GATA-4 Activates Transcription Via Two Novel Domains That Are Conserved within the GATA-4/5/6 Subfamily*

(Received for publication, December 26, 1996, and in revised form, January 21, 1997)

Edward E. Morrisey, Hon S. Ip, Zhihua Tang, and Michael S. Parmacek‡

From the Department of Medicine, University of Chicago, Chicago, Illinois 60637

GATA-4 is one of the earliest developmental markers of the precardiac mesoderm, heart, and gut and has been shown to activate regulatory elements controlling transcription of genes encoding cardiac-specific proteins. To elucidate the molecular mechanisms underlying the transcriptional activity of the GATA-4 protein, structure-function analyses were performed. These analyses revealed that the C-terminal zinc finger and adjacent basic domain of GATA-4 is bifunctional, modulating both DNA-binding and nuclear localization activities. The N terminus of the protein encodes two independent transcriptional Activation Domains (amino acids 1–74 and amino acids 130–177). Amino acid residues were identified within each domain that are required for transcriptional activation. Finally, we have shown that regions of *Xenopus* GATA-5 and -6 corresponding to Activation Domains I and II, respectively, function as potent transcriptional activators. The identification and functional characterization of two evolutionarily conserved transcriptional Activation Domains within the GATA-4/5/6 subfamily suggests that each of these domains modulates critical functions in the transcriptional regulatory programs encoded by GATA-4, -5, and -6 during vertebrate development. As such these data provide novel insights into the molecular mechanisms that control development of the heart.

The GATA family of zinc finger transcription factors plays an important role in transducing nuclear events that modulate cell lineage differentiation during vertebrate development (for review see Refs. 1–3). Six GATA family members have been identified in vertebrate species, each of which is expressed in a developmentally regulated, lineage-restricted pattern (4–18). Within the GATA family, members of the GATA-4/5/6 subfamily of transcription factors are expressed in an overlapping pattern in the extra-embryonic endoderm, precardiac mesoderm, embryonic and adult heart, and gut epithelium (7, 12, 14, 19, 20). Functionally important GATA-binding sites have been identified in multiple cardiac-specific transcriptional regulatory regions, and overexpression of GATA-4 has been shown to transactivate these cardiac-specific transcriptional regulatory elements in noncardiac cells (21–23). In addition, expression of antisense GATA-4 transcripts in pluripotent P19 embryonal carcinoma cells blocks retinoic acid-inducible expression of genes encoding cardiac-specific myofibrillar proteins (24); whereas injection of GATA-4 mRNA into *Xenopus* oocytes results in the premature expression of genes encoding cardiac-specific contractile proteins (20). Moreover, GATA-4 has been implicated in regulating the differentiation of the extra-embryonic endoderm as cystic embryoid bodies derived from GATA-4−/− ES cells exhibit gross defects in formation of visceral and parietal endoderm (25). Thus, GATA-4 may play an important role in the transcriptional program(s) that control the differentiation of embryonic tissues (heart and possibly gut) as well as extra-embryonic tissues.

Despite their critical role in regulating lineage-specific gene expression, relatively little is currently understood about the molecular mechanisms that regulate the transcriptional activity of each GATA family member. Each vertebrate GATA factor contains two conserved type IV Cys-X2-Cys-X17-Cys-X2-Cys zinc fingers that recognize, and bind to, a related sequence motif (WGATAR) that is present in the transcriptional regulatory regions of multiple lineage-specific genes (8, 26–28). NMR analyses of the chicken GATA-1 C-terminal zinc finger and adjacent basic domain bound to DNA have revealed that the C-terminal zinc finger interacts with the major groove of DNA, and the basic domain lies in physical contact with the minor groove of the target sequence (29). Structure-function analyses of the mouse GATA-1 and human GATA-3 proteins, which control critical steps in erythroid and lymphoid development, respectively (30, 31), revealed single independent transcriptional Activation Domains within the N-terminal regions of both proteins (28, 32, 33). The GATA-1 Activation Domain (aa1–66) is an acidic, serine-rich domain with sequence homology to other acidic Activation Domains such as that present in the herpes virus transcriptional activating protein, VP16 (32–34).

Given its unique pattern of expression and highly divergent structure from the previously examined GATA-1/2/3 subfamily of proteins, it was of interest to examine the molecular mechanisms that control the transcriptional activity of the GATA-4 protein. In this report, structure-function analyses were employed to map the domains responsible for transcriptional activation of the murine GATA-4 protein. These studies revealed that the C-terminal zinc finger and basic domain is bifunctional.

