

DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site

Thomas Czerny, Gotthold Schaffner, and Meinrad Busslinger¹

Research Institute of Molecular Pathology, Dr. Bohr-Gasse 7, A-1030 Vienna, Austria

Previous DNA-binding studies indicated that an intact paired domain is required for interaction of the transcription factor BSAP (Pax-5) with DNA. We have now identified a subset of BSAP recognition sequences that also bind to a truncated BSAP peptide lacking 36 carboxy-terminal amino acids of the paired domain. Sequence comparison of this class of BSAP-binding sites made it possible to unequivocally align all known BSAP-binding sites and to deduce a consensus sequence consisting of two distinct half sites. We propose here a model for the paired domain–DNA interaction in which the paired domain is composed of two subdomains that bind to the two half-sites in adjacent major grooves on the same side of the DNA helix. The existence of these half sites and of the two paired domain subregions was directly demonstrated by methylation interference analysis and by in vitro mutagenesis of both the paired domain and its recognition sequence. Both half-sites contribute to the overall affinity of a given BSAP-binding site according to their match with the consensus sequence. However, none of the naturally occurring BSAP-binding sites completely conform to the consensus sequence. Instead, they contain compensatory base changes in their half-sites that explain the versatile and seemingly degenerate DNA sequence recognition of Pax proteins. Domain swap experiments between BSAP and Pax-1 demonstrated that the sequence specificity of the BSAP paired domain is determined by both its amino- and carboxy-terminal subdomains. Moreover, mutations affecting only one of the two subdomains restricted the sequence specificity of the paired domain. Such mutations have been shown previously to be the cause of mouse developmental mutants (*undulated*, *Spotch*, and *Small eye*) and human syndromes (Waardenburg's syndrome and aniridia) and may thus differentially affect the regulation of target genes by the mutated Pax protein.

[Key Words: Transcription factor BSAP; Pax proteins; paired domain; sequence recognition]

Received June 21, 1993; revised version accepted July 29, 1993.

Transcription factors play a central role in regulating gene expression by binding to specific control elements in enhancers and promoters. Their ability to recognize specific DNA sequences is specified by a variety of structurally distinct DNA-binding domains, each of which defines a separate transcription factor family (for review, see Harrison 1991; Pabo and Sauer 1992). One of these DNA-binding motifs is the conserved paired domain, which is characteristic of a small class of developmental control genes found in *Drosophila* and vertebrates (for review, see Gruss and Walther 1992; Noll 1993). Nine paired box-containing (*Pax*) genes have been identified so far in the mammalian genome (Walther et al. 1991; Stapleton et al. 1993). Three of them, *Pax-1*, *Pax-3*, and *Pax-6*, have been associated with the mouse developmental mutants *undulated* (*un*) (Balling et al. 1988), *Spotch* (Epstein et al. 1991), and *Small eye* (Hill et al. 1991), whereas mutations in the human *PAX-3* and *PAX-6*

genes cause Waardenburg's syndrome (WS) (Baldwin et al. 1992; Morell et al. 1992; Tassabehji et al. 1992) and aniridia (Glaser et al. 1992; Ton et al. 1991), respectively. This evidence, together with the fact that transcriptional activity has been directly demonstrated for some *Pax* gene products (Chalepakakis et al. 1991; Adams et al. 1992; Zannini et al. 1992), implies that Pax proteins are important regulators of early development.

The paired domain consists of a stretch of 128 amino acids that has been well conserved in evolution (Bopp et al. 1989). Pax proteins can be grouped into six different subfamilies according to their sequence similarity in this region (Walther et al. 1991). The paired domain shows no obvious resemblance to other known DNA-binding motifs, and its structure has not yet been resolved. Moreover, the DNA sequence recognition by the paired domain has been poorly characterized so far. A DNA sequence found in the *Drosophila even-skipped* promoter (referred to as the e5 site) has been used for studying the DNA-binding activity of the Paired, Pax-1, and Pax-3 proteins (Chalepakakis et al. 1991; Goulding et al. 1991;

¹Corresponding author.

Treisman et al. 1991). Gel shift assays using mutant oligonucleotides derived from this e5 site led to the identification of the pentanucleotide GTTCC as the DNA-binding core motif recognized by the paired domain (Chalepakidis et al. 1991). However, it is important to note that there is no genetic evidence for any interaction of a *Drosophila* paired domain-containing protein with the e5 site (Noll 1993).

Functional target sequences have been identified to date only for the subfamily comprising Pax-2, Pax-5 (BSAP), and Pax-8. While studying the developmental regulation of two pairs of nonallelic histone *H2A-2* and *H2B-2* genes in the sea urchin, we identified a transcription factor, TSAP, that binds to and regulates each of these four promoters (Barberis et al. 1989). Subsequently, we identified the B-cell-specific transcription factor BSAP as a mammalian homolog of this sea urchin protein (Barberis et al. 1990). Furthermore, we found that the *CD19* gene, which codes for a B-cell-specific surface protein, is a BSAP target gene (Kozmik et al. 1992). Biochemical purification and cDNA cloning revealed that BSAP is encoded by the *Pax-5* gene (Adams et al. 1992). Moreover, the recent cloning of TSAP cDNA confirmed that the paired domain of this sea urchin protein is highly related to that of BSAP (Pax-5) (T. Czerny, unpubl.). Both TSAP and BSAP generate a relatively large, 25-bp-long DNase I footprint on the five natural target sequences mentioned above (Barberis et al. 1989; Kozmik et al. 1992). However, only a degenerate consensus sequence could be deduced from these sequences, which furthermore lack the GTTCC motif (Barberis et al. 1989; Kozmik et al. 1992). Neither the identification of additional BSAP target sequences in the vicinity of immunoglobulin heavy-chain gene switch regions (Rothman et al. 1991; Waters et al. 1989; Williams and Maizels 1991; Liao et al. 1992) nor the demonstration that Pax-8, a close relative of BSAP (Pax-5), binds to and regulates the thyroglobulin (*Tg*) and thyroperoxidase (*TPO*) genes (Zannini et al. 1992) seemed to facilitate the definition of a better consensus sequence. Hence, degenerate DNA sequence recognition appears to be a characteristic feature of Pax proteins.

In this paper we have studied in detail how the paired domain of BSAP interacts with DNA. The starting point for this analysis was the observation that a truncated BSAP protein lacking 36 amino acids at the carboxyl terminus of its paired domain binds only to a subset of BSAP recognition sequences. This allowed us to unequivocally align all BSAP-binding sites and to define an improved consensus sequence that consists of two distinct half-sites. We then used in vitro mutagenesis to directly demonstrate that the BSAP paired domain and its recognition sequence exhibit a bipartite structure. Each half site is a major groove contact for one of the two subdomains of the paired domain. The overall affinity of a given BSAP-binding site is shown to depend on the strength of both interactions and to correlate with its match to the consensus sequence. As a consequence, deviation of one half site from its consensus sequence can be compensated for by a perfect match of the second half

site with its consensus motif. Compensatory base changes in the two half sites thus appear to be responsible for the versatile and seemingly degenerate DNA sequence recognition of Pax proteins.

Results

A functional BSAP-binding site is 18 bp long

Previous protein–DNA binding studies indicated that the BSAP recognition sequence is unusually large. BSAP protected 25–27 bp on each strand from DNase I digestion and contacted nucleotides on two successive turns of the DNA helix in methylation interference experiments (Barberis et al. 1990; Kozmik et al. 1992). To further define the BSAP recognition sequence, we systematically mutated every nucleotide of a high affinity binding site that originates from the sea urchin *H2A-2.2* gene (Fig. 1). Binding of BSAP to this panel of mutant oligonucleotides was tested by electrophoretic mobility shift assay (EMSA). As shown in Figure 1, most mutations throughout the binding site reduce BSAP binding. Interestingly, the most severe mutations are not clustered but, rather, are evenly spaced at positions 9, 12, 16, and 21. Moreover, the first and last mutations (positions 6 and 23) that reduce BSAP binding are 18 nucleotides apart from each other. This indicates that functional BSAP-binding sites may be highly complex, as they can span up to 18 bp. Furthermore, deletion of a single nucleotide (G residue 14) in the middle of the recognition sequence drastically reduces BSAP binding. Similarly, insertion of one nucleotide (A) between positions 13 and 14 prevents binding almost completely (last two lanes of Fig. 1). We conclude, therefore, that the topology between the left and right halves of the BSAP recognition sequence is critical for protein binding.

A BSAP peptide with an incomplete paired domain binds only to a subset of BSAP recognition sequences

Binding of BSAP to the *H2A-2.2* recognition sequence was shown previously to be strictly dependent on an intact paired domain (Adams et al. 1992). This finding is in apparent contradiction with the observations of Treisman et al. (1991) and Chalepakidis et al. (1991) that the *Drosophila* Paired and mouse Pax-1 proteins bind DNA even in the absence of the last 35 amino acids of the paired domain. These investigators used the e5 site for their studies; hence, a possible explanation for the discrepancy could be that distinct recognition sequences differ in their requirement for an intact paired domain. To examine this possibility, we have created a large panel of oligonucleotides based on BSAP-binding sites and tested whether they could bind full-length BSAP or the truncated protein BSAP 1–107 lacking 36 carboxy-terminal amino acids of the paired domain. The different BSAP recognition sequences originate from the sea urchin *H2A-2* and *H2B-2* genes, the human *CD19* gene, the *Drosophila even-skipped* promoter (e5; PRS-5) and from different regions of the immunoglobulin heavy-chain

