

An Automated Flow for Directed Evolution Based on Detection of Promiscuous Scaffolds Using Spatial and Electrostatic Properties of Catalytic Residues

Sandeep Chakraborty*

Department of Biological Sciences, Tata Institute of Fundamental Research, Mumbai, India

Abstract

The aspiration to mimic and accelerate natural evolution has fueled interest in directed evolution experiments, which endow or enhance functionality in enzymes. Barring a few de novo approaches, most methods take a template protein having the desired activity, known active site residues and structure, and proceed to select a target protein which has a pre-existing scaffold congruent to the template motif. Previously, we have established a computational method (CLASP) based on spatial and electrostatic properties to detect active sites, and a method to quantify promiscuity in proteins. We exploit the prospect of promiscuous active sites to serve as the starting point for directed evolution and present a method to select a target protein which possesses a significant partial match with the template scaffold (DECAAF). A library of partial motifs, constructed from the active site residues of the template protein, is used to rank a set of target proteins based on maximal significant matches with the partial motifs, and cull out the best candidate from the reduced set as the target protein. Considering the scenario where this 'incubator' protein lacks activity, we identify mutations in the target protein that will mirror the template motif by superimposing the target and template protein based on the partial match. Using this superimposition technique, we analyzed the less than expected gain of activity achieved by an attempt to induce β -lactamase activity in a penicillin binding protein (PBP) (PBP-A from *T. elongatus*), and attributed this to steric hindrance from neighboring residues. We also propose mutations in PBP-5 from *E. coli*, which does not have similar steric constraints. The flow details have been worked out in an example which aims to select a substitute protein for human neutrophil elastase, preferably related to grapevines, in a chimeric anti-microbial enzyme which bolsters the innate immune defense system of grapevines.

Citation: Chakraborty S (2012) An Automated Flow for Directed Evolution Based on Detection of Promiscuous Scaffolds Using Spatial and Electrostatic Properties of Catalytic Residues. PLoS ONE 7(7): e40408. doi:10.1371/journal.pone.0040408

Editor: Vladimir N. Uversky, University of South Florida College of Medicine, United States of America

Received: May 8, 2012; **Accepted:** June 6, 2012; **Published:** July 11, 2012

Copyright: © 2012 Sandeep Chakraborty. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by the Tata Institute of Fundamental Research (Department of Atomic Energy). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The author has declared that no competing interests exist.

* E-mail: sandeepchak@tifr.res.in

Introduction

Directed evolution experiments aspire to mimic and accelerate natural evolution. The methodology typically consists of random mutations [1], in vitro recombination [2–6], continuous evolution [7,8], intramolecular relocation of the N and C termini of a protein [9,10] and high-throughput screening [11,12]. While a few exceptional de novo approaches that endow function in proteins from scratch have been successful [13–17], most techniques target residues in the catalytic site or in its close vicinity for mutational purposes [18–20]. Directed evolution techniques have also sought out the remnants of secondary activities [21–23], with the rationale that the presence of a pre-existing catalytic scaffold increases the odds of success in endowing the activity to the protein [24–28].

Previous work by our group has established a computational method (CLASP) for the detection of active sites [29]. CLASP analysis of the catalytic residues in Class A β -lactamases identified the dichotomy in the proton abstraction mechanism [30]. Based on this idea, we proposed a method that enumerates the possible pathways for proton abstraction [31]. We also outlined a computational methodology to quantify promiscuous activities in

a wide range of proteins [32]. Jensen hypothesized that ancient enzymes were sparse and promiscuous [33], and this promiscuity formed the basis of the evolution of complex organisms through gene duplication and specialization [34,35]. This hypothesis germinated the idea of selecting promiscuous active sites to serve as a pre-existing scaffold for directed evolution.

In the current work, we present a computational method that selects a protein which has a significant match with a desired catalytic scaffold - Directed evolution using CLASP: an automated flow (DECAAF).

We create a library of partial motifs from the template scaffold and rank a set of proteins based on maximal significant matches with this library. Manual screening, considering further substantiating evidence, is used at this stage to cull the best candidate from the reduced set of proteins with significant congruence. For example, a particular protein would be more suitable if related proteins in the same superfamily are known to have the function under consideration.

The second part of the flow addresses the problem of predicting mutations that mirrors catalytic properties (efficiency, substrate specificity, etc.) of the template protein. In order to replicate the

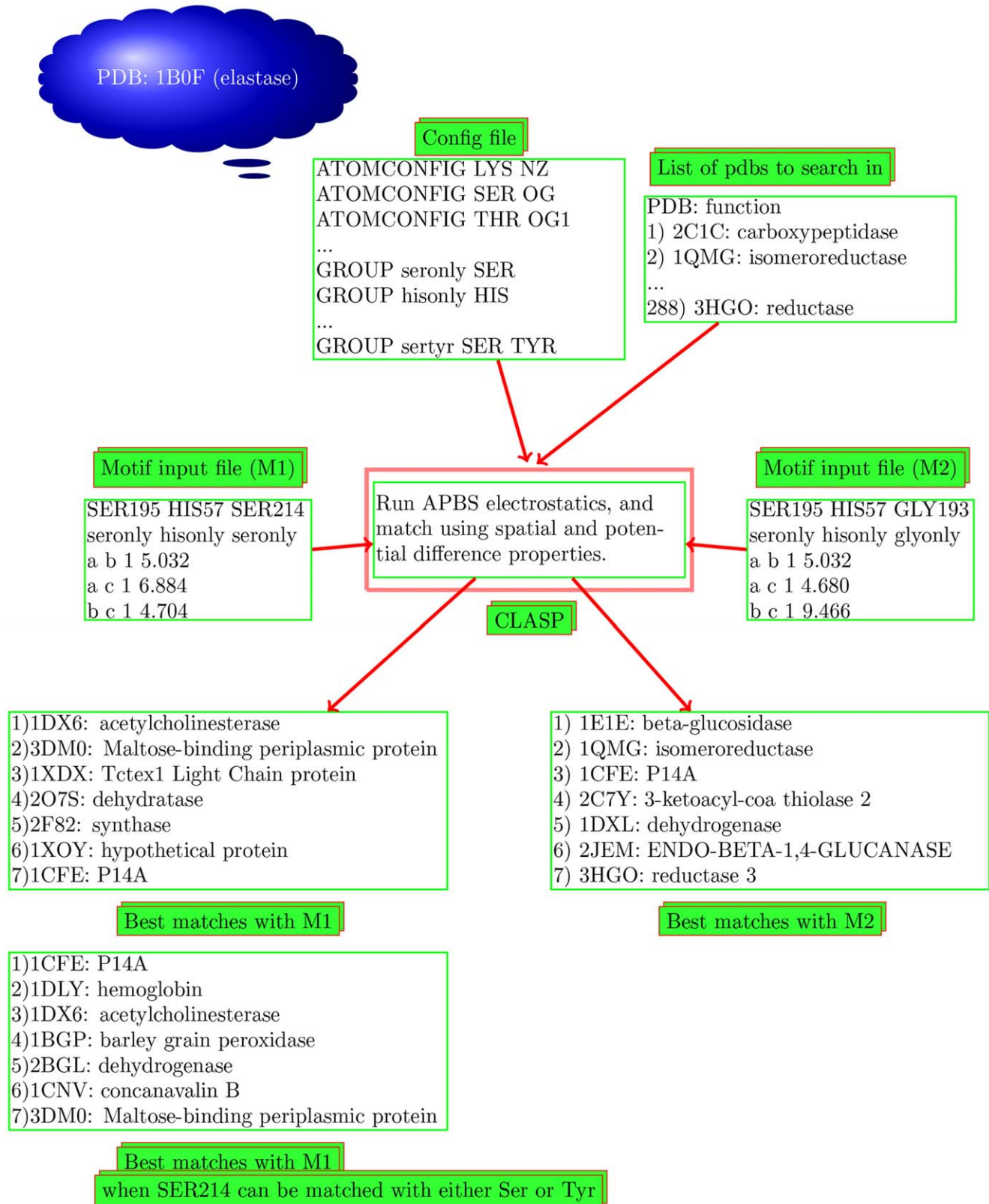


Figure 1. DECAAF flow to select a plant protein with a significant presence of an elastase-like scaffold. A pathogenesis related protein (PDB id: 1CFE) was selected as the best possible candidate. The ability to specify a set of residues to match a particular position allows this protein (PDB id: 1CFE) to score much better when a Tyr (Tyr36) is matched instead of a Ser (Ser120).
doi:10.1371/journal.pone.0040408.g001

template scaffold in the target protein, we first superimpose the template and target proteins based on the partial match. The residues in the target protein that are spatially close to the residues in the template motif, if they exist and are not part of the partial motif, are enumerated as possible candidates for site directed mutagenesis techniques to build on [18–20]. In contrast to the structural homology used by DECAAF, a recent method uses phylogenetic and sequence analyses to predict such amino acid substitutions [36].

