Journal of Molecular Graphics and Modelling 38 (2012) 101–111

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Prediction of ligand binding site by *insilico* approach in cold resistant protein isolated from cold resistant mutant of *Pseudomonas fluorescens*

Amit Kumar[b], Mahejibin Khan[a,*]

[a] *Central Food Technological Research Institute, (CSIR), Mysore, India*
[b] *Amity Institute of Biotechnology, Amity University Rajasthan Campus, Jaipur, India*

ARTICLE INFO

Article history:
Accepted 26 June 2012
Available online 7 July 2012

Keywords:
Cold shock domain
Cold resistant protein
Molecular modeling
Transcription regulation
Pseudomonas fluorescens

ABSTRACT

Cold shock proteins perform vital functions, such as mRNA masking, coupling of transcription to translation and developmental timing and regulation, which aids in survival of microbes in cold stress. *Pseudomonas fluorescens* is an ecologically important bacterium which helps in plant growth promotion. Since the cold tolerant mutant of the bacterium is able to grow at the temperature ranges from 30 to 4°C, it is of interest to study the process of its survival in the extreme temperatures. Therefore, this study is focused on the three dimensional structure and molecular modeling of cold resistant protein (CRP) from *P. fluorescens* to predict its molecular mechanism. Investigating the structure of CRP confirmed the presence of a conserved domain characteristic of the cold shock domain (CSD) family and a single nucleotide binding domain. When 3D structure of CRP was compared with the existing cold shock proteins, major deviations were found in the loop regions connecting the β2–β3, β3–β4 and β4–β5 sheets. Docking studies showed that CRP forms a significant clamp like structure at the substrate binding cleft which stabilizes the ligand. Therefore, it can be concluded that CRP has a strong affinity for the poly thymidine (poly T) stretch and can be considered a candidate for transcription regulation.

1. Introduction

All the groups of microorganisms, i.e. psychrophiles, mesophiles and thermophiles, synthesize cold shock proteins to counteract the effect of temperature downshift. The cold-shock response and cold shock proteins have been studied in detail using *Escherichia coli* and *Bacillus subtilis* as model system [1–3]. The first cold-shock protein CspA, was reported from *E. coli* and its homologues have been reported from both Grams positive and Gram negative bacteria. The CSP family of *E. coli* consists of nine homologus proteins, CspA to CspI, but among them only few are cold shock inducible.

Cold shock-induced proteins from a number of organisms, including *E. coli* [4], *B. subtilis* [5], *Thermotoga maritime*, *Streptococcus pyogenes*, *Mycobacterium tuberculosis*, *Pseudomonas aeruginosa*, *Pseudomonas fluorescens*, *Listeria monocytogene*, and *Yersinia pestis*, have been implicated in the adaptive processes required for cell viability at a low temperature [6]. The domain of CSP, involved in the nucleic acid binding is referred to as the cold shock domain (CSD). It consists of ∼70 amino acid residues and represented the prototype of CSD. CSPs are involved in a variety of important functions such as mRNA masking, coupling of transcription to translation and developmental timing and regulation [7,8]. Although a large amount of information has been accumulated, many features of the CSD protein family members are yet to be explored. Strikingly, the data that even the simplest members of the CSD super family, like the bacterial CSPs, have surprisingly diverse functions. Apart from its properties as a transcriptional activator [9], CspA of *E. coli* can destabilize RNA secondary structures [10] and possesses transcriptional antiterminator functions like its homologs, CspC and CspE [11].

Studies from Schindelin et al. [12], on the three-dimensional structures of CspA and CspB by 2D-NMR and X-ray revealed that the proteins surface, rich in aromatic and basic amino acid residues interact with nucleic acid. It was also found that these proteins contain sequence motifs typical of RNA-binding proteins. The interaction of CSP with single-stranded nucleic acid, has been verified experimentally by chemical shift perturbation analysis of complexes between *EcCspA* and ssDNA [13], *BsCspB* and ssDNA [14], and *TmCsp* and ssDNA and ssRNA, by ssRNA-binding gel-shift assays [10], tryptophan fluorescence quenching studies and site-directed mutagenesis [16,17]. An *in vitro* selection approach revealed the specific consensus sequences UUUUU, AGGGAGGGA and AAAUUU in *EcCspB, EcCspC* and *EcCspE*, respectively [18]. It has been found that thymine-based ssDNA templates containing 8 or more (dT) residues, bound with high affinity, and its binding is salt independent, whereas binding of cytosine-based

* Corresponding author at: Food Microbiology Department, CFTRI, Mysore, India.
E-mail address: mahejibin@gmail.com (M. Khan).

1093-3263/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmgm.2012.06.012

ssDNA templates is strongly salt dependent, indicating a large electrostatic component involved in the interactions. Zeeb et al. [14] reported that the crystal structure of *BsCspB*-TTTTTT shows aligned dimers. In contrast, the solution structure of *BsCspB* bound to a ssDNA heptathymidine fragment (dT7) displays a monomeric nucleoprotein complex, where the *BsCspB* conformation is almost invariant to oligonucleotide binding compared to the unbound conformation. Only minor reorientations were observed in loop β1–β2 and β3–β4 as well as of a few aromatic side chains involved in base stacking. Klass et al. [15] determined the crystal structure of *BsCspB* in complex with hexathymidine and heptathymidine and assessed that the affinity of *BsCspB* to oligonucleotides does not simply depend on the abundance of thymine in a ligand molecule but rather on preferential base binding to individual subsite. The largest preference for T was observed for sequence positions 2 and 6, C was preferred at position 3. Preferential binding of *BsCspB* to thymine or uracil containing sequences may be especially relevant for its function as antiterminator. Johnston et al. [19], studied the affinity of CspE for contiguous dT residues over the previously reported AT-rich sequences AATT, AATTT, AATTTT, AAATT, AAATTT, AAAATT and reported that CspE bind all of the oligonucleotides with much lower affinity than the dT-rich sequences.

Khan et al. [20] isolated cold resistant protein (CRP) from cold tolerant mutant of *P. fluorescens* and found that CRP is structurally similar to the major cold shock protein of *E. coli* CspA and contains an S1-like cold-shock domain that is also found in eukaryotes, prokaryotes, and archaea. CRP contained 10 ligand binding sites that preferentially bound poly-pyrimidine region of single-stranded RNA and DNA and therefore, thought to bind mRNA and regulate ribosomal translation, mRNA degradation, and the rate of transcription termination. This specific ssDNA-binding property of CSD was required for the binding of Y-box protein to the promoter's Y-box sequence, thereby regulating transcription.

For any given set of proteins, molecular docking is a widely used computational tool to study molecular recognition. If the compound binds, it is the spatial configuration which they adopt in their bound state and strength of interactions are strong or weak. It can also be of interest to determine if binding can be induced by mutation. Thus docking studies can aim to predict binding mode and binding affinity of a complex formed by two or more constituent molecules with known structures, estimate the binding mode for applications based on binding affinity of a ligand receptor complex. Even proteins of unknown function can be docked based on physical properties.

P. fluorescens, a member of family pseudomonadeaceae, demonstrates a great deal of metabolic diversity. The ability of this species to the co-inhabitant with plants in various niches affects the plant's growth in a positive way [21] and makes it an attractive target for this study. Although many species have been characterized that can survive in alpine regions, but no input has been made to study the mechanism of action of their survival in the cold afflicted regions [22]. In this study, we have proposed a feasible mechanism of the action of cold resistant protein for survival of *Pseudomonas* under cold stress conditions. The information generated here sheds light on the mechanism and functioning of CRP in *Pseudomonas* and it is possibly a necessity in the organism during the stress response.

2. Materials and methods

2.1. Sequence analysis

Protein sequence encoding CRP, from a cold tolerant *P. fluorescens* mutant (GenBank accession number: AAZ14100.1) was selected. Homologous protein sequences were retrieved by using protein Blast (http://www.ncbi.nlm.nih.gov/BLAST/) against Protein Data Bank (http://www.rcsb.org/pdb/home/homedo). The protein sequence of 1c9o (BC-CSP cold shock protein of *B. caldolyticus)*, 1csp (major cold shock protein of *B. Subtilis*), 1mjc (major cold shock protein of *E. coli*), 3i2z (cold shock protein E of *Salmonella typhimurium*) and 3cam (cold shock domain protein of *Neisseria meningitidis*) were retrieved from GENBANK database. Multiple sequence alignment for the obtained protein sequences was performed by ClustalW2 [23], to map the sequence conservation and variation in all the sequences. The alignment was represented by ESPript [24], a program which allows rapid visualization, via PostScript output of sequences aligned with ClustalW2.

