Mechanistic Insight of Tardigrade Anhydrobiosis via Protein Structural Analysis

Claire K. Zhang

Brentwood High School

5304 Murray Ln, Brentwood, TN 37027

I would like to thank Dr. Qi Zhang from Florida Atlantic University for his guidance on method selection and paper preparation.

This study does not involve non-human vertebrates or human subjects.

Table of Contents

Abstract ……3

Introduction ……4

Methods ……...6

Results ……7

Discussion ……11

Tables ……13

Figures ……..14

Bibliography ……..20

Appendices ……...22

Abstract

Tardigrades are microscopic animals exhibiting exceptional tolerance to extreme environmental stresses such as desiccation. Hence, they provide unique solutions to biological preservation, anti-aging, and interstellar travel. Recent studies show that various cytoplasmic and secretory abundant heat soluble proteins (CAHSs and SAHSs) enable tardigrades to survive complete dehydration and revive upon rehydration. Using bioinformatic tools, possible molecular mechanisms underlying the protective effect of CAHSs and SAHSs were investigated. The gene and protein sequences of 18 CAHSs and 23 SAHSs from three representative tardigrade species (i.e., *Ramazzottius varieornatus*, *Hypsibius exemplaries*, and *Paramacrobiotus metropolitanus*) were collected from public-accessible databases. Sequence analyses revealed that both CAHSs and SAHSs possess a highly conserved center region bounded by two terminal regions variable in lengths and amino acid compositions. Analysis using AI-based AlphaFold 3.0 showed that the aforementioned terminal regions were disordered in normal physiological conditions. As for the center regions, CAHSs' formed a long alpha-helix domain while SAHSs' folded into a mixture of alpha-helix and beta-sheets. Furthermore, the trimers and decamers of CAHSs and SAHSs were generated to simulate their structural changes upon dehydration. Intriguingly, CAHS trimers generally formed linear bundles and their decamers were mostly circular, likely providing mechanical support for intracellular structures. In contrast, SAHS trimers or decamers were largely spheric with differently exposed tails, likely filling extracellular spaces or behaving as chaperones. These results suggest different strategies taken by tardigrades to counteract intracellular hyperosmosis and extracellular exsiccation respectively, inspiring new strategies to preserve biomolecules, cells, and tissues in extreme environments.

Introduction

Tardigrades, commonly known as water bears or moss piglets, are microscopic invertebrates comprising the distinct phylum Tardigrada (McInnes et al., 2021). They have a segmented body with four pairs of legs and live in marine or freshwater environments as well as semi-terrestrial habitats [1]. Tardigrades are renowned for withstanding extreme environments including desiccation, extreme temperatures, and cosmic radiation (Goldstein, 2022). Moreover, they can remain lifeless in a vacuum for decades until rehydration (Arakawa, 2022) (**Figure 1**). Thus, tardigrades are great models for studying anhydrobiosis (i.e., an organism loses almost all its water and enters a state of reversible ametabolism) (KEILIN, 1959). Desiccation causes hyperosmosis and damages biomolecules including proteins, nucleic acids, and lipids, leading to cell deformation and eventually death. Notably, tardigrades have many protective measures to prevent such destruction.

Recent studies have found that two protein families, cytoplasmic and secretory abundant heat soluble proteins (CAHSs and SAHSs), serve as intracellular and extracellular protectants, respectively, during desiccation [5]. Initial research deemed that both were intrinsically disordered and vitrified upon dehydration to preserve biomolecules, organelles, and cellular apparatuses (Boothby et al., 2017). Structural studies of exemplary SAHSs and CAHSs suggest that those proteins possess defined tertiary structure at least in their center regions under hydrated conditions (Bino et al., 2024; Fukuda et al., 2017; Fukuda and Inoue, 2018). Moreover, *in vitro* and *in silico* tests (i.e., cell-based assays and molecular dynamic simulation, respectively) have indicated that CAHSs enhance cells' resilience to hyperosmotic challenge from within (Bino et al., 2024) whereas SAHSs stabilize cell surface biomolecules like membrane lipids(Lim et al., 2024). How can CAHSs and SAHSs with distinct sequences and different localizations (i.e., inside and outside of cells, respectively) collectively enable anhydrobiosis for tardigrades? The answer to this crucial question most likely lies in the structures of CAHSs and SASHs, especially the dynamic changes upon forced oligomerization by dehydration.

While empirical tests are indispensable, computational analyses can provide useful information like conserved sequences, structural estimation like distinct motifs, and mechanistic predictions like molecular interactions. Notably, more and more tardigrades have undergone whole-genome sequencing and transcriptomic profiling. Subsequently, their gene and protein sequences have been made available to the general public (Erözden et al., 2024; Li et al., n.d.). Furthermore, sequence-based structure modeling fueled by the latest artificial intelligence have become more reliable, exemplified by Google's latest AlphaFold 3.0 achieving over 97% accuracy (Abramson et al., 2024).

Here, gene and protein sequences of 18 CAHSs and 23 SAHSs from three representative tardigrade species (i.e., *Ramazzottius varieornatus*, *Hypsibius exemplaries*, and *Paramacrobiotus metropolitanus*) were retrieved from GenBank and/or UniProt. Protein sequence alignment has revealed that the center regions of both CAHSs and SAHSs are highly conserved whereas the nearby terminal regions are quite variable in lengths and amino acid compositions. AI-based AlphaFold 3.0 has shown that CAHSs' center regions tend to form one long alpha-helix and SAHSs' folds into a mixture of alpha-helix and beta-sheets. Interestingly, AlphaFold 3.0 has predicted very different trimer and decamer conformations between CAHSs and SAHSs, suggesting two very different mechanisms to counteract intracellular hyperosmosis and extracellular exsiccation. Particularly, CAHSs assemble into higher-order helix bundles to support internal structures while SAHSs congregate into larger spherical complexes to protect external biomolecules. More than a showcase about the power of advanced bioinformatic tools, this computational study provides mechanistic insight about anhydrobiosis and will likely inspire unorthodox solutions to complex biological questions.

Methods

Gene and protein sequence retrieval

All nucleotide and protein sequences used for this study were obtained from GenBank (https://www.ncbi.nlm.nih.gov/genbank/about/) and UniProt (https://www.uniprot.org/) using keyword searching (i.e., abundant heat soluble protein, CAHS, or SAHS) and filtered by selected tardigrade species (i.e., *Ramazzottius varieornatus*, *Hypsibius exemplaries*, and *Paramacrobiotus metropolitanus*). The GenBank access numbers of all protein sequences are listed in **Table 1**.

Protein sequence alignment and phylogenetic tree generation

All protein sequences alignments were performed using constraint-based multiple alignment tools (i.e., COBALT) (Papadopoulos and Agarwala, 2007)available from the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi). The default alignment parameters were used. To generate phylogenetic trees based on the sequence alignments, we used the ETE3 toolkit with default settings available from GenomeNet (https://www.genome.jp/tools-bin/ete). All annotation of the resulting phylogenetic trees was done manually using Google Slides.

Protein structure prediction and analysis

All monomers, trimers, and decamers of CAHSs and SAHSs were generated using AlphaFold 3.0 (Abramson et al., 2024) provided by Google (https://github.com/google-deepmind/alphafold3). The default settings were used to ensure a fair treatment to all proteins. Resulting structures were downloaded and visualized using UCSF Chimera program (Pettersen et al., 2004). All qualitative comparisons were carried out in Google sheets and Prism. The surface areas and volumes of all CAHS and SAHS monomers, trimers, and decamers were calculated by Chimera. Statistical analyses were done using Student *t*-test. Linear regression fittings were done using maximum likelihood estimation (MLE)

Results

Consensus regions in CAHSs and SAHSs are identified by protein sequence analysis

Well over a thousand species of tardigrades have been described to date (McInnes et al., 2021) (For the most recent checklist of Tardigrada species, visit https://iris.unimore.it/handle/11380/1178608#), all of which exhibit a wide range of dehydration-resistance. Three species (i.e., *Ramazzottius varieornatus*, *Hypsibius exemplaries*, and *Paramacrobiotus metropolitanus*) were selected for this study because of their exceptional tolerance, most representative, and near complete genome sequences (Goldstein, 2022; McInnes et al., 2021). Using keyword searching, 18 CAHSs (3 from *R. varieornatus,* 7 from *H. exemplaries*, and 8 from *P. metropolitanus*) and 23 SAHSs (11 from *R. varieornatus,* 4 from *H. exemplaries*, and 8 from *P. metropolitanus*) were retrieved from GenBank with their encoding gene sequences and corresponding protein sequences further validated by UniProt. For easy access, **Table 1** listed all 41 proteins with their GenBank IDs and annotated names.

Using COBALT with default settings, both CAHS and SAHS sequences were aligned. In CAHSs, a highly conserved center region of approximately 130 amino acid residues was discovered, surrounded by N- and C-terminal regions with varying lengths and amino acid compositions (**Figure 2a**). Notably, the most common residues (shown in red) are generally those with charged or polar side chains (e.g., Lysine or Glutamine, respectively), and they are distributed nearly identically among all CAHSs (**Figure S1** in **Appendix 1**).

In SAHSs, a conserved center region was also revealed by the alignment, measuring a length of approximately 100 amino acid residues (**Figure 2b**). However, this region is less consistent than that of CAHSs and is segmented into three subregions by very short and more variable sequences. Those subregions are about 20 ~ 40 amino acid residues long, composed of amino acids with charged, polar, and hydrophobic side chains (e.g., Glutamate, Threonine, and Isoleucine, respectively). The distribution of those consensus residues is less consistent in SAHSs than in CAHSs (**Figure S2** in **Appendix 1**).

By examining the ETE3-generated phylogenetic trees, it became clear that CAHSs are more conserved across different species, shown by the most closely related CAHSs being from different species (**Figure 3a**). Despite its conserved center regions, Pm CAHS 89226-like (*P. metropolitanus*) is phylogenetically distant from all other 17 CAHS. In case of SAHSs, the closest genes are always from the same species (**Figure 3b**). When plotting CAHSs and SAHSs together, it became clear that CAHSs and SAHSs are fundamentally different in primary protein sequences as they form two separate clusters (**Figure 3c**). Altogether, the primary sequence analysis suggests that CAHSs and SAHSs are two very different groups of proteins despite their shared names. Furthermore, their conserved center regions display very different compositions, which implicates their structural difference.

Distinct domains in CAHSs and SAHSs are revealed by AlphaFold 3.0

With the rapid advance of artificial intelligence, the accuracy of protein structure prediction has improved dramatically. The latest AI-based prediction program, AlphaFold 3.0 is exceptionally reliable in sequence-based modeling of protein structures, achieving 90 ~ 97% accuracy (Abramson et al., 2024). Using its default setting, the monomer structures of 18 CAHSs and 23 SAHSs were obtained (Please see **Appendix 2** for the collection of all CAHS and SAHS structures). Consistent with the sequence alignments, 17 CAHSs exhibit a long and highly common alpha-helix domain, which corresponds to their highly conserved center regions (He CAHS 89226, Rv CAHS3, He CAHS 94205, and Pm CAHS 2-like in **Figure 4a first row** and **Appendix 2**). The only exception is Pm CAHS 89226-like (*P. metropolitanus*), which was singled out in phylogenetic tree (**Figure 3a**). In fact, it is the only CAHS whose center region does not form the long alpha helix (**Figure 4a** and **Appendix 2**), indicating that Pm CAHS 89226-like should not be categorized to CAHS family.

In case of SAHS, their conserved center regions form several consecutive beta-sheets often decorated with an alpha-helix or coil (**Figure 4b first row** and **Appendix 2**). Unlike CAHSs, there is no structure outlier in SAHSs, which is consistent with the phylogenetic tree

(**Figure 3b**). Notably, the more variable N- and C-terminal regions of both CAHSs and SAHSs mostly end up as disordered tails with low predicted local distance difference test scores (plDDT) (**Figure 4** and **Appendix 2**), indicating a protective function in the event of extreme environmental stress like desiccation (KC et al., 2024).

