

2012

Conservation of the Three-Dimensional Structure in Non-Homologous Or Unrelated Proteins

Konstantinos Sousounis

Carl E. Haney

Jin Cao

Bharath Sunchu

Panagiotis A. Tsonis

University of Dayton, ptsonis1@udayton.edu

Follow this and additional works at: https://ecommons.udayton.edu/bio_fac_pub



Part of the [Biology Commons](#)

eCommons Citation

Sousounis, Konstantinos; Haney, Carl E.; Cao, Jin; Sunchu, Bharath; and Tsonis, Panagiotis A., "Conservation of the Three-Dimensional Structure in Non-Homologous Or Unrelated Proteins" (2012). *Biology Faculty Publications*. 10.
https://ecommons.udayton.edu/bio_fac_pub/10

This Article is brought to you for free and open access by the Department of Biology at eCommons. It has been accepted for inclusion in Biology Faculty Publications by an authorized administrator of eCommons. For more information, please contact frice1@udayton.edu, mschlangen1@udayton.edu.

REVIEW

Open Access

Conservation of the three-dimensional structure in non-homologous or unrelated proteins

Konstantinos Sousounis¹, Carl E Haney¹, Jin Cao², Bharath Sunchu³ and Panagiotis A Tsonis^{1*}

Abstract

In this review, we examine examples of conservation of protein structural motifs in unrelated or non-homologous proteins. For this, we have selected three DNA-binding motifs: the histone fold, the helix-turn-helix motif, and the zinc finger, as well as the globin-like fold. We show that indeed similar structures exist in unrelated proteins, strengthening the concept that three-dimensional conservation might be more important than the primary amino acid sequence.

Keywords: 3D protein structure, Conserved motifs, Unrelated proteins

Introduction

When the human genome was sequenced (as well as that of other mammals), it was estimated that there are approximately 25,000 genes encoding for proteins [1,2]. After being synthesized, proteins assume their three-dimensional structure by a specific arrangement of beta strands, alpha helices, turns, or loops. In many cases, a combination of these structural features creates certain motifs, exerting a particular function (i.e., DNA binding) that is quite conserved in proteins from virtually all organisms. Interestingly, the number of these motifs is much smaller than the number of genes. However, it has also been noted that some structural motifs show significant robustness even though no significant homology exists among them at the primary amino acid sequence. It seems that evolutionary constraints have limited the ability of proteins to become vastly different. Moreover, it has been shown that protein structures are three to ten times more conserved than the amino acid sequence [3]. Thus, a particular motif, i.e., a zinc-binding domain of very similar or virtually identical structure, can be found in many different proteins, which could also be unrelated to each other when function is concerned. Thus, it seems that evolution does favor conservation of structural motifs in proteins.

The purpose of this tutorial/review is to illustrate this diversity that exists in the function of structurally conserved protein motifs. For this reason, protein folds with

low homology in amino acid sequence and high structural similarity were used. The analysis for the obvious reason of space is not exhaustive and is focused on four specific protein structural folds: We have selected to present data with three different DNA-binding domains: the histone fold, the helix-turn-helix motif (HTH), and the zinc finger, as well the globin-like fold, part of an important protein in oxygen binding and transport. These four folds were chosen because they are ubiquitous in many different organisms and are well represented in many different proteins.

For our comparisons, an intensive search of the Vector Alignment Search Tool (VAST) [4,5], an algorithm to determine three-dimensional (3D) structure similarities according to geometric criteria, was done. A protein family was identified using a representative protein and, using VAST and the Molecular Modeling Database [6], dissimilar structure proteins were identified and annotated followed by root mean square deviation (RMSD) determination. The structures were then downloaded into Cn3D ('see in 3D') [7] for viewing the sequence alignment. The above are part of Entrez [8,9]. These structures were then aligned in PyMOL [10] for 3D viewing. The files for the PyMOL structures provided have been downloaded from the Protein Data Bank (PDB) [11]. The lower the RMSD means better structural alignment. Lower identity means that the two proteins do not share the same amino acids in the corresponding structural alignment. Though, depending on how big are the structures that we are comparing, the RMSD and sequence identity may vary. Small domains may

* Correspondence: ptsonis1@udayton.edu

¹Department of Biology, University of Dayton, Dayton, OH 45469-2320, USA
Full list of author information is available at the end of the article

contain always certain amino acids increasing the identity. On the other hand, big proteins may not align well and may increase the RMSD. For the present analysis, we chose to set the limits as follows: RMSD to be lower than 3.5 and amino acid identity to be lower than 25% in order to conclude that this pair of proteins has similar structures but dissimilar sequences.

Globin-like fold

Globin-like fold is an all-alpha protein fold normally consisting of six alpha helices [12]. The number of helices can be altered in different families of globin-like proteins. These helices are not randomly distributed in the protein, but they are oriented following standard helix-helix packing rules in order to form a globular structure. Globin-like fold is mostly known from hemoglobins (Figure 1) and myoglobins which play an important role in transferring oxygen to all the tissues of an organism with the help of heme groups which can bind oxygen reversibly. The heme-binding proteins are part of the actual family of globins [12].

The globin family was the first example that showed structural conservation even in different organisms [14-17] and led scientists' pursuit to prove that 3D structures of proteins are more conserved than their sequences. It turned out that globin-like folds exist in many proteins with different functions. Hemoglobins and myoglobins play a role in oxygen transport; cyanoglobins [18] bind to oxygen to help in cellular processes; phycocyanins and phycoerythrins [19,20] play a role in absorbing light; cytokines and immuno-globins [21,22] play a role in the immune system; and fibronectin [23] is part of the extracellular matrix. Natural selection kept the 3D structure of this fold intact [24] while utilizing it for different functions to meet other required organismal requirements.

We have compared pairs of functionally different proteins or proteins from organisms that diversified long ago.

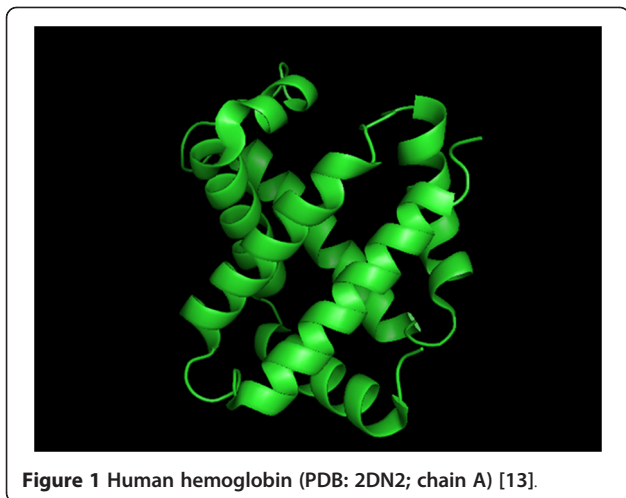


Figure 1 Human hemoglobin (PDB: 2DN2; chain A) [13].

