The EMBO Journal vol.4 no.8 pp.2003-2008, 1985

Nucleotide sequence of cDNA clones of the murine *myb* proto-oncogene

Thomas J.Gonda, Nicholas M.Gough, Ashley R.Dunn and Julie de Blaquiere

Melbourne Tumour Biology Branch, Ludwig Institute for Cancer Research, Post Office Royal Melbourne Hospital, Australia 3050

Communicated by J.Schlessinger

We have isolated cDNA clones of murine c-*myb* mRNA which contain ~2.8 kb of the 3.9-kb mRNA sequence. Nucleotide sequencing has shown that these clones extend both 5′ and 3′ to sequences homologous to the v-*myb* oncogenes of avian myeloblastosis virus and avian leukemia virus E26. The sequence contains an open reading frame of 1944 nucleotides, and could encode a protein which is both highly homologous, and of similar size (71 kd), to the chicken c-*myb* protein. Examination of the deduced amino acid sequence of the murine c-*myb* protein revealed the presence of a 3-fold tandem repeat of 52 residues near the N terminus of the protein, and has enabled prediction of some of the likely structural features of the protein. These include a high α-helix content, a basic region toward the N terminus of the protein and an overall globular configuration. The arrangement of genomic c-*myb* sequences, detected using the cDNA clones as probes, was compared with the reported structure of rearranged c-*myb* in certain tumour cells. This comparison suggested that the rearranged c-*myb* gene may encode a protein which, like the v-*myb* protein, lacks the N-terminal region of c-*myb*.

Key words: murine c-*myb*/nucleotide sequence/predicted protein/relationship to v-*myb*/N-terminal truncation

Introduction

Hemopoiesis provides a valuable experimental system for examining the control of cellular proliferation and differentiation, because hemopoietic cell lineages are well defined and because hemopoietic cells can both proliferate and differentiate under appropriate conditions *in vitro*. Furthermore, the availability of many leukemic cell lines facilitates comparison of these processes in normal and neoplastic cells. Since the *myb* proto-oncogene (c-*myb*) and its viral derivatives (v-*myb*) appear to be active primarily in hemopoietic cells, c-*myb* is an attractive candidate for a study of the role of proto-oncogenes in proliferation and differentiation.

Expression of c-*myb* is normally restricted to hemopoietic cells (Gonda *et al.*, 1982; Westin *et al.*, 1982; Sheiness and Gardinier, 1984); expression has been detected in all hemopoietic lineages (Gonda *et al.*, 1982; Westin *et al.*, 1982) and appears to be 'switched off' during differentiation (Gonda *et al.*, 1982; Westin *et al.*, 1982; Craig and Block, 1984; Gonda and Metcalf, 1984; Sheiness and Gardinier, 1984). This specificity of c-*myb* is probably reflected in the ability of retroviruses carrying v-*myb* oncogenes [avian myeloblatosis virus (AMV) and avian leukemia virus E26] to transform hemopoietic cells (Moscovici and Gazzolo, 1982; Radke *et al.*, 1982). Alterations to the c-*myb* gene are also associated with hemopoietic tumours: retroviral insertions have been detected in a class of lymphoid tumours (Shen-Ong *et al.*, 1984), and amplification of c-*myb* has been reported in a human myeloid leukemia (Pelicci *et al.*, 1984).

The products of the chicken c-*myb* gene and the v-*myb*^{AMV} gene have been identified as nuclear proteins of 75 kd and 45 kd, respectively (Klempnauer *et al.*, 1983, 1984). Moreover, nucleotide sequence data on v-*myb*^{AMV} and the homologous regions of chicken c-*myb* indicate that p45$^{v\text{-}myb}$ represents an internal portion of c-*myb*, since the v-*myb*^{AMV} sequence does not contain initiation or termination codons derived from the cellular gene (Klempnauer *et al.*, 1982). To examine further the relationship between the cellular *myb* gene product and its (transforming) viral derivatives, and to advance studies of the function of c-*myb* in the well-characterized murine hemopoietic system, we have isolated murine c-*myb* cDNA clones. The nucleotide sequence of these clones indicates the capacity to encode a protein both highly homologous and of similar size to chicken p75$^{c\text{-}myb}$;

