

The N-terminal fingers of chicken GATA-2 and GATA-3 are independent sequence-specific DNA binding domains

Paolo V.Pedone, James G.Omichinski¹,
Pascale Nony, Cecelia Trainor,
Angela M.Gronenborn¹, G.Marius Clore¹
and Gary Felsenfeld²

Laboratory of Molecular Biology, NIDDK, Building 5, National Institutes of Health, Bethesda, MD 20892-0530 and ¹Laboratory of Chemical Physics, NIDDK, Building 5, National Institutes of Health, Bethesda, MD 20892-0520, USA

²Corresponding author

P.V.Pedone and J.G.Omichinski contributed equally to this work

The GATA family of vertebrate DNA binding regulatory proteins are expressed in diverse tissues and at different times of development. However, the DNA binding regions of these proteins possess considerable homology and recognize a rather similar range of DNA sequence motifs. DNA binding is mediated through two domains, each containing a zinc finger. Previous results have led to the conclusion that although in some cases the N-terminal finger can contribute to specificity and strength of binding, it does not bind independently, whereas the C-terminal finger is both necessary and sufficient for binding. Here we show that although this is true for the N-terminal finger of GATA-1, those of GATA-2 and GATA-3 are capable of strong independent binding with a preference for the motif GATC. Binding requires the presence of two basic regions located on either side of the N-terminal finger. The absence of one of these near the GATA-1 N-terminal finger probably accounts for its inability to bind. The combination of a single finger and two basic regions is a new variant of a motif that has been previously found in the binding domains of other finger proteins. Our results suggest that the DNA binding properties of the N-terminal finger may help distinguish GATA-2 and GATA-3 from GATA-1 and the other GATA family members in their selective regulatory roles *in vivo*.

Keywords: DNA binding proteins/erythroid gene regulation/GATA proteins/zinc fingers

Introduction

GATA proteins are members of a zinc finger subfamily of DNA binding proteins that recognize the consensus motif (T/A)GATA(A/G) (Orkin, 1992). Six GATA family members have been identified in vertebrates (Evans and Felsenfeld, 1989; Tsai,S.F. *et al.*, 1989; Yamamoto *et al.*, 1990; Arceci *et al.*, 1993; Kelley *et al.*, 1993; Laverriere *et al.*, 1994). The proteins GATA-1, GATA-2 and GATA-3 are required for normal hematopoietic development in the mouse (Pevny *et al.*, 1991; Tsai,F.Y. *et al.*, 1994; Pandolfi *et al.*, 1995). GATA-1 expression is primarily restricted

to some of the cells of the myeloid lineage (erythroid cells, mast cells and megakaryocytes; Martin *et al.*, 1990; Romeo *et al.*, 1990; Whitelaw *et al.*, 1990; Yamamoto *et al.*, 1990), with the notable exception of abundant expression in Sertoli cells of the testis (Ito *et al.*, 1993; Yomogida *et al.*, 1994). GATA-2 is expressed in a wide variety of tissues, including hematopoietic progenitors, erythroid cells, mast cells, megakaryocytes, endothelial cells and embryonic brain cells (Yamamoto *et al.*, 1990; Dorfman *et al.*, 1992; Visvader and Adams, 1993; Leonard *et al.*, 1993; Mouthon *et al.*, 1993). GATA-3 is highly expressed in embryonic brain cells and T lymphoid cells but is also found in other tissues (Yamamoto *et al.*, 1990; Oosterwegel *et al.*, 1992; George *et al.*, 1994). The GATA-4, GATA-5 and GATA-6 genes are expressed in several non-hematopoietic cell types and they may be involved in regulating cardiogenesis and differentiation of gut epithelium (Arceci *et al.*, 1993; Kelley, 1993; Laverriere *et al.*, 1994).

All of the GATA proteins interact with their consensus target sequence via zinc fingers of the form Cys-X₂-Cys-X₁₇-Cys-X₂-Cys. The vertebrate GATA factors contain two adjacent zinc finger domains of this type, while fungal proteins have a single finger (Fu and Marzluf, 1990; Kudla *et al.*, 1990), more similar to the C-terminal vertebrate finger. The C- and N-terminal fingers (Cf and Nf) of the vertebrate proteins are clearly related and they are both well conserved among all the known homologs.

In several studies, primarily with the GATA-1 protein (Martin and Orkin, 1990; Yang and Evans, 1992; Omichinski *et al.*, 1993b), it has been shown that the principal determinant for DNA binding is the C-terminal finger. This finger, along with a basic C-terminal tail, has been shown to be necessary and sufficient for specific binding to the consensus GATA recognition sequence (Omichinski *et al.*, 1993a,b). A mutant protein containing only the N-terminal finger does not bind to this site (Martin and Orkin, 1990; Yang and Evans, 1992). The high degree of conservation of the N-terminal finger throughout vertebrate evolution, however, suggests that it likely plays some crucial role in the function of the GATA proteins. In fact, the N-terminal finger has been reported to contribute to full specificity and stability of binding to DNA (Martin and Orkin, 1990; Yang and Evans, 1992). High affinity binding to several double GATA sites, important for full activity of the promoters in which they occur, requires both the N- and C-terminal fingers of the GATA-1 protein (Trainor *et al.*, 1996).

In this report, we demonstrate that the isolated N-terminal fingers of the chicken GATA-2 and GATA-3 proteins (cGATA-2 and cGATA-3) are capable of specific high affinity binding to DNA. This binding is independent of the C-terminal finger and requires two basic arms that flank the zinc finger motif. The N-terminal finger binds

