Determinants of GATA-1 Binding to DNA

THE ROLE OF NON-FINGER RESIDUES*

Rodolfo Ghirlando‡ and Cecelia D. Trainor
From the Laboratory of Molecular Biology, NIDDK, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland 20892

Mammalian GATA transcription factors are expressed in various tissues in a temporally regulated manner. The prototypic member, GATA-1, is required for normal erythroid, megakaryocytic, and mast cell development. This family of DNA-binding proteins recognizes a consensus (A/T)GATA(A/G) motif and possesses homologous DNA binding domains consisting of two zinc fingers. The C-terminal finger of GATA-1 recognizes the consensus motif with nanomolar affinities, whereas the N-terminal finger shows a binding preference for a GATC motif, albeit with much reduced affinity ($K_d \approx \mu M$). The N-terminal finger of GATA-2 also shows a preference for an AGATCT binding site, with an increased affinity attributed to N- and C-terminal flanking basic residues ($K_d \approx \text{nm}$). To understand the differences in the binding specificities of the N- and C-terminal zinc fingers of GATA-1, we have constructed a series of swapped domain peptides. We show that the specificity for AGATAA over AGATCT arises from the C-terminal non-finger basic domain. Thus, the N-terminal finger binds DNA with an affinity crucial for gene expression (3, 5–8). Recent reports, however, indicate that the flanking or poorly conserved linker residues may play a key role in DNA recognition. A role for the flanking residues has been pointed out in the case of the N-terminal finger of GATA-2 (10). Like the N-terminal finger of GATA-1, this homologous finger shows a binding preference for GATC motifs. Furthermore, because of the additional basic region at the N-terminal arm, flanking residues have a higher binding affinity than for the native C-terminal finger.

Solution NMR studies of the complex formed between the C-terminal finger of cGATA-1 and the AGATAA DNA sequence highlight the role that the C-terminal basic arm plays in the high affinity recognition process (11). Briefly, the C-terminal basic arm residues wrap around and lie in the minor groove, forming specific contacts with the phosphates of bases C-13, T-25, and C-26 and the sugar portion of bases C-13, T-23, A-24, T-25, and C-26. The residues involved in these interactions all belong to the basic QTRNRK motif (residues Gln-209 to Lys-214 of cGATA-1). The importance of these basic linker residues for the high affinity interactions of the C-terminal finger of GATA-1 with DNA has been further emphasized in a recent solution structure of the fungal AREA-DNA complex (12). The basic C-terminal arm of AREA, which lacks the QTRNRK motif, wraps around the DNA. However, rather than lying in the minor groove, the arm runs parallel to the sugar phosphate backbone along the edge of the minor groove. It has been proposed that this accounts for the 1000-fold difference in affinity noted for the C-terminal finger of GATA-1 and the AREA single finger (12). Do these different modes of C-terminal arm binding influence the DNA binding specificity of the GATA zinc fingers? To address this issue and understand the different specificities observed for the homologous N- and C-terminal fingers of the GATA proteins, we have constructed a series of swapped domain peptides. We show that replacement of the C-terminal arms of the N-terminal zinc fingers of GATA-1 and -2 with the corresponding C-terminal arm of the C-terminal finger of GATA-1 results in affinities and specificities similar to those noted for the native C-terminal finger. Furthermore, we show that replacement of the LVSKRA residues (residues Leu-155 to Ala-160 of cGATA-1) in the basic arm of the N-terminal zinc finger with the corresponding QTRNRK residues from the basic arm of the C-terminal arm are sufficient for the selective recognition of the GATA motif. Replacement of the QTRNRK residues with LVSKRA in the C-terminal zinc finger does not result in the loss of specificity for the GATA motif; rather equal specificities are noted for both GATC and GATA motifs. These results highlight the role of the flanking residues in modulating the affinity and specificity of the GATA zinc fingers toward DNA.

DNA binding zinc fingers of the form CXXCX_{17}CNAC zinc finger motif (i.e. non-finger residues) in the specific recognition of DNA residues.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed. Tel.: 301-496-5889; Fax: 301-496-0201; E-mail: rodolfo@intra.niddk.nih.gov.

1 The abbreviations used are: c, chicken; GST, glutathione S-transferase; EMSA, electrophoretic mobility shift assay; dsDNA, double stranded DNA.
Expression and Purification of GATA Peptides