The evolution of serine β -lactamases (Blase) (divided into Classes A, C and D) and the related penicillin binding proteins (PBP) has fascinated and confounded researchers for decades [37–39]. The physiological relevance of these proteins remains enigmatic [40,41]. Surprisingly, few directed evolution experiments have reported success in enhancing deacylation in PBPs [42,43], the catalytic step that Blases use to hydrolyze β -lactams [44]. Even when successful, the activity in these PBP mutants remains much lower than the Blases (110-fold in [42] and 90-fold in [43]). Sequence alignment identified L158E as a potential mutation that would modify the active site in PBP-A from *T. elongatus* to replicate the active site of Class A Blases, but reported only a 90-fold enhancement in the activity [43]. Firstly, CLASP analysis revealed that the L158E mutant indeed gains potential congruence with the Class A Blase as compared to the wildtype PBP. Subsequently, we demonstrated that Pro159 and Asp160 hinders the approach of substrate, using our superimposition technique. This probably explains the ‘puzzling’ lack of significant improvement in spite of the fact that ‘in the L158E mutant, all the catalytic residues and the interactions characteristic of β -lactamases seem to be present’ [43]. Finally, we chose another PBP-5 from *E. Coli* [45], which is also known to have a similar loop present in Class A Blases and PBP-A from *T. elongatus* [43,46], and propose that a L153E mutant of this protein can gain significant β -lactamase activity, as the side chain resembles Class A Blase to a greater extent.

There have been several methods developed for identifying enzymes with a partial catalytic structure [47–49]. The use of ‘exact’ electrostatic properties in the search process at an early

stage is a key innovation of DECAAF when compared to these methods. Other related methods have used binding energy and energy minimization considerations at a later stage of the search [24,25,27]. Although, the CLASP signature for any function comprises of a few residues, it implicitly encodes the surroundings. Potential congruence implies a favorable milieu for the particular function (hydrophilicity, polarity, presence in a cleft, etc.). Consequently, DECAAF filters out unfeasible configurations at a much lower computational cost.

We have worked out the detailed steps in DECAAF for identifying an elastase-like protein in plants. Our motivation is to find an alternative for human neutrophil elastase (HNE) [50] in a chimeric enzyme that provides enhanced resistance to the Pierce’s disease causing Gram-negative pathogen *Xylella fastidiosa* [51]. The substitute should be preferably from an organism related to grapevines. P14A [52] from *Solanum lycopersicum* (tomato), a member of the PR-1 group of pathogenesis-related proteins [53], is obtained as a suitable candidate. The structural homology shared by P14A and a snake venom protein, which was previously demonstrated to be an elastase [54], suggests with increasing certitude that P14A might have elastase activity. Furthermore, protease function has also been associated with the pathogenesis related proteins [55]. We have also identified residues in P14A that do not match the corresponding residue in HNE, and thus predicts mutations in the P14A which should replicate the HNE scaffold. In case we do not find detectable activity in the wild type P14A, we believe that the identified P14A mutants might gain elastase function. Finally, we propose that a L153E mutant of PBP-5 from *E. coli* (PDBid:1NZO) is likely to show better results than the L158E mutant of PBP-A from *T. elongatus* (PDBid:2J9O), which achieved a 90-fold increase in β -lactamase activity. Such experimental validations will be the ultimate litmus test for the DECAAF flow.

Results

We demonstrate our method (DECAAF) by working out the steps in selecting a elastase-like protein from a set of plant proteins. Human neutrophil elastase (HNE) is a serine protease that is present in the primary granules of polymorphonuclear neutrophils, the other two being proteinase 3 and cathepsin G [56]. The bactericidal properties of HNE have been exploited to design an anti-microbial protein that targets the outer-membrane of the bacteria [57]. HNE is combined with cecropinB, which lyses gram-negative bacteria, in this chimeric protein using a flexible linker and provides enhanced resistance to grapevines against the Gram-negative pathogen *Xylella fastidiosa*. A protein from a plant related to grapevines that has elastase activity, if found, can substitute HNE in the anti-microbial enzyme.

1 Selecting the Best Candidate

We chose the HNE protein with PDBid:1B0F as the template protein. The active site residues comprised of the following residues - Ser195, His57, Asp102, Ser214 and Gly193. The keyword search for ‘plants’ in <http://www.pdb.org> was pruned for redundancy based on a 40% sequence similarity, and yielded 288 proteins (Table S1). From the active site residues ($N = 5$), we generated all possible motifs of size $k < N$ (3, 4 and 5 in this case). An example motif is shown in Fig. 1. By default each position in the match can be occupied by the same amino acid, i.e. we do not consider stereochemical equivalence at this stage.

Let us consider the case when the partial motifs have three residues - there are 10 such motifs ($\binom{5}{3} = 10$). M1 = (Ser195, His57, Ser214), M2 = (Ser195, His57, Gly193) ... M10 = (Ser195, His57,

Table 1. DECAAF SCORES (DScore) - best matches for motifs of length 3: The best matching protein (PDBid:1DX6) is from *Torpedo californica* (Pacific electric ray), an organism not related to plants.

PDB	Description	Dscore
1DX6	acetylcholinesterase	0.048
1CFE	pathogenesis-related protein P14A	0.06
1QMG	acetohydroxy-acid isomeroeductase	0.068
2EFJ	3,7-dimethylxanthine methyltransferase	0.069
1W1O	cytokinin dehydrogenase 1	0.076
3DM0	maltose-binding periplasmic protein	0.077
1E1E	beta-glucosidase	0.077
1AIR	pectate lyase c	0.084
2V6G	progesterone 5-beta-reductase	0.084
3HGO	12-oxophytodienoate reductase 3	0.087

The acetylcholinesterase protein figures in the set of proteins being queried since it has a plant alkaloid galanthamine bound in active site, and we used a keyword search for ‘plant’ in <http://www.pdb.org> to obtain the query set of 288 proteins. Hence, the best match was with the P14A protein from *Solanum lycopersicum* (tomato).

doi:10.1371/journal.pone.0040408.t001

Table 2. Expanding the set of residues that can match a single position in the input motif.

PDB:	Predicted residues			Pairwise Distances in Å			Pairwise Potential difference			Score
	a	b	c	ab	ac	bc	ab	ac	bc	
1B0F	Ser195	His57	Ser214	5	6.8	4.7	53.7	105.7	51.9	0
1CFE (Ser214 can be Ser only)	Ser49	His48	Ser120	0.1	-0.2	1.9	15.3	83.9	68.6	0.111
1CFE (Ser214 can be Ser or Tyr)	Ser49	His48	Tyr36	0.1	0.3	-0.3	15.3	88.6	73.3	0.023

When the position Ser214 in the input motif (Ser195, His57, Ser214) can be matched by a Ser or Tyr, Tyr36 has a much better spatial orientation with respect to Ser49 and His48 as compared to Ser120 in the P14A protein (PDBid:1CFE). The distances are specified in the reference protein (PDBid:1B0F) in Å. For the remaining, we show the deviation from the reference value. Potential differences are in units of kT/e (k is Boltzmann's constant, T is the temperature in K and e is the charge of an electron). They are absolute values and not shown as deviations.

doi:10.1371/journal.pone.0040408.t002

Asp102). We ran each motif (M1, M2, ... M10) on each protein in the target list, resulting in a score that quantifies the degree to which a motif exists in a protein based on spatial and electrostatic congruence [29]. Each run of a single motif on the list of 288 proteins took less than 30 mins on a simple workstation (2 GB Ram). Fig. 1 shows the best matches with M1 and M2.

We then compute a cumulative score from the runs for different motifs for each protein that reflects the likelihood of endowing elastase to that protein (Table 1). Manual inspection now chose the best protein from the reduced set of the proteins. The small number of proteins to be manually evaluated made this feasible. The best matching protein (PDBid:1DX6) is from *Torpedo californica* (Pacific electric ray), an organism not related to plants. The acetylcholinesterase protein is included in the set of proteins being queried since it has a plant alkaloid galanthamine bound in active site, and we used a keyword search for 'plant' in <http://www.pdb.org/> to obtain the query set of 288 proteins. The P14A (PDB id: 1CFE), the next best ranked protein, from *Solanum lycopersicum* (tomato) was therefore the desired target protein [52]. Table 2 shows the congruence of the partial motif (Ser195, His57, Ser214) from the HNE protein with the match from P14A.

