Cell Reports
Article

DNA Binding by GATA Transcription Factor Suggests Mechanisms of DNA Looping and Long-Range Gene Regulation

Yongheng Chen,[1,6,8] Darren L. Bates,[7,8] Raja Dey,[1] Po-Han Chen,[3,4] Ana Carolina Dantas Machado,[1] Ite A. Laird-Offringa,[2,4,5] Remo Rohs,[1,2] and Lin Chen[1,2,6,7,*]
[1]Molecular and Computational Biology Program, Departments of Biological Sciences and Chemistry
[2]Norris Comprehensive Cancer Center
[3]PhD Program in Genetic, Molecular, and Cellular Biology
[4]Department of Biochemistry and Molecular Biology
[5]Department of Surgery
University of Southern California, Los Angeles, CA 90089, USA
[6]Laboratory of Structural Biology and Drug Design, and Key Laboratory of Cancer Proteomics of the Chinese Ministry of Health, XiangYa Hospital, Central South University, Changsha, Hunan 410008, China
[7]Department of Chemistry and Biochemistry, University of Colorado at Boulder, Boulder, CO 80309, USA
[8]These authors contributed equally to this work
*Correspondence: linchen@usc.edu
http://dx.doi.org/10.1016/j.celrep.2012.10.012

SUMMARY

GATA transcription factors regulate transcription during development and differentiation by recognizing distinct GATA sites with a tandem of two conserved zinc fingers, and by mediating long-range DNA looping. However, the molecular basis of these processes is not well understood. Here, we determined three crystal structures of the full DNA-binding domain (DBD) of human GATA3 protein, which contains both zinc fingers, in complex with different DNA sites. In one structure, both zinc fingers wrap around a palindromic GATA site, cooperatively enhancing the binding affinity and kinetic stability. Strikingly, in the other two structures, the two fingers of GATA DBD bind GATA sites on different DNA molecules, thereby bridging two separate DNA fragments. This was confirmed in solution by an in-gel fluorescence resonance energy transfer analysis. These findings not only provide insights into the structure and function of GATA proteins but also shed light on the molecular basis of long-range gene regulation.

INTRODUCTION

The GATA-binding proteins are a group of structurally related transcription factors that bind to the DNA consensus sequence GATA. Members of the GATA protein family (GATA1-6) function as lineage-specific transcription factors for a number of cell types in the hematopoietic system (Molkentin, 2000; Patient and McGhee, 2002; Weiss and Orkin, 1995). For example, GATA1 is essential for erythroid and megakaryocytic development (Ferreira et al., 2005; Orkin et al., 1998); GATA2 plays an essential role in regulating the transcription of genes involved in the development and proliferation of hematopoietic (Tsai and Orkin, 1997) and endocrine cell lineages (Dasen et al., 1999); GATA3 is an important regulator of T cell development, including Th2 (George et al., 1994; Ho et al., 1991; Zheng and Flavell, 1997) and regulatory T cells (Wang et al., 2011), and plays an important role in endothelial cell biology (Song et al., 2009); and GATA4 regulates genes involved in embryogenesis and in myocardial differentiation and function (Watt et al., 2004).

The function of GATA proteins depends critically on two highly conserved zinc fingers and nearby basic regions (Shimizu et al., 2001). The C-terminal "C-finger" and its adjacent basic region are necessary and sufficient for GATA to bind its cognate sequence, WGATAR (W = A/T, R = A/G; Ko and Engel, 1993; Merika and Orkin, 1993; Omichinski et al., 1993; Visvader et al., 1995). The N-terminal "N-finger" can also bind DNA independently but has a preference for GATC core motifs (Martin and Orkin, 1990; Newton et al., 2001; Pedone et al., 1997). Both fingers participate in binding the palindromic GATA motif ATCWGATA (W = A/T), resulting in markedly increased affinity (Trainor et al., 1996). Recent chromatin immunoprecipitation sequencing (ChIP-seq) experiments on GATA1, GATA2, and GATA3 showed that GATA proteins mostly bind to a single GATA site ((A/T)GATAA) or a palindromic site (catctGATAAG; Fujiwara et al., 2009; Horiuchi et al., 2011; Wei et al., 2011; Yu et al., 2009).

GATA proteins contribute to transcriptional regulation by facilitating chromosome looping, thereby mediating long-range control of gene expression in the nucleus. For example, GATA1 and its cofactor FOG-1 directly occupy looped enhancers and target gene promoters at the β-globin locus (Vakoc et al., 2005). Similarly, GATA3 has been shown to be important for the establishment and/or maintenance of long-range chromatin interactions at the Th2 cytokine locus (Spilianakis et al., 2005). Moreover, exchange of GATA1 and GATA2 has been shown to mediate transitions in looped chromatin organization of the *KIT* gene,

which is expressed during early erythropoiesis (Jing et al., 2008). However, the molecular mechanism by which GATA and FOG proteins mediate loop formation remains unclear.

Structural studies based on both nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography have characterized the DNA-binding mechanisms of the GATA C-finger (Bates et al., 2008; Omichinski et al., 1993; Starich et al., 1998). In addition, the structure of the GATA1 N-finger bound to the FOG-1 zinc finger 1 has been characterized by NMR (Liew et al., 2005). However, the DNA-binding mechanisms of the GATA N-finger, the mechanisms by which the GATA double fingers bind to palindromic GATA sites, and the mechanisms by which GATA proteins loop DNA are not yet fully understood.

Here, we present the crystal structures of the full DNA-binding domain (DBD) of GATA3 bound to different DNA sequences. These structures reveal that GATA zinc fingers can bind DNA using two different modes: the wrapping mode, wherein the two zinc fingers wrap around a single palindromic GATA site, and the bridging mode, wherein the two zinc fingers bridge two distant sites on DNA and (in our case) sites on different DNA molecules. Our results reveal the structural basis of how GATA fingers can bind distinct *cis*-regulatory elements in different conformations, and how GATA proteins may participate in chromosome looping by bridging separated DNA molecules. These insights will help to elucidate gene regulation by GATA proteins and provide a basis for further studies of the complex functions of GATA proteins in vivo.

RESULTS

Crystallographic Study of GATA3 Tandem Zinc Fingers Bound to DNA

To better understand the mechanism by which GATA proteins recognize and bind DNA, we carried out crystallographic studies of the DBD of the human GATA3 protein (amino acids 260–370) bound to DNA. GATA DBD contains two zinc fingers that are highly conserved within the GATA family (GATA1-6). Recent studies showed that 40% of the GATA1 ChIP-seq peaks are palindromic GATA sites (Yu et al., 2009). Given the fact that GATA family members share a highly conserved DBD, it is reasonable to assume that all GATA proteins could bind palindromic GATA sites, although the binding preference and affinity may vary among different GATA members. To understand how GATA proteins bind these palindromic GATA sites, we tested different palindromic sequences with variable spacers between the two GATA sites in our crystallization attempts. Two DNA fragments crystallized successfully with the GATA3 DBD. In one DNA fragment, the two GATA sites are separated by 1 bp. This fragment, which is derived from the mouse GATA1 promoter, contains the consensus palindromic motif defined by ChIP-seq: CATCTGATAAG. This complex (complex 1) was solved at 2.8 Å resolution. In the other DNA fragment, the two GATA sites are separated by 3 bp. This DNA fragment crystallized with the GATA3 DBD in two different crystal shapes, one with a bar-shaped morphology and one with a plate shape. The bar-shaped crystal belongs to the $P2_1$ space group (complex 2), and the plate-shaped crystal belongs to the C_2 space group (complex 3). The bar-shaped crystal structure was solved at 2.65 Å, and the plate-shaped crystal structure was solved at 1.6 Å. In total, we solved three crystal structures of GATA3 DBD/DNA complexes. The structures of complexes 2 and 3 were solved by single-wavelength anomalous dispersion (SAD) phasing using the zinc anomalous signal. We also found a molecular replacement solution using previously published coordinates of the GATA3 C-finger (PDB 3DFX) as a partial search model. We solved the structure of complex 1 by molecular replacement using the N-finger and C-finger determined in complex 3 as partial search models. The X-ray crystallography statistics for all complexes are presented in Table S1. The crystal packing and related analyses of each complex are described in Figure S1.

