INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN 2278 – 0211 (Online)

Identification of Bacterial Isolates, PCR Amplification and In-Silico Analysis of Nitroreductase and Rubrerythrin Responsive Genes from *Shilajit*

Tanuja Mishra
Ph.D. Scholar, Department of Biotechnology, Eternal University, Baru Sahib (Sirmour), Himachal Pradesh, India
Prabhjot Kaur Gill
Assistant Professor,Department of Biotechnology, Eternal University, Baru Sahib (Sirmour), Himachal Pradesh, India
H. S. Dhaliwal
Professor, Department of Biotechnology, Eternal University, Baru Sahib (Sirmour), Himachal Pradesh, India
Arun Dev Sharma
Assistant Professor, PG Department of Biotechnology, Lyallpur Khalsa College Jalandhar, Punjab, India
Gobind Ram
Assistant Professor, PG Department of Biotechnology, Lyallpur Khalsa College Jalandhar, Punjab, India

Abstract:
16S rRNA gene sequence homology in combination with morphological, physiological and biochemical characteristics was used as the tools for identification of the selected isolates from Shilajit samples designated EUS1, EUS2 and EUS3 which were collected from mountains of Baru Sahib, Sirmour district of Himachal Pradesh. The isolated strains were identified as B. subtilis, B. cereus, B. clausii and were deposited in GenBank with accession number KP126518, KP126519, and KP126520 respectively. Furthermore, the present study was designed to analyze molecular modeling of proteins from Bacillus spp. (KP126518: BsNR, KP126519: BcRbr) and was carried out by subjecting to in silico analysis. Motif scan analysis revealed the presence of signature sequences that might be related to structure and enzymatic function. The 3D structure of Bacillus proteins were predicted by swiss-model server based on homology modeling method. The modeled structures were validated by various validation servers. Six probable active sites in B. subtilis (BsNR) and two in B. cereus (BcRbr) were predicted for target protein structure.

Keywords: *16S rRNA gene, 3D modeling, in-silico analysis, phylogenetic analysis, Shilajit*

1. Introduction

Shilajit is a pale-brown to blackish-brown exudates, of variable consistency, obtained from layer of rocks in many mountain ranges of the world, at altitudes of 1000 to 5000 meters. The Himalayan ranges of Indian subcontinent i.e. Uttarakhand, Himachal Pradesh, Kashmir, and Arunachal Pradesh are known to be potent spots for *Shilajit* (Jaiswal and Bhattacharya, 1992). *Shilajit* comprises of a variety of significant compounds namely dibenzo-alpha pyrones, phospholipids, triterpenes, phenolic acids of low molecular weight, fulvic acids (carrier molecules), humins, humic acids and trace elements (Fe, Ca, Cu, Zn, Mg, Mn, Mo, P) (Kong *et al*.,1987). The therapeutic properties of *Shilajit* are probably provided by the significant levels of fulvic acids that are known to be a strong antioxidant and likely have systemic effects as complement activator. The fulvic acid stimulates blood formation, energy production, prevents hypoxia by actively participating in the nutrient transport into deep tissues, and also helps to overcome chronic fatigue (Schepetkin *et al.,* 2009;Vucskits *et al*., 2010). It also works effectively as a cognition enhancer, an adaptogen, an antistress agent and a tonic for cardiac, gastric and nervous systems (Ghosal *et al*., 1989). *Shilajit* improves the ability to handle high altitudinal stresses (hypoxia, acute mountain sickness, high altitude cerebral edema, pulmonary edema, insomnia, tiredness, lethargy, lack of appetite, body pain, dementia etc.) and also stimulates the immune system (Meena *et al*., 2010). Apart from this, *Shilajit* find its application in treatment of various ailments namely diabetes (Bhattacharya, 1995), rheumatism, ulcers (Goel *et al*., 1990) and allergic reactions like mast cell degranulation (Ghosal *et al*., 1989).

The formation of this ancient panacea is attributed by the degradation of plant remains, by the action of various microorganisms, under high temperature and pressure for several years (Ghosal *et al*., 1976). On account of the fact that it is a degradation product, there is a need for isolation and characterization of microbial species associated with formation of crude *Shilajit* so as to explore various thermostable enzymes that could find its wide applications in the industries. Moreover, some enzymes may be associated with these microorganisms which make worthwhile contribution either in biodegradation of plant remains or in alleviating varied stresses in microorganisms and need to get explored.

Nitroreductases comprise a family of proteins with conserved sequences that were originally discovered in eubacteria and have been grouped together based on their sequence similarity (Bryant *et al*., 1981). These enzymes are capable of catalyzing the reduction of nitrosubstituted compounds using flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD) as prosthetic groups and nicotinamide adenine dinucleotide (NADH) or nicotinamide adenine dinucleotide phosphate (NADPH) as reducing agents. These proteins have recently raised enormous interest both

in the environmental engineering community (due to their central role in mediating nitro aromatic toxicity and their potential use in bioremediation and biocatalysis) and in the medical community (as agents used to activate prodrugs in directed anticancer therapies) (Bryant *et al*., 1991).
Rubrerythrins are non-haem, non-sulfur proteins that have been implicated in oxidative stress protection in anaerobic bacteria as well as archaea and first isolated from the anaerobic sulfate-reducing bacterium *Desulfovibrio vulgaris* (LeGall *et al.*, 1988). Rbr contains an oxo-bridged diiron site and a rubredoxin-like [$Fe(SCys)_4$] site within separate domains of a single subunit (deMare' *et al*., 1996). The diiron site is similar to those in a class of enzymes that activate dioxygen (Kurtz, 1997). However, no *in vivo* substrates for Rbr have been conclusively identified. *In vitro*, the fully reduced (all-ferrous) form of *D. vulgaris* Rbr reduces hydrogen peroxide much more rapidly than dioxygen, and purified Rbrs can serve as the terminal component of an NADH peroxidase by catalysing two-electron reduction of hydrogen peroxide (Coulter *et al*., 2000; Coulter and Kurtz, 2001).
Present study is a pioneer for isolation and characterization of microbial flora from *Shilajit*. Three bacterial strains were isolated from *Shilajit* oozing out from rocks. Morphological and biochemical characters were studied in the lab and then the strains were further characterized by using 16S ribosomal RNA gene sequencing. For the identification as well as the natural relationships between the microorganisms, 16S rRNA is considered to be most powerful method. 16S rRNA gene sequences of isolates were determined and compared with the sequences of different *Bacillus* species obtained from the EMBL and GenBank databases in order to clarify the sequence homology of different *Bacillus* species and then the enzymes associated with them were also identified using *in-silico* analysis. In order to make *Bacillus* strains commercially viable, genes from isolated strains needs to be characterized at molecular level by different *in-silico* approaches so that these strains may be genetically manipulated. In this study, sequence of *BcRbr* and *BsNR* were obtained to understand the functional behavior of the protein from bacterial isolates, and their structure (secondary and tertiary) as well as physiochemical properties, identification of motif/domain and catalytic active sites were predicted by using various bioinformatics' tools viz., homology modeling, validation and analysis. Identification of active sites and motifs might be useful to get insight into the molecular function at sequence-structure-function relationship. Homology modeling and *in-silico* characterization is an invaluable tool widely used in bioinformatics to predict 3D structure and manifestation at function level of unknown protein.

2. Experimental

2.1. Materials and Methods

2.1.1. Isolation of Bacteria

Shilajit samples were collected in zip lock polythene bags from different sources i.e. *Shilajit* (exudates) deposited over leaves, exuding out from mountain and deposited on rocks.
After serial dilution, the samples were spread onto Luria Bertani (LB) Agar plates (HiMedia) followed by incubation at 35±2ºC for 48 hours. Colonies were selected initially according to their morphology and isolation of well-defined single colonies resulted in three different bacterial strains as pure cultures.

2.1.2. Preliminary Identification of Isolates

Bacterial isolates were subjected to preliminary identification by morphological (colony shape, size,) and biochemical analysis: Indole, Methyl-red, Voges-Proskaeur's, citrate utilization and fermentation of various carbohydrates were tested using IMViC kit (HiMedia), while catalase activity was checked by pouring a few drops of hydrogen peroxide on isolates spreaded on LB agar plate. Haemolysis was noted by inoculating isolates on blood agar (5% blood added to autoclaved LB Agar) and examining the pattern of degradation of blood after incubation at 35±2ºC for 24 hours

2.1.3. DNA Extraction

Genomic DNA of the isolated bacterial strains was obtained, by growing pure cultures of the bacterial strains in LB broth for 24 hours followed by using CTAB method of DNA extraction (Wilson, 1987).

