Intact RNA binding domains are necessary for structure-specific DNA binding and transcription control by CBTF\textsuperscript{122} during \textit{Xenopus} development

Garry P. Scarlett, Stuart J. Elgar\textsuperscript{1}, Peter D. Cary, Anna M. Noble, Robert L. Orford\textsuperscript{2}, G. Geoffrey Kneale, and Matthew J. Guille\textsuperscript{3}

Institute of Biomedical and Biomolecular Sciences, University of Portsmouth, Portsmouth, PO1 2DT

\textsuperscript{1} Cardiff School of Biosciences, Cardiff university, Cardiff, CF10 3TL

\textsuperscript{2} NIMR, The Ridgeway, Mill Hill, London, NW7 1AA

\textsuperscript{3} Author for correspondence (e-mail: matthew.guille@port.ac.uk)

Running title: DNA binding by dsRBDs

Keywords: Double stranded RNA binding domain; DNA binding; RNA binding; A-form DNA; CCAAT box; ILF3
**Summary**

CBTF$^{122}$ is a subunit of the *Xenopus* CCAAT box transcription factor (CBTF) complex and a member of a family of dsRNA binding proteins that function in both transcriptional and post-transcriptional control. Here we identify a region of CBTF$^{122}$ containing the double stranded RNA binding domains (dsRBDs) that is capable of binding either RNA or DNA. We show that these domains bind A-form DNA in preference to B-form DNA, and that the −59 to −31 region of the GATA-2 promoter (an *in vivo* target of CBTF) adopts a partial A-form structure. Mutations in the RNA binding domains that inhibit RNA binding also affect DNA binding *in vitro*. In addition, these mutations alter the ability of CBTF$^{122}$ fusions with engrailed transcription repressor and VP16 transcription activator domains to regulate transcription of the GATA-2 gene *in vivo*. These data support the hypothesis that the dsRBDs of this family of proteins are important for their DNA binding both *in vitro* and *in vivo*. 
Introduction

The 122 KDa subunit of the *Xenopus* CCAAT box transcription factor CBTF\(^{122}\) (also known as 4F and ubp4) belongs to a family of double stranded RNA binding domain (dsRBD) containing proteins implicated in a variety of cellular processes. CBTF\(^{122}\) and its splice variant CBTF\(^{98}\) have been identified as subunits of CBTF, a transcription factor that activates the GATA-2 gene both in oocytes and zygotically at the start of gastrulation (1,2). Homologues of CBTF\(^{122}\) have been identified mainly through their affinity for RNA and include NF90 (also known as DRBP76 and NFAR 1), NF110, TCP80, ILF3 and MPP4 in humans (3), SPNR and ILF3 in mice (4) and p74 and ILF3 in rats (5). It is now clear that many of the human homologues arise from alternative splicing of transcripts from the ILF3 locus (6,7).

The functions assigned to these related proteins include the regulation of both transcription and mRNA stability. For example, the transcript levels of more than 90 genes, most of which are known to be regulated by type I interferons, were altered by expression of NF90, suggesting that NF90 has a role in mediating the antiviral response (8). However, it is unclear whether these changes are a result of transcriptional or post-transcriptional mechanisms. By contrast, specific regulation of IL-2 levels by NF90 is clearly dependent upon its stabilisation of IL-2 mRNA (9).

In transcriptional terms, the human homologues of CBTF\(^{98}\) and CBTF\(^{122}\), NF90 and NF110, have been shown to be capable of both transcription activation and repression.
depending upon which promoter they are acting at (3,10). Deletion analysis showed that transcription activation required the nuclear localisation signal and two dsRBDs (3). Mutations within the dsRBDs that abrogated RNA binding also inhibited transcription activation (10).

During *Xenopus* development, CBTF\textsuperscript{122} has been shown to be critical for activation of GATA-2 transcription both by *in vitro* (2) and *in vivo* methods (Scarlett *et al*., manuscript in preparation). GATA-2 is a transcription factor necessary for correct haematopoietic, neural and urogenital development in mice (11,12) and is also implicated in the formation of ventral mesoderm in *Xenopus* (13). Although CBTF is maternal, it does not activate GATA-2 transcription in embryos until the start of gastrulation and this is, at least in part, due to regulation of its nuclear translocation. CBTF is present in the oocyte nucleus where GATA-2 is initially transcribed. However, after oocyte maturation, CBTF activity is cytoplasmic and re-locates to the nucleus only at the gastrula stage when zygotic GATA-2 transcription commences (1). CBTF\textsuperscript{122} is anchored in the cytoplasm by RNA binding (14) and its movement from the cytoplasm to the nucleus is synchronous with the mass degradation of maternal RNA at the mid-blastula transition (MBT), suggesting a mechanism for the developmental stage specific control of CBTF activity based on RNA binding.

Analysis of the amino acid sequence of CBTF\textsuperscript{122} indicates a number of conserved domains (figure 1a). The N-terminal domain is predicted to contain ZFR type zinc fingers (15). CBTF\textsuperscript{122} also contains two dsRNA binding domains (dsRBDs) of the type found in
staufen and Xlrbpa (16,17). C-terminal to this is an arginine-glycine rich region (RGG domain), similar to those found in hnRNP A and U1, which are implicated in nucleic acid binding (18,19). There is also a poly-Q region which has been suggested to be involved in transcriptional activation in the context of other proteins (20). Finally, there is a nuclear localisation signal (NLS) located N-terminal to the dsRBDS (14,21) with a bipartite arrangement similar to the NLS of nucleoplasmin (22).

The RNA binding activity of CBTF^{122} and NF90, as well as of other dsRBD containing proteins, has been well characterised. A CBTF^{122}-MBP fusion binds to dsRNA and A-form DNA-RNA hybrids without detectable sequence specificity (21). The NMR structure of the staufen dsRBD has been solved (23), as has the crystal structure of the Xlrbpa dsRBD complexed with dsRNA (24). Saturation mutagenesis studies on the dsRBD of Xlrbpa demonstrated that the substitution of a conserved central Phe residue by Ala in both staufen and Xlrbpa dsRBDS leads to abolition of RNA binding activity (23,25). Reference to the published structure of the dsRBD-RNA complex demonstrates that this residue is important in stabilising a lysine side-chain that makes important backbone contacts adjacent to the major groove (24). It is this mutation that we have previously used to inactivate the dsRBDS of CBTF^{122} (14).

