



Article

Comparative Analysis of the HSP70 Protein Family Across Vertebrates Reveals Evolutionary Conservation, Functional Divergence, and Structural Insights

My Abdelmajid Kassem 

Plant Genomics and Bioinformatics Lab, Department of Biological and Forensic Sciences, Fayetteville State University, Fayetteville, NC 28301, USA; mkassem@uncfsu.edu

Abstract

Heat shock proteins of the 70 kDa family (HSP70s) are essential molecular chaperones that preserve proteostasis by assisting protein folding, transport, and degradation. Although the core HSP70 architecture is deeply conserved, the degree and functional significance of sequence divergence across vertebrates remain incompletely understood. Here, an integrative comparative analysis of HSP70 proteins from ten representative vertebrate species spanning mammals, birds, amphibians, and teleost fish was performed. Multiple sequence alignment, phylogenetic reconstruction, motif discovery, entropy-based conservation profiling, hydrophobicity analysis, and structural mapping reveal a strikingly conserved ATPase domain alongside a more variable substrate-binding domain and C-terminal region. Multiple Expectation Maximization for Motif Elicitation (MEME) motif analysis identifies both universally conserved motifs and lineage-specific elements, highlighting functional constraint as well as adaptive diversification. Structural projection onto the human HSP70 crystal structure (PDB 5AQV) demonstrates that conserved hydrophobic residues cluster in the protein core, whereas variable residues are predominantly surface-exposed. Together, these findings illuminate how evolutionary pressures shape both the conserved chaperone machinery and flexible regulatory regions of HSP70, and they establish a scalable analytical framework for comparative protein evolution studies.

Keywords: HSP70; vertebrates; comparative genomics; protein evolution; structural conservation; entropy analysis; MEME motifs; molecular chaperones

1. Introduction

Heat Shock Proteins (HSPs) are highly conserved and ubiquitous proteins that play crucial roles in maintaining cellular proteostasis [1,2]. Originally identified as proteins induced by elevated temperatures, HSPs are now known to be constitutively expressed in most organisms and to respond to a wide range of physiological and environmental stresses, including heat shock, oxidative stress, hypoxia, heavy metals, and infection. Acting primarily as molecular chaperones, HSPs facilitate the correct folding of nascent polypeptides, prevent protein misfolding and aggregation, assist in refolding denatured proteins, and target irreversibly damaged proteins for degradation [3–6]. Through these functions, HSPs play a central role in maintaining protein quality control and are essential for cellular homeostasis, organismal fitness, and survival under stress conditions [7–9].

Among the HSP families, the HSP70 group represents one of the most extensively studied molecular chaperone systems. HSP70 proteins possess a conserved modular



Received: 26 January 2026

Revised: 19 May 2026

Accepted: 21 May 2026

Published: 2 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

architecture consisting of an N-terminal nucleotide-binding domain (ATPase domain), a substrate-binding domain (SBD), and a flexible C-terminal region involved in regulatory interactions. Despite this conserved structural framework, HSP70 proteins display notable sequence variation across taxa, particularly in surface-exposed regions, terminal segments, and flexible loops [10]. Such variation is thought to contribute to functional specialization by modulating interactions with co-chaperones—including HSP40/DnaJ proteins and nucleotide exchange factors—as well as with diverse client proteins and regulatory partners [11,12]. Understanding how evolutionary forces shape this balance between structural conservation and sequence divergence is therefore critical for elucidating how HSP70 proteins adapt to different cellular and organismal contexts.

Comparative evolutionary analyses provide a powerful framework for identifying regions of proteins that are subject to strong functional constraints versus those that are more evolutionarily labile [13,14]. Highly conserved residues typically experience strong purifying selection because of their importance in maintaining protein stability, catalytic activity, or essential molecular interactions. In contrast, regions exhibiting higher sequence variability may reflect relaxed selective constraints, adaptive divergence, or lineage-specific functional specialization [15,16]. In protein families such as HSP70, comparative analysis across a broad phylogenetic spectrum can therefore help reveal how fundamental chaperone functions are preserved while still allowing evolutionary diversification to meet species-specific physiological demands.

Integrating sequence-level evolutionary analyses with structural information provides additional insight into the mechanistic basis of protein function. Residues that are both evolutionarily conserved and hydrophobic are often located within the interior of the protein, where they contribute to structural stability and domain integrity. Conversely, variable and hydrophilic residues are more frequently found on solvent-exposed surfaces, where they participate in transient interactions, regulatory processes, or post-translational modifications [17,18]. Quantitative approaches such as Shannon entropy analysis allow for the measurement of positional variability across multiple sequence alignments, whereas hydrophobicity profiling—such as through the Kyte–Doolittle scale—can identify regions likely to form structural cores or functional surfaces.

Recent advances in bioinformatics and structural biology have made it increasingly feasible to conduct integrative analyses that combine multiple sequence alignment, phylogenetic inference, domain annotation, evolutionary conservation metrics, and three-dimensional structural mapping. Tools such as MAFFT enable high-quality multiple sequence alignment of protein families [19], while resources such as the Pfam database provide robust identification of conserved protein domains using profile hidden Markov models [20]. When combined with experimentally determined or predicted protein structures deposited in the Protein Data Bank, these approaches allow evolutionary patterns to be interpreted within a structural and functional context. Such integrative analyses can reveal relationships between sequence conservation, structural organization, and functional constraint that may not be apparent from any single analytical method alone.

In this study, a comparative evolutionary analysis of the HSP70 protein family was performed across ten representative vertebrate species: *Homo sapiens* (human), *Mus musculus* (mouse), *Gallus gallus* (chicken), *Danio rerio* (zebrafish), *Xenopus laevis* (frog), *Oryzias latipes* (Japanese medaka), *Macaca mulatta* (rhesus macaque), *Bos taurus* (domestic cow), *Rattus norvegicus* (Norway rat), and *Canis lupus familiaris* (domestic dog) (Table 1). These species span major vertebrate lineages—including mammals, birds, amphibians, and teleost fish—and collectively represent hundreds of millions of years of evolutionary divergence. This phylogenetic breadth provides an opportunity to identify both deeply conserved structural features and lineage-specific sequence variation within the HSP70 family.

Table 1. Vertebrate HSP70 sequences analyzed in this study. List of species, protein accession numbers, and sequence lengths for the HSP70 proteins included in the comparative analysis.

Common Name	Scientific Name	Accession Number	Length (AA)
Human	<i>Homo sapiens</i>	NP_002145.3	641
Mouse	<i>Mus musculus</i>	AAC84169.1	641
Chicken	<i>Gallus gallus</i>	AAN18282.1	634
Zebrafish	<i>Danio rerio</i>	AAF70445.1	658
African Clawed Frog	<i>Xenopus laevis</i>	NP_001121147.1	647
Japanese medaka	<i>Oryzias latipes</i>	XP_011482368.1	639
Rhesus macaque	<i>Macaca mulatta</i>	UWI70517.1	641
Domestic cow	<i>Bos taurus</i>	NP_976067.3	641
Norway rat	<i>Rattus norvegicus</i>	AAA17441.1	641
Domestic dog	<i>Canis lupus familiaris</i>	NP_001003067.2	641

By integrating multiple sequence alignment, phylogenetic reconstruction, entropy-based conservation profiling, hydrophobicity analysis, motif discovery, and structural mapping, this study aims to characterize how evolutionary constraints are distributed across the HSP70 protein architecture. In doing so, the work seeks to clarify how a highly conserved molecular chaperone system preserves its essential biochemical functions while accommodating sequence variation that may contribute to regulatory flexibility and species-specific adaptation.

