therefore, the possibility to correlate the generated diversity with structural properties would help to identify in advance the random mutagenesis method that has the least number of "deleterious" mutations on the protein stability and the higher probability to introduce amino acid substitutions that may improve the fitness toward an expected property, *e.g.*, substitutions to charged amino acid residues to increase solubility in water. For this reason, we have expanded the capability of the server by introducing these new features. The new server (MAP$^{2.0}$3D) can correlate the mutational propensity at the amino acid level of a gene for 19 random mutagenesis methods (and now also for a user customized random mutagenesis method) with the crystallographic or homology modeled structure (if available in Protein Data Bank[22] format) of the target protein. MAP$^{2.0}$3D analyses the three-dimensional structure of the target proteins by calculating secondary structure elements, important local interactions (such as hydrogen bonds, hydrophobic contacts, salt bridges, disulfide bridges, solvent accessibility), and amino acid motilities from the crystallographic B-factors. Taken together, this information helps to identify biased amino acid substitutions that may improve stability and function of the protein.[23−25]

To correlate the sequence-based analysis to the structural data analysis, a new indicator, the residue mutability indicator "μ" (amino acid substitution probability leading to amino acid change at specific position) has been introduced (see Methods). The mutability indicator allows a rapid identification of mutagenic hot spots and easier comparison of experimental data to the predicted ones.

In this article, the new features of the MAP$^{2.0}$3D server are illustrated in detail by performing a study on three model proteins. The results of the MAP$^{2.0}$3D analysis are compared with the experimental results of protein engineering experiments reported in the literature. The three examples show possible uses of the server for computational prescreening of the target protein to evaluate and select a mutagenesis method for direct protein evolution.

RESULTS AND DISCUSSION

The use of the MAP$^{2.0}$3D server is illustrated by performing the analysis of three different enzymes evolved for different properties by using directed protein evolution. The first example describes how to decrease effects of mutational bias and to generate a mutant library with a higher fraction of active clones. The second and third examples show the usability of the server to analyze the influence of mutational preferences on the evolution of desirable properties. Outputs of the complete MAP$^{2.0}$3D analysis are provided as examples in the instruction link of the server (http://map.jacobs-university.de/instruction.html).

D-Amino Acid Oxidase. D-Amino acid oxidase (DAAO) is a flavin adenine dinucleotide (FAD) dependent flavoenzyme. DAAO catalyzes the dehydrogenation of D-amino acid to the corresponding α-keto acids, producing ammonia and hydrogen peroxide.[26,27] The high turnover rate, the stable FAD-binding, and the broad substrate specificity of DAAO from *Rhodotorula gracilis* (RgDAAO) make it an attractive catalyst for

dx.doi.org/10.1021/sb200019x | *ACS Synth. Biol.* 2012, 1, 139−150

Table 1. Summary of the MAP$^{2.0}$3D Analysis for the Oxidase, the Phytase, and the Aldolase, Targeting Different epPCR Methods for Random Mutagenesis

	RgDAAO (1st round)	RgDAAO (2nd round)	phytase	*N*-acetylneuraminic acid aldolase
epPCR method	Taq (+, G=A=C=T)	Taq (+, G=A, C=T)	Taq (+, G=A, C=T)	Taq (+, G=A=C=T)
average amino acid substitutions[a]	7.40	7.40	7.45	7.20
preserved amino acid substitutions[b]	24.53%	23.38%	25.40%	28.47%
codon diversity coefficient[c]	42.48	34.04	36.49	43.70
stop codons frequency[d]	2.30%	4.38%	4.69%	2.12%
Gly/Pro frequency[e]	20.58%	13.23%	11.60%	16.26%
charged amino acid diversity[f]	−0.34% (25.00%)	−2.62% (25.00%)	1.39% (19.21%)	5.00% (22.22%)
neutral amino acid diversity[g]	3.37% (27.99%)	4.47% (27.99%)	−4.14% (35.65%)	1.73% (26.94%)
aromatic amino acid diversity[h]	−3.19% (7.34%)	−0.23% (7.34%)	0.91% (5.79%)	−3.14% (8.08%)
aliphatic amino acid diversity[i]	−2.13% (39.67%)	−6.00% (39.67%)	−2.86% (39.35%)	−5.72% (42.76%)

[a]Average number of amino acid substitutions per residue. [b]$I_{\alpha\to pr}$: fraction of variants with preserved amino acid substitutions. [c]Codon diversity coefficient. [d]$I_{\alpha\to st}$: fraction of variants with stop codons. [e]$I_{\alpha\to gp}$: fraction of variants with Gly/Pro and chemical diversity generated by the mutagenesis methods presented as [f]$I_{\alpha\to ch}$: charged, [g]$I_{\alpha\to ne}$: neutral, [h]$I_{\alpha\to ar}$: aromatic, and [i]$I_{\alpha\to al}$: aliphatic amino acid diversity with the amino acid composition of the target protein sequence (in parentheses) and deviation from this composition after mutagenesis.

biotechnological applications such as biosensing (*i.e.*, the rapid and reliable detection of D-amino acid content in food specimens or of the neurotransmitter D-serine in the brain).[27] We performed MAP$^{2.0}$3D analysis on the *RgDAAO* to evaluate the amino acid diversity generated by random mutagenesis methods.

MAP$^{2.0}$3D Analysis. The sequence based MAP$^{2.0}$3D analysis was performed using the following descriptors: (i) protein structure indicators, (ii) amino acid diversity indicator with codon diversity coefficient, and (iii) chemical diversity indicator.

In Figure 1, the values for the stop codon indicator ($I_{\alpha\to st}$) and the Gly/Pro indicator ($I_{\alpha\to gp}$) for different random mutagenesis methods are reported. The two methods show opposite trends in the generation of stop codons (sequence truncation) and Gly/Pro (α helix destabilizers), *i.e.*, the higher the stop codons frequency the lower the Gly/Pro substitutions and *vice versa*.[17] The two epPCR methods (indicated in the Figure 1 with the black rectangles) were found to be more appropriate for the *RgDAAO* with the balanced frequencies of stop codons and Gly/Pro in comparison to other mutagenesis methods. In Table 1, the sequence-based analysis of the server for selected random mutagenesis methods is summarized. The first method, the balanced epPCR Taq-Pol (Mn^{2+}, balanced dNTP),[28] has a strong preference for specific nucleotide exchanges ~32% AT→GC (transition mutations), whereas the second method, the unbalanced epPCR Taq-Pol (Mn^{2+}, unbalanced dNTP),[29] is expected to produce more transversions (21.41% AT→TA) than transitions (14.45% AT→GC). Balanced epPCR was expected to generate a lower fraction of stop codons ($I_{\alpha\to st}$ = 2.30%) and higher Gly/Pro ($I_{\alpha\to gp}$ = 20.58%) content than the unbalanced epPCR ($I_{\alpha\to st}$ = 4.38% and $I_{\alpha\to gp}$ = 13.23%) (see Table 1). For both methods, an average of 7.4 amino acid substitutions per residue was calculated.

In Figure 2, cartoon representations of the RgDAAO crystallographic structure colored accordingly to $I_{\alpha\to gp}$ using the Jmol[30] visualization feature of the new server are shown. Out of 30% of the residues involved in helix formation, 51% have a higher $I_{\alpha\to gp}$ value (if α is equal to S, L, E, and D) with a prevalence of negatively charged residues (E and D, highlighted in stick format in Figure 2). In comparison to the unbalanced epPCR, the balanced epPCR method was observed with a higher probability of the charged residues substitution into

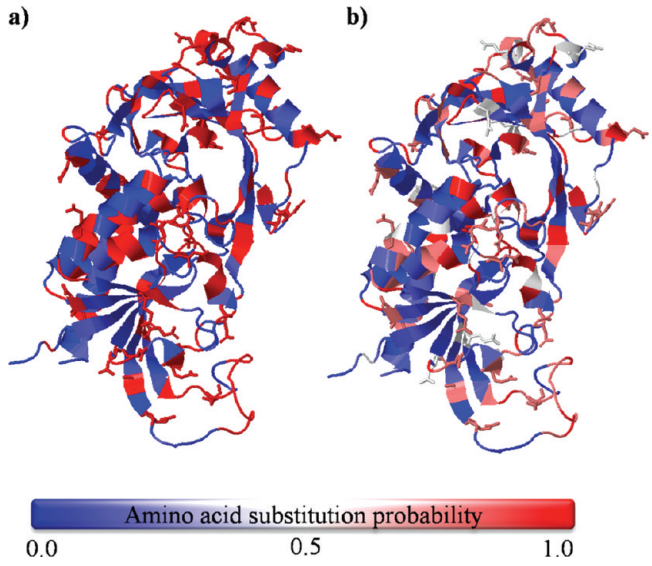


Figure 2. Gly/Pro amino acid substitutions mapping on RgDAAO structure for (a) epPCR (Taq ($MnCl_2$, G=A=C=T)) and (b) epPCR (Taq ($MnCl_2$, G=A, C=T)). For the balanced epPCR method (a) the red colored regions of RgDAAO structure indicate an overall higher probability of charged residues substitutions, mainly for negatively charged residue (in stick representation), into Gly/Pro than the unbalanced epPCR (b).