The human orthologue of CBTF^{122} has been shown to require intact RNA binding domains for transcription regulation, but on promoters that are not known to be its in vivo targets (10). We therefore tested whether an intact RBD was similarly required for
transcription regulation at the GATA-2 promoter in *Xenopus*, a well-characterised *in vivo* target of CBTF122.

**Experimental procedures**

*Embryo manipulation and western blotting*

Embryos were obtained as described by Smith and Slack (26) and staged according to Nieuwkoop and Faber (27). *In vitro* transcription and microinjection of mRNA was carried out as described previously (28,29). Injections were carried out at the 8-cell stage with 4 nl of RNA solution (37 pg RNA total). One embryo equivalent of freon extracted whole embryo lysate was assayed for CBTF122 fusion expression by SDS-PAGE and western blotting (30) using CBTF122 antibody (provided by B. Bass).

*Analysis of mRNA levels by real-time RT-PCR*

cDNA was synthesised from RNA extracted from dissected explants of injected embryos cultured until stage 11 using the method of Steinbach and Rupp (31). RNA expression was assayed as cDNA by PCR conducted on an ABI 7900HT sequence detection system using TAQMAN fluorescently labelled probes. Data were analysed using the ΔΔCt method (ABI) using ornithine decarboxylase (ODC) as a control gene and uninjected marginal zones as control tissue. The probe and primer sequences are available by request.

*Protein expression and purification*
The RBD and RBD+RGG regions of CBTF122 as well as the mutant RBD (mRBD) region were sub-cloned into the pGex2T (Amersham Pharmacia Biotech). The GST-fusion proteins were expressed in *E. coli* and purified by glutathione-sepharose and DNA-cellulose column chromatography as previously described (14). The purified proteins were assayed and quantified by SDS-PAGE and UV spectroscopy using the calculated extinction coefficient of each protein. Embryo extract containing the multi-subunit CBTF complex was prepared as described previously (2).

**Electrophoretic mobility shift assays of recombinant CBTF122 and endogenous CBTF**

Either RNA (Cruachem) or DNA (Invitrogen) oligonucleotides were annealed to form duplexes and end-labelled by T4 polynucleotide kinase (NEB) using α$^{32}$P ATP (32). The proteins were incubated with the appropriate nucleic acid probe for 15 minutes on ice in EMSA buffer (2) with the addition of 500ng poly dI-dC in the case of embryo extract, prior to separation of DNA-protein complexes on a 4% native polyacrylamide gel in 0.25X TBE. The gels were dried and the free DNA and DNA-protein complexes quantified using a phosphorimager (Molecular Dynamics). $K_d$s were estimated from the protein concentration required to bind 50% of the DNA.

**Dimethyl Sulphate interference footprinting**

Four picomoles of oligonucleotide were labelled at the 5’ end and annealed with the complementary sequence to form a double-stranded 55mer probe. The probe was then gel purified to remove residual single stranded oligonucleotide. The purified dsDNA was treated with 10% DMS (33), and the partially methylated probe used in an EMSA.
Separated species were transferred onto DE81 paper. After autoradiography the bound and free species were excised and eluted in 500µl of buffer (1M NaCl, 20 mM Tris pH 8.0, 1mM EDTA). 10µg of tRNA was added and the DNA was ethanol precipitated. Piperidine cleavage and further treatment was performed as described by Maxam and Gilbert (33), DNA fragments were separated on a denaturing 16% polyacrylamide gel and visualised by autoradiography.

**Circular dichroism**

An Applied Photophysics π*-180 instrument was flushed with nitrogen gas for all CD experiments. Cell path lengths of 1, 2, 4 and 10mm were used with sample concentrations of 15 to 60 µg/ml. All 29 bp duplexes were dissolved in 100mM KF, 5mM NaPO₄ buffer, pH 7.6 at 22 °C. Concentrations were determined by UV measurements at 260 nm coupled with snake-venom time course digestions to correct for hypochromic differences. Data were collected over the wavelength range 180-360 nm using adaptive sampling in conjunction with signal averaging in all cases. The instrument wavelength accuracy was ± 0.1nm, determined from the Xenon lines and the CD was calibrated from camphor sulphonylic acid at 290.5nm.

**A-form DNA predictive methods**

A-DNA propensity energy (APE) values for the duplex oligonucleotides were calculated using the method of Basham et al., (38). The method was modified to determine the percentage of each duplex in either A or B-form using the following classifications and rules. Class 1, strong A or B formers are base pairs that have the same sign and whose
sum is greater than 0.5. Class 2, neutral A or B formers are base pairs that are of opposite sign and whose sum is zero. Class 3, weak A or B formers are base pairs whose sum is less than 0.5 but not zero. Class 4, base pairs unclassified by Basham et al., because of lack of data. These are defined by their base composition as GC, CG or TA, AT. The following rules were applied to the above classifications; (1) All Class 1 base pairs remain as they are under all situations, conserved type. (2) If one or two consecutive class 2 or 3 are between two or more class 1 base pairs of the same type (A or B) then they both take on strong base pair type. (3) Class 2 neutrals adjacent to class 1 base pairs take on that type if the base pair on the other side to the class 2 neutral is of class 2 or class 3. (4) Class 4 base pairs between class 1 base pairs of the same type can take on the same character as the class 1 type. The predictions shown (table 3) show only the A and B percentage forms, there is 7% contribution which cannot be determined using the three base moving window model of Basham et al., because these involve the ends of the molecule. Also, regions where no bias for each form can be determined add to the percentage of uncertainty.