2.1.4. PCR Amplification of 16S rRNA Gene

PCR reaction was performed in a gradient thermal cycler (Applied Biosystems). The universal primers (Forward primer 5'-AGAGTTTGATCCTGGCTCAG-3' and reverse primer 5'CTTGTGCGGGCCCCCGTCAATTC-3') were used for the amplification of the 16S rRNA gene fragment.
The reaction mixture of 25µl consisted of 12.5µl 2X master mix (Promega India), 3µl of 10 pmoles each of the two universal primers, 3µl DNA and final volume was made up with autoclaved distilled water to 25 µl. Amplification was done by initial denaturation at 94ºC for 5 mins, followed by 34 cycles of denaturation at 94ºC for 1 min, annealing temperature of primers was 52ºC for 1 min and extension at 72ºC for 2 mins. The final extension was conducted at 72ºC for 8 mins.

2.1.5. Agarose gel Electrophoresis

1.2% agarose with ethidium bromide was used to analyze 8µl of the amplified reaction mixture by gel electrophoresis at 8V/cm and the reaction product was visualized under Gel doc/UV trans-illuminator.

2.1.6. 16S rRNA Gene Sequencing of Isolates

Amplified 16S rRNA genes (using the bacterial universal primers-27F 5'-AGAGTTTGATCCTGGCTCAG-3' and 1492R 5'GGTTACCTTGTTACGACTT-3') i.e. PCR products were resolved on agarose gel (1.2% w/v) and bands of ~1500 bp were excised using a sharp blade (Fig. 1). The DNA from the gel slices were eluted employing GeneJET™ Gel Extraction Kit, Fermentas according to the manufacturer's instructions. The sequencing PCR reaction was performed by Big Dye TM Chain termination reaction. The unincorporated dye terminators were removed from the sequencing reaction using Montage SEQ96 sequencing reaction clean up kit (Millipore). The purified sequencing products were transferred into the 96 well injection plates which were then placed in automated sequencer (Applied Biosystems).

2.1.7. Nucleotides Sequence Accession Numbers

The resulting 16S rRNA sequences have been deposited in the GenBank under the accession numbers KP126518 (EUS1), KP126519 (EUS2), and KP126520 (EUS3).

2.1.8. Computational Analysis

A comparison of the 16S rRNA gene sequence of the isolates with the non-redundant collection (GenBank) of sequences was performed using BLAST. A number of sequences of the genus *Bacillus* were aligned with 16S rRNA gene sequence of isolate with 99% sequence similarity. A multiple sequence alignment was then developed, for these homologous sequences using the algorithm described in Clustal W (Thompson *et al*., 1994). An evolutionary distance matrix was generated from these nucleotide sequences in the dataset. Further a phylogenetic tree was then drawn using the Neighbor joining method (Saitou and Nei, 1987) for 16S rRNA gene. A phylogenetic analysis was conducted using MEGA6 (Tamura *et al*., 2011).

2.1.9. Homology Modeling, Structural Analysis and Verification

The sequences of *BcRbr* (KP126519) and *BsNR* (KP126518) were retrieved from NCBI and subjected to ORF scan (www.ncbi.nlm.nih.gov) to identify the coding regions (exons) in the amplified regions. It was ascertained that 3D structure of *BcRbr* and *BsNR* proteins are not available in PDB database, hence an attempt has been made, in the present study, to determine the structure of genes from *B. cereus* and *B. subtilis*. 3D model was generated using protein structure homology modeling by using Swiss-model tool (www.expasy.org). The target EUS1 protein (64 amino acids) and EUS2 (51amino acids) as obtained from ORF scan was subjected to BLASTp against PDB (Protein Data Bank)database for obtaining most similar template. So the modeled tertiary structures of EUS1 and EUS2 were built using best sequence homology with templates as generated PDB-Blast. For *B. clausii* EUS 3, no template was detected in PDB-BLAST, hence not selected for further studies. The selected strains were then annotated as *BsNR* and *BcRbr* for EUS1 and EUS2, respectively. The 3D structures were subjected to energy minimization using Swiss-Pdb viewer. The structural analysis and verification were done using a variety of tools like PDBsum (www.ebi.ac.uk/pdbsum), VADAR (http://redpoll.pharmacy.ualberta.ca/vadar), Verify 3D (http://nihserver.mbi.ucla.edu/Verify_3D), and QMEAN (swissmodel.expasy.org/qmean). 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). To identify the likely biochemical function from 3D structure of *BcRbr* and *BsNR* modeled proteins, ProFunc sever of EMBL-EBI was used.

2.1.10. In Silico Analysis

The conserved motifs were identified by Multiple EM for Motif Elicitation (MEME) (http://meme.nbcr.net/meme4_6_1cgi-bin/meme.cgi). For motif analysis of predicted proteins, the parameters set were as follows; 3 as maximum number of motifs and 6 and 50 as minimum and maximum width, respectively. These selections were made in order to minimize the 'E-value' of the given parameter based on probability of finding an equally well conserved pattern in set of sequences. Physiochemical properties including molecular weight, theoretical isoelectric point (pI), amino acid composition, instability index and GRAVY etc. of predicted protein were calculated by Protparam tool at expacy (www.expacy.org). PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/) tools were used for secondary structure verification.

Super imposition of model and template

Chimera structure alignment tools were used to evaluate the structural similarity by superimposing the 3D protein models and templates. It calculates protein superposition using a modified quaternion approach. From a superposition of two structures, superimpose tools generate structure alignments with Tm Score and RMSD statistics. Chimera web servers support the submission of either PDB-formatted files or PDB accession numbers.

2.1.11. Catalytic Active Site Identification

CASTp (Catalytic Active Site Tool) was used for active site identification of 3D model (http://sts.bioengr.uic.edu/castp/calculation.php). These binding sites were further compared to active sites of the template.

3. Results and Discussion

The *Bacillus* genus is a heterogeneous group of Gram-positive, facultative anaerobic, endospore-forming bacteria. Most species are motile by peritrichous flagella and ubiquitous in nature (Graumann, 2007). The ability to produce endospores allows *Bacillus* to survive extreme environmental conditions. The *B. subtilis and B. cereus* species are haemolytic, a useful characteristic in differentiating them from *B. anthracis* (which is non-haemolytic). Colonies of *B. clausii* form filamentous margins that appear cream-white in color. *B. clausii* is alkaliphilic and non-haemolytic. The protease from *B. clausii* strain 221, the H-221 protease, was the first enzyme to be identified in an alkaliphilic *Bacillus* (Kageyama *et al*., 2007). The alkaliphilic nature of the organism has also proved it to be useful in preventing and treating various gastrointestinal disorders as an oral bacteriotherapy (Senesi *et al*., 2001).

All the 3 bacillus strains were found to be positive in citrate utilization, oxidase, urease, gelatin hydrolysis, catalase, lecithinase, motility, nitrate reduction, Voges-proskaeur and starch hydrolysis as well as they grew well in 5% and 7% NaCl at 42°C and aerobically, it gave negative reaction to indole. The isolates also ferment glucose, mannitol, sucrose as well as inulin and starch. The positive results indicate that the bacterial strains were able to ferment different sugars due to have specific enzyme that responsible for sugar fermentation and production of acid and gas and/or gas. Our results are in agreement with the previous reported data (Cihan *et al*., 2012).

The use of 16S rRNA gene sequences to study bacterial phylogeny and taxonomy has been reported most common housekeeping genetic marker used for a number of reasons. These reasons include (i) its presence in almost all bacteria, often existing as a multigene family, or operons (ii) the function of the 16S rRNA gene over time has not changed, suggesting that random sequence changes are a more accurate measure of time (evolution) and (iii) the 16S rRNA gene (1,500 bp) is large enough for informatics purposes (Patel, 2001).

Approximately 1500 bp of the 16S rRNA gene were sequenced and used for the identification of isolated bacterial strains (Fig.1). RDP classifier tool revealed that all the three isolates belong to genus *Bacillus*. The BLAST result showed that the 16S rRNA gene sequence of isolate EUS1 has 99% sequence similarity and having highest score i.e. 2769 bits with *B. subtilis.* Similarly the isolate EUS2 has 99% sequence similarity with score of 2767 bits with *B. cereus* and that of EUS3 has 99% sequence similarity with score of 2687 bits with *B. clausii*. A phylogenetic tree drawn on the basis of the representative sequences of the 16S rRNA gene of the 16 bacillus species formed 3 clusters. The *B. cereus* strains were very clearly segregated and had moderate bootstrap value of 55-69%. Since *B. subtilis* could not be clustered concretely and formed cluster with both *B. clausii* and *B. cereus* showing bootstrap value of 99%, it is indicative of less diversity and reflects that there is quite less genetic variability among the different clusters. Similarly, *B. clausii* formed moderate phylogenetic line which is supported by moderate bootstrap value of 62 (Fig. 2). Recently Cihan *et al* (2012) analyzed 115 endospore-forming isolates with the 16S rRNA gene sequence. All of them were phylogenetically clustered on the

basis of their individual 16S rRNA gene sequence homologies to their closest relatives. Comparison of the generated sequences with those in the GenBank database indicated that all of the identified isolates belonged to the genera *Bacillus*, *Brevibacillus, Paenibacillus* and *Thermoactinomycetes*. Similarly, Porwal *et al,* (2009) reported 16S rDNA sequences of 10 different *Bacillus* species, developed a phylogenetic frame work of 34 representative sequences, and identified *Bacillus* genera upto species level. In addition, using 16S rRNA gene sequence analysis, Ash *et al.,* (1991) described the presence of five phylogenetically distinct groups in the genus *Bacillus*, and Nielsen *et al.,* (1995) subsequently described a sixth group belonging to the alkaliphilic bacilli.

