

In-Silico Structural and Functional Characterization of a *V. cholerae* O395 Hypothetical Protein Containing a PDZ1 and an Uncommon Protease Domain

Avirup Dutta¹✉, Atul Katarkar²✉, Keya Chaudhuri³*

1 CSIR-SRF, Molecular and Human Genetics Division, CSIR - Indian Institute of Chemical Biology, Kolkata, West Bengal, India, **2** ICMR-SRF, Molecular and Human Genetics Division, CSIR - Indian Institute of Chemical Biology, Kolkata, West Bengal, India, **3** Chief Scientist, Molecular and Human Genetics Division, and Head Academic Affairs Division, CSIR - Indian Institute of Chemical Biology, Kolkata, West Bengal, India

Abstract

Vibrio cholerae, the causative agent of epidemic cholera, has been a constant source of concern for decades. It has constantly evolved itself in order to survive the changing environment. Acquisition of new genetic elements through genomic islands has played a major role in its evolutionary process. In this present study a hypothetical protein was identified which was present in one of the predicted genomic island regions of the large chromosome of *V. cholerae* O395 showing a strong homology with a conserved phage encoded protein. *In-silico* physicochemical analysis revealed that the hypothetical protein was a periplasmic protein. Homology modeling study indicated that the hypothetical protein was an unconventional and atypical serine protease belonging to HtrA protein family. The predicted 3D-model of the hypothetical protein revealed a catalytic centre serine utilizing a single catalytic residue for proteolysis. The predicted catalytic triad may help to deduce the active site for the recruitment of the substrate for proteolysis. The active site arrangements of this predicted serine protease homologue with atypical catalytic triad is expected to allow these proteases to work in different environments of the host.

Citation: Dutta A, Katarkar A, Chaudhuri K (2013) *In-Silico* Structural and Functional Characterization of a *V. cholerae* O395 Hypothetical Protein Containing a PDZ1 and an Uncommon Protease Domain. PLoS ONE 8(2): e56725. doi:10.1371/journal.pone.0056725

Editor: Eugene A. Permyakov, Russian Academy of Sciences, Institute for Biological Instrumentation, Russian Federation

Received: October 12, 2012; **Accepted:** January 14, 2013; **Published:** February 18, 2013

Copyright: © 2013 Dutta et al. 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: The study was funded by Council of Scientific and Industrial Research (CSIR), Govt. of India. AD and AK are recipients of fellowships from CSIR and ICMR, Govt. of India, respectively. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: keya.chaudhuri@gmail.com

✉ These authors contributed equally to this work.

Introduction

Vibrio cholerae, the most notable member of the *Vibrionaceae* family is the etiological agent of epidemic cholera, causing a severe and sometimes lethal diarrheal disease. *V. cholerae* is classified into two serogroups: O1 and nonO1. So far, the toxigenic strains of serogroups O1 and O139 have been found to cause cholera epidemics. There are two biotypes of *V. cholerae* O1, Classical and El Tor. There have been seven major pandemics since 1817. Isolates of the sixth pandemic were of O1 classical biotype [1].

The complete genome of *V. cholerae* classical biotype has been sequenced, which revealed that the genome is composed of two chromosomes, the large and the small chromosome [2]. Cumulatively 3875 genes have been identified. However, 1402 open reading frames, code for hypothetical proteins, the functions of which are not known.

V. cholerae infection is noninvasive. In this organism, the two major virulence factors cholera toxin (CT) and toxin coregulated pili (TCP) have been reported to be encoded on mobile genetic elements. Gene acquisition and other genomic alterations, by the mechanism of Horizontal gene transfer have always played a critical role in the adaptive evolution of prokaryotes. Genomic Islands (GIs) in prokaryotic genomes often contain horizontally transferred genetic materials as evident from the presence of

integrase, transposons, phage mediated genes, etc. in these islands [3–5]. These genomic islands are therefore of critical importance in the evolution of the prokaryotic genomes, their pathogenicity and other special function.

The *ctxAB* genes coding for CT are encoded on a filamentous bacteriophage CTXφ [6]. TCP, an essential colonization factor, was originally designated as part of a pathogenicity island named *Vibrio* pathogenicity island VPI, but this island has later on been proposed to be the genome of a filamentous phage, VPIφ [7]. Clinical trials on volunteers using vaccine strains of *V. cholerae* in which several toxin genes including the cholera toxin were eliminated were performed. Results of those trials showed mild to moderate diarrhea in the subjects clearly suggesting that there are yet to be determined virulence factors in the *V. cholerae* genome [8].

In order to survive distinct stress situations and prevent the accumulation of misfolded and aggregated proteins, all cells employ an efficient protein quality control system consisting of molecular chaperones [9,10] in order to prevent cellular malfunctions and even cell death [11,12]. The high temperature requirement A (HtrA) family of proteases are involved in the key aspects of protein quality control [13]. In *Escherichia coli* they have been reported to monitor the proper folding and the functioning of the proteins in cell envelope and the periplasm [14]. HtrA proteases consists of a chymotrypsin-like serine protease as their

catalytic domain with one or two C-terminal PDZ domains [15,16]. The PDZ domains are responsible for substrate binding and controlling protease function. In case of *E. coli*, three HtrA proteases, DegS, DegP and DegQ are responsible for the protein quality control [17]. Prokaryotic HtrAs have been reported to be involved in not only protein quality control but in pathogenicity as well [18–25]. A similar kind of HtrA - protease DO is present in *Vibrio cholerae* O395 which is a homologue of the DegQ protein of *Escherichia coli* H299. Studies have shown that *htrA* mutant in many Gram negative pathogens are attenuated in animal models and can act as live vaccines [14]. A vaccination study indicated that the purified recombinant DegQ protein acted as a protective immunogen conferring protection upon fish against infection by *V. harveyi* [26].

In the present study a hypothetical protein had been identified which was present in one of the predicted genomic island regions of the large chromosome of *V. cholerae* O395. This hypothetical protein showed strong homology with a conserved phage encoded protein. Homology modeling study indicated that the hypothetical protein was an unconventional and atypical serine protease belonging to HtrA protein family. The predicted 3D-model of the hypothetical protein revealed that it had a serine residue at its catalytic center which utilizes a single catalytic residue for proteolysis. The predicted catalytic triad may constitute the active site for the recruitment of the substrate for proteolysis. Recently revealed crystallographic structure of DegQ and DegP with higher order oligomers suggested that signaling cascade leading to protease activation of 12- and 24-mer HtrA complex was highly conserved and depended on precise positioning of PDZ1 domain upon substrate engagement. The active site arrangements of this predicted serine protease homologue with atypical catalytic triad is expected to allow these proteases to work in different environments of the host.

Results

Identification of genomic islands in *V. cholerae* O395

Co-ordinates of statistically significant horizontally acquired genomic segments of *V. cholerae* O395 were determined by *Design-Island* [27]. A customized Perl script was used to mark out the coding regions from the predicted Genomic Islands (GIs) using the protein table as the reference available at the NCBI database. The results showed that after the *refinement phase* the GIs covered ~44% of the large chromosome and ~41% of the small chromosome (Data not shown). *Design-Island* identified all the known GIs of *V. cholerae* Classical O395, such as CTX ϕ , VPI-1, VPI-2 [28–30]. Along with the known ones, a number of genomic segments, which has the potential of being GIs, were also identified. Some of these new segments were flanked by transposase or integrase genes or had phage or potential phage related genes. The Perl script developed for the visualization of the putative GIs used the coordinates obtained from the output of *Design-Island* to generate a circular map of each chromosome (Figure S1), the newly identified regions are shown in supplementary figures (Figure S2A & Figure S2B).

Our study revealed a distinct GI region in the large chromosome of *V. cholerae* Classical strain O395, which was absent in the El Tor strain N16961 of *V. cholerae*. This unique cluster consisted of a number hypothetical proteins, phage related proteins and other biosynthetic and transferase like proteins. Conserved domain analysis of these hypothetical proteins showed that many of these had domains of phage related proteins, clearly indicating the possibility of gene acquisition from phages. Among these hypothetical proteins one having locus tag VCO395_1035,

came up which did not show any hit with any of the conserved domains of known protein functions as determined by CDD search analysis. However this protein emerged as a potential periplasmic protein when checked for possible localization using the HSLpred [31], CELLO [32,33] and the SubLoc v1.0 servers [34] (vide Subcellular Localization section).

Structure Functional Analysis of the Protein

VCO395_1035

To determine the possible function of *V. cholerae* VCO395_1035, the sequence was subjected to comparative protein structure modeling using the target protein sequence as query for different servers described in materials and methods. Significant hits were obtained for the ModWeb server [35] which retrieved the crystal structure of the protease along with the PDZ1 domain of DegQ from *Escherichia coli* (PDB ID: 3STJ). The alignment coverage region for target residue (17–207) showed the 34% sequence identity with template 3STJ residue 152–309.