Gly/Pro (represented by the color code to define amino acid substitution probability in Figure 2). The mapping of charged amino acid substitution patterns on the structure of RgDAAO is reported in Figure 3 and was consistent with the latter observation of Gly/Pro substitution patterns. The balanced epPCR (Figure 3a) shows lower probability for charged amino acid substitutions than unbalanced epPCR (Figure 3b), which is opposite to the Gly/Pro substitution patterns for both methods (Figure 2). Hence, the amino acid substitutions of charged residues into residues unfavorable for forming molecular interactions result in destabilization of RgDAAO. For example, charged residues were found to be involved in molecular interactions such as salt bridges (15 out of 21 show more than 0.5 probability to be substituted into glycine) and side chain H-bonds (5 out of 26 with more than 0.5 probability for glycine substitutions). In Figure 4, charged residues involved in salt bridge formation with the amino acid diversity

dx.doi.org/10.1021/sb200019x | *ACS Synth. Biol.* 2012, 1, 139–150

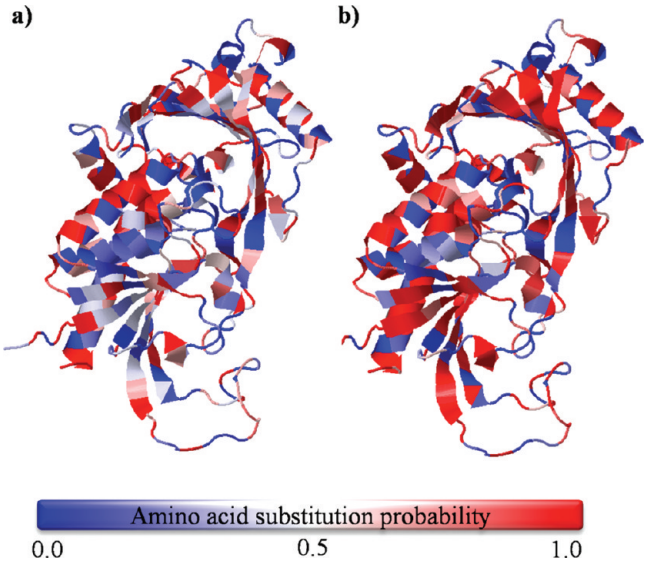


Figure 3. Amino acid substitutions mapping of charged residues (E, D, R, K, H) on RgDAAO for (a) epPCR (Taq ($MnCl_2$, G=A=C=T)) and (b) epPCR (Taq ($MnCl_2$, G=A, C=T)).

generated by the balanced (Figure 4a and b) and unbalanced epPCR (Figure 4c and d) methods are reported. The balanced epPCR method shows lower probabilities for substitution into charged residues when compared to the unbalanced epPCR method. The unbalanced epPCR is less transition biased (AT→GC), which results in higher probability of substitutions to charged residues (for E coded by GAG and D coded by GAC; a transition mutation leads often to a substitution into glycine (GGC, GGG)). These effects of mutagenesis methods due to mutational preferences might be minimized by codon optimization like for E using GAA and for D using GAT codon.

Using the new focused analysis feature of the server, the amino acid substitution patterns for active site residues (Y223, Y238, and R285) were also evaluated. Y223 and Y238 are involved in substrate binding and product release, while R285 forms a pair with the carboxylate portion of the substrate (arginine) in RgDAAO.[31] R285 has a very low residue mutability indicator ($\mu(285) < 0.3$) (*i.e.*, low probability of substitution leading to amino acid change) for both methods. Y223 and Y238 have $\mu(223/238) = 0.9$ and therefore have a higher probability to be substituted into another amino acid. For the balanced epPCR, Y223 and Y238 are preferentially substituted into charged ($\delta(223/238)_{Y\rightarrow ch} = 0.37$) and neutral ($\delta(223/238)_{Y\rightarrow ne} = 0.46$) amino acids. In the unbalanced epPCR, the chemical diversity at Y223/238 is more preserved ($\delta(223/238)_{Y\rightarrow ne} = 0.44$ and $\delta(223/238)_{Y\rightarrow ar} = 0.31$). The tendency toward the substitution of active site aromatic residues into chemically different amino acids might result in an increased number of inactive clones in the mutant library.

In summary, MAP[2.0]3D provides qualitative indication that the balanced epPCR method might be less beneficial (or of lower quality) than the unbalanced one in the directed evolution of *RgDAAO*.

RgDAAO Directed Evolution. In one directed evolution study by Pollegioni *et al.*,[32] the substrate specificity of RgDAAO was altered to formulate it as a biosensor for analytical determination of D-amino acid in biological samples. Two rounds of directed evolution were performed employing epPCR mutant libraries (balanced dNTP) followed by another round of directed evolution employing epPCR (unbalanced dNTP) for diversity generation. In the first round (1st set of epPCR, balanced) and the second round (2nd set of epPCR, unbalanced), 91% and 63%, respectively, of clones were reported to be inactive. The results of these experiments are in agreement with the predictions of the MAP[2.0]3D server. In fact, mutational preferences of the balanced method induce more structural destabilizing substitutions and resulted in a higher number of inactive clones than balanced epPCR. In addition, MAP[2.0]3D analysis suggests that most of the inactive

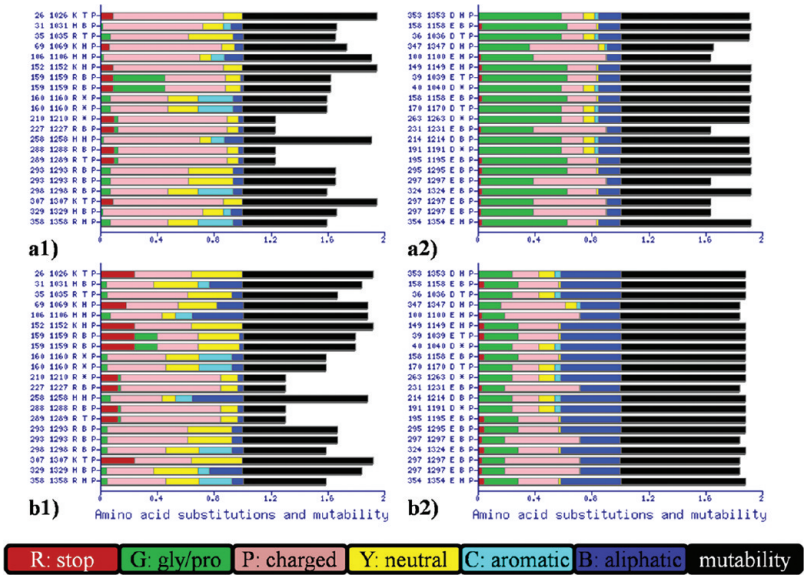


Figure 4. Chemical diversity and mutability of charged amino acid positions of RgDAAO (E, D, R, K, H) that are involved in salt bridges formation (a and b) for epPCR (Taq ($MnCl_2$, G=A=C=T)) and (c and d) for epPCR (Taq ($MnCl_2$, G=A, C=T)). *Y*-axis shows (i) residue sequence id, (ii) PDB id, (iii) residue name, (iv) secondary structure elements (H, α helix; B, beta bridge and extended strand; T, hydrogen bonded turn and bend; *, loop or irregular structure), and (v) amino acid category according to the chemical property of its side chain (P, charged; Y, neutral; C, aromatic; B, aliphatic) with stop codon (R) and Gly/Pro (G) as separate classes.

dx.doi.org/10.1021/sb200019x | *ACS Synth. Biol.* 2012, 1, 139–150

clones should be a result of substitutions into Gly/Pro (destabilizing amino acids), which can destabilize the secondary structure of a helix or weaken intramolecular interactions.