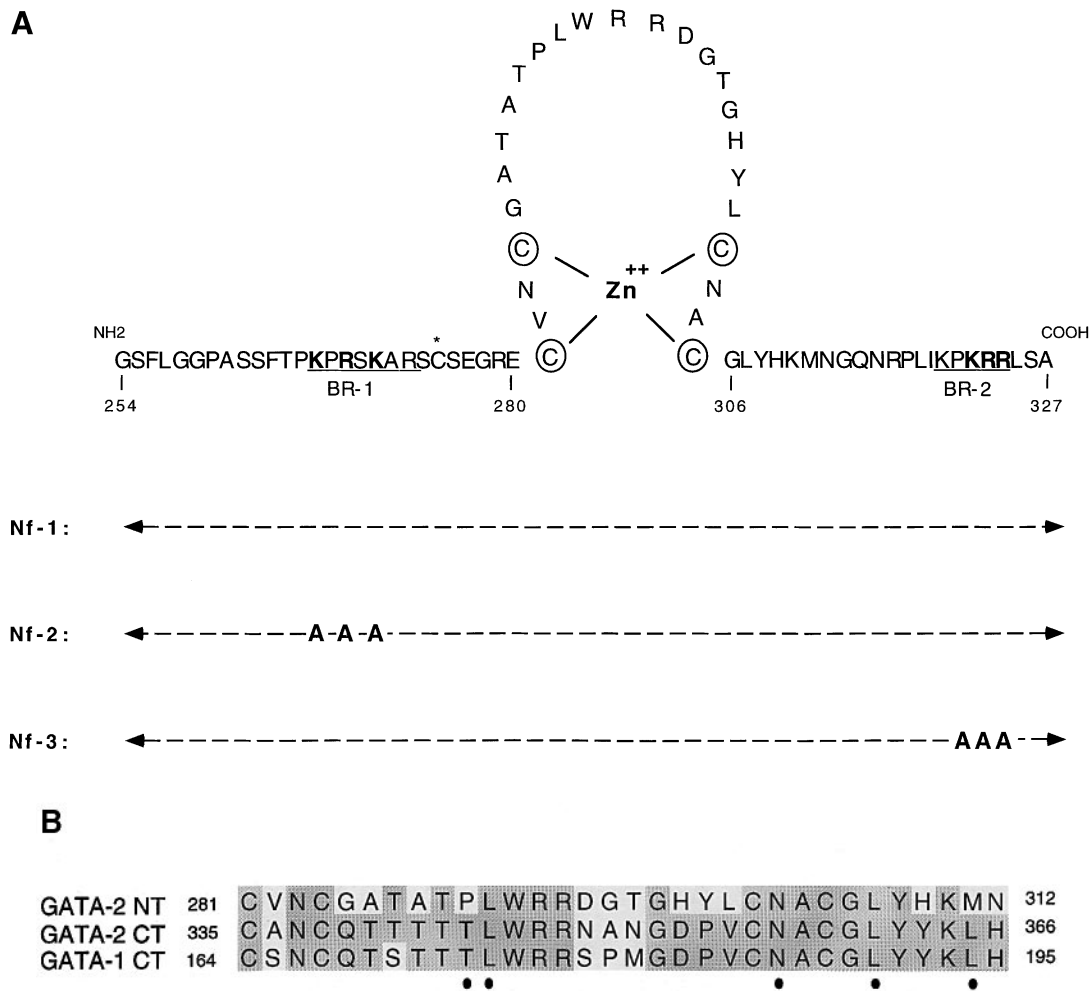


Fig. 1. (A) Schematic representation of the amino acid sequence of the N-terminal finger region of cGATA-2 and the peptides (Nf1, Nf2 and Nf3) used in the binding studies. The underlined sequences indicate the basic regions flanking the zinc finger motif (BR1 and BR2). The basic residues which have been mutated in the Nf2 and Nf3 peptides are indicated in bold. In all the peptides utilized, Cys275 (indicated with an asterisk) was substituted by a serine in order to increase the stability and ease of handling of the proteins. Control experiments, comparing Nf1 with a wild-type peptide, demonstrated that this substitution does not affect DNA binding. (B) Comparison of the sequence of the N- and C-terminal fingers of the cGATA-2 protein and the C-terminal finger of the cGATA-1 protein. Positions of amino acid identity within the finger regions are shaded. Closed circles indicate the residues of the cGATA-1 C-terminal finger which make base-specific contacts in the major groove with the sequence AGATAA.

weakly to the classical AGATAA consensus sequence, while it is able to bind with high affinity to an AGATCT motif. In contrast, the N-terminal finger of the chicken GATA-1 protein (cGATA-1), which lacks an adjacent N-terminal basic region, failed to display DNA binding to these sites.

Although some differences in binding specificity have been noted between the GATA family members, there are as yet no definitive explanations of how the different members discriminate among different sites. This issue is particularly relevant considering that different GATA factors display partially overlapping expression patterns and thus more than one GATA factor can be present in the same cell. Our finding that the N-terminal fingers of cGATA-2 and cGATA-3, but not that of cGATA-1, are able to bind independently to DNA represents a significant difference in the DNA binding domains of these proteins. This result suggests possible mechanisms by which these factors could distinguish among different DNA sequences.

Results

It has been shown recently that a stretch of basic amino acids located on either the N-terminal or C-terminal side of a zinc finger motif is necessary to stabilize binding of a single finger unit to the DNA (Omichinski *et al.*, 1993b; Pedone *et al.*, 1996). Because the N-terminal zinc finger (Nf) motif of the GATA-2 protein is flanked on both sides by stretches of basic amino acids (BR1 and BR2, Figure 1A), we wished to determine whether this zinc finger region is able to bind independently to DNA.

The sequence corresponding to residues 254–327 of the GATA-2 protein (Yamamoto *et al.*, 1990), which includes the zinc finger domain and the two flanking basic regions (Figure 1A), was subcloned and expressed in *Escherichia coli*. The recombinant protein (Nf-1) was purified and its identity confirmed by electrospray ionization mass spectrometry. (Note that in Nf-1, Cys275 was substituted by a serine in order to increase the stability and ease of handling of the protein. Control experiments, comparing

A

Probes:

A: AGGTTGCAGATAAACATTTT

B: AGGTTGCAGATCTTAATTTT

C: AGGTTGCAGATCTACATTTT

B

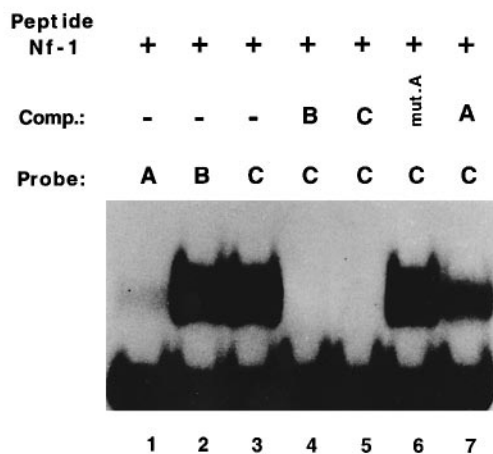


Fig. 2. Analysis of DNA binding of the Nf1 peptide. (A) Sequences of the different oligonucleotides utilized as probes in the binding experiments. The core consensus sequence in each oligonucleotide is underlined. (B) Gel mobility shift analysis of binding of the Nf1 peptide to the different probes. Where indicated, unlabeled oligonucleotides were added as competitor in a 100-fold excess. The sequence of the oligonucleotide mut. A, in which the GAT core sequence is mutated, is indicated in Materials and methods. Comp., competitor.

Nf-1 with a wild-type peptide, demonstrated that this substitution has no effect on DNA binding; data not shown.)

