In silico analysis and structure modelling of heat shock protein HSP70 from *Glaciozyma antarctica* PI12 as a model system to understand adaptation strategies of Antarctic organisms amid adverse climates

Wan Nur Shuhaida Wan Mahadi[1], Clemente Michael Wong Vui Ling[1], Hyun Park[2], Nur Athirah Yusof[1#]

1 Biotechnology Research Institute, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, MALAYSIA.
2 Division of Biotechnology, Korea University, Seoul, 02841, SOUTH KOREA.
#Corresponding author. E-Mail: nrathirah.yusof@ums.edu.my; Tel: +6088-320000 ext 8474; Fax: +6088-320993.

ABSTRACT The 70-kDa heat-shock proteins (HSP70) are integral components of the cell's folding catalysts. *Glaciozyma antarctica* PI12 is an obligate psychrophilic yeast that possesses six HSP70 genes in its genome. The functions of these HSP70s in *G. antarctica* in terms of similarities and differences are yet to be discovered. The purpose of this study is to determine the structure and function of HSP70 from *G. antarctica*, which will lead to understanding this organism's adaptation strategies through structural and functional annotation. In this study, we utilize the HSP70 genes derived from genome data of *Glaciozyma antarctica* PI12 isolated from the Casey Research Station to characterize and compare structural characteristics which may contribute to their adaptation strategies during global warming. Computational tools such as Expasy's ProtParam, MEGA 11, SWISS-MODEL, AlphaFold2, and SAVES were used to analyze all the genes via physicochemical analysis, phylogenetic study, homology modelling and structure validation, and superimposition of models. Results showed that reliable 3D models of HSP70 were successfully generated via the homology modelling approach using SWISS-MODEL and AlphaFold2 programs. The proposed model was evaluated as reliable with high confidence based on the structural stereochemical property, verification of protein patterns of non-bonded atomic interactions, compatibility of a 3D model with its amino acid sequence and determination of the protein's native fold. Among the new findings are the molecular signatures such as ionic, aromatic-aromatic, aromatic-sulphur and cation-pi interactions that are lesser in the buried residues when compared to their homologs. These interactions are important for maintaining structure stability, flexibility and packing in proteins. This may reflect the yeast response and adaptation strategies during the adverse climate. By studying the structural adaptations of HSP70 proteins in psychrophilic yeast, researchers can gain insights into how these proteins maintain their functionality in changing temperature conditions. This knowledge can inform the development of strategies to mitigate the impact of global warming on cold-adapted organisms and potentially guide the design of novel enzymes with improved thermal stability for biotechnological applications. In conclusion, this comprehensive study provides an in-depth understanding of the structural adaptation and evolution of HSP70 about their thermal resistance to global warming.

KEYWORDS: HSP70; Stress response; Antarctic Yeast; Structural characteristics; Homology modelling.
Received 15 May 2023 **Revised** 4 July 2023 **Accepted** 6 July 2023 **In press** 8 July 2023 **Online** 29 September 2023

Original Article

INTRODUCTION

According to the British Antarctic Survey (BAS), studies have shown that the west coast of the Antarctic Peninsula has been one of the most rapidly warming parts of the planet. However, in early 2022, heat waves were observed to cause a significant global warming effect, with temperatures over the eastern Antarctic ice sheet soaring 50 to 90 degrees above normal (Chown *et al.*, 2022). This has been a shocking phenomenon amongst scientists all over the world who are studying global warming and its impact on Antarctica. Global warming is causing physical changes in Antarctica and altering the natural distribution of the living environment (McBride *et al.*, 2021). All reports on the effects of global warming in Antarctica pointed out the importance of studying climate change to

enable scientists to predict more accurately the effects of climate change on the ecosystem, particularly regarding the resistance and resilience of biological organization in all major Antarctic environments; terrestrial, freshwater, and marine (Convey & Peck, 2019; Gutt *et al.*, 2021; McGaughran *et al.*, 2021; Yadav *et al.*, 2023). The different responses of all living organisms will determine their evolution or extinction to global warming. Parallel to the scientific goal of SCAR Research on Antarctic Threshold-Ecosystems Resilience and Adaptation (AnT-ERA), it is vital to understand the impact of environmental change on biodiversity and the consequences for adaptation and function. Therefore, it is extremely important to find biomarkers that can measure the effects of global warming as well as the ecosystem tolerance limits, which include resistance and resilience to environmental change.

Glaciozyma antarctica PI12 is an obligate psychrophilic yeast isolated from the Antarctic marine water near Casey Station (S66° 21′ 25″; E110° 37′ 09″) in the Southern Ocean, Antarctica. It grows optimally at 12°C and can survive below sub-zero temperatures (<0 °C) and the maximum growth temperature of 20°C (Boo *et al.*, 2013). The genome data has revealed a total of 7857 protein-coding genes (Firdaus-Raih *et al.*, 2018). Previously, Boo et al. (2013) analysed the expression levels of genes in *G. antarctica* PI12 that were exposed to cold shocks and heat shocks using quantitative real-time PCR (qPCR). Several related studies have provided some insights into the adaptation strategies of *G. antarctica* PI12 to thermal stress. Among the genes which are expressed differentially in *G. antarctica* PI12 when exposed to higher temperatures are 70 kDa heat-shock protein (HSP70) genes (Yusof *et al.*, 2021).

HSP70 are molecular chaperones abundantly present in all living organisms which play essential functions in protein folding, assembly, transport across biological membranes and degradation (Horwich & Fenton, 2020). More importantly, HSP70 is also involved in stress response due to their ability to be constitutively expressed or stress-inducible (Bakhos-Douaihy *et al.*, 2021; P. Mitra *et al.*, 2021; Rathor *et al.*, 2023). HSP70 functioned in refolding of protein during stress conditions and regulatory proteins management in signal transduction pathways (Mayer & Bukau, 2005). In such stressful conditions, HSP70s are activated and begin protecting the cell by minimizing denatured or abnormal proteins in the cell (Kabani & Martineau, 2008). The mechanism of work of HSP70 is by binding to proteins which are affected by heat stress and preventing their aggregation within cells. Upon stimulation and induction of cellular stress response, several post-translational and transcriptional mechanisms take place in which heat shock factors are present in inactive forms within cells. The heat shock factors will then become active via post-translational modification induced by heat stress, hence, increasing the transcription of HSP70. Heat shock messages should be efficiently translated to protect the cell against stress-induced denaturation. Therefore, the messenger RNA existing previously should be reserved for translation so that competition for the translation of new mRNA can be avoided. During heat shock, those mRNAs of HSP70, which are unstable at normal temperatures, become stable. Due to their importance, HSP70 are highly conserved in organisms, which makes it a highly potential biomarker to measure thermal stress (Daugaard *et al.*, 2007).

In a previous study of *G. antarctica*, the protein structures of two HSP70, gahsp70-1 and gahsp70-2, were shown to have high similarities with that of human and *E. coli* HSP70, respectively (Yusof *et al.*, 2021). The analysis of the structures showed the replacement of the β-sheet with a loop and several residue substitutions with alanine as well as less polar and charged residues at the ATP-binding domain. According to cold adaptation theories, loops and protein surface hydrophobicity contribute to higher flexibility which is crucial in adaptation to cold environments and enables ATP hydrolysis at low temperatures. However, the trend of thermal adaptation strategies in HSP70 is yet

to be determined by investigating the structures of all HSP70 present in *G. antarctica*. Studying the rest of the HSP70 from *G. antarctica* (gahsp70) is hence crucial to see the overall strategies implemented by the organism to survive environmental stresses.

Besides its importance in *G. antarctica*, the importance of HSP70, ranging from bacteria to eukarya, has been reported in several studies (Karunanayake & Page, 2021; Moss *et al.*, 2019; Rosenzweig *et al.*, 2019; Yusof *et al.*, 2022). One study showed that HSP70 was abundantly expressed in *Shewanella frigidimarina*, which is a psychrophilic bacterium, when exposed to heat stress at 28°C (García-Descalzo *et al.*, 2014). Another study on the Antarctic haloarchaea showed that *Halohasta litchfieldiae* and *Halorubrum lacusprofundi* had higher expression levels of HSP70 at high temperatures (Williams *et al.*, 2017). In addition, studies in *D. melanogaster*, *Caenorhabditis elegans*, rodents, insects, and humans have also indicated that the higher the expression level of HSP70, the better the thermal stress tolerance of cells (Cho & Park, 2019; Gong & Golic, 2006; Jin *et al.*, 2020; Kuennen *et al.*, 2011; Odunuga *et al.*, 2012; Silver & Noble, 2012). The gene expression pattern of HSP70 in different goat organs also exhibited a higher expression than that of the control during the peak heat stress period (Rout *et al.*, 2016).

In summary, previous studies have shown that HSP70 plays a crucial role in protecting cells from thermal stress and other forms of environmental stress. Its expression is induced in response to stress and is associated with enhanced survival and resistance to stress-induced cell death. The findings from these studies highlight the significance of HSP70 in cellular protection and adaptation to thermal stress. These findings also confirmed the importance of HSP70 in protecting cells against thermal and cold stress. Hence, HSP70 serves as a potential biomarker for determining the expression of mRNA inside *G. antarctica* cells upon exposure. Comparative analysis of the structures of all HSP70 present in the *G. antarctica* PI12 genome will shed a better understanding of the functions of all HSP70 in the organism. They may serve as potential biomarkers for detecting the effects of global climate change in the Antarctic Ocean.

Over the last few years, a technological revolution in computational biology has evolved and contributed to scientific knowledge by manipulating artificial intelligence to assist in the annotation of protein functions (Li *et al.*, 2023; Sadiqi *et al.*, 2022). In protein structure analysis, computational biology tools can predict the three-dimensional structure of a protein based on its amino acid sequence. This is done through techniques such as homology modelling, which compares the target protein sequence to known protein structures to generate a predicted structure, such as SWISS Modeller and AlphaFold2 software. These tools can also analyse the stability and interactions of proteins, providing insights into their functional properties. The AlphaFold2 software is reported to be able to accurately predict the three-dimensional protein structures, making protein function determination in cells much easier (Schauperl & Denny, 2022). Analysis of protein functions using bioinformatics tools and AI provides rapid and ideal information on the evolutionary relationships between the sequences, phylogenetic profiling, 3D structural conformation and comparative analysis of protein structures (Zaman *et al.*, 2022).

By studying the evolutionary relationships between protein sequences, researchers can identify proteins that are unique to Antarctic organisms or have undergone specific adaptations in response to extreme cold conditions. This information can help in understanding the molecular mechanisms underlying the resilience of Antarctic organisms to environmental stressors. Due to the relevance of the structural and functional properties of HSP70 in *G. antarctica*, this work aimed to assign functions to the rest of the HSP70 present in the genome of this species. We also compared the protein 3D structures provided by AlphaFold2 with SWISS-MODEL. The novelty of this research is

to provide crucial information regarding HSP70 that may decipher how it works in *G. antarctica*, which can allow us to determine the potential of HSP70 to be used as a biomarker in measuring the effects of global warming in the Antarctic.

METHODOLOGY

Physicochemical Analysis

The FASTA format of amino acid sequences of the *G. antarctica* HSP70 (Accession number: LAN_10_353, LAN_01_088, LAN_09_065, LAN_10_316, LAN_12_055, LAN_13_463, LAN_15_171, and LAN_16_202) were obtained from the *G. antarctica* genome database (https://nibm.my/glacier/) as the target template and used for further analyses. Physicochemical properties of the proteins, including molecular weight, amino acid composition, theoretical isoelectric point (pI), the total number of positive and negative residues, extinction coefficient (EC), instability index (II), aliphatic index (AI) and grand average of hydropathicity (GRAVY) were analyzed using Expasy's ProtParam tool (https://web.expasy.org/protparam/) (Gasteiger *et al.*, 2003). The identification of each gene is labelled in Table 1. The signal peptide was analyzed using SignalP-5.0 (https://services.healthtech.dtu.dk/service.php?SignalP-5.0) (Almagro Armenteros *et al.*, 2019). SOSUI server was used to determine whether it is a soluble or a transmembrane protein, while CYS_REC (http://linux1.softberry.com) was used to predict the presence of cysteine residues and their bonding patterns (Hirokawa *et al.*, 1998).

Subcellular Localization Identification

The subcellular localization was anticipated by utilizing the CELLO server (http://cello.life.nctu.edu.tw) (C.-S. Yu *et al.*, 2014). CELLO provides predictions for four main subcellular locations: cytoplasmic, secreted, potentially surface-exposed, and membrane (Tasneem *et al.*, 2023). The results were further cross-checked by using PSORTb (https://www.psort.org/psortb) servers which were used to predict the subcellular localization of the HSP70 proteins (N. Y. Yu *et al.*, 2010). PSORTb takes into account various features, including amino acid composition, signal peptides, and protein motifs, to make predictions. It provides refined localization subcategories and has been continuously improved to enhance prediction accuracy and coverage (Yu *et al.*, 2010).

Multiple Sequence Alignment and Phylogenetic Study

Multiple sequence alignments were done to compare all sequences of HSP70 from *G. antarctica* via Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) platform. Multiple sequence alignment (MSA) is crucial to understand the evolutionary relationships between sequences and analyzing their similarities and dissimilarities (Asir *et al.*, 2021). The phylogenetic study, on the other hand was done to BLAST the sequences through the gene database to observe evolutionary relationships via MEGA 11 (Tamura *et al.*, 2021). The selection of the most suitable template in the phylogenetic study of HSP70 in *G. antarctica* would depend on several criteria. One important criterion is the degree of sequence similarity between the HSP70 sequence of *G. antarctica* and the potential templates. Templates with higher sequence similarity to the target sequence are more likely to provide accurate and reliable phylogenetic information (Drini *et al.*, 2016).

