

Research article

Open Access

## On the origin of the histone fold

Vikram Alva, Moritz Ammelburg, Johannes Söding and Andrei N Lupas\*

Address: Department of Protein Evolution, Max-Planck-Institute for Developmental Biology, Spemannstr. 35, D-72076 Tübingen, Germany

Email: Vikram Alva - vikram.alva@tuebingen.mpg.de; Moritz Ammelburg - moritz.ammelburg@tuebingen.mpg.de;

Johannes Söding - johannes.soeding@tuebingen.mpg.de; Andrei N Lupas\* - andrei.lupas@tuebingen.mpg.de

\* Corresponding author

Published: 28 March 2007

Received: 3 November 2006

BMC Structural Biology 2007, 7:17 doi:10.1186/1472-6807-7-17

Accepted: 28 March 2007

This article is available from: <http://www.biomedcentral.com/1472-6807/7/17>

© 2007 Alva et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Histones organize the genomic DNA of eukaryotes into chromatin. The four core histone subunits consist of two consecutive helix-strand-helix motifs and are interleaved into heterodimers with a unique fold. We have searched for the evolutionary origin of this fold using sequence and structure comparisons, based on the hypothesis that folded proteins evolved by combination of an ancestral set of peptides, the antecedent domain segments.

**Results:** Our results suggest that an antecedent domain segment, corresponding to one helix-strand-helix motif, gave rise divergently to the N-terminal substrate recognition domain of Clp/Hsp100 proteins and to the helical part of the extended ATPase domain found in AAA+ proteins. The histone fold arose subsequently from the latter through a 3D domain-swapping event. To our knowledge, this is the first example of a genetically fixed 3D domain swap that led to the emergence of a protein family with novel properties, establishing domain swapping as a mechanism for protein evolution.

**Conclusion:** The helix-strand-helix motif common to these three folds provides support for our theory of an 'ancient peptide world' by demonstrating how an ancestral fragment can give rise to 3 different folds.

### Background

The organization of DNA into chromatin allows its compact and reversible packaging into the nucleus of a eukaryotic cell. The basic structural unit of chromatin is the nucleosome [1], which consists of 146 base pairs of double-stranded DNA wrapped around an octameric histone core complex [2]. The core complex is composed of two copies of each of the histone proteins H2A, H2B, H3, and H4, organized as a central (H3-H4)<sub>2</sub> tetramer flanked by two H2A-H2B dimers [3]. Despite low sequence similarity, all core histone subunits share a common fold; they are composed of three helices separated by two short strap loops and assemble into heterodimers by interleaving the

helices into the 'handshake motif' and juxtaposing the strap loops into short parallel  $\beta$ -bridges [3]. This fold may have arisen through the duplication of a primordial helix-strand-helix motif [4,5], consistent with the hypothesis that folded proteins arose by the combination of sub-domain-sized peptides, the so-called antecedent domain segments [6-8].

Archaea also wrap their DNA into nucleosome-like structures [9]; their constituent histone subunits assemble into tetramers, which may reflect an ancestral form of the central part of the eukaryotic nucleosome octamer, the (H3-H4)<sub>2</sub> tetramer [10]. Archaeal histone subunits are occa-

sionally duplicated on a single polypeptide chain [11], a form observed in eukaryotes only in the histone-like domain of the son of sevenless protein [12].

Bacteria also have nucleoid proteins with histone-like properties [13], but these belong to a different, unrelated fold. However, a homolog of archaeal single-chain histones was recently reported from the bacterium *Aquifex aeolicus* (1R4V) [14]. Further homologs appear in the genomes of a few, phylogenetically diverse bacteria. It thus seems likely that the histone fold originated in the common ancestor of eukaryotes and archaea and spread into some bacteria through lateral gene transfer.

In an all-against-all application of HHsearch [15] to the SCOP database (JS, unpublished results) we found an evolutionary relationship between histone proteins and the helical part of the extended AAA+ ATPase domain, the C-domain [16,17]. Based on this finding, we used sequence and structure comparisons to reconstruct in detail the evolutionary events that may have shaped the histone fold. Our results point to a common origin not only with the C-domain but also with the N-terminal substrate recognition domain of Clp/Hsp100 proteins [18]. The conserved element is a helix-strand-helix motif, which we propose gave rise divergently to these three different folds and thus represents an antecedent domain segment.

## Results

Homology between proteins is typically inferred from similarities in sequence and structure. Sequence similarity is the primary criterion for deducing a common origin, but for distant evolutionary events, sequences may have diverged beyond our ability to detect their relatedness. Structures diverge much more slowly and their similarity is therefore often used to identify such distant events. However, similar structures may have arisen convergently from different origins and their similarity thus frequently does not provide conclusive evidence of common ancestry. In this study we applied a new, highly sensitive method for sequence comparison based on profile Hidden Markov Models (HMMs) to identify distant homologs of histones on the basis of sequence similarity alone. Subsequently, we validated our findings through structure comparisons.

### HMM-HMM comparisons

We used HHpred [15,19], a sensitive HMM-to-HMM comparison method, to detect homologs of the histone fold by searching the SCOP25 database [20] with sequences from the three protein families with this fold: archaeal histones, nucleosome core histones and TBP-associated factors. As expected, these identified each other as their best matches with high statistical significance (Fig.

1). Remarkably, their subsequent matches were consistently to the helical part of the extended ATPase domains found in AAA+ proteins (the C-domain) [16]. Good matches to a third protein family, the N-terminal domain of Clp/Hsp100 proteins (Clp N-domain), were frequently obtained [18]. Reciprocal searches with a set of C-domain sequences confirmed the similarity of these protein families (Fig. 1).

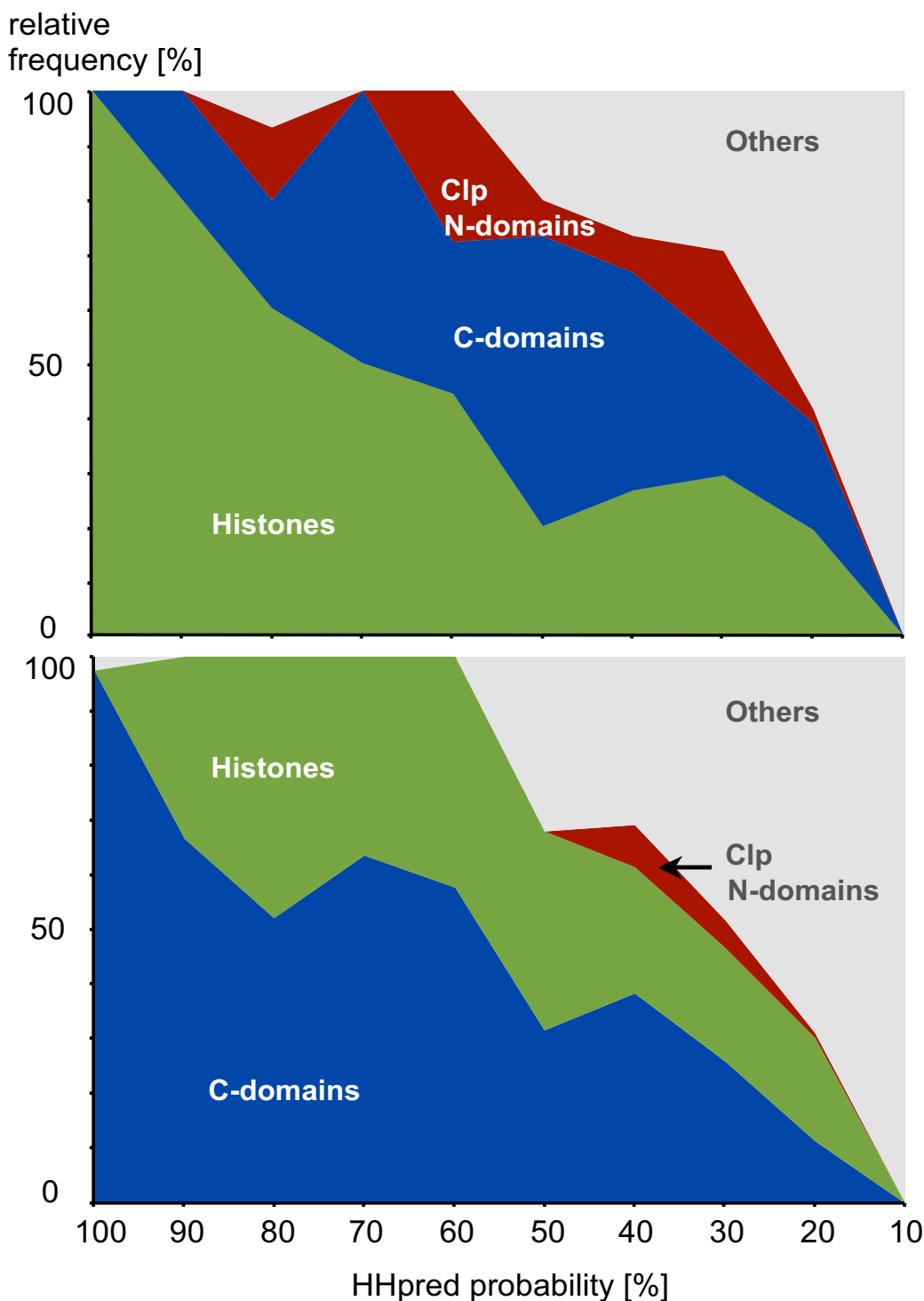
We found two high-scoring matches with other folds. These are an alanyl tRNA synthetase (1RIQ, a.203.1.1, identified by the histone entry 1JFI), and the zeta subunit of a plasmid maintenance system (1GVN, c.37.1.21, identified by two C-domains: 1LV7 and 1R7R). Subsequent analysis could not confirm these matches as homologs.

