

RESEARCH ARTICLE

# Coevolution in human small Heat Shock Protein 1 is promoted by interactions between the Alpha-Crystallin domain and the disordered regions

Vanesa Racigh<sup>1\*</sup>, Luciana Rodriguez Sawicki<sup>1,2</sup>, Facundo Nicolas Eric Bravo<sup>1</sup>, Maria Silvina Fornasari<sup>1\*</sup>

**1** Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Bernal, Argentina, **2** Instituto de Investigaciones Bioquímicas de La Plata, CONICET-UNLP, Facultad de Ciencias Médicas, La Plata, Argentina

\* [vanesa.racigh@unq.edu.ar](mailto:vanesa.racigh@unq.edu.ar) (VR); [silvina@unq.edu.ar](mailto:silvina@unq.edu.ar) (MS)



## OPEN ACCESS

**Citation:** Racigh V, Rodriguez Sawicki L, Bravo FNE, Fornasari MS (2025) Coevolution in human small Heat Shock Protein 1 is promoted by interactions between the Alpha-Crystallin domain and the disordered regions. PLoS One 20(5): e0321163. <https://doi.org/10.1371/journal.pone.0321163>

**Editor:** Alexander F. Palazzo, University of Toronto, CANADA

**Received:** December 23, 2024

**Accepted:** March 3, 2025

**Published:** May 5, 2025

**Copyright:** © 2025 Racigh et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data availability statement:** <http://hdl.handle.net/11336/248901>.

**Funding:** PUNQ 827- 2282/22, Universidad Nacional de Quilmes, PI: MSF Fundación Florencio Fiorini Grants for Research in Biomedical Sciences 2024: Effect of pathogenic variants on the self-regulation mechanism of human HSPB1 chaperone activity, PI: LRS

## Abstract

Human small Heat Shock Protein 1 (HSPB1) belongs to the Small Heat Shock Protein (sHSP) superfamily, a group of ATP-independent molecular chaperones essential for cellular stress responses and protein quality control. These proteins share a conserved domain organization, with a structured Alpha-Crystallin domain (ACD) flanked by disordered N-terminal and C-terminal regions (NTR and CTR). While the prevailing evolutionary hypothesis for the sHSP family suggests that the disordered regions evolved independently and at a faster rate than the ACD, this study provides, for the first time, evidence of coevolution between these regions in human HSPB1, introducing new insights into the evolutionary mechanisms that sustain critical regulatory interactions. By integrating evolutionary and structural approaches, we estimated evolutionary rates per region and position, analyzed the composition of key interacting motifs, and employed structural modeling with Alpha-Fold 2 to assess the prevalence of these interactions. Our findings reveal that while the disordered regions globally evolve faster than the ACD, specific motifs involved in regulatory interactions exhibit lower-than-average evolutionary rates, reflecting evolutionary constraints imposed by their functional importance. This coevolutionary mechanism may also extend to other small Heat Shock Proteins featuring interacting motifs in the NTR, CTR, or both, offering a new perspective for studying their molecular evolution. Furthermore, the analysis presented in this work could be applied to assess coevolution in other proteins with intrinsically disordered regions.

## Introduction

Small Heat Shock Proteins (sHSPs) are ATP-independent molecular chaperones that act as the first line of defense in the cellular chaperone network [1]. The human genome encodes ten sHSPs (HSPB1 to HSPB10), which can be ubiquitously expressed or exhibit tissue-specific expression patterns and become upregulated under conditions of cellular stress. Their primary

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

role is to interact with misfolded clients to prevent aggregation. Variants of these proteins with altered chaperone activity are associated with several diseases, including Parkinson's, Alzheimer's, and neuropathies [2].

Structurally, sHSPs consist of a conserved Alpha-Crystallin domain (ACD) flanked by disordered N-terminal (NTR) and C-terminal (CTR) regions [3]. Depending on the subtype, sHSPs can assemble into dynamic homo- and hetero-oligomers of varying sizes, stabilized by extensive contacts between the disordered regions and the ACD [1,4,5]. In human HSPB1 and HSPB5, under stress conditions, specific phosphorylation sites in the NTR trigger the disassembly of oligomers into smaller species, freeing the NTR from oligomeric contacts and enhancing its interaction with exposed hydrophobic regions of misfolded proteins [6–8]. Experimental studies indicate that the dimeric form of these sHSPs exhibits the highest chaperone activity, making dimer constructs the standard unit for assessing chaperone function [8,9]. Each ACD dimer presents three grooves: a central groove at the dimer interface and two lateral grooves, one on each subunit. These grooves serve as interaction sites for specific motifs within the disordered regions, such as the I/V-X-I/V motif in the CTR and the conserved SRLFDQXFG motif in the NTR [3,10,11]. In higher-order assemblies, the CTR often functions as a non-covalent cross-linker between dimers by binding the I/V-X-I/V motif into the lateral groove of a neighboring dimer [5,12,13].

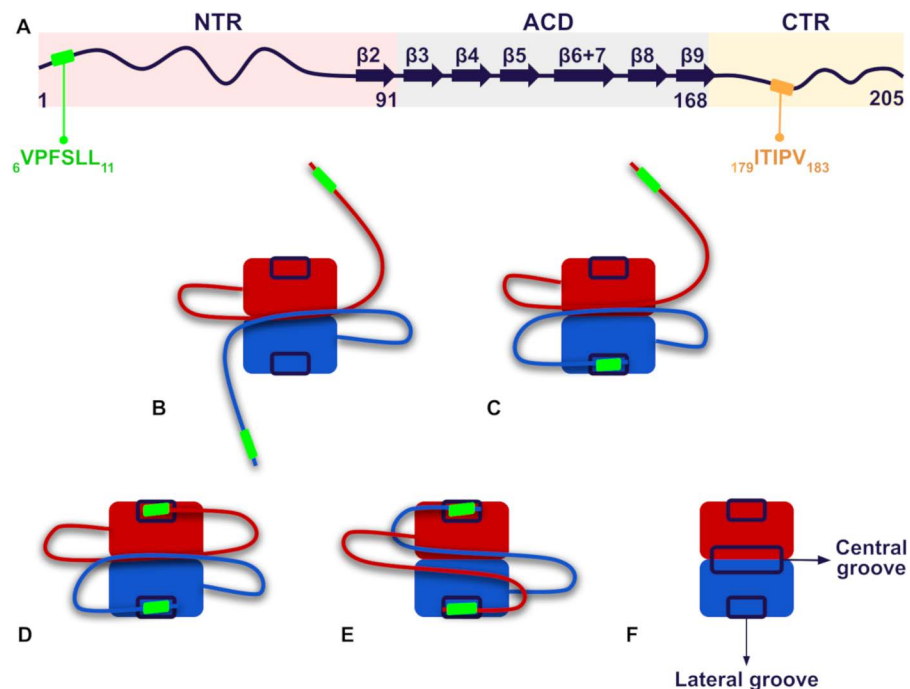
Recent work by Clouser et al. provides a detailed experimental description of the interactions between the NTR and the ACD of human HSPB1 [14]. They found, following the canonical human HSPB1 sequence numbering, that the  $_6\text{VPFSL}_{11}$  motif located at the NTR interacts with the lateral grooves in a dimeric construct (Fig 1). Interestingly, experimental results indicate that abolishing this interaction enhances in vitro human HSPB1 chaperone activity towards its natural client tau [15]. Additionally, the crystal structure of the oligomeric form of human HSPB1 reveals that an overlapping or extended I/V-X-I/V motif ( $_{179}\text{ITIPV}_{183}$ ) within the CTR also interacts with the ACD's lateral grooves [5]. Moreover, the substitutions P7R, P7S, P182A, and P182L within the  $_6\text{VPFSL}_{11}$  and  $_{179}\text{ITIPV}_{183}$  motifs are associated to Charcot-Marie-Tooth disease [16–21].

These findings underscore the crucial role of the interactions between the  $_6\text{VPFSL}_{11}$  and the  $_{179}\text{ITIPV}_{183}$  motifs in the disordered regions with the ACD in the self-regulation of human HSPB1 chaperone activity. Both oligomer formation and the reversible interactions between the NTR and the ACD at the dimeric level are fundamental to this regulatory process. Given the current evolutionary hypothesis suggesting that the disordered regions of sHSPs evolved independently and at a faster rate than the ACD [22], it is relevant to explore how these contacts have shaped the evolution of human HSPB1. To investigate the evolutionary importance of these interactions, we used a manually curated dataset of HSPB1 orthologs to estimate the global evolutionary rates for the NTR, ACD, and CTR and to calculate site-specific evolutionary rates by taking the human HSPB1 sequence as the reference. Furthermore, we derived evolutionary insights from models generated with AlphaFold 2 (AF2) [23]. Our results suggest that the ACD, NTR, and CTR of human HSPB1 did not evolve independently; instead, they must have coevolved to maintain the critical interactions necessary for regulating its chaperone activity.

## Results

### HSPB1 ortholog dataset characterization

The curated dataset comprises 474 protein sequences, 199 from invertebrates and 275 from vertebrates. No sequences from other kingdoms, including Plantae, Fungi, Eubacteria, and Protista, remained in the dataset after curation and filtering. Among vertebrates, the dataset



**Fig 1. Schematic representation of human HSPB1 structure and the interactions between the distal segment of the NTR and the lateral grooves of the ACD via the  ${}^6\text{VPFSLL}_{11}$  motif in the dimer, as described by Clouser et al. [14].** A. Primary and secondary structure of human HSPB1, indicating NTR, ACD, and CTR regions, as well as the  ${}^6\text{VPFSLL}_{11}$  and  ${}^{179}\text{ITIPV}_{183}$  motifs. B-E. The figures illustrate five possible interaction modes of the NTR's distal region within the human HSPB1 dimer: B. both NTRs are free in solution; C. the NTR of one chain interacts with the lateral groove of the same chain while the NTR of the other subunit remains free; D. both NTRs are bound to the lateral grooves of their respective chain; and E. the NTR of each chain is in contact with the lateral groove of the opposite chain. In all modes, the CTR is omitted from the representation, while the global conformation of the NTR is schematically represented, with the  ${}^6\text{VPFSLL}_{11}$  motif shown as a green rectangle. F. Lateral grooves that interact with the distal segment in the ACD dimer and the central groove that contacts other NTR regions.

