The C-terminal Subdomain Makes an Important Contribution to the DNA Binding Activity of the Pax-3 Paired Domain

(Received for publication, July 25, 1997)

Kyle J. Vogan‡§ and Philippe Gros‡¶

From the Department of Biochemistry, McGill University, Montreal, Quebec H3G 1Y6, Canada

The recognition of DNA targets by Pax-3 is achieved through the coordinate use of two distinct helix-turn-helix-based DNA-binding modules: a paired domain, composed of two structurally independent subdomains joined by a short linker, and a paired-type homeodomain. In mouse, the activity of the Pax-3 paired domain is modulated by an alternative splicing event in the paired domain linker region that generates isoforms (Q$^+$ and Q$^-$) with distinct C-terminal subdomain-mediated DNA-binding properties. In this study, we have used derivatives of a classical high affinity paired domain binding site (CD19-2/A) to derive an improved consensus recognition sequence for the Pax-3 C-terminal subdomain. This new consensus differs at six out of eight positions from the C-terminal subdomain recognition motif present in the parent CD19-2/A sequence, and includes a $\Delta$-TT-3' dinucleotide at base pairs 15 and 16 that promotes high affinity binding by both Pax-3 isoforms. However, with a less favorable guanine at position 15, only the Q$^-$ isoform retains high affinity binding to this sequence, suggesting that this alternative splicing event might serve to stabilize binding to suboptimal recognition sequences. Finally, mutagenic analysis of the linker demonstrates that both the sequence and the spacing in this region contribute to the enhanced DNA-binding properties of the Pax-3/Q$^-$ isoform. Altogether, our studies establish a clear role for the Pax-3 C-terminal subdomain in DNA recognition and, thus, provide insights into an important mechanism by which Pax proteins achieve distinct target specificities.

Pax-3 belongs to a family of transcription factors that regulate a variety of developmental processes in vertebrates and invertebrates (1). In the mouse, Pax-3 is expressed in a subset of neuroectodermal and mesodermal lineages (2), and is required for proper neural tube closure and neural crest cell migration (3), and for the development of particular skeletal muscle lineages (4, 5). To interact with target genes, Pax-3 contains a paired domain, a conserved 128-amino acid DNA-binding domain (6, 7) composed of two structurally distinct HTH$^1$-based subdomains joined by a short linker region, as revealed by crystallographic studies with a related family member, Drosophila Prd (8). In addition to the paired domain with its bipartite structure, Pax-3, like Prd, also contains a second conserved DNA-binding domain, a paired-type homeodomain, which folds into a classical HTH-based structure (9) and which binds cooperatively as a dimer to specific palindromic DNA sequences (10). To complement the distinct DNA binding activities of the isolated domains, the paired domain and homeodomain of Prd can also cooperate to recognize specific composite binding sites with high affinity (11). Consequently, the interaction of Pax-3 with DNA targets has the potential to be quite complex, with up to three distinct HTH-based modules contributing to target site selection.

Within Pax proteins, the N-terminal subdomain of the paired domain is highly conserved (12), and it interacts with a well defined 10–12-bp consensus sequence through a series of base-specific major groove and minor groove contacts (8). In contrast, the C-terminal subdomain in these proteins is more variable (12), and its precise involvement in DNA recognition has not been extensively characterized. In the case of Prd, the C-terminal subdomain does not contact the 15-bp recognition sequence used for crystallization (8), and appears to be dispensable both for DNA binding by Prd in vitro (11) and for Prd function in vivo (13). However, deletion studies have shown that this subdomain is required for the recognition of a class of naturally occurring target sequences by members of at least three Pax subfamilies (14), and in the case of human PAX6 (15, 16) and the Caenorhabditis elegans Pax-6 homologue mab-18 (17), point mutations in this subdomain lead to specific developmental disorders, suggesting that the C-terminal subdomain plays an important role in DNA recognition by some Pax proteins in vivo.

Thus far, selection experiments in vitro with the Pax-3 paired domain have yielded a consensus recognition sequence which, by analogy with Prd, does not include a C-terminal subdomain recognition motif (18, 19), raising the question of whether the C-terminal subdomain in Pax-3 contributes at all to sequence-specific DNA binding. In a preliminary study addressing this question, we recently described the DNA-binding properties of two alternatively spliced isoforms of Pax-3 (called Q$^+$ and Q$^-$), which differ by the presence or absence of a single glutamine residue in the paired domain linker region (20). This alternative splicing event was found to have no noticeable effect on the DNA-binding properties of the N-terminal subdomain of Pax-3; however, the novel isoform lacking the glutamine was found to exhibit a higher affinity for representative paired domain recognition sequences bearing consensus elements for both subdomains (20). This finding suggested that, in at least one naturally occurring Pax-3 isoform, the C-terminal subdomain is competent to bind DNA. However, the inability of either isoform of Pax-3 to bind to the Pax6-5a consensus (20), a C-terminal subdomain recognition sequence derived using an
alternatively spliced isoform of Pax6 defective in N-terminal subdomain binding (21), suggested that the C-terminal subdomain of Pax-3 may exhibit sequence preferences distinct from other Pax family members, and moreover, that such differences might be important for the discrimination of genomic targets by Pax-3 and other Pax proteins in vivo.

