IJPSR (2020), Volume 11, Issue 6 (Research Article)

INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES AND RESEARCH

Received on 12 July 2019; received in revised form, 09 December 2020; accepted, 17 April 2020; published 01 June 2020

RELATIVE STRUCTURAL ANALYSIS OF LytTR DOMAIN OF AgrA PROTEIN INVOLVED IN BACTERIAL QUORUM SENSING

Hriday Kr. Basak [1], Abhik Chatterjee [1] and Ayon Pal [*2]

Department of Chemistry [1], Department of Botany (Microbiology & Computational Biology Laboratory) [2], Raiganj University, Raiganj - 733134, West Bengal, India.

Keywords:

Drug resistance, Quorum sensing, AgrA protein, Homology modeling, LytTR domain

Correspondence to Author:
Dr. Ayon Pal

Associate Professor, Microbiology & Computational Biology Laboratory, Department of Botany, Raiganj University, Raiganj - 733134, West Bengal, India.

E-mail: ayonpal.ruc@gmail.com

ABSTRACT: Resistance to antimicrobials by pathogenic bacteria is a global menace. An attractive pathway to resolve this problem of resistance may be achieved by targeting the bacterial quorum sensing (QS) process. Fighting bacterial infections by interfering with their command language and disrupting virulence expression could serve as a viable alternative. One of the current strategies is to attenuate virulence gene expression of *agr* quorum sensing system of some Gram-positive bacteria through inhibition of AgrA_ P2/P3 interaction. Tertiary structures of AgrA proteins of some pathogenic bacteria like *Listeria monocytogenes, Chlamydia trachomatis, Enterococcus faecalis, Streptococcus pyogenes* and *Macrococcus canis* showing antimicrobial resistance are not yet available in structural databases like PDB. These proteins play a key role in AgrA quorum sensing. 3D structures of these proteins are essential to determine most of their functions, and for the development of agents that can be explored as therapeutic agents. In order to derive structures of these AgrA proteins, homology modeling approach was employed. The modeled proteins were validated by several methods, and the physico-chemical properties of these refined protein structures were predicted using *in-silico* methods. The compositional and structural differences of the LytTR domain of AgrA proteins between the pathogenic and non-pathogenic bacteria were analyzed in our study. We have also identified the cleft and cavities in these structures and have explored the potential binding sites. The structural features of these sites have also been studied to have a better understanding of screening/designing inhibitors.

INTRODUCTION: Nowadays, the treatment of bacterial infections such as bacteremia, urinary tract infections, infective endocarditis, skin and soft tissue infections, pneumonia, toxic shock syndrome, septic arthritis, and osteomyelitis is a major global challenge.

QUICK RESPONSE CODE

DOI: 10.13040/IJPSR.0975-8232.11(6).2828-39

This article can be accessed online on www.ijpsr.com

DOI link: http://dx.doi.org/10.13040/IJPSR.0975-8232.11(6).2828-39

Bacteria are continuously developing resistance to current antibacterial compounds, and the problem is becoming more wide-spread [1-4] in third world countries where medical services are not always up to the mark, and monitoring drug abuse is quite impossible [5]. It is estimated that bacterial resistance can increase mortality and morbidity by two-fold [6]. It has been discovered that a large number of bacterial species produce many virulence factors and form biofilms, which are regulated by the cell to cell communication process called quorum sensing (QS) [7-9]. An attractive pathway to resolve the problem of bacterial resistance to current antibacterial agents is targeting bacterial QS [10-14].

Fighting bacterial infections by interfering with their command language or QS and disrupting virulence expression could serve as an alternative way to inhibit growth [15-18]. The accessory gene regulator (*agr*), a well-studied QS system, controls the expression of a series of virulence-associated protein genes in some Gram-positive bacteria [19, 20]. At the *agr* locus, there are two divergent transcription units driven by promoters *agr*P2 and *agrP3*. *agr*P2 drives the synthesis of RNA II, which encodes Agr ABCD, the structural components of the QS system, while *agr*P3 leads to the synthesis of RNA III which encodes delta–hemolysin [20] but also acts as a regulatory RNA that controls the expression of a series of virulence genes transcriptionally or translationally [21]. One of the current strategies to attenuate virulence gene expression of *agr* QS system of Gram-positive bacteria is through inhibition of AgrA_ P2/P3 interactions. The availability of good quality tertiary structure of LytTR domain of AgrA protein is necessary for the development of such therapeutic agents. Although the number of AgrA protein sequences has increased exponentially, the number of experimentally determined protein structures lags far behind. This is a general concern since the methods for determination of three dimensional structure of a protein are time consuming and expansive.

This predicament can be overcome by exploiting tertiary protein structure prediction using the approach of homology modeling, which has the ability to determine protein structure from sequence data with an accuracy that is comparable to experimental methods. Homology modeling relies on the fact that during evolution, homologous sequences tend to have conserved structures and sequences having identity, less than 30% can have different structure [22]. The 3D structure of a given protein sequence can be predicted using homology modeling provided an X-ray or NMR structure is available based on an alignment with one or more identified protein structure [23]. In this study, a comprehensive *in-silico* analysis and homology modeling studies on LytTR domain of AgrA proteins of some pathogenic and non-pathogenic bacteria have been performed. The non-pathogenic bacteria have been included in this study to resolve the structural differences present, if any, within the bacteria having significantly different lifestyles. 3D structures of DNA binding LytTR domain of these proteins are not yet available.

MATERIALS AND METHODS:

Data Collection: LytTR domain of AgrA protein sequences of five pathogenic bacteria and five non-pathogenic bacteria given in **Table 1** were retrieved from the GenBank database of NCBI (National Centre for Biotechnology Information) (http://www.ncbi.nlm.nih.gov/), a freely accessible resource of protein sequences and functional information.

TABLE 1: THE PROTEIN SEQUENCE RETRIEVED FROM THE GenBank DATABASE OF NCBI

	Gene name	Organism	Accession number of the target protein	Length of the modelled protein (LytTR domain)
Pathogenic bacteria	Accessory gene regulator protein A	*Listeria monocytogenes* SLCC2378	CBY71828.1	1-102=102
	Accessory gene regulator protein A	*Chlamydia trachomatis*	CRH93505.1	148-237=90
	Accessory gene regulator protein A	*Enterococcus faecalis* Fly1	EEU78551.1	145-242=98
	DNA-binding response regulator	*Streptococcus pyogenes*	WP_014635338.1	79-168=90
	DNA-binding response regulator	*Macrococcus canis*	WP_086043251.1	141-238=98
Non-pathogenic bacteria	Accessory gene regulator protein A	*Lysinibacillus sphaericus* C3-41	ACA41473.1	116-212=97
	DNA-binding response regulator	*Eubacterium plexicaudatum*	WP_004058878.1	141-239=99
	DNA-binding response regulator	*Carnobacterium gallinarum*	WP_034560381.1	148-238=91
	DNA-binding response regulator	*Floricoccus penangensis*	WP_070788284.1	139-236=98
	DNA-binding response regulator	*Paenibacillus typhae*	WP_090712524.1	141-238=98

