Involvement of the N-finger in the Self-association of GATA-1*

(Received for publication, January 12, 1998, and in revised form, August 3, 1998)

Joel P. Mackay†, Kasper Kowalski, Archa H. Fox, Robert Czolij, Glenn F. King, and Merlin Crossley‡

From the Department of Biochemistry, University of Sydney, Sydney, New South Wales 2006, Australia

Zinc fingers are recognized as small protein domains that bind to specific DNA sequences. Recently however, zinc fingers from a number of proteins, in particular the GATA family of transcription factors, have also been implicated in specific protein-protein interactions. The erythroid protein GATA-1 contains two zinc fingers: the C-finger, which is sufficient for sequence-specific DNA-binding, and the N-finger, which appears both to modulate DNA-binding and to interact with other transcription factors. We have expressed and purified the N-finger domain and investigated its involvement in the self-association of GATA-1. We demonstrate that this domain does not homodimerize but instead makes intermolecular contacts with the C-finger, suggesting that GATA dimers are maintained by reciprocal N-finger–C-finger contacts. Deletion analysis identifies a 25-residue region, C-terminal to the core N-finger domain, that is sufficient for interaction with intact GATA-1. A similar subdomain exists C-terminal to the C-finger, and we show that self-association is substantially reduced when both subdomains are disrupted by mutation. Moreover, mutations that impair GATA-1 self-association also interfere with its ability to activate transcription in transfection studies.

The transcription factor GATA-1 recognizes WGATAR motifs (where W represents A/T and R represents A/G), which are found in the control regions of the vast majority of erythroid genes (for reviews see Refs. 1 and 2). In particular, GATA-1 binds to elements within both the promoters and the locus control regions of the α- and β-globin-like genes, and it has been proposed that it may be involved in communication between these distant enhancer-like elements and the globin promoters (3–5). The biological importance of GATA-1 has been demonstrated by knockout experiments in mice, which showed that it is strictly required for erythroid cell development (6, 7). Other members of the GATA family, such as GATA-2 and -3, also play essential roles in the hematopoietic system (8–10). Other members of the GATA family, such as GATA-2 and -3, also play essential roles in the hematopoietic system (8–10).

The GATA-1 protein contains two adjacent zinc finger domains, encoded by successive exons (12–14). We refer to these fingers as the N-terminal finger, or N-finger (NF) and the C-terminal finger, or C-finger (CF), according to their relative positions in the protein. The two domains each contain four cysteines and are about 50% identical in amino acid sequence (Fig. 1). The CF, together with a C-terminal extension (encoded in the next exon), has been shown to be both necessary and sufficient for the recognition of WGATAR motifs (15, 16). In contrast, the exact role of the NF has proved more difficult to define. Early studies, focusing on GATA-1’s ability to transactivate promoters in nonerythroid cells, indicated that the NF was not essential (15). However, more physiologically relevant experiments investigating GATA-1’s ability to induce erythroid maturation have revealed a critical role for this finger (17).

While the NF of GATA-1 cannot bind DNA in isolation (unlike the NFs of GATA-2 and -3; Ref. 18), it does play a role in DNA binding. Biochemical studies have demonstrated that the NF increases the stability of GATA-1-DNA complexes (15), influences the specificity of binding (19), and is instrumental in the recognition of double WGATAR motifs, which occur in a number of important erythroid promoters (including GATA-1’s own promoter) (20, 21). The GATA double zinc finger domain is also implicated in specific protein-protein interactions; it has been shown to mediate the formation of complexes with Fog (22), Sp1, and EKLF (23) and is implicated in GATA self-association (24–28). It is known that only the NF (and not the CF) is required for formation of the complex with Fog, whereas in the cases of Sp1, EKLF, and the self-association of GATA-1, both the NF and the CF are implicated. It is also likely that the NF plays a role in GATA-1’s interaction with the LIM domain protein RBNT2/LMO2 (29), although in this case the specific contributions of the NF and the CF have not been delineated.

The structure of the CF bound to DNA has been determined using NMR spectroscopy (16), but relatively little is known about the NF. In the present work, we have expressed and purified the NF domain, characterized its metal-dependent folding properties, and investigated its role in GATA-1 self-association. We show that it does not homodimerize in isolation but instead mediates GATA-1 self-association by interacting with the CF. We identify a minimal 25-amino acid subdomain that is able to interact with intact GATA-1. This subdomain, which we refer to as the NF-tail, lies immediately downstream of the core zinc-binding region of the NF and contains an Asn-Arg-Pro-Leu motif followed by a short basic stretch. An analogous subdomain is contained in the CF and has previously been implicated in GATA-1 homodimerization (26). We show that when either subdomain is disrupted by mutation, the self-association is reduced. We have also tested the effect of these mutations on the ability of GATA-1 to activate transcription from single GATA and multi-GATA site promoters. In both cases, mutations that interfere with self-association reduce the ability of GATA-1 to activate transcription, but the effect is more dramatic on multi-GATA site promoters. These results suggest that the ability of GATA factors to self-associate may...
play an important role in the regulation of genes containing multiple GATA sites.

