PLOS ONE

RESEARCH ARTICLE

The PCI domains are "winged" HEAT domains

Eleanor Elise Paul, Assen Marintchev*

Department of Physiology & Biophysics, Boston University School of Medicine, Boston, Massachusetts, United States of America

* amarint@bu.edu

Abstract

The HEAT domains are a family of helical hairpin repeat domains, composed of four or more hairpins. HEAT is derived from the names of four family members: huntingtin, eukaryotic translation elongation factor 3 (eEF3), protein phosphatase 2 regulatory A subunit (PP2A), and mechanistic target of rapamycin (mTOR). HEAT domain-containing proteins play roles in a wide range of cellular processes, such as protein synthesis, nuclear transport and metabolism, and cell signaling. The PCI domains are a related group of helical hairpin domains, with a "winged-helix" (WH) subdomain at their C-terminus, which is responsible for multi-subunit complex formation with other PCI domains. The name is derived from the complexes, where these domains are found: the 26S Proteasome "lid" regulatory subcomplex, the COP9 signalosome (CSN), and eukaryotic translation initiation factor 3 (eIF3). We noted that in structure similarity searches using HEAT domains, sometimes PCI domains appeared in the search results ahead of other HEAT domains, which indicated that the PCI domains could be members of the HEAT domain family, and not a related but separate group, as currently thought. Here, we report extensive structure similarity analysis of HEAT and PCI domains, both within and between the two groups of proteins. We present evidence that the PCI domains as a group have greater structural similarity with individual groups of HEAT domains than some of the HEAT domain groups have among each other. Therefore, our results indicate that the PCI domains have evolved from a HEAT domain that acquired a WH subdomain. The WH subdomain in turn mediated self-association into a multi-subunit complex, which eventually evolved into the common ancestor of the Proteasome lid/CSN/eIF3.

Check for updates

OPEN ACCESS

Citation: Paul EE, Marintchev A (2022) The PCI domains are "winged" HEAT domains. PLoS ONE 17(9): e0268664. https://doi.org/10.1371/journal.pone.0268664

Editor: Monika Oberer, Karl-Franzens-Universitat Graz, AUSTRIA

Received: May 2, 2022

Accepted: August 17, 2022

Published: September 12, 2022

Copyright: © 2022 Paul, Marintchev. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting information files.

Funding: This work was supported by the National Institute of Health [GM134113 to A.M.]. 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.

Introduction

Helical repeat domains are widespread in eukaryotic genomes, and are involved in virtually every major cellular process. These encompass several families, including tetratricopeptide repeat (TPR), ankyrin repeat (ANK), armadillo (ARM), as well as the HEAT repeat domain family (reviewed in [1–4]). The acronym HEAT is derived from the names of several founding members of the family: huntingtin, eukaryotic translation elongation factor 3 (eEF3), protein phosphatase 2 regulatory A subunit (PP2A), and mechanistic target of rapamycin (mTOR) [5]. Of these families, ARM and HEAT repeat domains are only found in eukaryotes, and likely

arose more recently during evolution [2]. Another, smaller group of eukaryote-specific helical repeat proteins are the PCI domains. Most PCI domain-containing proteins are subunits of the **P**roteasome "lid" regulatory subcomplex, the **C**OP9 signalosome (**C**SN), or eukaryotic translation **i**nitiation factor 3 (e**I**F3) [6–9]. The Evolutionary Classification of Protein Domains (ECOD) database (http://prodata.swmed.edu/ecod/) [10, 11] groups the HEAT and PCI domains in the ARM-repeat H-group, along with multiple other helical repeat domain proteins.

Unlike most globular domains, helical repeat domains do not have a fixed size, with a distinct beginning and end. Instead, they consist of varying numbers of helical repeats. The larger domains have varying curvatures and tend to form solenoids. This plasticity likely contributed to the fast divergence of helical repeat domains in size, structure, sequence, and function. Sequence analyses have revealed that it is often impossible to detect statistically significant sequence similarity, even among members of the same domain family. The variable length, irregular shapes, and low/undetectable sequence similarity have made the analysis of the evolution of these domain families challenging. Sequence conservation is limited to the interfaces between adjacent helices and tends to be lower than for other domain families (reviewed in [1, 3, 5]).

The HEAT repeat domain family was first reported in 1995 [5]; these domains consist of a series of helical hairpins, each formed by two antiparallel helices, packing against each other and the two surrounding helical hairpins (Fig 1). The number of hairpins varies from as few as four to over 50 [5, 12–14].

Along with another domain, MPN/Mov34, the PCI domain was discovered as a structural unit found in three multi-subunit complexes, 26S Proteosome lid, CSN, and eIF3 [15–18]. PCI domains have very similar organization to the HEAT domains and the other helical repeat domains, but also have a "winged-helix" (WH) subdomain at their C-terminus. The WH subdomains mediate complex assembly with the WH subdomains of other PCI domain containing subunits. The 26S Proteosome lid, CSN, and eIF3 are multiprotein complexes with a common ancestor and similar organization, including the presence in each of them of six PCI domain-containing subunits: RPN3, RPN5, RPN6, RPN7, RPN9, and RPN12 in the Proteasome lid; CSN3, CSN4, CSN2, CSN1, CSN7, and CSN8 in the CSN; and eIF3l, eIF3a, eIF3c, eIF3e, eIF3m, and eIF3k in eIF3, respectively, as well as two MPN/Mov34 domain-containing subunits. While the MPN domains show high degree of sequence conservation, this is not the case for the PCI domains. For PCI domain-containing subunits in the Proteasome lid and the CSN complexes, the pairwise correspondence was possible to establish based on sequence similarity. However, similarity was only detected to two of their eIF3 counterparts (among CSN3, eIF3l, and RPN3, and among CSN3, eIF3k, and RPN12). For the others, the pairwise correspondence only became obvious when the structures of the Proteasome, CSN, and eIF3 became available. In each of these complexes, the WH subdomains form a ring in the core of the complex. No similarity was detected between the six groups of PCI domain proteins [6–8, 16–18]. Exhaustive PSI-BLAST [19] searches yielded additional pairs of PCI domains with significant sequence similarity; while using HHblits [20], we were able to detect statistically significant sequence similarity among all PCI domains, but not between PCI and HEAT domains, or between different groups of HEAT domains (data not shown). eIF3 in some groups of organisms has lost part of its subunits and has fewer than six PCI domains. For example, *Saccharomyces cerevisiae (S. cerevisiae)* has two of the PCI subunits: eIF3a and eIF3c; while *Schizosaccharomyces pombe (S. pombe)* has four: eIF3a, eIF3c, eIF3e, and eIF3m subunits [15, 21–26].