In this study, we have used derivatives of CD19-2/A, a classical paired domain recognition sequence bearing consensus elements for both subdomains (14), to define an improved consensus for the Pax-3 C-terminal subdomain, and to evaluate the contribution of this subdomain to the sequence-specific DNA binding activity of the Pax-3 paired domain. In addition, we have used these derivatives to further define differences in the DNA binding activities of the two alternatively spliced Pax-3 isoforms, and have used site-directed mutagenesis to probe the structural basis for these differences. Altogether, these studies establish a clear role for the Pax-3 C-terminal subdomain in DNA recognition, suggesting that this subdomain may exert an important influence on the selection of DNA targets by Pax-3 within the developing embryo.

**EXPERIMENTAL PROCEDURES**

**Expression of Recombinant Proteins in COS-7 Cells—**pMT2 expression plasmids carrying the full-length coding regions of human Pax6 and the two alternatively spliced isoforms of mouse Pax-3 have been described (20). Additional point mutations in the Pax-3 coding region were generated by PCR, as described below. For protein expression, plasmids (15 μg) were transiently transfected into COS-7 cells (1 × 10^6 cells/plate) by calcium phosphate co-precipitation (22). At 16 h after transfection, cells were treated for 1 min with 15% glycerol in HEPES-buffered saline (140 mM NaCl, 5 mM KCl, 0.75 mM NaH2PO4, 6 mM dextrose, and 50 mM HEPES, pH 7.05), incubated an additional 24 h with fresh medium, and then harvested in 1 ml of phosphate-buffered saline (154 mM NaCl, 2.7 mM KCl, 4.3 mM Na2HPO4, 1.4 mM KH2PO4, pH 7.3) with 15 μM sodium citrate. After washing once with phosphate-buffered saline, whole cell extracts were prepared by sonicating cells for 30 s in 250 μl of an ice-cold buffer containing 20 mM HEPES, pH 7.6, 150 mM NaCl, 0.5 mM dithiothreitol, 0.2 mM EDTA, 0.2 mM EGTA, 25 μM glyceraldehyde, and protease inhibitors. After centrifuging for 10 min at 4 °C to remove cellular debris, supernatants were isolated, frozen in small aliquots, and stored at −80 °C until use.

**Electrophoretic Mobility Shift Assays—**The parent CD19-2/A sequence (14) was synthesized as a double-stranded oligonucleotide (top strand: 5′-TGTTGTCAGCCTCAGTGCCC-3′; bottom strand: 5′-TATGGGGCAGCAGGCGTACCC-3′) with 5′ overhangs for subsequent labeling with [α-32P]dATP (DuPont, 3000 Ci/mmol) using the Klenow fragment of Escherichia coli DNA polymerase I (Pharmacia Biotech Inc.). Substituted derivatives of CD19-2/A were synthesized with complementary base changes in both strands of the double-stranded oligonucleotide, as described under “Results.” The other double-stranded oligonucleotides used in this study have been described (20, 23). Binding assays were typically performed using 0.5 μl of whole cell extract in a 20-μl reaction containing 10 mM Tris, pH 7.5, 50 mM KCl, 1 mM dithiothreitol, 2 mM spermidine, 2 mM bovine serum albumin, 10% glyceraldehyde, 100 fmol of radiolabeled oligonucleotide probe, 200 fmol of unlabeled probe, and 1 μg of either poly[dI-dC] (di-dC) or denatured salmon sperm DNA as a nonspecific competitor. After incubating for 30 min at room temperature, protein-DNA complexes were resolved from free probe by electrophoresis through a 4% nondenaturing polyacrylamide gel (29:1) buffered with 0.25 × TBE, using a constant voltage of 12 V/cm. Following electrophoresis, gels were dried and exposed to x-ray film for autoradiography or to a phosphorimaging plate for quantitation.