2.2. Molecular modeling of CRP and structural comparison of three dimensional model of CRP

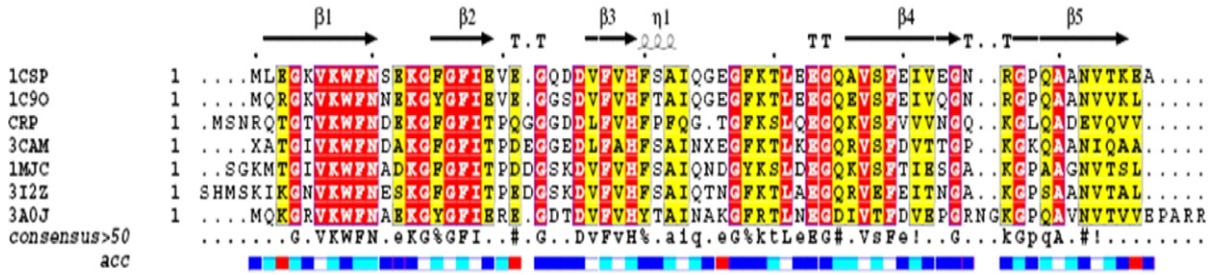
The CRP structure was modeled using the program MODELLER [25] that produced different conformations using three input files; the alignment file in PIR format, template coordinate file, and file containing sequence of the CSP to be modeled. MODELLER uses a new comparative modeling methodology to rapidly build structural models for protein sequences. Structural models for a sequence can be constructed from structures of one or more homologous protein in the same family. The stereochemistry of the theoretical model of CRP was done by PROCHECK [19] and RAMPAGE [26]. WHATIF [27], was used for verification and evaluation of the three dimensional structure of model or protein. Root mean square deviations were determined to contrast the difference in the domains of the protein using the program Pymol (http://www.pymol.org/pymol) and Superpose server (http://wishart.biology.ualberta.ca/SuperPose/). SuperPose server uniquely superimposes the structures that differ substantially in sequence, size or shape [28]. The template selected for comparison with CRP was 1csp and 1mjc. The disordered regions in the CRP, 1csp & 1mjc sequences have been predicted by using GlobPlot server (http://globplot.embl.de). The necessity of determining the disordered region of CRP arose due to the fact that disordered regions often contain motifs which aid in protein functioning [29].

2.3. Preparation of ligand structures

As reported earlier CRP binds with pyrimidine residues [19]. Therefore, in this study the ligands used were tetra- pyrimidine, penta- pyrimidine, and hexa- pyramidine. Their structures were prepared using Chemdraw [30]. The final structures were obtained in PDB format using the PRODRG server [31]. The small molecule topology generator PRODRG takes input from existing coordinates or various two-dimensional formats and automatically generates coordinates and molecular topologies suitable for X-ray refinement of protein–ligand complexes [31]. Ligands were constructed with explicit hydrogen atoms present and energy minimized using the 3D environment of RasMol 2.7.5. The best conformer obtained was based on energy minimization and geometry optimization.

2.4. Docking studies with potential ligand

Docking experiments were performed using the program HEX.5 [32] that employs spherical polar Fourier (SPF) correlations in conjunction with a soft molecular mechanics potential function, thus improving the rank obtained for low RMS docking orientations [33]. Relative stabilities were evaluated on the basis of free energy calculations done. The active site was defined based on shape and electrostatics post, docking processing was done using MM minimization. All the other parameters were kept as default.



PAIRWISE PERCENT (%) SEQUENCE IDENTITIES IN THE CSPs

	CRP	1mjc	1csp	3cam	3i2z	3a0j	1c90
CRP	100	56	49	59	52	39	51

Fig. 1. Multiple sequence alignment of the CRP with other cold shock proteins retrieved from NCBI-PDB database. Regions interacting with the ligands are enclosed in box. 1c9o: BC-CSP cold shock protein of *Bacillus caldolyticus*; 1csp: major cold shock protein of *Bacillus Subtilis;* 1mjc: major cold shock protein of *Escherichia coli*; 3i2z: cold shock protein E of *Salmonella typhimurium*; 3cam: cold shock domain protein of *Neisseria meningitidis*; CRP: a cold tolerant mutant of *Pseudomonas fluorescens;* 3a0j: cold shock protein 1 of *Thermus Thermophilus*. The pairwise alignment of the given PDBs is tabulated.

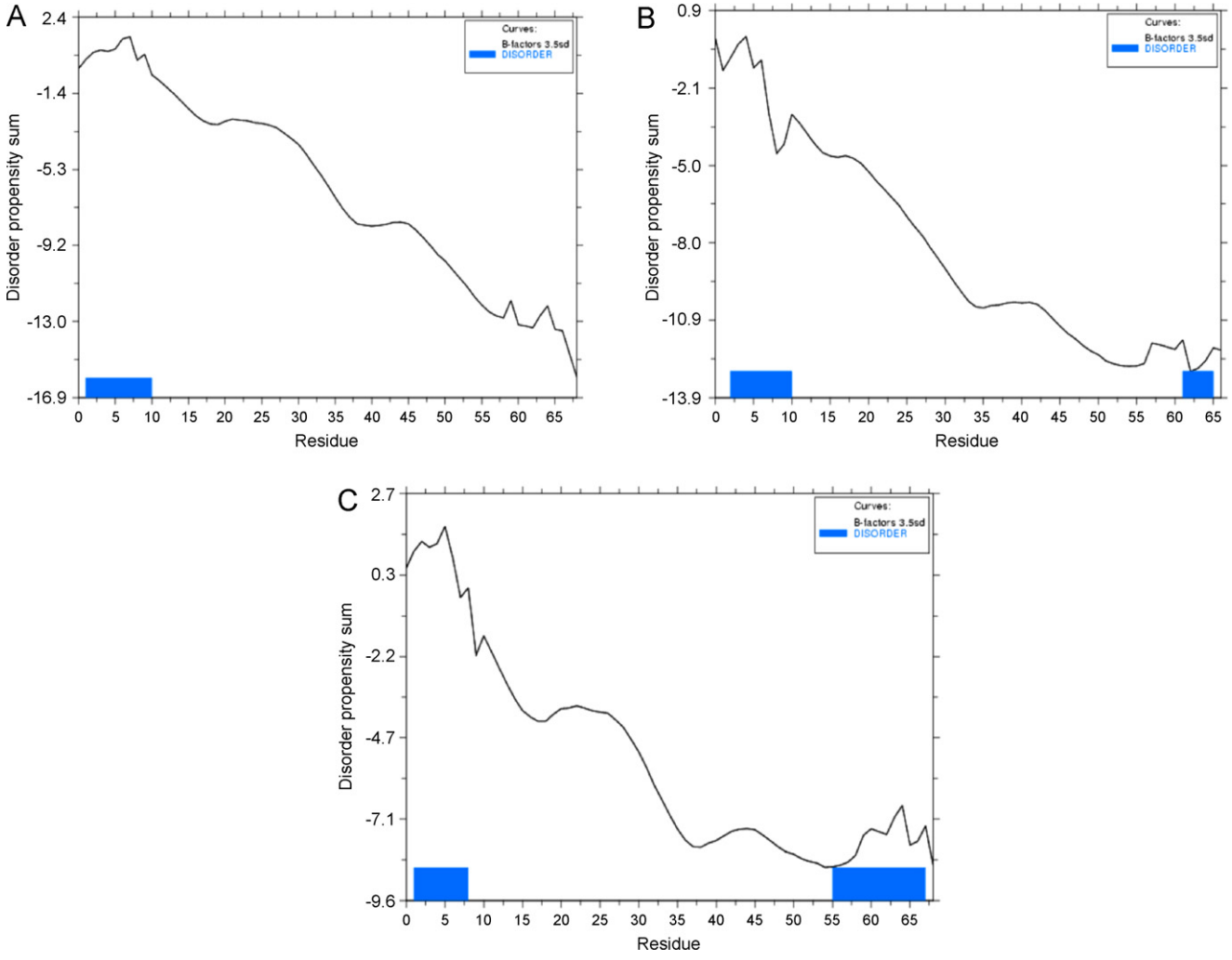


Fig. 2. The degree of disorder represented by a graphical representation of disorder propensity vs residue number for (A) CRP, (B) 1csp and (C) 1mjc. Images generated using the GLOBPLOT server. Disorder region of CRP: 1–10 residues (shown blue in color), disorder region of 1csp: 2–10 and 61–65 residues (shown blue in color), disorder region of 1mjc: 1–8 and 55–67 residues (shown blue in color). (For interpretation of references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Root mean square deviations in angstroms for the three dimensional structures of CSPs and CRP.