<https://doi.org/10.1371/journal.pone.0321163.g001>

includes 88 mammals, 124 fish, 26 birds, 26 reptiles, and 11 amphibians. The multiple sequence alignment (MSA) analysis shows that the length of ACD is highly uniform, with an average of  $77.0 \pm 0.5$  amino acids. The NTR is longer than the CTR ( $88.8 \pm 18.3$  and  $34.7 \pm 8.4$  respectively). The CTR lengths are similar in vertebrates ( $32.1 \pm 8.1$ ) and invertebrates ( $38.2 \pm 7.5$ ). In contrast, the NTR length shows more variability between vertebrates ( $99.5 \pm 16.6$ ) and invertebrates ( $74.0 \pm 6.5$ ). This difference is due to a low-complexity region of variable length present only in vertebrates, known as the inserted segment in human HSPB1 [14].

The analysis reveals a high conservation of ACD composition across all organisms. S1 Fig shows that the percentage of each type of amino acid in the ACD is similar for most amino acids in the vertebrate and invertebrate sets. The NTR contains a higher proportion of aromatic amino acids (TRP, TYR, and PHE), which are nearly absent in the CTR; this enrichment, although unusual for intrinsically disordered regions, is commonly associated with molecular recognition regions that facilitate protein-protein interactions [24,25].

The amino acid composition of the NTR is more variable between vertebrate and invertebrate sets, with a notable presence of hydrophobic and positively charged amino acids, especially in vertebrates. In contrast, the CTR exhibits a more uniform composition across different organism sets. It contains a higher proportion of polar and negatively charged amino acids, which may help maintain the protein in solution in the dimeric context. Although both regions are disordered, they exhibit distinct compositions linked to their specific roles.

Regarding the structured domain, human HSPB1 contains a single cysteine residue in the ACD (C137), which forms a disulfide bridge upon oxidation and helps to keep the dimer bonded, as shown by experimental data (PDB structure 2N3J) [26]. Cysteines form disulfide bonds that can influence the oligomeric equilibrium of sHSPs. Thus, it is intriguing that some organisms have cysteines in the NTR and the CTR in addition to the ACD, despite their low proportion.

### NTR and CTR motifs that interact with the ACD are conserved across all HSPB1 orthologs

We further explored the composition of the  ${}_6\text{VPFSL}_{11}$  motif in the NTR and the  ${}_{179}\text{ITIPV}_{183}$  motif in the CTR. S1 Table shows compositional data. Both motifs show distinct patterns between vertebrates and invertebrates. An essential aspect of this analysis is the consideration of proline 7 in the NTR, fully conserved across both vertebrates and invertebrates, as a reference point for defining the position of the  ${}_6\text{VPFSL}_{11}$  motif in invertebrates. In vertebrates, the  ${}_6\text{VPFSL}_{11}$  motif occurs in 27.6% of sequences, while the three most frequent alternatives, VPFSL, VPFTFL, and IPFTLL, account for 62.9% of sequences.

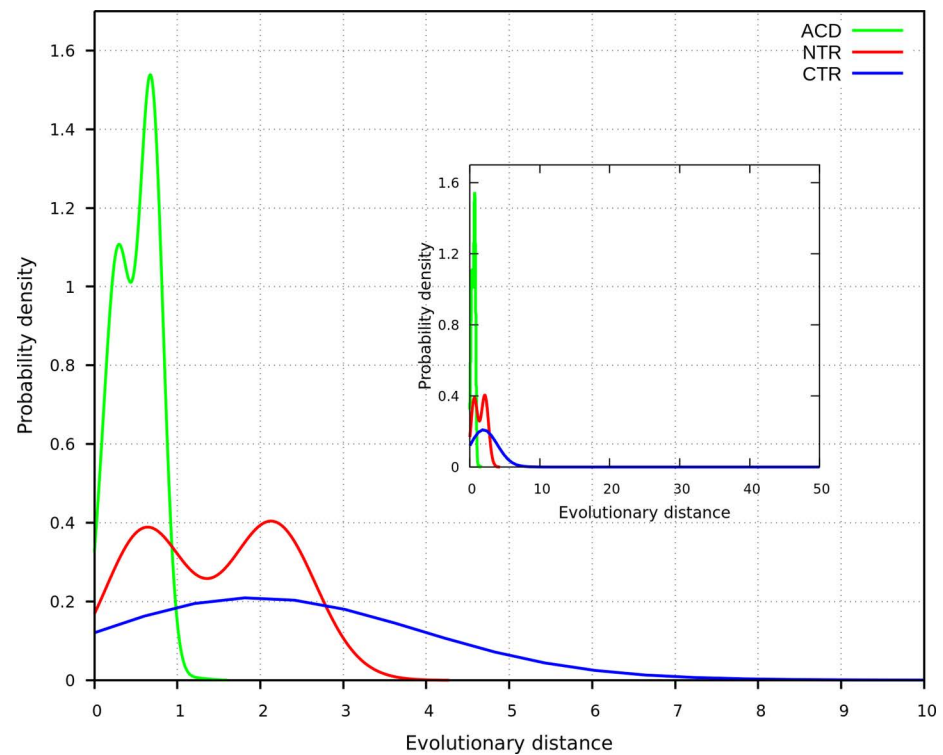
Among vertebrates, the  ${}_{179}\text{ITIPV}_{183}$  motif in the CTR is highly conserved. Alternatives of the I/V-X-I/V motif are present in 98.5% of the recruited sequences, while the extended form ITIPV is predominant, found in 56.0% of sequences. Notably, the non-extended form TTIPV occurs in 20.4% of sequences and is observed exclusively in fish species. In contrast, invertebrates display only non-extended forms of this motif. Proline 182 exhibits conservation in 87.5% of total sequences.

Analysis of the lateral grooves formed by the  $\beta 4$  and  $\beta 8$  strands within the ACD (positions  ${}_{109}\text{LTVKT}_{113}$  and  ${}_{153}\text{VSSSL}_{157}$ ) highlights significant conservation of key amino acids. In the  $\beta 4$  strand of vertebrates, leucine, valine, and lysine are the predominant amino acids, while in the  $\beta 8$  strand, valine, serine, and leucine are the most frequent ones. Although invertebrates show a more variable composition of the  $\beta 4$  and  $\beta 8$  strands, the residues primarily vary by others with chemically equivalent or similar side chains, suggesting preservation in the structural characteristics of the lateral grooves. This analysis indicates that the composition of the lateral grooves is highly conserved, and despite being located in disordered regions, the motifs interacting with the ACD are also highly conserved or exhibit variants that may fulfill roles similar to those of the  ${}_6\text{VPFSL}_{11}$  and  ${}_{179}\text{ITIPV}_{183}$  motifs in human HSPB1. Therefore, the interactions relevant to the self-regulation of human HSPB1 might be characteristic of all HSPB1 orthologs.

### Disordered regions show faster evolutionary rates than conserved ACD

As mentioned above, the current evolutionary hypothesis for the sHSP family suggests that the disordered regions evolved independently and at a faster rate than the ACD [22]. To assess this hypothesis, we divided sequences in the complete dataset into NTR, CTR, and ACD segments to create three subsets. These were employed to calculate the evolutionary distances for each pair of organisms. The underlying assumption behind this protocol is that the evolutionary time elapsed between each pair of organisms is equivalent across the three subsets, as they derive from the complete set of orthologous sequences. Consequently, if the entire protein were under the same evolutionary constraints, a similar distribution of evolutionary distances would be expected for the three regions.

As shown in Fig 2, probability density functions calculated for the evolutionary distances of the NTR, CTR, and ACD display distinct patterns. This observation indicates that these regions have evolved at different rates. The distribution for the ACD (green curve) exhibits



**Fig 2. Probability densities of evolutionary distances for the disordered regions and the ACD.** The inset displays the full range of distance values.

<https://doi.org/10.1371/journal.pone.0321163.g002>

a sharp peak at low evolutionary distances, likely due to the structural constraints limiting the variability of the ACD. In contrast, the distributions for the NTR and CTR are broader, suggesting that these disordered regions are less constrained and have accumulated more sequence changes over time. Furthermore, the NTR distribution is slightly shifted towards lower distances relative to the CTR, implying that this region may be under stronger evolutionary pressure.

Moreover, the NTR distribution appears bimodal corresponding the first peak to intra-group comparisons within vertebrates and invertebrates. It is associated with shorter evolutionary distances, suggesting recent divergence within each group. The second peak, arising from comparisons between vertebrate and invertebrate sequences, reflects higher evolutionary distances and more divergence between these lineages. A similar pattern is observed in the ACD curve, though the distance between the peaks is smaller, highlighting the higher conservation of this domain across species. The inset within Fig 2 shows the full range of evolutionary distances. Distributions are centered within a limited range, although some sequences in the CTR encompass significantly higher distances, reflecting more extensive divergence in that region.