### Analysis of sequence and structure conservation

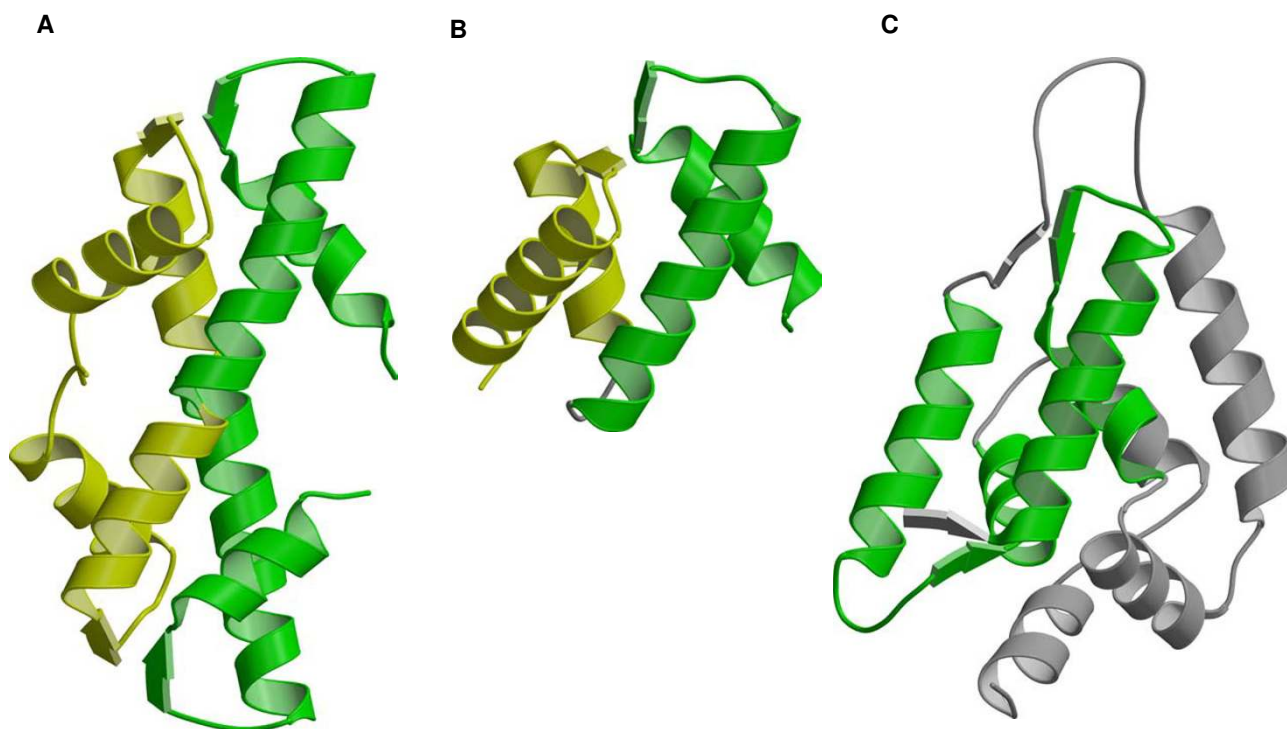
The surprising aspect of these findings is that histones, C-domains and Clp N-domains belong to three different folds (Fig. 2A–C). Histones are dimeric, interleaved helical bundles, as described in the Background section. C-domains are four-helix bundles composed of two consecutive helix-strand-helix motifs [17]. Clp N-domains, finally, are multihelical domains formed by the repetition of a 4-helical motif [21]. Although these three protein families have different topologies, they all incorporate two copies of the helix-strand-helix motif, which engages in the formation of a short parallel  $\beta$ -bridge. In the histone dimer, the  $\beta$ -bridge is formed by the association of one helix-strand-helix motif from each monomer, in the C-domain by the association of the two motifs consecutive in the polypeptide chain, and in the Clp N-domains by the association of each motif with an N-terminal strand of the symmetry-related motif.

The similarities detected by HMM-to-HMM comparison are limited to these helix-strand-helix motifs. Histones and C-domains both contain two consecutive copies of the motif and can be aligned over essentially their entire length (Fig. 3A). Clp N-domains contain two motifs decorated by two helices and each motif has its best matches to the C-terminal motif of histones and C-domains (Fig. 3A). The sequence alignment shows extensive similarity in the hydrophobic patterns of the three folds, but no highly conserved residues other than two Alanines in the core of the second helix-strand-helix motif, which allow for close packing interactions at the crossover point between the helices.

A structural comparison of the three folds shows that C-domains can be superimposed onto one half of the histone fold with root-mean-square deviations (rmsd) of around 1.5Å (Table I). The main difference between the two folds lies in the fact that the two helix-strand-helix motifs of C-domains are connected by a hinge region,



**Figure 1**  
**Results of HHpred searches of the SCOP25 database with histone sequences and C-domains.** The relative frequencies of SCOP families encountered in the searches are plotted against the HHpred probabilities as described in the Methods (searches with histones – top panel; searches with C-domains – bottom panel). Histones (SCOP a.22) are colored in green, C-domains (SCOP c.37.1.20; also includes the misclassified a.49.1.1) in blue, Clp N-domains (SCOP a.174) in red and others in gray.



**Figure 2**

**The structure of histones, C-domains and Clp N-domains.** (A) The histone of *Methanothermobacter feravidus* (1B67); the N-terminal helix-strand-helix motif in each subunit is colored yellow and the C-terminal motif green. (B) The C-domain of the helicase RuvB (1IN4); the motifs are colored as in the histone subunits. (C) Clp N-domain of ClpA (1K6K); the two helix-strand-helix motifs are colored green.

while they are continuous in histones, requiring dimerization to form the hydrophobic core (Fig. 3B). The similarity between histones and Clp N-domains is also in the range of 1.5Å rmsd, but extends only over the C-terminal helix-strand-helix motif of histones.