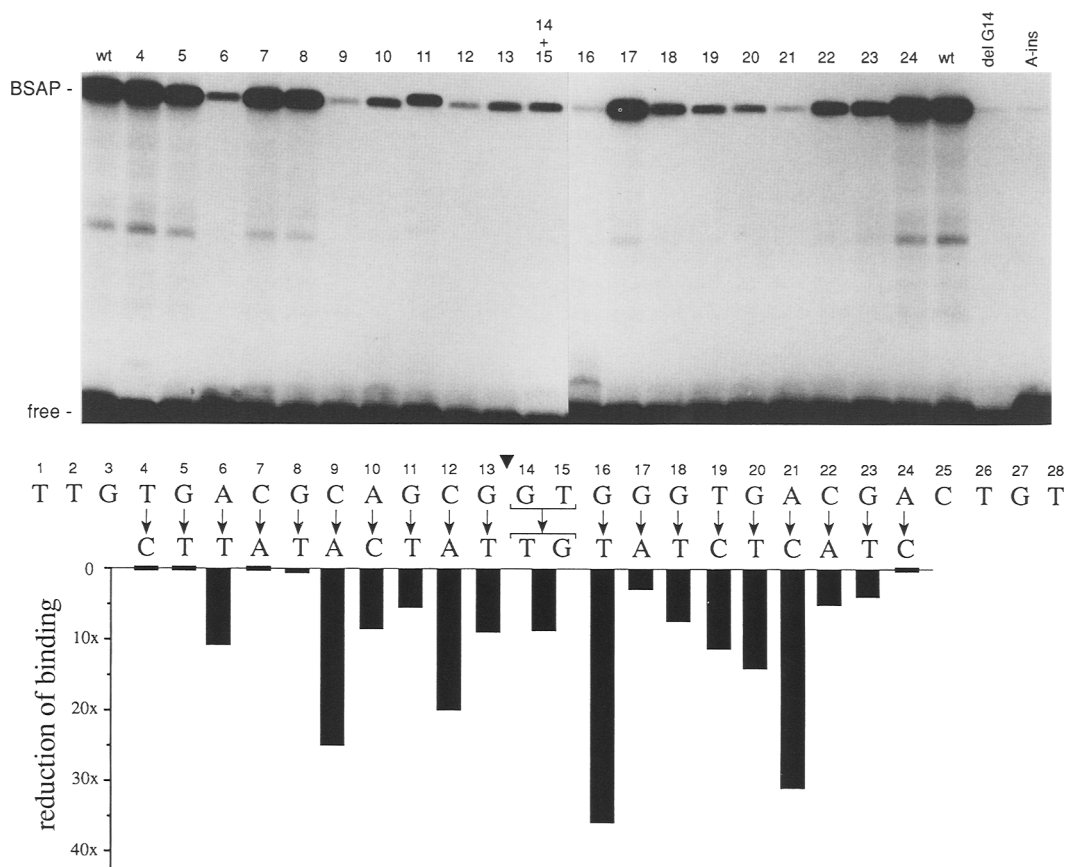


Figure 1. Identification of nucleotides critical for BSAP binding by saturation mutagenesis of a BSAP recognition sequence. Single point mutations were introduced into the BSAP-binding site of the sea urchin *H2A-2.2* gene (Barberis et al. 1989), and these mutant oligonucleotides were analyzed by EMSA using in vitro-translated BSAP protein. The oligonucleotide denoted A-ins contains an additional A residue between the two guanines at positions 13 and 14 (indicated by an arrowhead). The radioactivity in the BSAP complex was quantitated on a PhosphorImager (Molecular Dynamics), and the reduction of BSAP binding was plotted below each point mutation.

gene locus (for references, see legend to Fig. 2). As shown by the EMSA analysis in Figure 2, full-length BSAP, either in vitro translated or assayed in BJA-B nuclear extracts, interacted with all of these sites. Three sequences (e5, α S-1, and S γ 1) were bound with low affinity, whereas all other sites were recognized with medium or high affinity, in agreement with previous analyses (Barberis et al. 1990; Adams et al. 1992; Kozmik et al. 1992). In contrast, the truncated peptide BSAP 1–107 was able to bind only 6 of the 13 DNA probes (H2A-2.1, 5'S γ 2a, and wild-type and mutant CD19-2 and e5 sites). A 10-fold higher molar concentration of the truncated peptide was needed to obtain equivalent binding to these sites, thus indicating that the truncated BSAP peptide recognizes DNA with an $\sim$ 10-fold lower affinity than full-length BSAP. Interestingly, the ability of an oligonucleotide to bind the truncated BSAP protein did not correlate with its affinity for full-length BSAP. These data also demonstrate that eliminating the last 36 amino acids from the paired domain restricts the sequence specificity of the truncated peptide. This change in sequence specificity explains the previous discrepancy between our DNA-

binding results (Adams et al. 1992) and other published data (Chalepakakis et al. 1991; Treisman et al. 1991).

A consensus recognition sequence and a model for the DNA-paired domain interaction of BSAP

For sequence alignment we divided the different BSAP recognition sequences into two classes (I and II) according to their ability to bind the truncated BSAP peptide (Fig. 3A). Comparison of the class II recognition sequences, which interact with BSAP 1–107, revealed that all of these sequences are highly conserved in their 3' half. The deduced consensus sequence for this 3' part is GCG G/T A/G AC G/C G/A and contains the pentanucleotide GGAAC, which was identified previously as a DNA-binding core motif for Pax-1 (Chalepakakis et al. 1991; complementary strand published: GTTCC). The class I BSAP recognition sequences, which do not bind BSAP 1–107, were aligned according to their match with the 3' consensus motif. As shown in Figure 3A, all of these sequences deviate from the 3' consensus motif at 1 to 4 positions. In particular, five of these seven sites con-

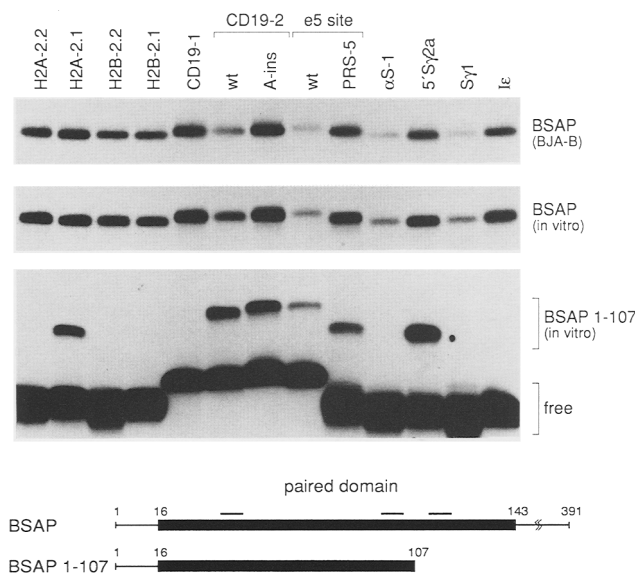


Figure 2. Differential sequence recognition by truncated and full-length BSAP proteins. Oligonucleotides containing different BSAP recognition sites were analyzed by EMSA for binding to the endogenous BSAP of BJA-B nuclear extracts, to in vitro-translated full-length BSAP, and to the truncated peptide BSAP 1–107. The latter peptide was used in a 10-fold higher molar concentration than full-length BSAP. Only the relevant part of the autoradiograph containing the protein–DNA complexes is shown for full-length BSAP. A schematic representation of the two BSAP peptides is shown below. Horizontal rules above the paired box denote the positions of the three α -helices postulated by secondary structure prediction (Bopp et al. 1989). The different BSAP-binding sites originate from the sea urchin *H2A-2* and *H2B-2* genes (Barberis et al. 1989), the human *CD19* gene (Kozmik et al. 1992), the *Drosophila even-skipped* promoter (e5, Treisman et al. 1991; PRS-5, Chalepakos et al. 1991) and from different regions of the immunoglobulin heavy-chain gene (α S-1, Waters et al. 1989; 5'Sy2a, Liao et al. 1992; Sy1, Williams and Maizels 1991; I ϵ , Rothman et al. 1991).

tain a guanine at position 21, which is absent in all class II sequences.

The 5' half of the BSAP-binding sites is conserved at five positions. A purine is present at position 6 in all sequences and a TG dinucleotide is preferentially located at positions 12 and 13. In addition, all class I sequences contain an invariant CA dinucleotide at positions 9 and 10, whereas the same two nucleotides are absent in class II binding sites.

In the past we have used methylation interference analysis to study the interaction of BSAP with its binding sites in the *H2A-2*, *H2B-2*, and *CD19* promoters (Barberis et al. 1990; Kozmik et al. 1992) and with the e5 and PRS-5 sequences (Adams et al. 1992). As summarized in Figure 3B, these methylation interference data indicate that BSAP contacts guanines in two clusters that are separated by one turn of the DNA helix. Hence, the paired domain of BSAP recognizes one side of the DNA helix by interacting with bases in two consecutive major grooves. Comparison of the methylation interference

data (Fig. 3B) with the consensus recognition sequence (Fig. 3A) demonstrates that both the 5' consensus sequence centered around the CA dinucleotide and the more extensive 3' consensus motif correspond to the two contact sites in the major groove. We therefore concluded that BSAP-binding sites exhibit a bipartite structure, with each half site being represented by the 5' or 3' consensus motif. Furthermore, we hypothesized that the paired domain might also be divided into two subdomains. As illustrated by the model in Figure 3C, the carboxy-terminal part of the paired domain would interact with the 5' half of the recognition sequence containing the invariant CA dinucleotide. Conversely, the amino-terminal region of the paired domain would contact the more extensive 3' consensus motif.

A consequence of this model is that the two individual interactions with the major groove both contribute to the total energy of the BSAP–DNA-binding reaction. Thus, either both contact sites could make a similar contribution to the overall affinity or one of the two contact sites might dominate over the other. The latter case is apparently encountered with the class II binding sites, which perfectly match the 3' consensus motif but lack the critical CA dinucleotide in the 5' contact site. According to our model (Fig. 3C), the 3' site should therefore make a greater contribution to the overall affinity. Dominance of the 3' site over the 5' site could explain why BSAP 1–107 is able to bind class II recognition sequences, as deletion of the carboxy-terminal 36 amino acids of the paired domain would weaken the interaction of BSAP with only the less important 5' contact site.