Comparative Sequence Analysis and Alignment

The hypothetical protein VCO395_1035 when aligned with *E. coli* DegQ, shared 25.7% identity and 40.7% similarity as shown in Figure 1A. DegQ contains a protease domain and two distinct domains, PDZ1 and PDZ2 at 258–349 and 355–445 amino acid residues respectively. The target sequence showed maximum conserved residue in the coverage of the PDZ1 domain of the protease chain. For the PDZ2 domain, the residues were showing low identity and similarity. PDB structure of 3STJ lacked the PDZ2 domain coordinate hence for the further modeling and analysis was restricted to Protease+PDZ1 domain. The first 241 residues were selected, in which conserved residues were aligned properly with the functionally essential regions of the protein template. The proposed alignment for homology modeling of VCO395_1035 is shown in Figure 1B.

Homology Modeling of VCO395_1035 and Validation

The three-dimensional structure of a hypothetical VCO395_1035 from DegQ of *Escherichia coli* (PDB ID: 3STJ chain A, at 2.6 Å resolution) was used as template for homology modeling. The Comparative modeling of VCO395_1035 was performed using a restrained-based approach implemented in MODELLER9v6 [36]. A set of 10 models for each target protein was constructed. The resulting three-dimensional models of VCO395_1035 were sorted according to scores calculated from discrete optimized protein energy (DOPE) scoring function [37]. The final model that shared the lowest Root Mean Square Deviation (RMSD), relative to the trace (C α atoms) of the crystal structure was selected. The final deviations in the protein structure geometry was regularized by energy minimization with the GROMOS96 force field [38] using Deep View [39] by applying 200 steps steepest descent algorithm and 200 steps conjugate gradients algorithm. The final model had 2 major loops, which arose due to insertion (Figure 1B). The two major loops, one from protease domain (residue 79–89; TPYQFQVGERL) and another from PDZ1 domain (residue 176–189; IIQPRFKPYAHLNANPL) were submitted on FALC-Loop webserver for predicting the local structure of loops [40]. The server was used to construct loop region and to refine unreliable loop region in homology modeling by employing an *Ab-initio* loop modeling method FALC (fragment assembly and analytical loop closure) of designed sequence [41]. The output modeled loop after gradient minimization of FALC which had low DFIRE energy, L-RMSD (C α RMSD of loop after superimposition of loop structures), A-RMSD (C α RMSD of loop

A

vib1035	1	MWQNNALTWPASAGSIQTTAESVTQQVGS	TMSAATGRLTNLQSDANLGRHPLSAEAEALL	60
EC_DegQ	163	RVGDFAVAVGNPFGLGQTATSGIVSALGRSGLNLE	GLENFIQTDASINRG-NS--GGALL	219
		: *:: . * **:::..... :*	: * . :*:**::.* * . ***	
vib1035	61	NLRGELNTFLNQGTVLSATPYQFQVGERLE	SGCYLSPANATKTAAKLRDLSDTHRPKGQ	120
EC_DegQ	220	NLNGELIGI--NTAILAPGGGSGVIG-----	FAIPSNMARTLAQQLIDFGEIKR--GL	268
		::* : : :*:.. .. :*	: *:* :.*** :* *::: :*	
vib1035	121	LYAVAIMVSTQSLGEFVSTLSVVTRAFPLPE	WCQCYRQAEAMSKQEAELHQPAGIIQPR	180
EC_DegQ	269	LGIKGTEMSADIAKAF--NLDVQRGAF-----	VSEVLPGSGSAKAGVKAGDI---	313
		* . :*:: * .*,* **	:*.. . : * ** *	
vib1035	181	FKPYAHLNANPLNDYFAAQGAQIATLES	SLASDASHVIGKLSALAQRANQLSEITATINA	240
EC_DegQ	314	---ITSLNGKPLNS-FAELRSRIATTEP-----	GTKVKLGLLRNGKPLEVEVTLDT	360
		: **::***. ** :.*** *,	*, * . * .: * :.*:::	
vib1035	241	LKSLSGSVYSIK-----	LSGTPESIATQLEQAAAPSTCPH	275
EC_DegQ	361	STSSASAEMITPALEGATLSDGQLKDGKGIK	IDEVVKGSPAAQAGLQKDDVIIGVNRD	420
		.* *. * . *	::*:* : * :: . ..	
vib1035	276	TIASV-----LISSQPQPFFEELLC	SH-----	297
EC_DegQ	421	RVNSIAEMRKVLAAPAIIALQIVRGNESI	YLLMR	455
		: *:	: : : : * : : : : :	

B

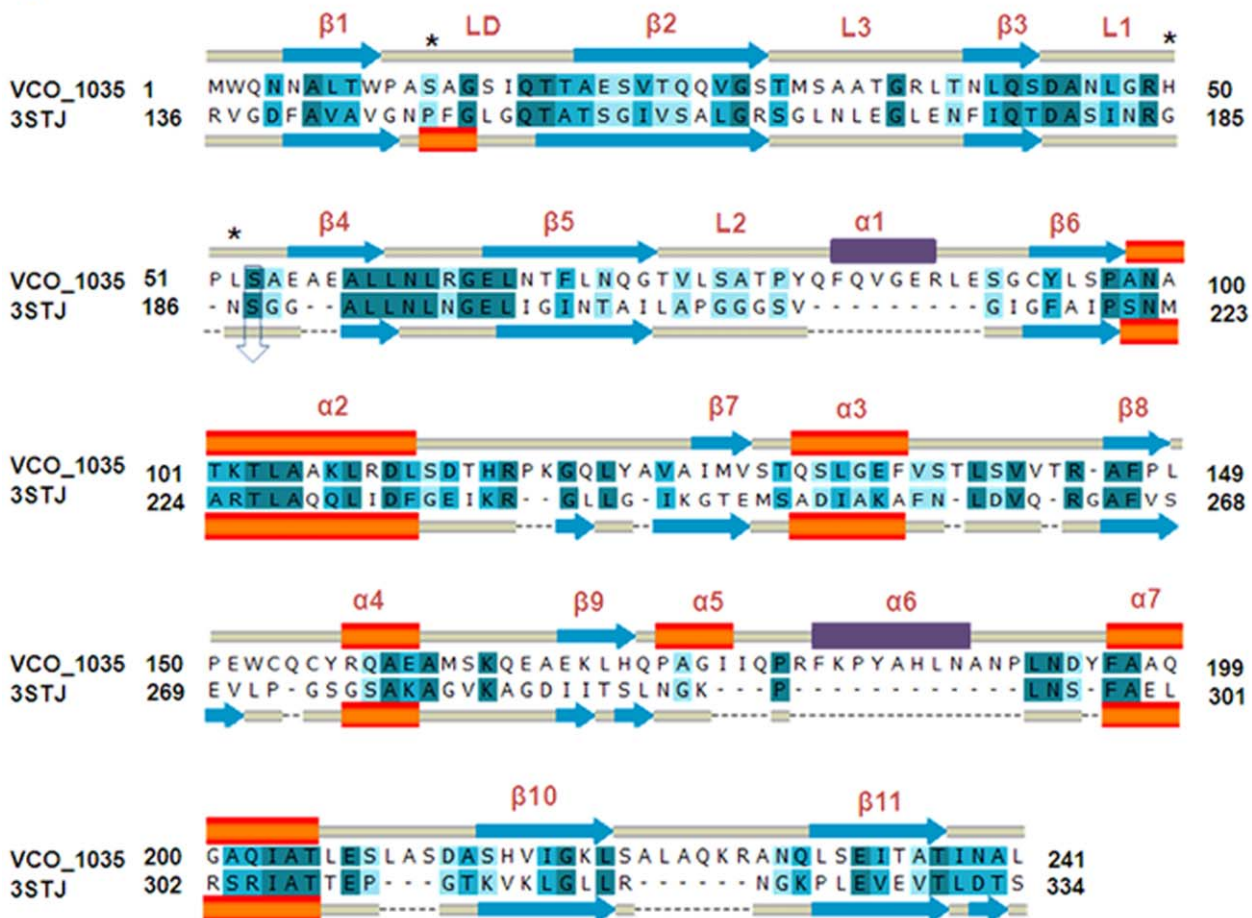


Figure 1. Sequence alignment of VCO395_1035 with *E. coli* DegQ. A. **Sequence alignment of the query (vib1035) and the *E. coli* DegQ (EC_DegQ).** The '*' indicate the conserved amino acids; ':' represents similar group of amino acids. B. **Sequence alignment used for 3D-modeling of VCO395_1035 using *E. coli* DegQ as template (PDB ID: 3STJ).** The blue arrows indicate β sheets, orange bars indicate helix and the yellow bars indicate loops. The deep blue color indicates identical amino acids; lighter blue colors indicate similar and weakly similar amino acids. The two major loop modeled to their corresponding secondary structure were shown in violet color. The predicted catalytic triad residue Ser12-His50-Leu52 indicated by '*' and conserved Ser53 residue with DegQ Ser187 which is one of catalytic triad residue of DegQ of *E. coli* indicated by down arrow.
doi:10.1371/journal.pone.0056725.g001

at the fixed framework) and C-RMSD (C α RMSD of loop of protein structure) was selected and complete loops assembled model further allowed for energy minimization with 100 steps steepest descent and 100 steps conjugate gradients. The final model was validated by using PROCHECK [42] and TM-align [43].