## Discussion

### Domain swapping as mechanism for protein evolution

The results presented here suggest an evolutionary link between histones and the C-domains of AAA+ proteins, despite differences in their topology. We propose 3D domain swapping as the mechanism that accounts for their structural differences. 3D domain swapping is a process by which two or more identical proteins exchange a domain to form interlocked oligomers [22], in which all of the packing interactions that stabilize the monomer are present. The swapped portions can range from a single secondary structure element to an entire domain. In the simplest case the native fold, normally constituted by a single 'closed' monomer, is reconstituted by two so-called 'open' monomers. This reciprocal swap leads to a homodimer, whereas the runaway domain swap, in which swapping propagates along an axis in an open-

ended manner, has been proposed to contribute to amyloid fibril formation [23-25].

Up to now, about 40 proteins have been shown to be able to undergo 3D domain swapping [26], and several studies indicate a physiological role of this mechanism in allostery and signal transduction [27-29]. A precondition is the presence of a flexible loop or hinge, about which the swapped elements can rotate in order to form a pair of 'open' monomers. The primary intervention by which 3D domain swaps have been engineered into monomeric proteins is through the shortening of the hinge, thus preventing the packing of part of the protein into its native location and forcing a swap, such as in domain 1 of lymphocyte antigen CD2 [30], staphylococcal nuclease [31], single-chain Fv fragments [32,33], in a 3-helix bundle designed by Ogihara *et al.* [34].

Our results suggest that such a shortening of the hinge region, which connects the two helix-strand-helix motifs of the AAA+ C-domain, led to a 3D domain swap. The event caused head-to-tail dimerization of monomers, which thereby recovered the lost interactions between the

