American Journal of Bioinformatics 1 (1): 30-40, 2012
ISSN 1948-9862

Use of Enterobacterial Repetitive Intergenic Consensus PCR in Detecting Target(s) of Hapalindole-T, From a Cyanobacterium, in *Escherichia Coli*: *In Silico* Validation

[1]Manoj K. Tripathi, [1]Maheep Kumar, [2]Vinay K. Singh,
[3]Gopal Nath and [1]Ravi K. Asthana
[1]Department of Botany, Faculty of Science,
[2]School of Biotechnology, Faculty of Science,
[3]Institute of Medical Sciences, Department of Microbiology,
Banaras Hindu University, Varanasi-221005, India

Abstract: Problem statement: Identification of newer biomolecule as well as targets are a rising concern because of increasing drug resistance in bacteria. We have isolated a broad spectrum antibiotic biomolecule Hapalindole-T from a cyanobacterium, *Fisherella* sp. growing on local *Azadirachta indica* tree bark. A model bacterium *E. coli* was screened spontaneously for *Hap-T* resistance. These resistant strains of *E. coli* were used to identify *Hap-T* target(s). **Approach:** These strains were subjected to Enterobacterial Repetitive Intergenic Consensus (ERIC) PCR analysis and compared with the sensitive one. We have used different bioinformatic tools like Clustalw, NJ Plot and Docking server. The Swiss-model server was used for homology modelling. Predicted 3D structure was refined by energy minimization and quality was assessed by Procheck. The model protein hailed from fimbrial biogenesis outer membrane usher protein (ADK89122.1). The interaction between the predicted structure of Model-1 protein and *Hap-T* biomolecule was analysed *in silico* using Autodock and Mopac parameters. **Results:** An additional band of DNA fragment (~500 base pairs) was found on agarose gel run after amplified genome of resistant strain. The results indicated that certain residues (Tyr-28), (Phe-54), (Leu-36) and (Val-52) were highly conserved and present in active sites. **Conclusion/Recommendations:** Thus understanding of microbial adhesion can act as an alternative approach in development of broad spectrum antibiotics.

Key words: Homology modelling, active sites, alternative approach, enterobacterial repetitive, resistant strain, highly conserved, active sites, model bacterium

INTRODUCTION

There is a rising concern of Multi-Drug Resistance (MDR) in clinical practices. In spite of launching synthetic/artificial antibiotics in the market, there is a need of screening natural drugs/lead molecules with newer target (s) from natural resources. Available antibiotics in the market are bactereostatic or bactericidal in nature with a aim to eradicate bacteria based on different modes of action such as cell-wall biosynthesis, inhibition of protein synthesis or DNA replication and repair leading to bacterial evolution. There is a serious concern regarding containment of spread of *Staphylococcus aureus* (MRSA) which led the scientists to think over alternative of the drugs already available with different modes of action (Mwangi *et al*., 2007). Whole genome sequencing of *Staphylococcus aureus* strain RN4220 revealed that virulence and general fitness of the pathogen is related with the genetic polymorphism (Nair *et al*., 2011). Targeting bacterial virulence is an alternative approach to the development of new antimicrobials (Marra, 2004) Virulence specific therapeutics would also avoid evolution in bacterial system, preventing them from pathogenesis (Cegelski *et al*., 2008). Therefore, Hapalindole-T isolated from a cyanobacterium, *Fischerella sp.* colonising *Azadirachta indica* (a medicinal tree) was subjected to drug targeting (Asthana *et al*., 2006). Increasing knowledge of proteomics, genomics and bioinformatics provided new dimensions in the screening of drug targets. Identifying the drug targets by screening resistant mutants of model bacterium against the target drug is one of the popular approaches. Thus the resistant strain can be analysed by

Corresponding Author: Ravi K. Asthana, Department of Botany, Faculty of Science, Banaras Hindu University, Varanasi-221005, India

using proteomics/genomics/bioinformatics. Report of repetitive elements in prokaryotic genomes and there distribution in eubacteria has application to fingerprinting of bacterial genomes (Sharples and Lloyd, 1990). The distribution of Enterobacterial Repetitive Intergenic Consensus (ERIC) elements known as Intergenic Repeat Units (IRU) are 126 base pairs having a highly conserved central inverted repeats and are located in extragenic regions and can be transcribed (Hulton *et al*., 1991). PCR analysis using primers to ERIC sequences, with bacterial genomic DNA as template demonstrates inter ERIC distances and patterns specific for bacterial species and strains limited to adjacent repeat elements within the limitations of polymerase extension (~500 bp) (Versalovic *et al*., 1991). Therefore, ERIC PCR amplification was done for both *Hap-T*S and *Hap-T*R strains of *Escherichia coli* and the product was run on agarose gel. The differing band was subjected to DNA sequencing and related protein was identified. Thus the present report encompasses annotation of the protein structure, comparative amino acid sequence analysis, *insilco* screening of active sites, docking with Hapalindole-T (*Hap-T*) molecule using different bioinformatic packages.

MATERIALS AND METHODS

Bacteria, media and hapalindole-T: *Escherichia coli* DH5α was routinely cultured and maintained in Luria Bertani broth (LB) and on LB agar plates (1.0 tryptone, 0.5% yeast extract and 1.0% NaCl) (37°C, 24-48 h). *Hap-T* ($C_{21}H_{23}N_2ClSO$, Mr 386, melting point 179-182°C) isolated from the biomass of *Fischerella sp.* a cyanobacterium colonizing Neem tree (an Indian medicinal plant) bark (Asthana *et al*., 2006) was used in the present investigation.

Screening of *E. coli* DH5α resistant (*Hap-T*^{R}) strain: The bacterial cells grown in LB broth were washed and suspended in Phosphate Buffered Saline (PBS) at concentration ~ 10^7 cells/mL. As a first step towards isolation of *Hap-T*^{R} mutants, the survival of the *E. coli* DH5α was checked at increasing concentrations of *Hap-T* (10-50 μg m L^{-1}). Spontaneously occurring *Hap-T*^{R} mutants were obtained by plating approximately 10^7 cells/mL on solid medium (Muller Hinton m edium) and few colonies growing on 50 μg mL^{-1} of *Hap-T* plate were selected. Few surviving colonies that remained were picked up and transferred to the plate containing 50 μg mL^{-1} *Hap-T*. These coloies were grown upto seven generation in *Hap-T* deficient medium. The mutant strain was maintained on the plate having 50 μg mL^{-1} of *Hap-T*.

