Int J Physiol Pathophysiol Pharmacol 2024;16(6):138-151
www.ijppp.org /ISSN:1944-8171/IJPPP0161908

Original Article

Bioinformatics analysis and alternative polyadenylation in Heat Shock Proteins 70 (HSP70) family members

Srishti Shriya, Ramakrushna Paul, Neha Singh, Farhat Afza, Buddhi Prakash Jain

Gene Expression and Signaling Lab, Department of Zoology, Mahatma Gandhi Central University Motihari, Motihari, Bihar 845401, India

Received November 11, 2024; Accepted December 21, 2024; Epub December 25, 2024; Published December 30, 2024

Abstract: Objective: The Heat Shock Protein 70 (HSP70) family is a highly conserved group of molecular chaperones essential for maintaining cellular homeostasis. These proteins are necessary for protein folding, assembly, and degradation and involve cell recovery from stress conditions. HSP70 proteins are upregulated in response to heat shock, oxidative stress, and pathogenic infections. Their primary role is preventing protein aggregation, refolding misfolded proteins, and targeted degradation of irreparably damaged proteins. Given their involvement in fundamental cellular processes and stress responses, HSP70 proteins are critical for cell survival and modulating disease outcomes in cancer, neurodegeneration, and other pathologies. The present study aims to understand domain architecture, physicochemical properties, phosphorylation, ubiquitination, and alternative polyadenylation site prediction in various HSP70 members. Method: SMART and InterProScan software were used for domain analysis. EXPASY Protparam, NetPhos 3.1 server DTU, and MUbisiDa were used for physicochemical analysis, phosphorylation, and ubiquitination site analysis, respectively. Alternative polyadenylation was studied using the EST database. Result: Domain analysis shows that coiled-coil and nucleotide-binding domains are present in some of the HSP70 members. Five HSP70 family members have alternate polyadenylation sites in their 3'UTR. Conclusion: The present work has provided valuable insights into their structure, functions, interactome, and polyadenylation patterns. Studying their therapeutic potential in diseases like cancer can be helpful.

Keywords: Heat Shock Protein, chaperones, protein folding, domain, alternative polyadenylation

Introduction

Protein folding occurs when polypeptide chains fold into their specific form, which is stable and functional; this folding occurs by forming a three-dimensional structure. The folding process of polypeptides occurs inside cells in the endoplasmic reticulum [1-3]. The endoplasmic reticulum has several types of machinery by which it folds properly and packages the protein, then sends it to the Golgi apparatus, called the warehouse of the cell. Protein production occurs on the polyribosome complex in the cytosol and bound ribosome on the endoplasmic reticulum surface. In the ribosome complex, when a newly synthesized protein is forming, a signal recognition particle (SRP) is attached to the end of the newly synthesized protein [4-7]. On the surface of ER, there is an SRP receptor, which binds with SRP, and this receptor generates a signal and transfers the polyribosome complex to the translocation channel [5, 8]. The translocation channel allows polyribosomes to come inside the lumen. Inside the lumen of the Endoplasmic Reticulum (ER), several chaperon proteins can help fold the nascent polypeptide into a complex one [9, 10].

The Heat Shock Protein 70 (HSP70) family is a highly conserved group of molecular chaperones essential for maintaining cellular homeostasis [11-14]. These proteins play critical roles in protein folding, assembly, and degradation, thereby mitigating cellular stress and facilitating recovery from various stress conditions [15, 16]. In humans, the HSP70 family comprises several members, each with distinct and overlapping functions. These include HSPA1A, HSPA1B, HSPA2, HSPA4, and others, each encoded by different genes but sharing a com-

https://doi.org/10.62347/CWPE7813

mon structural and functional framework. All HSP70 proteins share a conserved domain structure consisting of an N-terminal ATPase domain and a C-terminal substrate-binding domain. The ATPase domain binds and hydrolyzes ATP, which drives the conformational changes necessary for protein folding [17-20]. The substrate-binding domain interacts with unfolded or misfolded proteins to facilitate proper folding or degradation. HSP70s are upregulated in response to heat shock, oxidative stress, and pathogenic infections. In silico models have revealed that HSP70 proteins undergo significant conformational changes during their functional cycle. The transition between the open and closed states of the substrate-binding domain is crucial for their chaperone activity. Molecular dynamics simulations have provided insights into the dynamics of these conformational changes and their implications for protein-substrate interactions [21-24].

Their primary role involves the prevention of protein aggregation, refolding of misfolded proteins, and targeted degradation of irreparably damaged proteins [14, 20, 22]. Phylogenetic studies using sequence alignment and evolutionary tree construction have elucidated the evolutionary relationships among human HSP70 family members. These analyses suggest that the HSP70 family has undergone significant evolutionary divergence, leading to the specialization of different family members in distinct cellular functions [25]. HSP70 proteins could be targeted for therapeutic interventions in various diseases. For instance, inhibitors or modulators of HSP70 activity could be explored as potential treatments for excessive protein aggregation or impaired protein homeostasis [26-29].

The structural and functional insights gained from in silico analyses have significant implications for drug design. Small molecules that target specific domains or conformational states of HSP70 proteins could be developed to modulate their activity [30, 31]. This approach has potential applications in treating diseases related to protein misfolding and aggregation, such as neurodegenerative disorders and cancer [32-36]. The role of various HSP70 proteins has also been studied in multiple metabolic disorders like diabetes and hypertension [37-39].

Given their involvement in fundamental cellular processes and stress responses, HSP70 proteins are critical for cell survival and modulating disease outcomes in cancer, neurodegeneration, and other pathologies. In the present work, *insilico* analyses like domain architecture, physicochemical, phosphorylation, ubiquitination site prediction, and alternative polyadenylation analysis were performed in various HSP70 family members.

Materials and methods

The Heat Shock Protein 70 (HSP70) genes were analyzed using various parameters. For this, data was collected from different software and databases.

Retrieval of sequence of human HSP70 family members

cDNA and protein sequence of different HSP70 family members were retrieved from NCBI (National Centre for Biotechnology Information) and UniProt (Universal Protein Resources) database. Their interacting partners were studied in the UniProt database [40, 41] and STRING (Search tool for retrieval of interacting genes/proteins) tool [42, 43].

Domain analysis by SMART and InterProScan

SMART stands for simple modular architecture research tool [44, 45]. This database is used to identify and analyze different protein domains with the help of their sequence. InterPro scan is a bioinformatics database used for functional analysis of protein sequences and the presence of domains at specific sites [46, 47].

Physico-chemical analysis

EXPASY Protparam is a tool that provides different physical and chemical parameters of protein on a computational basis. The parameters include molecular weight, PI (Isoelectric point), amino acid composition, atoms composition, half-life, aliphatic index, etc. [48, 49].

Phosphorylation and ubiquitination site prediction

NetPhos 3.1 server DTU is a software that efficiently predicts eukaryotes' serine, threonine, and tyrosine phosphorylation sites by searching their gene names. NetPhos 3.1 server DTU detects the phosphorylation site of different HSP70 members. It also predicts their specific position, i.e., at which position phosphorylation

occurs in serine, threonine, and tyrosine residue [50, 51].

The MUbisiDa database provides comprehensive information about the ubiquitination site of different proteins in mammals. The ubiquitination site in different HSP70 members, their amino acid number, and their ubiquitination position were found in this database [52].

Alternative polyadenylation analysis by EST

EST database, a library of short reads generated from mRNA or cDNA, was used for insilico analysis of alternate 3'UTR sequence at 3' end of human *HSP70* family members. The genes with multiple polyA signals and cleavage sites generate transcripts with alternate 3' end, which the EST database can predict. The sequence of the longest possible 3'UTR is blasted in the EST database against Homo sapiens with high similarity. If multiple 3'UTRs exist, then multiple transcripts end at the same position in the graphical representation. The presence of shorter 3'UTR can be confirmed by taking the sequence of shorter UTR, and multiple adenine nucleotides are added at its 3' end and blasted in the EST database. If it shows similarity with the additional adenine nucleotides, it confirms the presence of transcript with this shorter 3'UTR in the EST library.