A

Histone fold	$\alpha 1$	$\beta 1$	$\alpha 2$ ----- $\alpha 2$	$\beta 2$	$\alpha 3$		
1b67_A (2)	-gelpI <b>A</b> PI <b>GRI</b> I <b>K</b> N <b>A</b> G <b>A</b> e-----	<b>R</b> <b>V</b> <b>S</b> - <b>D</b> <b>D</b> <b>A</b> <b>R</b> <b>I</b> <b>A</b> <b>L</b> <b>A</b> <b>K</b> <b>V</b> <b>L</b> -----	<b>E</b> <b>E</b> <b>M</b> <b>G</b> <b>E</b> <b>E</b> <b>I</b> <b>A</b> <b>S</b> <b>E</b> <b>A</b> <b>V</b> <b>K</b> <b>L</b> <b>A</b> <b>K</b> <b>H</b> <b>A</b> <b>G</b> <b>R</b> - <b>K</b> <b>T</b> <b>I</b> <b>K</b> ---	<b>A</b> <b>E</b> <b>D</b> <b>I</b> <b>E</b> <b>L</b> <b>A</b> <b>R</b> <b>K</b> <b>M</b> <b>f</b>			
1taf_A (19)	--pkd <b>A</b> Q <b>V</b> I <b>M</b> <b>S</b> <b>I</b> <b>L</b> <b>K</b> <b>E</b> <b>L</b> <b>N</b> <b>V</b> q-----	<b>E</b> <b>Y</b> <b>E</b> - <b>P</b> <b>R</b> <b>V</b> <b>V</b> <b>N</b> <b>Q</b> <b>L</b> <b>L</b> <b>E</b> <b>F</b> <b>T</b> -----	<b>F</b> <b>R</b> <b>Y</b> <b>V</b> <b>T</b> <b>S</b> <b>I</b> <b>L</b> <b>D</b> <b>D</b> <b>A</b> <b>K</b> <b>V</b> <b>Y</b> <b>A</b> <b>N</b> <b>H</b> <b>A</b> <b>R</b> <b>K</b> - <b>K</b> <b>T</b> <b>I</b> <b>D</b> ---	<b>L</b> <b>D</b> <b>D</b> <b>V</b> <b>R</b> <b>L</b> <b>A</b> <b>T</b> <b>E</b> <b>V</b> <b>t</b>			
1taf_B (3)	gssis <b>A</b> <b>E</b> <b>S</b> <b>M</b> <b>K</b> <b>V</b> <b>I</b> <b>A</b> <b>E</b> <b>S</b> <b>I</b> <b>G</b> <b>V</b> g-----	<b>S</b> <b>L</b> <b>S</b> - <b>D</b> <b>D</b> <b>A</b> <b>A</b> <b>K</b> <b>E</b> <b>L</b> <b>A</b> <b>E</b> <b>D</b> <b>V</b> -----	<b>S</b> <b>I</b> <b>K</b> <b>L</b> <b>K</b> <b>R</b> <b>I</b> <b>V</b> <b>Q</b> <b>D</b> <b>A</b> <b>A</b> <b>K</b> <b>F</b> <b>M</b> <b>N</b> <b>H</b> <b>A</b> <b>K</b> <b>R</b> - <b>Q</b> <b>K</b> <b>L</b> <b>S</b> ---	<b>V</b> <b>R</b> <b>D</b> <b>I</b> <b>D</b> <b>M</b> <b>S</b> <b>L</b> <b>K</b> <b>V</b> -			
1tzy_A (21)	glqfp <b>V</b> <b>G</b> <b>R</b> <b>V</b> <b>H</b> <b>R</b> <b>L</b> <b>L</b> <b>R</b> <b>K</b> <b>G</b> <b>N</b> <b>Y</b> ae----	<b>R</b> <b>V</b> <b>G</b> - <b>A</b> <b>G</b> <b>A</b> <b>P</b> <b>V</b> <b>Y</b> <b>L</b> <b>A</b> <b>A</b> <b>V</b> <b>L</b> -----	<b>E</b> <b>Y</b> <b>L</b> <b>T</b> <b>A</b> <b>E</b> <b>I</b> <b>L</b> <b>E</b> <b>L</b> <b>A</b> <b>G</b> <b>N</b> <b>A</b> <b>A</b> <b>R</b> <b>D</b> <b>N</b> <b>K</b> <b>K</b> - <b>T</b> <b>R</b> <b>I</b> <b>I</b> ---	<b>P</b> <b>R</b> <b>H</b> <b>L</b> <b>Q</b> <b>L</b> <b>A</b> <b>I</b> <b>R</b> <b>N</b> <b>d</b>			
1tzy_B (33)	rkesy <b>S</b> <b>I</b> <b>Y</b> <b>V</b> <b>K</b> <b>V</b> <b>L</b> <b>K</b> <b>Q</b> <b>V</b> hpd-----	<b>G</b> <b>I</b> <b>S</b> - <b>S</b> <b>K</b> <b>A</b> <b>M</b> <b>G</b> <b>I</b> <b>M</b> <b>S</b> <b>F</b> <b>V</b> -----	<b>N</b> <b>D</b> <b>I</b> <b>F</b> <b>E</b> <b>R</b> <b>I</b> <b>A</b> <b>G</b> <b>E</b> <b>A</b> <b>S</b> <b>R</b> <b>L</b> <b>A</b> <b>H</b> <b>Y</b> <b>N</b> <b>K</b> <b>R</b> - <b>S</b> <b>T</b> <b>I</b> <b>T</b> ---	<b>S</b> <b>R</b> <b>E</b> <b>I</b> <b>Q</b> <b>T</b> <b>A</b> <b>V</b> <b>R</b> <b>L</b> <b>l</b>			
1tzy_C (62)	rklpf <b>Q</b> <b>R</b> <b>L</b> <b>V</b> <b>R</b> <b>E</b> <b>I</b> <b>A</b> <b>Q</b> <b>D</b> <b>F</b> <b>K</b> <b>t</b> dl----	<b>R</b> <b>F</b> <b>Q</b> - <b>S</b> <b>S</b> <b>A</b> <b>V</b> <b>M</b> <b>A</b> <b>L</b> <b>Q</b> <b>E</b> <b>A</b> <b>S</b> -----	<b>E</b> <b>A</b> <b>Y</b> <b>L</b> <b>V</b> <b>G</b> <b>L</b> <b>F</b> <b>E</b> <b>D</b> <b>T</b> <b>N</b> <b>L</b> <b>C</b> <b>A</b> <b>I</b> <b>H</b> <b>A</b> <b>K</b> <b>R</b> - <b>V</b> <b>T</b> <b>I</b> <b>M</b> ---	<b>P</b> <b>K</b> <b>D</b> <b>I</b> <b>Q</b> <b>L</b> <b>A</b> <b>R</b> <b>R</b> <b>I</b> <b>r</b>			
1tzy_D (27)	--qgit <b>K</b> <b>P</b> <b>A</b> <b>I</b> <b>R</b> <b>R</b> <b>L</b> <b>A</b> <b>R</b> <b>R</b> <b>G</b> <b>G</b> <b>v</b> k-----	<b>R</b> <b>I</b> <b>S</b> - <b>G</b> <b>L</b> <b>I</b> <b>Y</b> <b>E</b> <b>E</b> <b>T</b> <b>R</b> <b>G</b> <b>V</b> <b>L</b> -----	<b>K</b> <b>V</b> <b>F</b> <b>L</b> <b>E</b> <b>N</b> <b>V</b> <b>I</b> <b>R</b> <b>D</b> <b>A</b> <b>V</b> <b>T</b> <b>Y</b> <b>T</b> <b>E</b> <b>H</b> <b>A</b> <b>K</b> <b>R</b> - <b>K</b> <b>T</b> <b>V</b> <b>T</b> ---	<b>A</b> <b>M</b> <b>D</b> <b>V</b> <b>V</b> <b>Y</b> <b>A</b> <b>L</b> <b>K</b> <b>R</b> <b>q</b>			
C-domain of AAA+	$\alpha 1$	$\beta 1$	$\alpha 2$	hinge	$\alpha 3$	$\beta 2$	$\alpha 4$
1in4_A (181)	t <b>V</b> <b>K</b> <b>E</b> <b>L</b> <b>K</b> <b>E</b> <b>I</b> <b>I</b> <b>K</b> <b>R</b> <b>A</b> <b>A</b> <b>S</b> <b>L</b> <b>M</b> <b>D</b> <b>V</b> -----	<b>E</b> <b>I</b> <b>E</b> - <b>D</b> <b>A</b> <b>A</b> <b>A</b> <b>E</b> <b>M</b> <b>I</b> <b>A</b> <b>K</b> <b>R</b> - <b>s</b> <b>r</b> <b>g</b> <b>t</b> -----	<b>P</b> <b>R</b> <b>I</b> <b>A</b> <b>I</b> <b>R</b> <b>L</b> <b>T</b> <b>K</b> <b>R</b> <b>V</b> <b>R</b> <b>D</b> <b>M</b> <b>L</b> <b>T</b> <b>V</b> <b>V</b> <b>K</b> <b>A</b> - <b>D</b> <b>R</b> <b>I</b> <b>N</b> ---				<b>T</b> <b>D</b> <b>I</b> <b>V</b> <b>L</b> <b>K</b> <b>T</b> <b>M</b> <b>E</b> <b>V</b> <b>l</b>
1r6b_A (350)	s <b>I</b> <b>E</b> <b>E</b> <b>T</b> <b>V</b> <b>Q</b> <b>I</b> <b>I</b> <b>N</b> <b>G</b> <b>L</b> <b>K</b> <b>P</b> <b>K</b> <b>Y</b> eahhdv----	<b>R</b> <b>Y</b> <b>T</b> - <b>A</b> <b>K</b> <b>A</b> <b>V</b> <b>R</b> <b>A</b> <b>A</b> <b>V</b> <b>E</b> <b>L</b> - <b>a</b> <b>v</b> <b>k</b> <b>y</b> <b>i</b> <b>n</b> <b>d</b> <b>r</b> <b>h</b> <b>l</b> <b>P</b> <b>D</b> <b>K</b> <b>A</b> <b>I</b> <b>D</b> <b>V</b> <b>I</b> <b>D</b> <b>E</b> <b>A</b> <b>G</b> <b>A</b> <b>R</b> <b>A</b> <b>R</b> <b>L</b> <b>M</b> <b>p</b> <b>v</b> <b>s</b> <b>k</b> <b>r</b> <b>k</b> <b>t</b> <b>t</b> <b>v</b> <b>n</b> <b>V</b> <b>A</b> <b>D</b> <b>I</b> <b>E</b> <b>S</b> <b>V</b> <b>V</b> <b>A</b> <b>r</b> <b>i</b>					
1lv7_A (326)	d <b>V</b> <b>R</b> <b>G</b> <b>R</b> <b>E</b> <b>Q</b> <b>I</b> <b>L</b> <b>K</b> <b>V</b> <b>H</b> <b>M</b> -- <b>r</b> <b>r</b> <b>v</b> -----	<b>P</b> <b>L</b> <b>A</b> p <b>d</b> i <b>d</b> <b>A</b> <b>A</b> <b>I</b> <b>I</b> <b>A</b> <b>R</b> <b>G</b> - <b>t</b> <b>p</b> <b>g</b> <b>f</b> <b>s</b> -----	<b>G</b> <b>A</b> <b>D</b> <b>L</b> <b>A</b> <b>N</b> <b>L</b> <b>V</b> <b>N</b> <b>E</b> <b>A</b> <b>A</b> <b>L</b> <b>F</b> <b>A</b> <b>A</b> <b>R</b> <b>G</b> <b>N</b> <b>K</b> - <b>R</b> <b>V</b> <b>V</b> <b>S</b> ---				<b>M</b> <b>V</b> <b>E</b> <b>F</b> <b>E</b> <b>K</b> <b>A</b> <b>K</b> <b>D</b> <b>K</b> <b>i</b>
1ny5_A (314)	r <b>K</b> <b>E</b> <b>D</b> <b>I</b> <b>I</b> <b>P</b> <b>L</b> <b>A</b> <b>N</b> <b>H</b> <b>F</b> <b>L</b> <b>K</b> <b>K</b> <b>F</b> <b>s</b> <b>r</b> <b>k</b> <b>y</b> <b>a</b> <b>k</b> <b>e</b> <b>v</b> <b>e</b> <b>g</b> <b>F</b> <b>T</b> - <b>K</b> <b>S</b> <b>A</b> <b>Q</b> <b>E</b> <b>L</b> <b>L</b> <b>S</b> <b>Y</b> <b>p</b> <b>w</b> <b>y</b> <b>g</b> <b>n</b> -----		<b>V</b> <b>R</b> <b>E</b> <b>L</b> <b>K</b> <b>N</b> <b>V</b> <b>I</b> <b>E</b> <b>R</b> <b>A</b> <b>V</b> <b>L</b> <b>F</b> <b>S</b> egk----				<b>F</b> <b>I</b> <b>D</b> --- <b>R</b> <b>G</b> <b>E</b> <b>L</b> <b>S</b> <b>C</b> <b>L</b> <b>V</b> ---
1g8p_A (220)	d <b>V</b> <b>E</b> <b>T</b> <b>R</b> <b>V</b> <b>E</b> <b>V</b> <b>I</b> <b>R</b> <b>R</b> <b>D</b> <b>T</b> <b>Y</b> <b>D</b>   <b>V</b> -----	<b>E</b> <b>A</b> <b>P</b> - <b>N</b> <b>T</b> <b>A</b> <b>L</b> <b>Y</b> <b>D</b> <b>C</b> <b>A</b> <b>A</b> <b>L</b> - <b>c</b> <b>i</b> <b>a</b> <b>l</b> <b>g</b> <b>s</b> -- <b>d</b> <b>g</b> <b>L</b> <b>R</b> <b>G</b> <b>E</b> <b>L</b> <b>T</b> <b>L</b> <b>L</b> <b>R</b> <b>S</b> <b>A</b> <b>R</b> <b>A</b> <b>L</b> <b>A</b> <b>E</b> <b>L</b> <b>E</b> <b>G</b> <b>A</b> - <b>T</b> <b>A</b> <b>V</b> <b>G</b> ---					<b>R</b> <b>D</b> <b>H</b> <b>L</b> <b>K</b> <b>R</b> <b>V</b> <b>A</b> <b>T</b> --
Clp-N motif		[ $\beta 2$ ]			$\alpha 1$	$\beta 1$	$\alpha 2$
1k6k_A (4)	-----	[ <b>q</b> <b>p</b> <b>t</b> ]-----	-----	-----	<b>Q</b> <b>E</b> <b>L</b> <b>E</b> <b>L</b> <b>S</b> <b>L</b> <b>N</b> <b>M</b> <b>A</b> <b>F</b> <b>A</b> <b>R</b> <b>A</b> <b>R</b> <b>E</b> <b>H</b> <b>R</b> <b>H</b> - <b>E</b> <b>F</b> <b>M</b> <b>T</b> ---		<b>V</b> <b>E</b> <b>H</b> <b>L</b> <b>L</b> <b>L</b> <b>A</b> <b>L</b> <b>L</b> <b>S</b> <b>n</b>
1k6k_A (82)	-----	[ <b>m</b> <b>l</b> <b>n</b> ]-----	-----	-----	<b>L</b> <b>S</b> <b>F</b> <b>Q</b> <b>R</b> <b>V</b> <b>L</b> <b>Q</b> <b>R</b> <b>A</b> <b>V</b> <b>F</b> <b>H</b> <b>V</b> <b>Q</b> <b>S</b> <b>S</b> <b>G</b> <b>R</b> - <b>N</b> <b>E</b> <b>V</b> <b>T</b> ---		<b>G</b> <b>A</b> <b>N</b> <b>V</b> <b>L</b> <b>V</b> <b>A</b> <b>I</b> <b>F</b> <b>S</b> <b>e</b>
1khy_A (8)	-----	[ <b>q</b> <b>p</b> <b>s</b> ]-----	-----	-----	<b>N</b> <b>K</b> <b>F</b> <b>Q</b> <b>L</b> <b>A</b> <b>L</b> <b>A</b> <b>D</b> <b>A</b> <b>Q</b> <b>S</b> <b>L</b> <b>A</b> <b>L</b> <b>G</b> <b>H</b> <b>D</b> <b>N</b> - <b>Q</b> <b>F</b> <b>I</b> <b>E</b> ---		<b>P</b> <b>L</b> <b>H</b> <b>L</b> <b>M</b> <b>S</b> <b>A</b> <b>L</b> <b>L</b> <b>N</b> <b>q</b>
1khy_A (85)	-----	[ <b>r</b> <b>l</b> <b>t</b> ]-----	-----	-----	<b>Q</b> <b>D</b> <b>L</b> <b>V</b> <b>R</b> <b>V</b> <b>L</b> <b>N</b> <b>L</b> <b>C</b> <b>D</b> <b>K</b> <b>L</b> <b>A</b> <b>Q</b> <b>K</b> <b>R</b> <b>G</b> <b>D</b> - <b>N</b> <b>F</b> <b>I</b> <b>S</b> ---		<b>S</b> <b>E</b> <b>L</b> <b>F</b> <b>V</b> <b>L</b> <b>A</b> <b>A</b> <b>L</b> <b>S</b>