Beyond advancing our understanding of HSP70 evolution, this study also demonstrates the value of combining complementary computational approaches for comparative protein analysis. The integrative framework presented here may therefore serve as a useful template for investigating evolutionary patterns in other conserved protein families of biomedical and evolutionary importance.

2. Results

2.1. Sequence Alignment Reveals Extensive Conservation of Vertebrate HSP70 Proteins

Multiple sequence alignment of HSP70 proteins from ten representative vertebrate species was generated using MAFFT v7. The alignment demonstrates a striking degree of sequence conservation across the vertebrate HSP70 family, particularly within the N-terminal ATPase (nucleotide-binding) domain (approximately residues 1–380), as illustrated in Supplementary Figure S1. Conserved residues are evident across mammals, birds, amphibians, and teleost fish, reflecting the essential role of ATP binding and hydrolysis in the HSP70 chaperone cycle. Despite the overall conservation, moderate sequence variation is observed in the substrate-binding domain (SBD) and C-terminal regions (approximately residues 381–641). These differences are most apparent among non-mammalian species such as *Xenopus laevis* and *Danio rerio*, where substitutions and short insertions occur primarily with predicted loop regions and solvent-exposed segments. Such regions are commonly associated with regulatory flexibility and interactions with co-chaperones or client proteins. The high overall alignment quality provides a robust foundation for subsequent phylogenetic, entropy, and structural analyses.

2.2. Phylogenetic Relationships of Vertebrate HSP70 Proteins

Phylogenetic relationships among vertebrate HSP70 proteins were reconstructed using the Maximum Likelihood (ML) method implemented in MEGA11 [21] based on the multiple sequence alignment generated with MAFFT (Figure 1). The resulting phylogeny illustrates the evolutionary relationships among HSP70 proteins from ten representative vertebrate species.

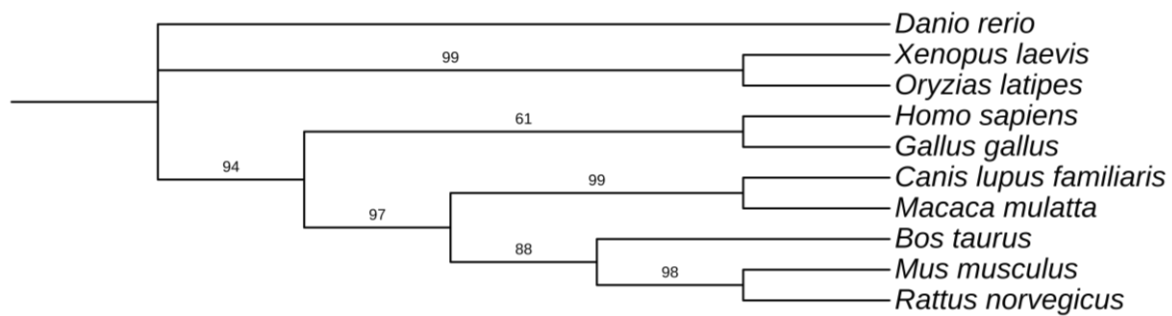


Figure 1. Maximum likelihood phylogenetic tree of vertebrate HSP70 proteins. Phylogenetic relationships among HSP70 proteins from ten representative vertebrate species reconstructed using the Maximum Likelihood method in MEGA11 (v11) [21] based on the MAFFT (v7.526) multiple sequence alignment. The tree was inferred using the JTT substitution model [22] with 1000 bootstrap replicates. Bootstrap support values are shown at internal nodes. Because the tree is based on a single conserved protein family, it should be interpreted as a representation of sequence relationships among HSP70 homologs rather than a strict species phylogeny.

Several mammalian sequences, including *Mus musculus*, *Rattus norvegicus*, *Bos taurus*, *Macaca mulatta*, and *Canis lupus familiaris*, cluster closely together, reflecting the strong sequence conservation of HSP70 proteins within mammals. The rodent proteins (*Mus musculus* and *Rattus norvegicus*) form a well-supported clade, consistent with their close evolutionary relationship. Other mammalian sequences, including *Bos taurus*, *Macaca mulatta*, and *Canis lupus familiaris*, branch within the broader mammalian cluster.

Non-mammalian vertebrates form distinct branches in the tree. The teleost fish (*Danio rerio* and *Oryzias latipes*) and amphibian (*Xenopus laevis*) sequences occupy separate positions relative to the mammalian sequences, reflecting their earlier divergence during vertebrate evolution. The avian sequence (*Gallus gallus*) also appears as a distinct lineage relative to the other vertebrate groups. Bootstrap support values derived from 1000 replicates indicate strong statistical support for several internal nodes in the tree.

Overall, the phylogenetic analysis highlights the high level of evolutionary conservation of HSP70 proteins across vertebrates while illustrating the divergence of HSP70 sequences among major vertebrate lineages.

2.3. Domain Architecture Is Conserved Across Vertebrates

Domain annotation performed using the Pfam database via HMMER confirmed that all analyzed sequences contain the two canonical domains characteristic of HSP70 proteins:

- PF00012: HSP70 nucleotide-binding domain (NBD) (~residues 1–380)
- PF00052: HSP70 substrate-binding domain (SBD) (~residues 381–641)

These domains were detected in all ten vertebrate sequences without evidence of lineage-specific domain insertions or deletions. The preservation of this domain architecture across species highlights the strong structural constraints acting on the HSP70 chaperone system. Functional diversification within the family therefore appears to arise primarily from sequence-level variation within domains or flexible terminal regions rather than from changes in core domain organization.

2.4. Residue Conservation Profiles Support Domain-Specific Evolutionary Constraints

Residue conservation was further examined using a normalized conservation score derived from the multiple sequence alignment (Figure 2). The conservation profile shows a pronounced plateau across the ATPase domain, indicating strong evolutionary constraint across this region.

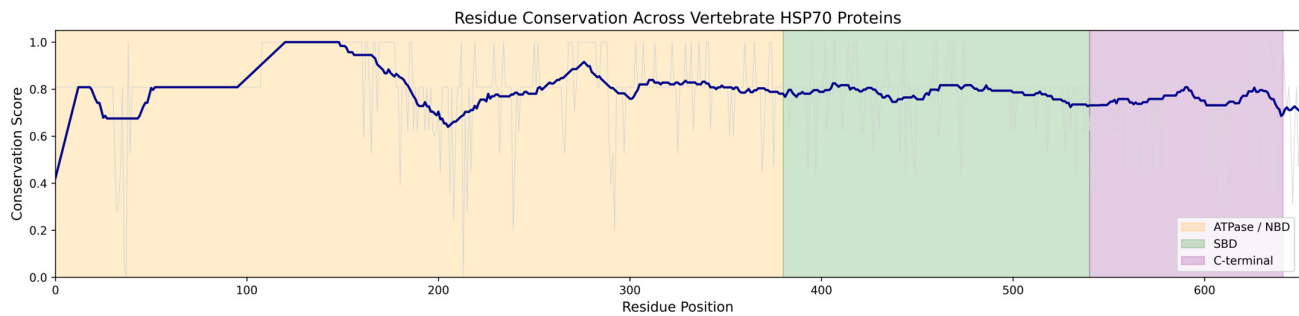


Figure 2. Residue conservation profile across vertebrate HSP70 proteins. Residue conservation derived from Shannon entropy calculated for each position in the multiple sequence alignment. The light gray trace represents raw conservation values, while the dark blue line indicates a smoothed conservation profile (25-residue window). Shaded regions denote approximate boundaries of the ATPase/nucleotide-binding domain (NBD), substrate-binding domain (SBD), and C-terminal region.