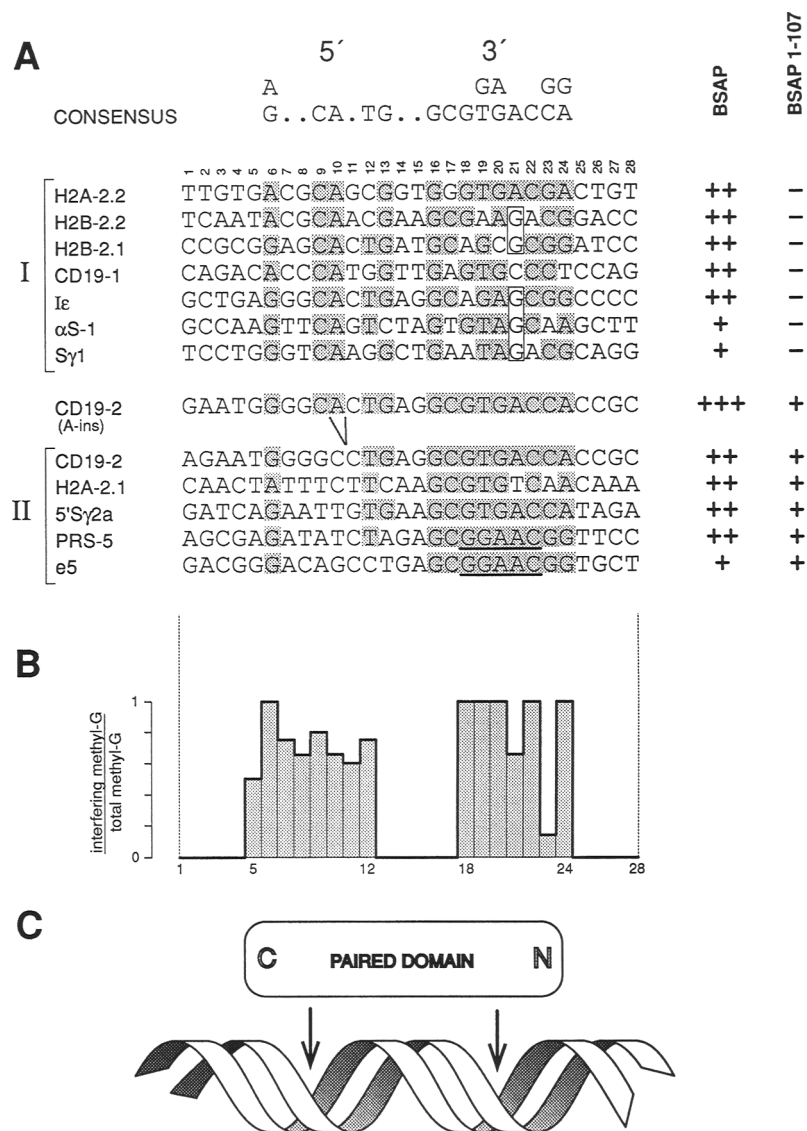
The carboxy-terminal region of the BSAP paired domain interacts with the 5' consensus motif

The sequence-specific interaction of the two paired domain subregions was directly demonstrated by methylation interference analysis. For this purpose we chose a high affinity binding site [CD19-2(A-ins)], which is bound by both full-length BSAP and BSAP 1–107. As shown in Figure 4, binding of full-length BSAP is prevented by methylation of two G residues in the 5' consensus motif and three G residues in the 3' consensus sequence. The truncated BSAP peptide lacking part of the carboxy-terminal paired domain, however, contacts only the three G residues of the 3' consensus sequence. This evidence directly demonstrates that the amino-terminal part of the paired domain interacts with the 3' consensus motif. It therefore follows that the carboxy-terminal paired domain sequences in full-length BSAP contact the 5' motif.

Chimeric recognition sequences confirm the bipartite structure of the BSAP-binding site

A further prediction of our model (Fig. 3C) is that improving the match of a binding site with the consensus sequence should increase its affinity for BSAP. Previously, we noted that the CA dinucleotide is missing in the 5' half of the CD19 site 2 because of lack of an A

Figure 3. Consensus recognition sequence and model for the DNA–paired domain interaction of BSAP. (A) BSAP recognition sequences. In each case, the sequence is shown for the entire DNase I-protected region. The orientation of the sequences corresponds to the natural orientation of the BSAP-binding sites in the *H2A-2.2*, *H2B-2.1*, *H2B-2.2*, *CD19*, and *I ϵ* promoters. Only the *H2A-2.1* sequence is shown in the opposite orientation as it occurs in the *H2A-2.1* promoter. The numbering of the positions is the same as in Fig. 1. Consensus residues are shaded, and the deduced consensus sequence is shown above the sequence compilation. The complementary strand of the pentanucleotide GTTCC (Chalepakidis et al. 1991) is underlined in the e5 and PRS-5 sequences. Binding of full-length BSAP and of the truncated peptide BSAP 1–107 is summarized at right. Most class I binding sites contain a G residue (boxed) at position 21 that is suboptimal for BSAP binding (see Fig. 5) and was therefore omitted from the consensus. (B) Compilation of methylation interference data (Barberis et al. 1990; Adams et al. 1992; Kozmik et al. 1992) were used to determine the ratio of interfering methylated G residues over total methylated G residues at each nucleotide position of the BSAP-binding site. G residues in both strands were considered. (C) Model of the paired domain–DNA interaction. In this model the amino- and carboxy-terminal regions of the paired domain contact two successive major grooves from the same side of the DNA helix.



residue at position 10 (Kozmik et al. 1992). Insertion of an adenine at this position generates a BSAP recognition sequence that completely conforms to the consensus sequence in both half sites (Fig. 3A). This A insertion increased the affinity of the CD19 site 2 by a factor of 30, thus creating the best BSAP-binding site known so far (Fig. 2; Adams et al. 1992).

We then generated chimeric recognition sequences by swapping entire 5' and 3' half sites between class I and class II sequences (Fig. 5). The H2A-2.2 site was selected as a class I BSAP recognition sequence with a strong 5' half site, whereas the e5 site was chosen as a class II representative with a strong 3' half site (Fig. 2). The H2A/e5 site contains the 5' half of the H2A-2.2 site linked to the 3' half of the e5 site and thus matches the consensus sequence better than the two original sequences. The complementary site e5/H2A, however, fits neither the 5' nor the 3' consensus sequences particularly well. As expected, the H2A/e5 site bound full-length BSAP with even higher affinity than the H2A-2.2

site and, furthermore, acquired the potential to bind BSAP 1–107 (Fig. 5). In contrast, the e5/H2A site was not able to bind either of the two BSAP proteins (Fig. 5).

The validity of our consensus sequence was investigated further by single point mutations. The H2A-2.2 site deviates from the 3' consensus sequence only at position 17 (G instead of C). Replacing G with C not only allowed binding to BSAP 1–107 but further increased the affinity of this sequence (H2A-17C) for full-length BSAP (Fig. 5). Most class I recognition sequences contain a G residue at position 21, which is absent from the 3' consensus motif of class II binding sites (Fig. 3). As these sites do not bind BSAP 1–107, we argued that a guanine at this position might decrease the affinity of the 3' contact site. To test this possibility, we introduced a guanine at position 21 of the H2A-17C sequence. Consistent with our theory, this single point mutation strongly decreased binding to BSAP 1–107 and reduced the affinity of the new site (H2A-17C/21G) for full-length BSAP (Fig. 5). Taken together, these data demonstrate that both

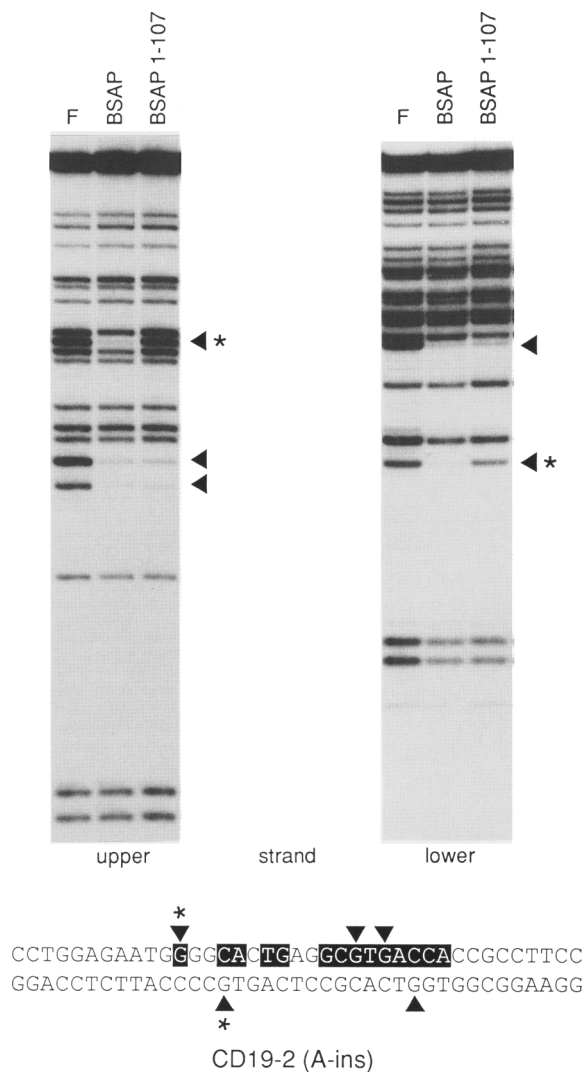


Figure 4. Methylation interference analysis of the DNA sequence interaction of full-length and truncated BSAP. The CD19-2 (A-ins) oligonucleotide shown at the bottom was cloned into the *Sa*II site of the plasmid pBend2 (Kim et al. 1989) and excised either as a *Bam*HI–*Bgl*II restriction fragment that was 3' end-labeled at the *Bam*HI site or as a *Nhe*I–*Bam*HI fragment with a 3' end-labeled *Nhe*I site. Both labeled DNA probes were used for methylation interference analysis (Barberis et al. 1989) with either full-length or truncated BSAP proteins that were expressed in transfected COP-8 cells. (F) Free DNA. The consensus sequence of Fig. 3 is highlighted by black overlay. Arrowheads point to methylated G residues that interfere with BSAP binding.

halves of the BSAP recognition sequence contribute to the overall affinity of BSAP binding. Moreover, there is a clear correlation between the match of a given binding site with the consensus sequence and its affinity for BSAP.

Identification of the minimal paired domain sequences required for DNA binding

Earlier we showed that a peptide consisting of the first

163 amino acids of BSAP is able to bind the H2A-2.2 site, a typical representative of class I sequences (Adams et al. 1992). To test whether the amino acids flanking the paired domain of BSAP 1–163 are essential for binding, we generated a peptide that consists only of paired domain sequences (BSAP 16–145; Fig. 6). This peptide, full-length BSAP, and BSAP 1–163 were synthesized in vitro (Fig. 7C) and then used in equivalent molar amounts for EMSA analysis with the H2A-2.2 probe (Fig. 7A). All three proteins bound with equal affinity to this sequence, thus indicating that the paired domain alone determines the DNA-binding properties of BSAP.