Isolation and quantification of genomic DNA: The cells of both the strains *Hap-T*^{S} and *Hap-T*^{R} were grown overnight in LB broth (37°C). Bacterial cultures (1.5 mL of each) were centrifuged (10,000 g × 10 min at room temperature). Cell pellets were washed twice and resuspended in TE buffer (10 mM Tris, 25 mM EDTA, pH 8.0) and vortexed. SDS (10%) was added followed by the addition of proteinase-K (100 μg mL^{-1}). This solution was mixed gently and incubated at 37°C for 1h. NaCl (5M) was added, vortexed and incubated (65°C) after adding cetyl trimethyl ammonium bromide (CTAB, 10%). The lysates were extracted with chloroform and Isoamyl Alcohol (IAA) (24:1). Aqueous phase was collected after centrifugation again extracted with tris saturated phenol, chloroform and IAA (25: 24: 1). Again aqueous phase was collected after centrifugation and RNAase was added (~ 30 μg mL^{-1}) and incubated 37°C. Equal volume of isopropanol was added to precipitate the DNA and centrifuged. The pelleted DNA was washed twice with 70% ethanol (chilled) and resuspended in milliQ water and stored at 4°C. Genomic DNA was quantified by UV/VISspectrophotometer (6715, Jenway, Germany) at 260 nm.

ERIC-PCR amplification: Primers used for ERIC-PCR were ERIC-1R (5'-ATGTAAGCTCCTGGGGATTCAC-3') and ERIC-2 (5'-AAGTAAGTGACTGGGGTGAGCG-3').
(Versalovic *et al*., 1991). PCR was done in a 50 μL reaction mixture containing 5 μL of 10X buffer, 1.2 mM dNTPs, 1.5 mM Mg Cl2, 5 Unit of *Taq* DNA polymerase (Fermentas Corporation Ltd. USA), 50 pmol of each primers and 60 ng of template DNA. The PCR amplification was carried out in Master Cycler epgradient (Eppendrof, Germany) according to the following protocol: An initial denaturation of 5 min, followed by 35 cycles of denaturation at 90°C for 30 sec, annealing at 50°C for 1min, extension at 72°C for 5 min and a final extension at 72°C for 15 min. Thereafter PCR product was examined on through horizontal electrophoresis in agarose gel electrophoresis (1.5%) containing ethidium bromide (0.5 μg mL^{-1}) at 65V for 6 h in 1X TAE buffer. The DNA finger prints on gel images were captured by gel-doc system (Bio-Rad, USA) and kept for further analysis.

Analysis of band patterns: DNA fingerprints of strains (*Hap-T*^{S} and *Hap-T*^{R}) were compared for similarity by visual inspection of band patterns. Two fingerprints were considered identical if the same number of bands at corresponding positions were observed, while variations in the band intensity were

not considered. The size of the bands was calculated by visual comparison to the molecular size marker run along with the PCRproducts. The differing band (DNA fragment) was excised with a sterile scalpel and the DNA was eluted directly using gel elution column (Qiagen, Germany) and stored in milliQ water.

Sequencing of ERIC-PCR amplified DNA fragment: Sequencing reactions were performed in an automated sequencer (Applied Biosystem, USA) according to manufacturer's manual.

Sequence analysis: Sequences of proteins from different organisms were aligned using Clustal W (Thompson *et al*., 1994) and phylogenetic tree was constructed using NJ plot method. A tree was inferred by Bootstrap phylogenetic inference using Tree view (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html). The conserved motifs present in these sequences were analyzed using BLOCKS and MEME (Multiple EM for Motif Elicitation) software version 3.5.7 (Bailey and Gribskov, 1998; Bailey *et al*., 2006). Proteins from diverse bacterial species were screened for motifs identification with residues minimum (3) and maximum (6) width for a maximum number of 20 motifs while rest of the parameters were kept at default.

Template identification and model generation: The sequence of ERIC PCR amplified gene was subjected to homology search with NCBI databases using BLASTN, TBLASTX and discontinuous MEGABLAST (Altschul and Lipman, 1990; Arnold *et al*., 1997; Thompson *et al.* 2009). The translated protein sequences were subjected to protein functional analysis using PFAM version 23.0 (Finn *et al*., 2006), PROSITE version 20.37 (Castro *et al*., 2006) and INTERPROSCAN version 4.4 (Quevillon *et al*., 2005). The protein structure identification was referred as Model-1 related with fimbrial biogenesis outer membrane usher protein. The Model-1 protein was used for template selection using advanced search option at Protein Data bank (http://www.pdb.org/pdb/home/home.do) and homology modelling was done using Swiss model http://swissmodel.expasy.org/, (Arnold *et al*., 2006) and Geno3D (http://geno3dpbil. ibcp.fr/). The rough model generated was subjected to energy minimization using the steepest descent technique to check the non-compatible contacts within protein. Computations were carried out in vacuo with the GROSMOS96 43B1 parameters set, implemented through Swiss-pdbViewer (http://expasy.org/spdv/). Model consistency and viability were appraised by PDBsum server (http://www.ebi.ac.uk/thorntonsrv/ databases/pdbsum/Generate.html).

Preparation of receptor (Model-1), loop refinement and evaluation: The backbone conformation of the rough model was inspected using the Phi/Psi Ramachandran plot obtained in the PROCHECK server (www.ebi.ac.uk/pdbsum). The results of Ramachandran plot indicated that the rough model generated had no any residue in the disallowed region, thus no need of loop refinement. The initial energy of protein was calculated (kcal/mol) using MMFF94x force field. The protein structure was subjected to energy minimization and its final energy was calculated.

Superimposition of target (*Hap-T*) and template (Outer membrane usher protein of *E. coli*): The structural superimposition of Cα trace of the template (Outer membrane usher protein of *E. coli*) and structure of Model-1 was performed using Combinatorial Extension of Polypeptides (http://www.cl.sdsc.edu). The root mean square deviation was calculated using chimera 1.5.2 (http://www.cgl.ucsf.edu/chimera/).

Active site determination: Active sites present in the Model-1 protein were identified using Q-SITEFINDER http://bmbpcu36.leeds.ac.uk/qsitefinder (Laurie and Jackson, 2005; 2006).

Ligand (*Hap-T*) preparation: The ligand was drawn using ISISdraw program (http://www.ch.cam.ac.uk/cil/SGTL/MDL/ISISdraw.html) and converted to threedimensional format (i.e., pdb) using Molekel Visualization Package (http://molekel.cscs.ch/wiki/pmwiki.php). The energy was calculated using docking server.

Docking studies: Docking calculations were carried out using Docking Server (Hazai *et al*., 2009) for Model-1 protein and *Hap-T* biomolecule (http://www.dockingserver.com). The MMFF94 force field (Halgren, 1996) was used for energy minimization of ligand molecule (*Hap-T*). Gasteiger partial charges were added to the ligand atoms. Non-polar hydrogen atoms were merged and rotatable bonds were defined. Essential hydrogen atoms, Kollman united atom type charges and solvation parameters were added with the aid of AutoDock tools and affinity (grid) maps of 20×20×20 Å grid points and 0.375 Å spacing were generated using the Autogrid program (Morris *et al*., 1998). AutoDock parameter set and distance-dependent dielectric functions were used in the calculation of the van der Waals and the electrostatic terms respectively. Docking simulations were performed using the Lamarckian Genetic Algorithm (LGA) and the Solis and Wets (1981) local search method. Initial position, orientation and torsions of the ligand molecules were

set randomly. Each docking experiment was derived from 10 different runs that were set to terminate after a maximum of 250000 energy evaluations. The population size was set to 150. During the search, a translational step of 0.2 Å and quaternion and torsion steps of 5 were applied.