Distinct conformation changes between CAHSs and SAHS during oligomerization

The fast and drastic loss of water during desiccation concentrates biomolecules and thus forces the formation of oligomers both inside and outside of cells. To simulate the process of oligomerization, the structures of CAHS and SAHS trimers and decamers were generated (**Second and third rows** in **Figure 4a** & **b**, and **Appendix 2**). CAHS trimers usually form linear bundles with each monomer offset from the other in a step-like fashion (**Figure 4a second row** and **Appendix 2**). CAHS decamers generally form rings with each monomer also slightly offset from the others (**Figure 4a third row** and **Appendix 2**). Meanwhile, the confidence of the disordered tails increases, and every tail wraps around the monomers more when they turn to trimers and decamers, which suggests that the variable terminal regions aid in polymerization. Again, the trimer and decamer structures of Pm CAHS 89226-like differed very much from other CAHSs (**Figure 4a**), agreeing with the notion that it was mistakenly annotated as a CAHS. Consistently, all other CAHSs such as He CAHS 89226 and Rv CAHS3 form stable monomers and trimers but become less stable as decamers, suggesting the structure or length of the tails impacted progressive oligomerization (**Figure 4a**). Another CAHS, Pm CAHS 2-like, has no tails and thus forms a less stable trimer and decamer (**Figure 4a**), consistent with the notion that the disordered tails may play a role in oligomerization.

SAHS trimers and decamers generally aggregated into clumps with differently exposed tails (**Figure 4b** and **Appendix 2**). The confidence of structure prediction was generally high, indicating that SAHSs oligomerization increased their structural rigidity. In the case of SASHs with little or no N-terminal regions, like SAHS9, SAHS10, and SAHS12, there was lower confidence in structural prediction and thus less stable trimers or decamers (**Figure 4b** and

Appendix 2), further implying that N-terminal tails may facilitate SAHS polymerization during desiccation.

Quantitative analyses suggest distinct protective mechanisms by CAHSs and SAHSs

Recent studies suggested that CAHSs or SAHSs collaborate with other protectants like sucrose and trehalose to preserve biomolecules and cellular structures (Boothby et al., 2017; KC et al., 2024; Sanchez-Martinez et al., 2023). If so, their surface area to volume ratios would remain high or even increase to absorb water, protectants, as well as other biomolecules during dehydration. In contrast, such ratios in CAHS and SAHS monomers, trimers and decamers unexpectedly decreased significantly except between SAHS trimers and decamers (**Figure 5**), which argues against the idea that CAHSs and SAHSs behave as absorbents.

As previously mentioned, longer C-terminal tails correlate to a lower prediction confidence in CAHSs whereas longer N-terminal tails correlate to a higher decamer confidence in SAHSs. For further investigation, linear regression fitting was used to analyze such correlations. After excluding Pm CAHS2-like (0 length C-terminal region) and Pm CAHS 89226-like (likely not a CAHS), there is no significant correlation between C-terminal tail length and prediction confidence (**Figure 6a**). Even including Pm CAHS2-like and/or Pm CAHS 89226-like did not increase R^2. Similarly, there is no significant correlation between N-terminal tail length and prediction confidence for SAHSs (**Figure 6b**). Thus, the disordered N- or C-terminal tails of SAHSs or CAHSs are unlikely to drive the oligomerization or stabilize them. Therefore, the distinct structures and oligomerization between SAHSs and CAHSs are likely determined by their distinct center regions.

Discussion

The fascinating ability of tardigrades to sustain and survive extreme environments like outer space ignites great interest in using them as model animals to study physiological mechanisms for biological preservation, anti-aging, and interstellar travel (McInnes et al., 2021). Importantly, their tolerance to anhydrobiosis might have predisposed the animals to survive other extreme conditions. As more and more CAHSs and SAHSs have been discovered in different species of tardigrades, it becomes clear that both are highly expressed proteins playing a key role in preserving intracellular and extracellular structures and functions during desiccation(Boothby et al., 2017; Erözden et al., 2024; Lim et al., 2024). Several studies have suggested that the disordered regions found in both CAHSs and SAHSs are functionally important against desiccation. However, it is puzzling how similar molecular mechanisms deal with very different needs for intracellular and extracellular protection against dehydration. Furthermore, significant differences in protein sequences and subcellular localizations between CAHSs and SAHSs also speak against the notion of a shared mechanism.

To tackle this challenge, we turned to public-accessible databases of tardigrade gene sequences and rapidly advancing bioinformatic tools like AI-based AlphaFold. As proteins' functions are largely determined by their amino acid sequences, we focused on structural analyses and *in silico* tests related to desiccation (e.g., oligomerization). First, our sequence alignments unveiled conserved center regions in both CAHSs and SAHSs, which are significantly different from each other (leading to separate clusters in the phylogenetic trees in **Figure 3c**. Second, AlphaFold 3.0 confidently predicted a long alpha-helix and a combination of beta-sheet with short helix for the center regions of CAHSs and SAHSs respectively, implying mechanistic and functional differences. Notably, both center regions form ordered structures and sit in the core of each whole protein structure. Thirdly, structural difference between CAHS and SAHS trimers or decamers provide further argument that they may utilize different strategies of protections during desiccation.

The oligomerization of CAHSs into circular rings and role in anhydrobiosis suggests that CAHS oligomers function as mechanical supports and scaffolding for keeping the shape of the cell during desiccation, preventing crystallization of solutes while keeping cellular structures in place. The oligomerization of SAHSs into globules suggests SAHSs oligomers function as chaperones and buffers to fill in extracellular space in the event of extracellular exsiccation. The significant decrease in the surface area to volume ratio of the CAHSs solidifies that the purpose of CAHSs are not to absorb or retain molecules and more likely to keep intracellular structural integrity. While the comparison between SAHS monomers and decamers show a significant difference and decrease in the surface area to volume ratio, the comparison between their trimers and decamers did not, indicating the possibility of SAHSs functioning as extracellular chaperones. Also, the lack of correlation between tail length and the prediction confidence for oligomerization is revealing, speaking against the idea that the disordered regions of CAHSs and SAHSs facilitate oligomerization or stabilize the oligomers. Therefore, the functional weight is more on the ordered center regions of CAHSs and SAHSs.

Due to the constraints of AlphaFold 3.0 software, a simplified model of desiccation that did not account for molecule density, ion concentration, or other extracellular and intracellular factors was used. There is also the possibility of other unaccounted errors due to using AI-based programs and computational software. Nevertheless, the fact that AlphaFold 3.0 consistently predicts differently structured center domains between CAHSs and SAHSs support the notion that its structure predictions are highly reliable.

As all the research discussed thus far was conducted using computational tools and data analysis, implementing these findings in the laboratory would be the next step. Also, other CAHSs and SAHSs that were not used in this study will be added to expand the findings and improve the accuracy of the analysis.

Tables

Table 1. List of CAHSs and SAHSs used in this study (annotated names with GenBank IDs)

	Ramazzottius varieornatus	***Hypsibius exemplaris***	**Paramacrobiotus metropolitanus**
CAHS	Rv CAHS1 (402783758)	He CAHS 94063 (1209747827)	Pm CAHS 94205-like (1) (2493494936)
	Rv CAHS2 (402783760)	He CAHS 86272 (1209747819)	Pm CAHS 94205-like (2) (2493480804)
	Rv CAHS3 (402783762)	He CAHS 94205 (1) (1209747816)	Pm CAHS 2-like (2493480067)
		He CAHS 94205 (2) (1209747825)	Pm CAHS 94205-like (3) (2493479718)
		He CAHS 77611 (1209747824)	Pm CAHS 107838-like (1) (2493469471)
		He CAHS 77580 (1209747822)	Pm CAHS 107838-like (2) (2493468852)
		He CAHS 89226 (1209747829)	Pm CAHS 86272-like (2493445383)
			Pm CAHS 89226-like (2493438951)
SAHS	Rv SAHS1 (402783754)	He SAHS 63681 (1210274786)	Pm SAHS 1-like (1) (2493482411)
	Rv SAHS2 (402783756)	He SAHS 53582 (1210274784)	Pm SAHS 1-like (2) (2493456970)
	Rv SAHS3 (1101394461)	He SAHS 33020 (1210274782)	Pm SAHS 1-like (3) (2493420974)
	Rv SAHS4 (1101395371)	He SAHS 68234 (1210274787)	Pm SAHS 1-like X2 (2493398470)
	Rv SAHS6 (1101395435)		Pm SAHS 1-like X1 (2493398467)
	Rv SAHS7 (1101395438)		Pm SAHS 1-like (4) (2493395038)
	Rv SAHS8 (1101395445)		Pm SAHS 1-like (5) (2493392926)
	Rv SAHS9 (1101395447)		Pm SAHS 1-like (6) (2493392853)
	Rv SAHS10 (1101395452)		
	Rv SAHS11 (1101395453)		
	Rv SAHS12 (1101395666)		

Figures

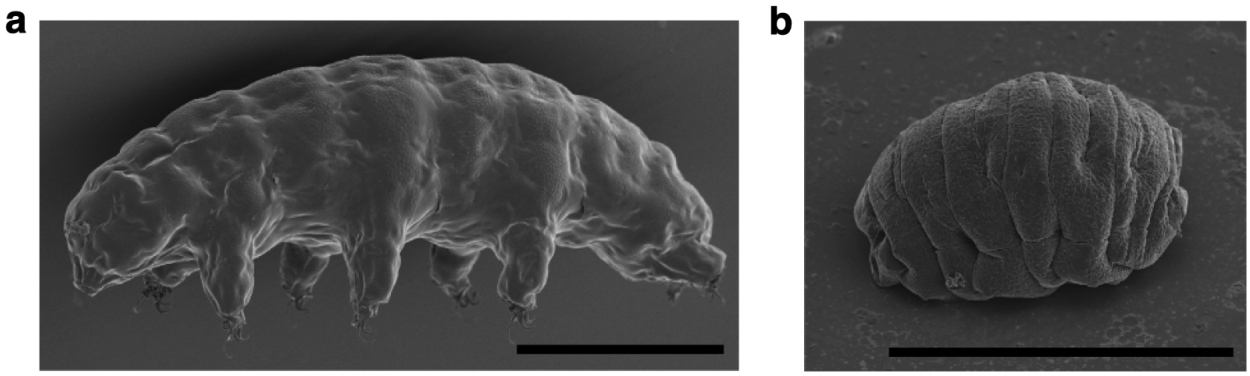


Figure 1. The extremotolerant tardigrade, *R. varieornatus*, sustains desiccation. Scanning electron microscopy images of *R. varieornatus* in hydrated (**a**) and dehydrated state (**b**). Scale bars, 100 μm. Adopted from reference [15]