GSTM-GATA-1 and GST-GATA-2 N-finger Fusion Proteins—The GST N-terminal fusions to cGATA-1 N-finger/C-terminal arm (peptide NC1) and cGATA-2 N-finger (peptide G19) were expressed in BL21(DE3) Escherichia coli (Novagen). The cells were grown at 37°C in LB broth containing 50 µg/ml ampicillin, 50 µg/ml carbenicillin, 50 µM Zn(OAc)₂ and induced with 0.5 mM isopropyl-1-thio-β-D-galactopyranoside for 3 h. Approximately 30 g of wet cells (from a 4-liter culture) were lysed exhaustively against degassed (i) 10 mM Tris (pH 7.4), 0.1 M NaCl, 0.5 mM Zn(OAc)₂, 1 mM EDTA, 0.5 mM phenylmethylsulfonyl fluoride, and 1 µM pepstatin A by sonication. The lysates were clarified by centrifugation and loaded on a Q-Sepharose® Fast Flow column (Amersham Biosciences) that had been equilibrated with the same buffer. The flow-through and rinse were combined and loaded on an S-Sepharose® Fast Flow column (Amersham Biosciences). A 0.1 to 1.0 M NaCl gradient in 50 mM Tris (pH 7.4), 0.5 mM Zn(OAc)₂, 1 mM EDTA, and 0.5 mM phenylmethylsulfonyl fluoride was used to elute the GST-GATA peptides. Fractions containing DNA binding activity, as monitored by EMSA, were pooled and further purified on a glutathione-Sepharose® 4B column (Amersham Biosciences) following the manufacturer's recommendations. Fractions containing DNA binding activity were pooled and stored at −80°C in 500-µl aliquots. GST fusion proteins were found to be >95% pure by SDS-PAGE (data not shown).

GATA-1 and GATA-2 Zinc Finger His₆ Peptides—The various histidine-tagged peptides were expressed in BL21(DE3) STAR E. coli (Invitrogen) in LB broth containing 30 µg/ml kanamycin and 50 µM Zn(OAc)₂. Expression, following induction with 0.5 mM isopropyl-1-thio-β-D-galactopyranoside, was carried out at 30°C overnight to yield ~40 g of wet cells (from a 10-liter culture). Purifications on Q- and S-Sepharose® Fast Flow columns were first carried out, as described above. Peptides were further purified on Hi-Trap Chelating HP (Amersham Biosciences) columns loaded with Ni²⁺. All peptides, except for N5, which expressed poorly, were found to be >90% pure by SDS-PAGE (data not shown). Following purification, the peptides were dialyzed exhaustively against degassed (i) 10 mM Tris (pH 7.5), 10 mM EDTA, and 14.3 mM 2-mercaptoethanol, (ii) water, and (iii) 0.005% (v/v) trifluoroacetic acid. They were lyophilized to yield 80–120 mg of a dried powder; lower yields were obtained in the case of peptide N5. Peptides were reconstituted with Zn²⁺ as described (11) and stored in 20-µl aliquots at −80°C. Experiments were usually performed right after zinc reconstitution as prolonged storage usually led to losses in the DNA binding activity.

Binding Site Selections Experiments

Binding site selection experiments were performed as described (13) using the GST N-terminal fusions to cGATA-1 N-finger/C-terminal arm (peptide NC1) and cGATA-2 N-finger (peptide G19). PCR amplification of the bound oligomers was carried out using Platinum Taq® (Invitrogen) in the presence of 1.5 mM MgCl₂ with annealing temperatures of 70°C. These conditions yielded a single DNA product having the expected size. All PCR products were purified on 5% acrylamide gels in Tris borate EDTA buffer. Three selection cycles using binding conditions
described (8) were carried out using starting oligomers (13) containing 12 degenerate nucleotides (N$_x$) or nine degenerate nucleotides with a central GAT site (N$_x$GATN$_y$). The final selection was carried out in the presence of an additional 50 mM NaCl. For each protein-DNA oligomer combination ~100 selected clones were sequenced.

**DNA Binding EMSA**

Radioactively labeled DNA probes were designed based on the selection experiments. The twelve double stranded probes used had the sequence, TACAGGNNNNNNNNGGTTGCG, with the sequences of the degenerate bases N described in Table III and subsequent figures. DNA binding experiments used to determine the dissociation constants were carried out as described (8). EMSA, used to assay the relative affinity of the various peptides (Fig. 1), were carried out in 5 mM HEPES, 5 mM Tris, and 0.5 mM EDTA. In these experiments sufficient protein was added in a manner such that losses in the sample well were non-existent. All data were quantitated on a Typhoon 8600 (Amersham Biosciences).