Recently, *Bacillus* genus have also received attention and considered an interesting genus to investigate for inhibitory substances, since it produces a large number of peptide antibiotics representing several different basic chemical structures (Bizani *et al*., 2005). Like lactic acid bacteria (LAB), some representatives of *Bacillus* spp., such as *B. subtilis*, *B. licheniformis* and *B. coagulans* are called 'GRAS' in food industry and agriculture (Yang *et al*., 2012).

In-silico analysis and 3D modeling

3 gene sequences designated as EUS1, EUS2, and EUS3, having accession numbers KP126518, KP126519 and KP126520 respectively were subjected to ORF scan at NCBI. Among 3, well defined template (PDB-BLAST) was retrieved only with 2 gene sequences (EUS1 designated as *BsNR* and EUS2 as *BcRbr* thereafter), hence selected for further studies. *BsNR* and *BcRbr* protein sequences were identified as nitroreductase and rubrerythrin, respectively, in PDB-NCBI BLAST analysis (Table 1).

Physiochemical properties of *BcRbr* andn *BsNR* proteins were analyzed by ProtParamtool as represents in Table 2. In biochemical properties, *BcRbr* and *BsNR* have isoelectric points 11.6 and 11.0 and molecular weights are 5.8 and 7.4 kDa, respectively. The computed pI above 7 indicates that the protein will precipitate in basic buffers.

The essential amino acid present in *BcRbr* and *BsNR* proteins sequence were arginine (14%, 15.6), subsequently alanine (6.2%, 9.8). The amount of negatively charged protein residues (Asp and Glu) is 3 for *BcRbr* and 4 for *BsNR*, while that of positive protein residues (Arg and Lys) are 8 for *BcRbr* and11 for *BsNR*.The values of aliphatic index, as well as GRAVY for *BcRbr* were 78 and -0.40 and on the other hand for *BsNR* were 85 and -0.20. GRAVY score for both the proteins indicated that overall protein is hydrophilic and soluble in nature.

The motifs present in the protein sequences were analyzed using MEME and showed thepresence of 3 motifs with regular expression and highly representative logo sequences (Fig. 3A,B). In *BcRbr* sequence, motif 1 and 3 corresponds to beta turn and the alpha-helix, whereas motif 2 corresponds to helix. In *BsNR* amino acid sequence, motif 1 corresponds to beta turn, motif 2 corresponds to sheet and motif 3 corresponds to helix. All these observations indicate signature sequential differences along with their important structural roles. Both *BcRbr* and *BsNR* stretches also contained some common active sites residues glycine, leucine, valine in different motifs. In addition some residues like phenylalanine, arginine and cysteine were detected in *BsNR* only and tryptophan and asparagine in *BcRbr* only. Motifs represent the functional motifs for *BsNR* and *BcRbr*, because these sequences contained various active site residues in modeled proteins. Nest analysis using ProFunc also revealed the presence of some structural motifs that are often found in functionally important regions of protein structures. Notably, one nest designated as Nest1, was located in *BsNR* whereas no nest was found in *BcRbr*. Nest 1 having Leu 31(A)-Ile-33(A) residue range was associated with beta turns.

The secondary structure prediction of *BcRbr* sequence by VADAR prediction server revealed a 75% frequency of helix, 24% for coil and turn, 0% for beta turn (Fig. 4). Whereas in *BsNR* sequence 19% frequency for helix, 53% for coil, 8% for turn and 23% for beta turn was detected. Secondary structure prediction tool like PSIPRED also validated the presence of major part as a helix in *BcRbr* and coil in *BsNR*.

Homology modeling, validation and structure analysis

To predict 3D structure of unknown *BcRbr* and *BsNR* proteins, homology modeling and *in-silico* characterization was carried out using various bioinformatics' tools. 3D structure provides valuable insight into molecular function and putative active sites identification. So computational homology modeling is an apt approach to predict 3D structure to get insight into the function of protein on the basis of one or more proteins of known structure. Moreover, in protein research, the 3D structure is a crucial step to understand enzyme kinetics, ligand-protein interaction, and structure based molecular design. So, 3D analysis is necessary for biological functional analysis and is further involved in the structure-function relationship. The suitable template protein structures of targets were selected by performing NCBI PDB-BLAST having three dimensional structural data banks. Results of PDB-BLAST revealed various templates out of which 3D structure of *BcRbr* and *BsNR* showed homology with templates of rubrerythrin (code 4D10A) and nitroreductase (code 3HOIA) which is suggestive of similar functions. Now swiss-model server for homology modeling was used to model protein structure by using above mentioned template structures. The rough model was saved as pdb.file and visualized by Swiss-pdb viewer and subjected to energy minimization. On subjecting to energy minimization, the rough *BcRbr* model with energy -1260.53 KJ/mol, gave stable model with energy value-2371.87 KJ/mol, whereas *BsNR* model with -1682.9 KJ/mol gave -2030.85 KJ/mol. Furthermore, the structural validation and quality assessment of model was analyzed by using various tools like PROCHEK, Verify 3D, VADAR and QMEAN and ResProx. The PROCHECK of PDBsum server was employed for evaluation of stereochemical quality of the predicted structure. The depicted model got validated with Ramachandran plot and showed 75% and 86% of amino acid residues within the most favored region. The Ramachandran plot confirmed the good quality of model. In the context of QMEAN global scores, QMEAN Z score was calculated as -2.78 and -1.46 for *BcRbr* and *BsNR* models, suggested model to be acceptable (Fig. 5A). VADAR server that included parameters like stereo packing quality index, 3D profile quality index, excluded were found to be good (Fig. 5B) using VADAR server. Verify 3D details represent the best verified and reliable model (Fig. 5C). Using ResProx server, the accuracy of model was good with statistics like: Z score of -0.25 for *BcRbr* and *BsNR* (Standard deviation of the $\chi 1$ angles), -1.34 for *BcRbr* and -1.67 for *BsNR* (Deviation of Θ angles), 0.98 for *BcRbr* and *BsNR* (Mean H-bond energy), -0.33 for *BcRbr* and *BsNR* (Percentage of bad bond length), indicates good quality structure**.** Hydropathy analysis revealed the presence of hydrophilic N and C terminals in the protein sequence (Fig. 5D).

The PDB file of successfully modeled structures of *BcRbr* and *BsNR* were further subjected to PDB sum server for structural motif assessment. Results of PDB Sum server revealed structural differences in *BcRbr* and *BsNR* modeled proteins. As shown in Figure 6, structural motif 1 helix (Leu24-Leu30), 3 gamma turns , 1 beta sheet and 1 beta hairpin (Ser15-Arg16,Val45-Asn46), 2 strands (Ser15-Arg16,Val45-Asn46), were present in 3D structure of *BsNR*, while 4 helix-helix interaction, 4 helix, and 5 beta turn were observed in *BcRbr*. The proportion of helices was more in *BcRbr* as compared to *BsNR* modeled protein, indicating associated role with large-deformation covalent bond stretching and their role in helix-helix interactions. Notably, gamma turn, beta sheet, beta hairpin and strands were observed only in *BsNR*. The 3D structures indicate their key role in protein-protein interaction, protein-ligand interaction, protein-folding, substrate binding and stability (Ackbarow *et al*., 2007). In *BcRbr* structure, it was observed that a set of amino acid sequence WGNLF of motif 1 and CNTRSR of motif 3 of *BcRbr* modeled protein were involved in formation of beta turn. A set of amino acids AVNTF from motif 1, LVIADQ from motif 2, SRLRWD from motif 3 involved in helices structure. In *BsNR*, amino acids KPHAPF from motif 1was observed in beta hairpin, whereas LVRFFA from motif 3 was observed in helix. Amino acid ISRIVR from motif 2

was involved in formation of gamma turn. In view of the high aliphatic index, a measure of the relative volume occupied by alanine, valine, isoleucine and leucine along with higher ratio (2.3:1) of (Glu+Lys) to (Gln+His) and high helix and low gamma turn content (Fig. 7), the *BcRbr* was presumed to possess thermostable property. A further detailed study on purification, optimization and characterization of *BcRbr* is underway. The structure of *BcRbr* contains one anti-parallel beta sheet that is formed because of the profound bulges present in beta strands. Hence the helix in *BsNR* is more extensively wrapped around by the beta sheet. Motif scan analysis also showed the presence of amino-acid rich profiles in *BcRbr* and *BsNR* 3D structures. In case of *BcRbr*, threonine and arginine rich profile (Thr28-Trp38 aa) was observed which mainly consists of beta turns. The *BsNR* protein showed phenylalanine rich profile (Leu20-Leu30 aa) in motif scan analysis which mainly consist of helix.