Validation of Homology Model of VCO395_1035

The quality of backbone conformation of model was assessed by PROCHECK for reliability [42]. The observed Psi-Phi pairs had, 82.7% of residues in most favored regions, 15.7% residues in additional allowed regions, 1.1% residues in generously allowed regions and 0.5% residues in disallowed regions as shown in Figure S3 and values shown in Table S1 indicated a good quality model.

The members of HtrA family (DegP, DegQ and DegS) protease exhibit highly extensive ordered secondary structure of α -helix and β -sheet. The final refined model of VCO395_1035 was superimposed with template by using TM-align server [43]. A calculated root-mean-square deviation (RMSD) value of 1.16 Å and TM-score of 0.797 was normalized by length of the template protein. The superimposition of model to the template was shown in Figure S4.

Characterization of Homology Model of VCO395_1035

The 3D model of VCO395_1035 using the template 3STJ, consisted of two domains, namely a protease domain and PDZ1 domain (Figure 2A). The 3D model of VCO395_1035 using the template 3STJ, consisted of two domains, namely a protease domain and PDZ1 domain (Figure 2A). In order to characterize the model, structural motif and mechanistically important loops were assigned to build the final 3D model of VCO395_1035. The final model consisted of 11 β -beta-sheets and 7 α -Helix, the details of which are presented in Table S2.

Protease domain (residue 1–111) consisted of 6 β -sheets arranged anti-parallel to form β -barrel like structure and their positions were stabilized by the corresponding loops which may take part in activation mechanism and active site formation (Figure 2B). PDZ1 domain of 3D-model VCO395_1035 (residue 112–241) consisted of 5 β -sheets and 5 α helix adopted a β -sandwich fold (Figure 2C). The flexible loop of PDZ1 domain of VCO395_1035 contained the highly conserved "carboxylate binding loop" (CBL) (residue 119–122).

Active Site

The protease domain of VCO395_1035 3D-model showed well-defined active site. The alignment of VCO395_1035 with active state DegQ clearly showed conserved active site containing Ser53 (Figure 3). The active site is formed by the proper adjustment of Ser53, Oxyanion hole and the S1 specificity pocket. The amide linkage between Gly48 and Arg49 of loop L1 enabled the Arg49 carbonyl oxygen to interact with the amide nitrogen of Ala13 of loop LD thus allowing the formation of Oxyanion hole. The orientation of the residues Leu47, Gln72, Gly73 and Thr79 form the shallow hydrophobic S1-specificity pocket. The residues which were actively participating in formation of active site

containing Oxyanion hole, S1 pocket and void were shown in Figure 4A, Table S3. The PDZ1 domain of VCO395_1035 containing the deep binding cleft, was formed by the Carboxylate binding loop (CBL), β 7-sheet and α 7-helix. The two hydrophobic pockets were formed P₀ and P₋₂. The residues involved in the formation of hydrophobic binding pockets were shown in Figure 4B, Table S3.

Docking Study

The protease domain and PDZ1 domain were predicted to be involved in substrate binding through the recognition of C-terminal residue of the substrate molecule. In order to check the mode of binding of the substrate molecule in the predicted 3D-model of VCO395_1035, two polyaniline oligopeptides from template (PDB ID: 3STJ) were selected. The protease domain was docked by seven residue polyaniline peptide molecule substrate and PDZ1 domain was docked by five residue polyaniline peptide molecule substrate. The docking was performed by Hex 5.0 software [44] using the reference of the template substrate molecule complex. The best dockpose was then refined and analyzed. The docking study showed active site of the protease domain interacted with substrate molecule by β -augmentation. The residues involved the specific binding of incoming substrate molecule with Ser53 as shown in Table S4. The C-terminal P₀ residue of substrate interacted with Ser53 and P₋₂ residue with the S1 specificity pocket (Figure 4C). The second peptide was bound to PDZ1 domain, the groove of PDZ1 domain was formed between α 7-helix and adjacent to β 7-strand, allowing the C-terminal ends of the substrate molecule to serve as an extra β -strand added to the β -sheets. The C-terminal P₀ residue of the polyaniline was bound to the P₀ pocket and P₋₂ pocket of PDZ1 active site by residue shown Table S4 and Figure 4D. After docking with the substrate molecule, RMSD deviation was calculated which showed that 3D-model had deviated from 1.16 Å to 1.18 Å, suggesting that the mode of binding of substrate molecule with respective binding site were feasible and correct. The docking pose of substrate molecule with respective to the template substrate were shown in Figure 4E & F.

Catalytic Triad in the Protease Domain

The Ser53 present in Loop L1 of VCO395_1035 was found to be conserved with the Loop L1 of the DegQ protease domain template (PDB ID: 3STI). This conserved Ser53 was retained in active site of the protease domain of VCO395_1035. The residues His50 and Leu52 of active site loop L1 were lined up in a one side of the active-site cleft, forming the catalytic triad with Ser12 of loop LD (Figure 5). To examine the role of catalytic triad, the 3D-model of protease domain was generated by utilizing inactive form of DegQ protease domain template (PDB ID: 3STI). On comparing C α distance between the catalytic triad molecules (Table 1) and active site arrangement of active and inactive form of the protease domain (Figure 6A–D), it was clear that the predicted Ser12-His50-Leu52 catalytic triad had an important role in the Oxyanion hole formation, and Ser53 rearrangement in protease active site directly exposed it to substrate molecule.

