

## Article

# Molecular Cloning, Bioinformatics Analysis and Expression of Small Heat Shock Protein Beta-1 (HSPB-1) from *Camelus dromedarius*, Arabian Camel

Manee M. Manee <sup>1,2,‡</sup>, Sultan N. Alharbi <sup>2,3,‡,\*</sup>, Abdulmalek T. Algarni <sup>1,2,‡</sup>, Waleed M. Alghamdi <sup>4</sup>, Musaad A. Altammami <sup>5</sup>, Mohammad N. Alkhayef <sup>3</sup> and Basel M. Alnafjan <sup>5</sup>

<sup>1</sup> National Center for Genomic Technology, King Abdulaziz City for Science and Technology, Riyadh 11442, Saudi Arabia; E-Mail: malmanee@kacst.edu.sa

<sup>2</sup> Joint Center for Genomics Research (JGCR), King Abdulaziz City for Science and Technology, Riyadh 11442, Saudi Arabia; E-Mail: malmanee@kacst.edu.sa

<sup>3</sup> National Center for Stem Cell Technology, King Abdulaziz City for Science and Technology, Riyadh 11442, Saudi Arabia; E-Mail: snharbi@kacst.edu.sa

<sup>4</sup> Institute of Innovation and Industrial Development, King Abdulaziz City for Science and Technology, Riyadh 11442, Saudi Arabia; E-Mail: wghamdi@kacst.edu.sa

<sup>5</sup> National Center for Biotechnology, King Abdulaziz City for Science and Technology, Riyadh 11442, Saudi Arabia; E-Mail: mtammami@kacst.edu.sa

\* Correspondence: snharbi@kacst.edu.sa; Tel.: +966114883555

‡ These authors contributed equally to this work.

Academic Editor: name

Version July 27, 2017 submitted to Int. J. Mol. Sci.

**Abstract:** Small heat shock protein beta-1 (HSPB-1) plays an essential role in protection of cells against environmental stress. This makes the study of its molecular, structural, and biological characteristics in a naturally wild type model essentially crucial. Despite the fact that the sequence information of the HSPB-1 gene for many mammalian species are available, the HSPB-1 gene of Arabian Camel (Arabian Camel HSPB-1) has not been characteristically and structurally studied. We cloned and functionally characterized a full-length of Arabian Camel HSPB-1 cDNA. It is composed of 791 bp with a 5'-untranslated region (UTR) of 34 bp, a 3'-UTR of 151 bp with a poly(A) tail, and an open reading frame (ORF) of 606 bp encoding a protein of 201 amino acids (accession number: MF278354). The tissue-specific expression analysis of Arabian Camel HSPB-1 mRNA was examined using quantitative real-time PCR (qRT-PCR); which suggested that Arabian Camel HSPB-1 mRNA was constitutionally expressed in all examined tissues of Arabian Camel, with the predominately level in the lung tissue. Piptide mass fingerprint mass spectrometry (PMF-MS) analysis of the purified Arabian Camel HSPB-1 protein confirmed the identity of this protein. Phylogenetic analysis showed that Arabian Camel HSPB-1 protein is grouped together with that of Bactrian Camel and Alpaca. Comprising the predicted 3D structure of Arabian Camel HSPB-1 protein with available protein 3D structure of HSPB-1 from Human confirmed the present of  $\alpha$ -crystallin domain and showed high similarities using super secondary structure prediction.

**Keywords:** *Camelus dromedarius*; Arabian Camel HSPB-1 gene; cDNA cloning; Sequence in silico analysis; PMF-MS; 3D Structure modeling; Bioinformatics.

## 1. Introduction

The one-humped Camel, *Camelus dromedarius* (also known as Arabian Camel), is one of the most important member of the camelidae family. Arabian Camel has played a major role in the

23 culture and way of life in the Arabian Peninsula over past couple thousand years [1]. This animal  
24 has acclimatized itself to live in the desert, and to survive in extreme environmental conditions by  
25 promoting the expression of several genes such as small heat shock genes, which encode a family of  
26 proteins known as small heat shock proteins (sHSPs) [2–6]. They act as crucial role in Arabian Camel  
27 defence from such conditions by protecting other proteins from irreversible aggregation [7].

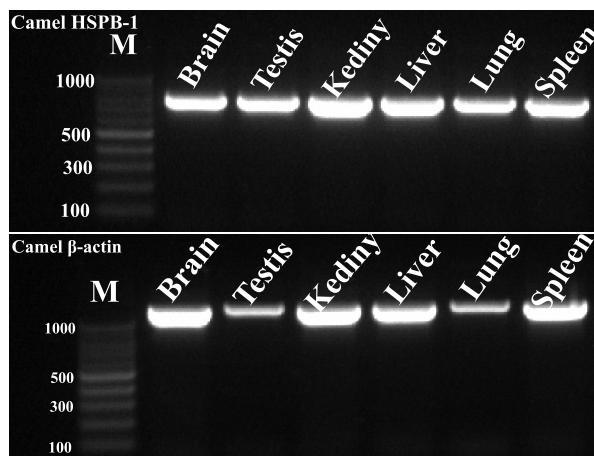
28 Small heat shock protein beta-1 (HSPB-1), as a typical member of the sHSPs family, is ubiquitously  
29 conserved ATP-independent protein, which is immensely preserved in a wide spectrum of organisms  
30 ranging from bacteria to eukarya [8,9]. Ten well-known members of the HSPB family (HSPB-1 to  
31 HSPB-10) have been well-studied in literature [10,11]. These proteins are molecular chaperones, which  
32 commonly have a low molecular weight ranging from 12–30 kDa, and generally distinguished by  
33 the presence of a typically conserved  $\alpha$ -crystallin domain (ACD) that is flanked by less conserved  
34 C-terminal extension (CTE) and N-terminal domain (NTD) [12–14]. The formation of a stable dimer  
35 interface between two contiguous monomers of small heat shock proteins' ACD facilitate the assembly  
36 of a large oligomers' subunits [15,16]. These molecular oligomers act as chaperones by binding to the  
37 unfolded proteins. Generally, the cellular concentration of many sHSPs is considerably increased in  
38 response to a variety of stresses, but they can also function fundamentally in many organisms and  
39 tissues [17].

40 Although HSPB-1 protein is highly conserved across species from bacteria to mammals, there  
41 are no reports about HSPB-1 protein from Arabian Camel. The aim of the present study is to clone  
42 and sequence a full-length of Arabian Camel HSPB-1 cDNA and its amino acid sequence as well as  
43 elucidate its protein structure. In addition, we examined the Arabian Camel HSPB-1 mRNA expression  
44 profile in six different tissues. We believe that the study of biochemical and biophysical aspects of  
45 Arabian Camel HSPB-1 gene is likely to provide molecular insights into *C. dromedarius* physiology as  
46 well as providing annotation of Camel HSPB-1 protein on which to advance further studies of Arabian  
47 Camel proteins.

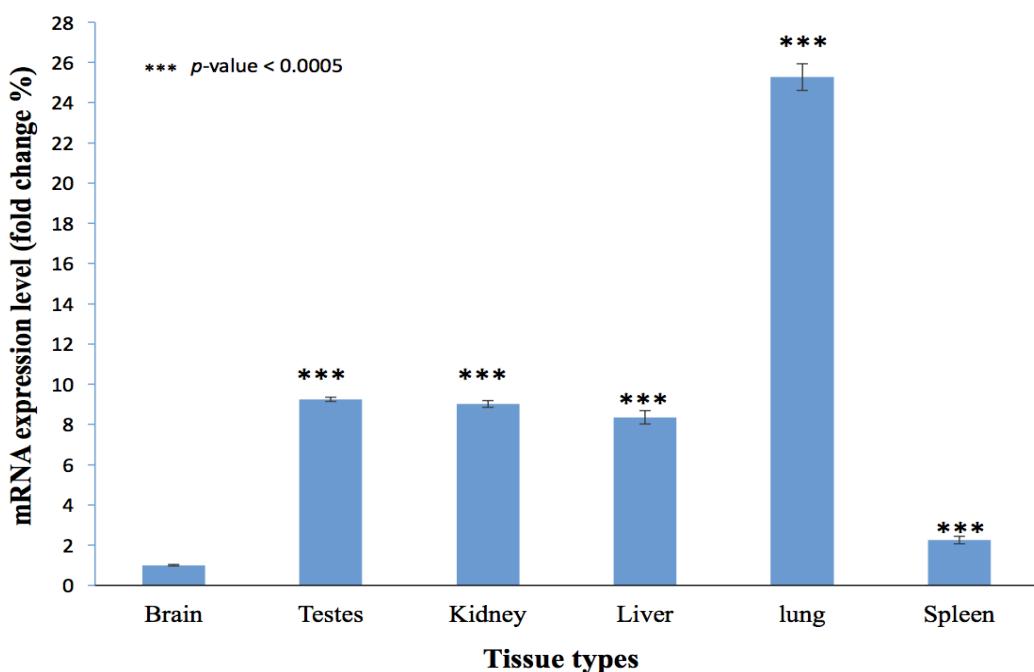
## 48 2. Results

### 49 2.1. Tissue-Specific Expression profile of Arabian Camel HSPB-1 mRNA

50 The expression of Arabian Camel HSPB-1 mRNA was found in all examined tissues of Arabian  
51 Camel (Figure 1), indicating its important role in cellular proteostasis. The specific primers (Table 8)  
52 were designed to amplify a single 759 bp for Arabian Camel HSPB-1 and 1140 bp for Arabian Camel  
53  $\beta$ -Actin genes (as endogenous control). In addition, the level of expression of Arabian Camel HSPB-1  
54 mRNA in six different tissues was studied using qRT-PCR. The qRT-PCR primers were designed to  
55 amplify 83 and 190 bp for Arabian Camel HSPB-1 and  $\beta$ -Actin, respectively. The maximum expression  
56 of Arabian Camel HSPB-1 mRNA was observed in the Arabian Camel lung; followed by nearly  
57 equally expressed in liver, kidney and testis, while the brain and spleen tissues revealed lowest levels  
58 of Arabian Camel HSPB-1 mRNAs. These observations may be significant in understanding the  
59 differential sensibilities of Arabian Camel tissues to environmental climate (Figure 2).