RESULTS

ERIC PCR amplification: Fingerprints obtained in agarose gel by ERIC PCR amplified product of *Hap-T*^{S} and *Hap-T*^{R} genomes consisted of multiple distinct band. An additional band of ~500 *base pairs* (*bp*) was present in the fingerprints obtained from the *Hap-T*^{R} strain of *E. coli* (Fig. 1). This DNA fragment was sequenced and out of 500bp only initial 201bp were translated as putative outer membrane usher protein using translation tool (http://expasy.org/tools/dna.html). Therefore, only 201bp were submitted to NCBI with the accession number HM625744. The sequence translated 66 amino acids, deduced as: CQCTISCWLTPFTKLMYRLVLLNFDLYSTSSSGD LLVEIKIAEYCPHSYQVPFSSAPLRHRPGRN. This amino acid sequence was deposited in GenePept (www.ncbi.nlm.nih.gov) and accession number was assigned as 'ADK89122.1' (Model-1). This led us to retrieve the full length outer membrane usher protein of *E. coli*. Homology was searched using BLASTX (search protein database using a translated nucleotide query) of NCBI and the sequence was traced with their similarity for fimbrial biogenesis outer membrane usher protein in different bacterial species.

Multiple sequence alignment: Multiple sequence alignment of HM625744/ADK89122.1 showed many amino acids as conserved in different bacterial species (Fig. 2) with highly conserved residues DLYPTSSSGDL and VPFSAVP in aligned part. The phylogenetic tree was constructed using NJ Plot tool showing the presence of two major clusters in aligned bacterial species (Fig. 3). However, our target protein (ADK89122.1) translated from DNA sequence HM625744_*Escherichia coli* was found to be present in cluster B2 which is close to *Shigella boydii* CDS and *E. coli* B088. The BLASTP similarity search of target protein also showed 77% sequence identity with *Shigella boydii* CDS and *E. coli* B088 (Table 1).

MEME analysis: Block diagram obtained after MEME analysis clearly revealed nine motifs in selected bacterial species. However, there was absence of motif1 with regular expression of [IV] [KT] EADG in target protein (Fig. 4). This indicates alteration in amino acid sequence during resistance. Six out of nine motifs are highly conserved. The most representative residues present in all bacterial species were motif 2, motif 3, motif 4 and motif 5 showed as shown in Fig. 5.

Homology modelling of membrane usher protein: Model-1 (ADK89122.1) was subjected to INTERPROSCAN to screen the family of the protein. The result showed that the protein belonged to fimbrial biogenesis outer membrane usher protein (PF00577). The hypothetical protein was submitted to SWISS MODEL Server (http://swissmodel.expasy.org/) (Arnold *et al*., 2006) for the prediction of 3-D model of the protein as a Model-1 (Fig. 6a) having two β-sheets and one turn. The protein was searched in the protein structure database (http://www.pdb.org/pdb/home/home.do) for finding the template structure which matched with Outer membrane usher protein of *E. coli* (Fig. 6b). The Model-1 protein is represented by 2 Beta sheets and three chains. Energy of the protein Model-1 and Outer membrane usher protein of *E. coli* was-997.113 and-15764.224 kcal/mol however, after energy minimization it was-1224.695 and-28351.635 kcal/molrespectively.

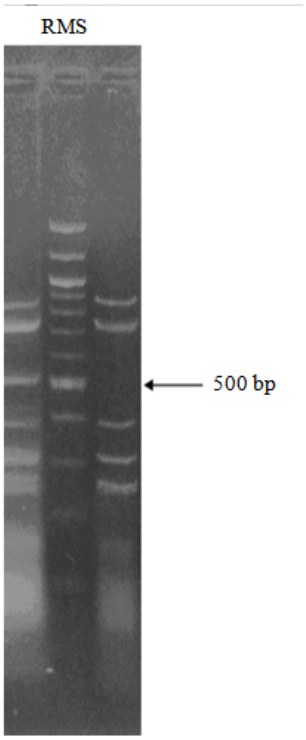


Fig. 1: Agarose gel (1.2%) electrophoresis of ERIC PCR amplified genome of *Hap-T*R (R) and *Hap-T*S (S) along with DNA marker (M) (1 kb) showing difference at ~500 bp DNA fragment

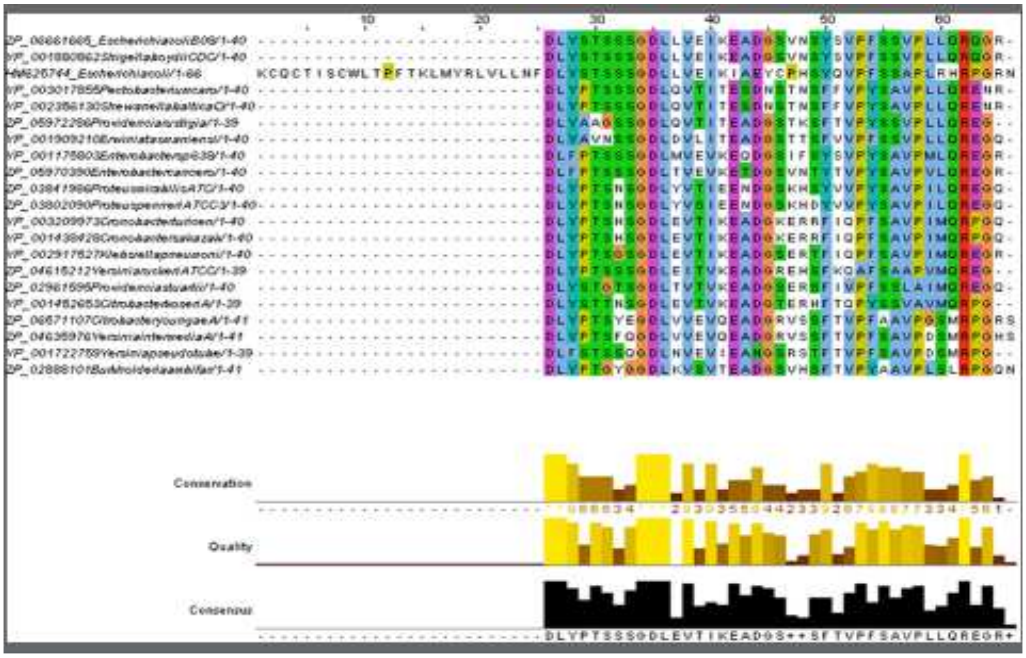


Fig. 2: Multiple sequence alignment output of Model-1 (Accession No. ADK89122.1) using Clustalw showing the conserved regiion among different bacterial species

