bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The PCI domains are "winged" HEAT domains

Eleanor Elise Paul[1] & Assen Marintchev[1*]

[1] Department of Physiology & Biophysics, Boston University School of Medicine, 700 Albany St. W336, Boston, MA 02118, USA

* Corresponding Author

E-mail: amarint@bu.edu (AM)

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract (300 words)

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: **h**untingtin, eukaryotic translation **e**longation factor 3 (eEF3), protein phosphatase 2 regulatory **A** subunit (PP2A), and mechanistic **t**arget 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 **P**roteasome "lid" regulatory subcomplex, the **C**OP9 signalosome (CSN), and eukaryotic translation **i**nitiation factor 3 (eIF3). We noted that in structure homology 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. Here, we report extensive structure homology 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 homology 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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Helical repeat domains are widespread in eukaryotic genomes, and 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: **h**untingtin, eukaryotic translation **e**longation factor 3 (e**E**F3), protein phosphatase 2 regulatory **A** subunit (PP2**A**), and mechanistic **t**arget of rapamycin (m**T**OR).[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].

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 sequence homology, even among members of the same domain family. Sequence conservation is limited to the interfaces between adjacent helices and tends to be lower than for other domain families. The variable length, irregular shapes, and low/undetectable sequence homology have made the analysis of the evolution of these domain families challenging (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, 10-12].

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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. The pairwise correspondence of some subunits in each of these complexes was possible to establish based on sequence homology, while for others, it 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 [6-8, 13-15]. eIF3 in some groups of organisms has lost part of its subunits and has fewer than six PCI domains. For example, *Saccharomyces cerevisiae* has two of the PCI subunits: eIF3a and eIF3c; while *Schizosaccharomyces pombe* has four: eIF3a, eIF3c, eIF3e, and eIF3m subunits [16-22].

The structure similarity between HEAT and PCI domains is illustrated in **Fig. 1** and has been noted previously [9], although no sequence homology can be detected. However, as pointed out above, there is no detectable sequence homology even between relatively close homologs within the HEAT and PCI families [8, 14].

When performing DALI [23] searches for structural homologs of HEAT proteins, 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 homology 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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Results

We reasoned that if structure homology between the PCI domains and individual groups of known HEAT domains is comparable to the structure homology 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 homology 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 homology 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 homology scores between PCI domains and at least one group of HEAT domains are comparable, or higher, than the lowest average homology 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**).

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 homology scores, we also used PCI domain structures with deleted WH subdomains. Finally, helical repeat domains pose unique challenges for structure homology 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 searches for HEAT and PCI domains with known structures and selected a representative set of structures with high-resolution and avoiding close relatives. We then subdivided the representative structures into groups based on structure homology scores. The HEAT domains fell into four groups, based on structure homology, as well as a couple of outlier

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

structures, which were considered as a fifth group. (**Supplementary Table S1**) The majority of the "founding" members of the HEAT domain family, including Huntingtin, eEF3, and PP2A [5], were in Group 1. The MA3, MIF4G, and W2 domains [10-12, 24-27] formed individual groups, Groups 2, 3, and 4, respectively. Finally, two of the selected proteins, including mTOR, one of the founding HEAT domain family members [5], were outliers and placed in Group 5. The PCI domains formed Group 6.

formed Group 6.

We found that the average homology score between Group 1 HEAT domains and PCI domains (Group 6), was 4.8 (where scores above 2 indicate homology). 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 homology score with the PCI domains (Group 6) was 4.6, whereas those with Groups 2 and 4 were both 3.9. The homology score of Group 5 with Group 3 was 4.5, comparable to that with Group 6 (PCI) (**Table 1,** The numbers in Fig. 3 plus intragroup Z-scores **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 homology 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 homology 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 homology 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).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discussion

The comparative structure homology analysis of HEAT and PCI domain structures presented here shows that the PCI domains are members of the HEAT domain family. In fact, they appear more closely related to the canonical HEAT domains than do the MA3, and W2 domains, while the canonical HEAT domains have similar structure homology with the MIF4G and PCI domains (**Fig. 3**). Sometimes described as atypical HEAT domains [10] 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 [28, 29]).