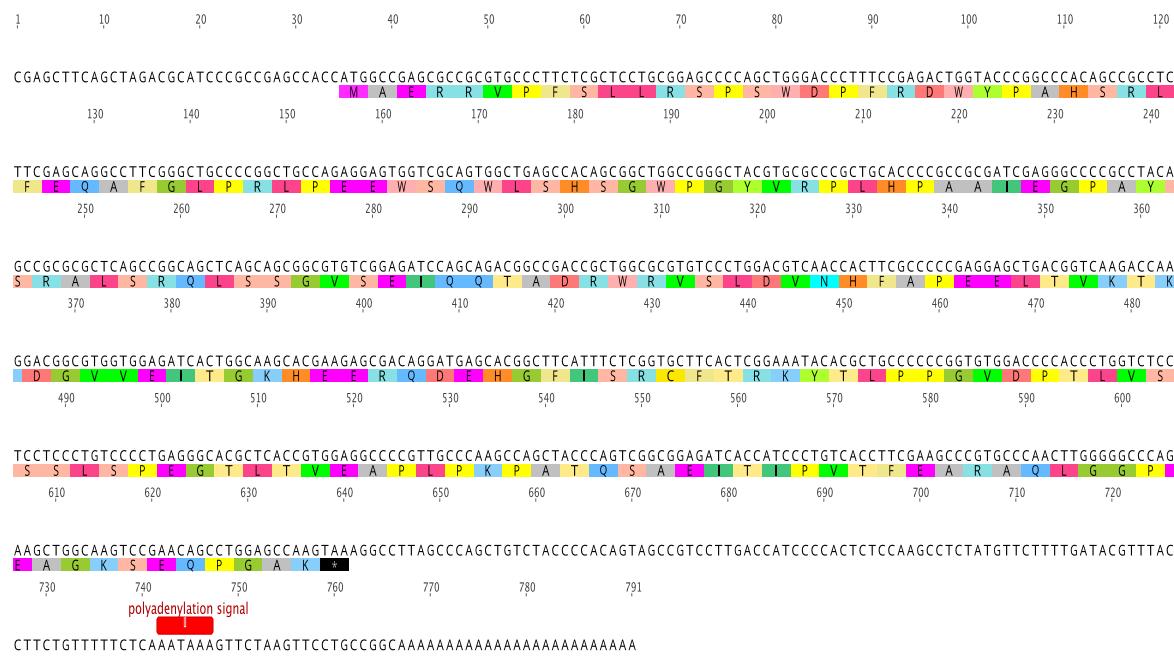
**Figure 1.** Agarose gel (1.2%) electrophoresis of PCR products for HSPB-1 and  $\beta$ -Actin Arabian Camel mRNAs, 1 kbp DNA molecular weight marker were used.



**Figure 2.** Arabian Camel HSPB-1 mRNA expression levels in different tissues. The results are expressed relative to  $\beta$ -Actin as endogenous control.

## 60 2.2. Characterization and sequence of full-length of HSPB-1 cDNA from Arabian Camel

61 The full-length of Arabian Camel HSPB-1 cDNA contained a 5'-untranslated region (UTR) of 34  
 62 bp, a 3'-UTR of 151 bp with typical polyadenylation signal (AATAAA), and with a poly(A) tail were  
 63 obtained and deposited as GenBank accession No. MF278354. The open reading frame (ORF) includes  
 64 606 bp; encoding a protein of 201 residues (Figure 3). The sequence indicated a length of 791 bp, and  
 65 revealed high statically significant similarity scores to many HSPB-1 nucleotide sequences from other  
 66 species. The Bactrian Camel (*Camelus bactrianus*) showed the highest homology score of 99% compared  
 67 to other mammalian organisms alongside Alpacas (*Vicugna pacos*); suggesting a close evolutionary  
 68 relationship. The other mammals shared a high identity score ranging from 82% to 92% as shown in  
 69 (Table 1).



**Figure 3.** Nucleotide and amino acid sequences of Arabian Camel HSPB-1 cDNA (GenBank Accession No.MF278354). The numbers above the nucleotide sequence to show the nucleotide positions. The stop codon is represented with an asterisk (\*). The putative polyadenylation signal is showed in red.

**Table 1.** Nucleotide homology of Arabian Camel HSPB-1 with other species.

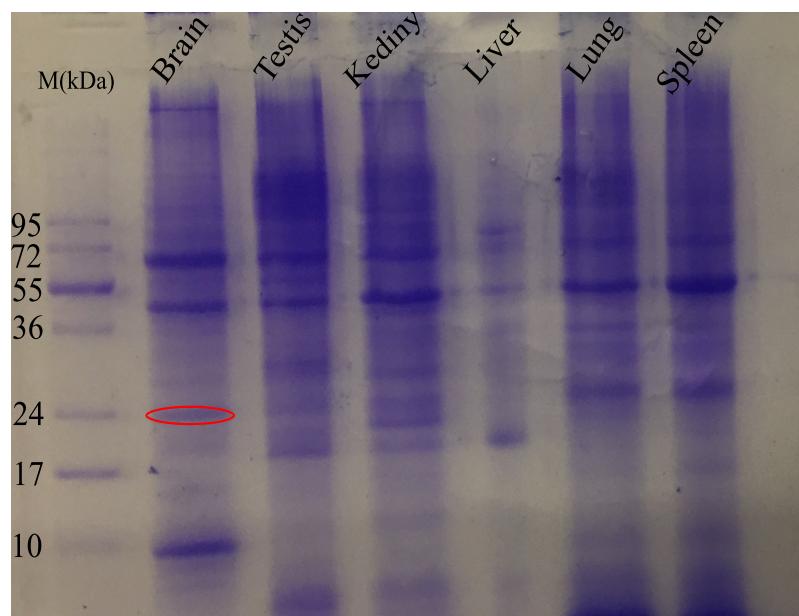
Species	Common name	Accession no.	cDNA length (bp)	Identity (%)
<i>Homo sapiens</i>	Human	BC000510.2	867	85
<i>Pan troglodytes</i>	Chimpanzee	XM_519162.5	947	85
<i>Mus musculus</i>	Mouse	NM_013560.2	913	82
<i>Sus scrofa</i>	Pig	NM_001007518.1	624	91
<i>Equus caballus</i>	Horse	XM_001504478.3	923	88
<i>Bos taurus</i>	Cattle	AB605262.1	672	92
<i>Camelus dromedarius</i>	Arabian Camel	MF278354	791	100
<i>Capra hircus</i>	Goat	XM_018040903.1	891	91
<i>Camelus bactrianus</i>	Bactrian Camel	XM_010972325.1	849	99
<i>Vicugna pacos</i>	Alpaca	XM_015236804.1	820	99
<i>Macaca mulatta</i>	Rhesus monkey	NM_001260949.2	893	85
<i>Ailuropoda melanoleuca</i>	Giant panda	NM_001304892.1	633	90
<i>Canis lupus familiaris</i>	Dog	NM_001003295.2	864	87
<i>Macaca fascicularis</i>	Crab-eating macaque	NM_001283885.1	845	85

### 2.3. Identification of Arabian Camel HSPB-1 protein by mass spectrometry

For peptide mass fingerprint mass spectrometry (PMF-MS), the targeted protein band (spot) was manually excised from the gel (Figure 4) and was subjected to MS analysis. Out of total trypsin-digested peptide masses of Arabian Camel HSPB-1 protein, seven peptides, which are covering 49% of whole protein sequence, were hit in NCBIproteins database (containing 4114420 sequences) by Mascot peptide fingerprint search engine with Arabian Camel HSPB-1 protein (accession no. XP\_010979037.1) with 125 score and  $p < 0.05$  (Figure 5).

The mass spectrum revealed several protonated ions  $[M+H]^+$  of the peptide fragments. As listed in (Table 2), the ions at 1031.89, 1178.04, 2314.72, 1479.23, 1619.27, 1798.42, and 1831.55 were the seven trypsin digested peptides corresponding to amino acids 21–28, 29–38, 39–57, 58–71, 76–90, 93–108, and

80 168-184. As interpreted in (Table 2), peptide mass profiles were obtained from NCBIprot database  
 81 search engine and amino acid sequence of individual peptides were identified from sequence of  
 82 Arabian Camel HSPB-1 protein from desired spot of Arabian Camel HSPB-1 protein of SDS-page.  
 83 The PMF-MS results were also homology with some other animals, the second best matching protein  
 84 received a score of 102 for Alpaca (accession no. XP\_015092290) HSPB-1 protein. The third and fourth  
 85 best matching proteins were scored with 100 and 80 for Bactrianus Camel (accession no. XP\_010970627)  
 86 and Pig (accession no. NP\_001007519.1) HSPB-1 proteins, respectively.



**Figure 4.** SDS-PAGE (15%) followed by staining with the Coomassie Blue as indicated in Section "Materials and Methods".

1 MAERRVPFSL LRSPSWDPFR DWYPAHSRLF EQAFGLPRLP EEWSQLSHS  
 51 GWPGYVRPLH PAAIEGPAYS RALSRQLSSG VSEIQQTADR WRVSLDVNHF  
 101 APEELTVKTK DGVVEITGKH EERQDEHGFI SRCFTRKYTL PPGVDPTLVS  
 151 SSLSPEGTLT VEAPLPKPAT QSAEITIPVT FEARAQLGGP EAGKSEQPGA  
 201 K

**Figure 5.** MLDI-TOF MS-derived peptides (red) matched sequence of Arabian Camel HSPB-1 protein

**Table 2.** Calculated and observed ions of peptide masses of Arabian Camel HSPB-1 protein.

Amino Acid positions Start-End	[M+H] <sup>+</sup>		Peptide Sequence
	Observed(m/z)	Calculated(m/z)	
21-28	1031.89	1030.46	DWYPAHSR
29-38	1178.04	1176.63	LFEQAFGLPR
39-57	2314.72	2313.1	LPEEWSQWLHSGWPGYVR
58-71	1479.23	1477.77	PLHPAAIEGPAYS
76-90	1619.27	1617.8	QLSSGVSEIQQTADR
93-108	1798.42	1796.93	VSLDVNHFAPEELTVK
168-184	1831.55	1829.95	PATQSAEITIPVTFEAR

87 2.4. Characterization of HSPB-1 protein from Arabian Camel

88 The protein sequence was compared with other mammalian HSPB-1 protein sequences using  
89 ClustalW alignment[18] as shown in (Figure 6). Results of multiple sequence alignment of Arabian  
90 Camel HSPB-1 protein showed two highly conserved domains about 85-residues (from 88-173): Alpha  
91 crystallin domain (ACD) and IbpA , which flanked by less conserved NTD and CTE across species. The  
92 comparative analysis of Arabian Camel HSPB-1 protein sequence showed a high similarity with other  
93 vertebrates' HSPB-1 proteins (Table 3). As expected, the highest homology was observed between  
94 Arabian Camel HSPB-1 protein and the one from Bactrian Camel (99%). The other vertebrates showed  
95 a high homology ranging from 86% to 95% as shown in (Table 3). The complete amino acid sequence  
96 of Arabian Camel HSPB-1 is showed in (Figure 3).

97 Based on the amino acid composition, the average calculated isoelectric point (pI) for Arabian  
98 Camel HSPB-1 protein using a computer algorithm [19] was found to be 6.162 (Figure 7), while its  
99 estimated molecular weight was 22.382 kDa. The basic, acidic, charged, polar and hydrophobic  
100 amino acids were 22(10.95%), 25(12.44%), 58(28.86%), 47(23.38%) and 65(32.34%), respectively. the  
101 hydrophobic and aromatic amino acids are overrepresented in the NTD, while polar and charged  
102 amino acids are underrepresented [20]. The instability of Arabian Camel HSPB-1 protein was calculated  
103 to be 64.99 which classified this protein as unstable. The molar extinction coefficient was found to be  
104  $39085 \pm 5\% \text{ cm}^{-1} M^{-1}$ . The amino acids composition of Arabian Camel HSPB-1 protein is showed in  
105 (Table 4).