The best variant obtained from the experiments was the triple mutant (T60A Q144R K152E) with broader substrate specificity. Amino acid substitution patterns calculated by the MAP$^{2.0}$3D server at these positions were also found in agreement with experimental results. All mutated positions were assigned by MAP$^{2.0}$3D with a high residue mutability value ($\mu(60/144/152) > 0.8$), *i.e.*, within mutagenic hotspots generated by mutagenesis methods. Q144R substitution was identified in the first round of the balanced epPCR. Q144 has a high probability to be substituted into charged residue ($\delta(144)_{Q\rightarrow ch} = 0.67$), and experimentally the Q144R ($\varphi(144)_{Q\rightarrow R} = 0.58$) substitution was found. In the second round of random mutagenesis with unbalanced epPCR, T60A ($\varphi(60)_{T\rightarrow A} = 0.58$) and K152E ($\varphi(152)_{K\rightarrow E} = 0.36$) were substituted. Both residues have a high preference to be substituted into aliphatic ($\delta(60)_{T\rightarrow al} = 0.6$) and charged residues ($\delta(152)_{K\rightarrow ch} = 0.7$), respectively.

In summary, the RgDAAO case illustrates how the MAP$^{2.0}$3D server can be used in developing efficient mutagenesis strategies before and during directed evolution experiments by, for example, the selection of the most efficient mutagenesis method for the target gene with the least unfavorable effects on its protein structure or function and enabling codon engineering. In this way, a gene can be synthesized prior to the directed evolution experiment to reduce highly destabilizing substitutions at key amino acid positions.

Phytase. Phytase is a class of phosphatase enzymes that catalyzes the hydrolysis of phytic acid (myoinositol hexakisphosphate) to release inorganic phosphorus in a usable form. Phytases have been used as a feed supplement since decades.[33] Application of phytases in industrial feed pelleting process requires high temperatures. For this reason, directed evolution methods have been used to increase thermal resistance of phytases while maintaining high activity at ambient temperature.[34]

MAP$^{2.0}$3D Analysis. MAP$^{2.0}$3D analysis was performed on the phytase *appA2* (full analysis is given in MAP$^{2.0}$3D server as an example). In comparison to other 18 random mutagenesis methods, epPCR Taq (+, G=A, C=T) was found to be the preferred choice for directed *appA2* evolution. In fact, as reported in Table 1, the sequence based MAP$^{2.0}$3D analysis shows frequency of stop codons $I_{\alpha\rightarrow st} = 4.69\%$ and substitutions into Gly/Pro $I_{\alpha\rightarrow gp} = 11.60\%$. An average of 7.45 amino acid substitutions per residue was calculated. The value of codon diversity coefficient was 36.49% and resulted in preserved amino acid substitutions $I_{\alpha\rightarrow pr} = 25.40\%$. Charged (19.21%) and aromatic (5.69%) residues were overrepresented with 1.39% and 0.91% deviation from their chemical distribution, respectively. The aliphatic (39.35%) and neutral (35.65%) residues were underrepresented with −2.86% and −4.14% deviation, respectively.

By using the structural data a different conclusion emerges in contrast to the sequence analysis alone. One rule of thumb, used to enhance the thermostability of an enzyme, is to increase the number of charged residues in the loop regions at the protein surface. The reduction of mobility of these flexible regions by strengthening with electrostatic and hydrogen bonding interactions usually has a stabilizing effect on the thermal stability.[35] Hence, the amino acid substitution patterns of charged residues were analyzed using the residue mutability indicator, the normalized B-factors (B′) as a residue flexibility indicator and the relative solvent accessibility (RSA) to differentiate exposed and buried residues. In Figure 5, the

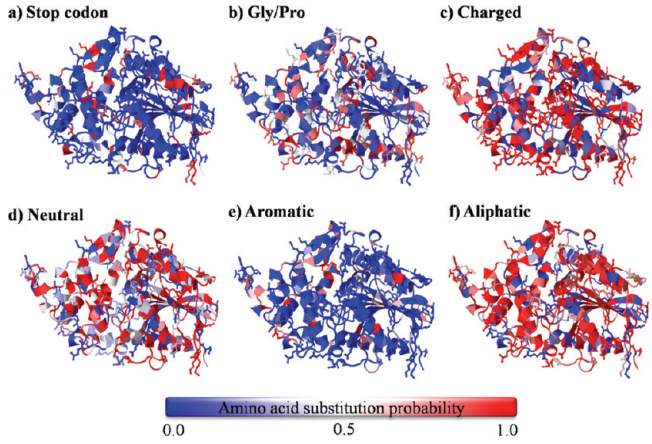


Figure 5. MAP$^{2.0}$3D analysis of amino acid substitutions probability of phytase *appA2* after being subjected to epPCR (Taq ($MnCl_2$, G=A, C=T) in cartoon representation; charged residues (D, E, H, K, R) shown in stick representation. The probability values increase from blue (lowest probability) to red (highest probability). Amino acids were grouped according to the chemical nature of their side chain: charged (c), neutral (d), aromatic (e), or aliphatic (f) with sequence interrupting (stop codons (a)) and structure destabilizing amino acids (glycine and proline (b)).

mapping of amino acid substitution patterns, generated by epPCR Taq (+, G=A, C=T), for different amino acid substitution classes (charged, neutral, aromatic, and aliphatic), stop codon, and Gly/Pro on the phytase *appA2* is reported with charged residues represented in stick representation. The high probability of charged residue substitutions into Gly/Pro, aliphatic, and neutral residues were observed in MAP$^{2.0}$3D analysis. In Figure 6, the detailed information of amino acid substitution patterns for charged residues is reported with three MAP$^{2.0}$3D structural indicators for the epPCR Taq (+, G=A, C=T) method. The experimentally determined mutations are highlighted with black rectangles in Figure 6. Most of the charged residues were found with mutability value $\mu > 0.6$, *i.e.*, high substitution probability to change into another amino acid. In Figure 6, the high probabilities were evident to substitute from charged residues into glycine or proline (α helix destabilizers), aliphatic, and neutral residues (less favorable to improve thermostability).

Phytase Directed Evolution. In one example, Kim *et al.* performed directed evolution on phytase *appA2* from *E. coli* to generate variants with increased thermostability by using epPCR with unbalanced dNTPs.[36] Two variants (K46E and K65E K97M S209G) with 20% improved thermostabilty were found after screening 5000 clones. Out of four positions, three resulted from charged residue substitutions that occurred at lysine residues.

MAP$^{2.0}$3D analysis of amino acid substitution patterns for these positions was in agreement with experimental findings with all four positions having a high mutability indicator value ($\mu > 0.8$) and relative solvent accessibility (RSA > 0.4). Furthermore, all lysine residues in the mutated positions have a probability for a nucleotide exchange that results in a stop codon. K46 and K97 have the same amino acid substitution