The sequence similarity between the N-terminal and C-terminal fingers of the GATA proteins (Figure 1B) suggested that the Nf might also bind to a sequence containing the GATA core motif. Furthermore, the N-terminal finger of GATA-1 has been shown to interact with sites containing the GATA consensus sequence when two sites occur in close proximity or overlap each other, although this binding requires the C-terminal finger (Trainor *et al.*, 1996). For these reasons, an oligonucleotide containing the AGATAA consensus (oligo A, Figure 2A), derived from the sequence of the chicken β - ϵ enhancer, was tested in a gel shift experiment. As shown in Figure 2B, the N-terminal finger of GATA-2 binds to oligo A, but only weakly (lane 1). This binding was demonstrated to be specific by competition experiments (data not shown). Because of the weakness of binding to this site, two additional oligonucleotides (oligo B and C, Figure 2A)

were tested. Oligo B contains the sequence AGATCTTA. This is a previously proposed consensus derived from sequences identified in binding experiments as having high affinity for GATA-2 and GATA-3, but lower affinity for GATA-1 (Ko and Engel, 1993). Oligo C contains only the AGATCT core present in oligo B and is otherwise identical to oligo A (see Figure 2A). Nf-1 binds tightly to both oligos B and C, as is clearly demonstrated by the gel shift analysis shown in Figure 2B (lanes 2 and 3). Binding to oligo C is completely competed for by a 100-fold molar excess of unlabeled oligo B or C (Figure 2B, lanes 4 and 5), but not by the same molar excess of an oligonucleotide mutated in the GAT central motif (mut. A, Figure 2B, lane 6). The same molar excess of oligo A competes only weakly for binding (Figure 2B, lane 7). These results indicate that the N-terminal finger of the GATA-2 protein is capable of specific, high affinity binding to a DNA sequence containing the motif AGATCT.

The affinity of the GATA-2 N-terminal finger for oligos A, B and C was measured using a gel mobility shift assay. Titration of the peptide with the different oligonucleotides is shown in Figure 3. Scatchard analysis of these data leads to an apparent dissociation constant of 2.8×10^{-8} M for oligo A, which contains the AGATAA motif, and 5.2×10^{-9} M for oligo C, which contains the AGATCT motif (note that the binding reactions were carried out at 0°C; see Materials and methods). No significant difference in the affinity of the peptide for oligos B and C was detected. For comparison, the binding affinity of the C-terminal finger of the cGATA-1 protein (residues 158–223 of cGATA-1) for oligos A and C was measured under the same binding conditions. The C-terminal finger of the GATA-1 protein binds with an apparent dissociation constant of 7.8×10^{-10} M to oligo A and 2.1×10^{-9} M to oligo C (see Table I).

Our results indicate that while the N-terminal finger of the GATA-2 protein does not bind as tightly to the classical consensus sequence AGATAA (in oligo A), it does interact with high affinity with the sequence AGATCT (in oligo C). The only differences between oligos A and C reside in the nucleotides at positions +1 and +2 with respect to the GAT core sequence (see Figures 2A and 4A). In order to test which of these bases were important for the differential affinity for the peptide Nf-1, two different point mutants of oligo C (+1: C→A and +2: T→A, Figure 4A) were used in binding experiments. As demonstrated by the gel shift experiments shown in Figure 4B (lane 3), substitution of the C at position +1 by an A drastically reduced binding to a level similar to that observed with oligo A. Conversely, substitution at position +2 does not have any effect on binding (Figure 4B, lane 4). These results indicate that it is the C in position +1 with respect to the core GAT that allows Nf-1 to distinguish between oligos A and C. In order to determine if any other bases could be allowed at position +1, two new sequences, oligos D and E (Figure 4A), were tested. These oligonucleotides have a T or a G at position +1 respectively. The results presented in Figure 4C clearly demonstrate that Nf-1 can recognize either the GATT or GATG motif. (Note that all of the oligonucleotides utilized as probes in the gel shift experiment shown in Figure 4C differ only at position +1.) The binding to these sequences is stronger