To avoid possible score bias due to the length of individual domains, we repeated the structure homology 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 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 and 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 domains probably resulted from the duplication of a MIF4G domain, and that the W2 domains, in turn,

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

resulted from the duplication of an MA3 domain, as opposed to 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 [30]. eIF5, eIF2Bε, and 5MP/BZW all have W2 domains closely homologous to the W2 domain of eIF4G [25, 29]. 5MP/BZW also has a predicted MA3 domain N-terminal to the W2 domain (ref. [29] and AlphaFold [31, 32]).

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 homology searches

The DALI Server [23] was used to search the RCSB PDB database [33] for homologous structures and obtain structural homology 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 [34] was used for structure visualization.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Structure classification

Any non-HEAT or PCI domain portions of the structures were deleted at this stage using PyMol [34]. Structure homology 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 S1**). Where DALI failed to detect homology, we did not use the pair in future analyses, because the structures are in fact homologous. 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, 10-12, 24-26, 35]. 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 S1** unambiguously assign these two structures to the correct group.

Analysis of structure homology and evolutionary relationships

We calculated average structure homology 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 [23]. 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 similar, homology 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 [34]. To account for the fact that different HEAT and PCI domains vary greatly in

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

size, we generated four-hairpin fragments from each structure in the dataset and repeated the analysis with those. MOLMOL [36] was used to create figures of structure alignments between HEAT and PCI domains.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

A. Intact structures

	Group 1 Classical HEAT	Group 2 MA3	Group 3 MIF4G	Group 4 W2	Group 5 HEAT Outliers	Group 6 PCI
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

	Group 1 Classical HEAT	Group 2 MA3	Group 3 MIF4G	Group 4 W2	Group 5 HEAT Outliers	Group 6 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

	Group 1 Classical HEAT	Group 2 MA3	Group 3 MIF4G	Group 4 W2	Group 5 HEAT Outliers	Group 6 PCI
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

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 [23] 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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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. doi: 10.1006/jsbi.2001.4392. PubMed 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. doi: 10.1006/jmbi.2001.4624. PubMed 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. doi: 10.1016/s0959-440x(99)80052-9. PubMed 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. doi: 10.1242/jcs.185710. PubMed 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. doi: 10.1038/ng1095-115. PubMed 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. doi: 10.1002/pro.5560070521. PubMed PMID: 9605331; PubMed Central PMCID: PMCPMC2144014.
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. doi: 10.1016/s0968-0004(98)01217-1. PubMed 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. doi: 10.1186/1471-2105-6-71. PubMed PMID: 15790418; PubMed Central PMCID: PMCPMC1274264.
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. doi: 10.1074/jbc.M405158200. PubMed PMID: 15180986.
10. 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. doi: 10.1016/j.str.2006.03.012. PubMed PMID: 16698552.
11. 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. doi: 10.1074/jbc.M311055200. PubMed PMID: 14681227.
12. 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. doi: 10.1016/s1097-2765(01)00167-8. PubMed PMID: 11172724.
13. 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. doi: 10.1038/nature14891. PubMed PMID: 26344199; PubMed Central PMCID: PMCPMC4719162.
14. 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. doi: 10.1016/j.str.2010.02.008. PubMed PMID: 20399188.
15. 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. doi: 10.1016/j.str.2013.04.002. PubMed PMID: 23623729; PubMed Central PMCID: PMCPMC3739965.
16. 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. doi: 10.1074/jbc.M010188200. PubMed PMID: 11134033.
17. 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

