American Journal of Bioinformatics Research 2017, 7(2): 49-58
DOI: 10.5923/j.bioinformatics.20170702.01

3D Structure Modeling of Major Subunit and Chaperone of CS3 Pili: Prediction of Residues Conferring Assembly of Fimbriae

Vajiheh Eskandari[1,*], Bagher Yakhchali[2], Seyed Shahriar Arab[3]

[1]Department of Biology, Faculty of Science, Zanjan University, Zanjan, Iran
[2]National Institute of Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran
[3]Department of Biophysics, Faculty of Biological Science, Tarbiat Modarres University, Tehran, Iran

Abstract The objective of the current study is to model the three dimensional (3D) structures of major subunit (CstH) and chaperone (CS3-1) of CS3 pili using computational methods, particularly comparative protein modeling. The generated models of CstH and CS3-1 have been deposited into the PMDB database under the ID numbers PM0079873 and PM0078481, respectively. The 3D structures were utilized in the molecular dimerization in order to identify the potential interaction sites in the CstH molecule. The interacting surface areas of the dimer molecules were determined using the PDBePISA server. The interaction affinities and electrostatic potential of the CstH subunit surfaces and its complexes were calculated by PPEPred server and APBS software, respectively. The results indicated that analysis of the 3D models presents an important insight into the permissive sites, binding mode and contact sites of the CstH protein and our method can be a useful tool for other experimental studies in the biology of the pili and their applications.

Keywords CstH, Homology modelling, Permissive site and *Molecular dimerization*

1. Introduction

Gram negative bacteria can produce surface hair-like structures referred to as pili or fimbriae. They comprise subunits called pilin which are assembled into pili with the aid of a periplasmic chaperone and an outer membrane transporter protein [1-3].

In the pilus assembly pathway, after translation and signal peptide cleavage of the precursor proteins, Pilin subunits interact with their cognate chaperones to ensure proper folding [2, 3]. Some fimbriae of the chaperone/usher assembly class, curl into non-fimbrial, capsule-like structures on the cell surface [4], while most such chaperone/usher pilin transport systems assemble into pili on the cell surface [3].

Despite low-sequence homology among Gram-negative bacterial *fimbriae*, all pilin subunits share an incomplete immunoglobulin (Ig)-like fold, which lacks the seventh, C-terminal, final antiparallel β-strand (strand G), creating a deep hydrophobic cleft that makes the subunits unstable outside the pili or chaperone–subunit complex. The periplasmic chaperone contains two Ig-like domains that meet at a right angle, giving the molecule an overall `boomerang' shape with a cleft between the two domains. Chaperones bind fimbrial subunits by insertion of their N-terminal strand G (β-strands G1) into the hydrophobic cleft of the pilin in a parallel orientation, so creating a non-classical Ig-like fold [5, 6]. Pilus assembly entails a similar principle of complementing the incomplete Ig-fold of the pilins. The assembly of subunits into pili carries out through a donor strand exchange reaction in which the inserted G1 β-strand of the chaperone, with a parallel directionality; into the subunit hydrophobic cleft is replaced by the N-terminal domain of a second subunit in the more energetically suitable anti-parallel orientation to complete the Ig-like fold and joining subunits into a fiber [3]. The N-terminal structural characterization of several fimbriae such as F1 capsule antigen from *Yersinia pestis* indicates that the N-terminal of capsule subunit contains a large unstructured region which is important in interacting with β-strand of the next subunit, presumably by adopting ‘new-structural’ state [4].

The fimbrial proteins also indicate a fair amount of sequence variability making them amenable for the acceptance of foreign sequences and displaying on the bacterial surface [7]. The CS3 fimbriae/pili expressed by most strains of enterotoxigenic *E. coli* afford several potential advantages which make it a suitable system for high-valence display of heterologous peptides on the

* Corresponding author:
Veskandari@znu.ac.ir (Vajiheh Eskandari)
Published online at http://journal.sapub.org/bioinformatics

bacterial cell surface [8-10].

The present study was conducted to model the three dimensional structure of the major subunit (CstH) and chaperone (CS3-1) of the CS3 pili and identify the CstH self-dimerization as well as CstH and CS3-1 hetero dimerization regions and *potential* CstH interface residues.

Knowledge on 3D structure of a protein provides important information for understanding its biochemical function and interaction properties in molecular detail which may have important applications. Therefore, it is plausible to *expect* that our molecular modeling efforts along with previous experimental [10] and literatures data can facilitate the identification of permissive foreign peptides insertion sites, *ligand-binding* sites along with rational designing of mutations in the protein subunit for the purpose of tightening and loosening of the pili structure.

2. Methods

2.1. Template Identification, Sequence Alignment, Model Generation and Assessment

The amino acid sequence of target proteins: CS3 fimbrial subunit A (CstH) and CS3 chaperone (CS3-1) were retrieved from the UniProtKB/Swiss-Prot database with accession numbers P15488 and P15483 respectively [11]. Modeling of CstH and CS3-1 were carried out sequentially: the target proteins were searched for their homologues through the Protein Data Bank (PDB) [12] using BLAST [13] and PSI-BLAST [14].

Generation of the 3D model for the CS3-1 chaperone was performed by Modeller 9v10 [15] based on the crystallographic structure of its homologue. The quality of the generated model was assessed using Modeller objective function, DOPE score [16] and stereochemical assessment using the Procheck program [17] and ProSA II-web server [18].

