Zinc Fingers – Folds for Many Occasions

Jacqueline Matthews and Margaret Sunde

School of Molecular and Microbial Biosciences, University of Sydney, NSW 2006

The zinc finger (ZnF) is probably the most versatile intracellular protein domain and it is certainly a commonly used one. ZnF-containing proteins form one of the most prevalent structural families in eukaryotes, comprising, about 2% of the proteins encoded by the human genome. ZnFs bind one or more zinc ions in a structural (rather than catalytic) manner, through cysteine, histidine and occasionally aspartate side chains. The term was originally applied to repeats of about 30 residues that were discovered in the *Xenopus* transcription factor TFIIIA (1), but the definition has subsequently been extended to many classes of small (less than about 100 residues) zinc-ligating domains.

Different classes of ZnFs differ by structure and function, as well as by the identity and spacing of their zinc-binding residues. There are currently at least 14 different well-characterised classes known, some of which are shown in **Fig. 1**. These different classes have a variety of roles within the cell, but they share the common feature of being able to mediate the interaction of proteins with other biomolecules, including DNA, RNA, other proteins or lipids. Many proteins contain multiple ZnFs, often from different classes, as well as a variety of other protein domains. Thus, a single ZnF protein may bring together a range of different molecules, usually with high specificities and interaction affinities. This article reviews the better-characterised ZnFs according to their structural and functional characteristics.

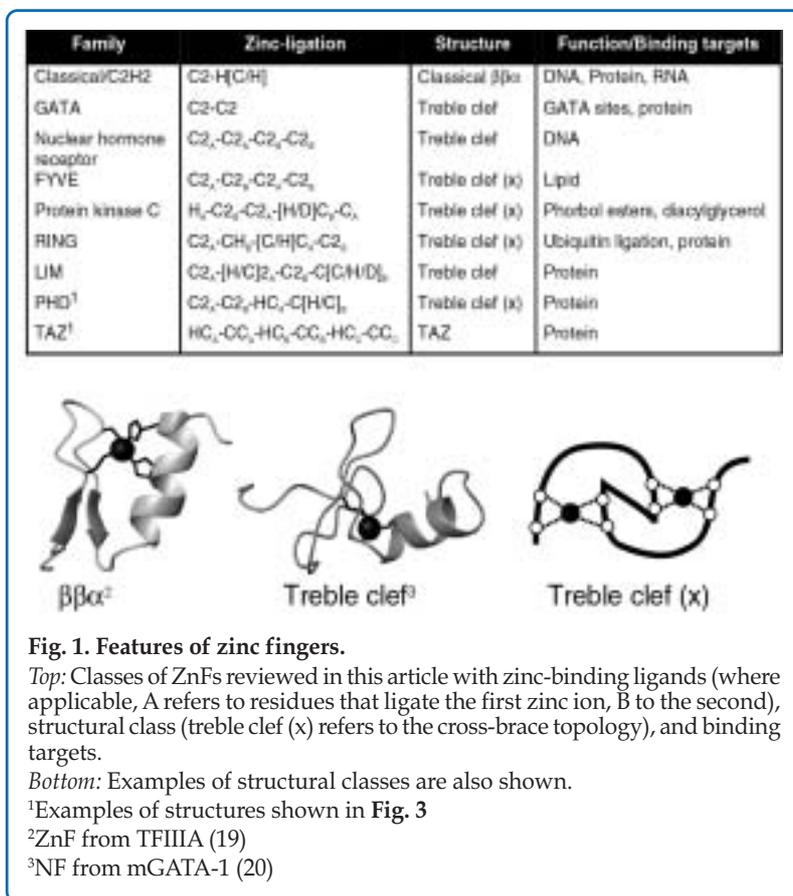## Classical/C2H2 zinc fingers: the original zinc finger

TFIIIA contains classical (or C2H2) ZnFs, which ligate zinc via pairs of cysteine and histidine residues. These domains constitute the majority of ZnFs in the human genome. Many proteins that contain classical ZnFs are involved in the regulation of gene expression, with the ZnFs often interacting with specific DNA sequences in the promoter or enhancer regions of target genes. Basic and hydrophobic residues that are found in the α-helix of these ββα structures appear to be the primary determinants of DNA-binding, making specific contacts with between two and four bases in the major groove of the DNA helix (2). In general, a single classical ZnF is not able to bind DNA, nor would the DNA sequence necessarily be long enough to define a useful transcriptional binding site. Instead, series of three or more C2H2 ZnFs, often separated by a conserved TGEKP linker, recognise DNA motifs in target genes.