Overall Structure of GATA Bound to a Palindromic DNA Site

In complex 1, the GATA3 DBD binds to a 20-mer palindromic GATA site, AATGTCCATCTGATAAGACG (binding sites are underlined; Figure 1A, lower panel). The crystal belongs to space group $P2_1$, wherein each asymmetric unit contains one GATA3 molecule bound to a double-stranded DNA (dsDNA) molecule (Figure 1A). The N-finger (amino acids 264–288, green) binds to the GATG site (CATC in complementary strand). The peptide region immediately following the N-finger (amino acids 289–317, green), which also serves as the linker to the C-finger (amino acids 318–342, magenta), tracks along one strand of the phosphodiester backbone and connects to the C-finger bound to the GATA site. The C-terminal basic region (amino acids 343–368, magenta) of the C-finger inserts deeply into the minor groove of the GATA site. Interestingly, the N and C termini of the GATA3 DBD come close in space in this palindrome-bound form, giving the protein the shape of a parallelogram (Figure 1A). The DNA essentially has a straight B-form, with a slightly widened major groove at the center of the palindromic site and an accompanying compression of the minor groove at the GATA site, presumably due to the insertion of the basic region of the C-terminal tail of the C-finger into the minor groove (Figure 1A). Our analysis demonstrates that the narrow minor groove correlates with the enhanced negative electrostatic potential (Figure S2). This effect of DNA shape-dependent electrostatic potential stabilizes the interaction with C-terminal arginines (Rohs et al., 2009). Together, the two fingers and their respective C-terminal extensions form a C-shaped clamp that wraps around the DNA along the major groove of the palindromic site, which leads to an extensive shape and electrostatic potential complementarity between the GATA DBD and its target site (Harris et al., 2012).

Structures of GATA Bridging Two Separate DNA Molecules

In an effort to crystallize GATA3 bound to palindromic GATA sites separated by different spacers, we unexpectedly solved two crystal structures of the GATA3 DBD bridging two dsDNA molecules. In both complexes, the GATA3 DBD was cocrystallized with a palindromic site with two GATA-binding sites separated by 3 bp, TTCCTAAATCAGAGATAACC (binding sites underlined; Figures 1B and 1C). One complex belongs to the $P2_1$ space group, where the asymmetric unit contains two GATA3 DBDs

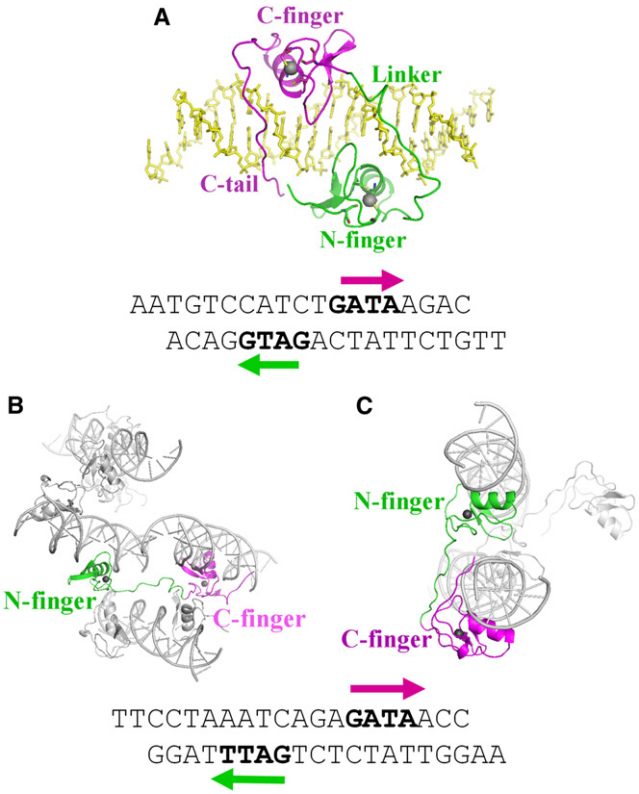


Figure 1. Overall Structures of GATA3/DNA Complexes

(A) Structure of complex 1. The N-finger and the linker are colored in green, the C-finger and C-tail are colored in magenta, and the DNA is colored in yellow.

(B) Structure of complex 2. Complexes of GATA3 N-/C-fingers bound to DNA from neighboring asymmetric units are shown to illustrate one GATA3 protein bridging two pieces of DNA. One N-finger is colored in green, the C-finger of the same GATA protein is colored in magenta, and all other molecules are colored in gray.

(C) Structure of complex 3. Complexes of GATA3 N-/C-fingers bound to DNA from neighboring asymmetric units are shown to illustrate DNA bridging by GATA3 protein; same color scheme as described in (B). The sequences of the DNA are listed below.

See also Figure S1 and Table S1.

and two dsDNA molecules (complex 2). In contrast to the one-spacer palindromic site, here the N-finger (green) and C-finger (magenta) of one GATA molecule bind to their cognate sites on two separate dsDNA molecules, which are parallel to each other in the crystal lattice (Figure 1B). Specifically, the C-finger binds to the GATA site of one dsDNA molecule, and the N-finger binds to the GATT site of another dsDNA molecule. The two DNA molecules are separated by 35 Å (axis to axis). In another complex, the protein and DNA are cocrystallized in the C_2 space group, with the asymmetric unit containing one copy of GATA3 DBD and one copy of dsDNA molecule (complex 3). Here again, the C-finger of the GATA3 DBD binds to the GATA site of one dsDNA molecule, and its N-finger binds to the GATT site on a separated, symmetry-related dsDNA molecule (Figure 1C). The two dsDNA molecules are arranged in a relative angle of ~30° in the crystal lattice. The minimal axis-to-axis distance between the two DNA molecules is 29 Å. In complexes 2 and 3, the relative orientations of the N-finger and C-finger are very different, reflecting the large conformational flexibility of the linker between the N-finger and C-finger and the different arrangement of the DNA molecules that can be accommodated by GATA3-mediated bridging.