Figure 2 shows the 3D structural conservation despite low sequence similarity. Structure is conserved in a monomeric hemoglobin of a trematode (PDB: 1H97) compared to a hemoglobin which is part of a large protein (3.6 million Da) from an annelid (PDB: 2GTL). In this case, the single hemoglobin from a trematode can bind and transport oxygen. However, it is structurally relevant to hemoglobins that are part of a 3.6 million-Da protein, an erythrocrucorin, which serves the same purpose but has more advantages such as resistance to oxidation and other cooperative binding properties [25,26]. Both proteins are part of the globin-like superfamily [12].

In the next example, structural conservation of a plant hemoglobin (PDB: 2GNW), which may play a role in binding free molecules that cause oxidation, and a globin-coupled sensor (PDB: 2W31), which plays a role in adapting the organism in the presence of oxygen via transmitted signals to a transmembrane protein, can be seen [27,28]. This example demonstrates how a globin-like fold has been used for different kinds of responses from scavenging hazardous active molecules to sense external stimuli and cooperate with other proteins to get the appropriate response. Both proteins are part of the globin-like superfamily [12].

Nitric oxide detoxification in *M. tuberculosis* occurs with the help of a truncated hemoglobin protein (PDB: 2GLN). Its structure is similar to an extracellular giant hemoglobin from an annelid (PDB: 2ZS1) that plays a role in binding oxygen [29,30].

Certain organisms absorb light through pigments. Allophycocyanin is a pigment and its structure is part of the phycobilisome family [12]. This structure (PDB: 1KN1) is similar to a protein that plays a role in regulating the sigma (σ) factor during transcription (PDB: 2BNL) and belongs to the Rsbr_N superfamily (VAST) [31,32]. This is an example of using the globin-like fold as a building block to make a larger structure like the N-domain of the rsbr to serve a different role.

The last example is from two organisms that evolved separately for many millions of years: a neuroglobin (PDB: 1OJ6) from *Homo sapiens* and a protoglobin (PDB: 2VEB) from archaea. The role of globin-like proteins in archaea is not yet fully determined. It is proposed to play a role in metabolism of the strictly anaerobic *M. acetivorans* and to be the building block of globin-coupled sensors. The structure is similar to the neuroglobin from humans which play an important role in regulating oxygen transport in neural tissues [33,34].

Histone fold

This motif is most commonly associated with histones but can also be found in a multitude of proteins such as DNA-binding transcription initiation factors which are functionally conserved in archaea and eukaryotes [35].

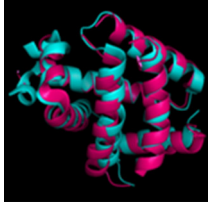

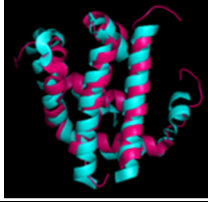
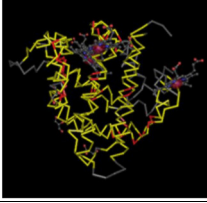

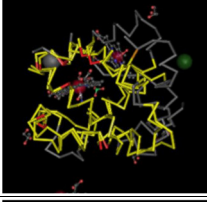
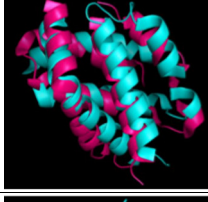
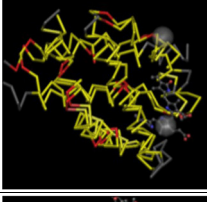


| PDB # - Function | RMSD - % Identity | Images from PyMOL and Cn3D | | References |
|---|------------------------------|--|---|------------|
| Example 1 - 2GTL_Chain A: <i>Lumbricus terrestris</i> (annelide) hemoglobin part of a 3.6million Dalton protein. Transports oxygen. - 1H97_Chain A: <i>Paramphistomum epiclitum</i> (trematode) monomeric hemoglobin. High affinity to oxygen. | RMSD: 2.3 Identity: 12.1% |  |  | 24, 25 |
| Example 2 - 2GNW_Chain B: Found in plants. Its role is not yet determined. <i>Oryza sativa</i> . - 2W31_Chain A: detects oxygen and transmits signal. <i>Geobacter sulfurreducens</i> . | RMSD: 3.2 Identity: 13.4% |  |  | 27, 26 |
| Example 3 - 2GLN_Chain A: nitric oxide scavenging. <i>Mycobacterium tuberculosis</i> . - 2ZS1_Chain A: extracellular giant Hb. Cooperative oxygen binding via inorganic cations. <i>Oligobrachia mashikoi</i> . | RMSD: 2.4 Identity: 6.7% |  |  | 28, 29, |
| Example 4 - 1KN1_Chain A: allophycocyanin, absorbs light, part of phycobilisomes and phycobilisome structural family. <i>Pyropia yezoensis</i> . - 2BNL_Chain C: Non heme, regulates s factor after environmental stress. <i>Bacillus subtilis</i> | RMSD: 2.9 Identity: 11.4% |  |  | 30, 31 |
| Example 5 - 2VEB_Chain A: Found in archae, role is not yet determined. <i>Methanosarcina acetivorans</i> . - 1OJ6_Chain A: A neuroglobin found in human brain. Binds to oxygen. <i>Homo sapiens</i> . | RMSD: 2.9 Identity: 12.7% |  |  | 32, 33 |

Figure 2 Comparison of structure and sequence similarity of sample globin-like fold proteins according to PDB number. First column: PDB number and a brief description of the protein. Second column: RMSD and amino acid sequence identity as defined by VAST. Third column: Left is the alignment of the two proteins taken by PyMOL. In the structure representation, the first protein is in pink, and the second, in cyan. Right is the alignment of the two proteins taken by Cn3D. In the sequence representation, red indicates the same amino acid, whereas yellow indicates differing amino acids. Fourth column: references.

Because of this functional conservation in archaea and eukaryotes, the histone fold is thought to be an ancient motif [36]. Interestingly, the pure functionality of the histone fold is not found in eubacteria [37]. As seen in Figure 3, the basic structure of the histone fold comprises a central alpha helix flanked on each side by two smaller helices.

Due to the hydrophobic nature of the histone fold, it is only stable within histone fold-to-histone fold dimers. Eukaryotic histones, for example, dimerize specifically with H2A dimerizing with H2B and H3 dimerizing with H4, thereby creating the basis of the histone octamer. Archaea histones appear to have less specificity in

dimerizing to a specific partner but, through dimerization, utilize the histone fold to produce a similar histone structure [38,39].