For CstH modeling, the query protein sequence was subjected to fold-recognition using GenTHREADER meta-servers [19] including: FUGUE [20], FFAS03 [21], mGenTHREADER [22] and Phyre-2 [23] servers. The structure- based sequence alignment was carried out to build homology model for CstH using Swiss-Model Alignment-Mode [24, 25]. Finally, the best model was selected based on the QMEAN Z-score [26], stereo-chemical assessment using the Procheck program, ProSA II-web server and the Verify-3D server [27] and also structural comparison with respect to the template.

The CstH and CS3-1 generated models were energetically optimized by applying the all atom *OPLS*-AA *force Field* available in GROMACS package-5.0. Molecular modeling (MD) simulations were carried out at a constant temperature (300 K) and pressure (1 bar) for 25 ns.

The consensus secondary structural elements of the proteins were obtained through the SOPM [28], GOR IV [29], PHD [30] and SIMPA96 [31] servers.

The information about solvent accessibility of CstH and CS3-1 models were obtained using the Accelrys Discovery Studio Visualizer 2.5.

Pro-origami [32] was used to generate the topology diagram from an uploaded PDB file. Protein topology diagrams are 2D representations of protein structure which are particularly useful in understanding and comparing protein folds.

2.2. Protein-protein Dimerization and Prediction of Protein-Protein Interaction Sites

The ternary Caf1M:Caf1:Caf1 (chaperone: subunit: subunit) complex (PDB ID: 1P5U) from *Yersinia pestis* were used as template for dimerization of CstH/CstH and CstH /CS3-1complexes. The 1p5u; pdb file was broken into two file; one file had two Caf1 subunits (Caf1:Caf1) and the other had the caf1 chaperone with caf1 subunit (Caf1:Caf1M). Models of protein complexes and interaction sites between CstH and CstH and also CstH- and Cs3-1 were retrieved by applying symmetry against Caf1:Caf1 and Caf1:Caf1M, respectively, using Swiss-Pdb Viewer. Molecular dynamics simulation of the complexes was carried out with the GROMACS as described above.

The surface interaction area of dimer models were calculated using the PDBePISA server [33]. The electrostatic potential surfaces for CstH/CstH and CstH/CS3-1 complexes were calculated by APBS [34] and displayed using PyMOL [35]. The binding affinities of dimer complexes were measured using a web server PPEPred [36].

2.3. Construction of Phylogenetic Tree

For generation of phylogenetic tree, amino acid sequences of various fimbriaes [37] were retrieved from Swiss-Prot database. Multiple sequence alignment was done using ClustalW [38]. Putative phylogenetic tree for multiple sequence alignment was performed by *TreeTop [39] Phylogenetic tree prediction online server*. In the bootstrap, multiple alignments were reassembled 100 times.

3. Results and Discussion

3.1. Identification of the Best Templates for CstH and CS3-1

Evolutionarily related proteins have similar sequences and naturally occurring homologous proteins have similar protein structure, so BLAST searching was performed for identifying significant similarity to one or more proteins with known 3D structure using CstH and CS3-1 sequences (Swiss-Prot: P15488 and P15483, respectively). Significant similarities were found with several Chaperone families for CS3-1. A Prosite analysis of query sequences also indicated the presence of Gram-negative pili assembly chaperone signature for CS3-1, but no hits were obtained for CstH. In addition, a PSI-BLAST search for CS3-1 against the

structures in Protein Data Bank (PDB) also found chaperone protein Caf1M (PDB ID: 1P5U_A) from *Yersinia pestis*, with lower E-values; 3e-49 and ~39% sequence identity as the most appropriate template for the CS3 chaperone, but no significant hits were obtained for the CstH. The CstH sequence was therefore submitted to the GenTHREADER meta-server for template identification. Fold-recognition servers; FFAS03 using jackal method, FUGUE, 3DPSSM, PHYRE2 and mGENETHREADER reported the crystal structures of F1 capsule antigen (Caf1) from *Yersinia pestis* (PDB entries: 3dbp_B, 1p5v_B, 1p5u_C, 1p5u_B and 3dos_B) as the best potential templates. Moreover the Fold-recognition servers also predicted the 3D models for the query sequence. As indicated in the Table 1, query coverage of the resulted models was not complete. Further investigation of all models showed that the N-terminal of the target protein was not modeled (data not shown). Due to the importance of pilin N-terminal in assembly of the pili, the results of the fold recognition methods were only used for template selection and the F1 capsule antigen (PDB entry; 3dbp_B) with higher scores in compare to other templates (Table 1), was selected as the appropriate template for CstH modeling.

3.2. Modeling of CstH and CS3-1

The 3D models of CstH and CS3-1 (Fig. 1) were *produced* using available crystal structures of Caf1 and Caf1M as structural templates by employing the computational methods. Since the CstH and CS3-1 target proteins and related templates belonged to the same family in which the structures and functions were well conserved, it seems that the functional information and overall structure similarities can overcome the problem of low-sequence identity (below 30%) between CstH and related template.

For CstH modeling, preliminary structure based sequence alignment (Fig. 2) generated by FFAS03 sever using 3dp_B templates followed by secondary structure comparison for improving alignment accuracy. The positions of the secondary structure elements of templates and target proteins were then superimposed onto the aligned sequences. As illustrated in figure 3, a good agreement was found between positions of the secondary structure elements of the templates and target proteins. CstH structural model was created using Alignment-Mode of Swiss-Model by submitting the regimented alignment. Resulted model was evaluated according to Q-MEAN Z-score and (N-terminal) query coverage for preliminary model selection. The model was more evaluated using protein validation tools such as ProSA II web-server and Procheck program for final selection. Properties of the final model are consistent with the standard scores despite relatively low sequence identity between the query and template (Table 2). The sequence alignment used to build CstH model is illustrated in Figure 2.