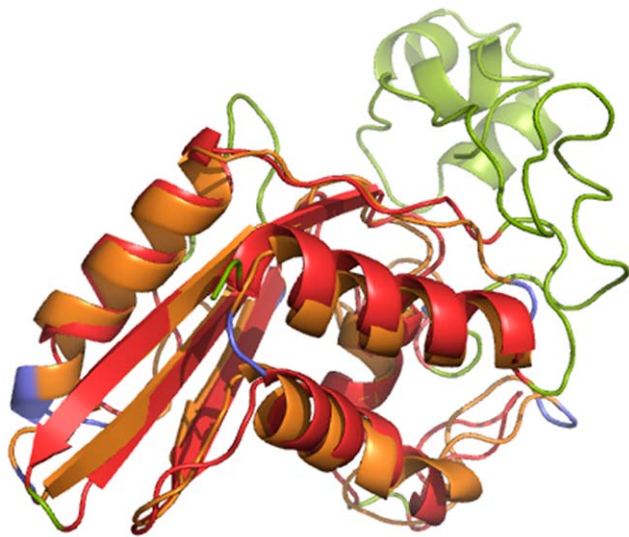


Figure 2. Superimposition of pseudochetoxin (PDB id: 1CFE, in green) and P14a (PDB id: 1B0F, in blue) and using TopMatch. The overlapping regions are colored red and orange for pseudochetoxin and P14a respectively. It is known that pseudochetoxin has elastase activity.

doi:10.1371/journal.pone.0040408.g002

P14A is a member of the PR-1 group of pathogenesis-related proteins [53]. While P14A has not been specifically known to have elastase activity, it is structurally homologous to a snake venom protein (Fig. 2) which was previously demonstrated to be an elastase [54]. Pathogenesis related proteins are known to possess proteolytic function [55]. Furthermore, it has been noted that "His48, Ser49 and His93 are in close proximity, so that the conservation of this group of residues could be taken as an indication of it being important as an active site in P14a" [52]. His48 and Ser49 are the residues predicted by DECAAF to be a part of the catalytic scaffold (Table 2 and Fig. 3).

A similar analysis using partial motifs of size $k = 4$ and $k = 5$ was done. We have not been able to design a reasonable algorithm for comparing scores with different motif lengths. The P14A protein does not have a good match when the input scaffold includes Asp102. Hence, any motif figuring the Asp102 was not found in P14A with significant congruence. We discuss this aspect subsequently.

2 Improving Matches through Stereochemical Equivalence

We extended the matches by allowing stereochemically equivalent residues to match in a motif. While stereochemical equivalence can be hardcoded for amino acids with similar properties, there are instances where residues with different properties occupy the same sequence and spatial location, and perform the same functionality. A well-known example is the equivalence of Ser130 and Tyr150 in Class A and C - lactamases respectively [58]. In the motif (M1 = Ser195, His57, Ser214), we allowed the position Ser214 to be matched by either a Ser or a Tyr. Table 2 shows that Tyr36 has a much better spatial orientation than the Ser120 in the P14A protein, although the potential differences are almost similar. Similarly extending the motif length ($n = 4$) allowed us to match Gly193 in the HNE protein to Ala51 in the P14A protein (Table 3 and Fig. 3). Gly and Ala both have non-polar, aliphatic R groups.

3 Superimposing the Partial Scaffolds

As mentioned previously, the P14A protein does not have a good match when the input scaffold includes Asp102. Although, unconventional serine proteases with variation in the catalytic triad have been observed in many cases, it is possible that P14A is not one them [59]. In such a scenario, we would have to resort to directed evolution techniques to impart elastase function to the protein. Table 4 shows the coordinates of the atoms before and after the transformations when the partial match - (Ser195/OG, His57/ND1, Ser214/OG, Gly193/N - PDBid:1B0F) and (Ser49/OG, His48/ND1, Tyr36/OH, Ala51/N - PDBid:1CFE) - are

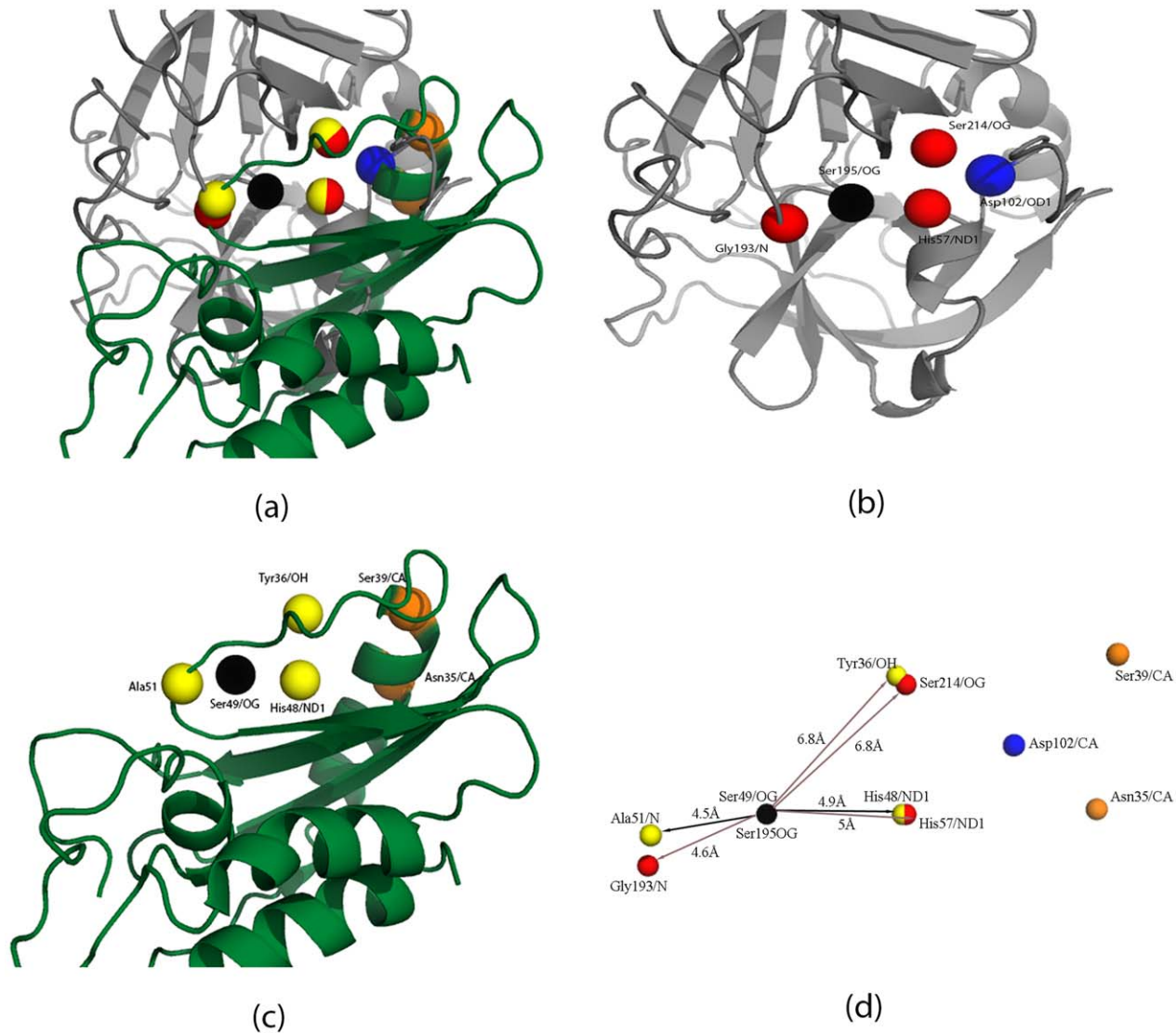


Figure 3. Superimposing proteins based on partial matches (Ser195, His57, Ser214 from PDBid:1B0F) and (Ser49, His48, Tyr36 from PDBid:1CFE). Ser195 from PDBid:1B0F and Ser49 from PDBid:1CFE coincide and is colored in black. (a) Superimposition of both proteins: Human neutrophil elastase (PDBid:1B0F) is in grey, and P14A (PDBid:1CFE) is in green. (b) Human neutrophil elastase (PDBid:1B0F) with active site atoms - Ser195/OG, His57/ND1, Ser214/OG, Gly193/N. (c) P14A protein (PDBid:1CFE) with predicted active site residues - Ser49/OG, His48/ND1, Tyr36/OH, Ala51/N. (d) Distance between pairs of residues in the partial matches. It can be seen that the Asp102 in PDBid:1B0F is close to Asn35 and Ser39 in PDBid:1CFE.

doi:10.1371/journal.pone.0040408.g003

superimposed. Note, that the superimposition algorithm uses the first three residues in the match. Fig. 3 shows the superimposed proteins. It can be seen that the pair (Gly193/N, Ala51/N), although not used for superimposition, are spatially close. The spatial position of the Asp102 in the HNE is in the vicinity of Asn35 and Ser39 in the P14A protein. Hence, it is possible that the Asn35Asp or Ser39Asp mutant might gain elastase function.

