Three dimensional modeling of C-terminal loop of CssA subunit in CS6 of Enterotoxigenic *Escherichia coli* and its interaction with the 70 KDa domain of Fibronectin

Raghunath Chatterjee[1, 3#], Abhisek Ghosal[2, 4#], Subrata Sabui[2], Nabendu Sekhar Chatterjee[2]*

[1]Biomedical Informatics Centers of ICMR, National Institute of Cholera & Enteric Diseases, P33 C.I.T. Road, Scheme XM, Beliaghata, Kolkata 700 010, India; [2]Division of Biochemistry, National Institute of Cholera & Enteric Diseases, Kolkata, India; [3]Present address: National Cancer Institute, NIH, Bethesda, MD USA 20852; [4]Present address: University of California, Irvine, VA Medical Center, 151, 5901 East 7th St, Long Beach, CA 90822; Nabendu Sekhar Chatterjee - Email: chatterjeens@icmr.org.in; [#]Authors contributed equally; *Corresponding author

Abstract:
Colonization factor CS6 of enterotoxigenic *Escherichia coli* (ETEC) helps to establish the adherence of CS6-expressing ETEC in the intestinal wall. CS6 is composed of two structural subunits, known as CssA and CssB. During CS6-expressing ETEC adherence in intestinal wall, 15 amino acid residues containing C-terminal region of CssA subunit, help to bind with N-terminal 70kDa domain of fibronectin (Fn). In this study, we have predicted a theoretical structural model for C-terminal domain of CssA by homology modelling using protein data bank (PDB) file, 1NTY-A as template (66.67% sequence identity) in Discovery Studio. The structural model of N-terminal region of Fn was also determined by homology modelling using PDB files 1FBR and 1E88 as templates. The structure of the model was also validated by Ramachandran plot. The energy minimization for Fn was performed in standard dynamic cascade using Steepest Descent algorithm followed by Adopted Basis NR algorithm in Discovery studio. The docking model between C-terminal domain and fibronectin were generated by using ClusPro algorithm. This docking study would be help for better understanding how CS6 interacts with fibronectin of intestinal extracellular matrix in the host during infection, and would be of great help towards subunit vaccine generation.

Keywords: Colonization factor CS6, ETEC, fibronectin, homology modelling, Ramachandran plot, docking model.

Background:
Enterotoxigenic *Escherichia coli* (ETEC), is one of the major causes of infantile diarrhoea and traveller's diarrhoea in different developing countries **[1, 2, 3]**. ETEC alone causes around 40-70% of the diarrhoeal incidence globally **[4]**. The essential step in pathogenesis is mediated by initial adherence followed by colonization in the intestine. Till date more than 25 proteinous appendages have been identified, collectively called colonization factor antigens which had been shown to be involved in adhesion in vivo and in vitro **[5]**. Among them, CS6 is one of the prevalent colonization factors worldwide **[1, 6]**. Genomic studies have suggested that CS6 is mainly composed of two equally expressed subunits, CssA and CssB in a stoichiometry of 1:1 **[7, 8]**. They are very tightly associated and cannot be separated from each other easily **[7]**. Functionally, CssA subunit was shown to bind to fibronectin (Fn-an extra cellular matrix protein of intestinal epithelial cells) by us (7) earlier whereas CssB has been shown to interact with glycosphingolipid **[9]**. Studies further showed that the C-terminal of CssA subunit of CS6 (from 112 to 126 amino acids) binds Fn in vitro **[7]** but residue interaction was not known. We reported purification of CS6 to a functionally active form while its organization (in

terms of both subunits) is not known. No template has been reported with high sequential homology till date to explore its organizational details. Lack of any three dimensional structures for CS6 thus makes it difficult to complement the functional details of the CS6-Fn interactions with structural insights. In this study, we have reported a model for CssA and its interaction with Fn by a systematic approach using different bioinformatics softwares. The prediction will help to explore the functional aspects of CS6 in greater detail in future.