Secondary Structure Prediction

The Self-Optimized Prediction Method with Alignment server-SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) was used to predict the studied protein's secondary structure (Combet *et al.*, 2000). SOPMA utilizes five independent algorithms to predict the secondary structure, including alpha-helix, beta-sheet, and random coil

regions (Zhao *et al.*, 2023). The result was cross-checked by using a PSI-blast-based secondary structure predicting the PSIPRED server (http://bioinf.cs.ucl.ac.uk/psipred/) (Altschul *et al.*, 1997). PSIPRED uses two consecutive feed-forward neural networks to predict the secondary structure of proteins (Bryson *et al.*, 2005). It has been shown to be effective in predicting secondary structure elements such as alpha-helices and beta-sheets (McGuffin *et al.*, 2000). These servers are important to decipher secondary structure as it is closely related to the stability, folding kinetics, and interactions of proteins. By accurately predicting the secondary structure, researchers can gain insights into the protein's folding pathway, stability, and potential functional regions (McGuffin *et al.*, 2000).

Comparative Homology Modelling and Structure Assessment

Homology modelling of the proteins was performed by the AlphaFold 2 and SWISS-MODEL servers, which aligned an input target with pre-existing templates to generate a series of predicted models (Arnold *et al.*, 2006; Jumper *et al.*, 2020). The most suitable template to build the 3D model was selected based primarily on sequence identity. The general steps for template selection in homology modelling are template identification at Protein Data Bank (PDB) and sequence alignment with the target protein. Stereochemical quality and accuracy of the predicted models were analyzed using PROCHECK's Ramachandran plot analysis, ERRAT, PROVE, Verify3D, all four available from the SAVES (https://saves.mbi.ucla.edu/), and ProQ software (Laskowski *et al.*, 1993). The Ramachandran plot analysis provided by PROCHECK identifies residues that may have unfavorable backbone conformations or structural irregularities (Mao *et al.*, 2023). ERRAT, PROVE, Verify3D, and ProQ, on the other hand, evaluate various aspects of the protein structure, including overall structural quality, non-bonded atom-atom interactions, stereochemical properties, and compatibility with experimental data (Hosseinian *et al.*, 2018). Structural analysis was performed, and model figures were generated and superimposed by the PyMOL v2.0 (Briggs, 2000).

Functional Annotation Analysis

To identify the conserved domain of the protein sequence, the Conserved Domain Search Service (CD Search) (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) from NCBI was used (Yang *et al.*, 2020). The protein sequence analysis and classification server InterProScan (https://www.ebi.ac.uk/interpro/search/sequence) was then used for the functional analysis of the protein. CD Search is a web-based tool that uses BLAST heuristics to detect structural and functional domains in protein sequences while InterProScan searches protein sequences against various databases, including Pfam, ProDom, and ProSiteProfiles, to retrieve conserved domains, motifs, and corresponding Gene Ontology (GO) terms (Jazi *et al.*, 2017).

Intraprotein interactions

Intraprotein interactions which analyze the ionic, aromatic-aromatic, aromatic-sulphur, and cation-pi interactions were analyzed using Protein Interaction Calculator, PIC (http://pic.mbu.iisc.ernet.in/job.html) (Tina *et al.*, 2007). The Protein Interaction Calculator (PIC) calculates various types of interactions within a protein structure, including disulfide bonds, hydrophobic interactions, ionic interactions, hydrogen bonds, aromatic-aromatic interactions, aromatic-sulfur interactions, and cation-pi interactions (Tina *et al.*, 2007). PIC determines these interactions based on standard, published criteria and provides a visualization of the identified interactions using RasMol and Jmol interfaces. The interaction cut-off values were set to default; ionic interactions (6 Å), aromatic-aromatic interactions (4.5 to 7 Å), aromatic-sulphur interactions (5.3 Å) and cation-pi interactions (6 Å). The structures of monomeric *G. antarctica* HSP70 were uploaded in PDB format (.pdb). The number of ionic interactions was detected for Arg, Lys, His, Asp and Glu pairs falling with 6 Å (default). The number of aromatic-aromatic interactions was calculated based on pairs of phenyl ring centroids that are separated between 4.5 to 7 Å. The number

of aromatic-sulphur interactions was detected from the interactions between the sulphur atoms of cysteine and methionine and the aromatic rings of phenylalanine, tyrosine and tryptophan within 5.3 Å. The amount of cation-pi interactions was detected when a cationic side chain (Lys or Arg) is near an aromatic side chain (Phe, Tyr, or Trp) within 6 Å separation. The chosen interaction cut-off values are based on established criteria determined based on empirical observations and experimental data analysis and comparison with known protein structures and interactions. The reference supporting the selection of these cut-off values is the original publication describing the PIC server (Tina *et al.*, 2007).

RESULT

Physicochemical Analysis

In this study, several different computational tools and servers were used to analyse the physicochemical characteristics and reliably predict the tertiary structure of the *G. antarctica* six HSP70 proteins. The HSP70 genes were retrieved from the *G. antarctica* genome database with amino acid lengths ranging from 603 to 782 (63.73 to 85.60 kDa) with no signal peptide (Table 1). To date, no PDB structure has been deposited yet into the PDB databank for all these HSP70.

Table 1. Retrieval of HSP70 from *G. antarctica* public database with hyperlinks and their features.

HSP70 Genes	Gene ID	Genbank Accession Num.	Amino Acid Length	Molecular Weight (Da)	Signal Peptide
gahsp70-1	LAN_10_353	MZ313862.1	672	72223.51	No
gahsp70-2	LAN_01_088	JF412505.1	627	67861.13	No
gahsp70-3	LAN_09_065	OQ625425	782	85597.88	No
gahsp70-4	LAN_10_316	OQ625426	603	63733.4	No
gahsp70-5	LAN_12_055	OQ625427	637	69661.49	No
gahsp70-6	LAN_16_202	OQ625428	635	69323.49	No

Among other findings, as tabulated in Table 2, the computed pI values for all the *G. antarctica* HSP70 proteins were below 7, indicating that they worked best at acidic pH. The extinction coefficient (EC) of all gahsp70s varied between 19035 to 64080, which concordances with the tyrosine, tryptophan and phenylalanine residues (Batabyal *et al.*, 2021). The grand average hydropathicity (GRAVY) of all the hsp70 were all in negative value except for gahsp70-4, corroborating most of those proteins are hydrophilic and highly soluble in water.

Table 2. Physicochemical properties of HSP70 proteins from *G. antarctica*.

HSP70	pI	Extinction Coefficient (EC)	Instability Index (II)	Aliphatic Index (AI)	Grand average of Hydropathicity (GRAVY)
gahsp70-1	5.8	19035	33.36	88.11	-0.320
gahsp70-2	4.8	35995	29.76	81.26	-0.414
gahsp70-3	4.8	64080	38.12	88.62	-0.328
gahsp70-4	5.3	23045	33.99	100.60	0.072
gahsp70-5	4.8	27390	38.17	85.56	-0.412
gahsp70-6	6.3	24660	34.57	84.65	-0.304

Theoretically, a negative GRAVY indicates that the protein is non-polar, and a positive value indicates that the protein is polar (Gasteiger *et al.*, 2005). The instability index (II), on the other hand,

showed that all the gahsp70s are stable proteins, as the value index is less than 40, indicating that they are stable in the test tube. The aliphatic index (AI) for the gahsp70 sequences ranged from 81.26 to 100.60. Interestingly, gahsp70-4 has the highest AI of 100.6, while others ranged between 81.26 to 88.62. A high aliphatic index indicates the increased thermostability of globular proteins (Ikai, 1980).

Table 3. Amino acid residues composition of HSP70 proteins from *G. antarctica*.

Amino Acid Composition	*gahsp70-1*		*gahsp70-2*		*gahsp70-3*		*gahsp70-4*		*gahsp70-5*		*gahsp70-6*	
	N	%	N	%	N	%	N	%	N	%	N	%
Ala (A)	71	10.6	64	10.2	91	11.6	72	11.9	49	7.7	62	9.8
Arg (R)	32	4.8	27	4.3	42	5.4	25	4.1	32	5.0	38	6.0
Asn (N)	28	4.2	32	5.1	24	3.1	17	2.8	25	3.9	22	3.5
Asp (D)	39	5.8	43	6.9	47	6.0	34	5.6	47	7.4	43	6.8
Cys (C)	2	0.3	2	0.3	5	0.6	2	0.3	1	0.2	4	0.6
Gln (Q)	28	4.2	18	2.9	21	2.7	17	2.8	25	3.9	20	3.1
Glu (E)	54	8.0	52	8.3	76	9.7	37	6.1	53	8.3	46	7.2
Gly (G)	51	7.6	53	8.5	42	5.4	40	6.6	54	8.5	42	6.6
His (H)	8	1.2	6	1.0	7	0.9	10	1.7	6	0.9	4	0.6
Ile (I)	44	6.5	40	6.4	39	5.0	28	4.6	37	5.8	31	4.9
Leu (L)	45	6.7	43	6.9	70	9.0	60	10.0	53	8.3	53	8.3
Lys (K)	53	7.9	48	7.7	57	7.3	33	5.5	48	7.5	48	7.6
Met (M)	10	1.5	11	1.8	10	1.3	4	0.7	10	1.6	10	1.6
Phe (F)	19	2.8	21	3.3	22	2.8	22	3.6	25	3.9	29	4.6
Pro (P)	25	3.7	26	4.1	43	5.5	32	5.3	30	4.7	23	3.6
Ser (S)	50	7.4	31	4.9	48	6.1	54	9.0	36	5.7	45	7.1
Thr (T)	43	6.4	52	8.3	53	6.8	40	6.6	43	6.8	53	8.3
Trp (W)	1	0.1	3	0.5	7	0.9	2	0.3	2	0.3	2	0.3
Tyr (Y)	9	1.3	13	2.1	17	2.2	8	1.3	11	1.7	9	1.4
Val (V)	60	8.9	42	6.7	61	7.8	66	10.9	50	7.8	51	8
Pyl (O)	0	0	0	0	0	0	0	0	0	0	0	0
Sec (U)	0	0	0	0	0	0	0	0	0	0	0	0
Negatively charged residues (Asp + Glu)	93	13.8	95	15.2	123	15.8	71	11.8	100	15.7	89	14.0
Positively charged residues (Arg + Lys)	85	12.6	75	12.0	99	12.7	58	9.6	80	12.6	86	13.5

Table 4. Solubility, cysteine occurrence pattern and probability of cysteine residue pairing in HSP70 proteins from *G. antarctica*.

HSP70	Solubility	Cysteine position	Status	Score
gahsp70-1	Soluble	Cys53	No S-S bond	-40.7
		Cys353	No S-S bond	-20.7
gahsp70-2	Soluble	Cys18	No S-S bond	-38.4
		Cys268	No S-S bond	-25.6
gahsp70-3	Soluble	Cys201	No S-S bond	-26.1
		Cys363	No S-S bond	-37.2
		Cys637	No S-S bond	-25.7
		Cys683	No S-S bond	-18.9
		Cys724	No S-S bond	-22.0
gahsp70-4	Soluble	Cys55	No S-S bond	-26.3
		Cys315	No S-S bond	-16.6
gahsp70-5	Soluble	Cys408	No S-S bond	-35.6
gahsp70-6	Soluble	Cys37	No S-S bond	-36.3
		Cys290	No S-S bond	-31.4
		Cys473	No S-S bond	-27.7
		Cys499	No S-S bond	-22.8

Analysis of the amino acid composition unveiled high amounts of alanine (average 10.3%), valine (8.35 %) and glutamic acid (7.93%) were the most abundant in gahsp70 (Table 3). Furthermore, all of them showed a trend of slightly higher negatively charged residues compared to positively charged residues. Table 4, on the other hand, showed that all gahsp70 are soluble and with the presence of one to several cysteine residues ranging from 2 to 5. However, no evidence was found for the existence of disulphide bonds. The covalent crosslinks produced by disulphide bridges affect protein folding and stability (Creighton, 1988).

Subcellular Localization Identification

The subcellular localization of a protein determines its functions and activities (Nachmias *et al.*, 2007). The functional properties, determination of interaction, and genome annotation of a protein are highly influenced by its subcellular location. Protein subcellular localization predicted from the CELLO and PSORTb server indicated that all the *G. antarctica* HSP70 protein belongs to the cytoplasmic portion which may reflect their function as molecular chaperones in a wide range of cellular processes (see Table 5).

Table 5. Subcellular localization identification of HSP70 proteins from *G. antarctica*.

HSP70	CELLO Prediction (Reliability Score)				PSORTb (Localization Score)			Localization
	Mitochondrial	Cytoplasmic	Extracellular	Plasma Membrane	Cytoplasmic	Cell Wall	Extracellular	
gahsp70-1	3.353	0.757	0.029	0.016	9.96	0.02	0.01	Cytoplasmic
gahsp70-2	0.628	3.322	0.024	0.017	9.96	0.02	0.01	Cytoplasmic
gahsp70-3	1.164	2.099	0.035	0.036	9.96	0.02	0.01	Cytoplasmic
gahsp70-4	0.676	0.922	0.112	0.303	9.65	0.17	0.17	Cytoplasmic
gahsp70-5	0.468	1.232	0.041	0.023	9.96	0.02	0.01	Cytoplasmic
gahsp70-6	1.663	0.953	0.033	0.030	9.65	0.17	0.17	Cytoplasmic

Multiple Sequence Alignment and Phylogenetic Study

All six genes of the *G. antarctica* HSP70 showed that they share similarities between 24.71 to 61.95, as shown in the percentage identity matrix (Table 6). The amino acids of gahsp70-4, -5 and -6 showed high similarities at 51.62 to 61.95% while gahsp70-1, -2 and -3 were similar at 24.71 to 26.6%. The tree topology showed that the HSP70s were basal to other HSP70 family members, indicating that they have been conserved throughout evolution (Figure 1). Interestingly, all 6 cytosolic *G. antarctica* HSP70s formed different clades. However, the clades still refer to the same group of HSP70 organisms ranging from prokaryotes to eukaryotes. This suggests the evolution of *G. antarctica* with multiple HSP70 family members that are closely related paralogs arise through duplication.

