



## ***In-silico* analysis and homology modeling of Antifreeze protein type 1, Choline dehydrogenase and Betaine aldehyde dehydrogenase of an extreme psychrophile: *Psychromonas ingrahamii* 37**

<sup>\*1</sup>AnushreeSrivastava, <sup>2</sup>Amita Kashyap

<sup>1,2</sup>Department of Biotechnology and Bioinformatics, BioAxis DNA Research Center (P) Ltd, Lucknow

\*Email: anushree.srivastava02@gmail.com

Received: September 23<sup>rd</sup> 2013, Accepted: October 20<sup>th</sup> 2013, Published: November 1<sup>st</sup> 2013

### **Abstract:**

Antifreeze proteins (AFPs) are the class of polypeptides, facilitating cold-adaptation in psychrophilic organisms, whereas, Choline dehydrogenase (CHDH), and Betaine-aldehyde dehydrogenase (BADH) catalyze the production of a common osmoprotectant, glycine betaine, which owes for organism's freeze tolerance as well. The present Bioinformatics study selects three proteins: type 1 AFP, CHDH, and BADH from an extreme psychrophile, *Psychromonas ingrahamii* 37 (*Pi37*) and performs a systematic *in-silico* analysis to shed light on the evolutionary as well as the functional aspects of these proteins. Multiple sequence alignment (MSA) shows type 1 *Pi*AFP with very low, whereas, *Pi*CHDH and *Pi*BADH with high residue conservation across BLASTp homologs, and phylogenetic trees generated through CLC workbench presented evolutionary perspective on query proteins. MSA identifies key regulatory residues/patterns in query proteins that are evolutionarily conserved across BLASTp organisms, which suggest their functional significance. Primary and secondary structure analysis has been performed. Conserved domain analysis predicts that type 1 *Pi*AFP, *Pi*CHDH, and *Pi*BADH belong to Band\_7\_5, GMC\_C-terminal, and ALDH\_F9\_TMBADH protein families, respectively, and their possible functions have been inferred. GMC oxidoreductase *signature 1* and 2, and ALDH *cysteine* and *glutamic acid* are catalytic sites in *Pi*CHDH and *Pi*BADH, respectively. Homology modeling of these protein and their domains provide a fairly good hypothetical models of the query proteins and their functional domains for the future investigations. Protein-protein interaction prediction, using STRING, yields functional partners of query proteins. The present work is an initial step towards comprehending the molecular basis of extreme adaptation of *Pi37*, and additional works are required, in light of the significant data and inferences presented here, for

complete understanding of the structure and function of these proteins, which may have bearing on the extreme psychrophily and halophily of *Pi37*.

**Keywords:** *Antifreeze proteins; computational analysis; homology modeling; osmoregulation; protein domains; protein-protein interaction, proteomic tools.*

### **Introduction:**

Psychrophiles have been studied primarily to understand biological mechanisms of adaptation to extremely cold conditions. In reaching for an understanding of how life processes work at extremes of temperature, most of the focus to date has been on the properties of enzymes of extremophiles<sup>1</sup>.<sup>2</sup> Recently, few studies on the sea ice bacterium *Pi37* have illustrated the species characterization<sup>3</sup>, whole genomics<sup>4</sup>, subfreezing growth<sup>6</sup>, and enzyme catalyzed cold-adaptation<sup>7</sup> of this extreme psychrophile. The extreme psychrophile, *Pi* was isolated from sea ice from the Arctic<sup>4</sup>. Strain 37<sup>T</sup> was considered to represent a novel species, which was provisionally named *Pi* sp.nov<sup>1</sup>. It grows exponentially with a doubling time of 240 hours at -12°C and may well grow at even lower temperatures<sup>6</sup>. It has been found that cold-hardy organisms possess some proteins and enzymes which are active at low temperatures<sup>7</sup>, and facilitate their survival in such extreme environments.

AFPs were initially characterized in marine fishes<sup>8,9</sup> where they protect their hosts from freezing by binding to, and preventing the growth of, sea ice crystals<sup>10</sup>. AFPs lower the freezing temperature of a solution containing ice below the melting point of the ice. This difference between the freezing and melting temperatures is called thermal hysteresis (TH) and is used as a measure of antifreeze activity. AFPs were subsequently found in freeze-tolerant organism<sup>11,12</sup> where, rather than preventing freezing, they stop ice crystals in frozen tissues from growing larger through the process of ice recrystallization (IRI)<sup>13</sup>. Another

feature of a typical psychrophile is its ability to produce an intracellular solute to counterbalance the osmotic effect of higher salinities in brine pockets where organisms reside that are encountered as sea ice freezes<sup>14</sup>

Bacteria and plants often accumulate choline, since it can easily be converted to glycine betaine, a common osmoprotectant<sup>15</sup>. While the conversion of choline to glycine betaine for the purpose of accumulating the latter as an osmoprotectant is considered biosynthetic, the conversion of choline to glycine betaine in this pathway is performed by the action of two dehydrogenases –CHDH and BADH. When the bacteria grow at normal osmolarity conditions, they degrade choline, using it as a carbon and nitrogen source. However, under inhibiting osmotic conditions enzyme activities, which lead to glycine betaine degradation, decrease, while enzyme activities that convert choline to glycine betaine either remain constant or increase. In this way, a high concentration of glycine betaine can be maintained in osmotically stressed cells<sup>16</sup>. Hence, the present computational analysis and homology modeling of type 1 *PiAFP*, *PiCHDH*, and *PiBADH* may provide essential insights into the molecular basis of cold-adaptation and osmoprotection in *Pi37*, and the information reported here may open avenues to further explorations of the grounds of unique adaptation of this bacterium.

#### Materials and Methods:

##### Data Mining and Sequence Analysis:

Amino acid sequences of type 1AFP, CHDH, and BADH from *Pi37* were obtained from protein sequence database of National Center for Biotechnology Information (NCBI) and blasted<sup>17</sup> against Protein Data Bank (PDB) entries to find out similar sequences. BLASTp yielded homologous sequences for each target protein sequences, which have been further used in multiple sequence alignment through CLC-Work Bench tool<sup>18</sup> to find out the degree of sequence conservation between target and homologous sequences. CLC-Work Bench tool facilitated phylogenetic analysis through dendrograms. Functional domains and motifs in query sequences were identified by SMART<sup>19, 20</sup> and Motif Scan<sup>21</sup> respectively. Significant functional partners of type 1AFP, CHDH, and BADH were identified by the prediction of protein-protein interaction using STRING<sup>22</sup>.

##### Protein Characterization and Topology Prediction:

Physico-chemical properties such as molecular weight, theoretical pI, total number of negatively (Asp+Glu) and positively (Arg+Lys) charged residues, extinction coefficients<sup>23</sup>, instability index<sup>24</sup>, aliphatic index<sup>25</sup>, and grand average of hydropathicity (GRAVY)<sup>26</sup> of the target proteins were computed using Expasy's ProtParam Proteomics server (<http://web.expasy.org/protparam/>). Secondary structures of type 1 *PiAFP*, *PiCHDH*, and *PiBADH* were calculated with SOPMA<sup>27</sup> (Self Optimized Prediction Method with Alignment) and disulfide bonds were predicted by the Cys\_REC<sup>28</sup> tool from Softberry.

##### Homology Modeling and Evaluation:

Tertiary structures of query proteins were predicted using CPH<sup>29, 30</sup> HH-pred<sup>31</sup> and Phyre<sup>32</sup>. One best representative PDB ID for each protein was selected and later used for structure validation. Different template sequences for *PiAFP*, *PiCHDH*, and *PiBADH* were identified and evaluated for their identities and expected value using Swiss Model. Stereochemical quality and accuracy of the selected models was evaluated with PROCHECK<sup>33</sup> by Ramachandran plot analysis<sup>34</sup>. Models were evaluated on the basis of G-factor, number of residues in core, allowed, generously allowed and disallowed regions. Annotia value<sup>35</sup>, Qmean value<sup>36-38</sup>, Errat value<sup>39</sup>, and D-fire Energies<sup>40</sup> for each template of each query protein were computed and analyzed. For each evaluation tool, one best model was selected and the best model for each target protein was selected on the basis of final evaluation. Finally, the 3D structure of each query protein was visualized with RASMOL (<http://openrasmol.org/>).