biological chemistry. 1998;273(29):18573-85. Epub 1998/07/11. doi: 10.1074/jbc.273.29.18573. PubMed PMID: 9660829.
18. 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. doi: 10.1074/jbc.272.43.27042. PubMed PMID: 9341143.
19. 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. doi: 10.1074/jbc.M107790200. PubMed PMID: 11705997.
20. 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. doi: 10.1074/jbc.M007236200. PubMed PMID: 11042177.
21. 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. doi: 10.1091/mbc.11.11.3993. PubMed PMID: 11071922; PubMed Central PMCID: PMCPMC15052.
22. 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. doi: 10.1128/MCB.18.8.4935. PubMed PMID: 9671501; PubMed Central PMCID: PMCPMC109077.
23. Holm L. Using Dali for Protein Structure Comparison. Methods Mol Biol. 2020;2112:29-42. Epub 2020/02/02. doi: 10.1007/978-1-0716-0270-6_3. PubMed PMID: 32006276.
24. 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. doi: S0022-2836(06)00603-6 [pii]

10.1016/j.jmb.2006.05.021. PubMed PMID: 16781736.
25. 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. doi: 10.1002/pro.5560040819. PubMed PMID: 8520487; PubMed Central PMCID: PMCPMC2143190.
26. 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. doi: 10.1016/s0065-3233(00)54007-8. PubMed PMID: 10829229.
27. 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. doi: 10.1016/j.jmb.2006.03.037. PubMed PMID: 16616930.
28. Marintchev A, Wagner G. Translation initiation: structures, mechanisms and evolution. Q Rev Biophys. 2004;37(3-4):197-284. PubMed PMID: 16194295.
29. 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. PubMed PMID: 16156639.
30. 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. doi: 10.1016/s0968-0004(00)01628-5. PubMed PMID: 10973054.
31. 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. doi: 10.1038/s41586-021-03819-2. PubMed PMID: 34265844; PubMed Central PMCID: PMCPMC8371605.
32. 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. doi: 10.1093/nar/gkab1061. PubMed PMID: 34791371; PubMed Central PMCID: PMCPMC8728224.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

33. 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. doi: 10.1093/nar/28.1.235. PubMed PMID: 10592235; PubMed Central PMCID: PMCPMC102472.

34. Schrödinger L, & DeLano, W. PyMOL. Retrieved from http://wwwpymolorg/pymol. 2020.

35. 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. doi: 10.1007/s13238-010-0070-6. PubMed PMID: 21204011; PubMed Central PMCID: PMCPMC4875318.

36. 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. doi: 10.1016/0263-7855(96)00009-4. PubMed PMID: 8744573.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 1. Structure alignment between the HEAT domain of CTF3 and the PCI domain of RPN12.

DALI [23]-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.