Since the function of histones and the histone fold are shared by archaea and eukarya, it is thought to have been derived from an early thermophile which initially utilized the histone fold to maintain the integrity of DNA under thermal stress. This increased integrity would have also brought about the added benefit of genome compaction which would have required a mechanism to unwind and transcribe those genes and thus the appearance of proteins such as TATA box-binding proteins and transcription initiation factors which also

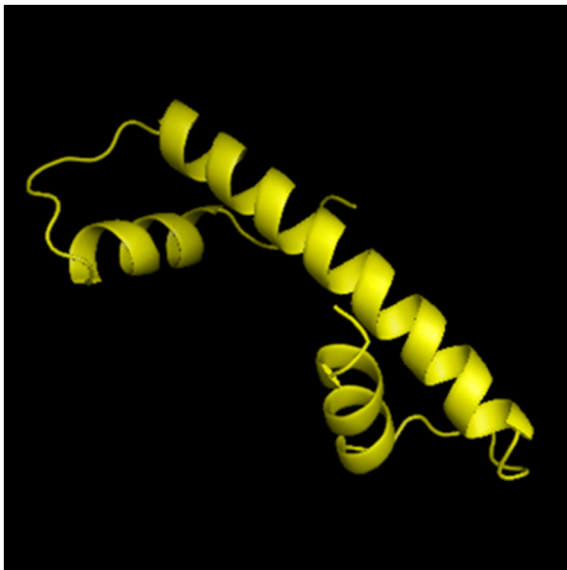


Figure 3 The typical histone fold. It consists of one central helix flanked on each side by a shorter alpha helix (PDB: 1HTA) [38].

utilize the histone fold and are functionally conserved in both eukaryotic and archaea organisms [35,40,41]. Since the packaging and protection of DNA is paramount along with the ability to transcribe DNA when needed, the numerous essential interactions have caused the histone fold to be conserved [42].

From a molecular point of view, the histone fold is thought to have evolved from the helix-strand-helix (HSH) motif where duplication caused two helices to merge, forming a larger central helix [36,43]. Alva et al. demonstrated how this could occur by shortening the HSH strand which led to a 3D swap and caused the dimerization of two HSH motifs. This dimerization recovered the interactions between the HSH motifs due to the strand shortening and thereby causing the histone fold [43].

As mentioned previously, eubacteria do not appear to contain the histone fold motif. They do, however, contain proteins which have histone-like proteins. The most ubiquitous of these proteins is the HU protein (H for histone-like and U from the U93 strain of *Escherichia coli*, in which it was identified from). HU proteins are essential in maintaining the nucleoid structure and are involved in all DNA-dependent functions [44]. Interestingly, the HSH-type motif is found in HU proteins of eubacteria which also have histone-like functionality [42,45]. Looking at the structures of HU and the histone fold (Figure 4), one can easily identify similarities in the HSH with respect to the histone fold, thereby showing how the functionality of DNA binding has been conserved through different but similar means.



Figure 4 Comparison of the histone fold (PDB: 1HTA) [38] to eubacteria HU protein (PDB: 1MUL) [46]. Hot pink: histone fold, cyan: eubacteria HU protein. Notice the similarity in the helix-turn-helix and the size difference in the central helix.

Interestingly, some proteins have evolved a method to overcome the need of the dimerization of different proteins through a double histone fold. A double histone fold is essentially two histone folds occurring in a single peptide chain which can 'dimerize' with itself [47]. As seen in Figure 5, a great structural similarity between the H2A/H2B two-protein dimer has a great structural similarity to the single-protein Son of sevenless (Sos) protein [48,49]. With the double histone fold being so 'economical' by not needing to dimerize with another protein, it is not surprising that it was recently found in a virus where it is hypothesized to aid in the packaging and organization of DNA inside the capsid [50].

Due to the multiple interactions required of the histone fold, the selective pressures limit a large differentiation in sequence identity. For example, H3 and H4 histones are among the most highly conserved proteins in terms of sequence and length due to their specific interactions with DNA. H2A and H2B have regions which show greater variability but show great specificity to dimerizing with each other. Despite the conservation of the histone fold in the histone structure, these four core eukaryotic histones have little sequence similarity (15–20%) with one another [42]. Interestingly, even proteins such as the histone H2A/H2B and the cytoplasmic hSos [50] (Figure 5, example 4) which show strong structural similarity but do not seem to function as histones or DNA-binding factors still do not stray far from this sequence identity. This sequence similarity is seen in organisms which are obviously so evolutionary distant as archaea and eukaryotes [51,52]

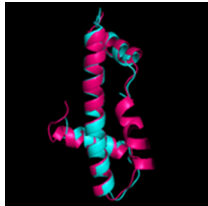
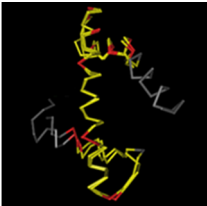
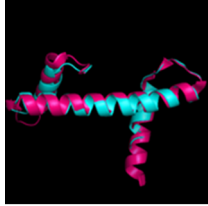
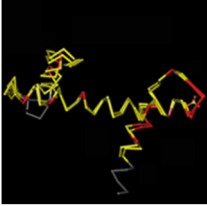
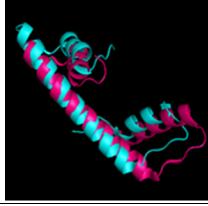
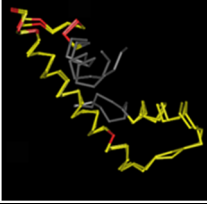
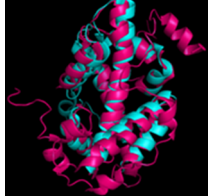
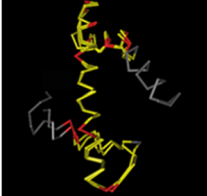
| PDB # - Function | RMSD - % identity | Images from PyMOL and Cn3D | | References |
|---|------------------------------|---|--|------------|
| Example 1 - 1BH8_Chain B: TATA binding protein associated factor (TAF)II28 in <i>Homo sapiens</i> - 1A7W: Histone HmfB from archaea <i>Methanothermus fervidus</i> | RMSD: 1.0 Identity: 22.6% |  |  | 37, 69 |
| Example 2 - 3R45_Chain A: Histone H3-like centromeric protein A in <i>Homo sapiens</i> - 1B6W: Histone HmfA from archaea <i>Methanothermus fervidus</i> | RMSD: 1.1 Identity: 24.6% |  |  | 37, 70 |
| Example 3 - 3AFA_Chain E: Histone 3 in <i>Homo sapiens</i> - 1H3O_Chain D: Transcription Initiation Factor (TFIID) in <i>Homo sapiens</i> | RMSD: 1.0 Identity: 12.3% |  |  | 71, 72 |
| Example 4 - 2JSS: Yeast (<i>Saccharomyces cerevisiae</i>) Histone H2A/H2B - 1Q9C: Double histone fold of <i>Homo sapiens</i> Son of sevenless (hSos) | RMSD: 1.9 Identity: 22.7% |  |  | 73, 74 |

Figure 5 Comparison of structure and sequence similarity of sample histone fold proteins according to PDB number. First column: PDB number and a brief description of the protein. Second column: RMSD and amino acid sequence identity as defined by VAST. Third column: Left is the alignment of the two proteins taken by PyMOL. In the structure representation, the first protein is in *pink*, and the second, in *cyan*. Right is the alignment of the two proteins taken by Cn3D. In the sequence representation, *red* indicates the same amino acid, whereas *yellow* indicates differing amino acids. Fourth column: references.

