

In silico Interactive Study of Periplasmic and Outer Membrane Proteins of Type IV Secreting System (T4ss) in *Helicobacter pylori*

Megha Vaidya*, PV Virparia and Hetalkumar Panchal

G. H. Patel P. G. Department of Computer Science & Technology (GDCST) Sardar Patel University, India

Abstract

Helicobacter pylori is gram-negative bacterial pathogen, associated with chronic gastritis and other severe gastro duodenal diseases such as peptic and gastric ulcers, gastric cancer, and mucosa-associated lymphoid tissue (MALT) lymphoma. One of the intriguing virulence factors of *H. pylori* is cag Pathogenicity Island (cagPAI) which codes for Type IV secreting system. Type IV secreting system is found in many strains of *H. pylori*. Present work focuses on protein interaction studies to clarify and confirm the interactions and assembly of proteins present in outer membrane and pilus structure of t4ss. During the interaction studies of t4ss we have focused on only CagC, Cag3 and CagM interactions based on stereochemical properties. It is found that the pattern of pilus formation follows the specific order of assembly and the order of assembly is CagC-C dimer formation followed by tetramer formation to form stable pilus assembly of t4ss. The order of interaction at the outer membrane is CagM interacts with CagC-C dimer and then with Cag3 to form sub-assembly of t4ss. This findings of structural interaction can be used to prevent onset of diseases caused by *H. pylori* by preventing the formation of t4ss, pilus assembly and ultimately the translocation of onco protein CagA in host organism.

Keywords: *Helicobacter pylori*; Gastroduodenal; Cag protein; Molecular docking

Introduction

Helicobacter pylori commonly known as *H. pylori* is gram-negative bacterial pathogen, associated with chronic gastritis and other severe gastroduodenal diseases such as peptic and gastric ulcers, gastric cancer, and mucosa-associated lymphoid tissue (MALT) lymphoma [1-3]. Almost 50% of the world's population is *H. pylori* infected and in some countries 100% population is found to be infected [4]. *Helicobacter pylori* are a highly recombining pathogen [5] and there are therapies present to control the infection. One of such therapy is triple therapy, including two antibiotics, amoxicillin and clarithromycin, and a proton pump inhibitor. However, this treatment may fail for several reasons and the main reason for failure was found to be *H. pylori* resistance to the antibiotics like clarithromycin, metronidazole, tetracycline, fluoroquinolones, and rifamycins [6]. Thus, it can be considered that *H. Pylori* resistance to the antibiotics has become an emerging issue and has rise a need for better understanding the pathogenic mechanism, identifying novel targets and medication options.

One of the intriguing virulence factors of *H. pylori* is cag pathogenicity island (cagPAI) [7-9]. Type I strains of *H. pylori* [10] that possess a functional cagPAI are particularly frequently associated with severe, gastric atrophy and cancer [9,11-13]. The Pathogenic island of *H. pylori* cagPAI is, 37 kb long genomic fragment containing 28 genes [3], encodes multiple structural components of a bacterial type IV secretion system (t4ss) as well as the 128 kDa effector protein, CagA [13]. CagA has been designated a bacterial onco protein [14] profoundly altering host cellular functions [15,16]. After *H. pylori* has adhered to a host cell, the Cag t4ss translocates CagA in to that cell [17], which suggest that t4ss is a mainly important component indirectly responsible for pathogenic nature of *H. pylori*.

Type IV secreting system is found in many strains of *H. pylori* and according to an important finding the cag-T4SS apparently does not inject its effector protein CagA randomly into target cells, but uses the $\alpha 5\beta 1$ integrin as a cellular receptor for the pilus-associated

adhesion CagL [18]. However, the mechanism of translocation and the requirements of the host cell in that event are not well understood. The T4SS consists of inner- and outer membrane-spanning Cag protein complexes and a surface-located pilus [19]. Till date many molecular studies are carried out to identify the internal arrangements of t4ss and its interaction with host receptors [20,21]. However the detail study of interacting cag Proteins to build outer membrane structure of t4ss is not yet available. In the present study we have focused on the characterization, structural analysis and interactive study of outer membrane and pilus proteins of *H. pylori* t4ss to get a better understanding of their mechanism to form t4ss; and to identify target proteins if possible, which could inhibit either the t4ss formation or inhibit CagA translocation.