The structure similarity between HEAT and PCI domains is illustrated in Fig 1 and has been noted previously [9], although no statistically significant sequence similarity could be

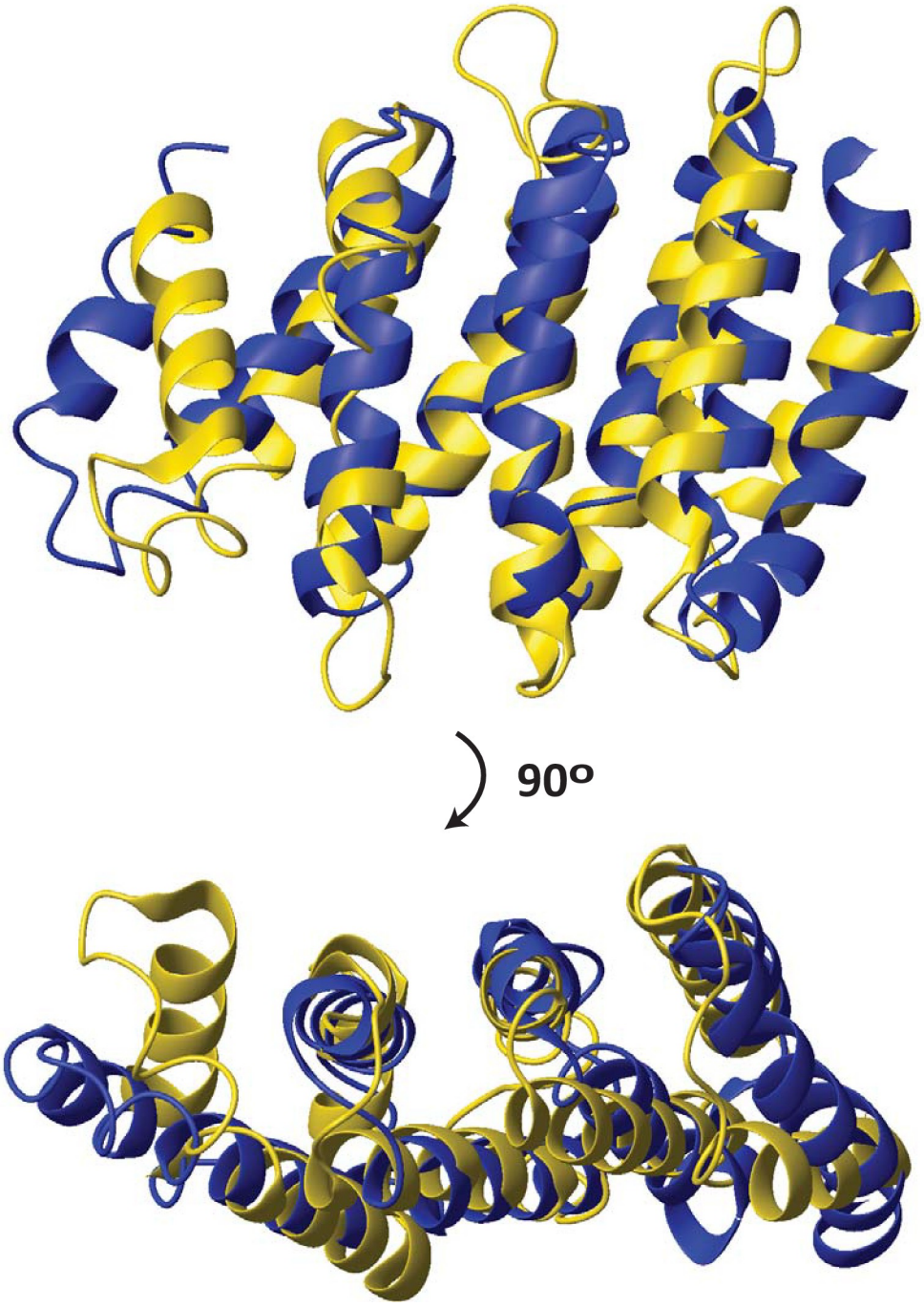

Fig 1. Structure alignment between the HEAT domain of CTF3 and the PCI domain of RPN12. DALI [27]-based structure alignment between the HEAT domain of *S. cerevisiae* CTF3 (6wuc.pdb, chain B) and the PCI domain of human RPN12 (5l4k.pdb, chain P). The structures are shown in ribbon. CTF3 is colored gold; RPN12 is colored blue. Only the aligned portions of the two structures are shown.

https://doi.org/10.1371/journal.pone.0268664.g001

detected. However, as pointed out above, there is no detectable sequence similarity among distant members of the HEAT family, either [8, 17]. It is of course, theoretically possible that the PCI domains and the different groups of HEAT domains with no detectable sequence similarity could be the product of convergent evolution because of the simple repetitive nature of the helical hairpins.

When performing DALI [27] searches with HEAT domains, we noticed that sometimes PCI domain structures appeared in the search results with higher scores that some of the known HEAT domains. This prompted us to investigate whether the PCI domains are a subset of the HEAT family, or the PCI and HEAT domains are two related but distinct families. Here we performed extensive comparative structure similarity analysis, which demonstrates that the PCI domains are a sub-family of the HEAT domains, which originated when a HEAT domain acquired a WH domain at its C-terminus.

Results

We reasoned that if structure similarity between the PCI domains and individual groups of known HEAT domains is comparable to the structure similarity among the groups of HEAT domains, then the PCI domains are a group of HEAT domains that acquired a WH subdomain during evolution, after the HEAT domains had diverged from other helical repeat domain families. Conversely, if all groups of HEAT domains have greater structure similarity to each other than to the PCI domains, then the PCI domains split earlier, before the common ancestor of all HEAT domains (Fig 2):

- If the PCI domains have lower structure similarity scores with all groups of HEAT domains than the scores among HEAT domain groups, then the PCI domains split off from their common ancestor with HEAT domains before the individual HEAT domain groups diverged from each other (Fig 2A).
- If the average similarity scores between PCI domains and at least one group of HEAT domains are comparable, or higher, than the lowest average similarity scores between at least two groups of HEAT domains, then the PCI domains are a group of HEAT domains that have acquired a winged helix (Fig 2B).

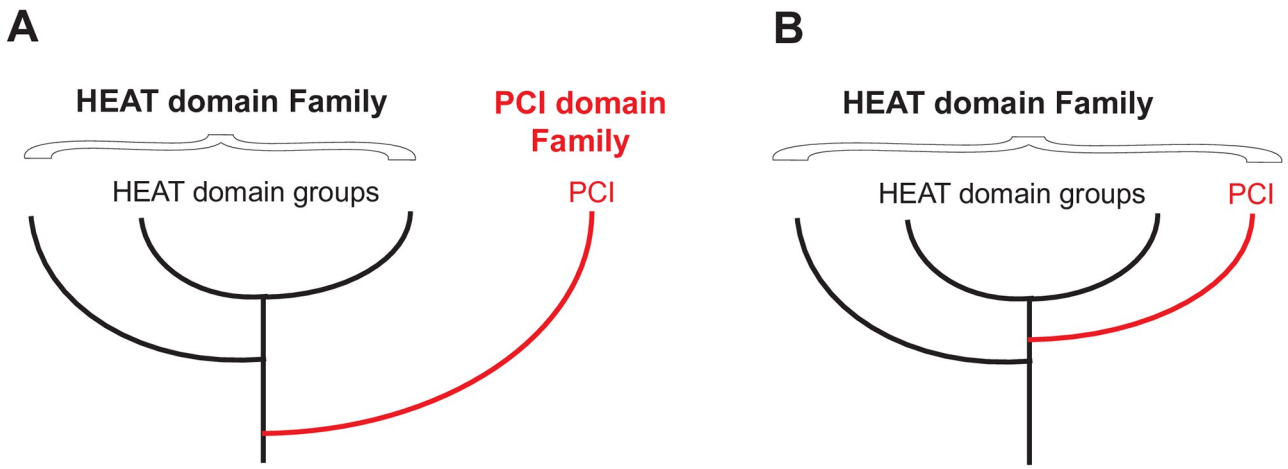

Fig 2. Possible alternatives for the evolution of PCI domains. A. PCI domains (red) diverged from HEAT domains before the last common ancestor of all HEAT domains and constitute a separate family of domains, as currently thought. **B**. The PCI domains are members of the HEAT domain family that acquired a WH subdomain after the last common HEAT domain ancestor.

https://doi.org/10.1371/journal.pone.0268664.g002

To account for the presence of a Winged Helix (WH) subdomain in PCI domains, but not in the HEAT domains, which could affect the structure similarity scores, we also used PCI domain structures with deleted WH subdomains. Finally, helical repeat domains pose unique challenges for structure similarity software, because they have a repetitive structure, a variable number of repeats, and varying curvature. To try to compensate partially for these factors, we also used four-hairpin fragments of HEAT and PCI domains.