Protein–DNA Interactions

The GATA DBD contains two zinc fingers, each of which can bind its respective target site in a modular fashion. In complex 1, the two fingers interact to bind adjacent sites in the palindromic motif cooperatively. In complexes 2 and 3, the two fingers bind their target sites on separated dsDNA molecules independently. Although the relative spatial arrangements of the N-finger and C-finger vary substantially in different DNA-binding modes, the detailed interactions between each zinc finger and its cognate DNA sites are largely conserved in different complexes. In all three complexes, the C-finger binds the consensus GATA site in a highly conserved mode that is also seen in the isolated C-finger/DNA complex (Bates et al., 2008). The conservation of the DNA-binding mechanisms by the C-finger includes not only the detailed interactions of the core zinc module (amino acids 316–349) with the major groove but also the binding of the C-terminal basic tail (amino acids 350–366) to the minor groove. The detailed interactions in the minor groove, however, vary slightly between different complexes. For instance, in complexes 2 and 3, Arg367 exhibits very weak density and thus is not included in the final model; however, in complex 1, Arg367 shows clear density and forms a hydrogen bond with both Thy14 and Thy10′ (Figure 2A). This difference could be due to the interactions with the N-finger that stabilize the conformation of the basic tail of the C-finger and facilitate its binding to DNA in the minor groove. As shown in Figure 2A, Arg277 of the N-finger interacts with the main chain of Lys368, while Met260 forms a van der Waals contact with Thr363 (not shown). In contrast, in complexes 2 and 3, where the N-finger is located on a separate DNA molecule, the basic tail of the C-finger is less well ordered in the tip region including residue Arg367.

Although the N-finger structure has been solved in complex with a zinc finger of the FOG1 protein by NMR spectroscopy (Liew et al., 2005), the mechanism by which the N-finger recognizes DNA has not yet been elucidated structurally. In all three structures, the zinc core module of the N-finger binds to the first three nucleotides (GAT) in a similar manner as observed for the C-finger core. For example, Arg276 and Asn286 interact with Gua13′, Thy9, and Ade8 (Figure 2B), in a manner almost identical to that observed for Arg330 and Asn340 in the C-finger (Figure 2C).

The major difference in DNA binding between the N-finger and the C-finger lies in the recognition of the fourth and variable position in their binding sites, GATN (where N refers to any nucleotide). The C-terminal basic tail of the C-finger inserts into the minor groove, and, most importantly, Arg365 forms a hydrogen bond with the carbonyl of Thy8′ and also engages in extensive van der Waals contacts with neighboring bases and sugar moieties (Figure 2D). In addition, the enhanced negative electrostatic potential in this region of the minor groove attracts the positive charge located in the arginine's guanidinium group (Figure S2). The electrostatic potential at this position is approximately −8.5 kT/e due to the electrostatic focusing effect in narrow minor

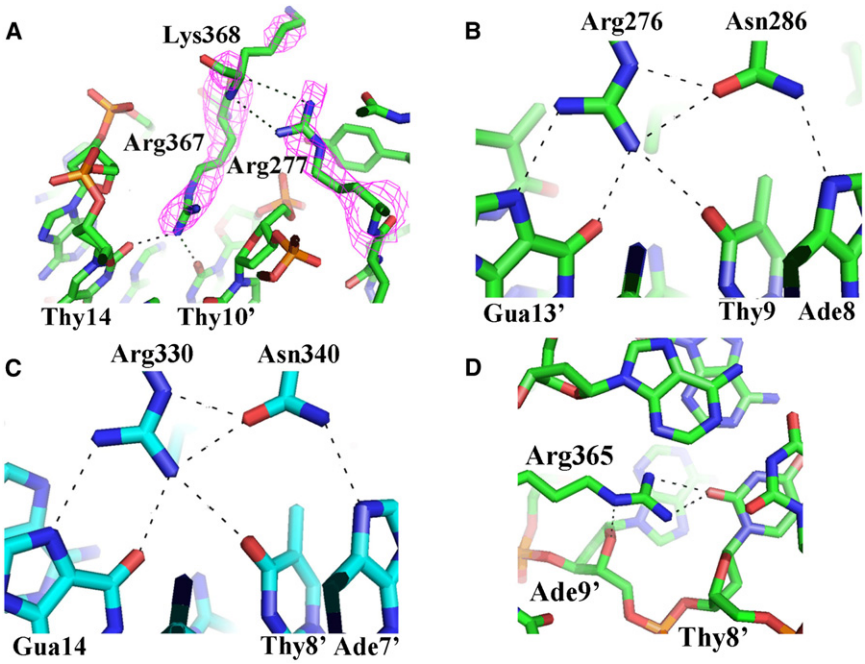


Figure 2. DNA Recognition by the GATA3 DBD

(A) Arg367 binds DNA only in the wrapping mode (complex 1). These interactions are stabilized through the interaction between the Arg277 guanidinium group and the Lys368 backbone, reducing the conformational flexibility of the C-terminal tip. The electron density is calculated from a composite omit map.

(B) Hydrogen-bonding interactions among Arg276, Asn286 (N-finger), and 3 bp (GAT) of the binding site at their major groove core. These hydrogen bonds, in particular the bidentate hydrogen bond of the Arg276 guanidinium group with guanine, lead to specific recognition of the base-pair identity.

(C) A network of hydrogen-bonding interactions, similar to the one shown in (B), including a bidentate hydrogen bond, is formed among Arg330, Asn340 (C-finger), and 3 bp (GAT) of the binding site at their major groove core.

(D) Arg365 forms hydrogen bonds with the minor groove edge of a base. Because these base readout interactions in the minor groove are not as specific as in the major groove, the negative electrostatic potential enhanced through the narrow minor groove stabilizes interactions with arginine residues (shape readout).

See also Figure S2.

groove regions (Rohs et al., 2009). Similar interactions are observed for the Arg367 of the C-finger in complex 1, including hydrogen bonds with Thy14 and Thy10′ and an electrostatic potential of approximately −9.1 kT/e (Figure S2A). Thus, both Arg365 and Arg367 play a key role in DNA binding through both base-specific hydrogen bonding (base readout) and shape-dependent electrostatic interactions (shape readout; Rohs et al., 2009, 2010). Most of the minor groove interactions by the C-finger described above are missing in the N-finger. The detailed DNA-binding interactions by the N-finger and C-finger in complex 1 are shown in Figure 3, which also represents most of the DNA-binding interactions by individual zinc fingers in complexes 2 and 3 (not shown). Overall, compared with the C-finger, the N-finger makes fewer contacts to DNA with both the bases and the phosphodiester backbone (Figure 3), which may explain the differences between the C- and N-fingers in DNA-binding affinity and specificity (Bates et al., 2008).

Linker Region

The linker region in all three crystal structures shows a large flexibility, as evidenced by high B-factors and a weak electron density. Although the electron density for the main chain of the linker region is visible in complex 1, the side chains are not defined (Figure S3A). This observation suggests that the linker region makes little contribution to DNA binding in the "wrapping" model. Between complexes 2 and 3, there are a total of three copies of the linker region, all of which show densities for the backbone but not the side chains (Figures S3B and S3C). A superposition of the N-finger of all three molecules shows that the linker region of these structures diverges at Ile302 and extends in different directions (Figure 4A). Similarly, a superposition of the C-finger shows different paths of the linker region converging at Thr316, immediately in front of the C-finger (Figure 4B). These structural analyses suggest that the linker region between the N-finger and the C-finger in GATA3 does not appear to have any conformational limitations. One of its functional roles appears to be providing a constraint on the spacing of the palindromic sites and the geometric distance between the DNA molecules bridged by GATA, while the flexibility of the linker region allows a wide variety of orientations of the bridged DNA molecules.

It is somewhat surprising that the linker region does not bind to DNA in the minor groove in complex 1 as does the C-terminal tail of the C-finger (C-tail). The linker region is similar to the C-tail at the sequence level because it has multiple basic residues (Figure S4). The major difference lies between P304 and K358. In the C-tail, the side chain of K358 interacts with the main chain of C321, and this interaction pulls the tail close to the minor groove; whereas in the linker region, P304 cannot interact with C267 and pushes the linker away from the minor groove (Figure 4C).

