The EMBO Journal vol.6 no.9 pp.2719–2725, 1987

The highly conserved amino-terminal region of the protein encoded by the v-*myb* oncogene functions as a DNA-binding domain

Karl-Heinz Klempnauer and Albrecht E.Sippel

Zentrum für Molekulare Biologie (ZMBH) der Universität Heidelberg, Im Neuenheimer Feld 282, D-6900 Heidelberg, FRG

Communicated by T.Graf

The retroviral oncogene v-*myb* encodes a 45 000 M_r nuclear protein ($p45^{v\text{-}myb}$) that is predominantly associated with the chromatin of transformed cells. It has previously been shown that $p45^{v\text{-}myb}$, when released from chromatin by salt-treatment, binds to DNA. To analyse the biochemical properties of $p45^{v\text{-}myb}$ in more detail we have expressed the v-*myb* coding region in *Escherichia coli*. Our results demonstrate that bacterially expressed *myb* protein has an intrinsic DNA-binding activity. Using two alternative strategies, (i) inhibition of DNA-binding by monoclonal antibodies and (ii) analysis of DNA-binding activities of partially deleted forms of the bacterial *myb* protein, we show that the DNA-binding domain is located in the amino-terminal region of the v-*myb* protein. This region has been highly conserved between *myb* genes of different species. Our results are therefore consistent with the hypothesis that DNA-binding is an important aspect of *myb* protein function.

Key words: v-*myb*/nuclear oncogenes/bacterial expression/DNA-binding domain

Introduction

The oncogene v-*myb* of avian myeloblastosis virus (AMV), an acutely oncogenic chicken retrovirus, encodes a 45 000-dalton nuclear protein ($p45^{v\text{-}myb}$) that has been implicated in the ability to AMV to transform cells of the myelomonocytic lineage both *in vivo* and *in vitro* (Souza *et al.*, 1980; Gonda *et al.*, 1981; Moscovici and Gazzolo, 1982; Rushlow *et al.*, 1982; Boyle *et al.*, 1983; Klempnauer *et al.*, 1983; Moscovici and Moscovici, 1983). $p45^{v\text{-}myb}$ is a truncated and mutated version of the 75 000-dalton protein ($p75^{c\text{-}myb}$) encoded by the chicken c-*myb* gene (Klempnauer *et al.*, 1982, 1983, 1984). c-*myb* expression has been detected primarily in immature hematopoietic cells and in growth-arrested fibroblasts following serum stimulation (Gonda *et al.*, 1982; Thompson *et al.*, 1986). c-*myb* has been conserved during evolution as homologous genes have been detected in man, mouse, chicken and the fruit-fly *Drosophila melanogaster* (Gonda and Bishop, 1983; Leprince *et al.*, 1983; Gonda *et al.*, 1985; Katzen *et al.*, 1985; Slamon *et al.*, 1986). Comparison of the nucleotide sequences of these genes has shown that a short region close to the amino terminus of the *myb* proteins has been conserved to a higher degree than the rest of the protein (Gonda *et al.*, 1985; Katzen *et al.*, 1985). This highly conserved domain is rich in positively or negatively charged amino acids and consists of three imperfect direct repeats of 50 amino acids each (Ralston and Bishop, 1983; Gonda *et al.*, 1985; Rosson and Reddy, 1986). Because of its high degree of conservation this domain is presumed to play an important role in *myb* protein function, although the nature of that function, at the moment, is not known. The conserved domain is also present at the amino terminus of the viral *myb* protein, however, part of the first of the three repeats has been lost during retroviral transduction and truncation of the v-*myb* gene. Biochemical analysis of proteins encoded by v-*myb* and c-*myb* has shown that they predominantly are associated with the chromatin (Boyle *et al.*, 1985; Klempnauer and Sippel, 1986; Klempnauer *et al.*, 1986). When released from chromatin by salt treatment *myb* proteins readily bind to DNA suggesting that they are DNA-binding proteins (Moelling *et al.*, 1985; Klempnauer and Sippel, 1986; Klempnauer *et al.*, 1986). However, it is not clear whether *myb* protein DNA-binding is mediated by other cellular proteins or whether DNA-binding is sequence specific. To facilitate a detailed analysis of the DNA-binding properties of *myb* proteins we have inserted the v-*myb* coding region into a plasmid vector that allows expression of foreign gene products in *Escherichia coli*. We here show that bacterially synthesized viral *myb* protein has an intrinsic general DNA-binding activity. Since no *myb* target gene that might contain a *myb*-specific binding site is known, it is presently not possible to analyse whether *myb* protein DNA-binding is sequence specific. As an alternative approach to analyse the role of DNA-binding in *myb* protein function we have mapped the general DNA-binding domain within the v-*myb* protein. Using two different experimental procedures, inhibition of DNA-binding by *myb*-specific monoclonal antibodies and analysis of DNA-binding activities of partially deleted versions of the v-*myb* protein, we have localized the DNA-binding domain within the amino-terminal region of the v-*myb* protein. Since this region has been highly conserved during evolution we propose that DNA-binding is important for *myb* protein function.

Results

Expression of v-myb in bacteria

In an effort to obtain large quantities of the protein encoded by the v-*myb* oncogene of AMV we have inserted the v-*myb* coding region into a plasmid vector designed to express foreign gene products in *E. coli*. Figure 1A outlines the contructions of the pVM2028 v-*myb* expression vector; details are described in Materials and methods. pVM2028 encodes an almost full-length *myb* protein that differs from authentic $p45^{v\text{-}myb}$ only at the amino terminus. In the bacterial protein the first seven amino acid residues (the first six of which are encoded by nucleotide sequences derived from the viral *gag* gene) are replaced by 13 amino acid residues derived from the tryptophan operon of *E. coli*. Bacteria that carry the pVM2028 expression plasmid synthesized a novel 45 000-dalton protein, referred to as $bp45^{v\text{-}myb}$, when transcription starting at the *trp* promoter was induced by β-indolacrylic acid (IAA). The identity of $bp45^{v\text{-}myb}$ was substantiated by several observations. (i) $bp45^{v\text{-}myb}$ reacted with several different *myb*-specific monoclonal antibodies (Evan *et al.*, 1984; Klempnauer *et al.*, 1986). A western blot analysis using the *myb*2-32 monoclonal antibody is shown in Figure 1B (lane 7) as an example. (ii) When the pVM2028 v-*myb* expression