The statistical analysis, as shown by the Kolmogorov-Smirnov (KS) test, supports these observations. The KS statistic values for the comparisons between NTR and ACD, CTR and ACD, and NTR and CTR are 0.60, 0.80, and 0.31, respectively, with p-values  $< 1e-16$  for all comparisons. These results confirm significant differences in the evolutionary rates of these regions. The broader and more right-shifted probability density distributions for the NTR and CTR, compared to the ACD, imply that these regions have accumulated more evolutionary

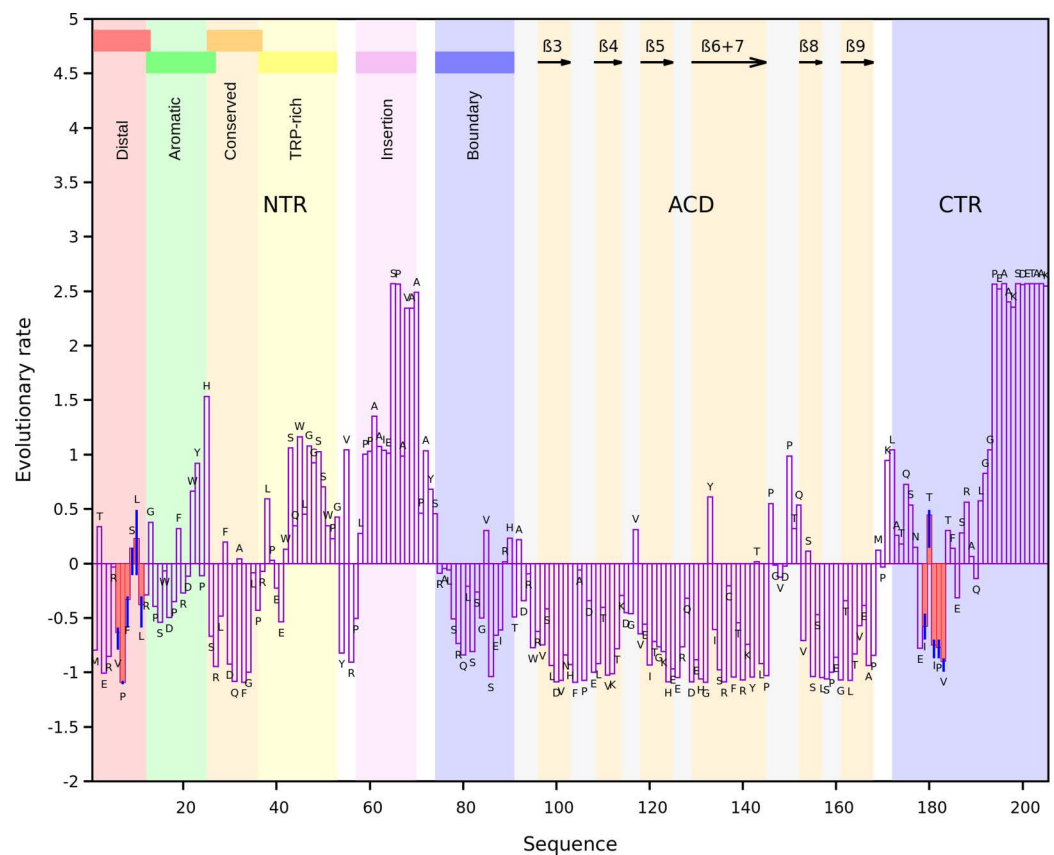


changes. Specifically, the higher evolutionary distances observed for the CTR imply that it has experienced the fastest rate of evolution, followed by the NTR, with the ACD being the slowest.

### Motifs in disordered regions interacting with the ACD show reduced evolutionary rates

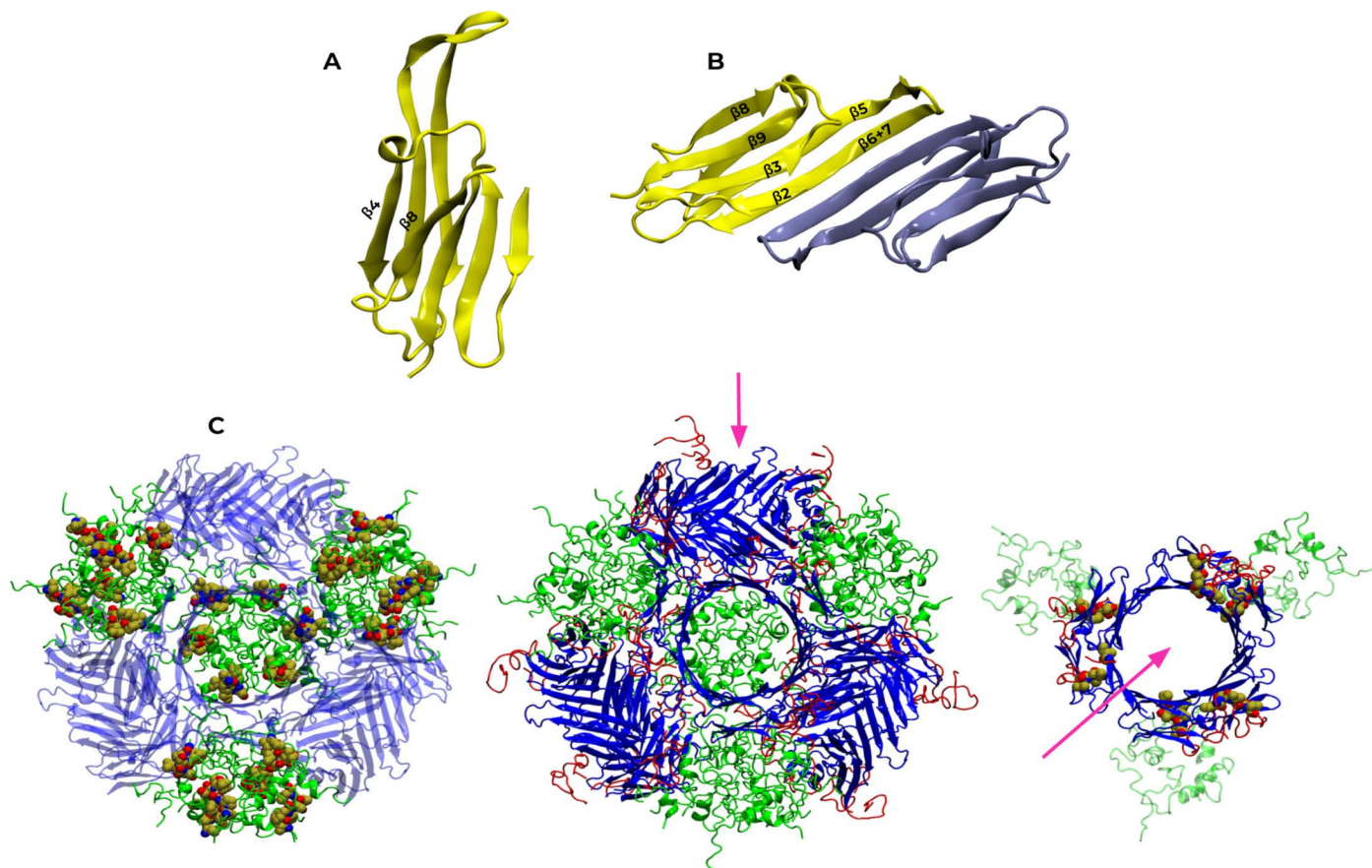
In light of the previous results, we estimated the evolutionary rates per site of the human HSPB1 sequence using the complete dataset. The profile obtained (Fig 3) reveals a clear distinction in the rates of evolution across different regions of the protein. Residues that constitute the ACD predominantly exhibit evolutionary rates lower than the average rate of the complete sequence, while the NTR and the CTR exhibit a more heterogeneous pattern.

Within the ACD, the  $\beta 4$  and  $\beta 8$  strands framing the lateral grooves (Fig 4A) show lower-than-average rates that may result from structural constraints to preserve the ACD structure and functional requirements, as these strands not only interact with motifs in the disordered regions but also with other proteins as it is the case of co-chaperone BAG3 and



**Fig 3. Per position evolutionary rate profile for the canonical human HSPB1 sequence estimated using the full ortholog dataset.** The background highlights the NTR, ACD, and CTR. The NTR consists of six segments [14]: distal (residues 1–13), aromatic (residues 12–27), conserved (residues 25–37), Trp-rich (residues 37 to 53), inserted (residues 57–70), and boundary (residues 74–91). Black arrows indicate the  $\beta$  strands of the ACD. Red highlights the  ${}^6\text{VPFSL}_{11}$  motif of the NTR and the  ${}^{179}\text{ITIPV}_{183}$  motif of the CTR. The QQ intervals, shown in blue for these motifs, represent the interquartile range (P25–P75) of the estimated evolutionary rates, capturing the range within which the central 50% of values fall, providing insight into their variability.

<https://doi.org/10.1371/journal.pone.0321163.g003>



**Fig 4. Structural overview of human HSPB1 highlighting the ACD and its oligomeric assembly.** A. ACD structure of human HSPB1 (PDB ID 4MJH, chain A), showing  $\beta 4$  and  $\beta 8$  strands forming the lateral groove. B. Top view of the ACD dimer, indicating the positions of the  $\beta$ -strands. C. 24-subunit oligomeric structure of human HSPB1 (PDB ID 6DV5) displaying four circular arrangements of three dimeric ACDs (blue) held together through extensive interactions mediated by the NTRs (green). The CTRs of each subunit within the dimers (red) are in contact with the lateral grooves of neighboring subunits. The pink arrow indicates the perspective of the top view on the right, showing one of the four arrangements of six subunits that make up the oligomer. This structure highlights the  $_{179}\text{ITIPV}_{183}$  motif of each subunit interacting with the lateral groove of the adjacent subunit. The left structure shows  $_{6}\text{VPFSLL}_{11}$  residues within the NTRs arrangement in sphere representation.