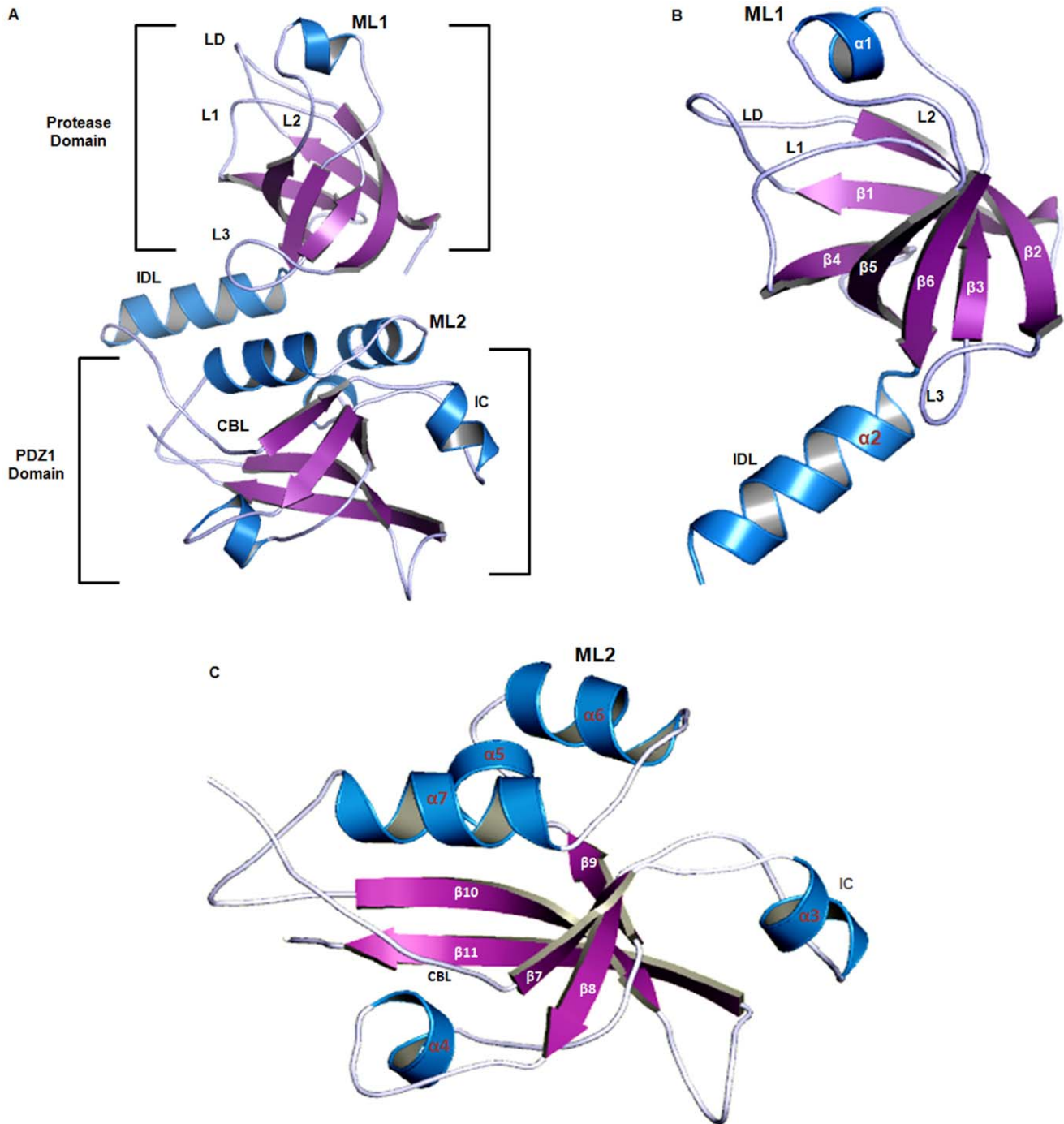


Figure 2. Characterization of Homology Model of VCO395_1035. A. The cartoon representation of 3D modeled structure of VCO395_1035 using PDB ID: 3STJ. Helix (blue), sheets (Purple) and loops (Sky Blue). B. The β -barrel like structure of protease Domain of VCO395_1035 showing active site loops LD: Activation loop, L1: Oxyanion loop, L2: Substrate specificity and L3: Regulatory loop along with interdomain linker (IDL) helix. ML 1: Modeled loop 1 in Protease domain (residue 79–89) on FALC-Loop server indicated as α 1-helix. C. The PDZ1 Domain of VCO395_1035, showing flexible carboxylate binding loop (CBL) and interacting clamp (IC). ML 2: Modeled loop 2 in PDZ1 domain (residue 176–189) on FALC-Loop server indicated as α 6-helix.
doi:10.1371/journal.pone.0056725.g002

Basic Trimeric Unit and Activation Mechanism

It is recognized that DegP of *E. coli* undergo substrate induced oligomer formation and the activation is of vital importance for HtrA protease regulation [45,46]. Recently the same mechanism was observed in the DegQ [17]. It is known that Protease and PDZ domain has the important role in the oligomerization. The

all HtrA protease exhibit a similar domain architecture and share a common trimeric building block, which are controlled by the conserved activation mechanism [47]. It had been observed that in presence of substrate formation of higher order 12-meric particles takes place while in absence of substrate, trimer formation occurred. Moreover, absence of the PDZ1 domain resulted in

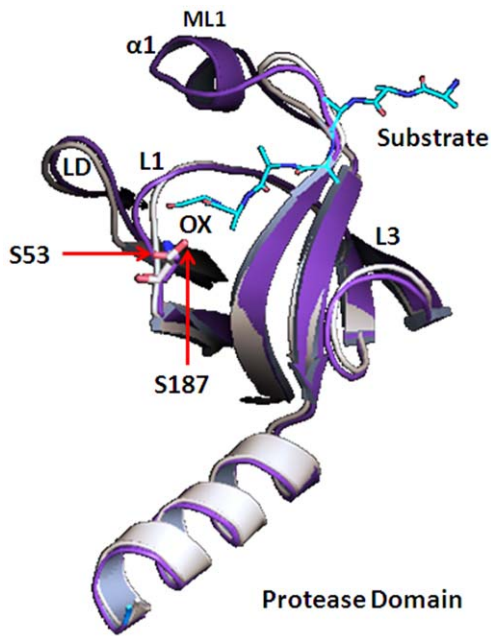


Figure 3. Structural alignment of protease domain. The cartoon representation of protease domain of model VCO395_1035 (magenta) aligned with template 3STJ (light orange) showing conserved Ser53 with DegQ Ser214 which is one of catalytic triad residue of DegQ along with substrate (cyan) bound to active site in Oxyanion hole (OX). doi:10.1371/journal.pone.0056725.g003

protease domain capable of forming basic trimer unit, but was unable to perform the proteolytic activity and underwent higher order oligomerization. It was also observed that only PDZ1 was essential to couple substrate binding with the formation of proteolytically active higher order DegQ oligomers.

In the present study the predicted 3D-model of VCO395_1035, contained the Protease+PDZ1 domain, the essential mechanistically important activation loop and structural motif important in the oligomerization of HtrA family protein. These were well retained in predicted 3D model of VCO395_1035. Hence it could be hypothesized that VCO395_1035 may undergo higher order oligomerization and similar activation mechanism, as found in highly conserved DegP/DegQ HtrA protease.

To study the activation mechanism, the basic trimeric unit of VCO395_1035 was built. Basic trimeric unit (Figure 7) was formed by the docking the monomer into the trimeric unit of template (PDB ID: 3STJ chain A, B & C). The spatial arrangement of trimer of VCO395_1035 resembled a planar triangle with centered protease and PDZ1 domains at the vertices. The peripheral PDZ1 domains contacted with each other through HtrA signature motif IC which was essential for higher order oligomer formation by mediating contact between juxtaposed trimers [46,48]. The interaction clamp comprised hydrophobic region residue 127–147 among which Ser129, Phe136, Leu140, Val142, Ala146 and Phe147 were conserved.

The activation of HtrA protease is known to be reversible process that could be triggered by distinct molecular signals. In DegS the substrate protein RseA signals the folding stress which are recognized and bound by the PDZ domain which are capable of inducing the rearrangement of sensor loop L3 which in turn re-modulate the activation domain into its functional state to cleave the substrate protein [49–51]. In DegP, substrate binding to the first PDZ1 domain induces the oligomer conversation from DegP6 to DegP12 and DegP24. This led to a repositioning and

immobilization of the PDZ1 in such a way so as to induce rearrangement of loop L3 and perform protease activity [47,48]. Similar mode of activation mechanism, as presumed for DegQ upon peptide binding to PDZ1, induces rearrangement of the protease loop L3 and stimulate the protease activity by activating the formation of catalytically active higher order oligomers [17]. The DegP and DegQ indicate the preserved intramolecular PDZ1→L3→LD/L1/L3 signaling constituent in regulating HtrA protease activity in both L2- and 24-meric HtrA oligomers [17]. To explore whether a similar PDZ1→L3→LD/L1/L3 protease activation cascade and molecular interplay between loop L3 and PDZ1 domain occurred in predicted 3D-model of VCO395_1035, the monomer and basic trimeric unit was scrutinized. Interestingly, it was observed that there was the flip in the position of Arg and Gly residue (In the DegQ Arg302 of PDZ1 form a hydrogen bond with carbonyl oxygen of Gly171 in loop L3). In the 3D-model of VCO395_1035, Arg37 in the loop L3 formed the hydrogen bond with carbonyl oxygen of Gly200 in α 7-helix of PDZ1 domain (Figure 8). The R37 of loop L3 interact with G200 of PDZ1 domain allowing Q26 of the loop L3 to interact with the residue I16 of loop LD in the adjacent protease. This may induce remodeling of the proteolytic sites and functional catalytic triad set up between S12 of loop LD and H50 & L52 of loop L1 (Figure 8). Hence the predicated model of VCO395_1035 indicated the preservation of intermolecular PDZ1→L3→LD/L1/L3 signaling event along with set up of catalytic triad. It was further hypothesized that like HtrA protease (DegP, DegQ), the loop L3 served as a molecular switch in regulating higher order oligomerization.

Subcellular Localization

The subcellular localization of VCO395_1035 was predicted using CELLO, an approach based on a two-level support vector machine (SVM) system [32,33]. The CELLO output gave significant reliability for outer membrane (1.493), periplasmic (1.477) and extracellular (1.426). SignalP [52] predicted it as a non-secretory protein. Localization study using the HSLpred [31] and the SubLoc v1.0 servers [34], both predicted it to be a periplasmic protein. This may be because of the fact that the PDZ domains of DegP proteins have been observed to be crucial for membrane localization [53–56]. Further, the lysine residues on the surface of PDZ domains in DegP has been reported to be essential for the lipid membrane attachment [48]. The presence of the lysine and arginine residues on the PDZ domain of the modeled 3D structure of the protein VCO395_1035 indicated that it may interact with the lipid membrane.

It has been well studied in *Escherichia coli* that the functionality of the three HtrA proteases (DegP, DegQ, and DegS) is regulated in the Cytoplasmic membrane via one transmembrane segment. To test this hypothesis and explore if the modeled 3D structure of VCO395_1035 might interact with the lipid membrane, the electrostatic potential of VCO395_1035 was generated by using in PyMOL [57]. The active site had greater positive charge than neutral charge. This mixed electrostatic potential around the active site of Protease domain and PDZ1 domain were assumed to be essential for attraction of C-terminal of substrate which is negatively charged (COO^-) and to perform proper binding of substrate into the active site (Figure 9A). The outer surface of the PDZ1 domain showed strong positive charge (Figure 9B) originating from the cluster of lysine and arginine residues, which might be the candidate site for membrane attachment [48]. The residue Lys-164, Lys-226 and Arg-227 were forming positive electrostatic potential as shown in the inset of Figure 9B.