We performed extensive structure similarity searches with HEAT and PCI domains with known structures, using the DALI server [27]. The searches with a HEAT domain structure consistently yielded PCI domains with structure similarity scores (Z-scores) comparable to those of other HEAT domains, and *vice versa*. The PCI domain structures were the only group that consistently appeared in the searches with high scores. All further analysis was focused on HEAT and PCI domain structures only; however, there could be other groups of domains with comparable structure similarity to HEAT domains. We selected a representative set of structures with high-resolution and avoiding close relatives, and then subdivided the representative structures into groups based on structure similarity scores. The HEAT domains fell into four groups, based on structure similarity, as well as a couple of outlier structures, which were considered as a fifth group. (**Table in** S1 Table) The "founding" members of the HEAT domain family, including Huntingtin, eEF3, and PP2A [5], were in Group 1. The MA3, MIF4G, and W2 domains [12–14, 28–31] formed individual groups, Groups 2, 3, and 4, respectively. Finally, two of the selected proteins, UTP10 and the FAT (**F**RAP, **A**TM, **T**RRAP) HEAT repeat domain of mTOR [32, 33], distinct from the "classical" HEAT repeat domain in the N-terminal portion of the protein [5], were outliers and placed in Group 5. The PCI domains formed Group 6.

We found that the average similarity score between Group 1 HEAT domains and PCI domains (Group 6), was 4.8 (where scores above 2 indicate similarity). This score was higher than those between Group 1 and Groups 2 (4.2) and 4 (3.6). Similar results were observed for the outlier HEAT domains (Group 5): average similarity score with the PCI domains (Group 6) was 4.6, whereas those with Groups 2 and 4 were both 3.9. The similarity score of Group 5 with Group 3 was 4.5, comparable to that with Group 6 (PCI) (Table 1, Fig 3A). This observation indicates that the PCI domains belong to the family of HEAT domains.

The same conclusion was reached when using a set of PCI domains with deleted WH subdomain, which tended to increase further the similarity scores between the PCI domains (Group 6) and the HEAT domain groups. The score between Group 1 and Group 6 was now 5.1, and that between Group 5 and Group 6 was now 4.8. Furthermore, the new score between Group 3 and Group 6 was now 4.6, comparable to that between Group 3 and Group 5 (4.5) (Table 1, Fig 3B).

Finally, we used a set of protein fragments with only four helical hairpins, so that all structures had the same size. Again, the average similarity score between Group 1 HEAT domains and PCI domains (Group 6), 4.6, was higher than those between Group 1 and Groups 2 (3.9) and 4 (3.7). Likewise, the average similarity score of Group 5 with the PCI domains (Group 6), 5.2, was higher than those with Group 2 (4.2), Group 3 (4.8), and Group 4 (3.5) (Table 1, Fig 3C). Thus, all three versions of the analysis yield the same conclusion: that the PCI domains are a group of HEAT domains that have acquired a WH subdomain (the scenario shown schematically in Fig 1B, above).

Structure-based sequence alignment of helical hairpins from PCI and HEAT domain structures showed a preference for hydrophobic side chains at positions buried at the interface between the two helices in the hairpin and, to a lesser extent, at positions facing adjacent hairpins, as expected (Fig 4). No obvious signature motifs could be observed at specific hairpin

Table 1. Structure similarity scores among groups of HEAT and PCI domains.

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
	Classical HEAT	MA3	MIF4G	W2	HEAT Outliers	PCI
A. Intact structures						
Vs. group 1	11.1 +/- 2.8	4.1 +/- 1.1	6.8 +/- 1.6	3.6 +/- 1.2	5.2 +/- 1.5	4.8 +/- 1.5
Vs. group 2	4.1 +/- 1.1	12.9 +/- 4.1	9.0 +/- 2.4	7.8 +/- 1.8	4.0 +/- 0.6	3.1 +/- 0.8
Vs. group 3	6.8 +/- 1.6	9.0 +/- 2.4	13.6 +/- 3.7	5.4 +/- 1.4	4.5 +/- 0.8	4.2 +/- 1.2
Vs. group 4	3.6 +/- 1.2	7.8 +/- 1.8	5.4 +/- 1.4	12.7 +/- 1.6	3.9 +/- 1.3	3.1 +/- 1.0
Vs. group 5	5.2 +/- 1.5	4.0 +/- 0.6	4.5 +/- 0.8	3.9 +/- 1.3	5.7 +/- 0.0	4.6 +/- 1.7
Vs. group 6 (PCI)	4.8 +/- 1.5	3.1 +/- 0.8	4.2 +/- 1.2	3.1 +/- 1.0	4.6 +/- 1.7	10.6 +/- 3.8
B. Wingless PCI						
Vs. group 1	11.1 +/- 2.8	4.1 +/- 1.1	6.8 +/- 1.6	3.6 +/- 1.2	5.2 +/- 1.5	5.1 +/- 1.5
Vs. group 2	4.1 +/- 1.1	12.9 +/- 4.1	9.0 +/- 2.4	7.8 +/- 1.8	4.0 +/- 0.6	3.1 +/- 0.6
Vs. group 3	6.8 +/- 1.6	9.0 +/- 2.4	13.6 +/- 3.7	5.4 +/- 1.4	4.5 +/- 0.8	4.6 +/- 1.3
Vs. group 4	3.6 +/- 1.2	7.8 +/- 1.8	5.4 +/- 1.4	12.7 +/- 1.6	3.9 +/- 1.3	3.3 +/- 1.1
Vs. group 5	5.2 +/- 1.5	4.0 +/- 0.6	4.5 +/- 0.8	3.9 +/- 1.3	5.7 +/- 0.0	4.8 +/- 1.8
Vs. group 6 (PCI)	5.1 +/- 1.5	3.1 +/- 0.6	4.6 +/- 1.3	3.3 +/- 1.1	4.8 +/- 1.8	7.9 +/- 2.8
C. Four-hairpin fragments						
Vs. group 1	9.1 +/- 2.4	3.9 +/- 1.0	6.4 +/- 1.6	3.7 +/- 1.0	6.3 +/- 1.7	4.6 +/- 1.2
Vs. group 2	3.9 +/- 1.0	10.8 +/- 3.2	7.6 +/- 3.0	7.4 +/- 1.8	4.2 +/- 0.5	3.1 +/- 0.6
Vs. group 3	6.4 +/- 1.6	7.6 +/- 3.0	10.3 +/- 4.5	4.8 +/- 1.3	4.8 +/- 1.5	4.2 +/- 1.1
Vs. group 4	3.7 +/- 1.0	7.4 +/- 1.8	4.8 +/- 1.3	15.5 +/- 3.6	3.5 +/- 0.8	3.0 +/- 0.7
Vs. group 5	6.3 +/- 1.7	4.2 +/- 0.5	4.8 +/- 1.5	3.5 +/- 0.8	5.6 +/- 0.0	5.2 +/- 2.5
Vs. group 6 (PCI)	4.6 +/- 1.2	3.1 +/- 0.6	4.2 +/- 1.1	3.0 +/- 0.7	5.2 +/- 2.5	7.6 +/- 2.3