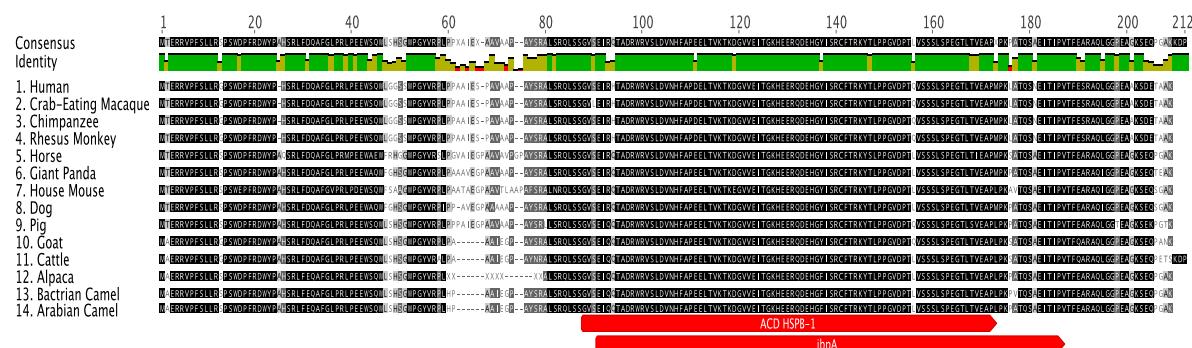
106 Protein structural flexibility has been predicted from amino acid sequence of Arabian Camel  
107 HSPB-1 protein using Karplus and Schulz method [21] in which the size of window was optimized to  
108 7 residues (Figure 8). The flexibility analysis showed that Arabian Camel HSPB-1 protein is more  
109 flexible at its C-terminal than N-terminal regions and thus possibly also the surface amino acids in this  
110 protein, which may be presented epitopes. In addition, Arabian Camel HSPB-1 protein sequence was  
111 used as a query to identify B cell epitopes using Kolaskar and Tongaonkar antigenicity method [22]  
112 (Figure 9). the results showed that the average antigenic tendency value of 1.027 for the protein with  
113 the minimum value of 0.876 and maximum of 1.192. This protein bears nine antigenic peptides with  
114 length range from 6-20 amino acids (Table 5). The results also revealed that the two regions from 8 to  
115 11 and 113 to 116 amino acids are the most preferred B cell epitope characteristics.

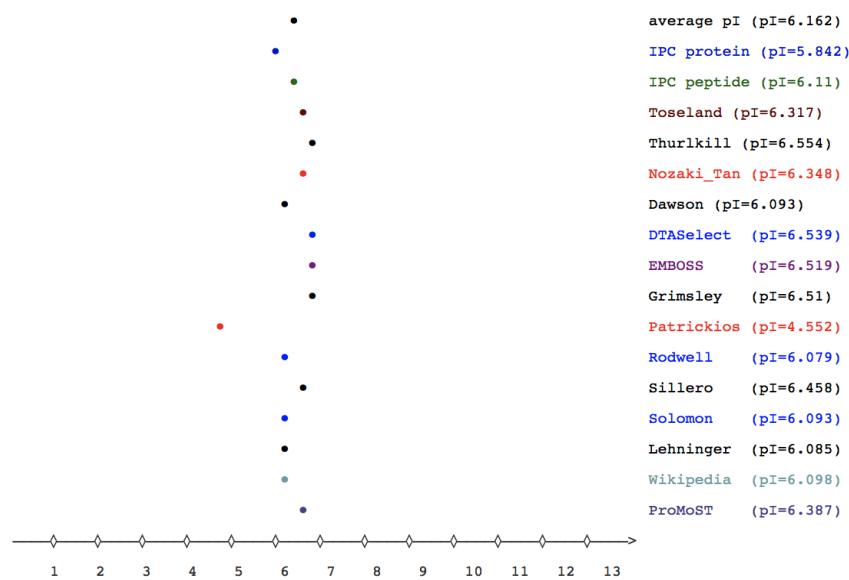
116 In order to predict the amino acid sites that are located on the surface of Arabian Camel HSPB-1  
117 protein, the Parker hydrophilicity tool [23] and the Emini surface accessibility prediction [24] were  
118 used. It is most likely those sites will increase the probability of predicting the antigenic regions as  
119 they are more accessible and hydrophilic than interior regions of the protein. The maximum surface  
120 probability value was found to be 5.823 from amino acid position 121 to 126 for Arabian Camel HSPB-1  
121 protein (Figures 10 and 11). In addition,  $\beta$ -turns structure in a protein are mostly hydrophilic and  
122 surface accessible in nature. Those  $\beta$ -turns were also predicted in Arabian Camel HSPB-1 protein using  
123 Chou and Fasman Beta turn prediction [25]. These results suggested that this protein is rich- $\beta$ -turns in  
124 the region between 80 to 175 residues, which is the region where  $\beta$ -strands are oriented in anti-parallel  
125 to form  $\beta$ -sheets (Figure 12).

126 GlobPlot server (<http://globplot.embl.deis>) was used in order to predict disordered and ordered  
127 (globular) regions within Arabian Camel HSPB-1 protein. In which ordered regions described as  
128 those have regular secondary structure ( $\alpha$ -helices and  $\beta$ -strands), while disordered regions are those  
129 lack such structures. The Russell/Linding [26] set was chosen in which  $\alpha$ -helices and  $\beta$ -strands  
130 structures are assigned as globular regions (GlobDoms), whereas random coils and  $\beta$ -turns structures  
131 as disordered regions. Residue ranges for found disordered regions (blue) and globular regions (green)  
132 are shown at the bottom of the graph (Figure 13).

**Table 3.** Amino acids homology of Arabian camel HSPB-1 with other species.

Species	Common name	Protein (Accession no.)	Protein length (bp)	Identity (%)
<i>Homo sapiens</i>	Human	NP_001531	205	86
<i>Pan troglodytes</i>	Chimpanzee	XP_519162.3	205	86
<i>Mus musculus</i>	Mouse	NP_038588	209	86
<i>Sus scrofa</i>	Pig	NP_001007519	207	92
<i>Equus caballus</i>	Horse	XP_001504528	209	87
<i>Bos taurus</i>	Cattle	NP_001020740	204	95
<i>Camelus dromedarius</i>	Arabian Camel	XP_010979037.1	201	100
<i>Capra hircus</i>	Goat	XP_017896392	201	95
<i>Camelus bactrianus</i>	Bactrian Camel	XP_010970627	201	99
<i>Vicugna pacos</i>	Alpaca	XP_015092290	197	94
<i>Macaca mulatta</i>	Rhesus monkey	NP_001247878.1	205	86
<i>Ailuropoda melanoleuca</i>	Giant panda	NP_001291821.1	207	90
<i>Canis lupus familiaris</i>	Dog	NP_001003295.2	206	89
<i>Macaca fascicularis</i>	Crab-eating macaque	NP_001270814.1	205	86

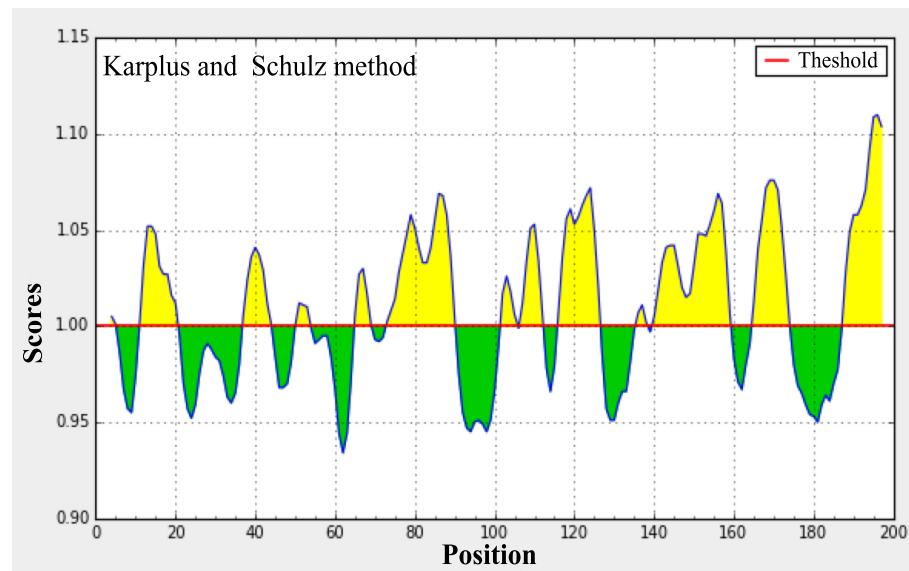
**Figure 6.** Multiple alignment of amino acid sequence of Arabian Camel HSPB-1 protein with other 13 mammalian species. Identical amino acids are in green color, typical ACD and ipbA domains are showed in red.



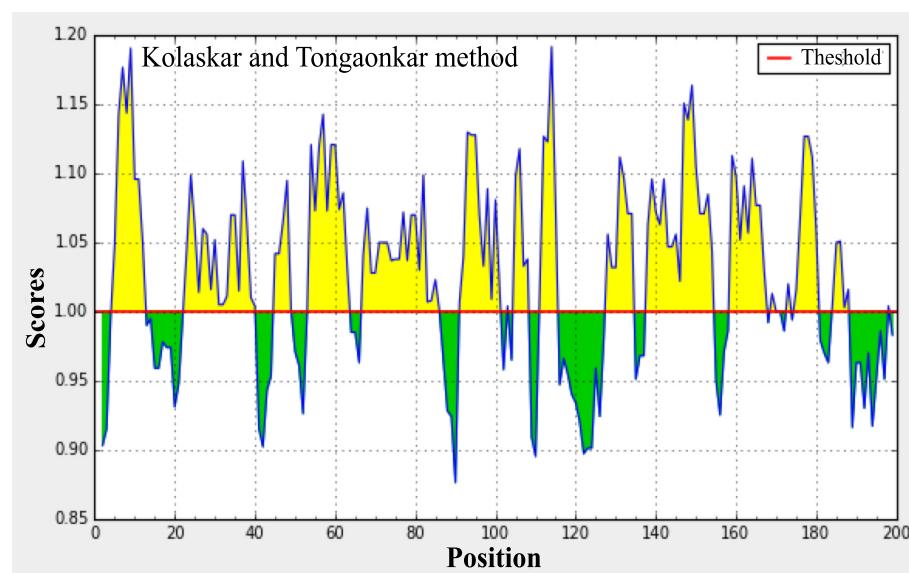
**Figure 7.** Isoelectric point (pI) of Arabian Camel HSPB-1 protein according to different scale calculation.