dx.doi.org/10.1021/sb200019x | *ACS Synth. Biol.* 2012, 1, 139–150

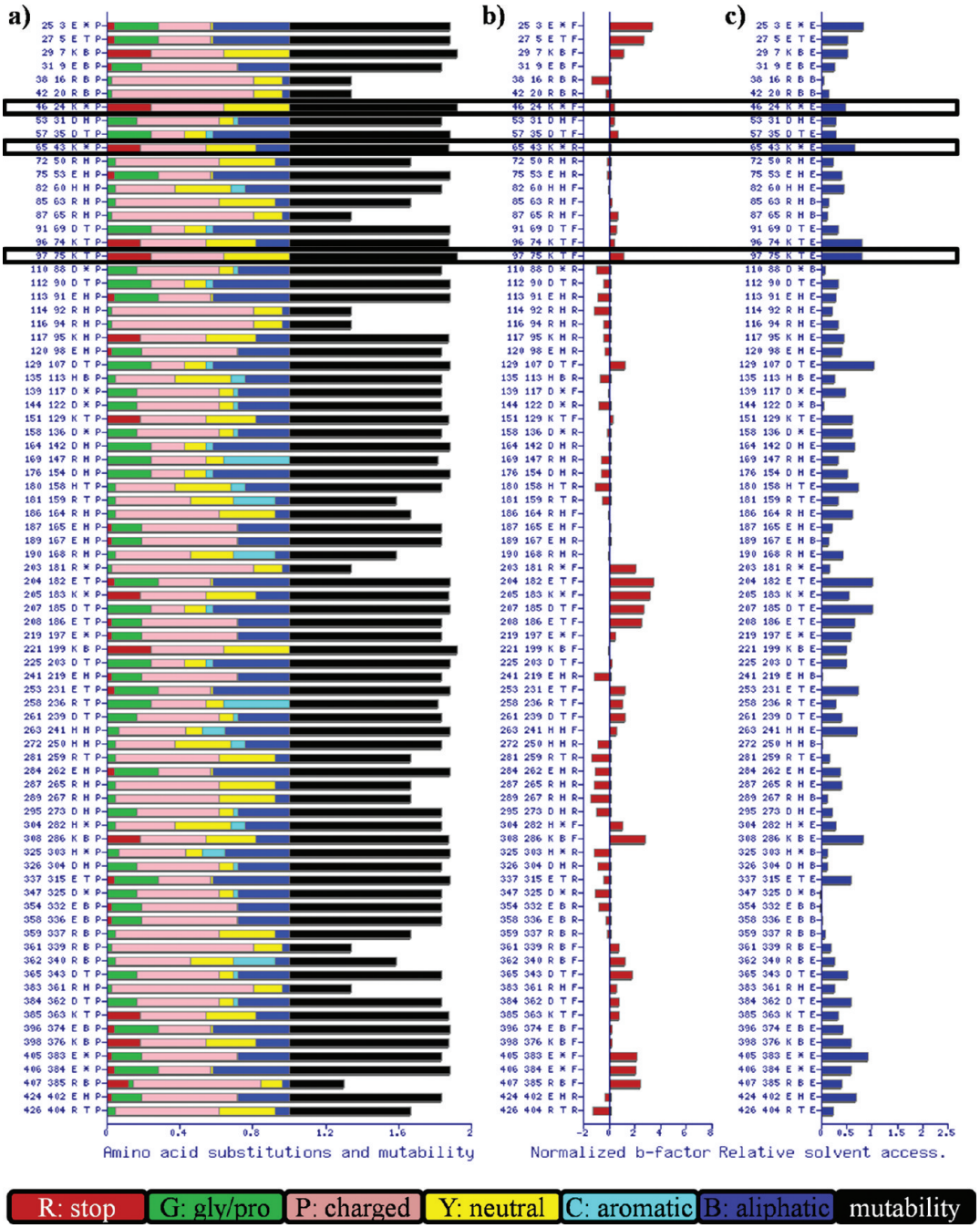


Figure 6. Amino acid substitution patterns for charged residues in phytase with the performance of following parameters: residue mutability, residue flexibility, and relative solvent accessibility of amino acids. The experimentally determined mutations are highlighted in black boxes. *Y*-axis shows sequence id, PDB id, amino acid name, and in (a) secondary structure elements (H, α helix; B, beta bridge and extended strand; T, hydrogen bonded turn and bend; *, loop or irregular structure), (b) normalized Cα B-factor to differentiate between flexible (F) and rigid (R) residues, and (c) relative solvent associability to identify exposed (E) or buried (B) residues.

patterns with a substitution preference to stop codon ($\delta(46/97)_{K\rightarrow st}$ = 0.24) but differ for charged ($\delta(46)_{K\rightarrow ch}$ = 0.40; $\varphi(46)_{K\rightarrow E}$ = 0.16) and neutral residues ($\delta(97)_{K\rightarrow ne}$ = 0.35; $\varphi(97)_{K\rightarrow M}$ = 0.24). K65 has different amino acid substitution values to change into residues with aliphatic ($\delta(65)_{K\rightarrow al}$ = 0.18), charged ($\delta(65)_{K\rightarrow ch}$ = 0.36; $\varphi(65)_{K\rightarrow E}$ = 0.12), and neutral ($\delta(65)_{K\rightarrow ne}$ = 0.27) side chains and $\delta(65)_{K\rightarrow st}$ = 0.18 for stop codon. S209 has a high probability to preserve the chemical property of its side chain and has high preference to neutral substitutions ($\delta(209)_{S\rightarrow ne}$ = 0.60). S209 substitution into glycine alone has probability $\varphi(209)_{S\rightarrow G}$ = 0.24. The mutations generated by using the epPCR Taq (+, G=A, C=T) mutagenesis method experimentally resulted in only 20% active clones in the library, and only 80 were found improved in thermal stability. Phytase *appA2* has a high helical content (42%), and substitutions into Gly/Pro residues might reduce thermal stability by destabilizing the structure and increasing the number of inactive clones. In general, amino acid substitutions of charged residues into aliphatic or neutral residues are less favorable to improve thermal stability.

***N*-Acetylneuraminic Acid Aldolase.** *N*-Acetylneuraminic acid aldolase (Neu5Ac aldolase) catalyzes the aldol con-

dx.doi.org/10.1021/sb200019x | *ACS Synth. Biol.* 2012, 1, 139–150

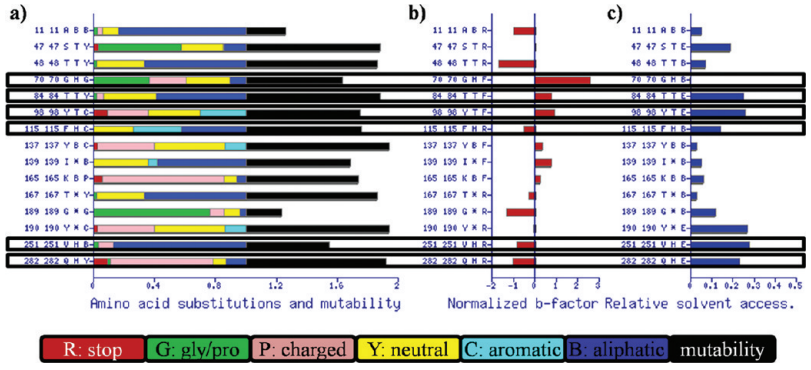


Figure 7. Amino acid substitution patterns for active site residues (A11, S47, T48, Y137, I139, K165, T167, G189, Y190) of Neu5Ac aldolase and experimentally determined mutations (1st generation, Y98H, P115L; 2nd generation, V251I; 3rd generation, G70A, T84S, Q282L) in the boxes for random mutagenesis method: epPCR (Taq ($MnCl_2$, G=A=C=T). *Y*-axis representations are the same as described in Figure 6.

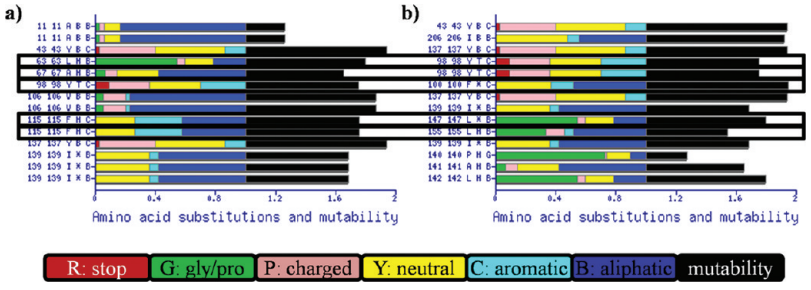


Figure 8. Amino acid substitution patterns for active site residues (A11, Y137, I139) of Neu5Ac aldolase and mutations (1st generation, Y98H and P115L; highlighted in the box frames) involved in hydrophobic interactions. Panels a and b show the interaction partners for hydrophobic interaction. *Y*-axis representations are the same as described in Figure 4.

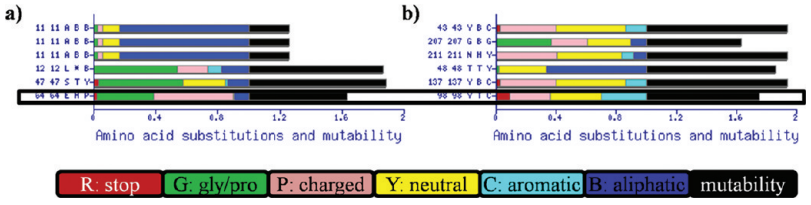


Figure 9. Amino acid substitution patterns for active site residues (A11, S47, T48, Y137) of Neu5Ac aldolase and mutation (1st generation, Y98H; highlighted in a black box) involved in side chain hydrogen bond. Panels a and b show the interaction partners for side chain hydrogen bond. *Y*-axis representations are the same as described in Figure 4.

densation of *N*-acetyl-D-mannosamine and pyruvate to give *N*-acetyl-D-neuraminic acid (D-sialic acid).[37] Neu5Ac aldolase is used in the synthesis of sialic acid, a complex sugar with many pharmaceutical applications.

