

Multiple modes of RNA recognition by zinc finger proteins

Traci M Tanaka Hall

Zinc finger proteins are generally thought of as DNA-binding transcription factors; however, certain classes of zinc finger proteins, including the common C_2H_2 zinc fingers, function as RNA-binding proteins. Recent structural studies of the C_2H_2 zinc fingers of transcription factor IIIA (TFIIIA) and the CCCH zinc fingers of Tis11d in complex with their RNA targets have revealed new modes of zinc finger interaction with nucleic acid. The three C_2H_2 zinc fingers of TFIIIA use two modes of RNA recognition that differ from the classical mode of DNA recognition, whereas the CCCH zinc fingers of Tis11d recognize specific AU-rich sequences through backbone atom interaction with the Watson–Crick edges of the adenine and uracil bases.

Addresses

Laboratory of Structural Biology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA

Corresponding author: Hall, Traci M Tanaka (hall4@niehs.nih.gov)

Current Opinion in Structural Biology 2005, **15**:367–373

This review comes from a themed issue on
Nucleic acids
Edited by Dinshaw J Patel and Eric Westhof

Available online 13th May 2005

0959-440X/\$ – see front matter
© 2005 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.sbi.2005.04.004

Introduction

The classical C_2H_2 ‘zinc finger’ proteins were identified 20 years ago as modular nucleic acid recognition elements [1]. Although most noted for their role as DNA-binding transcription factors, C_2H_2 zinc fingers were identified in transcription factor IIIA (TFIIIA), which was first shown to be the protein component that associates with 5S rRNA within a 7S particle in *Xenopus* oocytes [2]. Later, its DNA-binding properties were suggested when TFIIIA was shown to regulate expression of the 5S rRNA gene [3]. More recently, the CCCH class of zinc finger protein has been shown to bind to RNA targets [4]. This review will focus on recent structural work illuminating our understanding of C_2H_2 and CCCH zinc finger protein interaction with RNA. Readers may also be interested in the interaction of the CCHC zinc knuckle domain of a retroviral nucleocapsid protein with a model RNA representing the packaged dimeric RNA genome. These results, from D’Souza and Summers [5[•]], suggest how

the nucleocapsid protein may interact with RNA bases that become unpaired when the genomic DNA forms dimers.

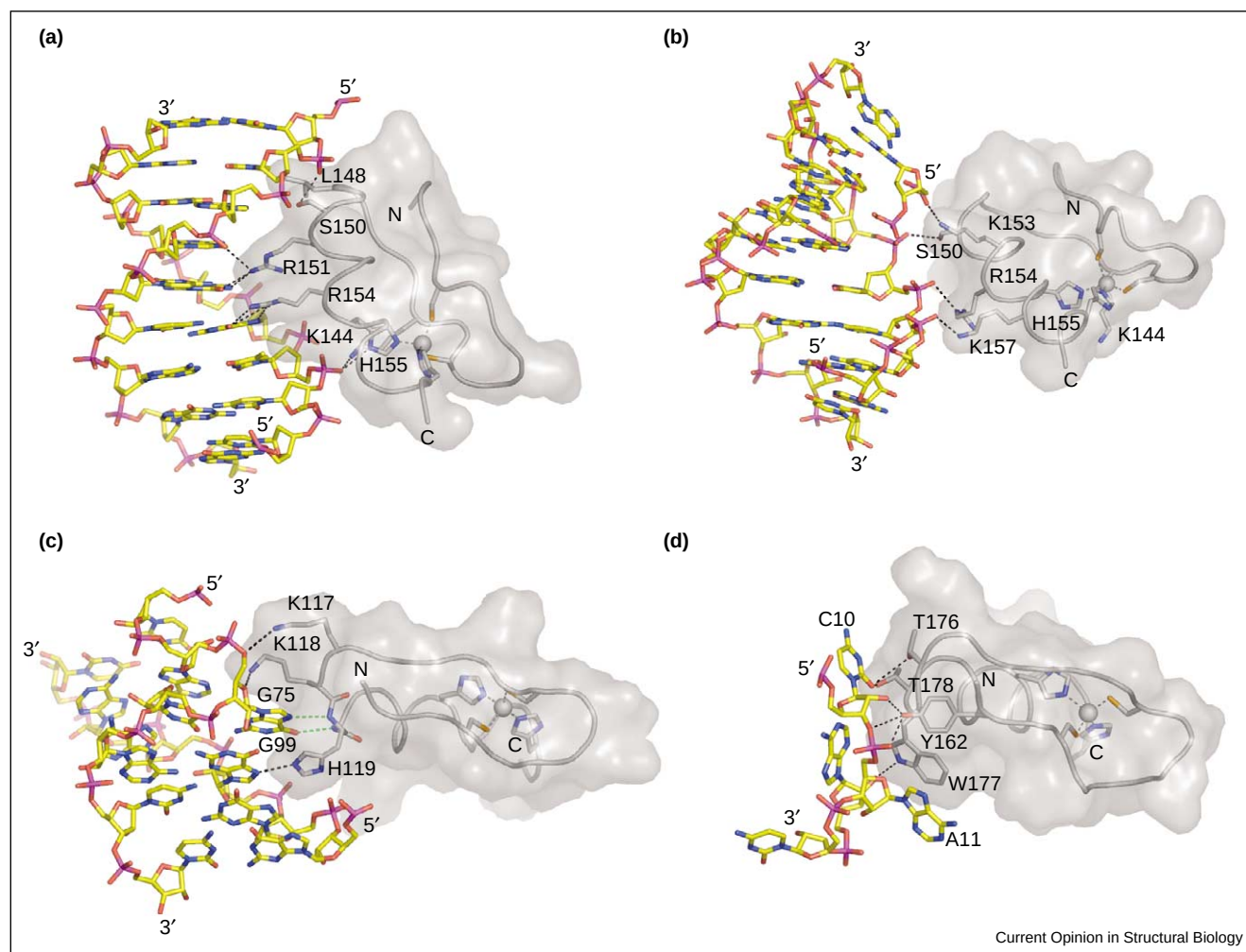
RNA binding by C_2H_2 zinc finger proteins