Table 6. Percentage identity matrix of the aligned amino acid sequences of the six HSP70.

Percentage Identity Index	gahsp70-1	gahsp70-2	gahsp70-3	gahsp70-4	gahsp70-5	gahsp70-6
gahsp70-1	100.00	26.60	26.26	29.38	30.14	25.98
gahsp70-2	26.60	100.00	24.71	30.14	29.42	30.48
gahsp70-3	26.26	24.71	100.00	45.26	47.93	43.67
gahsp70-4	29.38	30.14	45.26	100.00	60.13	51.62
gahsp70-5	30.14	29.42	47.93	60.13	100.00	61.95
gahsp70-6	25.98	30.48	43.67	51.62	61.95	100.00

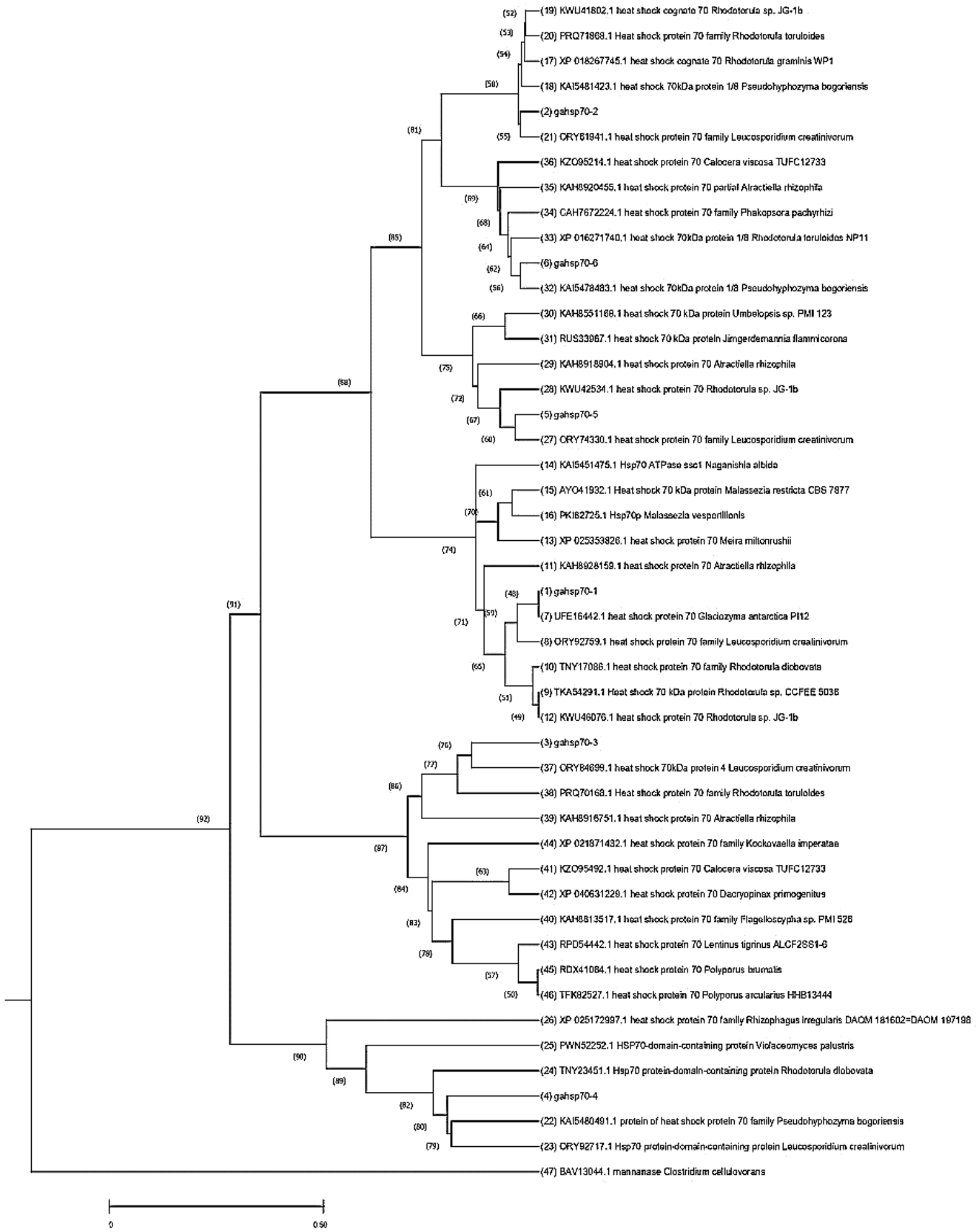

Figure 1. A phylogenetic tree was generated based on the alignment of the six Hsp70 amino acid sequences with HSP70 proteins from other organisms. A mannanase from *Clostridium cellulovorans* was used as an outlier. The tree was constructed using the neighbour-joining method. The numbers close to the individual branches represent the percentage of the 1000 bootstrap iterations supporting the branch.

Secondary Structure Prediction

A protein function is highly associated with its structure. The secondary structures of a protein are comprised of helices, sheets, turns, and coils. The secondary structures of the gahsp70s, obtained from the SOPMA server (Figure 2), demonstrated that the alpha helices were found to be most predominant (35.49 to 49.49%), random coil (29.02 to 42.29%), extended strand (13.94 to 19.20%), and beta-turn (3.20 to 8.32%) as shown in Table 7. A similar result was found from the PSIPRED server (Figure S1) validated the secondary structure result.

Table 7. Secondary structure prediction of HSP70 proteins from *G. antarctica*.

HSP70	Alpha Helix		Extended Strand		Beta Turn		Random coil	
	N	%	N	%	N	%	N	%
gahsp70-1	308	45.83	129	19.20	40	5.95	195	29.02
gahsp70-2	266	42.42	119	18.98	42	6.70	200	31.90
gahsp70-3	387	49.49	109	13.94	25	3.20	261	33.38
gahsp70-4	214	35.49	101	16.75	33	5.47	255	42.29
gahsp70-5	270	42.39	122	19.15	53	8.32	192	30.14
gahsp70-6	256	40.31	121	19.06	47	7.40	211	33.23

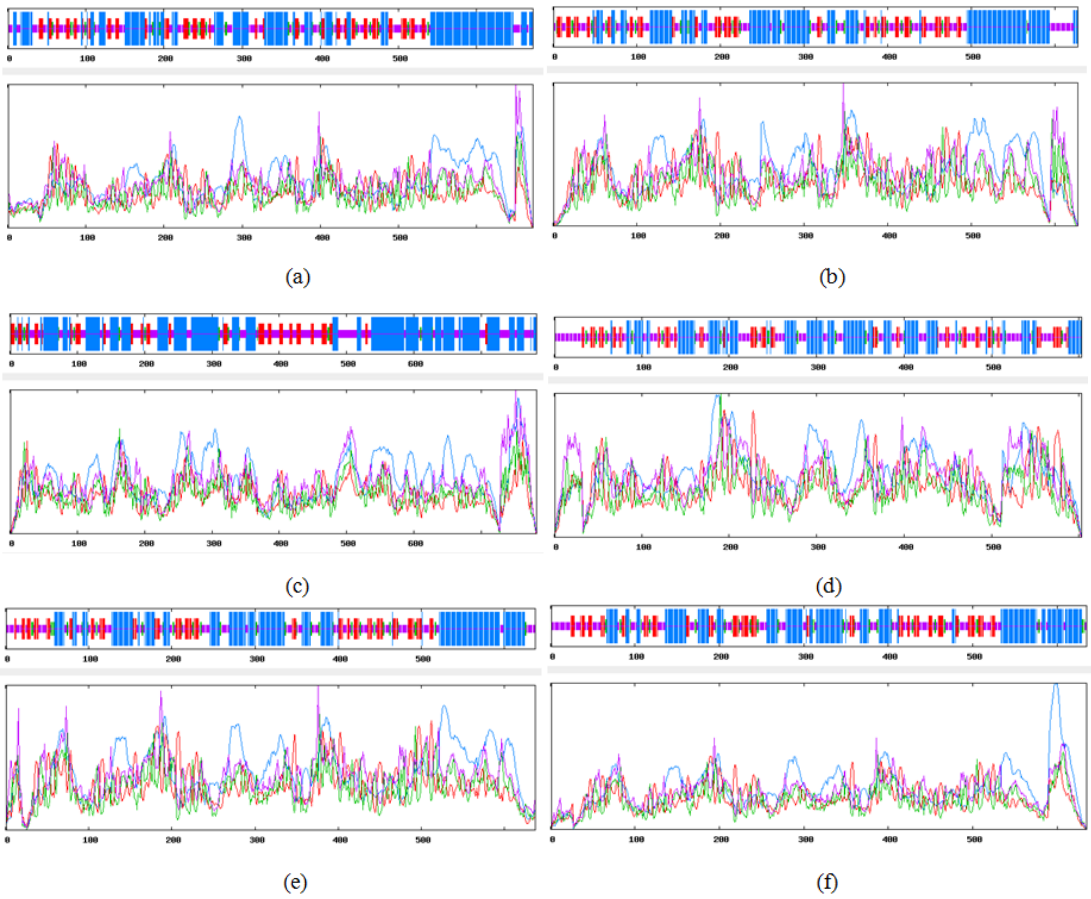

Figure 2. SOPMA secondary structure prediction diagram based on amino acid annotation. The blue colour indicates alpha helix, red indicates extended strands, green indicates beta-turn and purple indicates random coil. (a) gahsp70-1, (b) gahsp70-2, (c) gahsp70-3, (d) gahsp70-4, (e) gahsp70-5, and (f) gahsp70-6.

Comparative Homology Modelling and Structure Assessment

3D models of gahsp70 were built via two different platforms, Alphafold2 and SWISS-MODEL, to compare and choose the best model to represent the gahsp70 and uploaded to the database

subsequently. Different template models were chosen by default which has the highest similarity to each gahsp70 sequence, respective to Alphafold2 and SWISS-MODEL algorithm (Table 8). Verification of the results, using different tools, invariably indicated a good quality of the proposed models (Table 9). The Ramachandran plot analysis, where a good model would be expected to have over 80% of residues in the most favoured regions, suggested a good quality for all models from both modelling platforms.

Overall ERRAT quality factor value, expressed as the percentage of the protein for which the calculated value falls below the 95% rejection limit, was 86.4 to 96.3% for AlphaFold2 and 86.2 to 94.6% for SWISS-MODEL. The value indicates non-bonded atomic interactions, with higher scores equal to higher quality. The generally accepted range is >50 for a high-quality model; hence all models are in good quality range. Verify3D analysis of the models revealed that models generated via AlphaFold2 had an average 3D-1D score ≥0.2, ranging from 83.7 to 94.2%, whereas SWISS-MODEL ranged from 80.8 to 95.4%. A good model in Verify3D should have at least 80% of the amino acids with a score ≥ 0.2 in the 3D-1D; hence all models are of high quality. Most of the models at the resolution of 2 Å or better have RMS Z-scores that we consider good (between −2.0 and 2.0), in which all generated models are within that range. The LGscore indicated "very good" (value >5.0) for all gahsp70 models for both modelling platforms. All these validation tools strongly suggested that both proposed 3-D models could be accepted as reliable with high confidence. Figure 3 shows the gahsp70 3D models generated from AlphaFold2 and SWISS-MODEL.

Superimposition of the generated 3D models via AlphaFold2 and SWISS-MODEL with their respective templates showed values below 1 which indicated high similarity between the models. Meanwhile, the models generated from AlphaFold2 and SWISS-MODELs were also superimposed, yielding Root Mean Square Deviation (RMSD) (Å) between 0.754 to 1.379. This indicates that models generated from both platforms are highly similar to each other. Figure 3 shows the gahsp70 3D models generated from AlphaFold2 and SWISS-MODEL. Figure S2 shows the superimposed gahsp70 3D models generated from AlphaFold2 and SWISS-MODEL.

Table 8. Structural templates in the Protein Data Bank (PDB) and AlphaFold2 libraries.

Template	[PDB ID] Protein Name, Organisms	[AlphaFold2 ID] Protein Name, Organisms
gahsp70-1	[7KO2] DnaK, *Escherichia coli* K-12	[A0A1Y2G6S2] HSP70, *Leucosporidium creatinivorum*
gahsp70-2	[7SQC] HSP70, *Chlamydomonas reinhardtii*	[H9LJG2] HSP70, *Glaciozyma antarctica*
gahsp70-3	[3D2E] HSP70, *Homo sapiens, Saccharomyces cerevisiae*	[A0A1Y2FL88] HSP70, *Leucosporidium creatinivorum*
gahsp70-4	[6SR6] Putative HSP, *Thermochaetoides thermophila* DSM 1495	[M7WDV4] HSP70, *Rhodosporium toruloides* NP11
gahsp70-5	[5E84] 78 kDa glucose-regulated protein, *Homo sapiens*	[A0A1Y2ERW4] Endoplasmic reticulum chaperone BiP, *Leucosporidium creatinivorum*
gahsp70-6	[7SQC] HSP70, *Chlamydomonas reinhardtii*	[A0A511K7S9] HSP70, *Rhodosporium toruloides*

Table 9. Assessment of the predicted three-dimensional structures of the six HSP70.