(Figure 5, examples 1 and 2). This may be due to the hydrophobic residue interactions required in all six helices of a histone fold dimer [39].

Helix-turn-helix motif

HTH motif consists of an α -helix, a turn, and a second α -helix which is often called the 'recognition' helix as the part of the HTH motif that fits into the DNA major groove. There are several positions significant to keep the HTH structure rather than to specify contacts with the DNA, while the amino acid residues in other positions are usually varied to determine the specificity of DNA-protein interactions [53]. This motif is found in many DNA-binding domains and transcriptional factors such as homeotic proteins. This sequence, which is conserved in many organisms for related proteins, was used to discover a large number of DNA-binding proteins [54]

Winged helix-turn-helix (wHTH, Figure 6) shares the same original ancestor as that of HTH in evolutionary history; it is also a DNA-binding domain which binds to

specific DNA sequences. The wHTH is formed by a three-helix bundle (α_1 , α_2 , α_3) and a three- or four-

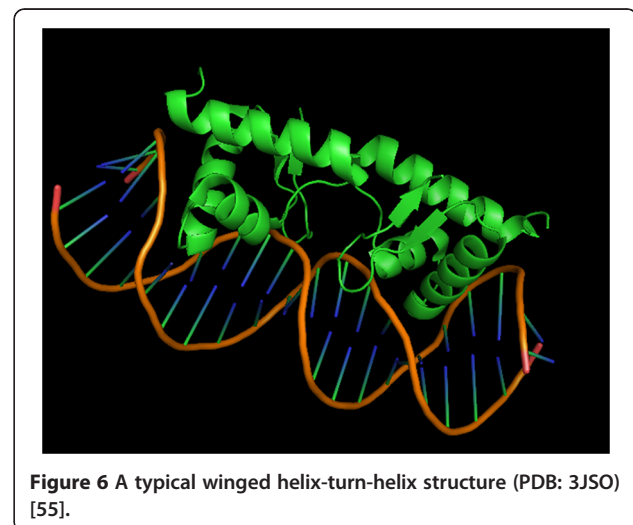


Figure 6 A typical winged helix-turn-helix structure (PDB: 3J50) [55].

strand beta-sheet. The $\alpha 2$ and $\alpha 3$ helices are similar to those of the HTH motif except that wHTH has beta-sheet wings on the ends of HTH parts. Many repressor DNA-binding domains like LexA, arginine, Rex, ArsR, and MarR form a wHTH structure.

Figure 7 shows five examples of HTH comparisons of different proteins. All of them show high structural similarity and low sequence identity. In addition, the examples compare HTH motifs from different organisms that do different functions.

An ancestral archaea homolog of the N-terminal of the transcription factor II E subunit a (PDB: 1Q1H) [56] folds as a wHTH. This domain has a groove which is negatively charged. Thus, it cannot bind to negatively

charged DNA as in vitro experiments show. Though, it promotes interactions with other proteins. This domain has structural similarities with a catabolite gene activator protein (PDB: 1RUN) [57], a protein that is known to bind DNA. This example clearly illustrates that natural selection chose structures to have different roles than the dominant ones. Cro repressor from the λ phage (PDB: 1D1L) [58] forms a dimer by two antiparallel b-stands in order to bind to DNA. This protein has structural similarities with the bacterial Fis protein (PDB: 3JRH) [59] which binds to DNA with no sequence specificity.

Transcriptional regulators can be triggered to function by different signals from the environment. Signals that

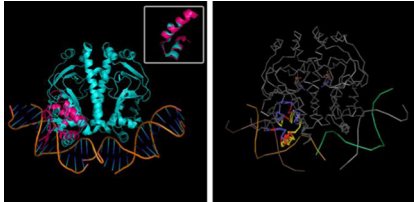
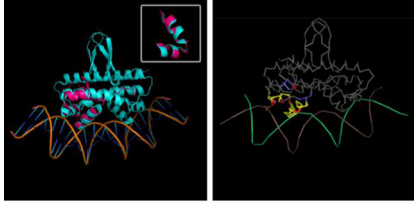
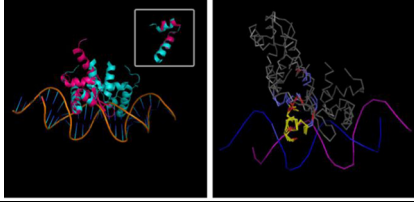
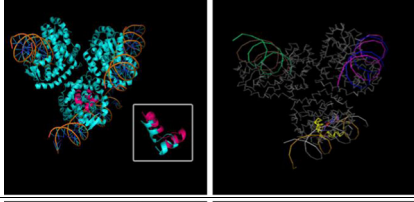
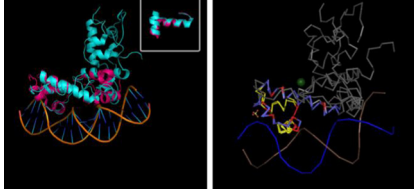
| PDB # - Function | RMSD - % Identity | Images from PyMOL and Cn3D | References |
|--|------------------------------|--|---|
| Example 1 - 1Q1H_Chain A: The alpha subunit of transcription factor TFIIIE homolog from archae <i>Sulfolobus solfataricus</i> . - 1RUN_Chain A: DNA binding domain of <i>Escherichia coli</i> regulatory proteins which belongs to catabolite activator protein family. | RMSD: 1.0 Identity: 22.7% |  | 47, 48 |
| Example 2 - 1D1L_Chain A: Cro repressor of <i>Enterobacteria phage lambda</i> . - 3JRH_Chain B: <i>Escherichia coli</i> protein from Fis family which binds to unspecific DNA. | RMSD: 2.0 Identity: 21.4% |  | 50, 49 |
| Example 3 - 1R1T_Chain A: Negatively allogetic regulated DNA binding of SmtB protein in presence of metals. <i>Synechococcus elongatus</i> . - 1ZLK_Chain A: Hypoxia – induced DosR protein. <i>Mycobacterium tuberculosis</i> . | RMSD: 1.5 Identity: 20.5% |  | 51, 52 |
| Example 4 - 1RES_Chain A: DNA-binding domains of <i>Escherichia coli</i> $\gamma\delta$ resolvase. - 1Z9C_Chain F: Chimeric winged helix-turn-helix (wHTH) DNA-binding domain of OhrR-ohrA complex. <i>Bacillus subtilis</i> . | RMSD: 1.4 Identity: 10.7% |  | 53, 54 |
| Example 5 - 3OIO_Chain A: Bacterial regulatory helix-turn-helix proteins of AraC family from <i>Chromobacterium violaceum</i> . - 1XS9_Chain A: MarA <i>Escherichia coli</i> regulatory helix-turn-helix protein that binds DNA as a monomer. | RMSD: 2.2 Identity: 20.7% |  | ⁵⁵ Chang et al., deposited in PDB, not published |