*This work was supported in part by Public Health Service Grant 1RO1HL51145 and an American Heart Association Grant-In-Aid (to M. S. P.). 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.

†Established Investigator of the American Heart Association. To whom correspondence should be addressed: Dept. of Medicine, University of Chicago, MC 6088, 5841 S. Maryland Ave., Chicago, IL 60637. Tel.: 312-702-2679; Fax: 312-702-2681.

The abbreviations used are: aa, amino acid(s); PCR, polymerase chain reaction; cTnC, cardiac troponin C; DDB, DNA binding domain; GH, growth hormone; EMSA, electrophoretic mobility shift assay; Act I and Act II, Activation Domain I and II, respectively.

8515
modulating DNA-binding and nuclear localization activities. In addition, two N-terminal transcriptional Activation Domains, which are not conserved in the GATA-1/2/3 subfamily of transcription factors, were identified. Mutational analyses revealed specific amino acid residues within each domain that are required for transcriptional activity. Finally, we have shown that both Activation Domains I and II are functionally conserved within the recently identified *Xenopus* GATA-5 and -6 proteins (20). Elucidation of conserved transcriptional Activation Domains within the GATA-4/5/6 subfamily provides new insights into the molecular program regulating cardiac-specific gene expression and further clarifies the molecular evolution of the GATA family of transcription factors in vertebrates.

**EXPERIMENTAL PROCEDURES**

**Plasmids and Site-directed Mutagenesis**—The pMT2-GATA-4 expression plasmid encoding the murine GATA-4 protein has been described previously (4). Of note, DNA sequence analyses revealed several sequencing errors within the previously reported 5’-end of the murine GATA-4 cDNA (4). The corrected GATA-4 cDNA has an initiation codon located 18 base pairs 5’ of the previously reported initiation codon and a shift in reading frame. The corrected deduced amino acid sequence has recently been reported (14). The p-124tntGH reporter plasmid contains 125 base pairs 5’ of the murine GATA-4 gene and was subcloned into the pBluescript II SK+ expression vector (31) to generate the p-124tntGH plasmid (27). The pGAL4 plasmid was subcloned into the pBluescript II SK+ expression vector. The series of plasmids generated included the following: pGAL4/2A, pGAL4/Q3A, pGAL4/S4A, pGAL4/62A, pGAL4/W172A, pGAL4/A249A, pGAL4/Q149A, pGAL4/305A, pGAL4/155A, pGAL4/Y162A, and pGAL4/172A (the suffix of each name corresponds to the point mutation of the encoded murine GATA-4 Activation Domain).

**Cell Transfections, and Growth Hormone Assays**—NIH 3T3 and COS-7 cells were grown as described previously (22). NIH 3T3 cells were co-transfected with 25 μg of the indicated expression plasmid, 2.5 μg of the indicated GH reporter plasmid, and 1 μg of the pMSVβgal reference plasmid using Lipofectin reagent as described previously (22). Cells were harvested 48 h following transfection, and the medium from each plate was assayed for growth hormone using a commercially available radioimmunoassay kit (Nichols Institute). In addition, to assess transfection efficiencies β-galactosidase assays were performed on cell lysates as described previously (37). Each co-transfection experiment was repeated in duplicate at least three times. Results are expressed as normalized growth hormone ± S.E.

**Electrophoretic Mobility Shift Assays (EMSAs)**—Nuclear extracts were prepared 48 h following transient transfection of COS-7 or NIH 3T3 cells with 15 μg of the indicated transcriptional enhancers (27) contained in the indicated expression plasmids according to the procedure of Andrews and Fuller (38). EMSAs were performed as described previously (35). Equal amounts of protein were loaded in each lane. The following radiolabeled double-stranded synthetic oligonucleotides were utilized: CEF-1, 5’ CGACGCTGGAT-TACAGGGGAG C (the GATA-binding site is underlined); GAL4, 5’ GACCCGAATCTGCTCCCGGAG C. To confirm the specificity of each nuclear protein complex, specific and nonspecific unlabeled competitor oligonucleotides were included in each binding reaction as described previously (35).

**Immunohistochemistry**—To identify nuclear localization signal(s) within the murine GATA-4 protein, NIH 3T3 cells plated on glass coverslips were transiently transfected with 15 μg of the indicated expression plasmid. 48 h post-transfection, cells were fixed with 3.7% formaldehyde and incubated with a rabbit polyclonal antiserum raised against the murine GATA-4 protein (the gift of D. Wilson, Washington University). Previous experiments have demonstrated that this polyclonal antiserum recognizes the murine GATA-4 protein but not the GATA-1, -2, or -3 proteins (12). GATA-4 protein was detected with a fluorescein isothiocyanate-conjugated goat anti-rabbit IgG secondary antibody, and the slides were viewed with a Zeiss Axioshot microscope.