Functional analyses of 3D model proteins (*BcRbr* and *BsNR*) were carried out by ProFunc server of EMBL-EBI which depicted 3 components like cellular component (CC), biological process (BP) and biochemical function (BF). Various gene ontology (GO) terms were retrieved like intracellular region in CC category; primary metabolic process in BP category and metal ion binding, catalytic activity in BF category. Nevertheless, the *BcRbr* and *BsNR* 3D models were analyzed by comparing the model and template structure by comparing root-mean-squared deviation (RMSD) which reflects the results of a structural superimposition of the target (models) and the template structures by using superpose and chimera servers. The results of superimposition are shown in Fig. 7 (left panel). The overall RMSD were 0.795 for *BcRbr* and 0.786 for *BsNR*,obtained after superimposition of the two structures indicated the reliability of the structure.

Active site identification and metal detection

Bioinformatics' methods for detection of active ligand binding regions in the protein structure have increasingly become an area of interest. Generally, ligand binding site prediction method analyses the protein surface for binding clefts/pockets. The ligand binding sites are usually in the largest interacting cavity having major active binding surface which are defined by energetic criteria. The binding clefts/cavities of target protein were analyzed by CASTp binder site prediction tool (Fig. 8). A total of 2 possible binding clefts were predicted in *BcRbr* and 6 in *BsNR* by CASTp analysis. Prominent binding clefts are shown in Figure 8 along with active site residues where 6 active sites in *BsNR* (major cavity) contains residues like 19-ARG, 20-GLY, 21-CYS, 22-GLNand 24-LEU while 2 active sites (major cavity) in *BcRbr* contains residues like 18-PHE, 19-PRO and 21-LEU etc, suggesting that these core residues may be implicated in *BsNR* and *BcRbr* enzymatic (catalytic) functions. In addition, the site volume for every predicted binding site was also treated into account. The site volume gives information about the selected predicted binding cleft. The common residues and its position in *BcRbr* and *BsNR* proteins are also screened. No common residues were observed in *BcRbr*. However, in *BsNR* it was observed that 22-GLN, 42-ARG, 27-PHE, 41-VAL, 4-LEU, 5-SER and 30-LEU shared two common predicted sites. In addition, some active sites residues from major cavity 1 also formed an important part of structure motifs. For example, in *BcRbr* active site residues 18-PHE, 19-PRO, 21-LEU and 22-VAL were present in helix, while 31-VAL, 32-CYS, 33-ASN, 39-GLY and 40-ASN were detected in beta turns. In *BsNR* active site residues like 19-ARG, 20-GLY, 21-CYS and 22-GLN were present in gamma turns, 24-LEU and 27-PHE were detected in helix, while 39-PRO, 40-LEU, 41-VAL and 42-ARG were present in beta turns. Taking cognizance of results of MEME and CASTp-site finder, it was observed that residues LEU-21, VAL-22 of motif 2 were present in major cavity 1 in the 3D structure of *BcRbr* and ARG-19, GLY-20, CYS-21 from motif 2, LEU-24, PHE-27, LEU-40, VAL-41,ARG-42 from motif 3, suggesting that it may form an important binding pocket in the active site and/or participate in catalysis.

4. Conclusion

To conclude, the isolates from *Shilajit* sample designated EUS1, EUS2 and EUS3 were identified using morphological and biochemical studies which were further characterized by using 16S ribosomal RNA gene as *B. subtilis*, *B. cereus* and *B. clausii* and were deposited in GenBank with accession number KP126518, KP126519 and KP126520 respectively. *In-silico* analysis of genes was carried out by motif analysis and conserved domain search. NCBI-BLAST-p analysis showed that *BsNR* and *BcRbr* have novel proteins sharing homology with nitroreductase and rubrerythrin, respectively. MEME analysis revealed that presence of motifs, whereas GRAVY score depicted that *BsNR* and *BcRbr* proteins are hydrophilic. The 3D structure was generated by swiss-model server based on homology modeling. The selected model was refined by energy minimization by Swiss-pdb viewer. The stereochemical qualities of the predicted model were validated by PROCHECK server and Ramachandran plot analysis, which suggested that predicted model, is satisfactory. Active site analysis revealed the presence of five ligand binding sites. Functional analysis by ProFunc showed that BcAP protein was primarily involved in catalytic processes. Further studies on bacterial isolates,their purification and characterization are also underway to provide better insight to the function of these genes in bacteria.

5. Acknowledgments

We acknowledge honorable Vice Chancellor, Dr HS Dhaliwal and Department of Biotechnology, Eternal University, for providing necessary research facility.

6. Legend to Figure

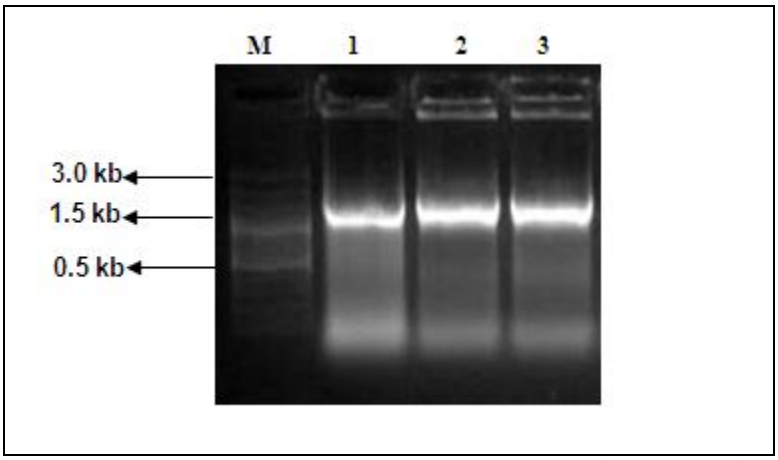


Figure 1: Amplified product of 16S rRNA gene. Lane M: 100bp plus ladder, Lane 1-3: Amplified product of EUS1, 2 and 3 (bacterial isolates)