The Krüppel-like ZnFs (KLFs, named for the *Drosophila* transcription factor Krüppel) are a subfamily of this type. Members of the Sp/KLF family of mammalian transcription factors possess, at or near their C-termini, three KLFs that bind to GC or CACCC boxes (3). The DNA-binding residues within KLFs vary between different fingers, imparting different sequence specificities. To a certain extent, it is possible to predict the DNA-binding site of a particular KLF on the basis of sequence. Alternatively, it is possible to design DNA-binding proteins with altered specificity by making mutations in a ZnF. In principle, artificial transcription factors comprising series of selected ZnFs can be designed to bind any target DNA sequence, with the view to controlling specific transcriptional events *in vivo* (reviewed in ref. 4).

Classical ZnFs have been shown to interact with RNA and DNA/RNA heteroduplexes with high affinity, although no biological significance of heteroduplex interactions has yet been established (5). One example of this is Wilm's Tumour protein 1 (WT1), a KLF protein that can bind to both DNA and RNA, and is thought to play an important role in the maturation of mRNA (6). Deletions and mutations in *WTI* give rise to an aggressive tumour of the kidney that usually occurs in children.

As well as binding to nucleic acids, some classical fingers can also bind to proteins, often via other zinc fingers. For example, the Ikaros family of transcription factors contain two

Fig. 1. Features of zinc fingers.
Top: Classes of ZnFs reviewed in this article with zinc-binding ligands (where applicable, A refers to residues that ligate the first zinc ion, B to the second), structural class (treble clef (x) refers to the cross-brace topology), and binding targets.
Bottom: Examples of structural classes are also shown.
[1] Examples of structures shown in **Fig. 3**
[2] ZnF from TFIIIA (19)
[3] NF from mGATA-1 (20)

clusters of zinc fingers. The N-terminal cluster, containing four ZnFs, mediates binding to DNA. The C-terminal cluster, containing two ZnFs, mediates protein dimerisation. The latter event can be either homodimerisation or heterodimerisation with all other family members, resulting in different DNA-binding and transcriptional activities. Ikaros-family proteins are important in the regulation of lymphoid cell development (7). Mutant forms of Ikaros, in which only the dimerisation ZnFs are present, can interfere with normal Ikaros function and have been implicated in specific childhood leukemias.

Although classical ZnFs require multiple domains in order to bind DNA, they can mediate protein:protein interactions through a single ZnF. For example, the transcriptional regulator FOG-1 (friend of GATA) contains nine ZnFs. Four of these are C2H2 ZnFs and five are variants in which the final histidine is replaced by cysteine (CCHC ZnFs; ref. 8), but which have a conserved classical structure (9). Most of these CCHC ZnFs in FOG-1 can independently bind to the erythroid transcription factor GATA-1 (**Fig. 2**). At least 50 human proteins contain a single CCHH or CCHC ZnF. It seems likely that these ZnFs mediate protein:protein interactions, rather that contacting DNA, although it is possible that they may interact with other biomolecules.

GATA-type zinc fingers: the multitasking zinc finger

A number of ZnF classes share a common core fold known as the treble clef motif (10): this motif is typified by GATA ZnFs. These ZnFs contain a zinc knuckle (a β-hairpin containing the sequence Cys-x-x-Cys) followed by a β-hairpin and an α-helix. GATA ZnFs are named for their ability to recognise GATA-containing DNA sequences, which are found in the promoter regions of erythroid-specific genes. Despite the fact that most human GATA ZnFs are found as tandem pairs in GATA-family proteins, single GATA-type ZnFs, with an accompanying basic region, are sufficient for DNA-binding.

