

Journal of Computational Biology and Bioinformatics Research

Volume 7 Number 1, May 2015

ISSN 2141-2227



*Academic
Journals*

ABOUT JCBBR

The **Journal of Computational Biology and Bioinformatics Research (JCBBR)** (ISSN 2141-2227) is published Monthly (one volume per year) by AcademicJournals.

Journal of Computational Biology and Bioinformatics Research (JCBBR), is a peer reviewed journal. The journal is published per article and covers all areas of the subject such as: Genome annotation, Comparative genomics, Analysis of gene expression, Structural Bioinformatics Approaches.

Submission of Manuscript

Please read the **Instructions for Authors** before submitting your manuscript. The manuscript files should be given the last name of the first author

[Click here to Submit manuscripts online](#)

If you have any difficulty using the online submission system, kindly submit via this email jcbbbr@academicjournals.org.

With questions or concerns, please contact the Editorial Office at jcbbbr@academicjournals.org.

Editor

Dr. Hong Bo Guo

*Northwest A & F University
Yangling,
China.*

Nicholas E. Korres

*Environmental Research Institute
Dept. of Civil and Environmental Engineering
University College Cork
Ireland.*

Dr. Devesh Kumar

*Babasaheb Bhimrao Ambedkar University,
Lucknow,
India.*

Dr. K. Thangavel

*Periyah University
India.*

Prof. Paul S. Marley

*Department of Crop Protection
Faculty of Agriculture
Institute for Agricultural Research
Ahmadu Bello University
Zaria,
Nigeria.*

Prof. Lenore M. Martin

*Department of Cell and Molecular Biology
University of Rhode Island
Kingston, RI
USA.*

Editorial Board

Dr. Eduardo Alberto Castro

*INIFTA (CONICET – UNLP)
Buenos Aires,
Argentina.*

Dr. Jayanthi Manicassamy

*Pondicherry University
India.*

Prof. Gang Yang

*Northeast Forestry University & Dalian Institute of
Chemical Physics
Harbin,
China.*

Jaiswal Nivedita

*Department of Biochemistry
Faculty of Science
Banaras Hindu University
Varanasi,
India*

Prof. Songjun Li

*Key Laboratory of Pesticide & Chemical Biology
Ministry of Education
College of Chemistry
Central China Normal University
Wuhan,
China.*

Dr. Dimitrius Leonardo Pitol

*University of São Paulo
Cell Biology and Histology
Brazil.*

Khubaib Ahmed Qureshi

*Hamdard University
Karachi,
Pakistan.*

Hamid Jafarzadeh

*Mashhad Faculty of Dentistry
Dental Research Center
Iran.*

Ranjit Prasad Bahadur

*Indian Institute of Technology
Kharagpur,
India.*

Alok Shekhar

*Yeshwant College of IT, BI & BT
Department of Bioinformatics
India.*

Lalit Ponnala

*Cornell University
Ithaca, NY
USA.*

Oliviero Carugo

*University of Pavia
Italy.*

Andrea Masotti

*Gene Expression - Microarrays Laboratory
Children's Hospital Bambino Gesù.
Italy*

Dr. Richard Emes

*University of Nottingham
UK.*

Oswaldo Andrade Santos-Filho

*Fundação Oswaldo Cruz
Rio de Janeiro - RJ
Brazil.*

Qi Dai

*Institute of Biomedical Engineering and
Instrumentation
Hangzhou Dianzi University
China.*

Dr. Asad U. Khan

*Interdisciplinary Biotechnology Unit
Aligarh Muslim University (AMU)
Aligarh,
India.*

Osman Radwan

*University of Illinois
Urbana-Champaign
USA.*

Anthony Hunt

*Department of Bioengineering and Therapeutic
Sciences
University of California
San Francisco,
USA.*

Dr. Issam El Naqa

*Washington University
St. Louis School of Medicine
USA.*

Dae-Ki Kang

*Dongseo University
South Korea.*

Dr. Xiao-Qin Xia

*Vaccine Research Institute of San Diego
USA.*

Simone Linz

*Department of Computer Science
University of California
Davis,
USA.*

Dr. Ivan Y. Torshin

*Computational Center of the Russian Academy of
Sciences
Russia.*

Dr. Daniele Dell'Orco

*Car von Ossietzky University of Oldenburg
Department of Biology and Environmental Sciences,
Biochemistry of Neurosensory Processes
Germany*

Zindoga Mukandavire

*National University of Science and Technology
Bulawayo,
Zimbabwe.*

Dr. Ahmed Fazary

*School of Pharmacy
College of Medicine
National Taiwan University
Taipei,
Taiwan.*

Dr. Prem Raj B. Joseph

*Department of Biochemistry and Molecular
Biophysics
Sealy Center for Structural Biology and Molecular
Biophysics
University of Texas Medical
Galveston, TX
USA.*

Dr. Hamid Alinejad-Rokny

*School of Electrical Engineering and Computer
Science
The University of Newcastle
NSW,
Australia.*

Dr. Peng Che

*Institute of Intelligent Machines
Chinese Academy of Sciences
Hefei,
China.*

Dr. Yupeng Wang

*Cornell University
628 Rhodes Hall
Ithaca, NY
USA.*

Dr. Vishal N. Koparde

*Bioinformatics Computational Core Laboratories
Center for Study of Biological Complexity
Virginia Commonwealth University
Richmond, VA
USA.*

Dr. Mai S. Mabrouk

*Department of Biomedical Engineering
Misr University for Science and Technology
Cairo University
Egypt.*

Dr. Marius Mihasan

*Alexandru Ioan Cuza University
Faculty of Biology
Biochemistry and Molecular Biology Laboratory
Iasi,
Romania.*

Instructions for Author

Electronic submission of manuscripts is strongly encouraged, provided that the text, tables, and figures are included in a single Microsoft Word file (preferably in Arial font).