In contrast, conservation values decline gradually across the substrate-binding domain and C-terminal region, reflecting increased sequence variability. This pattern mirrors the entropy analysis and reinforces the conclusion that the catalytic core of HSP70 is highly conserved, while peripheral regions exhibit greater evolutionary flexibility.

Together, the conservation and entropy analyses provide complementary quantitative measures of evolutionary constraint across the HSP70 protein family.

2.5. Entropy Analysis Quantifies Residue-Level Evolutionary Constraints

To quantify sequence variability across the alignment, Shannon entropy was calculated for each alignment position. The resulting entropy profile (Figure 3) reveals clear differences in evolutionary constraint across the protein.

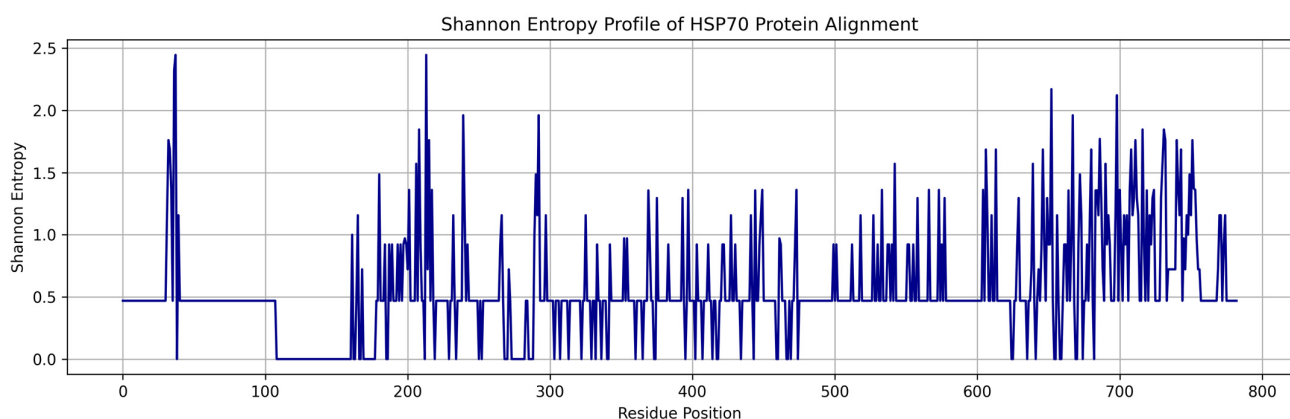


Figure 3. Shannon entropy profile of vertebrate HSP70 proteins. Shannon entropy was calculated for each position in the multiple sequence alignment to assess sequence conservation. Lower entropy values indicate highly conserved residues, whereas higher values reflect increased sequence variability across vertebrate HSP70 sequences.

Residues within the ATPase domain exhibit consistently low entropy values, indicating strong evolutionary conservation. These residues correspond to structural elements surrounding the ATP-binding pocket and the catalytic machinery required for ATP hydrolysis. Such low-entropy positions are characteristic of sites under strong purifying selection due to their essential functional roles.

In contrast, elevated entropy values are observed in the substrate-binding domain and C-terminal region, where sequence variation is more frequent. Many of these high-entropy positions correspond to predicted loop regions or solvent-exposed surfaces, suggesting that

evolutionary diversification primarily affects flexible regions involved in protein–protein interactions rather than the catalytic core of the protein.

2.6. Hydrophobicity Profiles Reveal Structural Organization of the HSP70 Core

The Kyte–Doolittle hydrophobicity scale was used to calculate position-specific hydrophobicity across the alignment. The resulting hydrophobicity profile (Figure 4) displays alternating hydrophobic and hydrophilic segments characteristic of globular proteins. Regions of elevated hydrophobicity correspond primarily to structural elements within the ATPase domain, where hydrophobic residues contribute to the formation of the protein’s internal structural core. Conversely, regions with lower hydrophobicity are frequently located within loops or surface-exposed regions, consistent with their accessibility to solvent and involvement in intermolecular interactions.

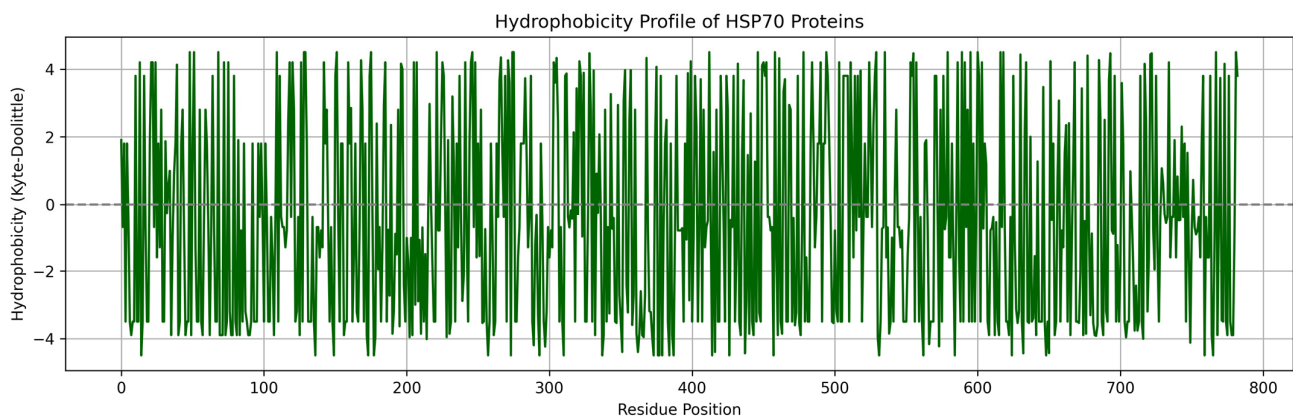


Figure 4. Hydrophobicity profile of vertebrate HSP70 proteins. Average hydrophobicity calculated for each alignment position using the Kyte–Doolittle scale. The dashed horizontal line indicates neutral hydrophobicity (0). Variations in hydrophobicity correspond to alternating hydrophobic and hydrophilic regions along the protein sequence.

Statistical comparison of mean hydrophobicity values across domains revealed that residues within the nucleotide-binding domain exhibit significantly higher average hydrophobicity than those within the substrate-binding domain ($p < 0.05$, two-sample t -test), supporting the interpretation that the ATPase domain forms a more tightly packed structural core.

2.7. Structural Mapping Links Sequence Conservation to Three-Dimensional Organization

To place the sequence-based analyses in structural context, conservation and hydrophobicity patterns were mapped onto the crystal structure of human HSP70 (PDB: 5AQV) using py3Dmol (Figure 5). This visualization reveals that residues exhibiting both low entropy and high hydrophobicity cluster within the interior of the ATPase domain, forming a conserved structural core surrounding the ATP-binding site.

In contrast, residues with higher entropy values are predominantly located on the protein surface, particularly within the substrate-binding domain and C-terminal region. These surface-exposed regions likely contribute to interactions with client proteins and co-chaperones, providing structural flexibility while preserving the essential catalytic machinery of the protein. The structural mapping therefore provides spatial confirmation of the evolutionary patterns identified in the sequence-based analyses.