**Mutagenesis—**Amino acid substitutions of the glutamine residue at position 75 of the paired domain on the background of the Pax-3/Q isoform were introduced by PCR using a primer overlapping the start codon (P_Met: 5′-TGTGGATATCGATGACCACGCTG-3′) and a degenerate mutagenic primer overlapping the glutamine codon (Q75-mut: 5′-GGTGTTGTCACCNHCTTTGGTTG-3′, where H = A, C, or T). A second overlapping Pax-3 DNA fragment was amplified using primer pairs P1 (5′-GGTGCTCTTCAAGTCCTG-3′ and P2 (5′-GGTGCTCCTGAGCAATTGTTG-3′). These amplification products were purified on gel, mixed, denatured at 94 °C for 2 min, and then incubated at 72 °C for 5 min with Vent™ DNA polymerase in a reaction mixture without primers to generate hybrid DNA fragments. The hybrid products were then amplified using primers P3_Met and P3_J, and a 329-bp Smal fragment carrying the mutagenic segment was purified and used to replace the corresponding Smal fragment in the Pax-3 pMT2 expression plasmid. To screen for clones carrying the desired mutations, individual plasmids were isolated and sequenced with primer P3_J. In this manner, we identified seven independent clones in which the CAG codon corresponding to Gln-75 in the Pax-3/Q isoform had been replaced with CCC (Gly), GCC (Ala), GTG (Val), TTG (Leu), TCG (Ser), ACG (Thr), and AAC (Asn). A similar strategy was used to introduce a series of amino acid substitutions of the valine residue at position 75 of the paired domain on the background of the Pax-3/Q isoform. To generate the desired mutations, we used both a degenerate (V75-mut: 5′-CAGGCGTT-GTCNTCTTGGTTGTC-3′, where K = G or T) and two nondegenerate (V75R: 5′-GCCTTGTCCTTGGTTGTC-3′; V75R: 5′-GCTTGGCTCCTTGGTTGTC-3′) oligonucleotide primers, along with the other primers described above. This method allowed us to isolate clones in which the GCG codon corresponding to Val-75 in the Pax-3/Q isoform had been replaced with CCG (Pro), CTG (Ile), TCG (Ser), CAG (Glu), AAC (Asn), and AGG (Arg).

All amplification reactions were performed using Vent™ DNA polymerase (New England Biolabs) to improve the fidelity of polymerization. In addition, the 329-bp Smal fragment was fully sequenced in all the mutant clones to ensure that no additional mutations were introduced during any of the cloning steps.

**RESULTS**

**Evaluation of the DNA-binding Specificities of the Pax-3 and Pax6 C-terminal Subdomains—**Selection of optimal binding sites in vitro from a pool of random oligonucleotides (SELEX) is a method that is commonly used to derive consensus recognition sequences for sequence-specific DNA-binding proteins (24). In experiments of this type, the paired domain of Pax-3 selects a relatively short consensus (18, 19), similar to the consensus defined using the paired domain of Drosophila Prd (Ref. 11; Fig. 1). Based on the three-dimensional structure derived for the Prd/DNA co-crystal (8), this consensus, like the one defined for Prd, does not extend far enough to interact with the C-terminal subdomain. In contrast, the Pax6 paired domain selects a significantly longer consensus in vitro, with recognition elements for both subdomains (Ref. 25; Fig. 1). In addition, an alternatively spliced isoform of Pax6 with a 14-amino acid insertion in helix 3 of its N-terminal subdomain (Pax6-5a) selects a novel consensus composed of two tandemly repeated half-sites, each optimized to interact exclusively with the Pax6 C-terminal subdomain and with the extended linker that joins the two subdomains (Ref. 21; Fig. 1). Importantly, a comparison of the consensus recognition sequences defined using these alternatively spliced isoforms reveals that the Pax6-5a isoform exhibits sequence preferences at several posi-
nous Pax proteins expressed from the transfected plasmids.

We began by assessing the affinity of Pax6 for the parental CD19-2/A sequence and for the three singly substituted derivatives of this sequence at each position (Fig. 2, second column from right). Each lane represents a binding reaction performed using an equal amount of whole cell extract and radiolabeled probe, and thus, a direct comparison of the relative intensity of the protein-bound complexes provides an estimate of the affinity of Pax6 for each derivative. As shown, Pax6 is highly sensitive to substitutions at base pairs 16, 17, 18, and 20, exhibiting the following sequence preferences: position 16, T/C; position 17, G/A; position 18, C/A; and position 20, T/C. Importantly, these same base pairs were identified as critical in selection experiments with a Pax6 paired domain-glutathione S-transferase fusion protein (Ref. 25; Fig. 1); moreover, there is good agreement between the consensus sequence derived for the Pax6 C-terminal subdomain by SELEX and the consensus we define using the CD19-2/A derivatives, supporting the validity of the approach we have used. The one notable difference is seen at base pair 17, where we observe a marked reduction in binding when the guanine in the parental sequence is replaced with thymine, whereas guanine and thymine were found with roughly equal frequency at this position among the sequences selected by the Pax6 paired domain (25). Finally, weaker preferences for particular nucleotides are also observed at each of the other base pairs in the C-terminal subdomain recognition motif, suggesting that these positions exert a more subtle influence on Pax6 DNA recognition.