**MATERIALS AND METHODS**

**Plasmids, Oligonucleotides, and Mutagenesis**—The region encompassing exon 4 of murine GATA-1 (nucleotides 656–843, GenBank accession number X17563) and encoding the entire NF domain (NF200–248) was amplified by PCR using primers MC287 (5'-GATGCAGGGCCAGATGTGTGTA-3') and MC288 (5'-CGGAAATCATGCCCCTTGCTGTCACTA-3'). The resulting fragment was cut with BamHI and EcoRI and cloned into pGEX-2T (Amersham Pharmacia Biotech), thereby creating in frame fusions with the glutathione S-transferase gene. C-terminal truncation mutants NF200–248 and NF200–243 were similarly constructed with the primers NF248 and A17 (5'-CAGAGATACCTGCTCCGGCCG-3') or A18 (5'-CGGAAATCATGCCCCTTGCTGTCACTA-3'). The resulting fragments were cloned into pET3a (Novagen) and expressed in E. coli BL21 (DE3) bacteria. The specific bands were purified, and a second polymerase chain reaction was carried out with primers MC287 and A17, respectively. Mutation of Cys260 to Ala was achieved by amplification with primers A10 (5'-CGTGACCCCTATGGGAGGAGGCTTAC-3') and MC288. Pro205-Thr was generated by overlap polymerase chain reaction. The two initial reactions used MC287 and A21 (5'-CGGGCAGGTTGTGCAATGGCCTGTC-3') and A20 (5'-GACAAGGCATCTACACTGTGGCGG-3') and MC288. The specific bands were purified, and a second polymerase chain reaction was carried out with primers MC287 and MC288. Sequencing was carried out to verify the mutagenesis.

**Preparation of Recombinant N-finger Fragments and Fusion Proteins**—Escherichia coli BL21 (DE3) bacteria were transformed with the recombinant plasmid and propagated in order to produce recombinant material. Typically, 4 liters of culture were grown at 37 °C to an _A_ _opt_ of around 0.6, at which time expression of the GST-NF fusion proteins was induced by the addition of isopropyl-β-D-thiogalactoside to a final concentration of 1 mM. After 3–5 h of incubation, the bacteria were harvested by centrifugation, washed twice with a buffer containing 10 mM Tris, 50 mM NaCl, 0.5% Nonidet P-40 (NP-40), and 1 mM TCEP, and then resuspended in 10 ml of 10 mM Tris (pH 8.0), 50 mM NaCl, 0.5% Nonidet P-40, and 1 mM TCEP. After 20 min, the samples were loaded onto a Glutathione-agarose affinity column. The columns were washed, and then the fusion proteins were released from the columns by the addition of thrombin (100 units) to each column. The eluate was collected and lyophilized. It was then resuspended in water and subjected to reverse phase HPLC on a semi-preparative Vydac C18 column. The eluate was collected and lyophilized. The major absorbance peaks were then identified using positive ion electrospray ionization mass spectrometry. The purity of the fusion proteins was verified by reverse phase HPLC. The purity of the NF200–248 fusion protein was verified by size exclusion chromatography on a Superase 6 column (Pharmacia Biotech) and by the presence of a single band on SDS-PAGE.

**Gel Retardation Experiments**—Gel retardation experiments were carried out essentially as described by Crossley et al. (26). The oligonucleotides used in gel retardation experiments contained a GATCC flanking site and a 20-bp 5‘-biotinylated GATA-1 oligonucleotide (GATCTCCGGCAACTGATAAGGATTCCCTG) and a 20-bp NF-κB oligonucleotide (GGATCCGCCATGGAGGCCAGAGAGGTGA). The resulting fragment was cut with BamHI and EcoRI and ligated into pET3a (Novagen) and expressed in E. coli BL21 (DE3) bacteria. The specific bands were purified, and a second polymerase chain reaction was carried out with primers MC287 and A17 (5'-CAGAGATACCTGCTCCGGCCG-3') or A18 (5'-CGGAAATCATGCCCCTTGCTGTCACTA-3'). The resulting fragments were cloned into pET3a (Novagen) and expressed in E. coli BL21 (DE3) bacteria. The specific bands were purified, and a second polymerase chain reaction was carried out with primers MC287 and A17, respectively. Mutation of Cys260 to Ala was achieved by amplification with primers A10 (5'-CGTGACCCCTATGGGAGGAGGCTTAC-3') and MC288. Pro205-Thr was generated by overlap polymerase chain reaction. The two initial reactions used MC287 and A21 (5'-CGGGCAGGTTGTGCAATGGCCTGTC-3') and A20 (5'-GACAAGGCATCTACACTGTGGCGG-3') and MC288. The specific bands were purified, and a second polymerase chain reaction was carried out with primers MC287 and MC288. Sequencing was carried out to verify the mutagenesis.

**Circular Dichroism Spectroscopy**—Circular dichroism spectra were recorded on a Jasco J-720 spectropolarimeter using a 1-mm quartz cuvette. NF200–248 (25 µM) was dissolved in a buffer containing Tris (10 mM) and 2-carboxyethylphosphine (TCEP, 250 µM) at pH 8.0. _ZnCl_ _2_ or _CoCl_ _2_ were made up in water at a concentration of 20 mM. CD spectra were recorded following the addition of _ZnCl_ _2_ or _CoCl_ _2_ to the sample at a final concentration of 10 mM. The pH was adjusted to 5.5, and a CD spectrum of the sample was recorded prior to the sedimentation equilibrium experiment to ensure that the peptide was folded; this spectrum was identical to that shown in Fig. 2A (dotted line). Sedimentation equilibrium data were collected in a double sector cell as absorbance versus radius scans (0.001 cm increments). Scans were collected at 4-h intervals and compared to ensure that the sample reached equilibrium. Analysis of the data was carried out using the NOLNIN software (31), and the final parameters were determined by a nonlinear least squares fit of the data to a model incorporating a single, nonassociating species. The goodness of fit was determined by examination of the residuals derived from the fit. The partial specific volume of NF200–248 (0.726 ml g⁻¹) was determined from the amino acid sequence (32), and the solvent density was taken to be 1.00 g ml⁻¹.