**RESULTS**

**Localization of a Bipartitional Domain That Mediates DNA-binding Activity and Encodes a Nuclear Localization Signal**—To define the minimal domain within the GATA-4 protein that confers DNA-binding activity in vitro, EMSAs were performed using nuclear extracts prepared from COS-7 cells transiently transfected with expression plasmids encoding GATA-4 deletion mutants and a radiolabeled oligonucleotide corresponding to the functionally important CEF-1/GATA motif from the murine cardiac troponin C (cTnC) transcriptional enhancer. Consistent with our previous reports (22, 35), recombinant full-length GATA-4 protein bound specifically to the radiolabeled CEF-1 oligonucleotide (Fig. 1B, lane 3). We have demonstrated previously that the lowest mobility complex (arrow) is super-shifted with GATA-4 plus tetracer and is abolished when unlabeled specific competitor CEF-1 oligonucleotide is added to the binding reactions (22). Both the pG4/199–323 deletion mutant (Fig. 1A), containing both zinc fingers and the adjacent C-terminal basic domain, and the pG4/251–323 deletion mutant (Fig. 1A), containing only the C-terminal zinc finger and basic domain, bound the radiolabeled CEF-1 oligonucleotide (Fig. 1B, lanes 4 and 6, arrows). In contrast, the pG4/199–302 GATA-4 deletion mutant (Fig. 1A), containing both zinc fingers, but lacking the C-terminal basic domain, failed to bind DNA (Fig. 1B, lane 5). Of note, Western blot analyses performed with GATA-4 specific antiserum and nuclear
nuclear extracts prepared from the transfected COS-7 cells confirmed that the 199–302 deletion mutant was produced (data not shown). These results demonstrate that the C-terminal zinc finger and the adjacent basic domain are necessary and sufficient to confer DNA-binding activity to the GATA-4 protein.

To identify the amino acid residues that target the murine GATA-4 protein to the nucleus, a series of expression plasmids encoding GATA-4 deletion mutants were co-transfected with the p-124cTnCGH reporter plasmid into NIH 3T3 cells. Consistent with our previous studies (22), co-expression of GATA-4 with the cTnC reporter plasmid resulted in an approximately 100-fold induction in GH activity as compared with co-transfection of the cTnC reporter plasmid with the negative control plasmid, pMT2 (Fig. 3A, row 1). Similarly, co-transfection with the pG4/Δ19 expression plasmid, which lacks the N-terminal 19 amino acids of GATA-4, transactivated the GH reporter plasmid approximately 100-fold (Fig. 3A, row 2). However, co-expression of deletion mutants lacking 36 or 49 N-terminal amino acids with the cTnC GH reporter plasmid resulted in 60 and 90% reductions, respectively, in GH activity (Fig. 3A, rows 3 and 4). Moreover, a GATA-4 deletion mutant lacking 76 N-terminal amino acids (pG4/Δ76) failed to activate the p-124GH reporter plasmid above basal levels (Fig. 3A, row 5). Deletion analysis of the C terminus of the murine GATA-4 protein revealed that both the pG4/Δ420C and pG4/Δ405C expression plasmids, encoding proteins harboring C-terminal truncation of 20 and 35 amino acids, respectively, transactivated the cTnC GH reporter plasmid to levels that were comparable with those obtained with the full-length GATA-4 expression plasmid (compare Fig. 3A, rows 1, 8 and 9). In contrast, co-transfection with the pG4/Δ390 protein, containing a deletion of 50 C-terminal amino acids, resulted in an 80% reduction in transcriptional activity (Fig. 3A, lane 7). These data suggested that both N- and C-terminal regions of the murine GATA-4 protein may be involved in regulating its transcriptional activity.