Results

HSP70 family members

HSP70 family proteins play a crucial role in folding and stabilizing the proteins. In different databases like NCBI and UniProt, 14 members of the human HSP70 family (**Table 1**) have been reported with diverse biological roles. HSPA1A, HSPA1B, and HSPA1L have the same amino acid number (641), while HSPA1A and HSPA1B have the same molecular weight (70052 Dalton). The analysis of the Interacting partner of HSP70 (by STRING and UniProt) provides information about how many different genes interact with HSP70 members and are responsible for their involvement in various biological pathways. The gene symbol, functions, and interacting partners of various HSP70 family members are represented in **Table 1**.

Domain architecture of HSP70 family proteins

HSP70 members have three functional domains: One N-terminal nucleotide-binding domain having ATPase activity, a middle domain that is protease sensitive, and another C-terminal substrate binding domain, which binds with substrate molecule [14, 53]. The Domain analysis of HSP70 members was performed through SMART and InterProScan databases. Two main domains, i.e., Coiled-coil and NBD domain, were present in different HSP70 members but only in some. HSPA1A, HSPA1B, HSPA6, and HSPA8 have only Coiled-coil domains. Both Coiled-coil and NBD domains are present in HSPA4 and HSPA5. HSPA14 has only an NBD domain. The domain architecture of HSP70 members is represented in **Figure 1**.

Characterization of physicochemical analysis of HSP70 family proteins

The physicochemical analysis of proteins is important to understand their functions because, based on these properties, it determines whether the protein can interact with another molecule [54, 55]. The physiochemical analysis of HSP70 was performed through the EXPASY Protparam tool, which provides a comprehensive physiochemical property like molecular weight, number of amino acids by which protein is formed, their Isoelectric point (PI), negative and positive charged residue, their aliphatic index, etc. A detailed analysis of the physicochemical properties of HSP70 members is given in **Table 2**.

Prediction of phosphorylation and ubiquitination sites on HSP70 proteins

Phosphorylation and ubiquitination play significant roles in the regulation and functions of proteins [56, 57]. Phosphorylation confers protein stability and is also responsible for conformational changes in protein by regulating its catalytic activity. Phosphorylation occurs at Serine (S), Threonine (T), and Tyrosine (Y) residues. Phosphorylation sites in various HSP70 members at S, T, and Y residues are represented in **Table 3**. Ubiquitination occurs mainly at Lysine (K) residue. It is a posttranslational modification and affects protein degradation, cellular localization, and protein-protein interactions. The ubiquitination site determination in HSP70 members was performed with MuBiSiDa, and it was observed that there are various Lysine (K) residues present in the

Table 1. HSP70 family members

S. No.	Gene Symbol	Function	Domain	Interaction/Interacting partner
01	*HSPA1A*	Stabilizes existing proteins against aggregation and mediates the folding of newly translated proteins in the cytosol and organelles. Involved in the ubiquitin-proteasome pathway through interaction with the AU-rich element RNA-binding protein 1. It is also present in the primary histocompatibility complex class III region.	Coiled-coil	TERT, TRIM5, METTL21A, DNAAF2, PRKN, FOXP3, NOD2, DNAJC9, ATF5, RNF207, HSF1, NAA10, HSP40, HSP90 and HDAC4, NEDD1 (via NBD) with BAG1, BAG2, BAG3 and HSPH1/HSP105, SMAD3, DNAJC8, NLRP12
02	*HSPA1B*	Stabilizes proteins against aggregation and mediates the folding of newly translated proteins in the cytosol and organelles. Involved in the ubiquitin-proteasome pathway through interaction with the AU-rich element RNA-binding protein 1. Also present in the major histocompatibility complex class III region.	Coiled-coil	CHCHD3, DNAJC7, IRAK1BP1, PPP5C and TSC2, TERT, TRIM5, METTL21A, PRKN, FOXP3, NOD2, DNAJC9, ATF5, NAA10, HSP40, HSP90 and HDAC4, NEDD1, BAG1, BAG2, BAG3 and HSPH1/HSP105, SMAD3, DNAJC8
03	*HSPA1L*	Stabilizes existing proteins against aggregation and mediates the folding of newly translated proteins in the cytosol and organelles. The gene is located in the major histocompatibility complex class III region.		HS71L, BAG4, TRI38, NFKB1, P53
04	*HSPA2*	Protection of the proteome from stress, folding, and transport of newly synthesized polypeptides, protein quality control system, ensuring the correct folding of proteins, the re-folding of misfolded proteins, and controlling the targeting of proteins for subsequent degradation. Plays a role in spermatogenesis. In association with SHCBP1L, it may participate in maintaining spindle integrity during meiosis in male germ cells.		HSPBP1, GRPEL1, S100A1, HSPA9, ABHD15, GAKHSPH1, HSP90AA1, DNAJB12, THSD4, POX2, FAM73B, STIP1, HSF2, DNAJC6, HSP90AB1, HSPA1A, KLHL34, TSSK1B, RNF34, VWA2, HSPA2, ILK, THAP4, DNAJB14, HSPA8
05	*HSPA4*	Ubiquitous expression in testis, esophagus, and other organs.	NBD, Coiled-coil	HSP74, APBP2, ATRAP, PO6F2, HSP7C, EGFR
06	*HSPA5*	Plays a role in the folding and assembly of proteins in the ER. It interacts with the transmembrane stress sensor proteins PERK, IRE1, and ATF6 during stress or infection. It acts as a repressor of the unfolded protein response (UPR) and plays a role in cellular apoptosis and senescence.	NBD, Coiled-coil	TMEM132A and TRIM21, DNAJC10, DNAJB9/ERdj4, ERN1/IRE1, MX1, METTL23, CEMIP, PCSK4, CIPC, CCDC88B, INPP5K, MANF, LOXL2, CLU, CCDC47, CLN3, KIAA1324, CASP7
07	*HSPA6*	Ensuring the correct folding of proteins, the re-folding of misfolded proteins, and controlling the targeting of proteins for subsequent degradation.	Coiled-coil	HSP76, RFA1, BAG4, SAHH2, CV015-2, FLNA-2, KT222, COMD6, PRAP1, TERF1, LEG7, PPIB
08	*HSPA7 (pseudogene)*	ATPase activity, binding with misfolded protein, and ubiquitin protein ligase binding site help in vesicle-mediated transport.		
09	*HSPA8*	Binds to nascent polypeptides to facilitate correct folding. Functions as an ATPase in the disassembly of clathrin-coated vesicles during transport of membrane components through the cell.	Coiled-coil	
10	*HSPA9*	Function in cell proliferation, stress response, and maintenance of the mitochondria.		GRP75, GRPE1, P53, P73, NELFB, HSC20, TCTP, EGFR, A4D2J0
11	*HSPA12A*	Broad expression in the brain and 16 other tissues.		SORL, HS12B
12	*HSPA12B*	It may be involved in susceptibility to atherosclerosis. Alternative splicing results in multiple transcript variants encoding different isoforms.		TNS1, STUB1, TRAP1, HSPA12B, MIB1, HSF4, DNAJC10, GRPEL1, S100A1, LALBA, TNS3, GAK, TENC1, GRPEL2, HSP90AA1, BAG3, DNAJC6, HSP90AB1, PTEN, DNAJC1, SUGT1, PODXL, DNAJC2, TPTE2, MIB2, TPTE
13	*HSPA13*	Role in the processing of cytosolic and secretory proteins, in removing denatured or incorrectly-folded proteins.		HSP13, SGTA, UBQL1, Q24JT5, UBQL2, SGTB
14	*HSPA14*	A component of the ribosome-associated complex (RAC) is a complex involved in folding or maintaining nascent polypeptides in a folding-competent state.	NBD	NUP37, HSPA14, DNAJC2, DNAJB6, DNAJB1, HSPA12A, HSPA12B, BAG3, HSPA4, HSPH1, HSPA4L