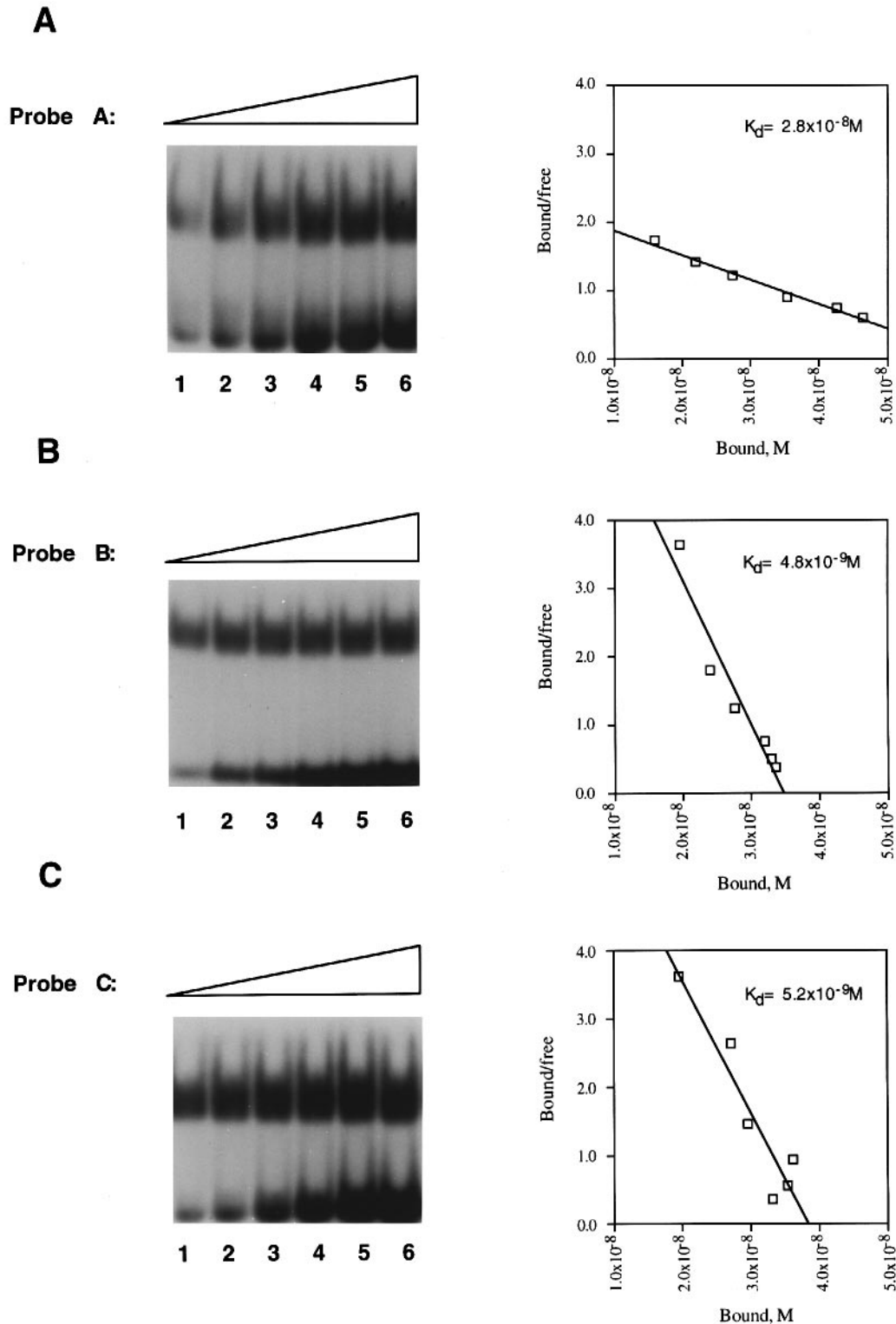


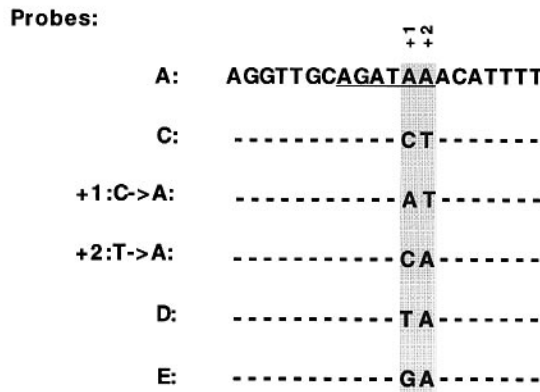
Fig. 3. Titration of the Nf1 peptide with oligos A–C. **(A)** Gel mobility shift titration of the Nf1 peptide with oligo A. Five pmol peptide were incubated in a volume of 20 μ l with 0.5, 0.75, 1, 1.5, 2 and 2.5 pmol duplex oligo A (lanes 1–6). **(B and C)** Titration of the peptide Nf1 with the oligos B and C respectively. In lanes 1–6 each sample contained, in a volume of 20 μ l, 1.7 pmol peptide and 0.5, 0.75, 1, 1.5, 2 and 2.5 pmol duplex respectively. All the reactions were performed as described in Materials and methods, but in the absence of poly(dI-dC). On the right of each gel is shown a Scatchard analysis of the binding data. The ratio of bound to free DNA is plotted versus the molar concentration of bound DNA in the reaction mixture. The value of the calculated dissociation constant is indicated. All numerical values were obtained by computer quantitation of the image using a Molecular Dynamics Phosphorimager.

than to a GATA motif, but somewhat weaker than to the GATC site.

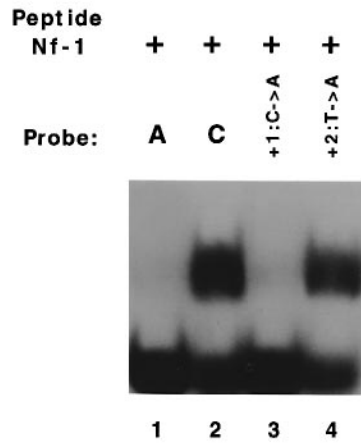
In order to test the role of the basic stretches of amino

acids which flank the zinc finger domain in stabilizing DNA binding of peptide Nf-1, two mutant peptides (Nf2 and Nf3, see Figure 1A) were tested for binding activity.

A



B



C

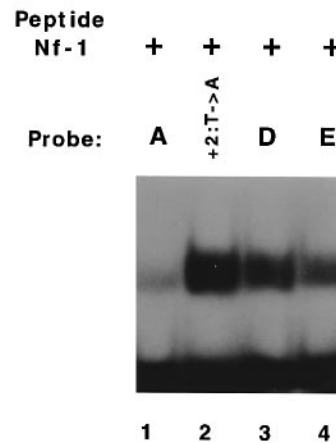


Fig. 4. Analysis of binding of the Nf1 peptide to the oligonucleotides shown in (A). (A) Sequences of the different oligonucleotides utilized as probes in the binding experiments. The shaded area highlights the nucleotides at positions +1 and +2 with respect to the GAT core sequence. The lines indicate the bases which are identical in all the different oligonucleotides. (B) Analysis of binding of the peptide Nf1 to oligos A, C, +1: C→A and +2: T→A. Oligos +1: C→A and +2: T→A are point mutants of oligo C at positions +1 and +2 respectively. (C) Analysis of binding of the peptide Nf1 to oligonucleotides which differ at position +1 with respect to the GAT core motif.

In both mutants, alanine residues were substituted for either lysine or arginine. In Nf2 the substitutions were K267A, R269A and K271A in BR1, while in peptide Nf3 the substitutions were K322A, R323A and R324A in BR2 (Figure 1A). As shown in Figure 5, both of the mutant peptides failed to bind to oligo C, even when they were used in a molar concentration 10 times higher than that used with Nf-1. No DNA binding was detected with the mutant peptides even when oligo A was used as the probe (data not shown). These results indicate that both basic arms contribute to high affinity binding of the N-terminal finger of the cGATA-2 protein to the DNA.

Interestingly, while the DNA binding domain, spanning the two zinc finger motifs and the amino acids in between them (including BR2), is highly conserved among the different GATA proteins, the basic region located upstream of the N-terminal finger (BR1) is present in cGATA-2 and

Table I. Dissociation constants of the N-terminal finger of the cGATA-2 protein (peptide Nf1 and the C-terminal finger of the cGATA-1 protein for oligos A and C)

	Dissociation constant (K_d)	
	Oligo A	Oligo C
cGATA-2 N-terminal finger	2.8×10^{-8}	5.2×10^{-9}
cGATA-1 C-terminal finger	7.8×10^{-10}	2.1×10^{-9}

cGATA-3 but not in cGATA-1 or the other family members (Figure 6A). For this reason, one might predict that the N-terminal finger of cGATA-3 would bind to DNA in a manner similar to that of the cGATA-2 N-terminal finger, while the equivalent region of the cGATA-1 protein would

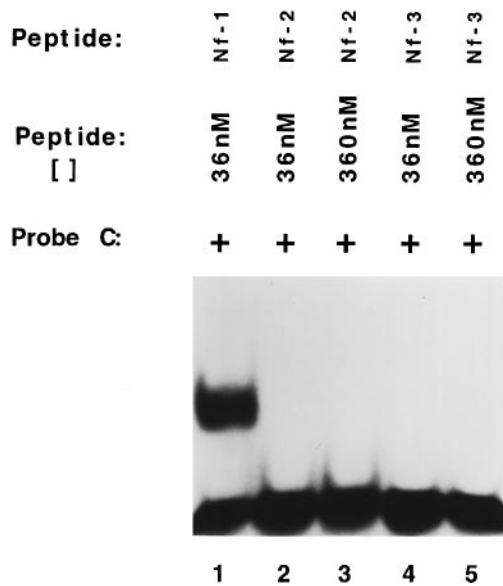


Fig. 5. Analysis of the role of the basic stretches of amino acids which flank the zinc finger domain in stabilizing DNA binding. Gel mobility shift analysis of DNA binding of the Nf1, Nf2 and Nf3 peptides. Peptide [], final concentration of the peptide in the binding reaction.

not. In order to test these predictions, peptides containing the cGATA-3 and cGATA1 N-terminal fingers (residues 232–310 for cGATA-3 and 78–156 for cGATA-1; Yamamoto *et al.*, 1990; Evans and Felsenfeld, 1989) were prepared. These peptides comprised the zinc finger region and the conserved basic arms at the C-termini of each (BR2) and both extend 32 amino acids on the N-terminal side of the finger domain. As predicted, the N-terminal finger of cGATA-3 binds with high affinity to oligo C, while the cGATA-1 N-terminal finger failed to display DNA binding (Figure 6B). We could not detect any DNA binding of the cGATA-1 N-terminal finger even when the AGATAA-containing oligonucleotide (oligo A) was used as probe (data not shown). When probes differing at position +1 were used in a gel shift analysis (Figure 7), the cGATA-3 N-terminal finger demonstrated the same specificity as the cGATA-2 N-terminal finger. These results indicate that the N-terminal fingers of cGATA-2 and cGATA-3, but not cGATA-1, are able to act as independent DNA binding domains and are consistent with the observation that two basic regions are necessary to stabilize binding of Nf to the DNA. This demonstrates a very clear difference in the DNA binding domains of the different GATA proteins.