For CS3-1 modeling, the primary sequence of CS3-1 was aligned to the 1p5u_A using Clustal-W at sequence identities and similarity of 39% and 65%, respectively (Fig. 4). When the *sequence identity* is *at least* 40%, the protein can be *modeled* using the *homology modeling programs [40, 41]*. The CS3-1 model was created using Modeller 9v10 (Fig. 1B). The best model was selected based on the DOPE score. The DOPE score of the model (−19720) was similar to the DOPE score of the template (−21504). Additionally, the model was further validated using Procheck and Prosa II-web server (Table 2).

The molecular dynamics (MD) simulations was performed on generated models for 22 ns and RMSD versus time was plotted to evaluate the conformational stability of the proteins during the MD simulation (Fig. 5). As shown in the RMSD profile (Fig. 5A), the RMSD analysis of CS3-1 simulation showed that the simulated system was reached to equilibrium after 8 ns. The final construct, after 25 ns molecular dynamic (MD) simulation, was deposited in PMDB [42] with PMDB-ID: PM0078481.

But the RMSD of the CstH model couldn't get stable value within ~25 ns. To study the conformational fluctuation of the CstH, the root mean square fluctuation (RMSF) was analyzed (Fig. 5B). The RMSF profile indicated that the CstH fluctuate within the range of ~0.1–1.4 nm in the entire simulation period and the maximum level of fluctuation was located in the residues ~1–22. This flexibility is probably due to the *unstructured region of the CstH N-terminal.* Deviation in the conformational stability of the CstH model was further analyzed by GROMOS96 force field implemented via Swiss-Pdb Viewer and compared to its PDB template. The results showed a good structural quality score for the model consisting residues 22-146, within the range of quality score for its template; caf1 (Low GROMOS energies obtained for CstH and caf1). Therefore it seems, if unstructured N-terminus acquired folded state using homo or hetero-dimer formation then the protein complexes subjected to MD simulation it is possible to obtain an appropriate conformation, and observation of a stable or relatively stable conformation in the trajectory could be expected. Our MD results (dimerization section) indicate that this refolding strategy has been successful in MD simulations. So one mere from CstH homodimer was introduced as CstH model (Fig. 1) and deposited in PMDB with PMDB-ID: PM0079873.

Table 1. CstH templates properties

Server	Template PDB IDs	Query coverage	RMSD (iterative)	SCORE	%Identity
FFAS03	3dpb_B	94.5	0.33(133)	-12.1[a]	23
	3dos_B	93	0.46(130)	-11.2	22
	1p5u_B	94.5	0.35(133)	-10.540	22
	1p5u_C	86	1.12(118)	-12.5	23
FUGUE	hs1p5vb	90	1.7(127)	4.60[b]	
PHYRE2	1p5v_b	86	0.59(111)	0.0055[c]	26
mGENETHREADER	1p5v_B	88	-	0.002[d]	-

[a]According to the FFAS03 benchmark, scores lower than −9.5 exhibit less than 3% of false positives [21].
[b]Z-score ≥ 4.0 give likely 95% confidence [20].
[c]e-value of *0.005* corresponding to more than 95% *fold recognition* confidence [23].
[d]p value cut-offs (< 0.001, < 0.01 and < 0.05) estimated coverage of sequences with assigned structures [45].

Table 2. CstH and CS3-1 models properties

Property	CstH Model	CS3-1 Model
Query coverage	99.3%	99%
Backbone RMSD from template[a]	2.06 Å	3.76 Å
ProsaII Z-score[b]	-3.43	-5.28
% Modeled regions	98.6	100
Procheck validation results[c]		
Most favored regions	82.2	87.2
Additionally allowed regions	12.4	12
Generously allowed regions	3.9	0.5
Disallowed regions	1.6	0.0

[a]A model can be considered as reliable or accurate model when its RMSD is less than 3 – 4 Å [46].
[b]ProsaII Z-score close to -5 is confidence.
[c]Percentage of residues present in the four different regions of the Ramachandran plot

(A)

(B)

Figure 1. 3D structure of templates and models. Visualization of the 3D structures of: **(A)** left image, Caf1 template and right image, CstH model **(B)** left image, Caf1M template and right image, CS3-1 model. The images prepared by using the *Discovery Studio* 2.5 software

```
2    AGPTLTKELALNVLSPAALDATWAPQDNLTLSNTGVSNT--LVGVLTLSN    49

1    ADLTASTTVTVTVVEPARITLTYKEGAPITIMDNGNIDTELLVGTLTLGG    50

50    TSIDTVSIASTNVSDTSKN-GTVTFAHETNNSASFATTISTDNAN---IT    95

51    YK-TGTTSTSVNFTDAAGDPMYLTFTSQDGNNHQFTTKVIGKDSRDFDIS    99

96    LDKNAGNTIVKTTNGSQLPTNLPLKFITTEGNEHLVSGNYRANITIT    142

100    PKVNGENLVGDDVVLATGSQDFFVRSIGSKGGK-LAAGKYTDAVTVT    145
```