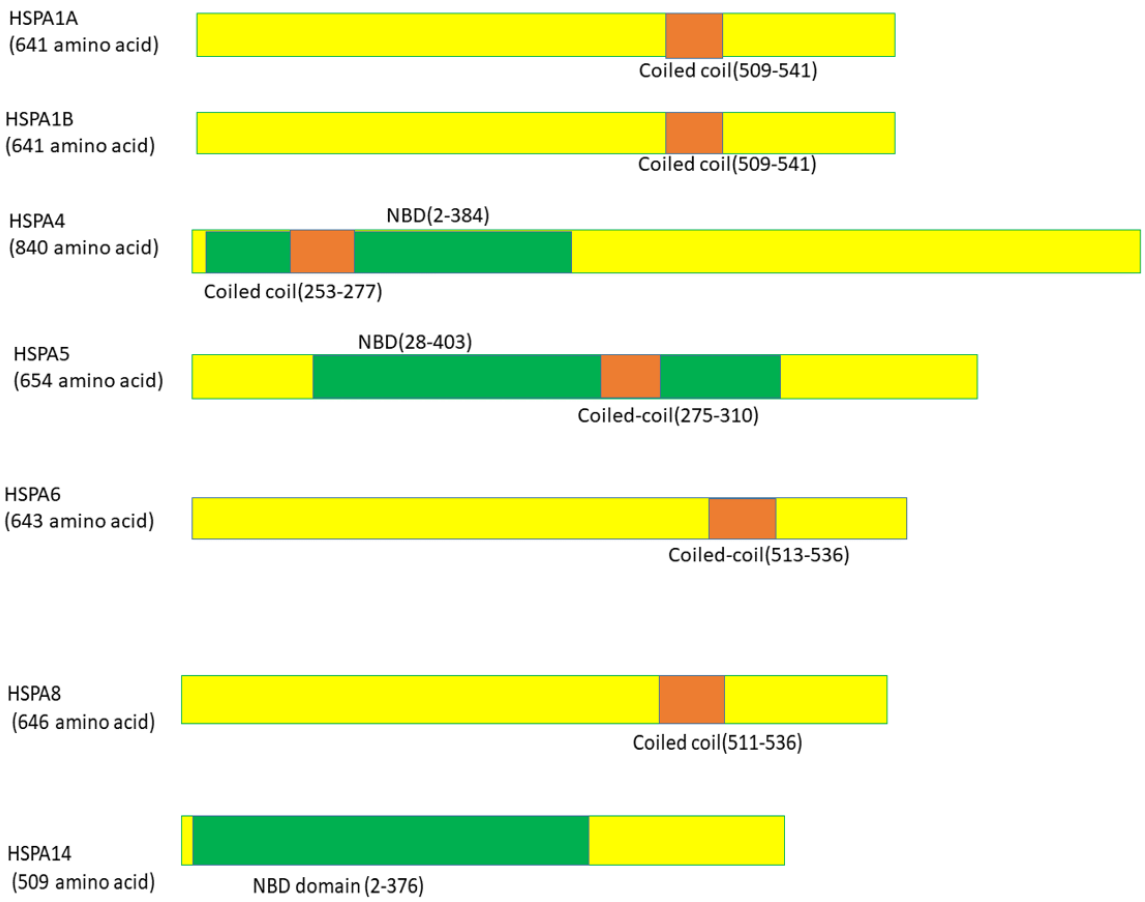


Figure 1. Domain architecture of Human HSP70 family proteins.

Table 2. Physicochemical properties of Human HSP70 members

Gene	MW	Amino acid	PI	Positive charge residue	Negative charge residue	Aliphatic index
HSPA1A	70052.23	641	5.47	(Arg + Lys): 81	(Asp + Glu): 92	85.23
HSPA1B	70052.23	641	5.47	(Arg + Lys): 81	(Asp + Glu): 92	85.23
HSPA1L	70375.05	641	5.75	(Arg + Lys): 82	(Asp + Glu): 90	87.96
HSPA2	70020.97	639	5.55	(Arg + Lys): 84	(Asp + Glu): 94	81.82
HSPA4	94330.92	840	5.10	(Arg + Lys): 110	(Asp + Glu): 137	74.86
HSPA5	72332.96	654	5.07	(Arg + Lys): 89	(Asp + Glu): 111	85.70
HSPA6	71028.14	643	5.81	(Arg + Lys): 84	(Asp + Glu): 94	80.86
HSPA7	40244.45	367	7.72	(Arg + Lys): 48	(Asp + Glu): 47	78.45
HSPA8	70898.09	646	5.37	(Arg + Lys): 82	(Asp + Glu): 95	81.52
HSPA9	73680.50	679	5.87	(Arg + Lys): 88	(Asp + Glu): 95	82.64
HSPA12A	74978.38	675	6.32	(Arg + Lys): 78	(Asp + Glu): 83	84.81
HSPA12B	75687.56	686	8.81	(Arg + Lys): 83	(Asp + Glu): 75	83.63
HSPA13	51927.46	471	5.52	(Arg + Lys): 47	(Asp + Glu): 57	99.30
HSPA14	54794.43	509	5.41	(Arg + Lys): 54	(Asp + Glu): 65	96.01

sequence of different HSP70 members, which are the sites of ubiquitination. Ubiquitination sites in various HSP70 members are summarized in **Table 4**.