TFIIIA in *Xenopus* oocytes acts as an essential RNA polymerase III transcription factor for the expression of the 5S rRNA gene, binds to 5S rRNA to form a 7S ribonucleoprotein particle that stabilizes the RNA until it is needed for ribosome assembly and facilitates nuclear export of the 5S rRNA [2,3,6]. TFIIIA contains nine C_2H_2 zinc fingers, which are used to recognize RNA and DNA targets — 5S rRNA and the 5S rRNA gene [1]. The C_2H_2 zinc finger is a module of approximately 30 amino acid residues, with two cysteine and two histidine residues that coordinate a zinc ion. It folds into a small domain comprising two β strands followed by an α helix [7].

Biochemical and structural studies have shown that binding to the 5S rRNA gene internal control region utilizes all but the fourth and sixth zinc fingers of TFIIIA (see, for example, [8–11]). TFIIIA binds to three elements within the 5S rRNA gene internal control region: an 11 base pair ‘box A’ sequence, a 3 base pair ‘intermediate element’ (IE) sequence and a 10 base pair ‘box C’ sequence (Figure 1a). These elements are distributed pseudo-symmetrically, with approximately one double-helical turn between each of the elements. Perturbation of the distance between the elements reduces transcription activity, suggesting that the spacing is important for recognition [12]. However, the distance between the recognition elements was perplexing; if one assumes that the TFIIIA zinc fingers wrap around the major groove, as seemed typical of C_2H_2 zinc fingers, nine zinc fingers would not be sufficient to span the expected binding site. The crystal structure of the first six zinc fingers of TFIIIA in complex with a DNA sequence containing the IE and box C sequences provided insight into how TFIIIA recognizes the widely spaced elements and why the spacing is important [11]. The first three zinc fingers were observed to bind to the box C sequence, wrapping around the major groove of the DNA, and the fifth zinc finger was bound to the IE element, but zinc fingers 4 and 6 did not interact with the DNA. Instead, they acted as non-DNA-binding spacers to allow recognition of the separated elements. Finger 4 traversed the region between the IE and box C sequences. With finger 6 acting as another spacer, zinc fingers 7–9 could be modeled to interact with the box A element.