The **cover letter** should include the corresponding author's full address and telephone/fax numbers and should be in an e-mail message sent to the Editor, with the file, whose name should begin with the first author's surname, as an attachment.

Article Types

Three types of manuscripts may be submitted:

Regular articles: These should describe new and carefully confirmed findings, and experimental procedures should be given in sufficient detail for others to verify the work. The length of a full paper should be the minimum required to describe and interpret the work clearly.

Short Communications: A Short Communication is suitable for recording the results of complete small investigations or giving details of new models or hypotheses, innovative methods, techniques or apparatus. The style of main sections need not conform to that of full-length papers. Short communications are 2 to 4 printed pages (about 6 to 12 manuscript pages) in length.

Reviews: Submissions of reviews and perspectives covering topics of current interest are welcome and encouraged. Reviews should be concise and no longer than 4-6 printed pages (about 12 to 18 manuscript pages). Reviews are also peer-reviewed.

Review Process

All manuscripts are reviewed by an editor and members of the Editorial Board or qualified outside reviewers. Authors cannot nominate reviewers. Only reviewers randomly selected from our database with specialization in the subject area will be contacted to evaluate the manuscripts. The process will be blind review.

Decisions will be made as rapidly as possible, and the journal strives to return reviewers' comments to authors as fast as possible. The editorial board will re-review manuscripts that are accepted pending revision. It is the goal of the AJFS to publish manuscripts within weeks after submission.

Regular articles

All portions of the manuscript must be typed double-spaced and all pages numbered starting from the title page.

The Title should be a brief phrase describing the contents of the paper. The Title Page should include the authors' full names and affiliations, the name of the corresponding author along with phone, fax and E-mail information. Present addresses of authors should appear as a footnote.

The Abstract should be informative and completely self-explanatory, briefly present the topic, state the scope of the experiments, indicate significant data, and point out major findings and conclusions. The Abstract should be 100 to 200 words in length. Complete sentences, active verbs, and the third person should be used, and the abstract should be written in the past tense. Standard nomenclature should be used and abbreviations should be avoided. No literature should be cited.

Following the abstract, about 3 to 10 key words that will provide indexing references should be listed.

A list of non-standard **Abbreviations** should be added. In general, non-standard abbreviations should be used only when the full term is very long and used often. Each abbreviation should be spelled out and introduced in parentheses the first time it is used in the text. Only recommended SI units should be used. Authors should use the solidus presentation (mg/ml). Standard abbreviations (such as ATP and DNA) need not be defined.

The Introduction should provide a clear statement of the problem, the relevant literature on the subject, and the proposed approach or solution. It should be understandable to colleagues from a broad range of scientific disciplines.

Materials and methods should be complete enough to allow experiments to be reproduced. However, only truly new procedures should be described in detail; previously published procedures should be cited, and important modifications of published procedures should be mentioned briefly. Capitalize trade names and include the manufacturer's name and address. Subheadings should be used. Methods in general use need not be described in detail.

Results should be presented with clarity and precision. The results should be written in the past tense when describing findings in the authors' experiments. Previously published findings should be written in the present tense. Results should be explained, but largely without referring to the literature. Discussion, speculation and detailed interpretation of data should not be included in the Results but should be put into the Discussion section.

The Discussion should interpret the findings in view of the results obtained in this and in past studies on this topic. State the conclusions in a few sentences at the end of the paper. The Results and Discussion sections can include subheadings, and when appropriate, both sections can be combined.

The Acknowledgments of people, grants, funds, etc should be brief.

Tables should be kept to a minimum and be designed to be as simple as possible. Tables are to be typed double-spaced throughout, including headings and footnotes. Each table should be on a separate page, numbered consecutively in Arabic numerals and supplied with a heading and a legend. Tables should be self-explanatory without reference to the text. The details of the methods used in the experiments should preferably be described in the legend instead of in the text. The same data should not be presented in both table and graph form or repeated in the text.

Figure legends should be typed in numerical order on a separate sheet. Graphics should be prepared using applications capable of generating high resolution GIF, TIFF, JPEG or Powerpoint before pasting in the Microsoft Word manuscript file. Tables should be prepared in Microsoft Word. Use Arabic numerals to designate figures and upper case letters for their parts (Figure 1). Begin each legend with a title and include sufficient description so that the figure is understandable without reading the text of the manuscript. Information given in legends should not be repeated in the text.

References: In the text, a reference identified by means of an author's name should be followed by the date of the reference in parentheses. When there are more than two authors, only the first author's name should be mentioned, followed by 'et al'. In the event that an author cited has had two or more works published during the same year, the reference, both in the text and in the reference list, should be identified by a lower case letter like 'a' and 'b' after the date to distinguish the works.

Examples:

Abayomi (2000), Agindotan et al. (2003), (Kelebeni, 1983), (Usman and Smith, 1992), (Chege, 1998;

1987a,b; Tijani, 1993,1995), (Kumasi et al., 2001)
References should be listed at the end of the paper in alphabetical order. Articles in preparation or articles submitted for publication, unpublished observations, personal communications, etc. should not be included in the reference list but should only be mentioned in the article text (e.g., A. Kingori, University of Nairobi, Kenya, personal communication). Journal names are abbreviated according to Chemical Abstracts. Authors are fully responsible for the accuracy of the references.