Figure 2. The sequence alignment between CstH and Caf1 (PDB ID: 3dpb_B) which retrieved from FFAS03 server and used to build the CstH model

```
Sec.Cons.      ccccccchhhhhheccccheeeccccccceeeeccccccc--eeeeeeecc
Sec.Str.       ccccccceeeeeeccccceeeeeccccccceeeeccccccceeeeeeeecc
Query          AGPTLTKELALNVLSPAALDATWAPQDNLTLSNTGVSNT--LVGVLTLSN
Template       ADLTASTTVTVTVVEPARITLTYKEGAPITIMDNGNIDTELLVGTLTLGG
Sec.Cons.      ccceeeeeeeccccccccceeeeeeccccccceeeeeecccccce---ee
Sec.Str.       cc-ceeeeeeeccccccccceeeeeeccccccceeeeeeccccccceeeccc
Query          TSIDTVSIASTNVSDTSKN-GTVTFAHETNNSASFATTISTDNAN---IT
Template       YK-TGTTSTSVNFTDAAGDPMYLTFTSQDGNNHQFTTKVIGKDSRDFDIS
Sec.Cons.      ecccccceeeeeccccccccccceeeeeccccceeeccccccceeeee
Sec.Str.       ccceeeeeccccccccccceeeeeccccceeecc-cccceeeeeeee
Query          LDKNAGNTIVKTTNGSQLPTNLPLKFITTEGNEHLVSGNYRANITIT
Template       PKVNGENLVGDDVVLATGSQDFFVRSIGSKGGK-LAAGKYTDAVTVT
```

Figure 3. Comparison of CstH and Caf1 (3dpb_b) protein pairs at the secondary structure level. The secondary structure of Caf1 was retrieved from protein data bank (PDB). Secondary structural elements were shown in three states: 'E': indicating b-sheet; 'H': helix; and 'C': coil

```
CS3-1     -N------NITTQKFEAILGATRVIYHLDGNGESLRVKNPQISPILIQSKVMDEG-SKDN 52

Caf1M      ANSAQPDIKFASKEYGVTIGESRIIYPLDAAGVMVSVKNTQDYPVLIQSRIYDENKEKES 60

           *         ::::::: . :* :*:** **. *   : ***.*   *:****:: **. . *:.

CS3-1     AD-FIVTPPLFRLDAKRETDIRIVMVNGLYPKDRESLKTLCVRGIPPKQGDLWAN---NE 108

Caf1M      EDPFVVTPPLFRLDAKQQNSLRIAQAGGVFPRDKESLKWLCVKGIPPKDEDIWVDDATNK 120

           * *:***********:.....:**. ..*::*:*:**** ***:*****: *:*.:     *:

CS3-1     KEF-----VGMKLNVSINTCIKLILRPHNLPKLDINSEGQIEWGIRDGNLVAKNKTPYYF 163

Caf1M      QKFNPDKDVGVFVQFAINNCIKLLVRPNELKGTPIQFAENLSWKVDGGKLIAENPSPFYM 180

           ::*       **: ::.:**.****::**::*      *:    ::.* : . *:*:*:*:* :*:*:

CS3-1     TIVNASFNGKALKTPGTLGPYEQKLYTLPSKISVSGLVKWEIIGDLG--ESSETKKFNI 220

Caf1M      NIGELTFGGKSIPS-HYIPPKSTWAFDLPKGLAGARNVSWRIINDQGGLDRLYSKNVTL 238

           .* : :*.**:: :    : * .    : **. :: :   *.*.**.* *   :    :*:..:
```

Figure 4. CLUSTALW multiple sequence alignment between CS3-1 and Caf1M

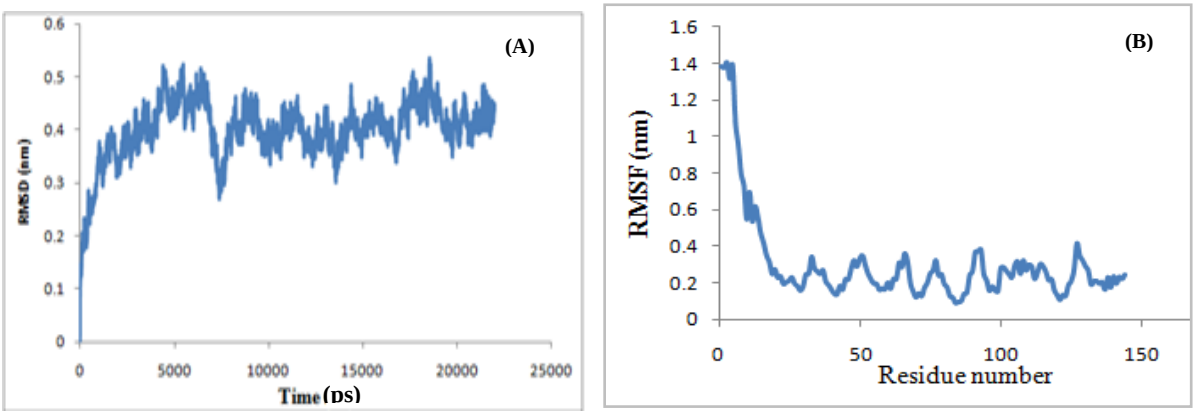

Figure 5. Structural stability assessment during the MD simulations: **(A)** Root mean square deviation (RMSD) plot for modeled CS3-1 during molecular dynamic simulation. The fluctuations of the modeled structure reached plateau after ~8ns simulation. (**B)** Root-mean-square fluctuations of the CstH molecule during the trajectory period of simulation: there is a high flexibility in N-Terminal of the CstH protein

3.3. Comparison of CS3 Fimbrial Subunit A with F1 Capsule Antigen Caf1

As shown in Figure 6, each monomer of CS3 pili has a fold consisting of one α- helix and ten β-strands (β1-β2-α1-β3-β4-β5-β6-β7-β8-β9-β10), which is compatible with 3dpb_b template. We also investigated the structural relation between the target and template through phylogenetic analysis (Fig. 7). The major subunit of several pilus were used for phylogenetic analysis. Inspection of the phylogenetic tree indicated a close relation between the CS3 major subunit (CstH) and F1 capsule antigen (Caf1) that explains conserved amino acids across protein families sharing the same fold, resulting in an identification of the Caf1 protein as the appropriate candidate for accurate modeling of the CstH. In the model building, accurate alignment of the model structure with the selected template is a crucial step and poor alignments are commonly responsible for inaccurate models. As expected, the model is quietly super-imposable with the template (Table 2) and the root-mean-square deviations (RMSD) using Cα atoms being 2.11 Å. These results show a strong structural resemblance of template and model and confirm the accuracy of the generated model for CS3.