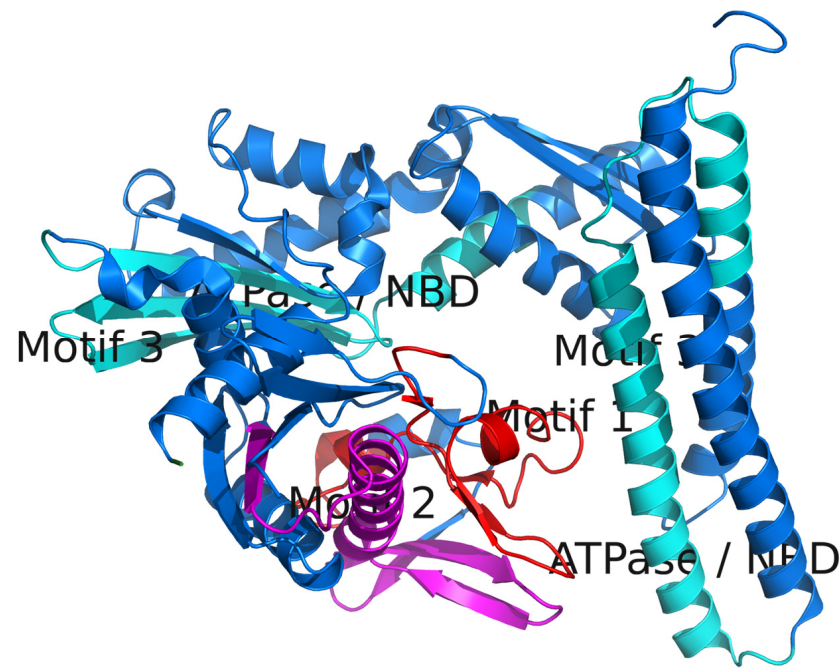


Figure 5. Structural mapping of conservation and hydrophobicity onto the human HSP70 structure. The three-dimensional structure of human HSP70 (PDB: 5AQV) was visualized using PyMOL (v2.5) [23]. Residues exhibiting low Shannon entropy and higher hydrophobicity values are predominantly located within the ATPase (nucleotide-binding) domain, forming the structural core of the protein. In contrast, residues with greater variability are mainly found in surface-exposed regions, particularly within the substrate-binding domain and C-terminal segment. The ATPase domain is shown in blue, whereas the substrate-binding domain and C-terminal region are shown in cyan. Conserved motifs identified in this study are highlighted using distinct colors corresponding to those shown in the structure. Motif locations correspond to conserved sequence regions identified using MEME (v5.5) [24].

2.8. Conserved Motif Analysis Identifies Core Functional Elements

Motif discovery using the Multiple Expectation Maximization for Motif Elicitation (MEME) suite [24] identified six statistically significant conserved motifs across the analyzed vertebrate HSP70 protein sequences (Figure 6; Table 2). The identified motifs were widely distributed among the examined taxa, demonstrating strong evolutionary conservation of functionally important regions within the HSP70 protein family. Several motifs were consistently detected across most vertebrate sequences, particularly within regions corresponding to the ATPase domain and substrate-binding regions, which are known to play essential roles in ATP binding, hydrolysis, protein folding, and chaperone-mediated client interactions. The MEME E-value represents the expected number of motifs with equal or greater statistical significance that would be identified in a dataset of the same size by chance; therefore, lower E-values indicate stronger statistical support for motif conservation across sequences.

Although the overall motif architecture was highly conserved among vertebrate species, minor variation in motif occurrence and arrangement was observed in the *Homo sapiens* sequence under the selected MEME parameters. Such variation may reflect sequence divergence relative to the consensus motifs identified by MEME rather than complete absence of the corresponding functional regions. Because MEME detects statistically enriched sequence patterns, differences in local amino acid composition can influence motif recognition sensitivity even when structurally related residues remain present. Closely related mammalian sequences, including *Macaca mulatta*, retained motif distributions highly sim-

ilar to other vertebrate taxa, supporting the overall conservation of HSP70 functional organization across evolution.

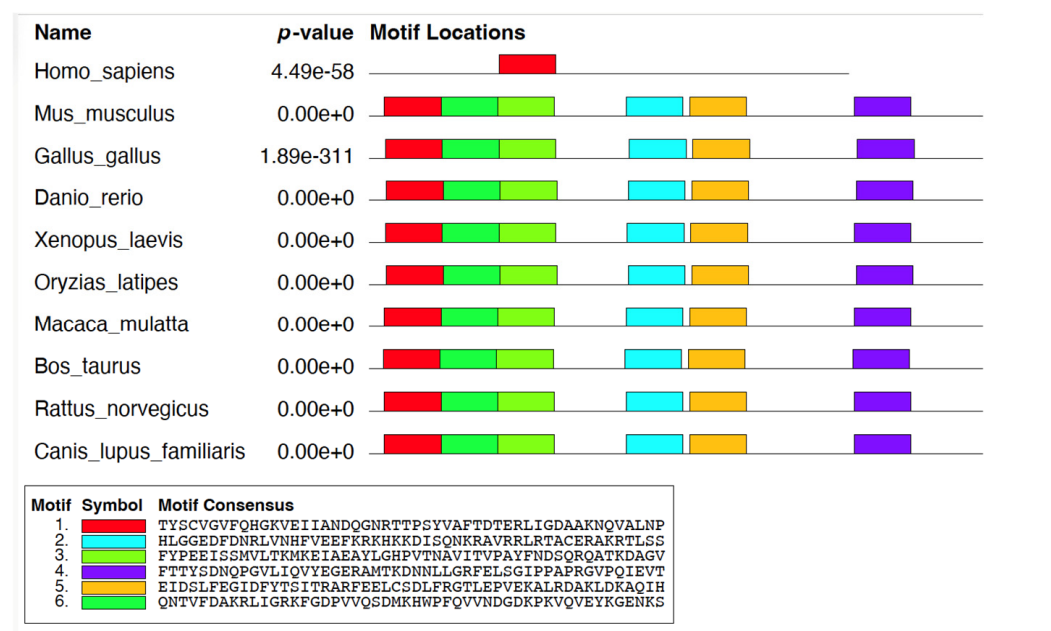


Figure 6. Conserved motif architecture and distribution across ten vertebrate HSP70 protein sequences identified using MEME [24]. Colored blocks represent conserved motifs detected within each protein sequence. Most motifs are highly conserved across vertebrate taxa, although variation in motif occurrence and arrangement is observed in *Homo sapiens* compared with other analyzed species.

Table 2. Motif statistics obtained using MEME [24], including motif width, occurrence frequency, consensus sequences, and predicted functional regions.

Motif ID	Width (aa)	Sites	Consensus Sequence (Representative Region)	Predicted Region
MEME Motif 1	50	10	TYSCVGVFQHGKVEIANDQGNRTTPSYVAFTDTER-LIGDAAKNQVALNP	ATPase domain (NBD)
MEME Motif 2	50	9	HLGGEDFDNRLVNHFEVEFKRKHKKDIS-QNKRAVRRLRTACERAKRTLSS	Interdomain linker/C-terminal region
MEME Motif 3	50	9	FYPEEISSMVLTKMKEIAEAY-LGHPVINAVITVPAYFNDSQRQATKDAGV	ATPase domain (NBD)
MEME Motif 4	50	9	FTTYSNQPGLVLIQVYEGERAMTKDNNLLGRFEL-SGIPPAPRGVVPQIEVT	Substrate-binding domain (SBD)
MEME Motif 5	50	9	EIDSLFEGIDFYTSITRARFEELCSDLFRGTLE-PVEKALRDAKLDAQIH	Interdomain linker
MEME Motif 6	50	9	QNTVFDKRLIGRKFGDPVVQSDMKHWPQVQVNDGDKPKVQVEYKGENKS	ATPase domain (NBD)

Sequence logos for each conserved motif (Figure 7) revealed strong positional conservation among several hydrophobic, polar, and charged amino acid residues that contribute to structural stability and chaperone activity. Highly conserved residues were particularly prominent within motifs associated with ATPase and substrate-binding functions, emphasizing the functional importance of these regions. Collectively, the identified motifs highlight the coexistence of deeply conserved catalytic elements and moderately variable sequence regions within vertebrate HSP70 proteins. The structural locations of these motifs within the HSP70 protein are illustrated in Figure 5, providing spatial context for their evolutionary conservation patterns.