Examples:

Chikere CB, Omoni VT and Chikere BO (2008). Distribution of potential nosocomial pathogens in a hospital environment. *Afr. J. Biotechnol.* 7: 3535-3539.

Moran GJ, Amii RN, Abrahamian FM, Talan DA (2005). Methicillinresistant *Staphylococcus aureus* in community-acquired skin infections. *Emerg. Infect. Dis.* 11: 928-930.

Pitout JDD, Church DL, Gregson DB, Chow BL, McCracken M, Mulvey M, Laupland KB (2007). Molecular epidemiology of CTXM-producing *Escherichia coli* in the Calgary Health Region: emergence of CTX-M-15-producing isolates. *Antimicrob. Agents Chemother.* 51: 1281-1286.

Pelczar JR, Harley JP, Klein DA (1993). *Microbiology: Concepts and Applications*. McGraw-Hill Inc., New York, pp. 591-603.

Short Communications

Short Communications are limited to a maximum of two figures and one table. They should present a complete study that is more limited in scope than is found in full-length papers. The items of manuscript preparation listed above apply to Short Communications with the following differences: (1) Abstracts are limited to 100 words; (2) instead of a separate Materials and Methods section, experimental procedures may be incorporated into Figure Legends and Table footnotes; (3) Results and Discussion should be combined into a single section.

Proofs and Reprints: Electronic proofs will be sent (e-mail attachment) to the corresponding author as a PDF file. Page proofs are considered to be the final version of the manuscript. With the exception of typographical or minor clerical errors, no changes will be made in the manuscript at the proof stage.

Fees and Charges: Authors are required to pay a \$550 handling fee. Publication of an article in the Journal of Computational Biology and Bioinformatics Research is not contingent upon the author's ability to pay the charges. Neither is acceptance to pay the handling fee a guarantee that the paper will be accepted for publication. Authors may still request (in advance) that the editorial office waive some of the handling fee under special circumstances

Copyright: © 2015, Academic Journals.

All rights Reserved. In accessing this journal, you agree that you will access the contents for your own personal use but not for any commercial use. Any use and or copies of this Journal in whole or in part must include the customary bibliographic citation, including author attribution, date and article title.

Submission of a manuscript implies: that the work described has not been published before (except in the form of an abstract or as part of a published lecture, or thesis) that it is not under consideration for publication elsewhere; that if and when the manuscript is accepted for publication, the authors agree to automatic transfer of the copyright to the publisher.

Disclaimer of Warranties

In no event shall Academic Journals be liable for any special, incidental, indirect, or consequential damages of any kind arising out of or in connection with the use of the articles or other material derived from the JCBBR, whether or not advised of the possibility of damage, and on any theory of liability.

This publication is provided "as is" without warranty of any kind, either expressed or implied, including, but not limited to, the implied warranties of merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications does not imply endorsement of that product or publication. While every effort is made by Academic Journals to see that no inaccurate or misleading data, opinion or statements appear in this publication, they wish to make it clear that the data and opinions appearing in the articles and advertisements herein are the responsibility of the contributor or advertiser concerned. Academic Journals makes no warranty of any kind, either express or implied, regarding the quality, accuracy, availability, or validity of the data or information in this publication or of any other publication to which it may be linked.

ARTICLES

**Computational Sequence Analysis And *In Silico* Modeling Of A Stripe
Rust Resistance Protein Encoded By Wheat *TaHSC70* Gene**
Zarrin Basharat

Full Length Research Paper

Computational sequence analysis and *in silico* modeling of a stripe rust resistance protein encoded by wheat *TaHSC70* gene

Zarrin Basharat

Microbiology and Biotechnology Research Lab, Department of Environmental Sciences, Fatima Jinnah Women University, 46000, Pakistan.

Received 3 March 2013; Accepted 27 April 2015

***TaHSC70* gene of *Triticum* sp. is an associate of the heat shock protein family and plays a significant role in stress-related and defense responses induced by contagion with stripe rust fungus through a Jasmonic acid dependent signal transduction pathway. Hence, understanding molecular structure and function of the protein coded by this gene is of paramount importance for plant biologists working on stripe rust. The present study was aimed at sequence and *in silico* structural analysis of Hsp70 protein coded by this gene, through comparative modeling approach. Validation of the overall folds and structure, errors over localized regions and stereo chemical parameters was carried out using PDBSum server. Structure was a monomer with seven sheets, 1 β - α - β unit, 12 hairpins, 13 β -bulges, 29 strands, 21 helices, 16 helix-helix interactions, 44 β -turns and 1 Y-turn. Two major domains were detected belonging to Hsp70 family while neural network analysis revealed protein to be highly phosphorylated at serine and threonine residues.**

Key words: *TaHSC70*, Hsp70, Stripe rust, homology modelling, wheat.