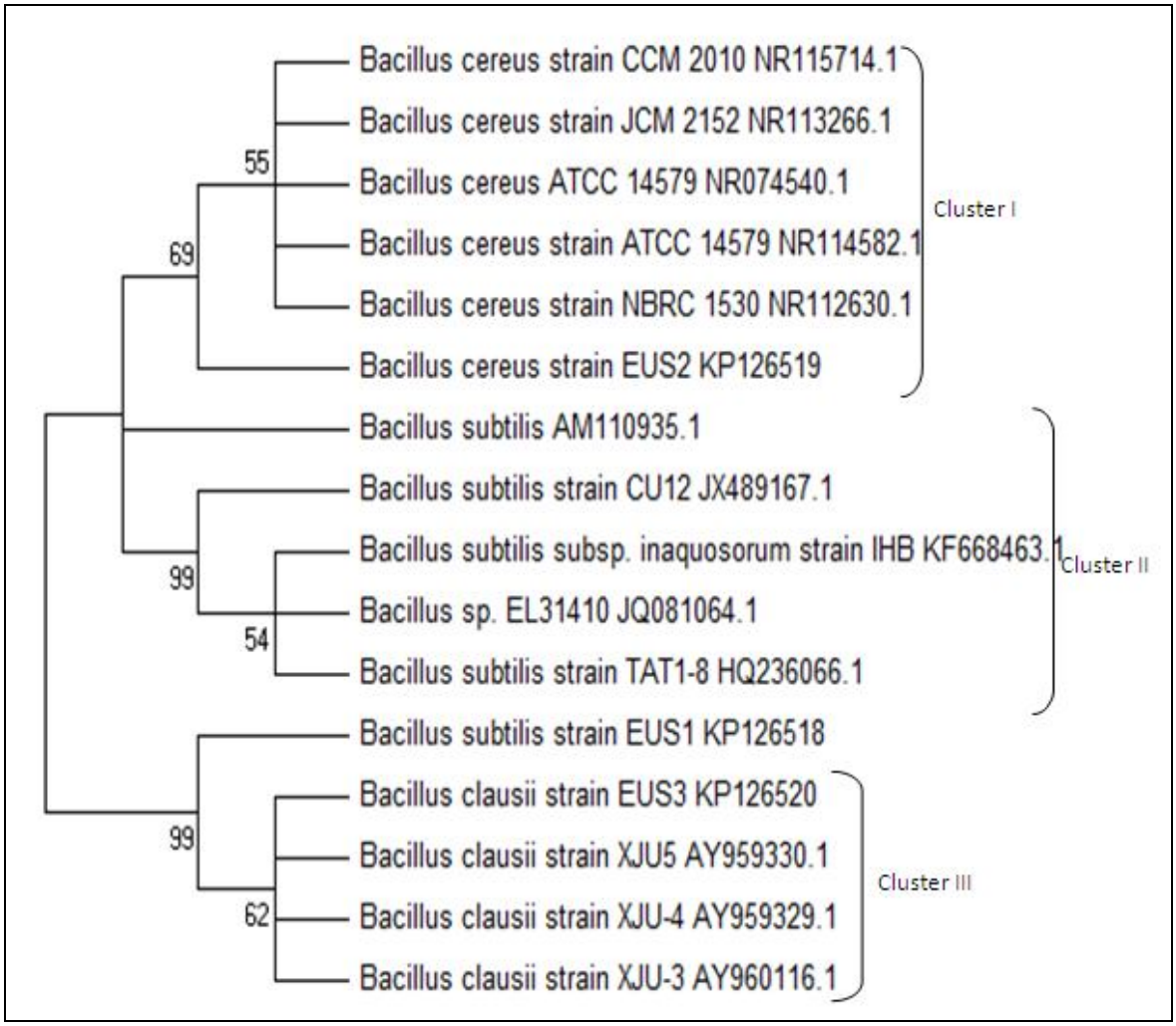


Figure 2: phylogenetic tree based on the 16s rRNA gene sequences between isolates belonging to genus Bacillus. The tree was generated by using the Neighbor-Joining method. The bootstrap consensus tree inferred from 500 replicates is taken and branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analysis involved 16 nucleotide sequences. All positions containing gaps and missing data were eliminated. Nucleotide sequence accession numbers are indicated with strains

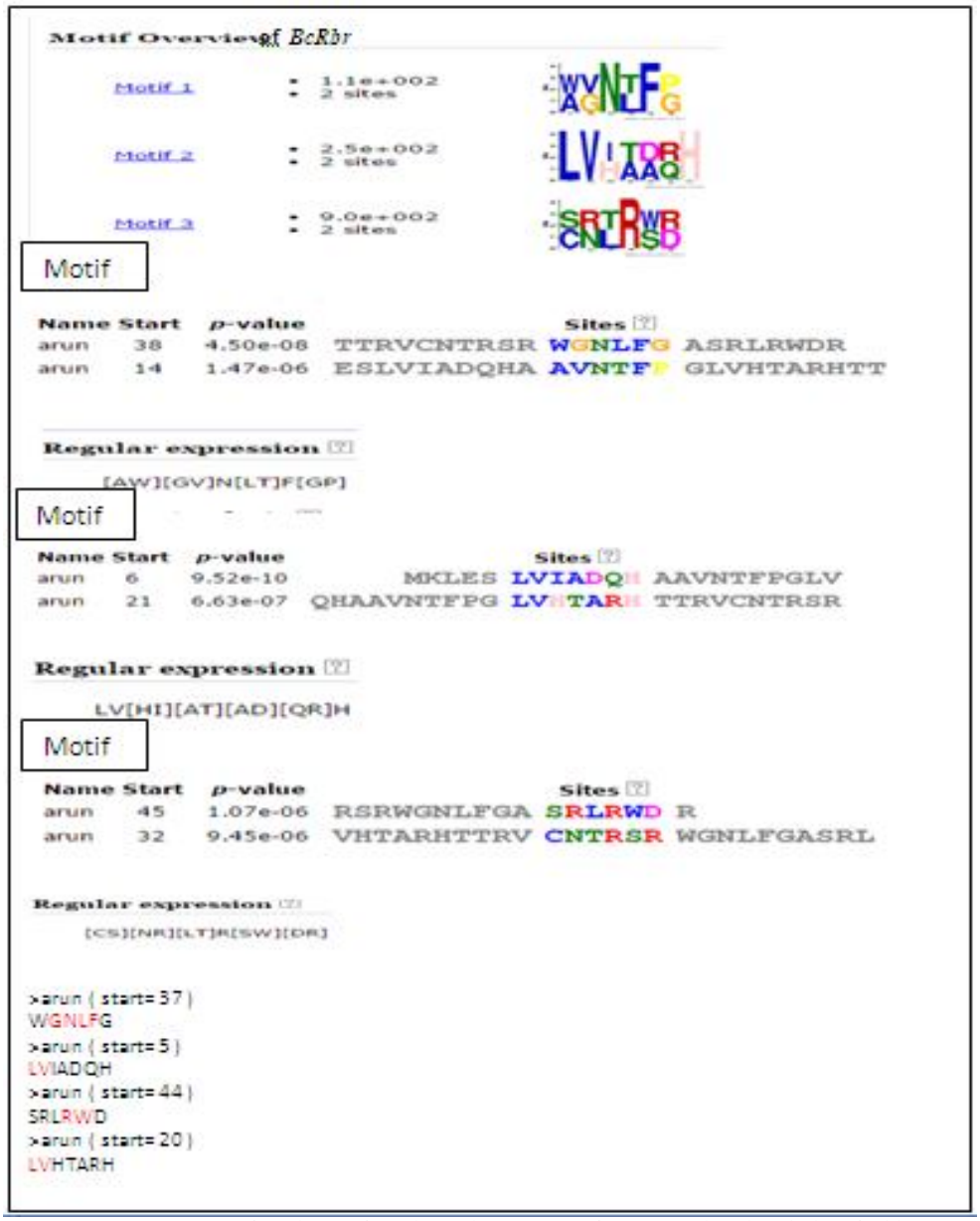


Figure 3A: Protein motif analyses of BcRbr and BsNR. Motifs were analyzed and identified by MEME.

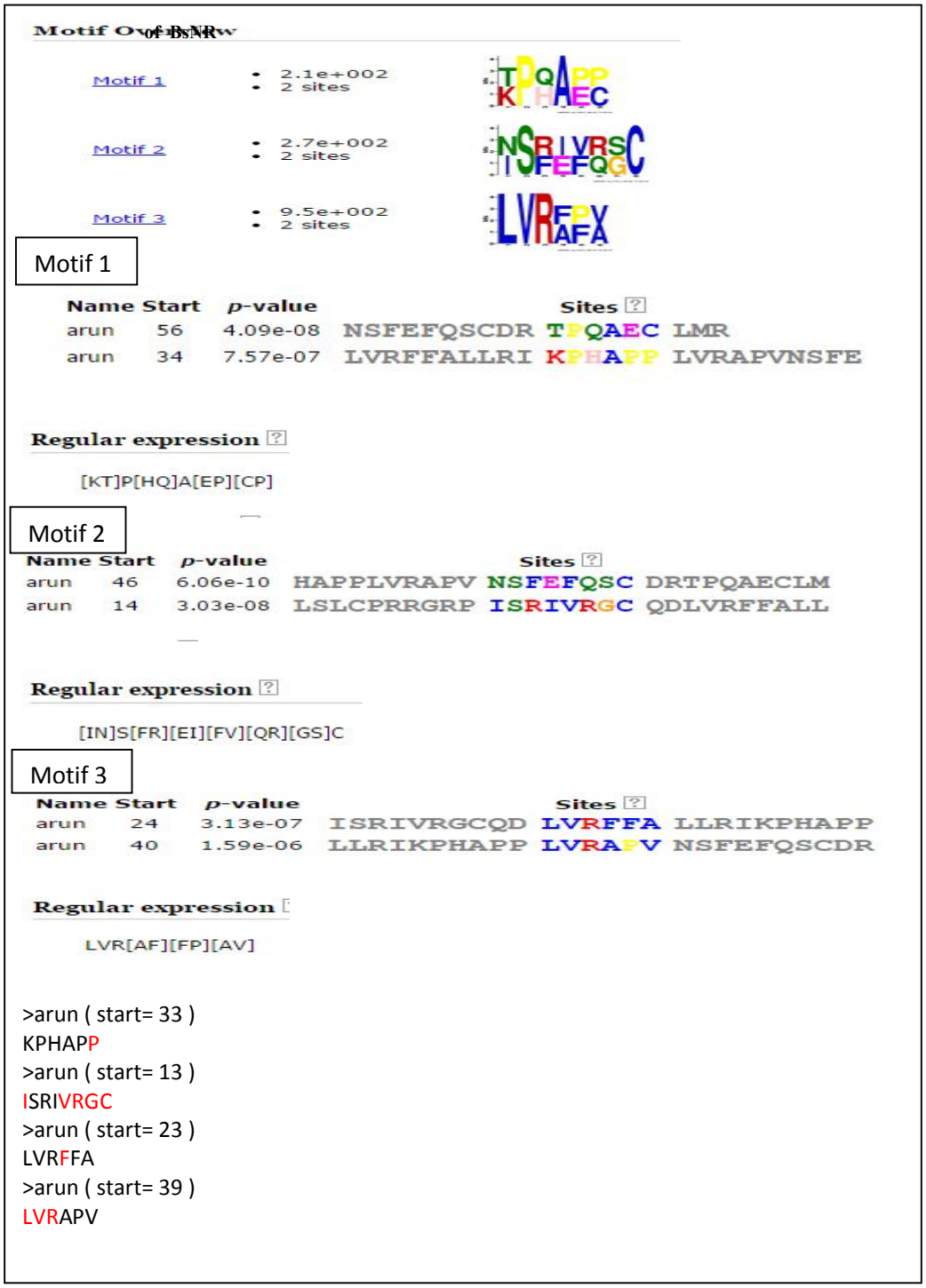


Figure 3B: Protein motif analyses of BcRbr and BsNR. Motifs were analyzed and identified by MEME.