<https://doi.org/10.1371/journal.pone.0321163.g004>

client proteins [3,27]. The CTR generally exhibits higher-than-average evolutionary rates, except in positions adjacent to and within the  $_{179}\text{ITIPV}_{183}$  motif, whose rates fall below the average. Residues T180 and P182 that correspond to the variable position in the I/V-X-I/V motif exhibit opposite evolutionary rates: T180 has a high rate, whereas P182 shows a low rate. This may result from the absence of the extended motif form in invertebrates and fish, leading to T180's lack of conservation compared to P182's high prevalence across organisms.

In terms of NTR, it was divided into six segments (distal, aromatic, conserved, Trp-rich, inserted, and boundary) as defined by Clouser et al [14] (Fig 3). These authors used a phosphomimetic construct of human HSPB1 in their work. Hereafter, any reference to experimental results related to the dimeric form will pertain to this construct. Their study shows that four NTR segments interact with the ACD (distal, aromatic, conserved, and boundary). Interestingly, the evolutionary rates estimated for the corresponding positions show lower-than-average values despite being located in a disordered region. The distal segment, which contains the  $_{6}\text{VPFSLL}_{11}$  motif, engages with the lateral grooves of the ACD. The positions within this motif exhibit evolutionary rates below the sequence average, except for residues

$_{9}\text{SL}_{10}$ , which show slightly higher values due to their variability in the alignment. However, the QQ interval encompasses both negative and slightly positive values, indicating fluctuations around the sequence average rate. The conserved segment includes the  $_{26}\text{SRLFDQXFG}_{34}$  motif and interacts with the central groove. The aromatic segment also interacts with the ACD. In vertebrates, this region contains alternative sequences of a characteristic motif in the NTR of sHSPs,  $_{16}\text{WDPF}_{19}$  [28], involved in chaperone oligomerization [29,30]. Additionally, the boundary segment adopts an antiparallel  $\beta$ -sheet conformation ( $\beta 2$ ) with the  $\beta 3$  strand of the ACD within the same chain (Fig 4.B) [14]. Thus, the reduced evolutionary rates observed for these segments could originate from a differential selective pressure due to the functional importance of their interactions with the ACD, contrary to the behavior observed for the Trp-Rich and the inserted segments. These do not interact consistently with the ACD in the dimeric form. In particular, the insertion segment behaves as a solvated random coil [14]. That might explain the higher rates observed for positions within these segments, as solvent-exposed regions tend to evolve more rapidly than those with lower accessible surface areas [31]. Also, the length and composition of the inserted segment vary significantly among vertebrates, contributing to the higher evolutionary rates observed in this region.

Analysis of the rate profile must consider both contacts at the dimeric level and those involved in large oligomer formation. Therefore, we mapped the interactions in the 24-subunit oligomeric structure of human HSPB1 (PDB structure 6DV5). This structure is an arrangement of four groups of three dimer pairs connected through extensive NTR interactions (Fig 4C). The inserted segment and part of the boundary segment were not solved due to their high mobility [5]. In this structure, the NTRs do not establish interchain interactions with the ACD (S2 Table). On the other hand, contacts between the CTR and the ACD represent 10.5% of the total interchain interactions in the oligomer (S2 Table). The CTR of each chain engages the lateral groove of the neighboring one, positioning the  $_{179}\text{ITIPV}_{183}$  motif within its groove. Nevertheless, although the PDB structure constitutes a valuable source of information about the complex topology, it only captures a single conformation of the oligomer. Thus, the number and nature of the interactions in this structure might differ in a dynamic context.

Additional experimental evidence supports the significance of the interactions between the  $_{6}\text{VPFSL}_{11}$  and the  $_{179}\text{ITIPV}_{183}$  motifs with the lateral grooves. Phosphomimetic mutations (S15D, S78D, and S82D) alone do not prevent oligomerization, as the  $_{179}\text{ITIPV}_{183}$  motif still interacts with the lateral grooves [9], and substituting it with  $_{179}\text{GTGPG}_{183}$  is essential to isolate a dimeric form. Moreover, structural data from all oligomeric sHSPs in the PDB show that the conserved I/V-X-I/V motif interacts with the ACD's lateral grooves [11,32,33], while NTR arrangements vary significantly [1,33]. Oligomer polydispersity, diversity in NTR arrangements, variability in NTR length, and the lack of PDB structures for many sHSPs make it imprudent to assess  $_{6}\text{VPFSL}_{11}$  role in oligomerization based solely on the 24-mer structure. Notably, a human HSPB1 construct lacking residues 1-14 can still form oligomers [34], indicating oligomerization is not entirely dependent on this region. Conservation of the motif across orthologs and its role in regulating dimeric chaperone activity [15] suggest selective pressures likely constrain these positions to preserve this role.

## Structural modeling of human HSPB1 with AlphaFold 2 reveals insights on coevolution

Structural data is available for the NTR in the phosphomimetic dimer of human HSPB1 [14]. However, there is currently no available structure exhibiting the interaction between the  $_{6}\text{VPFSL}_{11}$  motif and the lateral grooves in the wild-type dimer. Therefore, we used AF2 to



generate 500 structures of this dimer to investigate whether this interaction is predicted and to evaluate the frequency of this contact across the models. We used this approach because, unlike our manually curated dataset of 474 orthologous sequences, AF2 does not limit its MSA to strict orthologs. Instead, it automatically recruits homologous sequences, including both orthologs and paralogs, using JackHMMER [35] and HHblits [36] to search UniRef90, BFD, and MGnify. Through this process, AF2 initially retrieved 10019 sequences but applied an internal filtering mechanism to reduce redundancy and maximize sequence diversity, capping the final MSA at 2048 sequences for computational efficiency [37]. This large-scale sequence sampling enables AF2 to leverage coevolutionary signals from a more divergent MSA, allowing us to assess the prevalence of the relevant interactions within this broader evolutionary context.

Analysis of the models reveals conservation in the structure of the ACD (residues 92 to 168) across all models. RMSD for backbone atoms was  $0.51 \pm 0.15$ , using as reference the crystal structure of a dimeric construct containing only the ACD of human HSPB1 (PDB ID 4MJH). On the contrary, the disordered regions adopt various conformations (S2 Fig). The average per residue estimate of confidence (pLDDT) exhibited values above 70 with low standard deviation for the residues forming the ACD, which indicates that this region is well modeled (S3 Fig). In particular, positions with values close to or exceeding 90 suggest high model accuracy. In contrast, pLDDT scores for the residues within the CTR were below 50, as expected due to their disordered nature [38]. On the other hand, the high pLDDT scores for the distal segment of the NTR containing the  ${}_6\text{VPFSL}_{11}$  motif align with AF2's capacity to identify conditionally folded intrinsically disordered regions. This interaction-driven structural stabilization of the NTR likely reflects the coevolutionary signals captured in the MSA used by AF2 [39].

Using the interaction fingerprint obtained from the model set, we assessed whether the  ${}_6\text{VPFSL}_{11}$  motif was in contact with the lateral grooves of the ACD. We analyzed each chain separately, and the contacts present were symmetric between the two chains. Our results reveal that in 98.2% of the predicted structures, the  ${}_6\text{VPFSL}_{11}$  motif was in contact with one of the lateral grooves of the dimer through Van der Waals and H-bond interactions (Fig 5). In the remaining 1.8%, the NTR was not in contact with the lateral grooves but instead adopted an extended conformation. None of the models showed the  ${}_{179}\text{ITIPV}_{183}$  motif of the CTR interacting with the grooves. For the  ${}_6\text{VPFSL}_{11}$  motif, 68.6% of the interactions were intrachain contacts, while 29.6% were interchain contacts. Both types of interactions are consistent with the quasi-ordered states that human HSPB1 can adopt [14].

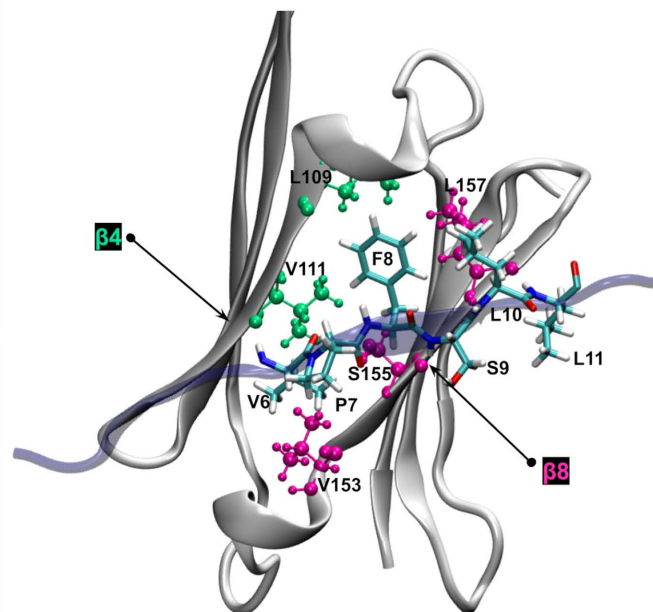
Further analysis to characterize residue pairs that constitute this interaction in the models reveals that the F8-S155 pair interacts consistently in nearly 100% of models displaying contacts, regardless of whether these interactions are intra or interchain. This observation aligns with findings by Baughman et al., who reported that substitutions at these specific positions (F8G and S155Q) in the dimer of human HSPB1 release its NTR enhancing its chaperone activity toward the natural client tau [15]. In a complementary analysis, we generated 25 additional models with AF2 configured to rely exclusively on MSA data, to observe how AF2 infers contacts relying only on sequence-derived evolutionary signals. During recycling, AF2 generates distograms as intermediate outputs, which predict the probabilistic distances between residue pairs and help refine structural predictions. Although these distograms are not final outputs, they provide insight into the algorithm assessment of inter-residue relationships. In these template-free models, the ACD structure of the dimers remained conserved.

