Proteins: Structure, Function, and Bioinformatics

WILEY

PROTEINS

SHORT COMMUNICATION

Large Extent of Convergent Evolution Towards the Double Histone Fold Revealed by Targeted Sequence and Structure Search Approach

Toshiko Miyake[1] | Anna Ranaudo[2] | Elena Sacco[3] | Claudio Greco[2]

[1]Department of Pharmaceutical and Pharmacological Sciences, University of Padova, Padova, Italy | [2]Department of Earth and Environmental Sciences, University of Milan-Bicocca, Milan, Italy | [3]Department of Biotechnologies and Biosciences, University of Milan-Bicocca, Milan, Italy

Correspondence: Claudio Greco (claudio.greco@unimib.it)

Received: 31 December 2024 | **Revised:** 9 August 2025 | **Accepted:** 27 October 2025

Keywords: DNA binding proteins | eukaryotes | globular protein folding | histone | molecular evolution

ABSTRACT

Histone proteins are key players in chromatin packaging. In eukaryotes, nucleosomal cores—the DNA packaging fundamental units—are formed by composition of histone dimers. The double histone fold is a protein structure where two consecutive regions, each featuring histone fold, come together to create a histone pseudodimer. Although regarded as an uncommon fold to date, in this study we show—by protein structure and sequence analyses—that the double histone fold is widespread in eukaryotes. Perspectives of such outcome are discussed in terms of novel directions that our results may open in diverse areas, from epigenetics to the design of DNA-binding proteins.

1 | Introduction

Histones are essential proteins that play a central role in the organization and regulation of DNA within eukaryotic cells. They serve as the primary structural components of chromatin, the complex of DNA and proteins that compacts genetic material to fit within the confines of the nucleus while maintaining accessibility for vital cellular processes. Histones achieve this by forming nucleosomes, the fundamental repeating units of chromatin, in which DNA is wrapped around an octamer composed of core histone proteins [1, 2]. The histone family includes five main types: H1 (a linker histone), H2A, H2B, H3, and H4. Core histones (H2A and H2B, H3 and H4) pair to form heterodimers, which subsequently assemble into an octamer, providing a scaffold around which DNA coils. The linker histone H1 stabilizes the structure by binding to the DNA between nucleosomes, further compacting the chromatin. Beyond their structural role, histones are dynamic regulators of gene expression and genome stability [3, 4]. They undergo a variety of post-translational modifications, such as acetylation, methylation, phosphorylation, and ubiquitination, which influence chromatin architecture and modulate interactions with transcription factors and other chromatin-associated proteins. These modifications constitute the histone code, an additional layer of epigenetic regulation that fine-tunes gene activity in response to developmental and environmental cues [5]. Histones are conserved across eukaryotic species, underscoring their fundamental importance to cellular life. Research into histone structure and function continues to uncover their intricate involvement in processes like DNA replication, repair, and transcriptional regulation, as well as their roles in health and disease [4].

Some histone homologs have developed a remarkable strategy to eliminate the need for dimerizing with their counterparts by utilizing a double histone fold. This structure consists of two histone folds within a single peptide chain, enabling it to effectively "dimerize" on its own [6–8]. The first examples of this fold described in literature regard the archaeal histone from *Methanopyrus kandleri* [9] and, quite strikingly, the N-terminal domain of a non-nuclear multidomain protein such as the Ras

activator Son of sevenless (Sos) protein [10]. As illustrated in previous studies by us and other authors, the structural resemblance between the two-protein H2A/H2B dimer and the single-protein human Son of sevenless 1 (hSos1) is very high, despite very low amino acid sequence identity [7, 10]. The efficiency of the double histone fold, which operates independently of a partner protein, likely explains its discovery in a virus, that is, the *Heliothis zea* virus 1. In this context, it is thought to play a role in organizing and packaging DNA within the capsid [6, 8, 11]; in fact, we based such functional assignment on targeted computational predictions, and subsequently this picture was confirmed in the case of other virus classes [12]. Our group also identified the presence of a double histone fold in the mammalian multidomain protein Cca3, and in the related "Similar to Cca3" (Cca3S) proteins [7]. The available data on the function of Cca3 proteins correlate with their preferential binding to DNA in a pre-replicative condensed G0/G1 state, which indicates that they are not involved in determining constitutive chromatin structure.

It is indeed interesting that, up to present times, there is no evidence of double histone fold proteins that have a structural key role in organizing eukaryotic chromatin; in general, the double histone fold appears to be rare; however renewed research can provide useful information on the function of such fold and, possibly, on chromatin biochemistry. The present investigation intends to deepen insights in such context, offering a perspective that gravitates around our novel finding: that the double histone fold is much more widespread than previously thought.