Average pairwise DALI [27] Z-scores +/- standard deviation (SD) between members of individual HEAT and PCI domain groups. The average intragroup Z-scores (in the diagonal) are highlighted in grey.

https://doi.org/10.1371/journal.pone.0268664.t001

positions within groups or among the entire alignment, at least with the set of hairpins used here.

Discussion

The comparative structure similarity analysis of HEAT and PCI domain structures presented here shows that the PCI domains are members of the HEAT domain family. The canonical HEAT domains (Group 1) have similar structure similarity with the MIF4G (Group 3) and PCI (Group 6) domains. And in fact, the PCI domains appear more closely related to the canonical HEAT domains than do the MA3 (Group 2) and W2 (group 4) domains (Fig 3). Sometimes described as atypical HEAT domains [12] and found mostly in eukaryotic translation initiation factors and proteins involved in translation regulation and ribosome biogenesis, the MA3 and W2 HEAT domains tend to be shorter, some with as few as four helical hairpins, often arranged in tandem, and likely evolved from a common ancestor containing a MIF4G, an MA3, and a W2 domains in a row, as observed in eIF4G and CBP80 (reviewed in [34–37]).

To avoid possible score bias due to the length of individual domains, we repeated the structure similarity analysis using four-hairpin fragments of the proteins, which corresponds to the size of the smallest HEAT domains. The results with four-hairpin fragments confirm the results obtained with the intact domains. Using the four-hairpin fragments has its own caveats, because, while for the PCI domains, using the last four hairpins before the WH subdomain ensures that corresponding fragments are used, this is not guaranteed to be the case for the

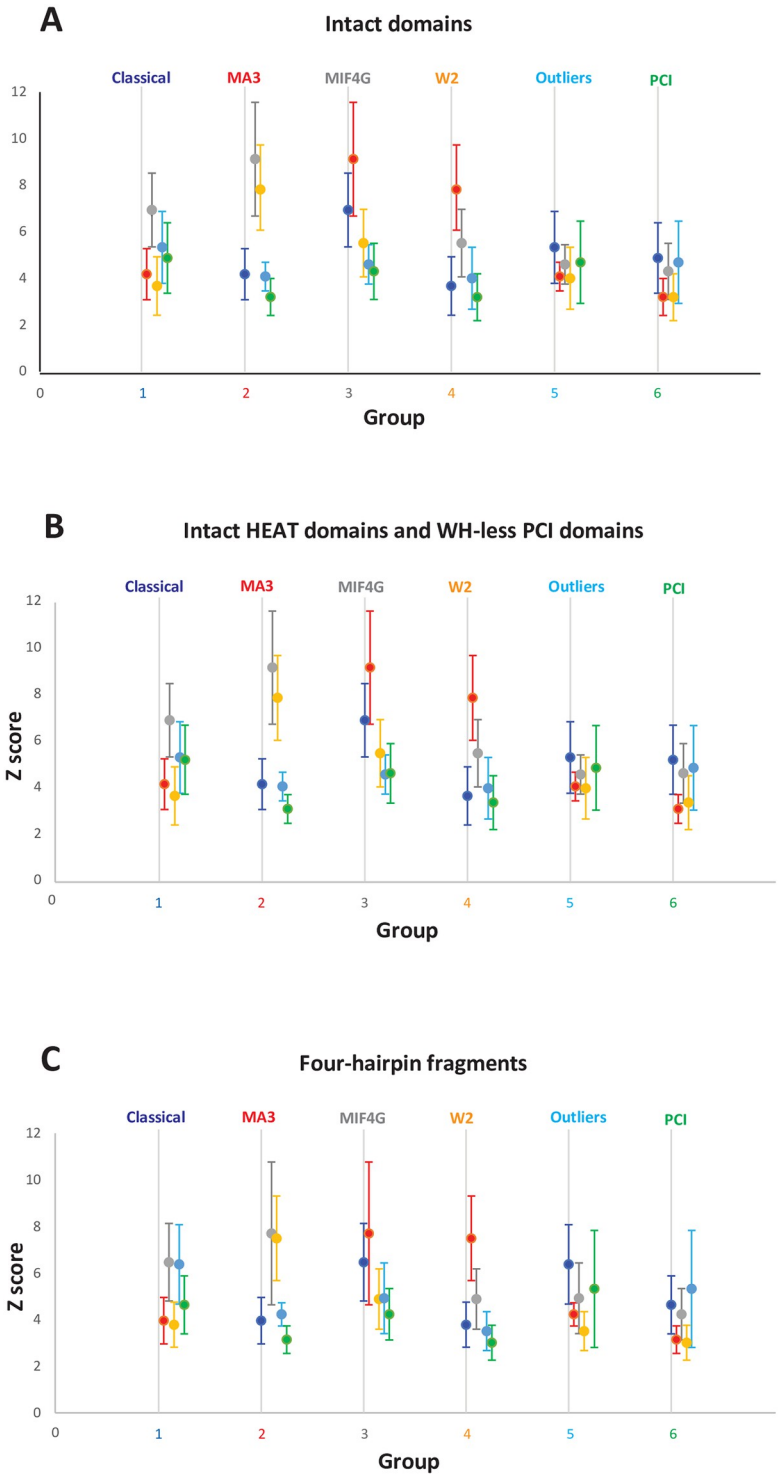

Fig 3. Structure similarity scores among groups of HEAT and PCI domains. Average pairwise DALI [27] Z-scores +/- standard deviation (SD) between members of individual HEAT and PCI domain groups. Groups are labeled and color-coded. The Z-scores of each group with the other five groups are shown on the Y-axis, with slight offset along the X-axis, and color-coded. Classic HEAT domains (Group 1) are navy. MA3 domains (Group 2) are red. MIF4G domains (Group 3) are grey. W2 domains (Group 4) are orange. Outlier HEAT domains (Group 5) are light blue. PCI domains are green. Intragroup Z-scores are not shown.