In turn, we employed the same panel of CD19-2/A derivatives to define a consensus for the Pax-3 C-terminal subdomain, using the alternatively spliced Q isoform of Pax-3 found in previous studies to exhibit enhanced C-terminal subdomain-mediated DNA binding activity (20). A preliminary comparison of the binding profiles of Pax6 and Pax-3 suggests that recognition of the CD19-2/A derivatives by the Pax-3 paired domain depends less critically on a contribution from the C-terminal subdomain, as fewer substitutions cause a significant reduction in binding (Fig. 2, far right column). Nevertheless, the affinity of Pax-3 for these sequences is sensitive to substitutions in this portion of the CD19-2/A recognition sequence, supporting the hypothesis that the Pax-3 C-terminal subdomain contributes in a sequence-specific fashion to DNA recognition. A second important finding is the similarity in the sequence preferences exhibited by Pax6 and Pax-3, most notably at positions 16, 18, and 19. This observation suggests that these proteins have conserved many of the same specific protein/DNA interactions, consistent with the substantial degree of sequence similarity between the C-terminal subdomains of these proteins (55% amino acid identity). At the same time, we also observe some clear differences in specificity between the two proteins. In particular, the 17T-, 20A-, and 20G-derivatives are all bound with moderate-to-high affinity by Pax-3, while Pax6 exhibits only weak binding to these sequences (Fig. 2), suggesting that sequence discrimination by the C-terminal subdomains of Pax-3 and Pax6 may also involve distinct protein/DNA contacts.

A more thorough analysis of the sequence preferences exhibited by Pax-3 indicates that in the parental CD19-2/A sequence, only positions 16 and 17 are optimized to interact with the Pax-3 C-terminal subdomain (Fig. 2); these base pairs together form a 5'-TG-3' dinucleotide that is also found in a number of naturally occurring Pax-5 target sequences, including H2A-2, 2, a site in the sea urchin histone H2A promoter region (14). However, in contrast to Pax-5, which shows significantly lower binding to a derivative of H2A-2.2 bearing a G→T substitution at position 17 (14), Pax-3 exhibits a similar affinity for both the

**DNA Recognition by the Pax-3 C-terminal Subdomain**

| position | parental sequence (CD19-2/A) | Pax-6 | Pax-3/Q- |
|----------|--------------------------------|-------|----------|
| 15       | G                              | A     | C        |
| 16       | T                              | G     | A        |
| 17       | G                              | C     | C        |
| 18       | C                              | C     | C        |
| 19       | C                              | A     | T        |
| 20       | A                              | T     | T        |
| 21       | G                              | C     | A        |
| 22       | A                              | T     | A        |

**Fig. 2. Derivation of a consensus recognition sequence for the Pax-3 C-terminal subdomain.** Electrophoretic mobility shift assays were performed with the Q isoform of Pax-3 (right column) and a panel of CD19-2/A derivatives bearing single nucleotide changes at each position along the C-terminal subdomain recognition motif (base pairs 15 to 22, according to the numbering system used in the published paired domain crystal structure). For comparison, a similar set of binding reactions with Pax6 are shown (second column from right). For each position analyzed, binding to the parental CD19-2/A sequence is shown along with binding to each of three singly substituted derivatives. The nucleotide occurring at each position in the parental CD19-2/A sequence is shown to the left for reference. For the purposes of comparison, each lane shows a binding reaction performed with an equal amount of protein extract and radiolabeled probe, and in all panels, only the shifted protein-DNA complexes are shown.
sequences shown. Comparing the affinities of Pax-3/Q to estimate of the affinity of Pax-3 for each of the DNA of the shifted complexes in Fig. 3 to derive an estimate of the relative affinities of concentration, generated Scatchard plots (data not shown), and used the slopes to assess separately the effects of favorable substitutions at base pair 18–22 of the recognition sequence (Fig. 2). Finally, the single most dramatic reduction in binding is caused by the substitution of the thymine at position 16 with guanine, implicating this base pair as a critical for DNA recognition by the Pax-3 C-terminal subdomain.

To summarize, we have tentatively defined an optimal recognition sequence for the Pax-3 C-terminal subdomain of 5′-TT(G/T)AATAT-3′, a consensus which differs at six out of eight positions from the parental CD19-2/A sequence. Moreover, the demonstration that the C-terminal subdomains of Pax-3 and Pax6 display similar but distinct sequence preferences suggests that the C-terminal subdomain may be an important specificity determinant for Pax proteins.

Measurement of the Affinity of Pax-3 for Optimized CD19-2/A Derivatives—Based on the results of Fig. 2, we designed a consensus oligonucleotide, P3-C-OPT (15T), to measure the cumulative effect of optimizing the 3′ portion of CD19-2/A for interaction with the Pax-3 C-terminal subdomain (Fig. 3A). In addition to making the most fully optimized consensus sequence, we also designed sequences CD19-2/A (15T) and P3-C-OPT (15G) to assess separately the effects of favorable substitutions at base pair 15 and at base pairs 18–22 of the C-terminal subdomain recognition sequence. Finally, we synthesized a doubly substituted derivative, CD19-2/A (15T16G), to evaluate the combined effect of a favorable G to T substitution at base pair 15 and an unfavorable T to G substitution at base pair 16.

