Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis

Ben Adams,[1,2] Petra Dörfler,[1] Adriano Aguzzi,[1] Zbynek Kozmik, Pavel Urbánek, Ingrid Maurer-Fogy,[3] and Meinrad Busslinger[4]

Institute of Molecular Pathology, A-1030 Vienna, Austria; [3]Ernst Boehringer Institute, A-1120 Vienna, Austria

BSAP has been identified previously as a transcription factor that is expressed at early, but not late, stages of B-cell differentiation. Biochemical purification and cDNA cloning has now revealed that BSAP belongs to the family of paired domain proteins. BSAP is encoded by the *Pax-5* gene and has been highly conserved between human and mouse. An intact paired domain was shown to be both necessary and sufficient for DNA binding of BSAP. Binding studies with several BSAP recognition sequences demonstrated that the sequence specificity of BSAP differs from that of the distantly related paired domain protein Pax-1. During embryogenesis, the BSAP gene is transiently expressed in the mesencephalon and spinal cord with a spatial and temporal expression pattern that is distinct from that of other *Pax* genes in the developing central nervous system (CNS). Later, the expression of the BSAP gene shifts to the fetal liver where it correlates with the onset of B lymphopoiesis. BSAP expression persists in B lymphocytes and is also seen in the testis of the adult mouse. All of this evidence indicates that the transcription factor BSAP may not only play an important role in B-cell differentiation but also in neural development and spermatogenesis.

[*Key Words*: Transcription factor BSAP; *Pax-5*; paired box recognition sequences; B-cell differentiation; CNS development; testis]

A family of regulatory proteins that are involved in the control of segmentation and specification of cell fate during *Drosophila* embryogenesis share the highly conserved motif known as the paired box. This domain of 128 amino acids was identified as a second homology region in the homeo box-containing proteins of the *Drosophila paired* and *gooseberry* genes (Bopp et al. 1986; Baumgartner et al. 1987). The products of the tissue-specific genes *pox meso* and *pox neuro* were subsequently shown to contain only a paired domain and to be localized in the nucleus (Bopp et al. 1989), suggesting that paired domain proteins are transcription factors. Recently, this hypothesis has been verified by demonstrating that the paired box encodes a novel DNA-binding motif (Goulding et al. 1991; Treisman et al. 1991) and that the murine paired domain protein Pax-1 is a transcriptional activator (Chalepakis et al. 1991).

The paired domain proteins of vertebrates are encoded by a multigene family that has been conserved in evolution. To date, eight paired box genes, *Pax-1–Pax-8*, have been isolated from the mouse genome (Walther et al.

1991 and references therein), two corresponding genes from the zebrafish (Krauss et al. 1991a,b), and four paired domain genes from human DNA (Burri et al. 1989; Ton et al. 1991). The paired domain proteins of the different species can be grouped into six different classes according to their sequence similarity in the paired box (Walther et al. 1991). The mouse *Pax* genes are expressed along the anteroposterior axis in the neural tube of the embryo (*Pax-2*, Nornes et al. 1990; *Pax-3*, Goulding et al. 1991; *Pax-6*, Walther and Gruss 1991; *Pax-7*, Jostes et al. 1991; *Pax-8*, Plachov et al. 1990), in the developing excretory system (*Pax-2*, Dressler et al. 1990; *Pax-8*, Plachov et al. 1990), or in the developing vertebral column (*Pax-1*, Deutsch et al. 1988). Mutations in the *Pax-1*, *Pax-3*, and *Pax-6* genes have been associated with the three mouse developmental mutants *undulated* (Balling et al. 1988), *splotch* (Epstein et al. 1991), and *small eye* (Hill et al. 1991), respectively. Moreover, genetic lesions in the human *Pax-3* and *Pax-6* genes were shown to cause Waardenburg's syndrome (Baldwin et al. 1992; Tassabehji et al. 1992) and aniridia (Ton et al. 1991). All of this evidence strongly implicates the Pax proteins as important regulators in early development.

The target genes and, hence, the function of the vertebrate Pax proteins are still unknown mainly because the vertebrate *Pax* genes have been cloned and identified

---

Adams et al.

solely on the basis of their sequence homology with *Drosophila* paired box genes (Walther et al. 1991 and references therein). In this report we describe the first biochemical purification of a paired domain protein, that is, the transcription factor BSAP, which was originally identified as a mammalian homolog of the sea urchin protein TSAP (tissue-specific activator protein). The transcription factor TSAP was shown to be responsible for the developmental and tissue-specific regulation of two nonallelic pairs of late histone *H2A-2* and *H2B-2* genes of the sea urchin. TSAP recognizes its four binding sites in these histone gene promoters with equal affinity, and yet these sites differ considerably in sequence, allowing the definition of only a degenerate consensus sequence (Barberis et al. 1989). Subsequently, a mammalian DNA-binding activity was identified that interacts with these four TSAP-binding sites in a manner indistinguishable from that of the sea urchin protein. Within the hematopoietic system, this mammalian factor is exclusively expressed in the B-lymphoid lineage from the pro-B cell up to the mature B-cell stage. However, its activity is undetectable in terminally differentiated plasma cells. This factor was shown to stimulate B-cell-specific transcription from a promoter containing a TSAP-binding site. Hence, we referred to it as BSAP (B-cell lineage-specific activator protein; Barberis et al. 1990). Recently, we have demonstrated that BSAP is involved in the regulation of the *CD19* gene, which codes for a B-lymphoid-specific transmembrane receptor involved in signal transduction (Kozmik et al. 1992).

Biochemical purification and cDNA cloning revealed that BSAP is a member of the paired domain protein family. It is encoded by the *Pax-5* gene and has been highly conserved between human and mouse. Mutational analysis identified the paired box as the only DNA-binding domain of BSAP. The distantly related Pax-1 protein was shown to recognize known BSAP-binding sites, although with lower affinities than BSAP. Expression of the BSAP/*Pax-5* gene was not only detected in B-lymphoid tissues, but also in specific regions of the developing central nervous system (CNS) and in testis of the adult mouse.

Results

Purification of the human BSAP protein

The purification of BSAP from the human B-cell line BJA-B is summarized in Figure 1A. BSAP-binding activity was monitored by electrophoretic mobility-shift assay (EMSA) using the TSAP-binding site of the sea urchin histone *H2A-2.2* gene as a probe. Quantitation of the DNA-binding activity according to Calzone et al. (1988) revealed that BSAP had to be purified by a factor of 5×10^4 starting from BJA-B nuclear extracts. Like many other transcripton factors, BSAP was found to be glycosylated (Jackson and Tjian 1988) and could therefore be enriched 250-fold by chromatography on a wheat germ lectin column. The eluted BSAP fraction was applied to a DNA affinity column consisting of multimerized H2B-
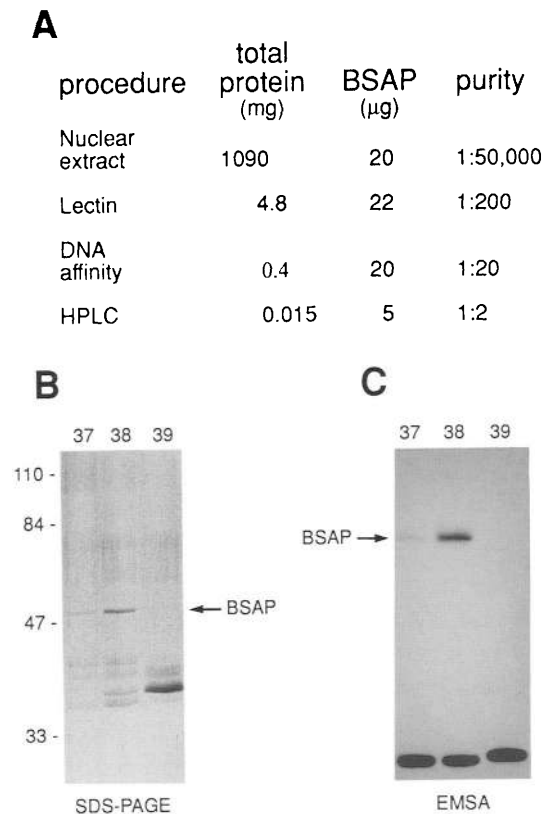

procedure	total protein (mg)	BSAP (µg)	purity
Nuclear extract	1090	20	1:50,000
Lectin	4.8	22	1:200
DNA affinity	0.4	20	1:20
HPLC	0.015	5	1:2

Figure 1. Biochemical purification of human BSAP. (*A*) Purification from 50 liters of BJA-B cells. Values are given for the nuclear extract, the eluate of the wheat germ lectin column, the eluate of the DNA affinity column, and fraction 38 of the HPLC reverse-phase column (total of 66 fractions). The amount of total protein was determined by the method of Bradford (1976) by using a BSA standard. The amount of BSAP was measured by quantitative gel-shift analysis according to Calzone et al. (1988), assuming that the 50-kD BSAP peptide interacts as a monomer with its recognition sequence (see Fig. 6A). The protein amounts in HPLC fraction 38 were estimated from the preparative SDS–polyacrylamide gel used as a last purification step before protein sequencing. (*B*) Analytical SDS–protein gel. An aliquot of fractions 37–39 of the HPLC reverse-phase column was analyzed by electrophoresis on a 10% SDS–polyacrylamide gel, and proteins were visualized by silver staining (Wray et al. 1981). The position and molecular mass (kD) of marker proteins are indicated to the *left*. (*C*) EMSA of HPLC fractions 37–39 with the *H2A-2.2* TSAP-binding site as probe (oligonucleotide 1 of Fig. 7D).

2.2 TSAP recognition sequences coupled to Sepharose. Chromatography on this DNA affinity resin resulted in an additional 10-fold purification of BSAP. However, repeated chromatography was unsuccessful because nonionic detergents, which are required to protect BSAP-binding activity in dilution, reduced the affinity for specific recognition sequences by more than one order of magnitude (B. Adams, unpubl.). Further purification of BSAP was achieved by high-performance liquid chromatography (HPLC) on a C8 reverse-phase column. A 50-kD polypeptide was eluted predominantly in fraction 38

and, to a lesser extent, in fraction 37, as revealed by SDS-PAGE and silver staining (Fig. 1B). The EMSA experiment of Figure 1C showed that BSAP-binding activity was present in these two fractions in a ratio that closely paralleled the abundance of the 50-kD peptide (Fig. 1C). This observation and the previous determination of a molecular mass of 50 kD for BSAP by UV crosslinking (Barberis et al. 1990) strongly suggested that the purified 50-kD peptide corresponded to BSAP.

BSAP is a paired domain protein

The 50-kD protein of the HPLC fraction 38 was isolated by SDS-PAGE and then subjected to trypsin digestion followed by purification and microsequencing of individual peptides. Reliable amino acid sequence data were obtained for the two peptides shown in Figure 2A. A computer search of the SwissProt data base revealed that the sequence of peptide 1 is present in the middle of the paired box of all *Drosophila* and mouse paired domain proteins known so far (for review, see Walther et al. 1991). The second peptide was only found in the carboxy-terminal half of the paired domain of Pax-2 (Dressler et al. 1990) and Pax-8 (Plachov et al. 1990), suggesting that BSAP belongs to the same subclass of Pax proteins. The limited amino acid sequence information suggested, therefore, that BSAP is a member of the growing family of paired domain proteins. This hypothesis was verified by the antibody experiment of Figure 2B. As the highly conserved paired domain has recently been

shown to specify a DNA-binding function (Chalepakis et al. 1991; Goulding et al. 1991; Treisman et al. 1991), we reasoned that polyclonal antibodies against the *Drosophila* Paired (Prd) protein might recognize conserved epitopes in the putative paired domain of BSAP and thus prevent DNA binding. As shown in Figure 2B, a Prd antiserum, but not a preimmune serum, specifically blocked the DNA-binding activity of BSAP, thus demonstrating that BSAP is a paired domain protein.

Cloning and identification of human BSAP cDNA

Degenerate oligonucleotides derived from the two BSAP peptides were used to isolate part of the BSAP paired domain by polymerase chain reaction (PCR) cloning from human BJA-B and mouse 70Z/3 cDNA, respectively. As shown by Northern blot analysis (Fig. 3A), the human PCR probe detected a transcript of ~10 kb in B cells (BJA-B) and pre-B cells (REH), both of which were shown previously to be positive for BSAP-binding activity (Barberis et al. 1990). No hybridization signal was observed with the plasma cell line RPMI 8226 and with HeLa cells in agreement with the absence of BSAP-binding activity in these cells. This analysis therefore indicates that BSAP is encoded by a ~10-kb-long mRNA.

Both the human and mouse PCR probes were used to screen BJA-B and 70Z/3 cDNA libraries, respectively. A human clone with a 3277-bp cDNA insert (hBSAP-1) and a mouse clone with a 926-bp cDNA insert (mBSAP-1) were isolated (Fig. 3D). As both cDNA clones covered only part of the 10-kb BSAP mRNA, we then asked whether they contained at least the entire protein-coding region. For this purpose, the two cDNA sequences were cloned into the expression plasmid pKW10. This vector drives expression of cDNA inserts from the strong cytomegalovirus (CMV) enhancer–promoter region and, furthermore, is able to replicate to high levels in polyoma large T-expressing COP-8 cells (Tyndall et al. 1981). Sense and antisense expression constructs were transfected into COP-8 cells, and whole-cell extracts were prepared for EMSA with the H2A-2.2 TSAP-binding probe. As shown in Figure 3B, the hBSAP-1 cDNA in the sense, but not in the antisense, orientation gave rise to a protein–DNA complex that comigrated with the endogenous BSAP activity of BJA-B extracts. Moreover, a single C → A mutation at the invariant position within the TSAP-binding site (Barberis et al. 1990) diminished binding of the endogenous (BJA-B) as well as exogenous (COP-8) activity to the same degree. We therefore concluded that the human cDNA clone hBSAP-1 contains the entire coding region for BSAP. In contrast, the mBSAP-1 cDNA directed the synthesis of only a partial BSAP polypeptide (P. Dörfler, data not shown). The complete coding sequence of mouse BSAP mRNA (mBSAP-2) was subsequently isolated from 70Z/3 cDNA by PCR cloning (Fig. 3D).