https://doi.org/10.1371/journal.pone.0268664.g003

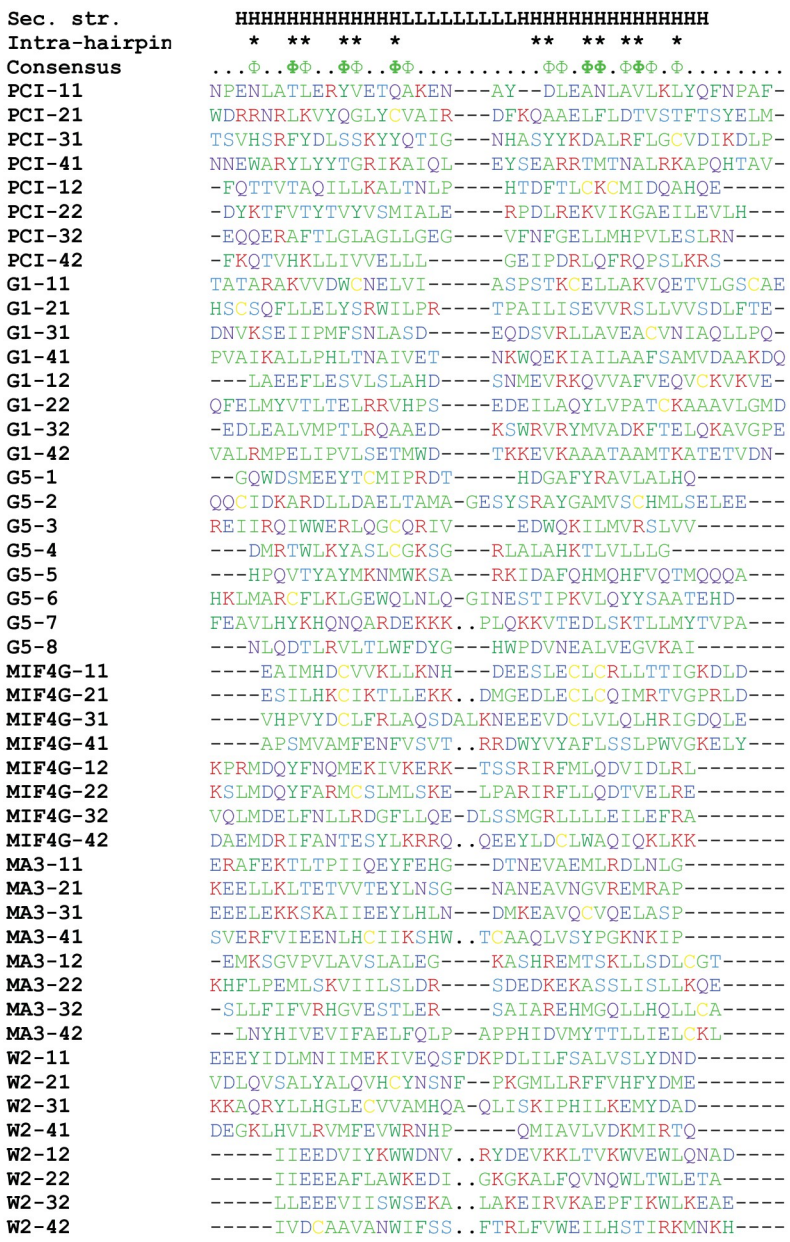

Fig 4. Structure-based sequence alignment of hairpins from HEAT and PCI domains. Eight hairpins per group were used for the alignment. Alignment viewer (https://alignmentviewer.org/) with Clustal residue coloring scheme was used, instead of conservation, because there were no detectable sequence conservation patterns, besides hydrophobic side chains at buried positions. Secondary structure is shown above the alignment: H, helix; L, loop. Positions of intra-hairpin contacts are marked with a "*" above the alignment. Note that side chains at the same position in the alignment may contact both the other helix in the hairpin and an adjacent hairpin, or contact one or the other in different hairpins, due to variations in inter-helix angles and orientations. Positions with hydrophobic side chains in >50% of sequences are labeled with "Φ"; positions with hydrophobic side chains in >75% of the sequences are labeled with "Φ" in bold. Gaps are marked with "-". Positions of deleted amino-acid sequences in the interhelix loop are marked with "..".

https://doi.org/10.1371/journal.pone.0268664.g004

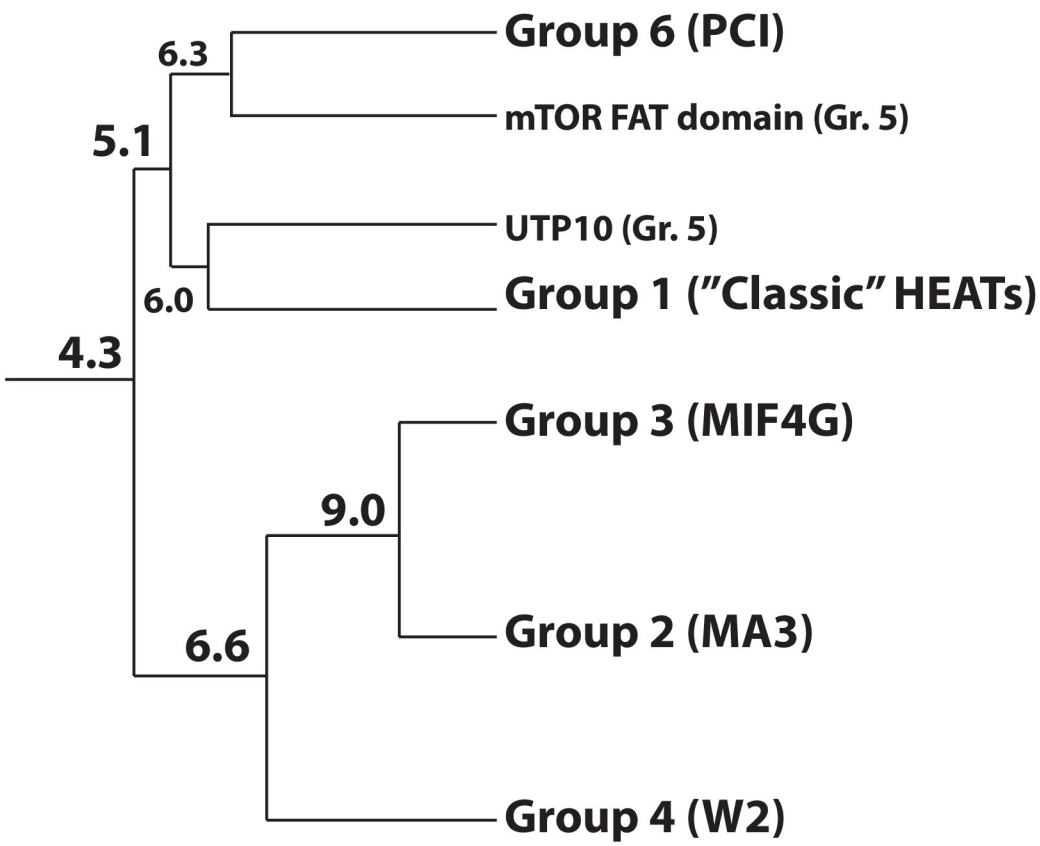

Fig 5. Possible evolutionary tree of HEAT and PCI domains. The dendrogram is based on average structure similarity Z scores from the DALI server [27] between individual groups of domains and branches, using intact HEAT domains and WH-less PCI domains. The outliers, UTP10 and the FAT domain of mTOR, which had been operationally grouped together as Group 5 in Fig 3, Table 1, and Table in S1 Table, had slightly lower structure similarity between each other (Z score 5.7) than their highest average Z scores with other groups (UTP10, Z score 6.0 with Group 1; and the mTOR FAT domain, Z score 5.9 with Group 6). Therefore, these two structures were not considered as a group in building the dendrogram and were added separately, instead, and are shown in smaller font.

https://doi.org/10.1371/journal.pone.0268664.g005

various HEAT domains, even if using the last four hairpins of each protein. However, the observation that the evolutionary trees and groups obtained using intact domains vs. four-hairpin fragments are very similar, supports the validity of this approach. Importantly, both the analysis using intact domains, and that using four-hairpin fragments yield the same results, which strongly supports the overall conclusions.