To obtain a quantitative measure of the affinity of Pax-3 for these derivatives, we performed saturation binding analyses (Fig. 3B). For each derivative, we quantitated the amount of protein-bound and free probe at each oligonucleotide concentration, generated Scatchard plots (data not shown), and used the slopes to derive an estimate of the relative affinities of Pax-3 for each of the target DNA sequences (28). The intensity of the shifted complexes in Fig. 3B also provides a qualitative estimate of the affinity of Pax-3 for each of the DNA sequences shown. Comparing the affinities of Pax-3/Q− for the parent sequence and for each of the partially optimized derivatives, we observe that both sets of substitutions lead to a substantial increase in binding (Fig. 3B, left column). In the case of CD19-2/A (15T), the single G to T substitution leads to a 7-fold increase in binding relative to the parent sequence; similarly, replacing the sequence of base pairs 18–22 with an optimal C-terminal subdomain recognition motif in P3-C-OPT (15G) leads to a 4-fold increase in binding relative to the parent sequence. In combination, these substitutions generate an optimal sequence, P3-C-OPT (15T), that is bound with a 12-fold higher affinity relative to the parent sequence, demonstrating that the increases caused by the separate substitutions are roughly additive. Finally, we observe that the favorable G to T substitution at base pair 15 cannot compensate for the loss in binding caused by an unfavorable T to G substitution at base pair 16, again implicating the thymine at base pair 16 as a critical specificity determinant for the Pax-3 C-terminal subdomain.

We sought in turn to address whether any of the substitutions in the C-terminal subdomain recognition motif could increase the affinity of the Q+ isoform for the CD19-2/A target sequence (Fig. 3B, right column). As shown previously (20), Pax-3/Q− binds to the parental CD19-2/A sequence with approximately a 5-fold lower affinity than the Pax-3/Q+ isoform. Strikingly, the single G to T substitution at base pair 15 leads to a large increase in binding; the affinity of Pax-3/Q− for this sequence is roughly 30-fold higher than its affinity for the parent sequence. In contrast, optimizing base pairs 18–22 leads to a more modest increase in binding by the Q+ isoform, raising the affinity of the Q+ isoform to a level similar to the affinity of the Q− isoform for the parent sequence, and in combination with the thymine at base pair 15, these changes to the 3′ portion of the recognition sequence do not significantly improve binding relative to the singly substituted 15T-derivative. Finally, as was the case with the Q− isoform, the enhanced binding caused by the G to T substitution at base pair 15 is sensitive to an unfavorable T to G substitution at base pair 16, demonstrating the critical importance of these base pairs for sequence recognition by both isoforms.
To summarize, these affinity measurements identify a 5'-TT-3' dinucleotide at base pairs 15 and 16 as critical for high affinity binding by both the Q and Q isoforms of Pax-3. In addition, favorable substitutions at more 3' positions further increase the affinity of Pax-3 for DNA; this is particularly evident for the Q isoform, which shows a significantly higher affinity than the Q isoform for derivatives bearing a suboptimal guanine at position 15. Finally, the ability to produce such dramatic changes in affinity solely by changing the sequence 3' of the N-terminal subdomain recognition motif demonstrates that the C-terminal subdomain makes a significant contribution to the DNA binding activity of the Pax-3 paired domain.

Site-directed Mutagenesis of the Pax-3 Paired Domain Linker Region—Having defined recognition sequences optimized to interact with the Pax-3 C-terminal subdomain, we proceeded to use site-directed mutagenesis of the linker region to probe the structural basis for the differences in the DNA-binding properties of the two Pax-3 isoforms. The additional glutamine residue at position 75 resides in a portion of the linker that is not well ordered in the published paired domain crystal structure (8), between the portion of the linker that contacts base pairs 12–14 in the minor groove (residues 69–71) and the first helix of the C-terminal subdomain (residues 79–88). In the absence of detailed information about the structure in this region, we reasoned that the lower affinity of the Q isoform for some sequences might be due to a disruption in spacing within the linker, or might be due to specific local perturbations in structure caused by the presence of the glutamine side chain. We addressed these two possibilities by first replacing the additional glutamine residue at position 75 in the Pax-3/Q isoform with a variety of amino acids of different size, charge, and hydrophobicity (Fig. 4). After expressing each mutant by transient transfection in COS-7 cells, the level of recombinant protein in each extract was normalized by performing saturation binding analyses, as described in Fig. 3, with an oligonucleotide probe (Nf3) isolated by in vitro selection using the Pax-3/Q paired domain (23). The resulting Scatchard plots (data not shown) allowed us to estimate both the relative affinity of each mutant for Nf3 (from the slope of the curve) and the number of protein binding sites in each extract (from the intercept of the curve with the x axis). By normalizing protein levels in this fashion, we were able to perform the assays shown in Fig. 4 using an equal amount of wild-type or mutant Pax-3 in each binding reaction.