Validation index		*gahsp70-1*		*gahsp70-2*		*gahsp70-3*		*gahsp70-4*		*gahsp70-5*		*gahsp70-6*	
		AF2	**SM**	**AF2**	**SM**	**AF2**	**SM**	**AF2**	**SM**	**AF2**	**SM**	**AF2**	**SM**
Ramachandran plot	Residues in the most favored region (%)	88.7	95.5	89.0	92.1	85.5	90.1	83.2	87.9	90.7	93.1	90.8	93.7
	Residues in additional allowed regions (%)	6.9	4.1	8.1	6.6	8.3	7.7	11.2	9.9	6.5	5.5	7.7	5.3
	Residues in generously allowed regions (%)	2.5	0.4	2.0	0.9	4.6	1.6	3.4	1.6	1.8	0.6	1.1	0.7
	Residues in disallowed regions (%)	1.9	0.0	0.9	0.4	1.6	0.7	2.3	0.6	0.9	0.8	0.4	0.3
ERRAT	Quality factor value	88.33	96.38	86.44	94.67	87.65	86.20	89.55	89.80	88.65	92.27	93.03	94.67
Verify3D	% 3D-1D >-0.2	88.99	92.88	92.98	94.20	83.76	89.04	85.74	80.80	88.70	88.87	92.76	95.41
PROVE	Z−score RMS (Å)	1.388	1.408	1.435	1.400	1.447	1.520	1.359	1.718	1.385	1.470	1.358	1.260
ProQ	LGscore	8.899	7.767	6.999	7.539	8.555	8.678	9.051	10.715	7.536	7.954	8.011	7.574
Superimposed RMSD (Å)	Between 3D model and template	0.482	0.086	0.573	0.049	0.594	0.111	0.600	0.105	0.613	0.097	0.917	0.090
	Template ID	A0A1Y2G6S2	7KO2	H9LJG2	7SQC	A0A1Y2FL88	3D2E	M7WDV4	6SR6	A0A1Y2ERW4	5E84	A0A511K7S9	7SQC
	Between AF2 and SM models	0.941		0.819		1.379		1.038		0.754		0.952	

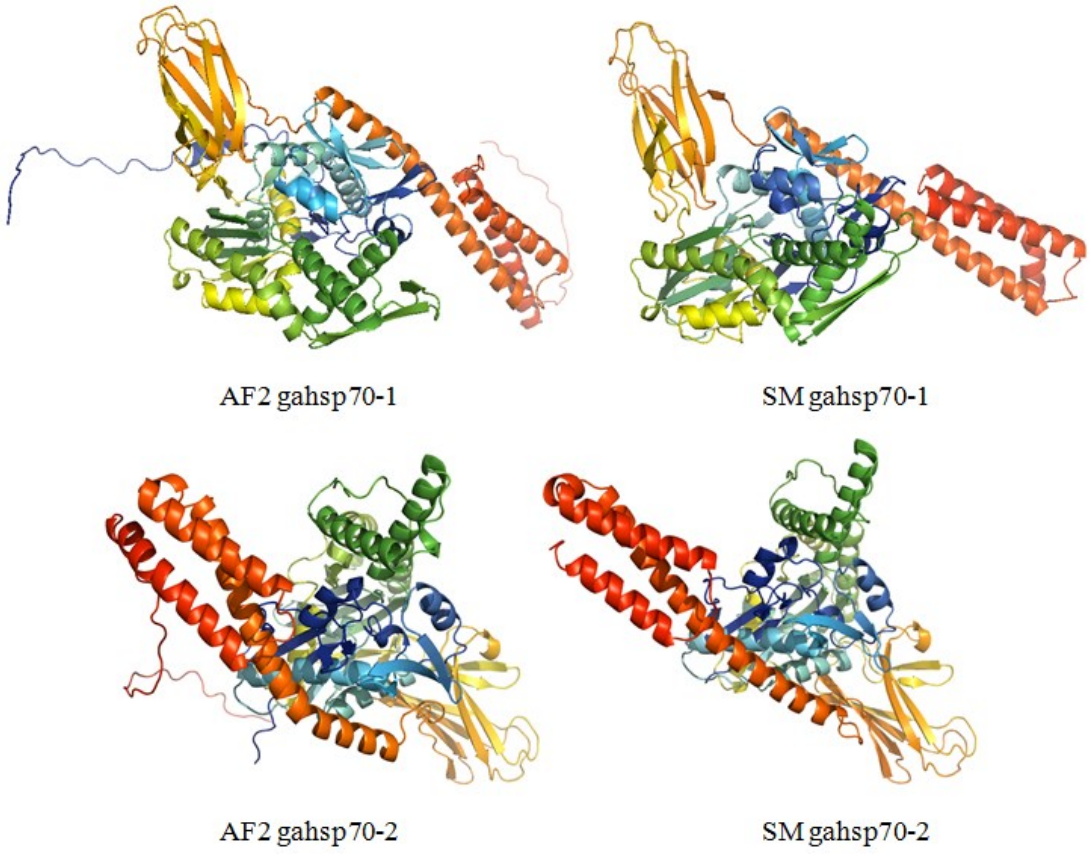

Figure 3. 3D homology model rendered by the AlphaFold2 (AF2) and SWISS-MODEL (SM) platforms for all six gahsp70, respectively. Each colour represents different chains on the protein structure.

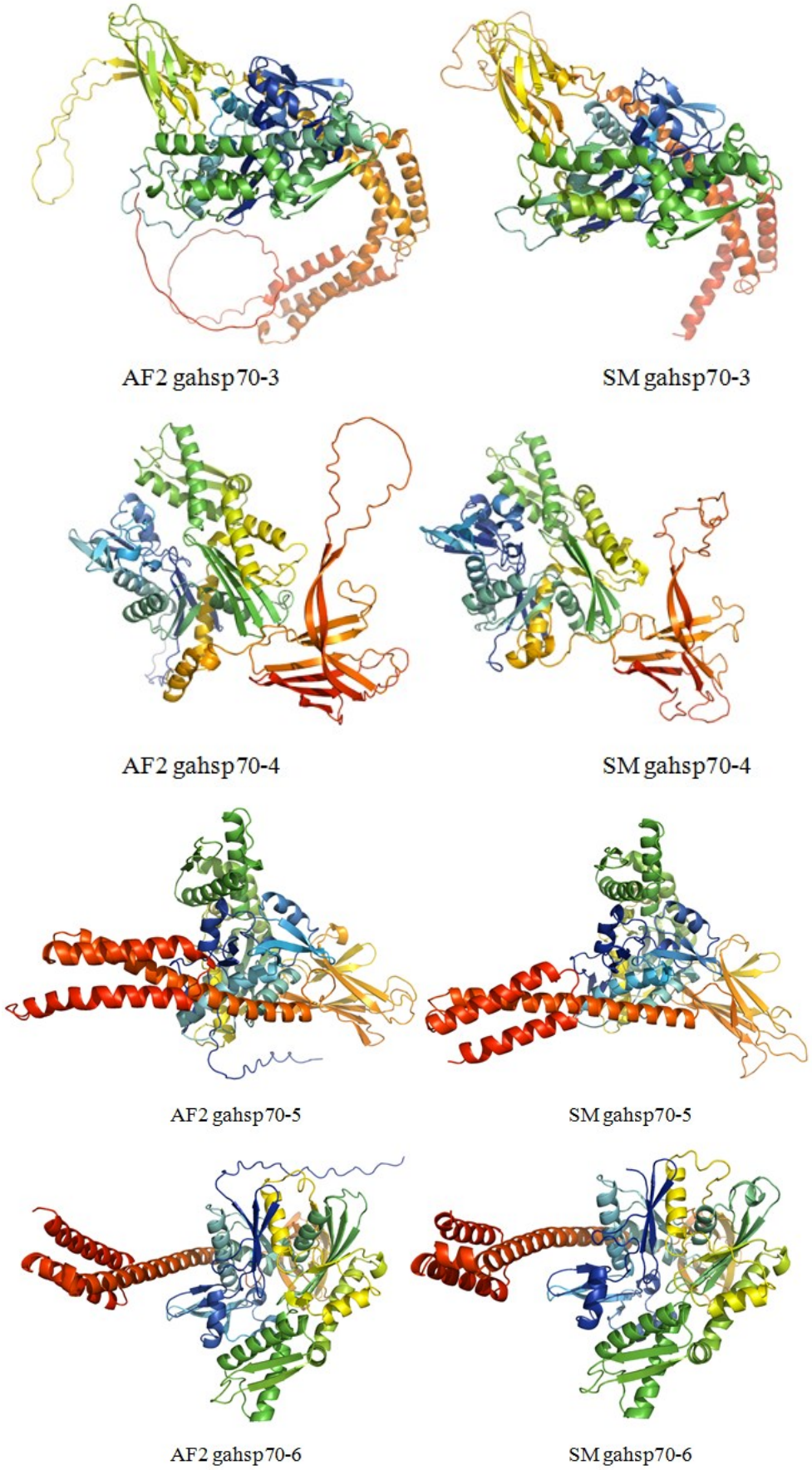

Figure 3 (Cont). 3D homology model rendered by the AlphaFold2 (AF2) and SWISS-MODEL (SM) platforms for all six gahsp70, respectively. Each colour represents different chains on the protein structure.

Between the two homology modelling platforms, the models generated from AlphaFold2 and SWISS-MODEL were further evaluated based on their identities and coverage versus respective templates (Table 10). The result showed that models generated via AlphaFold2 have greater confidence and coverage when compared to their counterparts generated via SWISS-MODEL. Models generated via AlphaFold2 produce a per-residue estimate of its confidence on a scale from 0 - 100. This confidence measure is called pLDDT and corresponds to the model's predicted score on the lDDT-Cα metric. Regions with pLDDT between 70 and 90 are expected to be modelled well (a generally good backbone prediction), as shown in the gahsp70 models generated (Figure 4). The percentages of residues covered in the generated models were all above 90%, higher than SWISS-MODEL generated models, which utilizes QSQE confidence higher than 0.7 as a cut-off value for generally reliable models. Furthermore, all of the gahsp70 models generated via AlphaFold2 also passed all the structural validation tests which indicated high-quality models (Figure S3).

Table 10. Comparison between gahsp70 models generated via AlphaFold and Swiss Model and their identities and coverage versus respective templates.

HSP70	Protein Length	AlphaFold2 (pLDDT >70)				Swiss Model (QSQE Confidence > 0.7)			
		Identities with Template (%)	Residues Location	Number of Residues	Percentage of Residues (%)	Identities with Template (%)	Residues Location	Number of Residues	Percentage of Residues (%)
gahsp70-1	672	89.70	40-646	606	90.18	60.07	40-631	591	87.95
gahsp70-2	627	97.90	6-591	585	93.30	73.84	6-588	582	92.82
gahsp70-3	782	72.20	3-748	745	95.27	40.98	2-644	642	82.10
gahsp70-4	603	57.50	31-603	572	94.86	36.83	31-589	558	92.54
gahsp70-5	637	92.60	21-637	616	96.70	66.33	21-621	600	94.19
gahsp70-6	635	88.70	25-635	610	96.06	60.49	25-625	600	94.49

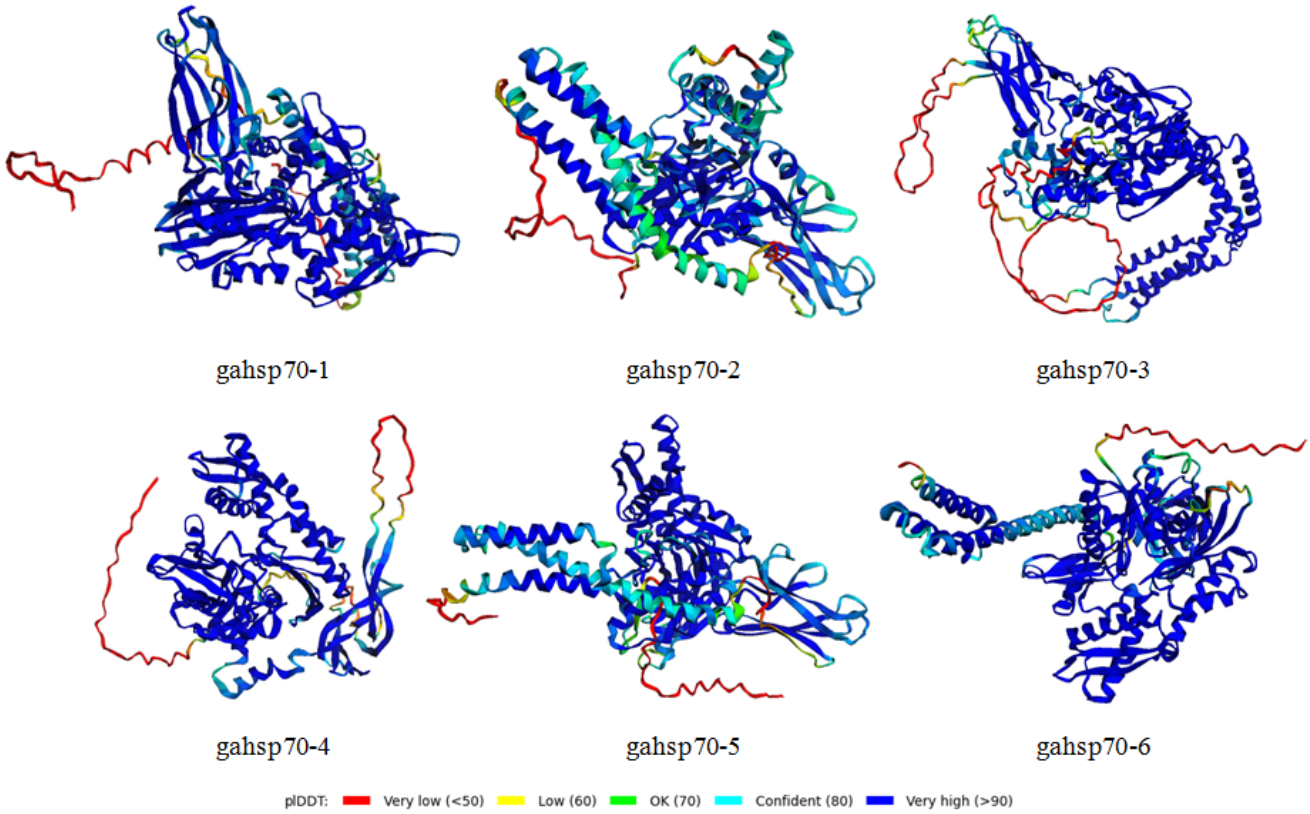

Figure 4. 3-D homology model rendered by the Alpha Fold 2 (AF2) programs for all six gahsp70 in which the structures are coloured according to the median of the pIDDT confidence scores; red: very low, yellow: low, green: OK, cyan: confident, and blue: very high.