Discussion

The GATA proteins share a similar architecture which in vertebrates includes two Cys₂-Cys₂ zinc fingers, only the C-terminal of which is essential for binding to the canonical site (T/A)GATA(A/G). There is sufficient conservation among members of the GATA family that this central DNA recognition sequence motif is shared by all. Since there are many DNA binding sites carrying this sequence and in many cells more than one member of the GATA family is expressed, this raises the question of how the

different GATA proteins are able to function differentially in activating individual genes.

Recent results indicate that despite their simplicity and common features, the individual GATA family members have developed a range of specificities that allows them to discriminate among binding sites. Random site selection experiments have revealed a variety of other binding motifs for the GATA proteins with affinities that in some cases equal or exceed that of the canonical site. Some of the results from these experiments appear to be mutually inconsistent. For example, one study (Ko and Engel, 1993) found that GATA-2 and GATA-3 bind well to sites with a central GATC, while another (Merika and Orkin, 1993) did not show selection of this sequence. In the case of GATA-1, one study (Merika and Orkin, 1993) found that the fourth position in GATA could be replaced by T or G, while another (Whyatt *et al.*, 1993) found T or C in the fourth position but not G. In part these differences may arise from variations in the selection and amplification procedures, but they also reflect the effect of the surrounding nucleotide sequence on binding. This has been discussed and partly analyzed in all of the papers cited.

The N- and C-terminal fingers of the GATA proteins have a high degree of homology (see Figure 1). Evidence from preceding studies strongly supports the view that the C-terminal finger is necessary and sufficient for binding to the canonical site, but that the N-terminal finger can provide considerable additional stability for certain more complex 'double' sites (Martin and Orkin, 1990; Yang and Evans, 1992; Trainor *et al.*, 1996). There is further evidence (Whyatt *et al.*, 1993) suggesting that GATA-1 can recognize a motif quite unrelated to the canonical one and that this depends on the presence of both the N-terminal and C-terminal fingers. Interpretation of all these results in terms of 'contributions' to binding by the individual fingers is difficult because of the possible effects arising from interactions between them.

For this reason, we have undertaken studies of the binding properties of individual fingers. In this paper, we have focused on the N-terminal fingers, which were previously thought to be incapable of displaying independent specific binding to DNA. In these experiments, it has been shown that the N-terminal fingers of cGATA-2 and cGATA-3, but not cGATA-1, are capable of binding with a pattern of sequence preference that is related to but distinct from that of the C-terminal fingers. Their ability to bind is dependent upon the presence of two basic regions, one on either side of the Cys₂-Cys₂ zinc finger domain. The data strongly suggest that the GATA-1 Nf does not bind because it lacks a basic region on the N-terminal side of the finger. This is consistent with the observation (Figure 5) that mutations in the GATA-2 Nf that alter the basic amino acids of region BR1 (mutant Nf-2) or BR2 (mutant Nf-3) abolish binding. Because the cGATA-4, cGATA-5 and cGATA-6 proteins, like cGATA-1, lack the basic arm located N-terminal of the zinc finger domain, one would predict that the N-terminal fingers of these proteins are incapable of independent high-affinity binding to DNA.

The cGATA-2 and cGATA-3 N-terminal fingers represent a novel variation on the principle governing binding of single fingers to DNA: in every case observed so far, a requirement has been shown not only for an intact finger

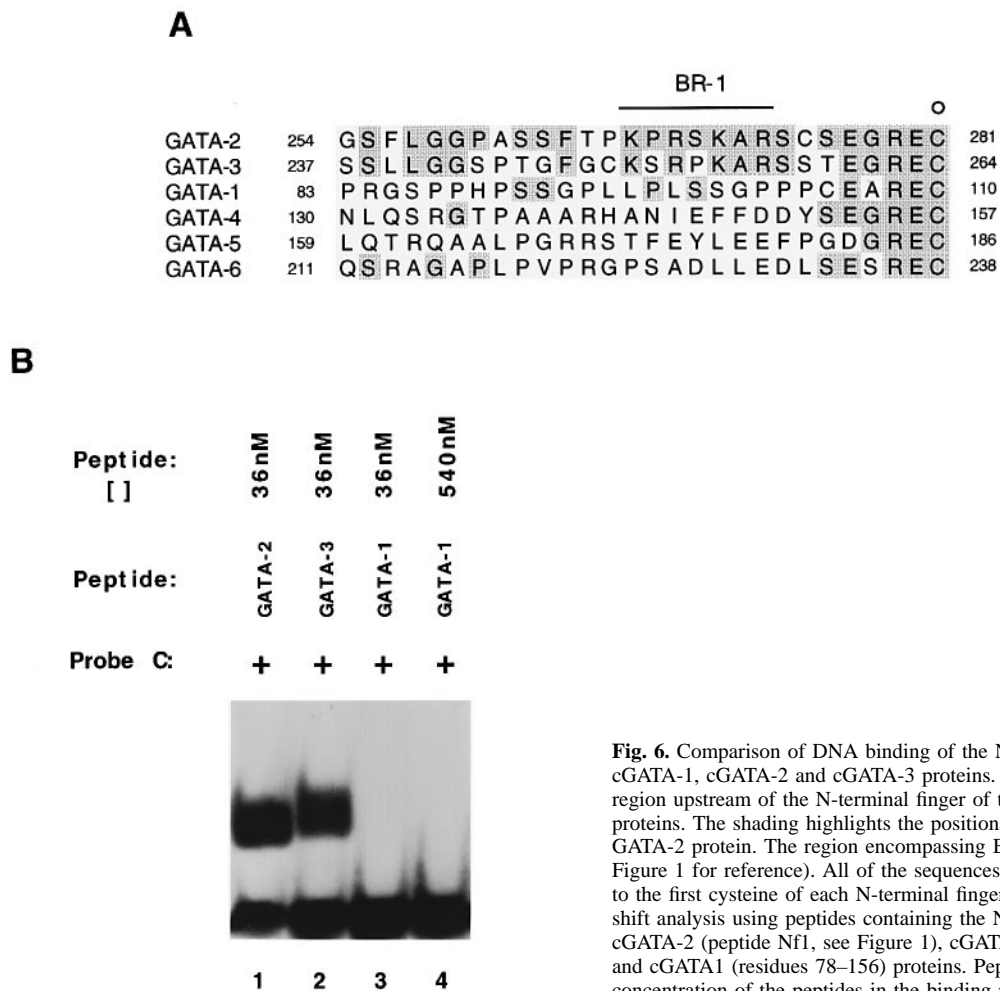


Fig. 6. Comparison of DNA binding of the N-terminal fingers of the cGATA-1, cGATA-2 and cGATA-3 proteins. **(A)** Comparison of the region upstream of the N-terminal finger of the different cGATA proteins. The shading highlights the positions of identity with the GATA-2 protein. The region encompassing BR1 is indicated (see Figure 1 for reference). All of the sequences are aligned with respect to the first cysteine of each N-terminal finger (o). **(B)** Gel mobility shift analysis using peptides containing the N-terminal fingers of the cGATA-2 (peptide Nf1, see Figure 1), cGATA-3 (residues 232–310) and cGATA1 (residues 78–156) proteins. Peptide [], final concentration of the peptides in the binding reaction.