The distogram shown in Fig 6 was generated by averaging the distograms from individual models, and standard deviations were calculated for each point. Residues 1-91, 92-168, and 169-205 of Chain A and their respective counterparts in Chain B (206-410) represent the

Intrachain interactions (341 out of 500 models)			
Interacting pair	Prevalence		Type
	$\text{{}_6\text{VPFSL}}_{11}$ Chain A/ Lateral groove Chain A	$\text{{}_6\text{VPFSL}}_{11}$ Chain B/ Lateral groove Chain B	
V6 - V111	100%	100%	VdW
F8 - F155	99.4%	100%	VdW
L11 - L157	99.4%	99.4%	H-bond
S9 - S155	99.4%	99.4%	H-bond
V6 - V153	61.5%	60.9%	VdW
F8 - L109	28.6%	29.4%	VdW
L10 - L157	1.5%	3.2%	VdW
P7 - S155	0.0%	0.6%	VdW
V6 - S155	0.0%	0.6%	VdW

Interchain interactions ( 148 out of 500 models)			
Interacting pair	Prevalence		Type
	$\text{{}_6\text{VPFSL}}_{11}$ Chain A/ Lateral groove Chain B	$\text{{}_6\text{VPFSL}}_{11}$ Chain B/ Lateral groove Chain A	
V6 - V111	100%	100%	H-bond
F8 - S155	100%	100%	VdW
L11 - L157	100%	100%	H-bond
S9 - S155	100%	100%	Vdw
V6 - V153	64.2%	60.1%	VdW
L10 - L157	5.4%	0.7%	VdW
F8 - L109	0.7%	0.0%	VdW

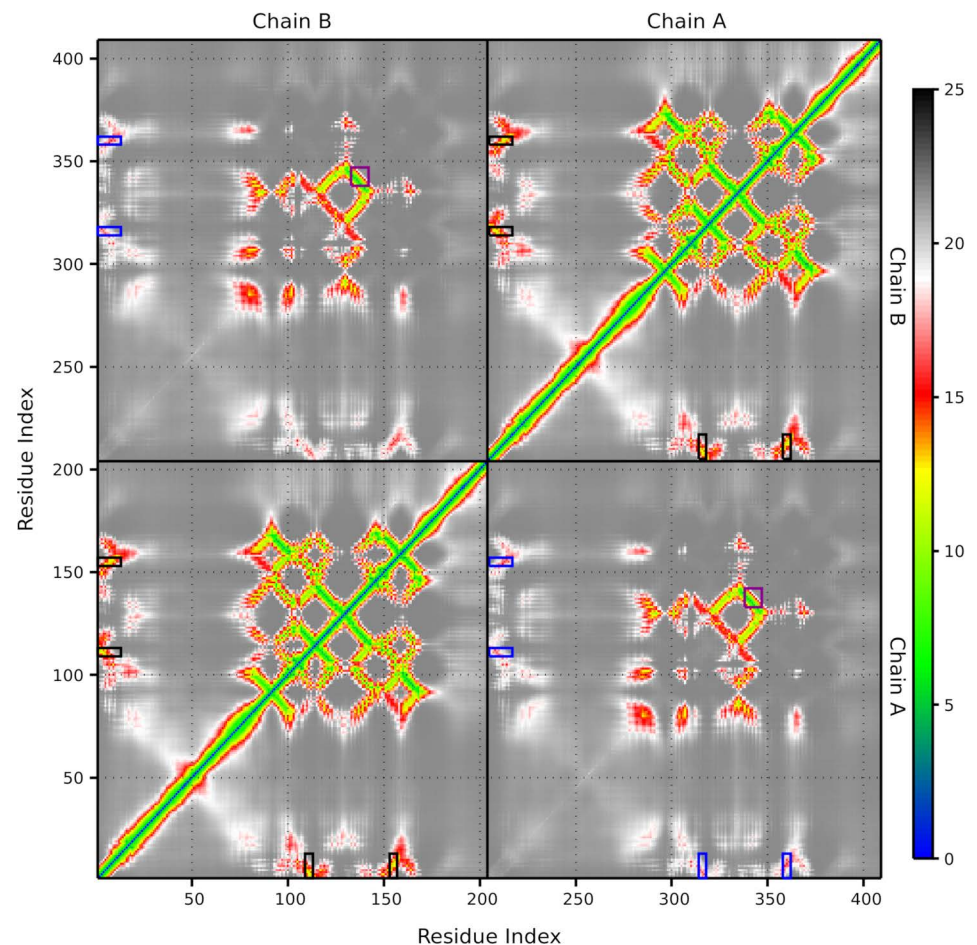


**Fig 5. Summary of intrachain and interchain interactions between the  $\text{{}_6\text{VPFSL}}_{11}$  motif and the lateral grooves of the ACD ( $\beta 4$  and  $\beta 8$  strands) in human HSPB1 dimer models.** The prevalence of interactions is shown as a percentage of the total number of models in which each type of contact (intra or interchain) occurs, along with the nature of the contact (Van der Waals or H-bond). The right panel highlights the residues involved in the interactions in the table.

<https://doi.org/10.1371/journal.pone.0321163.g005>

NTR, ACD, and CTR, respectively. The ACD forms the central block, with shorter intrachain distances ( $\sim 5$  Å) shown in green, reflecting conserved interactions within each ACD. The purple rectangles highlight the  $\beta 6/7$  strands (133–142), forming interchain contacts around  $\sim 5$  Å, consistent with the structured and conserved nature of the ACD dimer interface. Residues belonging to the CTR do not appear to establish contacts with the ACD. In contrast, the distal segment of the NTR of each chain (residues 1 to 13) comprising the  $\text{{}_6\text{VPFSL}}_{11}$  motif and the  $\beta 4$  and  $\beta 8$  strands (residues 109 to 113 and 153 to 157) are predicted to be in contact. Intrachain contacts between this segment and each strand are marked by black rectangles, with distances ranging from  $\sim 10$  Å to  $15$  Å. Interchain contacts, highlighted in blue, show distances of  $\sim 15$  Å to  $20$  Å. These values suggest transient or flexible associations rather than direct atomic interactions. It is important to note that these distances are not final, as seen in the relaxed models used for evaluating contact occurrence with Prolif. The standard deviation for distances corresponding to the ACD remains close to 0, indicating consistent predictions; in contrast, the NTR-lateral groove contacts exhibit deviations up to  $2.5$  Å, reflecting greater flexibility in these regions (data not shown).

Notably, even in this template-free setup, where AF2 utilized a more divergent MSA that includes paralogous sequences, the interaction between the NTR and the lateral groove was still predicted. This suggests that these contacts are strongly encoded in the evolutionary



**Fig 6. Average distogram generated from individual distograms produced by AF2 for human HSPB1 dimer models without including the PDB70 database.** Colors represent the probabilistic distances between residue pairs, with blue indicating short distances and red indicating long distances. Black and blue rectangles highlight the intra-chain and interchain contacts between the distal segment containing the  ${}_6\text{VPFSL}_{11}$  motif and the  $\beta 4$  and  $\beta 8$  strands that shape the lateral grooves. Purple rectangles emphasize the intersection of the  $\beta 6/7$  strands from each chain (residues 133–142). Dividing lines denote the areas corresponding to each chain, with residue index ranging from 1 to 205 for Chain A and 206 to 410 for Chain B.

<https://doi.org/10.1371/journal.pone.0321163.g006>

sequence information, regardless of whether the MSA is restricted to orthologs or includes a broader set of homologs.

## Discussion

In this work, we explore the implications of the interactions between the  ${}_6\text{VPFSL}_{11}$  motif of the NTR and the  ${}_{179}\text{ITIPV}_{183}$  motif of the CTR with the ACD in the evolution of human HSPB1. These contacts play a key role in its chaperone activity self-regulation: the  ${}_6\text{VPFSL}_{11}$  motif has a regulatory role at the dimeric level, while the  ${}_{179}\text{ITIPV}_{183}$  motif contributes to oligomer formation [14,15,40]. The current evolutionary hypothesis for the sHSPs superfamily suggests that the NTR and CTR evolved independently and at different rates than the conserved ACD [22]. Since residues involved in functionally relevant interactions are often evolutionarily constrained and tend to coevolve to maintain their roles [41], we conducted

evolutionary analysis and structural modeling to assess coevolution between the disordered regions and the ACD of human HSPB1.

Therefore, we worked with a manually curated set of orthologous sequences of human HSPB1. This dataset revealed that the amino acid composition of the ACD is conserved in both vertebrates and invertebrates, while that of the NTR and CTR exhibit higher variability. MSA analyses indicate that the composition of the lateral grooves of the ACD is highly conserved. Furthermore, the positions of the  ${}_6\text{VPFSL}_{11}$  and  ${}_{179}\text{ITIPV}_{183}$  motifs are either preserved in orthologs or replaced by amino acids with comparable physicochemical properties, likely maintaining their functional interaction with the ACD. This result is consistent with many motif-binding domains exhibiting weak specificity, interacting primarily with a small core of residues while tolerating substitutions that retain essential binding characteristics, thereby allowing critical interactions to persist despite evolutionary divergence [42,43].