**NMR Spectroscopy**—One-dimensional 1H NMR spectra were recorded on a Bruker AMX-600 spectrometer at 15 °C and consisted of 64 scans collected as 8192 data points over a spectral width of 7200 Hz. The recycle time of the water resonance of the water-decoupled spectrum was around 0.6, at which time expression of the GST-NF fusion proteins was induced by the addition of isopropyl-β-D-thiogalactoside to a final concentration of 1 mM. After 3–5 h of incubation, the bacteria were harvested by centrifugation, washed twice with a buffer containing protease inhibitors, and then frozen at −80 °C. For the NF200–248 used in the biophysical studies, the following purification protocol was carried out. The cell pellet was thawed, sonicated (2 × 4–30 s bursts), and centrifuged (15,000 rpm, 4 °C, 30 min), and the soluble fraction was retained. This fraction was passed down 4 × 10-ml gelatin–thiogalmine-agarose columns. The columns were washed, and then the fusion protein was released from the columns by the addition of thrombin (100 units) to each column. The eluate was collected and lyophilized. It was then resuspended in water and subjected to reverse phase HPLC on a semi-preparative Vydac C18 column. The major absorbance peaks (average recovery, 75%) were collected, and samples were then identified using positive ion electrospray ionization mass spectrometry. The purity of the fusion proteins was verified by reverse phase HPLC. The purity of the NF200–248 fusion protein was verified by size exclusion chromatography on a Superase 6 column (Pharmacia Biotech) and by the presence of a single band on SDS-PAGE.

**GST Pull-down Experiments with Full-length GATA-1 and Peptides**—GST-NF fusion proteins and mutant derivatives were prepared as above except that proteins were not released from the GSH affinity column. Typically, 2 µl of GSH beads containing 1 µg of fusion protein were used in these experiments. Experiments using *in vitro* translated 35S-labeled GATA-1 were carried out as described previously except that 0.5% rather than 0.2% Nonidet P-40 was used in the buffers (23, 26). Experiments with recombinant GATA-1 were carried out as follows. 10 µl of crude bacterial extract containing recombinant GATA-1 was incubated at 4 °C with 2 µl of beads containing the GST-NF fusion protein. After 20 min, the beads were collected by centrifugation and washed five times with binding buffer containing Tris (50 mM), NaCl (50 mM), β-mercaptoethanol (1 mM), bovine serum albumin (0.5 mg ml⁻¹), Nonidet P-40 (0.5%), and ZnSO₄ (100 mM). The bound material was then released by boiling in SDS loading buffer and subjected to electrophoresis and Western blotting. The presence of retained full-length GATA-1 was detected using the rat monoclonal N6 antibody and a secondary alkaline phosphatase-conjugated goat anti-rat antibody. The presence of alkaline phosphatase was revealed using nitro blue tetrazolium/5-bromo-1-chloro-3-indolyl phosphate (Promega) according to the manufacturer's instructions.

**CD Spectrum**—The CD spectrum of NF200–248 was investigated by sedimentation equilibrium methods on a Beckman Optima XL-A analytical ultracentrifuge. Experiments were carried out at loading concentrations of 96, 32, and 9.6 µM, using an An-60 rotor equipped with cells of 42.0, 52.0, and 62.0 mm path length at 20 °C. NF200–248 samples were made up as solutions in 10 mM Tris, 2 mM TCEP, and 1.1 molar equivalents of _ZnCl_ _2_; the pH was held initially at 8.5 for the addition of the _ZnCl_ _2_; and then reduced to 5.3. A CD spectrum of the sample was recorded prior to the sedimentation equilibrium experiment to ensure that the peptide was folded; this spectrum was identical to that shown in Fig. 2A (dotted line). Sedimentation equilibrium data were collected in a double sector cell as absorbance versus radius scans (0.001 cm increments). Scans were collected at 4-h intervals and compared to ensure that the sample reached equilibrium. Analysis of the data was carried out using the NOLNIN software (31), and the final parameters were determined by a nonlinear least squares fit of the data to a model incorporating a single, nonassociating species. The goodness of fit was determined by examination of the residuals derived from the fit. The partial specific volume of NF200–248 (0.726 ml g⁻¹) was determined from the amino acid sequence (32), and the solvent density was taken to be 1.00 g ml⁻¹.
and subjected to autoradiography. The anti-GATA-1 antibody used was N6 (Santa Cruz Biotechnology, Inc., Santa Cruz, CA).

Transfection Studies—Transactivations were carried out as previously reported (15). NIH3T3 cells were transfected with 2 μg of reporter plasmids and increasing amounts of GATA-1 expression vector as shown. The reporter plasmids contained minimal promoters with either one (M1α) or six (M6α) GATA sites upstream of a β-globin TATA box and growth hormone reporter gene. The GATA-1 expression plasmids contained normal or mutant murine GATA-1 cDNA inserted between the HindIII and NotI sites of RcCMV (Invitrogen). Growth hormone levels were measured 48 h after transfection using the Allegro kit from the Nichols Institute.