Figure 7. Sequence logos representing conserved amino acid motifs identified in vertebrate HSP70 proteins using MEME [24]. The height of each residue reflects its relative conservation at a given position, measured in bits of information content. Highly conserved residues indicate functionally or structurally important regions within the HSP70 protein family.

3. Discussion

Mapping sequence conservation and variability onto the three-dimensional structure of human HSP70 (PDB: 5AQV) provides an important structural context for interpreting evolutionary constraints. In this analysis, conserved and hydrophobic residues were found to cluster predominantly within the ATPase domain, forming the structural core of the protein, whereas more variable and hydrophilic residues were largely surface-exposed. This spatial distribution mirrors patterns observed in other molecular chaperone systems and reflects a fundamental principle of protein evolution: residues critical for maintaining structural integrity and catalytic activity tend to be highly conserved, while residues involved in regulatory or interaction functions exhibit greater evolutionary flexibility [25].

The strong conservation of residues within the nucleotide-binding domain (NBD) likely reflects the essential role of ATP binding and hydrolysis in driving the HSP70 chaperone cycle. Structural studies have demonstrated that conformational transitions within the ATPase domain regulate substrate binding and release, linking nucleotide state to client protein processing [25–27]. Consequently, even minor alterations in residues contributing to the catalytic core or nucleotide-binding pocket may disrupt the allosteric communication between domains that is essential for chaperone activity. The low entropy

values and high hydrophobicity observed in this region therefore support the interpretation that strong purifying selection maintains the structural and functional integrity of the ATPase machinery across vertebrates.

In contrast, the substrate-binding domain (SBD) and C-terminal regions exhibited increased sequence variability across species. These regions are known to mediate interactions with diverse client proteins and regulatory partners, including HSP40/DnaJ co-chaperones and nucleotide exchange factors [26,27]. Structural mapping indicates that many of these variable residues are located on solvent-exposed surfaces, consistent with the idea that evolutionary divergence is tolerated in regions involved in molecular recognition or regulatory modulation. Such variability may allow for species-specific adaptation of chaperone interaction networks without disrupting the conserved catalytic mechanism of the protein.

3.1. Motif Discovery Highlights Functional and Lineage-Specific Features

MEME [24] motif analysis revealed a high degree of conservation within vertebrate HSP70 proteins, particularly in regions associated with the ATPase domain and substrate-binding functions. Several conserved motifs were consistently detected across most analyzed species, reinforcing their association with essential biochemical processes such as ATP binding, ATP hydrolysis, interdomain communication, and maintenance of structural stability [26,27]. The strong positional conservation observed in the sequence logos (Figure 7) further supports the functional importance of these motifs as integral components of the HSP70 chaperone machinery [25].

Although the overall motif architecture was highly conserved across vertebrates, moderate variation in motif occurrence and arrangement was observed among certain species, particularly in the human HSP70 sequence under the selected MEME detection parameters. Given the strong conservation generally observed among vertebrate HSP70 homologs, this variation likely reflects localized sequence divergence or differences in motif detection sensitivity rather than complete evolutionary loss of functional regions. Many of the more variable motifs map to the substrate-binding domain and C-terminal regions, which are known to exhibit greater functional flexibility and involvement in co-chaperone interactions or client recognition [28,29].

Such motif organization parallels patterns described in other conserved protein families, where highly conserved motifs support catalytic and structural functions while more flexible sequence elements contribute to regulatory adaptation and interaction specificity [30,31]. In HSP70 proteins, diversification within peripheral or surface-exposed regions may facilitate adaptation of chaperone networks to lineage-specific cellular environments while preserving the conserved ATP-driven folding mechanism.

More broadly, the clustering of conserved residues within the structural core, combined with increased variability at surface-exposed positions, suggests that evolutionary divergence in HSP70 proteins primarily affects interaction interfaces rather than catalytic components. This structural organization enables the protein to maintain a stable functional scaffold while accommodating evolutionary changes that may refine regulatory interactions or substrate specificity.

Importantly, the integration of sequence conservation metrics with structural visualization provides a powerful framework for understanding how evolutionary pressures shape protein architecture. By combining entropy profiling, hydrophobicity analysis, motif discovery, and structural mapping, the present study demonstrates how complementary computational approaches can reveal functional constraints across multiple levels of protein organization. Such integrative strategies may therefore serve as a useful framework for investigating the evolution and functional diversification of other conserved protein families.

This work provides an integrative analytical framework that combines sequence-level evolutionary metrics with structural mapping to characterize functional constraints across protein families.

3.2. Broader Implications and Future Directions

Taken together, the results of this study reveal a coherent evolutionary pattern within vertebrate HSP70 proteins: a highly conserved catalytic core maintained by strong purifying selection, coupled with more variable surface regions that may enable functional diversification. These findings are consistent with broader evolutionary models demonstrating that protein structural cores evolve more slowly than solvent-exposed or regulatory regions [32,33].

Future work could extend this framework by incorporating a broader phylogenetic sampling, including non-vertebrate deuterostomes or invertebrate lineages, to better trace the evolutionary origins and diversification of HSP70 proteins. Additionally, integrating comparative sequence analysis with functional assays—such as co-chaperone binding experiments or client-protein interaction studies—would help clarify how sequence variation influences chaperone activity and cellular stress responses.

Ultimately, understanding how evolutionary constraints shape both the conserved catalytic machinery and the flexible regulatory regions of HSP70 proteins will provide deeper insight into how molecular chaperones adapt to diverse cellular environments while maintaining their essential roles in protein homeostasis.

4. Materials and Methods

4.1. Sequence Selection and Retrieval

To conduct a comprehensive comparative analysis of the HSP70 protein family across vertebrates, full-length amino acid sequences for ten representative species spanning major vertebrate clades were retrieved. These species were selected to encompass a broad evolutionary range, including mammals (*Homo sapiens*, *Mus musculus*, *Macaca mulatta*, *Bos taurus*, *Rattus norvegicus*, *Canis lupus familiaris*), a bird (*Gallus gallus*), an amphibian (*Xenopus laevis*), and two teleost fish (*Danio rerio* and *Oryzias latipes*). This taxonomic diversity allows for the identification of conserved and divergent features in HSP70 proteins that reflect both deep evolutionary constraints and lineage-specific adaptations. Protein sequences were retrieved from the NCBI Protein database [34] using curated accession numbers corresponding to canonical or highly expressed cytosolic isoforms of HSP70 (Table 1).

All sequences were downloaded in FASTA format and manually verified for completeness and annotation consistency. Only full-length sequences corresponding to canonical cytosolic HSP70 isoforms were retained to ensure comparability across species. These sequences served as the basis for downstream analyses, including multiple sequence alignment, phylogenetic reconstruction, motif discovery, and structural mapping.

4.2. Multiple Sequence Alignment

To investigate sequence conservation and divergence across vertebrate HSP70 proteins, multiple sequence alignment (MSA) was performed using MAFFT version 7.511, a widely used and highly accurate tool for aligning large protein datasets [19]. The alignment was conducted with the --auto strategy enabled, allowing MAFFT to automatically select the optimal alignment algorithm based on the size and similarity of the input sequences. This approach balances accuracy and computational efficiency by switching between progressive and iterative refinement algorithms as appropriate.