The MIF4G and MA3 domain groups are evolutionarily closer to each other than to the rest of the HEAT domains, while the W2 domains are closer to the MA3 domains that to the rest of HEAT domains, including the MIF4G group (Fig 3). These observations indicate that the MA3 and W2 domains could have resulted from the duplication of a MIF4G domain (Fig 5), as opposed to the breaking up of a long HEAT domain solenoid into three consecutive shorter HEAT domains. The propensity of these short HEAT domains to duplicate is illustrated by UPF2, which has three tandem MIF4G domains, and Pdcd4, which has two MA3 domains [38]. eIF5, eIF2Bε, and 5MP/BZW all have W2 domains closely homologous to the W2 domain of eIF4G [29, 37, 39, 40]. 5MP/BZW also has a predicted MA3 domain N-terminal to the W2 domain (refs. [37, 41] and AlphaFold [42, 43]). The much lower structure similarity scores of Group 1 HEAT domains with MA3 and especially W2 domains, compared to MIF4G

domains (Table 1, Fig 3), would then suggest different divergence rates of individual groups of domains. As stated in the Introduction, considering the simple, repetitive nature of the helical hairpins, convergent evolution cannot be excluded because there is no detectable sequence similarity between the different groups of domains. The structure of the mTOR FAT domain has average Z score 6.3 with WH-less PCI domains (Fig 5) and 5.9 with intact PCI domains (see **Table in** S1 Table), which is higher than with any of the groups of HEAT domains (the highest average Z score is 4.4 with Group 1, see **Table in** S1 Table). Therefore, it appears that the FAT domains could be more closely related to the PCI domains than to any of the current groups of HEAT domains. However, the absence of detectable sequence similarity makes it impossible to make any reliable conclusions.

In conclusion, this study offers insights into the evolution of the HEAT domain family in eukaryotes and shows that the PCI domains are in fact HEAT domains, which have acquired a WH subdomain at their C-terminus. This further expands the number of proteins containing a HEAT domain, and the already wide range of functions performed by these domains. It would be interesting to trace the origins and divergence of all helical repeat domain protein families as a whole, throughout the evolution of eukaryotes.

Methods

Structure similarity searches

The DALI Server [27] was used to search the RCSB PDB database [44] for similar structures and obtain structural similarity scores (Z-scores). The goal was to assemble a diverse set of HEAT and PCI domain structures, as well as identify potential new structures belonging to these families. Searches were initiated with several known HEAT domain and PCI domain structures. An initial diverse set of structures was assembled, eliminating structures with high similarity, as well as selecting high-resolution structures. Each structure and the corresponding structure-based sequence alignment were inspected manually, e.g., to confirm whether a putative PCI domain indeed contains the obligate WH subdomain, or that the DALI-generated automatic structure alignment does indeed define a contiguous helical hairpin structure segment. PyMol [45] was used for structure visualization.

Structure classification

Any non-HEAT or PCI domain portions of the structures were deleted at this stage using PyMol [45]. Structure similarity scores were calculated for all pairs of structures in the dataset, followed by a second round of removing highly similar structures, yielding the final dataset (**Table in** S1 Table). Where DALI failed to detect similarity, we did not use the pair in future analyses, because the structures are in fact similar. Using the pairs with a Z score 0, or with a Z score 2 (the minimum score considered statistically significant) did not affect the conclusions.

The evolutionary tree automatically generated by the DALI server served as a starting point for grouping the structures in the dataset. The HEAT domains fell into four groups, "classic" HEAT domains, MA3 domains, MIF4G domains, and W2 domains, all of which had previously been defined [5, 12–14, 28–30, 46]. Two HEAT domain structures were not part of any of the groups and were assigned a separate group (Group 5). The PCI domains formed one group (Group 6). The auto-generated dendrogram was correct in most cases, except grouping CTIF3 with Group 3 (MIF4G), instead of Group 1 ("classic" HEAT domains), and grouping the CBP80 MIF4G domain with Group 2 (MA3), instead of Group 3 (MIF4G). In both cases, the Z scores in **Table in** S1 Table unambiguously assign these two structures to the correct group.

Analysis of structure similarity and evolutionary relationships

We calculated average structure similarity scores (Z scores) and standard deviations for every pair of groups from the pairwise Z scores between all members of the two groups, obtained using the DALI server [27]. The average Z scores between groups of HEAT domains were compared to those between a HEAT domain group and the PCI domains, in order to determine whether or not at least one of the groups of HEAT domains shows greater, or at least comparable, similarity to the PCI domains than to at least one other group of HEAT domains.

To evaluate the contribution of the WH subdomain in the PCI domains, we repeated the analysis using "wingless" PCI domains. The WH subdomains were deleted from the corresponding structures in PyMol [45]. To account for the fact that different HEAT and PCI domains vary greatly in size, we generated four-hairpin fragments from each structure in the dataset and repeated the analysis with those. MOLMOL [47] was used to create figures of structure alignments between HEAT and PCI domains. The dendrogram in Fig 5 was generated using averaged DALI server [27] generated Z scores for intact HEAT domains and WH-less PCI domains. The outliers, UTP10 and the mTOR FAT domain, which had been operationally grouped together (Group 5) to simplify analysis, had relatively modest similarity between each other (Z score 5.7), which was comparable, and even slightly lower than their highest average Z scores: with Group 1 (UTP10, Z score 6.0) and with Group 6 (mTOR FAT, Z score 5.9), respectively. Therefore, these two structures were not considered as a group in building the dendrogram and were added separately, instead.

Structure-based sequence alignments

Exhaustive PSI-BLAST [19] and HHblits [20] searches yielded statistically significant sequence similarity between proteins belonging to the same group, but not between groups (data not shown). Therefore, we relied on structure alignments in aligning both consecutive hairpins from the same structure and from proteins belonging to different groups. Eight hairpins per group were used for the alignment. The structure alignments were done using the DALI server [27] (larger structures) and in PyMol [45] (hairpins) because the hairpins were too short for the DALI server to recognize statistically significant structure similarity. The structure-based alignment of corresponding hairpins from structures belonging to the same group (where sequence similarity could be observed), was consistent with the PSI-BLAST [19] and HHblits [19] based sequence alignments. We used the Clustal residue coloring scheme, instead of conservation, because there were no detectable sequence conservation patterns, besides hydrophobic side chains at buried positions. Fig 4 was generated with the help of Alignment viewer (https://alignmentviewer.org/).

Supporting information

S1 Table. Pairwise structure homology scores for all structures included in the analysis.
(XLSX)

Author Contributions

Conceptualization: Assen Marintchev.

Data curation: Assen Marintchev.