Materials and Methods

Protein sequence retrieval and analysis

Protein sequences of periplasmic and outer membrane proteins namely Cag3, CagC and CagM for *H. pylori* strain J99 were retrieved from NCBI (www.ncbi.nlm.nih.gov) with Accession number NP_223189.1, NP_223212.1 and NP_223203.1 respectively.

Physicochemical properties of protein sequences

Physicochemical properties of proteins were calculated using a ProPhyC. ProPhyC is standalone software developed at Bioinformatics research facility of our institute and is supported by Perl scripting

*Corresponding author: Megha Vaidya, G. H. Patel P. G. Department of Computer Science & Technology (GDCST) Sardar Patel University, Vallabh Vidyanagar – 388120, Gujarat, India, Tel: 91-2692-230389; E-mail: mbv26887@gmail.com

Received March 05, 2014; Accepted March 27, 2014; Published March 31, 2014

Citation: Vaidya M, Virparia PV, Panchal H (2014) *In silico* Interactive Study of Periplasmic and Outer Membrane Proteins of Type IV Secreting System (T4ss) in *Helicobacter pylori*. J Cancer Sci Ther 6: 105-109. doi:10.4172/1948-5956.1000257

Copyright: © 2014 Vaidya M, 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.

language and has the capacity to calculate several physicochemical properties of multiple proteins in a single run [22].

***In silico* protein modeling**

An *In silico* comparative modelling of protein Cag3 was done using remote homology modeling based web server CPHmodels [23]. Protein CagC was modeled using MODWEB server (<https://modbase.compbio.ucsf.edu/modweb/help.cgi?type=help>). MODWEB is a web server for automated comparative protein structure modeling that relies on MODPIPE for its functionality. ModPipe is a completely automated software pipeline that can calculate protein structure models for a large number of sequences with almost no manual intervention. MODWEB server is also present at The Protein Model Portal (PMP) as an option of interactive modeling server named Modweb server and here this server is used via PMP. Modeling of protein CagM was carried out by the online server BHAGEERATH: An Energy Based automated Protein Structure Prediction Server [24]. The modeled protein was visualized by PyMOL: An open-source Molecular Graphics tool [25].

Model refinement, validation and submission of modeled structures to PMDB

The modeled proteins were validated for the parameters included the covalent bond distances and angles, stereochemical validation and atom nomenclature using PROCHECK [26]. All the three modeled Proteins, after CASP test were deposited in the Protein Model Data Base [27].

Interactive studies of Cag3, CagC and CagM Proteins

Initially the protein interactions were studied based on protein structure complementarity using the online server Patch Dock [28]. Patch Dock is based on a rigid docking algorithm which allows protein surface flexibility for docking fit. In the beginning, individual interaction studies were carried out to predict correct interacting dimer. The dimers were then applied to further interactive study with respective third protein partner. Altogether sixteen different protein interactions were carried and analyzed. The study was based on structural complementarity only but the interactions need to be authenticated by analyzing the binding energy, hydrogen bonding and van der Waals interactions. FireDock: A web server for fast interaction refinement in molecular docking [29] was used for analyzing bonding interactions and binding energies for all sixteen interacting protein complexes.

Results

The protein sequence analysis reveals that Cag3 is a stable protein consisting total 481 amino acids and having a molecular weight equal to 55232.4 Dalt. The protein Cag3 is hydrophilic and has total 69 negatively charged residues and total 85 positively charged residues. Analysis of subcellular localization clarified it to be a membrane protein. CagC is a 115 amino acid, stable protein having a molecular weight equal to 11955.1 Dalt. The protein CagC is hydrophobic and has total 6 negatively charged residues and total 10 positively charged residues. Lastly CagM is a 376 amino acid, stable protein having a molecular

weight equal to 43726.0 Dalt. The protein CagM is hydrophilic and has total 52 negatively charged residues and total 64 positively charged residues.