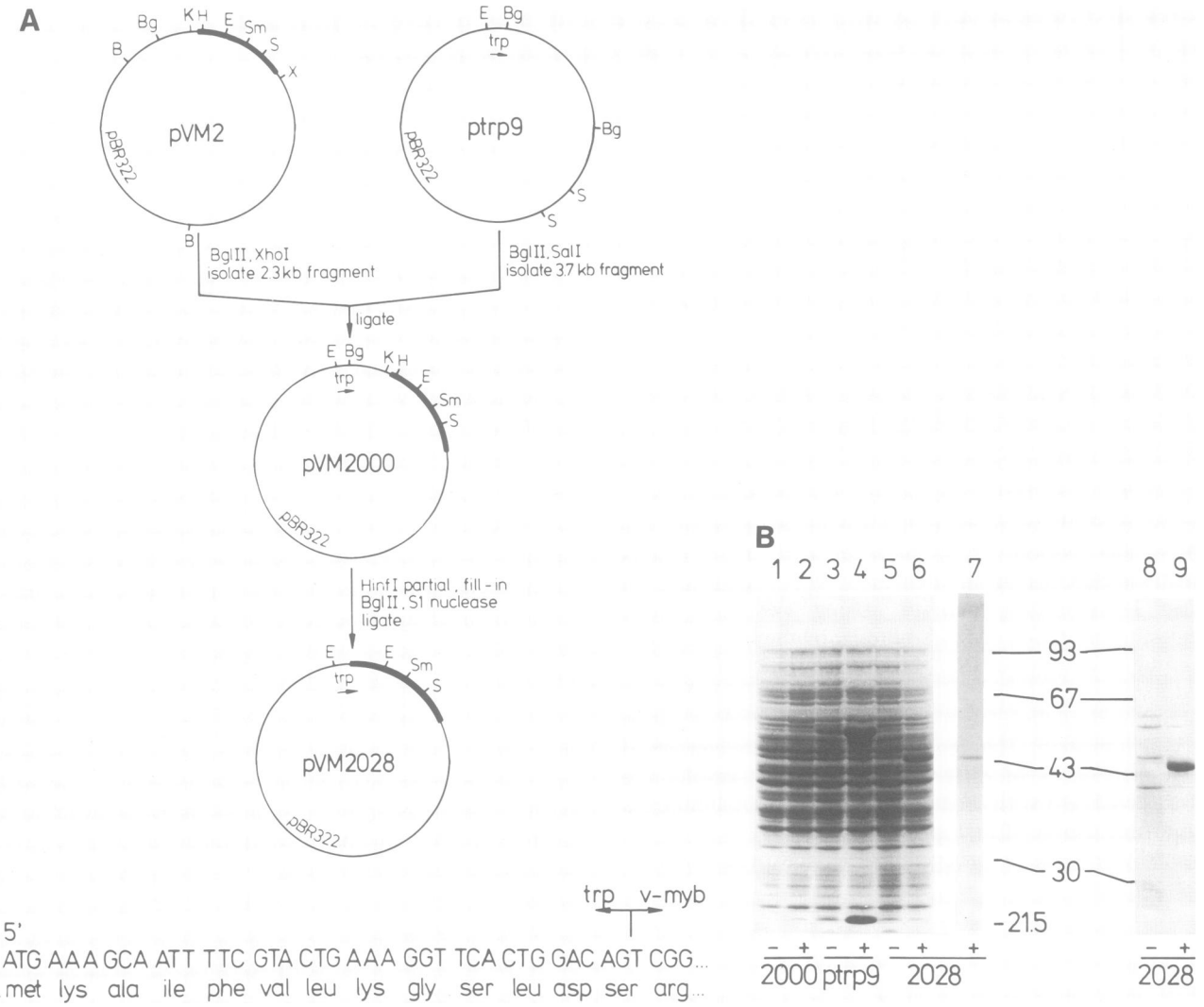

Fig. 1. Expression of v-*myb* in bacteria. **(A)** Construction of the pVM2028 *myb* expression vector. Details of the construction of plasmid pVM2028 are described in Materials and methods. The thick lines represent the v-*myb* gene. H marks a *Hin*fI restriction site, GAATC, located at the 5′ boundary of the v-*myb* coding region. The first nucleotide, G, of this site is identical with the last nucleotide of a TAG translational stop codon that precedes the v-*myb* coding region. The following three nucleotides of the *Hin*fI site, AAT, comprise the first codon of the v-*myb* coding sequence. Other restriction sites marked are: E; *Eco*RI, B; *Bam*HI, K; *Kpn*I, Sm; *Sma*I, S; *Sal*I, X; *Xho*I, Bg; *Bgl*II. trp marks the position of the tryptophan promoter; the direction of transcription is indicated by an arrow. pBR322 refers to sequences derived from pBR322. The partial nucleotide sequence and deduced amino acid sequence of pVM2028 is shown at the bottom. The sequence starts at the translational start codon of the trp−*myb* fusion protein encoded by pVM2028. The junction between sequences derived from the tryptophan operon and from v-*myb* is marked by arrows. The first codon of the v-*myb* coding region, AAT, is not present in pVM2028. The last codon shown, CGG, is identical with the second codon of the v-*myb* coding region. **(B)** Analysis of bacterial protein. Total protein from bacteria that carried various plasmids was fractionated in SDS−polyacrylamide gels and stained with Coomassie brilliant blue (**1−6**) or blotted onto nitrocellulose and stained with the *myb*-specific monoclonal antibody *myb*2-32 (**7**). Insoluble protein prepared from bacteria that carried the pVM2028 *myb* expression plasmid was fractionated in a SDS−polyacrylamide gel and stained with Coomassie brilliant blue (**8,9**). Each lane shows protein from bacteria contained in 50 μl (**1−7**) or 100 μl (**8,9**) of the original culture. − and + mark lanes that represent protein from bacteria grown in the absence or presence of IAA. The plasmids carried by the bacteria are indicated below the lanes. Sizes are indicated in kds.

plasmid was modified in order to produce truncated versions of bp45$^{v\text{-}myb}$ (as described in Materials and methods) bacteria that carried the modified expression plasmids no longer synthesized bp45$^{v\text{-}myb}$; instead smaller proteins of the expected sizes were detected (see Figure 5A). (iii) Antiserum raised against one of the truncated v-*myb* proteins immunoprecipitated authentic p45$^{v\text{-}myb}$ and p75$^{c\text{-}myb}$ (Klempnauer *et al.*, 1986). bp45$^{v\text{-}myb}$ was highly insoluble when bacteria expressing it were treated with lysozyme and extracted with buffers containing non-ionic detergent Nonidet P-40 (NP-40) and 0.4 M NaCl (see Figure 1B, lanes 8 and 9). The insolubility of bp45$^{v\text{-}myb}$ enabled us to separate it from most of the *E. coli* proteins as described in Materials and methods. Most of the experiments described here were performed using bacterially expressed *myb* protein partially purified in this manner.