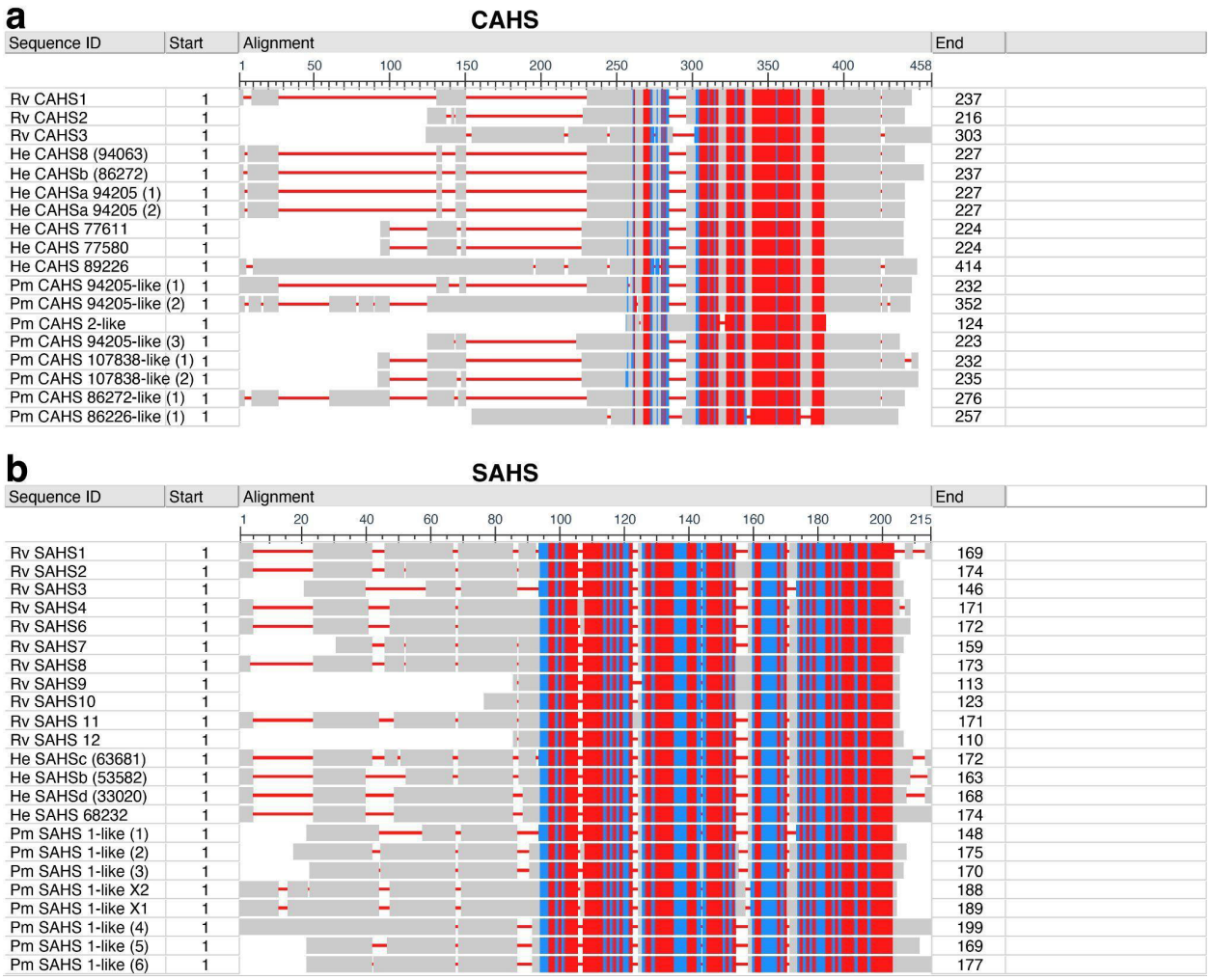


Figure 2. Sequence alignments of CAHSs (**a**) and SAHSs (**b**) generated by COBALT. Red encodes highly conserved amino acid residues, blue indicates less conserved ones, and gray represent highly variable ones.

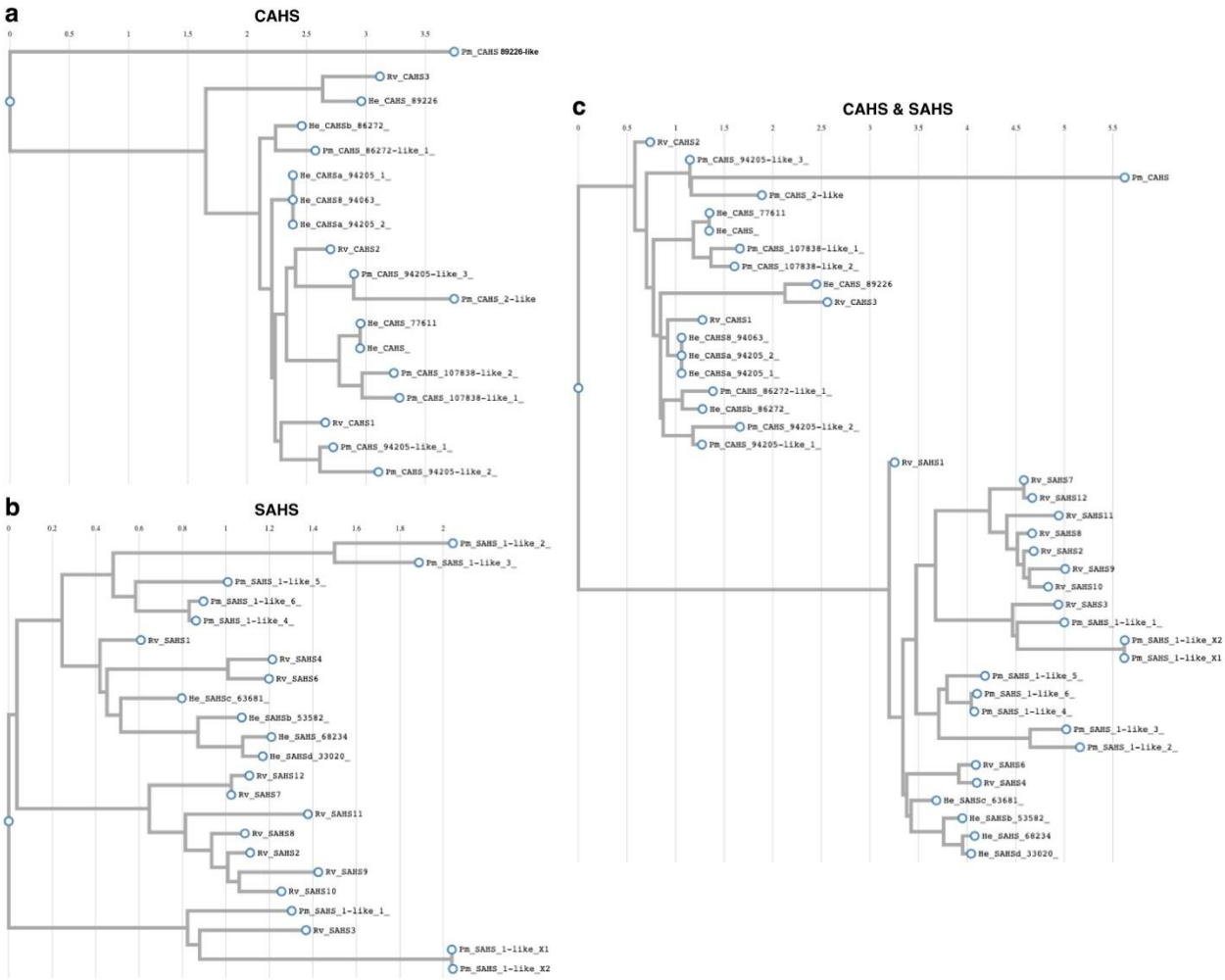


Figure 3. Phylogenetic trees of CAHSs (**a**), SAHSs (**b**), and CAHSs and SAHSs (**c**). Generated by ETE3. Prefixes: Rv, *R. varieornatus*; He, *H. exemplaries*; Pm, *P. metropolitanus*.

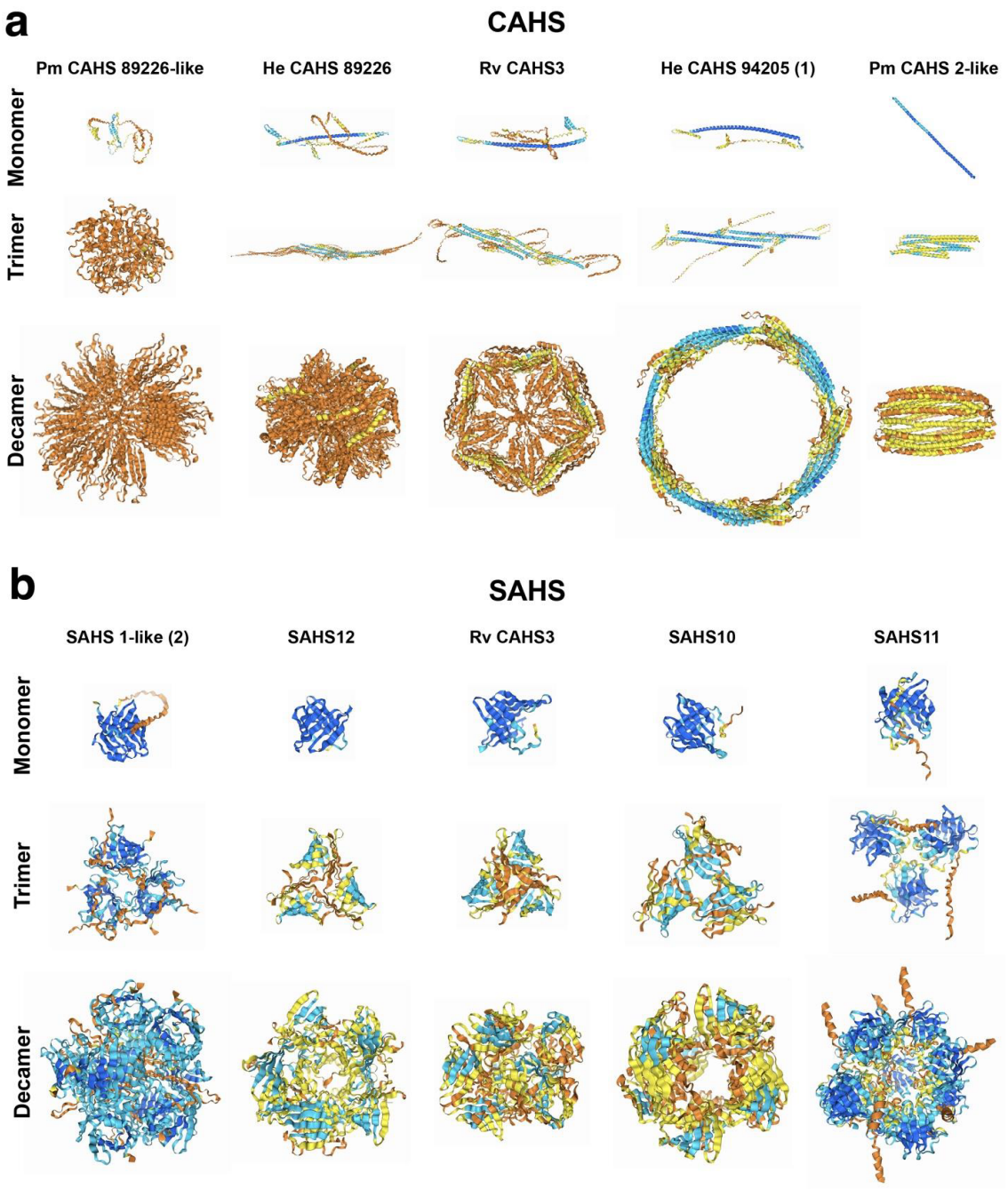


Figure 4. Selected structures of CAHS (**a**) and SAHS (**b**) monomers, trimers, and decamers are generated by AlphaFold 3.0. Dark blue indicates high confidence (i.e., plDDT > 90), cyan indicates confidence (i.e., 90 > plDDT > 70), yellow indicates low confidence (i.e., 70 > plDDT > 50), and orange indicates very low confidence (i.e., plDDT < 50). Generated by AlphaFold 3.0. plDDT, predicted local distance difference test score.

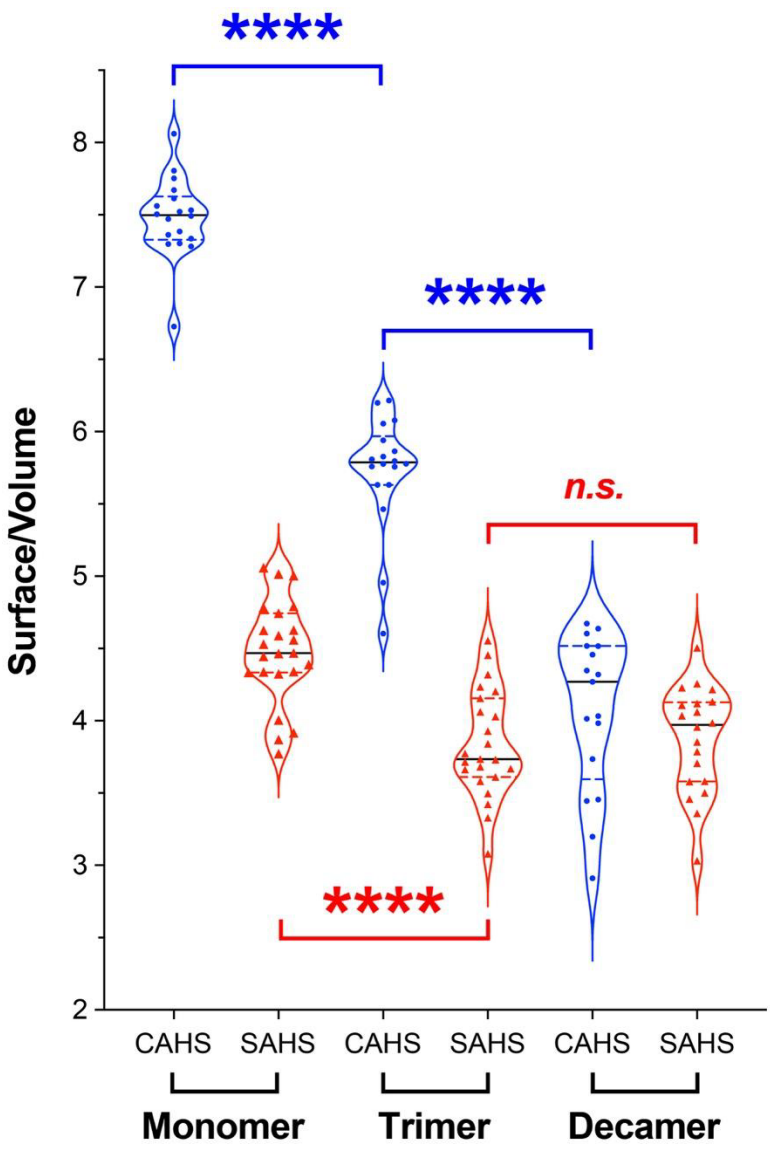


Figure 5. Surface area to volume ratio of all SAHSs and CAHSs, comparing monomers against trimers and trimers against decamers. Student *t*-test; ****, $p < 0.001$; n.s., $p = 0.3764$.