The protein model of Cag3 modeled using webserver CPHmodels shown 98.5% of the amino acid residues in favor and allowed regions of the Ramachandran plot and 1.5% of the amino acid residues in disallowed region. All the amino acid residues had main chain and side chain parameters within the range, with just one residue of main residue in worst region. The stereochemical assessments show that bond angle of more than 94% residues and bond length of 100% residues within the acceptable range. The protein model of CagC modeled using Modweb interactive modeling server shown 100% of the amino acid residues in favor and allowed regions of the Ramachandran plot. All the amino acid residues had main chain and side chain parameters within the range, with none of the residue inside or in worst region. The stereochemical assessments show that the bond angle of more than 92% residues and bond length of more than 99% residues within the acceptable range. According to validation of protein model of CagM shown 99% of the amino acid residues in favor and allowed regions of the Ramachandran plot. All the amino acid residues had main chain and side chain parameters within the range, with none of the residue inside or in worst region. The stereochemical assessments show that the bond angle of more than 98% residues and bond length of more than 89% residues within the acceptable range. Validation of all three protein models proved that all models are within the acceptable range and thus, the protein models were successfully submitted to a protein model database. The universally accessible protein model database PMID for Cag3, CagC and CagM are PM0078507, PM0078508 and PM0078316, respectively.

To identify the order of assembly and interactions of typeIV secreting system proteins, initially the individual protein interactive studies for the three proteins were carried out. Protein Cag3 was interacted with CagM and CagC, respectively; followed by the protein interaction study between CagC and CagM. The details of interaction are as below

Protein Cag3 is massive in compare to CagM and CagC so for interaction between Cag3, CagM and CagC; the larger protein Cag3 was taken as receptor and other interacting partner was taken as ligand. Likewise, CagM is bigger protein than CagC, thus CagM was taken as receptor and CagC as ligand. As per results of interactive study represented in Table 1, the interaction between CagM-CagC and Cag3-CagC shown lowest binding energy required for interaction having significant contribution of van der Walls attraction and hydrogen bond to the interactions. Thus, at the end of protein interaction study, two possible dimers CagM-CagC or Cag3-CagC seems to possibly exist.

According to the research in the past [30], the proposed hypothetical structure of the T4SS shows the pilus structure made of only single protein B2 coded by gene named VirB2 or HP0546. CagC is a synonym for B2 protein so interaction studies for two CagC monomers were performed to identify whether the CagC monomers interact to form pilus assembly or CagC-CagC dimers are formed and the dimers then

| Receptor | Ligand | Global energy (kcal/Mol) | van der Walls force –Attraction (Kcal/Mol) | Hydrogen bond (Kcal/Mol) |
|----------|--------|--------------------------|--|--------------------------|
| Cag3 | CagM | -30.29 | -45.65 | -3.06 |
| Cag3 | CagC | -57.11 | -67.8 | -5.52 |
| CagM | CagC | -58.84 | -50.09 | -6.6 |

Table 1: Individual protein interaction studies for all three proteins, to predict probable dimers.

directly interact to form pilus structure. The results of CagC-CagC interactions (Table 2) initially showed acceptable global binding energy for two CagC monomers forming the dimer. Further the dimer interaction for CagC show significant increase in binding energy and interactive forces energy. The CagC monomer interactions formed CagC dimer with a binding energy around -45 and hydrogen bond interaction contributing -0.24 Kcal/Mol to the binding energy but interactions of the CagC dimer to form CagC quaternary complex (Figure 1) shown an elevated increase in binding energy and hydrogen bond contribution rise to double around -120 and -2, respectively compare to dimer formations.