Table 3. Phosphorylation sites in various HSP70 members at S, T, and Y residues

Gene Name	Serine (S)	Threonine (T)	Tyrosine (Y)
HSPA1A	16, 40, 85, 106, 153, 254, 276, 277, 281, 286, 296, 307, 312, 340, 362, 400, 418, 494, 511, 537, 544, 551, 563, 579, 631	13, 38, 45, 47, 66, 111, 125, 145, 158, 211, 265, 273, 275, 278, 298, 313, 341, 425, 430, 450, 491, 495, 502, 585, 636	15, 41, 294, 371, 431, 525
HSPA1B	16, 40, 85, 106, 153, 254, 275, 276, 277, 281, 286, 296, 307, 312, 340, 362, 400, 418, 494, 511, 537, 544, 551, 563, 579, 631	13, 38, 45, 47, 66, 111, 125, 145, 158, 211, 265, 273, 278, 298, 313, 341, 425, 430, 450, 491, 495, 502, 585, 636	15, 41, 294, 371, 431, 525
HSPA1L	18, 42, 108, 155, 241, 256, 277, 278, 279, 288, 298, 387, 402, 420, 513, 546, 556, 565, 567, 581	15, 40, 47, 49, 68, 127, 147, 160, 213, 267, 275, 280, 300, 315, 343, 364, 427, 432, 452, 493, 497, 504, 618, 624, 636	17, 43, 290, 296, 342, 373, 433, 527, 626
HSPA2	17, 41, 97, 154, 242, 278, 279, 280, 284, 289, 299, 343, 365, 403, 497, 514, 530, 547, 622, 630	14, 39, 46, 48, 65, 67, 112, 126, 146, 159, 164, 214, 268, 276, 281, 301, 316, 344, 414, 421, 428, 430, 433, 453, 498, 505, 555, 634	16, 42, 291, 297, 374, 434, 528, 550, 595, 614
HSPA4	31, 40, 47, 54, 58, 76, 123, 131, 155, 258, 267, 287, 298, 323, 355, 384, 393, 403, 408, 414, 415, 448, 463, 471, 474, 475, 476, 491, 496, 503, 546, 552, 556, 575, 633, 641, 647, 692, 714, 715, 737, 756, 765, 777, 784, 828, 830	35, 63, 99, 104, 114, 149, 160, 225, 230, 300, 328, 364, 433, 468, 512, 538, 576, 649, 655, 732, 738, 776, 824	14, 89, 148, 336, 446, 454, 597, 624, 626, 660, 671, 723
HSPA5	4, 40, 64, 86, 300, 301, 311, 319, 354, 365, 406, 448, 452, 448, 452, 567, 571, 587, 588, 607, 632, 637	37, 69, 91, 124, 137, 151, 156, 166, 171, 184, 274, 321, 323, 338, 366, 428, 441, 445, 462, 473, 481, 518, 525, 527, 534, 561, 593, 643, 648	39, 65, 127, 396, 568
HSPA6	18, 42, 59, 87, 98, 127, 155, 213, 256, 277, 278, 279, 288, 298, 309, 314, 342, 364, 402, 489, 496, 513, 543, 554, 559, 634	15, 40, 47, 49, 68, 83, 115, 132, 147, 160, 267, 275, 280, 300, 315, 343, 414, 427, 429, 432, 452, 493, 497, 504, 626	17, 43, 296, 373, 433, 527, 594
HSPA7	18, 42, 59, 87, 98, 127, 155, 213, 256, 277, 278, 279, 288, 298, 309, 314	15, 40, 47, 49, 68, 83, 115, 132, 147, 150, 160, 267, 275, 280, 300, 315	17, 43, 296
HSPA8	16, 40, 85, 113, 121, 153, 275, 276, 277, 281, 286, 296, 329, 340, 362, 381, 385, 400, 489, 494, 511, 537, 538, 541, 544, 613, 633, 637, 638	13, 38, 45, 47, 64, 66, 111, 125, 140, 145, 158, 211, 265, 273, 278, 298, 313, 341, 418, 425, 427, 430, 450, 495, 502, 552, 641	15, 41, 288, 294, 371, 431, 525, 545, 611
HSPA9	3, 29, 34, 42, 48, 89, 148, 162, 164, 200, 253, 320, 321, 322, 337, 376, 378, 408, 444, 469, 473, 533, 550, 554, 627, 638, 639, 644, 657, 662, 664, 665, 667	22, 62, 87, 111, 116, 120, 177, 185, 192, 205, 294, 347, 356, 362, 449, 462, 466, 504, 539, 592, 594	46, 118, 568, 652
HSPA12A	9, 19, 23, 27, 37, 39, 47, 50, 68, 73, 190, 195, 226, 243, 258, 259, 267, 289, 303, 316, 348, 388, 408, 420, 428, 436, 437, 444, 458, 533, 587, 594, 605, 611, 632, 663, 665	18, 34, 45, 66, 99, 101, 106, 146, 150, 202, 277, 291, 320, 324, 474, 513, 536, 589, 614, 623, 631, 644, 651, 662	21, 122, 311, 339, 346, 360
HSPA12B	16, 17, 21, 25, 29, 44, 46, 58, 60, 72, 77, 80, 149, 192, 194, 230, 247, 262, 276, 279, 280, 291, 294, 347, 409, 413, 433, 434, 442, 449, 492, 539, 583, 593, 631, 672, 676, 680	32, 42, 70, 103, 110, 121, 200, 206, 296, 339, 375, 397, 428, 542, 662, 673	126, 316, 344, 351
HSPA13	10, 44, 69, 72, 84, 113, 136, 143, 185, 249, 280, 323, 324, 334, 335, 343, 350, 354, 357, 366, 429	6, 30, 40, 41, 52, 95, 108, 135, 274, 312, 372, 379, 428	42, 79, 97, 146, 162, 272
HSPA14	42, 55, 61, 74, 75, 122, 130, 174, 186, 202, 204, 206, 219, 241, 246, 264, 271, 273, 277, 284, 305, 338, 389, 407, 418, 435, 451, 487, 504	35, 63, 104, 108, 127, 188, 234, 274, 413, 420, 481, 491, 493, 497	444

Analysis of alternative polyadenylation in 3'UTR of human HSP70 mRNAs

Alternative polyadenylation is a process of generating transcripts with different 3'UTR ends using different polyadenylation sites. The transcript with alternate 3'UTR exhibits different stability and translation efficiency and possesses varying target sites for miRNA and cis-regulatory elements [58, 59]. Alternative polyadenylation (APA) in the different *HSP70* family members' mRNA was analyzed by the EST (Expressed Sequence Tag) database. Among 14 genes (As mentioned in **Table 1**) in this study, only five genes (HSPA1L, HSPA4, HSPA5, HSPA6, and HSPA9) showed the presence of multiple PAS and CS. It was observed that HSPA1L, HSPA4, HSPA5, and HSPA6 have two polyA sites, and HSPA9 has three polyA sites.

HSPA1L generates a transcript with two different lengths of 3'UTR viz: 320 bp and 420 bp (**Figure 2A**). A sequence of shorter UTR, i.e., 320 bp, was taken, and polyA tail was added at 3' end before EST blast. If adenine nucleotides attached to the sequence of shorter UTR show similarity in the EST blast, it confirms that shorter 3'UTR also exists in the EST library. **Figure 2B** depicts that the polyA tail also shows similarity, and the alignments are represented in **Figure 2C**. Further, for the HSP4 gene, a transcript with two different 3'UTR (1622 bp and 2000 bp) was confirmed in the EST database (**Figure 3**). In EST blast analysis with full-length 3'UTR of HSPA5, two putative polyadenylation sites were found at 353 bp and 1738 bp (**Figure 4A**). EST analysis with shorter UTR with and addition of polyA tail (like above) was performed, which shows similarity with shorter UTR and extended polyA tail (**Figure 4B**, **4C**). Another HSP70 member, HSPA6, has two alternate 3'UTRs of 220 bp and 300 bp (**Figure 5**). Interestingly, HSPA9 has three putative polyadenylation sites and generates transcripts with three different lengths of 3'UTR viz: 702 bp, 1155 bp, and 2100 bp. Both the shorter (702 bp) UTR and medium size UTR (1155 bp) have similarities when polyA tail is added at their 3' end (**Figure 6**).

Thus, five members of the HSP70 family have alternate 3'UTR as per the EST database. Their further validation and functional significance in different cellular contexts must be studied.

Discussion

Protein folding is an essential process inside the cell that makes the protein functionally active, and for this, many proteins are required, including molecular chaperones. The molecular chaperone helps properly fold the protein and prevents disaggregation [4, 9]. Changes in physiological conditions inside the cells create a stressful environment, which may alter the protein's overall confirmation and be responsible for the diseased state. The heat shock protein helps protect the cell and allows proper protein folding [60, 61]. In Humans majorly, five classes of Heat shock proteins are found, which are HSP40 (DNAJ), HSP60 (human chaperonin), HSP70, HSP90 (HSPC), and HSP110 (name of HSP given according to their molecular weight) [29, 61]. These HSP classes are involved in various biological pathways for the proper functioning of protein and proteostasis. The HSP70 and HSP40 mediates the appropriate folding of nascent polypeptides and inhibits disaggregation. The HSP70 family has fourteen members (HSPA1A-HSPA14). We have performed an Insilico analysis of HSP70 family members using different bioinformatical tools and databases. These databases and tools help analyze HSP-70 members in various aspects. The coiled-coil domain and NBD domain are present in some of the members of the HSP70 family. The coiled-coil domain, mainly involved in protein-protein interaction, might be responsible for the interaction of HSP70 members with other proteins in different cellular contexts and perform various functions, including protein folding [62]. Analysis of the physicochemical characteristics of the HSP70 members helps to study their aliphatic index, isoelectric point, etc., and can be helpful in studying the interaction and involvement of HSP70 proteins. Phosphorylation and ubiquitination of the proteins determine stability and ensure the interactome of the protein. Prediction of phosphorylation and ubiquitination sites will be helpful in studying HSP70 members in different cellular contexts.