##### Homology Modeling of Functional Domains:

Sections of the *PiAFP*, *PiCHDH*, and *PiBADH* sequence were submitted to the Phyre2 server<sup>32</sup>. Phyre2 uses the Hidden Markov method to generate alignments of a submitted protein sequence against proteins with published structures<sup>41</sup>. The resulting alignments are then used to produce homology-based models of the query sequence to predict its three-dimensional structure. In addition, Phyre2 uses an *ab-intio* folding simulation called Poing to model regions of a query with no detectable similarities to known structures<sup>42</sup>. Poing combines multiple templates of known structures to produce the final model of the query sequence. The model is judged to be accurate when over 90% of the submitted residues are modeled at >90% confidence<sup>32</sup>.

**Results and Discussion:****Phylogenetic Analysis**

Linear chains of type 1 *PiAFP* (Accession No.: YP\_942732), *PiCHDH* (Accession No.: ABM03823) and *PiBADH* (Accession No.: ABM03822), of 351, 562, and 481 amino acid length respectively, were retrieved from NCBI protein sequence database. The residues of query proteins were further subjected to BLASTp analysis which revealed sequence homology to target proteins. Eight organisms for type 1 *PiAFP* and 10 organisms for both *PiCHDH* and *PiBADH* were selected on the basis of maximum score, identity and least expected value. For type 1 *PiAFP*, *Paenibacillus lactis 154* shared 43% sequence identity with score 749 (greatest). For *PiCHDH* and *PiBADH*, *Shewanella frigidimarina NCIMB400* showed greatest sequence identity (87%) and score (2166). Phylogenetic analysis was conducted to explore sequences' shared evolutionary origin. MSA and Dendrograms (figure 1) generated by CLC work bench, revealed that type 1 AFP from *Paenibacillus lactis 154* is evolutionary closest to type 1 *PiAFP*, which is also validated by BLASTp analysis. Among 10 selected organisms, CHDH from *Moritella sp. PE36*, and BADH from *Bermanella marisrubri* and *Colwellia psychrerythraea 34H* are found to be phylogenetically closest to *PiCHDH* and *PiBADH* respectively.

Sequence analysis aims at finding important sequence similarities that would allow one to infer homology. This approach is used in predicting potential functions for a newly sequenced gene and its protein product and understanding their cellular roles routinely relies on family history<sup>43</sup>. As type 1 *PiAFP* shares limited sequence identity with other bacterial species, alignment of their sequences also yielded very low conservation (15% residues at 100% conservation) among residues. However, even for proteins that share a much lesser degree of identity and even a single amino acid residue is conserved in all the sequences, if they align unequivocally are indisputable homologs<sup>43</sup>. Unexpectedly, type 1 *PiAFP* is found to be phylogenetically closest to type 1 AFP of *Paenibacillus lactis 154* (43% sequence identity) in spite of the fact that *Paenibacillus lactis 154* is not a psychrophile but may be a mesophile with optimal growth temperature of 30-40°C<sup>44</sup> and belongs to a different taxonomic order than to a stenopsychrophile<sup>14</sup>: *Pi37*. Contrastingly, *PiCHDH*, and *PiBADH* showed high level of sequence identity, 72-87% and 74-87% respectively (with 55% and 72% residues conserved at 100% conservation,

respectively). Furthermore, *PiCHDH* has closest evolutionary relationship with CHDH of *Moritella sp. strain PE36*, which is a piezophile (thrives at high pressures) with optimal growth temperature of 15°C<sup>45</sup>. Also, *PiBADH* shows closest evolutionary link to BADH of a mesophile - *Bermanella marisrubri*<sup>46</sup> and a psychrophile - *Colwellia psychrerythraea 34H*. Like *Pi37* they all belong to Gamma-proteobacteria<sup>47</sup>.

Sequence or structural homology may not be strictly considered as evidence of common origin, as a logically consistent alternative does exist and involves convergence from unrelated sequences, which posits that sequence and structural similarities between proteins are observed because the shared features are strictly required for these proteins to perform their identical or similar functions<sup>48, 49</sup>. As convergence could only be a realistic possibility for deep relationships between proteins, which involve limited similarities<sup>43</sup>, homology between type 1 *PiAFP* and type 1 AFP of *Paenibacillus lactis 154* may be a product of functional convergence, which seems to be a plausible inference from the results of phylogenetic study of type 1 *PiAFP* of the present work. On the other hand, the higher the similarity between two or more sequences, the lower the probability that they have originated independently of each other and became similar merely by chance; therefore, it is concluded that the sequences in question must share a common ancestry<sup>43</sup>. Here, the high level conservation occurs over a long stretches of *PiCHDH* and *PiBADH*, sequences and their homologs, suggests that they are derived from the same ancestral residue and, typically, inherit their functions. Whenever statistically significant sequence or structural similarity between proteins or protein domains is observed, this is an indication of their divergent evolution from a common ancestor<sup>43</sup>. Conserved regions of an alignment are structurally and functionally important<sup>50</sup>. Identified conserved patterns (table 1) within the range of functional domains for each query protein, suggest that they significantly partake into their respective biological function across species yielded by BLASTp.

Figure: 1 Dendrograms have been generated for PiAFP pper], PiCHDH [mid], and PiBADH [below], sequences.

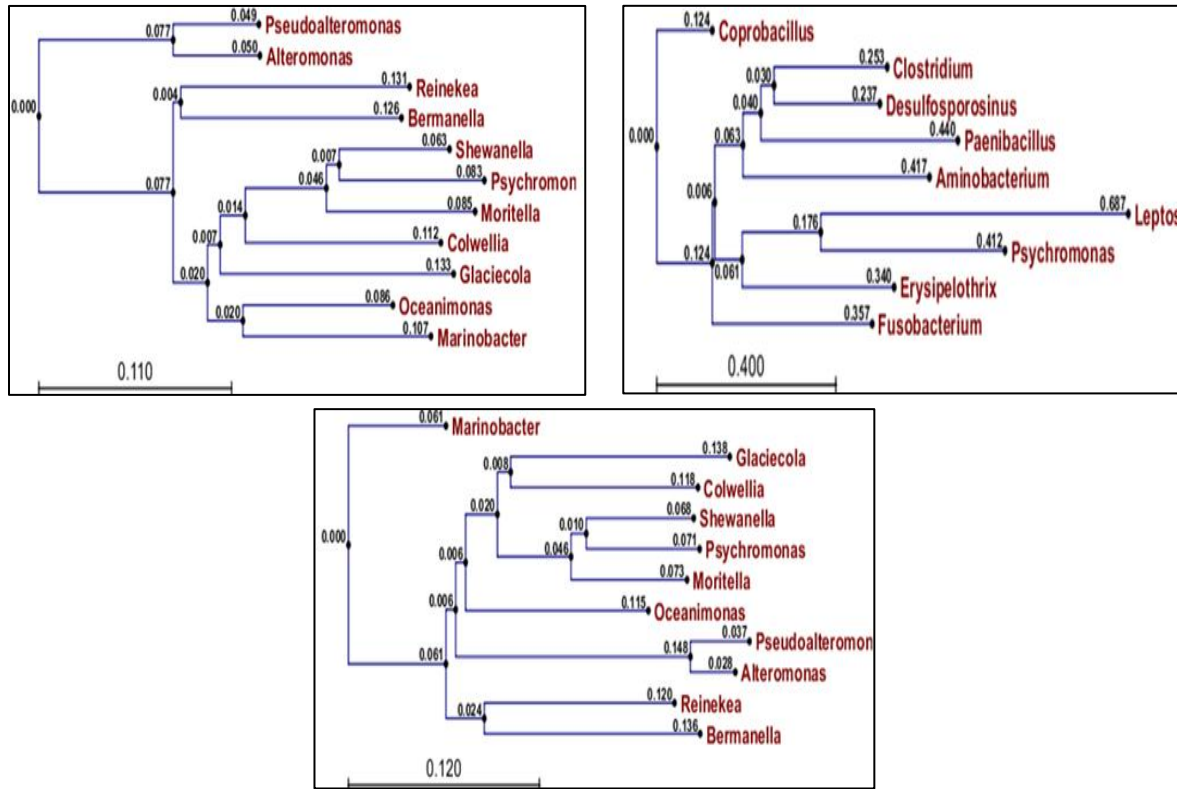


Table: 1 This table identifies key regulatory residues and patterns in query proteins that are evolutionarily conserved (C.) across BLASTp organisms. The residues and patterns reported here are at 100% conservation, yielded by CLC workbench tool.