4 Analyzing Results of Directed Evolution of Penicillin Binding Proteins to Enhance β -lactamase Activity

The technique for superimposing the active sites of proteins based on the partial matches has been applied to analyze the lower than expected gain in β -lactamase (Blase) activity of a penicillin binding protein (PBP) using site directed mutations [43]. Class A Blases are differentiated from other classes of Blases (C and D) by

virtue of the presence of a loop, which houses the critical Glu166 [46]. PBP-A from *T. elongatus* possesses a similar loop, and sequence alignment suggested that Leu158 occupies the position corresponding to Glu166 (of the Blase) in this protein. However, the mutant L158E showed a meager 50-fold gain in activity as compared to the wildtype. This gain in activity is corroborated by CLASP analysis of the mutant active site, which gains potential congruence in cognate pairs as compared to the Class A Blase catalytic residues (Table 5). This electrostatic congruence was missing in the wildtype (Table 5).

The superimposition of the two proteins, Class A Blase (PDBid:1E25) and L158E PBP mutant (PDBid:2J90) based on their partial matches - (Ser70/OG, Lys73/NZ, Ser130/OG) and (Ser61/OG, Lys64/NZ, Ser122/OG) respectively, suggested that the residues adjacent to Glu158 in the mutant protein (Pro159 and Asp160) are responsible for hindering substrate access (Fig. 4).

Table 3. Extending the partial match.

	a	b	c	d			
1B0F Motif	SER195	HIS57	SER214	GLY193			
1CFE motif	SER49	HIS48	TYR36	ALA51			
	ab	ac	ad	bc	bd	cd	
PDB	Pairwise Distances						
1B0F	5	6.8	4.6	4.7	9.4	11	
1CFE	4.9	6.8	4.5	5	9.1	10	
	Pairwise Potential difference						
1B0F	53.7	105.7	-111.1	51.9	-164.9	-216.8	
1CFE	15.3	88.6	-58.1	73.3	-73.4	-146.8	

Spatial and electrostatic potential difference (PD) congruence in cognate pairs in the human neutrophil elastase (PDBid:1B0F) and the pathogenesis related P14A protein (PDBid:1CFE) for an extended set of four residues in the active site.

doi:10.1371/journal.pone.0040408.t003

Pro159 and Asp160 are spatially close to Leu158/Glu158 in the wildtype/mutant, whereas in the Class A Blase there are no residues within 3 Å of Glu166 (Table 6). Several combinations of mutations failed to improve the activity any more than 90-fold. Pro159 was conserved in all mutants, and hence possibly continued to deny access to substrate in the mutations.

Subsequently, we chose PBP-5 from *E. Coli*, which is also known to have a similar loop like the Class A Blases and PBP-A [45]. It has been noted that a His151 occupies the same spatial position as Glu166, and is possible a good target for mutation – “spatial equivalence of carbonyl group of His-151 in PBP 5 with the carboxyl group of Glu-166 in TEM-1” [60]. However, Leu153 showed up as the best candidate for mimicking Glu166 when we

Table 4. Superimposing proteins based on partial matches (Ser195, His57, Ser214 from PDBid:1B0F) and (Ser49, His48, Tyr36 from PDBid:1CFE).

	Atom	Before			After		
		Coordinates	X	Y	Z	X	Y
PDB: 1B0F (template)	Ser195/OG	64.4	57	53.8	0	0	0
	His57/ND1	63.3	54.8	58.2	5	0	0
	Ser214/OG	63.6	50.6	56.1	5	4.7	0
PDB: 1CFE (target)	Ser49/OG	8.8	-6.3	-4.8	0	0	0
	His48/ND1	9.2	-2.4	-1.8	4.9	0	0
	Tyr36/OH	13.7	-1.6	-3.7	4.7	5	0

After applying linear and rotational transformations on both proteins, Ser195 and Ser49 lie at the center of the coordinate axis, His57 and His48 lie on the -X-Y axis (Y=0) and Ser214 and Tyr36 lie on the X-Y plane (Z=0).
doi:10.1371/journal.pone.0040408.t004

superimposed the Class A Blase and the PBP-5 (PDBid:1NZO) based on partial matches. Furthermore, the adjacent residues to Leu153 do not impede the substrate (Table 6 and Fig. 5). We hypothesize that a mutational study similar to the L158E of PBP-A, if applied to PBP-5 to generate L153E, might provide greater success in replicating Blase enzymatic efficiency in PBPs.

Discussion

Mutations resulting from the innate inaccuracies of biological processes and environmental factors accrue over millions of years and are a key determinant of evolution. Rapid technological advances in the last few decades have allowed enzyme engineers to apply similar mutation generating methods, and shrunk the timescales for evolving new enzymes to a few months [1–5]. Directed evolution is the generic term applied to experiments that

Table 5. Gain of potential congruence in L158E mutant of PBP-A.

	a	b	c	d	e						
1E25	Ser70/OG	Lys73/NZ	Ser130/OG	Lys234/NZ	Glu166/OE1						
2J7V	Ser61/OG	Lys64/NZ	Ser122/OG	Lys219/NZ	Leu158/CD2						
2J90	Ser61/OG	Lys64/NZ	Ser122/OG	Lys219/NZ	Glu158/OE1						
1NZO	Ser44/OG	Lys47/NZ	Ser110/OG	Lys213/NZ	Leu153/CD2						
	ab	ac	ad	ae	bc	bd	be	cd	ce	de	
1E25	PD	-125.6	22.4	-189.1	310.7	148.1	-63.5	436.3	-211.5	288.2	499.7
1E25	D	2.8	3.2	4.7	5.5	3.6	5.6	5	2.9	8	10
2J7V	PD	-134.5	38.7	-135.6	-112.4	173.3	-1	22.1	-174.3	-151.1	23.2
2J7V	D	3	3	4.4	7.1	3.1	5.4	5.8	2.9	8.7	10.1
2J90	PD	-162.3	-16.4	-198.3	247.7	146	-36	410	-181.9	264	445.9
2J90	D	2.7	3	4.5	4.5	3	5.3	4	2.8	6.7	8.6
1NZO	PD	-241.6	-68.8	-277.9	-188.2	172.8	-36.2	53.5	-209.1	-119.4	89.7
1NZO	D	3.1	4.2	6.3	6.5	5.1	6.7	5.5	2.7	10	12.1

Pairwise potential difference (PD) and distance (D) between cognate pairs in a Class A Blase (PDBid:1E25), a PBP-5 protein (PDBid:1NZO), the wildtype PBP-A (PDBid:2J7V) and the L158E PBP-A mutant (PDBid:2J90). The L158E PBP-A mutant has gains potential congruence with the Class A Blase, as compared to the wildtype PBP-A. Distances are in Å. See Methods section for units of potential.