To more finely map regions of the murine GATA-4 protein that are required to transactivate the cardiac-specific cTnC promoter, a series of expression plasmids encoding GATA-4 in-frame deletion mutants were co-transfected with the p-124GH reporter plasmid into NIH 3T3 cells. Given the lack of activity demonstrated by the pG4/Δ76 plasmid (Fig. 3A, row 5), it was surprising that co-transfection with either the pG4/Δ320–35 or pG4/Δ33–60 plasmids encoding GATA-4 in-frame N-terminal deletion mutants induced comparable levels of GH activity to those obtained with wild-type GATA-4 (Fig. 3B, rows 1–3). These data suggest either that in the context of the native protein aa 1–19 are absolutely required for transcriptional activity or, alternatively, that alterations in the secondary structure of the Δ76 deletion mutant mask the function of other transcriptional Activation Domains located within the murine GATA-4 protein (see below). Co-transfection with
Identification of a nuclear localization signal in the GATA-4 protein. To identify the amino acid residues that mediate nuclear localization of the GATA-4 protein, the pMT2-GATA4 (A), pG4/199–324 (B), pG4/199–302 (C), and the pG4/Δ199–302 (D) were transfected into NIH 3T3 cells, and the subcellular location of the recombinant protein was determined immunocytochemically using a rabbit polyclonal GATA-4-specific antiserum as described under “Experimental Procedures.” GATA-4 is located exclusively in the nuclei of NIH 3T3 cells transfected with expression plasmids encoding either the full-length GATA-4 protein (A), the two zinc fingers and basic domain (B), and the C-terminal zinc finger and basic domain (C). In contrast, diffuse localization throughout the nucleus and cytoplasm was demonstrated in a GATA-4 deletion mutant lacking both zinc fingers (D). These data suggest that amino acid residues residing within the C-terminal zinc finger and basic domain are sufficient to restrict intracellular localization of GATA-4 to the nucleus.

Identification of Independent Transcriptional Activation Domains—To identify independent transcriptional Activation Domains within the murine GATA-4 protein, a series of expression plasmids encoding chimeric proteins containing the 147-amino acid yeast GAL4 DNA-binding domain (GAL4-DBD) fused in-frame to regions of the GATA-4 protein were co-transfected into NIH 3T3 cells with the pGAL4 GH reporter plasmid. Consistent with previous analyses of both the GATA-1 and -3 proteins (28, 32), an expression plasmid encoding the GAL4-DBD fused to the full-length GATA-4 protein (including its DBD) failed to transactivate the GAL4 GH reporter plasmid (data not shown). However, a 320-fold increase in GH activity was demonstrated when the pGAL4/1–204 expression plasmid, containing the N-terminal 204 amino acids of GATA-4 fused in-frame to the GAL4-DBD, was co-transfected with the GH reporter plasmid. This experiment confirmed the deletion analyses (see Fig. 3A and B) demonstrating that the N terminus of the murine GATA-4 protein contains one or more transcriptional Activation Domains (Fig. 4A, row 2). Similarly, high levels of GH activity were demonstrated when the pGAL4/1–93 (150-fold) and pGAL4/93–204 (220-fold) expression plasmids were co-transfected with the pGAL4GH reporter plasmid (Fig. 4A, rows 3 and 5). In contrast, expression plasmids encoding GAL4-DBD/GATA-4 fusion proteins that included GATA-4 aa 59–119 or aa 334–440 (the entire C terminus of the protein beyond the zinc fingers and basic domain) failed to increase GH activity above basal levels (Fig. 4A, rows 4 and 6). Of note, the lack of GH activity demonstrated with the pGAL4/334–440 expression plasmid was not anticipated as a 90% reduction in GH activity was demonstrated with the pGAL4Δ323C plasmid (Fig. 3A, row 6). This suggested that either the C terminus of the murine GATA-4 protein is necessary but is not sufficient for transcriptional activation within the context of the native GATA-4 protein or that the secondary structure of one or both of the C-terminal deletion mutants was altered from that of the native GATA-4 protein. Taken together, these data demonstrated that GATA-4 contains at least two independent transcriptional Activation Domains both of which are located in the N terminus of the protein. In addition, they suggested that while the C terminus of the protein (aa 334–440) may be required for transcriptional activity of the native protein, it does not contain an independent transcriptional Activation Domain.