Results

Mutations in the dsRBDs of CBTF^{122}EN lead to a loss of transcriptional regulation

The engrailed transcription repression domain (EN) when fused to a transcription activator leads to transcription repression of the target gene (34). We have recently shown that the transcription inhibiting (CBTF^{122}EN) fusion form of CBTF^{122} gives rise to a double axis phenotype (arrowed, figure 1b) when expressed in Xenopus embryos. We confirmed that this effect was a result of changes at a transcriptional level and not a
failure of CBTF\textsuperscript{122}EN to be incorporated into the full CBTF complex, or an effect on mRNA turn-over, by testing the strong transcription up-regulating CBTF\textsuperscript{122} form, CBTF\textsuperscript{122}VP16. This protein gave the opposite, ventralised microcephalic phenotype when expressed in embryos dorsally (figure 1b, note the lack of eyes and head morphology), while CBTF\textsuperscript{122} expression alone gave rise to no obvious change in phenotype (Scarlett et al., manuscript in preparation).

Transcription activation by CBTF\textsuperscript{122} is regulated in part by its localisation in the cytoplasm through anchoring to RNA. The mutation F435/F559A in the dsRBDs leads to a loss of RNA binding. Such mutants localise early to the nucleus, and thus arrive at their site of action earlier than native CBTF\textsuperscript{122} (2). We investigated how this would affect transcriptional control by mutating the RBDs in the context of full length CBTF\textsuperscript{122}EN and CBTF\textsuperscript{122}VP16 fusions. These fusion proteins (CBTF\textsuperscript{122}mRBDEN and CBTF\textsuperscript{122}mRBDVP16) were expressed in embryos by injection of \textit{in vitro} transcribed mRNA alongside those injected with CBTF\textsuperscript{122}EN and CBTF\textsuperscript{122}VP16 mRNA. Although constructs with active dsRBDs resulted in the expected double axis and headless phenotypes, those containing a mutant form of the domains did not (table 1) suggesting a role for the dsRBDs in gene regulation.

We have shown CBTF\textsuperscript{122}EN to be a powerful, specific repressor of transcription from the GATA-2 promoter (Scarlett et al., manuscript in preparation). We therefore tested the ability of CBTF\textsuperscript{122}mRBDEN to repress GATA-2 transcription in a \textit{Xenopus} animal cap assay (35). In this assay, injection of CBTF\textsuperscript{122}EN mRNA reduces GATA-2 mRNA levels
by over 1000-fold whilst injection of CBTF\textsuperscript{122}mRBDEN mRNA reduces GATA-2 mRNA levels by only 10-fold (table 2). We confirmed by western blotting of whole embryo extract using CBTF\textsuperscript{122} antibody that the two proteins were expressed at comparable levels (figure 1c). Thus a single point mutation in each of the dsRBDs is almost sufficient to abolish repression by CBTF\textsuperscript{122}EN at the GATA-2 promoter. In order to understand how the RNA-binding mutations affected the ability of CBTF\textsuperscript{122} to act at the transcriptional level we then analysed how the binding of CBTF\textsuperscript{122} dsRBDs to nucleic acids was affected by these mutations in the dsRBDs \textit{in vitro}.

\textbf{The dsRBD domains of CBTF\textsuperscript{122} bind nucleic acids preferentially in the order dsRNA>dsDNA>ssRNA}

CBTF\textsuperscript{122} contains a single RGG domain of the type found in hnRNP A1 and U which is implicated in single stranded nucleic acid binding (18,19), and twin dsRBDs of the type found in Staufen and Xlrbpa (16). To determine the contribution of these different domains to the interaction with nucleic acids we expressed and purified CBTF\textsuperscript{122} fragments corresponding either to the dsRBD and RGG domains together, or the dsRBDs alone, as GST fusions (figure 1a). The predicted sizes of the expressed protein fragments were 41kDa and 28.5kDa respectively (CBTF\textsuperscript{122}(41), CBTF\textsuperscript{122}(28.5)), both contained the nuclear localisation signal. We purified these proteins to 95\% homogeneity and then tested their ability to bind a dsRNA homologue of the –69 to –33 CCAAT box containing region of the GATA-2 promoter by EMSA (figure 2a). The binding curves for CBTF\textsuperscript{122}(41)GST and CBTF\textsuperscript{122}(28.5)GST show that their affinities for RNA are
comparable, with $K_d$s of approximately 0.3 nM. Consequently the smaller 28.5kDa fragment was chosen for subsequent EMSA studies.

To estimate the relative affinity of the dsRBDs of CBTF$^{122}$ for ssRNA, dsRNA and dsDNA, data from EMSA experiments using CBTF$^{122}(28.5)$GST were quantitated and the percent bound probe plotted as a function of protein concentration (figure 2b). The estimated binding constants show that the dsRBDs have highest affinity for dsRNA, much lower affinity for dsDNA, and the lowest affinity for ssRNA with estimated $K_d$s of 0.3, 6.0, and 30 nM respectively. Previously a full length CBTF$^{122}$–MBP (maltose binding protein) fusion has been reported to have a $K_d$ of 0.27nM for dsRNA (21). Therefore, our data suggest that dsRNA binding activity resides entirely within the region of the dsRBDs represented in the CBTF$^{122}(28.5)$GST protein. Since this fragment bound to DNA, we then tested whether the RNA binding mutations that affected the ability of CBTF$^{122}$EN to control GATA-2 transcription also affected DNA binding.

**Mutations in the dsRBDs affect DNA binding activity**

The previously described mutated protein CBTF$^{122}$mRBD (2) contains two point mutations, F435A and F559A, known to be critical for RNA binding (23,36). However, it is possible that residues F435 and F559 are also important for the less understood DNA binding activity of this protein. We expressed and purified recombinant forms of both the native and mutant RNA binding domains, again as GST fusions. These proteins (CBTF$^{122}(28.5)$GST and CBTF$^{122}$mRBD(28.5)GST) were tested for their ability to bind to either DNA or RNA homologues of the –59 to –31 GATA-2 promoter sequence. In
agreement with our previous data (37) the CBTF\textsuperscript{122}mRBD(28.5)GST protein has a much reduced affinity for RNA duplex compared to CBTF\textsuperscript{122}(28.5)GST (figure 3a). However, its DNA binding activity is also inhibited (figure 3b). Since these data show that dsRNA and dsDNA binding by CBTF\textsuperscript{122} involve at least some of the same critical amino acids, we investigated the interaction of CBTF\textsuperscript{122}(28.5) with DNA further.