**RESULTS**

The C-terminal Arm of the GATA-1 Binding Domain Alters the Binding Specificity of the N-fingers of GATA-1 and -2—To determine the binding specificities of the N-terminal fingers of cGATA-1 and -2, we performed a series of binding site selection studies using DNA oligomers incorporating 12 random nucleotides with or without a central GAT motif. Published results have indicated that the N-terminal zinc finger of cGATA-2 binds to DNA with micromolar affinities (10), suggesting that selection experiments with this peptide were feasible. However, unlike GATA-2, the N-terminal zinc finger of GATA-1 does not bind with high affinity to DNA (8, 9). It has been shown that the basic C-terminal arm of the C-terminal zinc finger of GATA-1 is required for high affinity binding to DNA (3, 4). To determine whether the C-terminal arm could confer high affinity DNA binding to the N-terminal zinc finger, the basic arm of the N-terminal finger was exchanged for that of the C-terminal finger (peptide GST-NC1) for site selection experiments. Surprisingly, in addition to an increased binding affinity, changes in the DNA binding specificities were also observed. Three rounds of selection with this N-terminal finger/C-terminal arm construct show a distinct preference for AGATAA sites (Table I). Using both sets of initial random DNA oligomers, greater than 75% of the clones sequenced show that an AGATAA binding sequence is selected, followed by CGATAA (11%) and AGATAG (8%). This strong preference for a GATA motif reflects the binding preferences noted for the full-length cGATA-1 (13) and the C-terminal zinc finger (4, 14) and contrasts the reported GATC specificities of the N-terminal zinc finger (9, 14) suggesting a role for the C-terminal arm of the cGATA-1 DNA binding domain. It would have been instructive to carry out binding site selection experiments with the N-terminal finger of GATA-1, thus allowing for a direct comparison with the data in Table I. These experiments were not feasible. This peptide has a lower affinity for DNA than the N-terminal finger of GATA-2 (8, 9), and in our hands the latter peptide leads to a poorer definition of the sequence specificity (see below). Despite the lack of a direct comparison, data presented in Table III and Fig. 2 clearly demonstrate the role of the C-terminal arm in determining binding specificity.

Site selection experiments with the N-terminal finger of GATA-2 (peptide GST-G19) led to a poorer definition of the sequence specificity (Table II), leading to the following sequence preference: AGAT > GCGT > CGAT, with no definitive information on the nucleotides following the central GAT motif. These data differ from site selection experiments on the full-length cGATA-2 where a distinct preference for AGATAA and AGATCT is reported (13), reflecting the binding properties of the C-terminal finger of GATA-2. Binding studies on the N-terminal finger of GATA-2 have shown that this peptide has a preference for the GATC motif, binding with a $K_d$ of 5 nM to AGATCT and a $K_d$ of 28 nM to AGATAA (10). Even though this difference in affinity is reflected partially in the selection experiments, in which AGATC is selected over AGATAA in a 2.2:1 ratio (data not shown), an AGATCT site is not exclusively selected. This higher affinity for GATC motifs was further confirmed in a series of experiments in which the relative binding of the GATA-2 N-terminal finger for various DNA sites was determined (Fig. 2B). Unlike the fusion of the N-terminal fingers of GATA-1 and -2 to the C-terminal arm of GATA-1 (see below), it was not possible to determine the absolute binding affinities using a Scatchard analysis. This problem, essentially because of the micromolar affinities of interaction, was noted for many of the N-terminal finger constructs; it is for this reason that relative binding assays were performed. As in the case of the GATA-2 N-terminal finger, the GATA-1 N-terminal finger binds preferentially to GATC motifs (Fig. 2A). This specificity, noted previously by Newton et al. (9), indicates that the replacement of the basic arm of the N-terminal finger with that of the C-terminal arm of the GATA-1 C-terminal finger (peptide GST-NC1) significantly affects the DNA binding specificity.

**TABLE I**

<table>
<thead>
<tr>
<th>Oligomer selected</th>
<th>Frequency</th>
<th>Start = NNNNNGATNNNN</th>
<th>Oligomer selected</th>
<th>Frequency</th>
<th>Start = NNNNNNNNNNNNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVSVGATAAAVV</td>
<td>77/96</td>
<td>VVAGATAAAVV</td>
<td>79/105</td>
<td>VVAGATAAAVV</td>
<td></td>
</tr>
<tr>
<td>SVSVVGATAAAVV</td>
<td>11/96</td>
<td>VVCGATAAAVV</td>
<td>12/105</td>
<td>VVAGATAAVV</td>
<td></td>
</tr>
<tr>
<td>SVSVVGATAAAVV</td>
<td>6/96</td>
<td>VVAGATAAVV</td>
<td>11/105</td>
<td>VTVGATAAVV</td>
<td></td>
</tr>
<tr>
<td>SVSVAAATAAVV</td>
<td>2/96</td>
<td>VTVGATAAAVV</td>
<td>2/105</td>
<td>VVCGATAAVV</td>
<td></td>
</tr>
<tr>
<td>SVSVAGATAAAVV</td>
<td>10/96</td>
<td>VTVGATAAVV</td>
<td>105/105</td>
<td>VVAGATAAVV</td>
<td></td>
</tr>
</tbody>
</table>

![Fig. 2. The N-terminal zinc fingers of GATA-1 and GATA-2 bind preferentially to GATC motifs.](https://example.com/fig2.png)

Relative binding affinities, depicted as the fraction of bound DNA, for various dsDNA oligomers (Table III) were determined at a single peptide and DNA concentration. Experiments were carried out in the presence of 44 nM DNA and 4.9 nM of the N-terminal zinc finger (peptide NN4) (A) or 1.7 nM of the GATA-2 N-terminal zinc finger (peptide G19) (B). Essentially no binding of peptide NN4 to the GATA and GATC sites was noted. All data represent the average of at least three determinations, and error bars represent S.E. C, electrophoretic mobility shift assays comparing the binding specificity of peptide G19 for various dsDNA oligomers, corresponding to data presented in B.
of the C-terminal finger (peptide NC1) is responsible for the altered DNA specificity.