The quaternary complex formed by the CagC dimer interaction with another CagC dimer, was studied in detail to identify the interacting amino acids and their possible bond formations (Table 3). All the interacting atoms of interacting residues show the bond-distance within the range or close to the standard atom bonding range (Figure 2). Thus from the result of CagC-CagC interaction, it is convincingly evident that to form pilus assembly, CagC dimer interaction with another CagC dimer is more preferable than monomers interacts.

According to our former reported interacting results two possible dimers CagM-CagC or Cag3-CagC seems to possibly exist. But now it was clear that CagC interacts in dimer form so again interaction study was carried out taking CagM and Cag3 as receptors and taking CagC dimer as ligand to identify whether CagM is important in initial binding or Cag3 (Table 4).

According to comparison of results from Table 1 and 4, the interaction of receptor CagM with both the ligands CagC and CagC-C-dimer is consistent having the binding energy above -50 in both the case. While as in case of receptor Cag3 the binding energy obtained (-57) for Cag3-CagC interaction has significantly raise to positive value 0.45 for Cag3 and CagC-C-dimer interaction. This fluctuation of binding energy to positive value represents that the complex cannot be stable complex (Figure 3).

Similar to atomic interactive study of CagC-C-dimer, the detailed interactive study for CagM and CagC-C-dimer was carried out. The detail of interacting amino acids and their atom involved in bond forming are represented in Table 5 with bond length pictorially represented in Figure 4.

Discussion

So far from the analysis the order of interaction seems to be clear now. CagM initially interacts with CagC-C Dimer (Figure 3) with lowest binding energy followed by Cag3. Along with this it can be said that the interaction between CagC-C dimers are preferable over the CagC monomer interaction to forms Stable pilus assembly. With this it is also evident from Table 1 that all the monomers can interact to form dimers, independent of each other's presence. The above protein interaction studies clarify and confirms the interactions and assembly of proteins of outer membranes and pilus structure of t4ss. This structural interaction can be used to further target any one of these three easy to access proteins (as they are membrane protein n pilus

| Receptor | Ligand | Global energy (Kcal/Mol) | van der Walls force –Attraction (Kcal/Mol) | Hydrogen bond (Kcal/Mol) |
|----------|--------|--------------------------|--|--------------------------|
| CagC | CagC | -45.55 | -32.43 | -0.24 |
| CagC-C | CagC | -35.92 | -40.28 | -2.87 |
| CagC-C | CagC-C | -120.85 | -32.66 | -2.96 |

Table 2: Protein interaction study for CagC protein to predict the probable order of pilus assembly.

| Dimer-1 (Red) | Dimer-2 (Yellow) | Bond forming atoms of residues of Dimer-1 | Bond forming atoms of residues of Dimer-2 | Bond-Distance (Å) |
|---------------|------------------|---|---|-------------------|
| Ileu 48 | Ileu45 | CD1 | CD1 | 2.2 |
| Leu51 | Ileu45 | CD1 | CG2 | 1.3 |
| Val54 | Met90 | CG1 | CE | 2.9 |
| Tyr84 | Ala92 | OH | CB | 3.8 |
| Tyr84 | Val93 | O | CG1 | 3.9 |
| Gln46 | Ileu97 | NE2 | CD1 | 2.2 |
| Ileu 97 | Leu100 | CG2 | CD1 | 1.5 |
| Ileu91 | Gln46 | O | NE2 | 4.8 |

Table 3: Detail of bond formation by interacting atoms and residue of the CagC dimers with another CagC dimers for formation of quaternary complex.

| Receptor | Ligand | global energy (kcal/mol) | Vander walls force –Attraction (kcal/mol) | Hydrogen bond (kcal/mol) |
|----------|--------------|--------------------------|---|--------------------------|
| Cag3 | CagC-C Dimer | 0.45 | -4.16 | 0.0 |
| CagM | CagC-C Dimer | -53.00 | -32.49 | -1.52 |