2 | Results and Discussion

Prompted by the considerations illustrated in Section, we carried out targeted structure and sequence searches on protein databases, extending the scope of efforts we have carried out in the past in such context [6, 7].

2.1 | Structure-Based Search

With the aim of finding proteins featuring the double histone fold, we first queried the Dali server [13]. The choice of using Dali, a well-established program in structural bioinformatics for the search for structural similarities in proteins, lies in its ability to identify remote homologs, which may include proteins that underwent conformational changes during evolution. This is possible thanks to its underlying algorithm, based on a distance-matrix alignment, which allows for a flexible superimposition of protein structures [14]. The Dali server was queried with the N-terminal domain of the Sos protein (Protein Data Bank ID: 1Q9C, chain D) and the search was done across the PDB25, a representative subset of Protein Data Bank. Among the structures found, we analyzed the ones reported in Table 1, selected on the basis of criteria detailed in Section 4.

The PDB ID 1WWI corresponds to the crystal structure of the protein ttk003001566 which is not discussed in literature and poorly annotated in UniProt (https://www.uniprot.org/uniprotkb/Q5SI95/entry). This protein was found in the organism *Thermus thermophilus* and is currently still classified in the DUF1931 (Domains of unknown function) family [16]. By visually inspecting the superimposition of protein ttk003001566 onto the N-terminal domain of Sos, it is evident that the former assumes the double histone fold (Figure 1A). This protein belongs to the group of the recently identified bacterial histones [17] that only includes proteins with a single histone, with the sole exceptions of ttk003001566 itself and the protein AQ328 from *Aquifex aeolicus* (PDB ID: 1R4V), the latter being a double histone fold protein previously described by Qiu et al. [18]. This result we obtained with Dali reinforces previous more qualitative findings [12, 17], validating the use of computational methods to identify new members of the bacterial histone family; specifically, they suggest that the double histone fold is a conserved structural motif present in a broader range of bacterial proteins than previously recognized.

Another hit identified by Dali corresponds to the transcription initiation protein Spt3 homolog, the chain I in the experimental structure of the SAGA coactivator complex (PDB ID: 7KTR). The visual inspection of this chain clearly shows a double histone fold (Figure 1B). Interestingly, in the discussion of such structure [15], the Spt3 protein is presented as having a dimeric structure of histone folds; however, this may give rise to some ambiguity, as in literature it has been proposed that Spt3 proteins could actually dimerize *stricto sensu* [19]. In such context, the Dali's results here presented give us occasion to underline that in all known 3D structures encompassing the Spt3 protein [15, 20], the latter is present in the form of a histone pseudodimer, that is, it features a double histone fold.

2.2 | Sequence-Based Search

In the second part of the work, we performed sequence-based search by querying combinations of known dimerizing histone sequences as "chimeric" entries. As explained in the following, the results obtained evidence the presence of novel eukaryotic protein sequences featuring tandem of histones showing high degree of similarity with the query histone sequence, an outcome that has far reaching implications.

As first sequence search carried out by using protein BLAST webserver in default mode—that is, UniprotKB reference proteomes and SwissProt being the source databases [21, 22]—we

TABLE 1 | Structures retrieved from querying the Dali server (PDB25 search) with PDB ID 1Q9C (chain D) and analyzed in the present work.

PDB ID	Chain	*Z* score	rmsd (Å)	Lali	nres	%id	Reference
1WWI	A	9.8	3.2	128	148	13	No reference in literature
7KTR	I	8.9	3.1	137	243	15	Herbst et al. [15]

Abbreviations: %id = sequence identity, Lali = length of alignment, nres = number of residues.