RESULTS

The Folding of the N-finger Is Metal-dependent—A peptide containing amino acids 200–248 of the N-finger was initially chosen for characterization and is referred to as NF200–248 (for numbering, see Fig. 1). In the absence of metal, NF200–248 exhibits both CD and one-dimensional 1H NMR spectra (Fig. 2A, solid line, and Fig. 2B, upper part, respectively) characteristic of an unfolded peptide; viz. the CD spectrum displays a pronounced minimum below 200 nm, while the resonances in the 1H NMR spectrum are sharp and poorly dispersed. The addition of one molar equivalent of Zn2+ induces substantial changes in both cases, consistent with the formation of persistent secondary structure. The red shift of the minimum in the CD spectrum to 208 nm (Fig. 2A, dotted line) is diagnostic of the presence of α-helical secondary structure, and the wide chemical shift dispersion in the NMR spectrum (Fig. 2B, lower part) is suggestive of a folded protein. The addition of more than one molar equivalent of Zn2+ had no further effect on the CD spectrum (data not shown), establishing the stoichiometry of the interaction as 1:1. The general features of the CD spectrum are comparable with those observed previously for folded zinc finger peptides (see, for example, Refs. 33 and 34) and are indicative of a mixture of α and β secondary structure. Comparable spectra were observed for the constructs NF200–254 and NF200–243, consistent with those peptides folding in the same fashion.

The addition of Co2+ to NF200–248 induced a similar change in the CD spectrum (Fig. 2A, dashed line), indicating that the peptide folds similarly around the two metals. Consequently, changes observed in the ultraviolet (UV) and visible (vis) absorption spectra of NF200–248 following the addition of Co2+ were also used to probe both the nature and the geometry of the metal-peptide complex (note that the electronic configuration of Zn2+ precludes the observation of useful UV-vis spectra for this metal). The S → Co charge transfer bands observed at 310 and 370 nm in a UV-vis titration of NF200–248 with CoCl2 (Fig. 3A) are suggestive of cysteinate ligation (33, 34). In the visible region of the spectrum (Fig. 3B), intense d-d transition bands at 610, 700, and 740 nm (e.g. εmax ~ 600 M-1 cm-1) clearly show that the coordination mode is tetrahedral (35); i.e. because d-d transitions in tetrahedral complexes are formally allowed, molar extinction coefficients are typically greater than ~200 M-1 cm-1, while in octahedral complexes, for example, such transitions are formally forbidden, resulting in values for ε of ~10

m-1 cm-1 (36). Further, the wavelengths of the d–d transitions argue for a Cys4 coordination sphere. The presence of nitrogen ligands in Cys4His2 zinc fingers causes a substantial increase in the d–d energy gap compared with an all-sulfur coordination sphere (37), and in such complexes d–d bands are typically observed in the 600–650-nm region (see, for example, Refs. 33 and 34). Moreover, the shape of the binding curve (Fig. 3A, inset) suggests that the association constant for Co2+ binding is greater than ~10^8 M-1, which is substantially stronger than for interactions between Co2+ and Cys4His2 fingers (38). These conclusions are consistent with the configuration deduced in the solution structure of the CF of chicken GATA-1 (16), where the four homologous Cys residues are inferred from NMR data to coordinate the zinc in a tetrahedral arrangement.
The N-finger Can Interact with Full-length GATA-1—Having established that the NF is capable of folding in the anticipated fashion, we then focused on its ability to participate in protein-protein interactions and sought to define its role in the self-association of GATA-1. Previous experiments have focused on the GATA-1 CF and have demonstrated that it plays a significant role in the self-association of GATA-1 (26), but it has also been noted that the N-finger alone can mediate association with full-length GATA-1. However, the relative importance of the N-finger has not been studied in detail (26). We therefore carried out GST pull-down experiments to compare the ability of equal amounts of GST-NF200–254, GST-CF249–318, and GST-GATA-1 (full-length) to retain full-length in vitro translated 35S-labeled GATA-1. These three fusion proteins retained approximately equal amounts of GATA-1, while the control protein, GST alone, did not retain GATA-1 (Fig. 4A). Taken together with previous work on the CF, this result suggests that the NF also plays a significant role in the self-association of GATA-1.

Determination of a Minimal Region within the N-finger Domain That Is Involved in GATA-1 Self-association—In order to eliminate the possibility of a bridging event, where a separate mammalian protein present in reticulocyte lysates may interact with two GATA-1 molecules simultaneously, we further analyzed the binding of the N-finger to full-length GATA-1 using only recombinant proteins. Full-length recombinant GATA-1 was produced in bacteria and mixed with glutathione-agarose beads coated with GST-NF200–254. After extensive washing, the proteins retained on the beads were then subjected to electrophoresis and Western blotting to test for the presence of retained full-length GATA-1. As can be seen from Fig. 4B (lane 3), the GST-NF200–254 fusion was able to capture full-length GATA-1. In contrast, neither beads alone nor beads loaded with GST retained GATA-1 protein. This result confirms the previous result that the NF200–254 is able to associate with GATA-1 (Fig. 4A and Ref. 26) and indicates that it is a direct interaction.

The same GST pull-down assay was used to delineate the critical portion of the NF involved in the interaction. We first tested two deletion mutants: GST-NF200–48 and GST-NF200–243. As can be seen in Fig. 4B (lanes 4 and 5), the longer of the two constructs exhibited a substantially stronger interaction with full-length GATA-1; formation of a GATA-1-NF200–243 complex was only just detectable in the Western blot. This result suggests that either residues in the 244–254 region (see Fig. 1) directly contribute to GATA-1 self-association or that deletion of these residues has interfered with the proper folding of the domain. However, both CD and one-dimensional NMR experiments indicate that NF200–248 and NF200–243 are able to fold in the same manner as NF200–254 (data not shown). In addition, we have recently completed the NMR solution structure of NF200–243, demonstrating unequivocally that this subdomain is able to fold normally.3 We

3 J. P. Mackay, K. Kowalski, A. H. Fox, R. Czolij, G. F. King, and M. Crossley, unpublished results.
therefore conclude that residues 244–254 play a direct role in GATA-GATA associations.