INTRODUCTION

Stress impacts plants negatively and hinders proper activity. Stress protective roles in plants are played by Hsp70 family, which are induced in response to potential detrimental simulations (Efeoglu et al., 2009). *TaHSC70* demonstrates a decisive role in protecting plant cells against heat stress (Guo et al., 2014). Heat stress is one of the reasons behind pollen sterility, drying of stigmatic fluid/shrivelled seeds in wheat, pseudo-seed setting and empty endosperm pockets. The defence mechanisms of wheat to cope up with these conditions consists of heat responsive miRNAs, signalling molecules, transcription

factors and stress associated proteins like heat shock proteins (HSPs), antioxidant enzymes etc (Kumar and Rai, 2014). *TaHSC70* gene (70-kDa heat-shock cognate) is a constitutively expressed Hsp70 family member (Duan et al., 2011; Usman et al., 2014) in wheat. Furthermore, it is involved in protein-protein interactions, assisting the folding of de novo synthesized polypeptides and the import/translocation of precursor proteins (Feng et al., 2013; Wang et al., 2014). Heat shock proteins (HSPs) exist in nearly all living organisms (Feng et al., 2013). The major Hsps vary in molecular weights and are

E-mail: zarrin.iiui@gmail.com.

Author(s) agree that this article remain permanently open access under the terms of the [Creative Commons Attribution License 4.0 International License](#)

synthesized in eukaryotes belonging to six structurally distinct classes: Hsp100, Hsp90, Hsp70, Hsp60 (or chaperonins), ~17-30 kDa small Hsps and ~8-5 kDa ubiquitin (Safdar et al., 2012). Hsp70 family chaperones are considered to be the most highly conserved heat shock proteins (Jego et al., 2013). In plants, many Hsp70 proteins have been identified in different species (Daugaard et al., 2007). The Arabidopsis genome contains at least 18 genes encoding members of the Hsp70 family, Rice genome contains 32 (Sarkar et al., 2012), while, around 12 Hsp70 members have been found in the spinach genome (Guy and Li, 1998). The Hsp70 in wheat was reported by Duan et al. (2011) in expression profile analysis of the Arabidopsis and spinach. HSP70 has been observed to be increased in thermotolerant wheat variety so it is anticipated that HSP70 modulates the thermotolerance level of wheat (*Triticum aestivum*) pollen under heat stress (Kumar and Rai, 2014). This reveals that the over expression of Hsp70 genes correlates positively with the acquisition of thermo tolerance. HSPs are expressed in response to environmental stress conditions such as heat, cold and drought, as well as to chemical and other stresses (Daugaard et al., 2007) and results in enhanced tolerance to salt, water and high-temperature stress in plants (Alvim et al., 2001). However, the cellular mechanisms of Hsp70 function under stress conditions are not fully understood.

3D structure and conserved domain analysis can shed light on the function of a protein. The 3D structure of the wheat heat shock protein has not been modeled previously. Modeling is ground principally on alignment of query protein to the target (known structure or template). Prediction method may entail fold assignment, target-template alignment, model building followed by model evaluation (Marti-Renom et al., 2000). Comparative modeling approach has been utilized in this study to predict the three-dimensional structure of a given protein sequence (target) harnessing the bioinformatics tools. Functional analysis has also been attempted using a battery of computational tools and web servers.

MATERIALS AND METHODS

The 690 amino acid protein sequence encoded by the gene TaHSC70 with Accession ACT65562 was retrieved from the NCBI database.

Sequence analysis

Physicochemical properties of the protein were computed by ProtParam tool (<http://web.expasy.org/protparam/>). The parameters computed by ProtParam included the molecular weight, theoretical pI, instability index, aliphatic index, and grand average of hydropathicity (GRAVY). Subcellular localization of any protein aids understanding protein function. Prediction of subcellular localization of protein was carried out by CELLO v.2.5 (<http://cello.life.nctu.edu.tw/>). Phosphorylation profile analysis was

carried out using NetPhos 2.0 server (<http://www.cbs.dtu.dk/services/NetPhos/>).

Structure analysis

Blast (Altschul et al., 1990) search was performed with this query sequence against the Protein Data Bank (Berman et al., 2000). Query and template protein sequence were aligned using BioEdit program. Modeller (Fiser and Sali, 2003) was used to build a protein model using automated approach to comparative protein structure modeling by satisfaction of spatial restraints (Sali and Blundell, 1993; Eswar et al., 2008). The structure was energy minimized by SwissPDB viewer (Guex and Peitsch, 1997) using GROMOS96 force field and rendered in PYMOL (Delano, 2002). PDBSum analysis for secondary structure analysis was followed by PROCHECK (Laskowski et al., 1998) verification of the model by checking stereo chemical quality. Ramachandran plot (Ramachandran et al., 1963; Morris et al., 1992) was generated and the quality of the structure was computed in terms of percentage of residues in favourable regions, percentage of non Proline, glycine residues etc. ERRAT webserver (Colovos and Yeates, 1993) was also used to access quality of structure.

RESULTS AND DISCUSSION

Availability of plethora of quality tools and web servers has enabled computational biologists to perform reliable analysis of protein sequence and structure. The present study was aimed at sequence analysis and homology modeling of the wheat Hsp70 protein to shed light on its function.

Sequence analysis

Swiss protParam tool revealed the protein to be of ~73.5 kDa with theoretical pI value of 5.01. Total number of negatively charged residues (Asp + Glu) were 99 while total number of positively charged residues (Arg + Lys) were 82. The instability index was computed to be 29.0, classifying the protein as stable. Aliphatic index was found to be 86.33 while Grand average of hydropathicity (GRAVY) index was calculated as -0.272 demonstrating amino acid to be of soluble protein. CELLO results showed that the wheat Hsp70 protein is localized in the chloroplast. This is suggestive of the fact that chloroplast is the major site of function for wheat Hsp70 and the protein may be associated with the thermostability of chloroplast membranes. This can be allied to a study conducted by Bhadula and colleagues demonstrating association of 45 kD Hsps with heat stability of chloroplast membranes in a drought and heat resistant maize line (Bhadula et al., 2001).