**Table 4.** Chemical Composition of Arabian Camel HSPB-1 Protein.

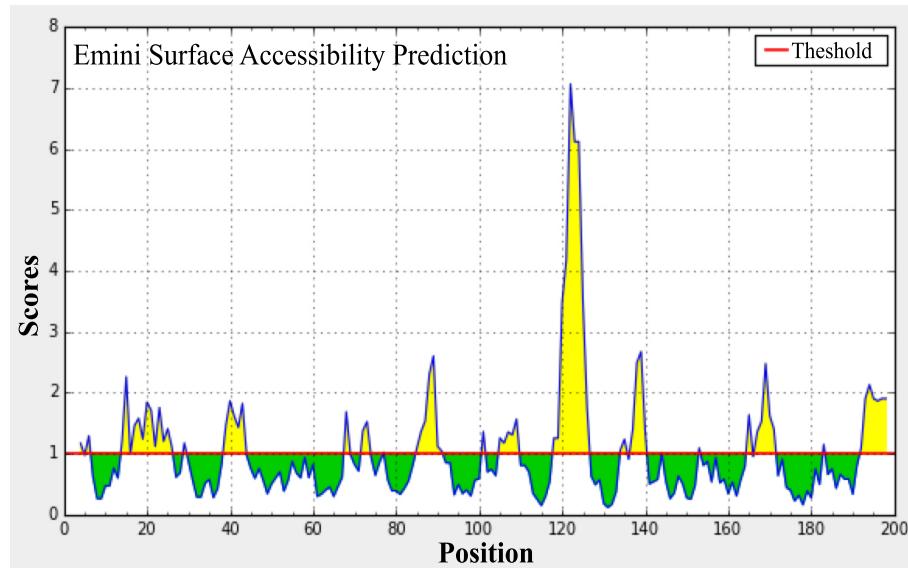
Amino Acid	Number Count	% by Frequency	% by Weight
A(Ala)	16	7.69	5.08
C(Cys)	1	0.5	0.46
D(ASP)	7	3.48	3.6
E(Glu)	18	8.96	10.38
F(Phe)	8	3.98	5.26
G(Gly)	14	6.97	3.57
H(His)	6	2.99	3.68
I(Ile)	6	2.99	3.03
K(Lys)	7	3.48	4.01
L(Leu)	17	8.46	8.58
M(Met)	1	0.5	0.59
N(Asn)	1	0.5	0.51
P(Pro)	21	10.45	9.11
Q(Gln)	9	4.48	5.15
R(Arg)	15	7.46	10.47
S(Ser)	20	9.95	7.78
T(Thr)	12	5.97	5.42
V(Vel)	12	5.97	5.31
W(Trp)	6	2.77	4.99
Y(Try)	4	1.99	2.92



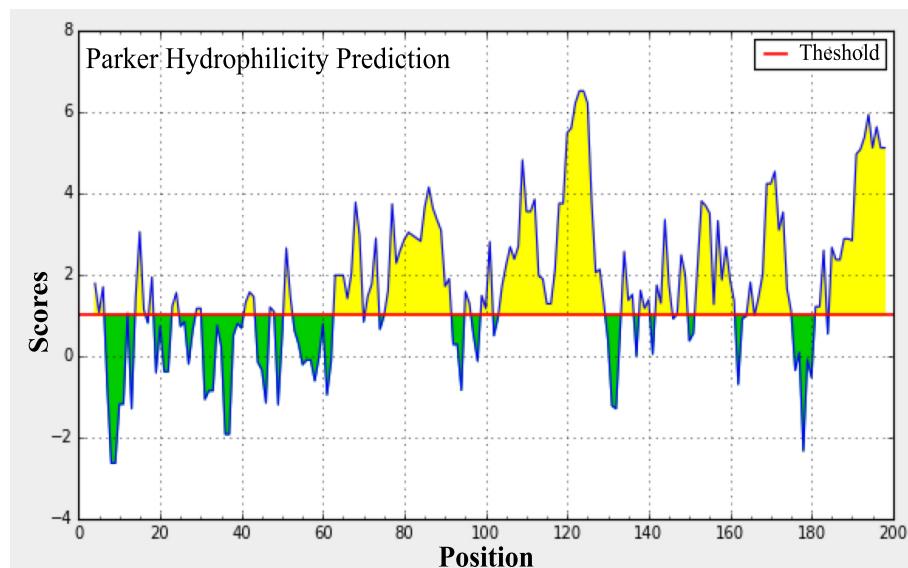
**Figure 8.** Karplus and Schulz flexibility prediction of Arabian Camel HSPB-1 protein. The x-axis and y-axis represent the position and score, respectively. The threshold is 1.0. The flexible regions of the protein are shown in yellow color, above the threshold value.



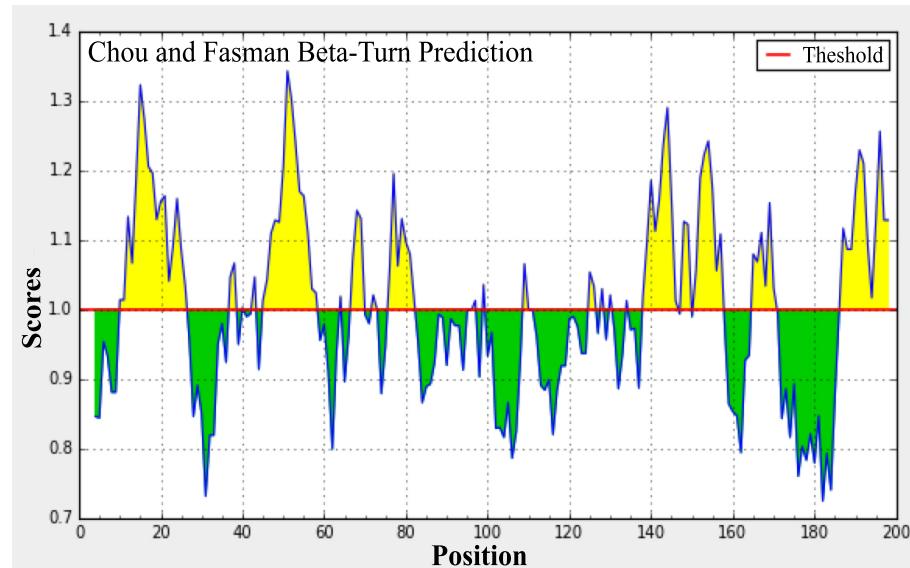
**Figure 9.** Kolashkar and Tongaonkar antigenicity prediction of the most antigenic regions of Arabian Camel HSPB-1 protein. The threshold value is 1.0. The regions above the threshold are antigenic, shown in yellow.



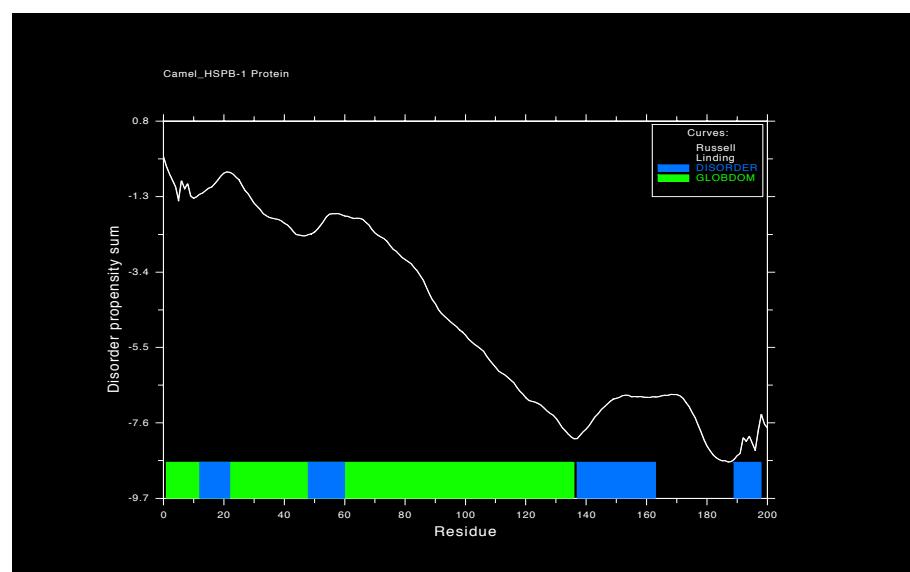
**Figure 10.** Emini surface accessibility prediction of Arabian Camel HSPB-1 protein. The threshold value is 1.000. The regions above the threshold are antigenic, shown in yellow



**Figure 11.** Parker hydrophilicity prediction of Arabian Camel HSPB-1 protein. The threshold is 1.0. The regions having  $\beta$ -turns in the protein are shown in yellow color, above the threshold value.



**Figure 12.** Chou and Fasman  $\beta$ -turns prediction of Arabian Camel HSPB-1 protein. The threshold is 1.00. The regions having  $\beta$ -turns in the protein are shown in yellow color



**Figure 13.** Glob Plot analysis. Blue boxes are disordered regions and green boxes are ordered regions in Arabian Camel HSPB-1 protein.

**Table 5.** Kolaskar and Tongaonkar antigenicity analysis.

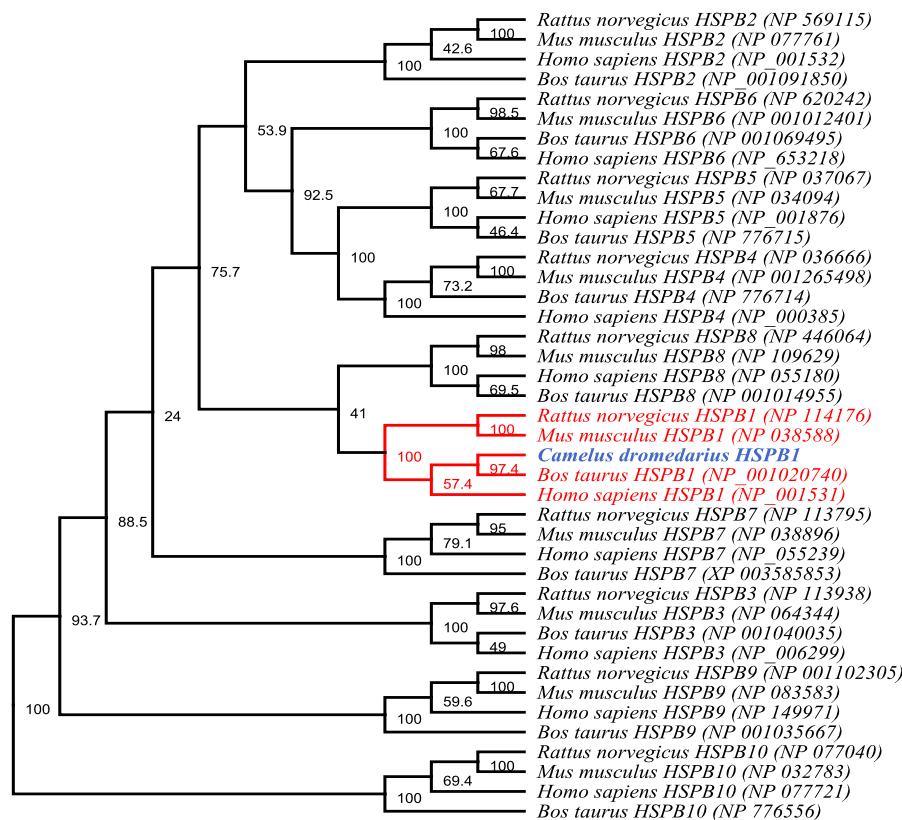
No.	Start	End	Peptide	Length
1	5	12	RVPLSLLR	8
2	23	40	YPAHSRLFEQAFGLPRLP	18
3	54	63	GYVRPLHPAA	10
4	67	86	PAYSRALSRQLSSGVSEIQQ	20
5	91	101	WRVSLDVNHFA	11
6	128	134	GFISRCF	7
7	138	154	YTLPPGVDP TLVSSLS	17
8	159	167	LTVEAPLPK	9
9	175	180	ITIPVT	6