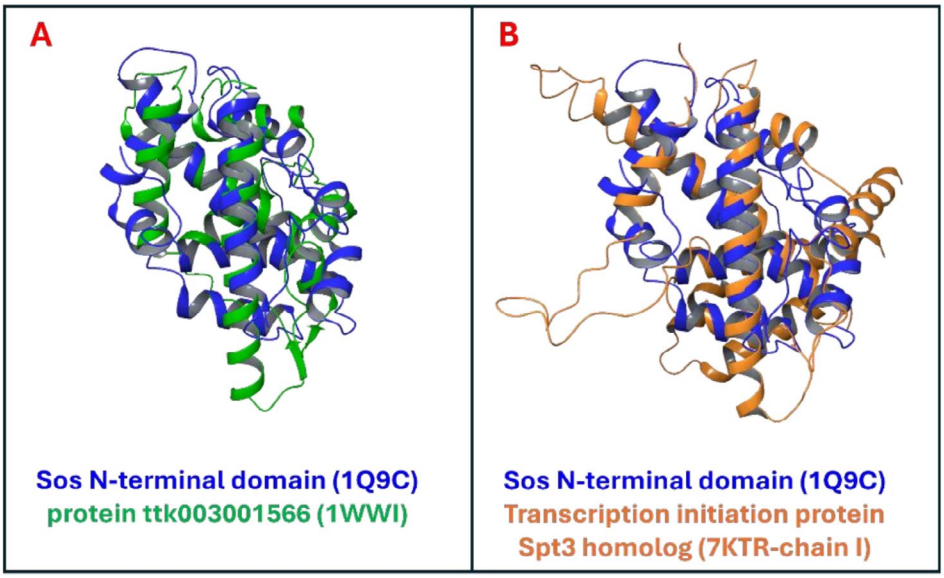

FIGURE 1 | (A) Superimposition of the protein ttk003001566 (in green, PDB ID: 1WWI, chain A) onto the N-terminal domain of Sos (in blue, PDB ID: 1Q9C, chain D). (B) Superimposition of the transcription initiation protein Spt3 homolog (in orange, PDB ID: 7KTR, chain I) onto the N-terminal domain of Sos (in blue, PDB ID: 1Q9C, chain D).

employed a "chimeric" query by joining together the full sequences of human histone H2B and H2A, in such relative order as reported in Data S1. The outcome of this search brings to light the existence of a group of proteins similar to each other in terms of length (ca. 230 aa) and expressed in fishes, which feature a tandem of sequences homologous to histones H2A and H2B, with overall sequence identity above 70% spanning both of the latter (see Table 2 and Data S1). Such sequence identity with standard histones is much higher than the corresponding degree of identity found in the case of Sos and Cca3 proteins, which was 30% at most [7]; this feature of the newly identified group of proteins, together with their single domain nature and annotation—issues more thoroughly discussed below—and their widespread conservation among fish species, may indicate that they have conserved the functional role of histones along evolution. Moreover, the linker region between the two histone folds in all such proteins appears to be long enough to allow the formation of a double histone fold, and this is actually confirmed by AlphaFold-based modeling [23–25]. In fact, Figure 2 presents the AlphaFold 3D models of two representative sequences, that is, the ones from *Lates calcarifer* and *Sphaeramia orbicularis*; it can be seen that the packing of the two sequences against each other to form a pseudodimer is predicted within a domain that is associated with high confidence of structure prediction. AlphaFold outcomes show that these proteins are composed of a single domain featuring a double histone fold structure, in full analogy with the abovementioned case of the *Heliotis zea* virus 1 double histone fold that has been functionally associated to DNA packaging [6, 8, 11]. Intriguingly, in the same BLAST search output we found tentative evidences of the presence of a double histone fold with similar features also in other eukaryotes, namely in algae: in fact, the JKP88DRAFT_157074 EMBL/GenBank/DDBJ whole genome shotgun ORF entry—part of preliminary genomic data—of *Tribonema minus* would correspond to a 232-aa protein having 61.3% sequence identity with the H2A-H2B "chimeric" query sequence. Also in this case, AlphaFold predicts a double histone fold with high confidence, within a single-domain protein (Figure 2).

The second sequence search by means of BLAST carried out in the present study was made by employing a "chimeric" input sequence again obtained by joining together the amino acid sequences of human histones H2B and H2A, but at this time the sequence of the latter preceded the one of the former in the query (see Data S1). Interestingly, also in this case several sequences were found, encompassing a tandem of two histone sequences on the same polypeptide chain (see Table 3 that reports a selection of BLAST entries). Several peculiar aspects in the novel sequences retrieved can be highlighted; first of all, these proteins are expressed by a much larger variety of organisms: not only fishes, but also insects, crustaceans and mammals. As far as their primary sequences are concerned, they tend to be slightly longer as compared to proteins shown in Table 2, which—in the hypothesis that they would form double histone folds—is somehow expected. In fact, in pseudodimers in which the H2B region precedes the H2A one in the primary sequence, the terminal portion at the end of the former is expected to be spatially close to the initiation of the sequence of the latter, due to the typical geometrical features of H2A-H2B histone heterodimers; therefore, in a pseudodimer the two histone folds in the same sequence can be joined together by relatively short linkers in such cases, which correspond to models shown in Figure 2. Conversely, when the H2B homology region follows that of H2A in the primary sequence, a longer linker between the two is needed in order for them to pack against each other, as the C-terminus of the H2A homology region does not directly face the N-terminus of the H2B homology region. AlphaFold modeling of one of the homologs listed in Table 3, that is, the one of the arthropod *Varroa destructor* is shown in Figure 3. It can be seen that a double histone fold is predicted, with high confidence; notice that, in Figure 3, the linker between the two histone folds is highlighted, for the sake of clarity.