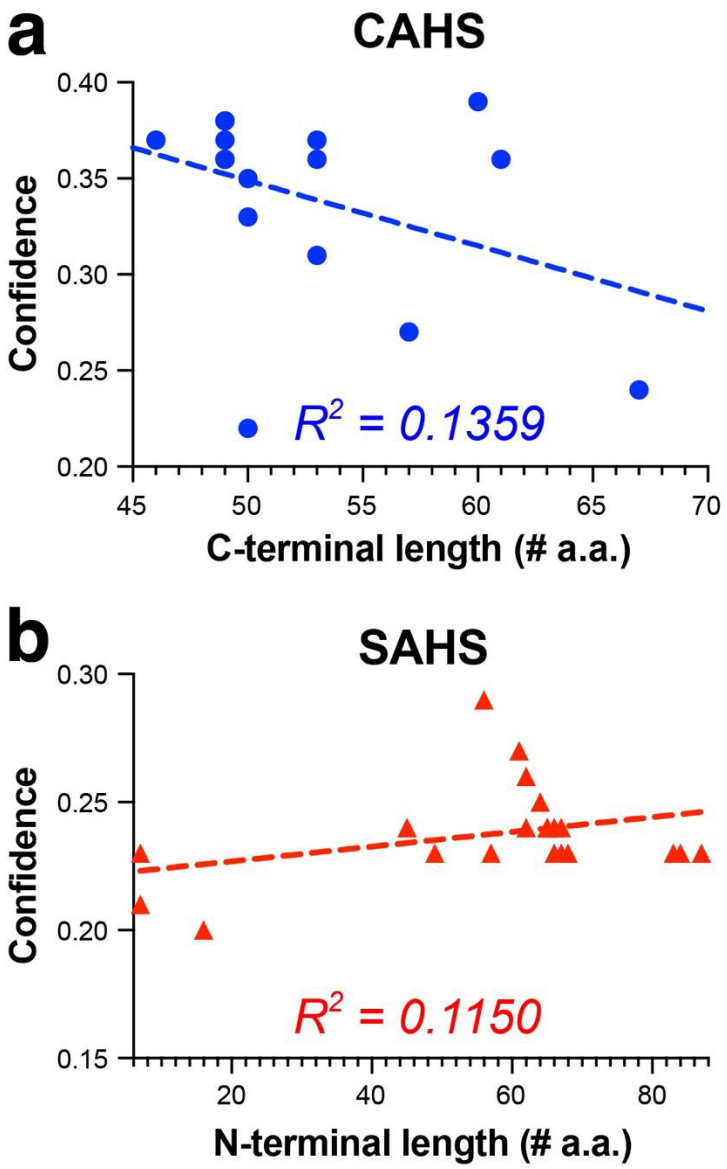


Figure 6. There is little correlation between the length of disordered terminal regions and the confidence predicting the decamer structure by AlphaFold 3.0. (**a**) scattering plot of CAHSs' C-terminal length (i.e., disordered region) and prediction confidence. Blue dots are data points of individual CAHSs and the blue dash line represents the best linear regression fitting. R^2 = 0.1359, and p = 0.1600. (**b**) scattering plot of SAHSs' N-terminal length (i.e., disordered region) and prediction confidence. Red triangles are data points of individual SAHSs and the red dash line represents the best linear regression fitting. R^2 = 0.1150, and p = 0.1134.

Bibliography

McInnes, S. J., Jørgensen, A., and Michalczyk, Ł. (2021) 20 years of Zootaxa: Tardigrada (Ecdysozoa: Panarthropoda). *Zootaxa*.

Goldstein, B. (2022) Tardigrades. *Nat Methods 19*, 904–905.

Arakawa, K. (2022) Examples of Extreme Survival: Tardigrade Genomics and Molecular Anhydrobiology. *Annu Rev Anim Biosci*.

KEILIN, D. (1959) The problem of anabiosis or latent life: history and current concept. *Proc R Soc Lond B Biol Sci 150*.

Boothby, T. C., Tapia, H., Brozena, A. H., Piszkiewicz, S., Smith, A. E., Giovannini, I., Rebecchi, L., Pielak, G. J., Koshland, D., and Goldstein, B. (2017) Tardigrades Use Intrinsically Disordered Proteins to Survive Desiccation. *Mol Cell 65*.

Bino, T., Goto, Y., Maryu, G., Arakawa, K., and Aoki, K. (2024) Possible roles of CAHS proteins from Tardigrade in osmotic stress tolerance in mammalian cells. *Cell Struct Funct 49*, 123–133.

Fukuda, Y., and Inoue, T. (2018) Crystal structure of secretory abundant heat soluble protein 4 from one of the toughest “water bears” micro-animals Ramazzottius Varieornatus. *Protein Science 27*.

Fukuda, Y., Miura, Y., Mizohata, E., and Inoue, T. (2017) Structural insights into a secretory abundant heat-soluble protein from an anhydrobiotic tardigrade, Ramazzottius varieornatus. *FEBS Lett 591*.

Lim, S., Reilly, C. B., Barghouti, Z., Marelli, B., Way, J. C., and Silver, P. A. (2024) Tardigrade secretory proteins protect biological structures from desiccation. *Commun Biol 7*.

Li, L., Ge, Z., Liu, S., Zheng, K., Li, Y., Chen, K., Fu, Y., Lei, X., Cui, Z., Wang, Y., Huang, J., Liu, Y., Duan, M., Sun, Z., Chen, J., Li, L., Shen, P., Wang, G., Chen, J., Li, R., Li, C., Yang, Z., Ning, Y., Luo, A., Chen, B., Seim, I., Liu, X., Wang, F., Yao, Y., Guo, F., Yang, M., Hua Liu, C., Fan, G., Wang, L., Yang, D., and Zhang, L. Multi-omics landscape and molecular basis of radiation tolerance in a tardigrade.

Erözden, A. A., Tavsanli, N., and Çalışkan, M. (2024, December 1) Advances in bioinformatic approaches to tardigrade phylogeny. *Comput Biol Chem*. Elsevier Ltd.

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C. C., O’Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A. I., Cowie, A., Figurnov, M., Fuchs, F. B., Gladman, H., Jain, R., Khan, Y. A., Low, C. M. R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E. D., Zielinski, M., Žídek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., and Jumper, J. M. (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature 630*, 493–500.

Papadopoulos, J. S., and Agarwala, R. (2007) COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics 23*, 1073–1079.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem 25*, 1605–1612.

KC, S., Nguyen, K. H., Nicholson, V., Walgren, A., Trent, T., Gollub, E., Romero-Perez, P. S., Holehouse, A. S., Sukenik, S., and Boothby, T. C. (2024) Disordered proteins interact with the chemical environment to tune their protective function during drying. *Elife 13*.

Boothby, T. C., Tapia, H., Brozena, A. H., Piszkiewicz, S., Smith, A. E., Giovannini, I., Rebecchi, L., Pielak, G. J., Koshland, D., and Goldstein, B. (2017) Tardigrades Use Intrinsically Disordered Proteins to Survive Desiccation. *Mol Cell 65*, 975-984.e5.

Sanchez-Martinez, S., Ramirez, J. F., Meese, E. K., Childs, C. A., and Boothby, T. C. (2023) The tardigrade protein CAHS D interacts with, but does not retain, water in hydrated and desiccated systems. *Sci Rep 13*.