**The dsRBDs of CBTF\textsuperscript{122} bind DNA in a structure dependent manner**

CBTF\textsuperscript{122} is known to bind dsRNA in a sequence independent manner (21). To test whether CBTF\textsuperscript{122}(28.5)GST showed any sequence specificity in binding DNA we designed five oligonucleotides, based around the –59 to –31 CCAAT box region of the GATA-2 promoter (P1 to P5, table 3). These oligonucleotides were then used in competition EMSA experiments against the P1 DNA probe. The results showed that P3, P4 and P5 competition was comparable with P1 self-competition, but P2 was found to compete very poorly (figure 3c).

Binding of P2 was disrupted despite the retention of the core CCAAT sequence to which CBTF binds. One possible explanation for this was that a structural rather than sequence-dependent binding mechanism was disrupted by the mutation. Since dsRBDs may be expected to have a preference for binding A-form structure, we initially used the method of Basham et al (38) to calculate the average A-DNA forming potential for each of the oligonucleotides used in this study (table 3). We refined this method to predict the percentage A and B-form for each duplex (see methods) this is shown as “predicted” in table 3. The P1 sequence was predicted to be strongly A-DNA forming. In particular, the
bases mutated in P2 mapped to a region of high A-DNA forming potential (−57 to −47). We therefore tested the hypothesis that A-DNA elements in the P1 probe are important for interactions of the dsRBDs with dsDNA by extending our study to the structure of the −59 to −31 region of the GATA-2 promoter.

Circular dichroism was used to examine the structure of P1 in solution. A-form DNA is characterised by a CD spectrum that has a strong positive band centred between 267-273 nm, and a negative band centred between 209–211. B-form DNA is characterised by a strong positive band centred between 273–278 nm, and two negative bands centred at approximately 206 nm and 250 nm. A spectrum of previously described B-form (C1), and A-form (C2) oligonucleotides (38) is shown in figure 4a.

Each base has different transition dipoles, hence the band centre positions and intensities vary for the same structures depending on composition. The spectrum (figure 4a) of the P1 oligonucleotide exhibits a negative band centred at 211 nm and a positive band centred at 268 nm. The CD results therefore show that oligonucleotide P1 has significant A-form structure. The CD spectrum of P1 could be the result of mixed regions of A-form and B-form DNA or a continuous structure sharing the properties of both. The A and B-form components were deconvoluted by measuring the signal intensity at 278 nm for C1 (100% B-form) and C2 (0% B-form) and plotted against percent B-form assuming a linear change between the two points (figure 4b).
The equation of the resulting graph was used to calculate percent B-form for P1 and P2. The data suggest that 35% of P1 and 51% of P2 is in the canonical B-form, assuming a two component model where the oligonucleotide duplex is composed solely of A and B-form. To test the validity of using a single wavelength to estimate the amount of B-form we reconstituted the P1 spectrum using the calculated 65% A and 35% B-form (figure 4c). The two curves fitted well, particularly over the important 250-310 nm range that reflects the contribution from the bases.

In light of the partial A-form nature of P1 we used oligonucleotides C1 and C2, as well as designing a further control oligonucleotide (C3) to test whether CBTF\textsuperscript{122(28.5)}GST bound preferentially to A-DNA (table 3). C2 is a known A-form oligonucleotide (38), while C3 is an alternating poly AT tract predicted to form a B-form structure. We then used these oligonucleotides for competition EMSA analysis. The data showed that the A-form DNA oligonucleotide C2 competed better than any of the oligonucleotides P1 to P5. In contrast, the control B-form oligonucleotides C1 and C3 competed poorly (figure 3c).

Having shown a correlation between A-form structure and the DNA binding affinity of CBTF\textsuperscript{122}, we investigated the multisubunit CBTF complex-DNA interaction and the relative importance of sequence and structure elements in the specific binding of CBTF to elements of the GATA-2 promoter.
Binding of the multimeric CBTF complex to DNA is mainly sequence not structure-dependent.

The binding of the CBTF<sup>122</sup>(28.5)GST fusion protein to DNA was unaffected by mutations in or around the CBTF consensus sequence, but dependent on A-form structure. To test the relative importance of the sequence and structure of the DNA in the binding of the intact CBTF complex, we used the same oligonucleotides we had previously used for binding studies on the dsRBDs, as competitors in EMSAs, not of purified CBTF<sup>122</sup> domains as before but of CBTF from whole cell extract. Binding of the full CBTF complex requires the presence of a CCAAT box in the sequence (1,2), we therefore included a new oligonucleotide (P6) lacking the CCAAT box (figure 5). The results show that oligonucleotides based around the CCAAT box compete well for the CBTF complex, and that sequences that lack the CCAAT box compete poorly for the complex, although an A-form oligonucleotide (C2) competed best of those oligonucleotides containing no CCAAT box (figure 5).

The binding of the intact CBTF complex to DNA was thus found to be dependent on specific sequences within the P1 region. To determine further those bases involved in these interactions, the binding was analysed by dimethyl-sulphate (DMS) interference footprinting. In this assay, the unbound fraction is considered to contain every species of methylated DNA, whilst the retarded band is enriched in species of methylated DNA that do not interfere with binding (39). It therefore identifies the extent of protein contacts with specific bases that are important for complex stability. Our results show that modification of purines between –56 to –40 (inclusive) reduces binding of CBTF to its
cognate sequence (figure 6). Thus the multisubunit CBTF makes extensive contacts over a sequence spanning 17 bases, including nucleotides well beyond the core CCAAT sequence.