Type 1 PiAFP													
C. Pat.	Region	C. Pat.	Region	C. Pat.	Region	C. Pat.	Region	C. Pat.	Region	C. Pat.	Region	C. Pat.	Region
M	1	K	43	L	66	Y	110	D	162	I	212	D	248
W	18	G	45	PF	79-80	G	120	S	165	D	218	AA	253-254
Q	30	D	49	E	83	D	128	E	184	L	228	N	256
V	33	G	54	NK	88-89	F	132	S	186	E	233	G	266
Q	37	L	58	WGT	97-99	VG	137-138	G	201	G	238	G	272
A	39	NIP	62-64	PI	102-103	F	143	F	207	Y	241		
PiCHDH													
C. Pattern	Region	C. Pattern	Region	C. Pattern	Region	C. Pattern	Region	C. Pattern	Region	C. Pattern	Region	C. Pattern	Region
M	1	GPL	143-145	GVG	288-290	CVR	424-426						
DYII	7-10	NNGN	148-151	NLQDHLEFYFQ	292-302	TREII	428-432						
GAGSAGCVLANRL	12-24	AGY	168-170	GLG	333-335	EIQPG	444-448						
ETGGSDK	36-42	DYN	175-177	TNHFESC	337-343	FVR	460-462						
IFI	44-46	QEGFG	181-185	FIRSK	345-349	ESAYHPSC	466-473						
MMPTALSIPMN	48-57	MHMT	187-190	QYHFLPAAMRYDG	358-370	LRVVDSSIFP	497-506						
AWQF	62-65	NLT	212-214	GFQ	378-380	PNGNLN	509-514						
MHCPRGKVLGGSSSINGMYYVRGHA	78-102	GSPH	259-262	GHNPKPSRG	384-392	PTIM	516-519						
LPYF	122-125	LQLSGIG	264-270	FNYL	409-412	AAD	524-526						
KAE	127-129												
PiBADH													
C. Pattern	Region	C. Pattern	Region	C. Pattern	Region	C. Pattern	Region	C. Pattern	Region	C. Pattern	Region	C. Pattern	Region
M	1	DFYYTR	133-138	KVSFTGE	222-228	TRV	287-289	GLAA	412-415				
RIL	70-72	CAGIGAWNYP	145-154	TGKKV	231-235	GDP	314-316	AHRVIH	425-430				
LLRE	78-81	QIACWK	156-161	VTMELGGKSPL	246-256	LIS	326-328	AGICW	434-438				

DTGKPWQEA	93-101	PALA	164-167	DAD	261-263	GG	352-353	MPVGGYK	449-455
FAG	116-118	FKPSEETP	174-181	NFYTQGE	275-281	EEIFGPVM	384-391	GRENG	460-464
QQ	127-128	GVFNV	199-203	CTN	283-285	RAN	405-407	TQ	472-473

[BLASTp organisms are: (for type 1 PiAFP) *Paenibacillus lactis* 154, *Clostridium phytofermentans* ISDg, *Coprobacillus* sp. 29\_1, *Erysipelothrix rhusiopathiae* ATCC 19414, *Desulfosporosinus meridiei* DSM 13257, *Aminobacterium colombiense* DSM 12261, *Leptospira weilii* str. 2006001853, *Fusobacterium* sp. 7\_1; (for PiCHDH and PiBADH) *Shewanella frigidimarina* NCIMB400, *Moritella* sp. PE36, *Colwellia psychrerythraea* 34H, *Oceanimonas* sp. GK1, *Glaciecola psychrophila* 170, *Marinobacter manganoxydans* Mn17-9, *Bermanella marisrubri*, *Reinekea blandensis* MED297, *Pseudoalteromonas atlantica* T6c, *Alteromonas* sp. SN2.]

### Physicochemical Characterization and Topology Prediction of Query Proteins:

The ExPasy's ProtParam tool provided useful data to determine the physicochemical properties of each target proteins presented in table 2 and 3. ProtParam program computed that type 1 PiAFP is may be an unstable protein while, PiCHDH and PiBADH are stable proteins, in vitro, with an instability index 40.72, 37.61, and 35.74 respectively. The instability index provides an estimate of the stability of a protein in a test tube<sup>24</sup>. According to this program, a protein whose instability index is <40, is predicted as stable; on the other hand, a value >40 predicts that the protein maybe unstable<sup>24</sup>. All query proteins are hydrophilic in nature (GRAVY index). Type 1 PiAFP, PiCHDH, and PiBADH have aliphatic index 84.44, 80.89, and 84.55 respectively. The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). It may be regarded as a positive factor for the increase of thermostability of globular proteins<sup>51</sup>. Isoelectric point (pI) is the pH at which the surface of protein is covered with charge but net charge of the protein is zero. At pI proteins are stable and compact<sup>52</sup>. Computation of pI values of type 1 PiAFP and PiBADH are 6.25 and 4.78 (pI<7) respectively, which indicates that these proteins are acidic. On the other hand, pI value for PiCHDH is 7.03 (pI>7) indicates that it is basic in

character. The extinction coefficient indicates how much light a protein absorbs at a certain wavelength. It is useful to have an estimation of this coefficient for following a protein which a spectrophotometer when purifying it<sup>53</sup>. Although ExPasy's ProtParam tool computes the extinction coefficient for a range of (276, 278, 279, 280 and 282 nm) wavelength, 280nm is favored because proteins absorb strongly there while other substances commonly in protein solutions do not<sup>61</sup>. Extinction coefficient of PiAFP type 1, PiCHDH, and PiBADH at 280nm is 37400, 71445, and 54235 respectively. The secondary structure prediction using SOPMA revealed that type 1 PiAFP and PiCHDH have more random coil while followed by  $\alpha$ -helices, extended strand, and  $\beta$ -turn, while PiBADH has mostly  $\alpha$ -helices, followed by random coil, extended strand, and  $\beta$ -turn (table 3). Presence of 8, 10, and 7 cysteine residues in type 1 PiAFP (at positions: 300, 303, 314, 317, 330, 333, 344, 347), PiCHDH (at positions: 18, 80, 121, 247, 254, 305, 343, 424, 473, 475), and PiBADH (at positions: 145, 159, 168, 283, 313, 375, 437) respectively, indicates the presence of disulfide (SS) bridges. CYS\_REC tool predicted 3 most probable SS bonds in both type 1 PiAFP, and PiCHDH at positions 300-344, 303-347, 314-317, and 80-473, 121-475, 247-254, respectively. In PiBADH, no probable SS bond has been predicted.

**Table: 2 Amino acid composition (%) of query proteins. Most abundant amino acid components are highlighted in bold.**

Protein	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
PiAFP	7.7	2.8	6.8	5.1	2.3	5.1	5.4	<b>8.3</b>	1.4	8.0	6.0	7.1	3.1	5.4	4.0	6.0	3.7	1.1	2.8	7.7
PiCHDH	8.4	4.4	6.0	5.5	1.8	5.0	5.3	<b>8.7</b>	3.4	5.3	8.4	6.2	2.3	3.6	4.6	6.2	3.6	1.4	3.2	6.6
PiBADH	9.4	3.7	3.7	5.3	1.4	4.3	7.2	<b>9.8</b>	1.4	6.6	6.8	4.3	2.7	4.3	4.9	4.7	7.4	1.2	2.9	8.0

**Table: 3 Primary and Secondary structure analysis of type 1 PiAFP, PiCHDH, and PiBADH sequences**

Categories	Type 1 PiAfp	PiCHDH	PiBADH
Molecular weight	38778.4	62142.4	52859.9
+ve Charged residues	(Arg + Lys): 35	(Arg + Lys): 60	(Arg + Lys): 39
-ve Charged residues	(Asp + Glu): 37	(Asp + Glu): 61	(Asp + Glu): 61
Extinction Coefficient [ $M^{-1} cm^{-1}$ ]	37400 [assuming all pairs of	71445 [assuming all pairs of	54235 [assuming all pairs of Cys