We then asked whether amino-terminal resections into the paired domain also lead to differential binding of

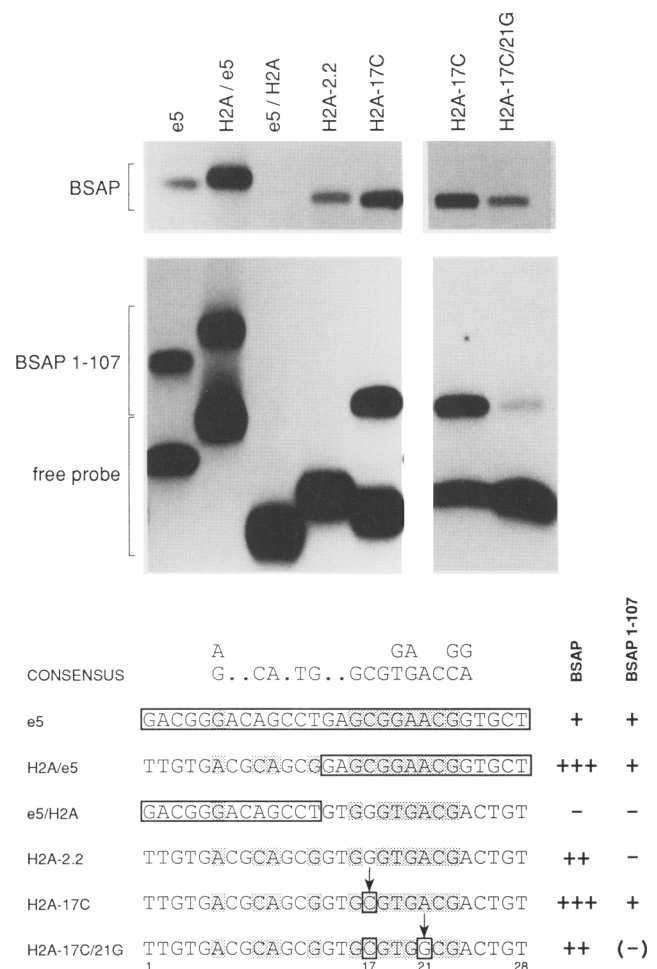


Figure 5. The affinity of chimeric BSAP-binding sites correlates with their match to the consensus sequence. Binding of in vitro-translated full-length and truncated BSAP peptides to the sequences shown at the bottom was investigated by EMSA. The truncated protein BSAP 1–107 was used in a 10-fold higher molar concentration than full-length BSAP. Only the relevant part of the autoradiograph containing the protein–DNA complexes is shown for full-length BSAP. The e5 sequence is boxed, and homologies with the consensus sequence are shaded. The G residue at position 21 (boxed) is present in most class I binding sites (see Fig. 3).

class I and II recognition sequences, as is seen with carboxy-terminal resections (Fig. 2). As shown in Figure 7B and summarized in Figure 6, amino-terminal deletion up to amino acid 30, which eliminates the first 14 amino acids of the paired domain, similarly reduced the affinity for both the H2A-2.2 and e5 sites by a factor of ~ 10 , because 10-fold higher concentrations of the truncated proteins had to be used to obtain the same level of binding as with BSAP 1–163. Further deletion up to position 37, which eliminates all sequences amino-terminal to the first predicted α -helix of the paired domain (Bopp et al. 1989), completely abolished binding to both the H2A-2.2 and e5 sites (Fig. 7B). These data map the amino-terminal boundary for the DNA-binding domain between amino acid residues 30 and 37 of BSAP.

The class I and II binding sites were defined according to their distinct requirements of carboxy-terminal paired domain sequences for DNA binding (Fig. 2). A detailed deletion analysis has previously indicated that even the most carboxy-terminal sequences of the paired domain are critical for interaction with the class I site of the H2A-2.2 gene (Adams et al. 1992). To delimit the minimal sequences required for binding to class II sites, we analyzed further carboxy-terminal resection mutants of the BSAP paired domain by EMSA with the e5 site (Fig. 7B). The peptide BSAP 1–97, lacking the second and third predicted α -helices of the paired domain (Bopp et al. 1989), still interacted with the e5 site as efficiently as BSAP 1–107. However, deletion of an additional eight amino acids (BSAP 1–89) dramatically reduced binding. In summary, the two classes of recognition sequences rely for binding to BSAP on the same amino-terminal

residues of the paired domain but exhibit a vastly different requirement for carboxy-terminal paired domain sequences.

Point mutations in the amino-terminal part of the paired domain preferentially affect binding to class II recognition sequences

The above result could be interpreted to mean that the amino-terminal amino acids of the paired domain are equally important for interaction with both types of binding sites. Alternatively, deletion analysis may be inadequate to study this question, as deletions can generate gross changes in the overall structure of a domain. We therefore introduced three point mutations into the amino-terminal part of the BSAP paired domain (Fig. 6) and analyzed their effect on DNA binding (Fig. 8). We substituted the arginine at position 38 in the first predicted α -helix with alanine (R38A), which constitutes a nonconservative amino acid change not reported so far. In contrast, the Gly–Ser replacement (G30S) at position 30 corresponds to the *un* mutation of the mouse *Pax-1* gene (Balling et al. 1988) and the Pro–Leu substitution (P32L) to a point mutation detected in the *PAX-3* gene of a patient with WS type I (Baldwin et al. 1992). BSAP peptides containing these mutations were all synthesized with equal efficiency in vitro (Fig. 8C). As shown in Figure 8A, the mutant protein P32L (WS) was completely unable to bind to the H2A-2.2 probe, whereas the other two substitutions, G30S (*un*) and R38A, reduced binding by more than a factor of 10 (Fig. 8A). In Figure 8B we analyzed the relative binding of the latter two mutants

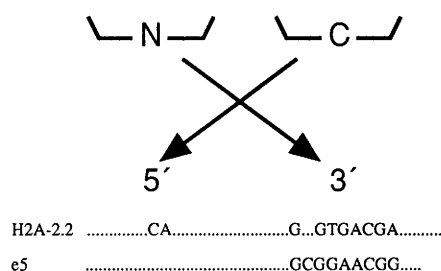
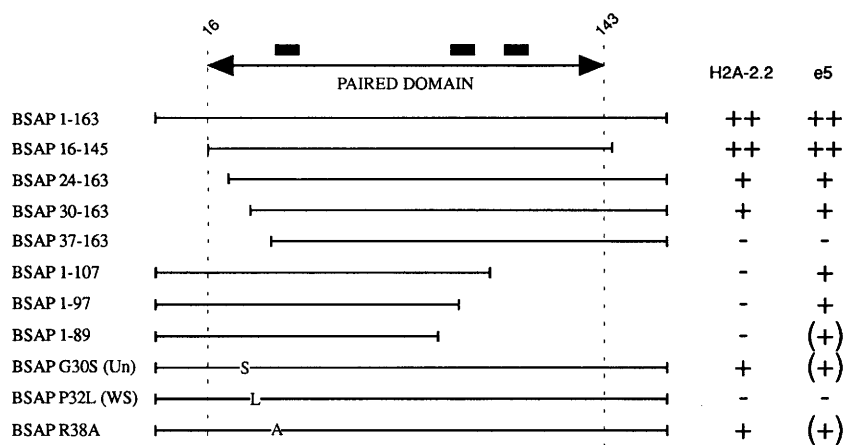


Figure 6. Summary of the in vitro mutagenesis of the BSAP paired domain. The first and last amino acids of each BSAP peptide are indicated. Numbering of amino acid residues refers to the sequence of full-length BSAP (Adams et al. 1992). The substitutions G30S and P32L correspond to the *undulated* (*un*) mutation of the mouse *Pax-1* gene (Balling et al. 1988) and to a point mutation detected in the *PAX-3* gene of a patient with Waardenburg's syndrome type I (WS, Brazil, Baldwin et al. 1992), respectively. Binding of the different BSAP peptides to the H2A-2.2 and e5 sites is summarized at right. The model of the paired domain–DNA interaction is diagramed at bottom.

DNA sequence recognition of BSAP (Pax-5)

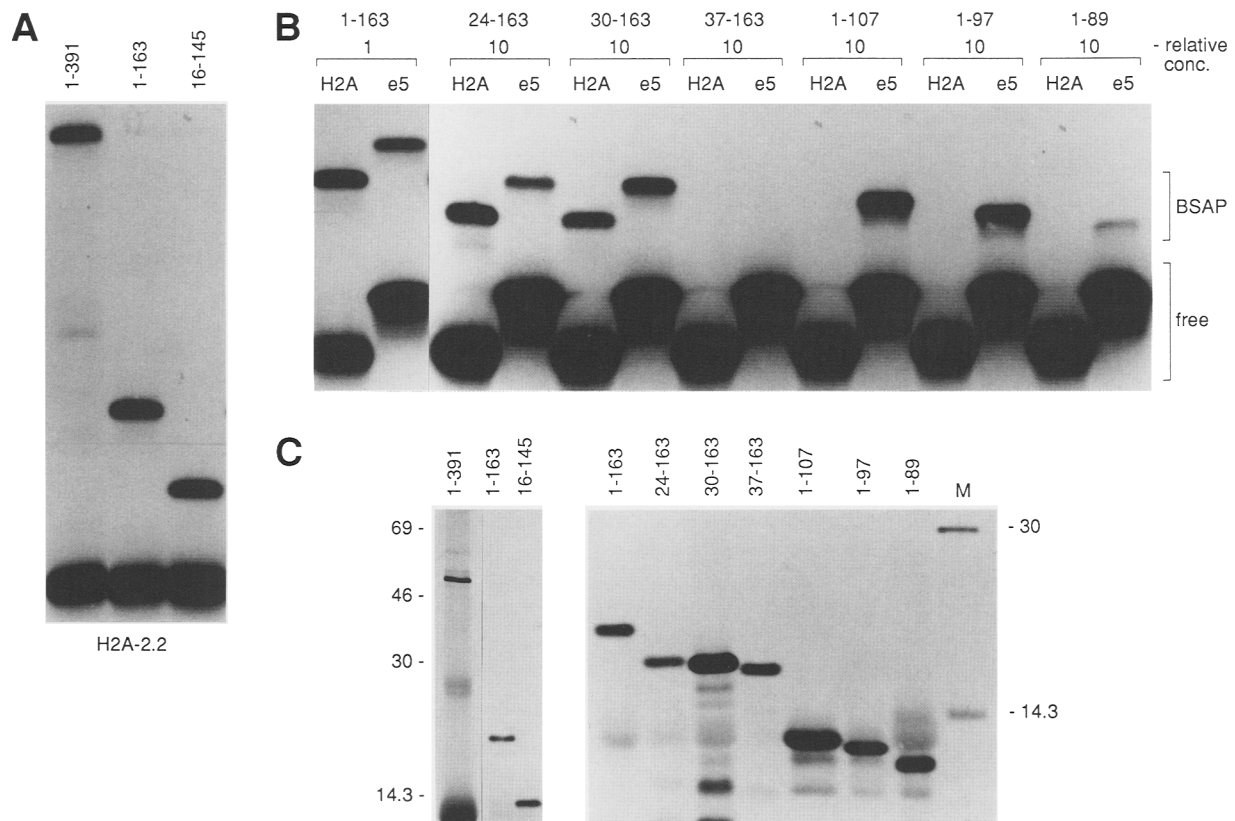


Figure 7. DNA binding to truncated versions of the BSAP paired domain. (A) The paired domain of BSAP is sufficient for high-affinity DNA binding. Full-length BSAP and the two peptides BSAP 1–163 and BSAP 16–145 were synthesized by in vitro translation (see panel C), and the same amount of each protein was analyzed by EMSA for binding to the H2A-2.2 site. (B) amino- and carboxy-terminal deletion mutants of the BSAP paired domain. All in vitro-translated deletion mutants were used in a 10-fold higher concentration than peptide BSAP 1–163 for EMSA analysis with the H2A-2.2 and e5 DNA probes. (C) Electrophoretic analysis of in vitro-translated proteins. BSAP peptides were synthesized by in vitro translation in the presence of [35 S]cysteine, and 1 μ l of each reaction was analyzed by 15% SDS-PAGE followed by autoradiography and quantitation on a PhosphorImager as described in Materials and methods. The positions of marker proteins (M, sizes in kD) are indicated at left and right.