DNA-Binding Analyses

Although the GATA N-finger and C-finger alone can bind DNA, a previous study showed that both fingers of GATA1 participate in binding palindromic GATA sites, resulting in markedly increased affinity (Trainor et al., 1996). Consistent with this biochemical finding, a recent ChIP-seq study showed that the palindromic GATA site accounts for 40% of GATA1 enrichment peaks, and the palindromic GATA site has an increased peak height compared with the peak average (Yu et al., 2009). Our structure (complex 1) shows that the two fingers of GATA3 bind opposite faces of the DNA, and the N-finger interacts with the C-terminal basic tail of the C-finger, thereby enhancing GATA–DNA interactions. These structural features suggest that the higher affinity of

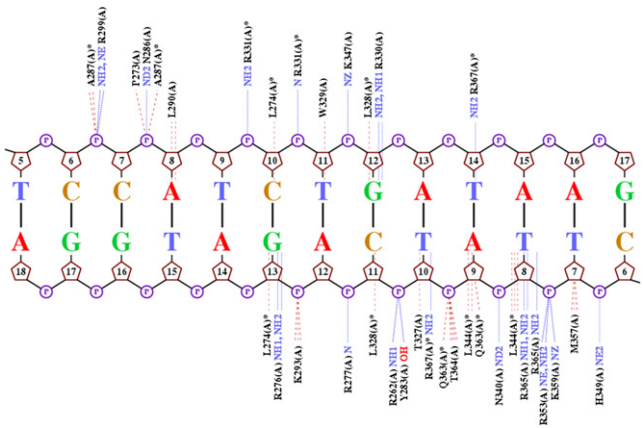


Figure 3. Schematic Representation of DNA Contacts with GATA3 Amino Acids in Complex 1 in the Palindromic Site

Hydrogen bonds are indicated by blue dotted lines, and hydrophobic contacts are indicated by red dashed lines. The protein/DNA contact diagram was generated by NUCPLOT (Luscombe et al., 1997). The contact map demonstrates that GATA3 contacts the DNA in both the major and minor grooves and employs base and shape readout mechanisms.

GATA DBD for the palindromic site compared with the single site is likely a result of more extensive protein/DNA contacts and the clamping mode of DNA binding by the two fingers of GATA, which could lead to enhanced kinetic stability of the protein-DNA complex (Stroud et al., 2002). To determine whether GATA3 also binds palindromic sites with higher affinity compared with single sites, we further analyzed the binding of GATA3 to DNA. We performed binding studies using surface plasmon resonance (SPR) on a Biacore 2000 instrument. DNA probes containing either a single or a palindromic GATA site were immobilized to the same low density on the surface of a biosensor chip. GATA3 protein was injected over the sensor chip surfaces to examine its interaction with the different DNA sites. A 1:1 interaction model with mass transport correction was applied to obtain the association (k_a) and dissociation (k_d) rates. The affinity was calculated from these values ($K_D = k_d/k_a$). Biacore analyses revealed that GATA3 binds to DNA containing a single GATA site at relatively low affinity with a fast off-rate (Figure 5A, top panel). In contrast, the binding of GATA3 to DNA containing the palindromic GATA site showed a markedly increased affinity, largely due to enhanced kinetic stability, as evidenced by the slower off-rate ($p < 0.05$; Figure 5A, bottom panel). In addition, the response units at equilibrium binding of GATA3 to a palindromic GATA site were 2-fold higher than the response obtained during binding to a single GATA site, further confirming that GATA3 has a stronger affinity for the palindromic GATA site over the single site (Figure 5A). Fitting of the data indicates that the GATA3 DBD binds the palindromic site with a K_D of ~6 nM, whereas it binds the single site with a K_D of ~60 nM (Figure 5B). Therefore, the binding affinity of GATA3 DBD for the palindromic site is ~10-fold higher than that of the single site.

DNA Looping by GATA in Solution

Our structures of complexes 2 and 3 reveal that GATA3 DBD can bridge two DNA molecules. This mode of DNA binding may have important implications for long-range gene regulation by GATA proteins (Jing et al., 2008; Spilianakis et al., 2005; Vakoc et al., 2005). However, in order to exclude potential crystal-packing effects, we had to determine whether GATA can bind two segments of DNA in solution. For this purpose, we used fluorescence (or Förster) resonance energy transfer (FRET) to monitor the proximity of DNA ends in solution. We synthesized a 40-mer DNA that contains the binding site for the C-finger (GATA) at one end and the N-finger (GATC) at the other end. The two ends of the DNA were labeled by fluorophore Cy3 (donor) and Cy5 (acceptor), respectively. The central region of the DNA is a stretch of single-stranded DNA (ssDNA) made of poly(dT). This design was chosen to reduce nonspecific binding and increase DNA flexibility. On the basis of our crystal structures (Figures 1B and 1C), we assume that simultaneous binding of GATA3 to both sites on the DNA would bring the two DNA ends together, thus allowing FRET to occur when the donor is excited. However, the detection of FRET signal in solution does not necessarily indicate DNA looping, because two different DNA molecules may be brought together by nonspecific interactions. To address this issue, we developed an in-gel FRET assay that enabled us to measure the FRET signal for a given complex with a well-defined mobility shift. As shown in Figure 6A (lanes 1–6), the FRET signal as indicated by the color change in the GATA3/DNA complex is reproducibly and significantly stronger than that obtained in free DNA, suggesting that the DNA ends are indeed closer in the GATA3-bound complex. The tight band and fast mobility suggest that the FRET signal most likely arises from GATA3-mediated looping rather than from binding of multiple DNA molecules. To test this further, we mutated the N-finger (R276E) to disrupt its DNA-binding activity. Compared with the wild-type protein, the mutant showed lower DNA-binding affinity, presumably due to the loss of avidity, and, more importantly, the FRET signal of the mutant complex was similar to that of free DNA, suggesting an open DNA structure (Figure 6A, lanes 7–12). These results are consistent with a model wherein the mutant GATA3 binds only one site with the C-finger in an open conformation (Figure 6B). Interestingly, the mutant-bound complex also showed slower mobility, consistent with a less compact structure and/or binding of two molecules of GATA3 as depicted in Figure 6B. The above studies strongly suggest that GATA3 can indeed loop DNA in solution.

DISCUSSION

Here we present a comprehensive analysis of the DNA recognition mechanisms employed by the GATA protein in the context of its full DBD, i.e., with both zinc fingers. Compared with previous studies involving isolated GATA zinc fingers (Bates et al., 2008; Liew et al., 2005; Omichinski et al., 1993; Starich et al., 1998), our study reveals aspects of DNA binding by GATA that have important implications for the function of GATA in transcriptional regulation.