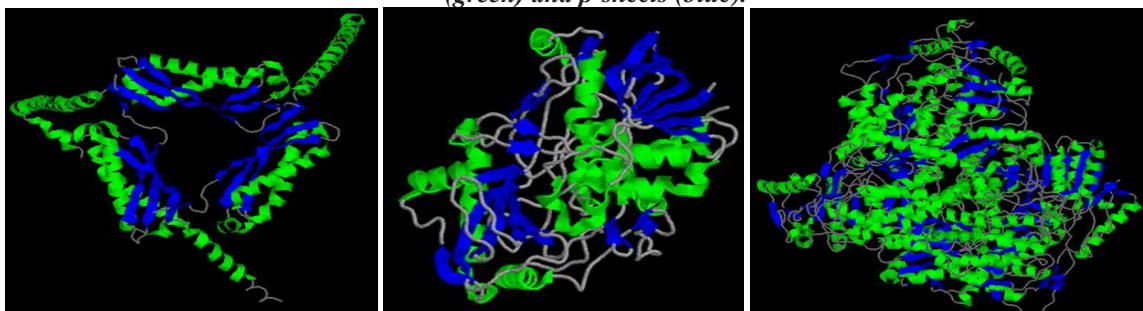
	Cys residues form cystines]	Cys residues form cystines]	residues form cystines]
<b>Half life</b>	30 h (mammalian reticulocytes, in vitro).>20 hours (yeast, in vivo). >10 h (E. coli, in vivo)	30 h (mammalian reticulocytes, in vitro).>20 hours (yeast, in vi >10 h (E. coli, in vivo)	30 h (mammalian reticulocytes, in vitro).>20 h (yeast, in vivo). >10 h (E. coli, in vivo)
<b>Instability Index</b>	40.72 [unstable]	37.61[stable]	35.74 [stable]
<b>Aliphatic Index</b>	84.44	80.89	84.55
<b>GRAVY value</b>	-0.141 [hydrophilic]	-0.382 [hydrophilic]	-0.088 [hydrophilic]
<b>Isoelectric point</b>	6.25	7.03	4.78
<b><math>\alpha</math>-Helix</b>	35.04%	27.76%	42.21%
<b>Extended Strand</b>	17.95%	18.86%,	17.42%
<b><math>\beta</math>-Turn</b>	7.12%	6.76%	7.58%
<b>Random Coil</b>	39.89%	46.62%	32.79%

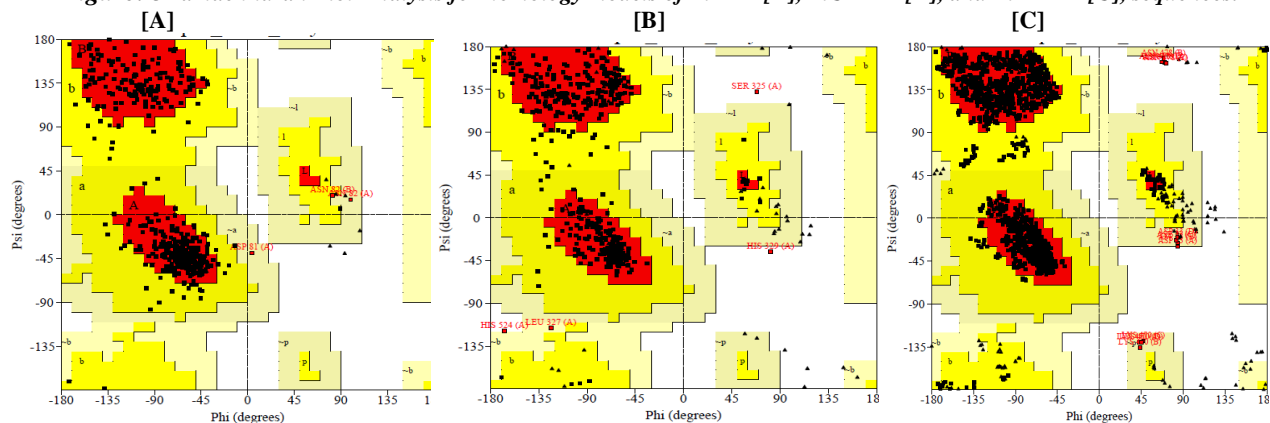
### Model Building and Evaluation:

Three dimensional structures of type 1 *PiAFP*, *PiCHDH*, and *PiBADH* have been modeled based on the experimentally solved structural homolog by using SWISS-MODEL server. Various templates were selected for each query protein and then examined for their quality and accuracy. Ramachandran plot analysis, Annolia value, Qmean value, Dfire energy, and Errat value each presented the best model for each query protein, on the basis of their respective parameters and scores. Tertiary structure analysis has also been used as step in model evaluation process. Final evaluation was done by analyzing the results of each evaluation tool, which yielded 3 PDB templates 3bk6 (crystal structure of a core domain of stomatin from hyperthermophilic archaeon *Pyrococcus horikoshii*), 3t37 (crystal structure of pyridoxine 4 oxidase from *Mesorhizobium loti*), and 1a4s (BADH from cod liver) as best structural models (also validated by Phyre2) for type 1 *PiAFP*, *PiCHDH*, and *PiBADH* respectively. Ramachandran plot analysis of all selected models was obtained by PROCHECK server. The PROCHECK suite of programs assess the "stereochemical quality" of a given protein structure. The aim of PROCHECK is to assess how normal, or conversely how unusual, the geometry of the

residues in a given protein structure is, as compared with stereochemical parameters derived from well-refined, high-resolution structures<sup>33</sup>. The best model for each query protein, in terms of stereochemical quality, showed overall G-factor value of -0.29, 0.08, and -0.12 for homology models of *PiAFP* type 1, *PiCHDH*, and *PiBADH* respectively, which indicates that geometry of the model corresponds to high probability confirmation with 87.5%, 90.0%, and 92.5% residues respectively in the core region of Ramachandran Plot. For type 1 *PiAFP*, no residue has been found in disallowed region whereas, *PiCHDH*, and *PiBADH* 0.7% and 0.2% residues are found in this region respectively. All the parameters of side chains and main chains are mostly in the better range. Other stereochemical parameters such as main chain lengths and bond angles are reported to be > 90% within the limits. Crystal structures for AFPs, CHDH, and BADH of variety of species are available in the PDB. And, there are currently no reported structural models for type 1 AFP, CHDH, and BADH from *Pi37*presentinPDB. Therefore, the developed 3D models of these proteins from *Pi37*reported here could serve as fairly good hypothetical models for these proteins and open up avenues to further investigation.

**Figure: 2 3D models for 3bk6 (Type 1 *PiAFP*) (left), 3t37 (*PiCHDH*) (mid), and 1a4s (*PiBADH*) (right).  $\alpha$ -helix (green) and  $\beta$ -sheets (blue).**



**Figure: 3Ramachdran Plot Analysis for homology models of PiAFP [A], PiCHDH [B], and PiBADH [C], sequences.****Table: 4 Validation parameters computed for the built 3D structures of targets PiAFP, PiCHDH, and PiBADH**

SelectedPDB IDs	Qmean Score	Dfire Energy	Errat Value (Quality Factor)
PiAFP 3bk6	0.621	-583.51	91.65
PiCHDH 3t37	0.791	-761.45	98.99
PiBADH 1a4s	0.921	-3220.67	95.183