region but also for an adjacent domain containing clusters of basic amino acid residues. This was first demonstrated for the GATA-1 Cf, in which a C-terminal basic region binding in the minor groove is required to stabilize the complex (Omichinski *et al.*, 1993a). A second example is the single Cys₂-His₂ finger of the GAGA binding protein, with two basic N-terminal regions, one binding in the major groove and the other in the minor groove (Omichinski *et al.*, 1997). This independence of binding carries with it an altered specificity: while both the GATA-2 and GATA-3 Nf are capable of binding independently to the AGATAA site, they do so with greatly reduced affinity relative to oligonucleotides containing GATC, GATG or GATT core sequences. In contrast, the GATA-1 Cf, although it also binds with high affinity to GATC (oligo C), shows an even greater preference for the GATA motif (oligo A, see Table I). It is striking that the most stringent requirement for high affinity binding of the GATA-2 and GATA-3 Nf is the absence of A at the fourth position. Mutations 3' of this position seem to have no observable effect on binding of the GATA-2 Nf (Figure 4).

It should be noted that the strong homology between the central regions of Cf and Nf includes residues that are important for interaction of the GATA-1 Cf with the sequence AGATAA (see Figure 1B), as determined by earlier structural studies (Omichinski *et al.*, 1993a). Because the Nf requires two distinct basic arms, one

located at its N-terminus and the other at its C-terminus, for high affinity binding, while the C-terminal finger utilizes only basic residues located at its C-terminus, one could predict that these two similar finger units will bind the DNA in a slightly different way. This may partly explain why the N-terminal finger recognizes a GATC motif better than a GATA motif. Our data suggest that the GATA protein DNA binding domain is a composite. In the case of the GATA-2 and GATA-3 proteins, this domain appears to be composed of at least five different elements: two finger units and three basic arms. Each of these elements can interact with the DNA and likely, depending on the DNA binding site, varying numbers of elements will be involved in recognition of the target sequence. This modular nature would certainly contribute to the versatility of these proteins in recognizing DNA target sites.

The unexpected ability of the N-terminal fingers to bind independently raises questions about the way in which GATA-2 and GATA-3 interact with their sites on DNA. It has been reported that *Elt-1*, a GATA protein from *Caenorhabditis elegans*, binds to a site containing a GATC core sequence only when both fingers are present and binds best when the site contains both GATC and GATA elements (Shim *et al.*, 1995). Preferential binding of GATA-2 (and GATA-3) to several sites which contain a GATC motif has also been reported (Ko and Engel, 1993).

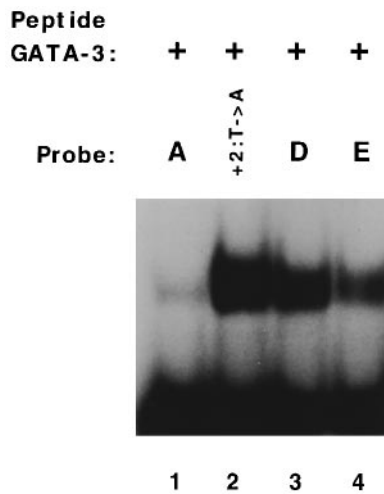


Fig. 7. Analysis of binding of the peptide encompassing the N-terminal finger of the cGATA-3 protein (residues 232–310) to oligonucleotides which differ at position +1 with respect to the GAT core motif.

As noted above, it must be kept in mind that two fingers acting in concert may have different properties from those of the separated fingers. Some earlier studies have addressed the contributions of the N-terminal finger in the context of the full-length protein. For example, deletion of the two basic regions on either side of the N-terminal finger of GATA-3, leaving the entire GATA-3 molecule otherwise intact, does not abolish binding to a target AGATAA sequence (Yang *et al.*, 1994), although inspection of the gel shift patterns suggests a decreased affinity; comparison with the effect of deleting the entire Nf suggests that the basic regions contribute most of the extra binding affinity. In other experiments (Merika and Orkin, 1993), the entire double finger region of hGATA-3 (excluding the N-terminal basic region) was substituted for the corresponding mGATA-1 region and binding was measured to a variety of sequence motifs previously identified by random site selection. This substitution conferred on the protein the sequence selective properties of wild-type hGATA-3 rather than those of mGATA-1, which might be consistent with a slightly different binding selectivity of the finger regions of the different GATA proteins. Finally, other experiments have shown that for some biological functions GATA-1, GATA-3 and GATA-4 are interchangeable (Blobel *et al.*, 1995) or that the C-terminal finger of GATA-1 or GATA-2 alone may suffice (Visvader *et al.*, 1995). This obviously does not preclude distinct roles in other circumstances for the individual GATA family members or for their N-terminal fingers.

The independent binding activities of the N- and C-terminal fingers raise the possibility that some simple sites may bind GATA-2 or GATA-3 in ways in which the N-terminal finger plays the major role. In the case of 'compound' sites involving partial or complete direct or inverted repeats of these motifs, the interaction capabilities of the N-terminus must also be considered and may help distinguish the biological activities of GATA-2 and GATA-3 from that of GATA-1. Structural analysis of these complexes will help to resolve such issues. In any case, it is clear that the rather small differences among the

GATA proteins and their components may be sufficient to give rise to considerable variation both in site selectivity and potential for transcriptional activation, as required by the multiple and diverse roles of these proteins during development.

Materials and methods

Cloning and purification of the peptides

DNA fragments encoding the different peptides were generated by PCR from plasmids containing the cDNA sequences of the different chicken GATA proteins. Oligonucleotides were synthesized on the basis of the published sequences (Evans and Felsenfeld, 1989; Yamamoto *et al.*, 1990). The following oligonucleotides were used as primers: primer 1, 5'-ACATGCCATGGGCAGCTTCTGGGGGGCC-3', and primer 2, 5'-CGGGATCCTTACGCTGACAGCCTTCGTTTGG-3', for the N-terminal finger of GATA-2; primer 3, 5'-CGGGATCCGAGCTGCTGCAACCCCCC-3', and primer 4, 5'-CGGAATTCTTACACCAGCAGGGCCTTTTGG-3', for the N-terminal finger of GATA-1; primer 5, 5'-ACATGCCATGGGACTTTTCCCCCAGCAG-3', and primer 6, 5'-CGGGATCCTTATGCAGACAGCCTTCTTGG-3', for the N-terminal finger of GATA-3.