Ultimate identification of the cDNA clones was achieved with polyclonal antibodies raised against the paired domain (amino acids 17–145) and the carboxy-terminal region (189–391) of the protein that is encoded
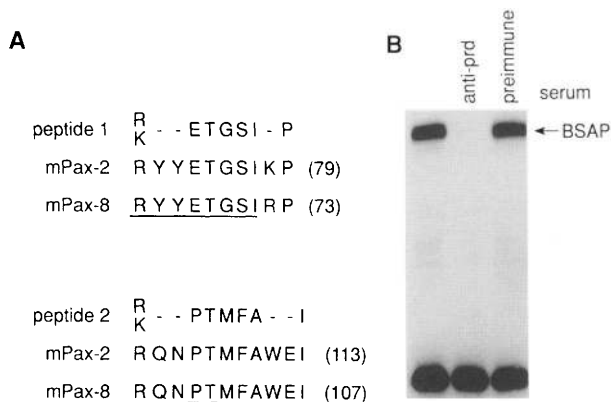

Figure 2. Identification of BSAP as a paired domain protein. (A) BSAP peptide sequences. The two peptide sequences obtained by the microsequencing of purified BSAP are aligned with the corresponding paired domain sequences of murine Pax-2 (Dressler et al. 1990) and Pax-8 (Plachov et al. 1990). The basic amino acid (R or K) that lies amino-terminal to the trypsin cleavage site is shown as the first residue. Dashes denote those positions where no amino acid residue could be unequivocally determined. The peptide sequences used to design degenerate PCR primers are underlined. (B) EMSA. Polyclonal rabbit anti-Paired (Prd) or preimmune serum (1 μl of a fivefold dilution) was added to the DNA-binding reaction containing partially purified BSAP (eluate of the lectin column) and the *H2A-2.2* TSAP-binding probe (oligonucleotide 1 of Fig. 7D).
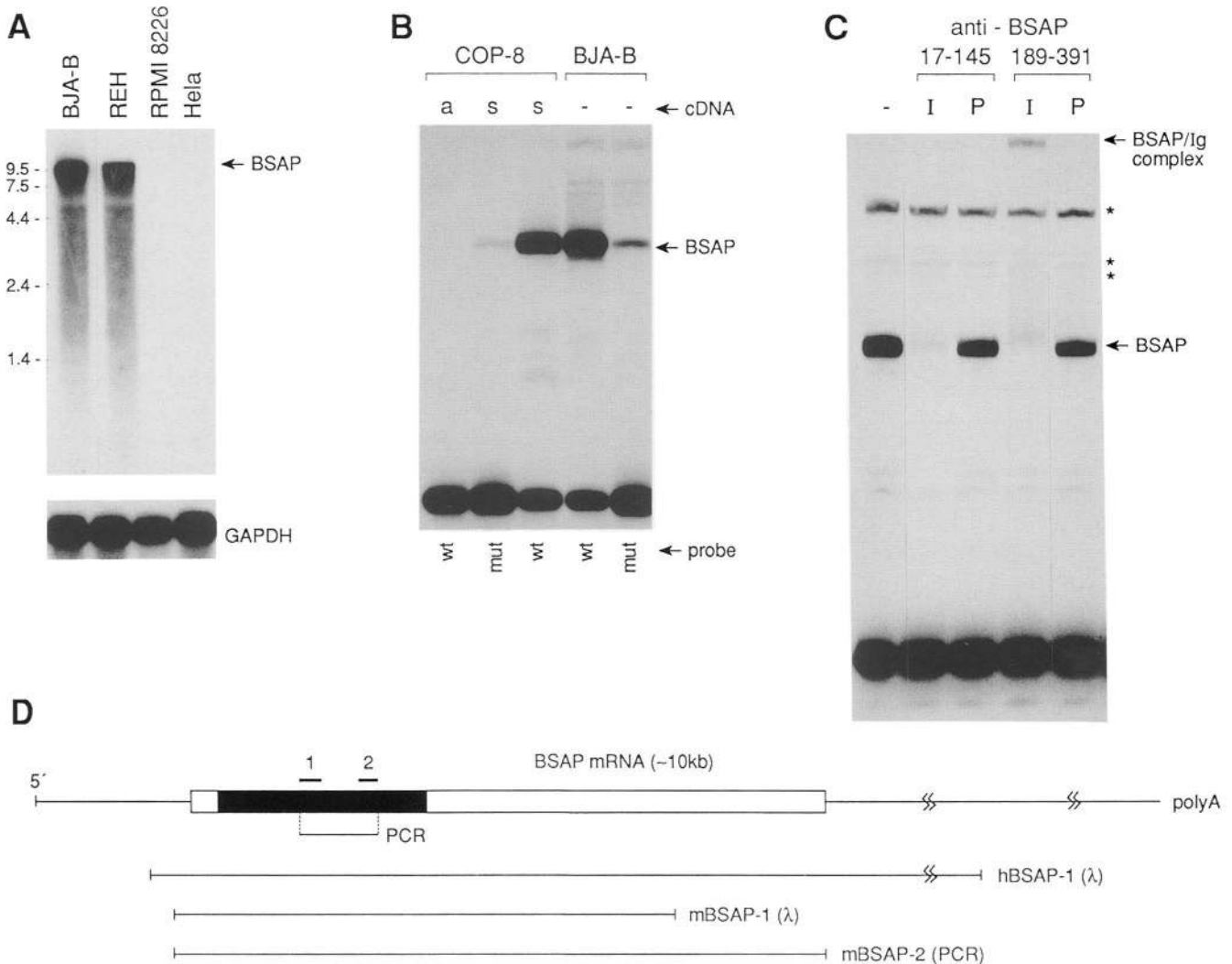
Adams et al.

Figure 3. Cloning and identification of human and mouse BSAP cDNA. (A) Northern blot analysis. Poly(A)$^+$ RNA (~5 μg) of the human cell lines BJA-B, REH, RPMI 8226, and HeLa was analyzed by Northern blot hybridization with the BSAP paired domain PCR probe. The positions and sizes (kb) of marker RNAs (BRL) are indicated to the *left*. The same filter was rehybridized with a rat glyceraldehyde phosphate dehydrogenase (GAPDH) cDNA probe. (B) Expression of BSAP protein in COP-8 cells. hBSAP-1 cDNA was cloned in the sense (s) or antisense (a) orientation into the expression vector pKW10, followed by transient transfection into COP-8 cells and preparation of whole-cell extracts. These COP-8 extracts and nuclear extracts of BJA-B cells were analyzed by EMSA with a wild-type (wt) or mutant (mut) *H2A-2.2* TSAP-binding probe (oligonucleotides 1 and 2 of Fig. 7D). (C) Interaction of specific antibodies with endogenous BSAP. Preimmune serum (P) or polyclonal rabbit antiserum (I; 1 μl of a fivefold dilution) raised against a BSAP peptide containing either amino acids 17–145 or 189–391 was added together with BJA-B nuclear extract and the *H2A-2.2* TSAP probe to the DNA-binding reaction (20 μl) before EMSA. Unspecific protein–DNA complexes are indicated by asterisks (*). (D) Schematic diagram of BSAP cDNA clones. The position of the mRNA 5' end relative to the BSAP-coding sequence (boxed) was determined by PCR amplification of 5' cDNA ends (P. Urbánek, unpubl.). The paired domain is marked by a solid box. The origin of the cDNA clones is indicated by λ (λgt10 cDNA library) or PCR (PCR cloning).

by the cDNA clone hBSAP-1. Both antibodies interacted specifically with the endogenous BSAP activity of BJA-B cells, as shown by EMSA (Fig. 3C). The anti-paired domain serum, but not the preimmune serum, prevented DNA binding of BSAP consistent with the result obtained with antibodies raised against the entire *Drosophila* Prd protein (Fig. 2B). In contrast, antibodies raised against carboxy-terminal sequences did not interfere with DNA binding but, instead, shifted the BSAP complex to a more slowly migrating position in the gel. Both

BSAP antisera did not affect Oct-1 and Oct-2 binding in control experiments (P. Dörfler, unpubl.). Taken together, these data therefore demonstrate that the hBSAP-1 cDNA clone codes for authentic BSAP.

BSAP is evolutionarily conserved and encoded by the Pax-5 *gene*

The DNA and deduced amino acid sequences of clone hBSAP-1 are shown in Figure 4. The first translation ini-

tiation site TCAAAATGG in this sequence fits well with the consensus sequence CCPuCCATGG at the two most critical positions for optimal initiation of translation (Kozak 1991) and, in addition, is immediately preceded by an in-frame stop codon. This initiation site is followed by an open reading frame (ORF) encoding a protein of 391 amino acids with a calculated molecular mass of 42.1 kD. The discrepancy between the calculated and apparent (50-kD) molecular mass of BSAP is most likely caused by post-translational modification such as O-linked glycosylation (see Fig. 1) and/or by anomalous migration of BSAP owing to a relatively high proline content (Fig. 4). Rapid amplification of cDNA ends (RACE) analysis (Frohman et al. 1988) revealed that ~100–200 nucleotides up to the correct 5' end of BSAP mRNA are missing in the cDNA sequence shown in Figure 4 (P. Urbánek, unpubl.). Moreover, the 3' sequences of the cDNA insert neither contain a polyadenylation signal nor a poly(A) tail, suggesting that only 2 kb of the ~8.6-kb-long trailer of BSAP have been cloned.

BSAP exhibits the following notable features. It contains a paired domain close to the amino terminus (amino acids 16–143) and the conserved octapeptide YSISGILG (amino acids 179–186), which deviates only at two amino acid positions from the sequence found in most other paired domain proteins (Burri et al. 1989; Jostes et al. 1991). BSAP lacks a classical paired-type homeo domain as it is found in a subclass of Pax proteins. However, it contains a region of 23 amino acids (229–251) that is homologous to the amino-terminal half of the homeo domain of Pax-3, Pax-6, and Pax-7, both with regard to its sequence (13–15 identical amino acids) and its position downstream of the paired domain. A similar homology has recently been identified in Pax-2 and Pax-8 (Krauss et al. 1991a; Walther and Gruss 1991). It remains to be seen whether this sequence conservation is functionally significant or represents only an evolutionary remnant of the homeo box in these proteins. BSAP also possesses two regions with a high proportion (48% and 35%) of serine and threonine residues and ends in a sequence that is both rich in serine/threonine as well as proline (39% in total; Fig. 4B). These features are reminiscent of the serine/threonine-rich *trans*-activation domains of Oct-1 and Oct-2 (Tanaka and Herr 1990).

Sequence analysis of the mouse BSAP cDNA clone showed that the human and mouse BSAP proteins differ only by three amino acid exchanges, whereas their coding sequences have diverged by 7.7% at the DNA level owing to silent site substitutions. Pax-6, for which the entire human and mouse sequences are also known, differs in a single amino acid position between the two species (Ton et al. 1991; Walther and Gruss 1991). These two findings strongly suggest that Pax proteins are subject to high evolutionary constraints even outside of their paired domains.

After completion of our biochemical purification and cDNA cloning of BSAP, Walther et al. (1991) described the isolation of the *Pax-5* gene from a mouse genomic library by cross-hybridization with a Pax-1 paired box DNA probe (Walther et al. 1991). However, no expression data were provided for this gene; furthermore, no corresponding cDNA could be isolated from mouse embryonic cDNA libraries. Hence, only the two exons coding for the paired domain of murine *Pax-5* were identified. Comparison with our mouse BSAP cDNA sequence revealed that the nucleotide sequences of these two exons of Pax-5 are identical with the corresponding BSAP cDNA sequence. We conclude, therefore, that BSAP is encoded by the *Pax-5* gene. Walther et al. (1991) have already demonstrated that *Pax-2*, *Pax-5* (BSAP), and *Pax-8* are closely related to one another in their paired domains, thus constituting a subclass of paired box proteins. As shown in Figure 5, the homology between these three proteins even extends into their central and carboxy-terminal sequences. However, BSAP is the shortest of all three proteins as it lacks a stretch of 23 amino acids of Pax-2 and a domain of 69 amino acids of Pax-8, both of which may be encoded by separate exons in their respective gene.

An intact paired domain is both necessary and sufficient for DNA binding of BSAP

Mutational analysis of the *Drosophila* Prd and mouse Pax-1 proteins has demonstrated recently that the paired domain encodes a novel DNA-binding motif (Chalepakis et al. 1991; Treisman et al. 1991). To test whether the paired domain alone is sufficient for DNA binding of BSAP, we have subjected the BSAP-coding sequences to a carboxy-terminal deletion analysis, expressed the truncated proteins in COP-8 cells, and analyzed their DNA-binding potential by EMSA (Fig. 6A). Deletion of carboxy-terminal sequences up to amino acid position 163 had no effect on the DNA binding of BSAP, demonstrating that the octapeptide and the homeo box homology region are dispensable for the interaction of BSAP with DNA. Further deletion up to position 107 in the paired domain completely abolished DNA binding. The same result was obtained with an internal deletion mutant lacking amino acids 122–186 (Fig. 6A). Western blot analysis with anti-paired domain antibodies demonstrated that the two nonbinding polypeptides were expressed in COP-8 cells as stable proteins and with efficiency similar to the DNA-binding polypeptides (Fig. 6A). We observed, however, that the three BSAP polypeptides terminating at amino acids 111, 116, and 121 in the paired domain were too unstable in transfected COP-8 cells to be detected by Western blot analysis (P. Dörfler, unpubl.). These three proteins, once synthesized by in vitro translation in rabbit reticulocyte lysates, also completely fail to bind to DNA (Fig. 6B). Taken together, these results clearly demonstrate that an intact paired domain is both necessary and sufficient for DNA binding of BSAP.