**Conserved Domain Analysis and Sequence Motifs:**

Band\_7 domain was identified in type 1 PiAFP sequence. Similarly, domains named Glucose-methanol-choline-N- and C-terminal (GMC\_OxRdtase\_N, GMC\_OxRdtase\_C) within PiCHDH sequence, and Aldehyde-dehydrogenase (Aldedh) within PiBADH sequence were identified. Their positions are mentioned in table 5. Specific hit reveals that band\_7\_5, GMC\_OxRdtase\_C, and NAD+-dependent 4-trimethylaminobutyraldehyde dehydrogenase, family 9A1 (ALDH\_F9\_TMBADH) are conserved domains in type 1 PiAFP, PiCHDH, and PiBADH sequences respectively. Specific hit is the top ranking RPS-BLAST hit that meets and exceeds a domain-specific E-value threshold. It represents a very high confidence that the query sequence belongs to the same protein family as the sequence used to create the domain model, and therefore a high confidence level for the inferred function of the protein sequence<sup>54</sup>. Thus, conserved domain search reveals that type 1 PiAFP, PiCHDH, and PiBADH belong to Band\_7, GMC\_OxRdtase, and Aldedh domain containing protein families respectively, which was confirmed by SMART tool as well. Band\_7\_5 is a subgroup of Band\_7 domain of flotillin (reggie) like proteins. This subgroup contains proteins similar to stomatin, prohibitin, flotillin, HlfK/C and podicin. Many of these Band\_7 domain-containing proteins are lipidraft-associated<sup>55-57</sup>. Lipid-rafts are specialized membrane-microdomains, compartmentalize cellular processes by

servicing as organizing centers for the assembly of signaling molecules, influencing membrane fluidity and membrane protein trafficking<sup>58</sup>. Microdomains formed from flotillin proteins may, in addition be dynamic units with their own regulatory functions. Prohibitin may act as a chaperone for the stabilization of mitochondrial proteins<sup>56-58</sup>. Stomatin and flotillins are integral membrane proteins with properties similar to the caveolins<sup>59, 60</sup>, which is also a family of integral membrane proteins<sup>61-63</sup> and has been shown to play a role in the integrin signaling that mediate the attachment between a cell and its surroundings<sup>64</sup>. Here, it is the subject of further investigation that whether the association between stomatin, flotillins, and caveolins has any implication for type 1 PiAFP's possible role in cell adhesion? There is an evidence that stomatin is linked to the cortical actin cytoskeleton, suggesting its role in cortical morphogenesis of the cell<sup>65</sup>, such as protrusions<sup>66</sup>, which involves the extension of cellarmembranes<sup>67</sup> and associated with cellular adhesion. Adhesion not only provides stability, but also nucleates and regulates signaling components that control actin reorganization and membrane endocytosis and delivery<sup>68</sup>. A previous report suggests the novel enviroctactic role of wild type AFP from *Marinomonas primoryensis* in adhesion that is, transiently binding the host bacteria to ice, which also plays an essential role in the survival of this bacterium. Furthermore, conserved domain analysis, through specific hit, confirms that

*Pi*CHDH and *Pi*BADH belong to GMC\_OxRdtase\_C and ALDH\_F9\_TMBADH protein families, respectively. GMC\_OxRdtase family is a family of enzymes with oxidoreductase activity, which is catalysis of an oxidation-reduction (redox) reaction<sup>65</sup>. On the other hand, ALDH\_F9\_TMBADH oxidizes betaine aldehyde, is involved in carnitine biosynthesis<sup>69</sup>. In *Listeria monocytogenes*, L-carnitine is important for osmoregulation<sup>70</sup>. It has been documented that CHDH converts choline to betaine aldehyde, while BADH converts the aldehyde to glycine betaine, which is a compatible solute and a common osmoprotectant<sup>16</sup>. In addition to their function of maintaining an osmotic equilibrium across the cell membrane, compatible solute is effective stabilizers of proteins and even whole cells<sup>71</sup>. Moreover, GMC\_OxRdtase are FAD flavoproteins<sup>72,73</sup> whose one of the essential functions is DNA repair. Here, additional research is required to discern the role, if any, of these flavoproteins in repairing the damage in DNA caused by extremely cold and hyper-saline conditions contributing towards unique cold-adaptation of *Pi37*. *Pi*CHDH contains conserved C-terminal domain, which is steroid-binding domain<sup>74, 75</sup>. Steroid hormone receptors are a group of DNA-binding transcription regulatory proteins in which the activity is modulated by the binding of steroid hormone<sup>76</sup>; these receptors initiate signal transduction for steroid hormone, which lead to changes in gene expression<sup>77</sup>. Such receptors are closely related to chaperones (namely heat shock protein hsp 90 and hsp 56), which are required to maintain their inactive (but receptive) cytoplasmic conformation. One major function of chaperones is to prevent both newly synthesized polypeptide chains and assembled subunits from aggregating into nonfunctional structures. Many chaperones are heat shock proteins because the tendency to aggregate increases as proteins is denatured by stress<sup>78</sup>; they help the defective-shaped proteins to acquire proper configuration and perform their biological functions<sup>79</sup>. So, in further studies, it is essential to examine that are there transcriptional modifications and protein conformational changes responsible to facilitate *Pi37* survival in -12°C, which is extremely low? Motif Scan yielded some putative functional motifs that may predict the post-translational modifications of query proteins and their respective functions. These putative sites, such as Amidation, N-glycosylation, Casein kinase II phosphorylation, Protein kinase C (PKC) phosphorylation, Tyrosine kinase (TK) phosphorylation site (listed in table: 6), need additional

biological evidence to determine their respective functions. RGD cell attachment motif in *Pi*CHDH implies its function in cellular adhesion. Active sites have also been reported in *Pi*CHDH and *Pi*BADH sequences. These active sites are GMC\_OxRdtase signature 1 and 2 in *Pi*CHDH. The first one is located in the N-terminal section of this enzyme, while the second one is located in the central section. The consensus patterns have been mentioned in figure 5 and 6. On the other hand, aldehyde dehydrogenase (Aldedh) cysteine and glutamic acid active sites have been found in *Pi*BADH at different positions and seem to be implicated for the catalytic activity of *Pi*BADH. These active sites have been located in the respective 3D models of domains (figure: 4 and 5).

#### **Homology Modeling of Functional Domains:**

Protein Homology/analogy Recognition Engine (Phyre 2)<sup>32</sup> generated the homology models of domains. 150aa of Band\_7 domain of type 1 *Pi*AFP, 293aa of GMC\_OxRdtase\_N, and 137aa of GMC\_OxRdtase\_C domains from *Pi*CHDH, and 484aa of Aldedh domain from *Pi*BADH were submitted to Phyre2 server. Phyre2 presented matches that are ranked by a raw alignment score that is based on the number of aligned residues and the quality of alignment. This in turn is based on the similarity of residue probability distributions for each position, secondary structure similarity and the presence or absence of insertions and deletions<sup>32</sup>. Five matches to band\_7 domain from type 1 *Pi*AFP with >95% confidence, and 20 matches with modeled structures of each GMC\_OxRdtase\_N and GMC\_OxRdtase\_C domains from *Pi*CHDH, and Aldedh domain from *Pi*BADH, with 100% confidence, have been yielded. Phyre2 presented top ranking model (PDB Ids) for each domain sequence submitted and these are: 3bk6 (crystal structure of a core domain of stomatin from *Pyrococcus horikoshii*) (99.8% confidence) for band\_7 domain from type 1 *Pi*AFP, 3q9t (crystal structure analysis of formate oxidase) and 3t37 (crystal structure of pyridoxine 4-oxidase from *Mesorhizobium loti*) (100% confidence each) for GMC\_OxRdtase\_C and GMC\_OxRdtase\_N domains respectively from *Pi*CHDH, and 1a4s (ALDH-like) (100% confidence) for Aldedh domain of *Pi*BADH.

Moreover, it validates the accuracy of 3D models for each query protein developed through homology modeling. GMC\_oxred\_N is ADP-binding domain with  $\beta\alpha\beta$ -fold<sup>72</sup>, which is evident from the 3D model of the domain. Although, the template structures for final model show only 15% identity to band\_7 domain of



type 1 PiAFP, 83% of the protein's residues were modeled at greater than 99% confidence. Here, it is important to note that even at very low sequence identities (<15%) models can be very useful as long as the confidence is high<sup>32</sup>. Phyre2 provided detailed template information which is very useful for glean information about the potential function of query proteins<sup>32</sup>. Functions of core domain of stomatin from *Pyrococcus horikoshii* have not been completely understood yet, however, literature documents that they mostly localize to membrane domains, and in many cases, they have been shown to modulate ion channel activity<sup>80</sup>. On the other hand, both formate oxidase and pyridoxine 4-oxidase belong to the family of

oxidoreductases. Pyridoxine 4-oxidase from nitrogen fixing symbiotic bacterium *Mesorhizobium loti* is significantly involved in degradation of Vitamin-B6<sup>81</sup>, and also employs other metabolic processes. Aldehyd domain of PiBADH is homologous to ALDH-like super and subfamily which is involved in oxidoreductase activity and metabolic processes that typically transform small molecules, but also include macromolecular processes such as DNA repair and replication, and protein synthesis and degradation<sup>82</sup>. Future studies should be conducted in light of above details to dig deep into the functional aspects of query proteins.