Two major domains were detected in the sequence HSPA9-Ssq1-like_NBD (residues: 51-427) and PLN03184 (residue: 21-688). HSPA9-Ssq1-like_NBD or nucleotide-binding domain of HSPA9 belongs to the heat shock protein 70 (Hsp70) family of chaperones that contribute to protein folding and assembly and degrada-

Table 1. Phosphorylation profile of analysed Hsp70 protein using neural network approach. Specific residue positions in the query protein are shown to be phosphorylated based on a significant score. *S* refers to phosphorylation on serine residue and *T* refers to phosphorylation on threonine residue.

Position	Context	Score	Prediction	Position	Context	Score	Prediction
6	ATFTSQVSA	0.043	.	3	--MATFTSQ	0.014	.
9	TSQVVSAMAG	0.686	*S*	5	MATFTSQVS	0.145	.
15	MAGASPCSCS	0.901	*S*	50	AMRVTCSEKV	0.569	*T*
17	GASPCSCSLF	0.012	.	61	IDLGTTNSA	0.051	.
19	SPSCSLFVS	0.038	.	62	DLGTTNSAV	0.115	.
23	SLFVSRRRP	0.414	.	75	GGKPTVITN	0.020	.
64	GTTNSAVAA	0.005	.	78	PTVITNAEG	0.173	.
88	RTTPSVVAY	0.035	.	85	EGQRTTPSV	0.206	.
118	NTFFSVKRF	0.939	*S*	86	GQRTTPSVV	0.982	*T*
138	AKQVSYNVV	0.068	.	93	VVAYTKGGE	0.251	.
166	AEEISAQVL	0.013	.	115	NPENTFFSV	0.745	*T*
178	VDDASKFLN	0.057	.	186	NDKITKAVV	0.071	.
199	YFNDSQRTA	0.793	*S*	191	KAVVTVPAY	0.058	.
226	PTAASLAYG	0.023	.	202	DSQRTATKD	0.212	.
252	TFDVSVLEV	0.423	.	204	QRTATKDAG	0.699	*T*
265	FEVLSTSGD	0.145	.	223	INEPTAASL	0.035	.
267	VLSTSGDTH	0.321	.	238	KNNETILVF	0.030	.
287	DWLASTFKN	0.024	.	248	LGGGTFDVS	0.215	.
319	KMELSTLTQ	0.026	.	266	EVLSTSGDT	0.560	*T*
327	QANISLPFI	0.093	.	270	TSGDTHLGG	0.076	.
346	EATLSRAKF	0.226	.	288	WLASTFKND	0.579	*T*
355	EELCSDLID	0.007	.	308	LQRLTEAAE	0.284	.
375	DAKLSVSNL	0.984	*S*	320	MELSTLTQA	0.056	.
377	KLSVSNLDE	0.959	*S*	322	LSTLTQANI	0.020	.
388	LVGGSTRIP	0.314	.	332	LPFITATAD	0.028	.
417	DEVVSLGAA	0.520	*S*	334	FITATADGP	0.110	.
443	VTPLSIGLE	0.264	.	344	HIEATLSRA	0.496	.
466	TLPTSKSEV	0.995	*S*	363	DRLKTPVNN	0.943	*T*
468	PTSKSEVFS	0.171	.	389	VGGSTRIPA	0.248	.
472	SEVFSTAAD	0.480	.	397	AVQETVRKI	0.429	.
480	DGQTSVEIN	0.959	*S*	402	VRKITGKDP	0.940	*T*
498	RDNKSLGSF	0.503	*S*	409	DPNVTNPD	0.183	.
501	KSLGSFRLD	0.077	.	440	LLDVTPLSI	0.013	.
530	NGILSVAAV	0.040	.	448	IGLETGGV	0.306	.
550	ITGASTLPK	0.291	.	454	GGVMTKIIP	0.172	.

tion of incompetent proteins. Typically, Hsp70s have a nucleotide-binding domain (NBD) which hosts nucleotide and a substrate-binding domain (SBD) which increases rate of ATP-hydrolysis. NBD site (17 residues), nucleotide exchange factor (NEF) co-chaperone interaction site (19 residues) for regulation of HSP70 and SBD interface (11 residues) existing on the conserved domain HSPA-9-Ssq1-like-NB were detected on the query protein.

Protein phosphorylation is a type of post-translational modification which can turn a protein on and off, thus modifying its function and activity. Phosphorylation

generally occurs on serine, threonine, tyrosine and histidine residues in eukaryotic proteins. Artificial neural networks have been extensively used in biological sequence analysis (Wu, 1997; Blom et al., 1999) for phosphorylation analysis. Regions of wheat Hsp70 sequence showed extensive phosphorylation on serine and threonine residues (Table 1) while no phosphorylation capability of tyrosine residues was predicted. This result is in accordance with the study conducted by May and Soll (2000) that chloroplast-destined precursor proteins are phosphorylated on serine or threonine residues. This finding can be further validated

Table 1. Contd.