Physico-chemical Characteristics: Different physicochemical properties such as amino acid composition, molecular weight, number of positive and negative ions, theoretical isoelectric point (pI), half-life, aliphatic index, instability index, extinction coefficient, grand average hydropathy (GRAVY) associated with the primary structure of LytTR domain of AgrA proteins were predicted using Expasy's ProtParam server (http://web.expasy.org/protparam/). Protein pI is calculated using pKa values of its amino acids. The pKa value of amino acids depends on their side chain and has a crucial role in defining the pH-dependent characteristics of a protein. The extinction coefficient serves as an indicator of how much light a protein absorbs at a certain wavelength. It is possible to calculate the molar extinction coefficient of a protein from the amino acid composition [24]. The stability of a protein in a test tube may be estimated using the instability index, where a value of less than 40 indicates a stable protein [25]. The aliphatic index is defined as the relative volume occupied by aliphatic side chains of a protein, which include alanine, valine, isoleucine, and leucine and contributes to the thermostability of globular protein [26]. The GRAVY value for a peptide or protein is the ratio of the sum of hydropathy values of all the amino acids to the total number of amino acid residues in the sequence [27].

Secondary Structure Prediction: SOPMA (Self Optimized Prediction Method with Alignment) [28] was employed for predicting the secondary structure of the LytTR domain of AgrA protein sequences considered for this study.

Model Building: In order to derive the tertiary structures of LytTR domain of AgrA proteins, the template was selected from PDB (Protein Data Bank) [29] by using BLASTp algorithm [30]. It was found that PDB ID: 4G4K of *Staphylococcus aureus* shared more than 40% identity with the queried proteins. The 3D structure of the proteins was constructed using Modeller 9.16 modeling tool [31].

Model Evaluation: The overall stereochemical property of the proteins was assessed with PROCHECK [32] by Ramachandran plot analysis [33]. Validation of generated models was further performed by VERIFY 3D [34], which ascertains the consistency of an atomic protein model with its own amino acid sequence as measured by a 3D profile. ERRAT [35] was also used for validation of generated models (it is a wave application which investigates the statistics of pairwise atomic interactions and is able to take into account six different noncovalently bonded atom-atom interactions: CC, CN, CO, NN, NO and OO).

ProSA [36] was used for validating the protein structures by comparing the global energy profile of the modeled protein to that of a unique set of good quality models. Modeled structures were compared with the template protein by the superimposition of the structures using Chimera [37] and MISTRAL tool [38]. Binding site of the modelled proteins were analysed using CASTp online server (http://sts.bioe.uic.edu/castp/index.html?2pk9) [39].

RESULTS AND DISCUSSION: The primary structure analysis **Table 2** showed that the theoretical pI (pH at which protein remains stable) of all the proteins were predicted to be greater than pI=7 which implies that the proteins can be considered as basic in nature. The computed pI values were found to range from 9.59 to 7.14. The total number of negatively charged residues (Asp+Glu) were comparatively lesser than the total number of positively charged residues (Arg+Lys) in *L. monocytogenes, C. trachomatis, E. faecalis, S. pyogenes, L. sphaericus, E. plexicaudatum, F. penangensis* and *S. typhae* indicating the intercellular nature of these proteins. Although the Expasy's ProtParam computes the extinction coefficient for a range of (276, 278, 279, 280, and 282 nm) wavelength, 280 nm is favored, because proteins absorb strongly at 280 nm.

Extinction coefficient values of all the proteins at 280 nm measured in water ranged from 2980 to 9065 $M^{-1}cm^{-1}$ with respect to the concentration of cysteine, tryptophan, and tyrosine. The extinction coefficient value of *L. monocytogenes, C. trachomatis, S. pyogenes,* and *M. canis* was 7575 $M^{-1}cm^{-1,}$ and that of *C. gallinarum* was 9096 $M^{-1}cm^{-1}$. Such a high value of extinction coefficient indicates the presence of a high concentration of cysteine, tryptophan, and tyrosine. The computed extinction coefficients are helpful in the quantitative study of protein-protein and protein-ligand interactions in solution [24].

The instability index values for the proteins were found to be 39.28, 37.66, and 34.69 for *M. canis, L. sphaericus* and *F. penangensis*, respectively. The results suggest *M. canis, L. sphaericus,* and *F. penangensis* as stable protein in a test tube. The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains, which include alanine, valine, isoleucine, and leucine and contributes to protein thermostability [26].

Aliphatic index values of all the proteins were found to be ranging from 87.47 to 114.67. Such a high aliphatic index value of all the proteins suggests that these proteins may be stable in a wide range of temperatures. The GRAVY index values of the proteins were found to range from -0.016 to -0.551. The low GRAVY index value of proteins indicates the possibility of better interaction with water.

The secondary structure of LytTR domain of AgrA proteins predicted by SOPMA reveals that extended strand dominates among secondary structure elements followed by alpha helix, random coils, and beta turns for all the sequences except for *P. typhae* where extended strand dominates followed by random coils, alpha-helix, and beta turns. Secondary structure features as predicted using SOPMA are represented in **Table 3**.

TABLE 2: PHYSICO-CHEMICAL CHARACTERS AS PREDICTED BY EXPASY's PROTPARAM

Organism	Sequence Length	Molecular Weight	pI	-R	+R	EC	II	AI	GRAVY
Listeria monocytogenes	102	12028.05	9.34	12	18	7575	49.86	87.84	-0.449
Chlamydia trachomatis	90	10534.22	9.04	9	12	7575	68.79	114.67	-0.093
Enterococcus faecalis	98	11473.23	8.61	14	16	4595	50.20	89.39	-0.364
Streptococcus pyogenes	90	10520.2	9.04	9	12	7575	62.96	113.56	-0.097
Macrococcus canis	98	11728.33	7.22	14	14	7575	39.28	92.45	-0.492
Lysinibacillus sphaericus	97	11471.34	7.96	15	16	4595	37.55	93.30	-0.264
Eubacterium plexicaudatum	99	11815.75	8.50	15	17	4595	53.38	87.47	-0.336
Carnobacterium gallinarum	91	10902.63	7.14	14	14	9065	49.22	93.08	-0.465
Floricoccus penangensis	98	11430.47	9.59	9	15	2980	34.69	106.43	-0.016
Paenibacillus typhae	98	11497.17	8.93	13	16	6085	48.29	77.55	-0.551