Global RMSD	CRP	1csp	CRP	1mjc
α Carbon	1.43	65	0.69	68
Backbone	1.37	260	0.63	272
All	1.88	443	0.92	459

3. Results and discussion

3.1. Sequence analysis

The homologous amino acid sequences of CRP were aligned according to Feng and Doolittle [34], as shown in Fig. 1.CRP showed 67%, 66%, 59%, 56%, 52% and 39% similarity with 1csp, 1c9o, 3cam, 1mjc, 3i2z and 3a0j respectively. ESPript output (Fig. 1) was obtained from sequences homologues to CRP. The overall secondary structure of the CRP was found to be conserved. Water accessibility of residues was also found based on the hydropathicity index of the protein. On the basis of sequence comparison it was observed that 70% of the residues present in the β-pleated were conserved. Residues that were critical for hydrophobic core formation in *E. coli* CSP have been identified earlier by three-dimensional structural analysis [35].

3.2. Structural comparison of three dimensional model of CRP

The superimposition of three-dimensional structures is an important method to evaluate the common 3D substructure of a set of molecules. The purpose of superimposition was to identify the stearic and electrostatic pattern of the ligand for which the receptors may have the same specificity. Comparison of crystal structures of the cold shock proteins from *E. coli* (1mjc) and *B. subtilis* (1csp) with CRP was performed using the Pymol suite and the back bones of three cold shock proteins were compared by determining their RMSD (root mean square deviation) values. RMSD values gave a direct indication of the dissimilarity of the protein in three dimensional spaces. The superimposition of CRP with 1mjc and 1csp ranged in RMSD values of 0.69–1.43. RMSD values for all Cα atoms indicating an overall conserved domains and a higher structural similarity with the template 1mjc. The RMSD values (Table 1) became more pronounced when all the atoms of the protein were considered. This indicated the possibility of conformational re-arrangements of the amino acids in three dimensional spaces. The GlobPlot server was used to determine the degree of disorder in the protein. The N terminal and C terminal end of the CSP protein are disordered whereas, an anomaly was found in case of CRP, it contained only one disordered region in the N terminal region as compared to 1csp and 1 mjc (Fig. 2).

The domain structure of CRP confirmed the presence of a conserved domain characteristic of the cold shock protein family. The conservation in the domain structure correlated with the sequence identity. Major deviations were found in the loop regions connecting the β2–β3, β3–β4 and β4–β5 sheets and the major fluctuation observed in three dimensional spaces was in loop 1 that plays an important role in binding of the substrate. Loops 2 and 3 of CRP structurally coincides with 1mjc (Fig. 3).

3.3. Electrostatic representation

The electrostatic surface representation was used to analyze the electrostatic tendency of the residues to form the substrate binding cleft. On the basis of comparison of surface charges of 1mjc, 1csp and CRP it was inferred that CRP had a uniform distribution of polar and hydrophobic residues across the binding cleft. All the three

Table 2
List of ligands with symbol for representation and interactive residues with docking energy.

S.no.	Symbol	Ligand	Interacting residues	Energy
1	A	d(CGCGTACCCG)	His33,Phe34,Pro35,Phe36,Gln37,Gly38,Thr39,Gly40,Lys42,Val55,Asn56,Gly57,Gln58,Gln62,Ala63,Asp6,Glu65, Val66	−460.42
2	B	d(GTGTACAC)	Trp11,Phe18,Gly27,Asp28,Asp29,Phe31,His33,Pro35, Phe36,Val55, Gly57,Gln58,Lys59,Gly60,Gln62, Ala63,Asp64,Glu65,Val66	−464.34
3	C	d(CCCCGCGG)	Trp11,Glu15,Lys16,Gly17,Phe18,Phe20,Asp28,Asp29,Phe31,His33,Gln58,Lys59,Gly60,Leu61	−386.80
4	D	d(CGCGCG)	Met1,Ser2,Asn4,Thr6,Gln24,Lys49,Val50,Ser51,Val65,Gln67, Val68, Val69	−353.71
5	E	d(CGCG)	Met1,Ser2,Asn3,Arg4,Gln5,Thr6,Pro23,Gln24,Gly25,Leu30, Phe52, Val54,Leu61	−328.83
6	F	d(TCGCGCGCCGA)	Asn3,Asp14,Glu15,Lys16,His33,Phe34,Pro35,Lys42, Val53, Val54, Val55, Asn56,Gln58,Gln62,Asp64, Glu65	−487.47
7	G	r(UCUUU)	Lys10,Trp11,Phe12,Asn13,Asp14,Glu15,Lys16,Gly17,Phe34,Gly40,Phe41,Lys42,Ser43,Gln45,Glu46	−586.33
8	H	d(CGGGTACCCG)	His33,Phe34,Pro35,Phe36,Gln37,Gly38,Thr39,Gly40,Lys42,Val55,Asn56,Gly57,Gln58,Gln62,Ala63, Asp64,Glu65,Val66	−467.08
9	I	r(UUUUU)	Asp14,Glu15,Gly40,Phe41,Lys42,Ser43	−364.31
10	J	d(TTTTTT)	Trp11,Phe12,Asn13,Asp14,Glu15,Lys16,Phe20,Phe31,His33,Phe34,Gly57,Gln58,Lys59,Gly60,Leu61,Gln62	−496.38
11	K	r(GUGGUCUCGAUGAGGCC)	Trp11,Phe12,Asn13,Asp14,Glu15,Lys16,Phe31,His33,Phe34, Pro35, Gly40, Phe41, Lys42, Ser43,Leu44, Gln45,Glu46, Val55, Asn56, Gln58, Gln62,Ala63,Asp64, Glu65	−939.32
12	L	r(GCUGCUGC)	Lys10,Trp11,Asp14,Glu15,Lys16,Gly17,Phe18,Gly19,Phe20,Ile21, Asp28,Asp29,Leu30,Phe31,His33,Phe34,Pro35,Gly57,Gln58, Lys59, Gly60,Leu61,Gln62	−720.43
13	M	d(TCGGCGCCGA)	Asn3,Asp14,Glu15,Lys16,His33,Phe34,Pro35,Lys42,Val53,Val54, Val55,Asn56,Gln58,Gln62,Asp64,Glu65	−491.34
14	N	r(AAAAAAA)	Lys10,Trp11,Phe12,Asn13,Asp14,Glu15,Phe20,Asp28,Asp29,Glu46	−662.25

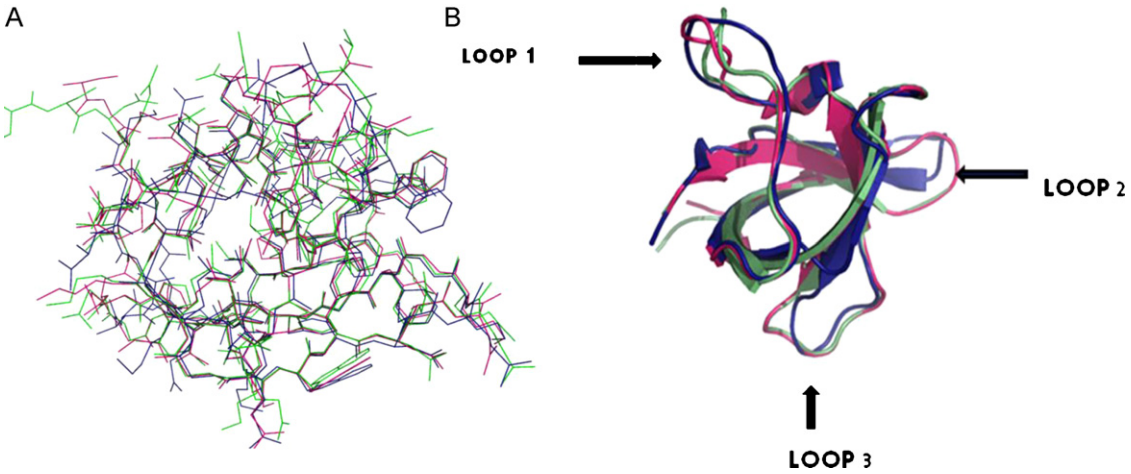


Fig. 3. (A) Line representation of CRP superimposed with 1CSP and 1MJC indicating the deviations of the residues in 3D space and (B) cartoon representation of superposed structures of 1mjc, 1csp and CRP with highly flexible loops of CRP being highlighted. The root mean deviations for the considered PDBs are tabulated in Table 1.

proteins 1mjc, 1csp and CRP form a clamp like arrangement near the substrate binding cleft (Fig. 4).