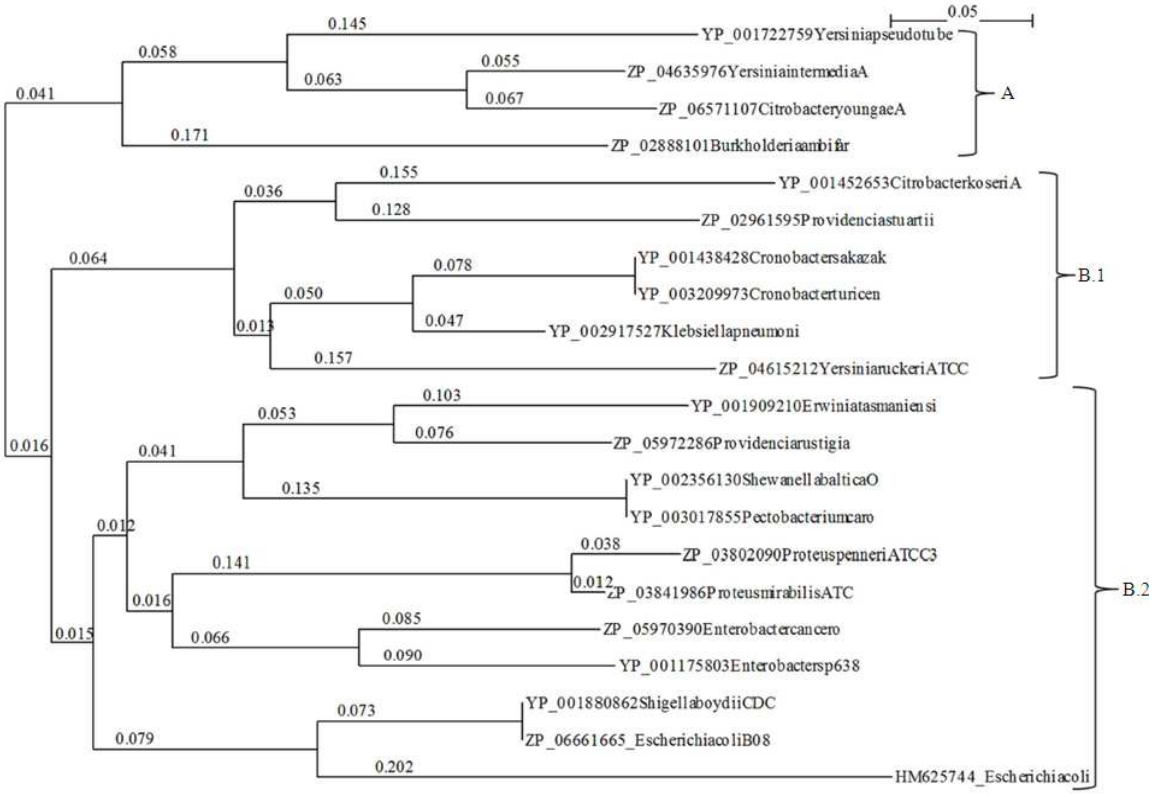


Fig. 3: Phylogenetic Tree inferences using NJ plot method

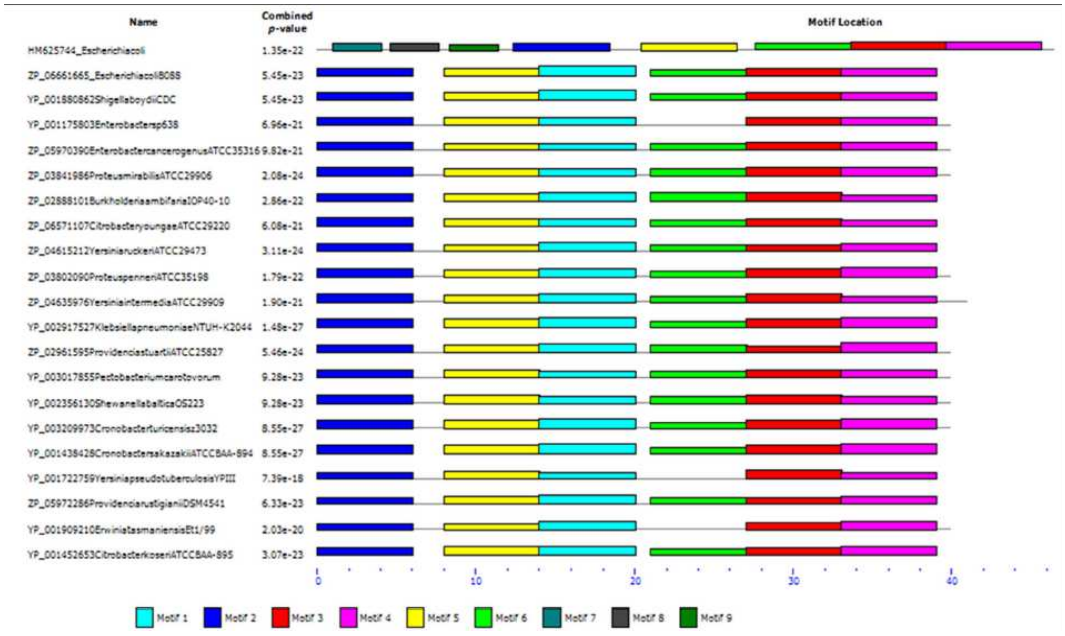


Fig. 4: Multiple Em for Motif Elicitation Study representing 9 motifs

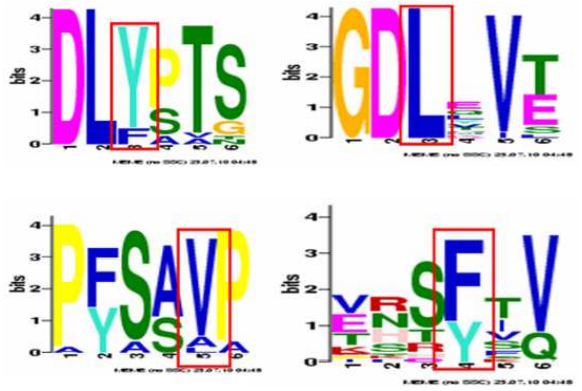
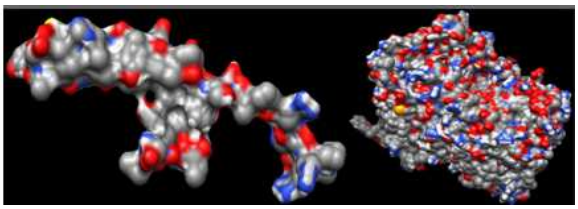