As all information for high affinity DNA binding of BSAP is contained within the paired domain, we then investigated, using the method of Hope and Struhl (1987), whether BSAP binds as a monomer or dimer to its recognition sequence. Full-length BSAP was coexpressed in COP-8 cells together with a truncated polypeptide

A

```
   1  AAAAAAAAGAAAAAAAAAGGCACAAAAAAGTGGAAACTTTTCCCTGTCCATTCCATCAAGTCCTGAAAAATCAAAATGGATTTAGAGAAA
                                                                                  M  D  L  E  K

  91  AATTATCCGACTCCTCGGACCAGCAGGACAGGACATGGAGGAGTGAATCAGCTTGGGGGGGTTTTTGTGAATGGACGGCCACTCCCGGAT
   6  N  Y  P  T  P  R  T  S  R  T │G  H  G  G  V  N  Q  L  G  G  V  F  V  N  G  R  P  L  P  D

 181  GTAGTCCGCCAGAGGATAGTGGAACTTGCTCATCAAGGTGTCAGGCCCTGCGACATCTCCAGGCAGCTTCGGGTCAGCCATGGTTGTGTC
  36  V  V  R  Q  R  I  V  E  L  A  H  Q  G  V  R  P  C  D  I  S  R  Q  L  R  V  S  H  G  C  V

 271  AGCAAAATTCTTGGCAGGTATTATGAGACAGGAAGCATCAAGCCTGGGGTAATTGGAGGATCCAAACCAAAGGTCGCCACACCCAAAGTG
  66  S  K  I  L  G  [R] Y  Y  [E][T][G][S][I] K [P] G  V  I  G  G  S  K  P  K  V  A  T  P  K  V

 361  GTGGAAAAAATCGCTGAATATAAACGCCAAAATCCCACCATGTTTGCCTGGGAGATCAGGGACCGGCTGCTGGCAGAGCGGGTGTGTGAC
  96  V  E  K  I  A  E  Y  K  [R] Q  N  [P][T][M][F][A] W  E  [I] R  D  R  L  L  A  E  R  V  C  D

 451  AATGACACCGTGCCTAGCGTCAGTTCCATCAACAGGATCATCCGGACAAAAGTA│CAGCAGCCACCCAACCAACCAGTCCCAGCTTCCAGT
 126  N  D  T  V  P  S  V  S  S  I  N  R  I  I  R  T  K  V │Q  Q  P  P  N  Q  P  V  P  A  S  S

 541  CACAGCATAGTGTCCACTGGCTCCGTGACGCAGGTGTCCTCGGTGAGCACGGATTCGGCCGGCTCGTCGTACTCCATCAGCGGCATCCTG
 156  H  S  I  V  S  T  G  S  V  T  Q  V  S  S  V  S  T  D  S  A  G  S  S  Y  S  I  S  G  I  L

 631  GGCATCACGTCCCCCAGCGCCGACACCAACAAGCGCAAGAGAGACGAAGGTATTCAGGAGTCTCCGGTGCCGAACGGCCACTCGCTTCCG
 186  G  I  T  S  P  S  A  D  T  N  K  R  K  R  D  E  G  I  Q  E  S  P  V  P  N  G  H  S  L  P

 721  GGCAGAGACTTCCTCCGGAAGCAGATGCGGGGAGACTTGTTCACACAGCAGCAGCTGGAGGTGCTGGACCGCGTGTTTGAGAGGCAGCAC
 216  G  R  D  F  L  R  K  Q  M  R  G  D  L  F  T  Q  Q  Q  L  E  V  L  D  R  V  F  E  R  Q  H

 811  TACTCAGACATCTTCACCACCACAGAGCCCATCAAGCCCGAGCAGACCACAGAGTATTCAGCCATGGCCTCGCTGGCTGGTGGGCTGGAC
 246  Y  S  D  I  F  T  T  T  E  P  I  K  P  E  Q  T  T  E  Y  S  A  M  A  S  L  A  G  G  L  D

 901  GACATGAAGGCCAATCTGGCCAGCCCCACCCCTGCTGACATCGGGAGCAGTGTGCCAGGCCCGCAGTCCTACCCCATTGTGACAGGCCGT
 276  D  M  K  A  N  L  A  S  P  T  P  A  D  I  G  S  S  V  P  G  P  Q  S  Y  P  I  V  T  G  R

 991  GACTTGGCGAGCACGACCCTCCCCGGGTACCCTCCACACGTCCCCCCCGCTGGACAGGGCAGCTACTCAGCACCGACGCTGACAGGGATG
 306  D  L  A  S  T  T  L  P  G  Y  P  P  H  V  P  P  A  G  Q  G  S  Y  S  A  P  T  L  T  G  M

1081  GTGCCTGGGAGTGAGTTTTCCGGGAGTCCCTACAGCCACCCTCAGTATTCCTCGTACAACGACTCCTGGAGGTTCCCCAACCCGGGGCTG
 336  V  P  G  S  E  F  S  G  S  P  Y  S  H  P  Q  Y  S  S  Y  N  D  S  W  R  F  P  N  P  G  L

1171  CTTGGCTCCCCCTACTATTATAGCGCTGCCGCCCGAGGAGCCGCCCCACCTGCAGCCGCCACTGCCTATGACCGTCACTGACCCTTGGAG
 366  L  G  S  P  Y  Y  Y  S  A  A  A  R  G  A  A  P  P  A  A  A  T  A  Y  D  R  H  *

1261  CCAGGCGGGCACCAAACACTGATGGCACCTATTGAGGGTGACAGCCACCCAGCCCTCCTGAAGATAGCCAGAGAGCCCATGAGACCGTCC
1351  CCCAGCATCCCCCACTTGCCTGAAGCTCCCCTCTTCCTCTCTTCCTCCAGGGACTCTGGGGCCCTTTGGTGGGGCCGTTGGACTTCTGGA
1441  TGCTTGTCTATTTCTAAAAGCCAATCTATGAGCTTCTCCCGATGGCCACTGGGTCTCTGCAAACCAATAGACTGTCCTGCAAATAACCGC
1531  AGCCCCAGCCCAGCCTGCCTGTCCTCCAGCTGTCTGACTATCCATCCATCATAACCACCCCAGCCTGGGAAGGAGAGCTTGCTTTTGTTG
1621  CTTCAGCAGCACCCATGTAAATACCTTCTTGCTTTTCTGTGGGCCTGAAGGTCCGACTGAGAAGACTGCTCCACCCATGATGCATCTCGC
1711  ACTCTTGGTGCATCACCGGACATCTTAGACCTATGGCAGAGCATCCTCTCTGCCCTGGGTGACCCTGGCAGGTGCGCTCAGAGCTGTCCT
1801  CAAGATGGAGGATGCTGCCCTTGGGCCCCAGCCTCCTGCTCATCCCTCCTTCTTTAGTATCTTTACGAGGAGTCTCACTGGGCTGGTTGT
1891  GCTGCAGGCTCCCCCTGAGGCCCCTCTCCAAGAGGAGCACACTTTGGGGAGATGTCCTGGTTTCCTGCCTCCATTTCTCTGGGACCGATG
1981  CAGTATCAGCAGCTCTTTTCCAGATCAAAGAACTCAAAGAAAACTGTCTGGGAGATTCCTCAGCTACTTTTCCGAAGCAGAATGTCATCC
2071  GAGGTATTGATTACATTGTGGACTTTGAATGTGAGGGCTGGATGGGACGCAGGAGCATCTGATCCCAGCCAAGGAGGGGCCTGAGGCT
2161  CTCCCTACTCCCTCAGCCCCTGCACCGGTGTTTTCTGAGGCATGCCCAGGTTCAGGTCACTTCGGACACCTGCCATGGACACTTCACCCA
2251  CCCTCCAGGACCCCAGCAAGTGGATTCTGGGCAAGCCTGTTCCGGTGATGTAGACAATAATTAACACAGAGGACTTTCCCCCACACCCAG
2341  ATCACAAACAGCCTACAGCCAGAACTTCTGAGCATCCTCTCGGGGCAGACCCTCCCCGTCCTCGTGGAGCTTAGCAGGCAGCTGGGCATG
2431  GAGGTGCTGGGGCTGGGGCAGATGCCTAATTTCGCACAATGCATGCCCACCTGTTGATCTAAGGGGCCGCGATGGTCAGGGCCACGGCCA
2521  AGGGCCACGGGAACTTGGAGAGGGAGCTTGGAGAACTCACTGTGGGCTAGGGTGGTCAGAGGAAGCCAGCAGGGAAGATCTGGGGGACAG
2611  AGGAAGGCCTCCTGAGGGAGGGGGCAGGAGAGCAGTGAGGAGCTGCTGTGTGACCTGGGAGTGATTTTGACATGGGGGTGCCAGGTGCCAT
2701  CATCTCTTTACCTGGGGCCTTAATTCCTTGCATAGTCTCTCTTGTCAAGTCAGAACAGCCAGGTAGAGCCCTTGTCCAAACCTGGGCTGA
2791  ATGACAGTGATGAGAGGGGGCTTGGCCTTCTTAGGTGACAATGTCCCCCATATCTGTATGTCACCAGGATGGCAGAGAGCCAGGGCAGAG
2881  AGAGACTGGACTTGGGATCAGCAGGCCAGGCAGGTCTTGTCCTGGTCCTGGCCACATGTCTTTGCTGTGGGACCTCAGACAAAACCCTGC
2971  ACCTCTTTGAGCCTTGGCTGCCTTGGTGCAGCAGGGTCATCTGTAGGGCCACCCCACAGCTCTTTCCTTCCCCTCCTCTCTCCAGGGAGC
3061  CGGGGCTGTGAGAGGATCATCTGGGGCAGGCCCTCCACTTCCAAGCAAGCAGATGGGGGTGGGCACCTGAGGCCCAATAATATTTGGACC
3151  AAGTGGGAAACAAGAACACTCGGAGGGGCGGGAATCAGAAGAGCCTGGAAAAAGACCTAGCCCAACTTCCCTTGTGGGAAACTGAGGCCC
3241  AGCTTGGGGAAGGCCAGGACCATGCAGGGAGAAAAAG  3277
```
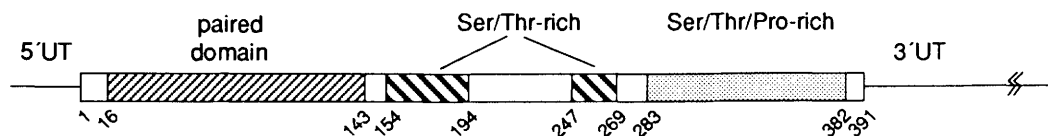
B