TABLE 2 | List of selected H2B-H2A-like double histone folds retrieved from BLAST searches.

Protein ID	Organism	Protein length	Sequence identity to human H2B-H2A "chimeric" sequence (%)
A0AA88QBL4_9TELE[a]	*Cirrhinus molitorella*	248	83.7
A0A8J7NWU4_ATRSP[a]	*Atractosteus spatula*	231	77.3
A0A667Z0M6_9TELE	*Myripristis murdjan*	232	74.1
A0A673AMV8_9TELE	*Sphaeramia orbicularis*	236	74.1
A0A672LVH8_SINGR	*Sinocyclocheilus grahami*	224	73.3
A0A672FYA7_SALFA	*Salarias fasciatus*	239	73.7
A0A3P9AVA4_9CICH	*Maylandia zebra*	247	73.7
A0A8C1MY98_CYPCA	*Cyprinus carpio*	227	72.9
A0A3Q3WY72_MOLML	*Mola mola*	227	73.4
I3K8S3_ORENI	*Oreochromis niloticus*	249	73.7
A0A4W6D208_LATCA	*Lates calcarifer*	242	73.3
A0A6G0I8L6_LARCR[a]	*Larimichthys crocea*	218	73.9
A0A3B5PQY2_XIPMA	*Xiphophorus maculatus*	245	73.3
A0A3Q3MRF1_9TELE	*Mastacembelus armatus*	229	72.9
A0A8C5GD79_GOUWI	*Gouania willdenowi*	252	74.9
A0A6Q2XGM9_ESOLU	*Esox lucius*	228	72.9
A0A674MJJ5_TAKRU	*Takifugu rubripes*	224	74.3
A0A673IN89_9TELE	*Sinocyclocheilus rhinocerous*	226	72.1
A0A667Z4X5_9TELE	*Myripristis murdjan*	247	72.9
A0A674DK55_SALTR	*Salmo trutta*	244	72.9
A0A671TSD9_SPAAU	*Sparus aurata*	245	73.5
A0A665T2Y3_ECHNA	*Echeneis naucrates*	245	73.7
A0A6G1QSB5_9TELE	*Channa argus*	219	73.3
A0A3Q1J4X1_ANATE	*Anabas testudineus*	240	74.5
A0A8C4B4E9_9TELE	*Denticeps clupeoides*	225	73.5
A0A8C9ZXR5_SANLU	*Sander lucioperca*	264	72.3
A0A3P9DGC1_9CICH	*Maylandia zebra*	244	76.4
A0A671L789_9TELE	*Sinocyclocheilus anshuiensis*	234	74.5
A0A8C9XJV8_SANLU	*Sander lucioperca*	233	72.3
A0A8C2X1W5_CYCLU	*Cyclopterus lumpus*	234	73.0
A0A8C6TPA6_9GOBI	*Neogobius melanostomus*	239	73.1
A0A9D3SW52_MEGAT[a]	*Megalops atlanticus*	216	73.1
A0A2U9B6T4_SCOMX	*Scophthalmus maximus*	217	73.5
A0A8P4FZ80_DICLA	*Dicentrarchus labrax*	239	72.9
A0A3Q3NHI0_9LABR	*Labrus bergylta*	235	73.5
A0A9Q0E4A4_9TELE[a]	*Muraenolepis orangiensis*	219	72.0

[a]Sequence derived from an EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is preliminary data.