CAHS and SAHS COBALT Alignment

```
Query_10001  1    MP[1]EKHVEqtVvEKTEQPG     ------HSSTHHAPAQRtVAREQEE    ---VVHKEFTHTDIRVPHIDAPPPI  60
Query_10002  1    --   ----------------     MSRDQGSTEYD-----A-NQRQEQH    -QEQHNTSYTHTDVRTNIPNIPAPF  43
Query_10003  1    --   ----------------[  1]SSRQNQQSSSQHSSSSQqGGQGGQG[68]NMESVRTGFTVTQDKN--LTVAAPN  117
Query_10004  1    MS[4]ESHMErnEkVVVNNSG     ------HAD---------VKKQQQQ    ---VEHTEFTHTEVKAPLIHPAPPI  54
Query_10005  1    MS[3]EKKVErtE-VVYGGDR     ------RVE---------GSASASA    ---EKTTNYTHTEIRAPMVNPLPPI  52
Query_10006  1    MS[4]ESHMErnEkVVVNNSG     ------HAD---------VKKQQQQ    ---VEHTEFTHTEVKAPLIHPAPPI  54
Query_10007  1    MS[4]ESHMErnEkVVVNNSG     ------HAD---------VKKQQQQ    ---VEHTEFTHTEVKAPLIHPAPPI  54
Query_10008  1    --   ----------------[  5]QESSYQYSDRSNNGQQQe----QQE    KKEVEHSSYTHTDVKVNMPNLIAPF  51
Query_10009  1    --   ----------------[  5]QESSYQYSDRSNNGQQQe----QQE    KKEVEHSSYTHTDVKVNMPNLIAPF  51
Query_10010  1    MA[2]ESKYErvEkVNVDADG[100]SNRQNSSSNVSSSDNSErASPASRN[69]QLESRRDGRTYTEDKE--LTIPAPV  237
Query_10011  1    MS[7]EAHIEhhQeVRKEGSG     ------ATQTQHA--------EKEH    ---TVHTGYTHTQVTAPLVAPAPPI  58
Query_10012  1    MS[3]ERHME--AeVCRGGAY[ 35]RDSAGSRQQQQEMRRDSaGSRQQQQ[77]EKEVVNETHVHTEIRAPLVAPPVAH  181
Query_10013       --   ----------------     -------------------------    -------------------------
Query_10014  1    --   ----------------     MSQHHHAHNYECYSHTS-GAKEDTH[ 4]EKEVHTTSYTHAEAKAPVLNVPDPL  53
Query_10015  1    --   ----------------[  7]MSMNPDAVFIPPEQQQQqGREVQQE    RKEAQQTSYMTSQVKVPMMHLPSPF  57
Query_10016  1    --   ----------------[  7]QQSKQQYSREQRDQREQr----EVQ    HSEVEQTSYMQTQIKVPIMKLPTPS  53
Query_10017  1    MT[2]EKKVVekVeVRTTGNA[ 39]CTERHEHHQHHHHEHDR---TCETH    ---EHSTNYTHTEVKAPVLNPPAPF  103
Query_10018  1    --   ----------------     -------------------------[73]VPSSVTESGTAPEVPK---NIPGPF  95

Query_10001  61   IAASAVGLAEEIVSHGFQASAARISGASTEVD    MRPSPKLAEEARRDAERYQKEHEMINRQAEATLQKKAEEYRHQT  136
Query_10002  44   ISTGVSGLGQQLVGEGFTASAARISGQSSETH    VQMTPEMEAEARKDRERYERELQAINERHQRDIEGKTEAYRKQA  119
Query_10003  118  IAAPIHSNLDLNLGGGARAEIT--AGTTVDLS[ 2]-----QRKDLGPEEYARYKAKVEQLARQDEQDAGMRAAQYREEV  188
Query_10004  55   ISTGAAGLAEEIVGQGFTASAARISGGTAEVH    LQPSAAMTEEARRDQERYRQEQESIAKQQEREMEKKTEAYRKTA  130
Query_10005  53   ISTGAAGLAQEIVGEGFTASATRISGAAATTQ    VLESQASREQAFKDQEKYSREQAAIARAHDKDLEKKTEEYRKTA  128
Query_10006  55   ISTGAAGLAEEIVGQGFTASAARISGGTAEVH    LQPSAAMTEEARRDQERYRQEQESIAKQQEREMEKKTEAYRKTA  130
Query_10007  55   ISTGAAGLAEEIVGQGFTASAARISGGTAEVH    LQPSAAMTEEARRDQERYRQEQESIAKQQEREMEKKTEAYRKTA  130
Query_10008  52   ISSSA-GLAQELVGEGFQASVSRITGASGELT    VIDTEAETEEARRDLEAKAREQELLSRQFEKELERKTEAYRKQQ  126
Query_10009  52   ISSSA-GLAQELVGEGFQASVSRITGASGELT    VIDTEAETEEARRDMEAKAREQELLSRQFEKELERKTEAYRKQQ  126
Query_10010  238  VAPQIHAHQQVNMSGGTSATIH--A--TTDLH    LASEAQINDMGPEEYERYRAKVEALARIHEDETSRKAAAYRNAV  309
Query_10011  59   IQTPR--MAEELVG-GFTGSAARYTATTGELN    IQPSERLLQEAARDSAAHSRESEAIAKAHERELEKKTEQYRKEA  131
Query_10012  182  VEVQG-GFGD--LAAGFTGSSARYTATSGEVF    VQPSPRALEEARRDSLSHQKEADSIARAHEQDLMKRSEKYQSET  254
Query_10013  1    ----MVGAREEL--QHGQAECSRRDALAQESD[12]ATETNVRRKGGYREEEARQKE-----ERQKHALKAKTEMYRTEA  77
Query_10014  54   TVTGIAGLGQTLVGEGFAASAVRASGESERVT    VLPNEHLTEEARRDQEKYQKEKEAIAERQKHALEAKTEAYRKEA  129
Query_10015  58   MSTGF-S-AQELIGEGFQASISRISAVSEDMR    QLDSPELDEEARKDMAAKQREHDMLTQQYEKEVERKTEAYRKLQ  131
Query_10016  54   IIT-V-PLAQELVGEGFQASISRISGVSEEIQ    IVENAQLEEETRRDYEAKQREQEQLQLAFEKEVEKKTEAYRKQQ  127
Query_10017  104  VVTSASGLAEEIVSEGFSASAARVAGEATATA    VCETAASSKQAAVDRERYEREKAAIARQHEKELEKKTESYRKQA  179
Query_10018  96   ISEEDSIHSRESHDAVIVGHLPNNGHTLKKME[ 3]ITKSPDSISMQSGSNISRTSQQALPGATGTKEVHERVTS----K  170
```

```
Query_10001  137  EAEAEKIRRELEKQHERDIQFRKDLIDQTIEKQKREVDLEAKMAKRELDREAQLAKEALERSRMATNVEVTLDTAAGHTV  216
Query_10002  120  EQEAERLRKELEKQHQRDIEFRKSLVQGTIENQKRQVELEAQLAKRELDREARLATQALDQSKMATDVQVNFDSAVGHTV  199
Query_10003  189  ERDAELIRQILERQHIRDLEFRKEMVENQVNRQEREIQLEAEYAMRALELERNAAKEALESAKAQTNVNVKVESAIGTTV  268
Query_10004  131  EAEAEKIRKELEKQHARDVEFRKDLIESTIDRQKREVDLEAKMAKRELDREGQLAKEALERSRLATNVEVNFDSAAGHTV  210
Query_10005  129  EAEAEKIRKELEKQHARDVEFRKDLVESAIDRQKREVDLEAKYAKKELEHERELAMNALEQSKMATNVQVQMDTAAGTTV  208
Query_10006  131  EAEAEKIRKELEKQHARDVEFRKDLIESTIDRQKREVDLEAKMAKRELDREGQLAKEALERSRLATNVEVNFDSAAGHTV  210
Query_10007  131  EAEAEKIRKELEKQHARDVEFRKDLIESTIDRQKREVDLEAKMAKRELDREGQLAKEALERSRLATNVEVNFDSAAGHTV  210
Query_10008  127  EVETEKIRKELEKQHLRDVEFRKELMEQTIENQKRQIDLEARYAKKELERERNKVKRVLERSKFHTDIQVNMEAAAGSTH  206
Query_10009  127  EVETEKIRKELEKQHLRDVEFRKELMEQTIENQKRQIDLEARYAKKELERERNKVKRVLERSKFHTDIQVNMEAAAGSTH  206
Query_10010  310  EADAELIRQTLERQHMRDIEFRKDLVESSVDRQQQEIRLEAEYAMRALEQERVNARAALDQAMASTNIDVNIDSAIGTTH  389
Query_10011  132  EAEAEKIRKELEKQHERDIEFRKGLIDSAIERQKREVELEAKMAKRELEREAQLAKEALERSKLATNVEVNFDSAVGHTA  211
Query_10012  255  EAEALKIRKEMEKQHERDIAFRKSLIDSAVDRQIREVDLEARMAKRELLREAQIAKEQLDRASAATTVEVDLKTAVGHTH  334
Query_10013  78   EQQAEKIPKEVEKQHGRDVGFRQDLVSETIAAQKKQVDLEAQLAKKL--------------------------------  124
Query_10014  130  EQQAEKIRKEMEKQHERDVGFRQELVSETIAAQKKQVDLEAQMAKKQLEREGQAAREALEKSKMTSNVEVNFDTAIGHAV  209
Query_10015  132  EAEADKIRKELEKQHLRDVEFRKDMVEMAIDNQKRQIDLESRYAKKEMDRERTRCRLMLEQQKFHSDISVNLEASAAHTE  211
Query_10016  128  EIEAEIIRKELEKQHIRDVEFRKELVEHAIENQKRQIDIESRYAKKELERERAKARMMLEKQKFHSDISVNLESHAANTQ  207
Query_10017  180  EAEAEKIRKELEKQHARDVEFRKDILETTIDRQKKEVELEAKMAKKELEHERQLALEALERSKMATNIEVNFESAAGRTV  259
Query_10018  171  APDTASIRSGSNISRASDQEVANKIEKTSIH--------DMPTGKSDPEYSAKTEQKVKDDGRKVTKTEEHVQDDMGHAT  242

Query_10001  217  SGGT--TV-SSVDKVETVRE[ 4]  237
Query_10002  200  SGAT--TV-SQSEKVTQSKH      216
Query_10003  269  SKGA---IqTSADKSSTTKT[18]  303
Query_10004  211  SGGT--TV-STSDKMEIKRN      227
Query_10005  209  SGGT--TV-SEHTEVHDGKE[12]  237
Query_10006  211  SGGT--TV-STSDKMEIKRN      227
Query_10007  211  SGGT--TV-STSDKMEIKRN      227
Query_10008  207  SGSSsvAV-SESEKFQTNN-      224
Query_10009  207  SGSSsvAV-SESEKFQTNN-      224
Query_10010  390  SQGR---VtTTSESRTSQAR[ 8]  414
Query_10011  212  SVGT--TV-SESESVTRDVR[ 4]  232
Query_10012  335  SAGV--TA---SETVMRDVR[ 3]  352
Query_10013       --------------------
Query_10014  210  SGST--TVeSEVNKLN----      223
Query_10015  212  TSGQ--CV-SESQKFTENNR[ 4]  232
Query_10016  208  AGEQlqVV-SESEKFQSNQK[ 9]  235
Query_10017  260  SEGT--VV-SEHTDIVHPRM      276
Query_10018  243  TKKKfeEAaPDESRV-----      257
```

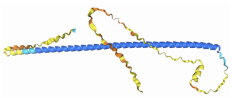
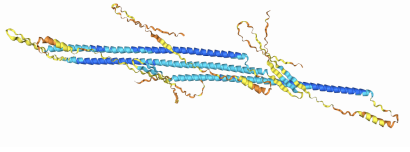
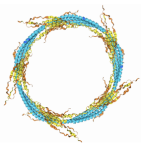
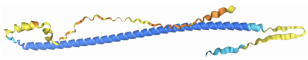
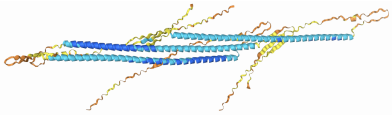
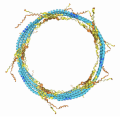
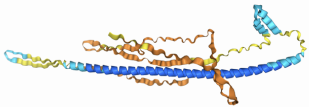
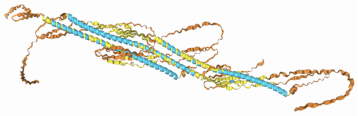
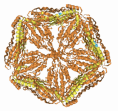
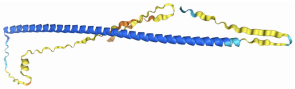
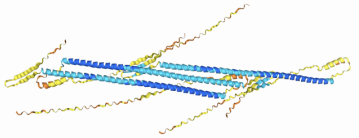
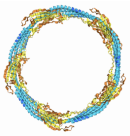
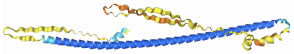
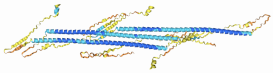
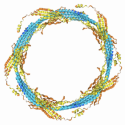
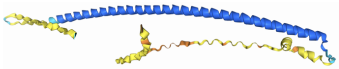
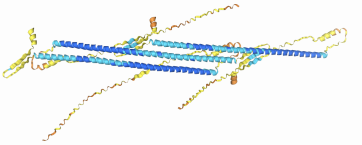
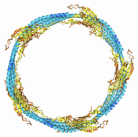
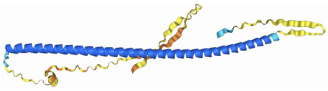
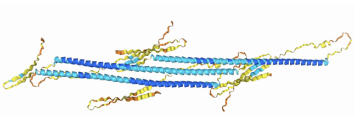
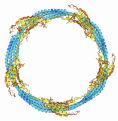
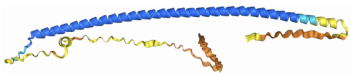
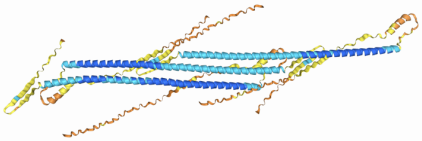
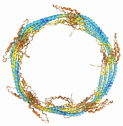
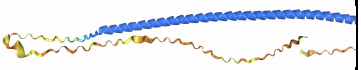
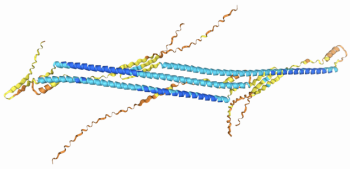
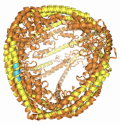
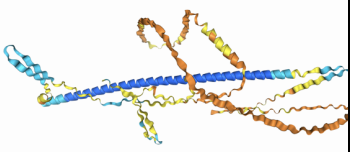
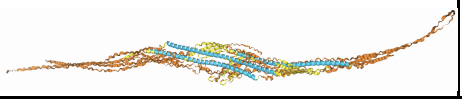
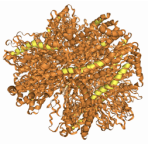
Figure S1. COBALT sequence alignment result for 18 CAHSs (in the same order as Figure 2a)