Table 4: Protein interaction details for receptors Cag3 and CagM with CagC-C dimer as ligand.

| CagM Green) | CagC-C dimer- (Blue) | Bond forming atoms of residues of Dimer-1 | Bond forming atoms of residues of Dimer-2 | Bond-Distance (Å) |
|-------------|----------------------|---|---|-------------------|
| Lys198 | Thr80 | N2 | OG1 | 2.2 |
| Asn206 | Ileu74 | OD1 | CD1 | 1.3 |
| Ser196 | Glu110 | OG | OE1 | 2.9 |
| Asn207 | Gln78 | OD1 | OE1 | 3.8 |
| Phe255 | Asp73 | O | OD2 | 3.9 |
| Lys371 | Lys103 | N2 | N2 | 2.2 |
| Ileu6 | Gln78 | CD1 | O | 1.5 |
| Ala3 | Gln78 | CB | OE1 | 4.8 |

Table 5: Detail of bond formation by interacting atoms and residue of the trimer formed by CagM and CagC-C dimers interaction.

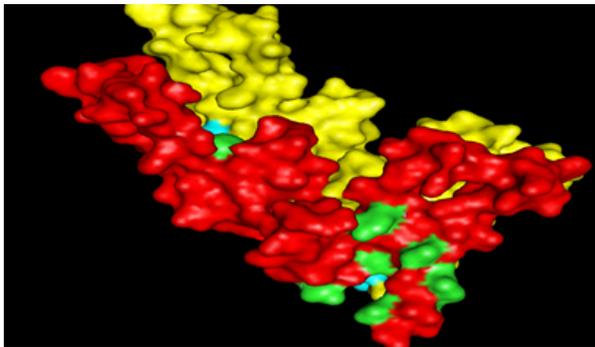


Figure 1: Represents structure complementary base interaction of CagC dimer (red color) with another CagC dimer (yellow) to form quaternary complex, with interacting regions represented with green and light blue patches.

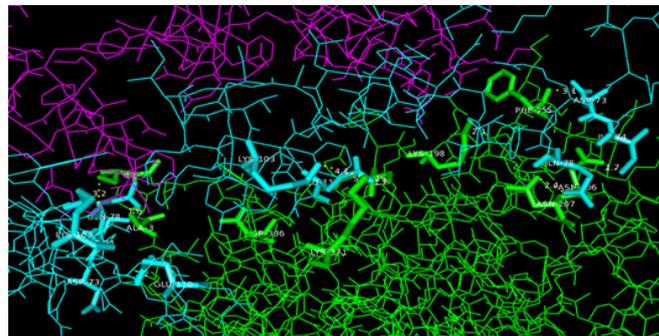


Figure 4: Represents interaction of CagC-C Dimer (both polypeptides colored blue and pink, respectively) with CagM (Green), with interacting residue represented with green and light blue while distance represented by dotted lines.

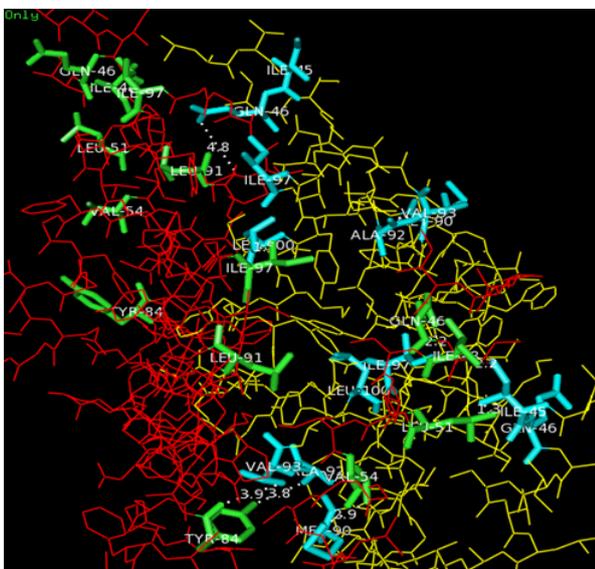


Figure 2: Represents interaction of CagC dimer (red color) with another CagC dimer (yellow), with interacting residue represented with green and light blue while distance represented by dotted lines.