doi:10.1371/journal.pone.0040408.t005

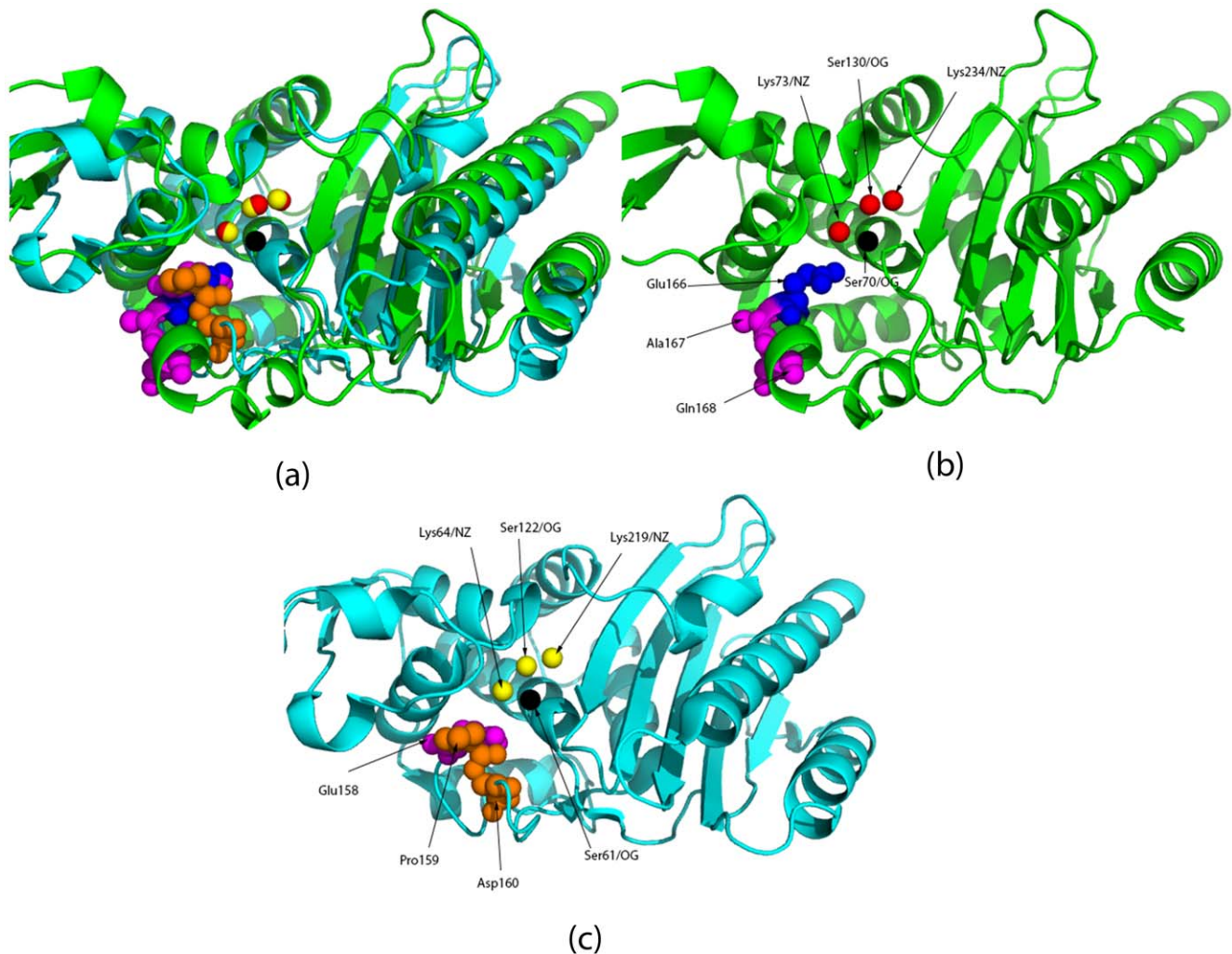


Figure 4. Superimposition of Class A Blase (PBDId:1E25 in green) and L158D PBP-A mutant (PBDId:2J90 in light blue) based on their partial matches - (Ser70/OG, Lys73/NZ, Ser130/OG) and (Ser61/OG, Lys64/NZ, Ser122/OG) respectively. (a) It is seen that their alignments aimed at superimposing the partial matches results in a good superimposition of the complete proteins. **(b)** Class A Blase (PBDId:1E25) - it can be seen that the Glu166/OE1 has no steric hindrance from the adjacent residues (Ala167 AND Gln168). **(c)** Mutant L158D PBP-A (PBDId:2J90) - neighboring residues, Pro159 and Asp160, can be seen obstructing access to substrate.
doi:10.1371/journal.pone.0040408.g004

mimic and accelerate natural evolution [13–20,61]. Several groups have reviewed the current developments in directed evolution in detail [62–65].

Promiscuity, the catalysis of reactions distinct from the one the protein has evolved to perform, is another characteristic feature of enzymes that has shaped the evolution of primitive cells [33,66,67]. Gene duplication unshackled such ‘leaky’/‘messy’ domains from selection pressures, and subsequent specialization culminated into new enzymes [34,35,68]. Identifying pre-existing scaffolds in proteins is a critical step as it implies better odds of success when attempting to endow catalytic activity to proteins [24–28].

The current work has tried to exploit this prospect of a promiscuous active site to serve as a pre-existing scaffold for directed evolution. Previously, we have defined a quantitative measure of promiscuity in a wide range of proteins [32]. The spatial and electrostatic congruence of the active site residues in a heme cytochrome C peroxidase and a Zn²⁺ carboxypeptidase was vindicated by the fact that carboxypeptidase is known to gain pyruvate oxidase catalysis abilities on the replacement of Zn²⁺ with Cu²⁺ [69]. This germinated the idea of defining a generic

methodology that would seek an ‘incubator’ protein for any enzymatic activity with known 3D structure and active site residues.

We present a computational method that selects a protein which has a significant match with a desired catalytic scaffold - **Directed evolution using CLASP: an automated flow (DECAAF)**. Barring the rare chance of finding a protein matching the complete motif, the problem at hand is to find a protein with the best partial matches. This has been solved using heuristics that generates a library of partial motifs and results in a cumulative score reflecting the likelihood of endowing the desired function to that protein. The intractability of the problem is evident, and the many methods are inclined to remain in a local minima [70]. Hence the final step might involve manual screening of a few best matching proteins. Expert eyes can also figure out other supporting facts - for example, it might be known that a related protein from the same superfamily possesses the function, instilling further confidence before one embarks on laborious mutational work.

The possibility that the selected protein does not possess the activity, despite the presence of a subset of the catalytic motif in its

Table 6. Atoms close to the Glu166/OE1 in Class A β -lactamase after superimposition of the partial motifs.

1E25			2J7V			2J9O			1NZO		
D	Residue	Atom	D	Residue	Atom	D	Residue	Atom	D	Residue	Atom
0	Glu166	OE1	1.6	Pro159	C	1.6	Pro159	C	1.4	LEU153	CD2
1.3	Glu166	CD	1.6	Pro159	CA	1.6	Pro159	CA	2.7	LEU153	CG
2.2	Glu166	OE2	1.8	Pro159	O	1.9	Glu158	OE1	2.8	HOH475	O
2.4	Glu166	CG	2.4	Leu158	CD1	1.9	Pro159	O			
2.7	Glu166	CB	2.7	Asp160	N	2.6	Asp160	N			
3	Glu166	CA	2.7	Pro159	N	2.7	Glu158	CA			
3.1	HOH2076	O	2.7	Leu158	CA	2.7	Pro159	N			
			2.8	Pro159	CB	2.8	Glu158	CB			
			2.9	Leu158	CB	2.8	Glu158	CD			
			3.1	Leu158	CG	2.9	Pro159	CB			
			3.2	Leu158	C	3.1	Glu158	C			
			3.4	HOH2146	O	3.3	HOH2234	O			

PDBid:1E25 is the Class A Blase, PDBid:2J7V is the wildtype PBP-A, PDBid:2J9O is the L158E PBP-A mutant and PDBid:1NZO is the PBP-5 protein. It can be seen that in the Class A Blase (PDBid:1E25) and PBP-5 there are no atoms within a radius of about 3 Å of Glu166/OE1. In the mutant L158E PBP, both Pro159 and Asp160 are spatially close to Glu158, and can be seen to be obstructing access to substrate (Fig. 4). Distances are in Å, and have a cut-off till the first water molecule is encountered. doi:10.1371/journal.pone.0040408.t006

structure, leads to the next logical question - can one predict mutations that would bestow function? We have addressed this requirement by superimposing the template and target proteins based on the partial matches, and identified target residues that are in the vicinity of 'unmatched' residues of the template protein. Such residues can be the target of site-directed methods [18–20].

This superimposition technique has been used to analyze the results of an experiment intended to transform a penicillin binding protein (PBP) into a serine β -lactamase (Blase). PBPs are involved in the synthesis and remodeling of bacterial cell wall [71]. Owing to the lack of a counterpart in mammalian organisms, PBPs are the target of the β lactam family of antibiotics (penicillin, ampicillin, etc.) which are structurally homologous to the D-alanyl-D-alanine moiety present on peptidoglycan precursors. These drugs acylate the nucleophilic serine residue and, unlike the peptide substrates, form long-lived complexes. Blases, which share three highly conserved motifs with PBPs (the SXXX tetrad, the SXN triad, and KTG triad) are similarly acetylated, but have evolved to gain significantly higher rates of deacylation resulting in the regeneration of the original uncomplexed enzyme [37–39]. This sets up the scene for a perfect directed evolution experiment - what are the mutations needed to transform a PBP into a Blase? However, this task has proved to be surprisingly difficult [42,43].

The serine Blases are divided into three classes (A, C and D) based on sequence homology [72]. Of these, the Class A Blases have evolved to use a Glu-X-X-Asn motif present on a loop (the loop) for catalysis [46]. PBP-A from *T. elongatus*, a thermophilic unicellular rod shaped cyanobacterium, possesses a similar loop, and sequence alignment suggested that Leu158 occupies the position corresponding to Glu166 in this protein [73]. However, the L158E PBP-A mutant could only achieve a meager 90-fold gain of activity, even when aided with other adjacent mutations [43]. This was 'puzzling' since 'in the L158E mutant, all the catalytic residues and the interactions characteristic of β -lactamases seem to be present' [43].

We have corroborated that the electrostatic properties of the active site residues in the L158E mutant now mirrors the Class A Blase catalytic site properties, explaining the increase in activity.