In the erythroid transcription factor GATA-1, the C-terminal ZnF appears to be the primary determinant of binding to GATA sites. Residues in the α-helix and basic tail region make most of the contacts with DNA. The N-terminal finger appears to contribute to stability of DNA-binding at double GATA sites, but has also been shown recently to bind to variant GATC sites (11). Both the N-fingers and C-fingers of GATA-1 can mediate interactions with a variety of other proteins, including the ZnFs of the ubiquitous transcription factor SP1 and the hematopoietic transcriptional regulator LMO2.

In other cases, the two fingers can instead have differential selectivity for proteins; only the N-finger of GATA interacts with the CCHC ZnFs of FOG-1 (see above) to regulate hematopoiesis (8). GATA proteins seem to have many different and complex roles in regulating transcription that can, in part, be attributed to the different specificities of its ZnFs for binding to both DNA and proteins. What is probably the most remarkable thing about each of these ZnFs is that they can achieve these diverse binding activities and specificities within the framework of only about 30 residues (**Fig. 2**).

At least four separate mutations in the N-terminal ZnF of GATA-1 appear to cause inherited blood disorders, and
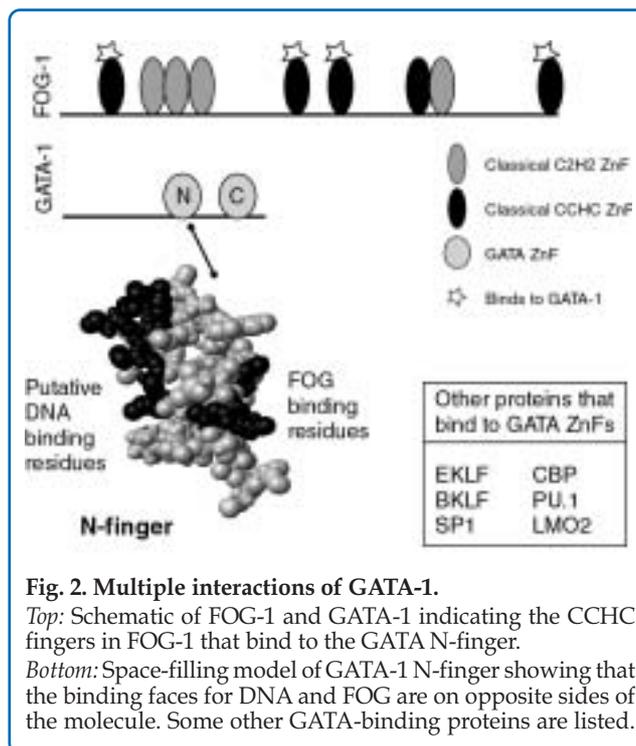

Fig. 2. Multiple interactions of GATA-1.
Top: Schematic of FOG-1 and GATA-1 indicating the CCHC fingers in FOG-1 that bind to the GATA N-finger.
Bottom: Space-filling model of GATA-1 N-finger showing that the binding faces for DNA and FOG are on opposite sides of the molecule. Some other GATA-binding proteins are listed.
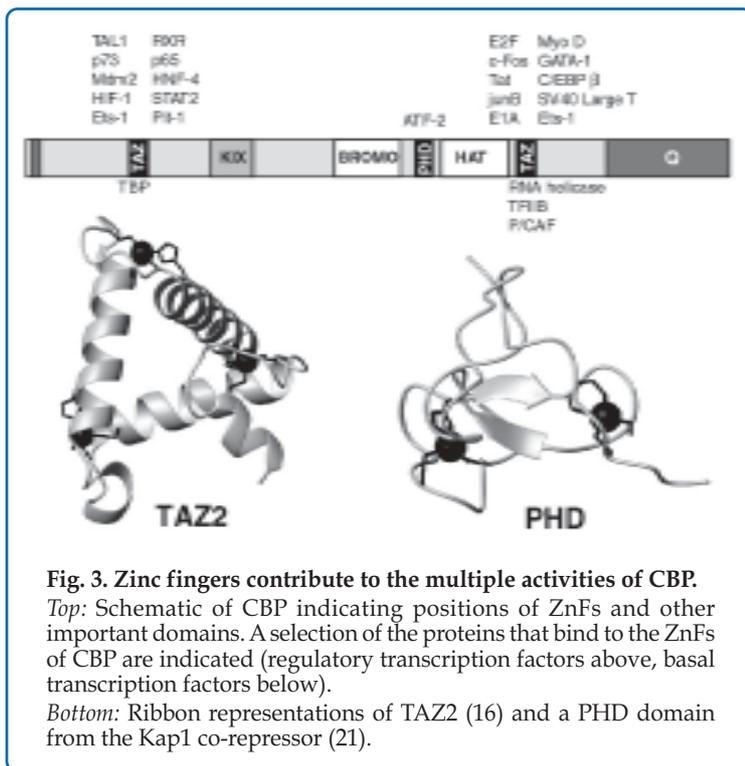
GATA-3 haploinsufficiency has been associated with human hypoparathyroidism, sensorineural deafness and renal anomalies (HDR) syndrome.