We then turned our attention to the "core" zinc-binding portion of the NF domain. It has previously been demonstrated that mutation of the first of the four zinc-coordinating cysteines in the CF zinc-binding region does not prevent self-association of GATA-1 (26). We therefore tested whether an equivalent mutation in the NF (Cys204 → Ala) affected the interaction. As shown in Fig. 4B (lane 6), this mutation had no effect on the interaction. Likewise, the mutation Pro213 → Thr had no effect on the interaction (lane 7). Suspecting that the zinc-binding region of this NF200–254 construct might not be strictly required for GATA-1 self-association, we next constructed a smaller GST fusion protein (NF224–254), which lacked the entire N-terminal half of the NF domain and tested its ability to interact with full-length GATA-1. As shown in Fig. 4B (lane 8), this region was capable of mediating the interaction, albeit less efficiently than the entire finger (NF200–254). This tail region therefore plays a key role in GATA-1 self-association. The deletion constructs NF224–248 and NF224–243 were also tested. The former displayed a weaker, but still detectable, interaction with GATA-1, while no interaction could be detected in the case of the latter construct. This confirms the substantial role played by residues 244–254 and demonstrates that an intact zinc finger is not required in both halves of a GATA-1 dimer for a high affinity interaction to take place. It is also of note that a similar subdomain, identified as sufficient to sequester GATA-1 in a GST pull-down assay, lies at the C-terminal end of the CF (residues 278–318) (26). Both subdomains contain a conserved Asn-Arg-Pro-Leu sequence followed by a short stretch of basic amino acids, and we refer to these regions as the NF-tail and CF-tail, respectively (see Fig. 1).

The N-finger Domain Does Not Self-associate—The results above indicate that the NF is involved in GATA-1 self-association, but they do not reveal the region within full-length GATA-1 with which the NF makes contact. Previous work has shown that GST-NF200–254 was able to capture either full-length GATA-1 or a peptide containing residues 229–333 (i.e., the critical NF-tail subdomain together with the entire CF; see Fig. 1) (26). This result suggests that either the NF directly homodimerizes (NF-NF) or it contacts the CF (NF-CF). Thus, GATA-1 dimers may be held together by “parallel” NF-NF and CF-CF contacts, or by an “antiparallel” NF-CF arrangement, where the NF of one monomer interacts with the CF of the second and vice versa. In order to shed light on the arrangement, we first used sedimentation equilibrium methods to test for an NF-NF interaction.

The reverse phase HPLC-purified NF200–248 peptide was used for analytical ultracentrifugation. A CD spectrum of the sample was recorded prior to the sedimentation experiments, to ensure that the domain was properly folded. Concentration versus radial distance profiles (Fig. 5) were obtained at three different rotor speeds, and a global nonlinear least squares analysis revealed that the data are well fitted by an ideal, single species model to the data. The reverse phase HPLC-purified NF200–248 peptide was used for analytical ultracentrifugation. A CD spectrum of the sample was recorded prior to the sedimentation experiments, to ensure that the domain was properly folded. Concentration versus radial distance profiles (Fig. 5) were obtained at three different rotor speeds, and a global nonlinear least squares analysis revealed that the data are well fitted by an ideal, single species model to the data.

The analytical ultracentrifugation data demonstrate that GATA-1 self-association does not involve an NF–NF interaction. We therefore tested for an NF–CF interaction using GST pull-down assays. Thus, we assayed immobilized GST-CF249–318 for its ability to capture the NF200–248 peptide. As shown in Fig. 6 (lane 4), GST-CF249–318 was able to retain NF200–248, while control reactions showed that this peptide was not retained by either beads alone or GST-loaded beads (not shown). Importantly, an additional control lane showed that the shortened NF peptide NF200–243, which has a reduced ability to capture full-length GATA-1 (see above), was only marginally retained by GST-CF249–318 (lane 5). Furthermore, GST-NF200–248 was unable to sequester NF200–248 from solution (data not shown), in agreement with the sedimentation equilibrium data.

This method was also used to probe the dependence of GATA-1 self-association on the presence of intact zinc finger units. As mentioned above, it is known that specific Cys mutations in either the NF (Fig. 4, lane 6) or the CF (26) do not completely abolish GATA-GATA self-association, consistent with the view that the NF- and CF-tails are major determinants of the interaction. Thus, for instance, the interaction may primarily involve NF-tail to CF-tail contacts. Alternatively, the data in Fig. 4 are also consistent with the view that the interaction involves contacts between the NF-core and the CF-tail and/or between the CF-core and the NF-tail. In order to test whether core-finger to tail interactions are involved or alternatively whether it is exclusively the tails that mediate self-
association, we repeated the GST pull-down experiment in the absence of zinc. As shown in Fig. 6 (lanes 2 and 3), the absence of zinc did indeed prevent the interaction between GST-CF249–318 and the NF. This result suggests that a folded NF is involved in GATA-1 self-association, consistent with a model where the core finger regions from each monomer interact with the tail regions in the dimerization partner.