MAP$^{2.0}$3D Analysis. On the basis of the sequence based analysis of the MAP$^{2.0}$3D server, the balanced epPCR method (Taq ($MnCl_2$, G=A=C=T) was found suitable for directed evolution of Neu5Ac aldolase (summarized in Table 1). For this method, the value of the codon diversity coefficient was 43.12, which resulted in $I_{\alpha\rightarrow pr}$ = 28.47% preserved amino acid substitutions with an average of 7.20 amino acid substitutions per residue. The frequency for stop codons was $I_{\alpha\rightarrow st}$ = 2.12%, and for Gly/Pro substitutions $I_{\alpha\rightarrow gp}$ = 16.26% was reported. The structure based analysis was focused on active site residues (A11, S47, T48, Y137, I139, k165, T167, G189, Y190) using the new option of the MAP$^{2.0}$3D server to restrict the analysis to selected amino acids. Figure 7 shows the expected amino acid substitutions for active site residues and, highlighted in boxes, experimentally determined mutation positions[37] (G70, T84, Y98, F115, V251, E282). With the exception of residues A11 and G189, the other active site residues have a residue mutability value (μ > 0.6). The values of the RSA and B′ indicate that A11 and G189 are buried in the protein active site and highly rigid. The residue I139, another aliphatic residue of the active site, resulted in a moderately high preference of substitution into a neutral amino acid ($\delta(139)_{I\rightarrow ne}$ = 0.36). The active site residues Y137 and Y190 have high residue mutability value ($\mu(137/190)$ = 0.94) and substitute into charged ($\delta(137/190)_{K\rightarrow ch}$ = 0.37) or neutral ($\delta(137/190)_{K\rightarrow ne}$ = 0.46) amino acids. S47 has mutability value μ = 0.88 with a substitution probability $\varphi(47)_{S\rightarrow G}$ = 0.6 to change into glycine. K165 has preference ($\mu(165)$ = 0.73) to substitute into charged residues ($\varphi(165)_{K\rightarrow R/K/E}$ = 0.26).

Figures 8 and 9 show the analysis of hydrophobic contacts and hydrogen bonds for active site residues and experimentally determined mutation positions, respectively. The results of the analysis highly suggest an involvement of A11 in hydrophobic interactions with Y43 or I206 (see Figure 8) and a side-chain hydrogen bond formation with Y43 or G207 or N211 (see Figure 9). In short, the substitution spectra analysis of active site residues (A11, S47, T48, Y137, I139, K165, T167, G189, Y190) indicates that the chemical environment of active site

dx.doi.org/10.1021/sb200019x | *ACS Synth. Biol.* 2012, 1, 139–150

residues is not substantially modified by the epPCR random mutagenesis method.

Neu5Ac Aldolase Directed Evolution. Neu5Ac aldolase was engineered applying epPCR with balanced dNTP for a complete reversal of enantioselectivity by Wada *et al.*[37] Three rounds of random mutagenesis resulted in a variant more effective toward both D/L enantiomeric substrates (3-deoxy-L/D-manno-2-octulosonic acid). The two mutation positions (Y98H and F115L, see Figure 7) from the first round of random mutagenesis were found to be involved in hydrophobic interactions (Figure 8, from Y98 to L63/A67/F100 and from F115 to L147/L155) and in side chain hydrogen bonds (Figure 9, from E64 to Y98) formation in wild type. Y98H and F115L were present outside the active site, partially exposed to the solvent with relative solvent accessibility value (RSA = 0.26), moderately flexible with normalized B-factor (B′ = 0.91) and "variable" amino acid substitutions with residue mutability indicator $\mu(98/115)$ = 0.75. The substitutions at these positions into comparatively more hydrophobic residues resulted in increased activity of the wild type aldolase. In MAP$^{2.0}$3D, Y98 is preferably preserved or substitute into charged ($\delta(98)_{Y\rightarrow ch}$ = 0.27; $\varphi(98)_{Y\rightarrow H}$ = 0.25) or neutral ($\delta(98)_{Y\rightarrow ne}$ = 0.33) residues, whereas F115 shows a slightly higher preference toward aliphatic substitution ($\delta(115)_{F\rightarrow al}$ = 0.42; $\varphi(115)_{F\rightarrow L}$ = 0.33) and cannot be substituted into charged residues ($\delta(115)_{F\rightarrow ch}$ = 0.00). The substitution at V251 residue was obtained in the second round of directed evolution experiment and resulted in partially inverted enantiomeric preference of the enzyme. The position was more conserved in MAP$^{2.0}$3D analysis with $\mu(251)$ = 0.54 or substituted into more hydrophobic residues. The third generation mutations (G70A, T84S, Q282L) resulted in a complete reversal of enzymatic enantioselectivity for use in the synthesis of both D- and L-sugars. G70 has a high flexibility (B′ = 2.59) and low probability to be substituted into an aliphatic residue ($\delta(70)_{G\rightarrow al}$ = 0.10; $\varphi(70)_{G\rightarrow A}$ = 0.03). Thr84 is a part of a turn with a high flexibility (B′ = 0.77) and exposure to the solvent (RSA = 0.25) with the residue mutability $\mu(84)$ = 0.87. T84 has a high preference for being substituted by aliphatic residues ($\delta(84)_{T\rightarrow al}$ = 0.58) and to a lesser extent by a "neutral" amino acid ($\delta(84)_{T\rightarrow ne}$ = 0.34; $\varphi(84)_{T\rightarrow S}$ = 0.13). Q282 is a part of a helix, is rigid (B′ = −1.01) but partially exposed to the solvent (RSA = 0.23). Q282 has a high preference to be substituted into a charged residue ($\delta(282)_{Q\rightarrow ch}$ = 0.67) and very low probability for aliphatic ($\delta(282)_{Q\rightarrow al}$ = 0.13; $\varphi(282)_{Q\rightarrow L}$ = 0.13) or neutral ($\delta(282)_{Q\rightarrow ne}$ = 0.08) substitution.

In the case of aldolase, the MAP$^{2.0}$3D analysis shows also a good agreement with experimental results. The variability in amino acid substitution patterns for active site residues resulted in exploring more sequence space for catalytic activity of the enzyme and resulted in obtaining a high fraction of beneficial mutations in first generation.

Conclusions. In this article, we introduced the MAP$^{2.0}$3D server and its use to assist the design of directed evolution experiments. MAP$^{2.0}$3D correlates the traditional sequence based MAP indicators with the structural information of the target protein. The combined information can help to improve the chances to find functional and stable enzyme variants. MAP$^{2.0}$3D helps to guide the directed evolution experiments by focusing the analysis on a set of residues that are important for specific enhancement of enzymatic properties such as to improve substrate specificity by targeting residues located in or near the active site or to enhance thermal stability or water solubility of proteins by increasing the number of charged amino acid substitutions. The new structure oriented features of the MAP$^{2.0}$3D server have been applied to the analysis of three different proteins (phytase, oxidase, and aldolase), and the predicted results were compared with the experimental results. The results of RgDAAO analysis indicate that the selection of the random mutagenesis method by the prescreening of the generated library can help to elucidate the effects of mutational bias on the structural environment of the protein and how these effects can be optimized. The analysis of phytase and Neu5Ac aldolase illustrate how the structural analysis features included in the MAP$^{2.0}$3D server can now assist to correlate the effect of mutational biases with protein structural environment and to evolve desired property. In this way, MAP$^{2.0}$3D server facilitates the *in silico* prescreening of the target gene and can also promote an increase of the active population in random mutagenesis libraries, thereby decreasing screening efforts and increasing the probability for obtaining desirable mutations even in the small mutant library.

METHODS

Mutational Probability and Statistics. The MAP$^{2.0}$3D server performs statistical analysis on a given nucleotide sequence based on the mutational spectra of different random mutagenesis methods that were slightly elaborated as follows to be used in the analysis.[17] First, insertions and deletions with an occurrence frequency between 0.8% and 13.9% were neglected, and remaining nucleotide substitution frequencies were scaled proportionally to 100%. Second, mutations in upper and lower DNA strands were considered to occur with equal frequency. The scaled mutational frequencies are used in the analysis to calculate the probability of amino acid substitutions resulting from one nucleotide exchange in one codon of the gene. The analysis is performed as follows. Consider a gene coding for a protein of L amino acids. For each nucleotide of a codon (named as *X*,*Y*,*Z*) in the gene sequence, the corresponding single nucleotide substitutions *X*′,*Y*′,*Z*′ (with $\{X, Y, Z, X', Y', Z' \in \{A, T, G, C\} \mid X' \neq X, Y' \neq Y, Z' \neq Z\}$) are considered. For each one of the 19 random mutagenesis methods, matrix **P** (in eq 1) gives the 16 mutational probability values for the given nucleotide substitution into another three (*e.g.*, *X*→*X*′). The values of matrix **P** have been already reported in Table 1 of our previous publication.[17]

$$\mathbf{P} = \begin{pmatrix} A\rightarrow A & A\rightarrow T & A\rightarrow G & A\rightarrow C \\ T\rightarrow A & T\rightarrow T & T\rightarrow G & T\rightarrow C \\ G\rightarrow A & G\rightarrow T & G\rightarrow G & G\rightarrow C \\ C\rightarrow A & C\rightarrow T & C\rightarrow G & C\rightarrow C \end{pmatrix} \quad (1)$$