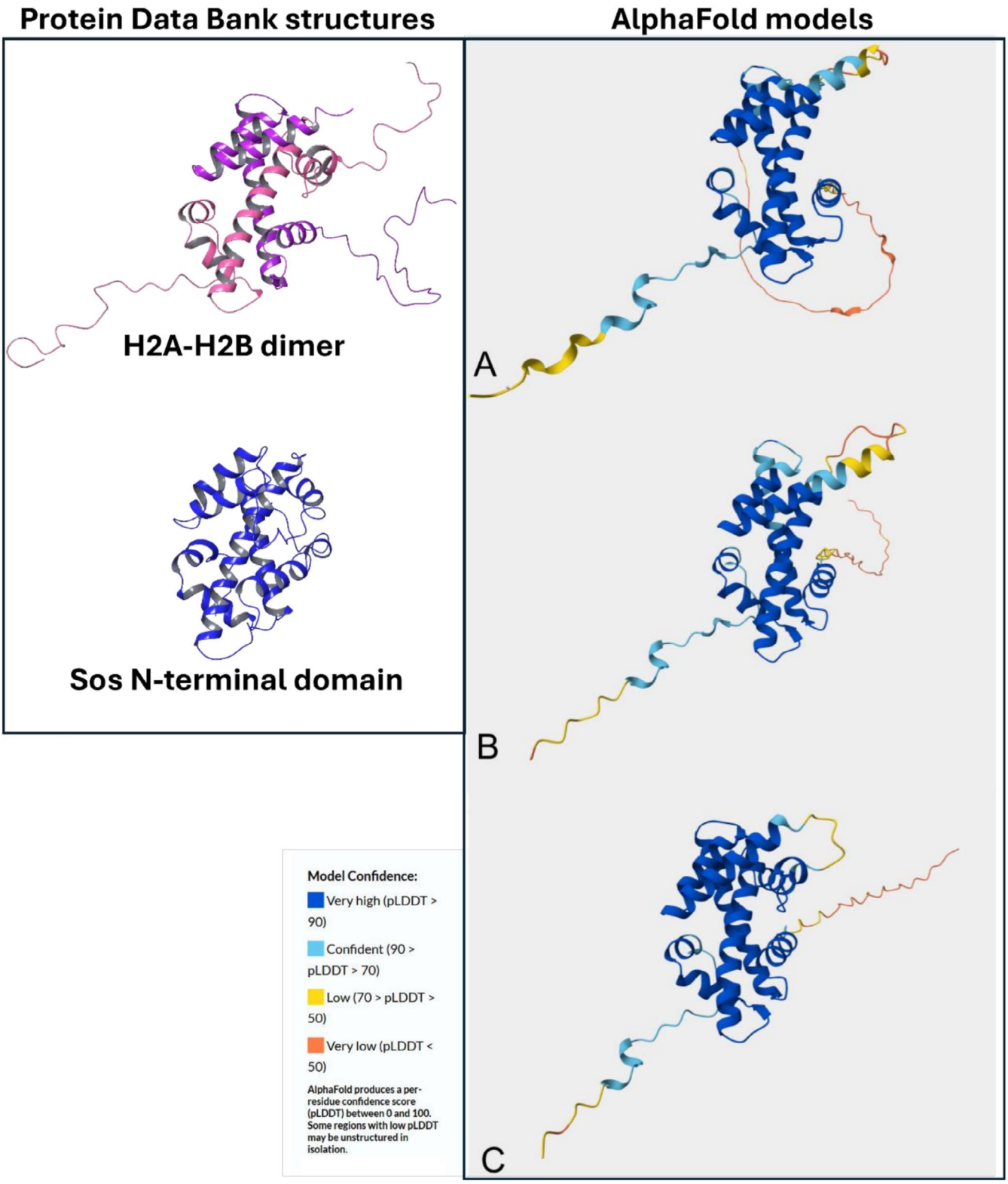

FIGURE 2 | Left: Protein Data Bank structures of the human histone H2A-H2B (PDB ID: 1KX5, chains C [H2A, pink] and D [H2B, purple]) and of the N-terminal domain of the Sos protein (PDB ID: 1Q9C, chain D). Right: AlphaFold models for the H2B-H2A double histone folds from: *Lates calcarifer* (A); *Sphaeramia orbicularis* (B); *Tribonema minus* (C). Color legend is shown within the figure (mid-left); notice that, in the last model, the linker region between the two histone folds is easily recognizable in the up-right part of the model, as a low confidence (yellow) region.

Finally, we also performed targeted BLAST searches using as a query a "chimeric" protein sequence formed by the amino acid sequences of human histones H3 and H4 (query sequence and raw BLAST output in Data S1). In this case we also found several > 200 aa long entries featuring tandem histone folds on the same sequence, which were predicted to correspond to single domain double histone fold proteins using AlphaFold-based modeling (sequence identity shared with query between 97% and 65%;

TABLE 3 | List of selected H2A-H2B-like double histone folds retrieved from BLAST searches.

Protein ID	Organism	Protein length	Sequence identity with human H2A-H2B "chimeric" sequence (%)
A0A6P6ITF9_PUMCO	*Puma concolor*	244	86.5
A0A1S3A9T1_ERIEU	*Erinaceus europaeus*	244	85.7
A0AA88M0B0_CHASR[a]	*Channa striata*	247	85.3
A0A7M7J940_VARDE	*Varroa destructor*	245	84.6
A0AAD4NF77_9BILA[a]	*Ditylenchus destructor*	276	80.5
A0A7R9BUF3_9CRUS	*Notodromas monacha*	282	73.9
A0A1D2N8U2_ORCCI[a]	*Orchesella cincta*	280	66.2
A0A9L0IRM1_EQUAS	*Equus asinus*	232	64.5

[a]Sequence derived from an EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is preliminary data.