It is known that genetic information is encoded in the DNA, which is transcribed in the form of mRNA and further translated up to the protein level. Newly synthesized pre-mRNA undergoes several steps to become a mature mRNA. These steps are mainly splicing, capping, and polyadenylation, and they are translated to the

Table 4. Ubiquitination sites in various HSP70 members

S. No.	Gene Name	Ubiquitination site	Number of amino acids
1.	HSPA1A and HSPA1B	K102; K550; K71; K559; K108; K348; K597; K628; K112; K126; K56; K361; K88; K423; K328	641
2.	HSPA1A and HSPA1B	K102; K550; K71; K559; K108; K348; K597; K628; K112; K126; K56; K361; K88; K423; K328	641
3.	HSPA1L	K425; K509; K528; K514; K161; K453; K363	641
4.	HSPA2	K510; K188; K351; K454; K109; K503; K360; K328	639
5.	HSPA4	K686; K609; K748; K679; K557; K194; K221; K697; K754; K573; K638; K388; K568; K517; K785; K719; K798; K704; K794; K437; K711; K53; K430; K126; K543; K124; K360	840
6.	HSPA5	K548; K341; K139; K371; K97; K524; K377; K465; K164; K119; K186; K327; K153; K114; K353; K82; K126; K574; K214; K602; K269	654
7.	HSPA6	K327; K509; K359; K321; K350; K453; K502	643
8.	HSPA7	K327; K321	367
9.	HSPA8	K137; K3; K71; K512; K583; K524; K597; K601; K246; K56; K88; K361; K539; K423; K328	646
10.	HSPA9	K653; K138; K368; K646; K121	679
11.	HSPA12A	No data for human	675
12.	HSPA12B	No data for human	686
13.	HSPA13	No data for human	471
14.	HSPA14	No data for human	509

HSPA1L:

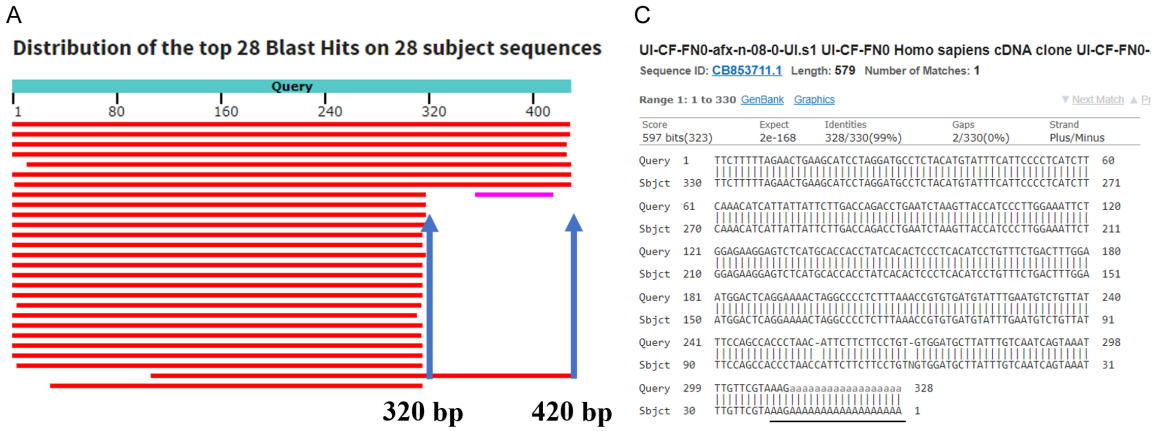


Figure 2. EST blast of human HSPA1L 3'UTR. Short and long 3'UTR (320 bp and 420 bp) are marked by blue arrow (A). EST blast of shorter 3'UTR (320 bp) of HSPA1L with additional adenine nucleotides. The similarity with the polyA tail is marked in box (B). Alignment of one EST constructs to show matching with polyA tail of shorter UTR of HSPA1L (C).

HSPA4:

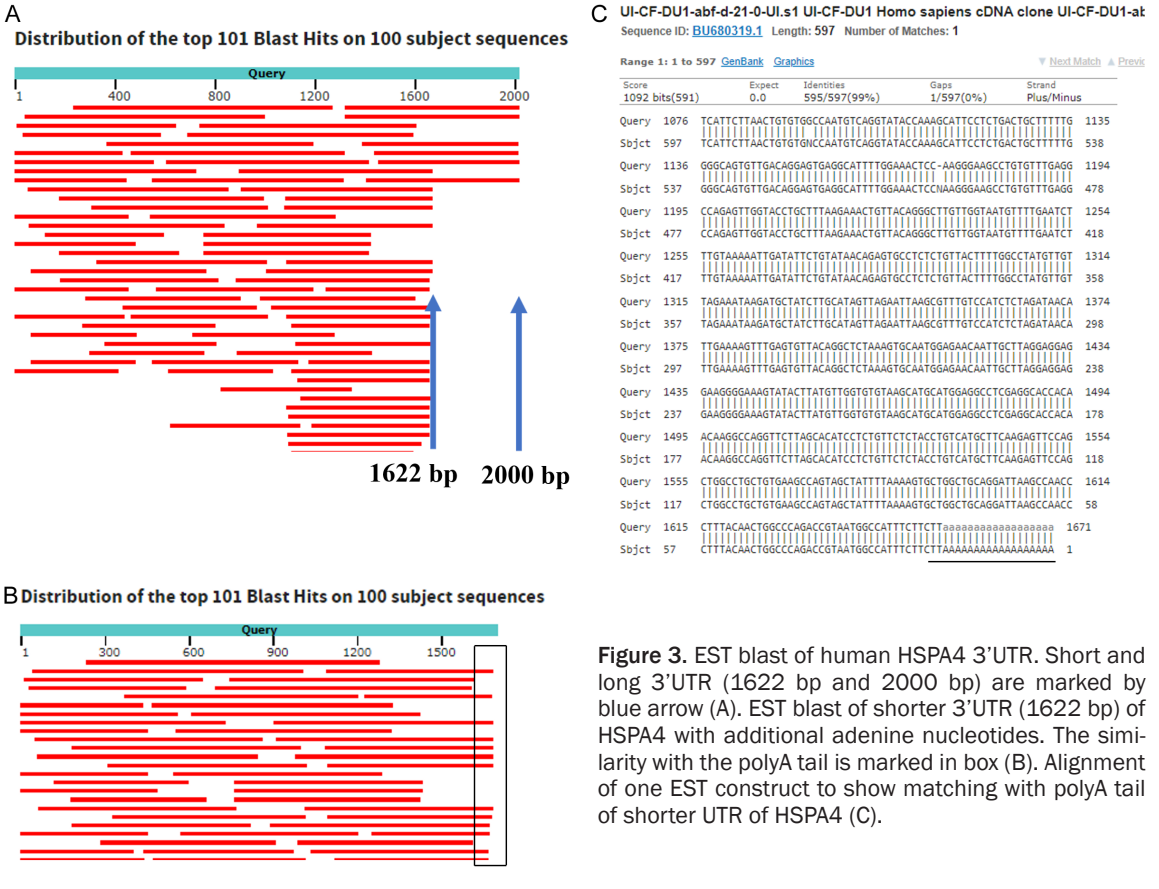
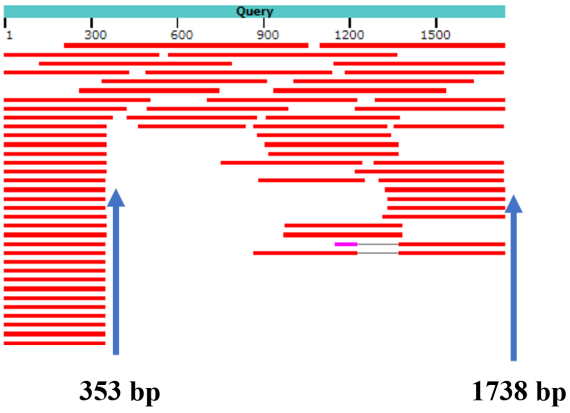


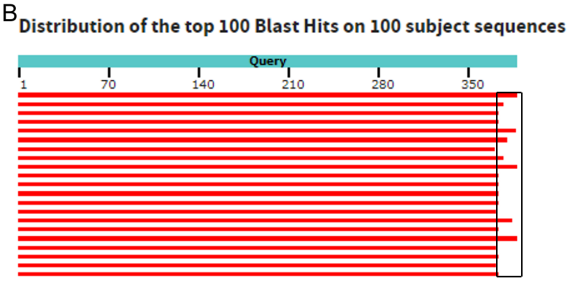
Figure 3. EST blast of human HSPA4 3'UTR. Short and long 3'UTR (1622 bp and 2000 bp) are marked by blue arrow (A). EST blast of shorter 3'UTR (1622 bp) of HSPA4 with additional adenine nucleotides. The similarity with the polyA tail is marked in box (B). Alignment of one EST construct to show matching with polyA tail of shorter UTR of HSPA4 (C).