DNA-binding of bacterially expressed v-myb protein

Subnuclear localization experiments have shown that *myb* proteins are associated mainly with the chromatin and, when released from nuclei by salt treatment, bind to DNA cellulose (Boyle *et al.*, 1985; Klempnauer and Sippel, 1986; Klempnauer *et al.*, 1986). It is not clear, however, whether DNA binding is mediated by other cellular factors or whether *myb* proteins themselves have intrinsic DNA-binding activities. To determine whether bacterially expressed v-*myb* protein binds to DNA we performed the DNA−cellulose chromatography experiment shown in Figure 2. Insoluble, partially purified bp45$^{v\text{-}myb}$ was first solubilized by

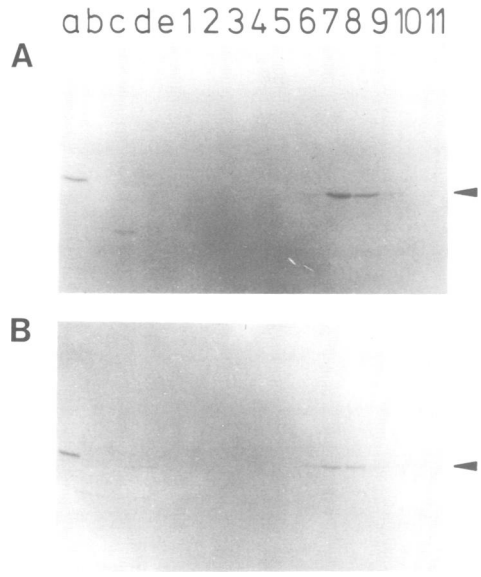

Fig. 2. DNA−cellulose chromatography of *myb* proteins. Protein extracted by 0.4 M NaCl from ~10^8 nuclei of [^{35}S]methionine-labeled BM2 cells was mixed with ~2 μg of soluble bacterially expressed *myb* protein, obtained from a 1 ml culture of bacteria carrying pVM2028, and chromatographed on native calf thymus DNA−cellulose. Samples representing one quarter of the total protein applied to the column were removed before mixing and were analysed individually. **Panel A:** the distribution of radiolabeled p45$^{v\text{-}myb}$ was determined by immunoprecipitates with polyclonal *myb*-specific rabbit antiserum followed by electrophoresis of the immunoprecipitation in an SDS−polyacrylamide gel and autoradiography. **Lane a** shows the sample containing p45$^{v\text{-}myb}$, **lane b** shows the sample containing bacterially expressed *myb* protein, prior to mixing. **Lane c** represents the flow through and **lanes d** and **e** represent the pooled wash fractions, respectively. Numbered lanes mark individual fractions of a NaCl gradient. The NaCl concentration in fraction 7 is ~300 mM. **Panel B:** fractions were electrophoresed in a 10% SDS−polyacrylamide gel, blotted onto nitrocellulose and *myb* protein was visualized by indirect immunostaining using the *myb*-specific monoclonal antibody *myb*2-32. **Lanes a** and **b** are reversed relative to **panel A**−**lane a** (of **panel B**) shows the sample containing only bacterially expressed *myb* protein, **lane b** shows the sample containing radiolabeled p45$^{v\text{-}myb}$. All other lanes are as in **panel A**. The arrows indicate the positions of p45$^{v\text{-}myb}$ (**panel A**) and bp45$^{v\text{-}myb}$ (**panel B**).

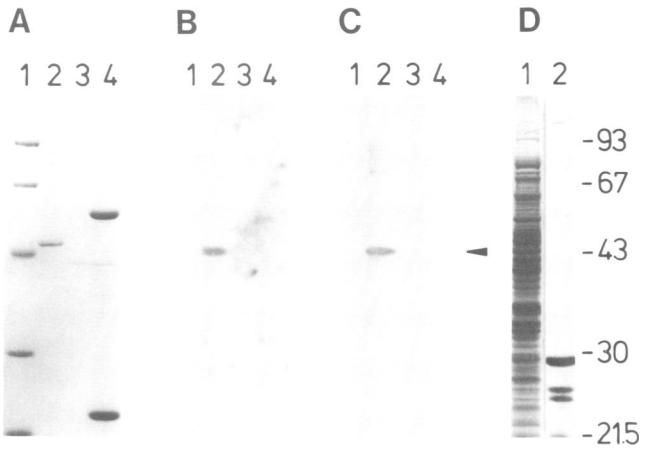

Fig. 3. Bacterially expressed *myb* protein binds to DNA. **Panels A−C**: insoluble protein, prepared from bacteria that carried the plasmids pVM2028 (**2,3**) or ptrp9 (**4**) and were grown in the presence (**2,4**) or absence (**3**) of IAA, was analysed by electrophoresis in SDS−polyacrylamide gels. **Lane 1** shows a set of different proteins used as mol. wt markers, namely (from top): phosphorylase B; bovine serum albumin; ovalbumin; carbonic anhydrase and soybean trypsin inhibitor. Proteins were visualized by staining with Coomassie brilliant blue (**A**). An identical gel was blotted onto nitrocellulose and the blot was incubated with ^{32}P end-labeled DNA fragments and analysed by autoradiography (**B**). The same blot was then analysed by indirect immunostaining using the *myb*-specific monoclonal antibody *myb*2-32 (**C**). The arrow marks the position of bp45$^{v\text{-}myb}$. The two intense bands visible in **lane 4** of **panel A** are the *trpD* (upper band) and the *trpE* (lower band) gene products. **Panel D**: total bacterial protein (from 50 μl of culture) was fractionated in a SDS−polyacrylamide gel and stained with Coomassie brilliant blue (**1**) or blotted onto nitrocellulose, incubated with ^{32}P labeled DNA fragments and analysed by autoradiography (**2**). Sizes are indicated in kds.

dissolving it in 6 M urea and diluting the resulting solution 50-fold, as described for several other proteins (Hager and Burgess, 1980). After dilution, bacterially expressed v-*myb* protein remained soluble and was co-chromatographed on native calf thymus DNA−cellulose together with authentic p45$^{v\text{-}myb}$ salt-extracted from nuclei of [^{35}S]methionine-labeled AMV-transformed myeloblasts. Radiolabeled p45$^{v\text{-}myb}$ was detected by immunoprecipitation and autoradiography, whereas bacterially expressed bp45$^{v\text{-}myb}$ was visualized by immunoblotting using the *myb*-specific monoclonal antibody *myb*2-32. Control experiments showed that the amount of p45$^{v\text{-}myb}$ extracted from myeloblasts is very low compared with the amount of bacterially expressed *myb* protein used in this experiment (compare lanes a and b of Figure 2B); immunoblotting therefore primarily revealed the bacterial protein. Figure 2A illustrates that authentic v-*myb* protein binds to DNA−cellulose and is eluted from the column at ~300 mM NaCl, as we have demonstrated before (Klempnauer and Sippel, 1986). Bacterially expressed *myb* protein showed essentially the same behavior, both in the presence (as shown in Figure 2B) and in the absence of nuclear extract (data not shown). This suggested (i) that the v-*myb* protein itself has an intrinsic DNA-binding activity and (ii) that replacement of seven amino-terminal amino acids of p45$^{v\text{-}myb}$ by 15 amino acids derived from the *E. coli trp* operon does not grossly affect DNA-binding of the protein.