In order to generate the peptide Nf1, we substituted Cys275 of the GATA-2 protein with a serine by PCR-mediated mutagenesis (White, 1993). This substitution increased the stability and ease of handling of the peptide. The peptide showed the same binding properties as the wild-type and so we used it in all the experiments in this paper. The plasmid including the coding sequence of the Nf1 peptide was used as a template to generate by PCR the mutants Nf2 and Nf3. For cloning of the coding sequence for peptide Nf2, we used primer 2 and primer 7, 5'-ACATGCCATGGGCAGCTTCTGGGGGGCCGGCCTCCAGCTTACCCCGCCAGCAAGCGCGGCCAGATCCTCATCAGAAAG-3'. Primer 1 and primer 8, 5'-CGGGATCCTTACGCTGACAGAGCAGCGGGTTAATGAGAGTCCGGT-3', were used to generate Nf3. All of the PCR products, except that encoding the N-terminal finger of the GATA-1 protein, were digested with the restriction enzymes *NcoI* and *BamHI* and cloned into a *NcoI/BamHI*-digested pET-11D (Novagen) expression vector. The PCR fragment encoding the cGATA-1 N-terminal finger (residues 78–156) was, instead, subcloned into the *BamHI* and *EcoRI* restriction sites of the pGEX-2T expression vector (Pharmacia) in-frame with a thrombin cleavage site and the protein was expressed as a glutathione S-transferase (GST) fusion. The peptide including the C-terminal finger of the cGATA-1 protein (residues 158–223 of cGATA-1) has already been described (Omichinski *et al.*, 1993a,b).

All proteins were expressed in BL21(DE3) cells. Cells were grown overnight at 37°C, after which protein expression was induced for 4 h with 0.5 mM isopropyl-β-D-thiogalactoside (IPTG). The cGATA-1 N-terminal finger-GST fusion protein was purified using a published procedure (Smith and Johnson, 1988). The fusion protein was then cleaved with thrombin (1 U/ml) for 4 h at room temperature. Excess thrombin was removed by addition of benzamidine-Sepharose beads (Pharmacia). In order to purify the other peptides, cells were harvested and resuspended in 50 mM Tris buffer, pH 8.0, 5 mM EDTA, 5 mM benzamidine and 5 mM dithiothreitol (DTT). The cells were then lysed by passage through a French press and centrifuged at 100 000 g for 1 h. The supernatant was purified over two ion exchange columns as previously described (Omichinski *et al.*, 1993a,b). All of the peptides, including the cGATA-1 N-terminal finger, were then subjected to further purification on a C-4 reversed phase (Vydac) HPLC column with a 1–100% acetonitrile gradient in 0.05% (v/v) aqueous trifluoroacetic acid and their identity verified by electrospray ionization mass spectroscopy. The lyophilized proteins were reconstituted with 1.1 equivalents of zinc and the final pH slowly adjusted to 6.0 with NaOH.

Gel mobility shift analysis

Unless otherwise specified, 714 fmol of each purified peptide were incubated for 10 min on ice with 0.2 pmol labeled duplex oligonucleotide in the presence of 25 mM HEPES, pH 7.9, 50 mM KCl, 6.25 mM MgCl₂, 0.5 mM DTT, 0.1 mM EDTA, 0.1% NP40, 5% glycerol (binding buffer) and 200 ng poly(dI-dC). After incubation, the mixture was loaded onto a 5% polyacrylamide gel (19:1 acrylamide/bisacrylamide) and run in 0.5× TBE at 4°C. The sequences of the oligonucleotides used in the binding studies were: oligo A, 5'-AGGTTGCAGATAAACATTTT-3'; oligo B, 5'-AGGTTGCAGATCTTAATTTT-3'; oligo C, 5'-AGGTTGCAGATCTACATTTT-3'; oligo mut. A, 5'-AGGTTGCACTGAAACAT-

TTT-3'; oligo +1: C→A, 5'-AGGTTGCAGATATACATTTT-3'; oligo +2: T→A: 5'-AGGTTGCAGATCAACATTTT-3'; oligo D, 5'-AGGTTGCAGATTAACATTTT-3'; oligo E. 5'-AGGTTGCAGATGAACATTTT-3'.

Acknowledgements

We thank Drs Adam Bell and Robert G.Martin for critical reading of the manuscript and helpful discussions. We thank Drs Kazuyasu Sakaguchi and Hiroshi Sakamoto for mass spectral analysis. The cDNA clones of cGATA-2 and cGATA-3 were kindly provided by Dr James Douglas Engel.

References

Arcenci,R.J., King,A.A., Simon,M.C., Orkin,S.H. and Wilson,D.B. (1993) Mouse GATA-4: a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart. *Mol. Cell. Biol.*, **13**, 2235–2246.

Blobel,G.A., Simon,M.C. and Orkin,S.H. (1995) Rescue of GATA-1 deficient embryonic stem cells by heterologous GATA-binding proteins. *Mol. Cell. Biol.*, **15**, 626–633.

Dorfman,D.M., Wilson,D.B., Bruns,G.A. and Orkin,S.H. (1992) Human transcription factor GATA-2. Evidence for regulation of preendothelin-1 gene expression in endothelial cells. *J. Biol. Chem.*, **267**, 1279–1285.

Evans,T. and Felsenfeld,G. (1989) The erythroid-specific transcription factor Eryf1: a new finger protein. *Cell*, **58**, 877–885.

Fu,Y.H. and Marzluf,G.A. (1990) nit-2, the major positive-acting nitrogen regulatory gene of *Neurospora crassa*, encodes a sequence-specific DNA-binding protein. *Proc. Natl Acad. Sci. USA*, **87**, 5331–5335.

George,K.M., Leonard,M.W., Roth,M.E., Lieuw,K.H., Kioussis,D., Grosveld,F. and Engel,J.D. (1994) Embryonic expression and cloning of the murine GATA-3 gene. *Development*, **120**, 2673–2686.

Ito,E., Toki,T., Ishihara,H., Ohtani,H., Gu,L., Yokoyama,M., Engel,J.D. and Yamamoto,M. (1993) Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis. *Nature*, **362**, 466–468.

Kelley,C., Blumberg,H., Zon,L.I. and Evans,T. (1993) GATA-4 is a novel transcription factor expressed in endocardium of the developing heart. *Development*, **118**, 817–827.

Ko,L.J. and Engel,J.D. (1993) DNA-binding specificities of the GATA transcription factor family. *Mol. Cell. Biol.*, **13**, 4011–4022.

Kudla,B., Caddick,M.X., Langdon,T., Martinez-Rossi,N.M., Bennett,C.F., Sibley,S., Davies,R.W. and Arst,H.N.,Jr (1990) The regulatory gene *areA* mediating nitrogen metabolite repression in *Aspergillus nidulans*. Mutations affecting specificity of gene activation alter a loop residue of a putative zinc finger. *EMBO J.*, **9**, 1355–1364.