The recent crystal structure of zinc fingers 4–6 of TFIIIA in complex with a minimal 5S rRNA has provided

Figure 2



DNA and RNA recognition by C_2H_2 zinc finger proteins. **(a)** The fifth zinc finger of TFIIIA recognizes bases in the major groove of 5S rRNA promoter DNA. The backbone of the fifth zinc finger of TFIIIA (Leu133–Gly161) is shown as a cartoon trace. Zinc-coordinating sidechains, DNA-interacting sidechains and the DNA (C6–G14 and C51–G59) are shown as stick representations with blue nitrogen, red oxygen, orange sulfur and magenta phosphorus atoms. **(b)** The fifth zinc finger of TFIIIA recognizes the phosphate groups of 5S rRNA. The backbone of the fifth zinc finger of TFIIIA (Leu133–Gly161) is shown as a cartoon trace. Zinc-coordinating sidechains, RNA-interacting sidechains and the RNA (G66–A74 and A101–U109) are shown as stick representations colored as in (a). Not shown are interactions of Lys144 and His155 with backbone atoms of C10 in loop A of the 5S rRNA. **(c)** The fourth zinc finger of TFIIIA recognizes primarily an extra-helical base in loop E of 5S rRNA. The backbone of the fourth zinc finger of TFIIIA (Met104–Gln131) is shown as a cartoon trace. Zinc-coordinating sidechains, RNA-interacting sidechains and the RNA (U73–C79 and G97–U102) are shown as stick representations colored as in (a). **(d)** The sixth zinc finger of TFIIIA recognizes loop A of 5S rRNA. The backbone of the sixth zinc finger of TFIIIA (Gly161–Gln189) is shown as a cartoon trace. Zinc-coordinating sidechains, RNA-interacting sidechains and the RNA (C10–A13) are shown as stick representations colored as in (a). Trp177 also forms a stacking interaction with A11. Consistent with this observation, an aromatic residue at this position was shown to be important for 5S rRNA recognition [14]. In all panels, the zinc atom is a grey sphere, zinc coordination is represented by grey dashed lines, hydrogen bonds and salt bridges with sidechain atoms are represented by black dashed lines, hydrogen bonds with mainchain atoms are represented by green dashed lines and a transparent molecular surface is superimposed on the protein structure.

that perturb the structure of the RNA in a small region, although not affecting the structure of any particular recognition element, could affect the relative positioning of those elements and hence be deleterious to binding.

RNA binding by CCCH zinc finger proteins

Tis11d is one of three closely related human proteins that are part of the family of CCCH zinc finger proteins

(reviewed recently in [16]). These zinc fingers are characterized by three cysteine residues and one histidine residue that coordinate the zinc ion and form, in general, a Cys- X_8 -Cys- X_5 -Cys- X_3 -His sequence. The prototypic CCCH zinc finger protein is the related protein tristetraprolin (TTP), also known as Tis11, Nup475 or ZFP36. In the zinc finger region, Tis11d is closely related to TTP in sequence (49 out of 70 amino acid residues are

identical) and is likely to be similar in structure to the better-studied TTP protein.

TTP was first identified as a gene that is rapidly upregulated following stimulation of cells with insulin, serum or phorbol esters [17–19]. The presence of tandem copies of the then-putative zinc-coordinating residues in this 36 kDa protein suggested that it functioned as a transcription factor. However, astute detective work following the generation of TTP knockout mice by Perry Blackshear's laboratory led to the identification of a role for the tandem zinc fingers as an RNA-binding region [4,20].

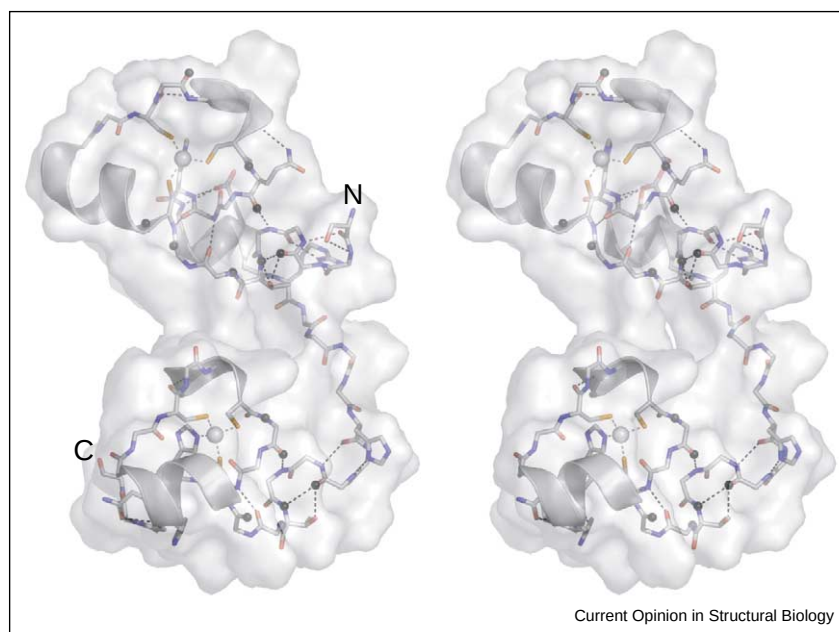
The TTP knockout mice appear normal at birth, but quickly develop a complex syndrome comprising arthritis, wasting, dermatitis and early death. The Blackshear laboratory recognized that this phenotype resembled closely that observed in earlier models of tumor necrosis factor (TNF)- α excess. Thus, they administered anti-TNF- α antibodies to the TTP knockout mice soon after birth and this treatment prevented essentially all of the abnormalities associated with TTP deficiency [20].