Position	Context	Score	Prediction	Position	Context	Score	Prediction
586	NQADSVVYQ	0.861	*S*	461	IPRNTTLPT	0.968	*T*
625	IAGGSTQNM	0.062	.	462	PRNTTLPTS	0.285	.
652	YNQTSAGGA	0.459	.	465	TTLPTSKE	0.572	*T*
658	GGAGSTDAE	0.174	.	473	EVFSTAADG	0.083	.
669	PGAGSTSSG	0.884	*S*	479	ADGQTSVEI	0.504	*T*
671	AGSTSSGKG	0.859	*S*	538	VDKGTGKKQ	0.305	.
672	GSTSSGKGP	0.833	*S*	545	KQDITITGA	0.061	.
689	DFTDSN---	0.016	.	547	DITITGAST	0.125	.
				551	TGASTLPKD	0.216	.
				580	DAIDTKNQA	0.863	*T*
				591	VVYQTEKQL	0.856	*T*
				626	AGGSTQNMK	0.067	.
				631	QNMKTAMEA	0.208	.
				651	MYNQTSAGG	0.090	.
				659	GAGSTDAET	0.593	*T*
				663	TDAETEPGA	0.092	.
				670	GAGSTSSGK	0.434	.
				687	DADFTDSN-	0.043	.

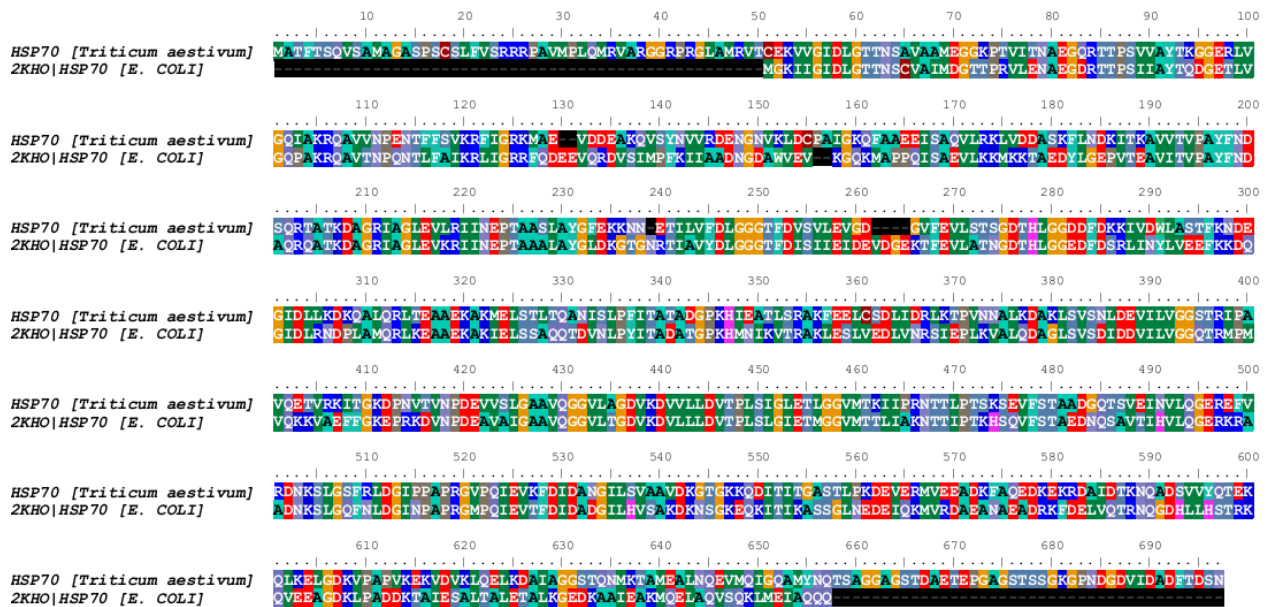


Figure 1. Aligned sequences of query and target protein visualized in BioEdit.

in the Lab and also tested for glycosylation that can further deepen our insight of the post translational modifications associated with wheat Hsp70.

Structure analysis

Homology modeling has gained popularity due to increas-

ing accuracy of the predictions using computational tools. For homology modelling, the suitable template structure selected was based X-ray structure of *E-coli* HSP70 protein (PDB ID:2KHO) (Bertelsen et al., 2009), having 55% identity with the query sequence and an E value of zero. Sequence was aligned to observe the residue conservation (Figure 1). Then, MODELLER was used to generate 3D structure. Predicted structure was a