Laverriere,A.C., MacNeill,C., Mueller,C., Poelmann,R.E., Burch,J.B.E. and Evans,T. (1994) GATA 4/5/6, a subfamily of three transcription factors transcribed in developing heart and gut. *J. Biol. Chem.*, **269**, 23177–23184.

Leonard,M., Brice,M., Engel,J.D. and Papayannopoulou,T. (1993) Dynamics of GATA transcription factor expression during erythroid differentiation. *Blood*, **82**, 1071–1079.

Martin,D.I.K. and Orkin,S.H. (1990) Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. *Genes Dev.*, **4**, 1886–1898.

Martin,D.I.K., Zon,L.I., Mutter,G. and Orkin,S.H. (1990) Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages. *Nature*, **344**, 444–446.

Merika,M. and Orkin,S.H. (1993) DNA-binding specificity of GATA family transcription factors. *Mol. Cell. Biol.*, **13**, 3999–4010.

Mouthon,M.A., Bernard,O., Mitjavila,M.T., Romeo,P.H., Vainchenker,W. and Mathieu-Mahul,D. (1993) Expression of tal-1 and GATA-binding proteins during human hematopoiesis. *Blood*, **81**, 647–655.

Omichinski,J.G., Clore,G.M., Schaad,O., Felsenfeld,G., Trainor,C., Appella,E., Stahl,S.J., and Gronenborn,A.M. (1993a) NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science*, **261**, 438–446.

Omichinski,J.G., Trainor,C., Evans,T., Gronenborn,A.M., Clore,G.M. and Felsenfeld,G. (1993b) A small single-finger peptide from the erythroid transcription factor GATA-1 binds specifically to DNA as a zinc or iron complex. *Proc. Natl Acad. Sci. USA*, **90**, 1676–1680.

Omichinski,J.G., Pedone,P.V., Felsenfeld,G., Gronenborn,A.M. and Clore,G.M. (1997) The solution structure of a specific GAGA factor/

DNA complex reveals a new twist in classical zinc finger/DNA recognition. *Nature Struct. Biol.*, **4**, 122–132.

Oosterwegel,M., Timmerman,J., Leiden,J. and Clevers,H. (1992) Expression of GATA-3 during lymphocyte differentiation and mouse embryogenesis. *Dev. Immunol.*, **3**, 1–11.

Orkin,S.H. (1992) GATA-binding transcription factors in hematopoietic cells. *Blood*, **80**, 575–581.

Pandolfi,P.P., Roth,M.E., Karis,A., Leonard,M.W., Dzierzak,E., Grosveld,F.G., Engel,J.D. and Lindenbaum,M.H. (1995) Targeted disruption of the GATA-3 gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis. *Nature Genet.*, **11**, 40–44.

Pedone,P.V., Ghirlando,R., Clore,G.M., Gronenborn,A.M., Felsenfeld,G. and Omichinski,J.G. (1996) The single Cys2-His2 zinc finger domain of the GAGA protein flanked by basic residues is sufficient for high-affinity specific DNA binding. *Proc. Natl Acad. Sci. USA*, **93**, 2822–2826.

Pevny,L., Simon,M.C., Robertson,E., Klein,W.H., Tsai,S.F., D'Agati,V., Orkin,S.H. and Costantini,F. (1991) Erythroid differentiation in chimeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature*, **349**, 257–260.

Romeo,P.H., Prandini,M.H., Joulin,V., Mignotte,V., Prenant,M., Vainchenker,W., Marguerie,G. and Uzan,G. (1990) Megakaryocytic and erythrocytic lineages share specific transcription factors. *Nature*, **344**, 447–449.

Shim,Y.H., Bonner,J.J. and Blumenthal,T. (1995) Activity of a *C. elegans* GATA transcription factor, ELT-1, expressed in yeast. *J. Mol. Biol.*, **253**, 665–676.

Smith,D.B. and Johnson,K.S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene*, **67**, 31–40.

Trainor,C.D., Omichinski,J.G., Vandergon,T.L., Gronenborn,A.M., Clore,G.M. and Felsenfeld,G. (1996) A palindromic regulatory site within vertebrate GATA-1 promoters requires both zinc fingers of the GATA-1 DNA-binding domain for high-affinity interaction. *Mol. Cell. Biol.*, **16**, 2238–2247.

Tsai,F.Y., Keller,G., Kuo,F.C., Weiss,M., Chen,J., Rosenblatt,M., Alt,F.W. and Orkin,S.H. (1994) An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*, **371**, 221–226.

Tsai,S.F., Martin,D.I., Zon,L.I., D'Andrea,A.D., Wong,G.G. and Orkin,S.H. (1989) Cloning of the cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature*, **339**, 446–451.

Visvader,J. and Adams,J.M. (1993) Megakaryocytic differentiation induced in 416B myeloid cells by GATA-2 and GATA-3 transgenes or 5-azacytidine is tightly coupled to GATA-1 expression. *Blood*, **82**, 1493–1501.

Visvader,J.E., Crossley,M., Hill,J., Orkin,S.H. and Adams,J.M. (1995) The C-terminal zinc finger of GATA-1 or GATA-2 is sufficient to induce megakaryocytic differentiation of an early myeloid cell line. *Mol. Cell. Biol.*, **15**, 634–641.

White,B.A. (1993) *Methods in Molecular Biology*, Vol. 15, *PCR Protocols. Current Methods and Applications*. Humana Press, Totowa, NJ.

Whitelaw,E., Tsai,S.F., Hogben,P. and Orkin,S.H. (1990) Regulated expression of globin chains and the erythroid transcription factor GATA-1 during erythropoiesis in the developing mouse. *Mol. Cell. Biol.*, **10**, 6596–6606.

Whyatt,D.J., deBoer,E. and Grosveld,F. (1993) The two zinc finger-like domains of GATA-1 have different DNA binding specificities. *EMBO J.*, **12**, 4993–5005.

Yamamoto,M., Ko,L.J., Leonard,M.W., Beug,H., Orkin,S.H. and Engel,J.D. (1990) Activity and tissue-specific expression of the transcription factor NF-E1 multigene family. *Genes Dev.*, **4**, 1650–1662.

Yang,H.Y. and Evans,T. (1992) Distinct roles for the two cGATA-1 finger domains. *Mol. Cell. Biol.*, **12**, 4562–4570.

Yang,Z., Gu,L., Romeo,P.H., Bories,D., Motohashi,H., Yamamoto,M. and Engel,J.D. (1994) Human GATA-3 *trans*-activation, DNA-binding, and nuclear localization activities are organized into distinct structural domains. *Mol. Cell. Biol.*, **14**, 2201–2212.

Yomogida,K., Ohtani,H., Harigae,H., Ito,E., Nishimune,Y., Engel,J.D. and Yamamoto,M. (1994) Developmental stage- and spermatogenic cycle-specific expression of transcription factor GATA-1 in mouse Sertoli cells. *Development*, **120**, 1759–1766.

Received on December 17, 1996