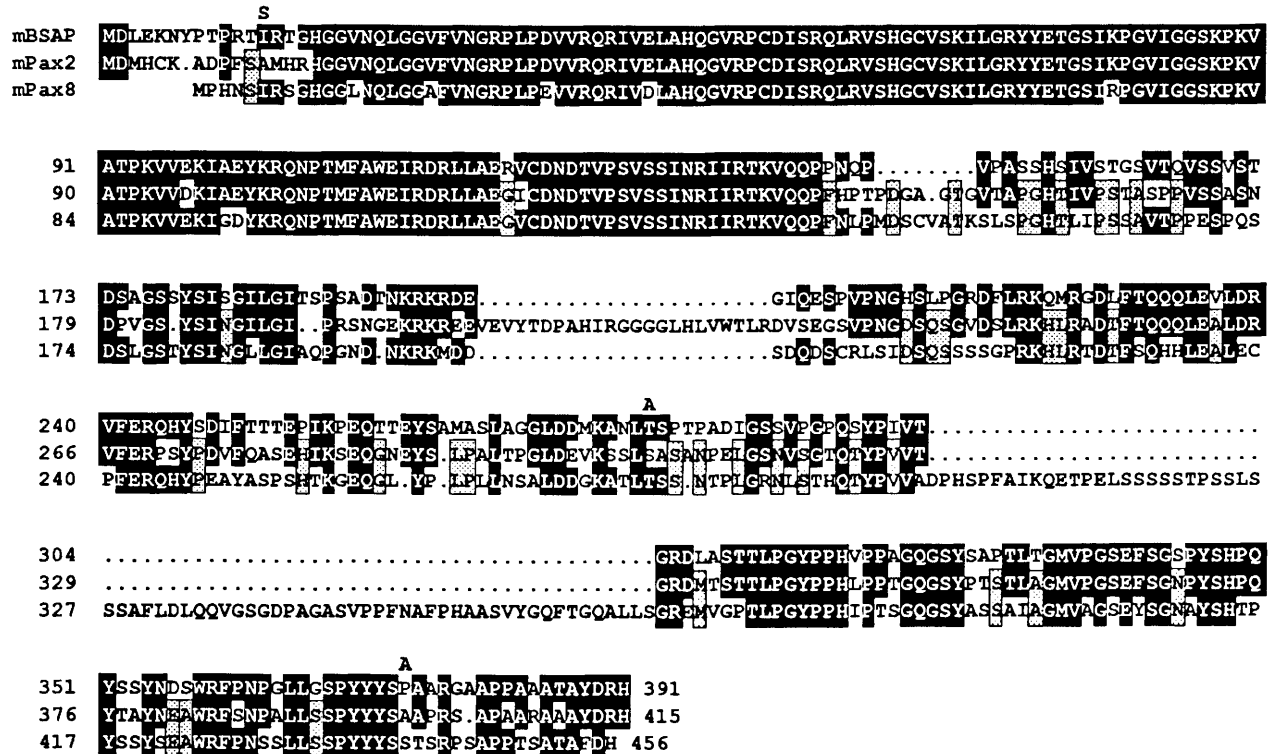Figure 4. (*See facing page for legend.*)

```
                          S
mBSAP    MDLEKNYPTPRTIRTGHGGVNQLGGVFVNGRPLPDVVRQRIVELAHQGVRPCDISRQLRVSHGCVSKILGRYYETGSIKPGVIGGSKPKV
mPax2    MDMHCK.ADPFSAMHRHGGVNQLGGVFVNGRPLPDVVRQRIVELAHQGVRPCDISRQLRVSHGCVSKILGRYYETGSIKPGVIGGSKPKV
mPax8        MPHNSIRSGHGGLNQLGGAFVNGRPLPEVVRQRIVDLAHQGVRPCDISRQLRVSHGCVSKILGRYYETGSIRPGVIGGSKPKV

 91      ATPKVVEKIAEYKRQNPTMFAWEIRDRLLAERVCDNDTVPSVSSINRIIRTKVQQPPNQP........VPASSHSIVSTGSVTQVSSVST
 90      ATPKVVDKIAEYKRQNPTMFAWEIRDRLLAEGICDNDTVPSVSSINRIIRTKVQQPFHPTPDGA.GTGVTAPGHTIVPSTASPEVSSASN
 84      ATPKVVEKISDYKRQNPTMFAWEIRDRLLAEGVCDNDTVPSVSSINRIIRTKVQQPFNLPMDSCVATKSLSPGHTLIPSSAVTPPESPQS

173      DSAGSSYSISGILGITSPSADTNKRKRDE.......................GIQESPVPNGHSLPGRDFLRKQMRGDLFTQQQLEVLDR
179      DPVGS.YSINGILGI..PRSNGEKRKREEVEVYTDPAHIRGGGGLHLVWTLRDVSEGSVPNGDSQSGVDSLRKHIRADIFTQQQLEALDR
174      DSLGSTYSINGLLGIAQPGNDNKRKMDD.......................SDQDSCRLSIDSQSSSSGPRKHIRTDIFSQHHLEALEC

240      VFERQHYSDIFTTTEPIKPEQTTEYSAMASLAGGLDDMKANLTSPTPADIGSSVPGPQSYPIVT.........................
266      VFERPSYPDVFQASEHIKSEQGNEYS.LPALTPGLDEVKSSLSASRNPELGSNVSGTQTYPVVT.........................
240      PFERQHYPEAYASPSHTKGEQGL.YP.LPLLNSALDDGKATLTSS.NTPLGRNLSTHQTYPVVADPHSPFAIKQETPELSSSSSTPSSLS

304      ..............................................GRDLASTTLPGYPPHVPPAGQGSYSAPTLTGMVPGSEFSGSPYSHPQ
329      ..............................................GRDMTSTTLPGYPPHLPPTGQGSYPTSTLAGMVPGSEFSGNPYSHPQ
327      SSAFLDLQQVGSGDPAGASVPPFNAFPHAASVYGQFTGQALLSGREMVGPTLPGYPPHIPTSGQGSYASSAIAGMVAGSEYSGNAYSHTP

                          A
351      YSSYNDSWRFPNPGLLGSPYYYSPAARGAAPPATATAYDRH 391
376      YTAYNEAWRFSNPALLSSPYYYSAAPRS.APAARATAYDRH 415
417      YSSYSEAWRFPNSSLLSSPYYYSSTSRPSAPPTSATAFDH  456
```

Figure 5. Amino acid sequence comparison of mBSAP, mPax-2, and mPax-8. The deduced amino acid sequence of the cDNA clones mBSAP-1 and mBSAP-2 (see Fig. 3D) was aligned with the protein sequence of mouse Pax-2 (Dressler et al. 1990) and Pax-8 (Plachov et al. 1990). Amino acids residues that are identical to the BSAP sequence are highlighted by black overlay. Additional homologies between mPax-2 and mPax-8 are indicated by stippled boxes. The three amino acid substitutions present at positions 13, 282, and 374 in human BSAP are shown above the mouse BSAP sequence. The mPax-2 sequence was corrected at positions 115 (D), 116 (R), and 119 (A), and the mPax-8 sequence at positions 158 (T), 263 (Y), and 340 (D) according to our own PCR cloning and sequencing data (Z. Kozmik, unpubl.). The composite cDNA sequence of clone mBSAP-1 and mBSAP-2 has been submitted to GenBank (accession no. M97013).

consisting of either the amino-terminal 220- or 163-amino-acid residues of BSAP. Only small and large BSAP complexes, but no complexes of intermediate size, were detected by EMSA (Fig. 6A, lanes a + b, a + d). This finding clearly indicates that short and long BSAP polypeptides do not form heterodimers and, hence, that BSAP binds to DNA as a monomer.

BSAP and Pax-1 recognize the same DNA sequences with different affinities

We next asked whether the previously characterized BSAP-binding sites are also recognized by Pax proteins that are distantly related to BSAP/Pax-5. For these experiments we have chosen Pax-1, which differs from BSAP in the paired domain at 38 of 128 amino acid positions (Walther et al. 1991). Both the Pax-1 and BSAP proteins were expressed in transiently transfected COP-8 cells and then analyzed by EMSA for specific binding to the seven recognition sequences shown in Figure 7D. The binding experiments of Figure 7A clearly demonstrate that both BSAP and Pax-1 interact with the sites present in the sea urchin *H2A-2.2* and human *CD19* genes, as well as with the e5 sequence of the *Drosophila even-skipped* promoter. Moreover, both proteins generated an identical methylation interference pattern on three of these recognition sequences analyzed (Fig. 7B).

To assess and compare the DNA-binding affinities of BSAP and Pax-1 for the different sites, we have performed saturation binding assays based on quantitative gel-shift measurements (Calzone et al. 1988; for details, see Materials and methods). Briefly, increasing amounts

Figure 4. Sequence and structural organization of human BSAP cDNA. (*A*) Entire nucleotide and deduced amino acid sequence of the cDNA clone hBSAP-1. The two peptide sequences obtained by the microsequencing of purified BSAP are highlighted by black overlay. The paired domain is boxed, and the octapeptide (amino acids 179–186) and homeo box homology region (amino acids 229–251) are underlined. The cDNA sequence has been submitted to GenBank (accession no. M96944). (*B*) Schematic diagram of the structural organization of the BSAP-1 cDNA. 5'- and 3'-Untranslated sequences (5' UT and 3' UT, respectively) are indicated by a line, and the long ORF with its subdomains and corresponding amino acid positions is shown by a boxed region.
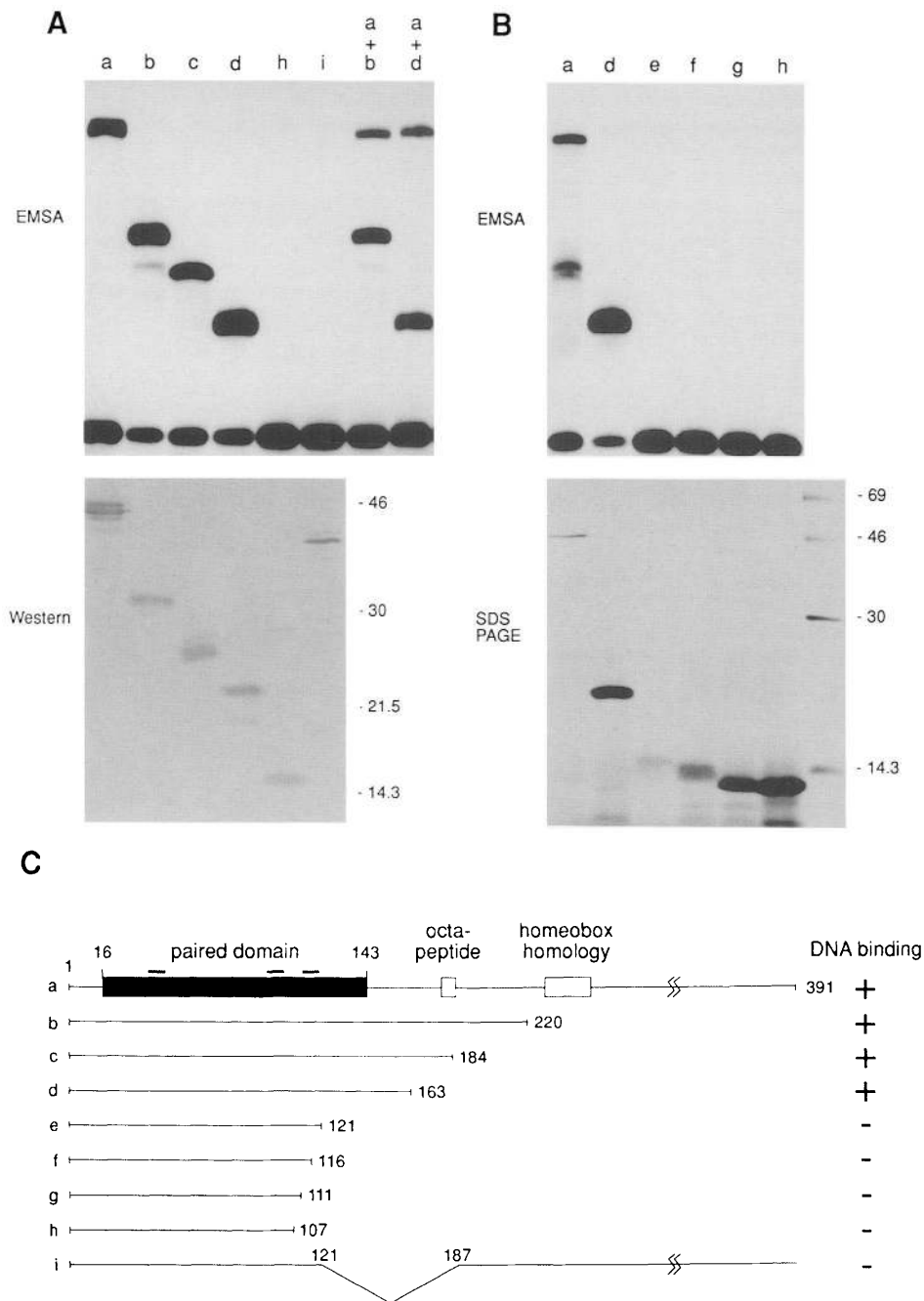
Adams et al.

Figure 6. Identification of the DNA-binding domain of BSAP. (*A*) DNA-binding analysis of in vivo-expressed BSAP peptides. COP-8 cells were transiently transfected with expression constructs directing the synthesis of the BSAP peptides shown in *C*, and COP-8 cell extracts were subsequently analyzed by EMSA for specific binding to the *H2A-2.2* TSAP-binding site (*top*). The same COP-8 cell extracts were analyzed by 15% SDS-PAGE followed by immunoblotting with affinity-purified anti-paired domain BSAP antibodies as described in Materials and methods (*bottom*). The positions of marker proteins (sizes in kD) are indicated to the *left*. (*B*) DNA-binding analysis of in vitro-translated BSAP peptides. The BSAP mutants diagramed in *C* were in vitro-translated in the presence of [^{35}S]methionine by programming a rabbit reticulocyte lysate with the same amount of capped SP6 RNA transcripts for each construct followed by 15% SDS-PAGE using a ^{14}C-labeled protein size marker (*bottom*). The same proteins were assayed for DNA binding by EMSA by using the H2A-2.2 TSAP probe (*top*). (*C*) Schematic representation of carboxy-terminal deletion mutants of BSAP. hBSAP-1 cDNA was resected with exonuclease III and cloned into the expression vector pKW10 containing stop codons in all three reading frames downstream of the *Eco*RI site. Numbers refer to amino acid positions of BSAP. The paired domain, octapeptide, and homeo box homology region are indicated by boxes. Bars above the paired box denote the position of the three α-helices postulated by secondary structure prediction (Bopp et al. 1989).