To more precisely map the number of independent Activation Domains located within the N terminus of the murine GATA-4 protein, a second series of expression plasmids encoding GAL4-DBD/GATA-4 fusion proteins were transiently co-transfected into NIH 3T3 cells with the pGAL4GH reporter plasmid. Further analysis of the GATA-4 domain spanning aa 1–93 (see Fig.
revealed that co-expression of a fusion protein containing GATA-4 aa 32–61, which spans an evolutionarily conserved serine- and proline-rich region (see Fig. 5, upper panel), or a fusion protein spanning GATA-4 aa 19–85 failed to increase GH activity above basal levels (Fig. 4B, rows 2 and 3). In fact, deletion of the N-terminal 10 amino acids of GATA-4 completely abolished the ability of the GAL4-DBD/GATA-4 fusion protein to transactivate the GAL4GH reporter plasmid demonstrating that these 10 amino acid residues are required for functional activity of this N-terminal domain (Fig. 4B, row 4). In contrast, co-expression of a fusion protein containing GATA-4 aa 1–74 with the pGAL4GH reporter plasmid resulted in an approximately 250-fold increase in GH activity (Fig. 4B, row 5). However, further C-terminal truncation of GATA-4 to aa 1–70 resulted in an approximately 90% decrease (12-fold residual activity) in the ability of the GAL4-DBD/GATA-4 fusion protein to transactivate the GH reporter plasmid (Fig. 4B, row 6). Once again, EMSAs revealed that each of the chimeric proteins that failed to activate transcription retained the capacity to bind to the GAL4 motif (Fig. 4D and data not shown). These analyses demonstrate that amino acids 1–74, heretofore designated Activation Domain I, are necessary and sufficient to independently activate transcription in NIH 3T3 cells. Activation Domain I (aa 1–74) has a calculated pI of 7.16 (Fig. 5, upper panel). It contains eight proline residues, eight serine residues, four tyrosine residues, and three threonine residues. This domain has been evolutionarily conserved in GATA-4 proteins from the human, mouse, and frog (Fig. 5, upper panel). Moreover, high level amino acid sequence identity was identified between this region of the murine GATA-4 protein and regions within the recently identified chicken and Xenopus GATA-5 and -6 transcription factors (12, 20) (Fig. 5, upper panel). Within Activation Domain I, subdomains spanning amino acids 1–11, 35–47, and 52–57 demonstrate high
level sequence identity with regions located in the N terminus of the GATA-5 and -6 proteins, respectively (Fig. 5, upper panel, gray boxes). Interestingly, each of these regions contains a conserved tyrosine residue (Fig. 5, upper panel, arrowheads), as well as conserved serine and proline residues. Given the deletion analysis presented in Fig. 4B, which suggested that aa 70–74 are critical to the function of this domain, it is noteworthy that both aa residues 70 and 71 are serine residues. Overall, the proline-rich nature and neutral charge of this domain suggested that this region of the GATA-4 protein may be subclassified into the proline-rich family of transcriptional Activation Domains (34). Of note, protein sequence analysis software failed to detect even low level sequence homology between this domain of the GATA-4 protein and the GATA-1, -2, or -3 proteins.
peptides including the previously characterized transcriptional Activation Domains of GATA-1 and -3 (28, 32, 33) (data not shown).

Further analysis of the domain spanning aa 93–203 of the murine GATA-4 protein (see Fig. 4C, row 1) revealed that a fusion protein including GATA-4 aa 93–155 failed to increase GH activity above basal levels (Fig. 4C, row 2). Similarly, co-expression of a fusion protein containing GATA-4 aa 151–204 resulted in a 90% reduction (21-fold versus 225-fold induction) in GH activity compared with that obtained with the following co-transfection of the pGALA/93–203 plasmid (Fig., 4C, rows 1 and 3). In contrast, the pGALA/130–177 plasmid transactivated the GH reporter plasmid approximately 125-fold above levels obtained with the pGALA negative control plasmid (Fig. 4C, row 4). Once again, EMSAs confirmed that each of these fusion proteins retained the capacity to bind to the GAL4 motif (Fig. 4D and data not shown). Thus, amino acids 130–177, heretofore designated Activation Domain II, are sufficient to produce high level transcriptional activity in this heterologous system.

Activation Domain II (aa 130–177) has a calculated pI of 6.20 (Fig. 5, lower panel). It contains three proline, five serine, four tyrosine, and one glutamine residues. It is conserved across species demonstrating 43% and 38% sequence identity to the chicken and Xenopus GATA-4 proteins, respectively (10, 12) (Fig. 5, lower panel). High grade amino acid sequence identity was demonstrated between GATA-4 Activation Domain II and three regions (aa 145–150, 156–166, and 172–176) located at the N terminus of the Xenopus and chicken GATA-5 and -6 proteins (Fig. 5, lower panel, conserved subdomains are shown in gray). The conserved GREDDYG motif (aa 145–150) bears no identifiable sequence homology to Activation Domain I. In contrast, the conserved proline- and serine-rich GSYSSPYPPAYM motif (aa 156–166) bears low level sequence homology to the SSSPYYVPT subdomain (aa 34–42) identified within Activation Domain I. Once again, within each of these subdomains proline, serine, and tyrosine residues (Fig. 5, lower panel, arrowheads) are conserved throughout the GATA-4/5/6 subfamily of transcription factors. As with Activation Domain I, sequence homology could not be detected between GATA-4 Activation Domain II and any region of the GATA-1, -2, or -3 proteins (data not shown).