It is known that GATA can bind palindromic sites with high affinity and that such binding may be functionally important (Trainor et al., 1996; Yu et al., 2002). However, it was not known until recently how prevalent this DNA-binding mode of GATA is in vivo. A genome-wide analysis of GATA1-binding sites using

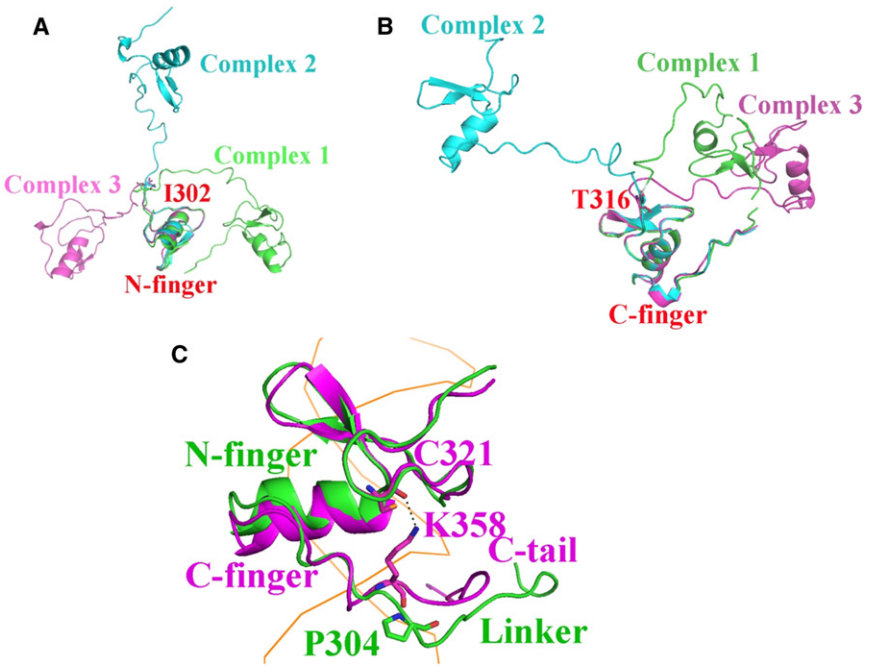


Figure 4. Superposition of Zinc Finger Motifs Displaying Various Linker Conformations

(A) Superposition of the N-fingers from the three complexes.

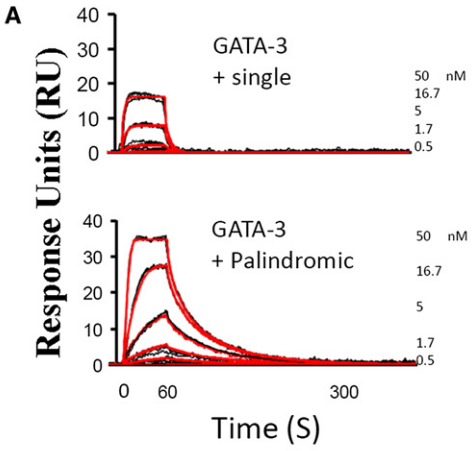
(B) Superposition of the C-fingers from the three complexes.

(C) Superposition of the N-finger and C-finger in complex 1. P304 pushes the linker region away from the minor groove, whereas K358 pulls the C-tail close to the minor groove through its interaction with C321. The DNA is visualized as an orange ribbon.

See also Figures S3 and S4.

ChIP-seq revealed that 40% of GATA1-binding sites follow the palindromic consensus motif. Moreover, ChIP-seq peaks from palindromic sites are significantly higher than those from isolated GATA1 sites (Yu et al., 2009), consistent with the higher-affinity binding of GATA1 to palindromic sites (Trainor et al., 1996). These observations suggest that binding of GATA1 to palindromic sites plays a significant role in vivo (Yu et al., 2009). Although a similar analysis of GATA3 has not yet been reported, it is possible that GATA3 also binds palindromic GATA sites. Our structure of the GATA3 DBD bound to the palindromic site represents an example of such an important DNA-binding mode by GATA proteins, which is also, more generally, a novel mode of DNA binding by a zinc finger protein. Compared with classical Cys_2His_2 zinc finger proteins, such as Zif268 and TFIIIA, the GATA proteins have a longer linker region, which allows more complex and versatile DNA-binding patterns (Stroud and Chen, 2003). Zif268 and TFIIIA have shorter linkers and use multiple zinc finger domains to wrap around DNA and track along the major groove (Elrod-Erickson et al., 1998; Wuttke et al., 1997). Our structure is also different from most other Cys_4 zinc finger proteins, such as nuclear receptor proteins. Although most nuclear receptor proteins also have two Cys_4 zinc finger motifs, they use one for DNA binding and the other for dimerization. In addition, nuclear receptors form dimers to bind the same side of the DNA (Bain et al., 2007). In our structure, which exhibits the wrapping mode (complex 1), the N-finger binds to one side of DNA in the major groove and then tracks along the minor groove without DNA binding. The C-finger binds to the major groove on the other side of the double helix and then tracks along the minor groove with extensive DNA binding. Finally, the end of the C-tail reaches the N-finger and completes the wrapping architecture. Several structural features may account for the enhanced binding affinity of GATA to the palindromic sites. First, the N-finger and C-finger bind to adjacent sites on the DNA, making extensive protein-DNA contacts. Second, the binding of a zinc finger induces conformational changes of the DNA, and such changes may facilitate the binding of the second zinc finger to the adjacent site. Third, the N-finger and C-finger interact directly on palindromic sites, and these protein–protein interactions may lead to cooperative binding of the two fingers to DNA, similar to the cooperative binding of GATA C-fingers to adjacent sites (Bates et al., 2008). Finally, the N- and C-fingers, in conjunction with their C-terminal extensions, wrap around DNA in the palindromic complex, and such encirclement of DNA by the protein usually leads to enhanced kinetic stability of the protein/DNA complex (Stroud et al., 2002). Our Biacore analysis suggests that GATA3 DBD indeed binds the palindromic site with a significantly slower off-rate than the single site.

The GATA family of proteins contains two zinc fingers that can each bind to a separate target site. Although complex 1 reveals that the two fingers can function jointly to recognize the palindromic sites, the majority of GATA1 response elements (60%) are comprised of single GATA sites (Yu et al., 2009). Binding of these isolated GATA sites by the N-finger and C-finger of GATA proteins could lead to chromosome looping or interchromosome interactions. This is a very attractive model of transcriptional regulation by GATA proteins, given the increasing evidence suggesting that GATA proteins have a role in long-range gene regulation (Jing et al., 2008; Spilianakis et al., 2005; Vakoc et al., 2005). Complexes 2 and 3 reveal that GATA3 DBD can indeed bridge two dsDNA molecules, thereby providing a direct mechanism of chromosome looping. The structures presented here are crystal structures of a single protein bridging two separate DNA fragments. Two lines of evidence support our conclusion that the DNA bridging by GATA proteins is not a crystal-packing artifact: first, we observed the DNA bridging in two crystal structures with different overall architectures; second, in-gel FRET experiments provided further evidence that the GATA DBD could bridge DNA in solution. We would like to provide the functional relevance of this DNA bridging in cells, however, there is no method, to our knowledge, that can provide direct evidence that a GATA protein can bridge two



B

Target	k_a (M^{-1}, s^{-1})	k_d (s^{-1})	K_D (M)
Single	$(8.26\pm2.78) \times 10^{6}$	$(4.93\pm1.04) \times 10^{-1}$	$(5.97\pm2.33) \times 10^{-8}$
Palindromic	$(1.38\pm0.48) \times 10^{7}$	$(9.12\pm5.36) \times 10^{-2}$	$(6.61\pm3.11) \times 10^{-9}$

Figure 5. Biacore Analyses

(A) Sensorgrams show a kinetic analysis of GATA3 DBD with either a single GATA site (top) or a palindromic site (bottom). Black lines represent triplicate injections performed in random order over the indicated DNA surface. A 1 min association was followed by a 5 min dissociation phase. Red lines represent the global fit of data sets using CLAMP.