To substantiate our conclusion that the v-*myb* protein has an intrinsic DNA-binding activity, partially purified bp45$^{v\text{-}myb}$ was electrophoresed in a SDS−polyacrylamide gel, blotted onto nitrocellulose and incubated with radiolabeled DNA. As shown in Figure 3, bp45$^{v\text{-}myb}$, but not any of the other proteins presents on the blot, bound DNA. Note that one of these proteins, the 25 000-dalton trpE protein, has the same amino-terminus as bp45$^{v\text{-}myb}$, yet it is inactive in DNA binding. It is therefore unlikely that we have artificially linked a DNA-binding domain to v-*myb* sequences by joining the v-*myb*-coding region to the start of the *trpE* gene. Binding of DNA to bp45$^{v\text{-}myb}$ in this assay was strongly dependent on the salt concentration. Binding was decreased ~4-fold at 150 mM NaCl and was virtually undetectable when the NaCl concentration was raised above 300 mM (data not shown). To illustrate the specificity of this DNA-binding assay we analysed total bacterial protein by this procedure (Figure 3D). Our results show that only a few of the *E. coli* proteins bind DNA under these conditions.

Mapping the DNA-binding domain of viral myb protein

To address the role of DNA-binding in *myb* protein function we intended, as a first step, to map the location of the DNA-binding domain within the v-*myb* protein, especially with respect to amino acid sequences at the amino terminus of the protein that have been highly conserved throughout evolution. We first salt-extracted p45$^{v\text{-}myb}$ from nuclei of [^{35}S]methionine-labeled AMV-transformed myeloblasts and incubated it with *myb*-specific

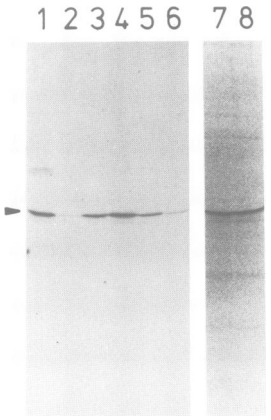

Fig. 4. Inhibition of DNA-binding of p45$^{v\text{-}myb}$ by monoclonal antibodies. Proteins extracted from ~5 × 10^7 nuclei of [^{35}S]methionine-labeled BM2 cells by 0.4 M NaCl were incubated with 10 μg of the monoclonal antibody *myb*2-2 (**1**–**3**) or CB100-18 (**4**–**6**) before they were analysed by DNA–cellulose chromatography. p45$^{v\text{-}myb}$ present in the original nuclear extract (**1**,**4**), not bound to DNA cellulose (**2**,**5**) or eluted from DNA–cellulose by 0.5 M NaCl (**3**,**6**) was visualized by immunoprecipitation with polyclonal *myb*-specific rabbit antiserum followed by electrophoresis in a SDS–polyacrylamide gel and autoradiography. **Lanes 7** and **8**: immunoprecipitation of p45$^{v\text{-}myb}$ from nuclear extract of 4 × 10^7 [^{35}S]methionine-labeled BM2 cells by monoclonal antibody CB100-18 (**7**) or *myb*2-2 (**8**). The arrow marks the position of p45$^{v\text{-}myb}$.

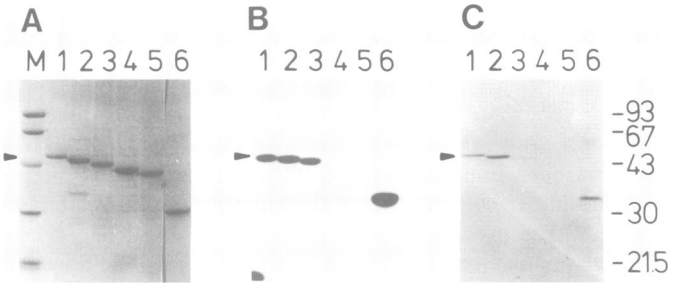

Fig. 5. DNA-binding activities of truncated *myb* proteins. Insoluble proteins prepared from bacteria that carried the plasmids pVM2028 (**1**), pVM2001 (**2**), pVM2016 (**3**), pVM2022 (**4**), pVM2013 (**5**) or pVM2062 (**6**) and were grown in the presence of IAA, were fractionated in an SDS–polyacrylamide gel and stained with Coomassie brilliant blue (**A**). Identical gels were blotted onto nitrocellulose and incubated with ^{32}P-labeled nick-translated pUC18 DNA (**B**), or stained with the *myb*-specific monoclonal antibody CB100-18 (**C**). **Lane M** of **panel A** shows a set of proteins used as mol. wt markers. Sizes are indicated in kds.

monoclonal antibodies binding to different parts of the v-*myb* protein. The epitope of the CB100-18 antibody lies close to the amino terminus of p45$^{v\text{-}myb}$, whereas the *myb*2-2 epitope is located in the middle of the protein (see Figure 7). p45$^{v\text{-}myb}$–antibody complexes were then analysed by DNA–cellulose chromatography. Antibody *myb*2-2 did not interfere with binding of the v-*myb* protein to DNA, whereas CB100-18 antibodies substantially inhibited binding of the protein to DNA (Figure 4). A control experiment showed that both antibodies reacted efficiently with the v-*myb* protein (lanes 7 and 8 of Figure 4). One possible explanation of this experiment is that amino acid sequences that are involved in DNA binding lie close to the amino terminus of the v-*myb* protein. Alternatively, it is also possible that the CB100-18 epitope is actually far away from the DNA-binding site and that CB100-18 antibodies interfere with DNA binding by causing a conformational change over some distance in the

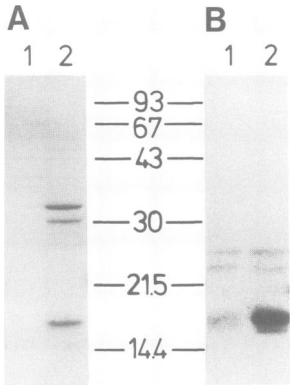

Fig. 6. The conserved amino-terminal domain of viral *myb* protein is sufficient for DNA binding. Insoluble proteins prepared from bacteria that carried the plasmid pVM2074 and were grown in the absence (**1**) or presence (**2**) of IAA, were fractionated in a SDS–polyacrylamide gel and stained with Coomassie brilliant blue (**A**). An identical gel was blotted onto nitrocellulose and incubated with ^{32}P-labeled nick-translated pUC18 DNA (**B**). Sizes are indicated in kds.