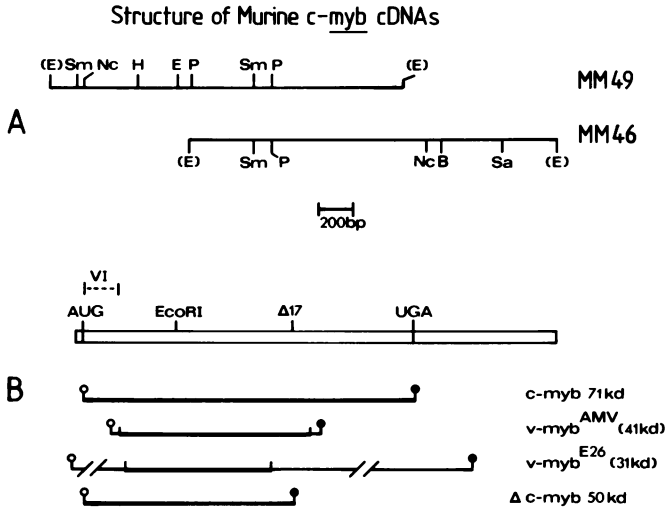

Fig. 1. (A) Restriction maps of the murine c-*myb* cDNA clones MM46 and MM49, showing cleavage sites for E, *Eco*RI; B, *Bgl*II; H, *Hind*III; Nc, *Nco*I; P, *Pst*I; Sa, *Sac*I; Sm, *Sma*I. (E) indicates *Eco*RI sites present in the cloning vector (pJL3) immediately adjacent to the GC tails which join the cDNA insert to the vector. **(B)** Salient features of the deduced c-*myb* mRNA sequence and relationships between *myb*-encoded polypeptides. The open box represents the mRNA sequence – the potential initiation (AUG) and termination (UGA) codons for the c-*myb* protein are indicated as is the location of the 17-base deletion in MM46 (see text). The bar marked VI indicates those sequences which, in the rearranged c-*myb* gene found in certain lymphoid tumours (ABPLs), are separated by a viral insertion from the remainder of the gene (see text). The diagram is not meant to suggest that the viral insertion in the c-*myb* gene occurs within the coding sequences. The lines beneath the open box represent predicted or observed *myb*-encoded proteins, from top: the predicted murine c-*myb* protein; the AMV-encoded protein p45$^{v\text{-}myb}$; the E26-encoded protein p135$^{gag\text{-}myb\text{-}ets}$; and the predicted product of a c-*myb* gene bearing the 17-base deletion. The thick portions of the lines indicate the *myb*-encoded parts of these proteins while the thin portions represent viral or other cellular sequences. The predicted mol. wt. of the *myb*-encoded regions of each protein are indicated at the right of each line, and the positions of initiation and termination codons are indicated by open and closed circles, respectively.