B

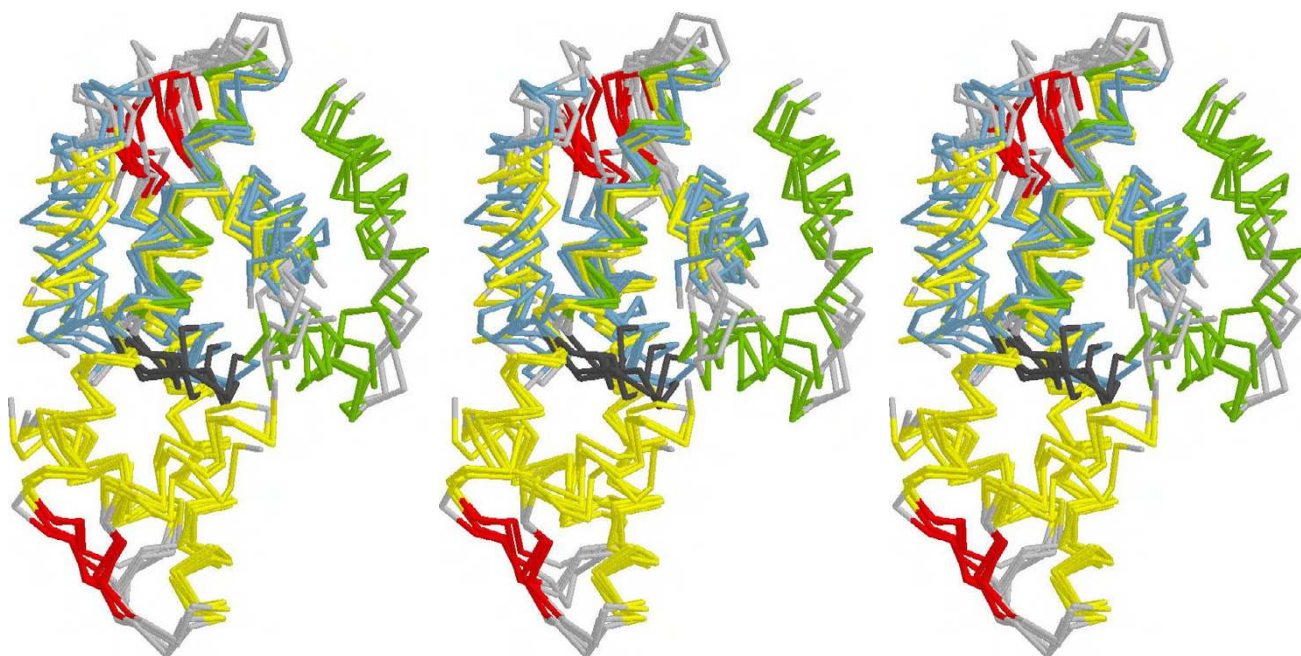
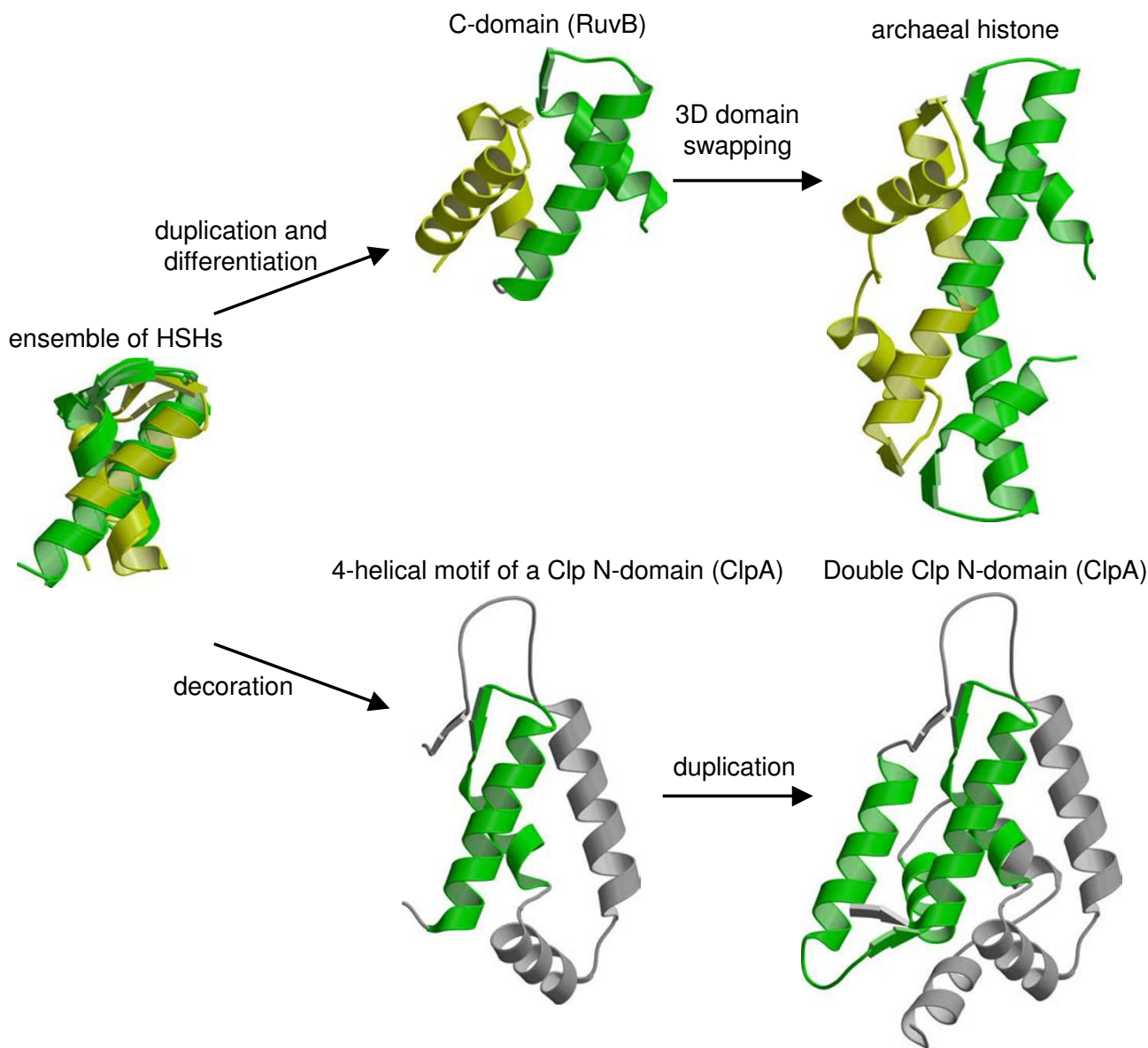