Specialized zinc fingers: finding the right zinc finger for the job

Not all ZnFs multitask to the extent of the GATA ZnFs. The nuclear hormone receptor proteins use the more restricted activities of their ZnFs to elicit changes in the expression of target genes. Hormones such as vitamin D, 9-*cis* retinoic acid and thyroid hormone diffuse into the cell and bind to non-ZnF domains on the cognate receptor protein. Loaded receptors are translocated into the nucleus, where the ZnF domains from two hormone receptor molecules dimerise and bind to DNA elements that are either palindromic or made up of directly repeated elements. Hormone receptor ZnFs bind two zinc atoms through two CCCC motifs. Structures of these domains in the presence of DNA show that the α-helix of the N-terminal motif lies in the DNA major groove.

Both homodimers and heterodimers of different receptor molecules can bind to DNA; different dimers bind to half-sites with different spacings (between one and five intervening bases; 12). A number of mutations in different nuclear hormone associated receptors are linked with diseases including familial rickets, early mature onset diabetes, and androgen insensitivity syndrome.

Protein kinase C ZnFs bind small ligands such as phorbol esters and diacylglycerol, which is important in signal transduction. The Protein Kinase C ZnF family of proteins, along with three other families (RING, PHD and FYVE ZnFs) are also members of the treble clef superfamily, but adopt a more complex "cross-brace" topology in which two zinc atoms are bound by alternate pairs of Cys/His ligands (**Fig. 1**). FYVE domains are specific recognition domains for the lipid phosphatidylinositol-3-phosphate

• *Zinc Finger Versatility*

Fig. 3. Zinc fingers contribute to the multiple activities of CBP.
Top: Schematic of CBP indicating positions of ZnFs and other important domains. A selection of the proteins that bind to the ZnFs of CBP are indicated (regulatory transcription factors above, basal transcription factors below).
Bottom: Ribbon representations of TAZ2 (16) and a PHD domain from the Kap1 co-repressor (21).

family, which also contains a PHD domain – a class of ZnF that mediates protein-protein interactions in multi-protein complexes and may be involved with chromatin remodelling (17; **Fig. 3**). The three ZnF domains in CBP/p300 help to give these widely expressed tumour suppressor proteins the ability to interact with a surprisingly large number of transcription factors (18; **Fig. 3**).

In this article, we have only touched on some of the diverse biological roles of ZnFs. It should be clear that they display a wide variety of functionalities that have evolved from a fairly simple, but stable motif of a few residues ligating a metal ion. The large number of different ZnF folds is likely to reflect a convenient mechanism of stabilisation for small domains in a reducing environment where disulfide bonds do not form readily. They are certainly a popular motif in nature, and have evolved to fill many functional niches within cells with many recently discovered roles in addition to our initial, somewhat naive, view of ZnFs as a DNA-binding scaffold.

(PtdIns3P), a component of eukaryotic membranes (13). Proteins that contain a FYVE domain are usually found in cell membranes and have functions in vesicular trafficking, signal transduction and phagocytosis.