TABLE 3: CALCULATED SECONDARY STRUCTURE ELEMENTS BY SOPMA

	Secondary structure									
Organism	**Alpha helix**	**310 helix**	**Pi helix**	**Beta bridge**	**Extended strand**	**Beta turn**	**Bend region**	**Random coil**	**Ambiguous states**	**Other states**
Listeria monocytogenes	29.41%	0.00%	0%	0.00%	37.25%	5.88%	0.00%	27.45%	0.00%	0.00%
Chlamydia trachomatis	28.89%	0.00%	0%	0.00%	32.22%	7.78%	0.00%	31.11%	0.00%	0.00%
Enterococcus faecalis	29.59%	0.00%	0%	0.00%	36.73%	9.18%	0.00%	24.49%	0.00%	0.00%
Streptococcus pyogenes	32.22%	0.00%	0%	0.00%	34.44%	6.67%	0.00%	26.67%	0.00%	0.00%
Macrococcus canis	26.53%	0.00%	0%	0.00%	39.80%	7.14%	0.00%	26.53%	0.00%	0.00%
Lysinibacillus sphaericus	27.84%	0.00%	0%	0.00%	40.21%	8.25%	0.00%	23.71%	0.00%	0.00%
Eubacterium plexicaudatum	29.29%	0.00%	0%	0.00%	37.37%	7.07%	0.00%	26.26%	0.00%	0.00%
Carnobacterium gallinarum	31.87%	0.00%	0%	0.00%	34.07%	4.40%	0.00%	29.67%	0.00%	0.00%
Floricoccus penangensis	28.57%	0.00%	0%	0.00%	39.80%	10.20%	0.00%	21.43%	0.00%	0.00%
Paenibacillus typhae	26.53%	0.00%	0%	0.00%	36.73%	9.18%	0.00%	27.55%	0.00%	0.00%

A comparative study of the primary structure of the proteins revealed the absence of certain amino acids. Tryptophan was found to be absent in all the proteins. Besides this alanine and methionine were absent in *C. trachomatis, S. pyogenes* and *M. canis*. Proline and threonine were also found to be absent in *E. faecalis* and *L. sphaericus*, respectively. Each amino acid has its own individual physicochemical

characteristic to perform a specific function in the protein. The percentage of polarity, charge, aliphatic, and aromatic nature of proteins vary based on their location and function. The frequency of the different amino acids was estimated using ProtoParam **Table 4** in which lysine was found to be the most abundant amino acid in *L. monocytogenes* (14.706%), *L. sphaericus* (11.340%), *E. plexicaudatum* (10.340%) and *P. pyphae* (10.204%) whereas serine was the most frequent in *C. trachomatis* (13.333%), *E. faecalis* (12.245%) and *S. pyogenes* (13.333%). Isoleucine was found to be maximum in *M. canis* (14.286%) and *F. penangensis* (17.347%).

TABLE 4: PERCENTAGE OF AMINO ACIDS PRESENT IN LytTR DOMAIN OF AgrA PROTEINS

Amino acids	*Listeria monocytogenes*	*Chlamydia trachomatis*	*Enterococcus faecalis*	*Streptococcus pyogenes*	*Macrococcus canis*	*Lysinibacillus sphaericus*	*Eubacterium plexicaudatum*	*Carnobacterium gallinarum*	*Floricoccus penangensis*	*Paenibacillus typhae*
Ala	1.961	0.000	1.020	0.000	0.000	3.093	2.020	2.198	2.041	3.061
Cys	1.961	2.222	2.041	2.222	2.041	2.062	3.030	2.198	1.020	2.041
Asp	4.902	4.444	4.082	4.444	8.163	6.186	7.071	3.297	5.102	7.143
Glu	6.863	5.556	10.204	5.556	6.122	9.278	8.081	12.088	4.082	6.122
Phe	5.882	4.444	6.122	4.444	7.143	8.247	8.081	4.396	10.204	7.143
Gly	3.922	3.333	4.082	3.333	3.061	4.124	4.040	4.396	3.061	5.102
His	4.902	2.222	4.082	2.222	7.143	3.093	4.040	5.495	2.041	3.061
Ile	8.824	11.111	10.204	10.000	14.286	10.309	9.091	8.791	17.347	8.163
Lys	14.706	5.556	10.204	5.556	8.163	11.340	10.101	9.890	13.265	10.204
Leu	8.824	13.333	7.143	13.333	7.143	5.155	6.061	8.791	7.143	7.143
Met	2.941	0.000	3.061	0.000	0.000	3.093	4.040	3.297	2.041	3.061
Asn	5.882	2.222	5.102	2.2222	6.122	6.186	7.071	6.593	10.204	7.143
Pro	0.980	3.333	0.000	3.333	2.041	1.031	1.010	2.198	2.041	1.020
Gln	1.961	6.667	1.020	6.667	1.020	3.093	1.010	1.099	2.041	3.061
Arg	2.941	7.778	6.122	7.778	6.122	5.155	7.071	5.495	2.041	6.122
Ser	6.863	13.333	12.245	13.333	9.184	5.155	4.040	4.396	9.184	7.143
Thr	4.902	2.222	3.061	2.222	4.082	0.000	2.020	1.099	2.041	4.082
Val	5.882	6.667	7.143	7.778	3.061	10.309	9.091	7.692	3.061	5.102
Trp	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Tyr	4.902	5.556	3.061	5.556	5.102	3.093	3.030	6.593	2.041	4.042

In order to predict the 3D structure of LytTR domain of AgrA proteins, pairwise sequences alignment between target and template was performed using ClustalW (BLOSUM) [40]. **Fig. 1** depicts the target-template sequence alignments obtained using ClustalW. The modeling of three-dimensional structures of proteins was performed using the homology modeling program, Modeller 9.16. Model structures are given in **Fig. 2**.

Ramachandran plots for all the modeled proteins have been illustrated in **Fig. 3**. The result reveals that more than 95% of the residues were found to be in the most favored regions for all the proteins except for *E. plexicaudatum* and *C. gallinarum* where 93.5% and 92.9% of the residues respectively were in the most favored regions. The values assigned by the Ramachandran plot indicate that the predicted models are of good quality.