**Mutational Analyses of Transcriptional Activation Domains I and II**—The demonstration that specific amino acid residues within Activation Domains I and II were conserved across species and within each member of the GATA-4/5/6 subfamily of transcription factors suggested that some, or all, of these conserved amino acid residues might mediate critical functions required for transcriptional activity of the GATA-4, -5, and -6 proteins. To determine whether these conserved amino acid residues affected the function of Activation Domains I and II, a series of expression plasmids was generated encoding chimeric proteins in which the yeast GAL4-DBD was fused in-frame to the murine GATA-4 Activation Domains I or II, respectively, containing single point mutations. Each expression plasmid was transiently co-transfected with the pGAL4GH reporter plasmid into NIH 3T3 cells, and the observed GH activity was compared with that obtained with the pGAL4/1–74 (encoding Activation Domain I) or pGAL4/130–177 (encoding Activation Domain II) expression plasmids, respectively. Functional analysis of Activation Domain I revealed that multiple discrete point mutants within Activation Domain I including Q3A, P26A, H28A, Y38A, P40A, Y53A, and Q55A resulted in greater than 80% reduction in transcriptional activity compared with that obtained with the native murine GATA-4 Activation Domain I (Fig. 6A, lanes 3, 5, 6, and 9–12). In contrast, the Y2A and S4A mutations had relatively little effect on transcriptional activity (Fig. 6A, lanes 2 and 4), whereas mutation of aa P36A resulted in a 65% reduction in transcriptional activity (Fig. 6A, lane 8). EMSAs revealed that each of the encoded fusion proteins were expressed and retained the capacity to bind to the yeast GAL4-DBD probe (data not shown). Thus, most, but not all, of the evolutionarily conserved amino acid residues in Activation Domain I are required for transcriptional activation in this heterologous system.

Functional analysis of Activation Domain II revealed that only one point mutant (W172A) decreased transcriptional activity greater than 90% compared with the control Activation Domain II expression plasmid (Fig. 6B, lane 9). In addition, the Y158A and Y162A mutations decreased transcriptional activity by greater than 60% (Fig. 6B, lanes 6 and 7), whereas the E147A and Y149A mutants had a modest affect on transcriptional activity decreasing GH activity by 45 and 35%, respectively (Fig. 6B, lanes 2 and 4). Finally, the Q148A, S157A, and Y165A mutations failed to decrease transcriptional activity (Fig. 6B, lanes 3, 5, and 8). Once again, EMSAs revealed that
The data are expressed as relative GH activity with the pGAL4Act I (aa 1–74) plasmid, encoding Activation Domain I. The relative GH activity observed was compared with that obtained following co-transfection of the pGAL4/GH reporter plasmid in NIH 3T3 cells. As shown in Fig. 5, the pGAL4/GH reporter plasmid transactivated the GH reporter plasmid 165- and 184-fold, respectively (Fig. 7, lanes 3 and 4). Similarly, the pGAL4/XG6Act I (Xenopus GATA-5 aa 1–80) and pGAL4/XG5Act II (Xenopus GATA-5 aa 100–160) expression plasmids encoding chimeric proteins spanning the putative Activation Domains I and II of the Xenopus GATA-5 protein transactivated the GH reporter 220- and 66-fold, respectively (Fig. 7, lanes 5 and 6). This level of GH activity was comparable with that obtained following co-transfection of the pGAL4/1–204 GATA-4 expression plasmid in this experiment (Fig. 7, lane 2). Taken together, these data demonstrate that the N-terminal regions of the GATA-4, -5, and -6 proteins contain two independent, evolutionarily conserved transcriptional activation domains.

**DISCUSSION**

GATA-4 appears to play an important role in directing cell lineage-specific gene expression during development of the vertebrate heart (20–23, 25, 39). In this report, we have performed structure-function analysis of the murine GATA-4 protein and mapped the regions of the protein that control DNA binding, nuclear localization, and transcriptional activation. These studies revealed that the C-terminal zinc finger and adjacent basic domain of GATA-4 is bifunctional, modulating both DNA-binding and nuclear localization activities. In addition, the N-terminus of the GATA-4 protein contains two independent, transcriptional activation domains that share no identifiable amino acid sequence homology to the transcriptional activation domains identified previously in GATA-1 and -3, respectively (28, 32, 33). Mutational analyses defined specific amino acid residues within each of these domains that are required for transcriptional activation. Finally, we have shown that both Activation Domain I and II of the murine GATA-4 protein have been conserved across species and within each member of the recently identified GATA-4/5/6 subfamily of zinc finger transcription factors.