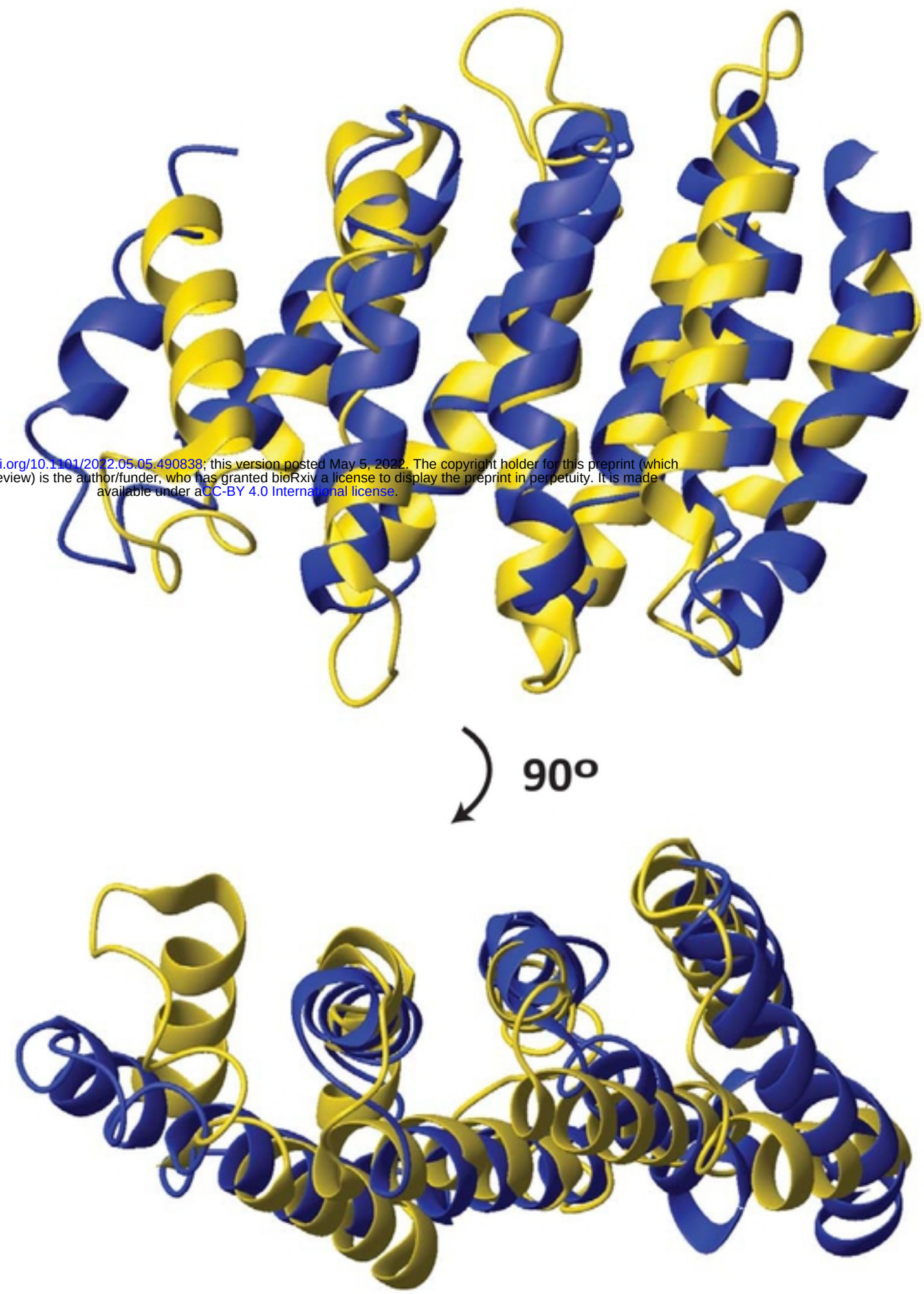
Figure 2. Possible alternatives for the evolution of PCI domains.

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 HEAD domain ancestor.

Figure 3. Structure homology scores among groups of HEAT and PCI domains.