**Table: 5 Functional domains at their respective positions in query protein sequences.**

Protein	Protein Domain by SMART	Position	Conserved Domain
Ty 1 PiAFP	Band_7	31-252	Band_7_5
PiCHDH	GMC_OxRdtase_N	6-298	GMC_OxRdtase_C
	GMC_OxRdtase_C	388-524	
PiBADH	Aldehyd	15-477	ALDH_F9_TMBADH

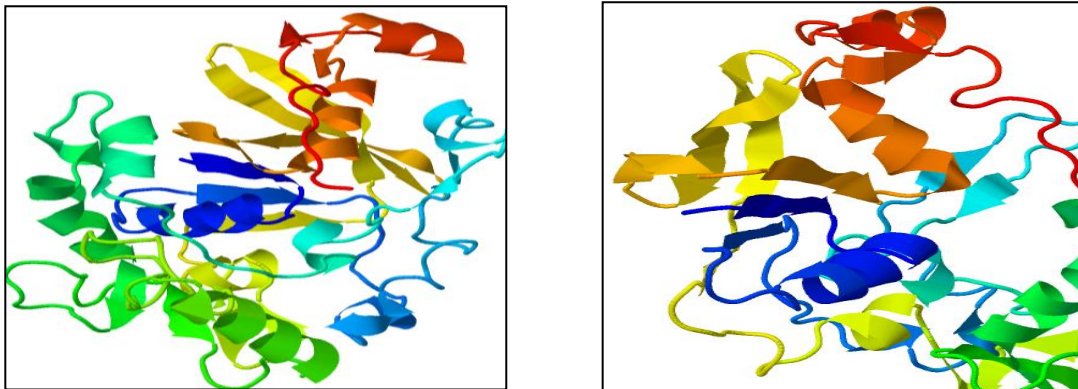
**Table: 6 Functional sites in type 1PiAFP, PiCHDH, and PiBADH sequences at their respective positions**

Sequence Motif	Position			Status
	Type 1 Pi AFP	Pi CHDH	Pi BADH	
<i>Amidation site</i>	347-350	227-230, 529-532,	231-234,352-355	?(weak match)
<i>N-glycosylation site.</i>	177-180, 283-286	212-215, 330-333	18-21, 285-288,407-410, 463-466	?(weak match)
<i>Casein kinase II phosphorylation site.</i>	260-265, 266-271, 334-339,	2-5, 67-70, 413-416, 453-456, 463-466	64-67, 102-105,108-111, 137-140, 215-218, 278-281, 382-385, 395-398, 445-448	?(weak match)
<i>N-myristoylation site</i>	123-128, 260-265, 266-271, 334-339	12-17, 39-44, 87-92, 165-170, 194-199,268-273, 288-293, 333-338	118-123, 143-148, 195-200, 211-216, 241-246, 274-279,412-417, 443-448	?(weak match)
<i>PKC phosphorylation site.</i>	57-59, 165-167,243-245, 335-337	40-42, 58-60, 190-192,474-476, 550-552	52-54, 94-96, 137-139,180-182, 231-233, 242-244, 467-469	?(weak match)
<i>RGD cell attachment motif.</i>	-	139-141	-	?(weak match)
<i>TK phosphorylation site.</i>	230-237	462-469	-	?(weak match)
<i>GMC_ OxRdtase. S.1</i>	-	83-106	-	!(active site)
<i>GMC_ OxRdtase S. 2</i>	-	256-270	-	!(active site)
<i>Aldehyd cysteine</i>	-	-	276-287	!(active site)
<i>Aldehyd glutamic acid.</i>	-	-	248-255	!(active site)

**Figure: 4** Developed homology model of Band\_7 domain (3bk6) (left) of type1PiAFP in rainbow color sequence (from N-C terminal). Aldedh domain (1a4s) in PiBADH sequence, indicates Cystein (Cys283, green color random coil) and Glutamic Acid (Glu249, green color  $\beta$ -sheet) active sites (middle), and GMC\_oxred\_C domain (3q9t) (right) in PiCHDH sequence in rainbow color sequence (from N-C terminal).



**Figure: 5** Developed homology model of GMC\_oxred\_N domain (3t37) (left) Figure in left indicates Signature 1 Active site from GLY 83 to ASP 106 (left) (green a-helix, random coil to sky-blue random coil). The consensus pattern of Signature 1 active site is: [GA]-[RKNC]-x-[LIVW]-G (2)-[GST](2)-x-[LIVM]-[NH]-x(3)-[FYWA]-x(2)-[PAG]-x(5)-[DNESHQA]<sup>72</sup>. A section of developed model of GMC\_oxred\_N domain indicates Signature 2 Active site from GLY 256 to GLY 270 (right). The consensus pattern of Signature 2 active site-[GS]-[PSTA]-x(2)-[ST]-[PS]-x-[LIVM](2)-x(2)-S-G-[LIVM]-G<sup>72</sup> (orange a-helix)



#### **Protein-Protein Interaction Prediction:**

The function of a protein may be extrapolated by knowing the function of binding partners<sup>83</sup>. Here, protein-protein interaction prediction revealed that Spermidine synthase is the highest scorer functional partner of type 1PiAFP. For PiCHDH, BADH and betI are the highest scorer functional partners. They both have shown stronger interaction with each other as well. Moreover, phospholipase A (1), phosphatidylserine decarboxylase (psd), and a substrate binding region of ABC-type glycine betaine transport system, are also found to be functionally significant to PiCHDH and to one another. For PiBADH, CHDH and betI have shown strongest interaction (table 7). The interaction between proteins is fundamental to a broad spectrum of biological functions<sup>84</sup>. So, to fully

understand the proper function of a protein, it is essential to study the functions of interacting partners of that protein.

Spermidine synthase is found to be the highest scorer functional partner of type 1 PiAFP. Previous studies have shown that overexpression of spermidine synthase enhances tolerance to multiple environmental stresses including chilling, freezing, salinity, hyperosmosis in plants<sup>85</sup>. Moreover, spermidine synthase catalyzes the reaction by which spermidine is synthesized, which is a polyamine, and another study shows that Polyamines can be implicated in cold adaptation in an Antarctic bacterium *Pseudoalteromonas haloplanktis TAC125*, which has been evidenced by detection of spermidine synthase<sup>86</sup>. This strongly suggests that spermidine synthase may

aid type 1 *PiAFP* to respond against extremely cold-conditions. The closest association of *PiCHDH* with *BADH* and of *PiBADH* with *CHDH* is apparent, as they both interact with each other in same chemical pathway at consequent steps which lead to the synthesis of a common osmoprotectant – glycine betaine. Also, they both interact with betaine inhibitor (*betI*). Phospholipase A1, which is another closely associated functional partner of *PiCHDH*, is an integral membranephospholipase, and has been most thoroughly studied in *Escherichia coli* where it participates in the secretion of bacteriocins<sup>87</sup>.