To confirm the role of the C-terminal basic arm of GATA-1 in the exclusive selection of a GATA motif, we measured the affinities of peptide NC1 for a series of GATA and GATC DNA oligomers (Table III). The majority of the sequences selected, namely AGATAA, CGATAV, and AGATAG (Table I), are bound with affinities of 3–5 nM, suggesting that the selection experiments are carried out under conditions such that only high affinity ($K_d < 10$ nM) interactions are observed. The frequency of the sequences selected does not mirror the experimentally determined binding affinities. For example, peptide NC1 binds to a TGATAA motif with a $K_d$ of 3.2 nM, yet such a motif is selected with a frequency of less than 2% in one of the selection experiments. Furthermore, peptide NC1 shows a 10-fold lower affinity for the CGTAG motif even though a single clone containing this sequence was obtained (Table I). It is clear, nevertheless, that a distinct preference for the GATA motif exists with affinities similar to those observed for the binding of the C-terminal finger to AGATAA (4) and other high affinity sites (8). To further determine the role of the C-terminal arm in establishing a high affinity preference toward GATA motifs, a GATA-2 N-terminal zinc finger/GATA-1 C-terminal arm peptide (G20) was prepared. Unlike the GATA-2 N-finger (see this work and Ref. 11), this peptide shows a distinct high affinity preference for GATA motifs (Table III), in a manner similar to the analogous peptide NC1. Overall, these results demonstrate that fusion of the C-terminal basic arm of GATA-1 to the N-terminal zinc fingers of GATA-1 or -2 alters their binding specificity from GATC to that observed for the full-length protein or C-terminal zinc finger. Moreover, the binding affinities are similar to values noted for both the C-terminal zinc finger and the full-length GATA-1 (4, 8).

The Conserved QTRNRK C-terminal Arm Residues Are Sufficient to Alter the Specificity of the C-terminal Zinc Finger—Structural studies of the complex formed between the C-terminal zinc finger of cGATA-1 and the AGATAA DNA sequence show that the C-terminal basic arm forms a number of contacts with the sugar and phosphate backbone of bases T-23, A-24, and T-25 (i.e. equivalent to the ATA in AGATAAA (11)). The highly conserved non-zinc finger QTRNRK residues (i.e. Gln-209 to Lys-214) are involved in this interaction. A BLAST analysis of the C-terminal arm of GATA-1, spanning residues Gln-196 to Arg-223, shows that the QTRNRK sequence is conserved as QTR-RK in the GATA-1 and -4 families or as QTRNR– in the GATA-2 and -3 families. This conservation may account, in part, for the GATA motif binding specificity of the various members of the GATA-1, -2, and -3 (13–15) and possibly -4 families, because the specificity of the full-length protein is usually determined by the specificity of the C-terminal zinc finger. Interestingly, the corresponding residues on the N-terminal zinc finger arm, namely LVSKRA residues Leu-155 to Ala-160 of cGATA-1, are conserved in GATA-1 but not as conserved among the other GATA members.

To evaluate the contribution of the QTRNRK residues on the C-terminal arm toward the DNA binding specificity, a number of constructs were designed (Fig. 1) such that (i) the basic LVSKRA residues of the arm of the N-terminal finger were replaced with QTRNRK (peptide NN6), and (ii) the QTRNRK residues of the arm of peptide NC1 were replaced with LVSKRA (peptide NC3). These constructs were prepared by site-directed mutagenesis in two steps, involving the interconversion of RNRK to Skra, generating peptides NN5 and NC2 respectively, followed by the interconversion of the remaining QT and LV (peptides NN6 and NC3). Because of technical considerations resulting from the low DNA binding affinities observed for many of these peptides, single point binding analyses were carried out. To ensure that data were reproducible, assays were carried out in triplicate. Furthermore, experiments were designed such that the binding of a single peptide to various DNA binding sites was compared, rather than comparing the binding of different peptides to particular DNA sequences. This avoids the problem associated with the differential recovery of binding activity for the various peptides. Peptide NN5 binds preferentially to AGATCT and AGATCA motifs, albeit with a very low affinity (Fig. 3A). Note that relative binding assays are carried out with sufficient peptide such that sample losses are not observed. Sample losses and nonspecific binding were observed with increasing concentrations of NN5. This nonspecific binding, partially because of the impurity of this particular peptide, was of concern for experiments with peptide NN6 in which the LVSKRA residues of the N-terminal finger are changed to QTRNRK. Accordingly, in addition to these mutations, the basic KGKKRR C-terminal arm residues are appended to create peptide NN8. Even though these residues do not alter the binding specificity of the C-terminal zinc finger (4, 16) and are not observed in the structure of the complex with DNA (11), these residues reduce nonspecific DNA binding. Unlike NN4, which binds to GATC motifs, this peptide shows a distinct preference for GATA motifs, and particularly (A/T)GATAA (Fig. 3B), as noted for NC1