The overall G-factors of the modeled proteins were found to lie between -0.08 to 0.02. As the values are greater than the acceptable value -0.50, this suggests that the modeled structures are acceptable. VERIFY 3D showed that more than 80% of the residues have scored greater than 0.2 in the 3D/1D profile for *E. faecalis*, *S. pyogenes*, *M. canis*, *E. plexicaudatum*, *F. penangensis* and *P. typhae.*

This implies that these models are compatible with their sequence. The modeled structures were also validated by another structure verification program, ProSA-web. ProSA Z-score values for all the modeled proteins were found to be within the range of scores typically found for native proteins of similar size, showing the good quality of the models. A plot showing the ProSA Z-score is presented in **Fig. 4**.

```
4G4K              ---ETIELKRGSNSVYVQYDDIMFFESSTKSHRLIAHLDNRQIEFYGNLKELSQLDDR-F
Macrococcus       ---DKLEIKSGSQTLFIDYDDIIFFESSPNPHRIILHLNNRVIEFYGSLKDYETIDER-F
Enterococcus      ---KLFKFKDGDTIRSVDMTTIIFFESSSVSHKIVLHLENGEIEFYGSLKEIEEQSND-F
Lysinibacillus    ---KKFVQRIADKIISVNFDEIMFFESSP-LHKVVLHLDNRQVEFYGKLKDVEKEYEG-F
Paenibacillus     ---KKFQTKSGDKMISVDYNDILFFETSPQLHKIILHAKNRQVEFYGKLKDILEMDSR-F
Carnobacterium    ----------GDKVIHVPYDEILFFESSPKVHKIIMHLENRQVEFYGKLKDLENLGAN-F
Listeria          MQKYFTFKVSDKKIIHELLDDILFFETAPTIHKVILHGKNRQVEFYGKLKNIEKMLDESF
Eubacterium       --ANFSVKAN-EKVFTVDYDDILFFETSPNVHKIILHCKNRQMEFLGKIKEIEKEVDGRF
Floricoccus       ---KMIDIKIGAEIKFIPLKQIMFFETSPISHKLILNLTNDKIEFYGKLKNIINLSDD-F
Chlamydia         ----------KSQIQVPFSDLLYIETSLIPHKLILYSTKQRVEFYGQLSEIIEQDDR-L
Streptococcus     ----------KSQIQVPFSDLLYIETSLIPHKLILYSTKQRVEFYGQLSEIVEQDDR-L
                                    ::::*::   *:::    :  :** *.:.:        :

4G4K              FRCHNSFVVNRHNIESIDSKERIVYFKNKEHCYASVRNVKKI
Macrococcus       FRCHHSYIINKDHITSVNKKERIVHFTSGDHCYISIRNLKKL
Enterococcus      YRCHKSYLINRKHISKVIKSERIVEMSNGERCLVSVRAMKNL
Lysinibacillus    IRCHNSFVVNIDNIQEVDFKRKEILMKNGEICYGSARMIKAV
Paenibacillus     YRCHNSFVVNKDNIAEVDTKNREIRMTNGEICYASSRFLGGL
Carnobacterium    YRCHHSYVVNRENILEIKKKEREIIMTNEEVCYASMRYLKGL
Listeria          YRCHRSYIVNKKNIHELDTTKGVVKMSNGENCYASSKLIKSL
Eubacterium       YRCHRSFLVNKDNIREIDFHKRVIYMVNGDECLISVRMMKGL
Floricoccus       IKVHSSFVVNKNSISSIDYKNFEINFINGQKCFASRRFIKNL
Chlamydia         FQCHRSFVVNPYNISSIDRSERLVYLKGGLSCIVSRLKIRSL
Streptococcus     FQCHRSFVVNPYNISSIDRSERLVYLKGGLSCIVSRLKIRSL
                   : * *:::*   * .:   .  : : .   *  *   :  :
```

FIG. 1: AMINO ACID SEQUENCE ALIGNMENT OF TEMPLATE (4G4K) AND TARGET PROTEINS BY CLUSTALW

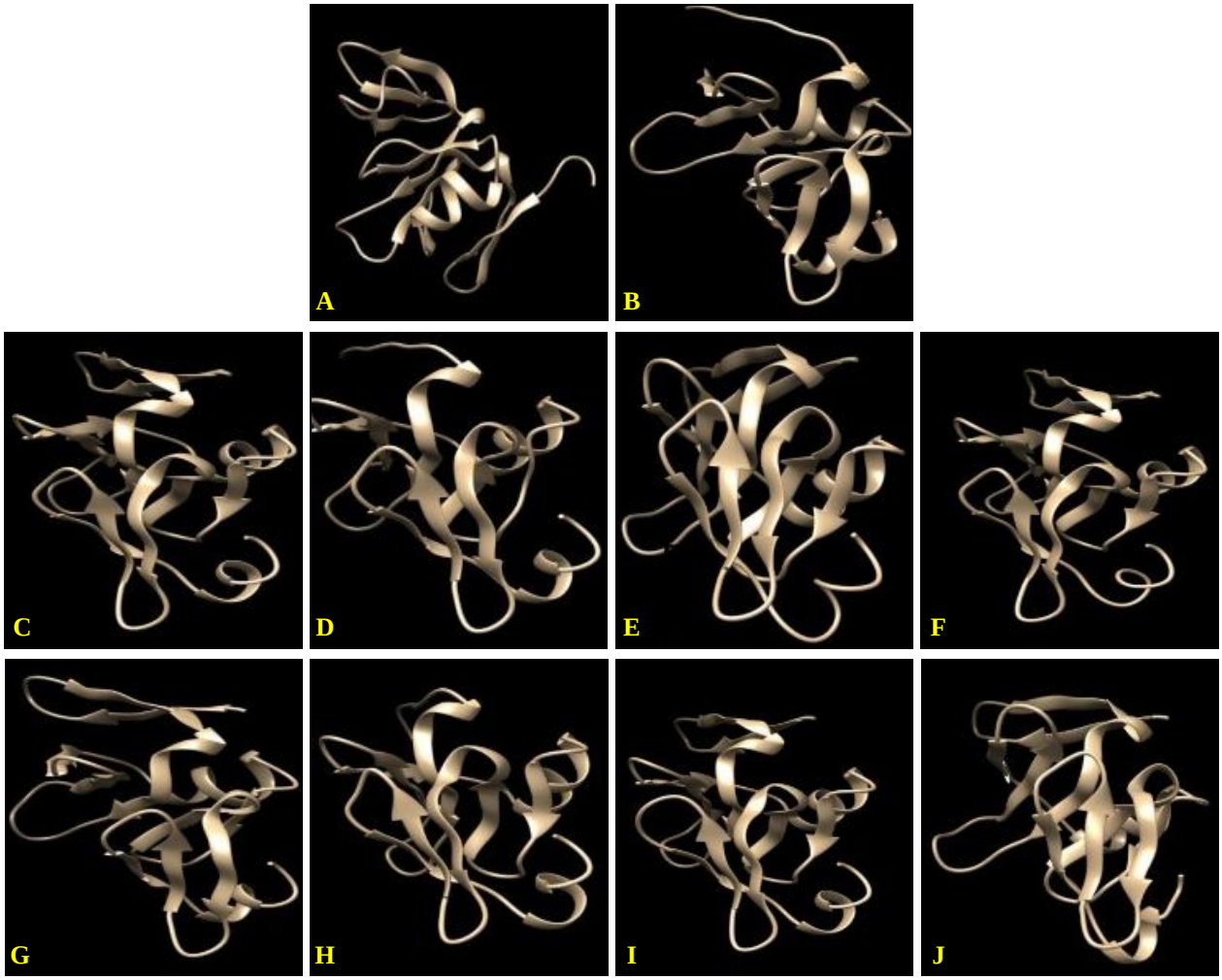

FIG. 2: MODELLED STRUCTURE OF LytTR DOMAIN OF AgrA PROTEINS. (A) *L. MONOCYTOGENES* (B) *C. TRACHOMATIS* (C) *E. FAECALIS* (D) *S. PYOGENES* (E) *M. CANIS* (F) *L. SPHAERICUS* (G) *E. PLEXICAUDATUM* (H) *C. GALLINARUM* (I) *F. PENANGENSIS* (J) *P. TYPHAE*