The input dataset consisted of ten full-length HSP70 protein sequences, each representing a distinct vertebrate species. These sequences varied slightly in length due to

species-specific insertions, deletions, or isoform differences. The alignment process introduced gaps where necessary to optimize residue matching across conserved domains, particularly the nucleotide-binding domain (NBD) and substrate-binding domain (SBD), which are known to be highly conserved across species [35].

The resulting MSA was visually inspected to confirm alignment quality, with particular attention to conserved motifs and domain boundaries. The alignment was exported in FASTA format and used as input for downstream analyses, including phylogenetic tree construction, motif discovery, entropy and hydrophobicity profiling, and structure-function mapping. By using a robust and automated alignment strategy, high-quality residue-level comparisons were achieved, enabling the identification of conserved evolutionary features and functionally important regions within the HSP70 protein family.

4.3. Phylogenetic Tree Construction

To investigate the evolutionary relationships among the HSP70 protein sequences analyzed in this study, a phylogenetic tree was constructed using the Maximum Likelihood (ML) method implemented in MEGA11 (v11) [21]. The amino acid sequences obtained from NCBI [34] were first aligned using the MAFFT sequence alignment program. The resulting alignment was then used as input for phylogenetic reconstruction.

Maximum Likelihood is a statistically robust method for phylogenetic inference that estimates the tree topology most likely to have produced the observed sequence data under a specified evolutionary model. In this analysis, amino acid substitutions were modeled using the Jones–Taylor–Thornton (JTT) substitution model [22], which describes empirically derived substitution rates among amino acids during protein evolution.

The ML analysis was performed assuming uniform rates among sites, and gaps or missing positions were treated using the complete deletion approach. The initial tree for the heuristic search was generated automatically using the Neighbor-Joining/BioNJ method, followed by optimization of branch lengths using the Nearest-Neighbor Interchange (NNI) algorithm to identify the topology with the highest likelihood. Statistical support for internal branches was assessed using bootstrap analysis with 1000 replicates.

The resulting phylogenetic tree was visualized using the MEGA11 (v11) Tree Explorer interface [21]. Species names were displayed in italicized format, and the final tree was arranged in a rectangular layout to improve readability. The reconstructed phylogeny reflects broad vertebrate evolutionary relationships, with mammalian HSP70 sequences clustering together and more distantly related vertebrate taxa (birds, amphibians, and fish) occupying more basal positions in the tree. This phylogenetic framework provides evolutionary context for subsequent analyses of sequence conservation, motif distribution, and structural variation across the HSP70 protein family.

4.4. Domain Annotation Using Pfam

To identify conserved protein domains within the HSP70 sequences across vertebrates, domain-level annotation was performed using HMMER (v3.3), a widely used suite for sequence analysis based on profile hidden Markov models (HMMs) [36]. Specifically, the HMMER web API hosted by the European Bioinformatics Institute (EMBL-EBI) [37] was used to access the hmmscan functionality, enabling efficient annotation of sequences against the Pfam protein families database [20].

Each of the ten HSP70 protein sequences in FASTA format was submitted to the hmmscan endpoint via HTTP POST requests, using Python (v3.10) scripts [38] and the requests library. The target database for scanning was Pfam-A, the manually curated subset of Pfam, which includes high-confidence models for protein domain families. Pfam

domains are defined using multiple sequence alignments and HMMs, making them suitable for detecting remote homologs and conserved structural or functional regions [20].

Upon submission, the HMMER server returned results in JSON format, which were programmatically parsed to extract key domain information, including Domain name and accession (e.g., PF00012: HSP70 family), Start and end positions within the protein sequence, E-values representing the statistical significance of the match, and domain descriptions.

Domain boundaries for the HSP70 nucleotide-binding domain (PF00012) and substrate-binding domain (PF00052) were identified using HMMER hmmscan against the Pfam-A database. Across the analyzed sequences, the nucleotide-binding domain (NBD) spans approximately residues 1–380, whereas the substrate-binding domain (SBD) spans approximately residues 381–641.

4.5. Entropy Analysis

To quantitatively assess evolutionary conservation across the aligned HSP70 protein sequences, Shannon entropy was calculated at each column of the multiple sequence alignment. Shannon entropy provides a measure of variability or uncertainty in a set of symbols—in this case, amino acid residues at a given alignment position.

The entropy H for each alignment position was calculated using the standard formula:

$$H = -\sum_{i=1}^n p_i \log_2 p_i$$

where p_i is the observed frequency of amino acid i at a given position, and n is the total number of unique amino acids observed at that site. Positions with low entropy (i.e., close to 0) indicate strong conservation, suggesting functional or structural importance. In contrast, high entropy values denote greater variability, which may reflect relaxed selective constraints or adaptation to species-specific contexts.

Entropy profiling was performed using a custom Python (v3.10) script that parsed the aligned FASTA file generated by MAFFT. The alignment was read as a matrix, and amino acid frequencies were computed for each column. Only standard amino acids (excluding gaps or ambiguous residues) were included in the calculation to avoid inflation of entropy values due to poor alignment quality. The resulting entropy profile was plotted along the sequence length to visually distinguish conserved core regions from hypervariable sites, often corresponding to loop regions, termini, or insertion segments.

Shannon entropy is widely used in comparative sequence analysis to infer the functional importance of residues and domains and has been previously applied to both prokaryotic and eukaryotic systems [39]. In the context of HSP70, entropy analysis provides an important link between evolutionary constraints and biophysical properties of the protein, complementing hydrophobicity and structural mapping efforts.

4.6. Hydrophobicity Profiling

Hydrophobicity profiling was employed to explore the distribution of hydrophobic and hydrophilic residues along the aligned HSP70 protein sequences. This analysis helps identify regions that are likely to form buried hydrophobic cores or surface-exposed hydrophilic loops, both of which are critical for understanding protein folding, stability, and function [40,41]. Hydrophobicity values were assigned based on the Kyte–Doolittle scale [40], a widely used empirical metric that ranks amino acids by their relative hydrophobicity. In this scale, hydrophobic residues such as Ile, Val, and Leu are assigned positive values, whereas polar and charged residues like Asp, Glu, and Lys are assigned negative values. The scale has historically been useful in predicting transmembrane domains, folding cores, and functional surfaces [42].

For each alignment position (i.e., column), the mean hydrophobicity score was calculated by averaging the Kyte–Doolittle values of the amino acids present at that position across all aligned sequences. Gaps and ambiguous residues were excluded from the calculation to avoid skewed scores. When interpreted alongside the entropy analysis, the hydrophobicity profile adds a complementary layer of insight: highly conserved hydrophobic regions are typically essential for structural integrity, while hydrophilic, variable regions may be implicated in species-specific functions or post-translational modifications. This biophysical mapping was crucial for correlating sequence-based features with functional constraints and domain architecture and later integrated with structural models to provide a spatial interpretation of evolutionary and hydrophobic patterns.

4.7. Motif Discovery Using MEME

To identify conserved sequence motifs within vertebrate HSP70 proteins, motif discovery analysis was performed using the Multiple Expectation Maximization for Motif Elicitation (MEME) Suite version 5.5.9 [24]. The aligned HSP70 amino acid sequences were submitted to the MEME web server using the classic discovery mode.

The analysis was configured to identify up to six conserved motifs with motif widths ranging from 8 to 50 amino acids. The zero or one occurrence per sequence (zoops) model was applied, allowing motifs to occur once or not at all in each sequence. Statistical significance of identified motifs was evaluated using MEME E-values, where lower E-values indicate stronger support for motif conservation across the analyzed sequences.