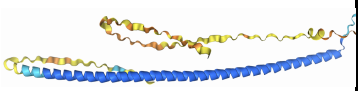
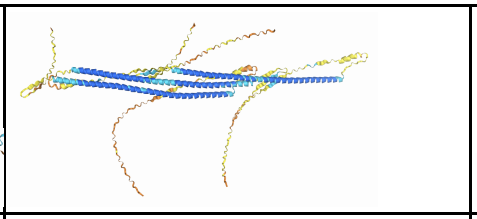
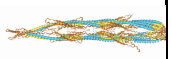
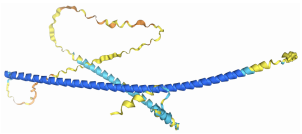
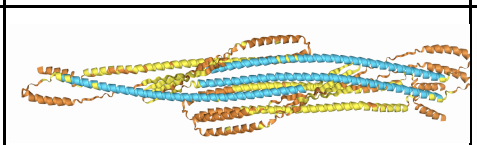
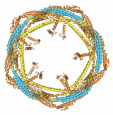
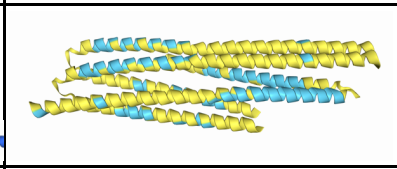
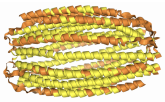
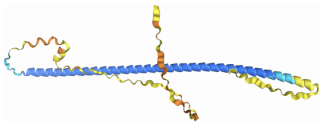
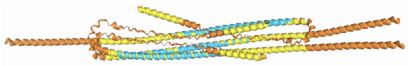
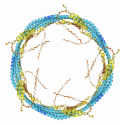
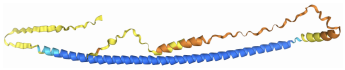
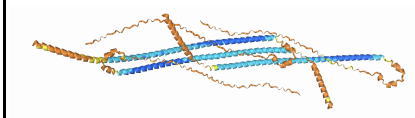
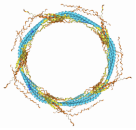
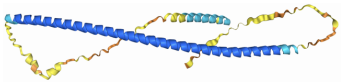
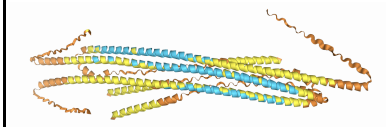
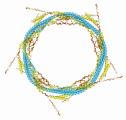
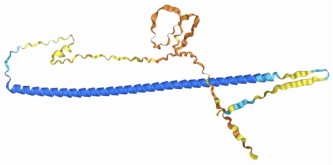
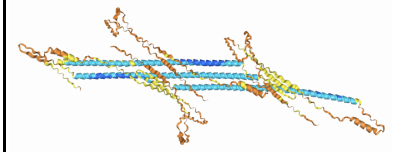
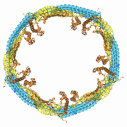
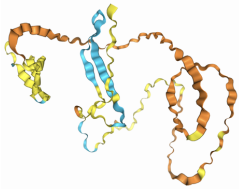
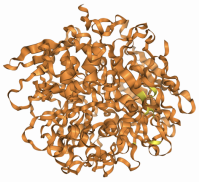
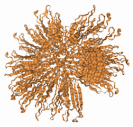
```
Query_10001  1    MSRA    AVAIALLGCVVAAYGAPA[2]HDDAKAEWTGKSWMGKWES--TDRIENFDAFISALGLP--LEQYG-GNHKT  70
Query_10002  1    MHRF    VLALVVFAGAAIVWAADD[2]HEEG-VEWTGKPWMGKWESD-PSKDENVEEFKKKLQLPM-SHSEMNKNSKV  72
Query_10003  1    ----[ 3]FAAILVMGFAAVAVAH--   -----------VPLGKYETT--GRMENTEAFSVASGTPH-------VNDTI  50
Query_10004  1    MSPA    LVRVILCGLVVTIAARP-   HDESKAQWTGKPWLGKWESI-DGTPENWEAFVKAANIPPkDQALYNGKQKT  71
Query_10005  1    MSSA    LVAAILCGLAVIIAARP-   HDESKAQWTGKPWLGKWEST-DKTPENWEAFVKAANIEPkYQSLYSGKQKA  71
Query_10006  1    ----    -------MASKQIQDADI[2]GEEG-AEWTGKPWLGKWCSI-PEKDHNLEAYLAAMGVDM-SHPNMKKDQPV  61
Query_10007  1    MKC-    ILALALFAGVAVVWAGDD[2]HEEG-VDWTSKPWLGKWESI-PEKDENLVEFLKKLNVPM-DHSKMNATVKV  71
Query_10008  1    ----    ------------------   -------------------------------------M-SHPEVNLNSTV  12
Query_10009  1    ----    ------------------   ----------------------------MGMRSADIPL-SHPHLKHNNRV  21
Query_10010  1    MRRF    LRNLLIFASAAVCLAEHD[2]-HEEGAEWTGKPWLGKWVSV-PEKDVNVLNFITEIGVAA-SHPELPPIVTV  72
Query_10011  1    ----    ------------------   -------------------------------------M-NHPNMKKDQPV  12
Query_10012  1    MSRT    IVALILLGLAALAAADHH[2]HG-AEKEWAGKAWLGKWVS--TDRSENWDAFVEALGLP--LAAYG-GNHKT  69
Query_10013  1    MARL    FVAVALFGVVAFAAAE--   -----KEWTGKTWLGSWAS--TDRAENWEAFVDALGLP--SDQYPREVQRT  62
Query_10014  1    MARF    LVALALFGVVAMTAAS--   -GDAPKEWSGKPWLGKFVAEvSDKSENWEAFVDALGLP---DQYPRAQLKT  67
Query_10015  1    MARF    LVALALFGVVAMTAAT--   -GDAPKEWSGKPWLGKFVAEvTDKSENWEAFVDALGLP---EQFGRAPVKT  67
Query_10016  1    ----[ 2]FMRIICLGTLAIAAIEGA[2]----------DAFLGKYEST--DRRENTQAFAAALKQDD-------IHAKI  54
Query_10017  1    ----[ 6]VIAAVCFVGAHAAVGRAH[3]QGGSSDENVAHMWLGKWESE-QGSEENFDQLLDAIKEAF----PYYTNKKI  73
Query_10018  1    ----[ 1]LLSVCILGMVAASVTSVP[5]TAQSGHTDHSHAWLGKWEST-KEGEENLAQLLDQIRDAI----PHYTDKKV  70
Query_10019  1    MTST[15]VVTAVLPADLASKAAVNS[2]TNDAVDKPALLIPFGKFEAT--DQTENFANYRSSLNVEFkGFSAKHLKGKV  88
Query_10020  1    MTST[16]VVTAVLPADLASKAAVNS[2]TNDAVDKPALLIPFGKFEAT--DQTENFANYRSSLNVEFkGFSAKHLKGKV  89
Query_10021  1    MLQF[19]IKAFLLIALVAAVSARPA[6]EQNEIAVDADHPWLGKWESI-DGRQENFANFVNALGFAH-----YTAEHKV  92
Query_10022  1    ----[ 2]GTFAIIFGLIACVIGHGQ[1]DPKDIAADPAHPWIGRWESI-DGRHENFENFIKHLGLAH-----YGADNKV  66
Query_10023  1    ----[ 2]FKAFLLIALVAAVSGRPA[5]EQNEIAADADHPWIGKWESI-DGRQENFANFINALGFAH-----YTAEHKV  70

Query_10001  71   FHKIWKE--GDHYHHQISVPDKNY--KNDVNFKLNEEGTTQHNNT-EIKYKYTEDG   GNLKAEVHVPSR-NKVIHDEY  141
Query_10002  73   WIHHYKK--GDEYHHKIIINDAHY--KNDIVFKLGQESAGSYNGS-SFSVKYEDKD[4]GSVHYTGTKEQSlDKTINNVF  148
Query_10003  51   YNEFTQN--GDEYVHKFSVPSANY--HQELPFKLGEQRTGQYRGK-DFTYKYYMDE   GHLKAHFDGNNK---HIKHSY  119
Query_10004  72   LLKYWKEagEDHYHVQTSFPGTEH--KMETSFKMGQEGTLSHDGV-DLKYVCTEDG   EQLITKINIPSK-NQETIVTY  144
Query_10005  72   IITIYKE-gDSHYHAQMTFPGTDH--KKEWDFKIGQEGTYSMDGT-EVKYVYTENG   DQLDSKLNIPSK-NTEMTHTY  143
Query_10006  62   TLHTFKK--GDEYHHKIVVKEAGY--INDVIFRLGQETPGSYNGQ-QFTVNYEEQG   GALVGTIKYPAH-NKVIHNTY  132
Query_10007  72   HLNHYKK--GDDYHHKIIVKEAEY--KNDVVFKLGQESAGSYNGS-SFTVKYEDKD[4]GTIHYTGTKEQSlDKTINNEY  147
Query_10008  13   LVNHLKK--GDEYHHKIIIKEYYT---NHVVYKLGEQSPGSYDGL-SYSVKYGEKD[4]GTAHYTGTKDQRlNITMHNVY  87
Query_10009  22   WVNTYKK--GDEYHHKIIIKEAGY--TNDVVFKLGQESAGSHNGS-SFSLKYEDKD[4]GTVHRTGTKEQPlDKTINNVF  97
Query_10010  73   LVNHYKK--GDEYHQRLRVKEVADldDHDIVYKLGQETKNVFNGT-TFSVKYDEKD   DALVGQVMLPSN-NATYKNEF  145
Query_10011  13   TLQTFKK--GDKYHHKIVVEEAGY--INDVIFRLGRETPGSYNGQ-QITVNYEEQG   GALVGTVKYPAH-NKVIHNTY  83
Query_10012  70   VHKLWKE--GDHYHHQIIIADKSY--KQDIQFKLGEEGRTAHNGT-EVTFKYTEVG   DNLQNEVKIPSK-NKTISDSY  140
Query_10013  63   IHTIYKQ--GDKYHHEVSIPSKNF--KKAIEYTLGTETDVQHGPH-TIKLKYTEDG   EKLVADVQIPSK-NKQIHDIY  133
Query_10014  68   IHSFYKQ--GEHYHHILSLPDKNI--NKDIEFTLGQEVEIKHGEH-SLKIKYFEDG   NKLVADVSIPAK-GKSIHDVY  138
Query_10015  68   IQKIYKQ--GDHYHHIFALPDKNF--EKDIEFTLGQEVEIKQGEH-IAKTKYSEDG   EKLVADVSIPTK-GKTIRSEY  138
Query_10016  55   INEFSAS--GDEYHHKFSVPSKNY--VQDLPFKLGEERQAAFNGT-NYKYKYTLEG   DSLKANFDLPDR---QVDQMF  123
Query_10017  74   IHDFVRK-sDDDYVHKVKIGQGED--QYQVSFKLNEEGTLHKAGApEVKFVYKETS[1]NKLVVQQRVASK-NIVLEEIY  147
Query_10018  71   THEYIEE--GEQFLHKVKIVDGGQ--EYEVRFKLNEEATFHLGNEpDIKYVYKEEG[1]NKLTVHMKIPSK-NKELKECY  143
Query_10019  89   QHEFSRI-nSTTYAHAFWIVDTPY--KQNLPFELGREYPHSYNGT-AFKYRYYQEP[3]-GLHAVFHVPADnPLPIEHLY  163
Query_10020  90   QHEFSRI-nSTTYAHAFWIVDTPY--KQNLPFELGREYPHSYNGT-AFKYRYYQEP[3]-GLHAVFHVPADnPLPIEHLY  164
Query_10021  93   WHKLWKE--GDHYHHRIKVPEKNY--KLDVEFKLGEEGTGSFNNT-NFKYKYTEEG   KDLHVEVNLLTH-NKVINDIY  163
Query_10022  67   YHKFWKE--DDHFHHGIAVTEKDF--KKFVEFKLGEEGTLTWNNT-EFKYKYTEEN   KDLHVEVKVPAK-DKVVHDTY  137
Query_10023  71   WHKLWKE--GDHYHHRIKVPEKNY--KLDVEFKLGEEGTGSFNNT-NFKYKYTEEG   KDLHAEINLLTH-NKVITDLY  141
```