3.4. Dimerization Interface of Subunit-Subunit Homodimer and Subunit-Chaperone Heterodimer

A pilus is a polymer of protein subunits that, before assembly, form transient complexes with a chaperone in the periplasm. Then, during pilus assembly, the chaperone is replaced by another subunit [1, 3].

In this study, dynamically stability of dimer interface in CstH/CstH and CstH-/CS3-1 dimer complexes were measured by estimating the root mean square deviation (RMSD) during MD simulation. Plots of RMSD versus simulation time (Fig. 8) shows that the 15 ns simulation time is sufficient for equilibration of all systems, thereby providing best evidence for introduction of the intact mere of CstH homodimer as model.

Evaluation of the protein-protein interface residues of the final dimer complexes using the PDBePISA server revealed that the N-terminal of one CstH molecule interacts with C-terminal of other molecule in CstH/CstH complex. The N-terminal and C-terminal amino acids and some other residues of CstH subunit contribute in CstH-Cs3-1 interface (Table 3).

The interaction of CstH subunit with CS3-1 chaperone or another CstH subunit were also analyzed by electrostatic potential (ESP) calculations and PPEPred server to examine the interface surfaces and binding affinity of protein- protein complexes, respectively. ESP represented dimer region clefts (encircled with an arrow) of CstH/CstH and CstH/CS3-1 were completely engulfed by a hydrophobic and mostly hydrophilic residues (Fig. 9). Hydrophobic interactions and electrostatic complementarity were critical factors for high-affinity binding [43]. The hydrophilic residues were also investigated with the Discovery Studio software. The results confirmed that the most residues of models for CstH and CS3-1 were exposed to the solvent (Fig. 10), which is important in protein-protein interaction [44]. The measured binding affinities were -15.87 and -21.50 kcal/mol for CstH/CstH and CstH/CS3-1 complexes, respectively. Thus, EPS and binding affinity analysis also revealed that the N-terminal and C-terminal of the CstH molecule were in interaction interfaces (Table 3), especially in CstH/CstH complex, which were in agreement with literatures [3, 5, 6] and again confirmed the generated models.

Table 3. Interface and inaccessible residues of CstH in the CstH complexes

NO	Complex	Interfacing Residues
1	CstH-CstH	A2, G3, P4, T5, L6, T7, K8, E9, L10, A11, L12, N13, V14, L15, S16, P17, A18, A19. H129, L130, V131, S132, G133, N134, Y135, R136, A137, N138, I139, T140, I141, T142, S143, T144, I145, K146.
2	CstH-CS3-1	S16, P17, A18, A19, L20, D21, A22, W24, P26, Q27, D28, N29, L30, V42, L47, N49, I52, S56, A58, S59, N61, K67, N68, G69, T70, F73, F83, T86, D97, K98, N99, A100, G101, N102, T103, I104, V105, K106, T107, L117, I120, N138, I139, T140, I141, T142, S143, T144, I145, K146

Note: residues involved in the formation of H-bond are underlined and in each cell, the orientation of sequences are similar to the orientation of molecules in complexes structures

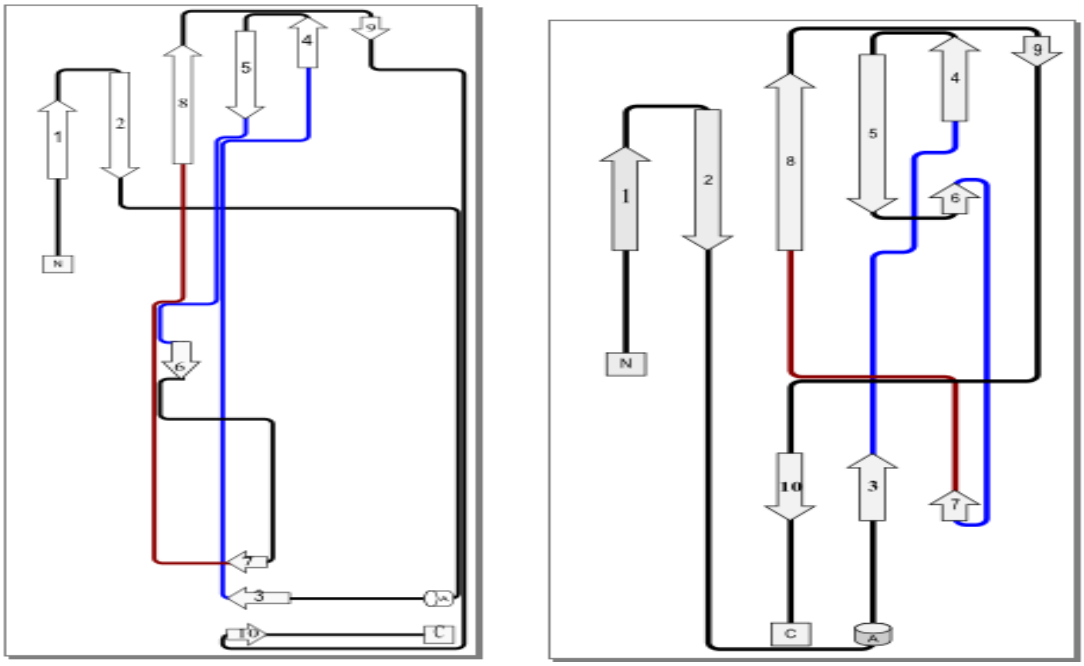

Figure 6. Topology diagrams of the CstH model (left image), and Caf1 (3dpb_b) template (right image)