**Table II**

**Binding site selection for the N-terminal finger of GATA-2 (peptide G19)**

Summary of the binding selection experiments after three rounds of selection. Nucleotides labeled S represent G, C, whereas V represents A, G, C. In the first experiment only two instances of the palindromic site, AGATCT, were noted. Unlike GATA-1 (Table I), a large number of selected sequences had double GATA sites. In the first experiment, 48 of the 107 sequences had double GATA sites, and consensus sequences for the oligomer-defined GATA are shown. The second selection experiment had 21 double sites.

<table>
<thead>
<tr>
<th>Oligomer selected</th>
<th>Frequency</th>
<th>Oligomer selected</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGATV</td>
<td>103/123</td>
<td>AGATV</td>
<td>103/123</td>
</tr>
<tr>
<td>GGATV</td>
<td>10/124</td>
<td>GGATV</td>
<td>10/123</td>
</tr>
<tr>
<td>CGATV</td>
<td>7/123</td>
<td>CGATV</td>
<td>7/123</td>
</tr>
<tr>
<td>TGATN</td>
<td>3/123</td>
<td>TGATN</td>
<td>3/123</td>
</tr>
</tbody>
</table>

**Table III**

**GATA-1 and GATA-2 N-finger constructs fused to the C-terminal arm of GATA-1 bind preferentially to GATA motifs**

DNA binding affinities of peptides NC1 and G20 for the sequences indicated, in the context of TACGGAGNNNNNNNGGGTTGCG. Dissociation constants were determined by Scatchard analysis and represent the mean of the number of experiments shown in parentheses. Standard deviations are indicated. In certain cases it was not possible to determine the affinity. Accordingly, minimum $K_d$ estimates are given based on the protein and DNA loading concentrations.

<table>
<thead>
<tr>
<th>DNA oligomer</th>
<th>GATA-1 Nf/Carm (peptide NC1)</th>
<th>GATA-2 Nf/Carm (peptide G20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A GAT AA</td>
<td>2.8 ± 1.3 (5)</td>
<td>4.4 ± 0.7 (3)</td>
</tr>
<tr>
<td>C GAT AA</td>
<td>5.1 ± 2.3 (5)</td>
<td>2.1 ± 0.6 (3)</td>
</tr>
<tr>
<td>A GAT AG</td>
<td>4.6 ± 1.5 (7)</td>
<td>2.5 ± 0.8 (3)</td>
</tr>
<tr>
<td>T GAT AA</td>
<td>3.2 ± 0.4 (8)</td>
<td>3.4 ± 2.0 (4)</td>
</tr>
<tr>
<td>C GAT AG</td>
<td>33 ± 12 (5)</td>
<td>14.9 ± 2.1 (3)</td>
</tr>
<tr>
<td>A GAT CT</td>
<td>71 ± 41 (3)</td>
<td>&gt;100 (2)</td>
</tr>
<tr>
<td>A GAT TG</td>
<td>94 ± 25 (3)</td>
<td>12.8 ± 3.7 (4)</td>
</tr>
<tr>
<td>A GAT GG</td>
<td>&gt;300 (2)</td>
<td>&gt;300 (1)</td>
</tr>
<tr>
<td>G GAT CA</td>
<td>&gt;300 (2)</td>
<td>&gt;300 (1)</td>
</tr>
<tr>
<td>C GAT CA</td>
<td>&gt;300 (2)</td>
<td>&gt;300 (1)</td>
</tr>
<tr>
<td>G GAT CG</td>
<td>&gt;300 (2)</td>
<td>&gt;300 (1)</td>
</tr>
<tr>
<td>A GAT CA</td>
<td>&gt;300 (2)</td>
<td>&gt;300 (1)</td>
</tr>
</tbody>
</table>

$K_d$ nM

C. D. Trainor, unpublished data.
Relative binding affinities, depicted as the fraction of bound DNA, for various dsDNA oligomers (Table III) were determined at a single peptide concentration, it would appear as though the relative affinities have also increased substantially (see Fig. 2 and Fig. 3). Evidence that the KGKKRR residues are not involved in this switch of DNA binding specificity will be presented below.