Methodology:
The amino acid sequence of the CssA subunit of CS6 was extracted from ETEC 4266 strain (GeneBank accession number EF451566). The secondary structure of CssA was determined by Discovery Studio (Acclereys, USA). We choose the C-terminal domain of CssA subunit of CS6 protein because our previous study experimentally proves that 15 amino acid (NYTSGDKEIPPGIYN) containing C-terminal region of CssA bind to 70 kDa N-terminal fibronectin domain **[7]**. The probable structure of the binding region of CssA was assigned by homology modeling using the crystal structure corresponding to protein data bank (PDB) (http://www.rcsb.org/pdb/) file

BIOINFORMATION

open access

1NTY-A as template (66.67 % sequence identity), following a standard method in Discovery Studio (Acclereys, USA). For loop refinement and energy minimization we used the Disocovery studio module 'smart minimizer' with 0.001 minimizing RMS gradient and 2000 minimizing steps. After loop refinement and energy minimization the predicted structure was validated by examining the root mean square deviation (RMSD) value by Ramachandran plot. The probable structure of Fn was also determined by homology modeling. Although earlier ELISA data showed that CS6 binds to both proteolytic domains Fn and that binding was enhanced in presence of both domains [7]. We tried to determine the probable interacting region of CS6 with the N-terminal region of Fn by computational modeling. Structure of N-terminal region of Fn was determined by homology modeling using PDB (http://www.rcsb.org/pdb/) files 1FBR and 1E88 as templates. Energy minimization for Fn was performed in standard dynamic cascade using Steepest Descent algorithm followed by Adopted Basis NR algorithm in Discovery studio. For energy minimization, 5000 steps were included with an RMS gradient of 0.01and an Adopted Basis NR algorithm minimization was performed to an RMS gradient of 0.0001. After inclusion of heating-cooling and equilibration steps the final most stable structure was obtained based on the DOPE score. The predicted structure of Fn was also validated by Ramachondran plot. The docking model between the C-terminal domain of CssA and N-terminal domain of Fn homology models were extracted using ClusPro software [10]. The energetically most stable model that fits with experimental observations in ELISA data from our previous work was chosen as the final model. Potential energy for interaction of the C-terminal loop with different motifs of Fn was determined by Discovery studio.

RTEIATKNFP*VSTTI*SKSFFAPEPQIQPSFGKNVGKEG*GLLFSVSLT*

VPENVS*QVTVYPVY*DEDYGLG*RLV*NTADDSQSIIYQIVDDKG*RKML*

*KD*HGAEVTPNQQ*ITFRAL*NYTSGDKEIPPGIYNDQVMVGYYVN
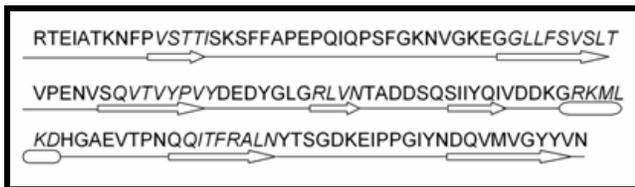
Figure 1: Schematic diagram of the secondary structure of CssA of CS6. The secondary structure was predicted theoretically by Discovery studio from the amino acid sequence derived from DNA sequence. The barrel and arrow denote the α-helix and β-sheet respectively.
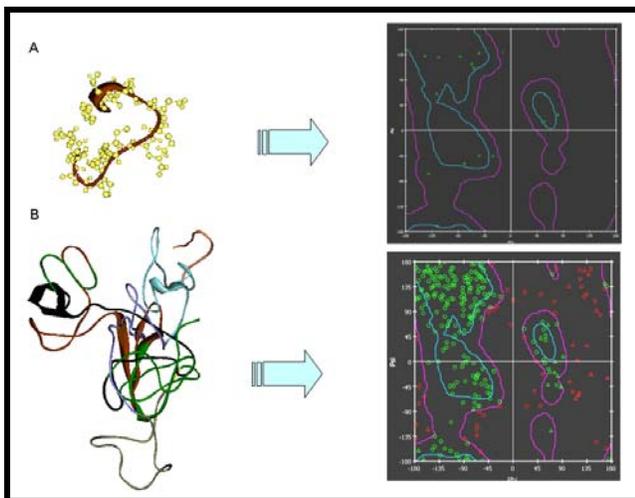