As shown in the top panel, both of the wild-type isoforms and all of the mutants bind with a similar affinity to Nf3, an observation confirmed by the slopes of the Scatchard plots used to normalize protein levels (data not shown). Similarly, as shown in the bottom panel, the wild-type and mutant forms of Pax-3 all bind with a similar affinity to the optimized CD19-2/A derivative P3-C-OPT (15T), suggesting that the favorable interactions with the 5'-TT-3' dinucleotide at base pairs 15 and 16 are not affected by substitutions at position 75. However, with probes that discriminate between the two naturally occurring Pax-3 isoforms (middle panels), we observe that none of the substitutions of the extra glutamine residue at position 75 restore binding to the level of the Q isoform, including substitution with valine, the amino acid present at position 75 in the absence of the additional glutamine residue. This finding demonstrates that proper spacing in this portion of the linker region is critical for high affinity binding by the C-terminal subdomain, independent of the nature of the side chain at position 75.

To further evaluate the structural basis for the enhanced binding of the Q isoform, we also tested the effect of replacing the valine at position 75 in the Pax-3/Q isoform with a number of different amino acids (Fig. 5). Among different members of the Pax family, a valine is found at this position in all paired...
domains described to date, with the exception of Prd, where an isoleucine is found (29), and Pax-4, where a leucine is found (12). This high degree of conservation suggests that this side chain may be play an important structural role in Pax proteins, though the conformation that this portion of the paired domain adopts when the C-terminal subdomain is bound to DNA is still unknown (8). As shown in the upper and lower panels in Fig. 5, both naturally occurring isoforms of Pax-3 and each of the mutant forms of Pax-3 bind with a similar affinity to N35' and to the optimized CD19-2/A derivative P3-C-OPT (15T), again demonstrating that changes in this portion of the linker do not affect binding to these high affinity Pax-3 recognition sequences. However, the introduction of different side chains at this position does affect the ability of these mutants to bind to recognition sequences that distinguish between the two naturally occurring Pax-3 isoforms (Fig. 5, middle panels). In particular, substitution of the valine at position 75 with either proline, serine, or glutamine reduces binding to a level similar to the affinity of Pax-3/Q isoform for these sequences; in addition, substitutions with either isoleucine or arginine also lead to a moderate reduction in binding. Among the substitutions tested, only the Val → Lys mutant retains high affinity binding to these sequences. These results indicate that in addition to spacing, the nature of the side chain at position 75 is important for maintaining a structure that can interact optimally with the portion of the recognition sequence 3′ to base pairs 15 and 16. Therefore, these findings support the possibility that the enhanced C-terminal subdomain-mediated DNA binding activity of the Pax-3/Q isoform might lead to the recognition of a distinct set of target genes by this alternatively spliced isoform in vivo.

### DISCUSSION

The ability of tissue-specific transcription factors to recognize specific DNA sequences within the regulatory regions of target genes is an important means of controlling gene expression. Faced with the enormous complexity of the genetic material present in living cells, DNA-binding proteins rely on a number of distinct mechanisms to achieve target specificity in vitro (30). Among these is the use of complex DNA-binding domains, such as POU domains and zinc fingers, that achieve specificity through the coordinate use of multiple, structurally independent DNA-binding subdomains (30). The bipartite paired domain, which is composed of two distinct HTH-based subdomains joined by a short linker (8), is a well conserved DNA-binding structure present in a family of transcription factors controlling diverse developmental processes in a wide variety of vertebrate and invertebrate species (1). The two helical subdomains and the extended linker region allow a single paired domain to interact with up to 20 base pairs of DNA, and together provide a large number of base-specific major groove and minor groove contacts that contribute to the recognition of specific DNA sequences as potential targets for regulation (6, 8, 14). However, with nine distinct Pax genes in mammals and seven in Drosophila, understanding how individual paired domains achieve distinct target specificities is an important step toward identifying the downstream targets of these transcription factors during development.

While the binding specificity of the N-terminal subdomain has been defined for several members of the Pax family (8, 11, 18, 19, 25), the role of the C-terminal subdomain in DNA recognition has not been characterized as extensively. Significantly, the selection of optimal binding sites with isolated paired domains in DNA recognition, including the Pax-3 paired domain, has often failed to define a consensus that extends far enough to include a recognition motif for the C-terminal subdomain (8, 11, 18, 19, 25). However, many naturally occurring paired domain recognition sequences identified within candidate target promoters are longer than the consensus sequences defined in vitro, and require a contribution from both subdomains for recognition (14). A clear example of this is seen with Pax-2, where selection from a pool of random oligonucleotides defined a 13-bp consensus lacking a C-terminal subdomain recognition motif (25), whereas the purification of Pax-2-bound complexes from native chromatin defined a 23-bp consensus with recognition motifs for both subdomains (31). Therefore, in some instances, the role of the C-terminal subdomain in DNA recognition by Pax proteins may have been masked by the tendency of the N-terminal subdomain to dominate over the C-terminal subdomain in standard in vitro selection protocols.