Fig. 5: Active sites showing in red circled portion of highly representative motifs

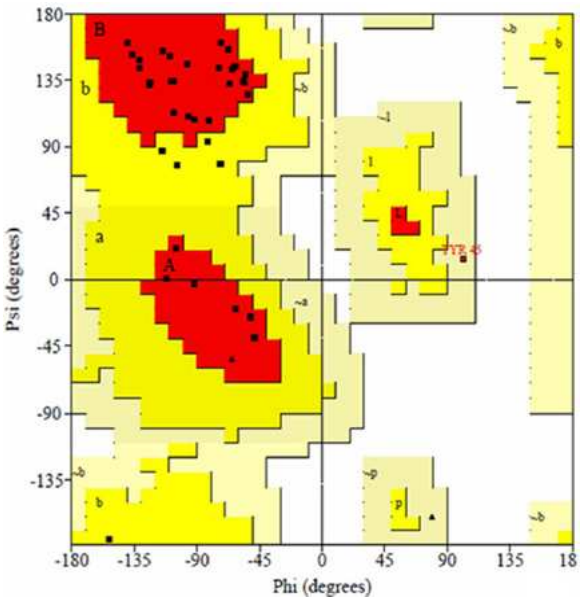


(a) (b)

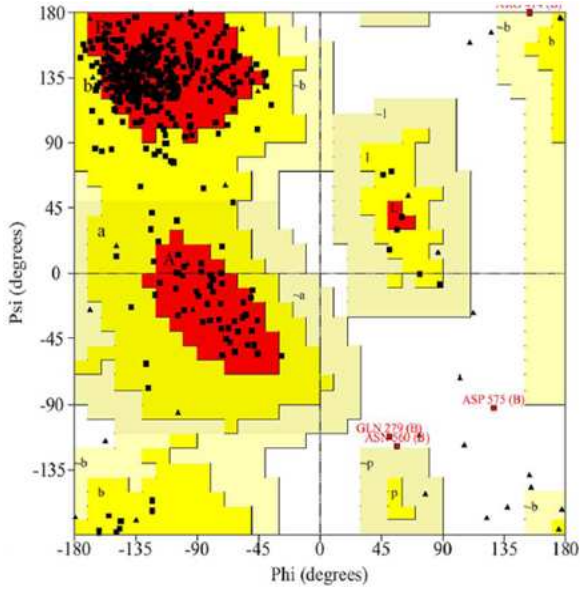
Fig. 6 a): The 3D structure of Model-1 (Accession No. ADK89122.1) obtained from Swiss model Workplace, b) the 3D model of full length outer membrane usher protein (obtained after homology modelling)

Validation of model-1 and its superimposition with the template (Outer membrane usher protein of *E. coli*): The stereochemistry of the constructed Model-1 protein and outer membrane usher protein of *E. coli* was subjected to energy minimization and the stereochemical quality of the predicted structure assessed. The deciphered model of protein got validated with Ramachandran plot as depicted in Fig. 7a and b. Most of the residues in Model-1 (82.8%) were placed in the core region of the Ramachandran plot and 15.2 as well as 3% in additional allowed region and in generous allowed region respectively. However, there is no residue in disallowed region suggesting best possible stability in the model. No difference was found in reconstructed Ramachandran Plot after energy minimization. Similarly, Ramachandran plot of outer membrane usher protein of *E. coli* showed, 84.0% of the residues in core region, 14.8% in additional allowed region, 0.5% in generous allowed and 0.7% in disallowed region. The plot after energy minimization showed a minor difference of residues i.e., 84.3% in most favoured, 14.7% in additional allowed, 0.2% in generous allowed and again 0.7% in disallowed region.

The superimposition of Model-1 protein with the template is shown in Fig. 8. The weighted Root Mean Square Deviation (RMSD) of Cα trace between the template and the final refined model was calculated as 0.351 Å with a significant Z-score of 3.1.



(a)



(b)

Fig. 7: Ramachandran Plots after energy minimization of identified protein model-1 (a) and outer membrane usher protein (b)

Active site identification of Model-1: Ten binding sites in both Model-1 and outer membrane usher protein of *E. coli* were obtained using Q-sitefinder (Fig 9a and b). Among these sites of Model-1 (Fig. 9a) residues of site 1 (TYR-28, LEU-36, VAL-52 and PHE-54), site 2 (TYR-28 and PHE-54) site 4 (PHE-54) and site 8 (TYR-28) were found to be conserved in other bacteria as represented in multiple sequence alignment (Fig. 2).

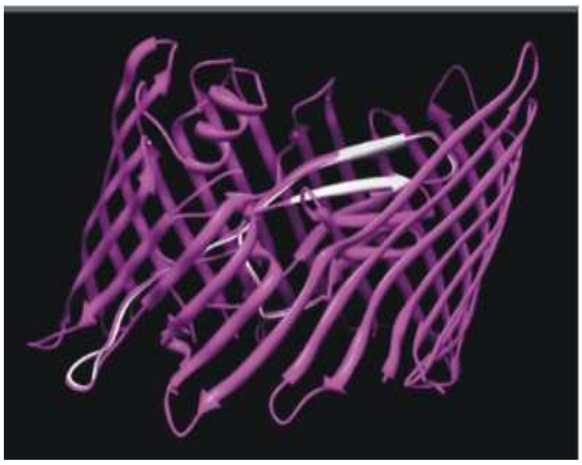
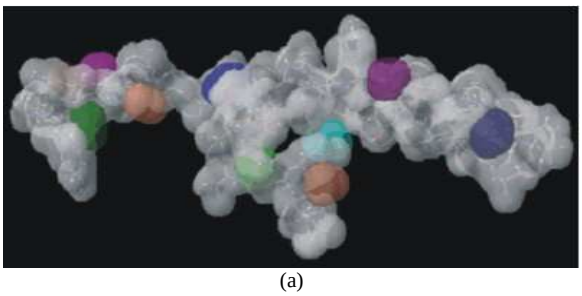
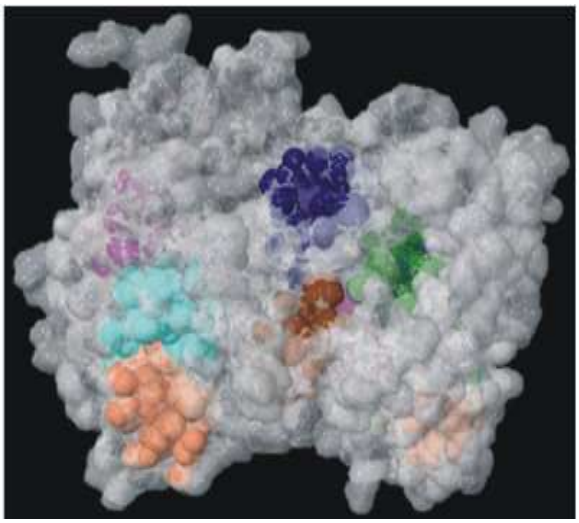