to representative sites of both classes of recognition sequences (H2A-2.2 and e5). For better comparison we adjusted the concentrations of the mutant and wild-type proteins to yield similar complex formation with the H2A-2.2 probe. Under these conditions the point mutations G30S (*un*) and R38A further reduced BSAP binding to the e5 site by factors of 8 and 15, respectively. These data demonstrate that single point mutations in the amino-terminal part of the paired domain affect BSAP binding more strongly to class II recognition sequences (e5) than to class I sites (H2A-2.2), which is consistent with our model of the paired domain–DNA interaction (Fig. 3C).

The DNA-binding specificity of BSAP is determined by the amino- and carboxy-terminal halves of the paired domain

Our previous finding that the Pax-1 protein binds BSAP recognition sequences with lower affinity than BSAP (Adams et al. 1992) allowed us to identify, by domain

swap experiments, sequences in the BSAP paired domain that determine its DNA-binding specificity. For this purpose, the amino-terminal sequences of Pax-1 were fused in the conserved middle part of the paired domain to the corresponding sequences of BSAP to generate the protein Pax1/BSAP (Fig. 9). Conversely, the fusion protein BSAP/Pax1 consists of amino-terminal sequences of BSAP linked at the same position in the paired domain to Pax-1. Chimeric and parental proteins were synthesized in vitro, and the same amount of each protein was subjected to EMSA analysis with both the e5 and H2A-2.2 DNA probes (Fig. 9). BSAP interacts with the H2A-2.2 site ~50-fold better than Pax-1, in agreement with previous data (Adams et al. 1992). Both chimeric proteins bind to the H2A-2.2 probe with a six- to eightfold higher affinity than Pax-1, thus indicating that the DNA-binding specificity of BSAP resides in both the amino- and carboxy-terminal halves of the paired domain. These data further imply that both the amino- and carboxy-terminal regions of the paired domain cooperate with each other to bring about high affinity binding of BSAP.

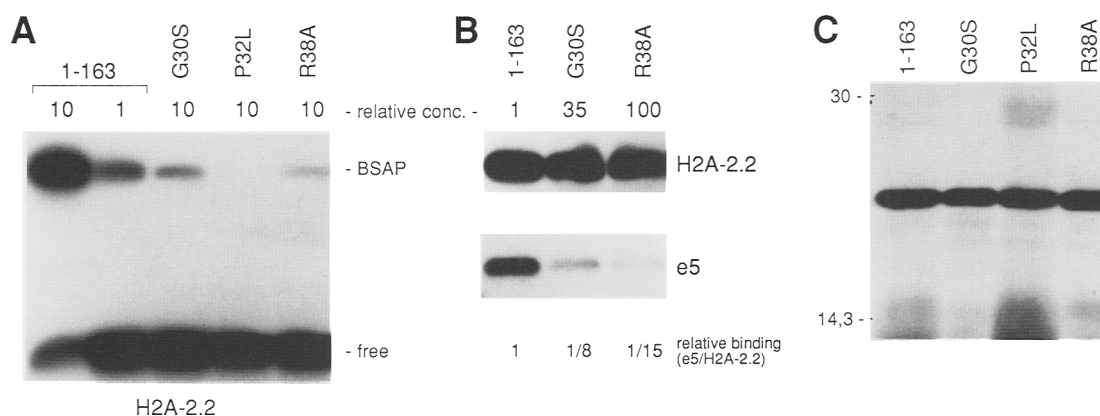


Figure 8. Amino acid substitutions in the amino-terminal part of the paired domain preferentially reduce binding to class II recognition sequences. (A) Wild-type and mutated BSAP peptides (see also Fig. 6) were synthesized *in vitro* and used for EMSA analysis with the H2A-2.2 probe at the relative molar concentrations indicated. (B) Preferential reduction of binding to the e5 recognition sequence. The concentrations of the different *in vitro*-translated BSAP peptides were adjusted to ensure equivalent protein binding to the H2A-2.2 DNA probe. The same relative concentrations were then used for EMSA analysis with the e5 probe. Only the relevant part of the autoradiographs containing the BSAP–DNA complexes are shown. The ratio of relative binding to the two sites was calculated for each mutation by quantitating the radioactivity of the BSAP signals on a PhosphorImager. (C) SDS-PAGE analysis of the *in vitro*-translated BSAP peptides. The positions of marker proteins (sizes in kD) are indicated at left.

Our model in Figure 3C predicts that improving the DNA-binding potential of the carboxy-terminal part of the paired domain in the chimeric protein Pax1/BSAP should have little or no effect on the binding of class II recognition sequences. As shown in Figure 9, Pax1/BSAP and Pax-1 bind to the e5 site with nearly the same affinity in clear contrast to the observation with the H2A-2.2 probe. This result supports our model of the paired domain–DNA interaction. Because BSAP/Pax1 and BSAP bind to the e5 site with higher affinity than Pax-1, it seems that the amino-terminal part of the BSAP paired domain is optimized for interaction with the 3' consensus motif.

Pax proteins of different subfamilies exhibit distinct sequence specificities

The different Pax proteins can be divided into six distinct subfamilies according to their sequence similarity in the paired domain (Walther et al. 1991). The question therefore arises whether the consensus recognition sequence deduced for BSAP (Fig. 3) is also valid for Pax proteins of other subfamilies. To answer this question, we analyzed binding of representative members of these subfamilies to the CD19-2 probe containing the A insertion (A-ins) (Fig. 10A). This site was chosen because of its perfect match with the consensus sequence (Fig. 3). As shown in Figure 10A, other paired domain proteins (Pax-1, Pax-6, Pax-3, Gsb) also bind to this recognition sequence. However, their affinity for this site is lower than that of BSAP, thus indicating that members of distinct Pax subfamilies differ in their sequence recognition. The consensus sequence in Figure 3 therefore appears to be optimized for the subfamily comprising BSAP.

To gain further insight into the sequence specificity of

the different Pax proteins, we analyzed binding of these proteins to our panel of BSAP recognition sequences (Fig. 10B). Two subfamilies of Pax proteins, represented by Pax-3 and the *Drosophila* proteins Pox neuro (Poxn), Paired (Prd), and Gooseberry (Gsb), bind exclusively to class II binding sites. This behavior and the low affinity of these proteins for the CD19-2 (A-ins) sequence is similar to that seen with BSAP 1–107 (Fig. 10A). Thus, proteins of these two subfamilies may interact only with the 3' consensus sequence. Representatives of two other subfamilies, Pax-1 and Pax-6, show differential interaction with members of both classes of BSAP-binding sites and thus exhibit novel DNA sequence specificities. Interestingly, all of the Pax proteins analyzed bind to the three recognition sequences, which were engineered to completely match the BSAP consensus sequence.

Discussion

The paired domain is a novel DNA-binding motif that characterizes a unique family of transcription factors (Chalepakakis et al. 1991; Treisman et al. 1991; Adams et al. 1992). In this report we subjected the paired domain of the transcription factor BSAP (Pax-5), one of the mammalian Pax proteins, to a detailed structure–function analysis by *in vitro* mutagenesis and DNA-binding assays using a large panel of naturally occurring BSAP-binding sites. These experiments allowed us to deduce a nonpalindromic consensus recognition sequence for BSAP and to put forward the following model of the paired domain–DNA interaction (Fig. 3). The paired domain binds to its recognition sequence from one side of the DNA helix and thereby interacts with two successive major grooves. The BSAP recognition sequence is divided into two halves, each corresponding to one major

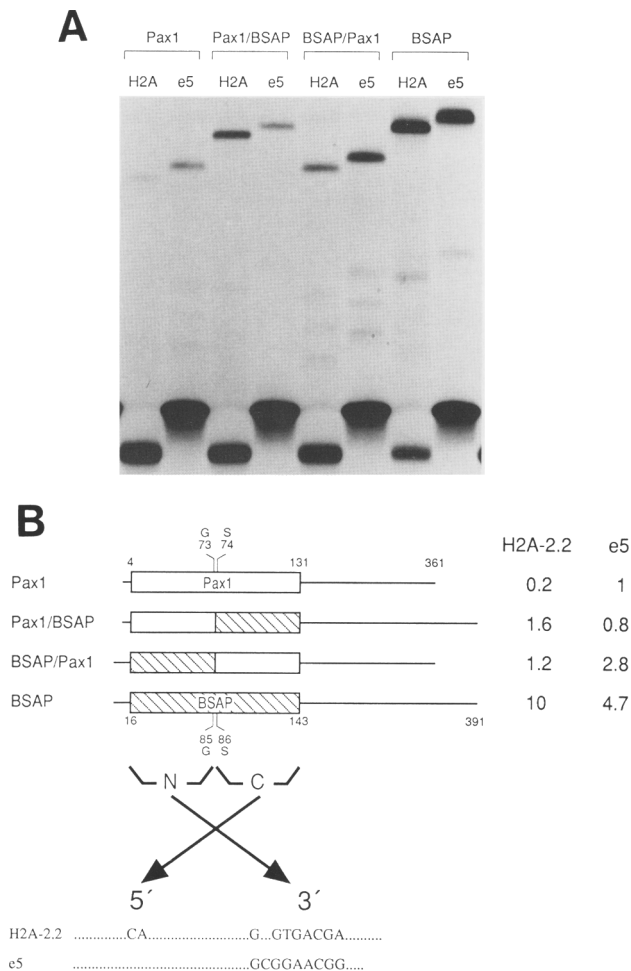


Figure 9. The DNA-binding specificity of BSAP resides in both halves of the paired domain. Pax-1, BSAP, and the chimeric proteins BSAP/Pax1 and Pax1/BSAP were translated in vitro and quantitated by SDS-PAGE and autoradiography. The same molar amount of each protein was used for EMSA analysis with the H2A-2.2 and e5 DNA probes (A). The chimeric proteins Pax1/BSAP and BSAP/Pax1 are schematically diagrammed, in B, together with the model of the paired domain–DNA interaction. Relevant amino acid positions of BSAP (Adams et al. 1992) and Pax-1 (Chalepakakis et al. 1991) are indicated. Relative binding of the different proteins to the H2A-2.2 and e5 sites is indicated at the far right and was determined by quantitating the autoradiographic signals shown in panel A on a PhosphorImager. The level of Pax-1 binding to the e5 site was given a value of 1. As BSAP binding already reached saturation levels in the experiment shown, the relative binding values of BSAP were determined in a separate EMSA experiment using five times lower concentrations of each protein.