RING fingers are relatively common in the human genome and have recently been shown to play roles in some types of protein ubiquitination. It is well established that polyubiquitination targets misfolded proteins for degradation (see *Australian Biochemist*, March 1999); however it is becoming apparent that ubiquitination plays a role in the regulation of other cellular processes. Ubiquitin (Ub) is transferred onto target proteins by E2 ubiquitin-conjugating enzymes, via additional proteins known as E3 Ub ligases. Many E3s are RING-finger proteins, and the RING finger plays a role in the transfer of Ub from the E2 to its target.
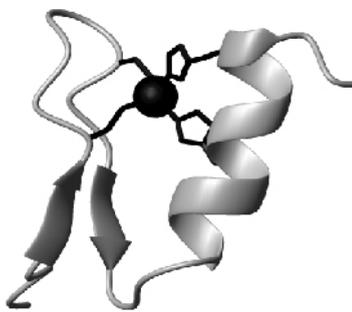
The breast-cancer-associated protein BRCA1 has ubiquitination activity that is markedly enhanced by the binding to one of its many binding partners, BARD. This interaction is mediated by RING domains from both proteins. Other RING-finger proteins have implied roles in the assembly of multi-protein complexes and thereby participate in a wide variety of cellular functions, including signal transduction, transcriptional regulation, immunoglobulin gene rearrangement and DNA repair (14). Several diseases are associated with defects in RING-finger proteins, including familial breast cancers.

There are a number of ZnFs that appear to interact exclusively with proteins (15), such as the LIM domain, reviewed by Susan Brown in the *Australian Biochemist*, April 2001. A recently identified ZnF that is remarkable for the wide range of partner proteins with which it can interact is the TAZ (transcriptional adaptor ZnF) domain. This domain binds three zinc atoms to form a novel triangular domain, where each zinc atom is found in a loop region between two antiparallel a-helices (16; **Fig. 3**). TAZ domains are unique to the CBP/p300 transcriptional regulator
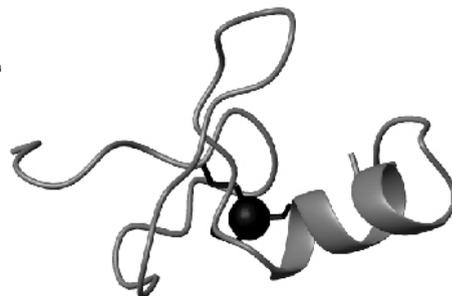
References