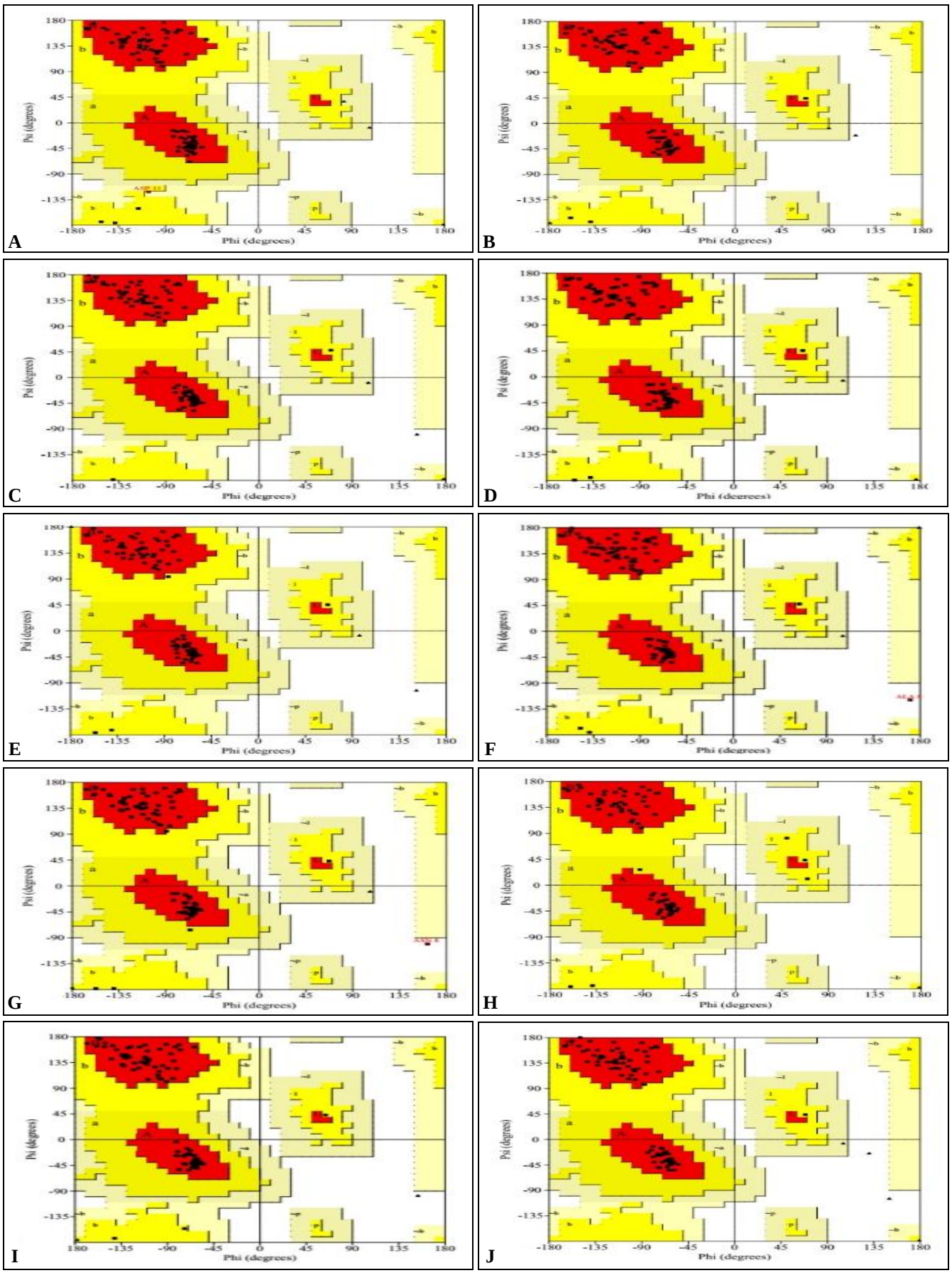

FIG. 3: RAMACHANDRAN'S MAP OF LytTR DOMAIN OF AgrA PROTEINS. (A) *L. MONOCYTOGENES* (B) *C. TRACHOMATIS* (C) *E. FAECALIS* (D) *S. PYOGENES* (E) *M. CANIS* (F) *L. SPHAERICUS* (G) *E. PLEXICAUDATUM* (H) *C. GALLINARUM* (I) *F. PENANGENSIS* (J) *P. TYPHAE*