3.4. Substrate docking with the active site

Docking of the substrate to CRP was performed using HEX, to find the binding site for the polynucleotide stretch. The solutions structures for the possible confirmations were determined for each of the substrate, out of which the one with minimum energy and least stearic hindrance was selected for each of the substrate given in Table 2. The algorithm exhaustively searched the entire rotational and translational space of the ligands with respect to the receptor to generate ligand binding sites. The observations indicated that the protein forms a significant binding with ligand K (r(GUGGUCUGAUGAGGCC)) with the energy value of −939.32 kJ/mol and the substrate binding pocket was composed of a stretch of polar as well as few aromatic amino acids residues mainly phenylalanine and tryptophan. The residues constituting the ligand binding site were TRP11, PHE12, ASN13, ASP14, GLU15, LYS16, PHE31, HIS33, PHE34, PRO35, GLY40, PHE41, LYS42,

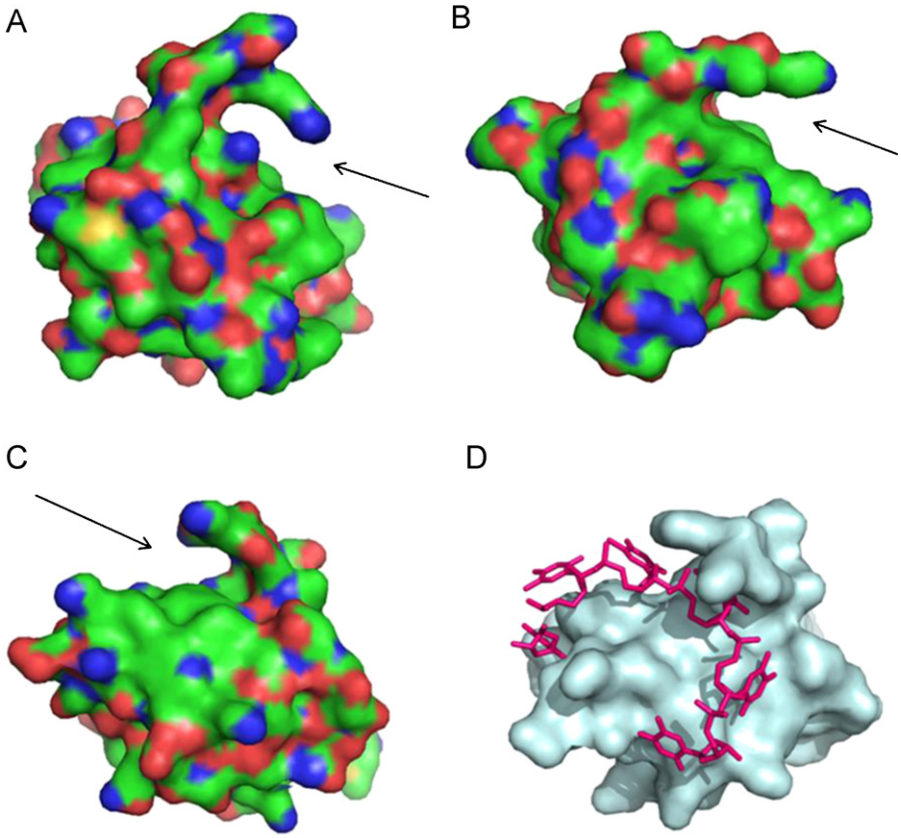


Fig. 4. Electrostatic surface potential for 1csp (A), 1mjc (B) and crp (C); the arrow indicates the significant clamp like arrangement in structures. (D) Surface diagram of CRP with ligand hexathymidine [in stick representation] in the substrate binding cleft. The polar regions of structure are shown in red and blue in color and green represented the hydrophobic region of the structure.

Table 3
Interaction of CRP mutants with substitution of leucine (L) and serine (S) residues at position 12, 18, 20, 31 and 34 of phenylalanine (F) residues with chain of oligonucleotide (GUGGUCUGAUGAGGCC) showing the energy values as *E*.