Structure Similarity Assessment

Protein structure alignment was estimated using RMSD to evaluate the similarity between the superimposed model to the reference protein structure. Since the RMSD values between template models and sequence models were relatively close, this would imply that the generated models are structurally similar. The low RMSD values ranging from 0.48 to 0.91Å indicated the high structural similarity between the templates and models. In addition, the high identities with templates ranging from 57.5 to 97.9% implied the high conservation level in the HSP70 structures. This shows that the 3D structures of reference proteins and models are structurally almost identical. The high structural similarity to the HSP70 proteins reflected a high level of evolutionary conservation among the HSP70 protein family members from a wide range of taxonomically different organisms. This may also imply a high level of functional conservation as molecular chaperones and folding catalysts.

Comparison of Structural Interactions

To unveil the thermal adaptations in *G. antarctica* HSP70, the HSP70 structure models were compared to their mesophilic homologs. Since all of the structure models share a high level of sequence identity to their homologs (Table 10), ranging from 57.5 to 97.9% identical, comparative structural analysis at the atomic level provides some perspectives on protein flexibility and thermal adaptation. Protein flexibility is strongly related to thermal adaptation; hence, the adaptive variation in structural flexibility reflects the species' adaptation ability to different temperatures (Dong *et al.*, 2018; Teilum *et al.*, 2011). The secondary structures of HSP70 contain mostly their residues in the α-helix (Table 11). Compared to their mesophilic homologs, all *G. antarctica* HSP70 have less percentage of alpha helix except for gahsp-1 and gahsp-2. When compared to their mesophilic homologs, the distribution of β-sheet in *G. antarctica* HSP70 was lesser in gahsp-1, gahsp-2 and gahsp-5 by 1.22, 0.52 and 0.01 %, respectively. Interestingly, when compared to their mesophilic counterparts, most of the *G. antarctica* HSP70 contained more extended strands by 0.2, 0.32, 0.97 and 0.37% in gahsp-2, gahsp-4, gahsp-5 and gahsp-6, respectively. While gahsp-1, gahsp-2 and gahsp-5 showed lesser percentage of random coil contribution compared to their homologs, the other *G. antarctica* HSP70 (gahsp-3, gahsp-4 and gahsp-6) had a higher percentage of coils by 1.85, 2.1 and 1.63 %, respectively.

Table 11. Amino acid compositions in coil, helix, and strand of psychrophilic HSP70 from *G. antarctica* and mesophilic HSP70 from other organisms. The difference in amino acid compositions is indicated in the ( ).

	α-Helix (%)	**Random Coil (%)**	**β-Sheet (%)**	**Extended strand (%)**
gahsp-1	45.83 (2.63%)	29.02	5.95	19.20
A0A1Y2G6S2	43.20	29.75 (0.73%)	7.17 (1.22%)	19.88 (0.68%)
gahsp-2	42.42 (0.53%)	31.90	6.70	18.98 (0.2%)
H9LJG2	41.89	32.10 (0.2%)	7.22 (0.52%)	18.78
gahsp-3	49.49	33.38 (1.85%)	3.20 (0.06%)	13.94
A0A1Y2FL88	50.75 (1.26%)	31.53	3.14	14.57 (0.63%)
gahsp-4	35.49	42.29 (2.1%)	5.47 (0.37%)	16.75 (0.32%)
M7WDV4	38.28 (2.79%)	40.19	5.10	16.43
gahsp-5	42.39	30.14	8.32	19.15 (0.97%)
A0A1Y2ERW4	42.88 (0.49%)	30.61 (0.47%)	8.33 (0.01%)	18.18
gahsp-6	40.31	33.23 (1.63%)	7.40 (0.87%)	19.06 (0.37%)
A0A511K7S9	43.18 (2.87%)	31.60	6.53	18.69

The ion pairs in *G. antarctica* HSP70 and their mesophilic homologs were identified using a distance cut-off of 6 Å. Instead of just counting the number of ion pairs, only a critical ion pair, that is, a polypeptide that bridged a distant region of 10 residues or more, was counted. The critical ion

pairs are most likely important for the stability of the protein structure in the folded state (Bae & Phillips, 2004). The number of ion pairs in *G. antarctica* HSP70s is either more or the same as their homologous counterparts. The number of critical ion pairs is 49–49 for gahsp-1–A0A1Y2G6S2, 41–45 for gahsp-2–H9LJG2, 51–61 for gahsp-3–A0A1Y2FL88, 29–40 for gahsp-4–M7WDV4, 35–35 for gahsp-5– A0A1Y2ERW4 and 42–42 for gahsp6–A0A511K7S9 (Table 12). Among the most prominent reason that has been attributed to the greater stability of protein structures is increased buried surface area (Munson *et al.*, 1996). Analysis of the stability of HSP70 was evaluated based on the presence of ion pairs between buried residues. Interestingly, all *G. antarctica* HSP70s have none or a lesser number of ionic pairs compared to their homologous counterparts. Although gahsp-1 and gahsp-5 have the same number of ion pairs as their homologs, no ion pair was detected in the buried residues. In addition, gahsp-2, gahsp-3 and gahsp-4 have lesser ion pairs compared to their homologs, and no ion pair was detected in their buried residues. Meanwhile, gahsp-6 has the same number of ion pairs as its homolog, including the ion pair in its buried residues.

Table 12. Total ionic interactions in psychrophilic HSP70 from *G. antarctica* and homologous HSP70 from other organisms. The ionic interaction between buried residues is in bracket ().

Protein	Ionic interactions within 6 Å
gahsp-1	49 (0)
A0A1Y2G6S2	49 (2)
gahsp-2	41 (0)
H9LJG2	45 (1)
gahsp-3	51 (0)
A0A1Y2FL88	61 (6)
gahsp-4	29 (0)
M7WDV4	40 (0)
gahsp-5	35 (0)
A0A1Y2ERW4	35 (1)
gahsp-6	42 (1)
A0A511K7S9	42 (1)

Aromatic interactions in protein structures have been shown to play key roles in structure stability (Lanzarotti *et al.*, 2011). Although the identity of *G. antarctica* HSP70 and their homologs are highly similar, the numbers of aromatic interactions are mostly very much lower in psychrophiles than mesophiles (Table 13). The number of aromatic–aromatic interactions is 10–11 for gahsp-1–A0A1Y2G6S2, 10–11 for gahsp-2–H9LJG2, 18–17 for gahsp-3–A0A1Y2FL88, 9–9 for gahsp-4–M7WDV4, 15–17 for gahsp-5– A0A1Y2ERW4 and 20–19 for gahsp6–A0A511K7S9. Meanwhile, the number of aromatic–sulphur interactions is 5–9 for gahsp-1–A0A1Y2G6S2, 4–3 for gahsp-2–H9LJG2, 6–6 for gahsp-3–A0A1Y2FL88, 2–3 for gahsp-4–M7WDV4, 3–2 for gahsp-5– A0A1Y2ERW4 and 7–6 for gahsp6–A0A511K7S9. As for the number of cation–pi interactions, there are 6–8 for gahsp-1–A0A1Y2G6S2, 8–11 for gahsp-2–H9LJG2, 14–22 for gahsp-3–A0A1Y2FL88, 9–7 for gahsp-4–M7WDV4, 7–8 for gahsp-5– A0A1Y2ERW4 and 13–14 for gahsp6–A0A511K7S9.

Aromatic interactions in the buried residues are important in stabilizing the inner core protein regions (Sivasakthi et al., 2013). Overall, most of the *G. antarctica* HSP70s have lesser aromatic interactions in the buried residues when compared to their homologs. In the aromatic–aromatic interactions analysis, all *G. antarctica* HSP70s have lesser buried aromatic residues, except for gahsp-2 which has the same number as its homolog. Moreover, in the aromatic–sulphur interactions, only gahsp-2 has additional pair of aromatic interactions in the buried residues, gahsp-5 has the same number of aromatic interactions as its homolog, while others have lesser aromatic interactions compared to their homologs. Furthermore, analysis of the cation–pi shows that only gahsp-3 and

gahsp-5 have lesser aromatic interaction in the buried residues, while gahsp-1, gahsp-4 and gahsp-6 have more aromatic interaction, and gahsp-2 has the same number of aromatic interaction when compared to their homologs.

Table 13. Aromatic interactions in psychrophilic HSP70 from *G. antarctica* and mesophilic HSP70 from other organisms. The aromatic interactions between buried residues are in bracket ().

Interactions	Aromatic–aromatic	Aromatic–Sulfur	Cation–Pi
gahsp-1	10 (1)	5 (3)	6 (1)
A0A1Y2G6S2	11 (4)	9 (8)	8 (0)
gahsp-2	10 (5)	4 (2)	8 (1)
H9LJG2	11 (5)	3 (1)	11 (1)
gahsp-3	18 (3)	6 (2)	14 (0)
A0A1Y2FL88	17 (7)	6 (3)	22 (3)
gahsp-4	9 (7)	2 (1)	9 (1)
M7WDV4	9 (8)	3 (3)	7 (0)
gahsp-5	15 (6)	3 (1)	7 (0)
A0A1Y2ERW4	17 (11)	2 (1)	8 (1)
gahsp-6	20 (8)	7 (3)	13 (3)
A0A511K7S9	19 (10)	6 (4)	14 (2)

DISCUSSION

Numerous recent studies have reported the importance of the HSP70 protein in an organism's resistance and resilience towards external stresses, particularly heat stress (Chand *et al.*, 2021; Day *et al.*, 2019; Hassan *et al.*, 2019; Kurashova *et al.*, 2020; Nava & Zuhl, 2020). HSP70 chaperones undergo coordinated movements to facilitate misfolded substrate binding and release. The domain layout of Hsp70 is well conserved across different species and homologues, which comprises a nucleotide-binding domain (NBD) and a substrate-binding domain (SBD) that is connected via a short linker. The NBD is made up of four subdomains that are organized into two lobes which results in the formation of a deep cleft (Flaherty *et al.*, 1990; Kityk *et al.*, 2012). The active site where ATP bind and hydrolysis occur is placed at the base of the cleft. ATP makes contact with all four subdomains and can facilitate subtle movements throughout the NBD. Typically, the rate of ATP hydrolysis is low with a turnover of 1 ATP molecule per half an hour (Mayer, 2018; Zuiderweg *et al.*, 2017). The SBD interacts with the substrate protein by engaging a short motif within the target polypeptide. This sequence usually consists of five hydrophobic amino acids that are flanked by charged residues (Rüdiger *et al.*, 1997). This type of motif is present in nearly all proteins, often occurring multiple times, and enables great versatility in substrate selection. In obligate psychrophiles, the HSP70 role is even more enhanced to prevent cellular degradation during temperature rise occur, such as global warming. This study is the first comprehensive description of six HSP70 from the Antarctic obligate psychrophilic yeast, *G. antarctica* PI12, via comprehensive computational analysis. Unlike other studies which focus only on the gene expression of HSP70, much attention is lacking on the structures of the proteins which reflect their functions in cells. This is the first report we know of where all HSP70 structures in a single species have been constructed. The highly conserved structures of HSP70 make them great targets for studying protein flexibility and thermal adaptation. In this study, the relationship between HSP70 gene expression and protein structure flexibility is elucidated, which explains the cold adaptation ability in the Antarctic yeast *G. antarctica*. The high expression induction in gahsp-1 and gahsp-2 is strongly correlated with the high protein flexibility compared to the more rigid protein that has lower induction. Hence, this study provides a wealth of information regarding how Antarctic organisms adapt at the molecular level when facing adverse climates such as global warming.

Before analyzing *G. antarctica* HSP70 3D protein structures for temperature adaptation, they should be analyzed for their elemental temperature-related properties in their amino acid compositions, evolutionary relationships among organisms, and secondary structures. Analysis of signal peptides shows that all *G. antarctica* HSP70 contain no signal peptide (Table 1). This reflects the function of HSP70 as the pivotal component of the cellular network of molecular chaperones and folding catalysts (Mayer & Bukau, 2005). The GRAVY (grand average of hydropathy) of *G. antarctica* HSP70 reveals that all gahsp were non-polar except for gahsp-4 (Table 2). The interactions between non-polar residues and water produce a hydrophobic effect that promotes stability in protein at low-to-moderate temperatures (Ang *et al.,* 2021). It is interesting to note that gahsp70-4 presents as a polar protein, which may, in turn, confer higher thermostability compared to other gahsp70s, a physiochemical adaptive mechanism for stress adaptability (Panja *et al.,* 2020). Besides that, the physicochemical analysis of *G. antarctica* HSP70 indicates the presence of a high number of acidic, negatively charged residues (Asp and Glu) as compared to positively charged residues (Arg and Lys) (Table 3). This finding is parallel to a study on cold-adapted subtilisin which revealed that the negatively charged residues on the protein surface are responsible for their cold adaptation (Almog *et al.,* 2008). Similar to HSP70 from other organisms, the *G. antarctica* HSP70 contain no disulfide bridges and is located in the cell cytoplasmic (Table 4 & 5). The *G. antarctica* HSP70 showed that they share similarities between 24.71 to 61.95 (Table 6). Subsequent phylogenetic analysis showed that *G. antarctica* HSP70 have high sequence similarity to HSP70 in various organisms (Figure 1). All the *G. antarctica* HSP70 fall in different clades, suggesting gene duplication that may occur by forming paralogous genes such as in the nematode, mammalian, and molluscan (Hess *et al.,* 2018; Kourtidis *et al.,* 2006; Nikolaidis & Nei, 2004). Besides that, since HSP70 family members are often upregulated upon various stresses such as thermal stress, that disrupt protein folding, the *G. antarctica* HSP70 may evolve and form duplication from the multiple acquisitions of stress inducibility, such as those observed within various taxa (E. Yu *et al.*, 2021).