Current paradigms suggest that transcription factors are modular proteins with specific domains encoding distinct functions (28, 32, 33). The demonstration that the conserved C-terminal zinc finger and basic domain of GATA-4 is necessary and sufficient to confer both DNA-binding activity and nuclear localization has not been recognized previously. Conservation...
of a single domain encoding both DNA-binding and nuclear localization activities represents an efficient mechanism whereby a single modular domain may encode a specific function (DNA-binding) and target its subcellular location (the nucleus). The fact that each of these domains is conserved in yeast single finger GATA proteins demonstrates that this important bifunctional domain has been preserved throughout ancient evolution. In contrast, computer homology searches revealed that the two N-terminal transcriptional activation domains identified within the murine GATA-4 protein only share amino acid sequence homology with the recently identified GATA-5 and -6 proteins and not with other GATA factors (data not shown). As discussed below, these data strongly suggest that the N terminus of GATA-4 encodes a novel function that is shared only with the closely related GATA-5 and -6 proteins. Moreover, the structural organization of GATA-4, -5, and -6 suggests that this subfamily of zinc finger transcription factors evolved via duplication and subsequent diversification of a common ancestral gene that displayed a structure similar to GATA-4. In support of this theory preliminary characterization of the murine GATA-4, -5, and -6 genes has revealed common intron-exon boundaries.

Although a great deal is currently understood about the function of the zinc finger DNA-binding domains of each GATA family member (28, 29, 32, 33), relatively little is currently understood about the molecular mechanisms by which each GATA family member activates transcription. In fact, it remains unclear whether the transcriptional activation domains in each GATA factor function as interchangeable modules or, alternatively, direct the unique cell lineage-specific developmental program encoded by each vertebrate GATA factor. With respect to this question, two transcriptional activation domains were identified within the murine GATA-4 protein as assessed by transient transfection analyses of GAL4-DDB/GATA-4 fusion proteins (see Fig. 4). In most cases these data confirmed transient transfection analyses of GATA-4 deletion mutants. However, the precise function of the C terminus of GATA-4 remains unclear as deletion analyses of the native protein suggested that this region of the protein is required for transcriptional activation (Fig. 3A, lane 6), whereas analyses of GAL4/GATA-4 fusion proteins revealed that an 334–440 do not contain an independent transcriptional activation domain (Fig. 4A, lane 6). This apparent inconsistency underscores the inherent limitation of these analyses as the lack of transcriptional activity demonstrated by either the GAL4/GATA-4 fusion pro-

2 H. Ip, E. Morrisey, and M. Parmacek, unpublished observations.

| Fold Activation |
|-----------------|
| pGAL4 | 1 |
| pGAL4/1-204 | 178 ± 15 |
| pGAL4/XG5Act I | 165 ± 8 |
| pGAL4/XG5Act II | 184 ± 7 |
| pGAL4/XG6Act I | 221 ± 23 |
| pGAL4/XG6Act II | 66 ± 2 |
development. Many studies have demonstrated that transcription factors are post-translationally modified and activated (or suppressed) in response to specific intracellular signals. For example, phosphorylation of the MADS box transcription factors, MEF2C and SRF, has been shown to enhance the DNA-binding and transcriptional activity of each of these proteins (53–55). In this regard it is noteworthy that GATA-1 is phosphorylated in vivo and that a single serine residue is differentially phosphorylated during the dimethyl sulfoxide-induced differentiation of MEL cells (56). Thus, it is striking that both Activation Domain I and II of the murine GATA-4 protein are centered around evolutionarily identical tyrosine and serine residues (Fig. 3).

Activation Domain I contains eight serine and four tyrosine residues, and Activation Domain II contains five serine and four tyrosine residues. Moreover, mutations of some of these tyrosine residues (Tyr-38, Tyr-53, Tyr-158, and Tyr-162), or deletion of serine 171, severely decreased transcriptional activity of each domain (Figs. 2B and 4). These data are consistent with the hypothesis that phosphorylation of one or more of these residues in the GATA-4, -5, and -6 proteins may play an important role in regulating their respective activities. Thus, it will be important to carefully characterize the role of phosphorylation in regulating the DNA-binding and transcriptional activity of the GATA-4/5/6 subfamily of transcription factors. The determination of whether GATA-4 is post-translationally modified in vivo may provide novel insights into the signal transduction pathways that regulate cardiac development.