**133 2.5. Phylogeny and classification of Arabian Camel HSPB-1 protein**

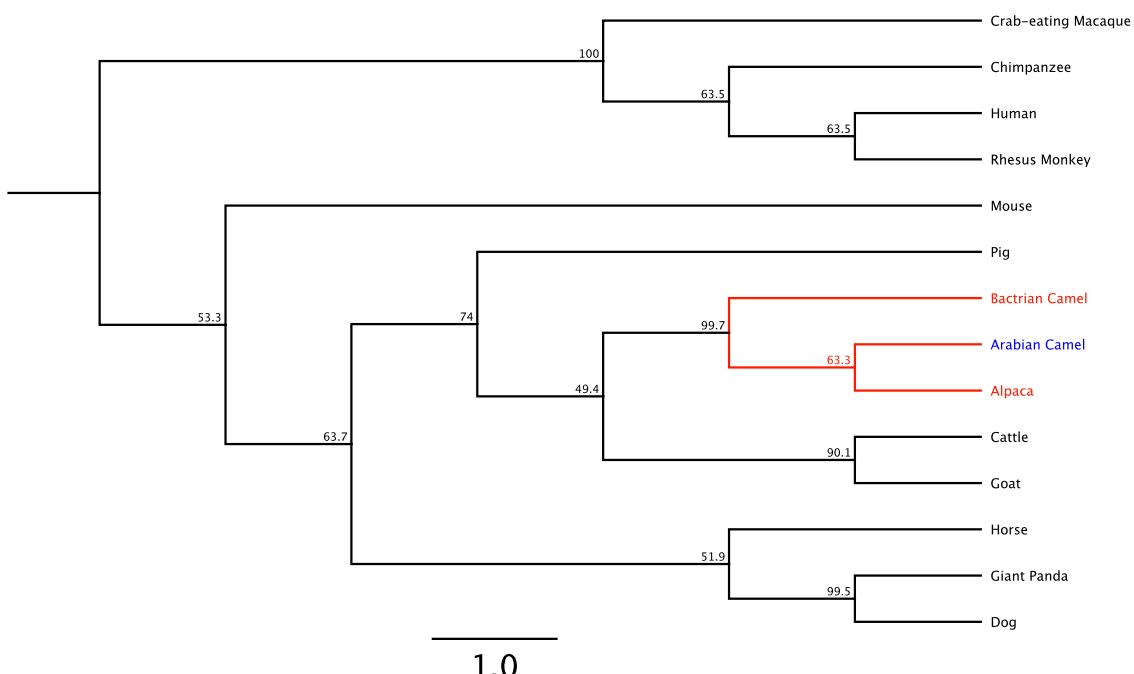
**134** After confirming the relationship of Arabian Camel HSPB-1 protein to the HSPB-1 family, we  
**135** constructed phylogenetic trees using Arabian Camel HSPB-1 protein sequence as a query to retrieve  
**136** 40 orthologues sequences derived from different vertebrate species (Table 6). Neighbour-joining  
**137** phylogenetic trees were constructed based on a multiple alignment of the HSPB-1 protein sequences  
**138** (Figure 14). The depicted topology showed the Arabian Camel HSPB-1 clustered closely with even-toed  
**139** ungulates' HSPB-1 into two distinct clades. Perhaps, this was caused by the gene duplication. In  
**140** addition, the evolution position of Arabian Camel was shown in a phylogenetic tree (Figure 15). The  
**141** Arabian camel HSPB-1 grouped more closely with the Bactrian Camel, Alpaca from Cattle, Goat and  
**142** further related with Pig.

**Table 6.** 40 orthologues sequences retrieved with Arabian Camel protein.

Protein family	Species	Accession no.	ACD	Length (bp)
<b>HSPB-1</b>	<i>Homo sapiens</i>	NP_001531	84-169	205
	<i>Mus musculus</i>	NP_038588	88-173	209
	<i>Rattus norvegicus</i>	NP_114176	88-173	206
	<i>Bos taurus</i>	NP_001020740	80-165	204
<b>HSPB-2</b>	<i>Homo sapiens</i>	NP_001532	66-148	182
	<i>Mus musculus</i>	NP_077761	66-148	182
	<i>Rattus norvegicus</i>	NP_569115	66-148	182
	<i>Bos taurus</i>	NP_001091850	66-148	182
<b>HSPB-3</b>	<i>Homo sapiens</i>	NP_006299	61-145	150
	<i>Mus musculus</i>	NP_064344	65-149	154
	<i>Rattus norvegicus</i>	NP_113938	60-147	152
	<i>Bos taurus</i>	NP_001040035	60-144	149
<b>HSPB-4</b>	<i>Homo sapiens</i>	NP_000385	63-145	173
	<i>Mus musculus</i>	NP_001265498	92-174	202
	<i>Rattus norvegicus</i>	NP_036666	63-145	173
	<i>Bos taurus</i>	NP_776714	63-145	173
<b>HSPB-5</b>	<i>Homo sapiens</i>	NP_001876	67-149	175
	<i>Mus musculus</i>	NP_034094	67-149	175
	<i>Rattus norvegicus</i>	NP_037067	67-149	175
	<i>Bos taurus</i>	NP_776715	67-149	175
<b>HSPB-6</b>	<i>Homo sapiens</i>	NP_653218	66-148	160
	<i>Mus musculus</i>	NP_001012401	66-148	162
	<i>Rattus norvegicus</i>	NP_620242	66-148	162
	<i>Bos taurus</i>	NP_001069495	66-148	164
<b>HSPB-7</b>	<i>Homo sapiens</i>	NP_055239	74-153	170
	<i>Mus musculus</i>	NP_038896	74-152	169
	<i>Rattus norvegicus</i>	NP_113795	74-152	169
	<i>Bos taurus</i>	XP_003585853	74-152	169
<b>HSPB-8</b>	<i>Homo sapiens</i>	NP_055180	84-170	196
	<i>Mus musculus</i>	NP_109629	84-170	196
	<i>Rattus norvegicus</i>	NP_446064	84-170	196
	<i>Bos taurus</i>	NP_001014955	84-170	196
<b>HSPB-9</b>	<i>Homo sapiens</i>	NP_149971	42-127	159
	<i>Mus musculus</i>	NP_083583	77-166	199
	<i>Rattus norvegicus</i>	NP_001102305	75-166	203
	<i>Bos taurus</i>	NP_001035667	46-128	157
<b>HSPB-10</b>	<i>Homo sapiens</i>	NP_077721	114-202	250
	<i>Mus musculus</i>	NP_032783	100-188	248
	<i>Rattus norvegicus</i>	NP_077040	100-1883	245
	<i>Bos taurus</i>	NP_776556	114-202	262



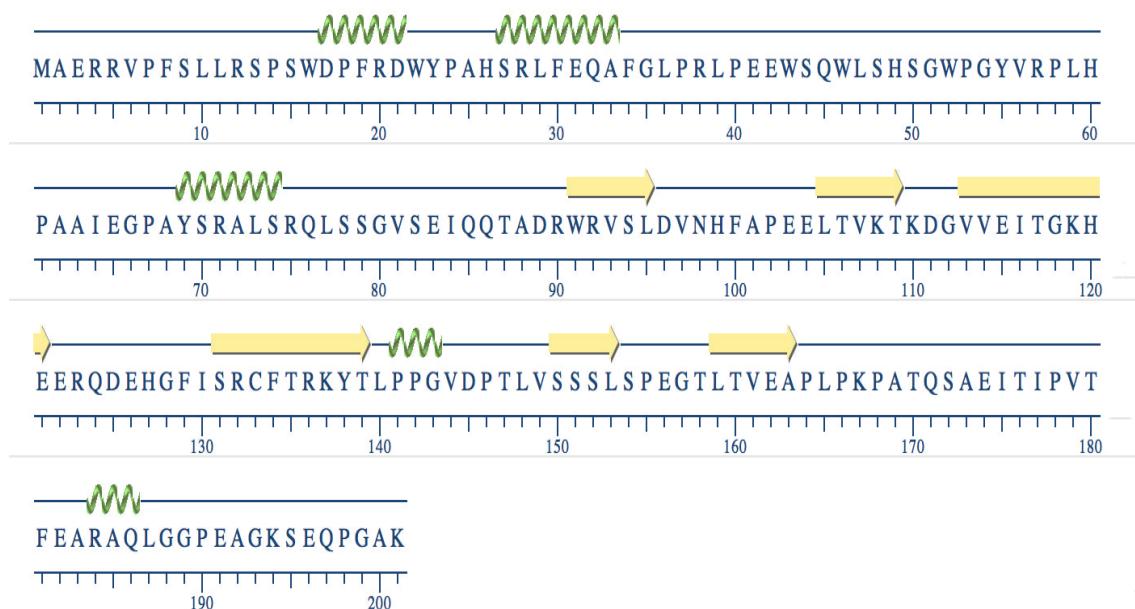
**Figure 14.** Phylogenetic tree shows the classification of Arabian Camel within sHSPB family.



**Figure 15.** Phylogenetic tree demonstrates the relationship of Camel HSPB-1 protein and protein sequences from other species. Maximum likelihood tree based on complete coding sequences. Values at nodes are bootstrapping  $\geq 49\%$  obtained from 1000 resampling of the data.

**143 2.6. Secondary and 3D Structures of Arabian Camel HSPB-1 protein**

**144** The primary structure of Arabian Camel HSPB-1 protein was used to predict its secondary  
**145** structure, which demonstrates the first level of protein folding. The predicted structure suggested  
**146** that Arabian Camel HSPB-1 protein composed of 5  $\alpha$ -helices and 6  $\beta$ -strands (Figure 16) in which the  
**147** 6  $\beta$ -strands forms a highly conserved ACD approximately 85-residues (from 88-173), which flanked  
**148** by less conserved NTD and CTE. The three-dimensional (3D) structure of this domain forms an  
**149** immunoglobulin-like  $\beta$ -sandwich fold in the C-terminal half of the Arabian Camel HSPB-1 protein  
**150** (Figure 17). The ACD domain mediates forming Arabian Camel HSPB-1 dimers via the anti-parallel  
**151** pairing of the same  $\beta$ -strand from two monomers.