S.no.	Mutant residues	Interacting residues	Energy
1	FFFFL	Arg4,Lys10,Trp11,Phe12,Asn13,Asp14,Glu15,Lys16,His33,**Leu34**,Pro35,Gln37,Gly40, Phe41,Lys42,Ser43,Leu44,Gln45,Glu46,Val55,Gln58, Gln62,Ala63,Asp64,Glu65	−968.27
2	FFFLL	Arg4,Lys10,Trp11,**Phe12**,Asn13,Asp14,Glu15,Lys16,His33,**Leu34**, Pro35, Gln37, Gly40, Phe41,Lys42,Ser43,Leu44,Gln45,Glu46,Val55,Gln58, Gln62,Ala63,Asp64,Glu65	−968.27
3	FFLLL	Arg4,Lys10,Trp11,**Phe12**,Asn13,Asp14,Glu15,Lys16,His33,**Leu34**,Pro35,Gln37,Gly40, Phe41,Lys42,Ser43,Leu44,Gln45,Glu46,Val55,Gln58,Gln62, Ala63,Asp64,Glu65	−970.56
4	FLFLF	Lys10,Trp11,**Phe12**,Asn13,Asp14,Glu15,Lys16,His33,**Phe34**,Pro35,Gln37,Gly40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46,Val55,Gln58,Gln62,Ala63, Asp64,Glu65	−881.72
5	FLLLF	Lys10,Trp11,**Phe12**,Asn13,Asp14,Glu15,Lys16,Gly17,His33,**Phe34**,Pro35,Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46,Val55,Asn56,Gln58,Gln62, Ala63,Asp64,Glu65	−893.07
6	FLLLL	Lys10,Trp11,**Phe12**,Asn13,Asp14,Glu15,Lys16,Gly17,Leu18,His33,**Leu34**,Pro35, Gly40, Phe41, Lys42,Ser43,Leu44,Gln45,Glu46,Val55,Asn56,Gln58, Gln62, Asp64, Glu65	−953.45
7	LFFFF	Trp11,**Leu12**,Asn13,Asp14,Glu15,Lys16,Phe31,His33,**Phe34**,Pro35,Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46,Val55,Asn56,Gln58,Gln62,Ala63, Asp64,Glu65	−888.00
8	LFLFL	Arg4,Lys10,Trp11,**Leu12**,Asn13,Asp14,Glu15,Lys16,His33,**Leu34**,Pro35, Gln37, Gly40, Phe41, Lys42,Ser43,Leu44,Gln45,Glu46, Gln62,Ala63,Asp64,Glu65	**−982.79**
9	LLFFF	Lys10,Trp11,**Leu12**,Asn13,Asp14,Glu15,Lys16,Gly17,Phe31,His33,**Phe34**,Pro35,Glys40 Phe41,Lys42,Ser43,Leu44,Gln45,Glu46,Val55,Asn56,Gln58, Gln62, Ala63, Asp64,Glu65	−913.72
10	LLLFF	Lys10,Trp11,**Leu12**,Asn13,Asp14,Glu15,Lys16,Gly17,Phe31,His33,**Phe34**,Pro35, Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46,Val55,Asn56,Gln58,Gln62,Ala63, Asp64, Glu65	−904.84
11	LLLFL	Arg4,Lys10,Trp11,**Leu12**,Asn13,Asp14,Glu15,Lys16,Gly17,His33,**Leu34**,Pro35,Gln37, Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46,Val55,Gln58, Gln62,Ala63,Asp64,Glu65	−977.64
12	LLLLF	Lys10,Trp11,**Ser12**,Asn13,Asp14,Glu15,Lys16,Gly17,His33,**Phe34**,Pro35,Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46,Val55,Asn56,Gln58,Gln62, Ala63,Asp64,Glu65	−894.13
13	LLLLL	Lys10,Trp11,**Leu12**,Asn13,Asp14,Glu15,Lys16,Gly17,His33,**Leu34**,Pro35,Glys40,Phe4,Lys42,Ser43,Leu44,Gln45,Glu46,Val55,Asn56,Gln58,Gln62, Ala63,Asp64,Glu65	−968.37
14	FLFFF	Trp11,**Phe12**,Asn13,Asp14,Glu15,Lys16,Phe31,His33,Phe34,Pro35,Glys40,Phe41,Lys4,Ser43,Leu44,Gln45,Glu46,Val55,Asn56, Gln62, Ala63, Asp64, Glu65	−898.34
15	FFLFF	Trp11,**Phe12**,Asn13,Asp14,Glu15,Lys16,Phe31,His33,Phe34,Pro35,Glys40,Phe41,Lys4,Ser43,Leu44,Gln45,Glu46, Val55,Asn56, Gln62, Ala63, Asp64, Glu65	−881.44
16	FFFLF	Lys10,Trp11,**Phe12**,Asn13,Asp14,Glu15,Lys16,Gly17,His33,Phe34,Pro35,Gly40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46, Val55, Asn56, Gln58, Gln62, Asp64, Glu65	−863.26
17	LFLLL	Trp11,**Leu12**,Asn13,Asp14,Glu15,Lys16,His33,**Leu34**,Pro35,Gln37,Glys40,Phe41,Lys4,Ser43,Leu44,Gln45,Glu46, Val55, Gln58	−966.29
18	LLFLL	Trp11,**Leu12**,Asn13,Asp14,Glu15,Lys16,Gly17,His33,**Leu34**,Pro35,Glys40,Phe41,Lys4,Ser43,Leu44,Gln45,Glu46, Val55, Asn56,Gln58, Gln62, Ala63, Asp64, Glu65	−974.14
19	FFFFS	Arg4,Lys10,Trp11,**Phe12**,Asn13,Asp14,Glu15,Lys16,**Phe31**,His33,**Ser34**,Pro35,Gln37, Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46,Val55, Gln58, Gln62,Ala63,Asp64,Glu65	−956.36
20	FFFSF	Lys10,Trp11,**Phe12**,Asn13,Asp14,Glu15,Lys16,His33,Phe34,Pro35,Gln37,Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46, Val55, Gln58, Gln62, Ala63, Asp64,Glu65	−864.75
21	FFFSS	Arg4,Lys10,Trp11,**Phe12**,Asn13,Asp14,Glu15,Lys16,His33,**Ser34**,Pro35,Gln37,Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46, Val55, Gln58, Gln62, Ala63, Asp64,Glu65	**−961.79**
22	FFSFF	Trp11,**Phe12**,Asn13,Asp14,Glu15,Lys16,**Phe31**,His33,Phe34,Pro35,Glys40,Phe41,Lys4,Ser43,Leu44,Gln45,Glu46, Val55,Asn56, Gln62, Ala63, Asp64, Glu65	−881.88
23	FFSSS	Lys10,Trp11,**Phe12**,Asn13,Asp14,Glu15,Lys16,His33,**Ser34**,Pro35,Gln37,Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46,Val55, Gln58, Gln62, Ala63, Asp64, Glu65	−946.48
24	FSFFF	Trp11,**Phe12**,Asn13,Asp14,Glu15,Lys16,Gly17,**Phe31**,His33,Phe34,Pro35,Glys40,Phe4,Lys42,Ser43,Leu44,Gln45,Glu46, Val55,Asn56, Gln58, Gln62, Ala63,Asp64,Glu65	−−899.88
25	FSFSF	Lys10,Trp11,**Phe12**,Asn13,Asp14,Glu15,Lys16,His33,Phe34,Pro35,Gln37,Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46, Val55, Gln58, Gln62, Ala63, Asp64,Glu65	−957.03
26	FSSSF	Lys10,Trp11,**Phe12**,Asn13,Asp14,Glu15,Lys16,Gly17,His33,Phe34,Pro35,Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46, Val55, Gln58, Gln62, Ala63, Asp64, Glu65	−939.75
27	FSSSS	Lys10,Trp11,**Phe12**,Asn13,Asp14,Glu15,Lys16,Gly17,His33,**Ser34**,Pro35,Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46, Val55,Asn56,Gln58,Gln62, Ala63,Asp64,Glu65	−932.35
28	SFFFF	Trp11,**Ser12**,Asn13,Asp14,Glu15,Lys16,**Phe31**,His33,Phe34,Pro35,Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46, Val55, Asn56, Gln58, Gln62, Ala63, Asp64, Glu65	−884.11
29	SFFFS	Arg4,Lys10,Trp11,**Ser12**,Asn13,Asp14,Glu15,Lys16,**Phe31**,His33,**Ser34**,Pro35,Gln37, Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46, Val55, Gln58, Gln62, Ala63, Asp64, Glu65	−956.68
30	SFSFS	Arg4,Lys10,Trp11,**Ser12**,Asn13,Asp14,Glu15,Lys16,**Phe31**,His33,**Ser34**,Pro35,Gln37, Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46, Val55, Gln58, Gln62, Ala63, Asp64, Glu65	−948.76
31	SFSSS	Lys10,Trp11,**Ser12**,Asn13,Asp14,Glu15,Lys16,Gly17,His33,**Ser34**,Pro35,Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46, Val55, Asn56,, Gln58, Gln62, Ala63, Asp64,Glu65	−940.36
32	SSFFF	Lys10,Trp11,**Ser12**,Asn13,Asp14,Glu15,Lys16,Gly17,**Phe31**,His33,Phe34,Pro35,Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46, Gln62, Ala63, Asp64, Glu65	−889.58
33	SSFSS	Lys10,Trp11,**Ser12**,Asn13,Asp14,Glu15,Lys16,Gly17,His33,**Ser34**,Pro35,Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46, Val55, Asn56,, Gln58, Gln62,Ala63,Asp64,Glu65	−936.63
34	SSSFF	Lys10,Trp11,**Ser12**,Asn13,Asp14,Glu15,Lys16,Gly17,**Phe31**,His33,**Phe34**,Pro35,Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46, Val55, Asn56, Gln58, Gln62, Ala63, Asp64, Glu65	−886.83
35	SSSFS	Lys10,Trp11,**Ser12**,Asn13,Asp14,Glu15,Lys16,Gly17,His33,**Ser34**,Pro35,Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46, Val55, Asn56, Gln58, Gln62,Ala63,Asp64,Glu65	−956.07
36	SSSSF	Lys10,Trp11,**Ser12**,Asn13,Asp14,Glu15,Lys16,Gly17,His33,**Phe34**,Pro35,Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46, Val55, Gln58, Gln62, Ala63, Asp64, Glu65	−865.02
37	SSSSS	Lys10,Trp11,**Ser12**,Asn13,Asp14,Glu15,Lys16,Gly17,His33,**Ser34**,Pro35,Glys40,Phe41,Lys42,Ser43,Leu44,Gln45,Glu46, Val55, Gln58, Gln62, Ala63, Asp64,Glu65	−933.20

The most crucial residues in both the Leucine and Serine mutant series are highlighted as bold.

SER43, LEU44, GLN45, GLU46,VAL55, ASN56, GLN58, GLN62, ALA63, ASP64,GLU65 (Fig. 5). It was observed that the hydrophobic aromatic amino acids were critically placed at the positions such that they enable steady π–π stacking interactions. Comparison of the CRP with CSPs from various microorganisms revealed that most of the amino acid residues were conserved except at 14th, 28th, 45th and 57th position corresponding to CRP (Fig. 1). It was evident from the data obtained that the ligand was located in stable interaction with these residues via hydrogen bonding. It has already been reported that the "under stress phase variations" occurs in micro-organisms resulting in the formation of poly T stretches in their genes. Phase variation is a specific adaptive mechanism that typically involves a rapid switch between two different states (ON and OFF) to enable individual microbial cells to vary expression of proteins. Bacterial pathogens such as *E. coli*, *N. meningitides* and *H. pylori* are the examples of organisms where phase variation affecting the expression of flagella, fimbriae and outer membrane proteins. Phase variation has also been observed in a number of non-pathogenic bacteria including *G. stearothermophilus*, *A. ferroxidans*, and *B. subtilis* [36].

3.5. Mutational study

The correlation between ligand binding and protein stability for CRP was assessed by mutational studies. Five residues (12, 18, 20, 31 and 34) in CRP were substituted with leucine and serine, both singly and in combination. These mutants were identified by five sequential one-letter codes designating the identity of residues at position 12, 18, 20, 31 and 34 respectively. For example, non-mutant CRP is denoted as FFFFF and double mutant with Leucine substitution at position 20 and 31 are denoted as FFLLF, like same if double mutant with Serine substitution at position 18 and 20 is denoted as FSSFF. We have investigated the affinity of all substitution mutants for single stranded r(GUGGUCUGAUGAGGCC) by