Mutation of both N- and C-finger Tail Subdomains Abrogates Self-association—While the above experiments assess the association of GATA-1 molecules in the absence of DNA, contacts between DNA-bound molecules of GATA-1 are of greater interest, since these interactions are most likely to be involved in establishing looped domains within chromatin. We therefore used the gel retardation assay to test whether the subdomain we had identified was essential for the self-association of DNA-bound GATA-1. As shown both in Fig. 7 and in previous studies (24, 26) where GATA-1 has been analyzed in vitro by gel retardation experiments, the majority of the protein is detected binding to DNA as monomers, while a small but consistent proportion self-associates to form dimers. Both the monomer and dimer band are indicated in Fig. 7, and the nonspecific band, which is not affected by an anti-GATA-1 monoclonal antibody, is marked by an asterisk.

In order to further characterize the role of the residues in the tail regions of the NF and CF, we sought a mutation that would disrupt the critical tail subdomain without compromising the ability of GATA-1 to bind DNA. Since the deletion of residues 244–248 severely impaired the ability of the NF to interact with full-length GATA-1 (see above), we first targeted these residues. In order to preserve the spacing between the fingers, we prepared a substitution mutant: Lys245-Lys246-Arg247 to Met-Val-Ile. When this protein was tested in the gel retardation assay (Fig. 7, lanes 3 and 4), it was found to bind DNA normally and to self-associate to form a homodimer, although dimer formation was slightly reduced (duplicate samples are shown in order to accentuate the subtle reduction in the intensity of the dimer band). The failure of this mutation to completely abrogate self-association suggested that the previously identified tail subdomain downstream of the C-finger (278–318) might still be functional (possibly interacting with the intact zinc-binding region of the NF). To test this possibility, we prepared a second mutation, which disrupted the basic stretch in CF-tail (Lys315–Lys316–Arg317 replaced by Ser); we also prepared a third protein, which was a double mutant combining both substitutions and thus disrupting both tail regions. As can be seen in lanes 5 and 6, mutation of the basic region in the CF-tail also reduced but did not abolish dimer formation. However, when both tail regions were mutated (lanes 7 and 8), dimer formation was more severely impaired. This result is consistent with the GST pull-down results, which indicate that both tails are capable of mediating the interaction.

Transactivation by GATA-1 Is Impaired by Mutations That Interfere with Self-association—We tested whether these mutations influenced the ability of GATA-1 to activate GATA site-dependent promoters in transfection studies. We compared the ability of the NF-tail, the CF-tail, and the double mutant to activate transcription from either a promoter containing a single GATA site (A) or of a similar reporter driven by a promoter containing six GATA sites (B). The amounts of GATA-1 expression plasmid are shown. Open circles, results obtained with normal GATA-1; closed squares, the effect of GATA-1 carrying the N-tail mutation; open squares, GATA-1 with the C-tail mutation; closed circles, GATA-1 with both N-tail and C-tail mutations.
DISCUSSION

Recently, considerable attention has been focused on interactions between transcription factors. In the case of GATA-family members, interest has centered on the double zinc finger domain, which, in addition to its recognized role in DNA binding, is now known to interact with a number of other proteins. The NFs of GATA family proteins are of particular interest as multifunctional protein domains. The NF of GATA-1, for instance, has been shown to stabilize DNA binding (15), to alter the specificity of binding (19, 21), and to mediate the formation of complexes with Fog, Sp1, EKLF, and GATA-1 itself (22, 23, 26). The NF of GATA-3 has been implicated in both transactivation and nuclear localization (39). The recent discovery of a splice variant of the liver/endothelium factor, GATA-5, which is missing the NF (40), further suggests that this finger plays a distinct role in GATA protein function. Similar differential splicing of the zinc finger domain of a Bombyx mori GATA factor to yield one- and two-finger proteins also suggests that its NF has some special function (41). Moreover, the functional importance of the NF domain has been directly demonstrated by specific disruption of this structure in vivo systems. Deletion of the NF of GATA-1 renders the transcription factor unable to drive erythroid maturation (17). Also, a number of well characterized Drosophila mutants, termed Pannier mutants, have been analyzed, and the causative mutations have been mapped to the region encoding the NF of a Drosophila GATA factor (42). Finally, a GATA-3 mutation that specifically disrupts the basic stretch in the critical NF tail of GATA-3 has been shown to create a dominant negative protein that interferes with the activity of normal GATA-3 in T cells (43).

As a first step toward understanding the various roles of this domain, we have expressed and purified the NF domain of murine GATA-1, analyzed its metal-dependent folding properties, and tested its role in mediating GATA–GATA interactions. Purified recombinant NF produced in bacteria is highly soluble and appears to fold readily by coordinating zinc, as judged by UV-vis, CD spectropolarimetry, and NMR spectroscopy. The purified domain coordinates zinc by means of four cysteine residues oriented in a tetrahedral array. The appearance of the CD spectrum is indicative of a mixture of α and β secondary structure, suggesting that, as predicted (16), the overall topology of the NF resembles that of the CF of GATA-1.