Figure 3

**Sequence and structure comparisons of histones, C-domains and Clp N-domains.** (A) Multiple sequence alignment of representative members of each fold. Residues in helices are colored yellow in histones, blue in C-domains, and green in Clp N-domains; residues in  $\beta$ -bridges are colored red. Structurally equivalent residues are shown in capital letters and residues forming the hydrophobic core are shown bold. The sequences are labeled by their PDB codes; the numbers in brackets refer to the residue number for the first residue in the alignment. (B) Global superposition of the proteins listed in panel A. The superposition was made using the archaeal histone HmfA (1B67) as a reference structure. Quantitative information on the results of the superposition is provided in Table I. The coloring is as in panel A; in addition, the hinge region of C-domains is highlighted in black.



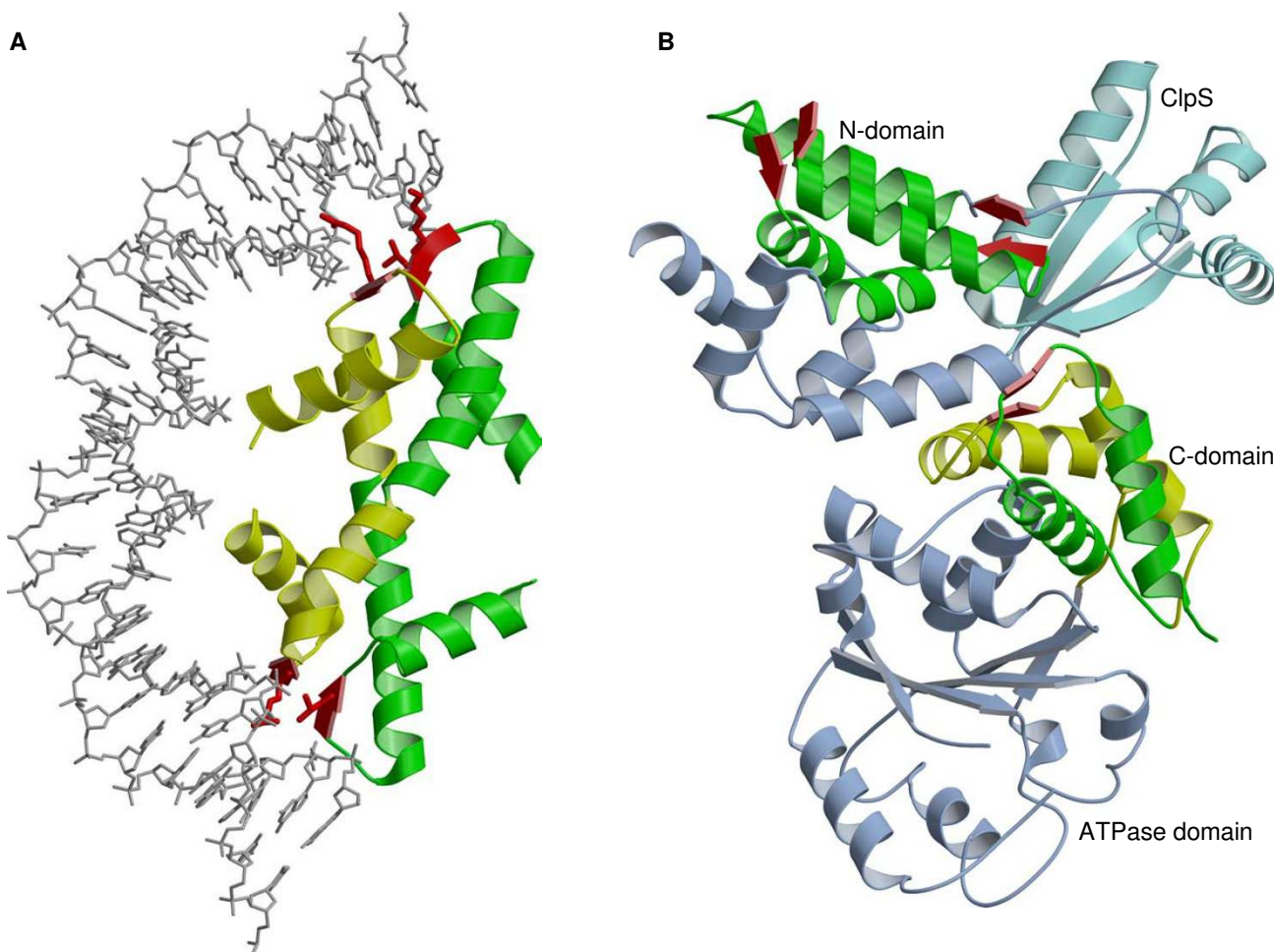
**Figure 4**  
**Evolutionary scenario for the origin of three folds from an ancestral helix-strand-helix motif.** The coloring and representative proteins are as in Fig. 2. The superimposed ensemble of helix-strand-helix motifs consists of motifs from the following proteins: yellow (1IN4: residues 181–212; 1B67: 4–33), green (1IN4: 216–251; 1B67: 134–166; 1K6K: 82–115).

two helix-strand-helix motifs, and resulted in the emergence of the histone fold (Fig. 4). Following the proposal that domain swapping might contribute to protein evolution [22,35], we present here the first concrete example.

**A primordial helix-strand-helix motif**

The helix-strand-helix motif, which is at the core of the similarity between histones and C-domains, is also found in Clp N-domains, which assume yet a third fold. Here, the motif is decorated with two C-terminal helices, and

two copies of this extended, 4-helical motif are fused in antiparallel orientation. Thus, three different folds appear to have been built from a common helix-strand-helix motif. One theory for the origin of folded proteins proposes that they arose by fusion and recombination from an ancestral set of peptides, which emerged in the context of RNA-dependent replication and catalysis (the 'RNA world') [6-8]. The helix-strand-helix motif would be such an ancestral peptide, which gave rise divergently to the Clp N-domain and the AAA+ C-domain through two



**Figure 5**  
**Involvement of the  $\beta$ -bridge in macromolecular interactions.** The coloring of helix-strand-helix motifs is as in Fig. 2, except the  $\beta$ -bridges are colored red. (A) The histone H3-H4 complex bound to DNA (1S32); residues of the  $\beta$ -bridges engaged in interactions with the phosphate backbone are shown in stick representation. (B) ClpS in complex with ClpA (1LZW, 1R6B); the ATPase domain is in light blue and ClpS in cyan.

independent events of duplication and fusion (Fig. 4). The C-domain then evolved into the histone fold by 3D domain swapping. This scenario extends a previous hypothesis on the origin of eukaryotic core histones, which proposed that they evolved from the duplication of a single helix-strand-helix motif [4,5].