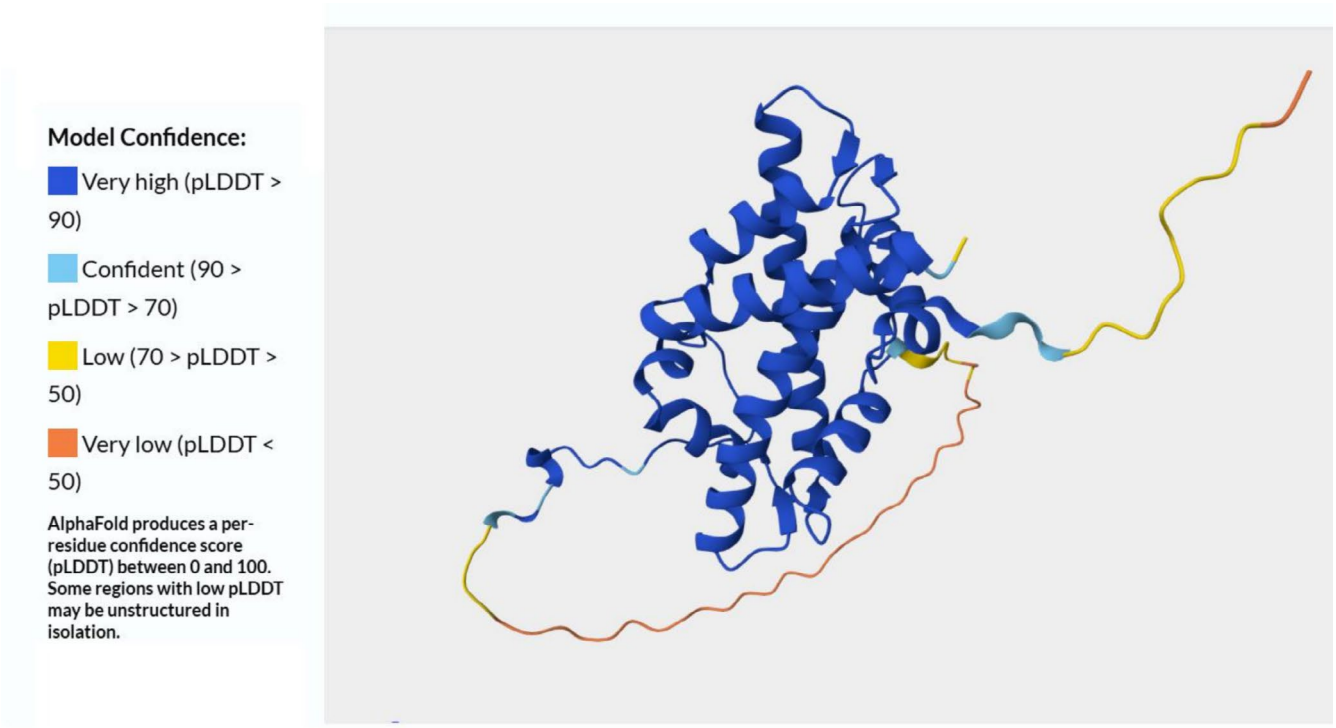

FIGURE 3 | AlphaFold model for the H2A-H2B double histone folds from *Varroa destructor*. Color legend is shown within the figure (left); notice that the linker region between the two histone folds is easily recognizable in the lower part of the model, as a very low confidence (orange) region.

AlphaFold results not shown). However, part of these proteins was found to be annotated as TATA box binding protein associated factor (TAF) histone-like fold domain-containing proteins [26]. Usually, such class of proteins feature a histone fold heterodimer, whereas we have pseudodimers in the cases found in the present study. Such annotation stems from automatic InterPro procedure based on the SMART approach [27, 28]; notably, the same automatic annotation procedure, when successful, always indicated histone or histone-like annotation in the case of the novel H2A/H2B pseudodimers described above. Since these are likely not proteins involved in structuring chromatin, we decided not to investigate further the H3/H4 pseudodimer case; still, the latter are novel as well as interesting cases from a structural point of view, and we are planning to study them in more detail in the future.