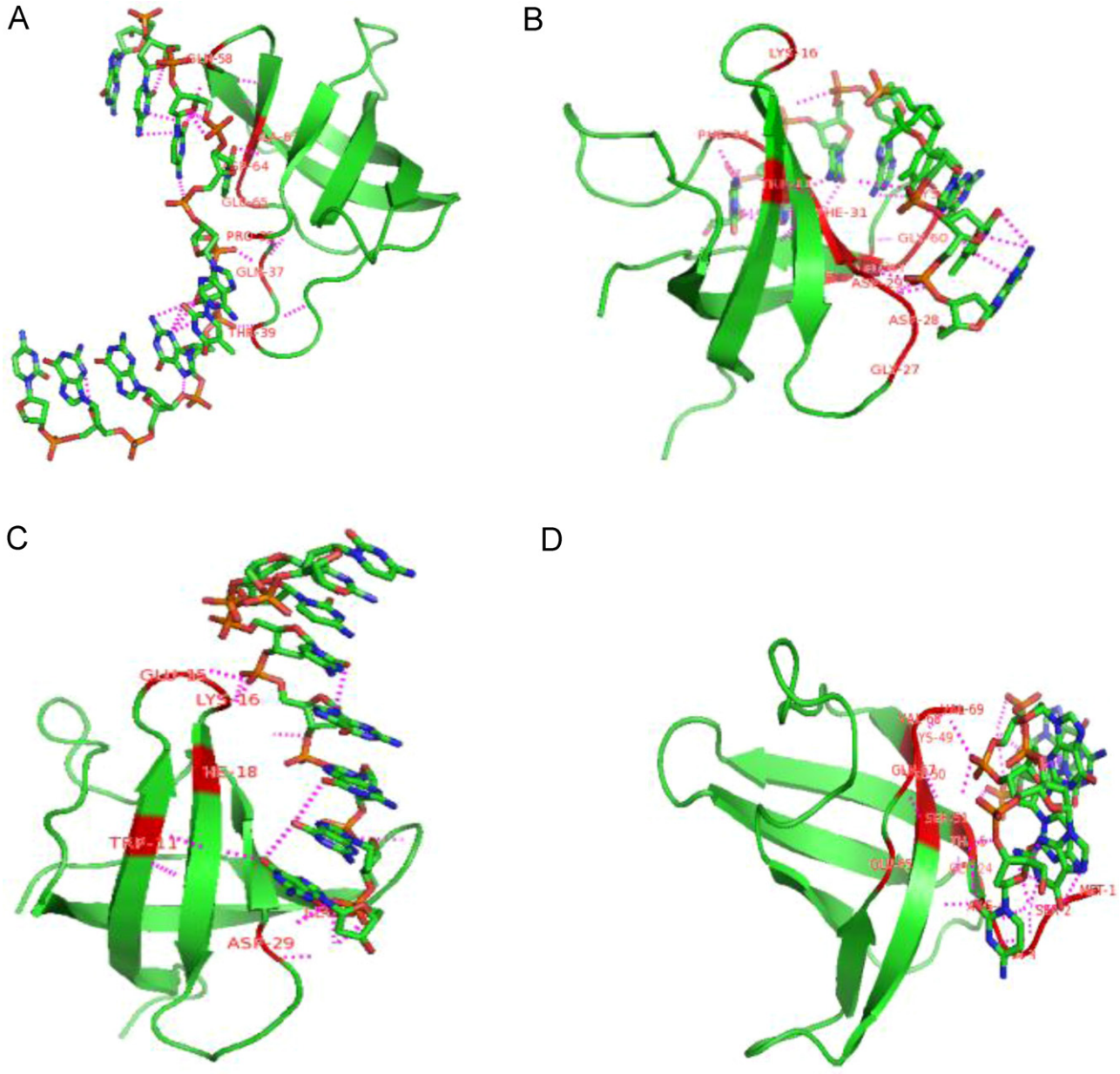


Fig. 5. Binding of ligands (A–N) to the CRP. Ligands are represented as sticks and proteins are represented as cartoon. All polar contacts are represented by magenta color and interactive residues by red color shown by pymol viewer. (For interpretation of references to color in this figure legend, the reader is referred to the web version of this article.)

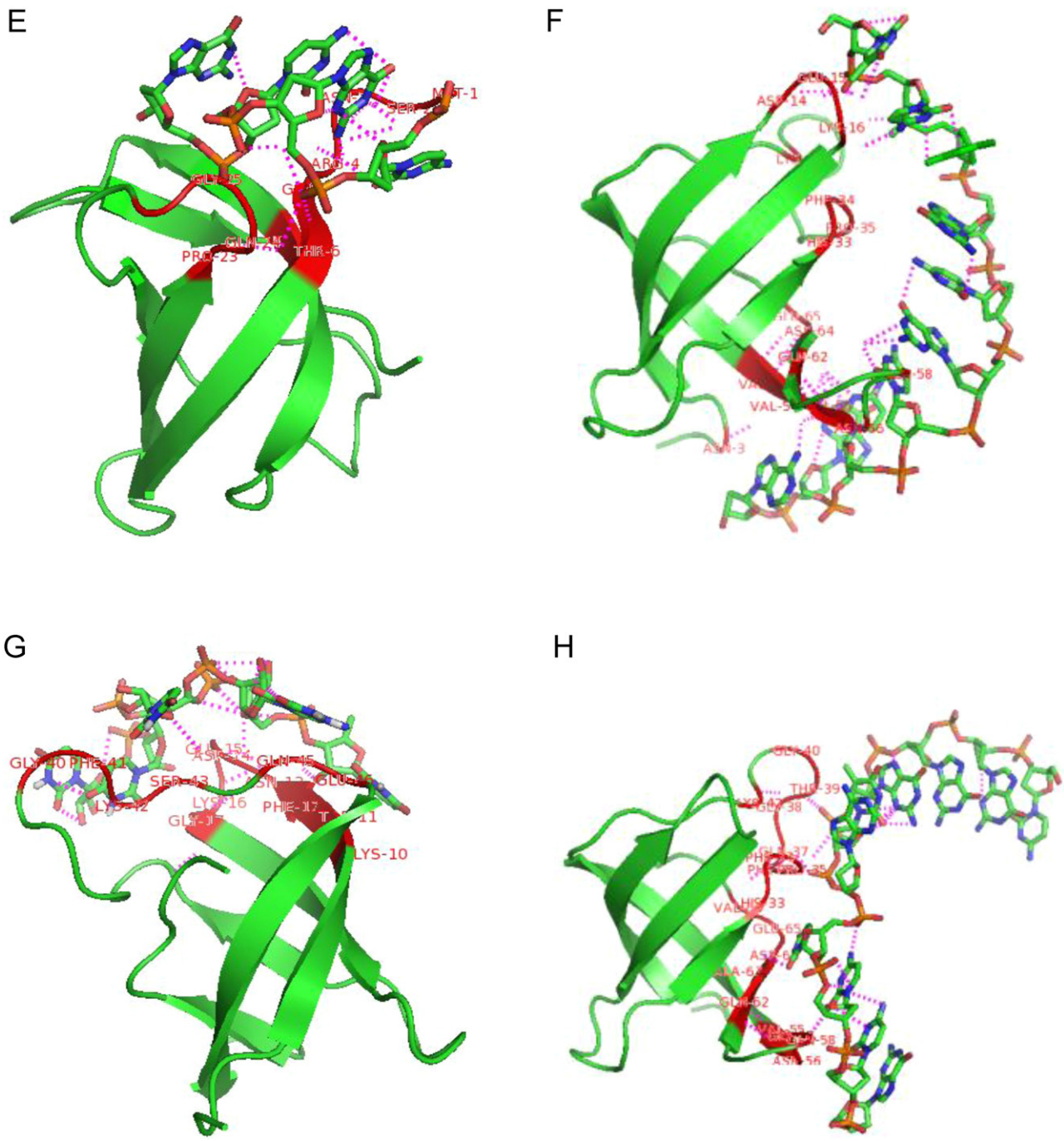


Fig. 5. (*continued*).