The flexibility of helices and strands is determined by the collective motion of many residues. The ability of proteins to undergo conformational changes for chemical modifications, ligand binding, and respond to environmental changes depends on the proteins' structural flexibility. In *G. antarctica* HSP70, the secondary structure is dominated by α-helix (Table 11). While alpha-helices are linked to structural flexibility due to their nonrigid bodies, β-sheets are associated with stability due to their sheet-like structures held together by hydrogen bonds (Emberly *et al.,* 2003; Marcos *et al.,* 2018). Meanwhile, extended strands help stabilize the original secondary structures and the resulting tertiary structures (Degrève *et al.,* 2014). Generally, random coil proteins are flexible compared with globular proteins (Alberts *et al.,* 2002). Compared to their mesophilic homologs, gahsp-1 and gahsp-2 have higher percentage values of α-helix. In addition, it is interesting to note that both gahsp-1 and gahsp-2 also have lesser amounts of random coil and β-turn. On the other hand, gahsp-3, gahsp-4, gahsp-5 and gahsp-6 have lesser percentage values of α-helix but more either in β-turns, random coils and extended strands. The distributions of helices and sheets determine the protein builds' rigidity and flexibility, whereby twists and bends are two dominant components that determine the structure's flexibility (Emberly *et al.,* 2004). Hence, the distribution of secondary structures in *G. antarctica* HSP70 must be composed of a justified proportion of helices, sheets, and coils to build proteins that are neither too rigid nor too flexible in achieving the final folds in low-energy environments.

Ionic interactions, attractive interactions that result between oppositely charged ions are known to be the strongest molecular interactions that occur in living systems. Protein stability is greatly determined and affected by ionic interactions. A study in malate dehydrogenase from various sources shows that salt bridges play an important role in structural stability at high temperatures,

pH and salt concentrations (ur Rehman *et al.*, 2022). Hence, salt bridges play an important role in thermo-stabilization, especially in hyper-thermophilic proteins (Jónsdóttir *et al.*, 2014). On the other hand, a reduced number of salt bridges is one of the contributors to a psychrophilic enzyme mannanase to adapt to low temperatures (Parvizpour *et al.*, 2017). Our analysis shows that all *G. antarctica* HSP70s have none or a lesser number of ionic pairs compared to their homologous counterparts. Compared to their mesophilic homologs, gahsp-1,-2,-3,-4 and -5 did not have any ion pairs detected in their buried residues. Consistent with the above findings on the role of ionic interactions in thermo-stabilization, the absence of ion pairs in *G. antarctica* HSP70s is important to ensure protein flexibility in a low-energy environment (Struvay & Feller, 2012). Notably, gahsp-6 has the same number of ion pairs as its homolog including the ion pair in its buried residues. This finding indicates that the ion pair in the buried residues in gahsp-6 may seem favorable, which forms across the hydrophobic core, facilitate water penetration, and hence maintain the required flexibility (Parvizpour *et al.*, 2017).

Aromatic rings produce non-covalent interactions that are highly stable due to the arrangement of the π-electrons which form a π-electron cloud. Aromatic rings in phenylalanine, tyrosine, tryptophan and histidine residues can greatly affect protein structure stability and are vital for molecular recognition processes (Lanzarotti *et al.*, 2020). In this study, we detailed the aromatic interactions; 1) Aromatic-aromatic or π-π interactions, 2) aromatic-sulphur and 3) cation-π. When two non-charged amino acids with aromatic side chains (Phe, Trp, Tyr) interacts, they create an aromatic-aromatic or π-π interaction which is known to be essential for protein structure stabilization and complexes (Makwana & Mahalakshmi, 2015). Here, we have carried out a detailed study to analyze the presence of buried aromatic interactions in HSP70 for the stability and flexibility of the protein structures. We report that *G. antarctica* HSP70s have lesser aromatic interactions in the buried residues when compared to their homologs. In RNA binding proteins, π–π interactions stabilize the core regions and determine the global conformational stability of these proteins (V Sivasakthi *et al.*, 2013). In small ubiquitin-like modifiers, the aromatic amino acids buried at the protein's core are vital for protein folding and stability (Chatterjee *et al.*, 2019). In the aromatic–aromatic interactions analysis, all *G. antarctica* HSP70s have lesser buried aromatic residues, except for gahsp-2 which has the same number as its homolog. As aromatic–aromatic interactions are vital for protein stability, the lesser number of this type of interaction could reflect the importance of protein flexibility in *G. antarctica* HSP70s.

Moreover, the interactions between sulfur-containing residues and aromatic residues in proteins are involved in stabilizing the tertiary structure of the protein or in structural changes (Gómez-Tamayo *et al.*, 2016). Our analysis shows that in the aromatic–sulfur interactions, only gahsp-2 has additional pair of aromatic interactions in the buried residues, gahsp-5 has the same number of aromatic interactions as its homolog, while others have lesser aromatic interactions compared to their homologs. A study on the occurrence of interactions involving Met and Cys side-chains in crystal structures of membrane proteins supported that the aromatic–sulfur interactions (Met–Met, Met–Phe, Met–Leu, and Cys–Phe), are stronger in magnitude than Phe–Phe interactions (Golledge *et al.*, 2019). The lesser number of aromatic–sulfur interactions in *G. antarctica* HSP70s may reflect the reduced stability in the protein structures which may contribute to the increase in flexibility. On the other hand, studies in psychrophilic isocitrate dehydrogenase (2UXR) reported the number of aromatic–sulfur interactions or pi–sulfur interactions is higher in psychrophilic isocitrate dehydrogenase than in mesophilic isocitrate dehydrogenase (D. Mitra & Das Mohapatra, 2021). This reflects the importance of aromatic–sulfur interactions which might play an important role in psychrophilic protein stabilization for extremely cold temperatures. Since gahsp-2

has additional pair of aromatic interactions in the buried residues compared to its mesophilic homolog, this also may reflect the importance of gahsp-2 in thermotolerance during heat stress.

As expected from the similarity with the aromatic-aromatic interactions, the Cation-π or His-aromatic or His-π interactions play a vital role in protein stabilization and contribute favorably to protein-protein binding and recognition (Dougherty, 2007). In a study of α-a cold-adapted amylase, an additional aromatic interaction introduced via mutagenesis demonstrated an increase of 50% in the protein stabilization energy (Gerday *et al.*, 2000). Our study shows that only gahsp-3 and gahsp-5 have lesser Cation-π interaction in the buried residues, while gahsp-1, gahsp-4 and gahsp-6 have more aromatic interaction and gahsp-2 has the same number of aromatic interactions when compared to their homologs. Similar to the aromatic-aromatic and aromatic–sulfur interactions, the Cation-π interaction contributes to protein stability; hence the increase of these aromatic interactions provides thermostability to protein structures to prevent protein misfolding and aggregation during high temperatures. In contrast, a study on a cold-adapted acetate kinase shows a reduced number of cation–π interactions in proteins from psychrophilic organisms to ensure sufficient flexibility (Tang *et al.*, 2014). A comparative study of cold-adapted endonucleases from *Aliivibrio salmonicida* shows that the cold-adapted enzyme has a lesser number of salt bridges, ionic and aromatic interactions in the intraprotein, which increase the protein flexibility (Michetti *et al.*, 2017). Similar features, such as few ionic and aromatic–aromatic interactions that contribute to altered dynamic properties, were also found in cold adapted lipases and esterases (Joseph *et al.*, 2008; Tutino *et al.*, 2009). The weakening of intramolecular interactions in gahsp-3 and gahsp-5 is consistent with the theory that protein flexibility assists activity in the cold. However, considering the presence of additional intramolecular forces in psychrophilic proteins, the proteins need to ensure local flexibility and rigidity interplay to ensure stability and activity (Pucci & Rooman, 2017). These findings are interesting as they are most likely to provide the importance of HSP70s during temperature shifts either from cold to hot or vice versa.

In our previous gene expression study of HSP70, gahsp1 and gahsp2 showed an increase in expression during hot and cold shock treatments which indicates the importance of Hsp70 in cell stress management in conditions of thermal stress (Yusof *et al.*, 2021). We may speculate that the high flexibility characteristics in gahsp1 reflect the importance of its high expression during thermal stress, while the increased stability in gahsp2 explains its lower induction when compared to gahsp1. The improved flexibility in gahsp-1 is possibly a strategy for cold-adapted proteins to maintain high catalytic activity at lower temperatures, which explains the need for its high expression at the gene level. On the other hand, gahsp-2 shows increase in its structural stability, which may explain the low induction in its gene expression during thermal shift. Our structure data may reflect the potential strong induction of gahsp-5 and the possibility of significant induction of gahsp-6 during thermal stress.

In summary, psychrophilic organisms, such as *Glaciozyma antarctica*, have evolved unique structural adaptations in their heat shock protein 70 (HSP70) proteins to survive in extremely cold environments (Yusof *et al.*, 2021). These structural differences contribute to the organism's ability to maintain protein homeostasis and protect cellular components from cold-induced damage. One of the observed structural differences in psychrophilic *G. antarctica* HSP70 proteins is the presence of residue substitutions in key domains involved in substrate binding and ATP hydrolysis. These substitutions affect the flexibility of loops that accommodate substrate binding and product release during the catalytic cycle. The increased flexibility of these loops in psychrophilic HSP70 proteins allows them to efficiently bind and fold proteins at low temperatures, where molecular motion is restricted (Manan *et al.*, 2019; Yusof *et al.*, 2019). This flexibility is crucial for the proper folding of

proteins in cold environments, as it enables the HSP70 proteins to adapt to the unique conformational requirements of their substrates (Yusof *et al.*, 2019).

Additionally, psychrophilic HSP70 proteins may exhibit changes in their membrane-interacting regions and hydrophobic core residues, which help maintain membrane fluidity and stability at low temperatures, allowing the organism to survive in extremely cold environments. The structural differences in psychrophilic HSP70 proteins also impact their functional properties. Comparative modelling studies have shown that psychrophilic HSP70 proteins, such as those from *G. antarctica*, exhibit higher expression levels and inducibility compared to their mesophilic counterparts (Yusof *et al.*, 2021). This increased expression of HSP70 proteins in response to thermal stress is crucial for the organism's survival in extremely cold environments, as it helps protect cellular proteins from denaturation and aggregation (Yusof *et al.*, 2021).

Furthermore, the structural adaptations in psychrophilic HSP70 proteins may also contribute to their enhanced stability and activity at low temperatures (Feller & Gerday, 1997). Although the crystallographic B-temperature factors of psychrophilic HSP70 proteins may not reveal significant differences compared to mesophilic homologues, the constraints imposed by crystal packing can obscure differences that occur in solution (Feller & Gerday, 1997). Therefore, it is likely that the observed structural differences in psychrophilic HSP70 proteins contribute to their enhanced stability and activity in the cold, allowing them to efficiently perform their chaperone functions and protect cellular proteins from cold-induced damage (Feller & Gerday, 1997). In conclusion, the observed structural differences in psychrophilic *G. antarctica* HSP70 proteins, including residue substitutions, changes in membrane-interacting regions, and modifications in hydrophobic core residues, contribute to the organism's stability.

One limitation of this study is that it primarily focuses on the structural and functional aspects of HSP70 proteins in *G. antarctica*. While these studies provide valuable information about the unique features of psychrophilic HSP70 proteins, they do not comprehensively address all potential adaptive mechanisms. Further research is needed to investigate other factors that may contribute to the survival of *G. antarctica* in extremely cold environments, such as the role of other chaperones, co-chaperones, or regulatory proteins. While the observed structural differences in psychrophilic *G. antarctica* HSP70 proteins provide valuable insights into their adaptive mechanisms, further research is needed to prove the adaptation in native environments and investigate additional avenues for understanding the comprehensive, adaptive mechanisms of these proteins in extremely cold environments.

CONCLUSION

To help better understand the functional biology of Hsp70 from *Glaciozyma antarctica* PI12, various computational tools were utilized to analyse the physicochemical properties, generate good quality homology models and annotate their functions based on structural properties. The question addressed in this study was whether HSP70s in *G. antarctica* PI12 have structural adaptation patterns that allow them to adapt during cold and heat stress. Our results provide the first comprehensive comparative structural analysis of all HSP70 presence in a single species, which in this case, a psychrophilic Antarctic yeast. The structural analysis of HSP70 from *G. antarctica* suggests several traits that allow thermal and cold adaptation ability in the Antarctic yeast *G. antarctica*. Since we have analysed the expression pattern of gahsp-1 and gahsp-2, we saw the pattern whereby the protein with high flexibility is strongly induced during thermal stress compared to the more rigid protein that has lower induction. Besides that, our computational investigations have demonstrated

that all six HSP70 protein structures in *G. antarctica* showed some similar and contrast patterns which probably are required to produce a balance correlation between local flexibility and rigidity. This intricate balance is crucial for protein folding, stability and activity (Isaksen *et al.*, 2016; Karshikoff *et al.*, 2015). Overall, these results will be useful in forecasting the suitability of HSP70 from *G. antarctica* as biomarkers to understand the thermal adaptation strategies amid adverse climates. Molecular analysis such as mutagenesis as well as quantitative measurements such as gene expression study that is literal to the current global warming scenario in Antarctica, is required to fully elucidate the molecular bases of the thermal stability and flexibility of psychrophilic proteins. The application of this study is crucial for future experimental design, including heat stress studies to support the in-silico findings.