Formal analysis: Eleanor Elise Paul, Assen Marintchev.

Funding acquisition: Assen Marintchev.

Investigation: Eleanor Elise Paul, Assen Marintchev.

Methodology: Assen Marintchev.

Project administration: Assen Marintchev.

Resources: Assen Marintchev.

Supervision: Assen Marintchev.

Validation: Eleanor Elise Paul, Assen Marintchev.

Visualization: Eleanor Elise Paul, Assen Marintchev.

Writing – original draft: Eleanor Elise Paul, Assen Marintchev.

Writing – review & editing: Eleanor Elise Paul, Assen Marintchev.

References

1. Andrade MA, Perez-Iratxeta C, Ponting CP. Protein repeats: structures, functions, and evolution. J Struct Biol. 2001; 134(2–3):117–31. Epub 2001/09/12. https://doi.org/10.1006/jsbi.2001.4392 PMID: 11551174.
2. Andrade MA, Petosa C, O'Donoghue SI, Muller CW, Bork P. Comparison of ARM and HEAT protein repeats. J Mol Biol. 2001; 309(1):1–18. Epub 2001/08/09. https://doi.org/10.1006/jmbi.2001.4624 PMID: 11491282.
3. Groves MR, Barford D. Topological characteristics of helical repeat proteins. Curr Opin Struct Biol. 1999; 9(3):383–9. Epub 1999/06/11. https://doi.org/10.1016/s0959-440x(99)80052-9 PMID: 10361086.
4. Yoshimura SH, Hirano T. HEAT repeats—versatile arrays of amphiphilic helices working in crowded environments? J Cell Sci. 2016; 129(21):3963–70. Epub 2016/11/02. https://doi.org/10.1242/jcs.185710 PMID: 27802131.
5. Andrade MA, Bork P. HEAT repeats in the Huntington's disease protein. Nat Genet. 1995; 11(2):115–6. Epub 1995/10/01. https://doi.org/10.1038/ng1095-115 PMID: 7550332.
6. Aravind L, Ponting CP. Homologues of 26S proteasome subunits are regulators of transcription and translation. Protein Sci. 1998; 7(5):1250–4. Epub 1998/05/30. https://doi.org/10.1002/pro.5560070521 PMID: 9605331.
7. Hofmann K, Bucher P. The PCI domain: a common theme in three multiprotein complexes. Trends Biochem Sci. 1998; 23(6):204–5. Epub 1998/06/30. https://doi.org/10.1016/s0968-0004(98)01217-1 PMID: 9644972.
8. Scheel H, Hofmann K. Prediction of a common structural scaffold for proteasome lid, COP9-signalosome and eIF3 complexes. BMC Bioinformatics. 2005; 6:71. Epub 2005/03/26. https://doi.org/10.1186/1471-2105-6-71 PMID: 15790418.
9. Wei Z, Zhang P, Zhou Z, Cheng Z, Wan M, Gong W. Crystal structure of human eIF3k, the first structure of eIF3 subunits. The Journal of biological chemistry. 2004; 279(33):34983–90. Epub 2004/06/08. https://doi.org/10.1074/jbc.M405158200 PMID: 15180986.
10. Cheng H, Liao Y, Schaeffer RD, Grishin NV. Manual classification strategies in the ECOD database. Proteins. 2015; 83(7):1238–51. Epub 2015/04/29. https://doi.org/10.1002/prot.24818 PMID: 25917548.
11. Cheng H, Schaeffer RD, Liao Y, Kinch LN, Pei J, Shi S, et al. ECOD: an evolutionary classification of protein domains. PLoS Comput Biol. 2014; 10(12):e1003926. Epub 2014/12/05. https://doi.org/10.1371/journal.pcbi.1003926 PMID: 25474468.
12. Bellsolell L, Cho-Park PF, Poulin F, Sonenberg N, Burley SK. Two structurally atypical HEAT domains in the C-terminal portion of human eIF4G support binding to eIF4A and Mnk1. Structure. 2006; 14(5):913–23. Epub 2006/05/16. https://doi.org/10.1016/j.str.2006.03.012 PMID: 16698552.
13. Boesen T, Mohammad SS, Pavitt GD, Andersen GR. Structure of the catalytic fragment of translation initiation factor 2B and identification of a critically important catalytic residue. The Journal of biological chemistry. 2004; 279(11):10584–92. Epub 2003/12/19. https://doi.org/10.1074/jbc.M311055200 PMID: 14681227.
14. Marcotrigiano J, Lomakin IB, Sonenberg N, Pestova TV, Hellen CU, Burley SK. A conserved HEAT domain within eIF4G directs assembly of the translation initiation machinery. Molecular cell. 2001; 7(1):193–203. Epub 2001/02/15. https://doi.org/10.1016/s1097-2765(01)00167-8 PMID: 11172724.

15. Asano K, Vornlocher HP, Richter-Cook NJ, Merrick WC, Hinnebusch AG, Hershey JW. Structure of cDNAs encoding human eukaryotic initiation factor 3 subunits. Possible roles in RNA binding and macromolecular assembly. The Journal of biological chemistry. 1997; 272(43):27042–52. Epub 1997/10/27. https://doi.org/10.1074/jbc.272.43.27042 PMID: 9341143.

16. des Georges A, Dhote V, Kuhn L, Hellen CU, Pestova TV, Frank J, et al. Structure of mammalian eIF3 in the context of the 43S preinitiation complex. Nature. 2015; 525(7570):491–5. Epub 2015/09/08. https://doi.org/10.1038/nature14891 PMID: 26344199.

17. Enchev RI, Schreiber A, Beuron F, Morris EP. Structural insights into the COP9 signalosome and its common architecture with the 26S proteasome lid and eIF3. Structure. 2010; 18(4):518–27. Epub 2010/04/20. https://doi.org/10.1016/j.str.2010.02.008 PMID: 20399188.

18. Querol-Audi J, Sun C, Vogan JM, Smith MD, Gu Y, Cate JH, et al. Architecture of human translation initiation factor 3. Structure. 2013; 21(6):920–8. Epub 2013/04/30. https://doi.org/10.1016/j.str.2013.04.002 PMID: 23623729.

19. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research. 1997; 25 (17):3389–402. Epub 1997/09/01. https://doi.org/10.1093/nar/25.17.3389 PMID: 9254694.

20. Remmert M, Biegert A, Hauser A, Soding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods. 2011; 9(2):173–5. Epub 2011/12/27. https://doi.org/10.1038/nmeth.1818 PMID: 22198341.

21. Akiyoshi Y, Clayton J, Phan L, Yamamoto M, Hinnebusch AG, Watanabe Y, et al. Fission yeast homolog of murine Int-6 protein, encoded by mouse mammary tumor virus integration site, is associated with the conserved core subunits of eukaryotic translation initiation factor 3. The Journal of biological chemistry. 2001; 276(13):10056–62. Epub 2001/01/13. https://doi.org/10.1074/jbc.M010188200 PMID: 11134033.

22. Asano K, Phan L, Anderson J, Hinnebusch AG. Complex formation by all five homologues of mammalian translation initiation factor 3 subunits from yeast Saccharomyces cerevisiae. The Journal of biological chemistry. 1998; 273(29):18573–85. Epub 1998/07/11. https://doi.org/10.1074/jbc.273.29.18573 PMID: 9660829.