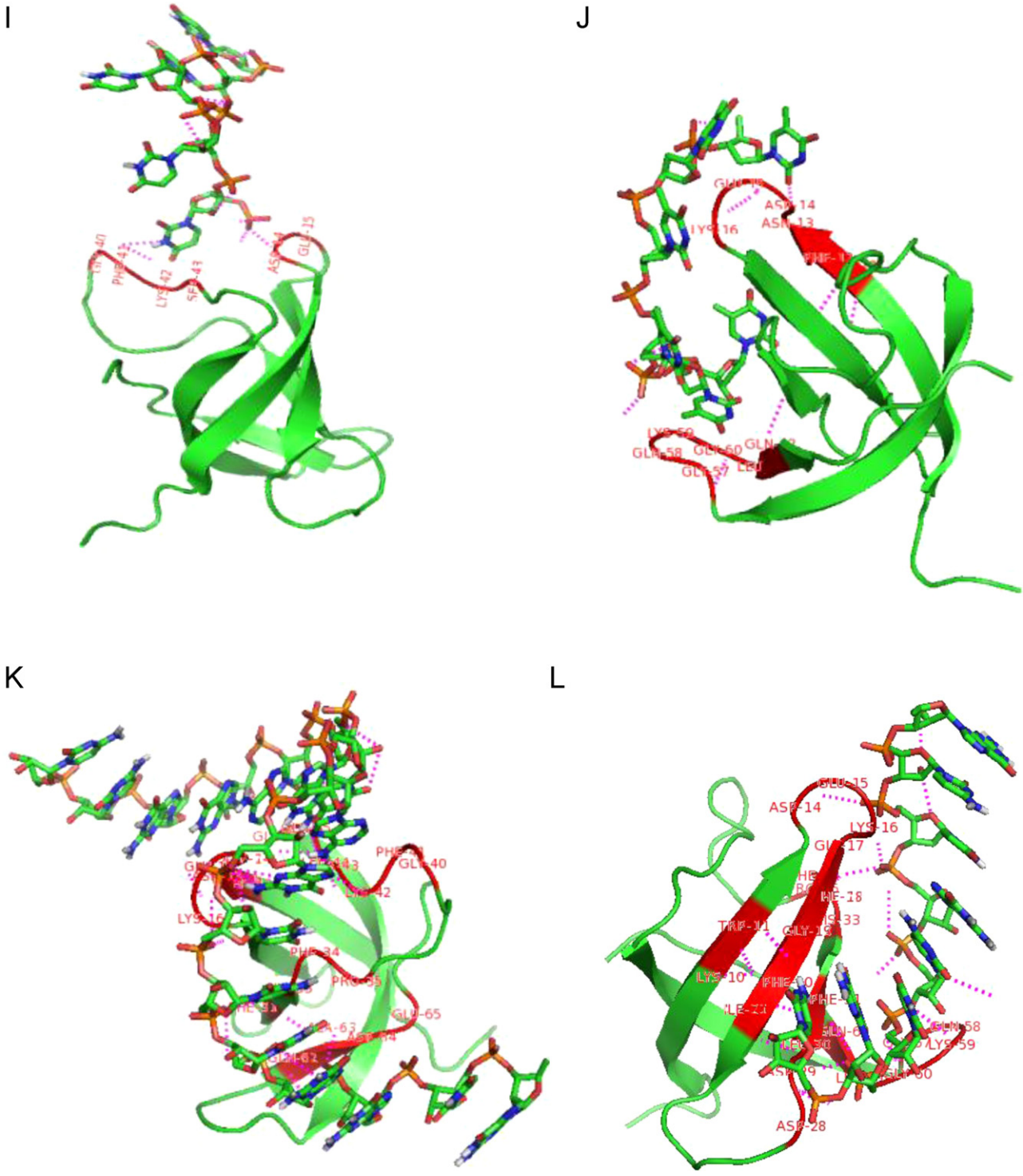


Fig. 5. (*continued*).

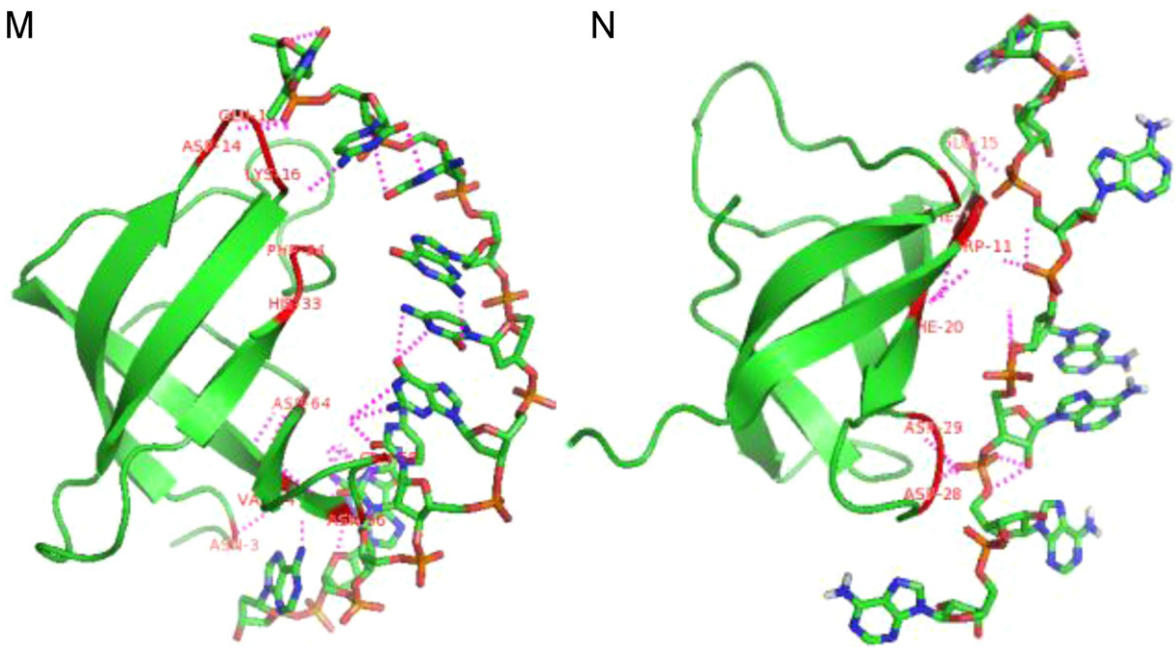


Fig. 5. (*continued*).

using HEX for docking study. The energy score of CRP mutants with single ribonucleotide is shown in Table 3. Mutant containing Leucine substitutions (LFLFL) has higher affinity for the oligonucleotide chain than Serine substitutions, suggesting that non-polar interactions contribute to binding. Phe12 and Phe34 were found to be the most crucial in both the Leucine and Serine mutant series.

In mutational studies, Leucine and Serine were selected because Leucine is a fairly large non-polar amino acid and the Phe → Leu mutation should destroys favorable aromatic interactions both within the protein and, potentially, between the protein and the bases of a single-stranded nucleic acid molecule [16]. Serine is a small and polar amino acid, quite different from phenylalanine. Although phenylalanine is one of the best β-sheet forming residues, Serine is still a reasonably good β-sheet former [37,38]. Mutation to Serine reduced the non-polar area of the surface patch, but not entirely removes the β-forming propensity. To investigate the role of the five phenylalanine aromatic residues as a whole, all the possible combinations of mutations were made (Table 3). A single Serine substitution may be destabilizing because it introduces a polar group among aromatic side chains, but multiple Serine substitutions may overcome this destabilization by interacting favorably with each other, as predicted by a study of residue-residue interactions in β sheets [39], or by interacting favorably with the solvent [16]. From the mutational studies, it was found that the solvent-exposed aromatic cluster in CRP is important for both protein stability as well as for oligonucleotide binding. Similar results were also reported in *E. coil* major cold shock protein CspA [40–42]. The surface of CspA has one very large, continuous non-polar patch of six aromatic residues. Stability of CspA was explained by "stability-function hypothesis" [43,44]. This theory states that the protein stability is not maximized, but rather that there is a balance between stability and function and residues which contribute to ligand binding or catalytic activity may be suboptimal for stability.

4. Conclusion

The CRP of *P. fluorescens* was found to have a single nucleotide binding domain based on the disorderliness of the residues. The protein forms a significant clamp like structure at the substrate binding cleft which stabilizes the ligand. Appropriate placements of the polar and aromatic residues have enabled the protein to interact suitably with the substrate and the aromatic cluster serves dual purpose in stabilizing the β-sheet structure of the protein and in nucleic acids binding. Mutational studies confirms that mutant containing leucine substitutions has higher affinity for the oligonucleotide chain than serine substitutions, and Phe12 and Phe34 are the most crucial in both the leucine and serine mutant. Based on all observations we can conclude that CRP has a strong affinity for the nucleotide stretch consisting mainly of thymidine and can be considered as a candidate for transcription regulation.