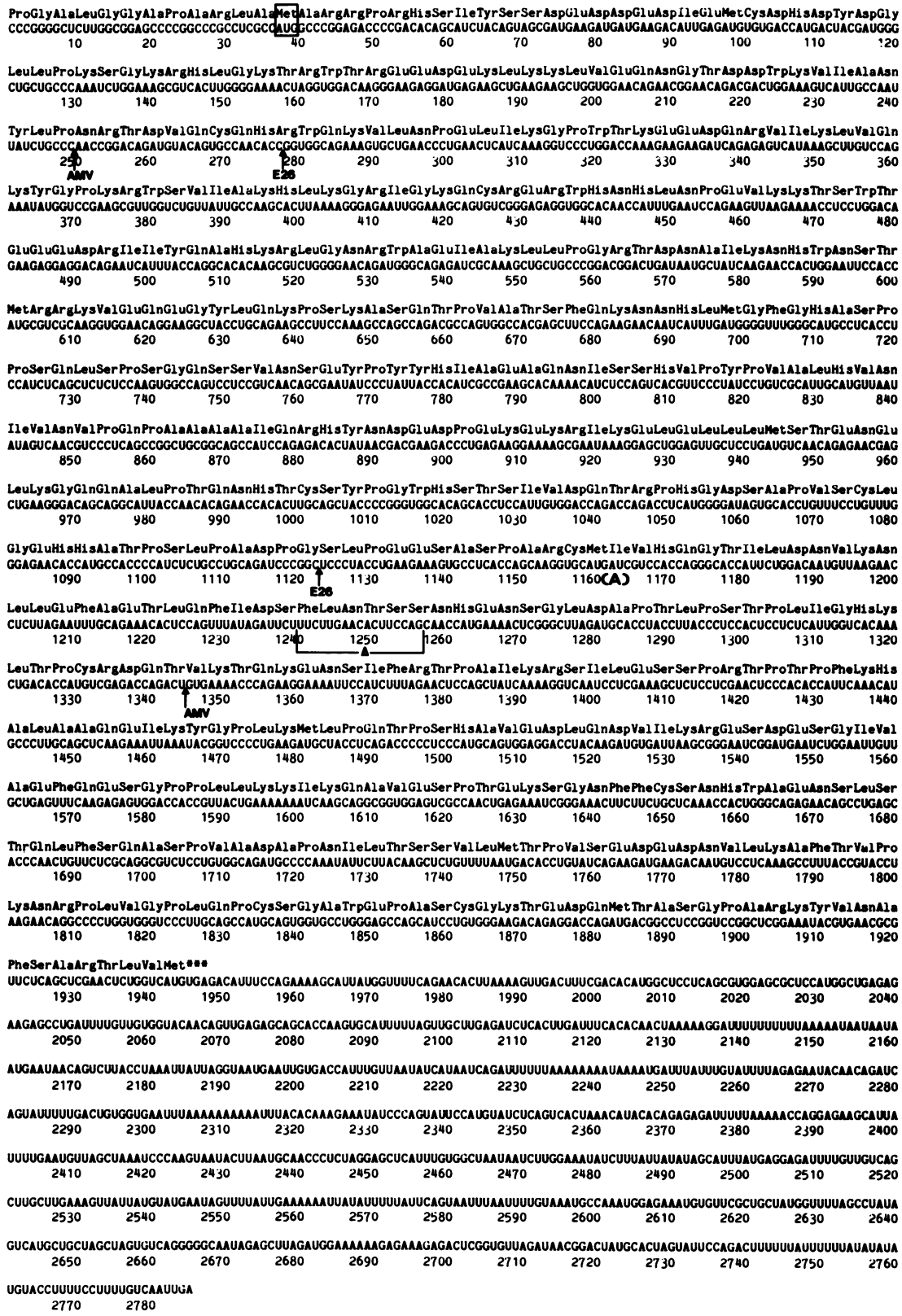

Fig. 2. Nucleotide sequence of 2.8 kb of the c-*myb* mRNA deduced from that of cDNA clones MM46 and MM49. The translation of this sequence in the long open reading frame, which extends from the start of the sequence to the termination codon (***) at position 1945, is shown. The potential initiation codon is boxed. The 17-base deletion in MM46 is indicated by the symbol △. Vertical arrows labelled AMV and E26 indicate the boundaries of the sequences homologous to the v-*myb* sequences of each virus (see text).

thus, these clones may contain the entire coding sequence of the c-*myb* mRNA. This has allowed us to compare the predicted amino acid sequences of the transforming v-*myb* proteins with the sequence of the c-*myb* protein, to predict several properties of the c-*myb* protein and to detect c-*myb* sequences in the murine genome.

Results and Discussion

A cDNA library, prepared using mRNA from the murine myeloid leukemia WEHI-3B, was screened by hybridization with a v-*myb* probe. Subsequent rounds of screening were carried out using the murine c-*myb* clones thus isolated, yielding a total of eight clones. The restriction maps of the two longest clones, MM46 and MM49 (Figure 1A), indicate that these clones span ~2.8 kb and thus represent a substantial portion of the c-*myb* mRNA sequence, which is ~3.9 kb (Mushinski *et al.*, 1983; Gonda and Metcalf, 1984). Since the 3′-most clone, MM46, does not contain a poly(A) tail, we cannot presently map our clones onto the 3.9-kb mRNA.

Nucleotide sequence and coding potential of c-myb mRNA

The nucleotide sequence of 2.8 kb of the c-*myb* mRNA deduced from the sequence of these cDNAs is shown in Figure 2. The longest open reading frame extends 1944 bases from the start of the sequence and contains a potential initiation codon at position 37; furthermore, the sequence around this codon (GCCAUGG) conforms to the 'consensus' sequence at eukaryotic initiation codons described by Kozak (1984). Translation initiated at this position would result in a protein of 636 amino acids which is highly homologous to the chicken c-*myb* protein (see below) and which would have a mol. wt. of 71 kd (see Figure 1B) similar to that (75 kd) reported for the chicken c-*myb* protein (Klempnauer *et al.*, 1983). However, we cannot presently rule out the possibility that the AUG codon at position 37 encodes an internal methionine residue and thus that the true initiation codon lies upstream of the region we have sequenced.

A substantial proportion of the overlap between MM46 and MM49 (see Figure 1A) was sequenced using M13 clones derived from both of these cDNAs. Only two differences were detected: a change from G to A at position 1161 (see Figure 2) and more importantly, a 17-base deletion in MM46 with respect to MM49 (and chicken c-*myb*) at position 1240 (Figures 2 and 1B). This deletion may simply be a cDNA cloning artefact, since it is not present in any of our other cDNA clones, although at least three of these probably contain the relevant region (data not shown). However, it could conceivably represent a transcript from an altered allele of c-*myb* in WEHI-3B; translation of such a mRNA would terminate at position 1263 giving rise to a protein of 50 kd (Figure 1B).

Relationship to other myb sequences

The predicted amino acid sequence of the murine c-*myb* protein was compared with the amino acid sequence predicted by the nucleotide sequence of part of the chicken c-*myb* gene (Klempnauer *et al.*, 1982). The alignment shown in Figure 3 reveals that the first amino acid of the sequenced portion of chicken c-*myb*, which corresponds to the first amino acid encoded by v-*myb*^{AMV} (Klempnauer *et al.*, 1982), corresponds to the asparagine located 72 amino acid residues from the putative murine c-*myb* initiation codon (Figure 2). Clearly, at least part of c-*myb* is highly conserved between chicken and mouse, since 82% of the residues shown in Figure 3 are identical in the two species. This is particularly striking in the case of the first 130 residues (Figure 3), where only one amino acid difference is observed.