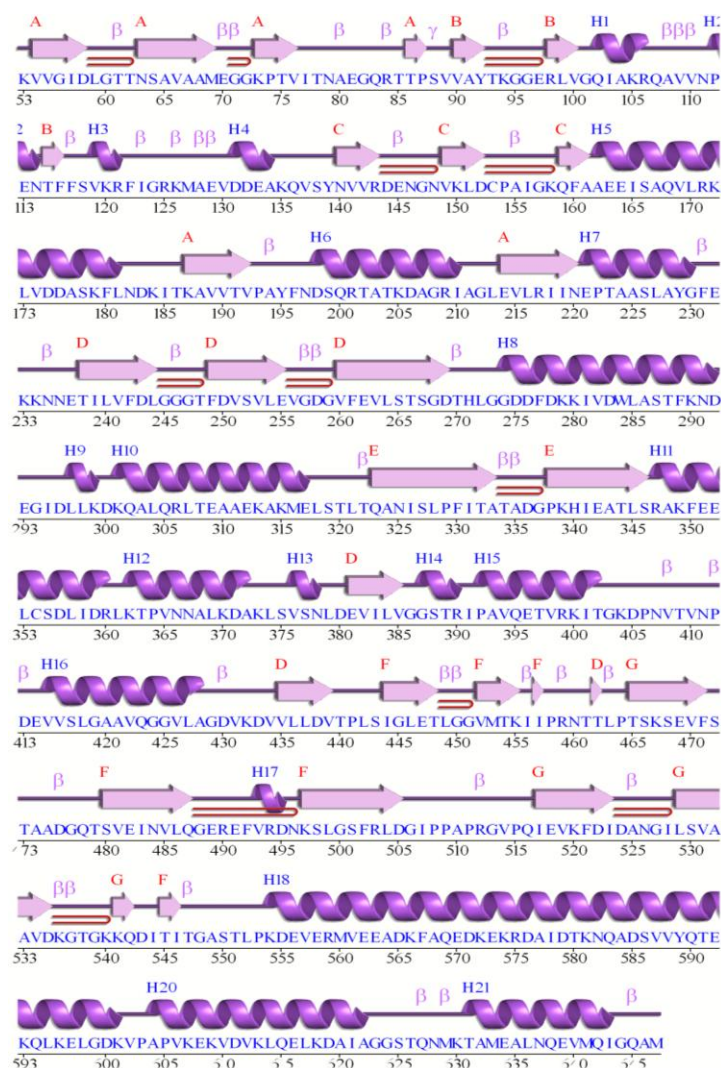


Figure 2. Secondary structure analysis of wheat Hsp70 protein. Helices (cylinder) labelled by H1, H2..., strands (arrow) by their sheets A,B... motifs by β , γ turn and (==) for beta hairpin.

monomer with molpdf score of 3662.76880, DOPE score value of -60344.10156 and a GA341 score of 1.00000. Protein consisted of 7 sheets, 1 beta alpha beta unit, 12 hairpins, 13 beta bulges, 29 strands, 21 helices, 16 helix-helix interacts, 44 beta turns and 1 gamma turn (Figure 2). Total number of bonds were 5216 while number of atoms were 5157. Structure validation of the predicted structures was done by feeding the predicting structure into the ERRAT protein verification server. The overall quality factor obtained was 74.671. The comparative peaks of DOPE scores of both template and model obtained from Modeller output demonstrate that there is no defect in the loop regions in the residues. So in the present case the loop refinement method was not required for the model (Figure 3). The validation of the model was carried out using Ramachandran plot calcula-

tions computed with the PROCHECK program. The Φ and Ψ distributions of the Ramachandran plots of non-Glycine, non-Proline residues are summarized in Figure 4. Altogether 99.2% of the residues were in favoured and allowed regions.

The overall G-factor used was computed as -0.1 which is good as compared to the typical value of -0.4. This is an initial attempt in modelling the structure of wheat Hsp70 and understanding its function. It is believed that this work has practical significance as it provides a foundation to not only the structure but also post translational modification of this protein. Post translational modification analysis can be further expanded to obtain new insights into the underpinnings of conformational changes in not only the cellular environment but also the chaperone itself. Structure can be utilized for interaction study with

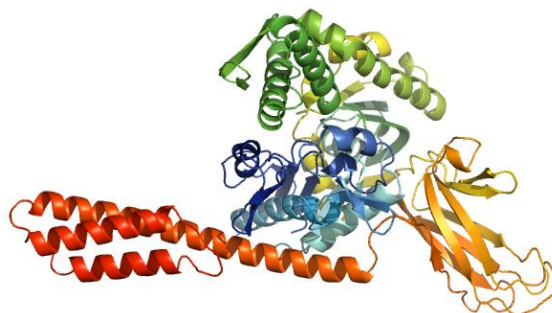


Figure 3. 3D structure of wheat Hsp70 protein visualized in PYMOL.

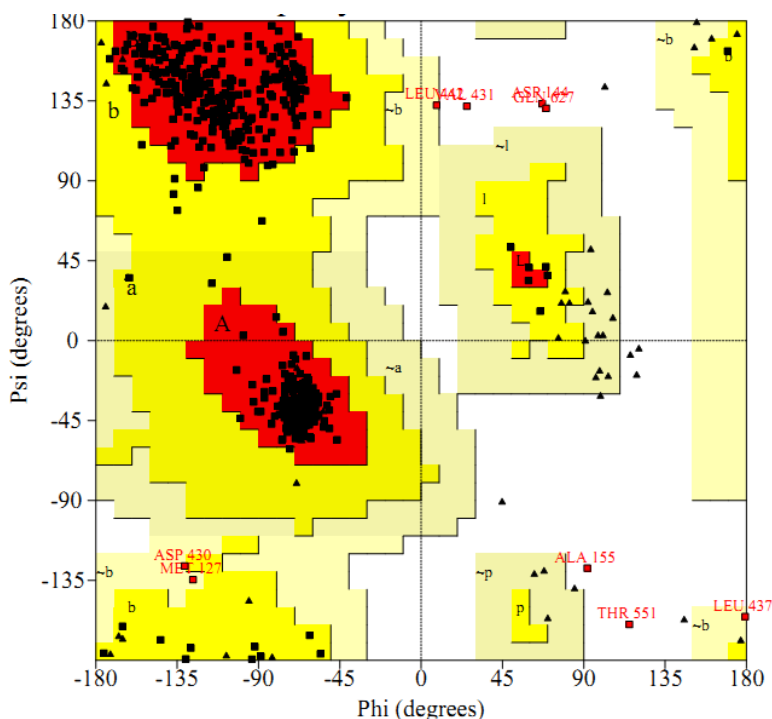


Figure 4. Ramachandran plot showing 94.0% residues in most favoured regions, 4.5% in additional allowed regions, 0.7% in generously allowed regions and only 0.8% in disallowed regions.

co-factors or other proteins/peptides in the cell to shed light on communication mechanism between chaperone and wheat cellular components under stress.