**Table: 7 Protein-protein interactions of three query proteins and scores represent the degree of interaction.**

Query Protein	STRING protein-protein interaction	Score
<i>Type 1 PiAFP</i>	Ping_0906 spermidine synthase	0.692
	Ping_0185 phage shock protein A	0.579
	Ping_1308 hypothetical protein	0.573
	Ping_3298 hypothetical protein	0.559
	Ping_3374 hypothetical protein	0.463
	Ping_0971 aminoimidazole riboside kinase	0.433
<i>PiCHDH</i>	Ping_2297 Na <sup>+</sup> -dependent transporter of the SNF family protein	0.996
	Ping_2071 <i>BADH</i>	0.992
	<i>betI</i>	0.800
	Ping_3493 Phospholipase A 1	0.800
	<i>Psd</i>	0.784
	Ping_2073 substrate binding region of ABC type glycine betaine transporter system	0.660
	Ping_0969-choline/carnitine/betainetransporter	0.616
	Ping_2647 hypothetical protein	0.574
	<i>tepB</i>	0.567
	Ping_1055 tryptophan synthase $\alpha$ -subunit	0.455
Ping_1131 hypothetical protein	0.996	
<i>PiBADH</i>	Ping_2072 <i>CHDH</i>	0.978
	<i>betI</i>	0.664
	Ping_2076 putative transporter, BCCT family protein	0.613
	Ping_2073 substrate binding region of ABC type glycine betaine transporter system	

Bacteriocin release is triggered by a lysis protein, followed by a phospholipase dependent accumulation of lysophospholipids and free fatty acids in the outer membrane<sup>87</sup>. A study by Svetoslav D et al<sup>88</sup> has shown that the production of bacteriocin from strain *ST712BZ* may be influenced by osmotic stress<sup>88</sup>. Another finding indicates a correlation between acquired osmotic stress resistances in lactic acid bacteria through bacteriocin production<sup>89</sup>. Here, further study is required to examine their interactive roles in order to facilitate *Pi37* its unique adaptive strategies.

Additionally, at low temperature, cells acquire fatty acids, required for the maintenance of membrane fluidity<sup>90</sup>, which also suggests an essential interactive role of phospholipase A1 in cold-adaptation. Phosphatidylserine decarboxylase and a substrate binding region of ABC-type glycine betaine transport system, which is a part of a high affinity multi-component binding-protein-dependent transport system involved in bacterial osmoregulation<sup>91</sup>, are also found to be functionally significant to *PiCHDH* and to one another.

### Conclusion:

The study of unique survival mechanisms of extremophilic microorganisms are of great interest to microbiological community. *Pi37* is an extreme psychrophile and not much has been understood about its adaptive strategies yet. The present work endeavors to provide evolutionary and functional insights through in-silico study of three proteins: type 1 *PiAFP*, *PiCHDH*, and *PiBADH* from *Pi37*, which may have bearing on the psychrophily and halophily of this bacterium. Phylogenetic analysis infers that evolutionary proximity between type 1 *PiAFP* and type 1 *AFP* of *Paenibacillus lactis 154*, with very low sequence identity, may be an outcome of functional convergence; on the other hand, high degree of sequence conservation and close phylogenetic relationship between *PiCHDH* and *CHDH* of *Moritella sp. PE36*, and between *PiBADH* and *BADH* of *Shewanella frigidimarina*, *BADH* of *Bermanella marisrubri*, and *BADH* of *Colwellia psychrerythraea 34H* suggest divergent evolution of these proteins from a common ancestor. MSA identifies some conserved patterns/residues in each query protein that are evolutionary preserved across BLASTp species and imply their functional significance. Physicochemical characterization studies provided essential data about the properties of target proteins, such as pI, EC, AI, GRAVY and Instability Index; it revealed that all query proteins are glycine-rich and hydrophilic in nature. Type 1 *PiAFP*, and *PiCHDH* have more coiling, whereas, *PiBADH* is  $\alpha$ -helical; as well as type 1 *PiAFP* and *PiCHDH* have been found to have 3 most probable SS bonds, while no SS bond is predicted in *PiBADH*. Present study found three fairly good representative homology models of type 1 *PiAFP* (3bk6), *PiCHDH* (3t37) and *PiBADH* (1a4s), which was again validated by homology modeling of domains through Phyre2. Although energy-minimization has not been performed here, the model information is significant and can further be utilized to conduct in-depth evolutionary, structural, and functional analysis of these proteins from *Pi37*.

Conserved domain analysis is integral to unearth functional aspects of proteins by knowing their family history, thus, significant inferences about possible roles of query proteins in environmental adaptations of *Pi37* have been drawn here, which need to be further analyzed in future studies. Some putative sites of posttranslational-modifications have been configured in all query proteins that may provide functional insights; and catalytic sites, such as GMC\_OxRdtase signature 1 and 2, and Aldedh cystein and glutamic acid in *PiCHDH* and *PiBADH*, respectively, have been confirmed and located in domain models. Further, the paper uses a standard resource (STRING) to predict different proteins that bind to the proteins of interest and utilizes the information to assume the possible functions of target proteins.

#### Acknowledgement:

Author is thankful to BioAxis DNA Research Center, Lucknow for the opportunity of research project and providing valuable support. The kind support from Mrs Amita Kashyap and Mr. Ashish Kumar Rai is gratefully acknowledged.

#### References:

1. Monica Riley, James T Staley, Antoine Danchin, Ting Zhang Wang, Thomas S Brettin, Loren J Hauser, Miriam L Land and Linda S Thompson, 2008, BMC Genomics, 9:2
2. Marx J-C, Collins T, D'Amico S, Feller G, Gerday C, 2006, Marine Biotech, 9:293-304.
3. Goodchild A, Raftery M, Saunders NF, Guilhaus M, Cavicchioli R, 2004a. J Proteome Res 3: 1164–1176
4. Monica Riley, James T Staley, Antoine Danchin, Ting Zhang Wang, Thomas S Brettin, Loren J Hauser, Miriam L Land and Linda S Thompson, 2008, BMC Genomics, 9:2
5. Sebastiana Angelaccio, Rita Florio, Valerio Consalvi, Guido Festa, Stefano Pascarella Serine, 2012, Int J Mol Sci; 13(2): 1314–1326.
6. Breezee, J.; Cady, N.; Staley, J. T. 2004, Microbial Ecology; Vol. 47 Issue 3, p300.
7. M K Chattopadhyay, RESONANCE, December 2007, pp. 26-30
8. Devries AL, Wohlschl De, 1969. Science 163: 1073.
9. Duman JG, Devries AL, 1974. Nature 247: 237–238.
10. Raymond JA, Devries AL, 1977, PNAS 74: 2589–2593.
11. Sidebottom C, Buckley S, Pudney P, Twigg S, Jarman C, et al. (2000) Nature 406: 256–256.
12. Worrall D, Elias L, Ashford D, Smallwood M, Sidebottom C, et al. (1998) Science, 282: 115–117.
13. Knight CA, Devries AL, Oolman LD, 1984, Nature 308: 295–296.
14. Karen Junge, Brent Christner, James T. Staley, 2011 Extremophiles Handbook, Springer.
15. Le Rudulier, D., Strom, A. R., Dandekar, A. M., Smith, L. T. (1984). Science 224:1064-1068.
16. Smith LT, Pocard JA, Bernard T, Le Rudulier D (1988). J Bacteriol 170(7); 3142-9.
17. SF Altschul, et al. 1990, J Mol Biol. 215:403.
18. <http://www.clcbio.com/index.php?id=28>, (27/10/2006)
19. Schultz et al., 1998, Proc. Natl. Acad. Sci. USA 95, 5857-5864
20. Letunic et al. (2012) Nucleic Acids Res.
21. [http://myhits.isb-sib.ch/cgi-bin/motif\\_scan](http://myhits.isb-sib.ch/cgi-bin/motif_scan)
22. <http://string-db.org/>
23. SC Gill, PH Von Hippel. Anal Biochem. 1989; 182:319.
24. K Guruprasad, et al. Protein Eng. 1990; 4:155.
25. Aikai. J Biochem. 1980; 88:1895.
26. J Kyte, RF Doolittle. J Mol Biol. 1982; 157:105.
27. C Geourjon, G Deléage. . 1995; Comput Appl Biosci 11:681
28. CYS\_REC.[http://sun1.softberry.com/berry.phtml?topic=cys\\_re c&group=help&subgroup=propt](http://sun1.softberry.com/berry.phtml?topic=cys_re c&group=help&subgroup=propt). (27/10/2006)
29. Nielsen M., Lundegaard C., Lund O., Petersen TN, 2010, Nucleic Acids Research, Vol. 38.
30. CPHmodels 2.0: X3M a Computer Program to Extract 3D Models. O. Lund, M. Nielsen, C. Lundegaard, P. Worning, Abstract at the CASP5 conference A102, 2002.
31. <http://toolkit.tuebingen.mpg.de/hhpred>
32. Kelley LA and Sternberg MJE. 2009, Nature Protocols 4, 363 - 371
33. RA Laskowski, et al. 1996, J Biomol NMR. ; 8:477.
34. GN Ramachandran, et al. 1963, J Mol Biol.; 7:95.
35. Melo, F. and E. Feytmans, 1998. J Mol Biol 277(5): 1141-1152.
36. Benkert, P., Tosatto, S.C.E. and Schomburg, D. 2008. Nucleic Acids Res 71(1):261-277
37. Benkert P, Künzli M, Schwede T. 2009. Nucleic Acids Res. 2009 Jul 1; 37 (Web Server issue):W510-4.
38. Benkert, P., Biasini, M. and Schwede, T. 2011. Bioinformatics
39. Colovos, C. and Yeates, T. O. (1993), Protein Science, 2: 1511–1519.
40. Zhou, H., and Zhou, Y. (2002). Protein Sci. 11:2714-2726.
41. Soding J (2005) Bioinformatics 21: 951–960.
42. Jefferys BR, Kelley LA, Sternberg MJE (2010) J Mol Biol 397: 1329–1338.
43. Koonin EV, Galperin MY. 2003, Sequence - Evolution - Function: Computational Approaches in Comparative Genomics. Boston: Kluwer Academic; Chapter 2, Evolutionary Concept in Genetics and Genomics.
44. Patsy Scheldeman, Karen Goossens, Marina Rodriguez-Diaz, Annelies Pil, Johan Goris, Lieve Herman, Paul De Vos, Niall A. Logan, Marc Heyndrickx., 2004, Int. J. of System. And Evol. Microbio. , vol. 54 no. 3 885-891.
45. Chris W. Michiels, High-Pressure Microbiology, ASM Press, 2008, Ch. 14, p. 239.
46. Jarone Pinhassi, María J. Pujalte, Javier Pascual, José M. González Itziar Lekunberri, Carlos Pedrós-Alió, David R. Arahal, 2009, IJSEM, vol. 59 no. 2 373-377
47. The UniProt Consortium, Reorganizing the protein space at the Universal Protein Resource (UniProt) Nucleic Acids Res. 40: D71D75 (2012).
48. Russell RB. 1998; J. Mol. Biol. 279:1211–1227.
49. Russell RB, Sasieni PD, Sternberg MJ. J. Mol. Biol. 1998; 282:903–918.