Consensus motif sequences, motif widths, occurrence frequencies, and motif distributions across species were extracted from the MEME output and used for comparative analysis and visualization.

4.8. Structure Mapping

To place the sequence-based analyses of conservation and hydrophobicity into a structural framework, structure mapping was performed using the resolved crystal structure of human HSP70 (PDB ID: 5AQV) [43]. This structure corresponds to the nucleotide-binding domain (NBD) of HSPA1A and serves as a reliable structural representative of the broader HSP70 family because of its high degree of sequence and structural conservation [44,45]. The three-dimensional structure was visualized and annotated using the PyMOL Molecular Graphics System (version 2.5, Schrödinger, LLC, New York, NY, USA) [23] and py3Dmol [46], enabling real-time rendering and annotation of structural features.

Because experimentally resolved structures are not available for all analyzed species, structural mapping was performed using the human HSP70 structure (PDB: 5AQV) as a representative model. Given the strong structural conservation reported across HSP70 homologs, this structure provides a suitable framework for visualizing conserved and variable residues across vertebrate species.

4.9. Computational Environment and Data Analysis

All computational analyses were performed using Python (v3.10) [38] within the Jupyter Notebook (v8.37) environment [47]. Sequence processing, phylogenetic data handling, and bioinformatics workflows were implemented using Biopython (v1.87) [48], while numerical computations and array-based analyses were conducted using NumPy (v2.2) [49] and SciPy (1.15) [50]. Data organization and tabular processing were performed using Pandas (v2.3) [51]. Statistical visualization and graphical representations were generated using Matplotlib (3.10) [52] and Seaborn (0.13) [53]. These open-source computational tools enabled reproducible analysis, visualization, and interpretation of sequence conser-

vation, entropy, hydrophobicity, motif distribution, and phylogenetic relationships across vertebrate HSP70 proteins.

5. Conclusions

This study presents a reproducible, integrative framework for investigating protein family evolution, combining multiple sequence alignment, domain annotation, entropy and hydrophobicity profiling, phylogenetics, structural visualization, motif discovery, and evolutionary conservation analysis. Applied to the HSP70 family across ten representative vertebrate species, these findings reveal a highly conserved molecular scaffold—particularly within the ATPase and substrate-binding domains—reflecting strong purifying selection and the essential role of HSP70 in protein homeostasis.

At the same time, lineage-specific sequence variation, particularly within surface-exposed loops and regulatory motifs [54], underscores how adaptive flexibility can be encoded within a structurally constrained framework. The identification of conserved and divergent motifs further supports the hypothesis that evolutionary innovation often occurs within peripheral or regulatory regions without disrupting the core chaperone machinery.

These findings reinforce the broader principle that protein evolution balances the dual pressures of functional constraint and adaptive diversification. Moreover, this approach is readily generalizable to other conserved protein families, offering a robust template for future comparative studies in molecular evolution, structural biology, and functional annotation. As large-scale protein datasets continue to grow, such integrative strategies will be increasingly vital for decoding the evolutionary logic underlying protein structure and function.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jgbg1020011/s1>. Figure S1: Multiple sequence alignment of vertebrate HSP70 proteins.

Funding: This research received no external funding.

Data Availability Statement: All protein sequences used in this study are publicly available in the NCBI Protein database under the accession numbers listed in Table 1. Alignment files, motif analyses, and scripts are available upon request.

Acknowledgments: The author gratefully acknowledges the open scientific resources, databases, and open-source software that made this study possible. Publicly available biological sequence repositories, computational tools, and bioinformatics platforms were essential for data acquisition, analysis, visualization, and interpretation. The author recognizes the broader scientific and open-source communities whose efforts in developing, maintaining, and freely distributing these resources continue to advance reproducible and accessible research in computational biology and molecular evolution.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

HSP70	Heat Shock Protein 70
ATP	Adenosine Triphosphate
NCBI	National Center for Biotechnology Information
CDS	Coding Sequence
aa	Amino Acid
ML	Maximum Likelihood
MSA	Multiple Sequence Alignment
MAFFT (v7.526)	Multiple Alignment using Fast Fourier Transform

MEME (v5.5)	Multiple EM for Motif Elicitation
Pfam	Protein Families Database
PDB	Protein Data Bank
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
NJ	Neighbor-Joining
BLAST	Basic Local Alignment Search Tool
FASTA	Fast-All file format (sequence format)
GRAVY	Grand Average of Hydropathicity
ORF	Open Reading Frame

References

1. Treweek, T.M.; Meehan, S.; Ecroyd, H.; Carver, J.A. Small heat-shock proteins: Important players in regulating cellular proteostasis. *Cell. Mol. Life Sci.* **2015**, *72*, 429–451. [\[CrossRef\]](#)
2. Lang, B.J.; Guerrero, M.E.; Prince, T.L.; Okusha, Y.; Bonorino, C.; Calderwood, S.K. The functions and regulation of heat shock proteins; key orchestrators of proteostasis and the heat shock response. *Arch. Toxicol.* **2021**, *95*, 1943–1970. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Hartl, F.U.; Hayer-Hartl, M. Molecular chaperones in the cytosol: From nascent chain to folded protein. *Science* **2002**, *295*, 1852–1858. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Young, J.; Agashe, V.; Siegers, K.; Hartl, F.U. Pathways of chaperone-mediated protein folding in the cytosol. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 781–791. [\[CrossRef\]](#)
5. Hartl, F.; Bracher, A.; Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **2011**, *475*, 324–332. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Balchin, D.; Hayer-Hartl, M.; Hartl, F.U. In vivo aspects of protein folding and quality control. *Science* **2016**, *353*, aac4354. [\[CrossRef\]](#)
7. Sorensen, J.G.; Kristensen, T.N.; Loeschcke, V. The evolutionary and ecological role of heat shock proteins. *Ecol. Lett.* **2003**, *6*, 1025–1037. [\[CrossRef\]](#)
8. Calderwood, S.K.; Khaleque, M.A.; Sawyer, D.B.; Ciocca, D.R. Heat shock proteins in cancer: Chaperones of tumorigenesis. *Trends Biochem. Sci.* **2006**, *31*, 164–172. [\[CrossRef\]](#)
9. Sherman, M.; Multhoff, G. Heat shock proteins in cancer. *Ann. N. Y. Acad. Sci.* **2007**, *1113*, 192–201. [\[CrossRef\]](#)
10. Mayer, M.P.; Bukau, B. Hsp70 chaperones: Cellular functions and molecular mechanism. *Cell. Mol. Life Sci.* **2005**, *62*, 670–684. [\[CrossRef\]](#)
11. Qiu, X.B.; Shao, Y.M.; Miao, S.; Wang, L. The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cell. Mol. Life Sci.* **2006**, *63*, 2560–2570. [\[CrossRef\]](#)
12. Kampinga, H.H.; Craig, E.A. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 579–592. [\[CrossRef\]](#)
13. Fay, J.C.; Wu, C.I. Sequence Divergence, Functional Constraint, and Selection in Protein Evolution. *Annu. Rev. Genom. Hum. Genet.* **2003**, *4*, 213–235. [\[CrossRef\]](#)
14. Worth, C.; Gong, S.; Blundell, T. Structural and functional constraints in the evolution of protein families. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 709–720. [\[CrossRef\]](#)
15. Conant, G.C.; Wolfe, K.H. Turning a hobby into a job: How duplicated genes find new functions. *Nat. Rev. Genet.* **2008**, *9*, 938–950. [\[CrossRef\]](#)
16. Nei, M.; Rooney, A.P. Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* **2005**, *39*, 121–152. [\[CrossRef\]](#)
17. Bogumil, D.; Dagan, T. Chaperonin-dependent accelerated substitution rates in prokaryotes. *Genome Biol. Evol.* **2010**, *2*, 602–608. [\[CrossRef\]](#)
18. Echave, J.; Spielman, S.J.; Wilke, C.O. Causes of evolutionary rate variation among protein sites. *Nat. Rev. Genet.* **2016**, *17*, 109–121. [\[CrossRef\]](#)
19. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [\[CrossRef\]](#)
20. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G.A.; Sonnhammer, E.L.L.; Tosatto, S.C.E.; Paladin, L.; Raj, S.; Richardson, L.J.; et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **2021**, *49*, D412–D419. [\[CrossRef\]](#)
21. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [\[CrossRef\]](#)
22. Jones, D.T.; Taylor, W.R.; Thornton, J.M. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **1992**, *8*, 275–282. [\[CrossRef\]](#)