References

[1] D.N. Ermolenko, G.I. Makhatadze, Bacterial cold-shock proteins, Cellular and Molecular Life Sciences 59 (2002) 1902–1913.
[2] S. Phadtare, M. Inouye, Sequence-selective interactions with RNA by CspB, CspC and CspE, members of the CspA family of *Escherichia coli*, Molecular Microbiology 33 (1999) 1004–1014.
[3] S. Phadtare, K. Yamanaka, M. Inouye, The cold shock response, in: G. Storz, R. Hengge-Aronis (Eds.), The bacterial stress responses, ASM Press, Washington DC, 2000, pp. 33–45.
[4] J. Goldstein, N.S. Pollitt, M. Inouye, Major cold shock protein of *Escherichia coli*, Proceedings of the National Academy of Sciences of the United States of America 87 (1990) 283–287.
[5] G. Willimsky, H. Bang, G. Fischer, M.A. Maraheil, Characterization of CspB, a *Bacillus subtilis* inducible cold shock gene affecting cell viability at low temperature, Journal of Bacteriology 174 (1992) 6326–6335.
[6] P.G. Jones, M. Inouye, The cold-shock response—a hot topic, Molecular Microbiology 11 (1994) 811–818.
[7] G. Sachetto-Martins, L.O. Franco, D.E. de Oliveria, Plant glycine-rich proteins: a family or just proteins with a comman motif, Biochimica et Biophysica Acta 1492 (2000) 1–14.
[8] J. Sommervilli, Activities of cold-shock domain proteins in translation control, Bioessays 21 (1999) 319–325.
[9] A. Brandi, C.L. Pon, C.O. Gualerzi, Interaction of the main cold shock protein CS7.4 (CspA) of *Escherichia coli* with the promoter region of hns, Biochimie 76 (1994) 1090–1098.
[10] W. Jiang, Y. Hou, M. Inouye, CspA, the major cold shock protein of *E. coli*, is an RNA chaperon, Journal of Biological Chemistry 272 (1997) 196–202.
[11] W. Bae, B. Xia, M. Inouye, K. Severinov, *Escherichia coli* CspA-family RNA chaperons are transcription antiterminators, Proceedings of the National Academy of Sciences of the United States of America 97 (2000) 7784–7789.
[12] H.Schindelin, A.M. Marahiel, U. Heinemann, Universal nucleic acid binding domain revealed by crystal structure of the *Bacillus subtilis* major cold shock protein, Nature 364 (1993) 164–168.

[13] K. Newkirk, W. Feng, W. Jiang, R. Tejero, S.D. Emerson, M. Inouye, G.T. Montelione, Solution NMR structure of the major cold shock protein (CspA) from *Escherichia coli*: identification of a binding epitope for DNA, Proceedings of the National Academy of Sciences of the United States of America 91 (1994) 5114–5118.
[14] M. Zeeb, K.E.A. Max, U. Weininger, C. Lçw, H. Sticht, J. Balbach, Recognition of T-rich single-stranded DNA by the cold shock protein Bs-CspB in solution, Nucleic Acids Research 34 (2006) 4561–4571.
[15] E.M. Klass, M. Zeeb, R. Bienert, J. Balbach, U. Heinemann, T-rich DNA single strands bind to a preformed site on the bacterial cold shock protein Bs-CspB, Journal of Molecular Biology 360 (2006) 702–714.
[16] B.J. Hillier, H.M. Rodriguez, L.M. Gregoret, Coupling protein stability and protein function in *Escherichia coli* CspA, Folding Des 3 (1998) 87–93.
[17] K. Schroder, P. Graumann, A. Schnuchel, T.A. Holak, M.A. Marahiel, Mutational analysis of the putative nucleic acid-binding surface of the cold-shock domain, CspB, revealed an essential role of aromatic and basic residues in binding of single-stranded DNA containing the Y-box motif, Molecular Microbiology 16 (1995) 699–708.
[18] S. Phadtare, M. Inouye, Sequence-selective interactions with RNA by CspB, CspC and CspE, members of the CspA family of *Escherichia Coli*, Molecular Microbiology 33 (1999) 1004–1014.
[19] D. Johnston, C. Tavano, Sue Wickner, N. Trun, Specificity of DNA binding and dimerization by CspE from *Escherichia coli*, The Journal of Biological Chemistry 281 (2006) 40208–40215.
[20] M.Khan, A. Kumar, R. Goel, Comparative protein modeling, prediction of conserved residue and active sites in cold resistant protein isolated from CRPF1, a cold tolerant mutant of *Pseudomonas fluorescens*, Current Microbiology 60 (2010) 428–434.
[21] L.S. Thomashow, D.M. Weller, Role of a phenazine antibiotic from *Pseudomonas fluorescens* in biological control of *Gaeumannomyces graminis* var. tritici, Journal of Bacteriology 170 (8) (1988) 3499–3508.
[22] E.C. Morris, L.L. Kinkel, K. Xiao, P. Prior, C.D. Sands, Surprising niches for the plant pathogen *Pseudomonas syringae*, Infection Genetics and Evolution 7 (2007) 84–92.
[23] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, ClustalW and ClustalX version 2, Bioinformatics 23 (2007) 2947–2948.
[24] P. Gouet, E. Courcelle, D.I. Stuart, F. Metoz, ESPript: multiple sequence alignments in PostScript, Bioinformatics 15 (1999) 305–308.
[25] A. Fiser, R.K. Do, A. Sali, Modeling of loops in protein structures, Protein Science 9 (2000) 1753–1773.
[26] K. Arnold, L. Bordoli, J. Kopp, T. Schwede, The SWISS-MODEL Workspace: a web-based environment for protein structure homology modelling, Bioinformatics 22 (2006) 195–201.
[27] G. Vriend, WHAT IF. A molecular modeling and drug design program, Journal of Molecular Graphics 8 (1990) 52–56.
[28] M. Rajarshi, G.H. Van Domselaar, H. Zhang, D.S. Wishart, Superpose: a simple server for sophisticated structural superposition, Nucleic Acids Research 32 (2004) 590–594.
[29] R. Linding, R.B. Russell, N.Victor, T.J. Gibson, GlobPlot: exploring protein sequences for globularity and disorder, Nucleic Acid Research 31 (2003) 3701–3708.
[30] N. Mills, ChemDraw Ultra 10.0, Journal of the American Chemical Society 128 (2006) 13649–13650.
[31] A.W. Schuttelkopf, M.F. Daan van, Aalten PRODRG: a tool for high-throughput crystallography of protein–ligand complexes, Acta Crystallographica D60 (2004) 1355–1363.
[32] G. Macindoe, L. Mavridis, V. Venkatraman, M.-D. Devignes, D.W. Ritchie, HexSever: an FFT-based protein docking server powered by graphics processors, Nucleic Acids Research 38 (2010) W445–W449.
[33] D.W. Ritchi, D.Kozakov, S. Vajda Accelerating, Focusing protein–protein docking correlation using multi-dimensional rotational FFT generating functions, Bioinformatics 24 (2008) 1865–1873.
[34] D. Feng, R.F. Doolittle, Progressive alignment and phylogenetic tree construction of protein sequences, Methods in Enzymology 183 (1990) 375–387.
[35] K. Yamanaka, L.Fang, M. Inouye, The CspA family in *Escherichia coli:* multiple gene duplication for stress adaptation, Molecular Microbiology 27 (1998) 247–255.
[36] T.M. Wassenaar, J.A. Wagenaar, A. Rigter, C. Fearnley, D.G. Newell, B. Duim, Homonucleotide stretches in chromosomal DNA of *Campylobacter jejuni* display high frequency polymorphism as detected by direct PCR analysis, FEMS Microbiology Letters 212 (2002) 77–85.
[37] D.L. Minor, P.S. Kim, Measurement of the beta-sheet forming properties of amino acid, Nature 367 (1994) 660–663.
[38] C.K. Smith, J.M. Withka, L. Regan, A thermodynamic scale for β-sheet forming tendencies of the amino acids, Biochemistry 33 (1994) 5510–5517.
[39] M.A.Wouters, P.M.G. Curmi, An analysis of side chain interactions and pair correlation within antiparallel β-sheets: the difference between backbone hydrogen-bonded and non-hydrogen-bonded residues pair, Protein 22 (1995) 119–131.
[40] D. Landsman, RNP-1, an RNA-binding motif is conserved in the DNA-binding cold shock domain, Nucleic Acid Research 20 (1992) 2861–2864.
[41] K. Newkirk, W. Feng, W. Jiang, R. Tejero, S.D. Emerson, M. Inouye, G.T. Montelione, Solution NMR structure of the major cold shock protein (CspA) form *Escherichia coli*: identification of a binding epitope for DNA, Proceedings of the National Academy of Sciences of the United States of America 91 (1994) 5114–5118.
[42] K. Shroeder, P. Graumann, A. Schnuchel, T.A. Holak, M.A. Marahiel, Mutational analysis of the putative nucleic-acid binding surface of the cold-shock domain, CspB, revealed an essential role of aromatic and basic residues in binding of single-stranded DNA containing the Y-box motif, Molecular Microbiology 16 (1995) 699–708.
[43] G. Schreiber, A.M. Buckle, A.R. Fersht, Stability and functions: two constraints in the evolution of barstar and other proteins, Structure 2 (1994) 945–951.
[44] B.K. Shoichet, W.A. Baase, R. Kuroki, B.W. Matthews, A relationship between protein stability and protein function, Proceedings of the National Academy of Sciences of the United States of America 92 (1995) 452–456.