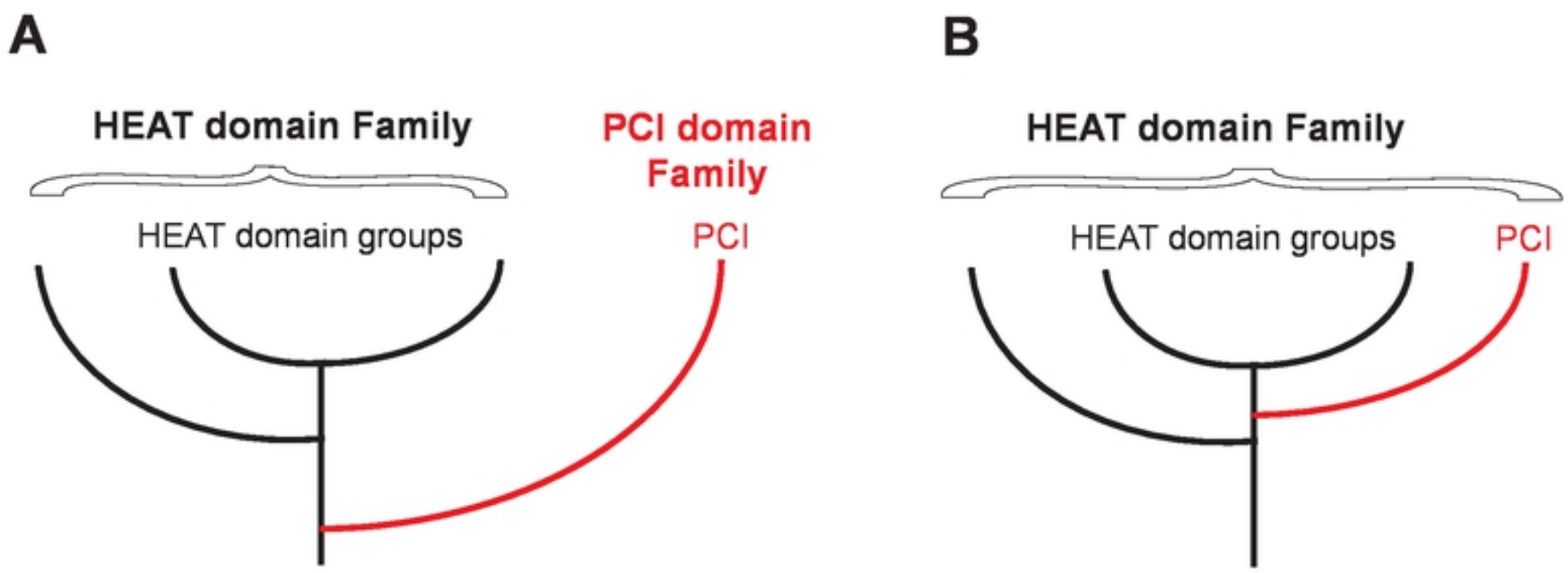
Average pairwise DALI [23] Z-scores +/- standard deviation (SD) between members of individual HEAT and PCI domain groups. Groups are labeled and color-coded. The Z-scores of each group with the other five groups are