Subsequent studies led to the discovery that TTP is an RNA-binding protein that binds to AU-rich elements in the 3' untranslated regions (UTRs) of TNF- α and

granulocyte-macrophage colony stimulating factor (GM-CSF) mRNAs [4,21] (Figure 1b). These AU-rich elements were discovered in the 3' UTRs of mRNAs encoding certain proto-oncogenes, cytokines and lymphokines, and have been shown to mediate rapid degradation of these messages [22,23]. Thus, TTP binding to the AU-rich elements of TNF- α and GM-CSF mRNAs induces the degradation of the messages and so regulates the level of protein expression. AU-rich elements are, as their name implies, rich in adenosine and uridine residues, and have been placed into categories based on the presence or absence of particular subsequences [24]. TTP binds to mRNAs that contain multiple, overlapping AUUUA pentamers. The determination of the solution structure of the tandem zinc fingers of Tis11d in complex with AU-rich element RNA identifies the minimal binding sequence for a tandem zinc finger and provides insight into the mechanism of mRNA degradation [25••].

The structure reveals that each zinc finger module binds to the sequence UAUU; thus, the tandem zinc fingers bind to UAUU-UAUU. This sequence is remarkably similar to the nonamer (UUAUUUAUU) that was identified as the functional AU-rich element motif in biochemical studies performed before TTP had been identified as an effector of AU-rich element mRNA degradation [26,27]. Zubiaga *et al.* [27] demonstrated that

Figure 3



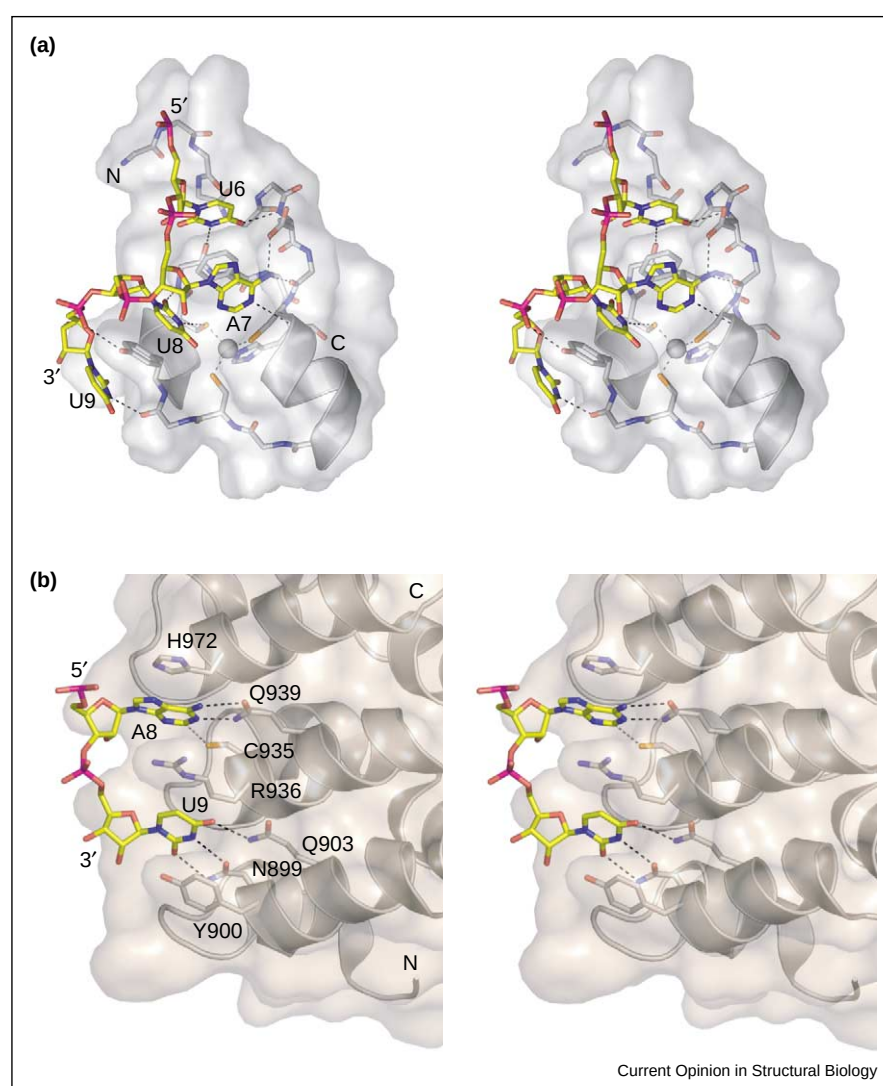
Zinc coordination and an extensive hydrogen-bond network provide the framework for sequence-specific RNA recognition by Tis11d. A stereo representation of the structure of Tis11d tandem CCCH zinc fingers (Ser151–Glu220) is shown. Secondary structure elements are presented in ribbon form. Mainchain atoms of the remainder of the protein, and sidechain atoms of zinc-coordinating residues and positions forming sidechain to mainchain hydrogen bonds are shown as stick representations with blue nitrogen, red oxygen and orange sulfur atoms. Grey dashed lines represent zinc coordination. Black dashed lines represent hydrogen bonds. Backbone positions that contact the RNA are indicated with black spheres. A transparent molecular surface is superimposed on the protein structure.