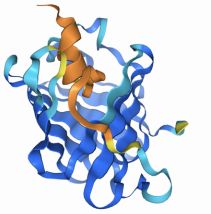
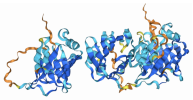
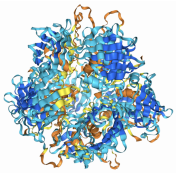
```
Query_10001  142  KVNGDELEKTYKVGDVTAKRWYKK---[4]  169
Query_10002  149  KLEGDHLVKTSTIEGVTMKRHYNKRQ-     174
Query_10003  120  KTEGDGYIKEYKTGNVMAKVWYKKSQQ     146
Query_10004  145  TATGDDLEQTFTSNGVTGKRWYKKIH-[1]  171
Query_10005  144  KVTGDELEHIFTSNGATGKKWYKKVNN[2]  172
Query_10006  133  KMDGQNLTKTSECEGVVHKRWYNKQQN     159
Query_10007  148  KVEGNQLVKTSTLEGVTHKRYYNKRN-     173
Query_10008  88   KLEGDRLLKSSTIDGVTLNCHHKRRI-     113
Query_10009  98   KLEGDHLVMTSTIDGVTMKRYYKTRT-     123
Query_10010  146  KVEGDFLVKTSDAHGIVHKRYYKRRN-     171
Query_10011  84   EMDGQNLAKTSECEGVVHKRWYNKQQN     110
Query_10012  141  VVKGDELEKTYKINDVVAKRWYKKHAH[5]  172
Query_10013  134  EVQGDTLTKTYKVGDVVAKRWFTREAN[3]  163
Query_10014  139  DVQGDQLIKSYKVGDVVAKKWFKKVAN[3]  168
Query_10015  139  EVQGDQLIKTYKTGDIVAKKWFKKVAN[9]  174
Query_10016  124  TFSGNEMVKTYKVGDVTAKVWFTKV--     148
Query_10017  148  NVQGDKIMKEYVSGDIRAKRIYRRMNH[1]  175
Query_10018  144  NVEGDKIVKNYESGNVKAKRIYKRVQN     170
Query_10019  164  TSSSEGFILTYKIGDMIAKRSYKRV--     188
Query_10020  165  TSSSEGFILTYKIGDMIAKRSYKRV--     189
Query_10021  164  HVDGDELVKTYKSGDVQAKRWYRRKSH[9]  199
Query_10022  138  HVEGEQLVKTYKVDNVEAKRWFAKATS[5]  169
Query_10023  142  HVEGDELVKTYKVGDVQAKRWYKRKSH[9]  177
```

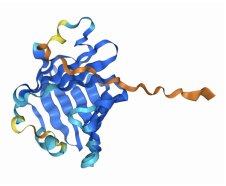
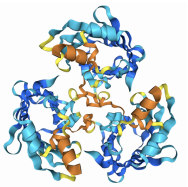
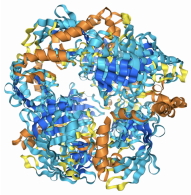
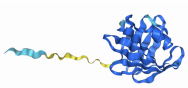
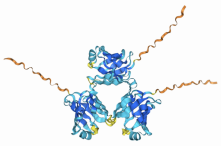
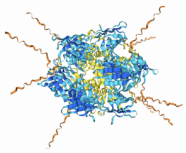
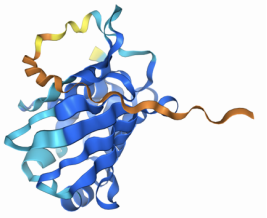
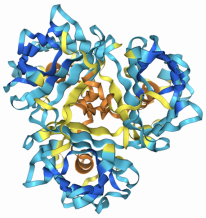
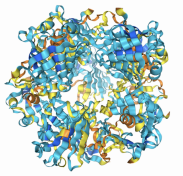
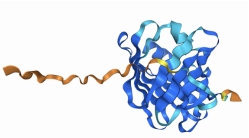
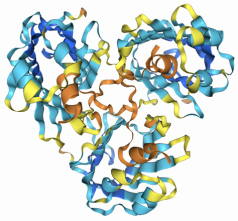
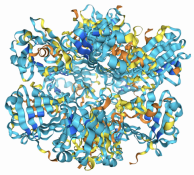
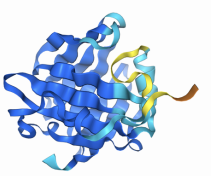
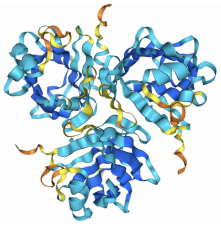
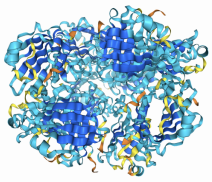
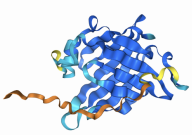
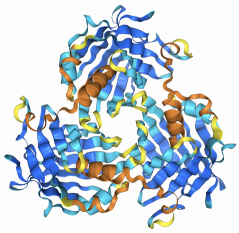
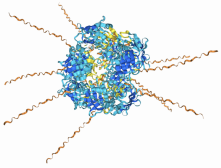
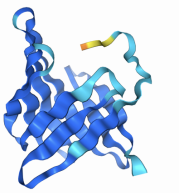
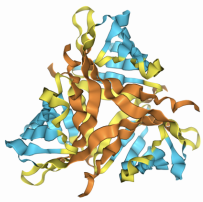
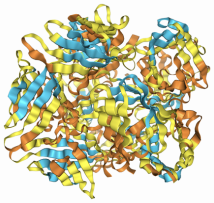
Figure S2. COBALT sequence alignment result for 23 SAHSs (in the same order as Figure 2b)

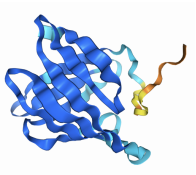
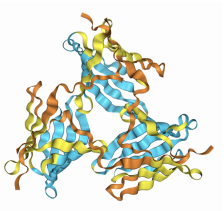
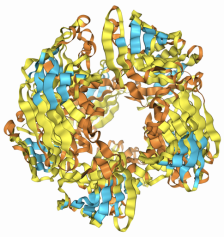
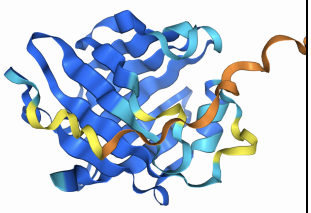
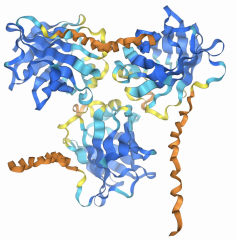
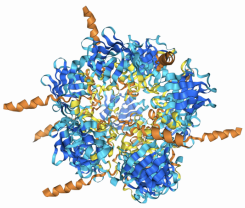
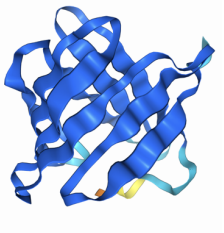
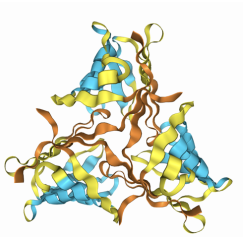
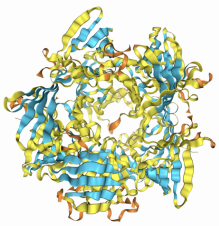
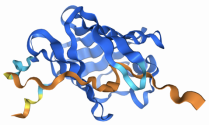
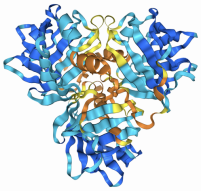
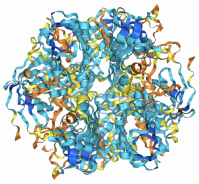
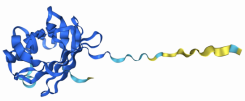
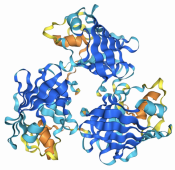
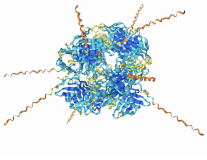
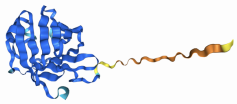
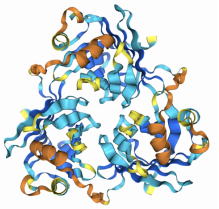
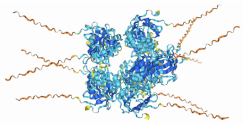
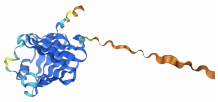
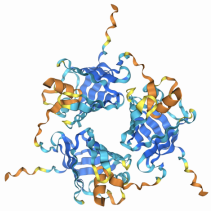
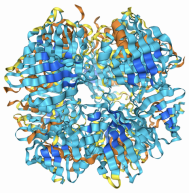
CAHS and SAHS AlphaFold 3.0 Structures

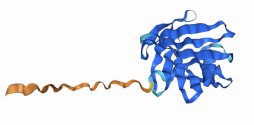
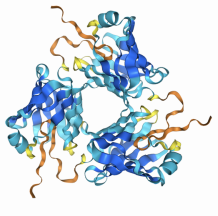
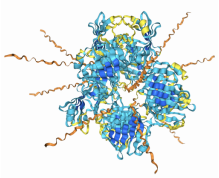
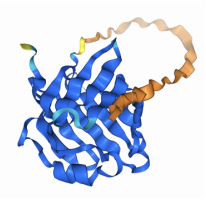
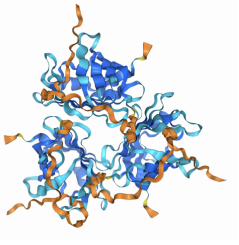
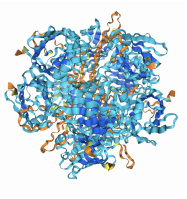
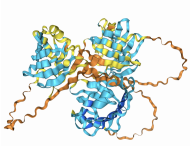
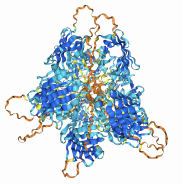
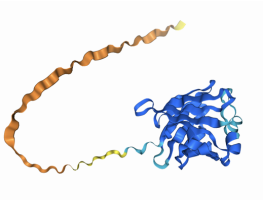
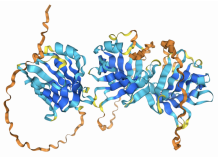
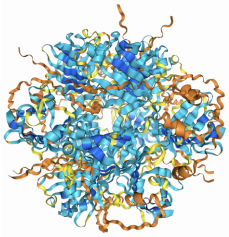
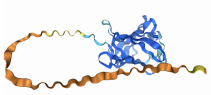
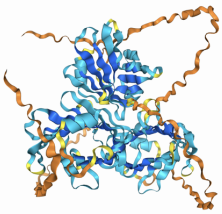
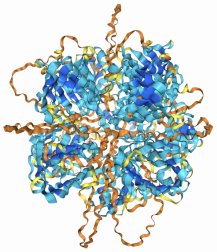
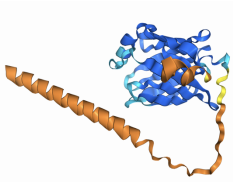
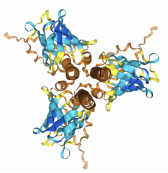
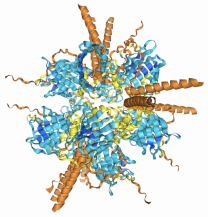
Name	GenBank ID	Monomer	Trimer	Decamer
Rv CAHS1	402783758			
Rv CAHS2	402783760			
Rv CAHS3	402783762			
He CAHS 94063	1209747827			
He CAHS 86272	1209747819			
He CAHS 94205 (1)	1209747816			
He CAHS 94205 (2)	1209747825			
He CAHS 77611	1209747824			
He CAHS 77580	1209747822			
He CAHS 89226	1209747829			