of a specific binding site were incubated with a constant amount of BSAP or Pax-1, and the concentration of protein-bound (PD_s) and free (D_s) sites was determined by EMSA. As shown in Figure 7C, these binding data were evaluated and plotted according to Scatchard analysis, which allowed the determination of the two parameters K_r and P_o (Emerson et al. 1985; Calzone et al. 1988). K_r corresponds to the negative slope of the curve, is defined
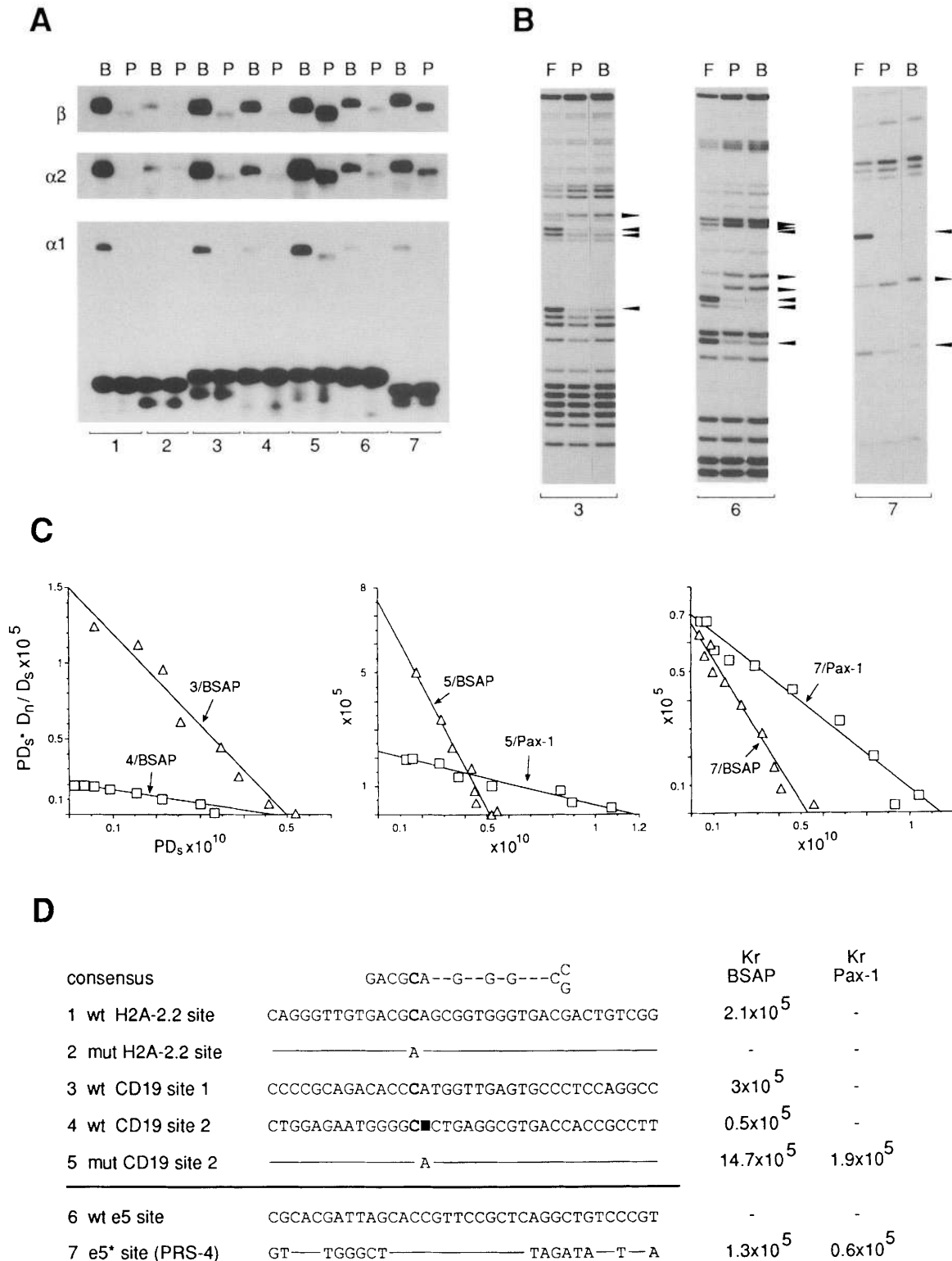

Figure 7. (See following page for legend.)

as the ratio of the specific to the nonspecific binding constant, and therefore directly measures the relative affinity of a protein for a given specific site. The K_r values of BSAP and Pax-1 were measured for five and two binding sites, respectively, and are listed in Figure 7D. P_o corresponds to the total concentration of active DNA-binding protein in the extract and is determined by the intersection of the curve with the x-axis. Scatchard analysis with sites 5 and 7 revealed a twofold higher concentration of active Pax-1 than BSAP in our COP-8 cell extracts (Fig. 7C).

A practical implication of the above analysis was that the concentration of the two proteins and of the specific DNA probes could be carefully adjusted in the EMSA experiment of Figure 7A to facilitate direct comparison of BSAP and Pax-1 binding to the different sites [see legend to Fig. 7A]. This comparison was particularly important for low affinity sites for which K_r values could not be accurately determined in our saturation binding assays. BSAP clearly bound to all seven sites with higher affinity than Pax-1, confirming the trend observed with the K_r values of sites 5 and 7. The difference between the two proteins was greatest for the binding site of the sea urchin $H2A$-2.2 gene (~20-fold) while it was smallest for the mutant version PRS-4 of the $Drosophila$ e5 sequence (~2-fold), which has recently been described as a high affinity site for Pax-1 (Chalepakis et al. 1991). Thus, BSAP and Pax-1 clearly differ in their sequence specificity. We have also identified two single point mutations with completely contrasting effects on protein binding. A single C → A mutation in the high affinity site of the $H2A$-2.2 gene, which drastically decreased BSAP binding in agreement with our previous analysis (Barberis et al. 1990), also reduced Pax-1 binding beyond the detection limit. In contrast, a single A insertion into the BSAP site 2 of the $CD19$ promoter (Kozmik et al. 1992) increased the affinity of BSAP for this site by a factor of ~30. The same A insertion also created the strongest binding site known so far for Pax-1. In summary, we have demon-strated that the distantly related BSAP and Pax-1 proteins are able to bind to the same sequences although with widely different affinities.

Expression of the BSAP/Pax-5 gene in B-lymphoid tissues and testis of the adult mouse

The expression of the BSAP gene was mapped in a panel of different B-lymphoid cell lines by RNase protection analysis. As a control, the level of mRNA coding for the ribosomal S16 protein was determined by S1 nuclease mapping (Fig. 8A). BSAP expression was only detected in cell lines representing the pro-B, pre-B, and mature B-cell stages of B-lymphoid differentiation. BSAP was, however, not expressed in plasma cells (Fig. 8A) and all other hematopoietic cell lines tested (P. Dörfler, unpubl.). This expression pattern fully agrees with our previous BSAP-binding studies (Barberis et al. 1990) and suggests that BSAP expression is regulated at the transcriptional level. These findings furthermore confirm that the BSAP gene is expressed within the hematopoietic system only at the early stages of B-cell differentiation.

In agreement with this conclusion, BSAP expression is high in spleen, lymph node, and blood, the three B-lymphoid tissues of the adult mouse that were analyzed by RNase protection assay (Fig. 8B). Lower levels of BSAP mRNA were detected in lung and salivary gland. However, these two tissues also contained detectable amounts of immunoglobulin κ constant (c_κ) gene transcripts, indicating that both tissues were contaminated with blood and that their BSAP mRNA therefore originated from B lymphocytes. We have recently demonstrated that BSAP is involved in the regulation of the gene coding for the B-lymphoid-specific transmembrane receptor CD19 (Kozmik et al. 1992). Consistent with this finding, the level of CD19 mRNA closely parallels that of the BSAP gene transcript in the three lymphoid tissues analyzed.

Testis was the only non-B-lymphoid tissue that

Figure 7. Comparison of the DNA-binding properties of BSAP and Pax-1. (*A*) EMSA. The hBSAP (B) and mPax-1 (P) proteins used for all experiments were expressed in transiently transfected COP-8 cells as described in Materials and methods. The sequences of DNA probes 1–7 are shown in *D*. Each of these oligonucleotides was 3'-end-labeled to the same high specific activity. For the experiment shown in panel α the same amount of BSAP or Pax-1 protein (final concentration, 0.5×10^{10} M) was added to the binding reaction (20 μl) containing 0.3 fmole of labeled double-stranded oligonucleotide and 1 μg of poly[d(I-C)]. The binding of BSAP to its high-affinity sites 1, 3, and 5 was not yet saturated under these conditions. A long-term autoradiographic exposure (panel α2) of the relevant part containing the protein–DNA complexes is shown above the short-term exposure of the entire gel (panel α1). A fivefold higher probe concentration was used for the experiment shown in panel β. (*B*) Methylation interference analysis. The oligonucleotides 3, 6, and 7 shown in *D* were cloned into the *Sal*I site of pSP64. The corresponding *Hind*III–*Eco*RI restriction fragments were end-labeled at the *Hind*III or *Eco*RI site, partially methylated by dimethylsulfate (DMS) at G residues, and used for methylation interference analysis with hBSAP (B) or mPax-1 (P) protein. (F) Free DNA. Arrowheads indicate methylated G residues that interfere with or enhance protein binding. (*C*) DNA-binding affinity measurements for hBSAP and mPax-1. The same amount of the BSAP or Pax-1 containing COP-8 extract was used for titration competition experiments with the indicated oligonucleotides as described in Materials and methods. The gel-shift data were quantitated and evaluated by Scatchard analysis as described in detail by Emerson et al. (1985) and Calzone et al. (1988). (PD$_s$) Molar concentration of specific protein/DNA complexes. (D$_s$) Molar concentration of free specific binding sites. (D$_n$) Molar concentration of poly[d(I-C)]. (*D*) Oligonucleotide sequences. The TSAP-binding site of the sea urchin *H2A-2.2* gene was published by Barberis et al. (1989), the two BSAP sites of the *CD19* gene by Kozmik et al. (1992), the wild-type (wt) e5 sequence of the *Drosophila even-skipped* promoter by Treisman et al. (1991), and the mutant e5 oligonucleotide PRS-4 by Chalepakis et al. (1991). The consensus BSAP-binding sequence is deduced from the four high affinity sites of the sea urchin *H2A-2* and *H2B-2* genes and from sites 1 and 2 of the human *CD19* promoter (see Barberis et al. 1989; Kozmik et al. 1992).
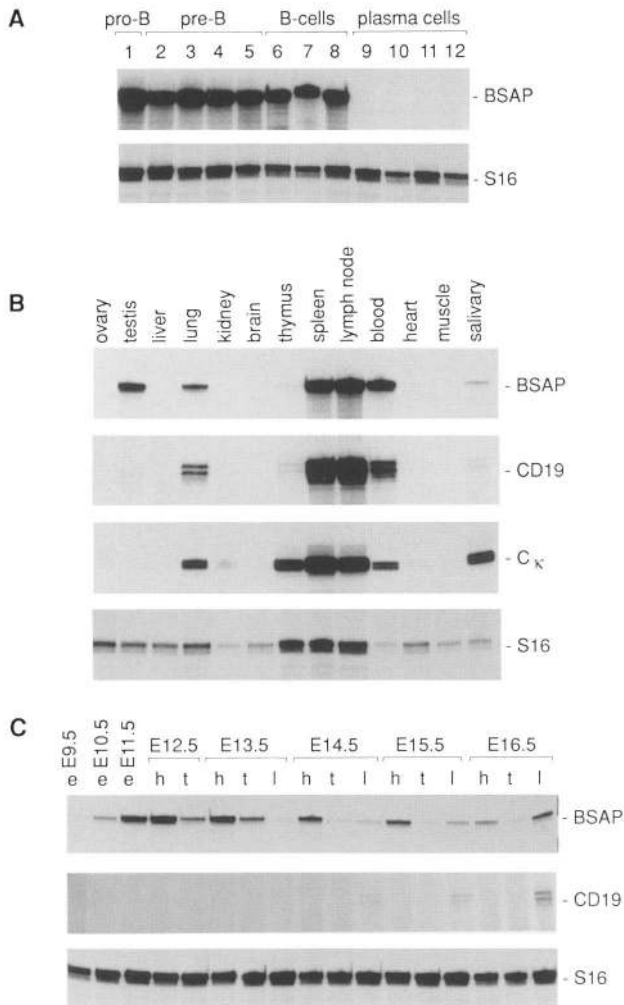
Figure 8. Expression of the BSAP/*Pax-5* gene during embryogenesis and in adult tissues of the mouse. (*A*) BSAP expression in B-lymphoid cells. Total RNA (10 μg) of the following cell lines was analyzed by RNase protection for BSAP transcripts and by S1 mapping for S16 mRNA. Lanes: (*1*) HAFTL-1sc1; (*2*) HAFTL-pBcl.6; (*3*) 18.81$_{2a}$; (*4*) PD31; (*5*) 70Z/3; (*6*) WEHI-231; (*7*) WEHI-279.1; (*8*) 38-C-13; (*9*) P3X63-Ag8.653; (*10*) J558L; (*11*) SP2/0; (*12*) S194. (*B*) BSAP expression in adult tissues of the mouse. Animals were bled prior to tissue dissection and RNA preparation. BSAP and CD19 mRNAs were mapped separately by RNase protection of 10 μg of total RNA; transcripts of the immunoglobulin c$_\kappa$ gene and of the S16 gene were detected by S1 nuclease analysis. (salivary) Salivary gland. (*C*) BSAP expression during mouse embryogenesis. Total RNA was prepared from whole embryos (e) up to day E11.5 p.c.; later embryos were dissected into head (h), trunk (t), and fetal liver (l) before RNA preparation. Total RNA (20 μg) was used for separate RNase protection analysis of BSAP and CD19 mRNA and for S1 nuclease mapping of ribosomal protein S16 gene transcripts.

showed a high level of BSAP gene expression. Neither the *CD19* nor the immunoglobulin κ light-chain genes were expressed in testis, thus ruling out blood contamination as a source for BSAP mRNA in this tissue. In summary, we conclude that the BSAP/*Pax-5* gene is ex-

pressed both in B lymphocytes and testis of the adult mouse. Moreover, *CD19* appears to be a strictly B-lymphoid-specific target gene of BSAP.

*Expression of the BSAP/*Pax-5 *gene
in the developing midbrain*

The BSAP gene is expressed in two waves during mouse embryogenesis, as shown by the RNase protection analysis of Figure 8C. BSAP gene transcripts start to accumulate between day 9.5 and 10.5 postcoitum (p.c.), reach maximal levels both in the head and trunk of the embryo between day 11.5 and 13.5, and then start to decline in abundance in these two parts of the body. This first wave of expression results entirely from transcription of the BSAP gene in the developing CNS (see below) and, interestingly enough, is not accompanied by expression of the target gene *CD19*. Later in embryogenesis, the site of *Pax-5* expression shifts to the fetal liver, where its expression gradually increases from day 13.5 onward in parallel with that of the *CD19* gene. This second expression pattern coincides nicely with the onset of B lymphopoiesis in fetal liver (Rolink and Melchers 1991).