AU-rich element potency is increased by combining multiple copies of the nonamer sequence, suggesting that binding of multiple copies of TTP to a target mRNA may potentiate the destabilization.

The RNA is recognized by the CCCH zinc finger modules through a familiar combination of hydrogen bond and stacking interactions with the bases, as seen with many RNA-binding proteins. However, the mode of RNA recognition is quite extraordinary: sequence-specific recognition is moderated mainly by protein mainchain

atom hydrogen bonds to the Watson–Crick edges of the bases (Figures 3 and 4a). The backbone architecture must therefore be very important for specificity. With only limited secondary structure elements, the zinc coordination, sidechain to mainchain hydrogen bonds and main-chain to mainchain hydrogen bonds form a network of interactions that produces the framework for sequence-specific RNA interaction (Figure 3). The sidechains that contact mainchain atoms in the Tis11d structure are relatively well conserved among the three human TTP-like proteins (eight out of eight identical in Tis11b;

Figure 4



Modular recognition of RNA by CCCH zinc finger proteins and Pumilio proteins. **(a)** The N-terminal zinc finger of Tis11d (Ser151–His178) bound to RNA is shown in stereo. Secondary structure elements are presented in ribbon form. Mainchain atoms of the remainder of the protein, sidechain atoms of zinc-coordinating residues and residues that contact the RNA, and the RNA (U6–U9) are shown as stick representations with blue nitrogen, red oxygen, orange sulfur and magenta phosphorus atoms. Black dashed lines represent hydrogen bonds. A transparent molecular surface is superimposed on the protein structure. **(b)** The second, third and fourth repeats of human Pumilio1 protein (Leu886–Leu986) bound to RNA are shown in stereo. Secondary structure elements are presented in ribbon form. Sidechain atoms of residues that contact the RNA and the RNA (A18 and U19) are shown as stick representations colored as in (a). A transparent molecular surface is superimposed on the protein structure.

six out of eight identical in TTP), but are less well conserved among all CCCH zinc finger proteins. It will be interesting to see how sequence specificity is achieved by other members of this class of zinc finger protein.

RNA recognition also appears to be modular, with each zinc finger recognizing a UAUU sequence. In fact, a single CCCH zinc finger can bind weakly but specifically to AU-rich element RNA [28^{*}]. This modular recognition is reminiscent of RNA binding by the Pumilio proteins that our laboratory studies (Figure 4b) [29]. Similar to the Tis11d structure, Pumilio recognizes its RNA partner through a combination of stacking interactions with the bases, and hydrogen-bond interactions with the Watson–Crick edges of mainly adenines and uracils. However, in Pumilio, two amino acid sidechains make sequence-specific interactions with the Watson–Crick edge and a single repeat of approximately 36 amino acid residues recognizes one RNA base. The Tis11d CCCH zinc finger modules are much more economical, with approximately 35 amino acid residues recognizing four bases. CCCH zinc finger proteins come in many forms [30]. Some have only one module, whereas others have up to six, and the spacing between domains and the inclusion of interceding domains with different functions can be quite variable. The structure of the Tis11d–RNA complex has provided a great leap in our understanding of this family of proteins, but there remains much to be learned.