Furthermore, we replicated the methodology used by Kriehuber et al. in their work to assess whether our dataset mirrors the evolutionary pattern they reported [22], even though it contains fewer divergent sequences. Assuming uniform evolutionary constraints would produce similar rates across all regions, our findings show distinct rates: the ACD evolves the slowest, followed by the NTR, with the CTR evolving the fastest. Although disordered regions collectively evolve faster than the ACD, individual positions do not evolve uniformly. In particular, sites involved in ACD interactions, such as the  ${}_6\text{VPFSL}_{11}$  and  ${}_{179}\text{ITIPV}_{183}$  motifs, evolve more slowly than the average of the complete sequence, likely due to selective pressures preserving these critical interactions. This result explains why the NTR, which is longer than the CTR, has a lower overall evolutionary rate, as it contains an increased number of ACD-interacting motifs.

Furthermore, models of human HSPB1 dimer generated with AF2 capture the interaction between the  ${}_6\text{VPFSL}_{11}$  motif and the lateral grooves of the ACD according to the quasi-ordered states described for this protein [14]. Notably, models generated without templates also predict these contacts. This consistency indicates that the coevolutionary-like relationships inferred by AF2 are grounded in evolutionary constraints, supporting their use as evidence of coevolution in our analyses.

Altogether, this evidence suggests that, although the disordered regions of HSPB1 exhibit higher substitution rates on average compared to the ACD, they have not evolved independently; instead, they have likely coevolved with the ACD to preserve essential interactions necessary for chaperone activity regulation. These findings align with studies on proteins with disordered regions, which often show evolutionary conservation at positions or motifs within these regions that interact with ordered domains [1,44,45].

Moreover, this evolutionary perspective provides a context for understanding why substitutions in conserved motifs within disordered regions, such as P7S and P7R variants in the  ${}_6\text{VPFSL}_{11}$  motif or the P182A and P182L variants in the  ${}_{179}\text{ITIPV}_{183}$  motif, are associated with Charcot-Marie-Tooth disease [16,21]. These substitutions likely alter the motif-ACD interaction, potentially affecting human HSPB1 chaperone function, as even a single amino acid change at a critical site can be enough to disrupt binding [43]. In this line, a study shows that P182, while not directly interacting with the ACD, restricts conformational flexibility in the  ${}_{181}\text{IPV}_{183}$  motif, facilitating its interaction with the lateral grooves. Substituting proline with leucine increases the motif's flexibility, reducing its binding affinity to the ACD [16].

Interactions between specific motifs from the disordered regions (I/V-X-I/V variants in the NTR and CTR) occur in other human paralogs [13,46,47]. Competition for the lateral grooves between these motifs has been reported for human HSPB5 [48], similar to the interactions seen with  ${}_6\text{VPFSL}_{11}$  and  ${}_{179}\text{ITIPV}_{183}$  described by Clouser et al. [14]. Given that various human paralogs feature interaction motifs in the NTR, CTR, or both [3], it is plausible that



coevolutionary processes driven by selective pressures to maintain critical interactions could also occur in other members of the sHSP family. Further studies integrating structural and evolutionary information with the impact of disease-associated variants on chaperone activity could improve our understanding of the functional diversity within the sHSP family.

## Materials and methods

### HSPB1 ortholog dataset creation

Homologous protein sequences were initially recruited from the UniprotKB database [49] with Uniprot BLASTp [50], using the canonical human HSPB1 (UniProt ID P04792) as a query. To ensure a comprehensive dataset, additional sequences were sourced by filtering according to specific taxonomic groups. The recruitment process followed the phylogenetic tree provided by Hedges et al. [51], focusing on major kingdoms: Animalia, Plantae, Eubacteria, Fungi, and Protista. When the maximum number of sequences was reached for a kingdom, further recruitment was conducted within subgroups of that kingdom to capture broader diversity. After recruitment, an initial filtering step was applied to remove sequences that were partial, hypothetical, or contained indeterminate residues. Duplicated sequences were also eliminated to avoid redundancy. The remaining sequences were aligned and a threshold of 30% identity and 40% coverage was applied using an in-house program that implements Biopython [52]. The resulting dataset was manually curated to compile a set of orthologous HSPB1 sequences for subsequent analysis. The importance of working with proteins coded by orthologous genes is related to studying the evolutionary history of the product of a single gene, ensuring that these are the same protein in different organisms and, consequently, perform the same function.

### Dataset characterization

Two subsets were generated from the ortholog dataset, one for vertebrates and another for invertebrates. The three sets were aligned using Clustal Omega within UGENE [53]. The amino acid composition and the observed substitutions of the  ${}_6\text{VPFSL}_{11}$  motif within the NTR and the  ${}_{179}\text{ITIPV}_{183}$  (I/V-X-I/V variant) in the CTR were analyzed from the alignments. Additionally, the amino acid composition of the  $\beta 4$  and  $\beta 8$  strands (positions  ${}_{109}\text{LTVKT}_{113}$  and  ${}_{153}\text{VSSSL}_{157}$ ) was examined.

### Structural and evolutionary analysis

The ortholog dataset was divided into three subsets, each corresponding to a region of the protein (NTR, ACD, and CTR). Each subset was subsequently aligned. Evolutionary distances within the NTR, CTR, and ACD datasets were calculated with the protdist program from the PHYLIP package with default parameters [54], following the protocol described by Kriehuber et al. [22]. To visualize the distribution of these evolutionary distances, we applied Kernel density estimation to the flattened distance matrices. Differences in evolutionary rates between these regions were statistically assessed through the Kolmogorov-Smirnov test implemented in Python. The evolutionary rate per site was calculated employing the Rate4Site program with default parameters [55]. The analysis was conducted on the HSPB1 ortholog dataset, with the canonical human HSPB1 sequence as the reference.

To assess the prevalence of the interaction between the  ${}_6\text{VPFSL}_{11}$  and  ${}_{179}\text{ITIPV}_{183}$  motifs and the lateral grooves of human HSPB1, first, the intra and interchain interactions in the 24-mer oligomer of human HSPB1 (PDB ID 6DV5) were mapped with the RING web server [56]. Subsequently, structural models of the human HSPB1 dimer were generated with



AF2 [23] on a local installation with A30 GPUs. The canonical sequence of human HSPB1 was used as input, along with the UniRef90, MGnify, and BFD sequence databases, and the PDB70 structural database. A total of 500 relaxed structures were generated (5 models with 100 predictions each) in PDB format. The files containing the structures were reformatted with pdb4amber into a PDB format compatible with the ProLIF toolkit [57] and stacked into a single file for processing with CPPTRAJ, which was also used to calculate the RMSD relative to the PDB structure 4MJH [46]. Both tools are part of the AmberTools package [58]. The combined file was then analyzed with the ProLIF toolkit to obtain an interaction fingerprint between the residues of the  ${}_6\text{VPFSL}_{11}$  and  ${}_{179}\text{ITIPV}_{183}$  motifs and those forming the  $\beta 4$  ( ${}_{109}\text{LTVKT}_{113}$ ) and  $\beta 8$  ( ${}_{153}\text{VSSSL}_{157}$ ) strands. A contact was defined based on ProLIF's detection of interactions between any residue in the motifs and any residue in the  $\beta 4$  or  $\beta 8$  strands. Per residue pLDDT values were extracted from the PDB files of the relaxed models using custom Python scripts. Structural models were visualized with VMD [59].

Next, to observe the contacts inferred by AF2 based solely on the MSA information, an additional set of 5 models of the HSPB1 dimer was generated, with 5 predictions per model (25 structures). Templates from the PDB70 database were not included in this prediction. The distance histograms (distograms) information generated by AF2 for each model were extracted from the.pkl files employing the dgram2dmap tool [60].

## Supporting information

**S1 Fig. Amino acid composition of the NTR, CTR, and ACD in the HSPB1 sequences from the full dataset, as well as the vertebrate and invertebrate subsets.**

(TIF)

**S2 Fig. Representative conformations of the disordered NTR and CTR in human HSPB1 models generated with AlphaFold 2.** Models A and B depict interchain and intrachain interactions between the NTR and the ACD, with the CTR remaining in solution. Model C shows a conformation where neither the NTR nor the CTR interacts with the ACD. This modeling aimed to determine whether interactions between the  ${}_6\text{VPFSL}_{11}$  motif and the ACD could be captured, regardless of the specific conformations adopted by the rest of the NTR. Models D and E display two perspectives of the same conformation from a dimer model generated using HSPB1's phosphomimetic sequence, modeled with ColabFold [61] and employing PDB 4MJH as a template. In the highest-ranked model among the five provided by the server, the distal (which adopts a  $\beta$ -sheet structure upon binding to the lateral grooves), aromatic, conserved, Trp-rich, inserted, and boundary segments adopt conformations consistent with the experimental description provided for each region by Clouser et al., while the CTR remains in solution.

(TIF)

**S3 Fig. Average pLDDT values for each residue across 500 models of the human HSPB1 dimer generated with AlphaFold 2.** Residues 1 to 205 correspond to Chain A, and residues 206 to 410 correspond to Chain B. Background colors highlight the residue ranges for the NTR, ACD, and CTR. Green error bars indicate the standard deviation for key residues across different regions: P7 (from the  ${}_6\text{VPFSL}_{11}$  motif), F29 (from the conserved  ${}_{26}\text{SRLFDQXFG}_{34}$  motif), L109 ( $\beta 4$  strand), V153 ( $\beta 8$  strand), and P182 (from the  ${}_{179}\text{ITIPV}_{183}$  motif).

(TIF)

**S1Table. Per-position percentage composition of amino acids in the sequence alignment of vertebrates, invertebrates, and the full dataset for the VPFSLL, I/V-X-I/V motifs, and the  $\beta 4$  and  $\beta 8$  strands.** For the motifs located in disordered regions, the three most frequent

alternative motifs in each alignment are indicated. Only the percentage of the predominant amino acids (up to three) is shown in each position.