The recent discovery of an alternatively spliced isoform of Pax-3 with a higher affinity for full-length paired domain recognition sequences provided the first evidence that the C-terminal subdomain might influence the DNA-binding properties of the Pax-3 paired domain (20). To follow up these observations, we have generated derivatives of CD19-2/A, a well studied paired domain recognition bearing consensus binding motifs for both subdomains (14), to further evaluate the role of the Pax-3 C-terminal subdomain in sequence discrimination. Since structural studies with Drosophila Prd had shown that the N-terminal subdomain makes contacts with base pairs 4 through 14 of the recognition sequence (8), and since the Pax-3 N-terminal subdomain had been shown to select a consensus very similar to Prd in vitro (18, 19), we tested the importance of sequences 3′ to this 10–12-bp N-terminal subdomain recognition motif for DNA recognition by Pax-3. Significantly, we find that the affinity of Pax-3 for derivatives of CD19-2/A is sensitive to single base substitutions in the C-terminal subdomain recognition motif. Moreover, a comparison between CD19-2/A (16G), the lowest affinity site identified in these studies (Fig. 2), and P3-C-OPT (15T), a sequence optimized to interact with the Pax-3 C-terminal subdomain (Fig. 3), reveals that the affinity of Pax-3 for these two sequences differs by nearly two orders of magnitude. Since the N-terminal subdomain recognition motif is identical between these two sequences, this result demonstrates that the C-terminal subdomain makes a significant contribution to the DNA binding activity of the Pax-3 paired domain.

In addition to establishing a clear DNA-binding function for the Pax-3 C-terminal subdomain, these studies also provide some interesting insights into the role of the C-terminal subdomain in contributing to paired domain DNA-binding specificity. As has been noted previously (14), the high degree of conservation in the N-terminal subdomain among paired domains leads to the recognition of a very similar core sequence by divergent family members, raising the question of how these proteins are able to achieve distinct DNA-binding specificities in vitro. Significantly, the C-terminal subdomain in these proteins shows a much higher degree of sequence diversity, raising the possibility that this more divergent DNA-binding structure might be important means of discriminating between DNA targets (14). Indeed, using our panel of singly substituted CD19-2/A derivatives, we do observe clear differences in specificity between the C-terminal subdomains of Pax-3 and Pax6, lending support to this hypothesis. The panel of CD19-2/A derivatives described herein may therefore serve as a useful tool for defining the distinct sequence preferences conferred by the C-terminal subdomain in other Pax family members.

A comparison of the behavior of the naturally occurring isoforms of Pax-3 on the CD19-2/A derivatives has also provided us with a better understanding of the effect of this alternative splicing event on Pax-3 DNA recognition. In particular, we find that the presence of a 5′-TT-3′ dinucleotide at base
pairs 15 and 16 significantly enhances DNA recognition by both isoforms. While the structural basis for these distinct sequence preferences is not known, we have shown that the recognition of these two base pairs is not affected by the presence of an additional residue at position 75 of the paired domain linker, or by the nature of the amino acid side chain at this position. We note also that in the paired domain crystal structure, the N-terminal portion of the linker is inserted into the minor groove, allowing residues 69–71 to interact with base pairs 12–14 of the recognition sequence (8). In the light of our observations, it is possible that residues 72–74 may extend these minor groove contacts and interact with base pairs 15 and 16, contributing to the specificity observed at these positions.

In contrast to the similar affinity of the two alternatively spliced isoforms of Pax-3 for derivatives of CD19-2/A bearing a 5′-TT-3′ dinucleotide at base pairs 15 and 16, the Q isoform of Pax-3 demonstrates a significantly higher affinity for sequences bearing a suboptimal guanine at position 15. In this respect, it is interesting to note that many natural targets of Pax-3 function in the developing embryo.

However, further studies will be needed to clarify the biological relevance of this alternative splicing event with respect to Pax-3 function in the developing embryo.

Despite significant advances in our understanding of the factors governing Pax-3/DNA interactions in vitro, little progress has been made thus far in the identification of actual targets of Pax-3 regulation in vivo. However, the evidence for a significant role for the C-terminal subdomain in DNA recognition, together with the previously established DNA binding activities of the homeodomain and N-terminal subdomain, suggests that the recognition of some genomic targets by Pax-3 may involve the coordinate use of all three HTH-based DNA-binding modules. Indeed, functional interdependence between the paired domain and homeodomain has been demonstrated recently for both Pax-3 (34–36) and Prd (11, 13, 37). Moreover, there is evidence that cooperative interactions with other homeodomain-containing proteins (10) or with members of the Ets family (38) may also contribute to the recognition of some genomic targets by Pax proteins, suggesting an additional mechanism influencing target selection by Pax-3 in vivo. Ultimately, a deeper knowledge of these and other mechanisms by which Pax-3 achieves DNA-binding specificity will be an important step toward defining the genetic pathways functioning downstream of Pax-3, and more broadly, will help us to understand how different Pax proteins are able to achieve unique target specificities in the developing embryo.