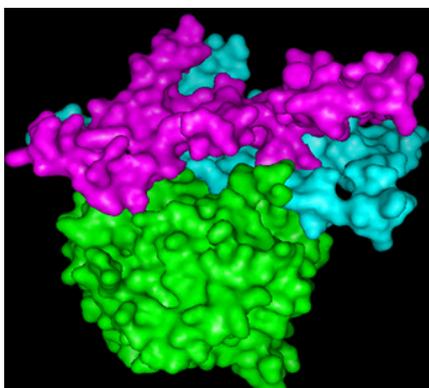


Figure 3: Represents structure complementarity base interaction of CagC-C Dimer (both polypeptides colored blue and pink, respectively) with CagM (Green); with interacting regions clearly visible in contact with each other.

protein assembly is extending external to bacterial cell) to hinder the formation of t4ss. To be specific the formation of CagC-C dimer could be targeted. If the formation of this CagC-C dimer could be prevented by competitive molecular docking, using either natural or synthetic compound as a competitor compound then the pilus assembly could be hindered and in turn the translocation of onco protein CagA may be prevented. CagC is protein coded by almost all the Type I strains of *H. pylori*. The hypothesis of disturbing pilus assembly of t4ss could be primarily confirmed by knocking out *cagC* gene. The pharmacophore analysis of three interacting proteins presented in this paper will help better understanding the assembly mechanism of *H. pylori* t4ss, as its structure is not clear yet.

Acknowledgement

We are heartily thankful to Prof. (Dr.) P.V. Virparia, Director, GDCST, S.P. University, V.V. Nagar, for providing us facilities for the research work. We are also thankful to DST-PURSE program and Center for Interdisciplinary Studies in Science and Technology (CISST), S.P. University, V.V. Nagar, Gujarat (India) for providing financial assistance in the form of fellowship.

References

- Cover TL, Blaser MJ (1992) *Helicobacter pylori* and gastroduodenal disease. *Annu Rev Med* 43: 135-145.
- Hunt RH (1996) The role of *Helicobacter pylori* in pathogenesis: the spectrum of clinical outcomes. *Scand J Gastroenterol Suppl* 220: 3-9.
- Labigne A, de Reuse H (1996) Determinants of *Helicobacter pylori* pathogenicity. *Infect Agents Dis* 5: 191-202.
- Matysiak-Budnik T, Mégraud F (1997) Epidemiology of *Helicobacter pylori* infection with special reference to professional risk. *J Physiol Pharmacol* 48 Suppl 4: 3-17.
- Ahmed N, Tenguria S, Nandanwar N (2009) *Helicobacter pylori*--a seasoned pathogen by any other name. *Gut Pathog* 1: 24.
- Mégraud F (2004) *H. pylori* antibiotic resistance: prevalence, importance, and advances in testing. *Gut* 53: 1374-1384.
- Censini S, Lange C, Xiang Z, Crabtree JE, Ghiara P, et al. (1996) *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc Natl Acad Sci U S A* 93: 14648-14653.
- Fischer W, Püls J, Buhrdorf R, Gebert B, Odenbreit S, et al. (2001) Systematic mutagenesis of the *Helicobacter pylori* *cag* pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. *Mol Microbiol* 42: 1337-1348.
- Wiedemann T, Loell E, Mueller S, Stoeckelhuber M, Stolte M, et al. (2009) *Helicobacter pylori* *cag*-Pathogenicity island-dependent early immunological response triggers later precancerous gastric changes in Mongolian gerbils. *PLoS ONE* 4: e4754.