ACKNOWLEDGEMENTS

We would like to thank University Malaysia Sabah for its strong support of research and innovation. We also would like to thank the research and innovation team at Biotechnology Research Institute for their administrative and technical support in this research and grant management.

REFERENCES

[1] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter P. 2002. *Molecular Biology of the Cell* (4th edition). New York: Garland Science.

[2] Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G. & Nielsen, H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, 37(4), 420–423.

[3] Almog, O., Kogan, A., Leeuw, M. de, Gdalevsky, G. Y., Cohen-Luria, R. & Parola, A. H. 2008. Structural insights into cold inactivation of tryptophanase and cold adaptation of subtilisin S41. *Biopolymers: Original Research on Biomolecules*, 89(5), 354–359.

[4] Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389–3402.

[5] Ang, D. L., Hoque, M. Z., Hossain, M. A., Guerriero, G., Berni, R., Hausman, J.-F., Bokhari, S. A., Bridge, W. J. & Siddiqui, K. S. 2021. Computational analysis of thermal adaptation in extremophilic chitinases: the achilles' heel in protein structure and industrial utilization. *Molecules*, 26(3), 707.

[6] Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 22(2), 195–201.

[7] Asir, F., Korak, T. & Öztürk, Ö. 2021. Three Msa Tools Analysis In Dna and Protein Datasets. *Middle East Journal of Science*, 2(7), 89-99.

[8] Bae, E. & Phillips, G. N. 2004. Structures and analysis of highly homologous psychrophilic, mesophilic, and thermophilic adenylate kinases. *Journal of Biological Chemistry*, 279(27), 28202–28208.

[9] Bakhos-Douaihy, D., Seaayfan, E., Demaretz, S., Komhoff, M. & Laghmani, K. 2021. Differential effects of STCH and stress-inducible Hsp70 on the stability and maturation of NKCC2. *International Journal of Molecular Sciences*, 22(4), 2207.

[10] Batabyal, D., Lord, H., Ahlstrom, B. & Wikström, M. 2021. Determination of the experimental extinction coefficient of therapeutic proteins using the Edelhoch method. *Biologicals*, 71, 42–47.

[11] Boo, S. Y., Wong, C. M. V. L., Rodrigues, K. F., Najimudin, N., Murad, A. M. A. & Mahadi, N. M. 2013. Thermal stress responses in Antarctic yeast, Glaciozyma antarctica PI12, characterized by real-time quantitative PCR. *Polar Biology*, 36(3), 381–389.

[12] Briggs, P. 2000. *CCP4 Newsletter on Protein Crystallography*. Number 38, April 2000. Daresbury Laboratory, Daresbury, Warrington, WA4 4AD, UK

[13] Bryson, K., McGuffin, L. J., Marsden, R. L., Ward, J. M., Sodhi, J. & Jones, D. T. 2005. Protein Structure Prediction Servers At University College London. *Nucleic Acids Research*, 33(Web Server issue), W36-W38.

[14] Chand, K., Iyer, K. & Mitra, D. 2021. Comparative analysis of differential gene expression of HSP40 and HSP70 family isoforms during heat stress and HIV-1 infection in T-cells. *Cell Stress and Chaperones*, 26, 403–416.

[15] Chatterjee, K. S., Tripathi, V. & Das, R. 2019. A conserved and buried edge-to-face aromatic interaction in small ubiquitin-like modifier (SUMO) has a role in SUMO stability and function. *Journal of Biological Chemistry*, 294(17), 6772–6784.

[16] Cho, M.-J. & Park, S.-Y. (2019). Carbon-dot-based ratiometric fluorescence glucose biosensor. *Sensors and Actuators B: Chemical*, 282, 719–729.

[17] Chown, S. L., Leihy, R. I., Naish, T. R., Brooks, C. M., Convey, P., Henley, B. J., Mackintosh, A. N., Phillips, L. M., Kennicutt II, M. C. & Grant, S. M. 2022. *Antarctic climate change and the environment: A decadal synopsis and recommendations for action*. Scientific Committee on Antarctic Research, Cambridge, United Kingdom. www.scar.org.

[18] Combet, C., Blanchet, C., Geourjon, C. & Deleage, G. 2000. NPS@: network protein sequence analysis. *Trends in Biochemical Sciences*, 25(3), 147–150.

[19] Convey, P. & Peck, L. S. 2019. Antarctic environmental change and biological responses. *Science Advances*, 5(11), eaaz0888.

[20] Creighton, T. E. 1988. Disulphide bonds and protein stability. *BioEssays*, 8(2-3), 57–63.

[21] Daugaard, M., Rohde, M. & Jäättelä, M. 2007. The heat shock protein 70 family: Highly homologous proteins with overlapping and distinct functions. *FEBS Letters*, 581(19), 3702–3710.

[22] Day, J., Passecker, A., Beck, H.-P. & Vakonakis, I. 2019. The Plasmodium falciparum Hsp70-x chaperone assists the heat stress response of the malaria parasite. *The FASEB Journal*, 33(12), 14611.

[23] Degrève, L., Fuzo, C. A. & Caliri, A. 2014. Extended secondary structures in proteins. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1844(2), 384–388.

[24] Dong, Y., Liao, M., Meng, X. & Somero, G. N. (2018). Structural flexibility and protein adaptation to temperature: Molecular dynamics analysis of malate dehydrogenases of marine molluscs. *Proceedings of the National Academy of Sciences*, 115(6), 1274–1279.

[25] Dougherty, D. A. 2007. Cation-π interactions involving aromatic amino acids. *The Journal of Nutrition*, 137(6), 1504S-1508S.

[26] Emberly, E. G., Mukhopadhyay, R., Tang, C. & Wingreen, N. S. 2004. Flexibility of β-sheets: principal component analysis of database protein structures. *Proteins: Structure, Function, and Bioinformatics*, 55(1), 91–98.

[27] Emberly, E. G., Mukhopadhyay, R., Wingreen, N. S. & Tang, C. 2003. Flexibility of α-helices: Results of a statistical analysis of database protein structures. *Journal of Molecular Biology*, 327(1), 229–237.

[28] Firdaus-Raih, M., Hashim, N. H. F., Bharudin, I., Abu Bakar, M. F., Huang, K. K., Alias, H., Lee, B. K. B., Mat Isa, M. N., Mat-Sharani, S., Sulaiman, S., Tay, L. J., Zolkefli, R., Muhammad Noor, Y., Law, D. S. N., Abdul Rahman, S. H., Md-Illias, R., Abu Bakar, F. D., Najimudin, N., Abdul Murad, A. M. & Mahadi, N. M. 2018. The Glaciozyma antarctica genome reveals an

array of systems that provide sustained responses towards temperature variations in a persistently cold habitat. *PLoS One*, 13(1), e0189947.

[29] Flaherty, K. M., DeLuca-Flaherty, C. & McKay, D. B. 1990. Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. *Nature*, 346(6285), 623–628.

[30] García-Descalzo, L., García-López, E., Alcázar, A., Baquero, F. & Cid, C. 2014. Proteomic analysis of the adaptation to warming in the Antarctic bacteria Shewanella frigidimarina. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1844(12), 2229–2240.

[31] Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. & Bairoch, A.(2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 31(13), 3784–3788.

[32] Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D. & Bairoch, A. 2005. Protein identification and analysis tools on the ExPASy server. *In*: Walker, J. M. (ed). *The Proteomics Protocols Handbook.* Springer Protocols Handbooks. Humana Press. https://doi.org/10.1385/1-59259-890-0:571. pp. 571–607.

[33] Gerday, C., Aittaleb, M., Bentahir, M., Chessa, J. P., Claverie, P., Collins, T., D'Amico, S., Dumont, J., Garsoux, G., Georlette, D., Hoyoux, A., Lonhienne, T, Meuwis, M. A. & Feller, G. 2000. 2000-Cold-adapted enzymes from fundamentals-Gerday. *Trends in Biotechnology*, 18(3), 103–107.

[34] Golledge, N. R., Keller, E. D., Gomez, N., Naughten, K. A., Bernales, J., Trusel, L. D. & Edwards, T. L. 2019. Global environmental consequences of twenty-first-century ice-sheet melt. *Nature*, 566(7742), 65–72.

[35] Gómez-Tamayo, J. C., Cordomí, A., Olivella, M., Mayol, E., Fourmy, D. & Pardo, L. 2016. Analysis of the interactions of sulfur-containing amino acids in membrane proteins. *Protein Science*, 25(8), 1517–1524.

[36] Gong, W. J. & Golic, K. G. 2006. Loss of Hsp70 in Drosophila is pleiotropic, with effects on thermotolerance, recovery from heat shock and neurodegeneration. *Genetics*, 172(1), 275–286.

[37] Gutt, J., Isla, E., Xavier, J. C., Adams, B. J., Ahn, I.-Young, Cheng, Colesie, C., Cummings, V. J., di Prisco, G., Griffiths, H., Hawes, I., Hogg, I., McIntyre, T., Meiners, K. M., Pearce, D. A., Peck, L., Piepenburg, D., Reisinger, R. R., Saba, G. K., Schloss, I. R.,Signori, C. N., Smith, C. R., Vacchi, M., Verde, C. & Wall, D. H. 2021. Antarctic ecosystems in transition–life between stresses and opportunities. *Biological Reviews*, 96(3), 798–821.

[38] Hassan, F., Nawaz, A., Rehman, M. S., Ali, M. A., Dilshad, S. M. R. & Yang, C. 2019. Prospects of HSP70 as a genetic marker for thermo-tolerance and immuno-modulation in animals under climate change scenario. *Animal Nutrition*, 5(4), 340–350.

[39] Hess, K., Oliverio, R., Nguyen, P. V., Le, D., Ellis, J., Kdeiss, B., Ord, S., Chalkia, D. & Nikolaidis, N. 2018. Concurrent action of purifying selection and gene conversion results in extreme conservation of the major stress-inducible Hsp70 genes in mammals. *Scientific Reports*, 8, 5082.

[40] Hirokawa, T., Boon-Chieng, S. & Mitaku, S. 1998. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, 14(4), 378–379.

[41] Horwich, A. L. & Fenton, W. A. 2020. Chaperonin-assisted protein folding: A chronologue. *Quarterly Reviews of Biophysics*, 53, e4.

[42] Hosseinian, S. H., Haddad-Mashadrizeh, A. & Dolatabadi, S. 2018. Simulation and Stability Assessment Of Anti-epcam Immunotoxin For Cancer Therapy. *Advanced Pharmaceutical Bulletin*, 8(3), 447-455.

[43] Ikai, A. 1980. Thermostability and aliphatic index of globular proteins. *The Journal of Biochemistry*, 88(6), 1895–1898.

[44] Isaksen, G. V., Åqvist, J. & Brandsdal, B. O. 2016. Enzyme surface rigidity tunes the temperature dependence of catalytic rates. *Proceedings of the National Academy of Sciences*, 113(28), 7822–7827.

[45] Jazi, M. A., Seyedi, S. M., Ebrahimie, E., Ebrahimi, M., Moro, G. D &, Botanga, C. J. 2017. A Genome-wide Transcriptome Map Of Pistachio (Pistacia Vera L.) Provides Novel Insights Into Salinity-related Genes and Marker Discovery. *BMC Genomics*, 18, Article number 627(2017).

[46] Jin, J., Zhao, M., Wang, Y., Zhou, Z., Wan, F. & Guo, J. 2020. Induced thermotolerance and expression of three key Hsp genes (Hsp70, Hsp21, and sHsp21) and their roles in the high temperature tolerance of Agasicles hygrophila. *Frontiers in Physiology*, 10, 1593.

[47] Jónsdóttir, L. B., Ellertsson, B. Ö., Invernizzi, G., Magnúsdóttir, M., Thorbjarnardóttir, S. H., Papaleo, E. & Kristjánsson, M. M. 2014. The role of salt bridges on the temperature adaptation of aqualysin I, a thermostable subtilisin-like proteinase. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1844(12), 2174–2181.

[48] Joseph, B., Ramteke, P.W. & Thomas, G. 2008. Cold active microbial lipases: some hot issues and recent developments. *Biotechnology Advances*, 26(5), 457–470.

[49] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A. and Bridgland, A. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589.

[50] Kabani, M. & Martineau, C. N. 2008. Multiple hsp70 isoforms in the eukaryotic cytosol: mere redundancy or functional specificity? *Current Genomics*, 9(5), 338–348.

[51] Karshikoff, A., Nilsson, L. & Ladenstein, R. 2015. Rigidity versus flexibility: the dilemma of understanding protein thermal stability. *The FEBS Journal*, 282(20), 3899–3917.

[52] Karunanayake, C. & Page, R. C. 2021. Cytosolic protein quality control machinery: Interactions of Hsp70 with a network of co-chaperones and substrates. *Experimental Biology and Medicine*, 246(12), 1419–1434.

[53] Kityk, R., Kopp, J., Sinning, I. & Mayer, M. P. 2012. Structure and dynamics of the ATP-bound open conformation of Hsp70 chaperones. *Molecular Cell*, 48(6), 863–874.

[54] Kourtidis, A., Drosopoulou, E., Nikolaidis, N., Hatzi, V. I., Chintiroglou, C. C. & Scouras, Z. G. 2006. Identification of several cytoplasmic HSP70 genes from the Mediterranean mussel (Mytilus galloprovincialis) and their long-term evolution in Mollusca and Metazoa. *Journal of Molecular Evolution*, 62, 446–459.

[55] Kuennen, M., Gillum, T., Dokladny, K., Bedrick, E., Schneider, S. & Moseley, P. 2011. Thermotolerance and heat acclimation may share a common mechanism in humans. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 301(2), R524–R533.