Fig. 8: Superimposition of Model-1 protein (Accession No. ADK89122.1) with the template (outer membrane usher protein) using using Combinatorial Extension of Polypeptides (http://www.cl.sdsc.edu)



(a)



(b)

Fig. 9: Predicted active sites of protein model-1 (a) and outer membrane usher protein (b)

Table 1: BLASTP similarity search for *E. coli* ADK89122.1

Organism name	Accession no.	Sequence identity (%)	Sequence e-value similarity (%)
Escherichia coli B088	ZP_06661665 72	77	2e-07
Shigella boydii CDC	YP_001880862 72	77	2e-07
Enterobactor sp. 638	YP_001175803 52	70	4e-04
Enterobacter cancerogenus ATCC35316	ZP_05970390 50	70	0.001
Proteus mirabilis ATCC29906	ZP_03841986 50	67	0.004
Burkholderia ambifariaI OP40-10	ZP_02888101 46	63	0.005
Citrobacter youngae ATCC29220	ZP_06571107 48	68	0.005
Yersinia ruckeri ATCC29473	ZP_04615212 51	69	0.009
Proteus penneri ATCC35198	ZP_03802090 47	65	0.011
Yersinia intermedia ATCC29909	ZP_04635976 48	65	0.014
Klebsiella pneumoniae NTUH-K2044	YP_002917527 50	62	0.022
Providencia stuartii ATCC25827	ZP_02961595 50	65	0.024
Pectobacterium carotovorum	YP_003017855 50	65	0.027
Shewanella baltica OS223	YP_002356130 51	63	0.028
Cronobacter turicensis z3032	YP_003209973 50	62	0.029
Cronobacter sakazakii ATCCBAA-894	YP_001438428 50	62	0.029
Yersinia pseudotuberculosis YPIII	YP_001722759 51	64	0.030
Providencia rustigianii DSM4541	ZP_05972286 51	61	0.031
Erwinia tasmaniensis Et1/99	YP_001909210 52	65	0.042
Citrobacter koseri ATCCBAA-895	YP_001452653 46	64	2e-09

Table 2a: Hydrophobic interaction of ligand (*Hap-T*) with model^{-1}protein (ADK89122.1)

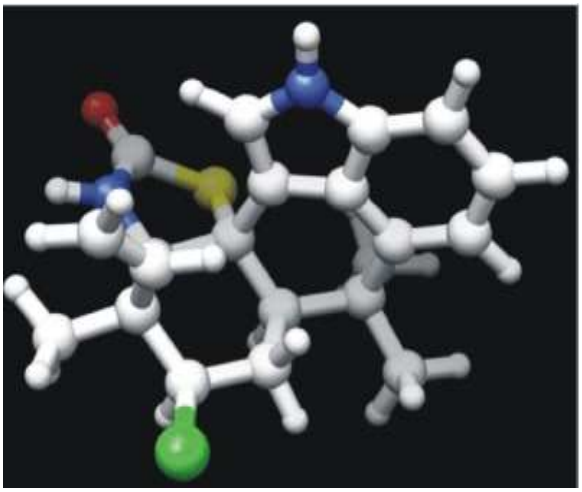
Interacting atom of ligand	Interacting amino acid of Model^{-1}
C14 (16) [3.51]	TYR28 (CD2, CE2)
C8 (8) [3.57]	LEU36 (CD1)
C18 (22) [3.65]	LEU36 (CD1)
C21 (26) [3.35]	VAL52 (CG1)
C18 (22) [3.18]	VAL52 (CG1)
C17 (21) [3.13]	PHE54 (CB, CD1)
C18 (22) [3.63]	PHE54 (CB, CD1)
C20 (24) [3.55]	PHE54 (CD1, CE1)

Table 2b: Hydrophobic interaction of Ligand (*Hap-T*) with the protein 2vqiB

Interacting atom of ligand	Interacting amino acid of 2vqiB
C17 (21) [3.34]	TYR284 (CB, CD1)
C18 (22) [3.60]	TYR284 (CD1)

The corresponding active sites were also found in protein outer membrane usher protein of *E. coli* (Fig. 9b) i.e., VAL (site 1), PHE, VAL and TYR (site-2, 3), LEU, PHE and TYR (site 4), TYR, LEU, VAL and PHE (site 5), TYR (site 6), VAL, PHE, TYR and LEU (site 8), VAL and PHE (site 9) and TYR, VAL and LEU (site 10).

Docking studies: There were ten possible active sites shown in different colours (Fig. 9a and b). Ligands were prepared (-6.70 kcal/mole) as depicted in Fig. 10. For docking studies, the Docking server was chosen because its algorithm allowed full flexibility of small ligands. It has been shown that one out of three predicted model structures was successfully reproduced and it included hydrophobic interaction with *Hap-T* an 294931294931d empirical evaluation of the binding free energy (Table 2a and b). Figure 11a and b represented interaction of ligand (*Hap-T*) and receptor molecule (Model-1 and outer membrane usher protein of *E. coli*).



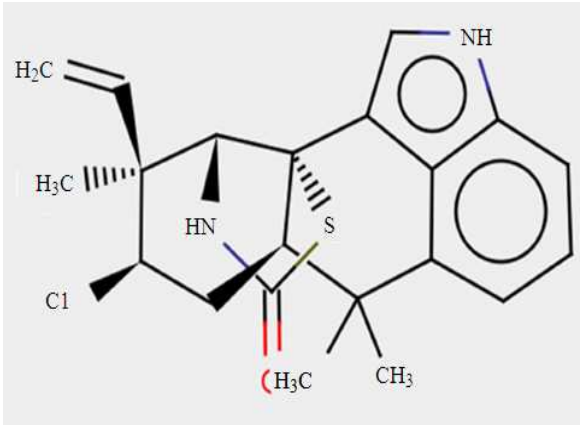
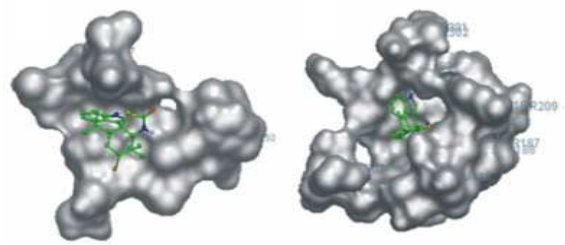


Fig. 10: Structure of Ligand (Hapalindole-T) using ISISdraw having 3D model and line draw



(a) (b)

Fig. 11: (a) Docking of Hapalindole-T with Model-1 and (b) outer membrane usher protein

Docking calculations for Model-1 and outer membrane usher protein with *Hap-T* estimated free energy of binding as -6.49 kcal/mol and -4.72 kcal/mol and their interaction surface amounted 584.801 and 695.603 respectively. The tyrosine residue of both the proteins interacts with carbon atom of ligands. However, phenylalanine, leucine and valine of Model-1, also interacted with *Hap-T*. These residues were prevalent in motifs.