(DOCX)

**S2 Table. Percentage of intra and interchain interactions involving the NTR (residues 1–91), ACD (residues 92–168), and CTR (residues 169–205) in the 24-mer structure of human HSPB1 (PDB ID 6DV5).** Values represent the proportion of interactions between and within each region relative to the total mapped interactions.

(DOCX)

## Acknowledgments

This work used computational resources from CCAD – Universidad Nacional de Córdoba (<https://ccad.unc.edu.ar/>), which is part of SNCAD – MinCyT, República Argentina.

## Author contributions

**Conceptualization:** Vanesa Racigh, Luciana Rodriguez Sawicki, María Silvina Fornasari.

**Data curation:** Vanesa Racigh, Facundo Nicolas Eric Bravo.

**Formal analysis:** Vanesa Racigh, María Silvina Fornasari.

**Funding acquisition:** Luciana Rodriguez Sawicki, María Silvina Fornasari.

**Investigation:** Vanesa Racigh, Luciana Rodriguez Sawicki, María Silvina Fornasari.

**Methodology:** Vanesa Racigh, María Silvina Fornasari.

**Supervision:** María Silvina Fornasari.

**Visualization:** Vanesa Racigh, Luciana Rodriguez Sawicki.

**Writing – original draft:** Vanesa Racigh, María Silvina Fornasari.

**Writing – review & editing:** Vanesa Racigh, Luciana Rodriguez Sawicki, Facundo Nicolas Eric Bravo, María Silvina Fornasari.

## References

- Haslbeck M, Weinkauff S, Buchner J. Small heat shock proteins: Simplicity meets complexity. *J Biol Chem.* 2019;294(6):2121–32. <https://doi.org/10.1074/jbc.REV118.002809> PMID: [30385502](#)
- Tedesco B, Ferrari V, Cozzi M, Chierichetti M, Casarotto E, Pramaggiore P, et al. The Role of Small Heat Shock Proteins in Protein Misfolding Associated Motoneuron Diseases. *Int J Mol Sci.* 2022;23(19):11759. <https://doi.org/10.3390/ijms231911759> PMID: [36233058](#)
- Klevit RE. Peeking from behind the veil of enigma: emerging insights on small heat shock protein structure and function. *Cell Stress Chaperones.* 2020;25(4):573–80. <https://doi.org/10.1007/s12192-020-01092-2> PMID: [32270443](#)
- Mymrikov EV, Riedl M, Peters C, Weinkauff S, Haslbeck M, Buchner J. Regulation of small heat-shock proteins by hetero-oligomer formation. *J Biol Chem.* 2020;295(1):158–69. <https://doi.org/10.1074/jbc.RA119.011143> PMID: [31767683](#)
- Nappi L, Aguda A, Nakouzi N, Lelj-Garolla B, Beraldi E, Lallous N, et al. Ivermectin inhibits HSP27 and potentiates efficacy of oncogene targeting in tumor models. *The Journal of Clinical Investigation.* 2020.
- Mühlhofer M, Peters C, Kriehuber T, Kreuzeder M, Kazman P, Rodina N, et al. Phosphorylation activates the yeast small heat shock protein Hsp26 by weakening domain contacts in the oligomer ensemble. *Nat Commun.* 2021;12(1):6697. <https://doi.org/10.1038/s41467-021-27036-7> PMID: [34795272](#)
- Kostenko S, Moens U. Heat shock protein 27 phosphorylation: kinases, phosphatases, functions and pathology. *Cell Mol Life Sci.* 2009;66(20):3289–307. <https://doi.org/10.1007/s00018-009-0086-3> PMID: [19593530](#)
- Sluzala ZB, Hamati A, Fort PE. Key Role of Phosphorylation in Small Heat Shock Protein Regulation via Oligomeric Disaggregation and Functional Activation. *Cells.* 2025;14(2):127. <https://doi.org/10.3390/cells14020127> PMID: [39851555](#)

9. Baughman HER, Clouser AF, Klevit RE, Nath A. HspB1 and Hsc70 chaperones engage distinct tau species and have different inhibitory effects on amyloid formation. *J Biol Chem*. 2018;293(8):2687–700. <https://doi.org/10.1074/jbc.M117.803411> PMID: [29298892](https://pubmed.ncbi.nlm.nih.gov/29298892/)
10. Shatov VM, Weeks SD, Strelkov SV, Gusev NB. The Role of the Arginine in the Conserved N-Terminal Domain RLFDQxFG Motif of Human Small Heat Shock Proteins HspB1, HspB4, HspB5, HspB6, and HspB8. *Int J Mol Sci*. 2018;19(7):2112. <https://doi.org/10.3390/ijms19072112> PMID: [30036999](https://pubmed.ncbi.nlm.nih.gov/30036999/)
11. Peters C, Haslbeck M, Buchner J. Catchers of folding gone awry: a tale of small heat shock proteins. *Trends Biochem Sci*. 2024;49(12):1063–78. <https://doi.org/10.1016/j.tibs.2024.08.003> PMID: [39271417](https://pubmed.ncbi.nlm.nih.gov/39271417/)
12. Pasta SY, Raman B, Ramakrishna T, Rao CM. The IXI/V motif in the C-terminal extension of alpha-crystallins: alternative interactions and oligomeric assemblies. *Mol Vis*. 2004;10:655–62. PMID: [15448619](https://pubmed.ncbi.nlm.nih.gov/15448619/)
13. Delbecq SP, Jehle S, Klevit R. Binding determinants of the small heat shock protein,  $\alpha$ B-crystallin: recognition of the “IXI” motif. *EMBO J*. 2012;31(24):4587–94. <https://doi.org/10.1038/emboj.2012.318> PMID: [23188086](https://pubmed.ncbi.nlm.nih.gov/23188086/)
14. Clouser AF, Baughman HE, Basanta B, Guttman M, Nath A, Klevit RE. Interplay of disordered and ordered regions of a human small heat shock protein yields an ensemble of “quasi-ordered” states. *Elife*. 2019;8:e50259. <https://doi.org/10.7554/eLife.50259> PMID: [31573509](https://pubmed.ncbi.nlm.nih.gov/31573509/)
15. Baughman HER, Pham T-HT, Adams CS, Nath A, Klevit RE. Release of a disordered domain enhances HspB1 chaperone activity toward tau. *Proc Natl Acad Sci U S A*. 2020;117(6):2923–9. <https://doi.org/10.1073/pnas.1915099117> PMID: [31974309](https://pubmed.ncbi.nlm.nih.gov/31974309/)
16. Alderson TR, Adriaenssens E, Asselbergh B, Pritišanac I, Van Lent J, Gastall HY, et al. A weakened interface in the P182L variant of HSP27 associated with severe Charcot-Marie-Tooth neuropathy causes aberrant binding to interacting proteins. *EMBO J*. 2021;40(8):e103811. <https://doi.org/10.15252/emboj.2019103811> PMID: [33644875](https://pubmed.ncbi.nlm.nih.gov/33644875/)
17. Rossor AM, Morrow JM, Polke JM, Murphy SM, Houlden H, et al. Pilot phenotype and natural history study of hereditary neuropathies caused by mutations in the HSPB1 gene. *Neuromuscul Disord*. 2017;27(1):50–6. <https://doi.org/10.1016/j.nmd.2016.10.001> PMID: [27816334](https://pubmed.ncbi.nlm.nih.gov/27816334/)
18. Fortunato F, Neri M, Geroldi A, Bellone E, De Grandis D, Ferlini A, et al. A CMT2 family carrying the P7R mutation in the N-terminal region of the HSPB1 gene. *Clin Neurol Neurosurg*. 2017;163:15–7. <https://doi.org/10.1016/j.clineuro.2017.09.012> PMID: [29031079](https://pubmed.ncbi.nlm.nih.gov/29031079/)
19. Echaniz-Laguna A, Geuens T, Petiot P, Péréon Y, Adriaenssens E, Haidar M, et al. Axonal Neuropathies due to Mutations in Small Heat Shock Proteins: Clinical, Genetic, and Functional Insights into Novel Mutations. *Hum Mutat*. 2017;38(5):556–68. <https://doi.org/10.1002/humu.23189> PMID: [28144995](https://pubmed.ncbi.nlm.nih.gov/28144995/)
20. Holmgren A, Bouhy D, De Winter V, Asselbergh B, Timmermans J-P, Irobi J, et al. Charcot-Marie-Tooth causing HSPB1 mutations increase Cdk5-mediated phosphorylation of neurofilaments. *Acta Neuropathol*. 2013;126(1):93–108. <https://doi.org/10.1007/s00401-013-1133-6> PMID: [23728742](https://pubmed.ncbi.nlm.nih.gov/23728742/)
21. Geuens T, De Winter V, Rajan N, Achsel T, Mateiu L, Almeida-Souza L, et al. Mutant HSPB1 causes loss of translational repression by binding to PCBP1, an RNA binding protein with a possible role in neurodegenerative disease. *Acta Neuropathol Commun*. 2017;5(1):5. <https://doi.org/10.1186/s40478-016-0407-3> PMID: [28077174](https://pubmed.ncbi.nlm.nih.gov/28077174/)
22. Kriehuber T, Rattei T, Weinmaier T, Bepperling A, Haslbeck M, Buchner J. Independent evolution of the core domain and its flanking sequences in small heat shock proteins. *FASEB J*. 2010;24(10):3633–42. <https://doi.org/10.1096/fj.10-156992> PMID: [20501794](https://pubmed.ncbi.nlm.nih.gov/20501794/)
23. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583–9. <https://doi.org/10.1038/s41586-021-03819-2> PMID: [34265844](https://pubmed.ncbi.nlm.nih.gov/34265844/)
24. Ho W-L, Huang J-R. The return of the rings: Evolutionary convergence of aromatic residues in the intrinsically disordered regions of RNA-binding proteins for liquid-liquid phase separation. *Protein Sci*. 2022;31(5):e4317. <https://doi.org/10.1002/pro.4317> PMID: [35481633](https://pubmed.ncbi.nlm.nih.gov/35481633/)
25. Pathak RK, Singh DB, Singh R. Introduction to basics of bioinformatics. *Bioinformatics*. 2022:1–15. <https://doi.org/10.1016/b978-0-323-89775-4.00006-7>
26. Rajagopal P, Liu Y, Shi L, Clouser AF, Klevit RE. Structure of the  $\alpha$ -crystallin domain from the redox-sensitive chaperone, HSPB1. *J Biomol NMR*. 2015;63(2):223–8. <https://doi.org/10.1007/s10858-015-9973-0> PMID: [26243512](https://pubmed.ncbi.nlm.nih.gov/26243512/)
27. Freilich R, Betegon M, Tse E, Mok S-A, Julien O, Agard DA, et al. Competing protein-protein interactions regulate binding of Hsp27 to its client protein tau. *Nat Commun*. 2018;9(1):4563. <https://doi.org/10.1038/s41467-018-07012-4> PMID: [30385828](https://pubmed.ncbi.nlm.nih.gov/30385828/)