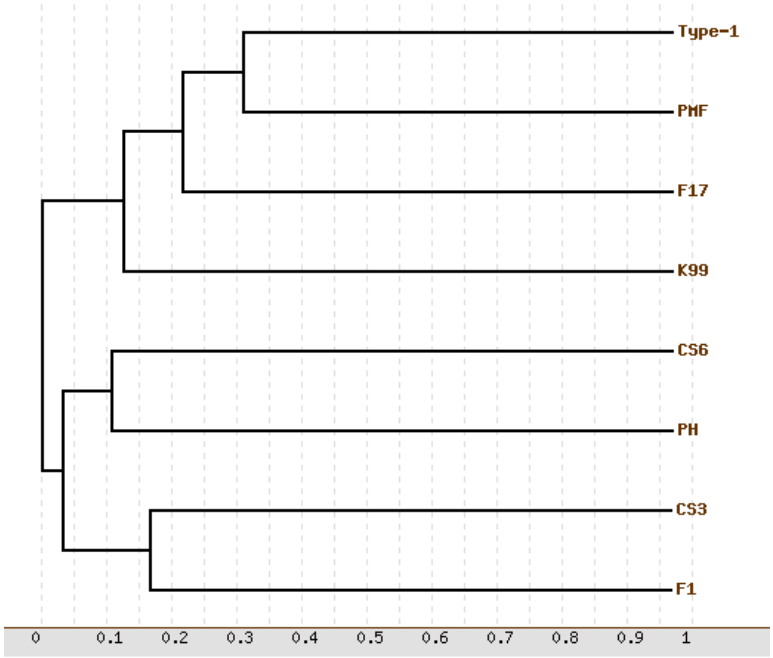

Figure 7. Phylogenetic relationship between some pilus family members. F1 is referred to F1 capsule antigen. The phylogenetic show the CS3 pili to be closest to the F1 capsule antigen

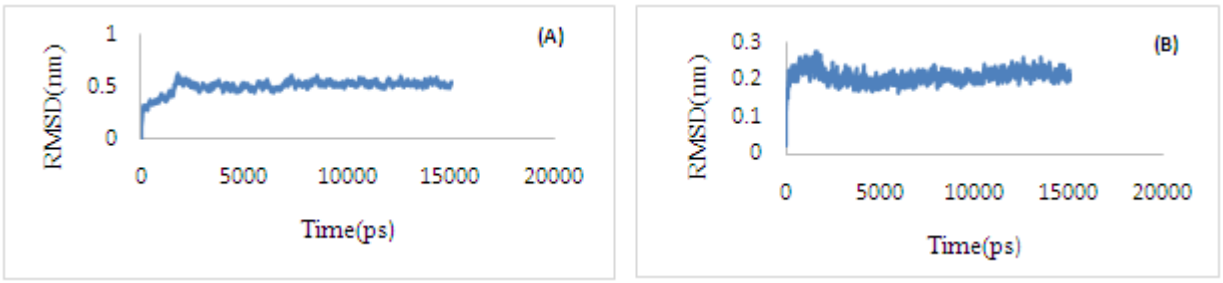

Figure 8. Root mean square deviation (RMSD) plot for complexes of CstH-CstH **(A)** and CstH-Cs3-1 **(B)** during molecular dynamic simulation