Consistent with previous results, we were able to show that the NF could interact with full-length GATA-1. Importantly, we showed that pure recombinant NF was capable of interacting with bacterially produced GATA-1. Thus, unlike the interaction between GATA-1 and Tal/SCL, which involves the red cell bridging proteins RBTN2/Lmo2 and Ldb1 (44), the GATA–GATA interaction is direct. We are currently using NMR spectroscopy to uncover the precise molecular details of these finger–finger interactions, but the results presented here provide important preliminary information regarding the general topology of the interaction. One critical portion of the NF lies immediately downstream of the zinc-binding region and consists of only 25 residues (Leu224–Ala244), with residues 244–254 clearly playing an important role. This subdomain contains an Asn-Arg-Pro-Leu motif followed by a number of basic residues. Our results suggest that this tail subdomain makes direct contacts with the core CF zinc-binding domain and that a similar “Asn-Arg-Pro-Leu + basic” subdomain downstream of the CF interacts with the zinc-binding core of the NF. Specific mutagenesis of the two subdomains revealed that both were involved in the dimerization of GATA-1, and only when both were disrupted was the self-association strongly inhibited. In this way, we propose that GATA-1 dimers could be held together by an antiparallel NF-CF, CF-NF arrangement, and only when both interfaces are impaired is self-association abrogated. In this model, the NF-tail of one GATA-1 molecule interacts with the CF zinc-binding core of a second GATA-1 molecule, while the CF-tail of the first molecule contacts the NF zinc-binding core of the second. The data provided here, together with previous work (26), suggest that either tail region is capable of mediating interaction with full-length GATA-1 even in the absence of its adjoining finger. While it is possible that the tail adopts a fully folded structure in isolation, it is perhaps more likely that it is rather flexible and that secondary structure is induced only when the tail contacts a second GATA molecule.

It is interesting to compare these results with previous work centered on the CF. It has been shown that, like the NF, the CF is also able to interact with GATA-1 (26). Unexpectedly, attempts to map the relevant subdomains within the CF revealed the presence of two distinct, overlapping domains. One subdomain, CF278–317, appears analogous to the tail subdomain we have identified in the NF in that it too lies immediately downstream of the zinc-binding region and contains an Asn-Arg-Pro-Leu motif followed by a short basic stretch. The second subdomain centers around the actual zinc-binding region of the CF. This result may be understood in light of our present model of an antiparallel association where the two subdomains of the CF interact with different regions of GATA-1; viz., the CF-tail may interact with the NF, while the CF contacts the NF-tail. Consistent with this, the disruption of both intermolecular interactions is required to abolish dimerization, while the mutation of a single contact region reduces the interaction affinity to a lesser degree.

The demonstration that the NF interacts directly with CF raises the question of whether it is able to make intra- as well as intermolecular contacts. In the experiments described here, intermolecular contacts must be involved, but intramolecular contacts may also play an important role in GATA-1 function and cannot be ruled out. It has recently been suggested that at double WGATAR sites in DNA the NF can adopt different orientations relative to the CF depending on the polarity of the two WGATAR motifs (21). It is thus possible that intramolecular interactions influence the positioning of the two fingers and play a significant role in determining the affinity of GATA-1 for this important subset of WGATAR sites. It appears from the gel shift experiments that only a small proportion of GATA-1 molecules are found as dimers, and it is possible that intramolecular contacts may predominate over intermolecular contacts in vivo at least in some circumstances. It will be interesting to investigate whether the proportion of GATA-1 molecules found in association varies at different binding sites or is influenced by post-translational modifications.

The exact biological role of GATA-1 self-association is not certain, but there are a number of appealing possibilities. The most simple effect of self-association is to increase the local concentration of GATA-1 and presumably therefore to increase its potency as a transcriptional activator. Self-association may not, indeed, be limited to the formation of dimers, and it is possible that, like Sp1, GATA proteins can aggregate to form higher order multimers and superactivate promoters (45–47). Indeed, confocal immunofluorescence studies of erythroid cells reveal that large assemblies of GATA-1 exist in the nucleus in vivo, although the function of these aggregates is not clear (27). Mutations that impaired the ability of GATA-1 to self-associate also reduced its ability to activate transcription, suggesting that proper assembly of a number of GATA-1 molecules at a promoter is required for full transactivation. It is notable that many erythroid promoters contain multiple GATA sites, and it
may be that GATA-1 self-association is involved in the ordered assembly of higher order complexes at these promoters. Additionally, self-association between molecules of GATA-1 bound at distant sites may mediate the formation of chromatin loops and the establishment of regions of active chromatin. It is of note that the communication between the bacterial NtrC enhancer and promoter is mediated by dimerization and subsequent multimerization of the NtrC protein (48). It is likely that looped domains in eukaryotic chromatin are also facilitated at least in part by the interaction of DNA-bound transcription factors. Finally, it is notable that numerous erythroid promoters and enhancers contain multiple GATA elements, and it may be that the self-association of GATA-1 plays a specific role in regulating this subset of genes.

Acknowledgments—We thank Stuart Orkin for helpful discussions, anti-GATA-1 antibodies, and expression plasmids; Menie Merika for assistance with analytical ultracentrifugation experiments.