v-*myb* protein resulting in inefficient DNA binding. Therefore, we mapped the location of the DNA-binding domain more precisely by analysing DNA-binding activities of partially deleted versions of bacterially expressed v-*myb* protein. Figure 5A shows a set of bacterially expressed v-*myb* proteins deleted from the amino terminus (lanes 2–5) or from the carboxy terminus (lane 6). The construction of the deletions is described in Materials and methods and the exact extent of the deletions is shown in Figure 7. The proteins were partially purified as insoluble material and analysed by the DNA-binding assay described in Figure 3. The result of this assay (Figure 5B) indicated that removal of up to 45 amino acids from the amino terminus does not interfere with DNA-binding of the v-*myb* protein. Deleting 68 amino acids, however, completely abolished the DNA-binding activity of the protein. By contrast, deletion of a large portion from the carboxy terminus did not destroy the DNA-binding activity. A Western blotting experiment of the same set of bacterial proteins using CB100-18 antibodies for immunostaining showed that only the full-length and the largest of the amino-terminally deleted proteins react well with the antibody (Figure 5C). This result supports our conclusion that the CB100-18 epitope lies very close to the DNA-binding site and suggests that sequences from the amino-terminal region of p45$^{v\text{-}myb}$ are critical for DNA binding. To determine how extensive these sequences are we created a large carboxy-terminal deletion by removing approximately two thirds of the amino acid sequence of p45$^{v\text{-}myb}$. The construction of the pVM2074 expression vector encoding the deleted protein is described in Materials and methods and the extent of the deletion is shown in Figure 7. This plasmid differs from the other expression vectors described here in that transcription starting at the *trp* promoter is directed towards the ampicillin resistance gene present on the plasmid. This clone was chosen because, in this orientation, the open reading frame present on the inserted DNA fragment (which itself does not contain an in-frame translational stop codon) is terminated three nucleotides beyond the end of the v-*myb*-coding sequence. A single amino acid residue is therefore fused to the amino-terminal domain of p45$^{v\text{-}myb}$; this residue is identical to the amino acid that, in p45$^{v\text{-}myb}$, resides at that position. Because IAA-induced transcription in the clone pVM2074 is directed towards the ampicillin resistance gene, expression of β-lactamase encoded by that gene is also induced by

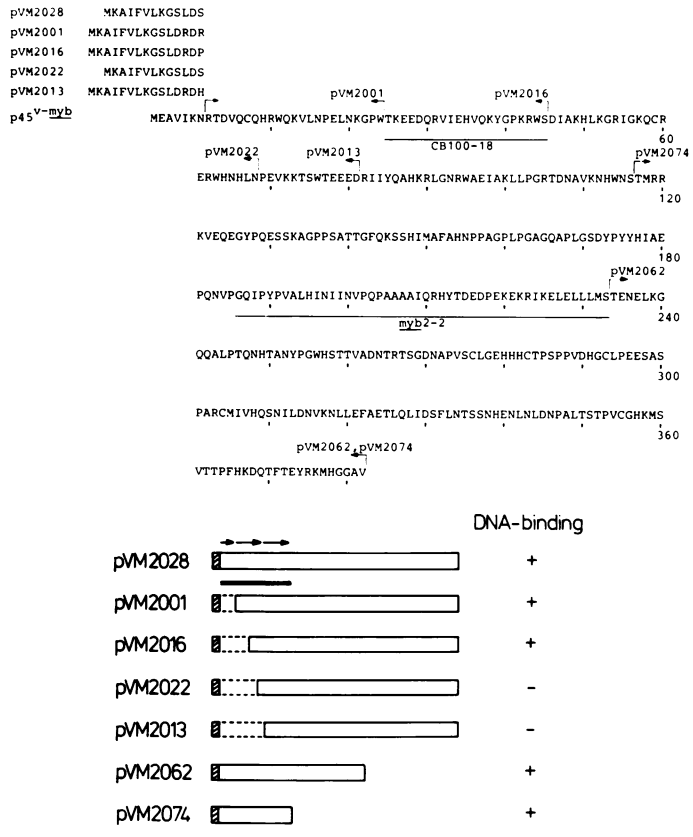

Fig. 7. The structure of bacterially expressed v-*myb* proteins. The deduced amino acid sequence of p45$^{v\text{-}myb}$ (Klempnauer *et al.*, 1982) is shown in the single-letter amino acid code at the top. The numbering starts at the first amino acid encoded by the v-*myb* open reading frame In p45$^{v\text{-}myb}$, this amino acid is preceded by six amino acid residues encoded by the viral *gag* gene (Schwartz *et al.*, 1983). In bacterially expressed v-*myb* proteins the first seven amino acid residues of p45$^{v\text{-}myb}$ are replaced by the sequences shown above. The amino-terminal deletions start after the seventh residue of p45$^{v\text{-}myb}$, as indicated by an arrow. In the carboxy-terminal deletions pVM2062 and pVM2074 the amino-terminal sequence is the same as in pVM2028. The extent of individual deletions is indicated by arrows pointing towards the deleted sequence. The lines indicate the approximate locations of epitopes of monoclonal antibodies. The position of the CB100-18 epitope was deduced from the results shown in Figure 5c. The leftward boundary of the region encompassing the *myb*2-2 epitope was inferred from the observation that a carboxy-terminal deletion extending to this position resulted in loss of the *myb*2-2 epitope (the construction of this deletion, pVM2060, has been described by Klempnauer *et al.*, 1986). The rightward boundary of the *myb*2-2 epitope is defined by clone pVM2062, which has not deleted this epitope (data not shown). The structure of bacterially expressed *myb* proteins and their DNA-binding activities are summarized schematically below. Amino acid sequences derived from ptrp9 or v-*myb* are shown as hatched or open boxes, respectively. Sequences not present in partially deleted proteins are shown by broken boxes. The black bar shows the highly conserved region at the amino terminus of p45$^{v\text{-}myb}$ and the arrows represent directly repeated sequence motifs in that region. Only part of the first repeat is present in the v-*myb* protein.

IAA, leading to the appearance of additional protein species (~30 000 daltons) in preparations of insoluble protein from bacteria that carry the pVM2074 plasmid (see Figure 6A). The 15 000-dalton protein visible in lane 2 of Figure 6 is the extensively truncated *myb* protein since it has the expected size and reacts with the *myb*-specific monoclonal antibody CB100-18 (data not shown). The DNA-binding assay (Figure 6B) showed that the 15 000-dalton protein still has retained the ability to bind to DNA, suggesting that the DNA-binding domain is confined to a region of ~120 amino acids at the amino terminus of the v-*myb* protein. Figure 7 summarizes our results with respect to the amino acid sequence of p45$^{v\text{-}myb}$.

In an attempt to define more precisely the sequences within the DNA-binding region that are necessary for DNA-binding we constructed a set of internal deletions of v-*myb* removing various amounts of sequence from the carboxy-terminal side of the DNA-binding region. We chose to construct internal deletions (with an identical piece from the carboxy terminus of p45$^{v\text{-}myb}$ retained in each deletion) in order to increase the size of the deleted proteins and thus avoid possible problems in the DNA-binding assay caused by background bands in the low mol. wt region of the gel. However, various attempts to express such internally deleted proteins in sufficient quantities failed, becaused these proteins are not stable in our *E. coli* expression system (data not shown, see Materials and methods for details).