In eq 2, the binary vectors **U** and **V** are then used to select a given probability (*f*) from the matrix **P**. The four elements of **U** and **V** correspond to the nucleotide (*A*, *T*, *G*, *C*) that can be selected by assigning a value of 1 or 0. **U** selects the original nucleotide, and **V** the mutated one. In eq 2, an example for the epPCR method with the Taq-Polymerase (unbalanced dNTPs) is given as matrix **P**. In this example, the mutational probability

dx.doi.org/10.1021/sb200019x | *ACS Synth. Biol.* 2012, 1, 139−150

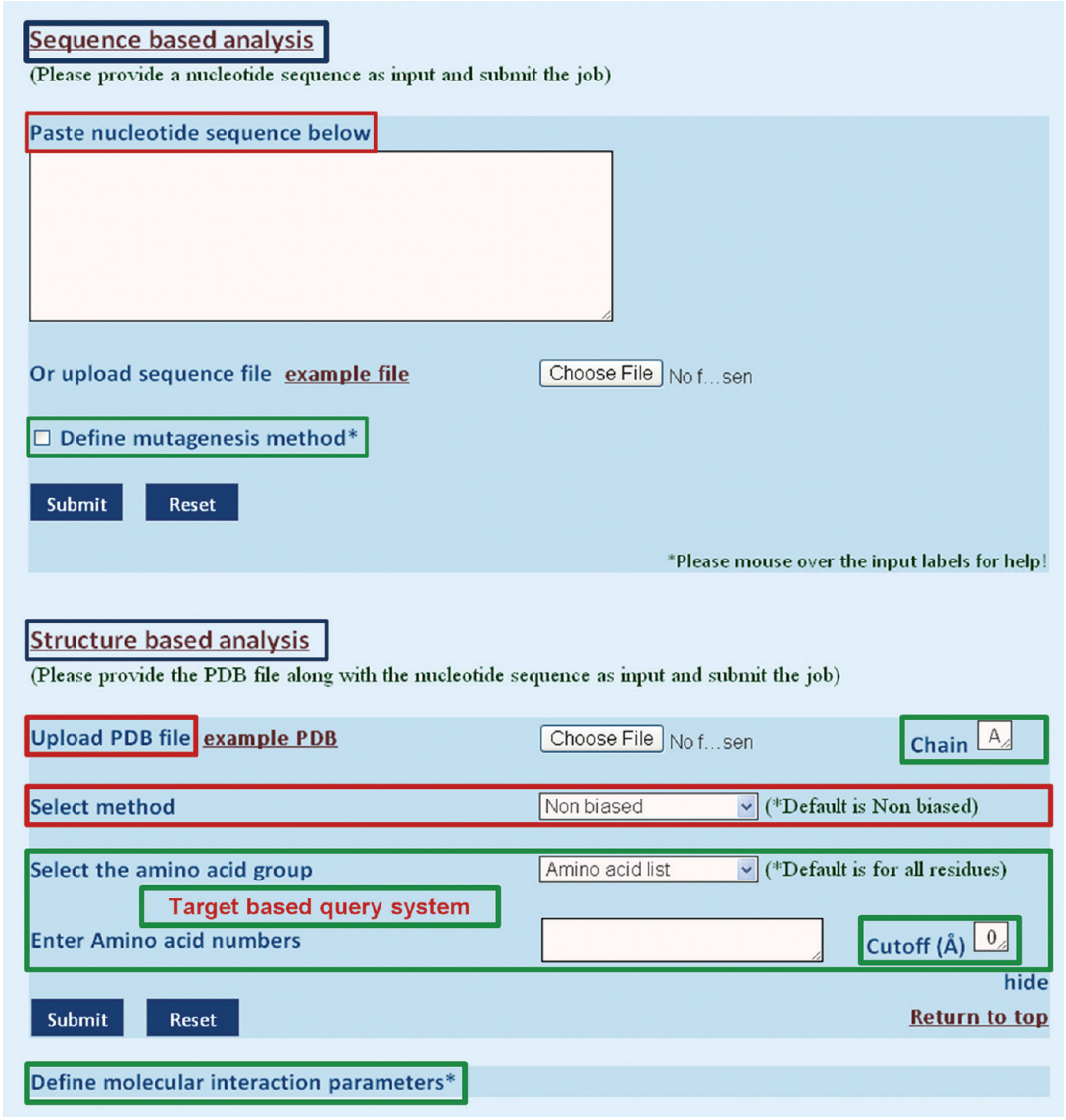


Figure 10. Query interface for MAP$^{2.0}$3D. Black boxes show two ways to query the server: (1) sequence based analysis that takes nucleotide sequence as an input (red box) and (2) structure based analysis, which takes protein coordinates (crystallographic structure or homology model), nucleotide sequence, and a random mutagenesis method as input (red boxes). The options given in the green boxes can be used to customize the query, for example: (1) 19 commonly used mutagenesis methods are included in the server as default, and a new method can be included by defining its mutational spectra, (2) selection of chain in the case of multi chain proteins, (3) restriction of the search for a group of amino acids either selecting the predefined groups based on (a) the chemical property of their side chain like charged, neutral, aromatic, and aliphatic, (b) the solvent accessible area such as buried or exposed, and (c) the given set of amino acids and definition of cut off (in Å) to include residues in the defined diameter of given residues in the analysis, and (4) alteration of the threshold used for the calculation of molecular interactions.

for the transformation of nucleotide $A \rightarrow T$ gives a value of $f = 9.7$.

$$f = \mathbf{UPV}^{\mathbf{T}} = (1\ 0\ 0\ 0)\begin{pmatrix} 0.0 & 9.70 & 19.34 & 16.14 \\ 9.70 & 0.0 & 16.14 & 19.34 \\ 4.82 & 0.0 & 0.0 & 0.0 \\ 0.0 & 4.82 & 0.0 & 0.0 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = 9.7 \quad (2)$$

By applying this procedure to each single nucleotide substitution in the codon, nine probability values (three for each nucleotide) are obtained. Each of these values gives the kth mutational probability ($f(i)^k_{\alpha\rightarrow\beta}$) that change the ith amino acid (α) expressed by the native codon into the one (β, which also comprises the stop codon) expressed by the mutated codon (*e.g.*, $X,Y,Z \rightarrow X',Y,Z$). Therefore, the 9 probabilities ($f(i)^k_{\alpha\rightarrow\beta}$) are summed to get the normalization factor (N_i) for the ith residue of the protein sequence:

$$N_i = \sum_{k=1}^{9} f(i)^k_{\alpha\rightarrow\beta} \quad (3)$$

Hence, the normalized probability for the substitution of amino acid $\alpha \rightarrow \beta$ is given by

$$\varphi(i)^k_{\alpha\rightarrow\beta} = \frac{f(i)^k_{\alpha\rightarrow\beta}}{N_i} \quad (4)$$

dx.doi.org/10.1021/sb200019x | *ACS Synth. Biol.* 2012, 1, 139–150

MAP indicators. Three indicators, protein structure indicator, amino acid diversity indicator, and chemical diversity indicator, are used to summarize the characteristics of a random mutagenesis method for the target gene on the amino acid level. The amino acid diversity for the substitution of amino acid $\alpha \rightarrow \beta$ in the protein sequence (L) is calculated by

$$\Delta_{\alpha\rightarrow\beta} = \frac{1}{L}\sum_{i=L}^{L} \varphi(i)_{\alpha\rightarrow\beta} \tag{5}$$

The amino acid diversities are summed together to calculate the values for MAP indicators:

$$I_{\alpha\rightarrow \boldsymbol{S}} = \sum_{r=1}^{r'} \Delta^{r}_{\alpha\rightarrow\beta(r)} \tag{6}$$

where $\boldsymbol{S}$ indicates different subset of amino acids or stop codons and r' represents the elements in these subsets. The chemical diversity indicator quantifies the generated chemical diversity by the random mutagenesis method. For this indicator, $\boldsymbol{S}$ consists of one of the subset of amino acids: charged (D, E, H, K, R; $\boldsymbol{S}$ = ch), neutral (C, M, S, P, T, N, Q; $\boldsymbol{S}$ = ne), aromatic (F, Y, W; $\boldsymbol{S}$ = ar), and aliphatic (G, A, V, L, I; $\boldsymbol{S}$ = al). For example, $I_{\alpha\rightarrow \mathrm{ch}}$ indicates the total probability of a given amino acid α to substitute into charged amino acids (ch) is calculated by $\Delta_{\alpha\rightarrow\beta(r)}$, where the substituted residue $\beta(r)$ can be one of the charged residues (E, D, R, K, and H), *i.e.* $r' = 5$. The protein structure indicator signifies the fraction of single nucleotide substitution resulting in protein structure/function-disrupting (stop codons; $\boldsymbol{S}$ = st) and likely destabilizing (glycine or proline; $\boldsymbol{S}$ = gp) amino acid substitutions. Finally, the amino acid diversity indicator measures the fraction of variants with preserved amino acid substitutions ($\boldsymbol{S}$ = pr) and average amino acid substitutions per residue. This is complemented by a codon diversity coefficient that measures the distribution of random mutations among the codons of the gene.