In this study we have deduced homology based on similarities in sequence and structure. We are aware that homology of proteins is an assumption inferred from heuristics, of which sequence similarity is generally accepted as the best indicator. Structural similarity alone, especially of small fragments, does not necessarily imply evolutionary divergence, since it may result from general biophysical constraints. Indeed, we find a number of  $\alpha$ -helical hairpins in the PDB with a high degree of structural

similarity to the helix-strand-helix motif (rmsds of less than 1.5Å); some examples include hairpins from fumarate reductase (1QLA\_A, residues 65–94) and tetracycline repressor-like protein (1T33\_A, residues 144–173). However, none of them show detectable sequence similarity to each other or to the proteins in our study. This shows that the constraints of structure on sequence variability are not sufficient to explain the observed sequence similarity between histones, C-domains, and Clp N-domains.

#### Functional implications

An interesting structural feature common to all three folds is the presence of one or two short, parallel  $\beta$ -bridges formed by the strands of the helix-strand-helix motifs. In histones, these  $\beta$ -bridges provide the main site of interaction with the phosphate backbone of DNA (Fig. 5). In Clp

**Table 1: Data for the superposition in Fig. 3**

PDB-ID	Name	SOURCE SPECIES	FOLD	SUBGROUP	NO. OF ALIGNED RESIDUES	RMSD TO HMFA [Å]
<u>1B67</u> A+B	HmfA	<i>Methanothermus fervidus</i>	Histone	Archaeal	124/124	0.00
<u>1TAF</u> A+B	TAFII42/62	<i>Drosophila melanogaster</i>	Histone	TBP-associated factors	124/132	1.19
<u>1TZY</u> A+B	H2A/H2B	<i>Gallus gallus</i>	Histone	Nucleosome core	118/129	1.33
<u>1TZY</u> C+D	H3/H4	<i>Gallus gallus</i>	Histone	Nucleosome core	114/134	1.10
<u>1IN4</u> A	RuvB	<i>Thermotoga maritima</i>	C-domain	AAA+: helicases	46/71	1.38
<u>1R6B</u> A	ClpA-D1	<i>Escherichia coli</i>	C-domain	AAA+: Clp-D1	52/86	1.70
<u>1LV7</u> A	FtsH	<i>Escherichia coli</i>	C-domain	AAA+: AAA	45/71	1.21
<u>1INY5</u> A	NtrC	<i>Aquifex aeolicus</i>	C-domain	AAA+: $\sigma$ 54	52/72	1.57
<u>1G8P</u> A	Mg chelatase	<i>Rhodobacter capsulatus</i>	C-domain	AAA+	a) 12/16( $\alpha$ 1) b) 47/54( $\alpha$ 2-4)	a) 0.59 b) 1.38
<u>1K6K</u> A	ClpA-N (1 <sup>st</sup> half)	<i>Escherichia coli</i>	Clp N-domain	ClpA	33/78	1.54
<u>1K6K</u> A	ClpA-N (2 <sup>nd</sup> half)	<i>Escherichia coli</i>	Clp N-domain	ClpA	33/62	1.40
<u>1KHY</u> A	ClpB-N (1 <sup>st</sup> half)	<i>Escherichia coli</i>	Clp N-domain	ClpB	33/80	1.13
<u>1KHY</u> A	ClpB-N (2 <sup>nd</sup> half)	<i>Escherichia coli</i>	Clp N-domain	ClpB	33/58	0.96

N-domains, one of the two  $\beta$ -bridges binds the adaptor molecule ClpS [18,21] (Fig. 5). Although the binding sites of the AAA+ C-domains have not been characterized yet, it thus seems attractive to propose that here also the single  $\beta$ -bridge formed in this domain represents the main binding site. C-domains play an important role in sensing the nucleotide bound by the AAA+ proteins [36-38] and are located close to the substrate-binding N-domains (Fig. 5), projecting radially at the circumference of the hexameric ring complex. We note in this context that C-domains are frequently rich in positively charged residues and that in the Lon protease, the C-domain has been implicated in interactions with DNA [39]. We propose that the helix-strand-helix motif served as a scaffold for the formation of parallel  $\beta$ -bridges. Ancestrally, these bridges bound proteins, but in a few C-domains they also acquired the ability to bind DNA, eventually leading to histones as proteins that only bind DNA at these sites.

## Conclusion

We have retraced the evolutionary events which may have shaped the histone fold and have found connections to two other folds; the N-terminal substrate recognition domain of Clp/Hsp100 proteins and the helical part of the extended AAA+ ATPase domain. These 3 folds contain a homologous helix-strand-helix motif, despite the differences in the topology, leading us to propose a scenario for the origin of these folds from a common ancestral helix-strand-helix motif through events of duplication, fusion and 3D domain swapping. The short functional parallel  $\beta$ -bridges formed by the strands of the helix-strand-helix motifs seem to be the evolutionary driving force for the conservation of this motif. Our findings provide additional support for our previously proposed hypothesis

that the diversity of today's folds might have arisen from an ancestral set of peptides.

## Methods

We obtained histone and Clp N-domain sequences from the ASTRAL compendium [40] as defined by the SCOP (version 1.71) [20] folds a.22 and a.174, respectively, and reduced the set to less than 25% pairwise identity at 90% length coverage using BLASTCLUST [41]. C-domains are not characterized as a separate fold in SCOP; we extracted their sequences from the 'extended AAA-ATPase' family (c.37.1.20) of the SCOP database by a procedure described by Ammelburg et al. [17] and also reduced this set to less than 25% pairwise identity.

We used these sequences to search the SCOP25 database for homologs with HHpred [15,19], at default parameters and a probability cutoff of 10%. The SCOP25 database is a version of SCOP filtered for a maximum of 25% pairwise sequence identity. For each group, we pooled all search results and tabulated the frequencies at which various SCOP families appeared at each probability, binned at 10% intervals.

The histone, C-domain and Clp N-domain structures were superimposed interactively in Swiss-PDB viewer [42]. We chose the archaeal histone HmfA (1B67) as the reference structure, as it made the highest number of connections both in sequence and structure searches. Quantitative information for the superimposition is listed in Table 1. The alignment in Fig. 3A reflects the structural superposition. The complex shown in Fig. 5B, consisting of ClpS, N-domain and the first AAA+ domain of ClpA, was generated by superimposing the N-domains of the structures



1R6B (N-domain and the AAA+ domains) and 1LZW (N-domain in complex with ClpS) from *E. coli*.

### Competing interests

The author(s) declare that they have no competing interests.

### Authors' contributions

VA and MA contributed equally to this work. JS discovered the similarity between histones and C-domains. VA, MA and AL conceived this study and VA and MA executed the analysis. VA, MA and AL wrote the paper. All authors read and approved the final manuscript.

### Acknowledgements

The authors thank Nick Grishin for discussions. This work was supported by institutional funds from the Max Planck Society.