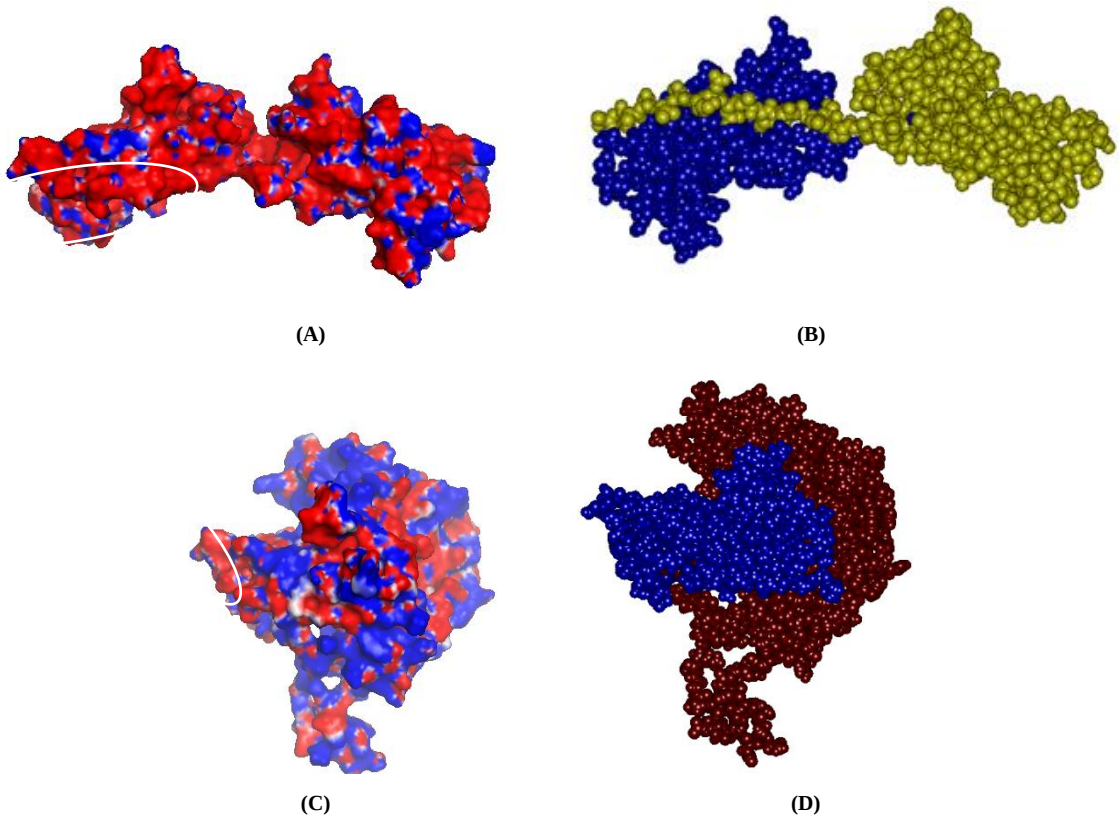

Figure 9. Electrostatic potential (ESP) surface for the CstH/CstH and CstH/CS3-1. Molecular surface representation of CstH/CstH Complex (blue, yellow) (B) and CstH/CS3-1 Complex (blue, brown) (D) and their respective EPS (A for CstH/CstH Complex and C for CstH/CS3-1 Complex) represented as blue (positive potential), red (negative potential) and white (hydrophobe). The encircled region represents their interface interacting regions. B and D images prepared with *Discovery Studio*, 2.5

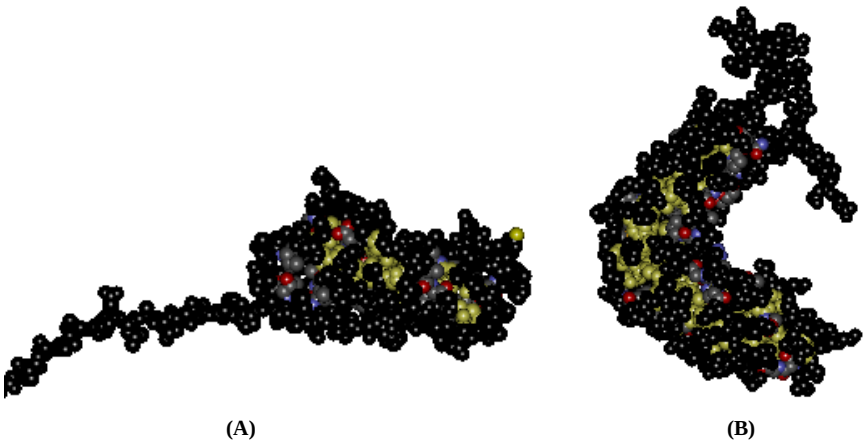

Figure 10. Solvent accessibility analyses of CstH and CS3-1 models. The *solvent accessible* surface and buried region were colored in black and yellow in CstH (A) and CS3-1(B) models. The images prepared with Discovery *Studio* 2.5 software

4. Conclusions

The understanding of protein structures provides useful information about the biochemical function and interaction properties of the protein and it is very important in medicine (such as, drug design), biotechnology (such as, novel enzymes design). The fact that experimentally methods such as Nuclear Magnetic Resonance (NMR) or X-ray crystallography which are used to determining protein structures are very expensive and time consuming. Therefore for minimizing the time and costs, bioinformatics methods are used for the prediction of the 3D structure of proteins. In this paper, we predicted the 3D structure and interface areas of CstH with a computational methods.

The CstH protein (major subunit of CS3 pili is a (relatively) small protein (146 amino acids) with no cysteine residues. These properties make CS3 fimbriae a good candidate for use as a carrier in the bacterial cell surface display.

Dimer complexes (CstH-CstH and CstH-CS3-1) revealed that homo and hetero-dimerization of CstH with CstH and CS3-1 mostly occupy its N-terminal and also C-terminal interacting sites, which are in agreement with fimbriae assembly mechanism [3, 4]. Additionally, PDBePISA and Swiss-Pdb Viewer EPS analysis of dimer complexes revealed details interaction sites of CstH model.

Our other finding in this study is that when there are low sequence identities between the targets and templates sequences, the best template can be identified with the fold recognition methods using multiple sequence alignment algorithms. The target and best template sequence alignment can be used for building of the model using Swiss - Model homology mode.

In summary, our study provides structural insights into the architecture of CstH protein, which can facilitate the identification of permissive foreign peptide insertion sites, *ligand-binding* site prediction and also rational engineering of the subunit for hybrid pili construction, in order to engineer tightening and loosening of the pili structure which is a great importance for many biotechnological applications.

REFERENCES

[1] Soto GE & Hultgren SJ, adhesins. Common themes and variations in architecture and assembly. *J Bacteriol.* (1999) 181:1059-71.

[2] Sauer FG, Pinkner JS, Waksman G & Hultgren SJ. Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation. *Cell.* (2002) 111:543-51.

[3] Nuccio SP & Baumler AJ. Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. *Microbiol Mol Biol Rev.* (2007) 71:551-75.

[4] Zavialov AV, Berglund J, Pudney AF, Fooks LJ, Ibrahim TM, MacIntyre S et al. Structure and biogenesis of the capsular F1 antigen from *Yersinia pestis*: preserved folding energy drives fiber formation. *Cell.* (2003) 113:587-96.

[5] Verger D, Bullitt E, Hultgren SJ & Waksman G. Crystal structure of the P pilus rod subunit PapA. *PLoS pathogens.* (2007) 3:73.

[6] Li YF, Poole S, Nishio K, Jang K, Rasulova F, McVeigh A et al. Structure of CFA/I fimbriae from enterotoxigenic *Escherichia coli*. *Proc Natl Acad Sci U S A.* (2009) 106:10793-8.