Because of the extensive homology between chicken and mouse c-*myb* sequences, and between chicken viral (Klempnauer *et al.*, 1982; Rushlow *et al.*, 1982; Nunn *et al.*, 1984) and cellular (Klempnauer *et al.*, 1982) *myb* sequences, the v-*myb* sequences can be readily mapped onto the murine c-*myb* sequence. As mentioned above and shown in Figure 2, the region homologous to

```
Mouse   NRTDVQCQHRWQKVLNPELIKGPWTKEEDQRVIKLVQKYGPKRWSVIAKHLKGRIGKQCR  131
        ********************************** ************************
Chicken NRTDVQCQHRWQKVLNPELIKGPWTKEEDQRVIELVQKYGPKRWSVIAKHLKGRIGKQCR   60

Mouse   ERWHNHLNPEVKKTSWTEEEDRIIYQAHKRLGNRWAEIAKLLPGRTDNAIKNHWNSTMRR  191
        ************************************************************
Chicken ERWHNHLNPEVKKTSWTEEEDRIIYQAHKRLGNRWAEIAKLLPGRTDNAIKNHWNSTMRR  120

Mouse   KVEQEGYLQKPSKASQTPVATSFQKNNHLMGFGHASPPSQLSPSGQSSVNSEYPYYHIAE  251
        *********  ***      * ***  *** * *  *   *   **    * ********
Chicken KVEQEGYLQESSKAGLPSATTGFQKSSHLMAFAHNPPAGPLPGAGQAPLGSDYPYYHIAE  180

Mouse   AQNISSHVPYPVALHVNIVNVPQPAAAAIQRHYNDEDPEKEKRIKELELLLMSTENELKG  311
         **      ************************************************** 
Chicken PQNVPGQIPYPVALHVNIVNVPQPAAAAIQRHYNDEDPEKEKRIKELELLLMSTENELKG  240

Mouse   QQALPTQNHTCSYPGWHSTSIVDQTRPHGDSAPVSCLGEHHA.TPSLPADPGSLPEESAS  370
        **********  *******   * **  ** **********  *** * * * *******
Chicken QQALPTQNHTANYPGWHSTTVADNTRTSGDNAPVSCLGEHHHCTPSPPVDHGCLPEESAS  300

Mouse   PARCMIVHQGTILDNVKNLLEFAETLQFIDSFLNTSSNHENSGLDAPTLPSTPLIGHKLT  430
        *********  ***************** *************  ** * * ***  ***
Chicken PARCMIVHQSNILDNVKNLLEFAETLQLIDSFLNTSSNHENLNLDNPALTSTPVCGHKMS  360

Mouse   PC....RDQTVKTQKENSI                                          445
           ***  ******
Chicken VTTPFHRDQAFKTQKENHV                                          379
```

Fig. 3. Homology between the predicted amino acid sequences (shown in single-letter code) of the murine c-*myb* product and part of the chicken c-*myb* product. Only the relevant portion of the murine sequence is shown here. The numbers at the right are the numbers of residues from the first methionine residue in the case of the murine sequence (see Figure 2); the residue numbers of the partial chicken sequence are those of Klempnauer *et al.* (1982). (The first residue shown corresponds to the first amino acid encoded by v-*myb*^{AMV}.)

v-*myb*^{AMV} extends from nucleotide 250, or 72 amino acids from the presumptive N terminus of the c-*myb* protein, to nucleotide 1344, which is 200 amino acids from the C terminus of the murine c-*myb* protein. Similarly, sequences homologous to v-*myb*^{E26} extend from nucleotide 278 of the murine c-*myb* sequence to nucleotide 1122. The sizes of, and relationships between, the various *myb*-encoded polypeptides are shown in Figure 1B.

Several points of interest emerge from these comparisons. First, both AMV and E26 *myb* sequences lack the N-terminal region of the presumptive c-*myb* protein; this is particularly striking since the *myb* coding sequences of the two viruses start only nine amino acid residues apart. Secondly, both v-*myb* proteins contain sequences homologous to the long highly conserved region shown in Figure 3, which is consistent with this region being of particular functional importance. In addition, it may be of significance that the protein encoded by an mRNA bearing the 17-base deletion described above would terminate near the ends of the *myb* sequences of AMV and E26; however, the effect of C-terminal truncation on the function of *myb* proteins remains to be elucidated.