(B) Association rate constants k_a, dissociation rate constants k_d, and equilibrium constants $K_D = k_a/k_d$ demonstrate the kinetics of GATA3 DBD binding to either a single or a palindromic GATA site. The SEM is indicated. The dissociation rates of single and palindromic sites are significantly different ($p < 0.05$).

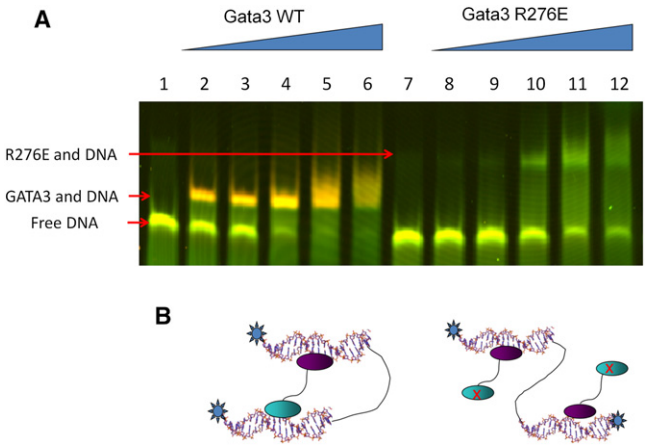


Figure 6. In-Gel FRET Analysis of DNA Looping by GATA3 in Solution

(A) GATA3 N-/C-fingers can bring two binding sites into close proximity despite their separation in sequence. The pseudo-colored image shows a superposition of the fluorescence of the Cy3-donor (green) and the fluorescence of the Cy5-acceptor (red) fluorophores. Labels to the left indicate the respective protein/DNA species. Lanes 1–6 show the binding of wild-type GATA3 to DNA. Lanes 7–12 show the binding of a GATA3 mutant R276E, which abolishes N-finger DNA binding.

(B) Cartoons represent models of the distinct overall architecture corresponding to the two protein shifts. The N-finger is displayed as a cyan oval, and the C-finger is shown as a purple oval.

separate DNA molecules in cells. Given that it is well known that GATA proteins mediate long-range gene regulation, we believe that our model represents at least one of the mechanisms by which GATA proteins loop DNA in cells. In addition to direct DNA bridging, GATA proteins may loop DNA through indirect mechanisms, such as via their cofactors (e.g., FOG proteins), which have also been shown to play important roles in mediating DNA loop formation (Vakoc et al., 2005). We recently showed that FOXP3, a transcription factor that is critical for the function of regulatory T cells, also bridges DNA to mediate DNA looping (Bandukwala et al., 2011). However, the mechanisms of DNA bridging by FOXP3 and GATA3 differ in two aspects: First, DNA bridging by FOXP3 occurs through a domain-swapped dimer, whereas GATA3 bridges DNA through two covalently linked DNA-binding motifs of the same protein. Second, the domain-swapped dimer of FOXP3 has a rigid structure that bridges two DNA molecules in a fixed orientation, whereas the flexible linker between the N- and C-fingers in GATA allows different orientations of bridged DNA molecules, as seen in complexes 2 and 3. These structural differences may reflect differences in long-range gene regulation by FOXP3 and GATA3. In spite of these differences, however, DNA bridging appears to be a common property of many transcription factors, including FOXP2 (Stroud et al., 2006), MEF2 (Guo et al., 2007), FOXP3 (Bandukwala et al., 2011), and GATA3 (as discussed here). Notably, these transcription factors are all implicated in lineage control during cellular differentiation. We thus propose that one mechanism whereby the epigenetic expression patterns of a given cell are controlled is specific folding of the three-dimensional structure of the genome by transcription factors that control a given lineage (Kalhor et al., 2012).

We have characterized two distinct DNA-binding modes of GATA transcription factors: (1) the wrapping mode, in which both zinc fingers synergistically enhance the binding affinity and kinetic stability; and (2) the bridging mode, in which a single GATA DBD bridges two pieces of DNA. Because the DBDs of GATA proteins are highly conserved, the structural features described here will likely hold true for all six GATA family members. The fact that GATA can bind DNA in two distinct modes with different binding affinity/kinetic stabilities and conformations has important implications for transcriptional regulation by this family of proteins. For example, varying protein concentrations of GATA proteins during development may not just affect their occupancy of DNA, but also could switch their DNA-binding mode and affect transcriptional networks (Georgescu et al., 2008). At low expression levels, the GATA proteins may preferentially bind to the palindromic sites, and at high expression levels, the GATA proteins may be able to bind both palindromic sites and single sites, leading to more intra- and interchromosomal interactions, either by themselves or through their interaction with cofactors such as the FOG protein. The different DNA-bound conformations of GATA proteins may recruit different cofactors or change the local chromosomal conformation, leading to different biological functions.

EXPERIMENTAL PROCEDURES

Sample Preparation and Crystallization

A sequence containing both the N- and C-terminal zinc fingers of human GATA3 (amino acids 260–370) was cloned into the pET-28a vector as a 6 × His-tagged fusion protein and was expressed in Rosetta (DE3) pLysS cells. The protein was first purified by Ni-NTA beads and then digested by thrombin protease to remove the His-tag. The protein was further purified by Mono S cation exchange and a Superdex 75 size exclusion column (Amersham Biosciences, Piscataway, NJ). The protein was then concentrated to ~40 mg/ml in 10 mM HEPES (pH 7.63), 5 mM β-mercaptoethanol, 0.5 μM zinc acetate, 100 mM NaCl, 200 mM NH_4-acetate, and 20% glycerol, and stored at −80°C. DNA was synthesized by Integrated DNA Technologies (Coralville, IA). The DNA sequences are listed in Figure 1.

The protein/DNA complex was prepared by mixing protein and DNA at a 1:1 molar ratio. Crystals were grown by the hanging-drop method at 18°C using a reservoir buffer of either 400 mM NH_4(OAc), 50 mM acetate (pH 4.7), and 18% polyethylene glycol (PEG) 4K (for complex 1), or 400 mM NH_4(OAc), 10 mM $Mg(OAc)_2$, 50 mM cacodylic acid (pH 6.33), and 15% PEG 4K (for complexes 2 and 3).

Data Collection and Structure Determination

Crystals were stabilized in the crystallization buffer with 25% (w/v) glycerol and flash frozen with liquid nitrogen for cryocrystallography. Data were collected at the Advanced Light Source BL8.2.1, BL8.2.2 beamline at Lawrence Berkeley National Laboratory. Data were reduced using HKL2000 (Otwinowski and Minor, 1997). Molecular replacement solutions for complexes 2 and 3 were found using the coordinates of the C-finger (Protein Data Bank ID Code [PDB] 3DFX; Bates et al., 2008) as a partial search model. Independent phases for complexes 2 and 3 were also obtained using SAD phasing by the zinc anomalous signal, which cross-validated the molecular replacement solutions. Complex 1 was determined solely by molecular replacement, using the N-finger and C-finger obtained from complex 3 as partial search models. Refinement were done using CNS refine (Brünger et al., 1998), CCP4 refmac5 (Collaborative Computational Project, Number 4, 1994), and Phenix.refine (Adams et al., 2002). Model building and analysis were carried out using Phenix.autobuild (Adams et al., 2002) and O (Jones et al., 1991). The statistics of the crystallographic analysis are presented in Table S1. Graphical representations of structure were prepared using PyMol (DeLano Scientific, San Francisco, CA).