Figure 7 Comparison of structure and sequence similarity of sample helix-turn-helix motif proteins according to PDB number. First column: PDB number and a brief description of the protein. Second column: RMSD and amino acid sequence identity as defined by VAST. Third column: Left is the alignment of the two proteins taken by PyMOL. In the structure representation, the first protein is in pink, and the second, in cyan. Right is the alignment of the two proteins taken by Cn3D. In the sequence representation, red indicates the same amino acid, whereas yellow indicates differing amino acids. Fourth column: references.

are not related with signal transduction cascades, which involve primarily phosphorylation or dephosphorylation of proteins, can involve smaller molecules like metals or oxygen. This is the case for SmtB (1R1T) [60], a cyanobacterial repressor protein that has reduced affinity for DNA in the presence of metals. The HTH motif of this repressor is structurally similar to the HTH motif of the bacterial DosR protein (PDB: 1ZLK) [61] which prolongs survival when the organism is left without oxygen.

OhrR is a bacterial protein (PDB: 1Z9C) [62] that has a HTH motif composed of eukaryotic-like wHTH, prokaryotic HTH motifs, and other helices. This protein is induced to function by oxidation of certain residues. This chimeric HTH motif is structurally similar to the HTH motif of a DNA-binding domain of a $\gamma\delta$ resolvase in *E. coli* (PDB: 1RES) [63].

Finally, the HTH motif from a bacterial transcriptional regulator, AraC-type (PDB: 3OIO), is structurally similar to that of the transcriptional activator MarA (PDB: 1XS9) [64] which is associated with the RNA polymerase and binds to DNA as a monomer.

Zinc finger motif

Zinc (Zn) fingers (see Figure 8) are small structural motifs whose structure is stabilized by a zinc ion, and they are the most common DNA- or RNA-binding motif in different proteins. There are different structural types of Zn fingers and are present in proteins that perform a broad array of functions such as replication and repair, transcription and translation, metabolism and signaling, cell proliferation, and apoptosis [65]. Zn fingers occupy 3% of the genes in the human genome [66]. The major part of structural stability of Zn fingers is provided by zinc coordination and by the conserved hydrophobic core that flanks the Zn binding site. There are a relatively small number of conserved residues present in Zn fingers [67].

Classical Cys2-His2 (C2H2) Zn fingers have about 30 amino acids in which 25 of the 30 amino acid residues form a loop around the central Zn ion and the 5 other amino acids form the linkers between the consecutive Zn fingers. It consists of two secondary structural units: The first one is an antiparallel beta-sheet, which contains the loop formed by the two cysteines, and the second one is an alpha helix containing the His-His. These two structural units are held together by the zinc atom. The Zn ion tetrahedrally coordinates to the conserved pairs of cysteines and histidines, and this coordination is vital for the maintenance of the overall structure of the Zn finger. The majority of the 30 amino acids are polar and basic residues which are important in nucleic acid binding. In addition to the conserved cysteines and histidines which are vital for the formation of the Zn finger fold, there are other conserved amino acids, notably Tyr, Phe, and Leu, which form a hydrophobic structural core of the folded structure [66].

In the example shown in Figure 9, each pair of the compared Zn fingers have less sequence similarity, sometimes bind to different types of molecules, may have different functions, may belong to different species, but exhibit a great structural overlap. This supports the notion that only few small numbers of conserved residues are required for the maintenance of the overall structure of the zinc finger.

Example 1 in Figure 9 shows two DNA-binding proteins: a DNA-binding domain (DBD) from the GAGA factor (PDB: 1YUJ) [69] and one of the zinc finger domains from zinc finger protein 692 (PDB: 2D9H), which belong to *D. melanogaster* and *H. sapiens*, respectively. The DBD of the GAGA factor uses only one zinc finger in contrast to other zinc finger proteins which commonly use more than two in order to have a good affinity for the DNA. They show a great structural similarity despite low sequence identity.

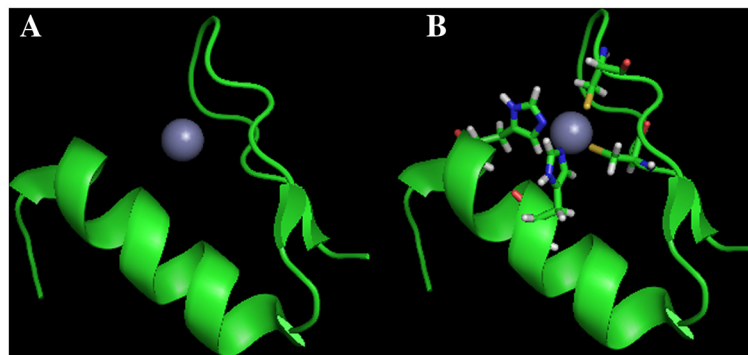


Figure 8 Structure of C2H2 zinc finger of transcription factor IIIA of *Xenopus laevis* (PDB: 2HGH, [68]). (A) Cartoon representation with zinc as a ball. (B) Includes the two cysteines and two histidines that interact with the zinc as sticks.


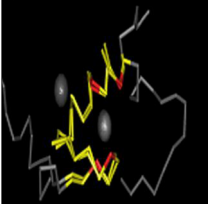

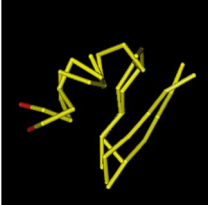

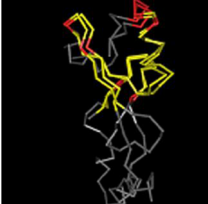
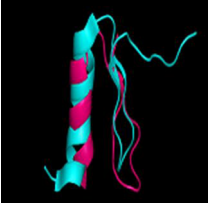
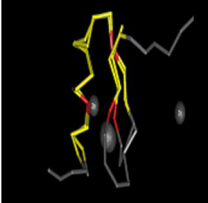