The spatial and temporal expression pattern of the BSAP gene was analyzed in the developing brain by in situ hybridization (Fig. 9). Serial sections of 11.5- and 14.5-day embryos were hybridized with antisense RNA probes corresponding to the central and carboxy-terminal-coding sequences of the mouse BSAP or control *Pax-8* gene. At day 11.5 of gestation, strong expression of the BSAP gene is detected in the posterior part of the mesencephalon. BSAP expression is confined in this region to the tegmental neuroepithelium on both sides of the fovea of the isthmus (Fig. 9B,C). Some expression of BSAP mRNA is also seen in the posterior end of the collicular neuroepithelium of the tectum (Fig. 9B). However, the BSAP gene is not expressed along the rhombencephalon in marked contrast to the *Pax-8* gene, which is expressed in the region of the pons (Fig. 9D; Plachov et al. 1990). At day 14.5, expression of the BSAP gene is much more restricted locally and is seen only in a small neuroepithelial region immediately centering around the fovea of the isthmus (Fig. 9F). As seen in the transverse section of Fig. 9G, BSAP gene expression is absent from both the roof and central floor plate of the mesencephalon. However, strong BSAP expression is seen laterally on both ventral sides of the third ventricle. As shown in the close-up image of Figure 9H, the BSAP gene is expressed in the cells of the subependymal matrix but not in the epithelial cells of the ependyma.

*Expression of the BSAP/*Pax-5 *gene
in the developing neural tube*

All murine *Pax* genes analyzed so far are expressed in the neural tube along the entire anterior–posterior axis, with the exception of *Pax-1*. The *Pax-5* gene also conforms to this rule, as shown by the in situ hybridization analysis of Figure 10. Initially, the neural tube consists of a single

neuroepithelial layer, the ventricular zone, which is radially arranged around the neural canal. The cells in the ventricular zone are mitotically active and generate the stem cell population of the CNS. A second zone, the mantle layer, forms between day 11 and 12 in a precise temporal and spatial pattern starting in the most ventral region of the basal plate (Fig. 10D, diagram). The mantle layer consists of differentiating postmitotic neurons that migrate radially away from the ventricular zone (Nornes et al. 1990 and references therein). Figure 10 shows the analysis of a longitudinal section through the thoracic part of the neural tube of a 12.5-day embryo, which was hybridized either with the antisense or the control sense BSAP RNA probe (Fig. 10B,C). Interestingly, BSAP expression, as detected with the antisense RNA probe, is confined to a cell layer that constitutes the border between the ventricular zone and the mantle layer.

Discussion

BSAP is encoded by the Pax-5 gene

BSAP has been characterized previously as a B-cell lineage-specific transcription factor that is expressed at early, but not late, stages of B-lymphoid differentiation (Barberis et al. 1990). Here, we describe the purification and cDNA cloning of BSAP from human BJA-B cells. Although only a partial cDNA clone corresponding to the 5' part of a ~10-kb transcript could be isolated, it was shown to code for authentic BSAP by four different criteria. First, the ORF of this cDNA clone contains the two peptide sequences obtained by microsequencing of BSAP (Fig. 4). Second, expression of the cDNA clone in COP-8 cells gives rise to a DNA-binding activity that is indistinguishable from the endogenous BSAP activity of BJA-B cells (Fig. 3A). Third, antibodies raised against the

Figure 9. (See facing page for legend.)

ORF specifically recognize endogenous BSAP (Fig. 3B). Fourth, the expression profile of the mRNA detected by the cloned cDNA completely correlates with the pattern of BSAP-binding activity in a large panel of hematopoietic cells (Fig. 8; P. Dörfler, unpubl.).

The BSAP protein contains a classical paired domain at its amino terminus and is therefore a member of the Pax protein family. BSAP is most closely related to the mouse Pax-2 and Pax-8 proteins and, hence, belongs to subclass III, as defined by Walther et al. (1991). Recently,

a third member of this subclass, *Pax-5*, was identified by screening a mouse genomic library with a *Pax-1* paired box probe. However, only the two exons coding for the conserved paired domain of Pax-5 could be identified in the absence of cloned cDNA (Walther et al. 1991). The cloning of mouse BSAP cDNA has revealed complete identity between these two exons of the *Pax-5* gene and the corresponding BSAP cDNA sequence, demonstrating that BSAP is encoded by the *Pax-5* gene (Fig. 5). Moreover, the human and mouse BSAP proteins have diverged
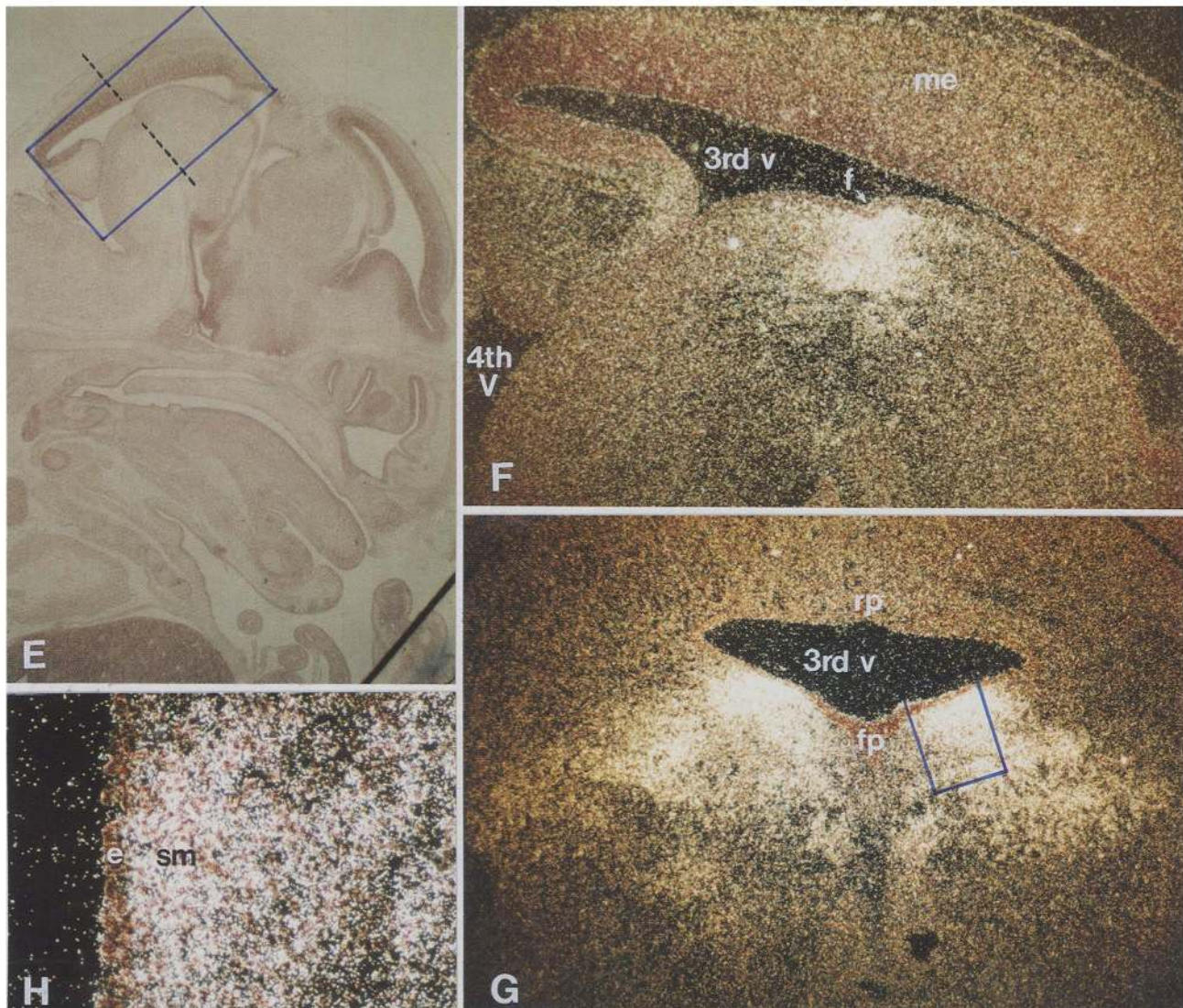
Figure 9. Expression of the BSAP/*Pax-5* gene in the developing midbrain. (*A*) Sagittal section through the brain of an 11.5-day mouse embryo; bright-field image. (*B*) The same section as in *A*, hybridized with the antisense BSAP RNA probe and photographed under dark-field illumination. (*C*) Sagittal paramedial section through mid- and hindbrain of a 11.5-day embryo, hybridized with antisense BSAP transcripts. (*D*) Adjacent section hybridized with an antisense Pax-8 RNA probe. (*E*) Sagittal paramedial section through the head of a 14.5-day embryo; bright-field image. (*F*) The same section hybridized with the BSAP RNA probe. Only the close-up view of the field boxed in *E* is shown. (*G*) Transverse section through the midbrain of a 14.5-day embryo along the plane indicated by a broken line in *E*, hybridized with the BSAP RNA probe. (*H*) Close-up view of the field indicated in *G*. (cn) Collicular neuroepithelium; (cp) choroid plexus; (e) ependymal layer; (f) fovea of the isthmus; (fp) floor plate; (me) mesencephalon; (mt) metencephalon; (my) myelencephalon; (p) pons; (rp) roof plate; (sm) subependymal matrix; (te) telecephalon; (tn) tegmental neuroepithelium; (v) ventricle.
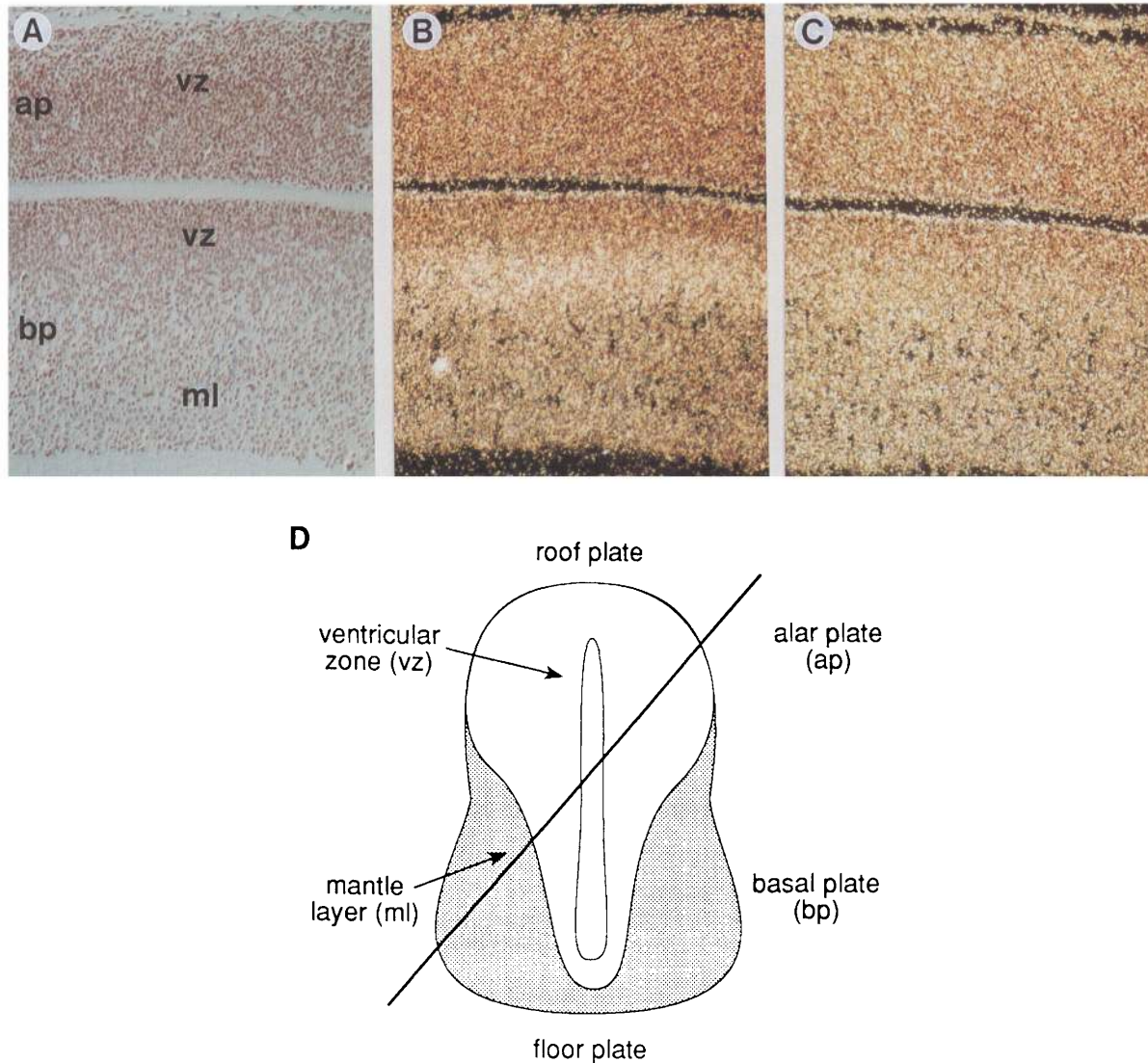
Adams et al.

Figure 10. Expression of the BSAP/*Pax-5* gene in the neural tube. (*A*) Longitudinal section through the neural tube of a 12.5-day mouse embryo, photographed under bright-field illumination. (*B*) In situ hybridization with the antisense BSAP RNA probe. Dark-field image of *A*. (*C*) In situ hybridization with the sense BSAP RNA probe. (*D*) Schematic illustration of a cross section through the neural tube of a 12.5-day embryo. The plane of the longitudinal section shown in *A–C* is indicated.

only at 3 of 391 amino acid positions. An equally high conservation between human and mouse is also observed for Pax-6 (Ton et al. 1991; Walther and Gruss 1991), suggesting that Pax proteins are subject to high evolutionary constraints even outside of their paired domains.