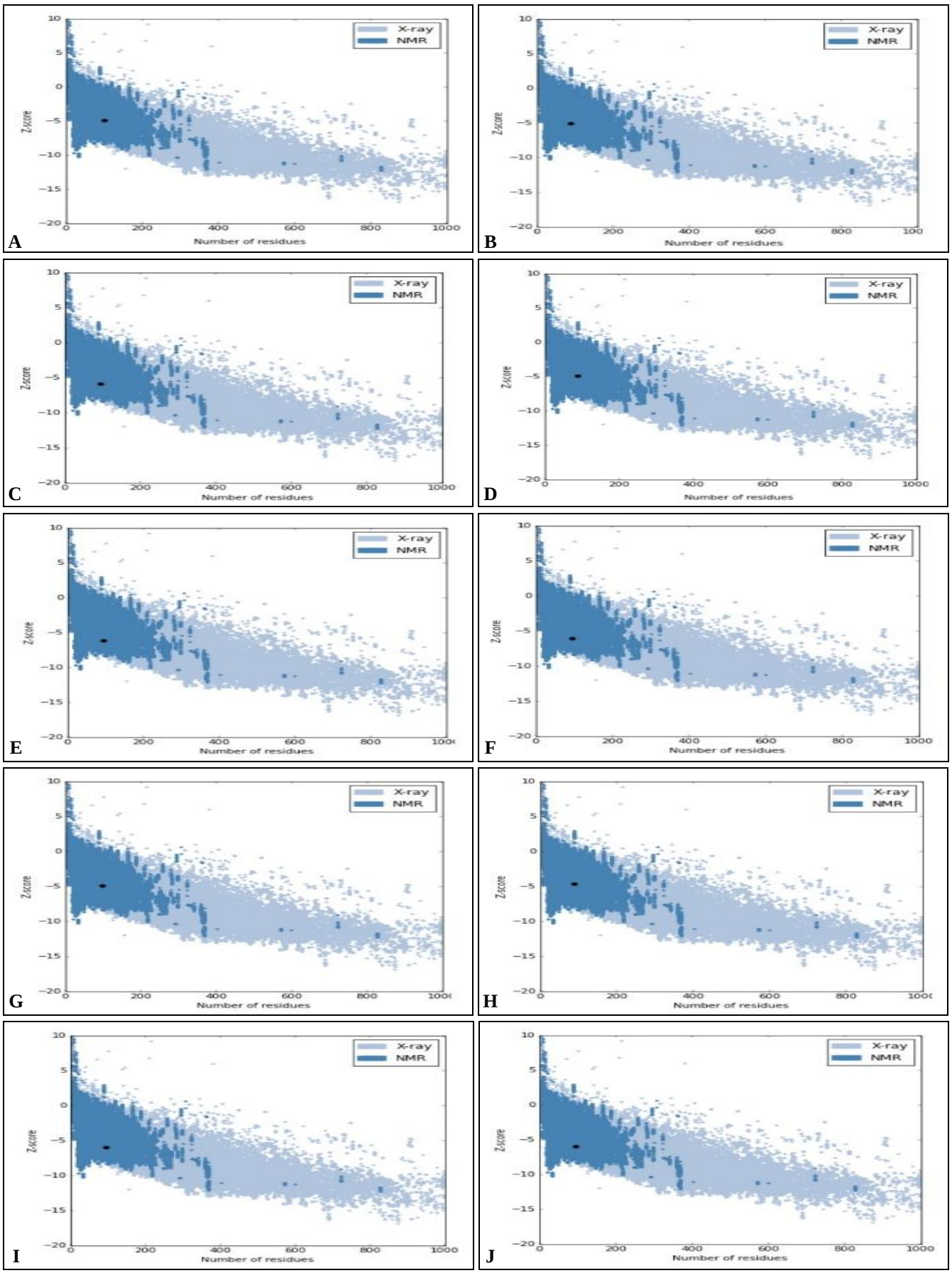

FIG. 4: PROSA-WEB Z-SCORE PLOT. (A) *L. MONOCYTOGENES* (B) *C. TRACHOMATIS* (C) *E. FAECALIS* (D) *S. PYOGENES* (E) *M. CANIS* (F) *L. SPHAERICUS* (G) *E. PLEXICAUDATUM* (H) *C. GALLINARUM* (I) *F. PENANGENSIS* (J) *P. TYPHAE*

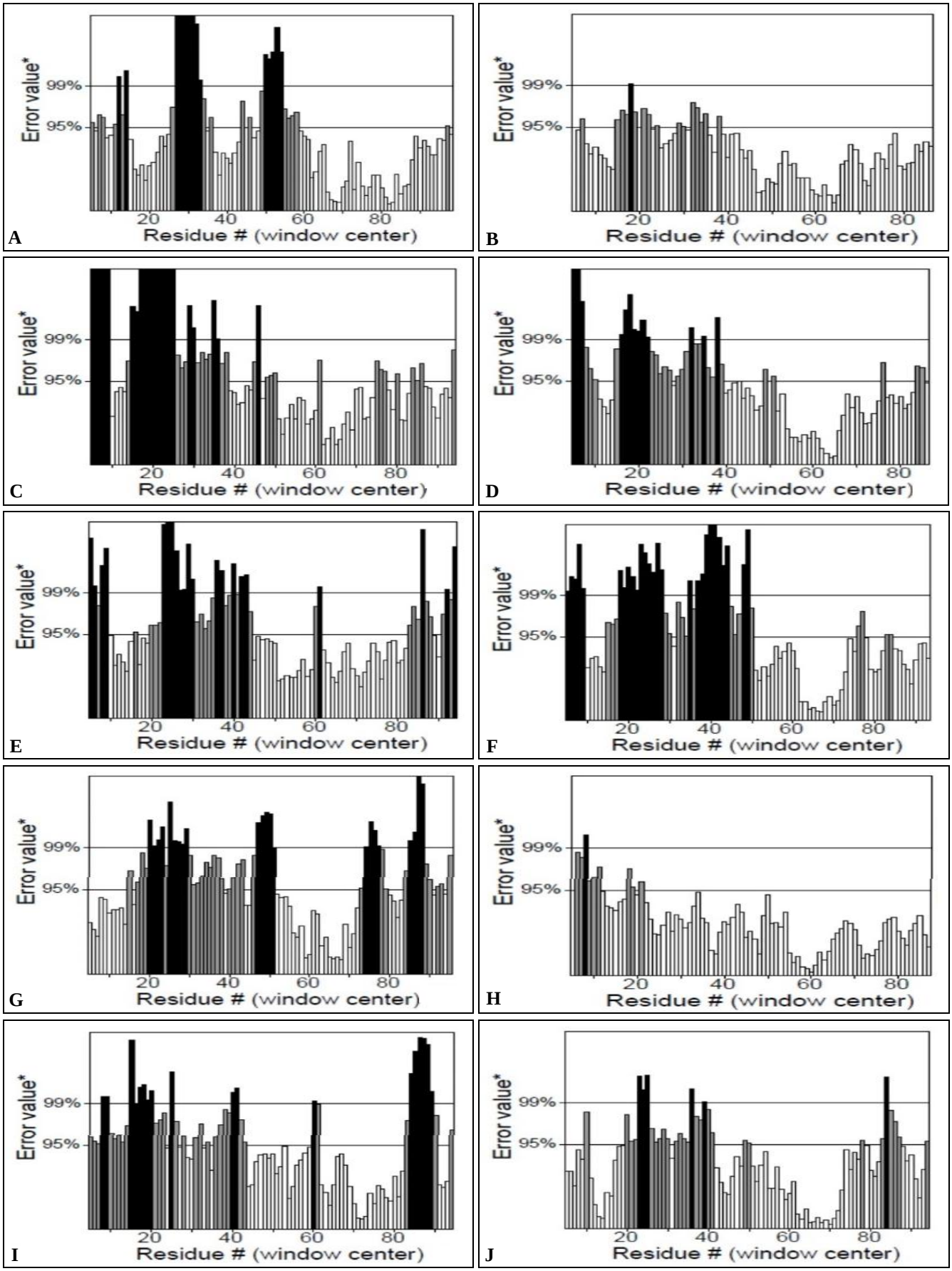

FIG. 5: ERRAT PLOT OF AgrA PROTEINS. (A) *L. MONOCYTOGENES* (B) *C. TRACHOMATIS* (C) *E. FAECALIS* (D) *S. PYOGENES* (E) *M. CANIS* (F) *L. SPHAERICUS* (G) *E. PLEXICAUDATUM* (H) *C. GALLINARUM* (I) *F. PENANGENSIS* (J) *P. TYPHAE*