HSPA5:

A



B



C

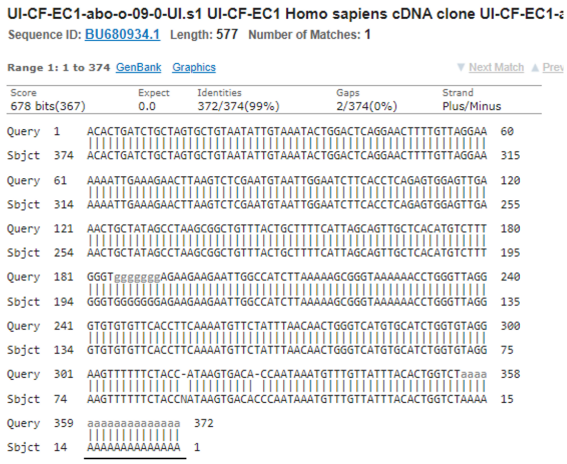
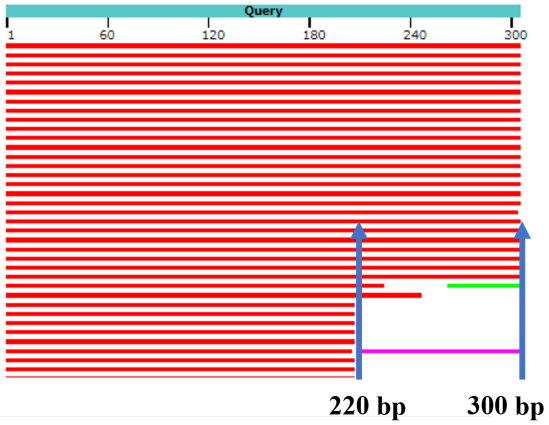


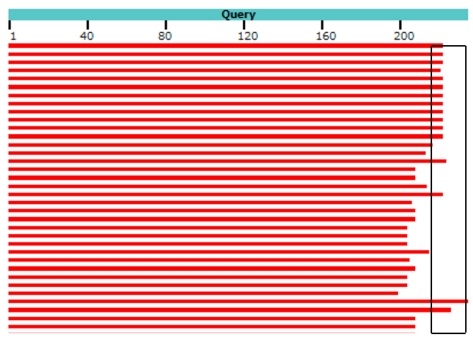
Figure 4. EST blast of human HSPA5 3'UTR. Short and long 3'UTR (353 bp and 1738 bp) are marked by blue arrow (A). EST blast of shorter 3'UTR (353 bp) of HSPA5 with additional adenine nucleotides. The similarity with the polyA tail is marked in box (B). Alignment of one EST construct to show matching with polyA tail of shorter UTR of HSPA5 (C).

HSPA6:

A



B



C

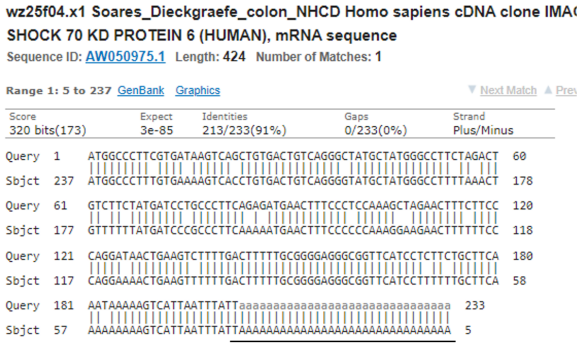


Figure 5. EST blast of human HSPA6 3'UTR. Short and long 3'UTR (220 bp and 300 bp) are marked by blue arrow (A). EST blast of shorter 3'UTR (220 bp) of HSPA6 with additional adenine nucleotides. The similarity with the polyA tail is marked in box (B). Alignment of one EST construct to show matching with polyA tail of shorter UTR of HSPA6 (C).