23. Bandyopadhyay A, Lakshmanan V, Matsumoto T, Chang EC, Maitra U. Moe1 and spInt6, the fission yeast homologues of mammalian translation initiation factor 3 subunits p66 (eIF3d) and p48 (eIF3e), respectively, are required for stable association of eIF3 subunits. The Journal of biological chemistry. 2002; 277(3):2360–7. Epub 2001/11/14. https://doi.org/10.1074/jbc.M107790200 PMID: 11705997.

24. Burks EA, Bezerra PP, Le H, Gallie DR, Browning KS. Plant initiation factor 3 subunit composition resembles mammalian initiation factor 3 and has a novel subunit. The Journal of biological chemistry. 2001; 276(3):2122–31. Epub 2000/10/24. https://doi.org/10.1074/jbc.M007236200 PMID: 11042177.

25. Crane R, Craig R, Murray R, Dunand-Sauthier I, Humphrey T, Norbury C. A fission yeast homolog of Int-6, the mammalian oncoprotein and eIF3 subunit, induces drug resistance when overexpressed. Mol Biol Cell. 2000; 11(11):3993–4003. Epub 2000/11/10. https://doi.org/10.1091/mbc.11.11.3993 PMID: 11071922.

26. Phan L, Zhang X, Asano K, Anderson J, Vornlocher HP, Greenberg JR, et al. Identification of a translation initiation factor 3 (eIF3) core complex, conserved in yeast and mammals, that interacts with eIF5. Molecular and cellular biology. 1998; 18(8):4935–46. Epub 1998/07/22. https://doi.org/10.1128/MCB.18.8.4935 PMID: 9671501.

27. Holm L. Using Dali for Protein Structure Comparison. Methods Mol Biol. 2020; 2112:29–42. Epub 2020/02/02. https://doi.org/10.1007/978-1-0716-0270-6_3 PMID: 32006276.

28. Bieniossek C, Schutz P, Bumann M, Limacher A, Uson I, Baumann U. The crystal structure of the carboxy-terminal domain of human translation initiation factor eIF5. J Mol Biol. 2006; 360(2):457–65. Epub 2006/06/20. https://doi.org/10.1016/j.jmb.2006.05.021 PMID: 16781736.

29. Koonin EV. Multidomain organization of eukaryotic guanine nucleotide exchange translation initiation factor eIF-2B subunits revealed by analysis of conserved sequence motifs. Protein Sci. 1995; 4 (8):1608–17. Epub 1995/08/01. https://doi.org/10.1002/pro.5560040819 PMID: 8520487.

30. Ponting CP, Schultz J, Copley RR, Andrade MA, Bork P. Evolution of domain families. Adv Protein Chem. 2000; 54:185–244. Epub 2000/06/01. https://doi.org/10.1016/s0065-3233(00)54007-8 PMID: 10829229.

31. Wei Z, Xue Y, Xu H, Gong W. Crystal structure of the C-terminal domain of S. cerevisiae eIF5. J Mol Biol. 2006; 359(1):1–9. Epub 2006/04/18. https://doi.org/10.1016/j.jmb.2006.03.037 PMID: 16616930.

32. Lovejoy CA, Cortez D. Common mechanisms of PIKK regulation. DNA Repair (Amst). 2009; 8(9):1004–8. Epub 2009/05/26. https://doi.org/10.1016/j.dnarep.2009.04.006 PMID: 19464237.

33. Yang H, Rudge DG, Koos JD, Vaidialingam B, Yang HJ, Pavletich NP. mTOR kinase structure, mechanism and regulation. Nature. 2013; 497(7448):217–23. Epub 2013/05/03. https://doi.org/10.1038/nature12122 PMID: 23636326.

34. Aravind L, Koonin EV. Eukaryote-specific domains in translation initiation factors: implications for translation regulation and evolution of the translation system. Genome Res. 2000; 10(8):1172–84. Epub 2000/08/25. https://doi.org/10.1101/gr.10.8.1172 PMID: 10958635.

35. Asano K. Evolution of eukaryotic translation initiation factors. In: Dubitzky W. W O, Cho K.-H., and Yokota H., editor. Encyclopedia of Systems Biology. New York, NY: Springer; 2013. p. 682–7.

36. Marintchev A, Wagner G. Translation initiation: structures, mechanisms and evolution. Q Rev Biophys. 2004; 37(3–4):197–284. https://doi.org/10.1017/S0033583505004026 PMID: 16194295.

37. Marintchev A, Wagner G. eIF4G and CBP80 share a common origin and similar domain organization: implications for the structure and function of eIF4G. Biochemistry. 2005; 44(37):12265–72. https://doi.org/10.1021/bi051271v PMID: 16156639.

38. Ponting CP. Novel eIF4G domain homologues linking mRNA translation with nonsense-mediated mRNA decay. Trends Biochem Sci. 2000; 25(9):423–6. Epub 2000/09/06. https://doi.org/10.1016/s0968-0004(00)01628-5 PMID: 10973054.

39. Asano K, Krishnamoorthy T, Phan L, Pavitt GD, Hinnebusch AG. Conserved bipartite motifs in yeast eIF5 and eIF2Bepsilon, GTPase-activating and GDP-GTP exchange factors in translation initiation, mediate binding to their common substrate eIF2. The EMBO journal. 1999; 18(6):1673–88. Epub 1999/03/17. https://doi.org/10.1093/emboj/18.6.1673 PMID: 10075937.

40. Singh CR, Watanabe R, Zhou D, Jennings MD, Fukao A, Lee B, et al. Mechanisms of translational regulation by a human eIF5-mimic protein. Nucleic acids research. 2011; 39(19):8314–28. Epub 2011/07/13. https://doi.org/10.1093/nar/gkr339 PMID: 21745818.

41. Singh CR, Glineburg MR, Moore C, Tani N, Jaiswal R, Zou Y, et al. Human oncoprotein 5MP suppresses general and repeat-associated non-AUG translation via eIF3 by a common mechanism. Cell Rep. 2021; 36(2):109376. Epub 2021/07/15. https://doi.org/10.1016/j.celrep.2021.109376 PMID: 34260931.

42. 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. Epub 2021/07/16. https://doi.org/10.1038/s41586-021-03819-2 PMID: 34265844.

43. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic acids research. 2022; 50(D1):D439–D44. Epub 2021/11/19. https://doi.org/10.1093/nar/gkab1061 PMID: 34791371.

44. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic acids research. 2000; 28(1):235–42. Epub 1999/12/11. https://doi.org/10.1093/nar/28.1.235 PMID: 10592235.

45. Schrödinger L, & DeLano, W. PyMOL. http://www.pymol.org/pymol. 2020.

46. Wei J, Jia M, Zhang C, Wang M, Gao F, Xu H, et al. Crystal structure of the C-terminal domain of the varepsilon subunit of human translation initiation factor eIF2B. Protein Cell. 2010; 1(6):595–603. Epub 2011/01/05. https://doi.org/10.1007/s13238-010-0070-6 PMID: 21204011.

47. Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph. 1996; 14(1):51–5, 29–32. Epub 1996/02/01. https://doi.org/10.1016/0263-7855(96)00009-4 PMID: 8744573.