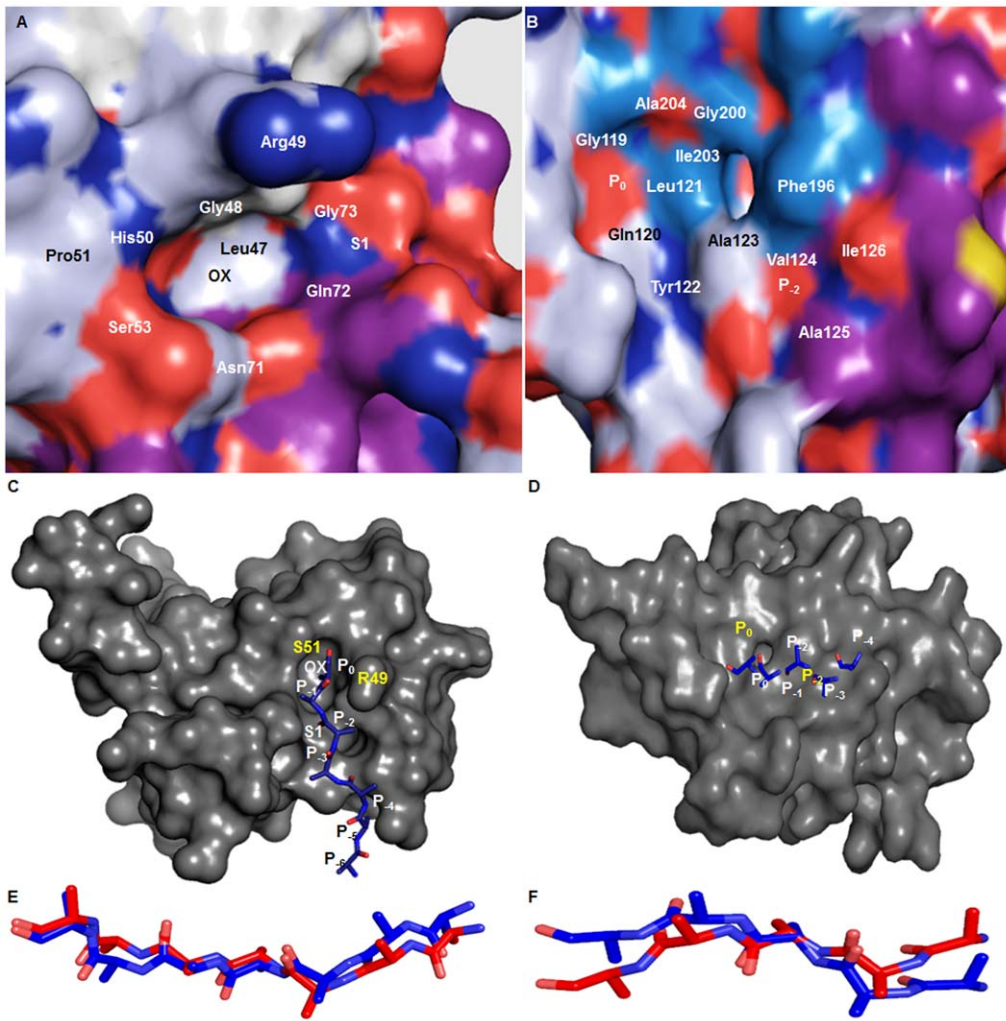


Figure 4. Active site and Protein-substrate interaction using Hex 5.0. A. The surface view of protease domain containing active site showing the oxyanion hole and properly oriented shallow S1 hydrophobic pocket. B. The surface view of PDZ1 containing hydrophobic binding groove formed by CBL and α 7-Helix showing shallow P₀ and P₋₂ substrate binding pocket. C. The C-terminal of poly-alanine peptide substrate (blue) docked into active side of protease domain. D. The C-terminal of poly-alanine peptide substrate (blue) docked into active side of PDZ1 domain via β -aggregation. E. The superimposition of substrate docked into the protease active site (blue) with respect to template (3STJ) substrate (red). F. The superimposition of substrate docked into the active site PDZ1 domain (blue) with respect to template (3STJ) substrate (red). doi:10.1371/journal.pone.0056725.g004

Discussion

In the present study a hypothetical protein VCO395_1035 was identified by *Design-Island* as a part of horizontally acquired region in the large chromosome of *V. cholerae* O395. This gene showed a strong homology with conserved phage protein. To determine the possible function this protein, comparative protein structure modeling was done.

The study showed that the protein VCO395_1035 had >30% sequence similarity to protease+PDZ1 domain of HtrA DegQ, however there was lack of the initial residues containing the LA loop in VCO395_1035 when aligned with DegQ.

In the DegQ protein of *E. coli* the function of the LA loop is still elusive. The LA loop and the subsequent loops contain two of the catalytic triad residues His82 and Asp112 (3STJ). However, proteins with mutations to the catalytic triad have been reported to be present in many enzyme families. It has been estimated that up to 15% of the members of all encoded enzyme families may have lost their catalytic activity [58]. In many cases the inactive

homologues are believed to have acquired alternative functions, such as competing with and antagonizing the active proteases, or otherwise regulating their function. Wrase et.al [59], recently showed in their study of *Legionella* HtrA DegQ homologue, deletion of LA loop did not affect formation of 12-mers in solution or proteolytic activity. There are several proteolytically active unconventional serine protease which having “serine only” configuration in the active site such as *Ochrobactrum anthropi* L-aminopeptidase D-Ala-esterase/amidase [60], *E. coli* Penicillin G acylase precursor [61,62], Glutaryl 7-aminocephalosporanic acid acylase precursor (GCA precursor) [63].

In the predicted 3D-model of VC1035_protPDZ1, simplest catalytic centre serine was discovered which is conserved and is utilized for proteolysis. Unlike a conventional catalytic triad which is usually composed of a Ser, His and an Asp residue the presence of another functionally active catalytic triad gives insight to the understanding of proteolytic mechanism and how serine protease preserved their mode of action.

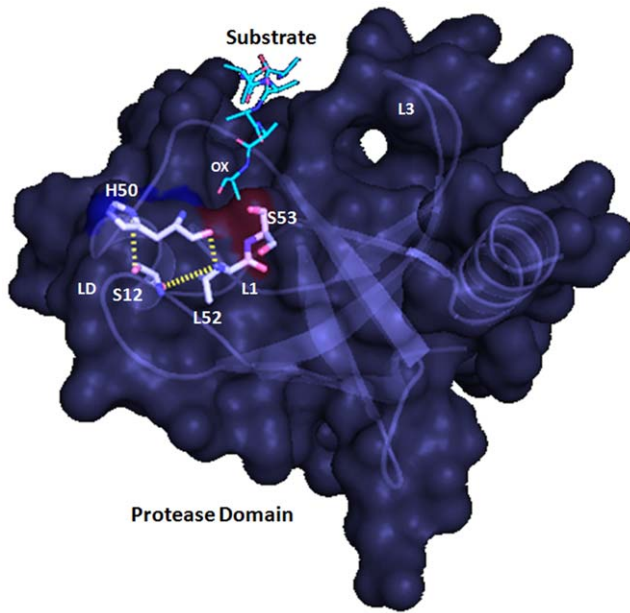


Figure 5. Predicted catalytic triad of model VCO395_1035. The cartoon and surface representation of model VCO395_1035 showing predicted catalytic triad residue Ser12-His50-Leu52 along with substrate (cyan) bound to active site in Oxyanion hole (ox). doi:10.1371/journal.pone.0056725.g005

The HtrA homologue from *E. coli* are under control of substrate-induced oligomer conversion and protease activation, irrespective of the presence of one or two PDZ domains [17,45,48]. Recently revealed crystallographic structure of DegQ and DegP with higher order oligomers suggested that signaling cascade leading to protease activation of 12- and 24-mer HtrA complex was highly conserved and depended on precise positioning of PDZ1 domain upon substrate engagement [17]. The present study revealed one type of serine protease homologue whose active site arrangements allowed these proteases to work in different environments of the host. Our homology modeling study and result analysis indicated that VCO395_1035, which has been annotated as a hypothetical protein, is predicted to be an unconventional serine protease of atypical HtrA homologue performing similar function.

Materials and Methods

Acquisition of Sequences

The complete genome sequences of *V. cholerae* O395, the O1 classical strain of Ogawa serotype isolated in 1964 from India was considered for the present study. The chromosomal sequences of

the organism were downloaded from the ftp server of NCBI database (<http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>).

Detection of putative GI using Design-Island

The program *Design-Island* developed in-house [27] was used for the identification of the putative GIs in the chromosomes of *V. cholerae* O395. *Design-Island* searches for islands in a prokaryotic chromosome using a probing window of varying size that slides over the entire chromosome. It uses an algorithm which is an unsupervised one and applies Monte Carlo’s statistical test on randomly selected segments on a genome. Precise statistical distribution theory then determines the reliable P-values for making the decision.