Moreover, superimposition of the reactive atoms of the catalytic residues in both the protein provides a possible explanation of the modest gains - the residues adjacent to Glu158 in the mutant PBP-A (Pro159 and Asp160) is seen to be restricting access to the catalytic cleft, in stark contrast to the Class A Blase. Since Pro159 was conserved in all of the mutants, it might be a key determinant for the less than expected gain of activity.

Next, we chose PBP-5 from *E. coli* since it possesses a loop similar to Class A Blases and PBP-A [45]. PBP-5 helps in maintaining the normal morphology of the cell and provides resistance to β -lactams antibiotics [74–76]. While it has been commented that His151 is the residue that corresponds to the Glu166 in Class A Blase [60], we noted that Leu153 is a better candidate after superimposition of the proteins using the technique detailed in this work. Also, unlike in PBP-A, the adjacent residues to Leu153 in PBP-5 have a conformation similar to that in Class A Blases, and seem to provide unrestricted access of substrate to the catalytic site. We hypothesize that the L153E PBP-5 mutant might achieve better gain of function as compared to the L158E mutant(s).

The use of 'exact' electrostatic properties in the search process at an early stage is a key innovation of DECAAF when compared to similar methods seeking partial catalytic structure [47–49]. Pruning based on electrostatic properties reduces considerable false positives compared to purely 3D matching methods. A few methods have used binding energy and energy minimization considerations, albeit at a later stage of the search [24,25,27]. Electrostatic congruence of a few active site residues implies a favorable milieu for the catalytic activity and thus encodes residues in the close vicinity as well. This allows DECAAF to prune unfeasible configurations at a much lower computational cost.

We describe in the details the steps (Fig. 1) involved in choosing a plant protein to substitute human neutrophil elastase (HNE) [50] in a chimeric anti-microbial enzyme designed to bolster the innate immune defense system in grapevines [51]. The flow identifies P14A [52] from *Solanum lycopersicum* (tomato) as a very significant match (Tables 1, 2, 3). P14A is a member of the PR-1 group of pathogenesis-related proteins [53]. While P14A has not been

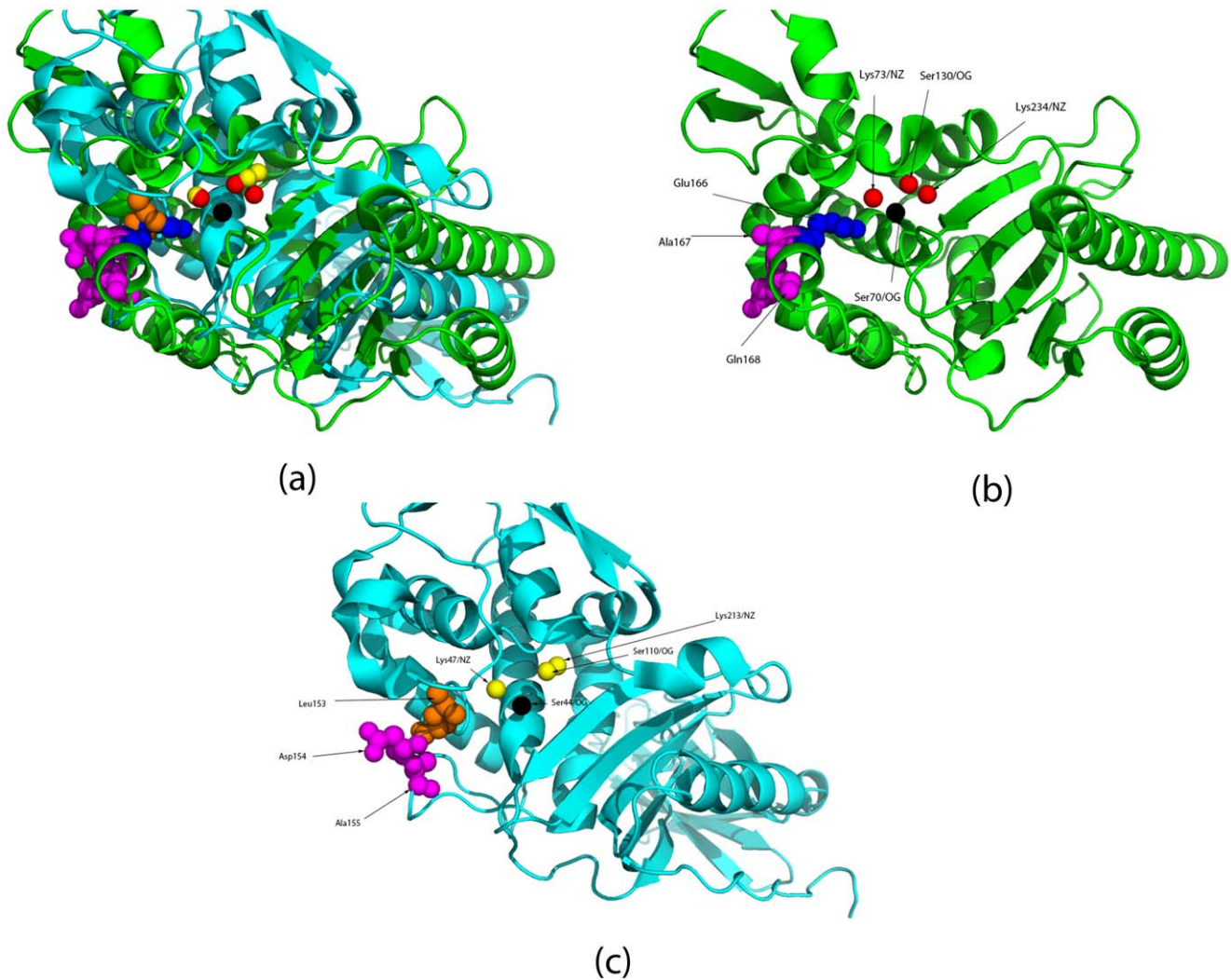


Figure 5. Superimposition of Class A Blase (PBDId:1E25 in green) and PBP-5 (PDBId:1NZO in light blue) based on their partial matches - (Ser70/OG, Lys73NZ, Ser130/OG) and (Ser44/OG, Lys47/NZ, Ser110/OG) respectively. (a) It is seen that the superimposition of PBP-5 and this Class A Blase is less significant than that with PBP-A. (b) Leu153 corresponds to the spatial location of Glu166, and it is seen that Asp154 and Ala155 do not restrict access to the catalytic cleft. doi:10.1371/journal.pone.0040408.g005

specifically known to have elastase activity, it is structurally homologous to a snake venom protein (Fig. 2), previously demonstrated to be an elastase [54]. Interestingly, the complete scaffold of the HNE does not exist in the P14A protein. While Ser195, His57, Ser214 and Gly193 from the input motif has an highly matching scaffold in P14A (Fig. 3), the spatial position of the Asp102 in the HNE is close to Asn35 and a Ser39 in the P14A protein when the proteins were superimposed. We intend to express this protein and test for its elastase activity. In case we do not find detectable activity, we believe that the Asn35Asp or the Ser39Asp P14A mutant might gain elastase function and thus validate the proposed rational design flow DECAAF.

Methods

1 Choosing a Consensus Group of Proteins

We start with a template protein with known 3D structure and a set of active site residues

$$(\Phi_{\text{residues}}, |\Phi_{\text{residues}}| = N) \text{ (Equation 1).}$$

Our goal is to select the best protein from a set of proteins ($\Phi_{\text{proteins}}, |\Phi_{\text{proteins}}| = M$) with known 3D structures (Equation 2), that has a domain in its structure which is significantly congruous to Φ_{residues}

$$\Phi_{\text{residues}} = \{R_1, R_2, \dots, R_N\} \quad (1)$$

$$\Phi_{\text{proteins}} = \{PDB_1, PDB_2, \dots, PDB_M\} \quad (2)$$

We first create partial motifs of size k from Φ_{residues} , for $k = >3, 4 \dots N$ (Equation 3). Let us denote the number of elements in ${}^n C_k$ by n .

$${}^n C_k \text{ motifs } \Phi = {}^n C_k \Phi_{\text{residues}} \text{ for } i = 3, 4 \dots N \quad (3)$$

CLASP analysis of each motif of size k results in n scores for each protein P_i (CLASPScore $_i$). We combine this score into a cumulative score Score (PDB $_k$) (Equation 4).

$$\text{Score}_i^k(\text{PDB}) = \sum \text{CLASPScore}_i \text{ for } i = 1 \dots n \text{ and} \quad (4)$$

$$k = 3, 4 \dots N$$

Now, for each k , we obtained the sorted set of scores for each protein (Equation 5).

$$\text{sorted}^k \Phi = \{ \text{PDB}_x, \text{PDB}_y, \text{PDB}_z, \dots, \text{PDB}_N \}$$

$$\text{such that, } \text{Score}(\text{PDB}_x) < \text{Score}(\text{PDB}_y), \quad (5)$$

$$\text{Score}(\text{PDB}_y) < \text{Score}(\text{PDB}_z) \dots$$

At this stage, we manually screen the proteins which occur in the top matches (**PDB_x**, **PDB_y**, **PDB_z** ...) in Φ^k_{sorted} for each k , and decide the best possible candidate.