Figure 2: A. Homology modeling of C terminal region of CssA. For homology modeling DH/PH domain of Trio (PDB: 1NTY-A) (66.67% sequence identity) was used. After loop refinement, energy minimization was performed using Smart Minimizer. The root mean square deviation (RMSD) from the template was 1.27 Å. The arrow indicates the ramachandran plot of the structure showing that more than 98% of residues are in permissible region. **B.** Homology modeling of N terminal region of Fn. 1FBR and 1E88 were used as template. The energy minimization for Fn was performed in standard dynamic cascade using Steepest Descent algorithm followed by Adopted Basis NR algorithm in Discovery studio. The RMSD of the derived structure was 1.67 Å with respect to the template. The arrow indicates the ramachandran plot of the structure showing more than 95% of residues are in the energetically permissible region of the Ramachandran diagram.

Result and Discussion:
Homology modeling of C-terminal domain of CssA:
In order to determine the probable model for the interaction of CS6 with Fn, homology modeling for both the N-terminal region of Fn and C-terminal region of CssA were performed. For homology modelling DH/PH domain of Trio (PDB: 1NTY-A) [11] was used as a template for the C-terminal region of CssA. The theoretical prediction of secondary structure suggested that the proposed binding region of CssA (NYTSGDKEIPPGIYN) assumes a loop in between two β sheets (**Figure 1**). Three dimensionally it was confirmed by homology modeling (**Figure 2A**). The RMSD of the predicted structure of the loop was 1.27 Å from the template. More than 98% of amino acids fall in the allowed region of Ramachandran plot (**Figure 2B**). Following denotes the energy content of the final structure of the C-terminal loop. Initial Potential Energy = -701.66 kcal/mol, Potential Energy of the final structure = -2517.75 kcal/mol, Van der Waals Energy = -205.03 kcal/mol, Electrostatic Energy = -2630.94 kcal/mol.
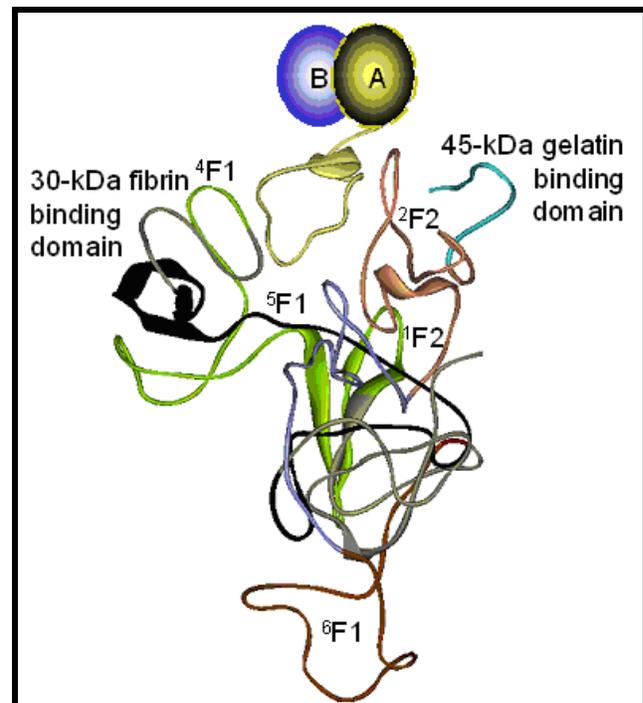

Figure 3: Representative model of interaction of the Carboxy-terminal loop of CssA and Fn. The template used for 30- kDa N terminal domain (PDB: 1FBR), which is made of ⁴F1 and ⁵F1. The template used for 45- kDa gelatin binding domain was made of ⁶F1, ¹F2 and ²F2. The groove region of Fn is between ⁵F1and ⁶F1 that interacts with the C-terminal loop of CssA.