10. Helena Enroth, Wolfgang Kraaz, Lars Engstrand, Olof NyreÅ'n, Tom Rohan (2000) Helicobacter pylori Strain Types and Risk of Gastric Cancer: A Case-Control Study. Cancer Epidemiol Biomarkers Prev 9: 981-985.
11. Figueiredo C, Machado JC, Pharoah P, Seruca R, Sousa S, et al. (2002) Helicobacter pylori and interleukin 1 genotyping: an opportunity to identify high-risk individuals for gastric carcinoma. J Natl Cancer Inst 94: 1680-1687.
12. Amieva MR, El-Omar EM (2008) Host-bacterial interactions in Helicobacter pylori infection. Gastroenterology 134: 306-323.
13. Hatakeyama M (2008) SagA of CagA in Helicobacter pylori pathogenesis. Curr Opin Microbiol 11: 30-37.
14. Ohnishi N, Yuasa H, Tanaka S, Sawa H, Miura M, et al. (2008) Transgenic expression of Helicobacter pylori CagA induces gastrointestinal and hematopoietic neoplasms in mouse. Proc Natl Acad Sci U S A 105: 1003-1008.
15. Hatakeyama M (2009) Helicobacter pylori and gastric carcinogenesis. J Gastroenterol 44: 239-248.
16. Lu HS, Saito Y, Umeda M, Murata-Kamiya N, Zhang HM, et al. (2008) Structural and functional diversity in the PAR1b/MARK2-binding region of Helicobacter pylori CagA. Cancer Sci 99: 2004-2011.
17. Olbermann P, Josenhans C, Moodley Y, Uhr M, Stamer C, et al. (2010) A Global Overview of the Genetic and Functional Diversity in the Helicobacter pylori cag Pathogenicity Island. PLoS Genet 6: e1001069.
18. Kwok T, Zabler D, Urman S, Rohde M, Hartig R, et al. (2007) Helicobacter exploits integrin for type IV secretion and kinase activation. Nature 449: 862-866.
19. Jiménez-Soto LF, Kutter S, Sewald X, Ertl C, Weiss E, et al. (2009) Helicobacter pylori type IV secretion apparatus exploits beta1 integrin in a novel RGD-independent manner. PLoS Pathog 5: e1000684.
20. Zhong Q, Shao SH, Cui LL, Mu RH, Ju XL, et al. (2007) Type IV secretion system in Helicobacter pylori: a new insight into pathogenicity. Chin Med J (Engl) 120: 2138-2142.
21. Kutter S, Buhrdorf R, Haas J, Schneider-Brachert W, Haas R, et al. (2008) Protein subassemblies of the Helicobacter pylori Cag type IV secretion system revealed by localization and interaction studies. J Bacteriol 190: 2161-2171.
22. Vaidya M, Patel M, Panchal H (2012) ProPhyC: Protein PhysicoChemical properties calculator. International Journal of Computer Science and Management Research 1: 980-983.
23. 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.
24. Jayaram B, Bhushan K, Shenoy SR, Narang P, Bose S, et al. (2006) Bhageerath: an energy based web enabled computer software suite for limiting the search space of tertiary structures of small globular. Nucleic Acids Res 34: 6195- 6204.
25. DeLano WL (2002) The PyMOL Molecular Graphics System." DeLano Scientific LLC, San Carlos, CA, USA.
26. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereo chemical quality of protein structures. J Appl Cryst 26: 283-291.
27. Castrignanò T, De Meo PD, Cozzetto D, Talamo IG, Tramontano A (2006) The PMDB Protein Model Database. Nucleic Acids Res 34: D306-309.
28. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res 33: W363-367.
29. Mashiach E, Schneidman-Duhovny D, Andrusier N, Nussinov R, Wolfson HJ (2008) FireDock: a web server for fast interaction refinement in molecular docking. Nucleic Acids Res 36: W229-232.
30. Olbermann P, Josenhans C, Moodley Y, Uhr M, Stamer C, et al. (2010) A global overview of the genetic and functional diversity in the Helicobacter pylori cag pathogenicity island. PLoS Genet 6: e1001069.