Discussion

Here we report that single point mutations in both dsRBDs of the protein CBTF\textsuperscript{122} prevent the formation of phenotypes we have previously found associated with the dominant negative and up-regulating fusion forms of this protein. Further, these mutations severely inhibit the GATA-2 transcription regulation activity of CBTF\textsuperscript{122EN} in embryos. Our \textit{in vitro} binding studies show that the dsRBDs can bind both RNA and DNA, and that the same point mutations inactivate both of these functions. The RBDs show a preference for binding to A-form DNA. This is likely to be important for the regulation of GATA-2 since an oligonucleotide corresponding to the –59 to –31 CBTF binding region of the GATA-2 promoter adopts a partial A-form structure. However, sequence and not structure is the prime requisite for binding of the complete CBTF complex to the GATA-2 promoter \textit{in vitro}.

Engrailed and VP16 fusions of CBTF\textsuperscript{122} proteins were used as a method of analysing the transcriptional regulation activity of this dual function protein in isolation. Opposing phenotypes were found to arise from CBTF\textsuperscript{122EN} and CBTF\textsuperscript{122VP16} expression when they were expressed in \textit{Xenopus} embryos (dorsalising and ventralising respectively); strongly suggesting that the effects are a consequence of down and up-regulation at the
transcriptional level. The mode of action of these fusions only requires the protein to be at the promoter site for transcription regulating complexes to be recruited by the EN and VP16 domains. In particular, the engrailed transcription repression domain would inhibit GATA-2 transcription if it were present at the GATA-2 promoter. Hence we infer that the mutations in the dsRBDs that prevent GATA-2 repression by CBTF^{122}EN most likely involve a failure of the protein to bind to the GATA-2 promoter in vivo.

These results are in agreement with a similar study performed in cultured cells using NF110 (a human orthologue of CBTF^{122} (10)), in which intact dsRBDs were also required for transcription control. RNA binding has previously been associated with transcription activity function, for example; an RNA–induced conformational change was proposed to be necessary for activation of the transcriptional regulatory role of NF110 (10) and RNA is also a co-factor of the steroid receptor coactivator (SRA) (40). We cannot formally exclude similar RNA-induced activation or co-activator function underlying the requirement for RNA binding for transcription regulation by CBTF^{122}. However, the fact that CBTF^{122} fusion forms are inactivated by mutations that affect RNA binding despite a dominant domain providing transcription regulation make a model in which RNA is required for transcription regulation per se unlikely.

Our data support an alternative hypothesis, that the dsRBDs act as both RNA and DNA binding motifs, since their mutation inhibits both RNA and DNA binding in vitro. Other proteins also contain RNA and DNA binding activities residing in the same amino acid sequence, for example the cold shock domain of YB-1 (41) and the homeo-domain like
motif of jerky (42). Further support for the proposal that dsRBDs also have a DNA binding role comes from Tn916 integrase, which has a structural domain highly related to the dsRBD that is capable of sequence specific DNA binding (43).

In the case of the dsRBDs of CBTF\textsuperscript{122}, binding to dsDNA is structure dependent, with a strong preference for A-form DNA. Although dsDNA is normally considered to exist in the B-form state, the CD data presented here suggest that the –59 to –31 region of the GATA-2 promoter used in this study adopts a partial A-form structure. Analysis of the CD spectrum of P1 indicates 65% of the DNA to be A-form. That much of this corresponds to the sequence encompassed by the –55 to –45 region of the promoter is suggested by predictive methods. This region is just upstream of the bulk of the specific contacts revealed by the footprinting data. Mutation of approximately half this region (–56 to –50) effectively abolishes binding of CBTF\textsuperscript{122}(28.5)GST, whereas mutations outside this region have little effect upon binding affinity. Together, these data argue for a two-component model of the P1 duplex (distinct regions of A and B helix), with the dsRBDs binding to the A-form region.

The crystal structure of the dsRBD from Xlrbpa complexed with dsRNA shows the dsRBD binding in the minor groove of an A-form helix, and specific contacts being made with a 2’-OH of the RNA by loop 2. It is possible that the adoption of an A-form structure by the P1 DNA sequence allows the dsRBDs of CBTF\textsuperscript{122} to bind the minor groove of dsDNA which is then similar to that of dsRNA. The lack of a 2’-OH in DNA is unlikely to abolish binding, but may reduce the binding affinity, as observed in our data.
Investigation of the DNA binding by CBTF$^{122}$ dsRBDs showed no clear evidence of sequence specificity outside the requirement for a sequence that may adopt an A-form helix. This is in contrast to the complete CBTF complex, for which our EMSA studies show clear a sequence requirement for the conserved CCAAT box region. An A-form structure alone is insufficient for binding by the intact complex, but our competition studies suggest that this structure makes some contribution to the overall affinity. CBTF makes sequence specific base contacts over a 17bp region that is highly conserved between GATA-2 promoters and distinct from those made by other CCAAT factors (44,45). It is likely that CBTF subunits other than CBTF$^{122}$ are required to contact the CCAAT core.

Overall, the data presented here add considerably to our knowledge of the CBTF-DNA interaction and the role of CBTF$^{122}$ within the multisubunit complex. Our results support a role for dsRBDs as DNA binding domains of functional significance both in vitro and in vivo.

Acknowledgements

We thank Dr Brenda Bass for the CBTF$^{122}$ antibody, and Colin Sharpe and James McClellan for discussions and for suggesting improvements to the manuscript. This work was supported by a BBSRC studentship to Stuart J. Elgar and a Wellcome Trust project grant.
References

1. Brewer, A. C., Guille, M. J., Fear, D. J., Partington, G. A., and Patient, R. K. (1995) *The EMBO Journal* **14**, 757-766

2. Orford, R. L., Robinson, C., Haydon, J., Patient, R. K., and Guille, M. J. (1998) *Mol. Cell. Biol.* **18**, 5557-5566

3. Reichman, T., Muniz, L., and Mathews, M. (2002) *Mol. Cell. Biol.* **22**, 342-356

4. Buaas, F. W., Lee, K., Edelhoff, S., Distech, C., and Braun, R. E. (1999) *Mamm Genome* **10**, 451-456

5. Coolidge, C. J., and Patton, J. G. (2000) *Nucleic Acids Res* **28**, 1407-1417

6. Duchange, N., Pidoux, J., Camus, E., and Sauvaget, D. (2000) *Gene* **261**, 345-353

7. Saunders, L. R., Jurecic, V., and Barber, G. N. (2001) *Genomics* **71**, 256-259

8. Krasnoselskaya-Riz, I., Spruill, A., Chen, Y. W., Schuster, D., Teslovich, T., Baker, C., Kumar, A., and Stephan, D. A. (2002) *AIDS Res Hum Retroviruses* **18**, 591-604