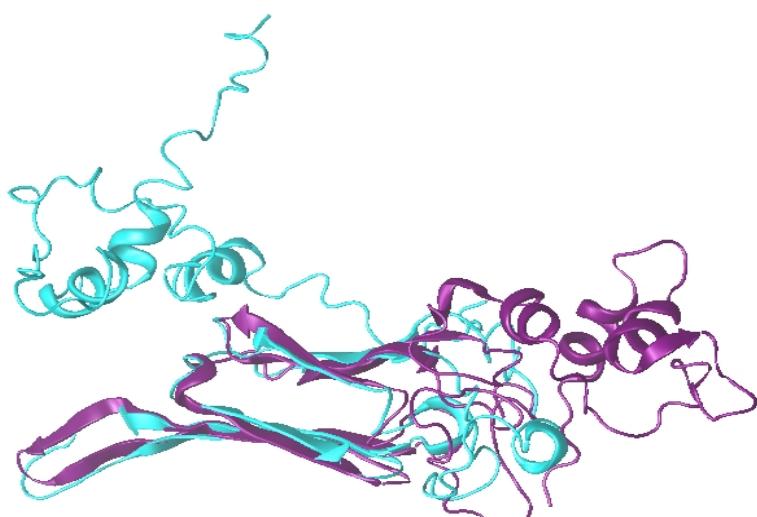


**Figure 16.** The secondary structure of Arabian Camel HSPB-1 protein.

**152** To construct 3D structural model of Arabian Camel HSPB-1 protein, we generated a homology  
**153** model of Arabian Camel HSPB-1 with Phyre2 server ([http://www.sbg.bio.ic.ac.uk/phyre2/html/  
\*\*154\*\* page.cgi?id=index](http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index)) (Figure 17). In this study, we employed Homo sapiens  $\alpha$ -B-crystallin chain V (PDB  
**155** ID: 2YGD) [27] as a template in which 86% of amino acids residues were modelled at > 90% confidence.  
**156** The 3D structural model consisted of 5  $\alpha$ -helices and 6  $\beta$ -strands. The ACD region is folded into a  
**157** compact of 6-anti-parallel-stranded formed two  $\beta$ -sheets. It has very similar fold and topology as those  
**158** from human. the structural similarity of Arabian Camel HSPB-1 with Human HSPB-1 was examined  
**159** by superimposing their structures using the Pymol program (<https://www.pymol.org/>) (Figure 18).  
**160** The RMSD (root mean square deviation) between Arabian Camel HSPB-1 and Human  $\alpha$ -B-crystallin  
**161** chain V structures was 3.134. The Q-score is another crucial parameter to assess the similarity of the  
**162** homologous structures which represents the quality of recognition and superimposition was 0.8435,  
**163** indicating high structural identity. The Z- and P-scores of the 3D structure of Arabian Camel HSPB1  
**164** and Human HSPB1 were 44.6 and 65.9, respectively.

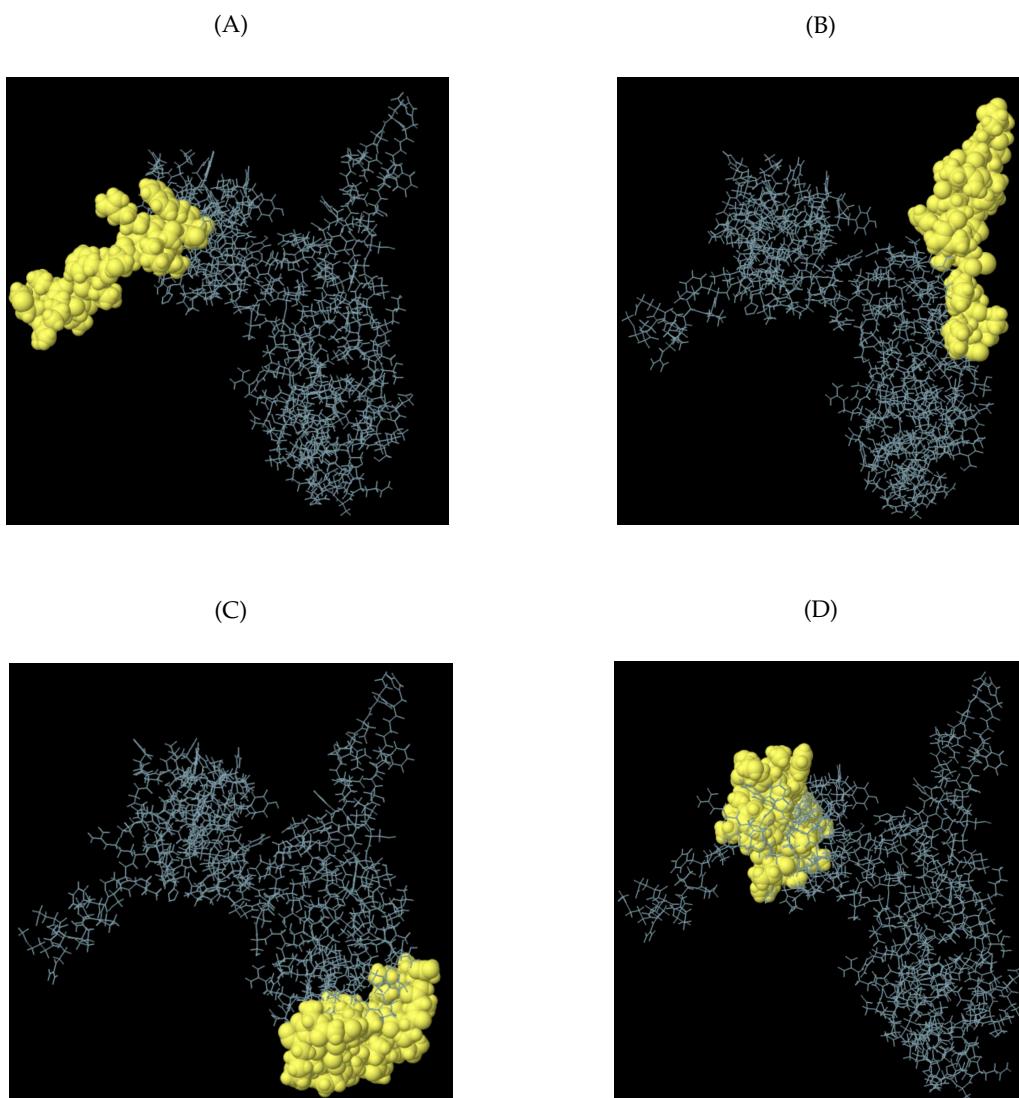


**Figure 17.** The 3D structure of Arabian Camel HSPB-1 protein.



**Figure 18.** Stereo ribbon representation of the predicted 3D structure model of Arabian Camel HSPB-1 (cyan) and the superimposition with Homo sapiens  $\alpha$ B-crystallin chain V (purple).

165 The epitope regions of Arabian Camel HSPB-1 protein based on its 3D structure were predicted  
166 using Ellipro server (<http://tools.iedb.org/ellipro/>) four discontinuous peptides were identified  
167 having score value  $> 0.7$ . The highest probability of a discontinuous epitope was computed as 78.5%.  
168 Amino Acids involved in discontinuous epitopes, their sequence location, number of amino acids and  
169 scores are given in (Table 7), while their positions on 3D structure of Arabian Camel HSPB-1 protein  
170 are illustrated in (Figure 19).



**Figure 19.** 3D representation of discontinuous epitopes (A to D) of Arabian Camel HSPB-1. The epitopes are shown in yellow surface and bulk of Arabian Camel HSPB-1 protein is shown in grey sticks.

**Table 7.** Predicted discontinuous antigenic epitopes of Arabian Camel HSPB-1 protein.

Start	End	Peptide	Peptide Length	Score	3D Structure
1	17	MAERRVPFSLLRSPSWD	17	0.785	A
119	137	KHEERQDEHGFISRCFTRK	19	0.753	B
166	201	PKPATQSAEITIPVTFEARALGGPEAGKSEQGAK	36	0.72	C
33	58	AFGLPRLPEEWSQWLSHSGWPGYVRP	26	0.706	D

### 171 3. Discussion

172 Small heat shock proteins (sHSPs) aid to keep protein homostasis by interplaying with unfolded  
 173 substrates to prevent cellular damage [10]. The ATP-independent chaperone HSPB-1 protein is a  
 174 typical example. The HSPB-1 protein is expressed in several tissue types, in stress-inducible conditions,  
 175 where it serves as a chaperone for partly folded cellular proteins. Thus, a full structural description of  
 176 Arabian Camel HSPB-1 gene is an important step toward understanding its mode of action.

The molecular characterization of Arabian Camel HSPB-1 gene is crucial for realizing the effect of exposure to different environmental factors on the health position of this animal. The present study aims to focus on the molecular characterization of small heat shock protein family mainly, HSPB-1 protein from *C. dromedarius*. We cloned Arabian Camel HSPB-1 cDNA(791 bp) using specific primers span the whole open reading frame; encoding 201 amino acids for protein with size of 22.382 kDa; that highly matches with several HSPB-1 protein sequences from other species (Table 1). Arabian Camel HSPB-1 cDNA sequence has been matched with other 13 mammalian HSPB-1 sequences in GenBank and submitted in NCBI database with the accession number MF278354. Our findings suggest that Arabian Camel HSPB-1 mRNA is highly expressed in lung followed by roughly equally expressed in testis, liver and kidney, and less expressed in brain and spleen as indicated by qRT-PCR. analysis (Figure 2).

There are two highly conserved domains in Arabian Camel HSPB-1 protein: ACD and IbpA domains that are localized near the C-terminal end of the protein. Multiple sequence alignment of Arabian Camel HSPB-1 protein (Figure 6) shows a highly conserved ACD approximately 85-residues (from 88-173) which flanked by less conserved NTD and CTE across species. The NTD is very diverse in protein sequence and therefore, largely responsible for the sequence variation of HSPB-1 protein between organisms. The amino acids such as Trp(W), Phe(F) and Pro(P) are overrepresented in the NTD. While the CTE involves the highly conserved IXI motif, which is thought to be important for inter-dimer contacts [28].

In general, mammalian HSPB-1 proteins exist as polydispersed oligomeric population and their full crystal structures have not been determined yet. Nevertheless, crystal structure of ACD of HSPB-1 protein indicates  $\beta$ -sheet rich immunoglobulin-like fold. The 3D structure of a protein supplies a useful understanding about its function. Comparative modeling is possible to predict the 3D structure of a protein based only on its primary structure. Therefore, we predicted a 3D structure of Arabian Camel HSPB-1 protein from which the amino acid sequence is known. The 3D structure of Arabian Camel ACD Domain forms an immunoglobulin-like  $\beta$ -sandwich fold in the C-terminal half of the Arabian Camel HSPB-1 protein (Figure 17). The ACD mediates forming Arabian Camel HSPB-1 dimers via the anti-parallel pairing of the same  $\beta$ -strand from two monomers.