BSAP contains three serine/threonine-rich regions, in the central and carboxy-terminal part, which may harbor a potential *trans*-activation function in analogy to the serine/threonine-rich *trans*-activation domains of Oct-1 and Oct-2 (Tanaka and Herr 1990). Previously, we have demonstrated by transient transfection experiment that the endogenous BSAP activity of B cells is able to stimulate transcription from a minimal promoter containing a BSAP recognition sequence (Barberis et al. 1990;

Kozmik et al. 1992). However, the same reporter genes could not be stimulated by ectopic expression of BSAP cDNA in heterologous cells despite the fact that BSAP-binding activity was readily detectable in the transfected cells (P. Dörfler, unpubl.). A possible explanation for this discrepancy could be that transcriptional activity of BSAP critically depends on a B-cell-specific modification and/or requires the interaction with another B-lymphoid-specific component of the transcription machinery.

DNA-binding properties of BSAP/Pax-5

The paired domain has recently been identified as a novel DNA-binding motif of the *Drosophila* Prd (Treis-

man et al. 1991) and the murine Pax-1 and Pax-3 proteins (Chalepakis et al. 1991; Goulding et al. 1991). Mutational analysis of BSAP has demonstrated that an intact paired domain was both necessary and sufficient for DNA binding of BSAP (Fig. 6). BSAP appears to differ in this respect from the *Drosophila* Prd and murine Pax-1 proteins, both of which bind to DNA even in the absence of the last 35 carboxy-terminal amino acids of their paired domain, albeit with reduced affinity (Chalepakis et al. 1991; Treisman et al. 1991). Many transcription factors, including steroid receptors, basic leucine zipper (bZIP), and basic helix–loop–helix (bHLH) proteins, bind as homo- or heterodimers to their recognition sequences (for review, see Jones 1990). BSAP clearly lacks a dimerization function, as no interaction between full-length and truncated forms of BSAP was observed in DNA-binding assays (Fig. 6). Paired domain proteins such as BSAP therefore bind as monomers to their target DNA. In agreement with this finding, the recognition sequences of BSAP are generally not palindromic (Barberis et al. 1990; Kozmik et al. 1992).

The paired domain has been well conserved in evolution, raising the question whether all Pax proteins are identical with regard to their sequence recognition. A comparison of BSAP and Pax-1 demonstrated that Pax-1 binds to known BSAP recognition sites and likewise that BSAP interacts with the recently characterized Pax-1-binding site derived from the e5 sequence of the *Drosophila even-skipped* promoter. Moreover, both proteins recognize the same G residues within these recognition sequences, as shown by methylation interference analysis (Fig. 7). However, quantitative gel-shift experiments revealed a clear difference in binding affinity between the two proteins, as BSAP binds more strongly than Pax-1 to all of the sites analyzed. These data therefore suggest that the Pax proteins of the different subclasses (Walther et al. 1991) exhibit different sequence-binding specificities due to amino acid changes in their paired domains.

We have previously shown that the DNA sequence recognition of BSAP is identical to that of the sea urchin transcription factor TSAP (Barberis et al. 1990). The recent cDNA cloning of this sea urchin factor confirmed that it is highly homologous in the paired domain to BSAP and thus belongs to the same subclass of Pax proteins (T. Czerny and M. Busslinger, unpubl.). For both proteins only a degenerate consensus recognition sequence (shown in Fig. 7D) could be deduced from the four high affinity sites present in the sea urchin histone *H2A-2* and *H2B-2* genes (Barberis et al. 1989) and the two binding sites identified in the human *CD19* promoter (Kozmik et al. 1992). Correct alignment of all six sequences was only obtained by comparing three different parameters, that is, primary sequence, DNase I footprint, and methylation interference pattern. This alignment predicts that insertion of an A residue into CD19 site 2 improves the homology to the consensus sequence by eliminating a 1-nuceotide gap and, consequently, should increase the affinity of this site for BSAP (Fig. 7D; Kozmik et al. 1992). This A insertion not only created

the highest affinity site for BSAP but also for Pax-1 (Fig. 7D). In this context it is, however, important to note the following two points. First, BSAP binds with relatively high affinity to the *Drosophila* e5 site and the derived PRS-4 sequence (Fig. 7); yet we were unable to unequivocally align these two sites with our BSAP consensus sequence. Second, Chalepakis et al. (1991) have recently defined a DNA-binding core motif (GTTCC) for Pax-1 by mutational analysis of the *Drosophila* e5 sequence. This pentanucleotide is not only absent from our highest affinity site for Pax-1, but also from all other binding sites. These two observations nicely document the degenerate sequence recognition of paired domain proteins and emphasize the limited value of consensus sequences and core motifs for Pax proteins.

Expression of the BSAP/Pax-5 gene in the developing CNS

All murine *Pax* genes, including the BSAP gene are transiently expressed in the developing CNS except for *Pax-1* (see introductory section). The expression of the BSAP gene is, however, delayed with regard to that of other *Pax* genes. BSAP expression peaks in the brain between embryonic day 11.5 and 13.5 and is locally restricted to the tegmental neuroepithelium at the posterior end of the mesencephalon (Fig. 9). The BSAP gene is not expressed in the rhombencephalon in contrast to the other two members of the same subclass, *Pax-2* and *Pax-8* (Fig. 9D; Nornes et al. 1990; Plachov et al. 1990). No BSAP gene transcripts were detectable in the neural tube at embryonic day 10.5 in contrast to Pax-2 (Nornes et al. 1990) and Pax-8 mRNA (A. Aguzzi, unpubl.); however, weak BSAP expression was observed at day 12.5 (Fig. 10). At this developmental stage the neural tube consists of the ventricular zone containing the mitotically active neuronal stem cells and of the mantle layer comprising the differentiating postmitotic neurons (Nornes et al. 1990). BSAP expression was confined to a cell layer at the border between these two zones (Fig. 10). In conclusion, the BSAP gene differs from the other *Pax* genes studied so far with regard to its spatial and temporal expression pattern in the developing CNS.

BSAP/Pax-5 expression in testis and B-lymphoid tissues

B lymphocyte development is initiated in the fetal liver at day 13 of gestation (Rolink and Melchers 1991). BSAP gene expression in this hematopoietic organ correlates nicely with the onset of B lymphopoiesis (Fig. 8C) in agreement with the fact that BSAP was originally identified as a B-cell-specific transcription factor expressed at early stages of B-cell differentiation (Barberis et al. 1990). Expression of the BSAP gene in B-lymphoid tissues persists into adult life in contrast to its transient expression in the CNS. Transcripts of the BSAP gene are localized to B-cell follicles in the adult spleen and lymph nodes, as shown by in situ hybridization (A. Aguzzi, unpubl.). Unexpectedly, the BSAP gene is also transcribed in the tes-

tis of the adult mouse (Fig. 8). The BSAP/*Pax-5* gene is therefore the first member of the *Pax* gene family reported to be expressed in testis (Walther et al. 1991 and references therein). The specific cells within the testis that express the BSAP gene remain to be identified.

What is the role of BSAP in B-cell differentiation? Previously, we have shown that known regulatory regions of immunoglobulin and class II major histocompatibility complex (MHC) genes lack high-affinity binding sites for BSAP, suggesting that this transcription factor may be involved in the regulation of a different set of B-lymphoid-specific genes (Barberis et al. 1990). Recently, we have identified the gene coding for the B-cell-specific transmembrane protein CD19 as a BSAP target. A high-affinity BSAP-binding site instead of a TATA box was shown to be located in the −30 region of this gene upstream of a cluster of heterogeneous start sites. Moreover, BSAP was shown to bind to this site in vivo in CD19-expressing B cells but not in plasma or HeLa cells (Kozmik et al. 1992). We now demonstrate that expression of the *CD19* gene does not only correlate with BSAP expression in adult B-lymphoid tissues but also during early B lymphopoiesis in fetal liver (Fig. 8). However, the *CD19* gene is neither expressed in the developing CNS nor in adult testis, demonstrating that *CD19* is a B-lymphoid-specific target gene of BSAP. This B-cell specificity may be achieved by regulating the accessibility of the *CD19* promoter at the chromatin level or by cooperation of BSAP with another B-cell-specific transcription factor. Given the enormous size of the BSAP trailer sequences, it is also conceivable that transcripts of this gene are differentially spliced in different tissues, giving rise to functionally distinct isoforms, only one of which may correspond to the transcription factor BSAP.

The *Pax-5* gene has been localized to mouse chromosome 4 close to the *mos* and *etl-2* genes (Walther et al. 1991). So far, no developmental mutants have been mapped to this chromosomal region. Mutations in the BSAP/*Pax-5* gene may be expected to create a pleiotropic phenotype in view of the expression of this gene in three distinct cell lineages. To test this hypothesis and to gain insight into the role of BSAP during early development, we are currently inactivating the BSAP gene in the mouse germ line by homologous recombination.

Materials and methods

Protein purification and microsequencing

Nuclear extracts were prepared from BJA-B cells, as described by Dignam et al. (1983), except that all buffers contained the following protease inhibitors: 1 mM PMSF, 0.5 µg/ml of leupeptin, 40 µg/ml of TLCK, and 5 µg/ml each of aprotinin, pepstatin, and antipain. Fifty milliliters of the 0.42 M NaCl extract (corresponding to a 12-liter culture) was loaded at a rate of 0.5 ml/min onto a 10-ml wheat germ lectin column (Sigma; 5 mg lectin/ml of Sepharose 6MB), which was equilibrated in buffer Z420 [420 mM KCl, 25 mM HEPES (pH 7.9), 12.5 mM MgCl$_2$, 10 µM ZnCl$_2$, 0.1 % NP-40, 20% glycerol]. The column was washed with 10 ml of buffer Z420 and 20 ml of buffer Z50 (as Z420, except for 50 mM KCl). Glycosylated proteins were eluted

with 0.3 M N-acetyl-glucosamine in buffer Z50. A BSAP-specific DNA affinity column was prepared by coupling a DNA fragment containing 16 multimerized H2B-2.2 TSAP-binding sites (Barberis et al. 1989) to cyanogen bromide-activated Sepharose 4B (Sigma). The eluate of the lectin column was applied to 1 ml of this DNA affinity resin, which was equilibrated in buffer A50 containing 50 mM KCl, 25 mM HEPES (pH 7.9), 5 mM EDTA, and 0.1% NP-40. The column was washed with 10 ml of buffer A50, and BSAP was eluted with buffer A500 (as A50, except for 500 mM KCl).

The eluted material from four different preparations (corresponding to ~50 liters of BJA-B cells) was concentrated 5- to 10-fold by lyophilization, precipitated with chloroform–ethanol, and dissolved in 100 µl of 6 M guanidine hydrochloride. This material was diluted fivefold into 1% TFA in water and loaded onto a C8 reverse-phase HPLC column (Nucleosil 300-5C8). A 60-ml linear gradient of acetonitrile (0–100% acetonitrile in 0.1% TFA) was applied at 1 ml/min. The BSAP peptide eluted in 58% acetonitrile and was purified further on a preparative 10% SDS–polyacryamide gel, followed by blotting onto an Immobilon-P membrane (Millipore) and Coomassie blue staining. The 50-kD band corresponding to ~5 µg of BSAP protein was excised and subjected to trypsin digestion. Individual peptides isolated by reverse-phase HPLC fractionation were sequenced on an Applied Biosystems pulsed liquid phase sequenator (type 477A). BSAP activity was followed throughout the entire purification procedure by EMSA. However, 50 µg of BSA was added as carrier protein to the binding reaction (20 µl) to titrate the nonionic detergent NP-40 away from BSAP.

PCR oligonucleotides

(a) GCGGAATTCAGPuTAPyTAPyGAPuACNGGNAGPyAT (BSAP peptide 1); (b) GCGGTCGACPuATPyTCCCANGCPuAACATNGTNGG (BSAP peptide 2); (c) GCGAGATCTCCATTTCATCAAGTCCTGAAA (5′ mBSAP); (d) GCGAAGCTTGGTCAPuTGNCGPuTCPuTANGCNGT (3′ mBSAP); (e) GCGGAATTCGCGCGATGGAGCAGACGTACG (5′ mPax-1); (f) GCGAAGCTTGTGGCTCTGTGAGAGGACAGC (3′ mPax-1); (g) GCGGGATCCAAAGCTGCGAGTGTCCCTCAG (5′ mPax-8); (h) GCGAAGCTTGCCTGGTTCTTCCCACTGT (3′ mPax-8).

RNA preparation and PCR cloning

Total RNA was prepared from mouse embryos, adult tissues, and B-lymphoid cell lines by the method of Chirgwin et al. (1979), and polyadenylated RNA was selected by three runs of oligo(dT)-cellulose chromatography. Poly(A)$^+$ RNA (~5 µg) from BJA-B cells, 70Z/3 cells, mouse kidney, and E10.5 embryos was transcribed into cDNA by Moloney muring leukemia virus (M-MuLV) reverse transcriptase (SuperScript; BRL), and 100 ng of cDNA was used for 35 cycles of PCR amplification (1 min at 94°C; 2 min at 55°C; 4 min at 72°C). The BSAP paired domain DNA probe was amplified from BJA-B cDNA with oligonucleotides a and b, the entire mBSAP-coding region from 70Z/3 cDNA with oligonucleotides c and d, mPax-1 from cDNA of day 10.5 embryos with oligonucleotides e and f, and mPax-8 from adult kidney cDNA with oligonucleotides g and h. All PCR clones were verified by DNA sequencing.