9. Shim, J., Lim, H., J, R. Y., and Karin, M. (2002) *Mol Cell* **10**, 1331-1344

10. Reichman, T. W., and Mathews, M. B. (2003) *RNA* **9**, 543-554

11. Tsai, F. Y., Keller, G., Kuo, F. C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F. W., and Orkin, S. H. (1994) *Nature* **371**, 221-226

12. Zhou, Y., Lim, K. C., Onodera, K., Takahashi, S., Ohta, J., Minegishi, N., Tsai, F. Y., Orkin, S. H., Yamamoto, M., and Engel, J. D. (1998) *EMBO J.* **17**, 6689-6700
13. Sykes, T. G., Rodaway, A. R. F., Walmsley, M. E., and Patient, R. K. (1998) *Development* **125**, 4595-4605

14. Brzostowski, J., Robinson, C., Orford, R., Elgar, S., Scarlett, G., Peterkin, T., Malartre, M., Kneale, G., Wormington, M., and Guille, M. (2000) *EMBO J* **19**, 3683-3693

15. Meagher, M. J., Schumacher, J. M., Lee, K., Holdcraft, R. W., Edelhoff, S., Disteche, C., and Braun, R. E. (1999) *Gene* **228**, 197-211

16. St Johnston, D., Brown, N. H., Gall, J. G., and Jantsch, M. (1992) *Proc. Natl. Acad. Sci. U S A* **89**, 10979-10983

17. Burd, C. G., and Dreyfuss, G. (1994) *Science* **265**, 615-621

18. Cobianchi, F., Karpel, R. L., Williams, K. R., Notario, V., and Wilson, S. H. (1988) *J. Biol. Chem.* **263**, 1063-1071

19. Kiledjian, M., and Dreyfuss, G. (1992) *EMBO J.* **11**, 2655-2664

20. Liberati, C., di Silvio, A., Ottolenghi, S., and Mantovani, R. (1999) *J. Mol. Biol.* **285**, 1441-1455

21. Bass, B. L., Hurst, S. R., and Singer, J. D. (1994) *Current Biology* **Vol.4**, 301-314

22. Robbins, J., Dilworth, S. M., Laskey, R. A., and Dingwall, C. (1991) *Cell* **64**, 615-623

23. Bycroft, M., Grunert, S., Murzin, A. G., Procter, M., and St Johnston, D. (1995) *EMBO J.* **14**, 4385-4391

24. Ryter, J. M., and Schultz, S. C. (1998) *EMBO J.* **17**, 7505-7513

25. Krovat, B. C., and Jantsch, M. F. (1996) *J. Biol. Chem.* **271**, 28112-28119

26. Smith, J. C., and Slack, J. M. W. (1983) *J. Embryol. Exp. Morphol.* **78**, 299-317
27. Nieuwkoop, P. D., and Faber, J. (1967) *Normal Table of Xenopus laevis, Daudin*, North Holland, Amsterdam

28. Guille, M. (1999) in *Molecular Methods in Developmental Biology* (Guille, M., ed) Vol. 127, pp. 111-125, Humana Press

29. Moore, W. M., and Guille, M. J. (1999) in *Molecular Methods in Developmental Biology*. (Guille, M. J., ed), Vol 127, pp. 99-111, Humana Press, New Jersey

30. Robinson, C., and Guille, M. (1999) in *Molecular methods in developmental biology* (Guille, M., ed) Vol. 127, pp. 89-97, Humana Press

31. Steinbach, O. C., and Rupp, R. A. W. (1999) in *Molecular Methods in Developmental Biology* (Guille, M., ed) Vol. 127, pp. 41-56, Humana Press, Totowa, N.J.

32. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982), Cold Spring Harbour, Laboratory Press

33. Maxam, A., and Gilbert, W. (1980) *Meth. Enzymol.* 65, 499-560

34. Jimenez, G., Paroush, Z., and Ish-Horowicz, D. (1997) *Genes Dev* 11, 3072-3082

35. Green, J. (1999) in *Molecular methods in developmental biology*. (Guille, M., ed) Vol. 127, pp. 1-13, Humana Press, Totowa, N.J.

36. Ramos, A., Grunert, S., Adams, J., Micklem, D. R., Proctor, M. R., Freund, S., Bycroft, M., St Johnston, D., and Varani, G. (2000) *EMBO J.* 19, 997-1009

37. Elgar, S. J. (2001) Ph.D thesis, University of Portsmouth

38. Basham, B., Schroth, G. P., and Ho, P. S. (1995) *Proc Natl Acad Sci U S A* 92, 6464-6468

39. Guille, M. J., and Kneale, G. G. (1997) *Molecular Biotechnology* 8, 35-52
40. Lanz, R. B., McKenna, N. J., Onate, S. A., Albrecht, U., Wong, J., Tsai, S. Y.,
Tsai, M. J., and O'Malley, B. W. (1999) Cell 97, 17-27

41. Izumi, H., Imamura, T., Nagatani, G., Ise, T., Murakami, T., Uramoto, H.,
Torigoe, T., Ishiguchi, H., Yoshida, Y., Nomoto, M., Okamoto, T., Uchiumi, T.,
Kuwano, M., Funa, K., and Kohno, K. (2001) Nucleic Acids Res 29, 1200-1207

42. Liu, W., Seto, J., Sibille, E., and Toth, M. (2003) Mol Cell Biol 23, 4083-4093

43. Connolly, K. M., Wojciak, J. M., and Clubb, R. T. (1998) Nat Struct Biol 5, 546-
550

44. Fleenor, D. E., Langdon, S.D., Decastro, C. M., and Kaufman, R. E. (1996) Gene
179, 219-223

45. Minegishi, N., Ohta, J., Suwabe, N., Nakaushi, H., Ishihara, H., Hayashi, N., and
Yamamoto, M. (1998) J. Biol. Chem. 273, 3625-3634
Figure legends