2 Superimposing Proteins Based on Partial Matches and Predicting Mutations

We apply transformations on the template and target proteins such that the partial matches overlap. The superimposition algorithm takes the first three atoms of the partial match from both the template and target proteins - let us refer to them as (A,B,C) and (P,Q,R). We apply both linear and rotational transformations on both the proteins such that A and P are at the center of the coordinate axis (coordinates = [0,0,0]), B and Q lies on the X-Y axis (i.e. Y coordinate is 0) and C and R lie on the X-Y plane (i.e. Z coordinate is 0). In short, after the transformations, (A,B,C) and (P,Q,R) both lie on the same plane (Z = 0) such that A and P are at the center, and B and Q lie on the x-y axis. Once the two structures are superimposed, we proceed to find the residues in the target protein which are spatially close to the unmatched residues in the template protein. In case the wild type

References

- Cirino PC, Mayer KM, Umeno D (2003) Generating mutant libraries using error-prone PCR. *Methods Mol Biol* 231: 3–9.
- Hall BG, Zuzel T (1980) Evolution of a new enzymatic function by recombination within a gene. *Proc Natl Acad Sci USA* 77: 3529–3533.
- Stemmer WP (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature* 370: 389–391.
- Zhao H, Giver L, Shao Z, Aholter JA, Arnold FH (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotechnol* 16: 258–261.
- Kolkman JA, Stemmer WP (2001) Directed evolution of proteins by exon shuffling. *Nat Biotechnol* 19: 423–428.
- Engler C, Gruetznier R, Kandzia R, Marillonnet S (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes. *PLoS ONE* 4: e5553.
- Esvelt KM, Carlson JC, Liu DR (2011) A system for the continuous directed evolution of biomolecules. *Nature* 472: 499–503.
- Johns GC, Joyce GF (2005) The promise and peril of continuous in vitro evolution. *J Mol Evol* 61: 253–263.
- Graf R, Schachman HK (1996) Random circular permutation of genes and expressed polypeptide chains: application of the method to the catalytic chains of aspartate transcarbamoylase. *Proc Natl Acad Sci USA* 93: 11591–11596.
- Guntas G, Kanwar M, Ostermeier M (2012) Circular Permutation in the Ω -Loop of TEM-1 β -lactamase Results in Improved Activity and Altered Substrate Specificity. *PLoS ONE* 7: e35998.
- Goddard JP, Reymond JL (2004) Enzyme assays for high-throughput screening. *Curr Opin Biotechnol* 15: 314–322.
- Stapleton JA, Swartz JR (2010) Development of an in vitro compartmentalization screen for high-throughput directed evolution of [FeFe] hydrogenases. *PLoS ONE* 5: e15275.
- Bolon DN, Mayo SL (2001) Enzyme-like proteins by computational design. *Proc Natl Acad Sci USA* 98: 14274–14279.
- Jiang L, Altho EA, Clemente FR, Doyle L, Rothlisberger D, et al. (2008) De novo computational design of retro-aldol enzymes. *Science* 319: 1387–1391.
- Faiella M, Andreozzi C, de Rosales RT, Pavone V, Maglio O, et al. (2009) An artificial di-iron oxo-protein with phenol oxidase activity. *Nat Chem Biol* 5: 882–884.
- Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, et al. (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* 329: 309–313.
- Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, et al. (2008) Kemp elimination catalysis by computational enzyme design. *Nature* 453: 190–195.
- Reetz MT, Carballeira JD (2007) Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat Protoc* 2: 891–903.
- Climic S, Ruiz-Perez L, Gonzalez-Pacanowska D, Prapunwattana P, Cho SW, et al. (1990) Saturation site-directed mutagenesis of thymidylate synthase. *J Biol Chem* 265: 18776–18779.
- Reetz MT, Carballeira JD, Peyralans J, Hobenreich H, Maichele A, et al. (2006) Expanding the substrate scope of enzymes: combining mutations obtained by CASTing. *Chemistry* 12: 6031–6038.
- Kimura M (1986) DNA and the neutral theory. *Philos Trans R Soc Lond, B, Biol Sci* 312: 343–354.
- Amitai G, Gupta RD, Taw k DS (2007) Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP J* 1: 67–78.
- Wroe R, Chan HS, Bornberg-Bauer E (2007) A structural model of latent evolutionary potentials underlying neutral networks in proteins. *HFSP J* 1: 79–87.
- Zanghellini A, Jiang L, Wollacott AM, Cheng G, Meiler J, et al. (2006) New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci* 15: 2785–2794.
- Dahiyat BI, Mayo SL (1997) De novo protein design: fully automated sequence selection. *Science* 278: 82–87.
- Malisi C, Kohlbacher O, Hocker B (2009) Automated scaffold selection for enzyme design. *Proteins* 77: 74–83.
- Georgiev I, Lilien RH, Donald BR (2008) The minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search

protein does not have the desired functions, mutating these residues might bestow function.

3 Tools

Adaptive Poisson-Boltzmann Solver (APBS) and PDB2PQR packages were used to calculate the potential difference between the reactive atoms of the corresponding proteins [77,78]. The APBS parameters were set as follows - solute dielectric: 2, solvent dielectric: 78, solvent probe radius: 1.4 Å, temperature: 298 K and zero ionic strength. APBS writes out the electrostatic potential in dimensionless units of kT/e where k is Boltzmann's constant, T is the temperature in K and e is the charge of an electron. We extensively integrated and used the freely available BioPerl [79] modules and Emboss [80] tools. 3D structure was aligned by the TopMatch server [81]. All protein structures were rendered by PyMol (<http://www.pymol.org/>).

Supporting Information

Table S1 Target set of 288 proteins. The keyword search for 'plants' in <http://www.pdb.org/> was pruned for redundancy based on a 40% sequence similarity, and yielded 288 proteins. (PDF)

Acknowledgments

I gratefully acknowledge B. J. Rao for technical discussions and infrastructural support. I am also indebted to Abhaya M. Dandekar from Plant Sciences Department, University of California, Davis, CA for technical inputs on the chimeric anti-microbial enzyme designed by their group [51]. I would like to thank Anindya S. Ghosh from Department of Biotechnology, IIT Kharagpur for discussions on PBP-5 from *E. coli*.

Author Contributions

Conceived and designed the experiments: SC. Performed the experiments: SC. Analyzed the data: SC. Contributed reagents/materials/analysis tools: SC. Wrote the paper: SC.