TABLE 5: DIFFERENT STRUCTURE VALIDATION PARAMETERS OF THE MODELLED PROTEINS

Validation	*L. monocytogenes*	*C. trachomatis*	*E. faecalis*	*S. pyogenes*	*M. canis*	*L. sphaericus*	*E. plexicaudatum*	*C. gallinarum*	*F. penangensis*	*P. typhae*
G-factor	-0.080	-0.080	-0.010	-0.020	-0.070	0.200	-0.060	0.020	-0.030	-0.020
Z-Score	-4.900	-5.060	-5.860	-4.910	-6.130	-6.050	-4.880	-4.630	-5.950	-5.990
RMSD	0.589	0.289	0.162	1.318	0.163	0.281	0.336	0.223	0.213	1.272

Structural comparison between the modeled proteins and template using root mean square deviation (RMSD) matric measures the difference between C-alpha atom positions between the modeled and template proteins. Smaller the deviation, the better is spatially equivalent of the modeled and template proteins. Superimposition of the protein structures was done following the online server MISTRAL and Chimera program. In MISTRAL, the RMSD values of two superimposed protein structures between template and model structures were found to range from 1.892 Å to 0.162 Å. RMSD value is maximum in the case of *L. monocytogenes* and minimum in the case of *E. faecalis* in MISTRAL. The RMSD values in Chimera are given in **Table 5**. RMSD values imply the good quality of the modeled structures.

ERRATA showed an overall quality factor of 68.085, 80.247, and 89.024 for *L. monocytogenes*, *C. trachomatis*, and *C. gallinarum*, respectively **Fig. 5**. The modeled 3D structures were submitted to the CASTp server. The protein structures when submitted to the server in its default proves radius of 1.4 Å, which generated 13, 11, 13, 9, 8, 16, 8, 15, 8, and 13 pockets for *L. monocytogenes*, *C. trachomatis*, *E. faecalis*, *S. pyogenes*, *M. canis*, *L. sphaericus*, *E. plexicaudatum*, *C. gallinarum*, *F. penangensis*, and *P. typhae* respectively. Area, volume, and residues of pocket 1 (largest pocket) of the modeled proteins are given in **Table 6,** which shows that the largest binding pockets are made up of preferably leucine, phenylalanine, isoleucine, glutamine, valine, and lysine.

TABLE 6: AREA, VOLUME, AND RESIDUES OF THE LARGEST POCKET

Organism	Pocket 1 (Largest)		
	Area (Å^2)	Volume (Å^3)	Residues
L. monocytogenes	40.219	16.046	Try4, Phe5, Leu19, Asp20, Leu56, Asp57, Phe60, Asn70
C. trachomatis	268.806	200.664	Gln3, Ile4, Gln5, Val6, Pro7, Phe8, Asp10, Leu11, Ile14, Leu23, Leu25, Ser27, Lys29, Gln30, Arg31, Val32, Phe34, Tyr35, Gly36, Gln37, Glu40, Ile41, Gln44, Leu48
E. faecalis	15.893	3.403	Phe3, Lys6, Phe42, Gly44, Glu48, Ile49
S. pyogenes	421.480	315.767	Lys1, Ser2, Gln3, Ile4, Gln5, Val6, Phe8, Asp10, Leu11, Tyr13, Ile14, Lys22, Leu25, Ser27, Lys29, Gln30, Arg31, Val32, Glu33, Phe34, Tyr35, Gly36, Gln37, Glu40, Ile41, Gln44, Leu48, Val56, Asn58
M. canis	6.352	1.374	Ser24, Ser25, Pro28, His29, Arg30, Ile31,
L. sphaericus	23.191	17.284	Arg57, His59, Asn60, Ser90, Arg92, Met93
E. plexicaudatum	47.050	33.300	Phe3, Ser4, Lys6, Phe42, Leu43, Gly44, Lys45, Glu48, Ile49
C. gallinarum	885.067	85.741	Val4, Ile5, His6, Val7, Pro8, Tyr9,Ile12, Phe15, Met26, Phe35, Leu42, Leu45, Phe49
F. penangensis	47.325	32.206	Tyr75, Lys76, Phe78, Arg92, Ile95, Lys96
P. typhae	66.348	50.938	Ile12, Val14, Asp18, Leu33, Ala35, Asn37, Arg38, Gln39, Val40

Our study suggests that the LytTR domain of AgrA proteins of the selected pathogenic organisms have different amino acid compositions from the non-pathogenic ones. An interesting finding was the absence of an aliphatic amino acid alanine from the LytTR domain of AgrA proteins of pathogenic bacteria like *C. trachomatis*, *S. pyogenes* and *M. canis*.

These organisms were also found to be devoid of methionine in the LytTR domain. The frequency of lysine was also found to vary within the LytTR domain between the pathogenic and the non-pathogenic organisms. Lysine content was found to be lower in pathogens such as *C. trachomatis*, *S. pyogenes*, and *M. canis* compared to the non-pathogens.

It was also observed that the LytTR domain of AgrA proteins of pathogenic bacteria possesses a higher amount of serine and leucine residues and lower amount of asparagine than that of the non-pathogenic bacteria. It was further analyzed that frequency of acidic residues such as aspartic acid, glutamic acid, asparagine, and glutamine in pathogenic bacteria is higher than the non-pathogenic bacteria **Table 4**. Thus, our comparative study suggests that there is a substantial difference in the amino acid composition of the LytTR domain of AgrA protein between the pathogenic and the non-pathogenic organisms.

CONCLUSION: The availability of a reliable 3D structure of a molecular target is essential for the development of therapeutics. In this study, ten LytTR domain of AgrA proteins were selected. Physico-chemical characterization suggests that the total number of negatively charged residues was comparatively lesser than the total number of positively charged residues in most of the proteins, which indicates the intercellular nature of these proteins. A higher value of the aliphatic index indicates higher stability of all the proteins in a wide range of temperatures.

All the proteins showed a very low GRAVY index value indicating the possibility of better interaction with water. Secondary structure analysis revealed that the extended strand dominated among secondary structure elements followed by an alpha helix, random coils, and beta turns for most of the sequences employed here.

The models were validated utilizing a variety of methods like Ramachandran plot, VERIFY 3D, ERRAT, and ProSA. Ramacandran plot showed that more than 92.8% residues were falling under the core region for all the proteins, which means that the predicted structures are stereochemically stable. The structural comparison showed that there is not much significant deviation of the structure of the modeled proteins from that of the template. Binding pockets of the modeled proteins were predicted using CASTp server. The comparative study suggests that there is a substantial difference in the amino acid composition of the LytTR domain of AgrA protein between the pathogenic and the non-pathogenic organisms. The findings of this study will add to the existing knowledge base of LytTR domain of AgrA protein structures, which will provide a reliable platform for designing of novel antibacterials targeting the *agr* quorum-sensing mechanism.