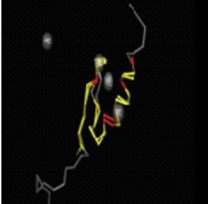
| PDB # - Function | RMSD - % Identity | Images from PyMOL and Cn3D | | References |
|--|------------------------------|--|---|---|
| Example 1 - 1YUJ_Chain A: structure of DNA binding domain of the GAGA factor(DNA binding protein) <i>Drosophila melanogaster</i> - 2D9H_Chain A: forth and fifth zf-C2H2 domains of zinc finger protein 692(DNA binding protein). <i>Homo sapiens</i> | RMSD: 0.8 Identity:16.0% |  |  | ⁵⁹ Zhang et al., deposited in PDB, not published |
| Example 2 - 3ODC_Chain B: Human zinc finger 2. Binds to DNA. <i>Homo sapiens</i> . - 1MHZ_Chain D: hydroxylase component of methane monooxygenase. No evidence that binds to DNA. <i>Methylosinus trichosporium</i> . | RMSD: 1.5 Identity: 3% |  |  | ^{60, 61} |
| Example 3 - 1MHZ_Chain D: hydroxylase component of methane monooxygenase <i>Methylosinus trichosporium</i> - 2INC_Chain A: Native Toluene/o-xylene Monooxygenase Hydroxylase <i>Pseudomonas stutzeri</i> | RMSD: 2.4 Identity: 21.0% |  |  | ^{60, 62} |
| Example 4 - 1U85_Chain A: tryptophan-containing CCHH zinc finger-Kruppel like factor 3-DNA binding protein. <i>Mus musculus</i> - 1UBD_Chain C: Human YY1 zinc finger domain which binds to DNA. <i>Homo sapiens</i> . | RMSD: 1.0 Identity: 15.8% |  |  | Cram et al., deposited in PDB, not published ⁶³ |
| Example 5 - 2VY4_Chain A: A splicing protein which Binds the 5 Splice Site of U12-Type intron. <i>Homo sapiens</i> . - 3MHH_Chain C: Transcriptional coactivator complex;has multiple roles on role in deubiquitination of histone H2B. <i>Saccharomyces cerevisiae</i> . | RMSD: 1.2 Identity: 23.1% |  |  | ^{64, 75} |

Figure 9 Comparison of structure and sequence similarity of sample zinc finger motif proteins according to PDB number. First column: PDB number and a brief description of the protein. Second column: RMSD and amino acid sequence identity as defined by VAST. Third column: Left is the alignment of the two proteins taken by PyMOL. In the structure representation, the first protein is in pink, and the second, in cyan. Right is the alignment of the two proteins taken by Cn3D. In the sequence representation, red indicates the same amino acid, whereas yellow indicates differing amino acids. Fourth column: references.

The hydroxylase domain from methane monooxygenase (PDB: 1MHZ) [70] contains a Zn finger which does not bind to DNA. Though, it is structurally very similar (RMSD: 1.5) and their sequence is very different (3%) from the human Zn finger 2 which binds to DNA (PDB: 3ODC) [71]. This is a good example to point out that structures are built up from extensively used raw materials (domains) like the Zn finger even if they are not going to be used as the majority of the other proteins in which these domains are found.

In the third example, and as a follow up from the previous one, the two proteins are monooxygenases

(PDB: 1MHZ, 2INC) [70,72] which belong to different species and have Zn finger domains whose structures overlap.

YY1 (PDB: 1UBD) [73] is a protein with four Zn fingers and is structurally similar to kruppel-like factor 3 (PDB: 1U85), which contains a Zn finger with tryptophan as shown in the fourth example.

Finally, the Zn finger in U11/U12 (PDB: 2VY4) [74], which is a RNA-binding protein, has a good structural overlap with SAGA protein (PDB: 3MHH) [75], which is a DNA-binding protein, in spite of the low sequence similarity. In addition, the role of SAGA is to deubiquitinate H2B

histone, so the affinity for DNA helps to dock to the nucleosome. This example was selected because these two different proteins bind to two different types of nucleic acids, have different functions, have low sequence identity, but exhibit a good overall structural similarity.

Concluding remarks

In this review, we have selected four protein motifs, which are present in several DNA-binding proteins and in oxygen-carrying and -transporting proteins. Using several comparisons, we show that these motifs exhibit an astonishing degree of structural conservation even though their primary sequence is not similar and even when they are involved in different functions. The examples underscore the importance of structure selection in evolution and a strategy of economy that nature is implementing. Much is to be learned when similar structures have evolved despite unrelated function. It will be interesting to determine how such similar structures have evolved and what could the possible ancestors be. Eventually, when all structures have been solved, evolution of protein structure will provide valuable information on protein function in general.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PAT conceived the idea, analyzed the data, and co-wrote the paper. KS, CEH, JC, and BS performed the search and analysis. KS co-wrote the paper. All authors read and approved the final manuscript.

Author details

¹Department of Biology, University of Dayton, Dayton, OH 45469-2320, USA.

²Department of Chemical Engineering, University of Dayton, Dayton, OH 45469-2320, USA. ³Department of Chemistry, University of Dayton, Dayton, OH 45469-2320, USA.