groove contact site. The paired domain also exhibits a bipartite structure. The amino-terminal subdomain recognizes the more extensive 3' consensus motif of the BSAP-binding site, whereas the carboxy-terminal part of the paired domain interacts with the 5' consensus motif. The two interactions with these half sites together determine the overall affinity of a given binding site.

This bipartite model of the paired domain–DNA interaction is supported by the following evidence. First, none of the naturally occurring BSAP-binding sites completely conform to the BSAP consensus sequence. Instead, these sites can be divided into two classes depending on whether their 5' (class I) or 3' (class II) half sites fully match the consensus sequence (Fig. 3). Second, BSAP peptides lacking up to 46 amino acids from the carboxy-terminal region of the paired domain are only capable of binding to class II recognition sequences. The 3' consensus motif of these sites appears to be optimized for interaction with the amino-terminal part of the paired domain that is still present in the truncated peptides. Class II sites show, however, only a rudimentary match to the 5' consensus motif, thus pointing to a weak interaction with the carboxy-terminal half of the paired domain. Consequently, deletion of this carboxy-terminal subdomain in the truncated peptides has only a relatively small effect on DNA binding (10-fold reduction). The 3' interaction site of class II recognition sequences is therefore dominant over the 5' site, in contrast to class I sequences where both half sites appear to contribute equally to the overall affinity. Third, methylation interference experiments with a truncated BSAP peptide lacking the carboxy-terminal region of the paired domain directly demonstrated that the remaining amino-terminal part of the paired domain exclusively contacts the 3' consensus motif of the BSAP-binding site. Fourth, our model predicts that class II recognition sequences rely primarily on the interaction with the amino-terminal region of the paired domain. Single amino acid substitutions in the amino-terminal part of the paired domain preferentially affected binding to these sites compared with class I recognition sequences. Fifth, the sequence specificity of the BSAP paired domain is determined by both its amino- and carboxy-terminal regions as shown by domain exchange experiments with Pax-1. Sixth, novel binding sites created by exchanging 5' and 3' half sites of class I and II recognition sequences exhibited binding specificities as predicted by our model. Furthermore, the binding affinities of these sites strictly correlated with the match to the BSAP consensus sequence.

How does the new consensus sequence (Fig. 3) compare with previously published consensus sequences for Pax proteins (Barberis et al. 1989; Chalepakakis et al. 1991)? By comparing the TSAP-binding sites in sea urchin histone H2A-2 and H2B-2 promoters we arrived at a degenerate consensus sequence that emphasizes the importance of the 5' half site for protein binding, as it contains all its conserved base pairs (Barberis et al. 1989). In contrast, Chalepakakis et al. (1991) defined the pentamer GTTCC as a DNA-binding core motif for Pax-1 by in vitro mutagenesis of the *Drosophila* e5 sequence. This pentamer is part of our 3' consensus motif (Fig. 3). In this context it is, however, important to note that even the improved BSAP consensus sequence (Fig. 3) may not greatly facilitate identification of binding sites based on sequence comparison alone. One reason for this is provided by the observation that binding sites with a complete match to the consensus sequence possess an excep-

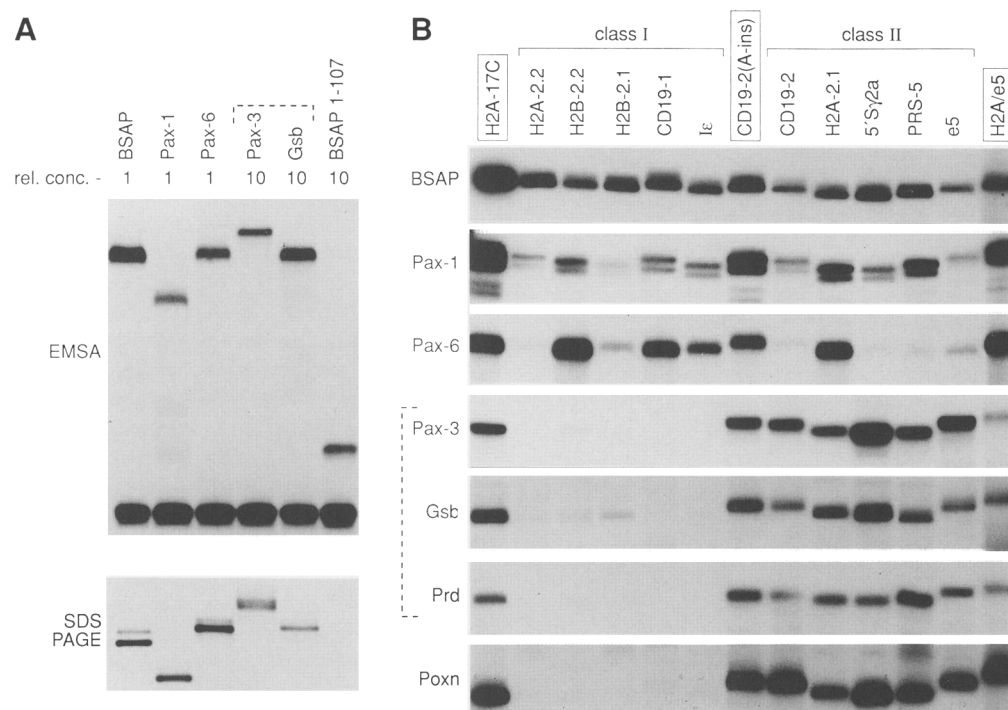


Figure 10. Pax proteins of different subfamilies exhibit distinct sequence specificities. (A) Comparison of DNA binding of different Pax proteins. The different Pax proteins were synthesized *in vitro* and analyzed by SDS-PAGE followed by autoradiography (*bottom*). Binding to the CD19-2 (A-ins) oligonucleotide was analyzed by EMSA (*top*) using the same molar amount of BSAP, Pax-1, and Pax-6 and 10-fold higher concentrations of Pax-3, Gsb, and BSAP 1-107. (B) Binding of the different paired domain proteins to a panel of BSAP recognition sequences. The different Pax proteins were expressed by transient transfection in COP-8 cells, and whole cell extracts were analyzed by EMSA for binding to the different BSAP-binding sites (see Fig. 3). No attempt was made to adjust the absolute protein concentrations of the different Pax proteins for the EMSA analysis in B in contrast to the experiment in A. Abbreviations of the *Drosophila* proteins: (Poxn) Pox neuro (Bopp et al. 1989); (Prd) Paired (Bopp et al. 1986); (Gsb) Gooseberry (BSH9; Baumgartner et al. 1987). The three BSAP-binding sites with the highest affinity are boxed, and members of the same subfamily of Pax proteins are indicated by brackets.

tionally high affinity for BSAP (Fig. 3; Adams et al. 1992). Such high affinities are apparently not required for BSAP binding *in vivo*, as all naturally occurring binding sites identified so far deviate from the consensus sequence by base changes in their 5' or 3' half sites. As a consequence, BSAP is able to interact with a panel of seemingly degenerate recognition sequences. This extraordinary flexibility of sequence recognition is thus a direct result of the bipartite structure of the paired domain.

The paired domains of BSAP (Pax-5), Pax-2, and Pax-8 are not only highly conserved but also recognize DNA in an identical manner (Kozmik et al. 1993), thus indicating that the BSAP consensus sequence holds true for this subfamily of Pax proteins. As shown in Figure 11, the Pax-8-binding sites identified in the *TPO* and *Tg* promoters (Zannini et al. 1992) conform to the BSAP consensus sequence. These two sites interact with BSAP like typical class I binding sites of medium and low affinity, respectively (see legend to Fig. 11). Members of other Pax subfamilies, however, exhibit different sequence specificities, as they bind only to subsets of BSAP recognition sequences with even reduced affinity. It is noteworthy that the high-affinity binding sites that match our con-

sensus sequence are bound by all the different Pax proteins. The BSAP consensus sequence therefore possesses the essential features of a general paired domain recognition sequence. Most Pax proteins bind to the class II binding sites that are optimized for the interaction with the amino-terminal part of the paired domain (Fig. 10B). In agreement with this, the paired domain sequences of Pax proteins are more highly conserved in the amino-

	A		GA	GG
CONSENSUS	G	.CA.TG	.GCGTGACCA	
TPO	ATGCCCACTCAAGCTTAGACAGG			
Tg	CTGCCCACTCAAGTGTCTTGAA			

Figure 11. Pax-8-binding sites conform to the BSAP consensus sequence. Shown is the alignment of the Pax-8 recognition sequences of *TPO* and *Tg* genes (Zannini et al. 1992) with the BSAP consensus sequence (Fig. 3). Compared with the H2A-2.2 site, TPO and Tg sites possess a 3- and 16-fold lower affinity for full-length BSAP, respectively. Both sites do not bind the truncated peptide BSAP 1-107 (T. Czerny, unpubl.).

terminal half than in the carboxy-terminal region (Walther et al. 1991). We therefore hypothesize that the recognition sequences of all Pax proteins resemble the conserved 3' motif of the BSAP consensus sequence but that their 5' motif is subfamily-specific, thus conferring high-affinity binding. In support of this, members of two subfamilies (Pax-3, Gsb, Prd, and Poxn) fail to interact with the 5' motif of BSAP recognition sequences (class I; Fig. 10B).