ACKNOWLEDGEMENT: The authors are deeply indebted to Late Prof. A. K. Bothra for his unending support and encouragement during the course of this work.

CONFLICTS OF INTEREST: No conflict of interest exists among the authors.

REFERENCES:

1. Levy SB and Marshall B: Antibacterial resistance worldwide: causes, challenges and responses. Nature Medicine 2004; 10: S122.
2. Levy SB: Antibiotic Resistance: Consequences of Inaction. Clinical Infectious Diseases 2001; 33(S3): S124-S129.
3. Levy SB: Factors impacting on the problem of antibiotic resistance. Journal of Antimicrobial Chemotherapy 2002; 49(1): 25-30.
4. Aiello AE and Larson E: Antibacterial cleaning and hygiene products as an emerging risk factor for antibiotic resistance in the community. The Lancet Infectious Diseases 2003. 3(8): 501-506.
5. Okeke IN: Antimicrobial resistance in developing countries. Part I: recent trends and current status. The Lancet Infectious Diseases 2005; 5(8): 481-93.
6. Holmberg SD, Solomon SL and Blake PA: Health and Economic Impacts of Antimicrobial Resistance. Reviews of Infectious Diseases 1987; 9(6): 1065-78.
7. Novick RP and Geisinger E: Quorum Sensing in Staphylococci. Annual Rev of Genet 2008; 42(1): 541-64.
8. Ng WL and Bassler BL: Bacterial quorum-sensing network architectures. Annual Review of Genetics 2009; 43(1): 197-22.
9. Williams P and Cámara M: Quorum sensing and environmental adaptation in *Pseudomonas aeruginosa*: a tale of regulatory networks and multifunctional signal molecules. Current Opinion in Microbiology 2009; 12(2): 182-91.
10. Chen G: A Strategy for Antagonizing Quorum Sensing. Molecular Cell 2011; 42(2): 199-09.
11. Swem LR: A quorum-sensing antagonist targets both membrane-bound and cytoplasmic receptors and controls bacterial pathogenicity. Molecular Cell 2009; 35(2): 143-53.
12. Galloway WRJD: Quorum sensing in gram-negative bacteria: small-molecule modulation of AHL and AI-2 quorum sensing pathways. Chemical Reviews 2011; 111(1): 28-67.
13. Suga H and Smith KM: Molecular mechanisms of bacterial quorum sensing as a new drug target. Current Opinion in Chemical Biology 2003; 7(5): 586-91.
14. Kalia VC and Purohit HJ: Quenching the quorum sensing system: potential antibacterial drug targets. Critical Reviews in Microbiology 2011; 37(2): 121-40.
15. Finch RG, Bcroft BW, Williams P and Stewart GS: Quorum sensing: a novel target for anti-infective therapy. Journal of Antimicrobial Chemotherapy 1998; 42: 569-71.

16. Habeck M: Stop talking at the back. Drug Discov Today 2003; 8: 279-80.
17. Smith RS and Iglewski BH: *Pseudomonas aeruginosa* quorum sensing as a potential antimicrobial target. The Journal of Clinical Investigation 2003; 112(10): 1460-65.
18. Hentzer M and Givskov M: Pharmacological inhibition of quorum sensing for the treatment of chronic bacterial infections. The Journal of Clinical Investigation 2003; 112(9): 1300-07.
19. Dunman PM: Transcription profiling-based identification of *Staphylococcus aureus* genes regulated by the agr and/or sarA loci. J of Bacteriology 2001; 183(24): 7341.
20. Novick RP: Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. The EMBO Journal 1993; 12(10): 3967-75.
21. Janzon L, Löfdahl S and Arvidson S: Identification and nucleotide sequence of the delta-lysin gene, hld, adjacent to the accessory gene regulator (*agr*) of *Staphylococcus aureus*. Mol and Genl Genet MGG 1989; 219(3): 480-85.
22. Liu X, Fan K and Wang W: The number of protein folds and their distribution over families in nature. Proteins: Structure, Function, and Bioinfor 2004; 54(3): 491-99.
23. Fiser A: Template-based protein structure modeling, in computational biology, D. Fenyö, Editor. Humana Press: Totowa, NJ 2010; 73-94.
24. Gill SC and von Hippel PH: Calculation of protein extinction coefficients from amino acid sequence data. Analytical Biochemistry 1989; 182(2): 319-26.
25. Idicula-Thomas S and Balaji PV: Understanding the relationship between the primary structure of proteins and its propensity to be soluble on over-expression in *Escherichia coli*. Protein Science 2009; 14(3): 582-92.
26. Ikai A: Thermostability and aliphatic index of globular proteins. The J of Biochemistry 1980; 88(6): 1895-98.
27. Kyte J and Doolittle RF: A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology 1982; 157(1): 105-32.
28. Geourjon C and Sopma DG: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 1995; 11(6): 681-84.
29. Bernstein FC: The protein data bank. A computer-based archival file for macromolecular structures. Eur J Biochem 1977; 80(2): 319-24.
30. Altschul SF: Basic local alignment search tool. J Mol Biol 1990; 215(3): 403-10.
31. Webb B and Sali A: Comparative protein structure modeling using MODELLER. Curr Protoc Bioinformatics, 2014; 47: 5.6.1-32.
32. Laskowski RA: PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography 1993; 26(2): 283-91.
33. Ramachandran GN, Ramakrishnan C and Sasisekharan V: Stereochemistry of polypeptide chain configurations. Journal of Molecular Biology 1963; 7(1): 95-99.
34. Eisenberg D, Lüthy R and Bowie JU: [20] VERIFY3D: Assessment of protein models with three-dimensional profiles, in Methods in Enzymology. Academic Press 1997; 396-04.
35. Colovos C and Yeates TO: Verification of protein structures: patterns of nonbonded atomic interactions. Protein science: a publication of the Protein Society 1993; 2(9): 1511-19.
36. Sippl MJ: Recognition of errors in three-dimensional structures of proteins. Proteins: Structure, Function, and Bioinformatics 1993; 17(4): 355-62.
37. Laskowski RA: AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996; 8(4): 477-86.
38. Micheletti C and Orland H: MISTRAL: a tool for energy-based multiple structural alignment of proteins. Bioinformatics 2009; 25(20): 2663-69.
39. Tian W: CASTp 3.0: computed atlas of surface topography of proteins. Nucleic Acids Research 2018; 46(W1): W363-W367.
40. Chenna R: Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Research 2003; 31(13): 3497-00.

How to cite this article:
Basak HK, Chatterjee A and Pal A: Relative structural analysis of LytTR domain of AgrA protein involved in bacterial quorum sensing. Int J Pharm Sci & Res 2020; 11(6): 2828-39. doi: 10.13040/IJPSR.0975-8232.11(6).2828-39.