1. Miller, J., McLachlan, A.D., and Klug, A. (1985) *EMBO J.* **4**, 1609-1614
2. Choo, Y., and Klug, A. (1997) *Curr. Opin. Struct. Biol.* **7**, 117-125
3. Turner, J., and Crossley, M. (1999) *Trends Biochem. Sci.* **24**, 236-240
4. Choo, Y., and Isalan, M. (2000) *Curr. Opin. Struct. Biol.* **10**, 411-416
5. Shi, Y., and Berg, J.M. (1995) *Science* **268**, 282-284
6. Little, M., Holmes, G., and Walsh, P. (1999) *Bioessays* **21**, 191-202
7. Georgopoulos, K. (2002) *Nature Rev. Immunol.* **2**, 162-174
8. Fox, A.H., Liew, C., Holmes, M., Kowalski, K., Mackay, J., and Crossley, M. (1999) *EMBO J.* **18**, 2812-2822
9. Liew, C.K., Kowalski, K., Fox, A.H., Newton, A., Sharpe, B.K., Crossley, M., and Mackay, J.P. (2000) *Structure Fold. Des.* **8**, 1157-166
10. Grishin, N.V. (2001) *Nucleic Acids Res.* **29**, 1703-1714
11. Newton, A., Mackay, J., and Crossley, M. (2001) *J. Biol. Chem.* **276** , 35794-35801
12. Rastinejad, F. (2001) *Curr Opin Struct Biol* **11**, 33-38
13. Gillooly, D.J., Simonsen, A., and Stenmark, H. (2001) *Biochem. J.* **355**, 249-258
14. Saurin, A.J., Borden, K.L., Boddy, M.N., and Freemont, P.S. (1996) *Trends Biochem. Sci.* **21**, 208-214
15. Mackay, J.P., and Crossley, M. (1998) *Trends Biochem. Sci.* **23**, 1-4
16. De Guzman, R.N., Liu, H.Y., Martinez-Yamout, M., Dyson, H.J., and Wright, P.E. (2000) *J. Mol. Biol.* **303**, 243-253
17. Aasland, R., Gibson, T.J., and Stewart, A.F. (1995) *Trends Biochem. Sci.* **20**, 56-59
18. Vo, N., and Goodman, R.H. (2001) *J. Biol. Chem.* **276**, 13505-13508
19. Foster, M.P., Wuttke, D.S., Radhakrishnan, I., Case, D.A., Gottesfeld, J.M., and Wright, P.E. (1997) *Nat. Struct. Biol.* **4**, 605-608
20. Kowalski, K., Czolij, R., King, G.F., Crossley, M., and Mackay, J.P. (1999) *J. Biomol. NMR* **13**, 249-262
21. Capili, A.D., Schultz, D.C., Rauscher, I.F., and Borden, K.L. (2001) *EMBO J.* **20**, 165-177
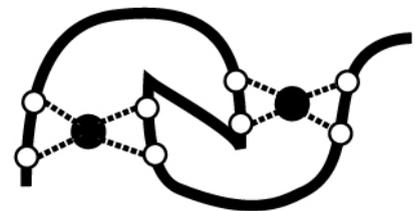
Family	Zinc-ligation	Structure	Function/Binding targets
Classical/C2H2	C2-H[C/H]	Classical $\beta\beta\alpha$	DNA, Protein, RNA
GATA	C2-C2	Treble clef	GATA sites, protein
Nuclear hormone receptor	$C2_A\text{-}C2_A\text{-}C2_B\text{-}C2_B$	Treble clef	DNA
FYVE	$C2_A\text{-}C2_B\text{-}C2_A\text{-}C2_B$	Treble clef (x)	Lipid
Protein kinase C	$H_A\text{-}C2_B\text{-}C2_A\text{-}[H/D]C_B\text{-}C_A$	Treble clef (x)	Phorbol esters, diacylglycerol
RING	$C2_A\text{-}CH_B\text{-}[C/H]C_A\text{-}C2_B$	Treble clef (x)	Ubiquitin ligation, protein
LIM	$C2_A\text{-}[H/C]2_A\text{-}C2_B\text{-}C[C/H/D]_B$	Treble clef	Protein
PHD[1]	$C2_A\text{-}C2_B\text{-}HC_A\text{-}C[H/C]_B$	Treble clef (x)	Protein
TAZ[1]	$HC_A\text{-}CC_A\text{-}HC_B\text{-}CC_B\text{-}HC_C\text{-}CC_C$	TAZ	Protein

$\beta\beta\alpha$[2]   Treble clef[3]   Treble clef (x)

Fig. 1

FOG-1

GATA-1

N C

Putative
DNA
binding
residues

FOG
binding
residues

N-finger

Classical C2H2 ZnF

Classical CCHC ZnF

GATA ZnF

Binds to GATA-1

Other proteins that
bind to GATA ZnFs

EKLF   CBP
BKLF   PU.1
SP1   LMO2

Fig. 2

TAL1   RXR   E2F   Myo D
p73   p65   c-Fos   GATA-1
Mdm2   HNF-4   Tat   C/EBP β
HIF-1   STAT2   junB   SV40 Large T
Ets-1   Pit-1   ATF-2   E1A   Ets-1