Structure of the predicted c-myb protein

A search of the NBRF protein sequence database for sequences related to c-*myb* was carried out using the Dayhoff 'SEARCH' program (Dayhoff, 1976). No striking homologies with proteins other than other *myb* gene products were revealed; however, several sequences from within the first 200 residues of the murine sequence each showed homology to several different regions of the (partial) chicken c-*myb* sequence. This suggested the presence of repeated sequences in murine c-*myb* and in fact a 2-fold repeat at the N terminus of v-*myb* (and chicken c-*myb*) has been reported by Ralston and Bishop (1983). Our analysis of the murine sequence revealed the presence of three tandem repeats of 51 or 52 residues (Figure 4A), the first of which begins 34 residues N-terminally of the boundary of v-*myb* homology (i.e., at residue 38). Comparison of the repeated sequences using the Dayhoff 'RELATE' program (Dayhoff, 1976) showed that the homology

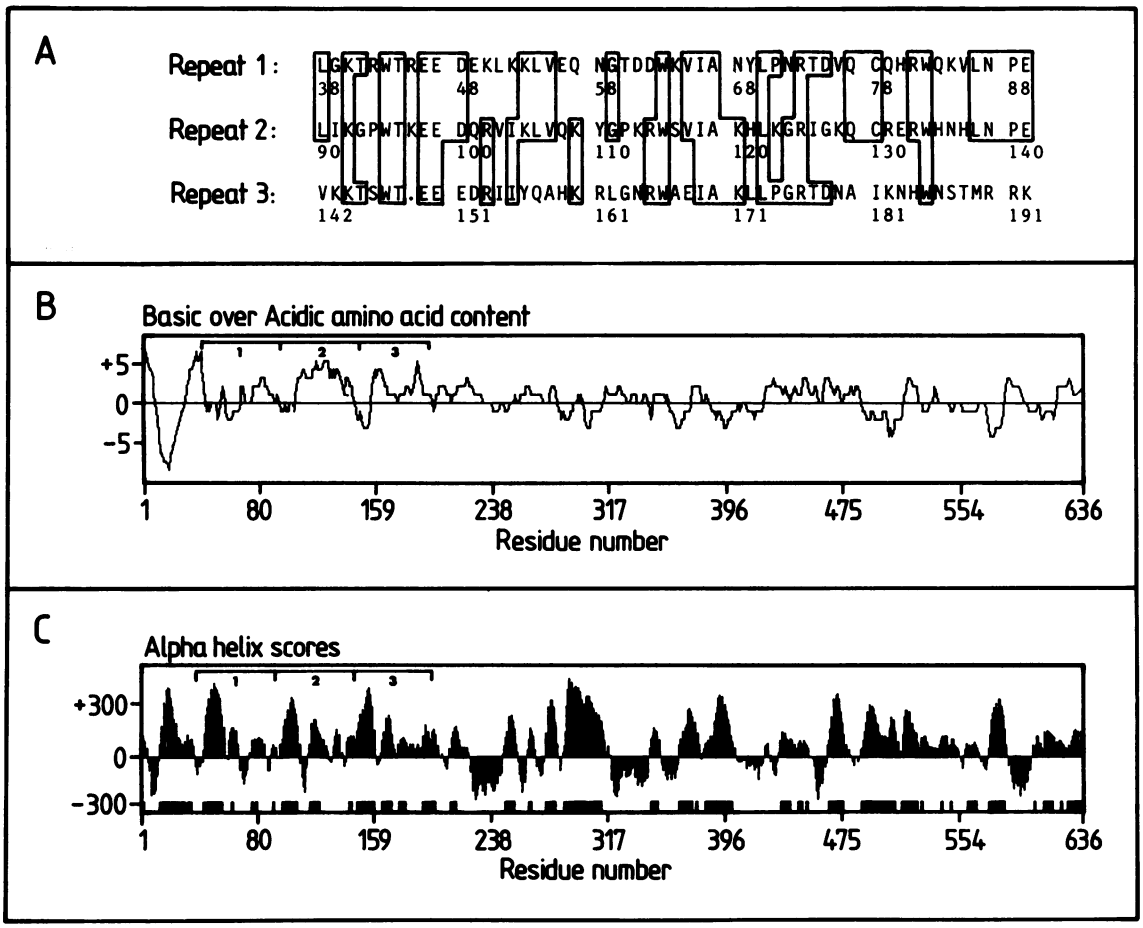

Fig. 4. Properties of the predicted c-*myb* protein. **(A)** Tandem repeats (shown in single-letter code) in the sequence. The residue numbers [starting from the first methionine (boxed in Figure 2)] in each repeat are shown below each repeat. Amino acids present in identical positions in two or all three of the repeats are boxed. The dot in repeat 3 indicates a gap in the sequence which was inserted to maximize homology. **(B)** Distribution of acidic and basic residues along the length of the protein. The graph was generated by plotting the net charge [the sum of arginine and lysine residues (+1) and glutamate and aspartate residues (−1)] of a region of 11 residues centered on each successive residue; this procedure was adopted in order to more clearly reveal the overall character of large regions of the protein. The bars at the top of the diagram indicate the position of the three repeats shown in **A.** **(C)** Potential α-helical sequences in the c-*myb* protein, according to the method of Garnier *et al.* (1978). The positions of the three repeats **(A)** are again shown at the top of the diagram.

between the three repeats scored at 5−9 standard deviations above the average score of all intra-sequence comparisons. This analysis showed that, as is apparent by inspection of Figure 4A, repeats 1 and 2 are more closely related to each other than either is to repeat 3: in the C-terminal portion of repeat 3, only the tryptophan residue is shared with repeats 1 and 2. It is probably significant that the repeats include almost all of the highly conserved 130 amino acid region mentioned above (i.e., the first 130 amino acids shown in Figure 3); in addition the three repeats probably have similar secondary structures (see below).