### References

- Kornberg RD, Thomas JO: **Chromatin structure; oligomers of the histones.** *Science* 1974, **184**(139):865-868.
- Luger K, Mader AVW, Richmond RK, Sargent DF, Richmond TJ: **Crystal structure of the nucleosome core particle at 2.8 Å resolution.** *Nature* 1997, **389**(6648):251-260.
- Arents G, Burlingame RW, Wang BC, Love WE, Moudrianakis EN: **The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix.** *Proc Natl Acad Sci USA* 1991, **88**(22):10148-10152.
- Arents G, Moudrianakis EN: **Topography of the histone octamer surface: repeating structural motifs utilized in the docking of nucleosomal DNA.** *Proc Natl Acad Sci USA* 1993, **90**(22):10489-10493.
- Arents G, Moudrianakis EN: **The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization.** *Proc Natl Acad Sci USA* 1995, **92**(24):11170-11174.
- Lupas AN, Ponting CP, Russell RB: **On the evolution of protein folds: are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world?** *J Struct Biol* 2001, **134**(2-3):191-203.
- Soding J, Lupas AN: **More than the sum of their parts: on the evolution of proteins from peptides.** *Bioessays* 2003, **25**(9):837-846.
- Fetrow JS, Godzik A: **Function driven protein evolution. A possible proto-protein for the RNA-binding proteins.** *Pac Symp Biocomput* 1998:485-496.
- Pereira SL, Grayling RA, Lurz R, Reeve JN: **Archaeal nucleosomes.** *Proc Natl Acad Sci USA* 1997, **94**(23):12633-12637.
- Bailey KA, Chow CS, Reeve JN: **Histone stoichiometry and DNA circularization in archaeal nucleosomes.** *Nucleic Acids Res* 1999, **27**(2):532-536.
- Fahrner RL, Cascio D, Lake JA, Slesarev A: **An ancestral nuclear protein assembly: crystal structure of the Methanopyrus kandleri histone.** *Protein Sci* 2001, **10**(10):2002-2007.
- Baxeianis AD, Arents G, Moudrianakis EN, Landsman D: **A variety of DNA-binding and multimeric proteins contain the histone fold motif.** *Nucleic Acids Res* 1995, **23**(14):2685-2691.
- Drlica K, Rouviere-Yaniv J: **Histonelike proteins of bacteria.** *Microbiol Rev* 1987, **51**(3):301-319.
- Qiu Y, Tereshko V, Kim Y, Zhang R, Collart F, Yousef M, Kossiakoff A, Joachimiak A: **The crystal structure of Aq\_328 from the hyperthermophilic bacteria Aquifex aeolicus shows an ancestral histone fold.** *Proteins* 2006, **62**(1):8-16.
- Soding J: **Protein homology detection by HMM-HMM comparison.** *Bioinformatics* 2005, **21**(7):951-960.
- Neuwald AF, Aravind L, Spouge JL, Koonin EV: **AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes.** *Genome Res* 1999, **9**(1):27-43.
- Ammelburg M, Frickey T, Lupas AN: **Classification of AAA+ proteins.** *J Struct Biol* 2006.
- Zeth K, Ravelli RB, Paal K, Cusack S, Bukau B, Dougan DA: **Structural analysis of the adaptor protein ClpS in complex with the N-terminal domain of ClpA.** *Nat Struct Biol* 2002, **9**(12):906-911.
- Soding J, Biegert A, Lupas AN: **The HHpred interactive server for protein homology detection and structure prediction.** *Nucleic Acids Res* 2005:W244-248.
- Andreeva A, Howorth D, Brenner SE, Hubbard TJ, Chothia C, Murzin AG: **SCOP database in 2004: refinements integrate structure and sequence family data.** *Nucleic Acids Res* 2004:D226-229.
- Maurizi MR, Xia D: **Protein binding and disruption by Clp/Hsp100 chaperones.** *Structure* 2004, **12**(2):175-183.
- Bennett MJ, Schlunegger MP, Eisenberg D: **3D domain swapping: a mechanism for oligomer assembly.** *Protein Sci* 1995, **4**(12):2455-2468.
- Janowski R, Kozak M, Jankowska E, Grzonka Z, Grubb A, Abrahamson M, Jaskolski M: **Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping.** *Nat Struct Biol* 2001, **8**(4):316-320.
- Guo Z, Eisenberg D: **Runaway domain swapping in amyloid-like fibrils of T7 endonuclease I.** *Proc Natl Acad Sci USA* 2006, **103**(21):8042-8047.
- Sambashivan S, Liu Y, Sawaya MR, Gingery M, Eisenberg D: **Amyloid-like fibrils of ribonuclease A with three-dimensional domain-swapped and native-like structure.** *Nature* 2005, **437**(7056):266-269.
- Liu Y, Eisenberg D: **3D domain swapping: as domains continue to swap.** *Protein Sci* 2002, **11**(6):1285-1299.
- Piccoli R, Di Donato A, D'Alessio G: **Co-operativity in seminal ribonuclease function. Kinetic studies.** *Biochem J* 1988, **253**(2):329-336.
- Gotte G, Bertoldi M, Libonati M: **Structural versatility of bovine ribonuclease A. Distinct conformers of trimeric and tetrameric aggregates of the enzyme.** *Eur J Biochem* 1999, **265**(2):680-687.
- Schymkowitz JW, Rousseau F, Wilkinson HR, Friedler A, Itzhaki LS: **Observation of signal transduction in three-dimensional domain swapping.** *Nat Struct Biol* 2001, **8**(10):888-892.
- Murray AJ, Lewis SJ, Barclay AN, Brady RL: **One sequence, two folds: a metastable structure of CD2.** *Proc Natl Acad Sci USA* 1995, **92**(16):7337-7341.
- Green SM, Gittis AG, Meeker AK, Lattman EE: **One-step evolution of a dimer from a monomeric protein.** *Nat Struct Biol* 1995, **2**(9):746-751.
- Kortt AA, Malby RL, Caldwell JB, Gruen LC, Ivancic N, Lawrence MC, Howlett GJ, Webster RG, Hudson PJ, Colman PM: **Recombinant anti-sialidase single-chain variable fragment antibody. Characterization, formation of dimer and higher-molecular-mass multimers and the solution of the crystal structure of the single-chain variable fragment/sialidase complex.** *Eur J Biochem* 1994, **221**(1):151-157.
- Perisic O, Webb PA, Holliger P, Winter G, Williams RL: **Crystal structure of a diabody, a bivalent antibody fragment.** *Structure* 1994, **2**(12):1217-1226.
- Ogihara NL, Ghirlanda G, Bryson JW, Gingery M, DeGrado WF, Eisenberg D: **Design of three-dimensional domain-swapped dimers and fibrous oligomers.** *Proc Natl Acad Sci USA* 2001, **98**(4):1404-1409.
- Kinch LN, Grishin NV: **Evolution of protein structures and functions.** *Curr Opin Struct Biol* 2002, **12**(3):400-408.
- Botos I, Melnikov EE, Cherry S, Khalatova AG, Rasulova FS, Tropea JE, Maurizi MR, Rotanova TV, Gustchina A, Wlodawer A: **Crystal structure of the AAA+ alpha domain of E. coli Lon protease at 1.9 Å resolution.** *J Struct Biol* 2004, **146**(1-2):113-122.
- Ogura T, Whiteheart SW, Wilkinson AJ: **Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases.** *J Struct Biol* 2004, **146**(1-2):106-112.
- Diemand AV, Lupas AN: **Modeling AAA+ ring complexes from monomeric structures.** *J Struct Biol* 2006.
- Lee AY, Hsu CH, Wu SH: **Functional domains of Brevibacillus thermoruber lon protease for oligomerization and DNA binding: role of N-terminal and sensor and substrate discrimination domains.** *J Biol Chem* 2004, **279**(33):34903-34912.

40. Brenner SE, Koehl P, Levitt M: **The ASTRAL compendium for protein structure and sequence analysis.** *Nucleic Acids Res* 2000, **28(1)**:254-256.
41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215(3)**:403-410.
42. Guex N, Peitsch MC: **SWISS-MODEL and the Swiss-Pdb-Viewer: an environment for comparative protein modeling.** *Electrophoresis* 1997, **18(15)**:2714-2723.

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