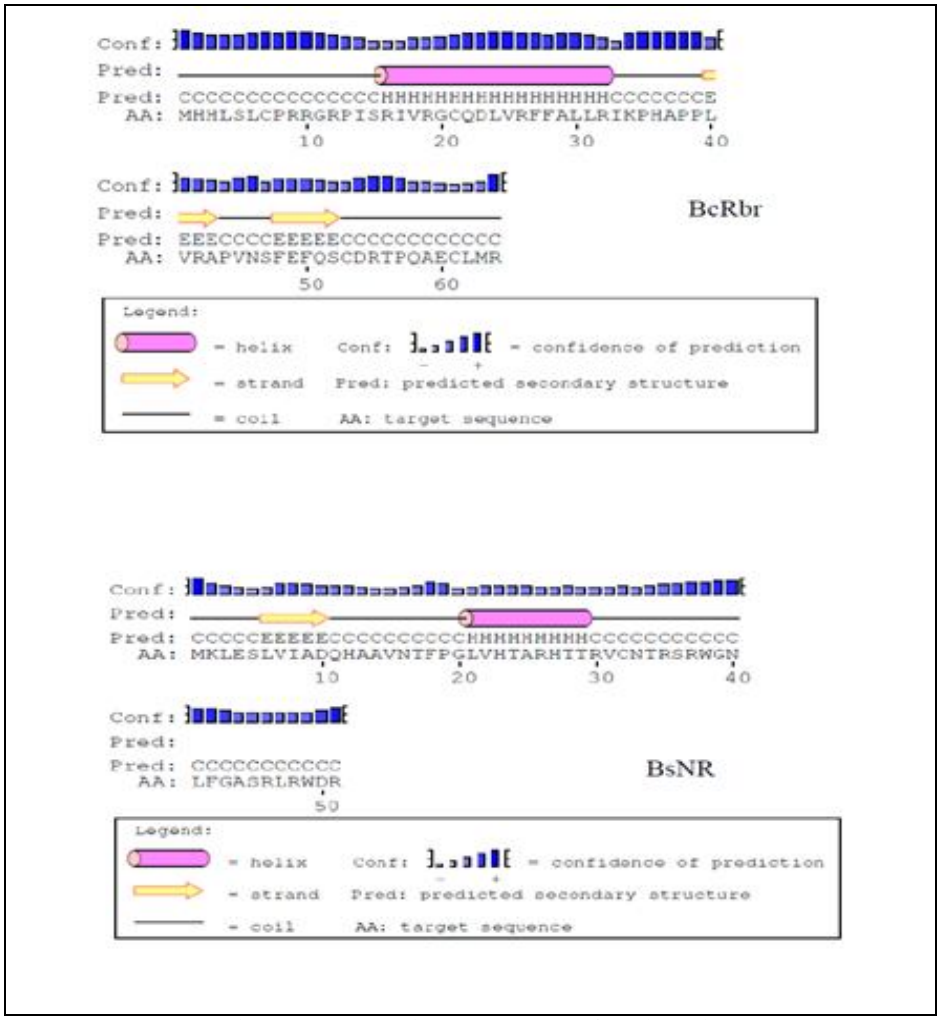


Figure 4: Secondary structure prediction PSIPRED of BcRbr and BsNR sequences

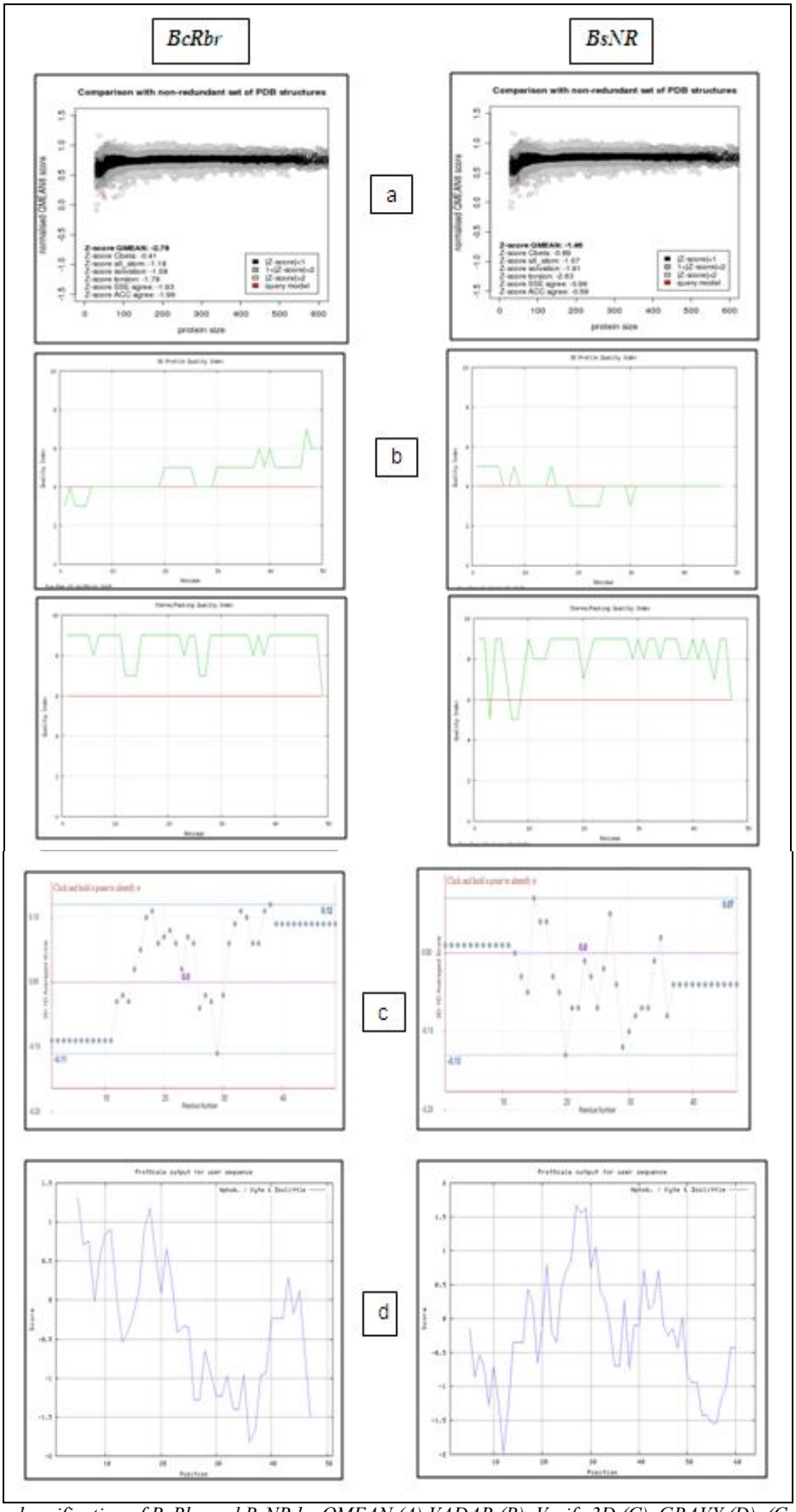
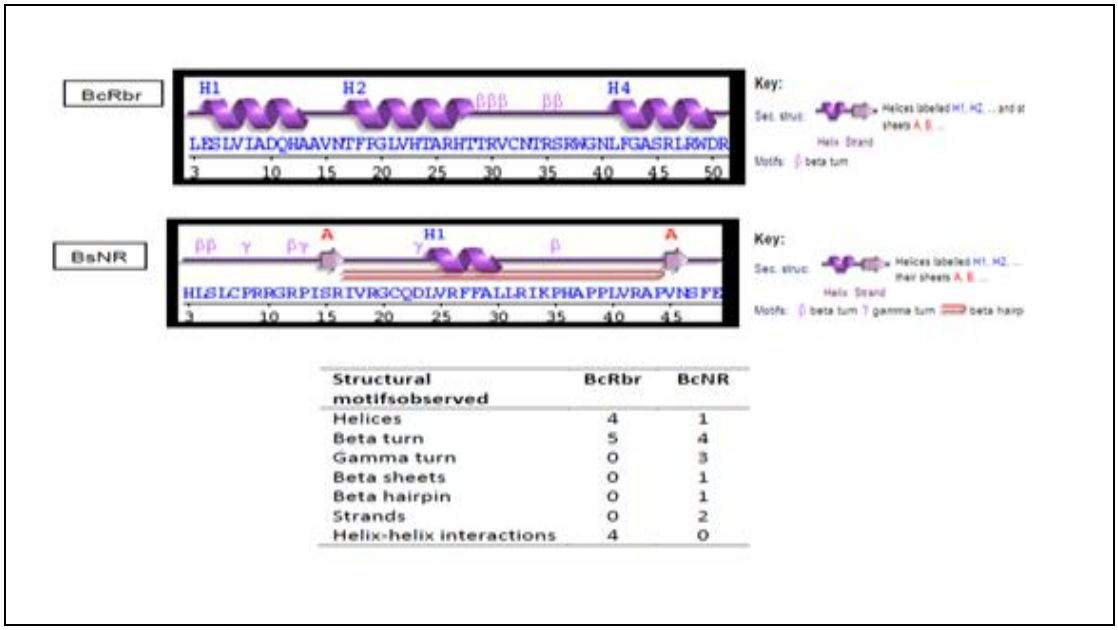