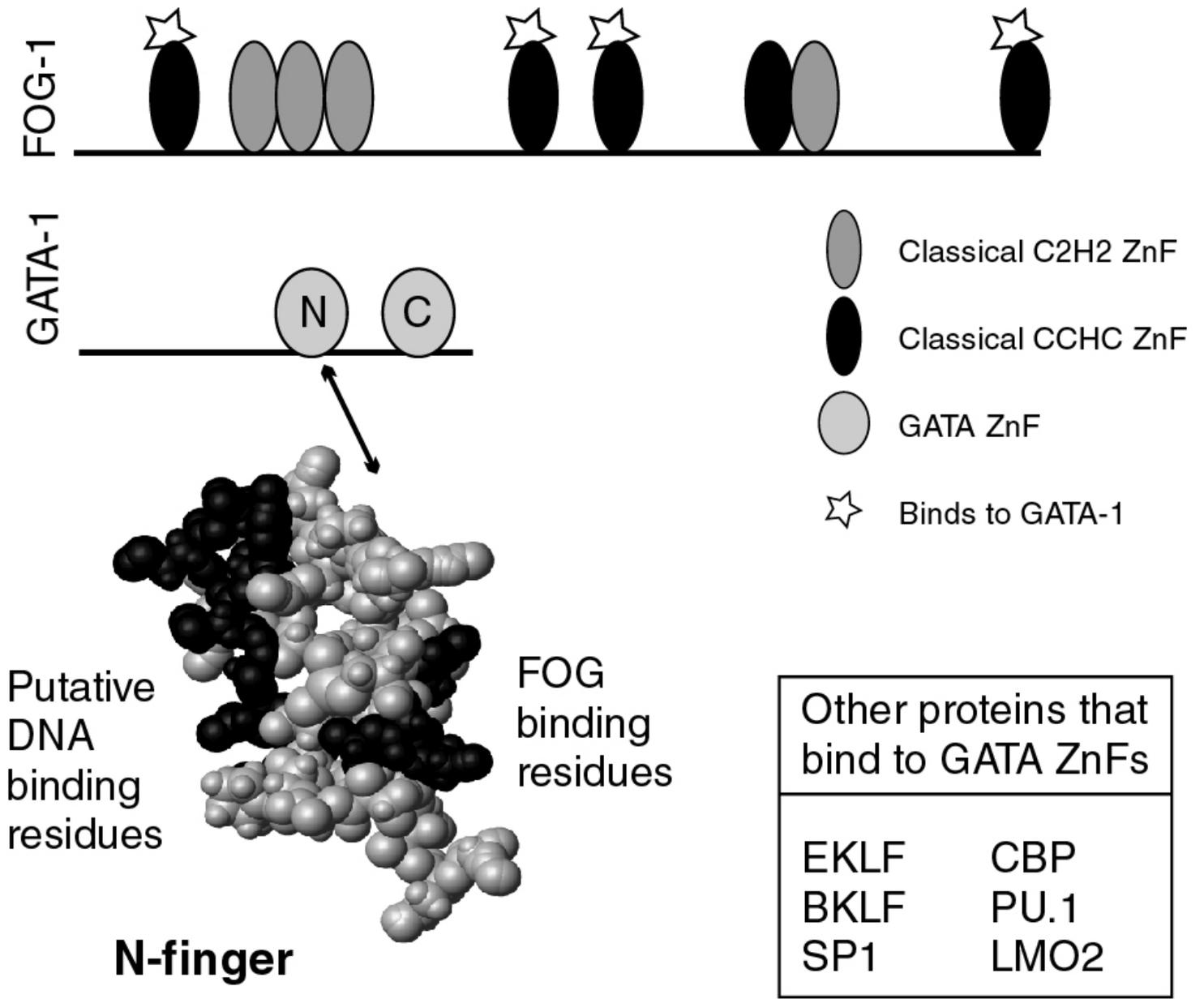| | TAZ | | KIX | | BROMO | PHD | HAT | TAZ | | Q |

TBP   RNA helicase
                                                               TFIIB
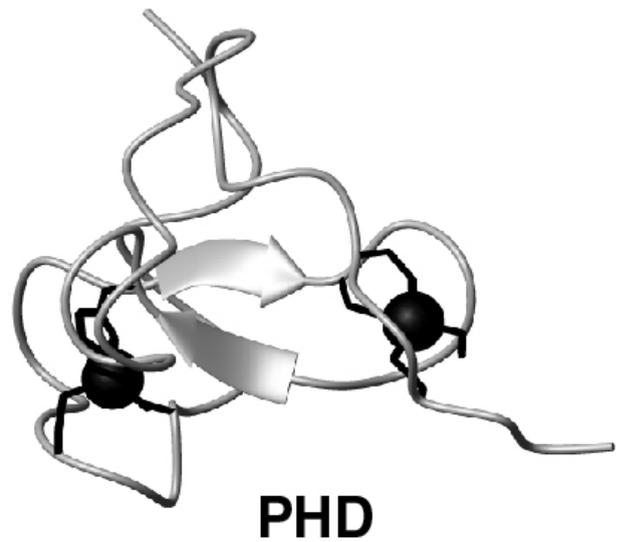   P/CAF

TAZ2   PHD

Fig. 3