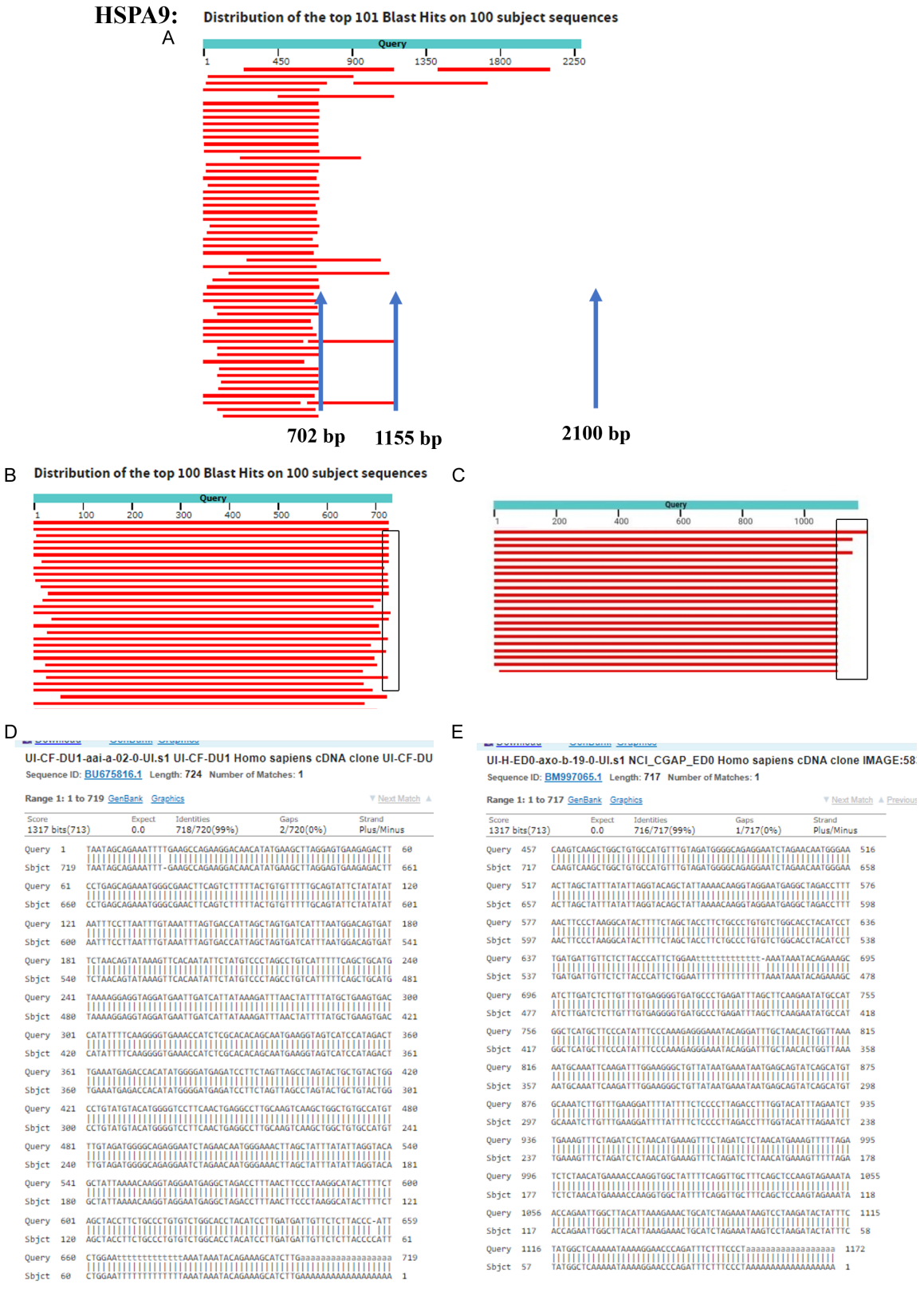


Figure 6. EST blast of human HSPA9 3'UTR. Short, middle, and long 3'UTR (702 bp, 1155 bp, and 2100 bp) are marked by blue arrow (A). EST blast of shorter 3'UTR (702 bp) of HSPA9 with additional adenine nucleotides. The similarity with the polyA tail is marked in box (B). EST blast of middle 3'UTR (1155 bp) of HSPA9 with additional adenine nucleotides. The similarity with the polyA tail is marked in box (C). Alignment of one EST construct to show matching with polyA tail of shorter and middle UTR of HSPA9 (D and E).

ribosome for protein synthesis [58, 63]. Capping is done at the 5' end of the transcript, and polyadenylation is done at the 3' end of the transcript by the addition of a stretch of adenosine residues (called poly(A) tail), which is essential for transcript stability, nuclear export, and translation. Heterogenous mRNA (HnRNA), generated by transcription, can form multiple transcripts with alternate 3' end by alternative polyadenylation. These transcripts with different 3'UTR lengths exhibit different stability and translation efficiency and have distinct binding sites for miRNA and various cis-regulatory elements [59]. Different pathophysiological conditions can affect the generation of the transcripts with alternate 3' end [64]. Five *HSP70* genes have more than one polyadenylation signal in their 3'UTR sequence. Thus, they can generate mRNA with different 3'UTR lengths. Further validation and studying the significance of this alternative polyadenylation of HSP70 members is a hot research topic.

Conclusion

HSP70 members possess various protein-protein interaction domains where the Coil-coil domain is prevalently present. This domain might be responsible for the interaction of HSP70 proteins with other proteins to perform proteostatic functions. Many phosphorylation and ubiquitination sites are present on HSP70 proteins, which might affect their location, stability, and functions during different cellular contexts. Five members of HSP70 show the presence of alternative polyadenylation signals, which may be responsible for their diverse functions and stability.

Overall, this bioinformatic analysis provides insights into the different cellular aspects and functions of HSP70 members, and further study might be useful to understand their role in various pathophysiological conditions.

Acknowledgements

We acknowledge Mahatma Gandhi Central University Motihari and IMS BHU for providing the necessary facilities for this work.

Disclosure of conflict of interest

None.

Address correspondence to: Buddhi Prakash Jain, Gene Expression and Signaling Lab, Department of Zoology, Mahatma Gandhi Central University Motihari, Motihari, Bihar 845401, India. E-mail: buddhiprakash@mgcub.ac.in

References

[1] Rosenzweig R, Nillegoda NB, Mayer MP and Bukau B. The Hsp70 chaperone network. Nat Rev Mol Cell Biol 2019; 20: 665-680.
[2] Nassar R, Dignon GL, Razban RM and Dill KA. The protein folding problem: the role of theory. J Mol Biol 2021; 433: 167126.
[3] Louros N, Schymkowitz J and Rousseau F. Mechanisms and pathology of protein misfolding and aggregation. Nat Rev Mol Cell Biol 2023; 24: 912-933.
[4] Acosta-Alvear D, Karagöz GE, Fröhlich F, Li H, Walther TC and Walter P. The unfolded protein response and endoplasmic reticulum protein targeting machineries converge on the stress sensor IRE1. Elife 2018; 7: e43036.
[5] Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walter P. The endoplasmic reticulum. Molecular Biology of the Cell. 4th edition. Garland Science; 2002.
[6] Jain BP. An overview of unfolded protein response signaling and its role in cancer. Cancer Biother Radiopharm 2017; 32: 275-281.
[7] Munro S and Pelham HR. A C-terminal signal prevents secretion of luminal ER proteins. Cell 1987; 48: 899-907.
[8] Neuhof A, Rolls MM, Jungnickel B, Kalies KU and Rapoport TA. Binding of signal recognition particle gives ribosome/nascent chain complexes a competitive advantage in endoplasmic reticulum membrane interaction. Mol Biol Cell 1998; 9: 103-115.
[9] Braakman I and Hebert DN. Protein folding in the endoplasmic reticulum. Cold Spring Harb Perspect Biol 2013; 5: a013201.
[10] Hartl FU, Bracher A and Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. Nature 2011; 475: 324-332.
[11] Dragovic Z, Broadley SA, Shomura Y, Bracher A and Hartl FU. Molecular chaperones of the Hsp110 family act as nucleotide exchange factors of Hsp70s. EMBO J 2006; 25: 2519-2528.
[12] Brocchieri L, Conway de Macario E and Macario AJ. hsp70 genes in the human genome: conservation and differentiation patterns predict a wide array of overlapping and specialized functions. BMC Evol Biol 2008; 8: 19.
[13] Ellis RJ. Molecular chaperones: assisting assembly in addition to folding. Trends Biochem Sci 2006; 31: 395-401.