Received: 10 May 2012 Accepted: 14 May 2012

Published: 2 August 2012

References

- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczyk J, Levine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, et al: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860–921.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, et al: **The sequence of the human genome.** *Science* 2001, **291**:1304–1351.
- Illergard K, Ardell DH, Elofsson A: **Structure is three to ten times more conserved than sequence—a study of structural response in protein cores.** *Proteins* 2009, **77**:499–508.
- Gibrat JF, Madej T, Bryant SH: **Surprising similarities in structure comparison.** *Curr Opin Struct Biol* 1996, **6**:377–385.
- Madej T, Gibrat JF, Bryant SH: **Threading a database of protein cores.** *Proteins* 1995, **23**:356–369.
- Ohkawa H, Ostell J, Bryant S: **MMDb: an ASN.1 specification for macromolecular structure.** *Proc Int Conf Intell Syst Mol Biol* 1995, **3**:259–267.
- Hogue CW: **Cn3D: a new generation of three-dimensional molecular structure viewer.** *Trends Biochem Sci* 1997, **22**:314–316.
- Hogue CW, Ohkawa H, Bryant SH: **A dynamic look at structures: WWW-Entrez and the Molecular Modeling Database.** *Trends Biochem Sci* 1996, **21**:226–229.
- Schuler GD, Epstein JA, Ohkawa H, Kans JA: **Entrez: molecular biology database and retrieval system.** *Methods Enzymol* 1996, **266**:141–162.
- Schrodinger LLC: *The PyMOL Molecular Graphics System*: Version; 2006:99.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The Protein Data Bank.** *Nucleic Acids Res* 2000, **28**:235–242.
- Murzin AG, Brenner SE, Hubbard T, Chothia C: **SCOP: a structural classification of proteins database for the investigation of sequences and structures.** *J Mol Biol* 1995, **247**:536–540.
- Park SY, Yokoyama T, Shibayama N, Shiro Y, Tame JR: **1.25 Å resolution crystal structures of human haemoglobin in the oxy, deoxy and carbonmonoxy forms.** *J Mol Biol* 2006, **360**:690–701.
- Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC: **A three-dimensional model of the myoglobin molecule obtained by x-ray analysis.** *Nature* 1958, **181**:662–666.
- Kendrew JC, Dickerson RE, Strandberg BE, Hart RG, Davies DR, Phillips DC, Shore VC: **Structure of myoglobin: a three-dimensional Fourier synthesis at 2 Å resolution.** *Nature* 1960, **185**:422–427.
- Perutz MF, Muirhead H, Cox JM, Goaman LC: **Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 Å resolution: the atomic model.** *Nature* 1968, **219**:131–139.
- Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North AC: **Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis.** *Nature* 1960, **185**:416–422.
- Trent JT 3rd, Kundu S, Hoy JA, Hargrove MS: **Crystallographic analysis of synechocystis cyanoglobin reveals the structural changes accompanying ligand binding in a hexacoordinate hemoglobin.** *J Mol Biol* 2004, **341**:1097–1108.
- Ficner R, Lobeck K, Schmidt G, Huber R: **Isolation, crystallization, crystal structure analysis and refinement of B-phycoerythrin from the red alga *Porphyridium sordidum* at 2.2 Å resolution.** *J Mol Biol* 1992, **228**:935–950.
- Schirmer T, Bode W, Huber R, Sidler W, Zuber H: **X-ray crystallographic structure of the light-harvesting biliprotein C-phycoyanin from the thermophilic cyanobacterium *Mastigocladus laminosus* and its resemblance to globin structures.** *J Mol Biol* 1985, **184**:257–277.
- Rozwarski DA, Gronenborn AM, Clore GM, Bazan JF, Bohm A, Wlodawer A, Hatada M, Karplus PA: **Structural comparisons among the short-chain helical cytokines.** *Structure* 1994, **2**:159–173.
- Williams AF, Barclay AN: **The immunoglobulin superfamily—domains for cell surface recognition.** *Annu Rev Immunol* 1988, **6**:381–405.
- Bork P, Doolittle RF: **Proposed acquisition of an animal protein domain by bacteria.** *Proc Natl Acad Sci USA* 1992, **89**:8990–8994.
- Aronson HE, Royer WE Jr, Hendrickson WA: **Quantification of tertiary structural conservation despite primary sequence drift in the globin fold.** *Protein Sci* 1994, **3**:1706–1711.
- Pesce A, Dewilde S, Kiger L, Milani M, Ascenzi P, Marden MC, Van Hauwaert ML, Vanfleteren J, Moens L, Bolognesi M: **Very high resolution structure of a trematode hemoglobin displaying a TyrB10-TyrE7 heme distal residue pair and high oxygen affinity.** *J Mol Biol* 2001, **309**:1153–1164.
- Royer WE Jr, Sharma H, Strand K, Knapp JE, Bhyravbhatla B: **Lumbricus erythrocrucorin at 3.5 Å resolution: architecture of a megadalton respiratory complex.** *Structure* 2006, **14**:1167–1177.
- Pesce A, Thijs L, Nardini M, Desmet F, Sisinni L, Gourlay L, Bolli A, Coletta M, Van Doorslaer S, Wan X, Alam M, Ascenzi P, Moens L, Bolognesi M, Dewilde S: **HisE11 and HisF8 provide bis-histidyl heme hexa-coordination in the globin domain of *Geobacter sulfurreducens* globin-coupled sensor.** *J Mol Biol* 2009, **386**:246–260.
- Smaghe BJ, Kundu S, Hoy JA, Halder P, Weiland TR, Savage A, Venugopal A, Goodman M, Premer S, Hargrove MS: **Role of phenylalanine B10 in plant nonsymbiotic hemoglobins.** *Biochemistry* 2006, **45**:9735–9745.
- Numoto N, Nakagawa T, Kita A, Sasayama Y, Fukumori Y, Miki K: **Structural basis for the heterotropic and homotropic interactions of invertebrate giant hemoglobin.** *Biochemistry* 2008, **47**:11231–11238.
- Ouellet Y, Milani M, Couture M, Bolognesi M, Guertin M: **Ligand interactions in the distal heme pocket of *Mycobacterium tuberculosis* truncated hemoglobin N: roles of TyrB10 and GlnE11 residues.** *Biochemistry* 2006, **45**:8770–8781.