We have also examined several predicted properties and structural features of the deduced c-*myb* protein sequences. Hydrophobicity analysis using the method of Kyte and Doolittle (1982) suggested that the c-*myb* protein is likely to be strongly hydrophilic along its entire length (data not shown). Examination of the distribution of acidic and basic amino acids (illustrated in Figure 4B) revealed a basic region of ~200 residues near the N terminus of the protein (approximately residues 30−230); the remainder of the protein shows a more even distribution of basic and acid residues. Several secondary structure prediction methods each suggested a high proportion of α-helix and a low proportion of β-sheet structure. For example, the method of Garnier *et al.* (1978) predicted that 46% of the protein would be in an α-helical configuration; the distribution of these predicted α-helical regions is shown in Figure 4C. The proportion of cysteine residues (1.6%) is fairly low compared with the average of 3.1% (Dayhoff, 1976) − these residues could potentially form five disulphide linkages. Taken together, the predicted properties of the c-*myb* protein suggest that the protein is likely to be globular in overall configuration.

It is of interest to consider the predicted structures of the tandem repeats discussed above and shown in Figure 4A. Figure 4C reveals that these repeats, including the more divergent third repeat (Figure 4A), are likely to have quite similar secondary structures, as indicated by the distribution of α-helical potential. Given their evolutionary conservation, their overall basic nature (Figure 4B) and considering the nuclear location of *myb*-encoded proteins (Klempnauer *et al.*, 1984; Boyle *et al.*, 1984) it is tempting to speculate that these repeats may be involved in binding of the c-*myb* protein to DNA.

Detection of genomic c-myb sequences

Examination of c-*myb* sequences in the murine genome using the cDNA clones as hybridization probes, revealed six major *Eco*RI restriction fragments (Figure 5), some of which (the 4.3-, 2.0-, 1.8- and 1.5-kbp fragments) also hybridized to v-*myb* probes (Shen-Ong *et al.*, 1984; T.J.Gonda, unpublished observations). Since each of the three probes used (MM46 and the

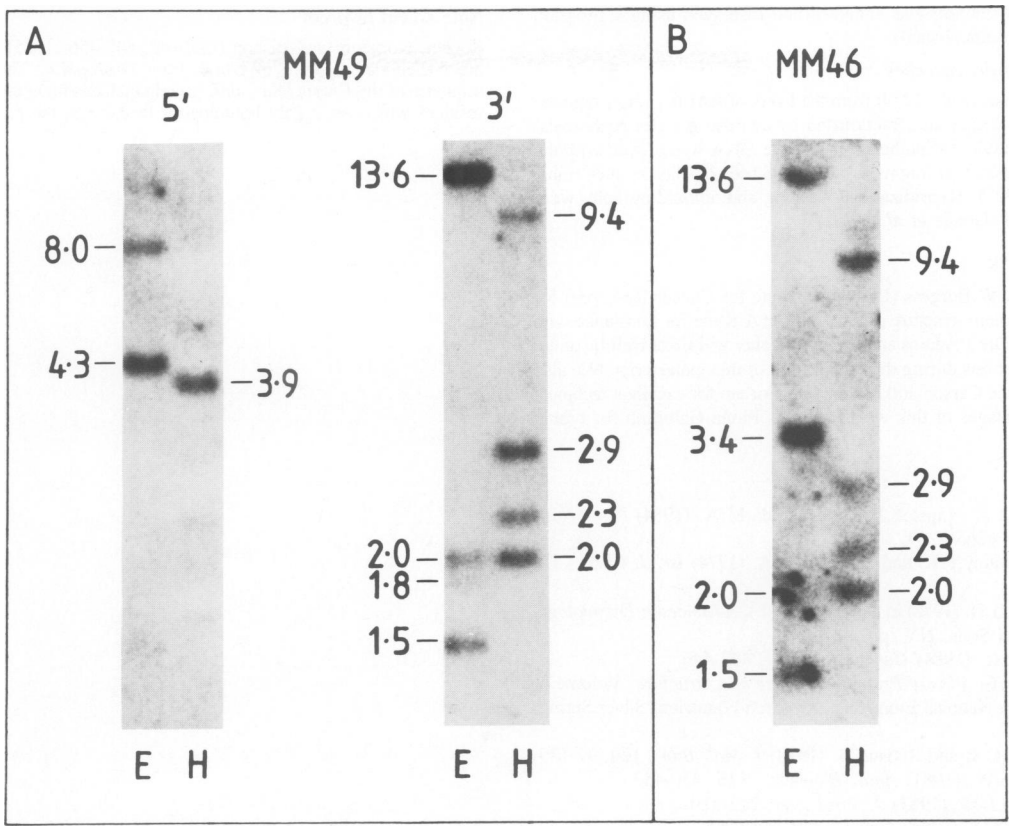

Fig. 5. Identification of murine genomic c-*myb* sequences, using as probes: **A**, the 5′ and 3′ *Eco*RI fragments of MM49; and **B**, MM46 (see Figure 1). Genomic DNA was cleaved with *Eco*RI (E) or *Hind*III (H) as indicated: the sizes, in kbp, of the major fragments detected are also shown.