To assess whether the converse is true, namely, whether the loss of the RNRK or QTRNRK residues results in the loss of specificity for GATA, peptides NC2 and NC3 were prepared. These peptides correspond to the GATA-1 N-terminal finger/C-terminal arm fusion peptide results in a mixed specificity for GATA and GATC motifs (Fig. 3, C and D). It therefore seemed possible that other C-terminal arm residues might be responsible for the retention of the GATA specificity. To validate this hypothesis, similar sets of mutants were tested for the C-terminal zinc finger of GATA-1. As noted in Fig. 4B, replacement of RNRK with SKRA (peptide CC12) results in the loss of binding specificity for the (C/A)GATAAG motifs and an increase in binding specificity for the GATC sites, particularly AGATCT and AGATCA. Similar observations are made for peptide CC13 in which the QTRNRK residues are changed to LVSKRA, highlighting the importance of these residues in establishing a GATA specificity.

Analysis of the solution structure of the C-terminal zinc finger of GATA-1 bound to DNA (11) shows that the α-helical residues of the zinc finger are also involved in the recognition of AGATAA. These residues, namely YKLH (Tyr-192 to His-195 of cGATA-1), are not all conserved in the N-terminal zinc finger, and, in a manner similar to the steroid and thyroid hormone receptors zinc fingers (18), it was thought that such residues might play a key role in the specific DNA recognition. The solution NMR structure of the GATA-DNA complex shows that Lys-193 interacts with the A6pG7pA8 phosphates on the forward strand, His-195 interacts with the T21pT22 phosphate on the reverse strand, and Leu-194 interacts directly with bases T23 and T22, representing the 3′ AA complement in AGATAA (11). Interestingly, residues KHL are also conserved in the AREA protein and exhibit similar properties in the AREA zinc-finger complex with CGATAG (12); Lys interacts with the C4pG5pA6 phosphates on the forward strand, His interacts with the T17pC18 phosphate on the reverse strand, and Leu interacts directly with the bases T19 and C18 suggesting that this leucine residue may be playing a role in the recognition of a GATA motif rather than a GATC. However, the corresponding HRLN residues (His-138 to Asn-141 of cGATA-1) on the N-terminal finger contain this same leucine residue, hinting that this residue may after all not be as crucial for GATA recognition over GATC. To test whether the context in which this leucine is found may play a role, a series of α-helical mutant peptides were constructed such that the HRLN residues of the N-terminal zinc finger were replaced (see Tables I and III). These data demonstrate that changing the LVSKRA residues with QTRNRK is sufficient to alter the DNA binding specificity of the N-terminal finger from GATC to GATA. Furthermore, based on the fractions of bound DNA and the peptide concentration, it would appear as though the relative affinities have also increased substantially (see Fig. 2A and Fig. 3B). Evidence that the KGKKRR residues are not involved in this switch of DNA binding specificity will be presented below.

To assess whether the converse is true, namely, whether the loss of the RNRK or QTRNRK residues results in the loss of specificity for GATA, peptides NC2 and NC3 were prepared. These peptides correspond to the GATA-1 N-terminal finger/C-terminal arm fusion (peptide NC1) in which the QTRNRK residues are either changed to QTSKRA (peptide NC2) or LVSKRA (peptide NC3). Relative binding assays show that these peptides bind equally well to AGATAA, TGATAA, AGATCT, and AGATCA. Similar observations are made for peptide CC13 in which the QTRNRK residues are changed to LVSKRA, highlighting the importance of these residues in establishing a GATA specificity.
with those of the C-terminal finger (peptides NN9 and NN10), and the YKLH residues of the C-terminal zinc finger were replaced with HRLN (peptides CC16 and CC17). Relative binding experiments with peptide CC16 show that both the binding specificity and affinities of the wild-type C-terminal finger are maintained (see Fig. 5A and Fig. 4A). In a similar fashion, binding experiments with peptide NN9 lead to a specificity very similar to that noted for peptide NN5 in which the SKRA residues of the N-terminal finger are mutated to RNRK (see Fig. 5B and Fig. 3A). These peptides recognize the AGATCA and AGATCT sites with low affinities and appear to show no preference for the other GATC motifs recognized by the N-terminal zinc finger (Fig. 2A). This lack of specificity may actually be a matter of sensitivity because of the very low binding affinities. Consequently, replacing the HRLN α-helical residues on the N-terminal zinc finger has no effect on the DNA binding specificity for these two types of binding sites. Interestingly, when both the HRLN and LVSKRA residues of the N-terminal zinc finger are changed to YKLH and QTRNRK, respectively (peptide NN10; see Fig. 5C), a mixed GATA and GATC specificity very similar to that noted for peptide NC3 (Fig. 3D) is observed. Thus, the C-terminal zinc finger α-helical residues, together with a QTRNRK arm in the context of an N-terminal zinc finger, shows the same specificity as for the GATA-1 N-terminal zinc finger/C-terminal arm in which QTRNRK are changed to LVSKRA. Therefore, although the C-terminal arm residues other than QTRNRK appear to contribute to the preference for GATA, the α-helical residues do not. The complementary experiment with peptide CC17 could not be carried out because of problems with protein expression.