Discussion

We have used a bacterial expression system to obtain large amounts of a protein encoded by the v-*myb* oncogene of AMV. The experiments described here demonstrate that the *myb* protein synthesized in bacteria binds to DNA. In this protein the first seven amino acids of authentic p45$^{v\text{-}myb}$ (the first six of which are encoded by the *gag* gene of AMV) have been replaced by 13 amino acid residues derived from the *E. coli* tryptophan operon. It is unlikely that this replacement grossly affects the DNA-binding activity of the v-*myb* protein, because, in DNA−cellulose chromatography experiments, bacterially expressed and authentic v-*myb* proteins behave identically [both proteins are released from the column at a salt concentration (~300 mM NaCl) which also releases the majority of DNA-binding proteins present in nuclear extracts of eucaryotic cells (data not shown)]. Furthermore, the v-*myb* gene of avian leukemia virus E26 lacks 27 nucleotides that are present at the 5′ end of the v-*myb* gene of AMV (and has replaced them by other sequences derived from the retroviral *gag* gene), yet the E26-*myb* protein binds to DNA (Nunn *et al.*, 1984; Moelling *et al.*, 1985). This suggests that the first nine amino acid residues encoded by the v-*myb* gene of AMV are dispensable for DNA-binding. This conclusion is extended by our observation that up to 45 amino acids can be removed from the amino terminus of the v-*myb* protein without abolishing its DNA-binding activity. Bacterially expressed v-*myb* protein is essentially pure with respect to other eucaryotic proteins; our results therefore strongly suggest that the DNA-binding activity is intrinsic to v-*myb* and does not require association with other proteins. This does not exclude the possibility that the v-*myb* protein interacts with other proteins and that such association may influence the ability of the protein to bind to DNA.

Inhibition of DNA binding of p45$^{v\text{-}myb}$ by *myb*-specific monoclonal antibodies and deletion analysis of bacterially expressed *myb* protein has enabled us to map the DNA-binding domain to a region of ~120 amino acid residues at the amino terminus of the v-*myb* protein. The observation that this region, even when expressed independently of other v-*myb* sequences, is active in DNA binding, supports the notion that it represents a relatively self-contained structural domain within the v-*myb* protein. The repetitive structure of the amino-terminal region of the v-*myb* protein superficially resembles the composite DNA-binding domain of the *Xenopus laevis* transcription factor TFIIIA (Miller *et al.*, 1985). Although there are structural differences between the DNA-binding regions of *myb* proteins and TFIIIA, such as the size of the repeated unit (50 amino acids in *myb* proteins and

30 amino acids in TFIIIA) and the absence, in *myb* proteins, of characteristic histidine and cysteine residues implicated in zinc-dependent stabilization of the TFIIIA DNA-binding domain, the interesting question is raised of whether both of the two complete repeats present in the v-*myb* protein represent distinct DNA-binding sites possibly acting in concert. The results presented here provide no clear answer. Our observations show that (i) monoclonal antibodies CB100-18 inhibit DNA binding although they bind to only one of the repeats and (ii) DNA-binding activity is lost as the first of the two repeats present in the protein is deleted (This results in a protein, encoded by pVM2022, that still contains the second of the two repeats but no longer binds DNA). While this clearly indicates that the first of the two repeats present in the v-*myb* protein is essential for DNA binding, we cannot rule out the possibility that the second repeat also interacts with DNA. If the two repeats would bind independently to DNA and one is removed (or blocked by an antibody) then the association constant will drop and no binding might be detected. This would be even more so if the two domains were involved in cooperative binding. We attempted to determine whether the second complete repeat of v-*myb* also contributes to DNA binding of the v-*myb* protein by constructing proteins that have deleted all or part of the second repeat but retain the first repeat. These attempts failed, however, because no stable protein was obtained in our expression system whenever sequences of the second complete repeat of v-*myb* were deleted (see Materials and methods).

The amino-terminal region of the v-*myb* protein has been highly conserved between *myb* genes of different species. In an extreme case, in the *myb* gene of *Drosophila melanogaster*, this region appears to be the only part of the gene that shows detectable homology to vertebrate *myb* genes (Katzen *et al.*, 1985). The observation that the DNA-binding domain lies within this highly conserved region predicts that the ability of *myb* proteins to bind to DNA also has been conserved throughout evolution. This seems to be the case, as we have recently shown that the c-*myb* proteins of chickens and humans bind to DNA (Klempnauer and Sippel, 1986; Klempnauer *et al.*, 1986). Furthermore, work not presented here has shown that the region of the *D. melanogaster* *myb* gene, which is homologous to vertebrate *myb* genes, also encodes a DNA-binding domain (C.Peters, A.E.Sippel, M.Vingron and K.-H.Klempnauer, in preparation). The observation that the DNA-binding domain of the v-*myb* protein is encoded by amino acid sequences that have been highly conserved during evolution leads us to speculate that DNA binding may be an important aspect of the molecular mechanism by which *myb* proteins act. This speculation is supported by the observation that DNA binding of *myb* proteins encoded by temperature-sensitive transformation-defective mutants of avian leukemia virus E26 appears to be temperature sensitive (Moelling *et al.*, 1985). However, it is obvious that further work needs to be done in order to understand the role of DNA binding in *myb* protein function. In the case of the v-*myb* and the c-*myb* proteins of chickens and humans the ability to bind to DNA is correlated with a predominant association of the proteins with chromatin (Klempnauer and Sippel, 1986; Klempnauer *et al.*, 1986) suggesting that *in vivo myb* proteins interact with cellular DNA and perhaps play regulatory roles at specific sites in the genome. Such a specific interaction possibly, but not necessarily, might be the result of sequence-specific DNA binding by *myb* proteins themselves. Because no genes or nucleotide sequences that might function as direct targets for *myb* proteins and thus might contain *myb*-specific binding sites are known, it is presently not possible to determine whether DNA binding by *myb* proteins is sequence specific. Experiments not presented here show that bacterially expressed v-*myb* protein does not bind equally well to any DNA sequence (H.Biedenkapp, A.E.Sippel and K.-H.Klempnauer, in preparation), suggesting that *myb* protein DNA binding exhibits sequence preference. Alternatively, it is possible that specific interaction of *myb* protein with cellular DNA results from association with other, as yet unknown, cellular proteins conferring specificity, or that *myb* protein function does not involve sequence-specific interaction. In any case, our observation that a DNA-binding domain has been conserved in *myb* proteins during evolution strongly emphasizes the importance of protein–DNA interaction in *myb* protein function. Finally, we cannot rule out the possibility that other, as yet undefined, biochemical activities, that are essential to *myb* protein function, also map to the amino-terminal domain of the v-*myb* protein. The existence of such activities, however, would not contradict our conclusions.

Materials and methods

Cell lines, radioactive labeling of cells and antisera

The BM2 line of AMV-transformed chicken myeloblasts was obtained from C.Moscovici. Cells were radiolabeled with [^{35}S]methionine as described previously (Klempnauer *et al.*, 1983). Polyclonal rabbit antiserum specific for the carboxy-terminal region of p45$^{v\text{-}myb}$ was prepared as described (Klempnauer *et al.*, 1983). The *myb*-specific mouse hybridoma lines *myb*2-2, *myb*2-32 and CB100-18 have been described (Evan *et al.*, 1984; Klempnauer *et al.*, 1986). Monoclonal antibodies were purified from supernatants of overgrown cultures by affinity chromatography using rabbit anti-mouse IgG antibodies coupled to Sepharose. For immunostaining, culture supernatant was used directly.