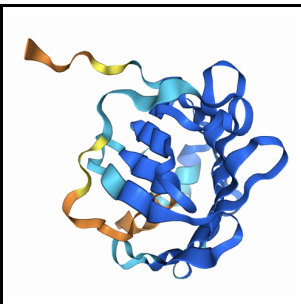
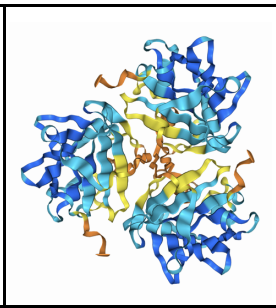
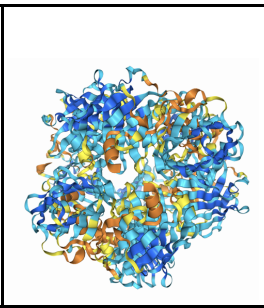
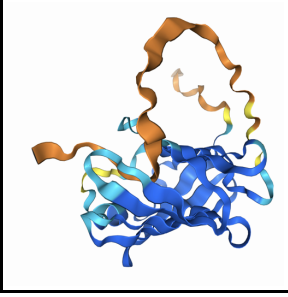
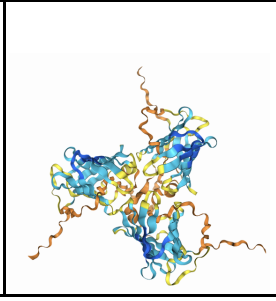
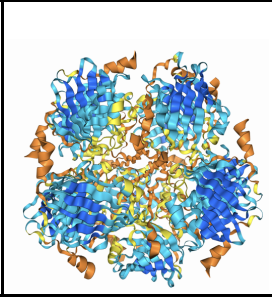
Pm CAHS 94205-like (1)	2493494936			
Pm CAHS 94205-like (2)	2493480804			
Pm CAHS 2-like	2493480067			
Pm CAHS 94205-like (3)	2493479718			
Pm CAHS 107838-like (1)	2493469471			
Pm CAHS 107838-like (2)	2493468852			
Pm CAHS 86272-like	2493445383			
Pm CAHS 89226-like	2493438951			

Name	GenBank ID	Monomer	Trimer	Decamer
Rv SAHS1	402783754			

Rv SAHS2	402783756			
Rv SAHS3	1101394461			
Rv SAHS4	1101395371			
Rv SAHS6	1101395435			
Rv SAHS7	1101395438			
Rv SAHS8	1101395445			
Rv SAHS9	1101395447			

Rv SAHS10	1101395452			
Rv SAHS11	1101395453			
Rv SAHS12	1101395666			
He SAHS 63681	1210274786			
He SAHS 53582	1210274784			
He SAHS 33020	1210274782			
He SAHS 68234	1210274787			

Pm SAHS 1-like (1)	2493482411			
Pm SAHS 1-like (2)	2493456970			
Pm SAHS 1-like (3)	2493420974			
Pm SAHS 1-like X2	2493398470			
Pm SAHS 1-like X1	2493398467			
Pm SAHS 1-like (4)	2493395038			

Pm SAHS 1-like (5)	2493392926			
Pm SAHS 1-like (6)	2493392853			

CAHS and SAHS Data

		Surface Area (angstrom^2)			Volume (angstrom^3)			Surface Area/Volume (angstrom^-1)			Confidence Score (pTM)			Tail Length
Name	GenBank ID	Monomer	Trimer	Decamer	Monomer	Trimer	Decamer	Monomer	Trimer	Decamer	1	3	10	(aa)
Rv CAHS1	40278375 8	1.95E+04	4.87E+04	1.38E+05	2.60E+04	8.43E+04	2.94E+05	7.50E-01	5.78E-01	4.67E-01	0.38	0.47	0.36	53
Rv CAHS2	40278376 0	1.73E+04	4.47E+04	1.22E+05	2.37E+04	7.52E+04	2.63E+05	7.30E-01	5.94E-01	4.64E-01	0.38	0.44	0.36	49
Rv CAHS3	40278376 2	2.41E+04	5.96E+04	1.20E+05	3.22E+04	1.03E+05	3.76E+05	7.47E-01	5.78E-01	3.20E-01	0.31	0.28	0.24	67
He CAHS 94063	120974782 7	1.84E+04	4.60E+04	1.14E+05	2.46E+04	7.84E+04	2.86E+05	7.49E-01	5.86E-01	3.98E-01	0.39	0.52	0.37	49
He CAHS 86272	120974781 9	1.89E+04	4.77E+04	1.27E+05	2.58E+04	8.19E+04	2.92E+05	7.30E-01	5.83E-01	4.35E-01	0.38	0.46	0.36	61
He CAHS 94205 (1)	120974781 6	1.83E+04	4.54E+04	1.25E+05	2.47E+04	7.89E+04	2.80E+05	7.38E-01	5.76E-01	4.46E-01	0.39	0.53	0.38	49
He CAHS 94205 (2)	120974782 5	1.82E+04	4.42E+04	1.15E+05	2.50E+04	8.09E+04	2.86E+05	7.28E-01	5.46E-01	4.01E-01	0.39	0.51	0.38	49
He CAHS 77611	120974782 4	1.86E+04	4.61E+04	1.26E+05	2.48E+04	7.93E+04	N/A	7.53E-01	5.81E-01	N/A	0.39	0.51	0.35	50
He CAHS 77580	120974782 2	1.87E+04	4.60E+04	1.04E+05	2.47E+04	7.94E+04	2.58E+05	7.56E-01	5.80E-01	4.03E-01	0.38	0.49	0.22	50
He CAHS 89226	120974782 9	3.38E+04	7.07E+04	1.30E+05	4.36E+04	1.43E+05	4.48E+05	7.75E-01	4.96E-01	2.91E-01	0.29	0.27	0.27	57
Pm CAHS 94205-like (1)	249349493 6	1.90E+04	4.82E+04	1.30E+05	2.47E+04	7.93E+04	2.82E+05	7.67E-01	6.08E-01	4.60E-01	0.38	0.51	0.31	53

Pm CAHS 94205-like (2)	249348080 4	3.01E+04	7.06E+04	1.61E+05	3.86E+04	1.23E+05	4.68E+05	7.81E-01	5.76E-01	3.44E-01	0.31	0.20	0.33	50
Pm CAHS 2-like	249348006 7	9.43E+03	4.85E+04	6.87E+04	1.40E+04	8.61E+04	1.52E+05	6.73E-01	5.63E-01	4.52E-01	0.47	0.19	0.16	0
Pm CAHS 94205-like (3)	249347971 8	1.79E+04	4.85E+04	1.17E+05	2.38E+04	8.61E+04	2.71E+05	7.52E-01	5.63E-01	4.32E-01	0.36	0.26	0.37	46
Pm CAHS 107838-like (1)	249346947 1	1.92E+04	5.10E+04	1.32E+05	2.62E+04	8.20E+04	2.93E+05	7.34E-01	6.22E-01	4.52E-01	0.36	0.42	0.37	53
Pm CAHS 107838-like (2)	249346885 2	1.97E+04	5.23E+04	1.31E+05	2.68E+04	8.43E+04	3.08E+05	7.36E-01	6.20E-01	4.27E-01	0.37	0.27	0.39	60
Pm CAHS 86272-like	249344538 3	2.25E+04	5.76E+04	1.29E+05	2.96E+04	9.51E+04	3.46E+05	7.61E-01	6.05E-01	3.74E-01	0.34	0.37	0.36	49
Pm CAHS 89226-like	249343895 1	2.12E+04	4.09E+04	1.01E+05	2.63E+04	8.89E+04	2.92E+05	8.06E-01	4.60E-01	3.46E-01	0.17	0.16	0.21	47
		Surface Area (angstrom^2)			Volume (angstrom^3)			Surface Area/Volume (angstrom^-1)			Confidence Score (pTM)			Tail Length
Name	GenBank ID	Monomer	Trimer	Decamer	Monomer	Trimer	Decamer	Monomer	Trimer	Decamer	1	3	10	(aa)
Rv SAHS1	40278375 4	8.26E+03	2.49E+04	7.62E+04	2.11E+04	6.33E+04	2.13E+05	3.92E-01	3.93E-01	3.58E-01	0.81	0.35	0.24	45
Rv SAHS2	40278375 6	9.79E+03	2.42E+04	8.95E+04	2.19E+04	6.71E+04	2.22E+05	4.47E-01	3.61E-01	4.03E-01	0.80	0.33	0.23	7
Rv SAHS3	110139446 1	8.03E+03	2.34E+04	7.72E+04	1.80E+04	5.41E+04	1.81E+05	4.46E-01	4.32E-01	4.26E-01	0.82	0.37	0.24	62

Rv SAHS4	110139537 1	9.70E+03	2.02E+04	6.73E+04	2.10E+04	6.55E+04	2.22E+05	4.62E-01	3.08E-01	3.03E-01	0.77	0.37	0.24	65
Rv SAHS6	110139543 5	9.24E+03	2.25E+04	7.57E+04	2.13E+04	6.58E+04	2.16E+05	4.34E-01	3.42E-01	3.50E-01	0.79	0.34	0.24	65
Rv SAHS7	110139543 8	7.58E+03	2.24E+04	7.58E+04	2.01E+04	6.12E+04	2.05E+05	3.77E-01	3.66E-01	3.71E-01	0.84	0.38	0.29	56
Rv SAHS8	110139544 5	9.38E+03	2.49E+04	8.83E+04	2.16E+04	6.66E+04	2.18E+05	4.34E-01	3.73E-01	4.06E-01	0.81	0.34	0.23	49
Rv SAHS9	110139544 7	6.99E+03	1.74E+04	6.08E+04	1.38E+04	4.32E+04	N/A	5.06E-01	4.03E-01	N/A	0.79	0.33	0.21	7
Rv SAHS10	110139545 2	7.54E+03	1.99E+04	3.94E+04	1.51E+04	4.70E+04	N/A	5.00E-01	4.23E-01	N/A	0.75	0.28	0.20	16
Rv SAHS11	110139545 3	9.54E+03	2.72E+04	8.53E+04	2.15E+04	6.55E+04	2.21E+05	4.45E-01	4.15E-01	3.85E-01	0.79	0.37	0.24	65
Rv SAHS12	110139566 6	6.55E+03	1.64E+04	6.27E+04	1.37E+04	4.35E+04	1.39E+05	4.79E-01	3.77E-01	4.50E-01	0.80	0.29	0.23	57
He SAHS 63681	121027478 6	8.60E+03	2.41E+04	7.81E+04	2.15E+04	6.57E+04	2.18E+05	4.01E-01	3.67E-01	3.58E-01	0.81	0.38	0.25	64
He SAHS 53582	121027478 4	8.97E+03	2.34E+04	8.25E+04	2.04E+04	6.29E+04	2.07E+05	4.39E-01	3.71E-01	3.98E-01	0.84	0.36	0.23	66
He SAHS 33020	121027478 2	9.09E+03	2.41E+04	8.38E+04	2.10E+04	6.47E+04	2.12E+05	4.32E-01	3.73E-01	3.96E-01	0.82	0.36	0.26	62
He SAHS 68234	121027478 7	9.92E+03	2.39E+04	7.64E+04	2.14E+04	6.67E+04	2.21E+05	4.63E-01	3.58E-01	3.46E-01	0.79	0.35	0.24	66
Pm SAHS 1-like (1)	249348241 1	8.27E+03	1.98E+04	7.71E+04	1.81E+04	5.66E+04	1.83E+05	4.56E-01	3.50E-01	4.22E-01	0.80	0.42	0.23	67
Pm SAHS 1-like (2)	249345697 0	9.96E+03	2.72E+04	8.44E+04	2.20E+04	6.70E+04	2.23E+05	4.53E-01	4.06E-01	3.79E-01	0.76	0.34	0.23	68

Pm SAHS 1-like (3)	249342097 4	9.70E+03	2.70E+04	8.56E+04	2.11E+04	6.42E+04	N/A	4.59E-01	4.20E-01	N/A	0.75	0.32	0.24	66.00
Pm SAHS 1-like X2	249339847 0	1.08E+04	3.16E+04	9.58E+04	2.28E+04	6.95E+04	2.33E+05	4.74E-01	4.55E-01	4.11E-01	0.71	0.32	0.23	83.00
Pm SAHS 1-like X1	249339847 6	1.09E+04	3.08E+04	9.72E+04	2.28E+04	6.92E+04	2.35E+05	4.78E-01	4.45E-01	4.13E-01	0.71	0.32	0.23	84.00
Pm SAHS 1-like (4)	249339503 8	1.23E+04	2.64E+04	1.06E+05	2.46E+04	7.94E+04	2.51E+05	5.01E-01	3.33E-01	4.23E-01	0.68	0.29	0.23	87.00
Pm SAHS 1-like (5)	249339292 6	8.42E+03	2.51E+04	7.40E+04	2.18E+04	6.53E+04	2.20E+05	3.87E-01	3.84E-01	3.36E-01	0.77	0.36	0.27	61
Pm SAHS 1-like (6)	249339285 3	9.70E+03	2.50E+04	9.38E+04	2.24E+04	6.79E+04	2.28E+05	4.33E-01	3.68E-01	4.12E-01	0.75	0.33	0.24	67