Molecular cloning of BSAP from BJA-B and 70Z/3 cDNA libraries

Poly(A)$^+$ RNA (5–10 µg) of BJA-B and 70Z/3 cells was transcribed into double-stranded cDNA by priming with p(dT)$_{15}$ as described (Gubler and Hoffman 1983). After *Eco*RI linker addi-

tion, the cDNA was cloned into bacteriophage λgt10. Recombinant phages (~2 × 10^6) of each library were screened with a ^{32}P-labeled BSAP paired domain DNA probe. Positive clones were rescreened and plaque purified. The EcoRI cDNA inserts were subcloned into the expression vector pKW10.

DNA sequencing

The human BSAP-1 cDNA was cloned in both orientations into the polylinker of pUC19 and resected with exonuclease III from the XbaI site. The sequence of the resected cDNA inserts was determined on an automated sequencer (Applied Biosystem, model 373A). Both strands of the mouse cDNA clones mBSAP-1 and mBSAP-2 were sequenced by the standard dideoxy technique.

S1 nuclease and RNase protection analyses

S1 nuclease and RNase protection analyses were carried out according to Busslinger et al. (1981) and Vitelli et al. (1988), respectively. The mouse CD19 SP6 probe has been described by Kozmik et al. (1992), and the mouse BSAP SP6 probe was generated by cloning a 195-bp PvuII–SalI fragment of mBSAP-1 cDNA in the antisense orientation into the polylinker of pSP64. The S1 probe for detecting murine S16 mRNA was obtained by 5'-end-labeling of a 425-bp EcoRV fragment of plasmid pS16cDNA (Wagner and Perry 1985). The S1 probe for detection of immunoglobulin κ transcripts was a 5'-end-labeled 210-bp StyI–KpnI fragment from the mouse $c_κ$ gene.

BSAP expression in COP-8 cells and in vitro protein–DNA-binding assays

The eukaryotic expression vector pKW10 is derived from pRK5 (R. Klein and D. Goeddel, unpubl.) by replacing the SV40 control region with a 380-bp Sau3A–EcoRI fragment of plasmid pPyoriAB (Weichselbraun et al. 1989) containing the replication origin of polyoma virus. The transcription unit of pKW10 consists of the CMV enhancer–promoter region, a leader sequence interrupted by an intron, a multiple cloning site, and the SV40 polyadenylation signal. cDNA expression plasmids (2–4 μg) were transfected into COP-8 cells (Tyndall et al. 1981) on 90-mm tissue culture dishes by use of the DEAE–dextran method. Two days later, the transfected cells were harvested and lysed in 50 μl of a buffer containing 420 mM NaCl, 20 mM HEPES (pH 7.9), 25% glycerol, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 0.5% NP-40, and the protease inhibitor cocktail described above. EMSAs and methylation interference experiments were performed as described by Barberis et al. (1989), except that 20 μg of BSA was also included into the binding reaction.

Quantitative electrophoretic gel-shift assays

COP-8 cell extracts were diluted to 0.1 μg/μl of total cellular protein in buffer D containing 20 mM KCl, 20 mM HEPES (pH 7.9), 20% glycerol, 1 mM EDTA, 1 mM DTT, 5 μg/ml of BSA, and 0.1 mM PMSF. One microliter of this diluted extract, 0.5 fmole of labeled probe, decreasing amounts of specific competitor DNA (starting with 1 pmole in twofold dilution steps), and 1 μg of the nonspecific competitor DNA poly[d(I-C)] were incubated at room temperature for 20 min in 20 μl of a buffer containing 10 mM HEPES (pH 7.9), 100 mM KCl, 4% Ficoll, 1 mM EDTA, and 1 mM DTT. Following gel electrophoretic separation, the two signals corresponding to protein-bound and free oligonucleotides were quantitated on a PhosphorImager (Molecular Dynamics), and the molar concentrations of bound (PD_s)

and unoccupied (D_s) specific sites were calculated, knowing the total amount of specific sites used in each reaction. These binding data were evaluated according to a modified Scatchard analysis, which also takes into consideration protein binding to the nonspecific competitor DNA poly[d(I-C)] (Emerson et al. 1985; Calzone et al. 1988). D_n, the concentration of poly[d(I-C)], is defined in terms of moles of base pairs.

Generation of polyclonal antibodies and Western blot analysis

The Escherichia coli expression vector pET-2a (Studier et al. 1990) was modified by replacing the short NdeI–BamHI fragment with an oligonucleotide encoding 6 histidine residues. Two PCR fragments containing hBSAP sequences for codons 17–145 or codons 189–391 were cloned into the modified vector pETH-2a. Recombinant protein was expressed in the E. coli strain BL21 (DE3, pLysS) according to Studier et al. (1990) and purified on a nickel–chelate affinity resin as described by Hochuli et al. (1988). Rabbits were immunized with ~150 μg of purified protein emulsified 1 : 1 with complete Freund's adjuvant. Boosts were given at weeks 2, 3, and 5 after the initial injection. Antisera were analyzed by ELISA, and specific antibodies were purified by affinity chromatography on Sepharose 4B coupled to purified antigen. Western blot analysis with BSAP antibodies was carried out as described (Chalepakis et al. 1991).

In situ hybridization analysis

In situ hybridization onto a cross section of mouse embryos was performed as described by Aguzzi et al. (1992). RNA probes were obtained by cloning a 700-bp AccI–HindIII fragment of the mouse BSAP cDNA (from codon 161 to the carboxyl terminus) and a 960-bp BalI–HindIII fragment of mouse Pax-8 cDNA (from codon 150 to the carboxyl terminus) into the polylinker of Bluescript.

Note added in proof

Recent EMSA and cross-competition experiments (T. Czerny and M. Busslinger, unpubl.) have indicated that BSAP is identical with the following, recently published DNA-binding activities, all of which interact with sequences either in or upstream of immunoglobulin heavy-chain gene switch regions: Sα-BP (Waters et al. 1989), complex 3 (Rothman et al. 1991), LR1 (Williams and Maizels 1991), NF-HB (Liao et al. 1992) and $NFS_μ$-

B_1 (Xu et al. 1992). However, of all the binding sites identified, only the 5'Sγ2a site (Liao et al. 1992) and the recognition sequence of complex 3 in the Iϵ promoter (Rothman et al. 1991) proved to be high-affinity BSAP-binding sites comparable to the *H2A-2.2* and *CD19* sites shown in Figure 7.

References

Aguzzi, A., K. Bothe, I. Anhauser, I. Horak, A. Rethwilm, and E.F. Wagner. 1992. Expression of human foamy virus is differentially regulated during development in transgenic mice. *New Biol.* **4:** 225–237.

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, M., 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 segmentation 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.

Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72:** 248–254.

Burri, M., Y. Tromvoukis, D. Bopp, G. Frigerio, and M. Noll. 1989. Conservation of the paired domain in metazoans and its structure in three isolated human genes. *EMBO J.* **8:** 1183–1190.

Busslinger, M., N. Moschonas, and R.A. Flavell. 1981. β^+ Thalassemia: Aberrant splicing results from a single point mutation in an intron. *Cell* **27:** 289–298.

Calzone, F.J., N. Theze, P. Thiebaud, R.L. Hill, R.J. Britten, and E.H. Davidson. 1988. Developmental appearance of factors that bind specifically to *cis*-regulatory sequences of a gene expressed in the sea urchin embryo. *Genes & Dev.* **2:** 1074–1088.

Chalepakis, 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.

Chirgwin, J.M., A.E. Przybyla, R.J. MacDonald, and W.J. Rutter. 1979. Isolation of biological active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18:** 5294–5299.

Deutsch, U., G.R. Dressler, and P. Gruss. 1988. *Pax 1*, a member of a paired box homologous murine gene family, is expressed in segmented structures during development. *Cell* **53:** 617–625.

Dignam, J.D., R.M. Lebovitz, and R.G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11:** 1475–1489.

Dressler, G.R., U. Deutsch, K. Chowdhury, H.O. Nornes, and P. Gruss. 1990. *Pax2*, a new murine paired-box-containing gene and its expression in the developing excretory system. *Development* **109:** 787–795.

Emerson, B.M., C.D. Lewis, and G. Felsenfeld. 1985. Interaction of specific nuclear factors with the nuclease-hypersensitive region of the chicken adult β-globin gene: Nature of the binding domain. *Cell* **41:** 21–30.

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* **57:** 767–774.

Frohman, M.A., M.K. Dush, and G.R. Martin. 1988. Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci.* **85:** 8998–9002.

Goulding, M.D., G. Chalepakis, U. Deutsch, J.R. Ersellius, and P. Gruss. 1991. Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. *EMBO J.* **10:** 1135–1147.

Gubler, U. and B.J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene* **25:** 263–269.

Hill, E.R., J. Favor, B.L.M. Hogan, C.C.T. 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.

Hochuli, E., W. Bannwarth, H. Döbeli, R. Gentz, and D. Stüber. 1988. Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate absorbent. *Bio-Technology* **6:** 1321–1325.

Hope, J.A. and K. Struhl. 1987. GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. *EMBO J.* **6:** 2781–2784.

Jackson, S.P. and R. Tjian. 1988. O-Glycosylation of eukaryotic transcription factors: Implications for mechanisms of transcriptional regulation. *Cell* **55:** 125–133.

Jones, N. 1990. Transcriptional regulation by dimerization: Two sides to an incestuous relationship. *Cell* **61:** 9–11.

Jostes, B., C. Walther, and P. Gruss. 1991. The murine paired box gene, *Pax7*, is expressed specifically during the development of the nervous and muscular system. *Mech. Dev.* **33:** 27–38.

Kozak, M. 1991. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J. Biol. Chem.* **266:** 19867–19870.

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.

Krauss, S., T. Johansen, V. Korzh, and A. Fjose. 1991a. Expression of the zebrafish paired box gene *pax[zf-b]* during early neurogenesis. *Development* **113:** 1193–1206.

Krauss, S., T. Johansen, V. Korzh, U. Moens, J.U. Ericson, and A. Fjose. 1991b. Zebrafish *pax[zf-a]*: A paired box-containing gene expressed in the neural tube. *EMBO J.* **10:** 3609–3619.

Liao, F., S.L. Giannini, and B.K. Birshtein. 1992. A nuclear DNA-binding protein expressed during early stages of B-cell differentiation intereacts with diverse segments within and 3' of the IgH chain gene cluster. *J. Immunol.* **148:** 2909–2917.

Nornes, H.O., G.R. Dressler, E.W. Knapik, U. Deutsch, and P.

Gruss. 1990. Spatially and temporally restricted expression of *Pax2* during murine neurogenesis. *Development* **109:** 797–809.

Plachov, D., K. Chowdhury, C. Walther, D. Simon, J.L. Guenet, and P. Gruss. 1990. *Pax8*, a murine paired box gene expressed in the developing excretory system and thyroid gland. *Development* **110:** 643–651.

Rolink, A. and F. Melchers.1991. Molecular and cellular origins of B lymphocyte diversity. *Cell* **66:** 1081–1094.

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 transcripts. *Mol. Cell. Biol.* **11:** 5551–5561.

Studier, W., A.H. Rosenberg, J.J. Dunn, and J.W. Dubendorff. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185:** 60–89.

Tanaka, M. and W. Herr. 1990. Differential transcriptional activation of Oct-1 and Oct-2: Interdependent activation domains induce Oct-2 phosphorylation. *Cell* **60:** 375–386.

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.

Ton, C.C.T., H. Hirvonen, H. Miwa, M.M. Weil, P. Monaghan, T. Jordan, V. van Heyningnen, 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 homebox-containing gene from the aniridia region. *Cell* **57:** 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.

Tyndall, C., G. La Mantia, C.M. Thacker, J. Favaloro, and R. Kamen. 1981. A region of the polyoma virus genome between the replication origin and late protein coding sequences is required in cis for both early gene expression and viral DNA replication. *Nucleic Acids. Res.* **9:** 6231–6250.

Vitelli, L., I. Kemler, B. Lauber, M.L. Birnstiel, and M. Busslinger. 1988. Developmental regulation of micro-injected histone genes in sea urchin embryos. *Dev. Biol.* **127:** 54–63.

Wagner, M. and R.P. Perry. 1985. Characterization of the multigene family encoding the mouse S16 ribosomal protein: Strategy for distinguishing an expressed gene from its processed pseudogene counterparts by an analysis of total genomic DNA. *Mol. Cell. Biol.* **5:** 3560–3576.

Walther, C. and P. Gruss. 1991. *Pax-6*, a murine paired box gene, is expressed in the developing CNS. *Development* **113:** 1435–1449.

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.H., K.U. Saikh, and J. Stavnezer. 1989. A B-cell-specific nuclear protein that binds to DNA sites 5′ to immunoglobulin Sα tandem repeats is regulated during differentiation. *Mol. Cell. Biol.* **9:** 5594–5601.

Weichselbraun, I., G. Haider, and E. Wintersberger. 1989. Optimal replication of plasmids carrying polyomavirus origin regions requires two high-affinity binding sites for large T antigen. *J. Virol.* **63:** 961–964.

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.

Wray, W., T. Boulikas, V.P. Wray, and R. Hancock. 1981. Silver staining of proteins in polyacrylamide gels. *Anal. Biochem.* **118:** 197–203.

Xu, L., M.G. Kim, and K.B. Marcu. 1992. Properties of B-cell stage specific and ubiquitous nuclear factors binding to immunoglobulin heavy chain gene switch regions. *Int. Immunol.* **4:**(in press).

Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis.

B Adams, P Dörfler, A Aguzzi, et al.

References	This article cites 51 articles, 19 of which can be accessed free at: **http://genesdev.cshlp.org/content/6/9/1589.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.**