Figure 1. Domain structure of CBTF\textsuperscript{122} and western blotting of CBTF\textsuperscript{122}EN or CBTF\textsuperscript{122}mRBDE\textsubscript{EN} injected embryos. (a) Schematic diagram of the domain organisation of the *Xenopus laevis* transcription factor CBTF\textsuperscript{122} and the two truncated proteins used in the EMSA studies. The amino acid numbers of the domains are given relative to the start of translation. The location of the two single phenylalanine to alanine point mutations used (F435A and F559A) are shown in each of the RBDs. (b) Embryos were injected with 37 pg of either CBTF\textsuperscript{122}EN (i) or CBTF\textsuperscript{122}VP16 (ii) mRNA and compared to uninjected embryos (iii). (c) Sets of 20 embryos were injected with 37 pg of the RNA shown and allowed to develop to stage 11. Embryo lysate was prepared and yolk proteins removed by freon extraction prior to analysis of the exogenous protein by western blotting using an antibody recognising CBTF\textsuperscript{122}.

Figure 2. Summary of EMSA data for the recombinant proteins CBTF\textsuperscript{122}(41)GST and CBTF\textsuperscript{122}(28.5)GST. The recombinant proteins CBTF\textsuperscript{122}(28.5)GST and CBTF\textsuperscript{122}(41)GST forms of the CBTF\textsuperscript{122} RNA binding domains were expressed in *E. coli* as fusions with Glutathione-S-Transferase. The purified proteins were used in EMSA assays with 0.3 nM of either the dsRNA, dsDNA or ssRNA version of the CBTF\textsuperscript{122} binding sequence as a probe. (a) EMSA data using the 36bp synthetic dsRNA probe for CBTF\textsuperscript{122}(41)GST and CBTF\textsuperscript{122}(28.5)GST. (b) EMSA data for CBTF\textsuperscript{122}(28.5)GST using the dsRNA, ssRNA, or dsDNA 36 bp probes. All data were analysed using a Molecular Dynamics phosphorimager. The proportion of bound probe at each protein concentration
was calculated using: % bound = bound/(free+bound)x100 and plotted against protein concentration. (c) The sequences of the probes used in the assays, the CCAAT sequence is boxed.

Figure 3. EMSA of CBTF\textsuperscript{122}mRBD(28.5)GST binding to RNA or DNA, and competition for the CBTF\textsuperscript{122}(28.5)GST-DNA complex by mutant probes. The wild-type (CBTF\textsuperscript{122}(28.5)GST) and mutant (CBTF\textsuperscript{122}mRBD(28.5)GST) forms of the CBTF\textsuperscript{122} RNA binding domains were expressed as fusions with Glutathione-S-Transferase. The purified proteins were used in EMSA assays with either the RNA or DNA version of the CBTF\textsuperscript{122} binding sequence as a probe. (a) 0.3nM RNA was mixed with 0, 0.15, 0.3, 0.45, 0.6nM CBTF\textsuperscript{122}(28.5)GST (lanes a1-a5) or CBTF\textsuperscript{122}mRBD(28.5)GST (lanes a6-a10). (b) 0.3 nM dsDNA oligonucleotide was incubated with 0, 3, 6, 9, 18nM CBTF\textsuperscript{122}(28.5)GST (lanes b1-b5) or CBTF\textsuperscript{122}mRBD(28.5)GST (lanes b6-b10). (c) 100nM CBTF\textsuperscript{122}(28.5)GST was bound to 0.3nM probe and analysed using EMSA. Cold competitor oligonucleotides (see table 3) were titrated in a range of 1, 5 and 10-fold excess over the probe. The percentage of inhibition of CBTF\textsuperscript{122} binding to the native probe was calculated at a 10-fold molar excess over probe.

Figure 4. Circular dichorism of P1, P2, C1 and C2 oligonucleotides. (a) CD Spectra from 205 nm to 360nm for :- (1) oligonucleotide C1 (B-form) (2) oligonucleotide P2 (3) oligonucleotide C2 (A-form) (4) oligonucleotide P1. All samples were in 100mM KF, 5mM NaH\textsubscript{2}PO\textsubscript{4} buffer pH 7.6 at 22°C. (b) Calibration curve derived from CD signal intensity at 278 nm and assuming C1 to be 100% B-form and C2 to be 0% B-form. The
equation defining the line was used to predict percent B-form for P1 and P2. (c) Comparison of the observed P1 CD spectrum with a calculated spectrum based on a duplex predicted to contain 65% A and 35% B-form.

Figure 5. Competition EMSA of the multisubunit CBTF-DNA complex by mutant probes. (a) In all cases end-labelled wild-type CCAAT probe (4 fmol, *) was used in a standard binding reaction to assay CBTF from crude embryo extracts. Competitor oligonucleotides (see table 3) were in the binding reactions at a 5-, 10- or 50-fold molar excess prior to the addition of embryo extracts and the subsequent separation of DNA-protein complexes using standard EMSA conditions. Samples were run alongside free probe (lane F) and in the absence of competitor (lane S). The specific CBTF complex is arrowed. Other complexes observed are non-specific and are a consequence of using crude embryo extracts, the only current source of the intact CBTF complex. The gel shown in the lower right panel has been run less far than the other three. (b) The percentage of inhibition of CBTF binding to the native probe was calculated when the EMSA reactions contained competitor at a 10-fold molar excess over probe. The data are means from three experiments.