DISCUSSION

In the present study, we have used *Hap-T*, a broad spectrum antibacterial biomolecule (Asthana *et al*., 2006) against a model bacterium *E. coli* DH5α and addressed the evolution of drug resistance. Computer modeling of the target molecule and its inslico interaction with *Hap-T* revealed involvement of fimbrial biogenesis outer membrane usher protein in resistance of the model bacterium. This interesting finding gets importance as the targeting of bacterial virulence factors instead of cellular processes so far has become an emerging area of drug development as reviewed by Kline *et al*. (2010). The importance of repeating sequences in bacteria led the foundation of identification of closely related strains (Sharples and Lloyd, 1990; Hulton *et al*., 1991; Versalovic *et al*., 1991; Thompson *et al*., 1994). Enterobacterial Repetitive Intergenic Consensus sequences (ERIC) also described as Intergenic Repetitive Units (IRU) differs from most other bacterial species repeats in being distributed across a wider range of species. Therefore, ERIC sequences may offer greater potential in comparative analysis of *Hap-T*^{S} and *Hap-T*^{R} strains by mapping the whole genome. It is reported to distinguish *Mycobacterium avium subsp. paratuberculosis* from closely related mycobacteria (Englund, 2003).

They have shown one additional band of approximately 650bp by IS900/ERIC PCR differentiating two closely related organisms. The ERIC PCR has been found to discriminate not only different strains of a species/subspecies but it can also discriminate between strains of the same serotype (Nath *et al*., 2010). We have also observed an additional band of ~500 bp in *Hap-T*R strain by this method analysing the whole genome of the bacterium (ref Fig. 1).

The differing band of the ERIC PCR was subjected to DNA sequencing and the translated amino acids in turn used in retrieving the Model-1 protein via different bioinformatic tools. Thus Model-1 protein (Accession No. ADK89122.1, ref Fig. 6a). To substantiate our findings we took the full length sequences of outer membrane usher protein (Fig. 6b) form database and selected as template for superimposition. Superimposition with the template (Fig. 8) was found to significant with RMSD of 0.351 Å. Outer membrane usher protein known to be involved in biogenesis of pilus in Gram-negative bacteria. The biogenesis of fimbrae (pili) requires a two component assembly and transport system which is composed of a periplasmic chaperone and outer membrane protein which has been termed as molecular usher. These proteins are outer membrane assembly platform where pili are assembled. The biogenesis of pili was inhibited by certain pillicides reducing pilus formation in uropathogenic *E. coli* because of point mutation in the pellicle binding site thus resulting in loss of virulence in pathogen (Pinkner *et al*., 2006; Cegelski *et al*., 2009). This indicates importance of virulence factor in as new alternative drug targets. Identification of Model-1 protein (Accession No. ADK89122.1) in the resistance strain adds significance of targeting bacterial virulence as an alternative approach to develop new antimicrobials in present case.

Multiple sequence alignment of Model-1 (ADK89122.1) highlighted the sequence conservation of amino acid residues among different microbial species as: Asp-26, Leu-27, Gly-34, Asp-35, Leu-36, Glu-42, Asp-44, Val-52, Pro-53, Phe-54, Ser-55, Val-57, Arg-62 and Gly-64 however, among these sequences, Model-1 showed Ile-42, Glu- 44 and Ala-57 in place of Glu-42, Asp-44 and Val-52 respectively (ref Fig. 2). These conserved sites were the probable drug targets. After docking studies it has been found that most of our interaction sites fall among the above mentioned sites. Based on these conserved motif study four most representative residue stretches revealed were [DLYPTS], [GDLVT], [PFSAVP] and [VRSFTV] (ref Fig. 5) however, in ADK89122.1, amino acid sequence used in the present study showed that motif 1 has ‘S’ in place of ‘P’ motif 2 contain ‘E’ in place of ‘T’ and motif 3 has ‘S’ and ‘A’ in place of ‘A’ and ‘V’ respectively. These stretches also contained active site residue Tyr-28,

in motif-1, Leu-36 in motif-2, Val-52 in motif-3 and Phe-54 in motif-4 in Model-1 protein (ref Fig. 4, 5 and 9). This conservation, however, were concomitant with differences and were sufficient enough to support the variations subsequently reflected at the structural and functional levels. Predicted protein Model-1 was refined and structure was established. Ramachandran Plot evaluated the stretches of Model-1 protein that there was no residue in disallowed region (Fig. 7a).

Phylogenetic tree traces the interrelationships of *E. coli* B008 and *Shigella boydii* CDC in class B2 showing homology with 72% identity containing 77% positivity (ref Table 1 and Fig. 3). This gene family of protein seemed to be strictly conserved and evolved from same ancestral species undergoing speciation. Motifs containing the signature sequences ([DLYPTS], [GDLVT], [PFSAVP] and [VRSFTV]) are well conserved and showed the docking site with *Hap-T* (ref Fig. 5, 11a and 11b).

CONCLUSION

The similarity in the chaperone/usher assembled pili in Gram-negative and sortase enzyme assembled Gram-positive bacteria are pathogenic determinants and paving to be promising targets for anti therapeutics (Maresso and Schneewind, 2008). Identification of Model-1 protein (Accession No. ADK89122.1), a member from the family of fimbrial biogenesis outer membrane usher protein in *Hap-T*^{R} strain as target indicated need of more deeper understanding of pili assembly in both the bacterial strain to develop a common broad spectrum new drug.

ACKNOWLEDGEMENT

We are grateful to the Head and Programme Coordinator of Centre of Advanced Study in Botany, Banaras Hindu University for laboratory facilities. We are also grateful to Prof. R. Raman, Department Zoology, Banaras Hindu University for providing the DNA sequencing facility. The valuable assistance regarding bioinformatics analysis and tools from Centre for Bioinformatics, School of Biotechnology, BHU is also gratefully acknowledged. Financial support from DST, New Delhi (Ref. No. SR/SO/PS-04/2006 dt. 03.07.2007) to RKA and MKT as JRF/SRF and from UGC-RGNFS JRF/SRF, New Delhi {F. 14-2(SC)/2008 (SAIII)} to MKT is gratefully acknowledged.

REFERENCES

Altschul, S.F. and D.J. Lipman, 1990. Protein database searches for multiple alignments. Proc. Natl. Acad. Sci., 87: 5509-5513.