31. Liu JY, Jiang T, Zhang JP, Liang DC: Crystal structure of allophycocyanin from red algae *Porphyra yezoensis* at 2.2-Å resolution. *J Biol Chem* 1999, **274**:16945–16952.
32. Murray JW, Delumeau O, Lewis RJ: Structure of a nonheme globin in environmental stress signaling. *Proc Natl Acad Sci USA* 2005, **102**:17320–17325.
33. Nardini M, Pesce A, Thijs L, Saito JA, Dewilde S, Alam M, Ascenzi P, Coletta M, Ciaccio C, Moens L, Bolognesi M: Archaeal protoglobin structure indicates new ligand diffusion paths and modulation of haem-reactivity. *EMBO Rep* 2008, **9**:157–163.
34. Pesce A, Dewilde S, Nardini M, Moens L, Ascenzi P, Hankeln T, Burmester T, Bolognesi M: Human brain neuroglobin structure reveals a distinct mode of controlling oxygen affinity. *Structure* 2003, **11**:1087–1095.
35. Pereira SL, Reeve JN: Histones and nucleosomes in Archaea and Eukarya: a comparative analysis. *Extremophiles* 1998, **2**:141–148.
36. Gangloff YG, Romier C, Thuault S, Werten S, Davidson I: The histone fold is a key structural motif of transcription factor TFIID. *Trends Biochem Sci* 2001, **26**:250–257.
37. Wong JT, New DC, Wong JC, Hung VK: Histone-like proteins of the dinoflagellate *Cryptothecodinium cohnii* have homologies to bacterial DNA-binding proteins. *Eukaryot Cell* 2006, **2**:646–650.
38. Decanniere K, Babu AM, Sandman K, Reeve JN, Heinemann U: Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon *Methanothermobacter fervidus*. *J Mol Biol* 2000, **303**:35–47.
39. Sandman K, Reeve JN: Archaeal histones and the origin of the histone fold. *Curr Opin Microbiol* 2006, **9**:520–525.
40. Tachiwana H, Kagawa W, Osakabe A, Kawaguchi K, Shiga T, Hayashi-Takanaka Y, Kimura H, Kurumizaka H: Structural basis of instability of the nucleosome containing a testis-specific histone variant human H3T. *Proc Natl Acad Sci USA* 2010, **107**:10454–10459.
41. Werten S, Mitschler A, Romier C, Gangloff YG, Thuault S, Davidson I, Moras D: Crystal structure of a subcomplex of human transcription factor TFIID formed by TATA binding protein-associated factors hTAF4 (hTAF(II)135) and hTAF12 (hTAF(II)20). *J Biol Chem* 2002, **277**:45502–45509.
42. Ramakrishnan V: The histone fold: evolutionary questions. *Proc Natl Acad Sci USA* 1995, **92**:11328–11330.
43. Alva V, Ammelburg M, Soding J, Lupas AN: On the origin of the histone fold. *BMC Struct Biol* 2007, **7**:17.
44. Grove A: Functional evolution of bacterial histone-like HU proteins. *Curr Issues Mol Biol* 2010, **13**:1–12.
45. Oberto J, Drlica K, Rouviere-Yaniv J: Histones, HMG, HU, IHF: Meme combat. *Biochimie* 1994, **76**:901–908.
46. Ramstein J, Hervouet N, Coste F, Zelwer C, Oberto J, Castaing B: Evidence of a thermal unfolding dimeric intermediate for the *Escherichia coli* histone-like HU proteins: thermodynamics and structure. *J Mol Biol* 2003, **331**:101–121.
47. Greco C, Fantucci P, De Gioia L: In silico functional characterization of a double histone fold domain from the *Heliothis zea* virus 1. *BMC Bioinformatics* 2005, **6**(Suppl 4):S15.
48. Sondermann H, Soisson SM, Bar-Sagi D, Kuriyan J: Tandem histone folds in the structure of the N-terminal segment of the ras activator Son of Sevenless. *Structure* 2003, **11**:1583–1593.
49. Zhou Z, Feng H, Hansen DF, Kato H, Luk E, Freedberg DI, Kay LE, Wu C, Bai Y: NMR structure of chaperone Chz1 complexed with histones H2A.Z-H2B. *Nat Struct Mol Biol* 2008, **15**:868–869.
50. Greco C, Sacco E, Vanoni M, De Gioia L: Identification and in silico analysis of a new group of double-histone fold-containing proteins. *J Mol Model* 2005, **12**:76–84.
51. Birck C, Poch O, Romier C, Ruff M, Mengus G, Lavigne AC, Davidson I, Moras D: Human TAF(II)28 and TAF(II)18 interact through a histone fold encoded by atypical evolutionary conserved motifs also found in the SPT3 family. *Cell* 1998, **94**:239–249.
52. Hu H, Liu Y, Wang M, Fang J, Huang H, Yang N, Li Y, Wang J, Yao X, Shi Y, Li G, Xu RM: Structure of a CENP-A-histone H4 heterodimer in complex with chaperone HJURP. *Genes Dev* 2011, **25**:901–906.
53. Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM: The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* 2005, **29**:231–262.
54. Pabo CO, Sauer RT: Protein-DNA recognition. *Annu Rev Biochem* 1984, **53**:293–321.
55. Zhang AP, Pigli YZ, Rice PA: Structure of the LexA-DNA complex and implications for SOS box measurement. *Nature* 2010, **466**:883–886.
56. Meinhart A, Blobel J, Cramer P: An extended winged helix domain in general transcription factor E/IE alpha. *J Biol Chem* 2003, **278**:48267–48274.
57. Parkinson G, Gunasekera A, Vojtechovsky J, Zhang X, Kunkel TA, Berman H, Ebright RH: Aromatic hydrogen bond in sequence-specific protein DNA recognition. *Nat Struct Biol* 1996, **3**:837–841.
58. Rupert PB, Mollah AK, Mossing MC, Matthews BW: The structural basis for enhanced stability and reduced DNA binding seen in engineered second-generation Cro monomers and dimers. *J Mol Biol* 2000, **296**:1079–1090.
59. Stella S, Cascio D, Johnson RC: The shape of the DNA minor groove directs binding by the DNA-bending protein Fis. *Genes Dev* 2010, **24**:814–826.
60. Eicken C, Pennella MA, Chen X, Koshlap KM, VanZile ML, Sacchetti JC, Giedroc DP: A metal-ligand-mediated intersubunit allosteric switch in related SmtB/ArsR zinc sensor proteins. *J Mol Biol* 2003, **333**:683–695.
61. Wisedchaisri G, Wu M, Rice AE, Roberts DM, Sherman DR, Hol WG: Structures of *Mycobacterium tuberculosis* DosR and DosR-DNA complex involved in gene activation during adaptation to hypoxic latency. *J Mol Biol* 2005, **354**:630–641.
62. Hong M, Fuangthong M, Helmann JD, Brennan RG: Structure of an OhrR-OhrA operator complex reveals the DNA binding mechanism of the MarR family. *Mol Cell* 2005, **20**:131–141.
63. Liu T, DeRose EF, Mullen GP: Determination of the structure of the DNA binding domain of gamma delta resolvase in solution. *Protein Sci* 1994, **3**:1286–1295.
64. Dangi B, Gronenborn AM, Rosner JL, Martin RG: Versatility of the carboxy-terminal domain of the alpha subunit of RNA polymerase in transcriptional activation: use of the DNA contact site as a protein contact site for MarA. *Mol Microbiol* 2004, **54**:45–59.
65. Krishna SS, Majumdar I, Grishin NV: Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res* 2003, **31**:532–550.
66. Klug A: The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu Rev Biochem* 2010, **79**:213–231.
67. Wolfe SA, Nekludova L, Pabo CO: DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct* 2000, **29**:183–212.
68. Lee BM, Xu J, Clarkson BK, Martinez-Yamout MA, Dyson HJ, Case DA, Gottesfeld JM, Wright PE: Induced fit and "lock and key" recognition of 5 S RNA by zinc fingers of transcription factor IIIA. *J Mol Biol* 2006, **357**:275–291.
69. Omichinski JG, Pedone PV, Felsenfeld G, Gronenborn AM, Clore GM: The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. *Nat Struct Biol* 1997, **4**:122–132.
70. Elango N, Radhakrishnan R, Froland WA, Wallar BJ, Earhart CA, Lipscomb JD, Ohlendorf DH: Crystal structure of the hydroxylase component of methane monooxygenase from *Methylosinus trichosporium* OB3b. *Protein Sci* 1997, **6**:556–568.
71. Langelier MF, Planck JL, Roy S, Pascal JM: Crystal structures of poly(ADP-ribose) polymerase-1 (PARP-1) zinc fingers bound to DNA: structural and functional insights into DNA-dependent PARP-1 activity. *J Biol Chem* 2011, **286**:10690–10701.
72. McCormick MS, Szazinsky MH, Condon KL, Lippard SJ: X-ray crystal structures of manganese(II)-reconstituted and native toluene/o-xylene monooxygenase hydroxylase reveal rotamer shifts in conserved residues and an enhanced view of the protein interior. *J Am Chem Soc* 2006, **128**:15108–15110.
73. Houbavijy HB, Usheva A, Shenk T, Burley SK: Cocystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc Natl Acad Sci USA* 1996, **93**:13577–13582.
74. Tidow H, Andreeva A, Rutherford TJ, Fersht AR: Solution structure of the U11-48 K CHHC zinc-finger domain that specifically binds the 5' splice site of U12-type introns. *Structure* 2009, **17**:294–302.
75. Samara NL, Datta AB, Berndsen CE, Zhang X, Yao T, Cohen RE, Wolberger C: Structural insights into the assembly and function of the SAGA deubiquitinating module. *Science* 2010, **328**:1025–1029.

doi:10.1186/1479-7364-6-10

Cite this article as: Sousounis et al.: Conservation of the three-dimensional structure in non-homologous or unrelated proteins. *Human Genomics* 2012 **6**:10.