Homology modeling of Fibronectin:
Since from our previous data we found that binding is mediated by both 30 kDa and 45 kDa regions, we used the PDB files corresponding to the common region of the domains to determine an approximate model for the interaction. Using 1FBR and 1E88 as template we determined the three dimensional structure of the ⁴F1, ⁵F1, ⁶F1, ¹F2, ²F2 motifs of Fn by homology modelling. The RMSD value of the region was 1.66 Å with respect to the template and more than 95% of the amino acids fall in the allowed region of the Ramachandran plot. After all the energy minimization steps the final energy content of Fn was as follows, Total Energy = -12078.49 kcal/mol, Potential Energy = -15326.90 kcal/mol, Van der Waals Energy = -1224.23kcal/mol, Electrostatic Energy = -19266.91 kcal/mol.

Docking of CssA and Fn in search of probable binding region:
In protein–protein docking by ClusPro [10] the probable binding region of the peptide was determined (**Figure 3**). The interaction energy was calculated using Discovery studio. The interaction energy of the peptide with the ⁵F1 motif of 30-kDa fibrin binding domain was -71.220 (kcal/mol). Binding also occurs to the connecting region of 30 and 45 kDa domains (interaction energy = -46.106 kcal/mol) (**Table 1 see Supplementary material**).

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 6(8): 307-310 (2011) 308 © 2011 Biomedical Informatics

BIOINFORMATION

Conclusion:
Model suggested C-terminal domain of CssA fits into the groove formed in 70-kDa domain of Fn. This is the first report of Fn binding bacterial protein of gram-negative bacteria. CssA did not show any significant sequence homology to the C-terminal binding repeats of the Fn binding proteins MSCRAMM (Microbial Surface Components Recognizing Adhesive Matrix Molecules), of many gram-positive species [12]. Unlike MSCRAMM, CssA is predicted to interact with the [5]F1 and the loop joining 30-kDa and 45-kDa domain strongly followed by [1]F2 domain of 45 kDa domain. Based on the computational modelling and binding studies, we propose that the predicted loop docks into the grove that is formed in between the fibrin binding (30-kDa) and gelatin binding (45-kDa) domains Fn. We also search CssA subunit protein sequence in different CS6-expressing ETEC strains which indicates most of the amino acid are conserved of CssA loop region (data not shown). However, this predicted model helps to improve visualization of the Fn binding with CS6 and may also provide a valuable information for developing a subunit vaccine against CS6-expressing ETEC, causing diarrhoeal disease.

References:
[1] Gaastra W & Svennerholm AM. *Trends Microbiol.* 1996 **4**: 444 [PMID: 8950814]
[2] Black RE. *Vaccine* 1993 **11**: 100 [PMID: 8438608]
[3] Sack RB. *Annu Rev Microbiol.* 1975 **29**: 333 [PMID: 1101807]
[4] http://www.who .int/whr/2004
[5] Qadri F *et al. Clin Microbiol Rev.* 2005 **18**: 465 [PMID: 16020685]
[6] Ghosal A *et al. J Clin Microbiol.* 2007 **45**: 3068 [PMID: 17596357]
[7] Ghosal A *et al. Infect Immun.* 2009 **77**: 2125 [PMID: 19237522]
[8] Wolf MK *et al. Clin Microbiol Rev.* 1997 **10**: 569 [PMID: 9336662]
[9] Jansson L *et al. PloS One.* 2009 **4**: e4487 [PMID: 19242561]
[10] Comeau SR *et al. Nucleic Acids Res.* 2004 **32**: W96 [PMID: 15215358]
[11] Skowronek KR *et al. J Biol Chem.* 2004 **279**: 37895 [PMID: 15199069]
[12] Perkins S *et al. J Biol Chem.* 2001 **276**: 44721 [PMID: 11568183]

Edited by P Kangueane
Citation: Chatterjee *et al.* Bioinformation 6(8): 307-310 (2011)

Supplementary material:

Table 1: Predicted binding energy for interaction of the peptide NYTSGDKEIPPGIYN of CssA to motifs of fibronectin.

Total interaction energy (in kcal/mol)						
Whole Fn	[4]F1	[5]F1	[6]F2	[1]F2	[2]F2	Fn groove
-110.989	-2.425	-71.220	0.00	13.299	-20.069	-46.106