The observation that the recently identified GATA-4/5/6 subfamily of proteins shares both conserved DNA-binding and transcriptional activation domains raises the question of whether (unlike the GATA-1, -2, and -3 proteins) this subfamily of transcription factors mediates redundant functions in the vertebrate embryo. However, several recent reports argue strongly that this is not the case. For example, Soudais and co-workers (25) observed that GATA-4 → ES cells exhibit gross defects in the formation of visceral and parietal endoderm. Moreover, we have observed that while GATA-4 and -6 are developmentally co-expressed in the precardiac mesoderm and embryonic heart (14), GATA-5 has a temporally and spatially distinct pattern of expression from that of GATA-4 and -6 during embryonic cardiac development (19). In addition, we have reported that the genes encoding the murine GATA-4, -5, and -6 proteins are each expressed in a unique cell lineage-restricted pattern in tissues including the lung, bladder, and vascular smooth muscle cells (14, 19). Taken together, these data strongly suggest that while GATA-4 and -6 may subserve partially, or completely, redundant functions in the developing vertebrate heart, each member of the GATA-4/5/6 subfamily of transcription factors performs a unique function during vertebrate development. As such, elucidation of the molecular mechanisms by which Activation Domains I and II function in different cellular contexts should provide novel insights into the transcriptional regulatory programs mediated by each member of this recently identified subfamily of zinc finger transcription factors.

Acknowledgments—We thank Jeffrey M. Leiden and M. Celeste Simon for helpful discussions and suggestions. We thank Lisa Gottschalk for expert preparation of illustrations and Amy Murphy for expert secretarial assistance.

REFERENCES
1. Orkin, S. H. (1992) Blood 80, 575–581
2. Simon, M. C. (1995) Nat. Genet. 11, 9–11
3. Weiss, M. J., and Orkin, S. H. (1995) Exp. Hematol. 23, 99–107
4. Arecei, R. J., King, A. A., Simon, M. C., Orkin, S. H., and Wilson, D. B. (1993) Mol. Cell. Biol. 13, 2235–2246
5. Dorfman, D. M., Wilson, D. B., Bruns, G. A. P., and Orkin, S. H. (1992) J. Biol. Chem. 267, 1279–1285
6. Evans, T., Reitman, I., and Felsenfeld, G. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 5976–5980
7. Heikinheimo, M., Scandrett, J. M., and Wilson, D. B. (1994) Dev. Biol. 164, 361–373
8. Ho, I. C., Vorbees, P., Marin, N., Oakley, B. K., Tsai, S. F., Orkin, S. H., and Leiden, J. M. (1991) EMBO J. 10, 1187–1192
9. Jelin, V., Bories, D., Eley, E. F., La, S., Crotti, C., Chretien, S., Mattei, M. G., and Romeo, P. H. (1991) EMBO J. 10, 1809–1816
10. Kelley, C., Blumberg, H., Zon, L. I., and Evans, T. (1993) Development 118, 317–327
11. Ko, L. J., Yamamoto, M., Leonard, M. W., George, K. M., and Orkin, S. H. (1989) Nature 339, 446–451
12. Wilson, D. B., Dorfman, D. M., and Orkin, S. H. (1990) Mol. Cell. Biol. 10, 4845–4846
13. Yamamoto, M., Ko, L. J., Leonard, M. W., Beug, H., Orkin, S. H., and Engel, J. D. (1990) Genes Dev. 4, 1650–1662
14. Pevny, L., Simon, M. C., King, A. A., Simon, M. C., Leiden, J. M., and Wilson, D. B. (1993) Dev. Biol. 159, 201–214
15. Kelley, C., Blumberg, H., Zon, L. I., and Evans, T. (1993) Development 118, 317–327
16. Chen, C. Y., and Schwartz, R. J. (1995) J. Biol. Chem. 270, 1251–1263
17. Ko, L. J., and Engel, J. D. (1993) EMBO J. 12, 3099–3101
18. Tsai, S. F., Martin, D., and Orkin, S. (1990) J. Biol. Chem. 265, 8077–8080
19. Tsai, S. F., Martin, D. I. K., Zon, L. I., D’Andrea, A. D., Wong, G. G., and Orkin, S. H. (1993) Nature 361, 491–502
20. Kelley, C., Blumberg, H., Zon, L. I., and Evans, T. (1993) Development 118, 317–327
21. Chen, C. Y., and Schwartz, R. J. (1995) J. Biol. Chem. 270, 1251–1263
22. Tsai, S. F., Martin, D. I. K., Zon, L. I., D’Andrea, A. D., Wong, G. G., and Orkin, S. H. (1993) Nature 361, 491–502
23. Crossley, M., and Orkin, S. H. (1994) J. Biol. Chem. 269, 16589–16596