Genetic lesions in Pax genes are associated with mouse developmental mutants and human syndromes (for review, see Gruss and Walther 1992; Stapleton et al. 1993). Some of these mutations directly affect the paired domain either by deleting part of it (Pax-3, Epstein et al. 1991; Morell et al. 1992; Tassabehji et al. 1992) or by substituting single amino acids (Pax-1, Balling et al. 1988; Pax-3, Baldwin et al. 1992; Epstein et al. 1993; Hoth et al. 1993; Tassabehji et al. 1993; Pax-6, Glaser et al. 1992). These mutations either abrogate DNA binding by the paired domain, such as the Waardenburg Brazil (WB) substitution (Fig. 8, P32L, Baldwin et al. 1992), or reduce DNA binding as exemplified by the undulated mutation (Fig. 8, G30S; Chalepakakis et al. 1991). All mutations of the latter class not only reduce DNA binding but also simultaneously restrict the target sequence specificity of the paired domain (Figs. 7 and 8; Chalepakakis et al. 1991), because they preferentially affect the interaction with one of the two half sites according to our bipartite model of the paired domain–DNA interaction. As a consequence, these latter mutations may affect the expression of only a subset of target genes. In support of this idea, amino acid substitutions were found at different positions within the Pax-3 paired domain in patients with phenotypically distinct subtypes of WS (Baldwin et al. 1992; Hoth et al. 1993; Tassabehji et al. 1993).

The paired domain alone could be synthesized as a stable protein that contains all of the information necessary for the DNA binding of BSAP (Fig. 7A). The question therefore arises whether the two subregions of the paired domain also represent separate or, rather, interdependent units. Peptides with large carboxy-terminal deletions in the paired domain could be expressed as apparently independent units still capable of low affinity DNA binding (Fig. 7B). Peptides with extensive amino-terminal deletions of the paired domain, however, lost their ability to bind DNA, thus indicating that the carboxy-terminal region of the paired domain is incapable of either proper folding or DNA binding in the absence of amino-terminal sequences. The two subregions of the paired domain are furthermore unable to bind DNA autonomously, as short oligonucleotides containing individual half sites of the BSAP recognition sequence were not bound by full-length BSAP (T. Czerny, unpubl.). The two subdomains are therefore best viewed as functionally interdependent units of the paired domain.

Most prokaryotic transcription factors bind to DNA by contacting two major grooves, similar to Pax proteins. These prokaryotic factors, however, bind as dimers to palindromic sequences and use helix–turn–helix (HTH)

motifs for sequence recognition (Harrison 1991; Pabo and Sauer 1992). In contrast, Pax proteins bind DNA as monomers and interact with nonpalindromic sequences (Adams et al. 1992). Pax proteins may also contact the minor groove of the DNA helix, as suggested by the following evidence. First, adenines, which are methylated at their N-3 position and thus interfere with protein binding in the minor groove, are preferentially located in the center of BSAP recognition sequences between the two major groove contact sites (Barberis et al. 1989, 1990). Second, the BSAP consensus sequence contains three conserved nucleotides in central positions (13, 16, and 17; Fig. 3). Third, mutations in the center of the H2A-2.2 recognition sequence reduce BSAP binding (Fig. 1). An alternative interpretation of these data is suggested by analogy to the phage 434 repressor–operator system, where the central base pairs in the operator site strongly influence the binding affinity although they are not contacted by the repressor (Koudelka et al. 1987). Similarly, the central base pairs of BSAP recognition sequences may also affect protein binding in an indirect, but sequence-dependent, manner.

Pou domain proteins are another family of eukaryotic transcription factors that utilize a bipartite DNA-binding domain to recognize DNA (Verrijzer et al. 1992 and references therein). The zinc finger proteins are examples of regulatory factors with multiple DNA-binding modules (for review, see Kaptein 1992). Even transcription factors containing two distinct DNA-binding domains exist, as exemplified by the Myb–Ets fusion protein of the avian leukemia virus E26 (Metz and Graf 1991) and by a subclass of Pax proteins that possess a homeo domain in addition to their paired domain (for review, see Gruss and Walther 1992; Noll 1993). Thus, the assortment of different DNA-binding domains or subdomains has led to the evolution of transcription factors with extremely versatile and flexible DNA sequence recognition.

Materials and methods

Oligonucleotides

All DNA probes have been described previously (Barberis et al. 1989; Adams et al. 1992; Kozmik et al. 1992), except for the following oligonucleotides: (1) tcgacTCTCAGCACCGC-ACGATTAGCACCGTCCGCTCAGGCTGTCCCg (e5); (2) tcgacGGGACAGCCTGAGCGAACGGTGCTAATCGTGCG-GTGCTGAGAg (e5); (3) tcgacTGCCAGCCAAGTTCAGTCT-AGTGTAGCAAGCTtcgg (αS-1); (4) tcgacccgaAGCTTGCTAC-ACTAGACTGAACTTGCTGGCag (αS-1); (5) tcgacCCAGG-GATCAGAATTGTGAAGCGTGACCATAGAAAGg (5'Sy2a); (6) tcgacCTTTCTATGGTCACGCTTCACAAATCTGATCC-CTGGg (5'Sy2a); (7) tcgacGGATCTCCTGGGTCAAGGCT-GAATAGACGCAGGGGAG (Sy1); (8) tcgaCTCCCCTGCGT-CTATTGACGCTTGACCCAGGAGATCCg (Sy1); (9) tcgaC-TTTTAGCTGAGGGCACTGAGGCAGAGCGGCCCTAGG (Ie); (10) tcgaCCTAGGGGCCGCTCTGCCTCAGTGCCCTC-AGCTAAAAG (Ie).

The following oligonucleotides were used for PCR-based mutagenesis or cDNA cloning: (11) GTGGCCGTGAATT-CACAAAAACC (G30S); (12) CCGGGAGTAGCCGTCAT-TCACA (P32L); (13) TCCTCTGGGCGACTACATCCGGGA

[R38A]; (14) GCGGAATTCGCGCGATGGAGCAGACGTACG (5' mPax-1); (15) GCGGGATCCCCCGATGGCCCCGGCCAGG (3' mPax-1); (16) GCGAGATCTAAACCTCGAGTTACCACC-CCC (5' mPax-1); (17) GCGAAGCTTGTGGCTCTGTGAGAGCAGAGC (3' mPax-1); (18) GCGGAATTCATAATTTGCGAGCGAAGCTGCC (5' mPax-3); (19) GCGGTCGACTCTCCTTCCAGGAGGAAGTAC (3' mPax-3); (20) GTGGGATCCGAGGCTGCCAACC (5' mPax-6); (21) CGCAAGCTTTACTGTAATCGAGGCCAGT (3' mPax-6).

In vitro mutagenesis of BSAP and Pax-1

Amino-terminal deletions were introduced into the BSAP paired domain by PCR cloning using 5' primers containing the translation initiation codon (GCGAAGCTTCGACCATGGG) linked to the next 7 codons downstream of the amino-terminal breakpoint in the human BSAP sequence (see Fig. 6). Likewise, carboxy-terminal deletions of the paired domain were generated with 3' primers containing the last 7 codons of the human BSAP sequence upstream of the deletion point linked in-frame to a stop codon (ATTCGAAGCG; noncoding strand). Single point mutations were introduced into the BSAP paired domain by oligonucleotide-directed mutagenesis according to the PCR-based method of Merino et al. (1992). Oligonucleotides 11, 12, and 13 were used as PCR primers to introduce the amino acid substitution G30S, P32L, and R38A into human BSAP cDNA, respectively. The chimeric Pax1/BSAP construct was generated by replacing the *EcoRI*–*Bam*HI fragment of the amino-terminal BSAP sequence with a PCR fragment that was obtained from mouse Pax-1 cDNA (Adams et al. 1992) with the primer pair 14/15 and that thus contains codons 1–73 of Pax-1. Oligonucleotides 16 and 17 were used to amplify a *Bgl*III–*Hind*III fragment that consists of mouse Pax-1 sequence from codon 74 to the carboxyl terminus. This PCR fragment was used to replace the corresponding *Bam*HI–*Hind*III fragment of BSAP cDNA in the chimeric BSAP/Pax1 construct. All PCR clones were verified by DNA sequencing.

PCR cloning of Pax-3 and Pax-6 cDNA

Poly(A)⁺ RNA isolated from E10.5 mouse embryos was transcribed into cDNA by Moloney murine leukemia virus reverse transcriptase (SuperScript; BRL), and 100 ng of these transcripts was used for PCR amplification of Pax-3 and Pax-6 cDNA with the oligonucleotide pairs 18/19 and 20/21, respectively. The Pax-3 cDNA was cloned as an *EcoRI*–*Sal*I fragment and the Pax-6 cDNA as a *Bam*HI–*Hind*III fragment into the expression vector pKW10 (Adams et al. 1992). The paired domain sequences of both cDNA clones were checked by DNA sequencing.

In vitro transcription and translation

The different BSAP cDNAs were cloned in the sense orientation downstream of the SP6 promoter of the expression vector pKW10. These templates were linearized with appropriate enzymes and transcribed in vitro into RNA using SP6 RNA polymerase and the cap analog 7mGpppG (Pharmacia). In vitro translation was carried out at 30°C for 40 min in a total volume of 50 μ l containing 20 μ M amino acids minus cysteine, 120 μ Ci of [³⁵S]cysteine (1100 μ Ci/mmol), 2 μ g of RNA, 40 units of RNasin, and 35 μ l of nuclease-treated rabbit reticulocyte lysate (Promega). Protein products were analyzed by electrophoresis on SDS–polyacrylamide gels and quantitated on a PhosphorImager.

Pax protein expression in COP-8 cells

The different Pax cDNAs were inserted downstream of the cytomegalovirus (CMV) enhancer–promoter region of the expression vector pKW10 and transiently transfected into COP-8 cells by the DEAE–dextran method, followed by the preparation of whole cell extracts as described (Adams et al. 1992).