Construction of myb expression vectors

All standard molecular cloning techniques were carried out according to published procedures (Maniatis *et al.*, 1982). The ptrp9 expression plasmid as derived by minor modifications only affecting sequences flanking the tryptophan operon (K.-H.Klempnauer, unpublished data) from the pVVI expression vector described by Nichols and Yanofsky (1983). ptrp9 contains a deleted version of the *E. coli* tryptophan operon in which the coding regions of the trp leader peptide and the *trpE* gene product are fused. Upon induction of *trp* expression by growing the bacteria in the presence of IAA the products of the fused leader–*trpE* gene and the *trpD* gene are detected as abundant proteins in bacterial extracts (see Figure 1B, lanes 3 and 4). A 2.3-kbp *Bgl*II/*Xho*I restriction fragment of plasmid pVM2 (Klempnauer *et al.*, 1982) containing the complete v-*myb* gene flanked by viral sequences was inserted into the ptrp9 expression vector, which had been digested with *Bgl*II and *Sal*I, to obtain plasmid pVM2000 (see Figure 1). In order to fuse the v-*myb* coding region to the open reading frame provided by the expression vector we digested pVM2000 partially with *Hin*fI, filled in 5′ overhanging restriction ends and isolated full-length linearized molecules by gel electrophoresis. Note that there is a *Hin*fI restriction site precisely at the start of the v-*myb* coding region. The DNA was then digested with *Bgl*II, the 5′ overhanging restriction ends were trimmed with nuclease S1 and the DNA molecules were ligated and transfected into *E. coli* HB101. To identify plasmid molecules that direct expression of *myb* proteins we plated individual colonies onto agar plates containing L-broth supplemented with 50 μg/ml ampicillin and 40 μg/ml IAA. After 12 h at 37°C the colonies were transferred to nitrocellulose as described (Kemp *et al.*, 1983) and stained with monoclonal antibody *myb*2-32 as described (Klempnauer and Sippel, 1986). Bacteria expressing *myb* proteins were then grown in liquid culture in L-broth supplemented with 50 μg/ml ampicillin and 40 μg/ml IAA for 12 h at 37°C and the total bacterial protein was analysed by electrophoresis in SDS–polyacrylamide gels, blotted onto nitrocellulose and stained with monoclonal antibody *myb*2-32. Two colonies expressing *myb* proteins with molecular weights of 45 000 and 40 000, respectively, were selected for further analysis. Plasmid obtained from one of these colonies (pVM2028) was shown by DNA sequence analysis to direct expression of an almost complete v-*myb* coding region (one amino acid residue located at the 5′ end of v-*myb* has been lost) fused to 15 amino-terminal amino acids derived from the *trp* operon. The second plasmid, pVM2022, encodes a partially deleted *myb* protein because the *trp* promoter was linked to a *Hin*fI site within the v-*myb* coding region (see Figure 7). A set of plasmids encoding amino-terminally truncated *myb* proteins was obtained by digesting plasmid pVM2000 with *Kpn*I, subsequent treatment with various amounts of nuclease *Bal*31 followed by digestion with *Bgl*II, filling-in of 5′ overhanging restric-

tion ends and ligation. Colonies expressing *myb* proteins of various sizes were identified as described above. We selected three clones (pVM2001, pVM2013 and pVM2016) for further study. The extent of the amino-terminal deletion in each clone was determined by nucleotide sequence analysis (see Figure 7). To produce a carboxy-terminally truncated *myb* protein we digested pVM2028 with *Sal*I, filled in 5'-overhanging restriction ends and re-ligated the DNA to obtain colonies which now expressed a 31 000-dalton protein upon induction by IAA. The plasmid from one of these colonies is referred to as pVM2062. To produce a second carboxy-terminally truncated *myb* protein we isolated the 0.5-kb *Eco*RI fragment of pVM2028 (containing the *trp* promoter and the amino-terminal region of v-*myb*) and cloned it into ptrp9 digested with *Eco*RI and *Bgl*II. One of the *Eco*RI sites of the insert fragment was linked to the *Bgl*II site of the plasmid by a short *Eco*RI/*Bam*HI DNA fragment obtained from the polylinker region of plasmid pUC18. We selected a plasmid (pVM2074) in which the direction of transcription from the *trp* promoter is pointing towards the ampicillin resistance gene of pBR322. In this clone translation of v-*myb* coding region is terminated 3 bp beyond the end of the inserted DNA fragment. To produce internal deletions starting at various positions 5' to the *Eco*RI site in v-*myb* and ending at the *Sal*I site in v-*myb* (see Figure 1) pVM2028 DNA was first linearized with *Sma*I, then digested with various amounts of nuclease *Bal*31 and finally cleaved with *Pst*I. The heterogenous set of fragments corresponding to the 3' portion of the ampicillin resistance gene and the 5' portion of v-*myb* was isolated and ligated to pVM2028 DNA cut with *Pst*I and *Sal*I. Ampicillin-resistant colonies were expected to express v-*myb* proteins with internal deletions extending from various positions in the N-terminal region of v-*myb* to the *Sal*I site. These colonies were screened with monoclonal antibodies specific for amino-terminally (CB100-18) or carboxy-terminally (*myb*2-37; Evan *et al.*, 1984) located epitopes and the sizes of the expressed proteins from colonies positive for both of the antibodies were determined. The majority of colonies produced deleted proteins at levels incompatible with performing the nitrocellulose DNA-binding assay. Those few clones that produced sufficient amounts of protein turned out not to have the expected deletions. A similar cloning experiment was performed in which the internal deletions ended at a *Pst*I site, located ~80 bp upstream of the originally used *Sal*I site. The results were similar in that the correctly deleted clones produced too low levels of protein.

Partial purification of bacterially expressed myb protein

Bacteria carrying *myb* expression vectors were grown overnight at 37°C in L-broth in the presence or absence of 40 μg/ml IAA. Bacteria were pelleted and suspended in one fifth of the original volume of 10 mM Tris-HCl, pH 7.8; 1 mM EDTA; 50 mM NaCl and 0.5 mg/ml lysozyme, incubated on ice for 30–60 min, freeze-thawed once and sonicated three times for 15 s each. Insoluble material containing the *myb* protein was washed three times with 10 mM Tris-HCl, pH 7.8; 1 mM EDTA; 0.5 M NaCl; 0.5% NP-40 and three times with 10 mM Tris-HCl, pH 7.8; 1 mM EDTA; 0.5% NP-40. All washes were performed at room temperature. The final pellet was dissolved in sample buffer and analysed by SDS–polyacrylamide gel electrophoresis. Usually 5–10 μg of bacterial *myb* protein were obtained from a 5 ml culture. Full-length bacterial *myb* protein to be used for DNA–cellulose chromatography was solubilized by dissolving the final insoluble pellet in 6 M urea (usually 100 μl for a pellet from a 5 ml culture) for 10 min at 37°C, clarifying and diluting it 50-fold with 10 mM Tris-HCl, pH 7.8; 1 mM EDTA; 50 mM NaCl; 0.5 mg/ml bovine serum albumin (BSA). This solution was kept on ice for at least 12 h before performing DNA–cellulose chromatography.