28. Lambert H, Charette SJ, Bernier AF, Guimond A, Landry J. HSP27 multimerization mediated by phosphorylation-sensitive intermolecular interactions at the amino terminus. *J Biol Chem*. 1999;274(14):9378–85. <https://doi.org/10.1074/jbc.274.14.9378> PMID: [10092617](#)
29. Lelj-Garolla B, Mauk AG. Roles of the N- and C-terminal sequences in Hsp27 self-association and chaperone activity. *Protein Sci*. 2012;21(1):122–33. <https://doi.org/10.1002/pro.761> PMID: [22057845](#)
30. Thériault JR, Lambert H, Chávez-Zobel AT, Charest G, Lavigne P, Landry J. Essential role of the NH2-terminal WD/EPF motif in the phosphorylation-activated protective function of mammalian Hsp27. *J Biol Chem*. 2004;279(22):23463–71. <https://doi.org/10.1074/jbc.M402325200> PMID: [15033973](#)
31. Tóth-Petróczy A, Tawfik DS. Slow protein evolutionary rates are dictated by surface-core association. *Proc Natl Acad Sci U S A*. 2011;108(27):11151–6. <https://doi.org/10.1073/pnas.1015994108> PMID: [21690394](#)
32. Weeks SD, Baranova EV, Heirbaut M, Beelen S, Shkumatov AV, Gusev NB, et al. Molecular structure and dynamics of the dimeric human small heat shock protein HSPB6. *J Struct Biol*. 2014;185(3):342–54. <https://doi.org/10.1016/j.jsb.2013.12.009> PMID: [24382496](#)
33. Basha E, O'Neill H, Vierling E. Small heat shock proteins and  $\alpha$ -crystallins: dynamic proteins with flexible functions. *Trends Biochem Sci*. 2012;37(3):106–17. <https://doi.org/10.1016/j.tibs.2011.11.005> PMID: [22177323](#)
34. Lelj-Garolla B, Mauk AG. Self-association of a small heat shock protein. *J Mol Biol*. 2005;345(3):631–42. <https://doi.org/10.1016/j.jmb.2004.10.056> PMID: [15581903](#)
35. Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics*. 2010;11:431. <https://doi.org/10.1186/1471-2105-11-431> PMID: [20718988](#)
36. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods*. 2011;9(2):173–5. <https://doi.org/10.1038/nmeth.1818> PMID: [22198341](#)
37. AlphaFold v2.3.0. [cited 31 Jan 2025]. Available: [https://github.com/google-deeppmind/alphafold/blob/main/docs/technical\\_note\\_v2.3.0.md](https://github.com/google-deeppmind/alphafold/blob/main/docs/technical_note_v2.3.0.md)
38. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Židek A, et al. Highly accurate protein structure prediction for the human proteome. *Nature*. 2021;596(7873):590–6. <https://doi.org/10.1038/s41586-021-03828-1> PMID: [34293799](#)
39. Alderson TR, Pritišanac I, Kolarić Đ, Moses AM, Forman-Kay JD. Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. *Proc Natl Acad Sci U S A*. 2023;120(44):e2304302120. <https://doi.org/10.1073/pnas.2304302120> PMID: [37878721](#)
40. Janowska MK, Baughman HER, Woods CN, Klevit RE. Mechanisms of Small Heat Shock Proteins. *Cold Spring Harb Perspect Biol*. 2019;11(10):a034025. <https://doi.org/10.1101/cshperspect.a034025> PMID: [30833458](#)
41. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, et al. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc Natl Acad Sci U S A*. 2011;108(49):E1293–301. <https://doi.org/10.1073/pnas.1111471108> PMID: [22106262](#)
42. Davey NE, Cyert MS, Moses AM. Short linear motifs - ex nihilo evolution of protein regulation. *Cell Commun Signal*. 2015;13:43. <https://doi.org/10.1186/s12964-015-0120-z> PMID: [26589632](#)
43. Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, et al. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev*. 2014;114(13):6733–78. <https://doi.org/10.1021/cr400585q> PMID: [24926813](#)
44. Palopoli N, Marchetti J, Monzon AM, Zea DJ, Tosatto SCE, Fornasari MS, et al. Intrinsically Disordered Protein Ensembles Shape Evolutionary Rates Revealing Conformational Patterns. *J Mol Biol*. 2021;433(3):166751. <https://doi.org/10.1016/j.jmb.2020.166751> PMID: [33310020](#)
45. Marchetti J, Monzon AM, Tosatto SCE, Parisi G, Fornasari MS. Ensembles from Ordered and Disordered Proteins Reveal Similar Structural Constraints during Evolution. *J Mol Biol*. 2019;431(6):1298–307. <https://doi.org/10.1016/j.jmb.2019.01.031> PMID: [30731089](#)
46. Hochberg GKA, Ecroyd H, Liu C, Cox D, Cascio D, Sawaya MR, et al. The structured core domain of  $\alpha$ B-crystallin can prevent amyloid fibrillation and associated toxicity. *Proc Natl Acad Sci U S A*. 2014;111(16):E1562–70. <https://doi.org/10.1073/pnas.1322673111> PMID: [24711386](#)
47. Sluchanko NN, Beelen S, Kulikova AA, Weeks SD, Antson AA, Gusev NB, et al. Structural Basis for the Interaction of a Human Small Heat Shock Protein with the 14-3-3 Universal Signaling Regulator. *Structure*. 2017;25(2):305–16. <https://doi.org/10.1016/j.str.2016.12.005> PMID: [28089448](#)

48. Woods CN, Janowska MK, Ulmer LD, Sidhu JK, Stone NL, James EI, et al. Activation mechanism of Small Heat Shock Protein HSPB5 revealed by disease-associated mutants. 2022. <https://doi.org/10.1101/2022.05.30.493970>
49. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* 2004;32(Database issue):D115–9. <https://doi.org/10.1093/nar/gkh131> PMID: 14681372
50. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10:421. <https://doi.org/10.1186/1471-2105-10-421> PMID: 20003500
51. Hedges SB, Kumar S. . The Timetree of Life. 2009. <https://doi.org/10.1093/oso/9780199535033.001.0001>
52. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics.* 2009;25(11):1422–3. <https://doi.org/10.1093/bioinformatics/btp163> PMID: 19304878
53. Okonechnikov K, Golosova O, Fursov M, UGENE team. Unipro UGENE: a unified bioinformatics tool-kit. *Bioinformatics.* 2012;28(8):1166–7. <https://doi.org/10.1093/bioinformatics/bts091> PMID: 22368248
54. Felsenstein J. PHYLIP (Phylogeny Inference Package). Department of Genome Sciences, University of Washington. Seattle: Distributed by Author; 2005.
55. Pupko T, Bell RE, Mayrose I, Glaser F, Ben-Tal N. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics.* 2002;18 Suppl 1:S71–7. [https://doi.org/10.1093/bioinformatics/18.suppl\\_1.s71](https://doi.org/10.1093/bioinformatics/18.suppl_1.s71) PMID: 12169533
56. Del Conte A, Camagni GF, Clementel D, Minervini G, Monzon AM, Ferrari C, et al. RING 4.0: faster residue interaction networks with novel interaction types across over 35,000 different chemical structures. *Nucleic Acids Res.* 2024;52(W1):W306–12. <https://doi.org/10.1093/nar/gkae337> PMID: 38686797
57. Bouysset C, Fiorucci S. ProLIF: a library to encode molecular interactions as fingerprints. *J Cheminform.* 2021;13(1):72. <https://doi.org/10.1186/s13321-021-00548-6> PMID: 34563256
58. Case DA, Aktulga HM, Belfon K, Cerutti DS, Cisneros GA, Cruzeiro VWD, et al. AmberTools. *J Chem Inf Model.* 2023;63(20):6183–91. <https://doi.org/10.1021/acs.jcim.3c01153> PMID: 37805934
59. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* 1996;14(1):33–8. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) PMID: 8744570
60. Wallner B, Amunts A, Naschberger A, Nystedt B, Mirabello C. dgram2dmap: Extraction, visualisation and formatting of distance constraints from AlphaFold distograms. 2022. <https://doi.org/10.1101/2022.12.08.519560>
61. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods.* 2022;19(6):679–82. <https://doi.org/10.1038/s41592-022-01488-1> PMID: 35637307