50. Steven M. Thompson, (2004) , Multiple Sequence Alignment and Analysis: Part I An Introduction to the Theory and Application of Multiple Sequence Analysis.
51. Ikai, A.J. 1980. J. Biochem. 88, 1895-1898.
52. K Sivakumar, S Bala Ji, Gangaradhakrishnan,(2007), J. Chem. Sci., Vol. 119, No. 5, pp. 571–579.
53. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; (2005), (In) John M. Walker (ed): The Proteomics Protocols Handbook, . pp. 571-607
54. Marchler-Bauer A, Bryant SH. CD-Search: protein domain annotations on the fly. 2004, Nucleic Acids Res. 1;32
55. Marchler-Bauer A et al. (2011), Nucleic Acids Res. 39 (D) 225-9.
56. Marchler-Bauer A et al. (2009), Nucleic Acids Res.37 (D) 205-10.
57. Marchler-Bauer A, Bryant SH (2004), Nucleic Acids Res.32 (W) 327-331.
58. Korade, Zeljka; Kenworthy, Anne K. (2008). "Lipid rafts, cholesterol, and the brain". Neuropharmacology 55 (8): 1265–73.
59. Salzer U, Mairhofer M, Prohaska R. Dynamic Cell Biology 2007; 1: 20-33.
60. Salzer U, Hunger U, Prohaska R. Advances in Planar Lipid Bilayers and Liposomes 2008; Vol. 6, 49-80.
61. Tang Z, Scherer PE, Okamoto T, Song K, Chu C, Kohtz DS, Nishimoto I, Lodish HF, Lisanti MP (January 1996). J. Biol. Chem. 271 (4): 2255–61.
62. Scherer PE, Okamoto T, Chun M, Nishimoto I, Lodish HF, Lisanti MP (1996). Proc. Natl. Acad. Sci. U.S.A. 93 (1): 131–5.
63. Williams TM, Lisanti MP (2004). Genome Biol. 5 (3): 214.
64. <http://en.wikipedia.org/wiki/Integrins>
65. [www.ebi.ac.uk/QuickGO/GTerm?id=GO:0016491](http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0016491)
66. Snyers, L., Umlauf, E., Prohaska, R. . (1998), J. Biol. Chem
67. Petrie, R. J., Doyle, A. D. & Yamada, K. M. Nat Rev Mol Cell Biol 10 538-49 (2009).
68. [http://www.cellmigration.org/resource/proteomics/data/bpix\\_t otal.shtml](http://www.cellmigration.org/resource/proteomics/data/bpix_t otal.shtml) (A resource supported by NIH/NIGMS funding for the Cell Migration Consortium, U54GM064346).
69. [www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=143409](http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=143409)
70. <http://www.ebi.ac.uk/interpro/entry/IPR007210>
71. Hans Jorg Kunte, Extremophiles; ©ELOSS, Vol 2.
72. Cavener DR, (1992). J. Mol. Biol.223 (3): 811–814.
73. Li J, Vrieling A, Brick P, Blow DM, 1993. Biochemistry 32(43): 11507–15.
74. Bannwarth M, Bastian S, Heckmann-Pohl D, Giffhorn F, Schulz GE (2004). Biochemistry 43 (37): 11683–90.
75. Vrieling A, Lloyd LF, Blow DM (1991). J Mol Biol 219 (3): 533–54.
76. Jan Carlstedt-Duke, Per-Erik Stromstedt, Bengt PerssonS, Ella Cederlund S, Jan-AkeGustafsson, and Hans Jornval, The journal of Biological Chemistry, Vol. 263, No. 14, Issue of May 15, pp. 6842-6846, 19B8
77. Dahlman-Wright K, Cavailles V, Fuqua SA, Jordan VC, Katzenellenbogen JA, Korach KS, Maggi A, Muramatsu M, Parker MG, Gustafsson JA, 2006, "International Union of Pharmacology. LXIV. Estrogen receptors". Pharmacol. Rev. 58 (4): 773–81.
78. Kris Pauwels and other (2007). Molecular Microbiology 64 (4): 917.
79. AmbilyNathI.VandLokaBharathi P.A, Extremophiles, vol.15 (2); 2011; 129-153.
80. Liudmilla Lapatsinaa, Janko Brandb, Kate Poolea, Oliver Daumkeb, Gary R. Lewin, (2012), European Journal of Cell Biology, 91 (4), pp. 240-245.
81. Yuan, B., Yoshikane, Y., Yokochi, N., Ohnishi, K. and Yagi, T. (2004), FEMS Microbiology Letters, 234: 225–230.
82. <http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0008152>
83. Golemis E. (2002) Protein-protein interactions : A molecular cloning manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. ix, 682 p. p
84. Alberts,B., Bray,D., Lewis,J., Raff,M., Roberts,K. and Watson,J.D. (1989) Molecular Biology of the Cell. 2nd edn, Garland, New York.
85. Kasukabe Y, He L, Nada K, Misawa S, Ihara I, Tachibana S. 2004, Plant Cell Physiol.; 45(6):712-22.
86. Piette, F., D'Amico, S., Struvay, C., Mazzucchelli, G., Renaut, J., Tutino, M. L., Danchin, A., Leprince, P. and Feller, G. (2010). Molecular Microbiology, 76: 120–132.
87. Snijder HJ, Timmins PA, Kalk KH, Dijkstra BW. 2003, Journal of Structural Biology , 141(2):122-131
88. TODOROV, Svetoslav D., DICKS, Leon M.T. 2007. Braz. J. Microbiol., vol.38, n.1
89. Effie Tsakalidou, Konstantinos Papadimitriou, 2011 Springer, pp. 287-288.
90. M K Chattopadhyay, 2006, Journal of Biosciences. 31(1), 157–165.
91. <http://www.ebi.ac.uk/interpro/entry/IPR007210>

\*\*\*\*\*