Local Chemical Diversity and Protein Structure Components. Two new sequence based indicators are introduced with the MAP$^{2.0}$3D server to complement the single amino acid structural analysis. The substitution probability of the ith amino acid (α) that leads to change in the amino acid (β) with the side chain of the same chemical nature is calculated by

$$\delta(i)_{\alpha\rightarrow \boldsymbol{S}} = \sum_{r=1}^{r'} \varphi(i)^{r}_{\alpha\rightarrow\beta} \tag{7}$$

where, x and r' represents the amino acid group and its members, respectively (as described for eq 6). The amino acid mutability of the ith amino acid (a special case of the eq 7 with $\acute{r} = 1$) is given by

$$\mu(i) = 1 - \varphi(i)_{\alpha\rightarrow\alpha} \tag{8}$$

where $\varphi(i)_{\alpha\rightarrow\alpha}$ is the normalized probability for the substitution does not lead to an amino acid change $(\alpha\rightarrow\alpha)$ at the ith residue. The local structure environment of the amino acid residue influences the acceptance of the amino acid substitutions.[23,24] The local structural environment of the protein comprises the secondary structure element, residue flexibility, and solvent accessibility. Intraprotein interactions contribute to define secondary structure elements and residue flexibility in a target protein and aid in understanding the molecular basis of the stability and activity of the protein.[38] To illustrate the effect of generated chemical diversity on the protein structural environment, these factors are mapped with amino acid substitution patterns.

The secondary structure elements are derived using DSSP,[39] while relative solvent accessibility (RSA) has been calculated by the number of water molecules in contact with the residue[39] divided by total surface area of the residue.[40] A threshold value of 0.16 is used to differentiate between exposed (RSA $\geq$ 0.16) or buried residues (RSA < 0.16). Crystallographic B-factors are used as indicators of the residue flexibility.[41] The B-factors of Cα atoms are normalized by

$$\mathrm{B}' = \frac{(\mathrm{B} - \langle \mathrm{B} \rangle)}{\sigma} \tag{9}$$

where $\langle \mathrm{B} \rangle$ is the average value for the Cα atom (after omitting first and last 3 residues) and σ is the standard deviation.[42] The relative B-factor values after normalization is employed to differentiate flexibility and rigidity of the residue.[43]

Finally, the new server calculates from the crystallographic protein structure, using criteria reported in literature, the following intraprotein interactions: disulfide bonds,[39] salt bridge,[44] hydrophobic interaction,[45] aromatic interaction,[46] and side chain hydrogen bond.[47] The default parameters are taken from the widely accepted primary literature for the calculation of molecular interactions and can be modified by the user.

MAP$^{2.0}$3D Server Description. MAP$^{2.0}$3D analysis was performed on gene sequence along with the 3D coordinates of target protein for a random mutagenesis method at a time. Figure 10 shows the query interface of the server available at http://map.jacobs-university.de/map3d.html.

The server is flexible to accept the gene sequence in commonly used sequence format (fasta, GenBank, GCG) or as the raw sequence. The 3D coordinates are accepted in PDB file format.[48] The protein sequence, after translation from the gene sequence, is aligned with protein sequence, extracted from protein coordinates, by using a Smith−Waterman algorithm[49] for local sequence alignment. For the complete analysis, the sequences should have appropriate identity (default $\geq$ 70%). In the case of multiprotein chain files, the analysis performs on the first chain or can be defined by the user. The analysis is performed on a user selected mutagenesis method (chosen among the MAP library of commonly used methods or as a feature of the server by directly introducing the values of the probability of transformation matrix **P**). By default the results include the analysis of all residues that can be changed by selecting a predefined group of amino acids (charged, neutral, aromatic and aliphatic or, accordingly to their relative solvent accessibility, exposed or buried) or by providing a set of amino acid residues, which can be extended to residues within a given range (in Å) from the given set of amino acids. Finally, the advanced user interface section allows changing parameters used for the calculation of molecular interactions.

MAP$^{2.0}$3D Output. Along with the sequence based MAP analysis indicators, the implemented indicators in MAP$^{2.0}$3D correlate the generated amino acid substitution patterns of random mutagenesis methods to the protein structure (by using the Jmol applet, http://www.jmol.org/) and include a residue mutability indicator and taking secondary structure elements, residue flexibility, relative solvent accessibility, and intraprotein interactions into account (see above). Generated results are also available to download for further use in text

format. The modified coordinate files (with amino acid substitution probabilities) in PDB format are also available as downloads.

Model Proteins. The enzymes selected for the analysis by MAP[2.0]3D are (1) D-amino acid oxidase from *Rhodotorula gracilis* (EC 1.4.3.1; EMBL-Bank AAB93974.1;[50] PDB id 1C0I[31]), (2) phytase from *Escherichia coli* (EC 3.1.3.2; EMBL-Bank AY496073.1;[51] PDB id 1DKP[52]), (3) *N*-acetylneuramine acid aldolase from *Escherichia coli* (EC 4.1.3.3; EMBL-Bank X03345.1;[53] PDB id 1NAL[54]). The sequence composition of the enzymes: (1) D-amino acid oxidase (1107 bases: A 19.96%; T 17.52%; G 31.17%; C 31.35%; 369 residues), (2) Phytase (1299 bases: A 24.25%; T 22.09%; G 27.25%; C 26.40%; 433 residues), and (3) *N*-acetylneuraminic acid aldolase (894 bases: A 24.38%; T 23.60%; G 27.07%; C 24.94%; 298 residues). Secondary structure of the enzymes: (1) D-amino acid oxidase (30% helical, 28% β sheet), (2) phytase (42% helical, 15% β sheet), (3) *N*-acetylneuraminic acid aldolase (50% helical, 13% β sheet).

AUTHOR INFORMATION

Corresponding Author

*Tel: +49 421 200-3144. Fax: +49 421 200-3249. E-mail: d.roccatano@jacobs-university.de.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the European Union seventh framework program (project "OXYGREEN", Project Reference 212281) for financial support.

REFERENCES

(1) Bornscheuer, U. T., and Pohl, M. (2001) Improved biocatalysts by directed evolution and rational protein design. *Curr. Opin. Chem. Biol. 5*, 137−143.

(2) Brakmann, S. (2001) Discovery of superior enzymes by directed molecular evolution. *ChemBioChem 2*, 865−871.

(3) Wong, T. S., Arnold, F. H., and Schwaneberg, U. (2004) Laboratory evolution of cytochrome p450 BM-3 monooxygenase for organic cosolvents. *Biotechnol. Bioeng. 85*, 351−358.

(4) Roccatano, D., Wong, T. S., Schwaneberg, U., and Zacharias, M. (2006) Toward understanding the inactivation mechanism of monooxygenase P450 BM-3 by organic cosolvents: a molecular dynamics simulation study. *Biopolymers 83*, 467−476.

(5) Wong, T. S., Zhurina, D., and Schwaneberg, U. (2006) The diversity challenge in directed protein evolution. *Comb. Chem. High Throughput Screening 9*, 271−288.

(6) Wong, T. S., Roccatano, D., and Schwaneberg, U. (2007) Steering directed protein evolution: strategies to manage combinatorial complexity of mutant libraries. *Environ. Microbiol. 9*, 2645−2659.

(7) Smith, J. M. (1970) Natural selection and concept of a protein space. *Nature 225*, 563−564.

(8) Olsen, M., Iverson, B., and Georgiou, G. (2000) High-throughput screening of enzyme libraries. *Curr. Opin. Biotechnol. 11*, 331−337.

(9) Tawfik, D. S., and Bershtein, S. (2008) Advances in laboratory evolution of enzymes. *Curr. Opin. Chem. Biol. 12*, 151−158.

(10) Shivange, A. V., Marienhagen, J., Mundhada, H., Schenk, A., and Schwaneberg, U. (2009) Advances in generating functional diversity for directed protein evolution. *Curr. Opin. Chem. Biol. 13*, 19−25.