Figure 5: Structural verification of BcRbr and BsNR by QMEAN (A) VADAR (B), Verify 3D (C), GRAVY (D). (Color figure online)



Structural motifsobserved	BcRbr	BcNR
Helices	4	1
Beta turn	5	4
Gamma turn	0	3
Beta sheets	0	1
Beta hairpin	0	1
Strands	0	2
Helix-helix interactions	4	0

Figure 6: Structure motif analysis of BcRbr and BsNR 3D structures generated by PDBsum server (Color figure online)

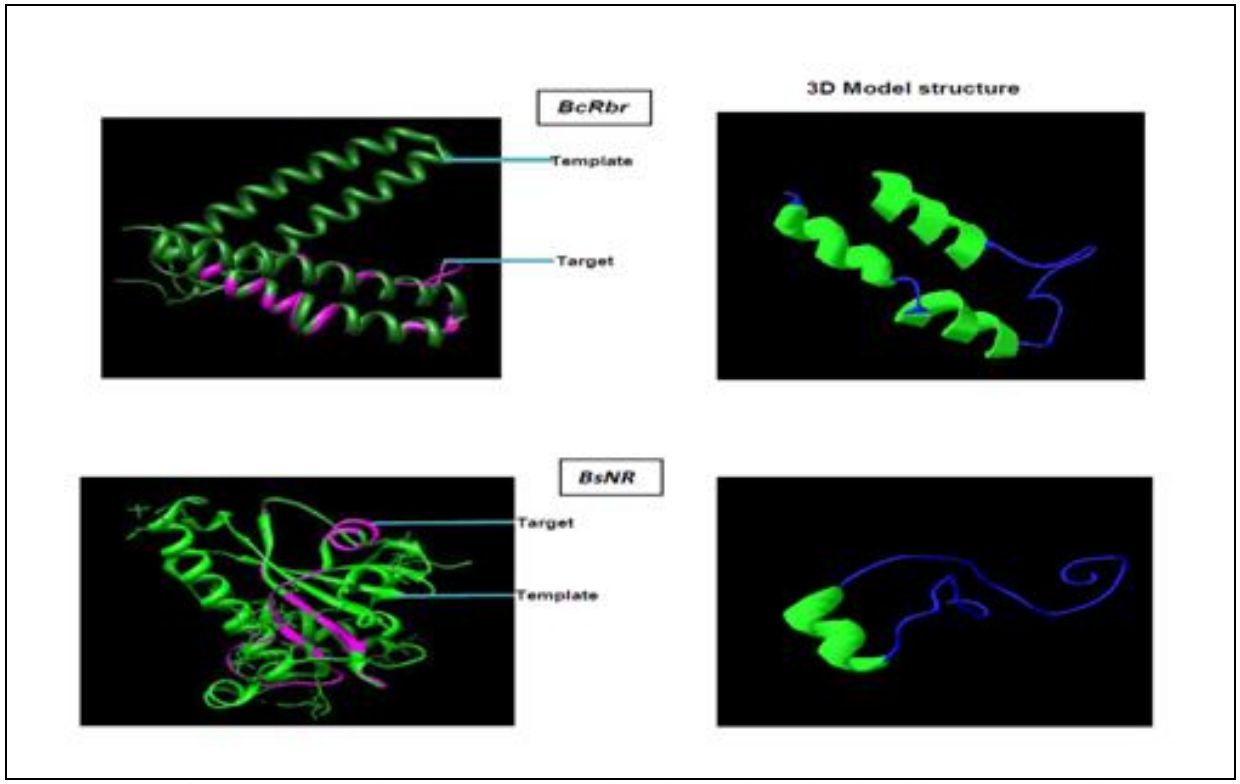
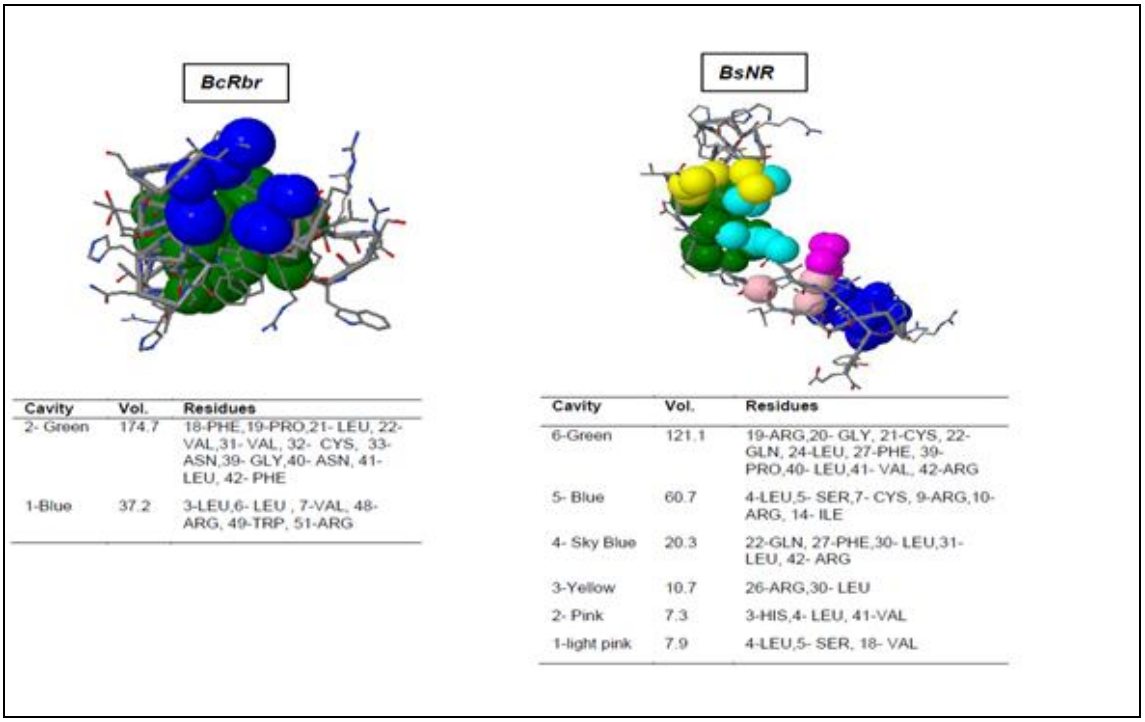


Figure 7: Superimposition of 3D structures BcRbr and BsNR (target and templates). The predicted models were generated by superpose and viewed by Swiss-PDB viewer. Left panel depicts superimposition and Right panel depicts the 3D model of BcRbr and BsNR presented as ribbon structure. Helices are in green, and coils are in blue color (Color figure online)



Cavity	Vol.	Residues
2- Green	174.7	18-PHE,19-PRO,21- LEU, 22-VAL,31- VAL, 32- CYS, 33-ASN,39- GLY,40- ASN, 41-LEU, 42- PHE
1-Blue	37.2	3-LEU,6- LEU , 7-VAL, 48-ARG, 49-TRP, 51-ARG

Cavity	Vol.	Residues
6-Green	121.1	19-ARG,20- GLY, 21-CYS, 22-GLN, 24-LEU, 27-PHE, 39-PRO,40- LEU,41- VAL, 42-ARG
5- Blue	60.7	4-LEU,5- SER,7- CYS, 9-ARG,10-ARG, 14- ILE
4- Sky Blue	20.3	22-GLN, 27-PHE,30- LEU,31-LEU, 42- ARG
3-Yellow	10.7	26-ARG,30- LEU
2- Pink	7.3	3-HIS,4- LEU, 41-VAL
1-light pink	7.9	4-LEU,5- SER, 18- VAL