[56] Kurashova, N. A., Madaeva, I. M. & Kolesnikova, L. I. 2020. Expression of HSP70 heat-shock proteins under oxidative stress. *Advances in Gerontology*, 10, 20–25.

[57] Lanzarotti, E., Biekofsky, R. R., Estrin, D. A., Marti, M. A. & Turjanski, A. G. 2011. Aromatic–aromatic interactions in proteins: beyond the dimer. *Journal of Chemical Information and Modeling*, 51(7), 1623–1633.

[58] Lanzarotti, E., Defelipe, L. A., Marti, M. A. & Turjanski, A. G. 2020. Aromatic clusters in protein–protein and protein–drug complexes. *Journal of Cheminformatics*, 12, Article number 30(2020).

[59] Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), 283–291.

[60] Li, D., Zaman, W., Lu, J., Niu, Q., Zhang, X., Ayaz, A., Saqib, S., Yang, B., Zhang, J., Zhao, H. & Lü, S. 2023. Natural lupeol level variation among castor accessions and the upregulation of lupeol synthesis in response to light. *Industrial Crops and Products*, 192, 116090.

[61] Makwana, K. M. & Mahalakshmi, R. 2015. Implications of aromatic–aromatic interactions: From protein structures to peptide models. *Protein Science*, 24(12), 1920–1933.

[62] Manan, F. A. A., Hong, W. W., Abdullah, J., Yusof, N. A. & Ahmad, I. 2019. Nanocrystalline cellulose decorated quantum dots based tyrosinase biosensor for phenol determination. *Materials Science and Engineering: C*, 99, 37–46.

[63] Mao, R., Zhang, H., Bie, L., Liu, L. & Gao, J. 2023. Million-atom Molecular Dynamics Simulations Reveal the Interfacial Interactions And Assembly Of Plant Psii-lhcii Supercomplex. *RSC Advances*, 10(13), 6699-6712.

[64] Marcos, E. S., Chidyausiku, T. M., McShan, A. C., Evangelidis, T., Nerli, S., Carter, L. M., Nivón, L. G., Davis, A., Oberdorfer, G., Tripsianes, K., Sgourakis, N. G. & Baker, D. 2018. De novo design of a non-local β-sheet protein with high stability and accuracy. *Nature Structural & Molecular Biology*, 25(11), 1028–1034.

[65] Mayer, M. P. & Bukau, B. 2005. Hsp70 chaperones: cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences*, 62(6), 670.

[66] Mayer, M. P. 2018. Intra-molecular pathways of allosteric control in Hsp70s. *Philosophical Transactions of the Royal Society B*, 373(1749), 20170183.

[67] McBride, M. C., Stokke, O. S., Renner, A. H. H., Krafft, B. A., Bergstad, O. A., Biuw, M., Lowther, A. D. & Stiansen, J. E. 2021. Antarctic krill Euphausia superba: spatial distribution, abundance, and management of fisheries in a changing climate. *Marine Ecology Progress Series*, 668, 185–214.

[68] McGaughran, A., Laver, R. & Fraser, C. 2021. Evolutionary responses to warming. *Trends in Ecology & Evolution*, 36(7), 591–600.

[69] McGuffin, L. J., Bryson, K. & Jones, D. T. 2000. The Psipred Protein Structure Prediction Server. *Bioinformatics*, 4(16), 404-405.

[70] Michetti, D., Brandsdal, B. O., Bon, D., Isaksen, G. V., Tiberti, M. & Papaleo, E. 2017. A comparative study of cold- and warm-adapted Endonucleases A using sequence analyses and molecular dynamics simulations. *PLoS One*, 12(2), e0169586.

[71] Mitra, D. & Das Mohapatra, P. K. 2021. Cold adaptation strategy of psychrophilic bacteria: an in-silico analysis of isocitrate dehydrogenase. *Systems Microbiology and Biomanufacturing*, 1(4), 483–493.

[72] Mitra, P., Deshmukh, A. S. & Choudhury, C. 2021. Molecular chaperone function of stress inducible Hsp70 is critical for intracellular multiplication of Toxoplasma gondii. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1868(2), 118898.

[73] Moss, S. M., Taylor, I. R., Ruggero, D., Gestwicki, J. E., Shokat, K. M. & Mukherjee, S. 2019. A Legionella pneumophila kinase phosphorylates the Hsp70 chaperone family to inhibit eukaryotic protein synthesis. *Cell Host & Microbe*, 25(3), 454–462.

[74] Munson, M., Balasubramanian, S., Fleming, K. G., Nagi, A. D., O'Brien, R., Sturtevant, J. M. & Regan, L. 1996. What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. *Protein Science*, 5(8), 1584–1593.

[75] Nachmias, B., Lazar, I., Elmalech, M., Abed-El-Rahaman, I., Asshab, Y., Mandelboim, O., Perlman, R. & Ben-Yehuda, D. 2007. Subcellular localization determines the delicate balance between the anti-and pro-apoptotic activity of Livin. *Apoptosis*, 12, 1129–1142.

[76] Nava, R. & Zuhl, M. N. 2020. Heat acclimation-induced intracellular HSP70 in humans: a meta-analysis. *Cell Stress and Chaperones*, 25(1), 35–45.

[77] Nikolaidis, N. & Nei, M. 2004. Concerted and nonconcerted evolution of the Hsp70 gene superfamily in two sibling species of nematodes. *Molecular Biology and Evolution*, 21(3), 498–505.

[78] Odunuga, O. O., Bollinger, S. A., Choi, K. H. & Polvadore, E. I. 2012. Caenorhabditis elegans Hsp70-1 expresses highly in bacteria, is sufficiently soluble, and has a catalytic constant similar to Hsc70 and BiP. *Protein Expression and Purification*, 82(1), 132–137.

[79] Panja, A. S., Maiti, S. & Bandyopadhyay, B. 2020. Protein stability governed by its structural plasticity is inferred by physicochemical factors and salt bridges. *Scientific Reports*, 10(1), 1822.

[80] Parvizpour, S., Razmara, J., Shamsir, M. S., Illias, R. M. & Abdul Murad, A. M. 2017. The role of alternative salt bridges in cold adaptation of a novel psychrophilic laminarinase. *Journal of Biomolecular Structure and Dynamics*, 35(8), 1685–1692.

[81] Pucci, F. & Rooman, M. 2017. Physical and molecular bases of protein thermal stability and cold adaptation. *Current Opinion in Structural Biology*, 42, 117–128.

[82] Rathor, P. K., Borza, T., Bahmani, R., Stone, S. L., Tonon, T., Yurgel, S., Potin, P. & Prithiviraj, B. 2023. Expression of a Heat Shock Protein 70 from the Brown Alga Ectocarpus sp. Imparts Salinity Stress Tolerance in Arabidopsis thaliana. *Journal of Applied Phycology*, 35(2), 803–819.

[83] Rosenzweig, R., Nillegoda, N. B., Mayer, M. P. & Bukau, B. 2019. The Hsp70 chaperone network. *Nature Reviews Molecular Cell Biology*, 20(11), 665–680.

[84] Rout, P. K., Kaushik, R. & Ramachandran, N. 2016. Differential expression pattern of heat shock protein 70 gene in tissues and heat stress phenotypes in goats during peak heat stress period. *Cell Stress and Chaperones*, 21(4), 645–651.

[85] Rüdiger, S., Germeroth, L., Schneider-Mergener, J. & Bukau, B. 1997. Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *The EMBO Journal*, 16(7), 1501–1507.

[86] Sadiqi, S., Hamza, M., Ali, F., Alam, S., Shakeela, Q., Ahmed, S., Ayaz, A., Ali, S., Saqib, S., Ullah, F. & Zaman, W. 2022. Molecular characterization of bacterial isolates from soil samples and evaluation of their antibacterial potential against MDRS. *Molecules*, 27(19), 6281.

[87] Schauperl, M. & Denny, R. A. 2022. AI-Based protein structure prediction in drug discovery: Impacts and challenges. *Journal of Chemical Information and Modeling*, 62(13), 3142–3156.

[88] Silver, J. T. & Noble, E. G. 2012. Regulation of survival gene hsp70. *Cell Stress and Chaperones*, 17(1), 1–9.

[89] Sivasakthi, V, Anbarasu, A. & Ramaiah, S. 2013. π–π Interactions in Structural Stability: Role in RNA Binding Proteins. *Cell Biochemistry and Biophysics*, 67, 853–863.

[90] Sivasakthi, V., Anitha, P., Kumar, K. M., Bag, S., Senthilvel, P., Lavanya, P., Swetha, R., Anbarasu, A. & Ramaiah, S. 2013. Aromatic-aromatic interactions: analysis of π-π interactions in interleukins and TNF proteins. *Bioinformation*, 9(8), 432 - 439.

[91] Struvay, C. & Feller, G. 2012. Optimization to low temperature activity in psychrophilic enzymes. *International Journal of Molecular Sciences*, 13(9), 11643–11665.

[92] Tamura, K., Stecher, G. & Kumar, S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, 38(7), 3022–3027.

[93] Tang, M. A. K., Motoshima, H. & Watanabe, K. 2014. Cold adaptation: structural and functional characterizations of psychrophilic and mesophilic acetate kinase. *The Protein Journal*, 33, 313–322.

[94] Tasneem, M., Gupta, S. D., Momin, M. B., Hossain, K. M., Osman, T. B. & Rabbi, M. F. 2023. In Silico Annotation Of a Hypothetical Protein From Listeria Monocytogenes Egd-e Unfolds A Toxin Protein Of The Type II Secretion System. *Genomics Inform*, 1(21), e7.

[95] Teilum, K., Olsen, J. G. & Kragelund, B. B. 2011. Protein stability, flexibility and function. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1814(8), 969–976.

[96] Tina, K. G., Bhadra, R. & Srinivasan, N. 2007. PIC: protein interactions calculator. *Nucleic Acids Research*, 35(suppl_2), W473–W476.

[97] Tutino, M. L., Parrilli, E., De Santi, C., Giuliani, M., Marino, G. & de Pascale, D. 2010. Cold-adapted esterases and lipases: A biodiversity still under-exploited. *Current Chemical Biology*, 4(1), 74–83.

[98] ur Rehman, M. F., Shaeer, A., Batool, A. I. & Aslam, M. 2022. Structure-function relationship of extremozymes. *In*: Kuddus, M. (Ed). *Microbial Extremozymes.* London: Academic Press (pp. 9–30).

[99] Williams, T. J., Liao, Y., Ye, J., Kuchel, R. P., Poljak, A., Raftery, M. J. & Cavicchioli, R. 2017. Cold adaptation of the Antarctic haloarchaea Halohasta litchfieldiae and Halorubrum lacusprofundi. *Environmental Microbiology*, 19(6), 2210–2227.

[100] Yadav, M., Gosai, H. G., Singh, G., Singh, A., Singh, A. K., Singh, R. P. & Jadeja, R. N. 2023. Major Impact of Global Climate Change in Atmospheric, Hydrospheric and Lithospheric Context. *In*: In: Singh, P., Ao, B. & Yadav, A. (eds). *Global Climate Change and Environmental Refugees.* Springer. (pp. 35–55)

[101] Yang, M., Derbyshire, M. K., Yamashita, R. A, & Marchler-Bauer, A. 2020. NCBI's conserved domain database and tools for protein domain analysis. *Current Protocols in Bioinformatics*, 69(1), e90.

[102] Yu, C.-S., Cheng, C.-W., Su, W.-C., Chang, K.-C., Huang, S.-W., Hwang, J.-K., & Lu, C.-H. 2014. CELLO2GO: a web server for protein subCELlular LOcalization prediction with functional gene ontology annotation. *PloS One*, 9(6), e99368.

[103] Yu, E., Yoshinaga, T., Jalufka, F. L., Ehsan, H., Mark Welch, D. B. & Kaneko, G. 2021. The complex evolution of the metazoan HSP70 gene family. *Scientific Reports*, 11(1), 17794.

[104] Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S. C., Ester, M., Foster, L. J. & Brinkman, F. S. L. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, 26(13), 1608–1615.

[105] Yusof, N. A., Charles, J., Wan Mahadi, W. N. S., Abdul Murad, A. M. & Mahadi, N. M. 2021. Characterization of inducible HSP70 genes in an Antarctic yeast, Glaciozyma antarctica PI12, in response to thermal stress. *Microorganisms*, 9(10), 2069.

[106] Yusof, N. A., Kamaruddin, S., Abu Bakar, F. D., Mahadi, N. M. & Abdul Murad, A. M. 2019. Structural and functional insights into TRiC chaperonin from a psychrophilic yeast, Glaciozyma antarctica. *Cell Stress and Chaperones*, 24(2), 351–368.

[107] Yusof, N. A., Masnoddin, M., Charles, J., Thien, Y. Q., Nasib, F. N., Wong, C. M. V. L., Murad, A. M. A., Mahadi, N. M. & Bharudin, I. 2022. Can heat shock protein 70 (HSP70) serve as biomarkers in Antarctica for future ocean acidification, warming and salinity stress? *Polar Biology*, 45(3), 371–394.

[108] Zaman, W., Ye, J., Ahmad, M., Saqib, S., Shinwari, Z. K., & Chen, Z. 2022. Phylogenetic exploration of traditional Chinese medicinal plants: A case study on Lamiaceae. *Pakistan Journal of Botany*, 54(3), 1033-1040

[109] Zhao, J., Zhang, S., Jiang, Y., Liu, Y. & Zhu, Q. 2023. Mutation Analysis Of Pathogenic Non-synonymous Single Nucleotide Polymorphisms (Nssnps) In Wfs1 Gene Through Computational Approaches. *Scientific Reports*, 13, Article number 6774.

[110] Zuiderweg, E. R. P., Hightower, L. E. & Gestwicki, J. E. 2017. The remarkable multivalency of the Hsp70 chaperones. *Cell Stress and Chaperones*, 22, 173–189.