Biosensor Analysis

Binding experiments were performed on a Biacore 2000 instrument (Biacore, Piscataway, NJ, USA). dsDNA oligos containing either a palindromic GATA site (5′-Bi-CTCCCGCTCGCTATCAGATAAGGCCTTAT-3′ and 5′-ATAAGG CCTTATCTGATAGCGAGCGGGAG-3′) or a single GATA site (5′-Bi-CTCCC GCTCGCTCAGAGATAAGGCCTTAT-3′ and 5′- ATAAGGCTTTATCTCTGA GCGAGCGGGAG-3′) were synthesized by Integrated DNA Technologies (Coralville, IA). One strand of each pair carried a 5′-biotin tag (Bi) to allow coating on streptavidin-coated sensorchips (SA chip; GE Healthcare). Two surfaces of comparable density were generated using DNA containing the palindromic GATA site and the single GATA site on two distinct flow cells of a single sensorchip. GATA3 was serially diluted in four 3-fold steps from 50 nM to 0.2 nM. The five concentrations of protein samples were injected at 20°C over the chip surface, using 1 min injections followed by a 5 min dissociation. Samples with different concentrations of protein were injected in random order and every injection was performed in triplicate within each experiment. All experiments were done at least three times. Data were processed using Scrubber and analyzed using CLAMP XP (Myszka and Morton, 1998). The data were fit globally using a simple 1:1 Langmuir interaction model with a correction for mass transport (Myszka et al., 1998). The results for differential protein/DNA-binding strengths were compared using Student's t test. Equal or unequal variance of the samples was determined using the F-test. Mean association and dissociation rates were used to calculate the equilibrium binding constants, and the SEM values in the k_a and k_d were used to compute the error in the mean K_D values reported in Figure 5B.

In-Gel FRET Analysis

The ssDNAs labeled with either Cy3 (containing a GATA site) or Cy5 (containing a GATC site) were synthesized by IDT and purified by high-performance liquid chromatography. A 40 bp oligonucleotide that annealed to both Cy3 and Cy5 oligos was also synthesized by IDT. The DNA was annealed at a 1:1:1 ratio of DNAs. The resulting probe, labeled at the 5′ and 3′ ends with the FRET-pair Cy3 and Cy5, respectively, consisted of two dsDNA regions at the ends and a 20 nt poly(dT) region at the center. The DNA was incubated with GATA3 or R276E mutant protein for 25 min. A native 6% (w/v) polyacrylamide gel in 0.5 × Tris borate EDTA buffer was used to resolve the free DNA from the protein/DNA complex. The gel was scanned with a Typhoon 8610 variable mode imager (Amersham Biosciences) at an excitation wavelength of 532 nm, which excited Cy3, and the fluorescence images were detected at emission wavelengths of 580 nm and 670 nm for Cy3 and Cy5, respectively. The Cy3 image was assigned in green, and the Cy5 image was assigned in red. The two images were superpositioned, resulting in the final image.

Computational Analysis

The electrostatic potential was calculated with DelPhi (Honig and Nicholls, 1995) based on the nonlinear Poisson-Boltzmann equation, which was solved in five focusing steps at physiologic ionic strength $I = 0.145$ (Rohs et al., 2009). The potential of the DNA was shown as an isopotential surface and plotted as a function of sequence in reference points at the center of the minor groove (Rohs et al., 2009). The DNA minor groove geometry was analyzed with the CURVES algorithm (Lavery and Sklenar, 1989).

ACCESSION NUMBERS

The atomic coordinates and structural factors have been deposited in the Protein Data Bank under ID codes 4HCA, 4HC7, and 4HC9.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2012.10.012.

LICENSING INFORMATION

ACKNOWLEDGMENTS

The authors thank Dr. Liang Guo, Dr. Aidong Han, and Grace Kim for experimental assistance and discussion, and staff members Corie Ralston, Peter Zwart, and Kevin Royal of the Advanced Light Source, Berkeley Center for Structural Biology, for help with data collection. L.C. is supported by National Institutes of Health (NIH) grants GM064642 and GM077320. R.R. is supported in part by grant IRG-58-007-51 from the American Cancer Society. Y.C. is partly supported by an NIH postdoctoral fellowship. Y.C., D.L.B., and L.C. designed the research; Y.C. and D.L.B. purified the protein and DNA; Y.C. and D.L.B. grew the crystals and collected diffraction data; Y.C., D.L.B., and R.D. solved and refined the structures; P.C. and I.A.L. conducted the Biacore experiments; D.L.B performed the in-gel FRET; A.C.D.M. and R.R. performed the computational work; and Y.C., D.L.B., and L.C. wrote the manuscript. All authors commented on and revised the manuscript.

Received: May 1, 2012
Revised: August 13, 2012
Accepted: October 1, 2012
Published: November 8, 2012

REFERENCES

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D Biol. Crystallogr. *58*, 1948–1954.

Bain, D.L., Heneghan, A.F., Connaghan-Jones, K.D., and Miura, M.T. (2007). Nuclear receptor structure: implications for function. Annu. Rev. Physiol. *69*, 201–220.

Bandukwala, H.S., Wu, Y., Feuerer, M., Chen, Y., Barboza, B., Ghosh, S., Stroud, J.C., Benoist, C., Mathis, D., Rao, A., and Chen, L. (2011). Structure of a domain-swapped FOXP3 dimer on DNA and its function in regulatory T cells. Immunity *34*, 479–491.

Bates, D.L., Chen, Y., Kim, G., Guo, L., and Chen, L. (2008). Crystal structures of multiple GATA zinc fingers bound to DNA reveal new insights into DNA recognition and self-association by GATA. J. Mol. Biol. *381*, 1292–1306.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Dasen, J.S., O'Connell, S.M., Flynn, S.E., Treier, M., Gleiberman, A.S., Szeto, D.P., Hooshmand, F., Aggarwal, A.K., and Rosenfeld, M.G. (1999). Reciprocal interactions of Pit1 and GATA2 mediate signaling gradient-induced determination of pituitary cell types. Cell *97*, 587–598.

Elrod-Erickson, M., Benson, T.E., and Pabo, C.O. (1998). High-resolution structures of variant Zif268 DNA complexes: implications for understanding zinc finger DNA recognition. Structure *6*, 451–464.

Ferreira, R., Ohneda, K., Yamamoto, M., and Philipsen, S. (2005). GATA1 function, a paradigm for transcription factors in hematopoiesis. Mol. Cell. Biol. *25*, 1215–1227.

Fujiwara, T., O'Geen, H., Keles, S., Blahnik, K., Linnemann, A.K., Kang, Y.-A., Choi, K., Farnham, P.J., and Bresnick, E.H. (2009). Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. Mol. Cell *36*, 667–681.

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.

Georgescu, C., Longabaugh, W.J.R., Scripture-Adams, D.D., David-Fung, E.-S., Yui, M.A., Zarnegar, M.A., Bolouri, H., and Rothenberg, E.V. (2008). A gene regulatory network armature for T lymphocyte specification. Proc. Natl. Acad. Sci. USA *105*, 20100–20105.