The program *Design-Island* runs in two phases, namely *first phase* and *refinement phase*. In the *first phase*, it identifies islands at different locations of the chromosome and to determine the stretches of those islands, and carries out statistical analysis using a probing window. This leads to the identification of some ‘putative GIs’ having varying sizes and locations in the chromosome that are identifiable with P-values generated using Monte-Carlo tests carried out at variable locations of the probing window with a fixed size. In the first phase, *Design-Island* was run using $P_0 = 0.05$, word size of 4 and initial window size of 5000 with consequent window increment of 500. 200 randomly selected fragments were tested for each window with a sliding window 500.

Following the *first phase*, *refinement phase* commences which takes random samples of genomic segments excluding the regions detected in the *first phase*. Some of the putative GIs identified in the *first phase*, are further refined into smaller segments containing horizontally acquired genes in the *refinement phase*. In this phase *Design-Island* was run with the same parameter values as used in the *first phase*, except for the initial window size, which was reduced to 2000 and the sliding window increased to 1000. The statistical analysis in the *refinement phase* is similar to that used in the *first phase* except the P_0 was set to 0.001. The results thus obtained were tabulated using customized Perl scripts where the cut-off E-value was set to 0.001.

Localization Study

The subcellular localization of the target protein was predicted using HSLpred [31], CELLO [32,33] (<http://cello.life.nctu.edu.tw/>), SubLoc v1.0 [34] and SignalP [52].

Template Selection for homology

The template selection for the homology modeling of the target protein was performed by submitting the amino acid sequence of the target protein in BLAST [64,65], PBD-BLAST [66], SWISS-Model [67], CPH models [68], 3D-JIGSAW [69], ESyPre3D [70], Geno3D [71], HHpred [72] and ModWeb servers [35].

Table 1. Comparison of the catalytic triad residues and active site arrangement of active and inactive form of the protease domain.

Model	Template	Catalytic Triad residue	Conformation	C α distance of triad (Å)		
				S-H	H-L	L-S
VCO395_1035 (protease)	3STI	Ser12, His50 and Leu52	Inactive	5.6	5.52	7.4
VCO395_1035 (protease+PDZ1)	3STJ	Ser12, His50 and Leu52	Active	4.61	5.12	5.18

doi:10.1371/journal.pone.0056725.t001

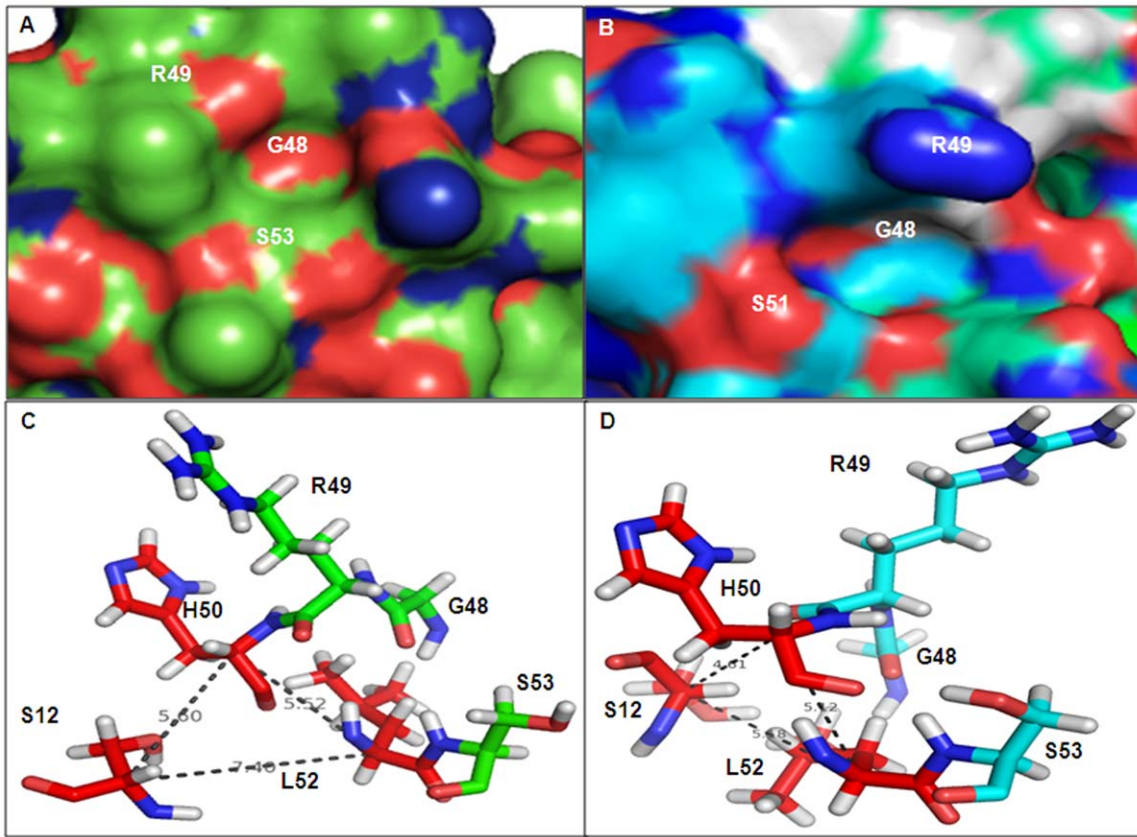


Figure 6. Organization of Active and Inactive form of serine containing proteolytic active site and Catalytic triad. A. Substrate binding site of inactive form of protease domain modeled using template 3STI. B. Substrate binding site of active form of protease domain modeled using template 3STJ. C. The orientation and C α distance between the catalytic triad molecules in the inactive form. D. The orientation and C α distance between the catalytic triad molecules in the active form. doi:10.1371/journal.pone.0056725.g006

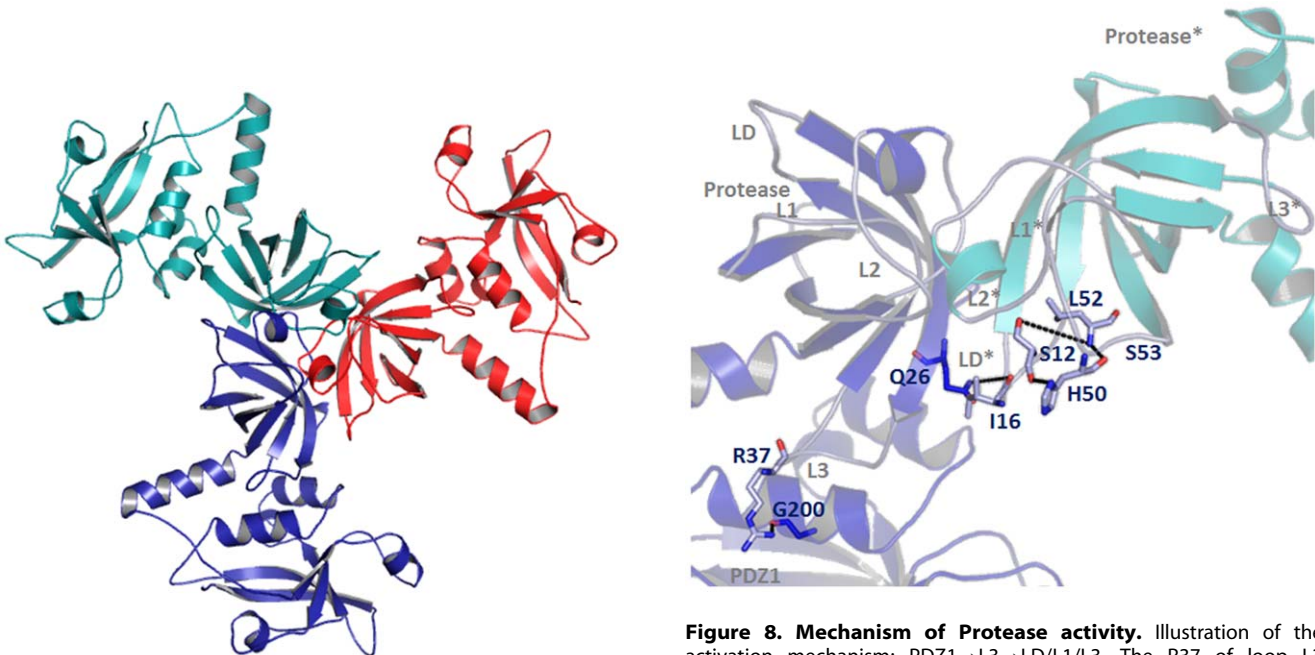


Figure 7. Basic Trimer Unit. Basic trimeric unit of the hypothetical protein (VCO395_1035) was formed by superimposing the monomer into the trimeric unit of template (PDB ID: 3STJ chain A, B & C). doi:10.1371/journal.pone.0056725.g007