two *Eco*RI fragments of MM49 – see Figure 1A) detected different sets of genomic fragments (Figure 5), there is probably only one c-*myb* locus in the murine genome. Furthermore, by comparing Figures 1A and 4, and taking into account which fragments also hybridize to v-*myb* probes, the order of the genomic *Eco*RI fragments can be deduced to be 5′-8.0, 4.3, 1.8, (2.0, 1.5), 13.6, 3.4-3′ (sizes in kbp) in the sense of the mRNA. This order is consistent with the structure of a clone of a portion of the murine c-*myb* gene which contains some of these fragments (Lavu *et al.*, 1984; Shen-Ong *et al.*, 1984). Note that of the two genomic *Eco*RI fragments detected with the MM49 5′ probe (Figure 5), only the 4.3-kbp fragment hybridizes to v-*myb* probes (Shen-Ong *et al.*, 1984; T.J.Gonda, unpublished observations). Thus at least some of the sequences encoding the N-terminal region of the c-*myb* protein (which are 5′ to the v-*myb*-related sequences – see Figure 2) are probably present in the 8.0-kbp genomic fragments (Figure 5).

This latter observation is of interest with regard to a class of lymphoid tumours (ABPLs) in which deleted Moloney leukemia virus proviruses have integrated in the vicinity of the c-*myb* gene, resulting in a high level of aberrant c-*myb* transcripts (Mushinski *et al.*, 1983). In these tumour cells, the proviruses have integrated toward the 5′ end of the 4.3-kbp *Eco*RI fragment (Shen-Ong *et al.*, 1984). Thus, the viral insertions have occurred within the c-*myb* gene between the sequences encoding the v-*myb*-related regions and those encoding the N-terminal region of the (predicted) c-*myb* protein (as indicated by 'VI' in Figure 2). Hence it is possible that, like the v-*myb* proteins, the proteins encoded by the aberrant c-*myb* transcripts in these tumours may lack the N-terminal region of the normal c-*myb* protein. If this proves to be the case, the possibility is raised that removal of the N-terminal region of the c-*myb* protein is essential, or even sufficient, to generate a transforming gene product.

Materials and methods

Isolation of c-myb cDNA clones

A cDNA library was constructed as previously described in the plasmid pJL3 (Gough *et al.*, 1984, 1985) using cytoplasmic polyadenylated RNA from WEHI-3B cells isolated as described in Gough (1983). Colonies were screened by hybridization to the v-*myb*-specific *Kpn*I-*Xba*I fragment of pVM2 (Klempnauer *et al.*, 1982) which had been radiolabelled by nick-translation (Rigby *et al.*, 1977). Hydridization was at 65°C in 6 x SSC, 5 x Denhardt's solution, 100 μg/ml denatured salmon sperm DNA and 0.1% SDS, and filters were washed at 65°C in 2 x SSC/0.5% SDS. A murine c-*myb* cDNA clone, MM1, was thus isolated and used to screen further colonies. One of the cDNA clones isolated in this second screen, MM49, was used for a third screen: this resulted in the isolation of several additional clones including MM46. Approximately 100 000 recombinants were screened in total. The orientation of and overlap between MM46 and MM49 were determined by restriction mapping and confirmed by hybridization of both clones to chicken genomic c-*myb* clones (Gonda and Bishop, 1983).

Nucleotide sequence determination

Restriction fragments of MM46 and MM49 were inserted into M13 vectors mp8, mp9, mp10 or mp11, either by isolation of specific restriction fragments from agarose gels or by 'shotgun' cloning of digests prepared with *Alu*I, *Hae*III, *Taq*I or *Hpa*II. Sequencing was performed by the 'dideoxy' chain termination procedure (Sanger *et al.*, 1977); reactions were electrophoresed on 0.2 mm thick thermostatistically regulated polyacrylamide gels (Garoff and Ansorge, 1981). Over 86% of the sequence, and 98% of the coding sequence, was determined by sequencing at least two different restriction fragments cloned in M13 vectors.

Nucleotide sequence data analysis

Sequencing gel readings were compiled and analyzed using modifications of computer programs described by Staden (1980). Amino acid sequence alignments were performed using the Dayhoff 'ALIGN' program (Dayhoff, 1976). Properties of the predicted c-*myb* protein were analyzed using the hydrophobicity prediction method of Kyte and Doolittle (1982) and the secondary structure prediction methods of Garnier *et al.* (1978), Chou and Fasman (1978) and Burgess *et al.*

(1979). The charge distribution was analyzed and displayed using a program developed by A.Kyne (unpublished).