- algorithm for computing partition functions over molecular ensembles. *J Comput Chem* 29: 1527–1542.
28. Lovell SC, Word JM, Richardson JS, Richardson DC (2000) The penultimate rotamer library. *Proteins* 40: 389–408.
 29. Chakraborty S, Minda R, Salaye L, Bhattacharjee SK, Rao BJ (2011) Active site detection by spatial conformity and electrostatic analysis-unravelling a proteolytic function in shrimp alkaline phosphatase. *PLoS ONE* 6: e28470.
 30. Dambion C, Raquet X, Lian LY, Lamotte-Brasseur J, Fonze E, et al. (1996) The catalytic mechanism of beta-lactamases: NMR titration of an active-site lysine residue of the TEM-1 enzyme. *Proc Natl Acad Sci USA* 93: 1747–1752.
 31. Chakraborty S (2012) Enumerating pathways of proton abstraction based on a spatial and electrostatic analysis of residues in the catalytic site. *PLoS ONE*, In Press.
 32. Chakraborty S, Rao BJ (2012) A measure of the promiscuity of proteins and characteristics of residues in the vicinity of the catalytic site that regulate promiscuity. *PLoS ONE* 7: e32011.
 33. Jensen RA (1976) Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* 30: 409–425.
 34. LEWIS EB (1951) Pseudoallelism and gene evolution. *Cold Spring Harb Symp Quant Biol* 16: 159–174.
 35. Taw k DS (2010) Messy biology and the origins of evolutionary innovations. *Nat Chem Biol* 6: 692–696.
 36. Cole MF, Gaucher EA (2011) Exploiting models of molecular evolution to efficiently direct protein engineering. *J Mol Evol* 72: 193–203.
 37. Ghuyens JM (1991) Serine beta-lactamases and penicillin-binding proteins. *Annu Rev Microbiol* 45: 37–67.
 38. Hall BG, Barlow M (2004) Evolution of the serine beta-lactamases: past, present and future. *Drug Resist Updat* 7: 111–123.
 39. Matagne A, Dubus A, Galleni M, Frere JM (1999) The beta-lactamase cycle: a tale of selective pressure and bacterial ingenuity. *Nat Prod Rep* 16: 1–19.
 40. Denome SA, Elf PK, Henderson TA, Nelson DE, Young KD (1999) *Escherichia coli* mutants lacking all possible combinations of eight penicillin binding proteins: viability, characteristics, and implications for peptidoglycan synthesis. *J Bacteriol* 181: 3981–3993.
 41. Bhullar K, Waglechner N, Pawlowski A, Koteva K, Banks ED, et al. (2012) Antibiotic resistance is prevalent in an isolated cave microbiome. *PLoS ONE* 7: e34953.
 42. Chesnel L, Zapun A, Mouz N, Dideberg O, Vernet T (2002) Increase of the deacylation rate of PBP2x from *Streptococcus pneumoniae* by single point mutations mimicking the class A beta-lactamases. *Eur J Biochem* 269: 1678–1683.
 43. Urbach C, Evrard C, Pudzaitis V, Fastrez J, Soumillion P, et al. (2009) Structure of PBP-A from *Thermosynechococcus elongatus*, a penicillin-binding protein closely related to class A beta-lactamases. *J Mol Biol* 386: 109–120.
 44. Matagne A, Lamotte-Brasseur J, Frere JM (1998) Catalytic properties of class A beta-lactamases: efficiency and diversity. *Biochem J* 330 (Pt 2): 581–598.
 45. Nicholas RA, Krings S, Tomberg J, Nicola G, Davies C (2003) Crystal structure of wild-type penicillin-binding protein 5 from *Escherichia coli*: implications for deacylation of the acyl-enzyme complex. *J Biol Chem* 278: 52826–52833.
 46. Banerjee S, Pieper U, Kapadia G, Pannell LK, Herzberg O (1998) Role of the omega-loop in the activity, substrate specificity, and structure of class A beta-lactamase. *Biochemistry* 37: 3286–3296.
 47. Savile CK, Janey JM, Mundor EC, Moore JC, Tam S, et al. (2010) Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* 329: 305–309.
 48. Chen CY, Georgiev I, Anderson AC, Donald BR (2009) Computational structure-based redesign of enzyme activity. *Proc Natl Acad Sci USA* 106: 3764–3769.
 49. Sandstrom AG, Wikmark Y, Engstrom K, Nyhlen J, Backvall JE (2012) Combinatorial reshaping of the *Candida antarctica* lipase A substrate pocket for enantioselectivity using an extremely condensed library. *Proc Natl Acad Sci USA* 109: 78–83.
 50. Macdonald SJ, Dowe MD, Harrison LA, Clarke GD, Inglis GG, et al. (2002) Discovery of further pyrrolidine trans-lactams as inhibitors of human neutrophil elastase (HNE) with potential as development candidates and the crystal structure of HNE complexed with an inhibitor (GW475151). *J Med Chem* 45: 3878–3890.
 51. Dandekar AM, Gouran H, Ibanez AM, Uratsu SL, Aguero CB, et al. (2012) An engineered innate immune defense protects grapevines from Pierce disease. *Proc Natl Acad Sci USA* 109: 3721–3725.
 52. Fernandez C, Szyperski T, Bruyere T, Ramage P, Mosinger E, et al. (1997) NMR solution structure of the pathogenesis-related protein P14a. *J Mol Biol* 266: 576–593.
 53. Stintzi A, Heitz T, Prasad V, Wiedemann-Merdinoglu S, Kaufmann S, et al. (1993) Plant 'pathogenesis-related' proteins and their role in defense against pathogens. *Biochimie* 75: 687–706.
 54. Bernick JJ, Simpson W (1976) Distribution of elastase-like enzyme activity among snake venoms. *Comp Biochem Physiol*, B 54: 51–54.
 55. Milne TJ, Abbenante G, Tyndall JD, Halliday J, Lewis RJ (2003) Isolation and characterization of a cone snail protease with homology to CRISP proteins of the pathogenesis-related protein superfamily. *J Biol Chem* 278: 31105–31110.
 56. Korkmaz B, Moreau T, Gauthier F (2008) Neutrophil elastase, proteinase 3 and cathepsin G: physico-chemical properties, activity and physiopathological functions. *Biochimie* 90: 227–242.
 57. Kunkel M, Vuyisich M, Gnanakaran G, Bruening GE, Dandekar AM, et al. (2007) Rapid clearance of bacteria and their toxins: development of therapeutic proteins. *Crit Rev Immunol* 27: 233–245.
 58. Lobkovsky E, Moews PC, Liu H, Zhao H, Frere JM, et al. (1993) Evolution of an enzyme activity: crystallographic structure at 2-A resolution of cephalosporinase from the ampC gene of *Enterobacter cloacae* P99 and comparison with a class A penicillinase. *Proc Natl Acad Sci USA* 90: 11257–11261.
 59. Ekici OD, Paetzel M, Dalbey RE (2008) Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. *Protein Sci* 17: 2023–2037.
 60. Davies C, White SW, Nicholas RA (2001) Crystal structure of a deacylation-defective mutant of penicillin-binding protein 5 at 2.3-A resolution. *J Biol Chem* 276: 616–623.
 61. Kuhn I, Harden P, Bauzon M, Chartier C, Nye J, et al. (2008) Directed evolution generates a novel oncolytic virus for the treatment of colon cancer. *PLoS ONE* 3: e2409.
 62. Nannemann DP, Birmingham WR, Scism RA, Bachmann BO (2011) Assessing directed evolution methods for the generation of biosynthetic enzymes with potential in drug biosynthesis. *Future Med Chem* 3: 809–819.
 63. Dalby PA (2011) Strategy and success for the directed evolution of enzymes. *Curr Opin Struct Biol* 21: 473–480.
 64. Lutz S (2010) Beyond directed evolution-semi-rational protein engineering and design. *Curr Opin Biotechnol* 21: 734–743.
 65. Antikainen NM, Martin SF (2005) Altering protein specificity: techniques and applications. *Bioorg Med Chem* 13: 2701–2716.
 66. James LC, Taw k DS (2003) Conformational diversity and protein evolution—a 60-year-old hypothesis revisited. *Trends Biochem Sci* 28: 361–368.
 67. Depristo MA (2007) The subtle benefits of being promiscuous: adaptive evolution potentiated by enzyme promiscuity. *HFSP J* 1: 94–98.
 68. Harada E, Nakagawa J, Asano T, Taoka M, Sorimachi H, et al. (2012) Functional evolution of duplicated odorant-binding protein genes, *Obp57d* and *Obp57e*, in *Drosophila*. *PLoS ONE* 7: e29710.
 69. Yamamura K, Kaiser ET (1976) Studies on the oxidase activity of copper(II) carboxypeptidase a. *J Chem Soc, Chem Commun* : 830–831.
 70. Gumulya Y, Sanchis J, Rectz MT (2012) Many pathways in laboratory evolution can lead to improved enzymes: how to escape from local minima. *ChemBiochem* 13: 1060–1066.
 71. Macheboeuf P, Contreras-Martel C, Job V, Dideberg O, Dessen A (2006) Penicillin binding proteins: key players in bacterial cell cycle and drug resistance processes. *FEMS Microbiol Rev* 30: 673–691.
 72. Bush K, Jacoby GA (2010) Updated functional classification of beta-lactamases. *Antimicrob Agents Chemother* 54: 969–976.
 73. Urbach C, Fastrez J, Soumillion P (2008) A new family of cyanobacterial penicillin-binding proteins. A missing link in the evolution of class A beta-lactamases. *J Biol Chem* 283: 32516–32526.
 74. Ghosh AS, Young KD (2003) Sequences near the active site in chimeric penicillin binding proteins 5 and 6 affect uniform morphology of *Escherichia coli*. *J Bacteriol* 185: 2178–2186.
 75. Sarkar SK, Dutta M, Chowdhury C, Kumar A, Ghosh AS (2011) PBP5, PBP6 and DacD play different roles in intrinsic β -lactam resistance of *Escherichia coli*. *Microbiology (Reading, Engl)* 157: 2702–2707.
 76. Ghosh AS, Chowdhury C, Nelson DE (2008) Physiological functions of D-alanine carboxypeptidases in *Escherichia coli*. *Trends Microbiol* 16: 309[317].
 77. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98: 10037–10041.
 78. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* 32: W665–667.
 79. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, et al. (2002) The bioperl toolkit: Perl modules for the life sciences. *Genome research* 12: 1611–1618.
 80. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16: 276–277.
 81. Sippl MJ, Wiederstein M (2008) A note on difficult structure alignment problems. *Bioinformatics* 24: 426–427.