(11) Turner, N. J. (2009) Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol. 5*, 567−573.

(12) Wong, T. S., Roccatano, D., Loakes, D., Tee, K. L., Schenk, A., Hauer, B., and Schwaneberg, U. (2008) Transversion-enriched sequence saturation mutagenesis (SeSaM-Tv+): a random mutagenesis method with consecutive nucleotide exchanges that complements the bias of error-prone PCR. *Biotechnol. J. 3*, 74−82.

(13) Dennig, A., Shivange, A. V., Marienhagen, J., and Schwaneberg, U. (2011) OmniChange: The sequence independent method for simultaneous site-saturation of five codons. *PLoS ONE 6*, e26222.

(14) Chica, R. A., Doucet, N., and Pelletier, J. N. (2005) Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol. 16*, 378−384.

(15) Zumarraga, M., Camarero, S., Shleev, S., Martinez-Arias, A., Ballesteros, A., Plou, F. J., and Alcalde, M. (2008) Altering the laccase functionality by in vivo assembly of mutant libraries with different mutational spectra. *Proteins 71*, 250−260.

(16) Vanhercke, T., Ampe, C., Tirry, L., and Denolf, P. (2005) Reducing mutational bias in random protein libraries. *Anal. Biochem. 339*, 9−14.

(17) Wong, T. S., Roccatano, D., Zacharias, M., and Schwaneberg, U. (2006) A statistical analysis of random mutagenesis methods used for directed protein evolution. *J. Mol. Biol. 355*, 858−871.

(18) Wong, T. S., Roccatano, D., and Schwaneberg, U. (2007) Are transversion mutations better? A Mutagenesis Assistant Program analysis on P450 BM-3 heme domain. *Biotechnol. J. 2*, 133−142.

(19) Rasila, T. S., Pajunen, M. I., and Savilahti, H. (2009) Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, Escherichia coli mutator strain, and hydroxylamine treatment. *Anal. Biochem. 388*, 71−80.

(20) Ditursi, M. K., Kwon, S. J., Reeder, P. J., and Dordick, J. S. (2006) Bioinformatics-driven, rational engineering of protein thermostability. *Protein Eng., Des. Sel. 19*, 517−524.

(21) Shoichet, B. K., and Beadle, B. M. (2002) Structural bases of stability-function tradeoffs in enzymes. *J. Mol. Biol. 321*, 285−296.

(22) Berman, H., Henrick, K., and Nakamura, H. (2003) Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol. 10*, 980−980.

(23) Zhang, H., Zhang, T., Chen, K., Shen, S., Ruan, J., and Kurgan, L. (2009) On the relation between residue flexibility and local solvent accessibility in proteins. *Proteins 76*, 617−636.

(24) Teilum, K., Olsen, J. G., and Kragelund, B. B. (2009) Functional aspects of protein flexibility. *Cell. Mol. Life Sci. 66*, 2231−2247.

(25) Gromiha, M. M., Oobatake, M., Kono, H., Uedaira, H., and Sarai, A. (1999) Role of structural and sequence information in the prediction of protein stability changes: comparison between buried and partially buried mutations. *Protein Eng., Des. Sel. 12*, 549−555.

(26) Pilone, M. S. (2000) D-Amino acid oxidase: new findings. *Cell. Mol. Life Sci. 57*, 1732−1747.

(27) Pollegioni, L., and Molla, G. (2011) New biotech applications from evolved D-amino acid oxidases. *Trends Biotechnol. 29*, 276−283.

(28) Lin-Goerke, J. L., Robbins, D. J., and Burczak, J. D. (1997) PCR-based random mutagenesis using manganese and reduced dNTP concentration. *Biotechniques 23*, 409−412.

(29) Vartanian, J. P., Henry, M., and Wain-Hobson, S. (1996) Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. *Nucleic Acids Res. 24*, 2627−2631.

(30) Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/

(31) Pollegioni, L., Diederichs, K., Molla, G., Umhau, S., Welte, W., Ghisla, S., and Pilone, M. S. (2002) Yeast D-amino acid oxidase: structural basis of its catalytic properties. *J. Mol. Biol. 324*, 535−546.

(32) Sacchi, S., Rosini, E., Molla, G., Pilone, M. S., and Pollegioni, L. (2004) Modulating D-amino acid oxidase substrate specificity: production of an enzyme for analytical determination of all D-amino acids by directed evolution. *Protein Eng. Des. Sel. 17*, 517−525.

(33) Rao, D. E., Rao, K. V., Reddy, T. P., and Reddy, V. D. (2009) Molecular characterization, physicochemical properties, known and potential applications of phytases: An overview. *Crit. Rev. Biotechnol. 29*, 182−198.

(34) Garrett, J. B., Kretz, K. A., O'Donoghue, E., Kerovuo, J., Kim, W., Barton, N. R., Hazlewood, G. P., Short, J. M., Robertson, D. E., and Gray, K. A. (2004) Enhancing the thermal tolerance and gastric

performance of a microbial phytase for use as a phosphate-mobilizing monogastric-feed supplement. *Appl. Environ. Microb. 70*, 3041−3046.
(35) Fields, P. A. (2001) Review: Protein function at thermal extremes: balancing stability and flexibility. *Comp. Biochem. Phys. A 129*, 417−431.
(36) Kim, M. S., and Lei, X. G. (2008) Enhancing thermostability of Escherichia coli phytase AppA2 by error-prone PCR. *Appl. Microbiol. Biotechnol. 79*, 69−75.
(37) Wada, M., Hsu, C. C., Franke, D., Mitchell, M., Heine, A., Wilson, I., and Wong, C. H. (2003) Directed evolution of N-acetylneuraminic acid aldolase to catalyze enantiomeric aldol reactions. *Bioorg. Med. Chem. 11*, 2091−2098.
(38) Gromiha, M. M., and Selvaraj, S. (2004) Inter-residue interactions in protein folding and stability. *Prog. Biophys. Mol. Biol. 86*, 235−277.
(39) Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 2577−2637.
(40) Chothia, C. (1976) The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol. 105*, 1−12.
(41) Peisajovich, S. G., and Tawfik, D. S. (2007) Protein engineers turned evolutionists. *Nat. Methods 4*, 991−994.
(42) Karplus, P. A., and Schulz, G. E. (1985) Prediction of chain flexibility in proteins - a tool for the selection of peptide antigens. *Naturwissenschaften 72*, 212−213.
(43) Yuan, Z., Zhao, J., and Wang, Z. X. (2003) Flexibility analysis of enzyme active sites by crystallographic temperature factors. *Protein Eng., Des. Sel. 16*, 109−114.
(44) Kumar, S., and Nussinov, R. (2002) Close-range electrostatic interactions in proteins. *ChemBioChem 3*, 604−617.
(45) Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol. 157*, 105−132.
(46) Burley, S. K., and Petsko, G. A. (1985) Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science 229*, 23−28.
(47) Overington, J., Johnson, M. S., Sali, A., and Blundell, T. L. (1990) Tertiary structural constraints on protein evolutionary diversity: templates, key residues and structure prediction. *Proc. R. Soc. London, Ser. B 241*, 132−145.
(48) Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res. 28*, 235−242.
(49) Smith, T. F., and Waterman, M. S. (1981) Identification of common molecular subsequences. *J. Mol. Biol. 147*, 195−197.
(50) Liao, G. J., Lee, Y. J., Lee, Y. H., Chen, L. L., and Chu, W. S. (1998) Structure and expression of the D-amino-acid oxidase gene from the yeast Rhodosporidium toruloides. *Biotechnol. Appl. Biochem. 27*, 55−61.
(51) Rodriguez, E., Han, Y., and Lei, X. G. (1999) Cloning, sequencing, and expression of an Escherichia coli acid phosphatase/phytase gene (appA2) isolated from pig colon. *Biochem. Biophys. Res. Commun. 257*, 117−123.
(52) Lim, D., Golovan, S., Forsberg, C. W., and Jia, Z. (2000) Crystal structures of Escherichia coli phytase and its complex with phytate. *Nat. Struct. Biol. 7*, 108−113.
(53) Ohta, Y., Watanabe, K., and Kimura, A. (1985) Complete nucleotide sequence of the E. coli N-acetylneuraminate lyase. *Nucleic Acids Res. 13*, 8843−8852.
(54) Izard, T., Lawrence, M. C., Malby, R. L., Lilley, G. G., and Colman, P. M. (1994) The three-dimensional structure of N-acetylneuraminate lyase from Escherichia coli. *Structure 2*, 361−369.