Guo, L., Han, A., Bates, D.L., Cao, J., and Chen, L. (2007). Crystal structure of a conserved N-terminal domain of histone deacetylase 4 reveals functional insights into glutamine-rich domains. Proc. Natl. Acad. Sci. USA *104*, 4297–4302.

Harris, R.C., Mackoy, T., Dantas Machado, A.C., Xu, D., Rohs, R., and Fenley, M.O. (2012). Opposites attract: shape and electrostatic complementarity in protein-DNA complexes. In Innovations in Biomolecular Modeling and Simulations, T. Schlick, ed. (London: The Royal Society of Chemistry), pp. 53–80.

Ho, I.C., Vorhees, P., Marin, N., Oakley, B.K., Tsai, S.F., Orkin, S.H., and Leiden, J.M. (1991). Human GATA-3: a lineage-restricted transcription factor that regulates the expression of the T cell receptor alpha gene. EMBO J. *10*, 1187–1192.

Honig, B., and Nicholls, A. (1995). Classical electrostatics in biology and chemistry. Science *268*, 1144–1149.

Horiuchi, S., Onodera, A., Hosokawa, H., Watanabe, Y., Tanaka, T., Sugano, S., Suzuki, Y., and Nakayama, T. (2011). Genome-wide analysis reveals unique regulation of transcription of Th2-specific genes by GATA3. J. Immunol. *186*, 6378–6389.

Jing, H., Vakoc, C.R., Ying, L., Mandat, S., Wang, H., Zheng, X., and Blobel, G.A. (2008). Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. Mol. Cell *29*, 232–242.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., and Chen, L. (2012). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nat. Biotechnol. *30*, 90–98.

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

Lavery, R., and Sklenar, H. (1989). Defining the structure of irregular nucleic-acids—conventions and principles. J. Biomol. Struct. Dyn *6*, 655–667.

Liew, C.K., Simpson, R.J.Y., Kwan, A.H.Y., Crofts, L.A., Loughlin, F.E., Matthews, J.M., Crossley, M., and Mackay, J.P. (2005). Zinc fingers as protein recognition motifs: structural basis for the GATA-1/friend of GATA interaction. Proc. Natl. Acad. Sci. USA *102*, 583–588.

Luscombe, N.M., Laskowski, R.A., and Thornton, J.M. (1997). NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. Nucleic Acids Res. *25*, 4940–4945.

Martin, D.I., 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.

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

Molkentin, J.D. (2000). The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. J. Biol. Chem. *275*, 38949–38952.

Myszka, D.G., and Morton, T.A. (1998). CLAMP: a biosensor kinetic data analysis program. Trends Biochem. Sci. *23*, 149–150.

Myszka, D.G., Jonsen, M.D., and Graves, B.J. (1998). Equilibrium analysis of high affinity interactions using BIACORE. Anal. Biochem. *265*, 326–330.

Newton, A., Mackay, J., and Crossley, M. (2001). The N-terminal zinc finger of the erythroid transcription factor GATA-1 binds GATC motifs in DNA. J. Biol. Chem. *276*, 35794–35801.

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

Orkin, S.H., Shivdasani, R.A., Fujiwara, Y., and McDevitt, M.A. (1998). Transcription factor GATA-1 in megakaryocyte development. Stem Cells *16* (*Suppl 2*), 79–83.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Patient, R.K., and McGhee, J.D. (2002). The GATA family (vertebrates and invertebrates). Curr. Opin. Genet. Dev. *12*, 416–422.

Pedone, P.V., Omichinski, J.G., Nony, P., Trainor, C., Gronenborn, A.M., Clore, G.M., and Felsenfeld, G. (1997). The N-terminal fingers of chicken GATA-2 and GATA-3 are independent sequence-specific DNA binding domains. EMBO J. *16*, 2874–2882.

Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S., and Honig, B. (2009). The role of DNA shape in protein-DNA recognition. Nature *461*, 1248–1253.

Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., and Mann, R.S. (2010). Origins of specificity in protein-DNA recognition. Annu. Rev. Biochem. *79*, 233–269.

Shimizu, R., Takahashi, S., Ohneda, K., Engel, J.D., and Yamamoto, M. (2001). In vivo requirements for GATA-1 functional domains during primitive and definitive erythropoiesis. EMBO J. *20*, 5250–5260.

Song, H., Suehiro, J.-i., Kanki, Y., Kawai, Y., Inoue, K., Daida, H., Yano, K., Ohhashi, T., Oettgen, P., Aird, W.C., et al. (2009). Critical role for GATA3 in mediating Tie2 expression and function in large vessel endothelial cells. J. Biol. Chem. *284*, 29109–29124.

Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R., and Flavell, R.A. (2005). Interchromosomal associations between alternatively expressed loci. Nature *435*, 637–645.

Starich, M.R., Wikström, M., Arst, H.N., Jr., Clore, G.M., and Gronenborn, A.M. (1998). The solution structure of a fungal AREA protein-DNA complex: an alternative binding mode for the basic carboxyl tail of GATA factors. J. Mol. Biol. *277*, 605–620.

Stroud, J.C., and Chen, L. (2003). Structure of NFAT bound to DNA as a monomer. J. Mol. Biol. *334*, 1009–1022.

Stroud, J.C., Lopez-Rodriguez, C., Rao, A., and Chen, L. (2002). Structure of a TonEBP-DNA complex reveals DNA encircled by a transcription factor. Nat. Struct. Biol. *9*, 90–94.

Stroud, J.C., Wu, Y., Bates, D.L., Han, A., Nowick, K., Paabo, S., Tong, H., and Chen, L. (2006). Structure of the forkhead domain of FOXP2 bound to DNA. Structure *14*, 159–166.

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., and Orkin, S.H. (1997). Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. Blood *89*, 3636–3643.

Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J., and Blobel, G.A. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. Mol. Cell *17*, 453–462.

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.

Wang, Y., Su, M.A., and Wan, Y.Y. (2011). An essential role of the transcription factor GATA-3 for the function of regulatory T cells. Immunity *35*, 337–348.

Watt, A.J., Battle, M.A., Li, J., and Duncan, S.A. (2004). GATA4 is essential for formation of the proepicardium and regulates cardiogenesis. Proc. Natl. Acad. Sci. USA *101*, 12573–12578.

Wei, G., Abraham, B.J., Yagi, R., Jothi, R., Cui, K., Sharma, S., Narlikar, L., Northrup, D.L., Tang, Q., Paul, W.E., et al. (2011). Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. Immunity *35*, 299–311.

Weiss, M.J., and Orkin, S.H. (1995). GATA transcription factors: key regulators of hematopoiesis. Exp. Hematol. *23*, 99–107.

Wuttke, D.S., Foster, M.P., Case, D.A., Gottesfeld, J.M., and Wright, P.E. (1997). Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: determinants of affinity and sequence specificity. J. Mol. Biol. *273*, 183–206.

Yu, C., Niakan, K.K., Matsushita, M., Stamatoyannopoulos, G., Orkin, S.H., and Raskind, W.H. (2002). X-linked thrombocytopenia with thalassemia from a mutation in the amino finger of GATA-1 affecting DNA binding rather than FOG-1 interaction. Blood *100*, 2040–2045.

Yu, M., Riva, L., Xie, H., Schindler, Y., Moran, T.B., Cheng, Y., Yu, D., Hardison, R., Weiss, M.J., Orkin, S.H., et al. (2009). Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. Mol. Cell *36*, 682–695.

Zheng, W., and Flavell, R.A. (1997). The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells. Cell *89*, 587–596.