[7] Klemm P & Schembri MA. Fimbrial surface display systems in bacteria: from vaccines to random libraries. *Microbiology.* (2000) 12:3025-32.

[8] Dong Z, Zhang Z, Li S, Zhang P & Huang C. CS3 fimbriae of enterotoxigenic *Escherichia coli* as chimeric expression carriers of heterologous antigenic determinants. *Sci China C Life Sci.* (1999) 42:128-35.

[9] Gao RK, Zhang ZS, Li SQ & Huang CF, Construction of a novel display vector deriving from CS3 fimbriae of human enterotoxigenic *Escherichia coli*. *Yi Chuan Xue Bao.* (2001) 28:971-9.

[10] Saffar B, Yakhchali B & Arbabi M. Development of a bacterial surface display of hexahistidine peptide using CS3 pili for bioaccumulation of heavy metals. *Curr Microbiol.* (2007) 55:273-7.

[11] Boutet E, Lieberherr D, Tognolli M, Schneider M & Bairoch A. UniProtKB/Swiss-Prot. *Methods Mol Biol.* (2007) 406:89-112.

[12] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H et al. The Protein Data Bank. *Nucleic Acids Res.* (2000) 28:235-42.

[13] McGinnis S & Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* (2004) 32:W20-25.

[14] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* (1997) 25:3389-402.

[15] Eswar N, Eramian D, Webb B, Shen MY & Sali A. Protein structure modeling with Modeller. *Methods Mol Biol.* (2008) 426:145-59.

[16] Shen MY & Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* (2006) 15:2507-24.

[17] Laskoswki RA, MM, Moss DS & Thorton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst.* (1993) 26:283-91.

[18] Wiederstein M & Sippl MJ. Prosa II-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* (2007) 35:W407-10.

[19] McGuffin LJ & Jones DT. Improvement of the GenTHREADER method for genomic fold recognition. *Bioinformatics.* (2003) 19:874-81.

[20] Shi J, Blundell TL & Mizuguchi K. FUGUE: sequence-structure homology recognition using

environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol.* (2001) 310:243-57.

[21] Jaroszewski L, Rychlewski L, Li Z, Li W & Godzik A. FFAS03: a server for profile--profile sequence alignments. *Nucleic Acids Res.* (2005) 33:W284-8.

[22] Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol.* (1999) 287:797-815.

[23] Kelley LA & Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc.* (2009) 4:363-71.

[24] Arnold K, Bordoli L, Kopp J & Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics.* (2006) 22:195-201.

[25] Kiefer F, Arnold K, Kunzli M, Bordoli L & Schwede T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* (2009) 37: D387-92.

[26] Benkert P, Biasini M & Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics.* (2011) 27:343-50.

[27] Eisenberg D, Luthy R & Bowie JU. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.* (1997) 277:396-404.

[28] Geourjon C & Deleage G. SOPM: a self-optimized method for protein secondary structure prediction. *Protein Eng.* (1994) 7:157-64.

[29] Garnier J, Gibrat JF & Robson B. GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol.* (1996) 266:540-53.

[30] Rost B & Sander C. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins.* (1994) 19:55-72.

[31] Levin JM & Garnier J. Improvements in a secondary structure prediction method based on a search for local sequence homologies and its use as a model building tool. *Biochim Biophys Acta.* (1988) 955:283-95.

[32] Stivala A, Wybrow M, Wirth A, Whisstock JC & Stuckey PJ. Automatic generation of protein structure cartoons with Pro-origami. *Bioinformatics.* (2011) 27:3315-6.

[33] Krissinel E & Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol.* (2007) 372:774-97.

[34] Baker NA, Sept D, Joseph S, Holst MJ & McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci.* (2001) 98:10037-41.

[35] DeLano WL. The PyMOL molecular graphics system (2002). http://www.pymol.org.

[36] Su Y, Zhou A, Xia X, Li W & Sun Z. Quantitative prediction of protein-protein binding affinity with a potential of mean force considering volume correction. *Protein Scie.* (2009) 18:2550-8.

[37] Sauer FG, Barnhart M, Choudhury D, Knight SD, Waksman G & Hultgren SJ. Chaperone-assisted pilus assembly and bacterial attachment. *Curr Opin Struct Biol.* (2000) 10:548-56.

[38] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. (1994) 22:4673–80.

[39] Brodskii LI, Ivanov VV, Kalaidzidis Ia L, Leontovich AM, Nikolaev VK, Feranchuk SI et al. GeneBee-NET: An Internet based server for biopolymer structure analysis. *Biokhimiia.* (1995) 60:1221-30. Russian.

[40] Doong S H. Protein homology modeling with heuristic search for sequence alignment, In *40th Annual Hawaii International Conference on System Sciences* (HICSS'07) 2007, 128-128.

[41] Xiang Z, Advances in homology protein structure modeling, *Curr Protein Pept Sci*, 7 (2006) 217-27.

[42] Castrignano T, De Meo PD, Cozzetto D, Talamo IG & Tramontano A. The PMDB Protein Model Database. *Nucleic Acids Res.* (2006) 34:D306-9.

[43] Fiorucci S & Zacharias M. Prediction of protein-protein interaction sites using electrostatic desolvation profiles. *Biophys J.* (2010) 98:1921-30.

[44] Jones S & Thornton JM. Analysis of protein-protein interaction sites using surface patches. *J Mol Biol*. (1997) 272:121-32.

[45] McGuffin LJ, Smith RT, Bryson K, Sorensen SA & Jones DT. High throughput profile-profile based fold recognition for the entire human proteome. *BMC bioinformatics.* (2006) 7:288.

[46] Rayan A. New tips for structure prediction by comparative modeling. *Bioinformation.* (2009) 3:263-7.