3 | Conclusions

To summarize and further dissect the implications of our results, in this study we demonstrated that—although previously overlooked—single domain proteins featuring the double histone fold and sharing high sequence identity with histones are rather common in eukaryotes. This observation is particularly intriguing; in fact, it is tempting to hypothesize that these novel double histone fold proteins could be involved in chromatin structural

organization, much like their counterparts in some archaea [9] and in certain viral particles [12]. This opens fascinating possibilities, providing a potential new layer of complexity to the histone code. For instance, the linker region between the two histone folds could undergo covalent modifications, effectively replacing the N- and C-termini of conventional histones, which are well-known targets for epigenetic modulation [29]. Our study evidenced that, far from being a rare occurrence in the life domain, the presence of histone pseudodimers is a shared feature of very diverse kinds of eukaryotic organisms, the only exception being—as far as we could see—plants. Deepening insights into the biological role of the linkage between histone folds to give place to double histone folds appears to be a promising research line, which may provide fresh perspectives in key areas, from enhancing our understanding of chromatin biochemistry to designing DNA-binding proteins.

4 | Methods

The Dali server (http://ekhidna2.biocenter.helsinki.fi/dali/; [13]) was used to search for proteins featuring the double histone fold. As a query, the N-terminal, histone domain of Son of Sevenless protein was used (Protein Data Bank ID: 1Q9C, chain D) and the search was done across the PDB25, a representative subset of Protein Data Bank. Dali's search output consists in a list of the protein structures labeled with their PDB ID and chain, and sorted by descending *Z*-score, which quantifies the significance of structural similarity. Moreover, the server also provides the protein structures superimposed to the query in .pdb format. After the exclusion of the many histones that were found from the search and of fusion proteins as well, the remaining structures were filtered by the number of residues (nres in Dali's output) and length of alignment (lali). Considering that the average length of core histones is around 100–130 aa [30], we limited our analysis to entries with nres ≥ 200, that is, at least two times the lower limit, and lali ≥ 100, as shorter sequence stretches do not allow to distinguish between similarities on a single histone fold or on two consecutive folds. We also considered only proteins with *Z*-score > 6. These criteria resulted in the detection of two proteins: PDB IDs 1WWI (chain A) and 7KTR (chain I), which we discussed in Section 2.1.

Sequence searches were carried out by using protein BLAST (i.e., blastp program) with UniprotKB reference proteomes and SwissProt being the source databases [21, 22] and with the following parameters: E-threshold = 10, Matrix = Auto-BLOSUM 62, gapped alignment with no filters applied on low-complexity regions. As query sequences for blastp searches, we employed "chimeric" histone sequences by joining together the full sequences of human histones, selected and reciprocally disposed as detailed in the Results and Discussion section and in Data S1.

3D models of selected proteins here shown are AlphaFold-based structures [23]; more specifically, structures were retrieved from the AlphaFold Protein Structure Database [24, 25], as queried by the blastp web interface for each of the entries found based on protein sequence searches.

Author Contributions

Toshiko Miyake: investigation, writing – original draft, methodology. **Anna Ranaudo:** writing – review and editing, investigation, methodology. **Elena Sacco:** methodology. **Claudio Greco:** conceptualization, investigation, writing – review and editing.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All data are included either in the main text, or in the Supporting Information.

Peer Review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/prot.70081.

References

1. K. Luger, A. W. Mader, R. K. Richmond, et al., "Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution," *Nature* 389 (1997): 251–260.

2. A. Ruiz-Carrillo, J. L. Jorcano, G. Eder, and R. Lurz, "In Vitro Core Particle and Nucleosome Assembly at Physiological Ionic Strength," *Proceedings. National Academy of Sciences. United States of America* 76 (1979): 3284–3288.

3. J. Pfluger and D. Wagner, "Histone Modifications and Dynamic Regulation of Genome Accessibility in Plants," *Current Opinion in Plant Biology* 10 (2007): 645–652.

4. W. Yao, X. Wu, and X. Wang, "Crossing Epigenetic Frontiers: The Intersection of Novel Histone Modifications and Diseases," *Signal Transduction and Targeted Therapy* 9 (2024): 232.

5. K. A. Janssen, S. Sidoli, and B. A. Garcia, "Recent Achievements in Characterizing the Histone Code and Approaches to Integrating Epigenomics and Systems Biology," *Methods in Enzymology* 586 (2017): 359–378.

6. C. Greco, E. Sacco, M. Vanoni, and L. de Gioia, "Identification and In Silico Analysis of a New Group of Double-Histone Fold-Containing Proteins," *Journal of Molecular Modeling* 12 (2005): 76–84.

7. C. Greco, P. Fantucci, and L. De Gioia, "In Silico Functional Characterization of a Double Histone Fold Domain From the Heliothis Zea Virus 1," *BMC Bioinformatics* 6, no. Suppl 4 (2005): S15.