Figure 6. DMS interference footprinting of the CBTF-DNA complex. A DNA probe corresponding to bases –66 to –22 from the GATA-2 promoter with either radio-labelled + or – strand underwent exposure to DMS prior to preparative EMSA using crude embryo extract. Free and bound species were recovered from the EMSA gel and cleaved; the resulting DNA fragments were resolved on a 16% denaturing polyacrylamide gel.
Methylated bases that interfere strongly with binding are marked by ●, and those that interfere partially by ○. The region corresponding to the CCAAT sequence is boxed.
### Table 1

| Injected RNA     | Number injected | % double axis | % headless |
|------------------|-----------------|---------------|------------|
| CBTF\(^{122}\)EN | 21              | 81            | 0          |
| CBTF\(^{122}\) mRBDEN | 38          | 0             | 0          |
| CBTF\(^{122}\) VP16 | 56            | 0             | 18         |
| CBTF\(^{122}\) mRBDVP16 | 66         | 0             | 1          |
| Uninjected       | 30              | 0             | 0          |

Table 1. Analysis of phenotypes arising from CBTF\(^{122}\)EN, CBTF\(^{122}\) mRBDEN, CBTF\(^{122}\) VP16, and CBTF\(^{122}\) mRBDVP16 mRNA injection into *Xenopus* embryos. The synthetic mRNAs (37pg) were injected into the marginal zones of 8-cell *Xenopus laevis* embryos and these were allowed to develop to stage 35. Embryos were scored blind for the formation of double axes or microcephaly.
Table 2

| Injected RNA       | Threshold | ΔCt   | ΔΔCt (injected-uninjected) | 2^{ΔΔCt} GATA2 fold change |
|--------------------|-----------|-------|---------------------------|---------------------------|
|                    | GATA2 | ODC   | (GATA2-ODC) |                     |                           |
| CBTF^{122}EN       | 37.40  | 19.74 | 17.66      | 10.36                | 0.00076                   |
| CBTF^{122}mRBDE    | 29.63  | 19.03 | 10.60      | 3.30                 | 0.1                       |
| Uninjected         | 26.83  | 19.53 | 7.30       | 0.00                 | 1                         |

Table 2. qRT-PCR determination of GATA-2 mRNA levels following CBTF^{122}EN or CBTF^{122}mRBDE mRNA injection into Xenopus embryos. Two cell embryos were injected in the animal pole with 37pg of synthetic RNA encoding either CBTF^{122}EN or CBTF^{122}mRBDE. The embryos were allowed to develop to stage 8 and the animal caps removed. These were cultured until sibling embryos reached stage 11. Relative amounts of GATA-2 mRNA were measured using real time RT-PCR. Data were analysed by the ΔΔCt method (ABI), where Ct is the threshold cycle and ΔΔCt is normalised for amount of input cDNA by comparison with the control gene (ODC) and levels relative to a calibrator sample (uninjected embryos).
Table 3. Oligonucleotides used in competition EMSA. Double stranded oligonucleotides were annealed and used in competition EMSA. Wild-type sequence of the –59 to –31 region of the GATA-2 promoter and mutants based around it are designated P1 to P5, the mutated bases are highlighted in bold. Control oligonucleotides are designated C1 to C3, APEs (Basham et al) are also shown. The percent of each duplex predicted to be A and B-form is also shown, this does not

| name | description | sequence                                      | APE   | Predicted  |
|------|-------------|-----------------------------------------------|-------|------------|
| P1   | -59 to -31  | GCGGAGGCTTGTGATTTGGCTGGCCCGGGGCGCCTCGAACACTAACCAGGCGGGCCC | -0.51 | 58%A, 24%B |
| P2   | -56 to -50mt| GCGTTTTAAAAGTATGGCTGGCCCGGGGCGCAAAATTTCACTAACCAGGCGGGCCC | +0.42 | 35%A, 52%B |
| P3   | -41 to -39mt| GCGGAGGCCTTTGTGATTGGAGATGCCCGGGGCGCCTCGAACACTAACCCTAACGCCCC | -0.2  | 69%A, 24%B |
| P4   | -44mt       | GCGGAGGCTTGTGATCGGCTGGCCCGGGGCGCCTCGAACACTAGGCACCGGGCC | -0.42 | 55%A, 26%B |
| P5   | -49 and -55mt| GCGGCGGGCTTTGTGATTGGCTGGCCCGGGGCGCCCGAGAACACTAACCAGGCGGGCCC | +0.16 | 31%A, 48%B |
| P6   | No CCAAT    | GCGGAGGCTTGTGTATTTCCTGGCCCGGGGCGCCTCGAACATAAGACCGGGGGCCC | -0.13 | 41%A, 34%B |
| C1   | B Form      | CGCGAAATTCGCCGCGCAATTTGGCGGCGCGGTAAAACGCGGGCC | +0.35 | 14%A, 73%B |
| C2   | A Form      | GTGCGCACCCTCCGGCGCGCGCGCGCGCGCGCGCGCGCGGGCGCGCGCGCGGGCC | -0.29 | 62%A, 31%B |
| C3   | (AT)\(_{14}\)A | ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT | +0.53 | 0%A, 93%B |
sum to 100% due to the inability to calculate these data for the duplex ends and for certain sequences.
Figure 1

(a) Diagram showing the regions of CBTF122 and its variants.

(b) Images showing morphological details:
   i. Second axis
   ii. Engrailed
   iii. Other morphological features

(c) Western blot images for Uninjected, CBTF122EN, and CBTF122mRBDEN.
Figure 2

(a) Fraction bound %

(b) Fraction bound %

(c) dsDNA CCAAT probe
CGA.CGCA.CGA.CGACAC.CAACC.ACC.GCG
GGT.GACG.ACG.GCGACG.ACCG.GCC

dsRNA CCAAT probe
CGA.CGCA.CGA.CGACAC.CAACC.ACC.GCG
GGT.GACG.ACG.GCGACG.ACCG.GCC

ssRNA probe
CUGA.CGCA.CGA.CGACAC.CAACC.ACC.GCG
GGT.GACG.ACG.GCGACG.ACCG.GCC
Figure 5
Figure 6

|            | Plus strand | Minus Strand |
|------------|-------------|--------------|
|            | Free        | Bound        | Free         |
| T          | G           | C            | T            |
| T          | G           | C            | A            |
| T          | T           | A            | C            |
| G          | G           | A            | C            |
| G          | A           | C            | C            |
| G          | G           | G            | G            |

CTGGCTGGCGGGAGGCTTGTGATTGGCTGGCCCGGTTATTAGCT
GACCGACCGCTGCCGAACAATAACCGGACCAGCGCCCAATAATCGA