REFERENCES

1. Weiss, M. J., and Orkin, S. H. (1985) Exp. Hematol. 23, 99–107
2. Orkin, S. H. (1995) Curr. Opin. Cell Biol. 7, 870–877
3. Crossley, M., and Orkin, S. H. (1993) Curr. Opin. Genet. Dev. 3, 232–237
4. Grroveld, F., Dillon, N., and Higgs, D. (1993) Baillieres Clin. Haematol. 6, 31–55
5. Engel, J. D. (1993) Trends Genet. 9, 304–309
6. Fujiwara, Y., Browne, C. P., Cumniff, K., Goff, S. C., and Orkin, S. H. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 12355–12358
7. Penney, L., Simon, M. C., Robertson, E., Klein, W. H., Tsai, S.-H., D’Agati, V., Orkin, S. H., and Costantini, F. (1991) Nature 349, 257–260
8. Pandolfi, P. P., Roth, M. E., Karis, A., Leonard, M. W., Dzierzk, E., Grosveld, F. G., Engel, J. D., and Lindenbaum, M. H. (1995) Nat. Genet. 11, 40–44
9. Ting, C. N., Olson, M. C., Barton, K. P., and Leiden, J. M. (1996) Nature 384, 474–478
10. Tsai, F.-Y., Keller, G., Kuo, F. C., Weiss, M. J., Chen, J., Rosenblatt, M., Alt, F. W., and Orkin, S. H. (1994) Nature 371, 221–226
11. Rehorn, K.-P., Thelen, H., Michelson, A. M., and Reuter, R. (1996) Development 122, 4023–4031
12. Tsai, S.-F., Martin, D. I. K., Zon, L. I., D’Andrea, A. D., Wong, G. G., and Orkin, S. H. (1989) Nature 339, 446–451
13. Tsai, S.-F., Strauss, E., and Orkin, S. H. (1991) Genes Dev. 5, 919–931
14. Evans, T., and Felsenfeld, G. (1989) Cell 58, 877–885
15. Martin, D. I. K., and Orkin, S. H. (1990) Genes Dev. 4, 1886–1898
16. Omichinski, J. G., Clare, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Apella, A., Stahl, S. J., and Grosveld, G. (1993) Science 261, 436–446
17. Weiss, M. J., Yu, C., and Orkin, S. H. (1997) Mol. Cell. Biol. 17, 1642–1651
18. Pedone, P. V., Omichinski, J. G., Nony, P., Trainor, C., Grosveld, G. M., and Felsenfeld, G. (1997) EMBO J. 16, 2874–2882
19. Wyhtrt, D. J., de Boer, E., and Grroveld, F. (1993) EMBO J. 12, 4993–5005
20. Yang, H.-Y., and Evans, T. (1992) Mol. Cell. Biol. 12, 4562–4570
21. Trainor, C. D., Omichinski, J. G., Vandergunn, T. L., Grosveld, G. M., Clare, G. M., and Felsenfeld, G. (1996) Mol. Cell. Biol. 16, 2238–2247
22. Tsang, A. P., Visvader, J. E., Turner, C. A., Fujiwara, Y., Yu, C., Weiss, M. J., Crossley, M., and Orkin, S. H. (1997) Cell 90, 109–119
23. Merika, M., and Orkin, S. H. (1995) Mol. Cell. Biol. 15, 2437–2447
24. Calligaris, R., Bottardi, S., Caglio, S., Apelleguia, I., and Santoro, C. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 11598–11602
25. Yang, H.-Y., and Evans, T. (1995) Mol. Cell. Biol. 15, 1353–1363
26. Crossley, M., Merika, M., and Orkin, S. H. (1995) Mol. Cell. Biol. 15, 2448–2456
27. Elefanty, A. G., Antoniou, M., Custodio, N., Carmona-Fonseca, M., and Grosveld, F. G. (1996) EMBO J. 15, 319–333
28. Haenlin, M., Cuda, Y., Blondevne, F., Heitzler, P., Lutz, Y., Simpson, P., and Ramain, P. (1997) Genes Dev. 11, 3096–3108
29. Osada, H., Grutz, G., Axelsson, H., Forster, A., and Rabbitts, T. H. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 9585–9589
30. Merika, M., and Orkin, S. H. (1993) Mol. Cell. Biol. 13, 3999–4010
31. Johnson, M. L., Correa, J. J., Yphantis, D. A., and Halvorson, H. R. (1981) Biophys. J. 36, 575–588
32. Perkins, S. J. (1986) Eur. J. Biochem. 157, 169–180
33. Burke, C. J., Sanyal, G., Bruner, M. W., Ryan, J. A., LaFemina, R. L., Robbins, H. L., Zett, A. S., Middaugh, C. R., and Cordingley, M. G. (1992) J. Biol. Chem. 267, 9639–9644
34. Frankel, A. D., Berg, J. M., and Pabo, C. O. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 4841–4845
35. Bertini, I., and Luchinat, C. (1984) Adv. Inorg. Biochem. 6, 71–111
36. Cotton, F. A., and Wilkinson, G. (1988) Advanced Inorganic Chemistry, 5th Ed., John Wiley & Sons, Inc., New York
37. Huheey, J. E. (1983) Inorganic Chemistry, 3rd Ed., Harper and Row, New York
38. Berg, J. M., and Merkle, D. L. (1989) J. Am. Chem. Soc. 111, 3759–3761
39. Yang, Z., Gu, L., Romeo, P. H., Bories, D., Motohashi, H., Yamamoto, M., and Engel, J. D. (1994) Mol. Cell. Biol. 14, 2201–2212
40. MacNeill, C., Ayres, B., Laverriere, A. C., and Burch, J. B. (1997) J. Biol. Chem. 272, 8396–8401
41. Drevet, J., Swevers, L., and Iatrou, K. (1995) J. Mol. Biol. 246, 43–53
42. Ramain, P., Heitzler, P., Haenlin, M., and Simpson, P. (1993) Development 119, 1277–1291
43. Smith, V. M., Lee, P. P., Szychowski, S., and Winoto, A. (1995) J. Biol. Chem. 270, 1518–1520
44. Wadman, I. A., Osada, H., Grutz, G. G., Agulnick, A. D., Westphal, H., Forster, A., and Rabbitts, T. H. (1997) EMBO J. 16, 3145–3157
45. Su, W., Jackson, S., Tjian, R., and Echols, H. (1991) Genes Dev. 5, 820–826
46. Pascal, E., and Tjian, R. (1991) Genes Dev. 5, 1646–1656
47. Mastrangelo, I. A., Courey, A. J., Wall, J. S., Jackson, S. P., and Hough, P. V. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 5670–5674
48. Wyman, C., Rombel, I., North, A. K., Busamamante, C., and Rustu, S. (1997) Science 273, 1658–1661