Figure 8: Major binding clefts in BcRbr and BsNR 3D structures

>Seq1 *BsNR*	MKLESLVIADQHAAVNTFPGLVHTARHTTRVCNTRSRWGNLFGASRLRWDR*
>Seq2 *BcRbr*	MHHLSLCPRRGRPISRIVRGCQDLVRFFALLRIKPHAPPLVRAPVNSFEFQSCDRTPQAECLMR*

Table 1: Protein sequence as retrieved from ORF finder tool of NCBI

Physiochemical properties	***BcRbr***	***BsNR***
Number of amino acids	51	64
Molecular weight	5847.6	7425.8
Theoretical pI	11.61	11.09
Total number of negatively charged residues (Asp + Glu)	3	4
Total number of positively charged residues (Arg + Lys)	8	11
Ext. coefficient	11000	250
Abs 0.1% (=1 g/l)	1.881, assuming all pairs of Cys residues form cysteine	0.034, assuming all pairs of Cys residues form cysteine
The estimated half-life is	30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo) >10 hours (Escherichia coli, in vivo)	30 hours (mammalian reticulocytes, in vitro) >20 hours (yeast, in vivo) >10 hours (Escherichia coli, in vivo)
The instability index (II)	49.50	71.48
Aliphatic index	78.43	85.31
Grand average of hydropathy (GRAVY)	-0.406	-0.208

Table 2: Physiochemical properties of BcRbr from Bacillus cereus and BsNR from Bacillus subtilis

7. References

i. Ackbarow, T., Chen, X., Keten, S. & Buehler, M.J. (2007) Hierarchies, multiple energy barriers, and robustness govern the fracture mechanics of α-helical and β-sheet protein domains. PNAS, 104, 16410-16415

ii. Ash, C., Farrow, J.A.E., Wallbanks, S., & Collins, M.D. (1991) Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small subunit-rRNA sequences.Lett.Appl. Microbiol., 13, 202–206

iii. Bhattacharaya, S.K. (1995) Shilajit attenuates streptozotocin induced diabetes mellitus and decrease in pancreatic islet superoxide dismutase activity in rats. Phytother. Res.,9, 41–44

iv. Bizani, D., Motta, A.S., Morrissy, J.A.C., Terra, R.M., Souto A.A., & Brandelli, A. (2005) Antibacterial activity of cerein 8A, a bacteriocin-like peptide produced by *Bacillus cereus*. Intl. Microbiol., 8, 125-131

v. Bryant, C., & DeLuca, M. (1991) Purification and Characterization of an Oxygen-insensitive NAD(P)H Nitroreductase from *Enterobacter cloacae*. J. Biol. Chem., 266, 4119-4125

vi. Bryant, D.W., McCalla, D.R., Leeksma M. & Laneuville P. (1981) Type Initroreductases of Escherichia coli. Can. J.Microbiol., 27, 81-86

vii. Cihan, A., Tekin, N., Ozcan, B., & Cokmus C. (2012) The genetic diversity of genus bacillus and the related genera revealed by 16S rRNA gene sequences and ARDA analyses isolated from geothermal regions of Turkey. Braz. J. Microbiol., 43, 309-324

viii. Coulter, E.D. & Kurtz, J.D.M. (2001) A role for rubredoxin in oxidative stress protection in *Desulfovibrio vulgaris*: catalytic electron transfer to rubrerythrin and two-iron superoxide reductase. Arch. Biochem. Biophys., 394, 76-86

ix. Coulter, E.D., Shenvi, N.V., Beharry, Z.M., Smith, J.J., Prickril B.C., & Kurtz, D.M., (2000) Rubrerythrin-catalyzed substrate oxidation by dioxygen and hydrogen peroxide. Inorg.Chim.Acta., 297, 231-241

x. deMare´, F., Kurtz, J.D.M., & Nordlund, P.(1996) The structure of *Desulfovibrio vulgaris* rubrerythrin reveals a unique combination of rubredoxin-like FeS4 and ferritin-like diiron domains. Nat Struct Biol, 3, 539–546

xi. Ghosal, S., Lal, J., Jaiswal, A.K., & Bhattacharya, S.K. (1993) Effects of Shilajit and its active constituents on learning and memory in rats. Phytother. Res., 7, 29–34

xii. Ghosal, S., Lal, J., Singh, S.K., Dasgupta, G., Bhaduri, J., Mukhopadhyay, M. & Bhattacharya, S.K. (1989) Mast cell protecting effects of shilajit and its constituents.Phytother. Res. 3:249–252

xiii. Ghosal, S., Reddy, J.P., & Lal, V.K. (1976) Shilajit I: Chemical constituents. J. Pharm. Sci., 65, 772–773

xiv. Goel, R.K., Banerjee, R.S. & Acharya, S.B. (1990) Antiulcerogenic and antiinflammatory studies with Shilajit. J. Ethnopharmacol., 29, 95-103

xv. Graumann, P.L. (2007) Cytoskeletal elements in bacteria. Annu. Rev. Microbiol,61, 589-618

xvi. Jaiswal, A.K., & Bhattacharya, S.K. (1992) Effects of Shilajit on memory, anxiety and brain monoamines in rats. Ind. J.Pharmacol., 24, 12-17

xvii. Kageyama, Y., Takaki, Y., Shimamura, S., Nishi, S., Nogi, Y., Uchimura, K., Kobayashi, T., Hitomi, J., Ozaki, K., Kawai, S. & others (2007) Intragenomic diversity of the V1 regions of 16S rRNA genes in high-alkaline protease-producing *Bacillus clausii* spp. Extremophiles, 11,597-603

xviii. Kong, Y.C., But, P.P.H., Ng, K.H., Cheng, K.F., Cambie R.C., & Malla S.B., (1987) Chemical Studies on a Nepalese Panacea-Shilajit (I). Pharma.Biol., 25, 179-182

xix. Kurtz, D.M. (1997) Structural similarity and functional diversity in diiron-oxo proteins.JBIC ,2, 159-167

xx. LeGall, J., Prickril, B.C., Moura, I., Xavier, A.V., Moura, J.J., & Hanh, H.B. (1988) Isolation and characterization of rubrerythrin, a non-heme iron protein from Desulfovibrio vulgaris that contains rubredoxin centers and a hemerythrin-like binuclear iron cluster. Biochem., 27, 1636-1642

xxi. Meena, H., Pandey, H.K., Arya M.C., & Ahmed, Z. (2010) Shilajit: A panacea for high-altitude problems. Int. J. Ayurveda, 1, 37-40

xxii. Nielsen, P, Fritze, D. & Priest, F.G. (1995). Phenetic diversity of alkaliphilic *Bacillus* strains: proposal for nine new species. Microbiol, 141, 1745–1761

xxiii. Patel, J.B. (2001) 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. Mol. Diagn., 6, 313-321

xxiv. Porwal, S., Lal, S., Cheema, S. & Kalia, V.C. (2009) Phylogeny in Aid of the Present and Novel Microbial Lineages: Diversity in Bacillus. PLoSONE, 4, 4438

xxv. Saitou, N., & Nei. M. (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol.,4, 406-425

xxvi. Schepetkin, I.A., Xie, G., Jutila, M.A., & Quinn, M.T., (2009) Complement-fixing activity of fulvic acid from Shilajit and other natural sources. Phytother. Res., 23, 373–384

xxvii. Senesi, S., Celandroni, F., Tavanti, A., & Ghelardi, E. (2001) Molecular characterization and identification of *Bacillus clausii* strains marketed for use in oral bacteriotherapy. Appl. Environ.Microbiol., 67, 834–839

xxviii. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol., 28, 2731-2739

xxix. Thompson, D., Higgins, D.G. & Gibson, T.J. (1994) Clustal W, improving the sensitivity of progressive multiple sequences alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res., 22, 4673–4680

xxx. Vucskits, A.V., Hullár, I., Bersényi, A ., Andrásofszky, E., Kulcsár, M. & Szabó J. (2010) Effect of fulvic and humic acids on performance, immune response and thyroid function in rats.J. Anim. Physiol. Anim. Nutr., 94, 721-728

xxxi. Wilson, K. (1987) Preparation of genomic DNA from bacteria. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K, (eds) Current Protocols in Molecular Biology, 2.4.1–2.4.5. John Wiley & Sons, New York.

xxxii. Yang, E., Fan, L., Jiang, Y., Doucette, C. & Fillmore, S. (2012) Antimicrobial activity of bacteriocin-producing lactic acid bacteria isolated from cheeses and yogurts. AMB Express, 2, 48