[14] Radons J. The human HSP70 family of chaperones: where do we stand? Cell Stress Chaperones 2016; 21: 379-404.
[15] Kubelka J, Hofrichter J and Eaton WA. The protein folding "speed limit". Curr Opin Struct Biol 2004; 14: 76-88.
[16] Wruck F, Avellaneda MJ, Koers EJ, Minde DP, Mayer MP, Kramer G, Mashaghi A and Tans SJ. Protein folding mediated by trigger factor and Hsp70: new insights from single-molecule approaches. J Mol Biol 2018; 430: 438-449.
[17] De Maio A. Extracellular Hsp70: export and function. Curr Protein Pept Sci 2014; 15: 225-231.
[18] Flaherty KM, DeLuca-Flaherty C and McKay DB. Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. Nature 1990; 346: 623-628.
[19] Morano KA. New tricks for an old dog: the evolving world of Hsp70. Ann N Y Acad Sci 2007; 1113: 1-14.
[20] Tavaria M, Gabriele T, Kola I and Anderson RL. A hitchhiker's guide to the human Hsp70 family. Cell Stress Chaperones 1996; 1: 23-28.
[21] Alderson TR, Kim JH and Markley JL. Dynamical structures of Hsp70 and Hsp70-Hsp40 complexes. Structure 2016; 24: 1014-1030.
[22] Mayer MP and Bukau B. Hsp70 chaperones: cellular functions and molecular mechanism. Cell Mol Life Sci 2005; 62: 670-84.
[23] Nitika, Porter CM, Truman AW and Truttmann MC. Post-translational modifications of Hsp70 family proteins: expanding the chaperone code. J Biol Chem 2020; 295: 10689-10708.
[24] Woo HJ, Jiang J, Lafer EM and Sousa R. ATP-induced conformational changes in Hsp70: molecular dynamics and experimental validation of an in silico predicted conformation. Biochemistry 2009; 48: 11470-7.
[25] Yu EM, Yoshinaga T, Jalufka FL, Ehsan H, Mark Welch DB and Kaneko G. The complex evolution of the metazoan HSP70 gene family. Sci Rep 2021; 11: 17794.
[26] Davis AK, Pratt WB, Lieberman AP and Osawa Y. Targeting Hsp70 facilitated protein quality control for treatment of polyglutamine diseases. Cell Mol Life Sci 2020; 77: 977-996.
[27] Evans CG, Chang L and Gestwicki JE. Heat Shock Protein 70 (Hsp70) as an emerging drug target. J Med Chem 2010; 53: 4585-602.
[28] Mouawad N, Capasso G, Ruggeri E, Martinello L, Severin F, Visentin A, Facco M, Trentin L and Frezzato F. Is it still possible to think about HSP70 as a therapeutic target in onco-hematological diseases? Biomolecules 2023; 13: 604.
[29] Ritossa F. Discovery of the heat shock response. Cell Stress Chaperones 1996; 1: 97-98.
[30] Nitzsche B, Höpfner M and Biersack B. Synthetic small molecule modulators of Hsp70 and Hsp40 chaperones as promising anticancer agents. Int J Mol Sci 2023; 24: 4083.
[31] Singh H, Almaazmi SY, Dutta T, Keyzers RA and Blatch GL. In silico identification of modulators of J domain protein-Hsp70 interactions in Plasmodium falciparum: a drug repurposing strategy against malaria. Front Mol Biosci 2023; 10: 1158912.
[32] Gupta A, Bansal A and Hashimoto-Torii K. HSP70 and HSP90 in neurodegenerative diseases. Neurosci Lett 2020; 716: 134678.
[33] Zhao K, Zhou G, Liu Y, Zhang J, Chen Y, Liu L and Zhang G. HSP70 family in cancer: signaling mechanisms and therapeutic advances. Biomolecules 2023; 13: 601.
[34] Venediktov AA, Bushueva OY, Kudryavtseva VA, Kuzmin EA, Moiseeva AV, Baldycheva A, Meglinski I and Piavchenko GA. Closest horizons of Hsp70 engagement to manage neurodegeneration. Front Mol Neurosci 2023; 16: 1230436.
[35] Hu B, Liu G, Zhao K and Zhang G. Diversity of extracellular HSP70 in cancer: advancing from a molecular biomarker to a novel therapeutic target. Front Oncol 2024; 14: 1388999.
[36] Zhang M and Bi X. Heat Shock Proteins and breast cancer. Int J Mol Sci 2024; 25: 876.
[37] Mulyani WRW, Sanjiwani MID, Sandra, Prabawa IPY, Lestari AAW, Wihandani DM, Suastika K, Saraswati MR, Bhargah A and Manuaba IBAP. Chaperone-based therapeutic target innovation: Heat Shock Protein 70 (HSP70) for type 2 diabetes mellitus. Diabetes Metab Syndr Obes 2020; 13: 559-568.
[38] Singh MK, Shin Y, Ju S, Han S, Choe W, Yoon KS, Kim SS and Kang I. Heat Shock Response and Heat Shock Proteins: current understanding and future opportunities in human diseases. Int J Mol Sci 2024; 25: 4209.
[39] Rodriguez-Iturbe B. Environmental stress and hypertension: the disregarded role of HSP70. J Hum Hypertens 2024; 38: 538-541.
[40] Coudert E, Gehant S, de Castro E, Pozzato M, Baratin D, Neto T, Sigrist CJA, Redaschi N and Bridge A; UniProt Consortium. Annotation of biologically relevant ligands in UniProtKB using ChEBI. Bioinformatics 2023; 39: btac793.
[41] MacDougall A, Volynkin V, Saidi R, Poggioli D, Zellner H, Hatton-Ellis E, Joshi V, O'Donovan C, Orchard S, Auchincloss AH, Baratin D, Bolleman J, Coudert E, de Castro E, Hulo C, Masson P, Pedruzzi I, Rivoire C, Arighi C, Wang Q, Chen C, Huang H, Garavelli J, Vinayaka CR, Yeh LS, Natale DA, Laiho K, Martin MJ, Renaux A and Pichler K; UniProt Consortium. UniRule: a unified rule resource for automatic annotation in

the UniProt Knowledgebase. Bioinformatics 2020; 36: 4643-4648.
[42] Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, Bork P, Jensen LJ and von Mering C. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res 2023; 51: D638-D646.
[43] Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ and von Mering C. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res 2021; 49: D605-D612.
[44] Letunic I, Khedkar S and Bork P. SMART: recent updates, new developments and status in 2020. Nucleic Acids Res 2021; 49: D458-D460.
[45] Letunic I and Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res 2018; 46: D493-D496.
[46] Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, Richardson L, Salazar GA, Williams L, Bork P, Bridge A, Gough J, Haft DH, Letunic I, Marchler-Bauer A, Mi H, Natale DA, Necci M, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A and Finn RD. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res 2021; 49: D344-D354.
[47] Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R and Hunter S. InterProScan 5: genome-scale protein function classification. Bioinformatics 2014; 30: 1236-1240.
[48] Walker JM. The Proteomics Protocols Handbook. Totowa, NJ: Humana Press; 2005.
[49] Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD and Hochstrasser DF. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol 1999; 112: 531-552.
[50] Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S and Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics 2004; 4: 1633-1649.
[51] Blom N, Gammeltoft S and Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol 1999; 294: 1351-1362.
[52] Chen T, Zhou T, He B, Yu H, Guo X, Song X and Sha J. mUbiSiDa: a comprehensive database for protein ubiquitination sites in mammals. PLoS One 2014; 9: e85744.
[53] Rutledge BS, Choy WY and Duennwald ML. Folding or holding?-Hsp70 and Hsp90 chaperoning of misfolded proteins in neurodegenerative disease. J Biol Chem 2022; 298: 101905.
[54] Du X, Li Y, Xia YL, Ai SM, Liang J, Sang P, Ji XL and Liu SQ. Insights into protein-ligand interactions: mechanisms, models, and methods. Int J Mol Sci 2016; 17: 144.
[55] Yu L, Zhang Y, Gutman I, Shi Y and Dehmer M. Protein sequence comparison based on physicochemical properties and the position-feature energy matrix. Sci Rep 2017; 7: 46237.
[56] Ardito F, Giuliani M, Perrone D, Troiano G and Lo Muzio L. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). Int J Mol Med 2017; 40: 271-280.
[57] Deng L, Meng T, Chen L, Wei W and Wang P. The role of ubiquitination in tumorigenesis and targeted drug discovery. Signal Transduct Target Ther 2020; 5: 11.
[58] Jain BP. The role of alternative polyadenylation in cancer progression. Gene Rep 2018; 12: 1-8.
[59] Yuan F, Hankey W, Wagner EJ, Li W and Wang Q. Alternative polyadenylation of mRNA and its role in cancer. Genes Dis 2019; 8: 61-72.
[60] Hu C, Yang J, Qi Z, Wu H, Wang B, Zou F, Mei H, Liu J, Wang W and Liu Q. Heat shock proteins: biological functions, pathological roles, and therapeutic opportunities. MedComm (2020) 2022; 3: e161.
[61] Morimoto RI, Kroeger PE and Cotto JJ. The transcriptional regulation of heat shock genes: a plethora of heat shock factors and regulatory conditions. EXS 1996; 77: 139-163.
[62] Carroni M, Kummer E, Oguchi Y, Wendler P, Clare DK, Sinning I, Kopp J, Mogk A, Bukau B and Saibil HR. Head-to-tail interactions of the coiled-coil domains regulate ClpB activity and cooperation with Hsp70 in protein disaggregation. Elife 2014; 3: e02481.
[63] Corbett AH. Post-transcriptional regulation of gene expression and human disease. Curr Opin Cell Biol 2018; 52: 96-104.
[64] Zhang Y, Liu L, Qiu Q, Zhou Q, Ding J, Lu Y and Liu P. Alternative polyadenylation: methods, mechanism, function, and role in cancer. J Exp Clin Cancer Res 2021; 40: 51.