## 4. Materials and Methods

### 4.1. Sample collection

Six different Camel tissue samples, including brain, lung, liver, kidney, testis and spleen, were obtained from male Arabian Camel slaughtered at the main slaughter-house located in Southern Riyadh. Tissue samples to be used for RNA analysis were instantly immersed in RNAlater<sup>®</sup> RNA Stabilization reagent (Qiagen, Ambion, Inc, USA) to prevent RNA degradation. The samples were then stored at -80°C until further use. While those other samples tissues to be used for protein analysis were transported on ice to the laboratory.

### 4.2. Total RNA isolation and cDNA synthesis

Samples of 50 mg of each preserved tissues were subjected for RNA isolation. The tissues were homogenized in RTL lysis based buffer (Qiagen) containing 1% 2-mercaptoethanol using steel beads (Sigma) and Tissue Lyser II (Qiagen). Nanodrop spectrophotometer (NanoDrop, ThermoScientific) was used to quantify samples at 260 nm and the quality of RNA samples were evaluated using denaturing SYBR safe agarose gel 1% electrophoresis.  $\approx$  2  $\mu$ g were transcribed to first-stranded cDNA by ImProm-II Reverse Transcription System (Promega, USA).

### 4.3. Examining Gene Expression by PCR and qPCR

Gene-specific primers (Table 8) were designed based on the data from the Arabian Camel genome project (<http://camel.genomics.org.cn/page/camel/index.jsp>). The PCR reaction mixture was carried

out in a final volume of 25  $\mu$ l, containing 12.5  $\mu$ l 2X GoTaq<sup>®</sup> Green Master Mix(Promega, USA), 1  $\mu$ l of 5 pmol of each primer, 2  $\mu$ l of cDNA. The PCR condition was 1 cycle at 94°C for 5 min, followed by 30 cycles at 94°C for 5 sec, 60°C for 30 sce, 72°C for 45 sec. The final extension step was carried out at 72°C for 10 min. The PCR products were then examined on 1.2% a garose gel stained with SYBR safe. In addition, in order to evaluate the level of relative expression of Arabian Camel HSPB-1 mRNA, six different Arabian Camel tissues were examined by using Fluorescent Quantitative Real-Time PCR (qRT-PCR) Detector ViiA 7 Real-Time PCR System.  $\beta$ -Actin mRNA were used as a house keeping gene control. In this experiment, Fast SYBR<sup>®</sup> Green Master Mix kit was utilized and gene-specific primer pairs were designed to amplify 83 bp length of Arabian Camel HSPB-1. The qRT-PCR reaction mixture were 10  $\mu$ l of Fast SYBR<sup>®</sup> Green Master Mix (Cat. No. 4385612, Applied Biosystems), 1  $\mu$ l of the forward primer, 1  $\mu$ l of the reverse primer, 3  $\mu$ l of Nuclease-free water and 5  $\mu$ l of cDNA target with a total volume of 20  $\mu$ l. Thermal Cycling Parameters were initial denaturation at 95°C for 3 min, amplification of 40 cycles at 95°C for 3 s and 60°C for 40 s.

**Table 8.** Gene-specific primers, the underlined bases represent the restriction sites used for cloning.

Usage	Primer Name	Primer Sequence 5' → 3'	Length (bp)
ORF-PCR	Camel HSPB1-F	CACGGATCCGAGGCCACCATGGCCGAG	759
	Camel HSPB1-R	CAC <u>CGCGCCGCTCAGCCGGCAGGA</u> ACTTAGAACT	
	$\beta$ -Actin-F	GATATTGCTCGCTCGTGGT	1140
	$\beta$ -Actin-R	TGGAACGTAACTAAGTCGCCT	
cDNA-RACE	cDNA-RACE-5'	CTTGGTCTTGACCGTCAGCTCCTC	283
	cDNA-RACE-3'	AGATCACCATCCCTGTCACCTTCGA	382
qRT-PCR	Camel HSPB1-qF	GTGTCGGAGATCCAGCAGAC	83
	Camel HSPB1-qR	TTCGTGCTTGCAGTGATCT	
	$\beta$ -Actin-qF	CCCATTGAGCATGGCATCGT	190
	$\beta$ -Actin-qR	GTAGATGGCACAGTGTGAG	

#### 4.4. Cloning and Sequencing of Arabian Camel HSPB-1 cDNA

Rapid-amplification of cDNA ends (RACE) was used to identify and isolate the 5'- and 3'-end of Arabian Camel HSPB-1 using RACE kits (Invitrogen, Carlsbad, CA, USA). Total RNA was annealed with 5'- and 3'-end primers (Table 8), and reversely transcribed respectively to respective 5'- and 3'-cDNA. The resulting first-stranded 5'- and 3'-cDNA were then utilized as templates in PCR. The cycling program was set for five cycles of 95 °C, 4 min; 5 cycles of 95 °C 15 s, 70 °C 15 s, 72 °C 3 min; 5 cycles of 95°C 15 s, 68°C 15 s, 72°C 3 min; 5 cycles of 95°C 15 s, 65°C 15 s, 72°C 3 min; 25 cycles of 95°C 15 s, 60°C 15 s, 72°C 3 min; 1 cycle of 72°C, 5 min. The purified nested PCR product was ligated into pcDNA5/FRT/TO GFP-tagged vector (a gift from Harm Kampinga ; Addgene plasmid No. 19487) [29], using BamHI (NEB R3136S) and NotI (NEB R3189S) restriction sites. Subsequently, 5  $\mu$ l of the ligation mixture was used as a template to transform chemically modified DH5 $\alpha$  competent cells (ThermoFisher Scientific). The cloned Arabian Camel HSPB-1 was sequenced using Applied Biosystems 3730xl DNA Analyzer platform (Applied Biosystems, Foster City, USA). The conditions of the Chain Termination PCR were: one cycle at 94°C for 35 s followed by 25 cycles at 94°C for 40 s, 50°C for 35 s and 60°C for 1 min.

#### 4.5. Protein extraction and quantitation

proteins from brain, testis, kidney, liver, lung and the spleen were extracted using RIPA lysis buffer. 4 mg from each tissue were homogenized in 4 ml of RIPA lysis buffer containing (5 M NaCl, 0.5 M EDTA, 1 M Tris-cl, NP-40, 10% soudium deoxycholate and 10% SDS) using steel beads (Sigma) and a Tissue Lyser II (Qiagen). lysates were then centrifuged at 14,000 rpm for 1 hour at 4°C. The supernatant fractions were then collected and total protein quantitation of each tissue was determined using BCA assay.

258 4.6. Arabian Camel HSPB-1 protein identification by LC-MS

259 25 µg of Arabian Camel protein lysates were subjected to one-dimensional sodium dodecyl sulfate  
260 polyacrylamide gel (1-D SDS-Page) using 4% stacking and 15% resolving polyacrylamide gels (1 mm  
261 thickness gel) running for 120 min. The 1-D SDS-Page then was consequently stained for overnight in  
262 a solution containing the mixture of Coomassie R-240, 40% methanol, 10% acetic acid. The gel was  
263 then destained in a solution containing 30% methanol and 10% acetic acid.

264 The excised band gel piece holding proteins with molecular weight of approximately 20-25 kDa  
265 was cut into cubes and incubated for 45 min in 300 µl of 1:1 mixture of 100 mM ammonium bicarbonate  
266 buffer with 50% acetonitrile and was vortexed for 10 min the supernatant was then discarded. The  
267 procedure was repeated until the stain was completely removed. 10 mM dithiothreitol (DTT) in 100  
268 mM ammonium bicarbonate buffer was added to the gel cubes in order to reduce Disulfide bonds and  
269 was incubated for 30 min at 56°C in an air thermostat. After rinsing in 100 µl of acetonitrile, 200 µl of  
270 50 mM iodoacetamide solution were added and incubated for 20 min at room temperature in the dark.

271 The gel cubes were then dehydrated twice with 100% acetonitrile for 10 min each then dried in a  
272 speed-vac for 10 min in order to make them ready for tryptic digestion. Trypsin (10ng/ul) solution  
273 was added to the dried gel cubes just enough to cover the gel cubes and incubated for 10 min at room  
274 temperature. Subsequently, 100 mM ammonium bicarbonate buffer was added until the gel cubes  
275 were immersed which was then incubated at 37°C for overnight. The digestion was then stopped  
276 by adding 20 µl of 5% formic acids. The digest solution (extracted peptides) was then transferred to a  
277 clear autosampler vial.

278 Millipore® Ziptips C18 pipette (Tip size:P10, Merck KGaA, Darmstadt, Germany) was used to  
279 prepare sample for Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass  
280 spectrometry. The Ziptip pipette was washed with 100% Methanol followed by 0.1% trifluoroacetic  
281 acid (TFA) solution. The tryptic-cleaved peptides mixture were then loaded onto the Ziptip  
282 pipette and then were desalted by 0.1% TFA. The loaded peptides were then eluted in 10 µl of  
283 β-Cyano-4-hydroxycinnamic acid which was used as matrix. 1 µl of aliquots were generally sampled  
284 directly from the digest supernatant for MS fingerprint analysis using Axima Performance® MALDI  
285 TOF/TOF Mass Spectrometer (Shimadzu Corporation, UK). The data were searched using the  
286 MASCOT search engine (<http://www.matrixscience.com>).

287 4.7. Phylogenetic tree

288 Using Arabian Camel HSPB-1 protein sequence, as a query we retrieved 40 small heat shock  
289 proteins sequences from NCBI Protein Database. The α-crystalline domain was verified in all the  
290 retrieved protein sequences using InterProScan [30] at (<https://www.ebi.ac.uk/interpro/>) (Table 6). To  
291 verify Arabian Camel HSPB-1 protein is distinctly related to the HSPB-1 proteins family we retrieved  
292 40 heat shock protein orthologues, which are conspicuously related to ten well-known small heat shock  
293 protein families known as HSPB-1-HSPB-10. To ensure the consistency of sampling all sHSPs proteins  
294 orthologues were retrieved from the same species. We used Arabian Camel HSPB-1 protein sequence,  
295 as a query to search the NCBI Protein Database for identified HSPB-1 proteins across diverse vertebrate  
296 species. Another sets of sHSP members were sampled from the same mammalian species to ensure the  
297 consistency of sampling. The accession numbers of protein members under study are listed in (Table 6).  
298 Consequently, the full length amino acid sequences, including Arabian Camel HSPB-1, were selected  
299 for multiple alignment using CLUSTALX 2.1 program [31]. A bootstrap re-sampling technique was  
300 utilized to ensure the robustness of the generated topological tree. Neighbor Joining (NJ) phylogenetic  
301 analysis was conducted in MEGA 7.0 [32]. Constructed topological trees were depicted and edited  
302 using FigTree v1.4.3. (<http://tree.bio.ed.ac.uk/software/figtree/>).