shown on the Y-axis, with slight offset along the X-axis, and color-coded. Classic HEAT 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.

Figure 1



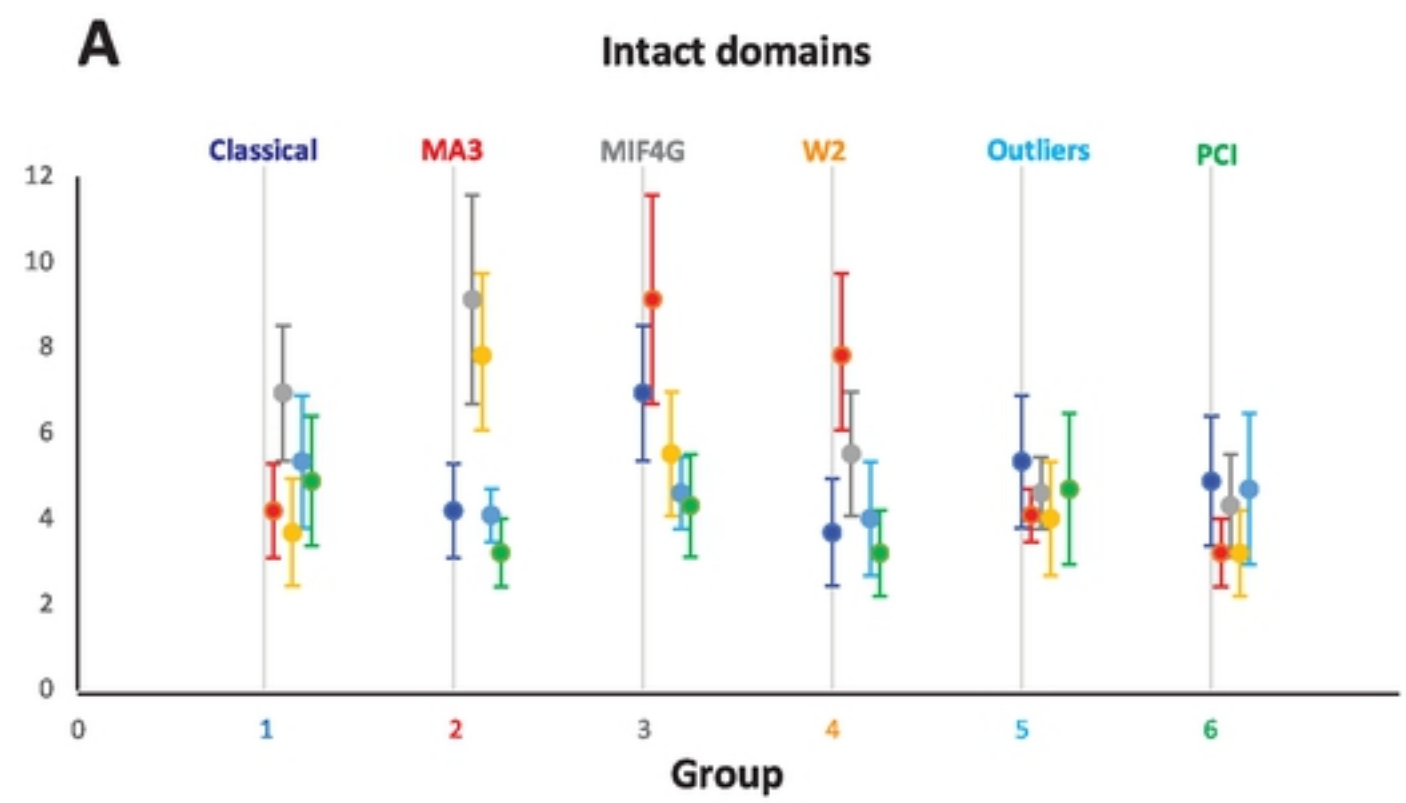
bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure

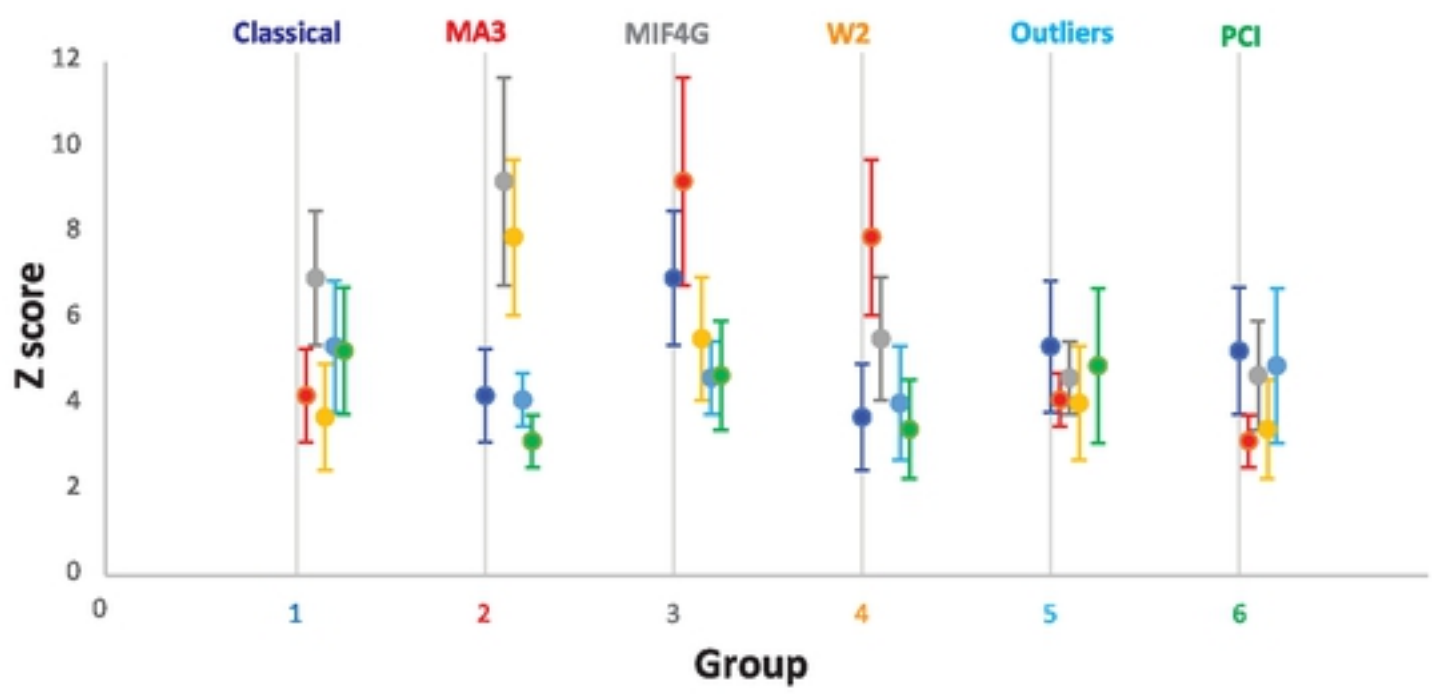


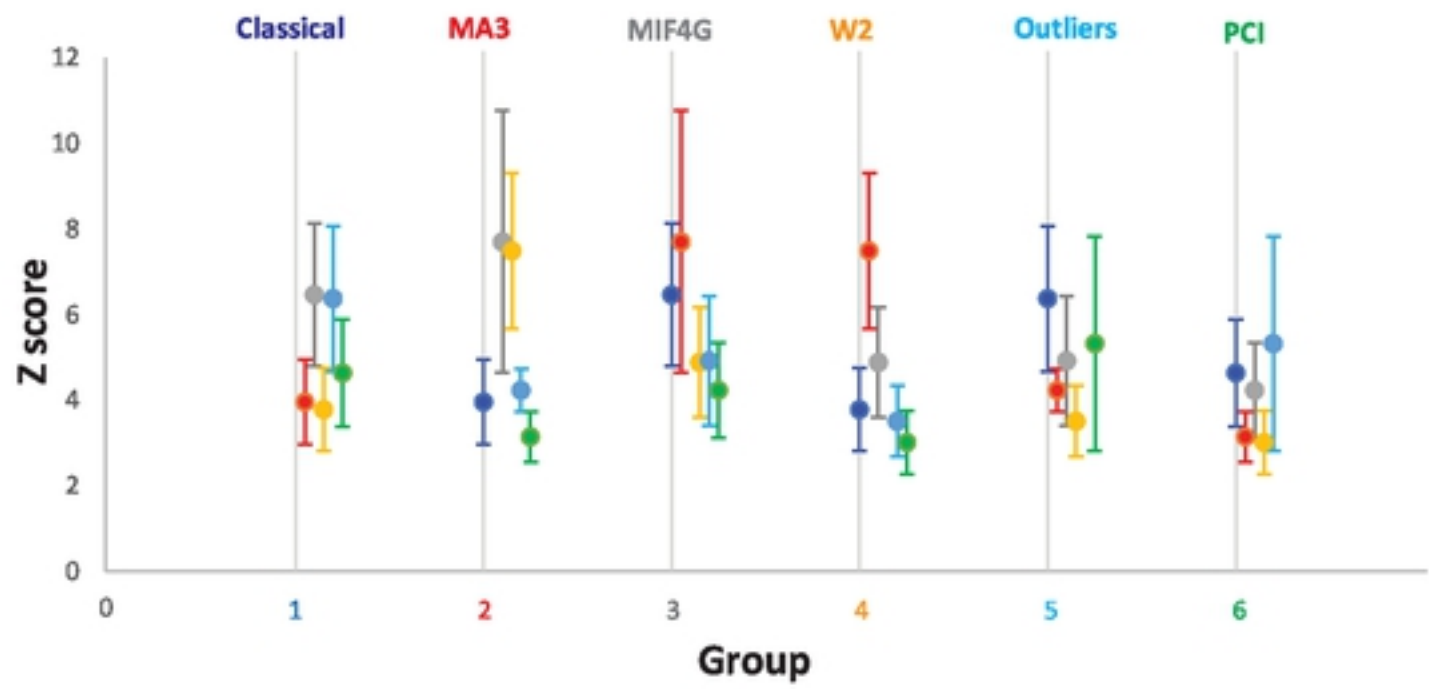
bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure



bioRxiv preprint doi: https://doi.org/10.1101/2022.05.05.490838; this version posted May 5, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.





Figure