23. Schrödinger, L. *The PyMOL Molecular Graphics System, Version 2.5*; Schrödinger, LLC: New York, NY, USA, 2021.
24. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The MEME Suite. *Nucleic Acids Res.* **2015**, *43*, W39–W49. [[CrossRef](#)]
25. Rohland, L.; Kityk, R.; Smalinskaitė, L.; Mayer, M.P. Conformational dynamics of the Hsp70 chaperone throughout key steps of its ATPase cycle. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2123238119. [[CrossRef](#)]
26. Liu, Y.; Gierasch, L.M.; Bahar, I. Role of Hsp70 ATPase Domain Intrinsic Dynamics and Sequence Evolution in Enabling its Functional Interactions with NEFs. *PLoS Comput. Biol.* **2010**, *6*, e1000931. [[CrossRef](#)]
27. Mayer, M.P.; Gierasch, L.M. Recent advances in the structural and mechanistic aspects of Hsp70 molecular chaperones. *J. Biol. Chem.* **2019**, *294*, 2085–2097. [[CrossRef](#)]
28. Metzger, D.C.H.; Hemmer-Hansen, J.; Schulte, P.M. Conserved structure and expression of hsp70 paralogs in teleost fishes. *Comp. Biochem. Physiol. Part D Genom. Proteom.* **2016**, *18*, 10–20. [[CrossRef](#)]
29. Freilich, R.; Arhar, T.; Abrams, J.L.; Gestwicki, J.E. Protein–Protein Interactions in the Molecular Chaperone Network. *Acc. Chem. Res.* **2018**, *51*, 940–949. [[CrossRef](#)]
30. Tokuriki, N.; Tawfik, D.S. Protein Dynamism and Evolvability. *Science* **2009**, *324*, 203–207. [[CrossRef](#)]
31. Schreiber, G.; Keating, A.E. Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol.* **2011**, *21*, 50–61. [[CrossRef](#)]
32. Chothia, C.; Lesk, A.M. The Relation Between the Divergence of Sequence and Structure in Proteins. *EMBO J.* **1986**, *5*, 823–826. [[CrossRef](#)]
33. Franzosa, E.A.; Xia, Y. Structural Determinants of Protein Evolution Are Context-Sensitive at the Residue Level. *Mol. Biol. Evol.* **2009**, *26*, 2387–2395. [[CrossRef](#)]
34. Sayers, E.W.; Beck, J.; Bolton, E.E.; Brister, J.R.; Chan, J.; Connor, R.; Feldgarden, M.; Fine, A.M.; Funk, K.; Hoffman, J.; et al. Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Res.* **2025**, *53*, D20–D29. [[CrossRef](#)]
35. Yu, E.-M.; Yoshinaga, T.; Jalufka, F.L.; Ehsan, H.; Mark Welch, D.B.; Kaneko, G. The complex evolution of the metazoan HSP70 gene family. *Sci. Rep.* **2021**, *11*, 17794. [[CrossRef](#)]
36. Eddy, S.R. Profile hidden Markov models. *Bioinformatics* **1998**, *14*, 755–763. [[CrossRef](#)]
37. Potter, S.C.; Luciani, A.; Eddy, S.R.; Park, Y.; Lopez, R.; Finn, R.D. HMMER web server: 2018 update. *Nucleic Acids Res.* **2018**, *46*, W200–W204. [[CrossRef](#)]
38. Van Rossum, G.; Drake, F.L. *Python 3 Reference Manual*; Python Software Foundation; Python Language Reference, version 3.10.; CreateSpace: Scotts Valley, CA, USA, 2009.
39. Bogumil, D.; Dagan, T. Cumulative impact of chaperone-mediated folding on genome evolution. *Biochemistry* **2012**, *51*, 9941–9953. [[CrossRef](#)]
40. Almeida, F.C.L.; Sanches, K.; Pinheiro-Aguiar, R.; Almeida, V.S.; Caruso, I.P. Protein Surface Interactions—Theoretical and Experimental Studies. *Front. Mol. Biosci.* **2021**, *8*, 706002. [[CrossRef](#)]
41. Almeida, F.C.L.; Pinheiro-Aguiar, R.; Vasconcelos, A.A. Importance of Surface Hydrophobic Clusters in Proteins. In *Repeat Amino Acid Sequences in Proteins*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2025.
42. Kyte, J.; Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132. [[CrossRef](#)]
43. Gowder, S.M.; Chatterjee, J.; Chaudhuri, T.; Paul, K. Prediction and Analysis of Surface Hydrophobic Residues in Tertiary Structure of Proteins. *Sci. World J.* **2014**, *2014*, 971258. [[CrossRef](#)]
44. Kityk, R.; Kopp, J.; Sinning, I.; Mayer, M.P. Structure and dynamics of the ATP-bound open conformation of Hsp70 chaperones. *Mol. Cell* **2012**, *48*, 863–874. [[CrossRef](#)]
45. Wisniewska, M.; Karlberg, T.; Lehtio, L.; Johansson, I.; Kotenyova, T.; Moche, M.; Schuler, H. Crystal Structures of the ATPase Domains of Four Human Hsp70 Isoforms: HSPA1L/Hsp70-hom, HSPA2/Hsp70-2, HSPA6/Hsp70B', and HSPA5/BiP/GRP78. *PLoS ONE* **2010**, *5*, e8625. [[CrossRef](#)]
46. Rego, N.; Koes, D. 3Dmol.js: Molecular visualization with WebGL. *Bioinformatics* **2015**, *31*, 1322–1324. [[CrossRef](#)]
47. Kluyver, T.; Ragan-Kelley, B.; Perez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S.; et al. Jupyter Notebooks—A publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*; Loizides, F., Schmidt, B., Eds.; IOS Press: Amsterdam, The Netherlands, 2016; pp. 87–90. [[CrossRef](#)]
48. Cock, P.J.A.; Antao, T.; Chang, J.T.; Chapman, B.A.; Cox, C.J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; et al. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423. [[CrossRef](#)]
49. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [[CrossRef](#)]
50. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [[CrossRef](#)]

51. McKinney, W. Data structures for statistical computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010. [\[CrossRef\]](#)
52. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [\[CrossRef\]](#)
53. Waskom, M.L. Seaborn: Statistical data visualization. *J. Open-Source Softw.* **2021**, *6*, 3021. [\[CrossRef\]](#)
54. Aziz, M.F.; Mughal, F.; Caetano-Anollés, G. Tracing the birth of structural domains from loops during protein evolution. *Sci. Rep.* **2023**, *13*, 14688. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.