The apparent DNA binding affinity of peptide NN8 (Fig. 3B) suggests a possible role for C-terminal arm residues other than QTRNRK in stabilizing the peptide-DNA interaction. Despite the clear preference for GATA motifs, the binding affinity appears to be unexpectedly lower than that observed for peptide CC13 (Fig. 4C). On the other hand, peptide NC3 (Fig. 3D) has an estimated affinity for GATA motifs that is similar to, if not slightly lower than, that observed for peptide NN8. This additional observation either discounts a possible role for non-QTRNRK residues or indicates that core finger residues, as noted under “Discussion,” stabilize non-QTRNRK residues on the C-terminal arm. However, care needs to be exercised in these comparisons; differences in the apparent affinity may arise partially from the differential recovery of binding activity, as well as differences in the stability of zinc coordination between the N-terminal and C-terminal fingers.

It has been demonstrated that swapping the C-terminal arm of the N-terminal zinc finger with that of the C-terminal zinc finger results in specificities and affinities similar to those noted for the C-terminal finger. In the case of the N-terminal finger we showed that mutation of the LVSKRA residues to QTRNRK was sufficient to change the specificity from GATC to GATA; however, the converse was not true for the C-terminal zinc finger. To assess what the contributions from the core finger may be, a C-terminal zinc finger/N-finger arm fusion peptide was constructed (peptide CN18). Relative binding assays with this peptide show a mixed GATA and GATC specificity (Fig. 5D), much like the N-terminal zinc finger/C-terminal arm construct in which the QTRNRK has been changed to LVSKRA (peptide NC3; see Fig. 3D). Thus, when the C-terminal arm required for the high affinity interaction is removed from both the N-terminal zinc finger/C-terminal arm fusion and the C-terminal zinc finger, both fingers show the same mixed specificity toward GATC and GATA sites.

**Discussion**

Flanking Residues Determine the DNA Binding Specificities of the GATA-1 Zinc Fingers—Even though the N- and C-terminal zinc finger domains of GATA-1 are quite homologous, they have different affinities and specificities toward DNA. The N-terminal zinc finger binds weakly to DNA with a preference for GATC sites, whereas the C-terminal zinc finger binds to GATA sites with high affinity (4, 9). The N- and C-terminal core finger residues of GATA-1 recognize the GAT motif on DNA, as expected from their homology, a result confirmed by selection experiments with the GATA-2 N-terminal finger, which shares the same core finger residues as the N-terminal finger of GATA-1. Only four residues of the GATA-1 C-terminal zinc finger, namely Thr-173, Leu-174, Arg-176, and Arg-177, appear to be involved in the recognition of the AGAT DNA (11). These core finger residues are conserved in the N-terminal zinc fingers of GATA-1 and -2 and are presumably responsible for...
concentration. Experiments were carried out in the presence of 44 nM DNA and 1.3

binding affinities, depicted as the fraction of bound DNA, for various dsDNA oligomers (Table III) were determined at a single peptide and DNA

The C-terminal end of the zinc finger do not determine the GATA versus GATC preference. Relative binding affinities, depicted as the fraction of bound DNA, for various dsDNA oligomers (Table III) were determined at a single peptide and DNA

Fig. 5. The α-helix residues on the C-terminal end of the zinc finger do not determine the GATA versus GATC preference. Relative binding affinities, depicted as the fraction of bound DNA, for various dsDNA oligomers (Table III) were determined at a single peptide and DNA

the recognition of the GAT motif, in a manner similar to the corresponding PL-RR residues of the AREA zinc finger (12). Although residues within the core of the N- and C-terminal fingers retain their GAT DNA binding characteristics, non-finger residues control the preference for a GATA or GATC motif on DNA. We show, here, that the C-terminal basic arm of the GATA-1 DNA binding domain bestows on the N-terminal zinc finger preferential binding to GATA motifs. We further show that the replacement of the poorly conserved LVS KRRA linker residues with the highly conserved C-terminal arm QTRNRK sequence is sufficient to bring about this change in DNA specificity. The replacement of the analogous residues from the arm of the N-terminal zinc finger (i.e. LVS KRRA) into the C-terminal finger causes an increase in affinity for GATA sites. Nevertheless, a strong preference for GATA sites remains, indicating that other residues within the C-terminal basic arm are also involved. These residues comprise portions of the flanking region other than the α-helix residues, as the α-helix residues are not involved in the differential distinction of GATA versus GATC. A number of hydrophobic clusters, in particular Thr-171, Leu-201, and Lys-204 and Thr-173, Gly-207, and Ile-208 stabilize the position of the C-terminal basic arm (11); it is possible that these intramolecular interactions are partially responsible for this asymmetrical complexity. Furthermore, these interactions, or lack thereof, may also contribute toward the different DNA binding affinities observed for peptides having N-terminal versus C-terminal core fingers.