REFERENCES

1. Noll, M. (1993) Curr. Opin. Genet. Dev. 3, 595–605
2. Goulding, M. D., Chalepakis, G., Deutsch, U., Erselius, J. R., and Gruss, P. (1991) EMBO J. 10, 1125–1127
3. Epstein, D. J., Vekemans, M., and Gros, P. (1991) Cell 67, 767–774
4. Boher, E., Franz, T., Arnold, H.-H., Gruss, P., and Tremblay, P. (1994) Development 120, 663–682
5. Tajbakhsh, S., Rocancourt, D., Cossu, G., and Buckingham, M. (1997) Cell 89, 127–138
6. Chalepakis, G., Fritsch, R., Fickenscher, H., Deutsch, U., Goulding, M., and Gruss, P. (1991) Cell 66, 873–884
7. Treisman, J., Harris, E., and Desplan, C. (1991) Genes Dev. 5, 594–604
8. Xu, W., Roulid, M. A., Jun, S., Desplan, C., and Pabo, C. O. (1995) Cell 80, 639–650
9. Wilson, D. S., Guenther, B., Desplan, C., and Kuriyan, J. (1995) Cell 80, 709–719
10. Wilson, D. S., Sheng, G., Lecuit, T., Dostatni, N., and Desplan, C. (1993) Genes Dev. 7, 2120–2134
11. Jun, S., and Desplan, C. (1996) Development 122, 2639–2650
12. Walther, C., Guenet, J.-L., Simon, D., Deutsch, U., Jostes, B., Goulding, M. D., Plachov, D., Balling, R., and Gruss, P. (1991) Genomics 11, 424–434
13. Bortztein, C., Fasano, L., Jun, S., Wang, S., Sheng, G., and Desplan, C. (1996) Development 122, 2673–2685
14. Czerny, T., Schaffner, G., and Busslinger, M. (1993) Genes Dev. 7, 2048–2061
15. Azuma, N., Nishina, S., Yanagisawa, H., Okuyama, T., and Yamada, M. (1996) Nature Genet. 13, 141–142
16. Tang, H. K., Chao, L.-Y., and Saunders, G. F. (1997) Hum. Mol. Genet. 6, 181–186
17. Chisholm, A. D., and Horvitz, H. R. (1995) Nature 377, 52–55
18. Chalepakis, G., and Gruss, P. (1995) Gene (Amst.) 162, 267–270
19. Epstein, J. A., Shapiro, D. N., Cheng, J., Lam, P. Y. P., and Maas, R. L. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 4213–4218
20. Vogan, K. J., Underhill, D. A., and Gros, P. (1996) Mol. Cell. Biol. 16, 6677–6686
21. Epstein, J. A., Glaser, T., Cai, J., Jepeal, L., Walton, D. S., and Maas, R. L. (1994) Genes Dev. 17, 2022–2034
22. Gorman, C., Padmanabhan, R., and Howard, B. H. (1983) Science 219, 551–553
23. Epstein, J. A., Lam, P., Jepeal, L., Maas, R. L., and Shapiro, D. N. (1995) J. Biol. Chem. 270, 11719–11722
24. Pollock, R., and Treisman, R. (1990) Nucleic Acids Res. 18, 6197–6204
25. Epstein, J., Cai, J., Glaser, T., Jepeal, L., and Maas, R. (1994) J. Biol. Chem. 269, 8355–8361
26. Koznik, Z., Wang, S., Dorrler, P., Adams, B., and Busslinger, M. (1992) Mol. Cell. Biol. 12, 2662–2672
27. Adams, B., Dorrler, P., Aguzzi, A., Koznik, Z., Urbének, M., Maurer-Fogy, L., and Busslinger, M. (1992) Genes Dev. 6, 1589–1607
28. Calzone, F. J., Thézé, N., Thiébaut, P., Hill, R. L., Britten, R. J., and Davidson, E. H. (1989) Genes Dev. 2, 1074–1088
29. Frigerio, G., Burri, M., Bopp, D., Baumgartner, S., and Noll, M. (1986) Cell 47, 735–746
30. Nelson, H. C. M. (1995) Curr. Opin. Genet. Dev. 5, 180–189
31. Phelps, D. E., and Dressler, G. R. (1996) J. Biol. Chem. 271, 7978–7985
32. Zannini, M., Francis-Lang, H., Plachov, D., and Di Lauro, R. (1992) Mol. Cell. Biol. 12, 4230–4241
33. Holst, B., Wang, Y., Jones, F. S., and Edelman, G. M. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 1465–1470
34. Underhill, D. A., Vogan, K. J., and Gros, P. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 3692–3696
35. Underhill, D. A., and Gros, P. (1997) J. Biol. Chem. 272, 14175–14182
36. Fortin, A. S., Underhill, D. A., and Gros, P. (1997) Hum. Mol. Genet. 6, 1781–1790
37. Fujikawa, M., Miskiewicz, P., Bai, L., Guldberg, A. A., Weir, M., and Goto, T. (1996) Development 122, 2697–2707
38. Fitzsimmons, D., Hodsdon, W., Wheat, W., Maira, S.-M., Wasylyk, B., and Hargman, J. (1996) Genes Dev. 10, 2198–2211