**303 4.8. Structure modeling**

**304** A secondary structure of Camel HSPB-1 protein sequence was generated using Geneious software  
**305** v10.0.3 [33]. Consequently, a 3D structure of Camel HSPB-1 protein containing 201 residues was  
**306** predicted after submitting the protein sequence to Phyre2 server (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>). The similarities between modeled Camel HSPB-1 and Human  
**307** HSPB-1 structure (PDB:2YGD) were superimposed using Pymol software. The quality of the  
**308** superimposed 3D structures was assessed using PDBe on (<https://swissmodel.expasy.org/interactive>).  
**309** The antigenicity, hydrophobicity and flexibility of camel HSPB-1 protein was predicted according to  
**310** the methods of Kolaskar, Parker and Karplus, respectively [21,22].  
**311**

**312** **Acknowledgments:** This work was supported by National Centre for Biotechnology, Life science and Environment  
**313** Research Institute, King Abdulaziz City for Science and Technology.

**314** **Author Contributions:** M.M.M., S.N.A. and A.T.A. conceived and designed the experiments; A.T.A., W.M.A.,  
**315** M.A.A., M.N.A., B.M.A and S.N.A performed the experiments; M.M.M. and A.T.A. analyzed the data; M.M.M.,  
**316** S.N.A. and A.T.A. contributed reagents/materials/analysis tools; M.M.M. and S.N.A wrote the paper.

**317** **Conflicts of Interest:** The authors declare that there is no conflict of interest for this article and there is no financial  
**318** employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received  
**319** or pending, royalties related to this manuscript.

**320 Abbreviations**

**321** The following abbreviations are used in this manuscript:

<b>322</b> sHSPs	Small heat shock proteins
HSPB-1	Small heat shock protein beta-1
5'-UTR	5'-untranslated region
3'-UTR	3'-untranslated region
ORF	Open Reading Frame
<b>323</b> qRT-PCR	quantitative Real-Time PCR
PMF-MS	Piptide mass fingerprint mass spectrometry
ACD	$\alpha$ -crystallin domain
NTD	N-terminal domain
CTE	C-terminal extension
1-D SDS-Page	one-dimensional sodium dodecyl sulfate polyacrylamide gel

**324 References**

- 325** 1. Fowler, M.E. *Medicine and surgery of camelids*; John Wiley & Sons, 2011.
- 326** 2. Monari, M.; Foschi, J.; Rosmini, R.; Marin, M.; Serrazanetti, G. Heat shock protein 70 response to physical  
**327** and chemical stress in Chamelea gallina. *Journal of Experimental Marine Biology and Ecology* **2011**, *397*, 71 –  
**328** 78.
- 329** 3. Wu, H.; Guang, X.; Al-Fageeh, M.; Cao, J.; Pan, S.; Zhou, H.; Zhang, L.; Abutarboush, M.; Xing, Y.; Xie, Z.;  
**330** Alshanqeeti, A.; Zhang, Y.; Yao, Q.; Al-Shomrani, B.; Zhang, D.; Li, J.; Manee, M.; Yang, Z.; Yang, L.; Liu, Y.;  
**331** Zhang, J.; Altammami, M.; Wang, S.; Yu, L.; Zhang, W.; Liu, S.; Ba, L.; Liu, C.; Yang, X.; Meng, F.; Wang,  
**332** S.; Li, L.; Li, E.; Li, X.; Wu, K.; Zhang, S.; Wang, J.; Yin, Y.; Yang, H.; Al-Swailem, A.; Wang, J. Camelid  
**333** genomes reveal evolution and adaptation to desert environments. *Nature Communications* **2014**, *5*.
- 334** 4. Saeed, H.; Shalaby, M.; Embaby, A.; Ismael, M.; Pathan, A.; Ataya, F.; Alanazi, M.; Bassiouny, K. The Arabian  
**335** camel *Camelus dromedarius* heat shock protein 90 $\alpha$ : cDNA cloning, characterization and expression.  
*International Journal of Biological Macromolecules* **2015**, *81*, 195–204.
- 336** 5. Elrobb, M.S.; Alanazi, M.S.; Khan, W.; Abduljaleel, Z.; Al-Amri, A.; Bazzi, M.D. Molecular Cloning  
**337** and Characterization of cDNA Encoding a Putative Stress-Induced Heat-Shock Protein from *Camelus*  
**338** *dromedarius*. *International Journal of Molecular Sciences* **2011**, *12*, 4214–4236.
- 339** 6. Xu, Q.; Schett, G.; Li, C.; Hu, Y.; Wick, G. Mechanical stress-induced heat shock protein 70 expression  
**340** in vascular smooth muscle cells is regulated by Rac and Ras small G proteins but not mitogen-activated  
**341** protein kinases. *Circulation Research* **2000**, *86*, 1122–1128.

- 343 7. Bakthisaran, R.; Tangirala, R.; Rao, C.M. Small heat shock proteins: Role in cellular functions and pathology. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2015**, *1854*, 291–319.
- 344 8. Waters, E.R.; Lee, G.J.; Vierling, E. Evolution, structure and function of the small heat shock proteins in plants. *Journal of Experimental Botany* **1996**, *47*, 325–338.
- 345 9. de Jong, W.W.; Leunissen, J.A.; Voorter, C.E. Evolution of the alpha-crystallin/small heat-shock protein family. *Molecular biology and evolution* **1993**, *10*, 103–126.
- 346 10. Haslbeck, M.; Franzmann, T.; Weinfurtner, D.; Buchner, J. Some like it hot: the structure and function of small heat-shock proteins. *Nature structural & molecular biology* **2005**, *12*, 842–846.
- 347 11. Kappé, G.; Franck, E.; Verschuur, P.; Boelens, W.C.; Leunissen, J.A.M.; de Jong, W.W. The human genome encodes 10  $\alpha$ -crystallin-related small heat shock proteins: HspB1–10. *Cell Stress & Chaperones* **2003**, *8*, 53.
- 348 12. Caspers, G.J.; Leunissen, J.A.; de Jong, W.W. The expanding small heat-shock protein family, and structure predictions of the conserved “ $\alpha$ -crystallin domain”. *Journal of molecular evolution* **1995**, *40*, 238–248.
- 349 13. de Jong, W.W.; Caspers, G.J.; Leunissen, J.A. Genealogy of the  $\alpha$ -crystallin—small heat-shock protein superfamily. *International journal of biological macromolecules* **1998**, *22*, 151–162.
- 350 14. Fu, X.; Jiao, W.; Chang, Z. Phylogenetic and biochemical studies reveal a potential evolutionary origin of small heat shock proteins of animals from bacterial class A. *Journal of molecular evolution* **2006**, *62*, 257–266.
- 351 15. Sudnitsyna, M.V.; Mymrikov, E.V.; Seit-Nebi, A.S.; Gusev, N.B. The role of intrinsically disordered regions in the structure and functioning of small heat shock proteins. *Current protein & peptide science* **2012**, *13*, 76–85.
- 352 16. Zhang, K.; Ezemaduka, A.N.; Wang, Z.; Hu, H.; Shi, X.; Liu, C.; Lu, X.; Fu, X.; Chang, Z.; Yin, C.C. A Novel Mechanism for Small Heat Shock Proteins to Function as Molecular Chaperones. *Scientific Reports* **2015**, *5*, 8811.
- 353 17. Bepperling, A.; Alte, F.; Kriehuber, T.; Braun, N.; Weinkauf, S.; Groll, M.; Haslbeck, M.; Buchner, J. Alternative bacterial two-component small heat shock protein systems. *Proceedings of the National Academy of Sciences* **2012**, *109*, 20407–20412.
- 354 18. Larkin, M.; Blackshields, G.; Brown, N.; Chenna, R.; McGettigan, P.; McWilliam, H.; Valentin, F.; Wallace, I.; Wilm, A.; Lopez, R.; Thompson, J.; Gibson, T.; Higgins, D. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947.
- 355 19. Kozlowski, L.P. IPC-Isoelectric Point Calculator. *Biology direct* **2016**, *11*, 55.
- 356 20. Kriehuber, T.; Rattei, T.; Weinmaier, T.; Bepperling, A.; Haslbeck, M.; Buchner, J. Independent evolution of the core domain and its flanking sequences in small heat shock proteins. *The FASEB Journal* **2010**, *24*, 3633–3642.
- 357 21. Karplus, P.; Schulz, G. Prediction of chain flexibility in proteins. *Naturwissenschaften* **1985**, *72*, 212–213.
- 358 22. Kolaskar, A.; Tongaonkar, P.C. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS letters* **1990**, *276*, 172–174.
- 359 23. Parker, J.; Guo, D.; Hodges, R. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* **1986**, *25*, 5425–5432.
- 360 24. Emini, E.A.; Hughes, J.V.; Perlow, D.; Boger, J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of virology* **1985**, *55*, 836–839.
- 361 25. Chow, P.; Fasman, G. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol* **1978**, *47*, 45–148.
- 362 26. Linding, R.; Jensen, L.J.; Diella, F.; Bork, P.; Gibson, T.J.; Russell, R.B. Protein disorder prediction: implications for structural proteomics. *Structure* **2003**, *11*, 1453–1459.
- 363 27. Braun, N.; Zacharias, M.; Peschek, J.; Kastenmüller, A.; Zou, J.; Hanzlik, M.; Haslbeck, M.; Rappsilber, J.; Buchner, J.; Weinkauf, S. Multiple molecular architectures of the eye lens chaperone  $\alpha$ B-crystallin elucidated by a triple hybrid approach. *Proceedings of the National Academy of Sciences* **2011**, *108*, 20491–20496.
- 364 28. Hilton, G.R.; Lioe, H.; Stengel, F.; Baldwin, A.J.; Benesch, J.L. Small heat-shock proteins: paramedics of the cell. In *Molecular Chaperones*; Springer, 2012; pp. 69–98.
- 365 29. Hageman, J.; Kampinga, H.H. Computational analysis of the human HSPH/HSPA/DNAJ family and cloning of a human HSPH/HSPA/DNAJ expression library. *Cell Stress & Chaperones* **2009**, *14*, 1–21.

- 394 30. Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.;  
395 Nuka, G.; Pesreat, S.; Quinn, A.F.; Sangrador-Vegas, A.; Scheremetjew, M.; Yong, S.Y.; Lopez, R.; Hunter, S.  
396 InterProScan 5: genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240.
- 397 31. Thompson, J. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment  
398 aided by quality analysis tools. *Nucleic Acids Research* **1997**, *25*, 4876–4882.
- 399 32. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for  
400 Bigger Datasets. *Molecular biology and evolution* **2016**, *33*, 1870–1874.
- 401 33. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.;  
402 Markowitz, S.; Duran, C.; Thierer, T.; Ashton, B.; Meintjes, P.; Drummond, A. Geneious Basic: an integrated  
403 and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*  
404 **2012**, *28*, 1647–1649.

405 **Sample Availability:** Samples of the compounds ..... are available from the authors.

406 © 2017 by the authors. Submitted to *Int. J. Mol. Sci.* for possible open access  
407 publication under the terms and conditions of the Creative Commons Attribution (CC BY) license  
408 (<http://creativecommons.org/licenses/by/4.0/>).