Conclusions

Zinc finger domains, although small, appear to be quite versatile. The C₂H₂ zinc finger scaffold can be used for base-specific recognition of the DNA major groove, backbone recognition of the RNA major groove, and almost customized RNA base and loop recognition. Our understanding of CCCH zinc finger–RNA recognition has just begun with the Tis11d–AU-rich element RNA structure. How versatile a protein backbone can be for RNA (and perhaps DNA or protein) recognition remains to be discovered.

Acknowledgements

I am grateful to my colleagues at the National Institute of Environmental Health Sciences (NIEHS) for critical comments on the manuscript, Rachelle Bienstock for providing a coordinate file for B-DNA, and my laboratory for their patience and advice as I prepared this article.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Miller J, McLachlan AD, Klug A: **Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes.** *EMBO J* 1985, **4**:1609–1614.
 2. Picard B, Wegnez M: **Isolation of a 7S particle from *Xenopus laevis* oocytes: a 5S RNA-protein complex.** *Proc Natl Acad Sci USA* 1979, **76**:241–245.
 3. Pelham HR, Brown DD: **A specific transcription factor that can bind either the 5S RNA gene or 5S RNA.** *Proc Natl Acad Sci USA* 1980, **77**:4170–4174.
 4. Carballo E, Lai WS, Blackshear PJ: **Feedback inhibition of macrophage tumor necrosis factor- α production by tristetraprolin.** *Science* 1998, **281**:1001–1005.
 5. D'Souza V, Summers MF: **Structural basis for packaging the •• dimeric genome of Moloney murine leukaemia virus.** *Nature* 2004, **431**:586–590.
- This paper describes the NMR structure of the nucleocapsid domain of the Gag polyprotein from Moloney murine leukemia virus in complex with an RNA representing the expected structure of a dimer of viral genomic RNA. The structure suggests that the zinc knuckle of the nucleocapsid protein binds to bases that become unpaired following dimerization.
6. Guddat U, Bakken AH, Pieler T: **Protein-mediated nuclear export of RNA: 5S rRNA containing small RNPs in *Xenopus* oocytes.** *Cell* 1990, **60**:619–628.
 7. Pavletich NP, Pabo CO: **Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å.** *Science* 1991, **252**:809–817.
 8. Hayes JJ, Tullius TD: **Structure of the TFIIIA-5S DNA complex.** *J Mol Biol* 1992, **227**:407–417.
 9. Liao XB, Clemens KR, Tennant L, Wright PE, Gottesfeld JM: **Specific interaction of the first three zinc fingers of TFIIIA with the internal control region of the *Xenopus* 5S RNA gene.** *J Mol Biol* 1992, **223**:857–871.
 10. Bogenhagen DF: **Proteolytic footprinting of transcription factor TFIIIA reveals different tightly binding sites for 5S RNA and 5S DNA.** *Mol Cell Biol* 1993, **13**:5149–5158.
 11. Nolte RT, Conlin RM, Harrison SC, Brown RS: **Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex.** *Proc Natl Acad Sci USA* 1998, **95**:2938–2943.
 12. Pieler T, Hamm J, Roeder RG: **The 5S gene internal control region is composed of three distinct sequence elements, organized as two functional domains with variable spacing.** *Cell* 1987, **48**:91–100.
 13. Lu D, Searles MA, Klug A: **Crystal structure of a zinc-finger-RNA •• complex reveals two modes of molecular recognition.** *Nature* 2003, **426**:96–100.
- This paper describes the crystal structure of the fourth through sixth zinc fingers of TFIIIA in complex with a core 5S rRNA. It reveals two modes of RNA recognition that differ from typical DNA recognition by C₂H₂ zinc fingers.
14. Theunissen O, Rudt F, Pieler T: **Structural determinants in 5S RNA and TFIIIA for 7S RNP formation.** *Eur J Biochem* 1998, **258**:758–767.
 15. Friesen WJ, Darby MK: **Phage display of RNA binding zinc fingers from transcription factor IIIA.** *J Biol Chem* 1997, **272**:10994–10997.
 16. Carrick DM, Lai WS, Blackshear PJ: **The tandem CCCH zinc finger protein tristetraprolin and its relevance to cytokine mRNA turnover and arthritis.** *Arthritis Res Ther* 2004, **6**:248–264.
 17. Lai WS, Stumpo DJ, Blackshear PJ: **Rapid insulin-stimulated accumulation of an mRNA encoding a proline-rich protein.** *J Biol Chem* 1990, **265**:16556–16563.
 18. Varnum BC, Ma QF, Chi TH, Fletcher B, Herschman HR: **The TIS11 primary response gene is a member of a gene family that encodes proteins with a highly conserved sequence containing an unusual Cys-His repeat.** *Mol Cell Biol* 1991, **11**:1754–1758.
 19. DuBois RN, McLane MW, Ryder K, Lau LF, Nathans D: **A growth factor-inducible nuclear protein with a novel cysteine/histidine repetitive sequence.** *J Biol Chem* 1990, **265**:19185–19191.
 20. Taylor GA, Carballo E, Lee DM, Lai WS, Thompson MJ, Patel DD, Schenkman DI, Gilkeson GS, Broxmeyer HE, Haynes BF *et al.*: **A pathogenetic role for TNF α in the syndrome of**