Arnold, K., L. Bordoli, J. Kopp and T. Schwede, 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics, 22: 195-201. DOI: 10.1093/bioinformatics/bti770

Arnold, S.F., T.L. Madden, A.A. Schaffer, J. Zhang and Z. Zhang *et al*., 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucl. Acids Res., 25: 3389-3402. DOI: 10.1093/nar/25.17.3389

Asthana, R.K., A. Srivastava, A.P. Singh, D. Singh and S.P. Nath *et al*., 2006. Identification of an antimicrobial entity from the cyanobacterium *Fischerella sp.* isolated from bark of *Azadirachta indica* (Neem) tree. J. Applied Phycol., 18: 33-39. DOI: 10.1007/s10811-005-9011-9

Bailey, T.L. and M. Gribskov, 1998. Methods and statistics for combining motif match scores. J. Comput. Biol., 5: 211-221. DOI: 10.1089/cmb.1998.5.211

Bailey, T.L., N. Williams, C. Misleh and W.W. Li, 2006. MEME: Discovering and analyzing DNA and protein sequence motifs. Nucl. Acids Res., 34: W369-W373. PMID: 16845028

Castro, E.D., C.J. Sigrist, A. Gattiker, V. Bulliard and P.S. Langendijk-Genevaux *et al*., 2006. ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucl. Acids Res., 34: W362-W365. DOI: 10.1093/nar/gkl124

Cegelski, L., G.R. Marshall, G.R. Eldridge and S.J. Hultgren, 2008. The biology and future prospects of antivirulence therapies. Nat. Rev. Microbiol., 6: 17-27. DOI: 10.1038/nrmicro1818

Cegelski, L., J.S. Pinkner, N.D. Hammer, C.K. Cusumano and C.S. Hung *et al*., 2009. Small-molecule inhibitors target *Escherichia coli* amyloid biogenesis and biofilm formation. Nat. Chem. Biol., 5: 913-919. DOI: 10.1038/nchembio.242

Englund, S., 2003. IS900/ERIC-PCR as a tool to distinguish *Mycobacterium avium* subsp. *paratuberculosis* from closely related mycobacteria. Vet. Microbiol., 96: 277-287. DOI: 10.1016/j.vetmic.2003.08.004

Finn, R.D., J. Mistry, B. Schuster-Bockler, S. Griffiths-Jones and V. Hollich *et al*., 2006. Pfam: Clans, web tools and services. Nucl. Acids Res., 34: D247-D2451. DOI: 10.1093/nar/gkj149

Halgren, T.A., 1996. Merck molecular force field. I. Basis, form, scope, parameterization and performance of MMFF94. J. Comput. Chem., 17: 490-519. DOI: 10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P

Hazai, E., S. Kovacs, L. Demko and Z. Bikadi, 2009. DockingServer: Molecular docking calculations online. Acta Pharm. Hung, 79: 17-21. PMID: 19526678

Hulton, C.S.J., C.F. Higgins and P.M. Sharp, 1991. ERIC sequences: A novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. Mol. Microbiol., 5: 825-834. DOI: 10.1111/j.1365-2958.1991.tb00755.x

Kline, K.A., K.W. Dodson, M.G. Caparon and S.J. Hultgren, 2010. A tale of two pili: Assembly and function of pili in bacteria. Trends Microbiol., 18: 224-232. DOI: 10.1016/j.tim.2010.03.002

Laurie, A.T.R. and R.M. Jackson, 2005. QQ-SiteFinder: An energy-based method for the prediction of protein–ligand binding sites. Bioinformatics, 21: 1908-1916. DOI: 10.1093/bioinformatics/bti315

Laurie, A.T.R. and R.M. Jackson, 2006. Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening. Curr. Protein Pept. Sci., 7: 395-406. DOI: 10.2174/138920306778559386

Maresso, A.W. and O. Schneewind, 2008. Sortase as a target of anti-infective therapy. Pharmacol. Rev., 60: 128-141. DOI: 10.1124/pr.107.07110

Marra, A., 2004. Can virulence factors be viable antibacterial targets? Expert Rev. Anti. Infect. Ther., 2: 61-72. DOI: 10.1586/14787210.2.1.61

Morris, G.M., D.S. Goodsell, R.S. Halliday, R. Huey and W.E. Hart *et al*., 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem., 19: 1639-1662. DOI: 10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

Mwangi, M.M., S.W. Wu, Y. Zhou, K. Sieradzki and H.D. Lencastre *et al*., 2007. Tracking the in vivo evolution of multidrug resistance in Staphylococcus aureus by whole-genome sequencing. Natl. Acad. Sci. USA., 104: 9451-9456. DOI: 10.1073/pnas.0609839104

Nair, D., G. Memmi, D. Hernandez, J. Bard and M. Beaume *et al*., 2011. Whole-genome sequencing of *Staphylococcus aureus* strain RN4220, a key laboratory strain used in virulence research, identifies mutations that affect not only virulence factors but also the fitness of the strain. J. Bacteriol., 193: 2332-2335. DOI: 10.1128/JB.00027-11

Nath, G., P. Maurya and A.K. Gulati, 2010. ERIC PCR and RAPD based fingerprinting of *Salmonella* Typhi strains isolated over a period of two decades. Infect. Genet. Evol., 10: 530-536. DOI: 10.1016/j.meegid.2010.02.004

Pinkner, J.S., H. Remaut, F. Buelens, E. Miller and V. Aberg *et al*., 2006. Rationally designed small compounds inhibit pilus biogenesis in uropathogenic bacteria. Natl. Acad. Sci. USA., 103: 17897-17902. DOI: 10.1073/pnas.0606795103

Quevillon, E., V. Silventoinen, S. Pillai, N. Harte and N. Mulder *et al*., 2005. InterProScan: Protein domains identifier. Nucl. Acids Res., 33: W116-W120. DOI: 10.1093/nar/gki442

Sharples, G.J. and R.G. Lloyd, 1990. A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. Nucl. Acids Res.,18: 6503-6508. DOI: 10.1093/nar/18.22.6503

Solis, F.J. and R.J.B. Wets, 1981. Minimization by random search techniques. Math. Oper. Res., 6: 19-30. DOI: 10.1287/moor.6.1.19

Thompson, J.D., D.G. Higgins and T.J. Gibson, 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acids Res., 22: 4673-4680. DOI: 10.1093/nar/22.22.4673

Thompson, K.E., Y. Wang, T. Madej, S.H. Bryant, 2009. Improving protein structure similarity searches using domain boundaries based on conserved sequence information. BMC Struct. Biol., 9: 33-33. DOI: 10.1186/1472-6807-9-33

Versalovic, J., Koeuth, T. and J.R. Lupski, 1991. Distribution of repetitive DNA sequences in eubacteria and application to finerpriting of bacterial enomes. Nucl. Acids Res., 19: 6823-6831. DOI: 10.1093/nar/19.24.6823