Conclusion

The wheat heat shock protein is one of the most important protein which provides the natural resistance against the stress due to stripe rust fungus. In the present work, sequence analysis has been conducted to shed light on post translational modification of Hsp70 domains associated with this protein and 3D structure study of the

protein. Computational study conducted can serve as a baseline source of information and can be further validated in the lab.

The validated protein model proposed in this study may be used further to dock with possible co-factors or relevant protein interactors to understand the potential mechanism of anti-stress and defense properties of this protein.

Conflict of interests

The authors did not declare any conflict of interest.

REFERENCES

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J. Mol. Biol.* 215 (3):403-410.
- Alvim FC (2001). Enhanced accumulation of BiP in transgenic plants confers tolerance to water stress. *Plant Physiol.* 126:1042-1054.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000). The Protein Data Bank. *Nucleic Acids Res.* 28 (1):235-242.
- Bhadula SK, Elthon TE, Habben JE, Helentjaris TG, Jiao S, Ristic Z (2001). Heat-stress induced synthesis of chloroplast protein synthesis elongation factor (EF-Tu) in a heat-tolerant maize line. *Planta* 212(3):359-366.
- Blom N, Gammeltoft S, Brunak S (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* 294(5):1351-1362.
- Colovos C, Yeates TO (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* 2(9):1511-1519.
- Daugaard M, Rohde M, Jäättelä M (2007). The heat shock protein 70 family: Highly homologous proteins with overlapping and distinct functions. *FEBS Lett.* 581(19):3702-3710.
- Duan YH, Guo J, Ding K, Wang SJ, Zhang H, Dai XW, Chen YY, Govers F, Huang LL, Kang ZS (2011). Characterization of a wheat HSP70 gene and its expression in response to stripe rust infection and abiotic stresses. *Mol. Biol. Rep.* 38(1):301-307.
- Efeoğlu B, Ekmekci Y, Cicek N (2009). Physiological responses of three maize cultivars to drought stress and recovery. *S. Afr. J. Bot.* 75(1):34-42.
- Eswar N, Eramian D, Webb B, Shen M, Sali A (2008). Protein structure modelling with MODELLER. *Methods. J Mol Biol.* 426:145-159.
- Feng PM, Chen W, Lin H, Chou KC. (2013). iHSP-PseRAAAC: Identifying the heat shock protein families using pseudo reduced amino acid alphabet composition. *Anal. Biochem.* 442(1):118-125.
- Fiser A, Sali A (2003). Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.* 374:461-491.
- Guex N, Peitsch MC (1997). SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis* 18(15):2714-2723.
- Guo M, Zhai YF, Lu JP, Chai L, Chai WG, Gong ZH, Lu MH. (2014). Characterization of CaHsp70-1, a Pepper Heat-Shock Protein Gene in Response to Heat Stress and Some Regulation Exogenous Substances in *Capsicum annuum* L. *Int. J. Mol. Sci.* 15(11):19741-19759.
- Guy CL, Li QB (1998). The organization and evolution of thenspinach stress 70 molecular chaperone gene family. *Plant Cell.* 10:539-556.
- Jego G, Hazoumé A, Seigneuric R, Garrido C. (2013). Targeting heat shock proteins in cancer. *Cancer Lett.* 332(2): 275-285.
- Kumar RR, Rai RD (2014). Can Wheat Beat the Heat: Understanding the Mechanism of Thermotolerance in Wheat (*Triticum aestivum* L.). *Cereal Res. Commun.* 42(1):1-18.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26:283-291.
- Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Šali A (2000). Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* 29(1): 291-325.
- May T, Soll J (2000). 14-3-3 proteins form a guidance complex with chloroplast precursor proteins in plants. *Plant Cell Online* 12(1):53-63.
- Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992). Stereochemical quality of protein structure coordinates. *Proteins* 12 (4): 345-364.
- Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 7:95-99.
- Safdar W, Majeed H, Ali B, Naveed I (2012). Molecular evolution and diversity of small heat shock proteins genes in plants. *Pak. J. Bot.* 44:211-218.
- Sali A, Blundell TL (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234 (3): 779–815.
- Sarkar NK, Kundhani P, Grover A (2013). Functional analysis of Hsp70 superfamily proteins of rice (*Oryza sativa*). *Cell Stress Chaperones* 18(4):427-437.
- Usman MG, Rafii MY, Ismail MR, Malek MA, Latif MA, Oladosu Y (2014). Heat Shock Proteins: Functions And Response Against Heat Stress In Plants. *Int. J. Sci. Technol. Res.* 3(11):204-218.
- Wang X, Gou M, Bu H, Zhang S, Wang G (2014). Proteomic analysis of Arabidopsis constitutive expresser of pathogenesis-related gene1 (Cpr30/cpr1-2) mutant. *Plant Omics J.* 7(3): 142-151.



Journal of Computational Biology and Bioinformatics Research

Related Journals Published by Academic Journals

- *Biotechnology and Molecular Biology Reviews*
- *African Journal of Biotechnology*
- *International Journal of Genetics and Molecular Biology*
- *Journal of Bioinformatics and Sequence Analysis*
- *International Journal of Biotechnology and Molecular Biology Research*

academicJournals