EMSA and methylation interference analysis

Complementary oligonucleotides were annealed and 3' end-labeled with Klenow DNA polymerase and [α -³²P]dCTP. One fmole of labeled DNA probe was incubated together with in vitro-translated BSAP peptide (1–5 μ l of the translation reaction) at room temperature in 20 μ l of a buffer containing 1 μ g of poly[d(I-C)], 10 mM HEPES (pH 7.9), 100 mM KCl, 4% Ficoll, 1 mM EDTA, and 1 mM DTT. Protein–DNA complexes were analyzed on a native 6% polyacrylamide gel (in 0.25 $\times$ TBE) and subsequently quantitated on a PhosphorImager (Molecular Dynamics). Methylation interference analysis was carried out as described (Barberis et al. 1989).

Acknowledgments

We are grateful to M. Noll (University of Zürich) for providing the *Drosophila pox neuro*, *paired* and *gooseberry* cDNAs, to R. Kurzbauer and I. Botto for DNA sequencing, to H. Tkadletz for graphical work, and to L. Ballou and M. Cotton for critical reading of the manuscript. This work was supported by the I.M.P. and, in part, by a grant from the Austrian Industrial Research Promotion Fund. T.C. was the recipient of a fellowship from the Austrian Science Foundation.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Adams, B., P. Dörfler, A. Aguzzi, Z. Kozmik, P. Urbánek, I. Maurer-Fogy and M. Busslinger. 1992. *Pax-5* encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes & Dev.* 6: 1589–1607.
- Baldwin, C.T., C.F. Hoth, J.A. Amos, E.O. da-Silva, and A. Milunsky. 1992. An exonic mutation in the *HuP2* paired domain gene causes Waardenburg's syndrome. *Nature* 355: 637–638.
- Balling, R., U. Deutsch, and P. Gruss. 1988. *undulated*, a mutation affecting the development of the mouse skeleton, has a point mutation in the paired box of *Pax 1*. *Cell* 55: 531–535.
- Barberis, A., G. Superti-Furga, L. Vitelli, I. Kemler, and M. Busslinger. 1989. Developmental and tissue-specific regulation of a novel transcription factor of the sea urchin. *Genes & Dev.* 3: 663–675.
- Barberis, A., K. Widenhorn, L. Vitelli, and M. Busslinger. 1990. A novel B-cell lineage-specific transcription factor present at early but not late stages of differentiation. *Genes & Dev.* 4: 849–859.
- Baumgartner, S., D. Bopp, M. Burri, and M. Noll. 1987. Structure of two genes at the *gooseberry* locus related to the *paired* gene and their spatial expression during *Drosophila* embryogenesis. *Genes & Dev.* 1: 1247–1267.
- Bopp, D., M. Burri, S. Baumgartner, G. Frigerio, and M. Noll. 1986. Conservation of a large protein domain in the segmen-

- tation gene *paired* and in functionally related genes of *Drosophila*. *Cell* **47**: 1033–1040.
- Bopp, D., E. Jamet, S. Baumgartner, M. Burri, and M. Noll. 1989. Isolation of two tissue-specific *Drosophila* paired box genes, *Pox meso* and *Pox neuro*. *EMBO J.* **8**: 3447–3457.
- Chalepakidis, G., R. Fritsch, H. Fickenscher, U. Deutsch, M. Goulding, and P. Gruss. 1991. The molecular basis of the *undulated/Pax-1* mutation. *Cell* **66**: 873–884.
- Epstein, D.J., M. Vekemans, and P. Gros. 1991. *Splotch* (*Sp^{2H}*), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of *Pax-3*. *Cell* **67**: 767–774.
- Epstein, D.J., K.J. Vogan, D.G. Trasler, and P. Gros. 1993. A mutation within intron 3 of the *Pax-3* gene produces aberrantly spliced mRNA transcripts in the *splotch* (*Sp*) mouse mutant. *Proc. Natl. Acad. Sci.* **90**: 532–536.
- Glaser, T., D.S. Walton, and R.L. Maas. 1992. Genomic structure, evolutionary conservation and aniridia mutations in the human *PAX6* gene. *Nature Genet.* **2**: 232–239.
- Goulding, M.D., G. Chalepakidis, U. Deutsch, J.R. Erselius, and P. Gruss. 1991. Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. *EMBO J.* **10**: 1135–1147.
- Gruss, P. and C. Walther. 1992. Pax in development. *Cell* **69**: 719–722.
- Harrison, S.C. 1991. A structural taxonomy of DNA-binding domains. *Nature* **353**: 715–719.
- Hill, R.E., J. Favor, B.L. Hogan, C.C. Ton, G.F. Saunders, I.M. Hanson, J. Prosser, T. Jordan, N.D. Hastie, and V. van Heyningen. 1991. Mouse *small eye* results from mutations in a paired-like homeobox-containing gene. *Nature* **354**: 522–525.
- Hoth, C.F., A. Milunsky, N. Lipsky, R. Sheffer, S.K. Clarren, and C.T. Baldwin. 1993. Mutations in the paired domain of the human *PAX3* gene cause Klein-Waardenburg syndrome (WS-III) as well as Waardenburg syndrome type I (WS-I). *Am. J. Hum. Genet.* **52**: 455–462.
- Kaptein, R. 1992. Zinc-finger structures. *Curr. Opin. Struct. Biol.* **2**: 109–115.
- Kim, J., C. Zwieb, C. Wu, and S. Adhya. 1989. Bending of DNA by gene-regulatory proteins: Construction and use of a DNA bending vector. *Gene* **85**: 15–23.
- Koudelka, G.B., S.C. Harrison, and M. Ptashne. 1987. Effect of non-contacted bases on the affinity of 434 operator for 434 repressor and Cro. *Nature* **326**: 886–888.
- Kozmik, Z., S. Wang, P. Dörfler, B. Adams, and M. Busslinger. 1992. The promoter of the CD19 gene is a target for the B-cell-specific transcription factor BSAP. *Mol. Cell. Biol.* **12**: 2662–2672.
- Kozmik, Z., R. Kurzbaue, P. Dörfler, and M. Busslinger. 1993. Alternative splicing of *Pax-8* gene transcripts generates isoforms with distinct transactivation properties and is regulated in early development. *Mol. Cell. Biol.* (in press).
- Liao, F., S.L. Giannini, and B.K. Birshtein. 1992. A nuclear DNA-binding protein expressed during early stages of B-cell differentiation interacts with diverse segments within and 3' of the IgH chain gene cluster. *J. Immunol.* **148**: 2909–2917.
- Merino, E., J. Osuna, F. Bolívar, and X. Sabarón. 1992. A general, PCR-based method for single and combinatorial oligonucleotide-directed mutagenesis on pUC/M13 vectors. *BioTechniques* **12**: 508–510.
- Metz, T. and T. Graf. 1991. Fusion of the nuclear oncoproteins v-Myb and v-Ets is required for the leukemogenicity of E26 virus. *Cell* **66**: 95–105.
- Morell, R., T.B. Friedman, S. Moeljopawiro, Hartono, Soewito, and J.H.J. Asher. 1992. A frameshift mutation in the HuP2 paired domain of the probable human homolog of murine *Pax-3* is responsible for Waardenburg syndrome type 1 in an Indonesian family. *Hum. Mol. Genet.* **1**: 243–247.
- Noll, M. 1993. Evolution and role of Pax genes. *Curr. Opin. Genet. Dev.* **3**: 595–606.
- Pabo, C.O. and R.T. Sauer. 1992. Transcription factors: Structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**: 1053–1095.
- Rothman, P., S.C. Li, B. Gorham, L. Glimcher, F. Alt, and M. Boothby. 1991. Identification of a conserved lipopolysaccharide-plus-interleukin-4-responsive element located at the promoter of the germ line ϵ transcript. *Mol. Cell. Biol.* **11**: 5551–5561.
- Stapleton, P., A. Weith, P. Urbánek, Z. Kozmik, and M. Busslinger. 1993. Chromosomal localization of seven PAX genes and cloning of a novel family member, PAX-9. *Nature Genet.* **3**: 292–298.
- Tassabehji, M., A.P. Read, V.E. Newton, R. Harris, R. Balling, P. Gruss, and T. Strachan. 1992. Waardenburg's syndrome patients have mutations in the human homologue of the *Pax-3* paired box gene. *Nature* **355**: 635–636.
- Tassabehji, M., A.P. Read, V.E. Newton, M. Patton, P. Gruss, R. Harris, and T. Strachan. 1993. Mutations in the PAX3 gene causing Waardenburg syndrome type 1 and type 2. *Nature Genet.* **3**: 26–30.
- Ton, C.C.T., H. Hirvonen, H. Miwa, M.M. Weil, P. Monaghan, T. Jordan, V. van Heyningen, N.D. Hastie, H. Meijers-Heijboer, M. Drechsler, B. Royer-Pokora, F. Collins, A. Swaroop, L.C. Strong, and G.F. Saunders. 1991. Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region. *Cell* **67**: 1059–1074.
- Treisman, J., E. Harris, and C. Desplan. 1991. The paired box encodes a second DNA-binding domain in the paired homeo domain protein. *Genes & Dev.* **5**: 594–604.
- Verrijzer, P.C., M.J. Alkema, W.W. van Weperen, H.C. Van Leeuwen, M.J.J. Strating, and P.C. van der Vliet. 1992. The DNA binding specificity of the bipartite POU domain and its subdomains. *EMBO J.* **11**: 4993–5003.
- Walther, C., J.L. Guenet, D. Simon, U. Deutsch, B. Jostes, M.D. Goulding, D. Plachov, R. Balling, and P. Gruss. 1991. Pax: A murine multigene family of paired box-containing genes. *Genomics* **11**: 424–434.
- Waters, S.J., K.U. Saikh, and J. Stavnezer. 1989. A B-cell-specific nuclear protein that bind to DNA sites 5' to immunoglobulin α tandem repeats is regulated during differentiation. *Mol. Cell. Biol.* **9**: 5594–5601.
- Williams, M. and N. Maizels. 1991. LR1, a lipopolysaccharide-responsive factor with binding sites in the immunoglobulin switch regions and heavy-chain enhancer. *Genes & Dev.* **5**: 2353–2361.
- Zannini, M., H. Francis-Lang, D. Plachov, and R. Di Lauro. 1992. Pax-8, a paired domain-containing protein, binds to a sequence overlapping the recognition site of a homeodomain and activates transcription from two thyroid-specific promoters. *Mol. Cell. Biol.* **12**: 4230–4241.



DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site.

T Czerny, G Schaffner and M Busslinger

Genes Dev. 1993, 7:

Access the most recent version at doi:[10.1101/gad.7.10.2048](https://doi.org/10.1101/gad.7.10.2048)

References

This article cites 38 articles, 12 of which can be accessed free at:
<http://genesdev.cshlp.org/content/7/10/2048.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