Figure 8. Mechanism of Protease activity. Illustration of the activation mechanism: PDZ1→L3→LD/L1/L3. The R37 of loop L3 interact with G200 of PDZ1 domain which allows Q26 of the loop L3 to interact with the residue I16 of loop LD in the adjacent protease shown in * mark. This may induce remodeling of the proteolytic sites and functional catalytic triad set up between S12 of loop LD and H50 & L52 of loop L1. doi:10.1371/journal.pone.0056725.g008

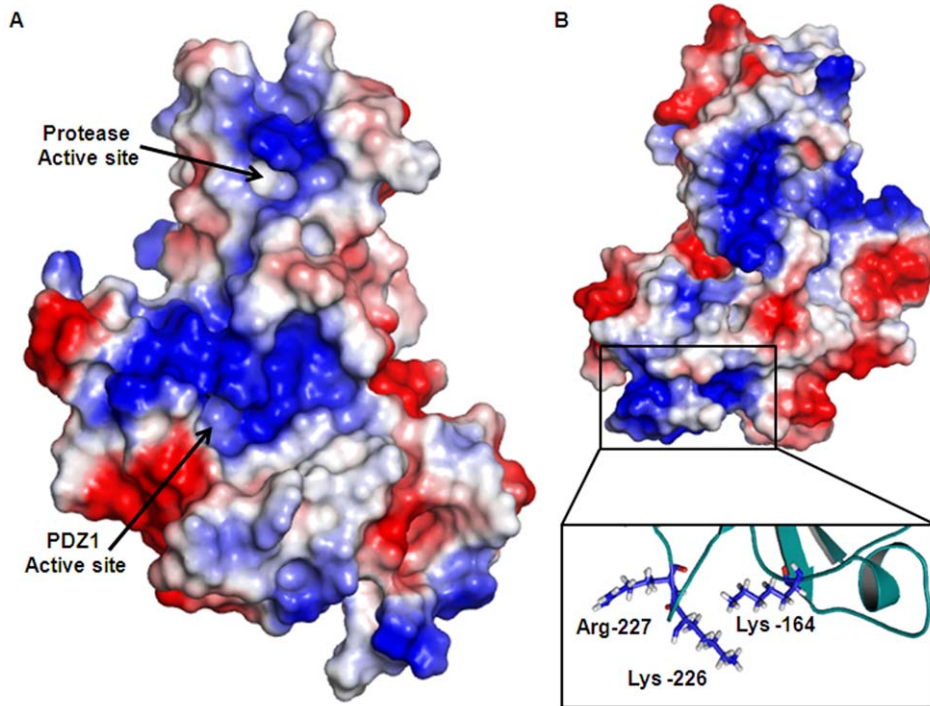


Figure 9. Surface electrostatic potential calculated by PyMOL. The positive charge shown in Blue and negative charge shown in Red. A. The active site of protease and PDZ1 domain showing more positively charge and less neutral environment. B. The outer surface of VCO395_1035 showing the blue patches spreads all over the molecule. The positive charge shown in rectangular frame is aggregated from Arginine and Lysine residue. The inset shows orientation of Arg-164 & 227 and Lys-226 residue in cartoon representation, which are predicted to interact with outer membrane.

doi:10.1371/journal.pone.0056725.g009

Alignment Study

The alignment study was performed by using CLSTALW [73], FUGE [74], T-Coffee [75] and MUSCLE [76,77] servers. During the alignment, the insertion of gaps were allowed in the region of final alignment in such a way that the secondary structure was not disturbed and first 241 amino acid residues of target were threaded into the Protease+PDZ1 domain (residue 136–334) template structure.

Model Construction and Validation

The three-dimensional structure of the target protein was performed using a restrained-based approach in MODELLER9v6 [36,78]. FALC-Loop: Protein Loop Modeling Server was used for predicting the local structure of loops [40]. The final Deviations in the protein structure geometry was regularized by energy minimization with the GROMOS96 force field [38] using Deep View [39]. The final model was validated by using PROCHECK [42] and TM-align [43].

Docking Study

The docking was performed using the Hex 5.0 software [44], with the reference of template complex with the substrate molecule. The electrostatic potential calculation, model visualization and image generation was performed using the PyMOL software [57] (www.pymol.org).

Supporting Information

Figure S1 Algorithmic flow-chart for generation of the circular map indicating GIs on the chromosome. (TIF)

Figure S2 Circular map representing an individual chromosome of *V. cholerae* O395 representing the region covered by the predicted GI. The map shows two circles representing the putative regions of the same chromosome in separate phases. The inner circle with regions marked in blue represents the predicted regions obtained in the first phase of the run by *Design-Island*. The outer circle with red regions represents the putative regions as predicted by *Design-Island* in the refinement phase or the second phase. *V. cholerae* O395 large chromosome. *V. cholerae* O395 small chromosome. (TIF)

Figure S3 Ramachandran plot for predicted 3D model of VCO395_1035 generated by PROCHECK. Most favored regions indicated in red, additional allowed in yellow, generously allowed in light yellow and disallowed regions indicated in white fields. (TIF)

Figure S4 Superposition of 3D-model of VCO395_1035. The superimposition model generated by PyMOL, where VCO395_1035 is shown in pink and the template 3STJ in blue. (TIF)

Table S1 PROCHECK report for the final model of VCO395_1035. (DOC)

Table S2 Characterization of 3D-model of VCO395_1035. (DOC)

Table S3 Residues involve in the active site formation. (DOC)