Immunoblotting and nitrocellulose DNA-binding assay

Immunoblotting was performed as described (Klempnauer and Sippel, 1986) except that 5% (w/v) of non-fat dry milk powder was used as blocking agent. Incubation of blots with radiolabeled DNA followed the procedure described by Bowen *et al.* (1980). The whole procedure was carried out at room temperature. Blots were rinsed briefly with water and were then washed three times with 6 M urea; 0.2% NP-40 (20 min each wash), followed by four washes (30 min each) with DNA-binding buffer (10 mM Tris-HCl, pH 7.8; 1 mM EDTA; 50 mM NaCl). The blots were then incubated for 30 min in DNA-binding buffer supplemented with ^{32}P-labeled DNA. In different experiments we used end-labeled as well as nick-translated DNA to obtain identical results. End-labeled DNA, usually *Hin*fI restriction fragments of pBR322 DNA labeled at the 3' end with [α-^{32}P]deoxynucleotides (Amersham, >3000 Ci/mmol) and Klenow fragment of *E. coli* DNA polymerase I, were used at ~40–80 ng/ml binding buffer. Nick-translated DNA, usually pUC18 DNA labeled to a specific activity of $\sim 5 \times 10^7$ c.p.m./μg, was used at 10–20 ng/ml binding buffer. After incubation with DNA the blots were washed three times (10 min each) with DNA-binding buffer lacking DNA and analysed by autoradiography.

Other procedures

Preparation of nuclei from BM2 cells, salt extraction of proteins and analysis of nuclear extract by DNA–cellulose chromatography were performed as described previously (Klempnauer and Sippel, 1986; Klempnauer *et al.*, 1986). Immunoprecipitations were carried out as described (Klempnauer *et al.*, 1983). The immunoprecipitations shown in Figure 4, lanes 7 and 8, were performed by a modified procedure, because CB100-18 antibodies do not bind to protein A of *Staphylococcus aureus*. After incubation of nuclear extract with monoclonal antibodies p45$^{v\text{-}myb}$–antibody complexes were precipitated by adding an excess of rabbit anti-mouse IgG antibodies (Dako) prior to incubation with *S. aureus* cells.

Acknowledgements

We thank Hannelore Arnold for expert technical assistance, Conny Bonifer, Horst Biedenkapp, Michael Koenen, Christoph Peters and Joachim Nowock for valuable discussions, advice and help, Brian Nichols and Charles Yanofsky for the plasmid pVVI and Gerard Evan and Mike Bishop for the *myb*-specific hybridoma cells *myb*2-2 and *myb*2-32. This work was supported by grants from the Bundesministerium für Forschung und Technologie (BCT 0381/5) and the Fonds der Chemischen Industrie.

References

Bowen,B., Steinberg,J., Laemmli,U.K. and Weintraub,H. (1980) *Nucleic Acids Res.*, **8**, 1–20.

Boyle,W.J., Lipsick,J.S., Reddy,E.P. and Baluda,M.A. (1983) *Proc. Natl. Acad. Sci. USA*, **81**, 4265–4269.

Boyle,W.J., Lampert,M.A., Li,A.C. and Baluda,M.A. (1985) *Mol. Cell. Biol.*, **5**, 3017–3023.

Evan,G.I., Lewis,G.K. and Bishop,J.M. (1984) *Mol. Cell. Biol.*, **4**, 591–595.

Gonda,T.J. and Bishop,J.M. (1983) *J. Virol.*, **46**, 212–220.

Gonda,T.J., Sheiness,D.K., Fanshier,L., Bishop,J.M., Moscovici,C. and Moscovici,M.G. (1981) *Cell*, **23**, 279–290.

Gonda,T.J., Sheiness,D.K. and Bishop,J.M. (1982) *Mol. Cell. Biol.*, **2**, 617–624.

Gonda,T.J., Gough,N.M., Dunn,A.R. and de Blaquiere,J. (1985) *EMBO J.*, **4**, 2003–2008.

Hager,A.D. and Burgess,R.R. (1980) *Anal. Biochem.*, **109**, 76–86.

Katzen,A.L., Kornberg,T.B. and Bishop,J.M. (1985) *Cell*, **41**, 449–456.

Kemp,D.J., Coppel,R.L., Cowman,A.F., Saint,R.B., Brown,G.V. and Anders,R.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3787–3791.

Klempnauer,K.-H. and Sippel,A.E. (1986) *Mol. Cell. Biol.*, **6**, 62–69.

Klempnauer,K.-H., Gonda,T.J. and Bishop,J.M. (1982) *Cell*, **31**, 453–463.

Klempnauer,K.-H., Ramsay,G., Bishop,J.M., Moscovici,M.G., Moscovici,C., McGrath,J.P. and Levinson,A.D. (1983) *Cell*, **33**, 345–355.

Klempnauer,K.-H., Symonds,G., Evan,G.I. and Bishop,J.M. (1984) *Cell*, **37**, 537–547.

Klempnauer,K.-H., Bonifer,C. and Sippel,A.E. (1986) *EMBO J.*, **5**, 1903–1911.

Leprince,D., Saule,S., de Taisne,C., Gegonne,A., Begue,A., Righi,M. and Stehelin,D. (1983) *EMBO J.*, **2**, 1073–1078.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual* . Cold Spring Harbor Laboratory Press, NY.

Miller,J., McLachlan,A.D. and Klug,A. (1985) *EMBO J.*, **4**, 1609–1614.

Moelling,K., Pfaff,E., Beug,H., Beimling,P., Bunte,T., Schaller,H.E. and Graf,T. (1985) *Cell*, **40**, 983–990.

Moscovici,C. and Gazzolo,L. (1982) In Klein,G. (ed.), *Advances in Viral Oncology*. Raven Press, NY, Vol. I, pp. 83–105.

Moscovici,M.G. and Moscovici,C. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1421–1425.

Nichols,B.P. and Yanofsky,C. (1983) *Methods Enzymol.*, **101**, 851–855.

Nunn,M., Weiher,H., Bullock,P. and Duesberg,P. (1984) *Virology*, **139**, 330–339.

Ralston,R. and Bishop,J.M. (1983) *Nature*, **306**, 803–806.

Rosson,D. and Reddy,E.P. (1986) *Nature*, **319**, 604–606.

Rushlow,K.E., Lautenberger,J.A., Papas,T.S., Baluda,M.A., Perbal,B., Chirikjian,J.G. and Reddy,E.P. (1982) *Science*, **216**, 1421–1423.

Schwartz,D.E., Tizard,R. and Gilbert,W. (1983) *Cell*, **32**, 853–869.

Slamon,D.J., Boone,T.C., Murdock,D.C., Keith,D.E., Press,M.F., Larson,R.A. and Souza,L.M. (1986) *Science*, **233**, 347–351.

Souza,L.M., Strommer,J.N., Hilyard,R.L., Komaromy,M.C. and Baluda,M.A. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5177–5181.

Thomspon,C.B., Challoner,P.B., Neiman,P.E. and Groudine,M. (1986) *Nature*, **319**, 374–380.

Received on December 4, 1986; revised on June 1, 1987