The QTRNRK residues, as evidenced by structural studies, dictate the positioning of the C-terminal basic arm around the DNA such that it lies in the minor groove (11). In the absence of these residues, such as in the fungal GATA protein AREA, the C-terminal basic arm runs parallel to the phosphate backbone of the minor groove resulting in lower DNA binding affinities (12). It is possible that the C-terminal arm of the C-terminal zinc finger of GATA-1 is unable to interact efficiently with the minor groove of GATC containing sites, resulting in lower affinities for non-consensus DNA binding sites. These studies also highlight the importance of the GATA-1 C-terminal arm residues Gly-207 and Lys-214 in positioning the arm into the minor groove (12); Lys-214, the last residue of the QTRNRK motif, anchors the basic arm by recognition of base T9, whereas Gly-207 provides the required flexibility. In the context of the N-terminal zinc finger, where the residue corresponding to Gly-207 is Arg-153, we show that the replacement of SKRA with RNRK is not sufficient to alter the DNA binding specificity from GATC to GATA. The whole QTRNRK motif is required for the selective recognition of GATA motifs, suggesting that the Lys-214 residue is partially, but not solely responsible, for the binding of the basic arm into the minor groove.

Roles of the Zinc Fingers and Linker Region—The GATA proteins influence gene expression in several ways. GATA-1 has a transactivation domain that is required for full activity of the protein and can stimulate the expression of reporter genes in heterologous cells (3, 5). GATA-1 can partially disrupt nucleosomes in vitro (20) and is therefore involved in the alteration of chromatin structure. We have shown previously (21) that GATA-1 bends DNA by 24° in a site-independent manner, indicating that it may play an architectural role in regulating transcription. In fact, GATA-1, a critical member of a large complex of transcription factors that includes SCL, LMO-2, and LBD-1 found in immature hematopoietic cells (22, 23), also interacts with many other transcription factors through its zinc finger domains and other regions. In this manner it influences lineage choice within hematopoietic cells (24). Inactivation of the GATA-1 gene is lethal in mice with failure of erythroid, megakaryocytic, and mast cell development (25). Rescue experiments in GATA-1 negative cells and mice have confirmed that both zinc fingers are essential for definitive erythropoiesis (26–31). Mutants with inactive C-terminal zinc fingers do not rescue embryonic lethality in GATA-1 knockout mice, highlighting the importance of the C-terminal zinc finger in DNA recognition (27). Rescue experiments with proteins containing similar mutations in the N-terminal zinc finger result in live mice that are severely impaired in definitive erythropoiesis (27). Whereas some of the observed phenotype is because of the inability of these mutant proteins to interact with the critical cofactor FOG (28, 32), N-terminal zinc finger mutants that retain the ability to interact with FOG but have impaired DNA binding also show defects in erythropoiesis in transgenic rescue
assays (33). Interestingly, such mutations in the N-terminal zinc finger of GATA-1 have been found in families with X-linked thrombocytopenia and thalassemia (33). It appears as though the importance of the DNA binding function of the N-terminal zinc finger is not restricted to the vertebrate GATA factors. A splice variant of the Drosophila GATA factor, Serpent, which contains two zinc fingers spaced in a fashion identical to the vertebrate GATA factors, has been identified recently (34). Although the single and double finger forms activate many genes equivalently, only the splice variant with two fingers is able to activate gcm, a gene that contains multiple palindromic GATA sites in its promoter (34). This activation is not dependent on the U-shaped protein, the Drosophila FOGER equivalent, indicating that binding of both fingers to DNA is essential.

The interaction of the N-terminal zinc finger of GATA-1 with DNA is important on DNA binding sites that contain double and multiple GATA binding motifs (8). The interaction of both GATA-1 fingers with these binding sites results in high inter-DNA bending (21), the relative orientation of the two fingers can be predicted for the various double GATA sites. Positioning of this linker region into the minor groove is not required for GAT recognition; thus the linker region may adopt DNA binding site-specific spatial trajectories. Interestingly, basic residues of this linker region have been implicated in the interaction of GATA-1 with other proteins (44) and itself (45, 46), suggesting that GATA-1-bound DNA may act as a nucleus for other cofactors or GATA-1 without the loss of DNA binding.

Acknowledgments—We are grateful to Gary Felsenfeld for excellent advice and encouragement. We acknowledge the reviewers for helpful comments.

REFERENCES
36. Ohneda, K., Shimizu, R., Nishimura, S., Muraoaa, Y., Takahashi, S., Engel, M.
Determinants of GATA-1 Binding to DNA: THE ROLE OF NON-FINGER RESIDUES
Rodolfo Ghirlando and Cecelia D. Trainor

doi: 10.1074/jbc.M306410200 originally published online August 26, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M306410200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 46 references, 30 of which can be accessed free at http://www.jbc.org/content/278/46/45620.full.html#ref-list-1