Detection of genomic myb sequences

DNA was isolated (Hughes *et al.*, 1979) from the livers of BALB/c mice, digested with restriction endonucleases and fractionated by agarose gel electrophoresis. After transfer to nitrocellulose (Southern, 1975), the DNA was probed with the relevant c-*myb* cDNA restriction fragments which had been labelled by nick-translation (Rigby *et al.*, 1977). Hybridization, washing and autoradiography were as previously described (Gonda *et al.*, 1981).

Acknowledgements

We wish to thank Dr A.W.Burgess (Ludwig Institute for Cancer Research) for invaluable advice on protein structure prediction, Dr A.Kyne for assistance with computer analysis, and Drs J.Adams and S.Cory (Walter and Eliza Hall Institute) for many helpful suggestions during the preparation of this manuscript. We also wish to thank Ms Suzanne Carson and Ms Pauline Gorban for excellent technical assistance in the early stages of this work, and Ms Fiona Galbraith for typing the manuscript.

References

Boyle,W.J., Lambert,M.A., Lipsick,J.S. and Baluda,M.A. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4265-4269.

Burgess,A.W., Ponnuswamy,P.K. and Scheraga,H.A. (1974) *Isr. J. Chem.*, **12**, 239-286.

Chou,P.Y. and Fasman,G.D. (1978) in Meister,A. (ed.), *Advances in Enzymology*, Vol. **47**, J.Wiley and Sons, NY, pp. 45-148.

Craig,R.W. and Block,A. (1984) *Cancer Res.*, **44**, 442-446.

Dayhoff,M.O. ed., (1976) *Atlas of Protein Sequence and Structure*, Volume **5**, Suppl. 2, published by National Biomedical Research Founation, Silver Spring, MD, pp. 3-9.

Garnier,J., Ogusthorpe,D.J. and Robson,B. (1978) *J. Mol. Biol.*, **120**, 97-120.

Garoff,H. and Ansorge,W. (1981) *Anal. Biochem.*, **115**, 450-457.

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

Gonda,T.J. and Metcalf,D. (1984) *Nature*, **310**, 249-251.

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.

Gough,N.M. (1983) *J. Mol. Biol.*, **165**, 683-689.

Gough,N.M., Gough,J., Metcalf,D., Kelso,A., Grail,D., Nicola,N.A., Burgess, A.W. and Dunn,A.R. (1984) *Nature*, **309**, 763-767.

Gough,N.M., Metcalf,D., Gough,J., Grail,D. and Dunn,A.R. (1985) *EMBO J.*, **4**, 645-653.

Hughes,S.H., Payvar,F., Spector,D., Schimke,R.T., Robinson,H.L., Payne,G.S., Bishop,J.M. and Varmus,H.E. (1979) *Cell*, **18**, 347-359.

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.

Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857-872.

Kyte,J. and Doolittle,R.F. (1982) *J. Mol. Biol.*, **157**, 105-132.

Lavu,S., Mushinski,J.F., Shen-Ong,G.L.C., Potter,M. and Reddy,E.P. (1984) *Curr. Top. Microbiol. Immunol.*, **113**, 37-40.

Moscovici,C. and Gazzolo,L. (1982) in Klein,G. (ed.), *Advances in Viral Oncology*, Vol. **1**, Raven Press, NY, pp. 83-106.

Mushinski,J.F., Potter,M., Bauer,S.R. and Reddy,E.P. (1983) *Science (Wash.)*, **220**, 795-798.

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

Pelicci,P.-G., Lanfrancone,L., Brathwaite,M.D., Wolman,S.R. and Dalla-Favera, R. (1984) *Science (Wash.)*, **226**, 1077-1080.

Radke,K., Beug,H., Kornfeld,S. and Graf,T. (1982) *Cell*, **31**, 643-653.

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

Rigby,P.W.J., Dieckman,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237-251.

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

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.

Sheiness,D. and Gardinier,M. (1984) *Mol. Cell. Biol.*, **4**, 1206-1212.

Shen-Ong,G.L.C., Potter,M., Mushinski,J.F., Lavu,S. and Reddy,E.P. (1984) *Science (Wash.)*, **226**, 1077-1080.

Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.

Staden,R. (1980) *Nucleic Acids Res.*, **8**, 3673-3694.

Westin,E.H., Gallo,R.C., Arya,S.K., Eva,A., Souza,L.M., Baluda,M.A., Aaronson,S.A. and Wong-Stahl,F. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 2194-2198.

Received on 20 May 1985

Note added in proof

Katzen, Kornberg and Bishop (*Cell*, **41**, 449-456, 1985) have recently reported the nucleotide sequence of c-*myb* from *Drosophila*. The predicted amino-acid sequence of the *Drosophila* c-*myb* protein includes three tandem repeats of 50–52 residues which are highly homologous to those in the predicted murine protein.