- cachexia, arthritis, and autoimmunity resulting from tristetraprolin (TTP) deficiency.** *Immunity* 1996, **4**:445-454.
21. Carballo E, Lai WS, Blackshear PJ: **Evidence that tristetraprolin is a physiological regulator of granulocyte-macrophage colony-stimulating factor messenger RNA deadenylation and stability.** *Blood* 2000, **95**:1891-1899.
 22. Caput D, Beutler B, Hartog K, Thayer R, Brown-Shimer S, Cerami A: **Identification of a common nucleotide sequence in the 3'-untranslated region of mRNA molecules specifying inflammatory mediators.** *Proc Natl Acad Sci USA* 1986, **83**:1670-1674.
 23. Shaw G, Kamen R: **A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation.** *Cell* 1986, **46**:659-667.
 24. Chen CY, Shyu AB: **AU-rich elements: characterization and importance in mRNA degradation.** *Trends Biochem Sci* 1995, **20**:465-470.
 25. Hudson BP, Martinez-Yamout MA, Dyson HJ, Wright PE: **Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d.** *Nat Struct Mol Biol* 2004, **11**:257-264.
This paper describes the solution structure of the tandem CCCH zinc finger protein Tis11d in complex with AU-rich element target RNA. The structure reveals that each zinc finger interacts with a UAUU sequence and achieves sequence-specific recognition using predominantly mainchain atoms.
 26. Lagnado CA, Brown CY, Goodall GJ: **AUUUA is not sufficient to promote poly(A) shortening and degradation of an mRNA: the functional sequence within AU-rich elements may be UUAUUUA(U/A)(U/A).** *Mol Cell Biol* 1994, **14**:7984-7995.
 27. Zubiaga AM, Belasco JG, Greenberg ME: **The nonamer UUAUUUAUU is the key AU-rich sequence motif that mediates mRNA degradation.** *Mol Cell Biol* 1995, **15**:2219-2230.
 28. Michel SL, Guerrero AL, Berg JM: **Selective RNA binding by a single CCCH zinc-binding domain from Nup475 (Tristetraprolin).** *Biochemistry* 2003, **42**:4626-4630.
This paper reports that a single CCCH zinc finger domain from TTP can bind weakly to a short AU-rich element RNA ($K_D \sim 5 \mu\text{M}$). This binding is selective as mutation of the adenosines to guanine or uridine abrogates binding ($K_D > 60 \mu\text{M}$). The single CCCH domain can also bind an equivalent DNA with $K_D \sim 15 \mu\text{M}$.
 29. Wang X, McLachlan J, Zamore PD, Hall TM: **Modular recognition of RNA by a human Pumilio-homology domain.** *Cell* 2002, **110**:501-512.
 30. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL *et al.*: **The Pfam protein families database.** *Nucleic Acids Res* 2004, **32**:D138-D141.
 31. DeLano W: **The PyMOL Molecular Graphics System.** San Carlos, CA, USA: DeLano Scientific; 2002.