8. K. Sousounis, C. E. Haney, J. Cao, et al., "Conservation of the Three-Dimensional Structure in Non-Homologous or Unrelated Proteins," *Human Genomics* 6 (2012): 10.

9. R. L. Fahrner, D. Cascio, J. A. Lake, and A. Slesarev, "An Ancestral Nuclear Protein Assembly: Crystal Structure of the Methanopyrus Kandlerihistone," *Protein Science* 10 (2001): 2002–2007.

10. H. Sondermann, S. M. Soisson, D. Bar-Sagi, and J. Kuriyan, "Tandem Histone Folds in the Structure of the N-Terminal Segment of the Ras Activator Son of Sevenless," *Structure* 11 (2003): 1583–1593.

11. G. Moro, L. Bonati, M. Bruschi, et al., "Computational Approaches to Shed Light on Molecular Mechanisms in Biological Processes," *Theoretical Chemistry Accounts* 117 (2007): 723–741.

12. V. Thomas, C. Bertelli, F. Collyn, et al., "Lausannevirus, a Giant Amoebal Virus Encoding Histone Doublets," *Environmental Microbiology* 13 (2011): 1454–1466.

13. L. Holm, "Dali Server: Structural Unification of Protein Families," *Nucleic Acids Research* 50 (2022): W210–W215.

14. L. Holm, "DALI and the Persistence of Protein Shape," *Protein Science* 29 (2020): 128–140.

15. D. A. Herbst, M. N. Esbin, R. K. Louder, et al., "Structure of the Human SAGA Coactivator Complex," *Nature Structural & Molecular Biology* 28 (2021): 989–996.

16. A. Bateman, P. Coggill, and R. D. Finn, "DUFs: Families in Search of Function," *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications* 66, no. 10 (2010): 1148–1152.

17. V. Alva and A. N. Lupas, "Histones Predate the Split Between Bacteria and Archaea," *Bioinformatics* 35, no. 14 (2019): 2349–2353.

18. Y. Qiu, V. Tereshko, Y. Kim, et al., "The Crystal Structure of Aq_328 From the Hyperthermophilic Bacteria Aquifex Aeolicus Shows an Ancestral Histone Fold," *Proteins* 62 (2006): 8–16.

19. C. Birck, O. Poch, C. Romier, et al., "Human TAFII28 and TAFII18 Interact Through a Histone Fold Encoded by Atypical Evolutionary Conserved Motifs Also Found in the SPT3 Family," *Cell* 94, no. 2 (1998): 239–249.

20. H. Wang, C. Dienemann, A. Stützer, H. Urlaub, A. C. M. Cheung, and P. Cramer, "Structure of the Transcription Coactivator SAGA," *Nature* 577 (2020): 717–720.

21. S. F. Altschul, T. L. Madden, A. A. Schaffer, et al., "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs," *Nucleic Acids Research* 25 (1997): 3389–3402.

22. C. O'Donovan and R. Apweiler, "A Guide to UniProt for Protein Scientists," *Methods in Molecular Biology* 694 (2011): 25–35.

23. J. Jumper, L. Evans, A. Pritzel, et al., "Highly Accurate Protein Structure Prediction with AlphaFold," *Nature* 596 (2021): 583–589.

24. M. Varadi, S. Anyango, M. Deshpande, et al., "AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space With High-Accuracy Models," *Nucleic Acids Research* 50 (2022): D439–D444.

25. M. Varadi, D. Bertoni, P. Magana, et al., "AlphaFold Protein Structure Database in 2024: Providing Structure Coverage for Over 214 Million Protein Sequences," *Nucleic Acids Research* 52 (2024): D368–D375.

26. M. Bando, S. Ijuin, S. Hasegawa, and M. Horikoshi, "The Involvement of the Histone Fold Motifs in the Mutual Interaction Between Human TAF1180 and TAF11221," *Journal of Biochemistry* 121 (1997): 591–597.

27. I. Letunic and P. Bork, "20 Years of the SMART Protein Domain Annotation Resource," *Nucleic Acids Research* 46 (2018): D493–D496.

28. I. Letunic, S. Khedkar, and P. Bork, "SMART: Recent Updates, New Developments and Status in 2020," *Nucleic Acids Research* 49 (2021): D458–D460.

29. G. Millán-Zambrano, A. Burton, A. J. Bannister, and R. Schneider, "Histone Post-Translational Modifications—Cause and Consequence of Genome Function," *Nature Reviews. Genetics* 23 (2022): 563–580.

30. A. R. Cutter and J. J. Hayes, "A Brief Review of Nucleosome Structure," *FEBS Letters* 589 (2015): 2914–2922.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** prot70081-sup-0001-Supinfo.docx.