Table S4 Residues involve in the substrate binding.
(DOC)**References**

- Chaudhuri K, Chatterjee SN (2009) Cholera Toxins: Springer. 322 p.
- Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, et al. (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406: 477–483.
- Hacker J, Blum-Oehler G, Muhldorfer I, Tschape H (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* 23: 1089–1097.
- Hacker J, Kaper JB (1999) Pathogenicity Islands and Other Mobile Virulence Elements. In: Kaper JB, Hacker J, editors. Washington, DC: Am. Soc. Microbiol. pp. 1–11.
- Hacker J, Kaper JB (2000) Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* 54: 641–679.
- Waldor MK, Mekalanos JJ (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272: 1910–1914.
- Karaolis DK, Somara S, Maneval DR Jr, Johnson JA, Kaper JB (1999) A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* 399: 375–379.
- Kaper JB, Lockman H, Baldini MM, Levine MM (1984) Recombinant nontoxigenic *Vibrio cholerae* strains as attenuated cholera vaccine candidates. *Nature* 308: 655–658.
- Gottesman S, Wickner S, Maurizi MR (1997) Protein quality control: triage by chaperones and proteases. *Genes Dev* 11: 815–823.
- Wickner S, Maurizi MR, Gottesman S (1999) Posttranslational quality control: folding, refolding, and degrading proteins. *Science* 286: 1888–1893.
- Macario AJ, Conway de Macario E (2005) Sick chaperones, cellular stress, and disease. *N Engl J Med* 353: 1489–1501.
- Selkoe DJ (2003) Folding proteins in fatal ways. *Nature* 426: 900–904.
- Clausen T, Southan C, Ehrmann M (2002) The HtrA family of proteases: implications for protein composition and cell fate. *Mol Cell* 10: 443–455.
- Pallen MJ, Wren BW (1997) The HtrA family of serine proteases. *Mol Microbiol* 26: 209–221.
- Kirk R, Clausen T (2010) PDZ domains as sensors of other proteins. In: Spiro S, Dixon R, editors. *Sensory Mechanisms in Bacteria: Molecular Aspects of Signal Recognition*: Caister Academic Press. pp. 231–254.
- Hansen G, Hilgenfeld R (2012) Architecture and regulation of HtrA-family proteins involved in protein quality control and stress response. *Cell Mol Life Sci*.
- Sawa J, Malet H, Krojer T, Canellas F, Ehrmann M, et al. (2011) Molecular adaptation of the DegQ protease to exert protein quality control in the bacterial cell envelope. *J Biol Chem* 286: 30680–30690.
- Antelmann H, Darmon E, Noone D, Veening JW, Westers H, et al. (2003) The extracellular proteome of *Bacillus subtilis* under secretion stress conditions. *Mol Microbiol* 49: 143–156.
- Cortes G, de Astorza B, Benedi VJ, Alberti S (2002) Role of the htrA gene in *Klebsiella pneumoniae* virulence. *Infect Immun* 70: 4772–4776.
- Ibrahim YM, Kerr AR, McCluskey J, Mitchell TJ (2004) Role of HtrA in the virulence and competence of *Streptococcus pneumoniae*. *Infect Immun* 72: 3584–3591.
- Jones CH, Bolken TC, Jones KF, Zeller GO, Hruba DE (2001) Conserved DegP protease in gram-positive bacteria is essential for thermal and oxidative tolerance and full virulence in *Streptococcus pyogenes*. *Infect Immun* 69: 5538–5545.
- Lewis C, Skovierova H, Rowley G, Rezuchova B, Homerova D, et al. (2009) *Salmonella enterica* Serovar Typhimurium HtrA: regulation of expression and role of the chaperone and protease activities during infection. *Microbiology* 155: 873–881.
- Mo E, Peters SE, Willers C, Maskell DJ, Charles IG (2006) Single, double and triple mutants of *Salmonella enterica* serovar Typhimurium degP (htrA), degQ (hhoA) and degS (hhoB) have diverse phenotypes on exposure to elevated temperature and their growth in vivo is attenuated to different extents. *Microb Pathog* 41: 174–182.
- Raivio TL (2005) Envelope stress responses and Gram-negative bacterial pathogenesis. *Mol Microbiol* 56: 1119–1128.
- Wilson RL, Brown LL, Kirkwood-Watts D, Warren TK, Lund SA, et al. (2006) *Listeria monocytogenes* 10403S HtrA is necessary for resistance to cellular stress and virulence. *Infect Immun* 74: 765–768.
- Zhang WW, Sun K, Cheng S, Sun L (2008) Characterization of DegQVh, a serine protease and a protective immunogen from a pathogenic *Vibrio harveyi* strain. *Appl Environ Microbiol* 74: 6254–6262.
- Chatterjee R, Chaudhuri K, Chaudhuri P (2008) On detection and assessment of statistical significance of Genomic Islands. *BMC Genomics* 9: 150.
- Faruque SM, Mekalanos JJ (2003) Pathogenicity islands and phages in *Vibrio cholerae* evolution. *Trends Microbiol* 11: 505–510.
- Murphy RA, Boyd EF (2008) Three pathogenicity islands of *Vibrio cholerae* can excise from the chromosome and form circular intermediates. *J Bacteriol* 190: 636–647.
- O'Shea YA, Finnan S, Reen EJ, Morrissey JP, O'Gara F, et al. (2004) The *Vibrio* seventh pandemic island-II is a 26.9 kb genomic island present in *Vibrio cholerae* El Tor and O139 serogroup isolates that shows homology to a 43.4 kb genomic island in *V. vulnificus*. *Microbiology* 150: 4053–4063.
- Garg A, Bhasin M, Raghava GP (2005) Support vector machine-based method for subcellular localization of human proteins using amino acid compositions, their order, and similarity search. *J Biol Chem* 280: 14427–14432.
- Yu CS, Chen YC, Lu CH, Hwang JK (2006) Prediction of protein subcellular localization. *Proteins* 64: 643–651.
- Yu CS, Lin CJ, Hwang JK (2004) Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci* 13: 1402–1406.
- Hua S, Sun Z (2001) Support vector machine approach for protein subcellular localization prediction. *Bioinformatics* 17: 721–728.
- Eswar N, John B, Mirkovic N, Fiser A, Ilyin VA, et al. (2003) Tools for comparative protein structure modeling and analysis. *Nucleic Acids Res* 31: 3375–3380.
- Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234: 779–815.
- Shen MY, Sali A (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci* 15: 2507–2524.
- van Gunsteren WF, Billeter SR, Eising A, Hünenberger PH, Krüger P, et al. (1996) *Biomolecular Simulations: The GROMOS 96 Manual and User Guide*. Zürich: Verlag der Fachvereine Hochschulverlag AG an der ETH Zurich.
- Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18: 2714–2723.
- Ko J, Lee D, Park H, Coutsias EA, Lee J, et al. (2011) The FALC-Loop web server for protein loop modeling. *Nucleic Acids Res* 39: W210–214.
- Lee J, Lee D, Park H, Coutsias EA, Seok C (2010) Protein loop modeling by using fragment assembly and analytical loop closure. *Proteins* 78: 3428–3436.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK - a program to check the stereochemical quality of protein structures. *J App Cryst* 26: 283–291.
- Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33: 2302–2309.
- Ritchie DW, Kemp GJ (2000) Protein docking using spherical polar Fourier correlations. *Proteins* 39: 178–194.
- Jiang J, Zhang X, Chen Y, Wu Y, Zhou ZH, et al. (2008) Activation of DegP chaperone-protease via formation of large cage-like oligomers upon binding to substrate proteins. *Proc Natl Acad Sci U S A* 105: 11939–11944.
- Krojer T, Garrido-Franco M, Huber R, Ehrmann M, Clausen T (2002) Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. *Nature* 416: 455–459.
- Krojer T, Sawa J, Huber R, Clausen T (2010) HtrA proteases have a conserved activation mechanism that can be triggered by distinct molecular cues. *Nat Struct Mol Biol* 17: 844–852.
- Krojer T, Sawa J, Schafer E, Saibil HR, Ehrmann M, et al. (2008) Structural basis for the regulated protease and chaperone function of DegP. *Nature* 453: 885–890.
- Hasselblatt H, Kurzbauer R, Wilken C, Krojer T, Sawa J, et al. (2007) Regulation of the sigmaE stress response by DegS: how the PDZ domain keeps the protease inactive in the resting state and allows integration of different OMP-derived stress signals upon folding stress. *Genes Dev* 21: 2659–2670.
- Walsh NP, Alba BM, Bose B, Gross CA, Sauer RT (2003) OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain. *Cell* 113: 61–71.
- Wilken C, Kitzing K, Kurzbauer R, Ehrmann M, Clausen T (2004) Crystal structure of the DegS stress sensor: How a PDZ domain recognizes misfolded protein and activates a protease. *Cell* 117: 483–494.
- Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8: 785–786.
- Mortier E, Wuytens G, Leenaerts I, Hannes F, Heung MY, et al. (2005) Nuclear speckles and nucleoli targeting by PIP2-PDZ domain interactions. *Embo J* 24: 2556–2565.
- Pan L, Wu H, Shen C, Shi Y, Jin W, et al. (2007) Clustering and synaptic targeting of PICK1 requires direct interaction between the PDZ domain and lipid membranes. *Embo J* 26: 4576–4587.
- Yan J, Wen W, Xu W, Long JF, Adams ME, et al. (2005) Structure of the split PH domain and distinct lipid-binding properties of the PH-PDZ supramodule of alpha-syntrophin. *Embo J* 24: 3985–3995.

Author Contributions

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

56. Zimmermann P, Meerschaert K, Reekmans G, Leenaerts I, Small JV, et al. (2002) PIP2-PDZ domain binding controls the association of syntenin with the plasma membrane. *Mol Cell* 9: 1215–1225.
57. Schrödinger LLC (2010) The PyMOL Molecular Graphics System, Version 1.3r1. 1.3 ed.
58. Pils B, Schultz J (2004) Inactive enzyme-homologues find new function in regulatory processes. *J Mol Biol* 340: 399–404.
59. Wrase R, Scott H, Hilgenfeld R, Hansen G (2011) The Legionella HtrA homologue DegQ is a self-compartmentizing protease that forms large 12-meric assemblies. *Proc Natl Acad Sci U S A* 108: 10490–10495.
60. Ekici OD, Paetzel M, Dalbey RE (2008) Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. *Protein Sci* 17: 2023–2037.
61. Choi KS, Kim JA, Kang HS (1992) Effects of site-directed mutations on processing and activities of penicillin G acylase from *Escherichia coli* ATCC 11105. *J Bacteriol* 174: 6270–6276.
62. Hewitt L, Kasche V, Lummer K, Lewis RJ, Murshudov GN, et al. (2000) Structure of a slow processing precursor penicillin acylase from *Escherichia coli* reveals the linker peptide blocking the active-site cleft. *J Mol Biol* 302: 887–898.
63. Kim Y, Kim S, Earnest TN, Hol WG (2002) Precursor structure of cephalosporin acylase. Insights into autoproteolytic activation in a new N-terminal hydrolase family. *J Biol Chem* 277: 2823–2829.
64. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410.
65. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
66. Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, et al. (2005) Protein database searches using compositionally adjusted substitution matrices. *Febs J* 272: 5101–5109.
67. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* 31: 3381–3385.
68. Nielsen M, Lundegaard C, Lund O, Petersen TN (2010) CPHmodels-3.0—remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Res* 38: W576–581.
69. Bates PA, Kelley LA, MacCallum RM, Sternberg MJ (2001) Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins Suppl* 5: 39–46.
70. Lambert C, Leonard N, De Bolle X, Depiereux E (2002) ESYPred3D: Prediction of proteins 3D structures. *Bioinformatics* 18: 1250–1256.
71. Combet C, Jambon M, Deleage G, Geourjon C (2002) Geno3D: automatic comparative molecular modelling of protein. *Bioinformatics* 18: 213–214.
72. Soding J (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21: 951–960.
73. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680.
74. Shi J, Blundell TL, Mizuguchi K (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol* 310: 243–257.
75. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302: 205–217.
76. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.
77. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797.
78. Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. *Protein Sci* 9: 1753–1773.